Computational Analysis of Collective Behaviors via Agent-Based Modeling

Lilian Weng and Filippo Menczer

Abstract Agent-based modeling (ABM) is a common computational analysis tool to study system dynamics. In the framework of ABM, the system consists of multiple autonomous and interacting agents. We can explore emergent collective patterns by simulating the individual operations and interactions between agents. As a case study, we present an experiment using an agent-based model to study how competition for limited user attention in a social network results in collective patterns of meme popularity. The 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. The model also builds upon empirical observations on how individual humans behave online. The combination of social network structure and finite agent attention is sufficient for the emergence of broad diversity in meme popularity and lifetime. The case study illustrates how one can analyze the kind of emergent human computation that makes some memes very popular.

1 Introduction

Agent-based modeling (ABM) is a class of computational analysis tools that is widely used for simulating system dynamics when the system consists of multiple autonomous and interacting individual components — named *agents*. Each agent follows its own decision-making processes according to a set of rules and contextual information from history, other agents, and possibly other environmental settings. The sets of behavioral rules can be identical for all agents (*homogeneous*) or different from agent to agent (*heterogeneous*). For example, in prisoner dilemma games, every agent follows the same strategy to negotiate; in an ecosystem, some agents play the roles of producers while others are consumers.

Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University Bloomington. e-mail: `{weng,fil}@indiana.edu`

Eric Bonabeau summarized the key characteristics of ABMs from the perspectives of *capturing emergent phenomena, natural descriptions of systems,* and *flexibility* [1]. The key point of ABM is to describe a system by setting up the behavioral strategies of its constituent agents. ABMs are often applied to validate individual-level configurations by comparing the patterns that result at the system level with empirical data. Model predictions are typically obtained by computational simulations, in which the outcomes of interactions between agents are repetitively calculated. This approach makes it possible to make predictions that reach beyond those derived by pure mathematical methods, when the model cannot be solved analytically.

Agent-based modeling has now obtained a central role in the study of natural systems. A large body of literature has been developing in the past few years about the internal characteristics of agents, their activities, connectivity, and multi-agent features [2]. Biological, ecological, human collaborative systems, and society can be naturally translated into an agent-based framework. ABM techniques are therefore employed in domains that include biology, ecology, cognitive science, epidemiology, and the social sciences. Let us consider a few domains to demonstrate the usage of ABMs in practice.

Social dynamics. The Axelrod model [3] investigated cultural dynamics by modeling individuals as nodes (agents) in networks in which whether a person interacts with another depends on the similarity between their statuses. Global convergence and the persistence of diversity are two important ingredients explored in this setting. Holme and Newman [4] proposed a model in which each agent is associated with an opinion. At each time step, agents either change their opinions to match neighbors, or re-wire links toward agents with similar opinions. The model can capture the dual process of social influence via opinion changes and selection via re-wiring of connections. Agent-based models have also been applied to the study of the birth and decline of scientific disciplines [5]. The evolution of disciplines is guided by social interactions among agents representing scientists. Disciplines emerge from splitting and merging of social communities in a collaboration network. This model is capable of reproducing various empirical observations about the relationships between disciplines, scholars, and publications. Many more models of social dynamics are reviewed in the literature [2].

Network Evolution. In the above examples, agents are connected in a network structure. This is often the case in ABMs, especially in the context of social systems. The edges in the network represent social relationships between pairs of individuals. Some models focus specifically on the local rules that regulate the growth and evolution of the network and lead to its observed global topology. Models have explored many different strategies of how an agent creates connections [6, 7, 8]. For example, the phenomenon of linking to well-connected nodes (e.g., people or Web pages) is described by preferential attachment mechanisms [8]. Other ingredients considered in network evolution models include homophily [9] and triadic closure [10, 11, 12].

Diffusion. Information and innovation spread on networks, and we can observe the cascades that ensue as agents are infected. The diffusion process is affected by the actions of agents and the underlying network structure. Watts studies the cascade sizes and vulnerability of the system to global cascades using a simple spreading process on random networks [13]. Pastor-Satorras and Vespignani simulated classical epidemic models on scale-free networks [8], revealing that infections always survive no matter how small the spreading rate [14]. Goetz *et al.* [15] proposed an agent-based model of blog dynamics, where each agent is associated with mechanisms capturing both the topology and temporal features.

2 Case Study: Meme Competition for Limited Attention

Here we present a case study using agent-based modeling to understand how limited human attention may constrain competition among ideas [16].

The advent of social media [17, 18] has lowered the cost of information production and broadcasting, boosting the potential reach of each idea or *meme* [19]. 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, increasing the competition among ideas for our finite attention. As a result, the dynamic of information is driven more than ever before by the economy of attention [20]. In this context one of the most challenging problems is the study of the competition dynamics of ideas, information, knowledge, and rumors [21, 22, 23, 24]. Studying limited user attention is motivated by the cognitive limit on the number of stable social relationships that one can sustain [25, 26]. However, it is hard to disentangle the effects of limited attention from many concurrent factors, such as the underlying network structure [13], the activity of users [27], homophily [9], and the intrinsic quality of the information [28].

We can think of the collective actions of many individuals, as they decide which information to propagate through the network, as a computational system that produces a concrete output, namely, a small number of very popular memes while the majority of memes go mostly unnoticed. An agent-based model allows to test micro-level hypotheses about this type of human computation: can certain individual behaviors be responsible for the patterns observed at the collective level?

