



# Competition among memes in a world with limited attention

L. Weng<sup>1</sup>, A. Flammini<sup>1</sup>, A. Vespignani<sup>2,3,4</sup> & F. Menczer<sup>1</sup>

<sup>1</sup>Center for Complex Networks and Systems Research School of Informatics and Computing Indiana University, Bloomington, USA, <sup>2</sup>Department of Health Sciences, Department of Physics and College of Computer and Information Sciences, Northeastern University, USA, <sup>3</sup>Institute for Quantitative Social Sciences, Harvard University, USA, <sup>4</sup>Institute for Scientific Interchange (ISI), Torino, Italy.

SUBJECT AREAS:  
INFORMATION THEORY  
AND COMPUTATION  
STATISTICAL PHYSICS,  
THERMODYNAMICS AND  
NONLINEAR DYNAMICS  
STATISTICS  
MODELLING AND THEORY

Received  
19 September 2011

Accepted  
8 March 2012

Published  
29 March 2012

Correspondence and  
requests for materials  
should be addressed to  
F.M. (fil@indiana.edu)

The wide adoption of social media has increased the competition among ideas for our finite attention. We employ a parsimonious agent-based model to study whether such a competition may affect the popularity of different memes, the diversity of information we are exposed to, and the fading of our collective interests for specific topics. Agents share messages on a social network but can only pay attention to a portion of the information they receive. In the emerging dynamics of information diffusion, a few memes go viral while most do not. The predictions of our model are consistent with empirical data from *Twitter*, a popular microblogging platform. Surprisingly, we can explain the massive heterogeneity in the popularity and persistence of memes as deriving from a combination of the competition for our limited attention and the structure of the social network, without the need to assume different intrinsic values among ideas.

Ideas have formidable potential to impact public opinion, culture, policy, and profit<sup>1</sup>. The advent of social media<sup>2</sup> has lowered the cost of information production and broadcasting, boosting the potential reach of each idea or *meme*<sup>3</sup>. However, the abundance of information to which we are exposed through online social networks and other socio-technical systems is exceeding our capacity to consume it. Ideas must compete for our scarce individual and collective attention. As a result, the dynamic of information is driven more than ever before by the economy of attention, first theorized by Simon<sup>4</sup>. Yet the processes that drive popularity in our limited-attention world are still largely unexplored<sup>5–15</sup>.

The availability of data from online social media has recently created unprecedented opportunities to explore human and social phenomena on a global scale<sup>16,17</sup>. In this context one of the most challenging problems is the study of the competition dynamics of ideas, information, knowledge, and rumors. Understanding this problem is crucial in a broad range of settings, from viral marketing to scientific discovery acceleration. Aspects of competition for limited attention have been studied through news, movies, and topics posted on blogs and social media<sup>10,11,13</sup>. The popularity of news decreases with the number of competing items that are simultaneously available<sup>8,18,19</sup>.

However, even in the simplified settings of social media platforms, it is hard to disentangle the effects of limited attention from many concurrent factors, such as the structure of the underlying social network<sup>7,13</sup>, the activity of users and the size of their potential audience<sup>19</sup>, the different degrees of influence of information spreaders<sup>20</sup>, the intrinsic quality of the information they spread<sup>21</sup>, the persistence of topics<sup>22,23</sup>, and homophily<sup>24</sup>. To compound these difficulties, social networks that host information diffusion processes are not closed systems; exogenous factors like exposure to traditional media and their reports of world events play important roles in the popularity and lifetime of specific topics<sup>10,25</sup>. Another example of our limited attention is the cognitive limit on the number of stable social relationships that we can sustain, as postulated by Dunbar<sup>26</sup> and recently supported by analysis of Twitter data<sup>27</sup>.

We propose an agent-based model to study the role of the limited attention of individual users in the diffusion process, and in particular whether competition for our finite attention may affect meme popularity, diversity, and lifetime. Although competition among ideas has been implicitly assumed as a factor behind, e.g., the decay in interest toward news and movies<sup>28,8,10</sup>, to the best of our knowledge nobody has attempted to explicitly model the mechanisms of competition and how they shape the spread of information. In particular, we show that a simple model of competition on a social network, without any further assumptions about meme merit, user interests, or explicit exogenous factors, can account for the massive heterogeneity in meme popularity and persistence.

## Results

Here we outline a number of empirical findings that motivate both our question and the main assumptions behind our model. We then describe the proposed agent-based toy model of meme diffusion and compare its

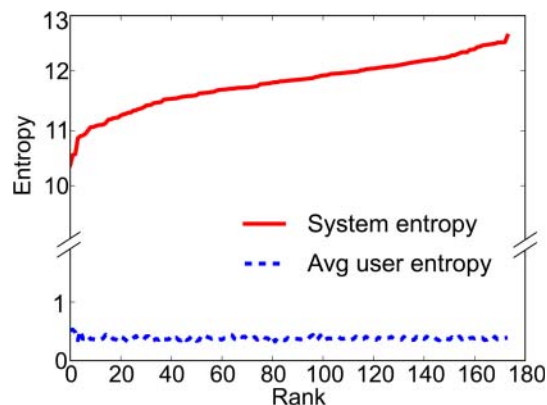


predictions with the empirical data. Finally we show that the social network structure and our finite attention are both key ingredients of the diffusion model, as their removal leads to results inconsistent with the empirical data.

We validate our model with data from *Twitter*, a micro-blogging platform that allows many millions of people to broadcast short messages through social connections. Users can “follow” interesting people, by which a directed social network is formed. Posts (“tweets”) appear on the screen of followers. People can forward (“retweet”) selected posts from their screen to their followers. Furthermore, users often mark their posts with topic labels (“hashtags”). Let us use these tags as operational proxies to identify memes. A retweet carries a meme from user to user. As a meme spreads in this way, it forms a cascade or diffusion network such as those illustrated in Fig. 1. We collected a sample of retweets that include one or more hashtags, produced by Twitter users over a specific period of time (see details in Methods section). This provides us with a quantitative framework to study the competition for attention in the wild.

**Limited attention.** We first explore the competition among memes. In particular, we test the hypothesis that the attention of a user is somewhat independent from the overall diversity of information discussed in a given period. Let us quantify *the breadth of attention of a user* through Shannon entropy  $S = -\sum_i f(i) \log f(i)$  where  $f(i)$  is the proportion of tweets generated by the user about meme  $i$ . Given a user who has posted  $n$  messages, her entropy can be as small as 0, if all of her posts are about the same meme; or as large as  $\log n$  if she has posted a message about each of  $n$  different memes. We can measure the diversity of the information available in the system analogously, defining  $f(i)$  as the proportion of tweets about meme  $i$  across all users. Note that these entropy-based measures are subject to the limits of our operational definition of a meme; finer or coarser definitions would yield different values.

In Fig. 2 we compare the daily values of the system entropy to the corresponding average user entropy. The key observation here is that

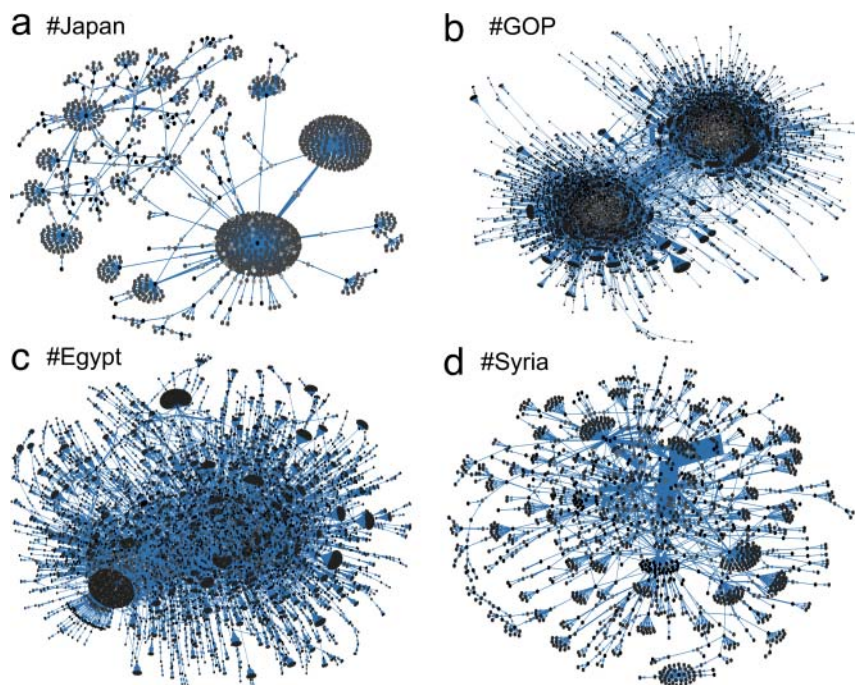


**Figure 2** | Plot of daily system entropy (solid red line) and average user breadth of attention (dashed blue line). Days in our observation period are ranked from low to high system entropy, therefore the latter is monotonously increasing.

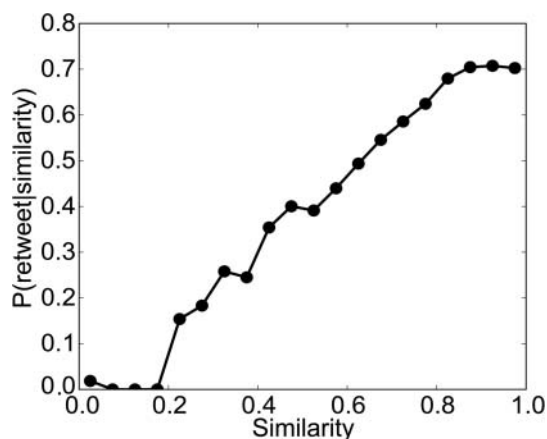
a user’s breadth of attention remains essentially constant irrespective of system diversity. This is a clear indication that the diversity of memes to which a user can pay attention is bound. With the continuous injection of new memes, this indirectly suggests that memes survive at the expense of others. We explicitly assume this in the information diffusion model presented later.

**User interests.** It has been suggested that topical interests affect user behavior in social media<sup>29,30</sup>. This is a potentially important ingredient in a model of meme diffusion, as an interesting meme may have a competitive advantage. Therefore we wish to explore whether user interests, as inferred from past behavior, are predictive of future behavior.

Let us consider every user in our dataset and any retweets they produce. When a user  $u$  emits a new retweet, we define her *interests*  $I_u$



**Figure 1** | Visualizations of meme diffusion networks for different topics. Nodes represent Twitter users, and directed edges represent retweeted posts that carry the meme. The brightness of a node indicates the activity (number of retweets) of a user, and the weight of an edge reflects the number of retweets between two users. (a) The **#Japan** meme shows how news about the March 2011 earthquake propagated. (b) The **#GOP** tag stands for the US Republican Party and as many political memes, displays a strong polarization between people with opposing views. Memes related to the “Arab Spring” and in particular the 2011 uprisings in (c) **#Egypt** and (d) **#Syria** display characteristic hub users and strong connections, respectively.



**Figure 3** | Relationship between the probability of retweeting a message and its similarity to the user interests, inferred from prior posting behavior.

as the set of all memes about which she has tweeted up to that moment. We also collect the set  $M_0$  of memes associated with the new retweet. The  $n$  most recent posts across all users prior to the new retweet are considered as a set of potential candidates that might have been retweeted, but were not. The corresponding sets of memes  $M_1, M_2, \dots, M_n$  are recorded ( $n = 10$ ). We compute the similarity

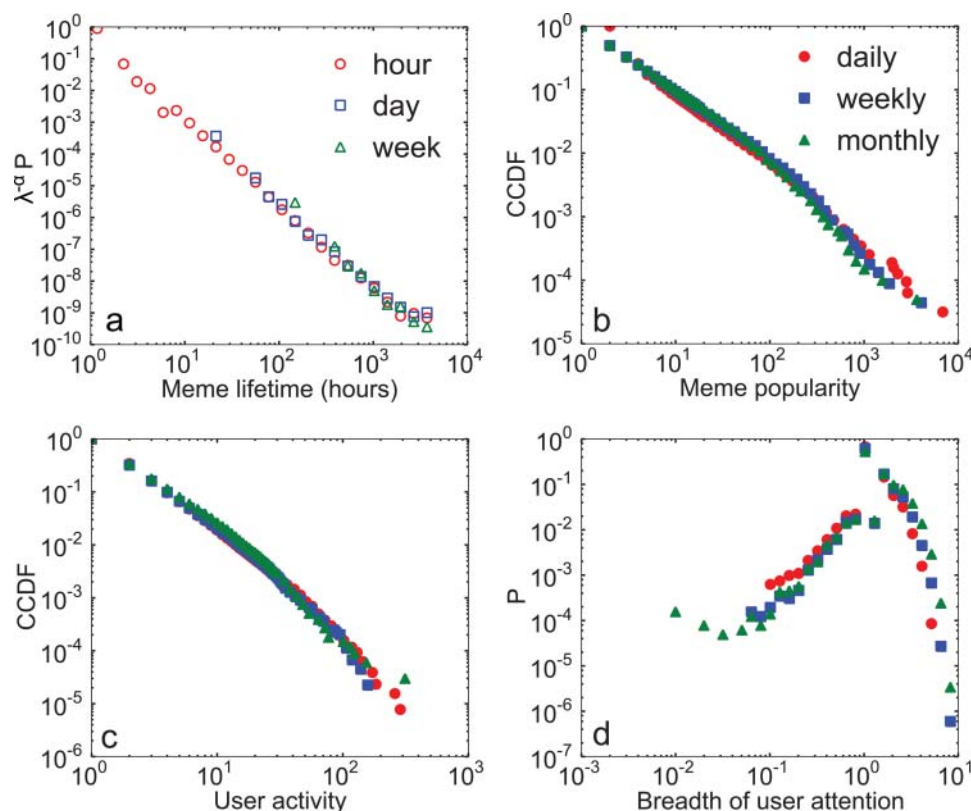
$sim(M_0, I_u), sim(M_1, I_u), \dots, sim(M_n, I_u)$  between the user interests and the actual and candidate posts, and recover the conditional probability  $P(\text{retweet}(u, M)|sim(M, I_u))$  that  $u$  retweets a post with memes  $M$  given the similarity between the memes and her user interests. We turn to the Maximum Information Path similarity measure<sup>31,32</sup> that considers shared memes but discounts the more common ones:

$$sim(M, I) = \frac{2 \log[\min_{x \in M \cap I} f(x)]}{\log[\min_{x \in M} f(x)] + \log[\min_{x \in I} f(x)]}$$

where  $x$  is a meme and  $f(x)$  the proportion of messages about  $x$ .

Fig. 3 shows that users are more likely to retweet memes about which they posted in the past (Pearson correlation coefficient  $\rho = 0.98$ ). This suggests that memory is an important ingredient for a model of meme competition, and we explicitly take this aspect into account in the model presented below.

**Empirical regularities.** In Fig. 4 we observe several regularities in the empirical data. We first consider meme lifetime, defined as the maximum number of consecutive time units in which posts about the meme are observed; meme popularity, defined as the number of users per day who tweet about a meme, measured over a given time period; and user activity, defined as the number of messages per day posted by a user, measured over a time period. These three quantities all display long-tailed distributions (Fig. 4(a,b,c)). The excellent collapse of the curves demonstrates that the distributions are



**Figure 4** | Empirical regularities in Twitter data. (a) Probability distribution of the lifetime of a meme using hours (red circles), days (blue squares), and weeks (green triangles) as time units. In the plot, units are converted into hours. Since the distributions are well approximated by a power law, we can align the curves by rescaling the  $y$ -axis by  $\lambda^{-\alpha}$ , where  $\lambda$  is the ratio of the time units (e.g.,  $\lambda = 24$  for rescaling days into hours) and  $\alpha \approx 2.5$  is the exponent of the power law (via maximum likelihood estimation<sup>33</sup>). This demonstrates that the shape of the lifetime distribution is not an artifact of the time unit chosen to define the lifetime. (b) Complementary cumulative probability distribution of the popularity of a meme, measured by the total number of users per day who have used that meme. This and the following measures were performed daily (filled red circles), weekly (filled blue squares), and monthly (filled green triangles). (c) Complementary cumulative probability distribution of user activity, measured by the number of messages per day posted by a user. (d) Probability distribution of breadth of user attention (entropy), based on the memes tweeted by a user. Note that the larger the number of posts produced, the smaller the non-zero entropy values recorded for users who focus on a small set of memes. This explains why the distributions for longer periods of time extend further to the left.



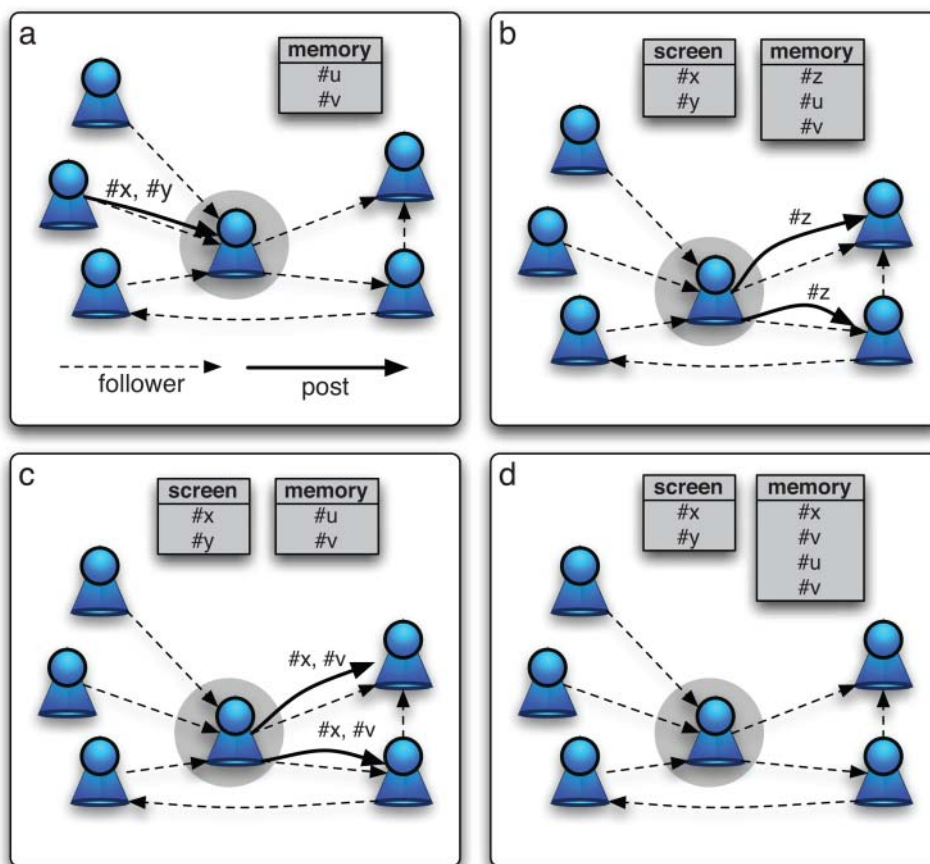
robust even if measured over different time units or observed over different periods of time. We further measure the breadth of user attention, defined earlier through the meme entropy. Although the entropy distribution is peaked, some users have broad attention while others are very focused (Fig. 4(d)). This distribution is also robust with respect to different periods of time.

All of these empirical findings point to extremely heterogeneous behaviors; some memes are extremely successful (popular and persistent), while the great majority die quickly. A small fraction of memes therefore account for the great majority of all posts. Likewise, a small fraction of users account for most of the traffic. These heterogeneities can in principle be attributed to a variety of causes. The broad distributions of meme popularity could result from a diversity in some intrinsic meme value, with “important” memes attracting more attention. Long-lived memes might be sustained exogenously by traditional media and real-world events. User activity and breadth of attention distributions could be a reflection of innate behavioral differences. What is, then, a minimal set of assumptions necessary to interpret this empirical data? One way to tackle this question is to start from a minimalist model of information spreading that assumes none of the above externalities. In particular we will explore to what extent the statistical features of memes and users can be accounted by the limited attention capacity of the users coupled with the heterogeneity of their social connections.

**Model description.** Our basic model assumes a frozen network of agents. An agent maintains a time-ordered list of *posts*, each about a specific *meme*. Multiple posts may be about the same meme. Users

pay attention to these memes only. Asynchronously and with uniform probability, each agent can generate a post about a new meme or forward some of the posts from the list, transmitting the corresponding memes to neighboring agents. Neighbors in turn pay attention to a newly received meme by placing it at the top of their lists. To account for the empirical observation that past behavior affects what memes the user will spread in the future, we include a memory mechanism that allows agents to develop endogenous interests and focus. Finally, we model *limited attention* by allowing posts to survive in an agent’s list or memory only for a finite amount of time. When a post is forgotten, its associated meme become less represented. A meme is forgotten when the last post carrying that meme disappears from the user’s list or memory. Note that list and memory work like first-in-first-out rather than priority queues, as proposed in models of bursty human activity<sup>34</sup>. In the context of single-agent behavior, our memory mechanism is reminiscent of the classic Yule-Simon model<sup>43</sup>, Cattuto30012007<sup>44</sup>).

The retweet model we propose is illustrated in Fig. 5. Agents interact on a directed social network of friends/followers. Each user node is equipped with a *screen* where received memes are recorded, and a *memory* with records of posted memes. An edge from a friend to a follower indicates that the friend’s memes can be read on the follower’s screen (#*x* and #*y* in Fig. 5(a) appear on the screen in Fig. 5(b)). At each step, an agent is selected randomly to post memes to neighbors. The agent may post about a new meme with probability  $p_n$  (#*z* in Fig. 5(b)). The posted meme immediately appears at the top of the memory. Otherwise, the agent reads posts about existing memes from the screen. Each post may attract the



**Figure 5 | Illustration of the meme diffusion model.** Each user has a memory and a screen, both with limited size. (a) Memes are propagated along follower links. (b) The memes received by a user appear on the screen. With probability  $p_n$  the user posts a new meme, which is stored in memory. (c) Otherwise, with probability  $1 - p_n$  the user scans the screen. Each meme *x* in the screen catches the user’s attention with probability  $p_r$ . Then with probability  $p_m$  a random meme from memory is triggered, or *x* is retweeted with probability  $1 - p_m$ . (d) All memes posted by the user are also stored in memory.



user's attention with probability  $p_r$  (the user pays attention to #x, #y in Fig. 5(c)). Then the agent either retweets the post (#x in Fig. 5(c)) with probability  $1 - p_m$ , or tweets about a meme chosen from memory (#v triggered by #y in Fig. 5(c)) with probability  $p_m$ . Any post in memory has equal opportunities to be selected, therefore memes that appear more frequently in memory are more likely to be propagated (the memory has two posts about #v in Fig. 5(d)). To model limited user attention, both screen and memory have a finite capacity, which is the time in which a post remains in an agent's screen or memory. For all agents, posts are removed after one *time unit*, which simulates a unit of real time, corresponding to  $N_u$  steps where  $N_u$  is the number of agents. If people use the system once weekly on average, the time unit corresponds to a week.

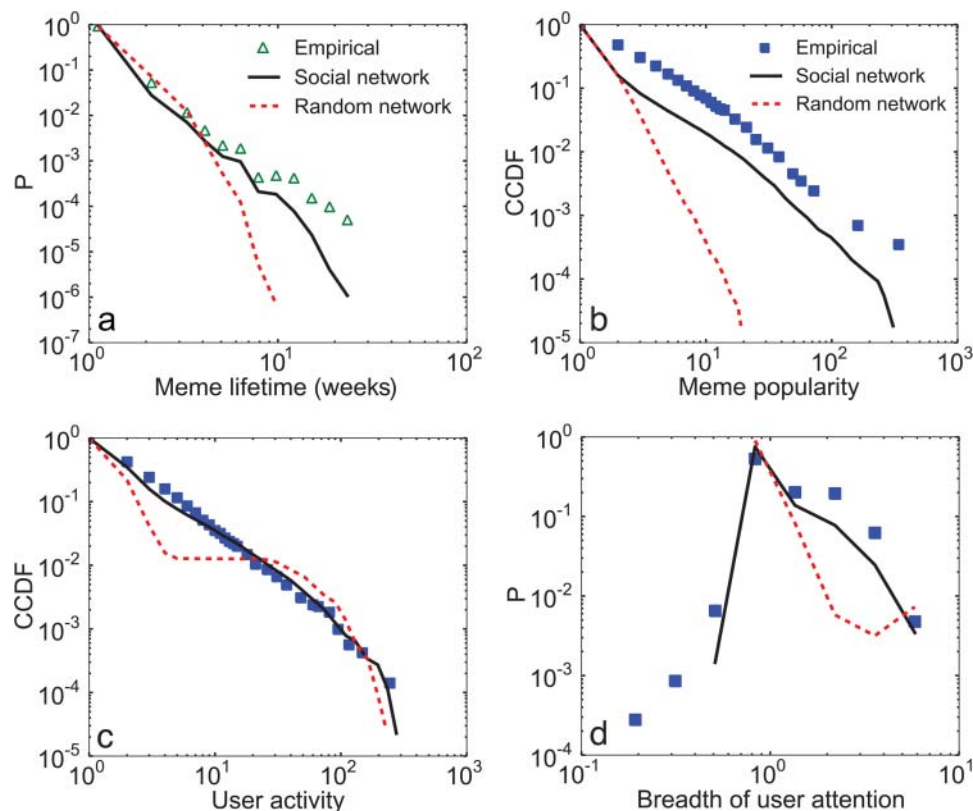
**Simulation results.** The model has three parameters:  $p_n$  regulates the amount of novelty that enters the system (number of cascades),  $p_r$  determines the overall retweet activity (size of cascades), and  $p_m$  accounts for individual focus (diversity of user interests). We estimated all three directly from the empirical data (see Methods).

The social network underlying the meme diffusion process is a critical component of the model. To obtain a network of manageable size while preserving the structure of the actual social network, we sampled a directed graph with  $10^5$  nodes from the Twitter follower network (details in Methods). The nodes correspond to a subset of the users who generated the posts in our empirical data. To evaluate the predictions of our model, we compare them with empirical data that includes only the retweets of the same subset of users. To study the role played by the network structure in the meme diffusion process, we also simulated the model on a random Erdős-Rényi (ER) network with the same number of nodes and edges. As shown in Fig. 6, the model captures the main features of the empirical distributions of meme lifetime and popularity, user activity, and breadth

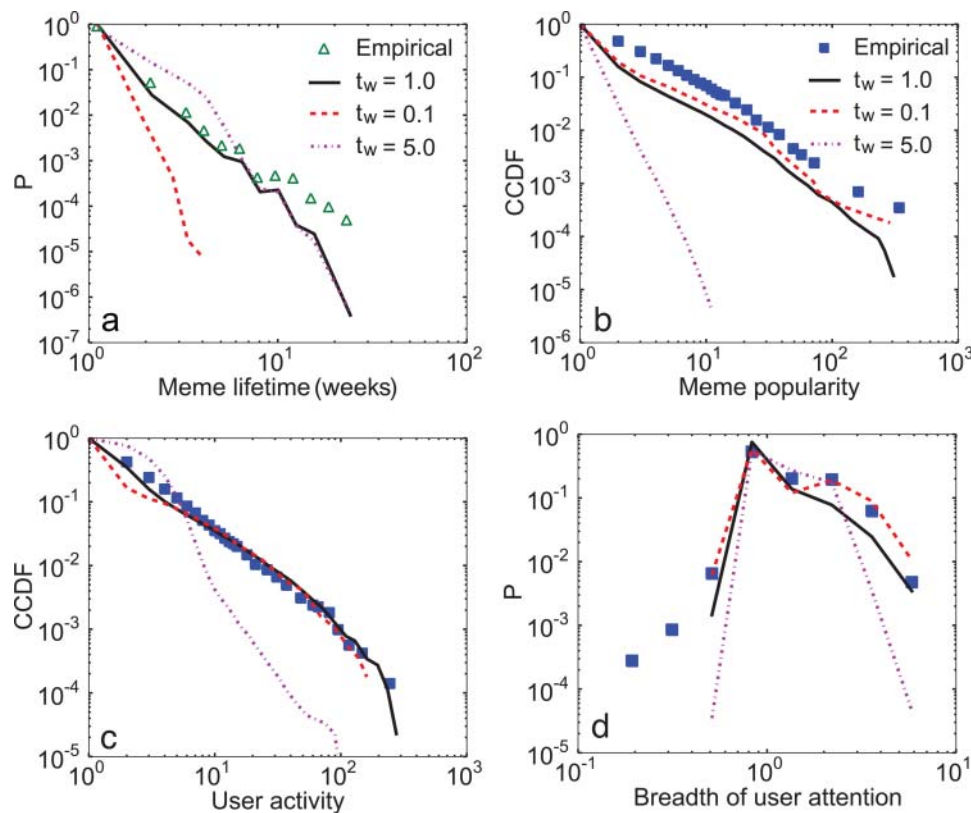
of user attention. The comparison with the corresponding distributions generated using the ER network shows that in general, the heterogeneity of the observed quantities is greatly reduced when memes spread on a random network. This is not unexpected. Consider for example meme popularity (Fig. 6(b)); the real social network has a broad (scale free, not shown) distribution of degree, with a consistent number of hub users who have a large number of followers. Memes spread by these users are likely to achieve greater popularity. This does not happen in the ER network where the degree distribution is narrow (Poissonian). The difference observed in the distribution of breadth of user attention, for both low and high entropy values (Fig. 6(d)), may be explained by the heterogeneity in the number of friends. Users with few friends may have low breadth of attention while those with many friends are exposed to many memes and thus may exhibit greater entropy.

The second key ingredient of our model is the competition among memes for limited user attention. To evaluate the role of such a competition on the meme diffusion process, we simulated variations of the model with stronger or weaker competition. This was accomplished by tuning the length  $t_w$  of the time window in which posts are retained in an agent's screen or memory. A shorter time window ( $t_w < 1$ ) leads to less attention and thus increased competition, while a longer time window ( $t_w > 1$ ) allows for attention to more memes and thus less competition. As we can observe in Fig. 7, stronger competition ( $t_w = 0.1$ ) fails to reproduce the large observed number of long-lived memes (Fig. 7(a)). Weaker competition ( $t_w = 5$ ), on the other hand, cannot generate extremely popular memes (Fig. 7(b)) nor extremely active users (Fig. 7(c)).

We also simulated our model without user interests, by setting  $p_m = 0$ . The most noticeable difference in this case is the lack of highly focused individuals. Users have no memory of their past behavior, and can only pay attention to memes from their friends.



**Figure 6** | Evaluation of model by comparison of simulations with empirical data (same panels and symbols as in Fig. 4). To study the role played by the network structure in the meme diffusion process, we simulate the model on the sampled follower network (solid black line) and a random network (dashed red line). Both networks have  $10^5$  nodes and about  $3 \times 10^6$  edges. (a) The definition of lifetime uses the week as time unit. (b,c,d) Meme popularity, user activity, and user entropy data are based on weekly measures.



**Figure 7 | Evaluation of model by comparison of simulations with empirical data (same panels and symbols as in Fig.4).** To study the role of meme competition, we simulate the model on the sampled follower network with different levels of competition; posts are removed from screen and memory after  $t_w$  time units. We compare the standard model ( $t_w = 1$ , solid black line) against versions with less competition ( $t_w = 5$ , dot-dashed magenta line) and more competition ( $t_w = 0.1$ , dashed red line). (a) The definition of lifetime uses the week as time unit. (b,c,d) Meme popularity, user activity, and user entropy data are based on weekly measures.

As a result, the model fails to account for low entropy individuals (not shown but similar to the random network case in Fig. 6(d)).

## Discussion

The present findings demonstrate that the combination of social network structure and competition for finite user attention is a *sufficient condition* for the emergence of broad diversity in meme popularity, lifetime, and user activity. This is a remarkable result: one can account for the often-reported long-tailed distributions of topic popularity and lifetime<sup>7,12,14,29</sup> without having to assume exogenous factors such as intrinsic meme appeal, user influence, or external events. The only source of heterogeneity in our model is the social network; users differ in their audience size but not in the quality of their messages.

Our model is inspired by the long tradition that represents information spreading as an epidemic process, where infection is passed along the edges of the underlying social network<sup>35–37,7,28,12</sup>.

In the context of social media, several authors explored the temporal evolution of popularity. Wu and Huberman<sup>8</sup> studied the decay in news popularity. They showed that temporal patterns of collective attention are well described by a multiplicative process with a single novelty factor. While the decay in popularity is attributed to competition for attention, the underlying mechanism is not modeled explicitly. Crane and Sornette<sup>10</sup> introduced a model to describe the exogenous and endogenous bursts of attention toward a video, by combining an epidemic spreading process with a forgetting mechanism. Hogg and Lerman<sup>38</sup> proposed a stochastic model to predict the popularity of a news story via the intrinsic interest of the story and the rates at which users find it directly and through friends. These models describe the popularity of a single piece of information,

and are therefore unsuitable to capture the competition for our collective attention among multiple simultaneous information epidemics. Although recent epidemiological models have started considering the simultaneous spread of competing strains<sup>39,40</sup>, our framework is the first attempt to deal with a virtually unbounded number of new “epidemics” that are continuously injected into the system. A closer analogy to our approach is perhaps provided by neutral models of ecosystems, where individuals (posts) belonging to different species (memes) produce offspring in an environment (our collective attention) that can sustain only a limited number of individuals. At every generation, individuals belonging to new species enter the ecosystem while as many individuals die as needed to maintain the sustainability threshold<sup>41</sup>.

Since Simon’s seminal paper<sup>4</sup>, the economy of attention has been an enormously popular notion, yet it has always been assumed implicitly and never put to the test. Our model provides a first attempt to focus explicitly on mechanisms of competition, and to evaluate the quantitative effects of making attention more scarce or abundant.

Our results do not constitute a proof that exogenous features, like intrinsic values of memes, play no role in determining their popularity. However we have shown that at the statistical level it is not necessary to invoke external explanations for the observed global dynamics of memes. This appears as an arresting conclusion that makes information epidemics quite different from the basic modeling and conceptual framework of biological epidemics. While the intrinsic features of viruses and their adaptation to hosts are extremely relevant in determining the winning strains, in the information world the limited time and attention of human behavior are sufficient to generate a complex information landscape and define a wide range



**Table 1 | Parameter settings for different simulations. “Avg user activity” is the average number of posts per user per time unit**

Simulation	Network	$t_w$	$p_r$	Avg user activity
Standard	Sampled	1.0	0.016	2.5
ER Network	Synthetic	1.0	0.029	2.1
Weak Competition	Sampled	5.0	0.205	2.3
Strong Competition	Sampled	0.1	0.001	2.7

of different meme spreading patterns. This calls for a major revision of many concepts commonly used in the modeling and characterization of meme diffusion and opens the path to different frameworks for the analysis of competition among ideas and strategies for the optimization/suppression of their spread.

## Methods

The data analyzed in this paper was obtained through Twitter’s public APIs. We collected more than 120 millions retweets from October 2010 to January 2011, involving 12.5 million distinct users and 1.3 million hashtags. Each post contains information about who generated and who retweeted it. As expected in a social network, the follower graph has scale-free degree distributions.

Due to the size of the empirical follower network, we sampled a manageable subset for our simulations. The sampling procedure was a random walk with occasional restarts from random locations (teleportation factor 0.15). Though no sampling method is perfect, the modified random walk is efficient in terms of API queries and reproduces the salient topological features of the sampled network<sup>42</sup>. The sampled network has  $10^5$  nodes and about  $3 \times 10^6$  edges. The empirical retweets generated by the users in the sample display trends similar to those from the entire dataset, therefore we expect the model predictions to be consistent not only with the sample but also with the full dataset.

The parameter  $p_n$  characterizes the probability of tweeting about a new meme. To estimate this parameter from the empirical data, we examine whether each hashtag has been observed in previous time units (weeks). The proportion of posts with new hashtags is approximately  $0.45 \pm 0.05$ . We thus set  $p_n = 0.45$  for all the simulations. For each simulation — standard model, model with underlying random network, and models with strong and weak competition — the parameter  $p_r$  is tuned to capture the average number of posted memes per user per unit time (Table 1). Finally, the parameter  $p_m$  represents the proportion of all memes tweeted by an individual that match the content of the memory. To estimate it from the empirical data, we compare each hashtag with those produced by a user in the previous time unit (week). Using the average value across all users ( $0.4 \pm 0.01$ ) we set  $p_m = 0.4$ .

- Davenport, T. H. & Beck, J. C. *The Attention Economy: Understanding the New Currency of Business* (Harvard Business School Press, 2001).
- Tapscott, D. & Williams, A. D. *Wikinomics: How Mass Collaboration Changes Everything* (Portfolio Hardcover, 2006).
- Dawkins, R. *The selfish gene* (Oxford University Press, 1989).
- Simon, H. Designing organizations for an information-rich world. In Greenberger, M. (ed.) *Computers, Communication, and the Public Interest*, 37–52 (The Johns Hopkins Press, Baltimore, 1971).
- Goldhaber, M. H. The attention economy and the net. *First Monday* **2** (1997).
- Morris, S. Contagion. *Rev. Econ. Studies* **67**, 57–78 (2000).
- Watts, D. J. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences* **99**, 5766–5771 (2002).
- Wu, F. & Huberman, B. A. Novelty and collective attention. *Proceedings of the National Academy of Sciences* **104**, 17599–17601 (2007).
- Falkinger, J. Attention economies. *Journal of Economic Theory* **133**, 266–294 (2007).
- Crane, R. & Sornette, D. Robust dynamic classes revealed by measuring the response function of a social system. *Proc. of the National Academy of Sciences* **105**, 15649–15653 (2008).
- Leskovec, J., Backstrom, L. & Kleinberg, J. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 497–506 (ACM, New York, NY, USA, 2009).
- Goetz, M., Leskovec, J., McGlohon, M. & Faloutsos, C. Modeling blog dynamics. In *Proc. Third International AAAI Conference on Weblogs and Social Media* (2009).
- Lerman, K. & Ghosh, R. Information contagion: an empirical study of the spread of news on digg and twitter social networks. In *Proc. Fourth International AAAI Conference on Weblogs and Social Media* (2010).
- Ratkiewicz, J., Fortunato, S., Flammini, A., Menczer, F. & Vespignani, A. Characterizing and modeling the dynamics of online popularity. *Phys. Rev. Lett.* **105**, 158701 (2010).

- Onnela, J.-P. & Reed-Tsochas, F. Spontaneous emergence of social influence in online systems. *Proceedings of the National Academy of Sciences* **107**, 18375–18380 (2010).
- Lazer, D. et al. Computational social science. *Science* **323**, 721–723 (2009).
- Vespignani, A. Predicting the behavior of techno-social systems. *Science* **325**, 425–428 (2009).
- Moussaid, M., Helbing, D. & Theraulaz, G. An individual-based model of collective attention. In *Proceedings of the European Conference on Complex Systems* (2009).
- Asur, S., Huberman, B. A., Szabo, G. & Wang, C. Trends in social media: Persistence and decay. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media* (2011).
- Romero, D. M., Galuba, W., Asur, S. & Huberman, B. A. Influence and passivity in social media. In *Proceedings of the 20th International Conference on World Wide Web (Companion Volume)*, 113–114 (ACM, 2011).
- Bakshy, E., Mason, W. A., Hofman, J. M. & Watts, D. J. Everyone’s an influencer: Quantifying influence on twitter. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining* (2011).
- Wu, F. & Huberman, B. J. A persistence paradox. *First Monday* **15** (2010).
- Romero, D. M., Meeder, B. & Kleinberg, J. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. In Srinivasan, S. et al. (eds.) *Proceedings of the 20th International Conference on World Wide Web* (ACM, 2011).
- Aral, S., Muchnik, L. & Sundararajan, A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences* **106**, 21544–21549 (2009).
- Lehmann, J., Gonçalves, B., Ramasco, J. J. & Cattuto, C. Dynamical classes of collective attention in twitter. In *Proc. 21st International World Wide Web Conference (WWW)* (2012).
- Dunbar, R. I. M. The social brain hypothesis. *Evolutionary Anthropology* **6**, 178–190 (1998).
- Gonçalves, B., Perra, N. & Vespignani, A. Validation of dunbar’s number in twitter conversations. *PLoS One* **6**, e22656 (2011).
- Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N. & Hurst, M. Cascading behavior in large blog graphs: Pattern and a model. Tech. Rep. 0704.2803, arXiv (2007). URL <http://arxiv.org/abs/0704.2803>.
- Ienco, D., Bonchi, F. & Castillo, C. The meme ranking problem: Maximizing microblogging virality. *Journal of Intelligent Information Systems* (Forthcoming).
- Yang, L., Sun, T. & Mei, Q. We Know What @You #Tag: Does the Dual Role Affect Hashtag Adoption? In *Proc. 21st International World Wide Web Conference (WWW)* (2012).
- Markines, B. et al. Evaluating similarity measures for emergent semantics of social tagging. In *Proc. Intl. World Wide Web Conf.*, 641–650 (2009).
- Markines, B. & Menczer, F. A scalable, collaborative similarity measure for social annotation systems. In *Proc. ACM Conf. on Hypertext and Hypermedia (HT)*, 347–348 (2009).
- Clauset, A., Shalizi, C. R. & Newman, M. E. J. Power-law distributions in empirical data. *SIAM Review* **51**, 661–703 (2009).
- Barabási, A.-L. & Albert, R. The origin of bursts and heavy tails in human dynamics. *Nature* **435**, 207–211 (2005).
- Goffman, W. & Newill, V. A. Generalization of epidemic theory: An application to the transmission of ideas. *Nature* **204**, 225–228 (1964).
- Daley, D. J. & Kendall, D. G. Epidemics and rumours. *Nature* **204**, 1118–1119 (1964).
- Bailey, N. *The Mathematical Theory of Infectious Diseases and its Applications* (Griffin, London, 1975), 2nd edn.
- Hogg, T. & Lerman, K. Stochastic models of user-contributory web sites. In *Proc. Third International AAAI Conference on Weblogs and Social Media (ICWSM)* (2009).
- Sneppen, K., Trusina, A., Jensen, M. H. & Bornholdt, S. A minimal model for multiple epidemics and immunity spreading. *PLoS One* **5**, e13326 (2010).
- Kerrer, B. & Newman, M. E. J. Competing epidemics on complex networks. Tech. Rep. 1105.3424, arXiv (2011).
- Pigolotti, S., Flammini, A. & Maritan, A. A stochastic model for the species abundance problem in an ecological community. *Physical Review E* **70**, 011916 (2004).
- Leskovec, J. & Faloutsos, C. Sampling from large graphs. In *Proc. of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 631–636 (ACM, 2006).
- Simon, H. A. et al. On a class of skew distribution functions. *Science* **42**, 425–440 (1955).
- Cattuto, C., Loreto, V., Pietronero, L. & Semiotic dynamics and collaborative tagging. *Proceedings of the National Academy of Sciences* **104**, 1461–1464 (2007).

## Acknowledgements

We thank Bruno Gonçalves, Michael Conover, and Jacob Ratkiewicz for assistance in data collection and discussions. We acknowledge Twitter for making the data available. This work was supported in part by NSF (grants IIS-0811994 and CCF-1101743), Lilly Endowment (Data to Insight Center grant), and the James S. McDonnell Foundation.



### Author contributions

LW, AF and FM performed empirical analysis and developed the model. LW prepared the figures. All authors contributed to model evaluation and to the manuscript.

### Additional information

**Competing financial interests:** The authors declare that they have no competing financial interests.

**License:** This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>

**How to cite this article:** Weng, L., Flammini, A., Vespignani, A. & Menczer, F. Competition among memes in a world with limited attention. *Sci. Rep.* 2, 335; DOI:10.1038/srep00335 (2012).