The design of agent operation strategies in our model in inspired by empirical observations about individual behaviors, outlined in § 2.1. We then describe an agent-based toy model of meme diffusion and compare its predictions with the empirical data in § 2.2. Finally, in § 2.3 we show that the social network structure and our finite attention are both key ingredients of the diffusion process, as their removal from the model leads to results inconsistent with the empirical observations.

2.1 Empirical Observations

We investigate this problem using a sample of data from *Twitter*, a micro-blogging platform that allows millions of people to broadcast short messages through social connections. Users post short messages (“tweets”), subscribe to (“follow”) people to receive their tweets, and forward (“retweet”) selected posts to their followers. Posts may contain special topic labels (“hashtags”), which we use to identify *memes* operationally. This provides us with a quantitative framework to study the competition for attention in the wild.

Here we outline two empirical findings that motivate both our question and the main assumptions behind our model [16]. First, the attention of a user is independent from the overall diversity of information discussed in a given period. A user’s daily breadth of attention (measured through Shannon entropy) remains roughly constant and bound irrespective of system diversity, which varies greatly day by day. Second, users are more likely to retweet memes about which they posted in the past, suggesting that user memory is an important component for modeling information diffusion [29, 30].

2.2 Model Description

We propose an agent-based model to simulate the retweeting behavior of Twitter users, which explicitly incorporates the above observations about *memory-based user interests* and *limited attention*.

Our basic model assumes a frozen friends/followers network of agents. An agent maintains two time-ordered lists of memes: a *screen* that stores received memes and a *memory* that records posted memes, capturing endogenous interests. Users pay attention to memes in these lists only. At each time step, an agent is randomly selected with uniform probability to transmit a few memes to neighboring agents. The selected agent can generate a new meme or forward some memes from the list and store the posted memes in memory. Neighbors in turn pay attention to a newly received meme by placing it at the top of their screens.

To model limited user attention, both screen and memory have a finite capacity so that memes only survive for a finite amount of time. Fig. 1 illustrates details of the model.

2.3 Simulation

To evaluate the simulation outcomes of the model, we measured several regularities in the empirical data. The meme lifetime and popularity display long-tailed distributions, meaning that a few memes gain huge popularity.

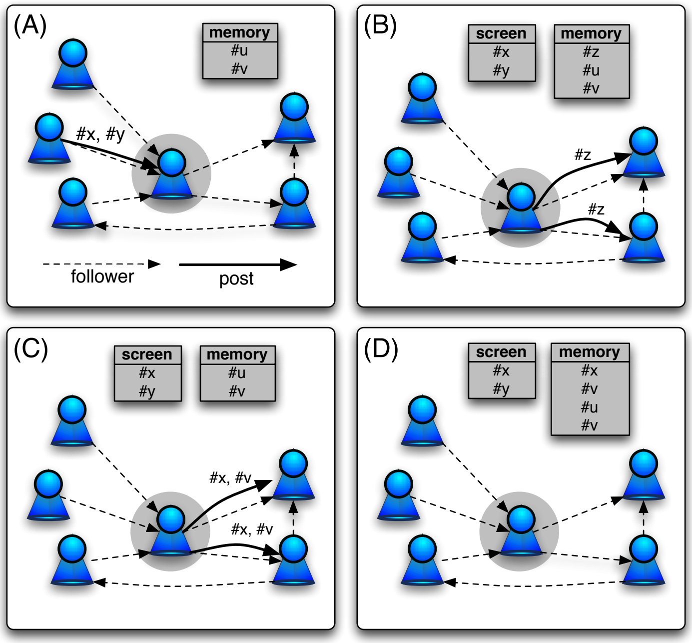

Fig. 1 Illustration of the meme diffusion model [16]. Tables show the memory and screen of the center user. (A) The memory contains memes `#u` and `#v`, that the center user has produced in the past. Memes `#x` and `#y` are received along follower links. (B) The received memes, `#x` and `#y`, appear on the screen. With probability p_n, the center user who is selected at this time step posts a new meme `#z` to his followers. (C) Otherwise, with probability $1-p_n$, the user scans the screen. Each meme h in the screen catches the user's attention with probability p_r; with probability p_m a random meme from memory is triggered, or h is retweeted with probability $1-p_m$. In the illustrated case, `#x` is retweeted and `#v` is triggered by `#y`. (D) All memes posted by the user are stored in memory. Parameters p_n, p_r and p_m are estimated from the empirical data.

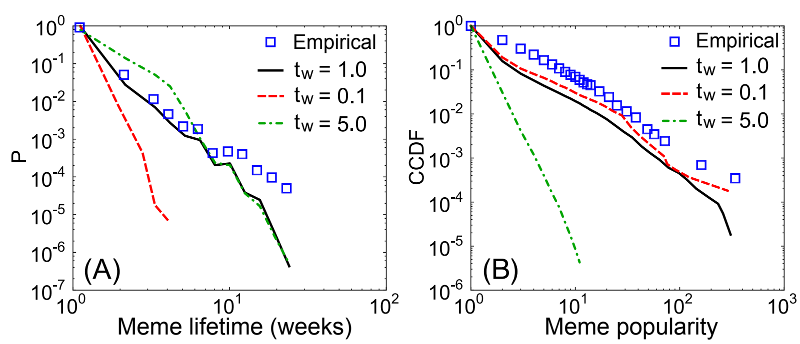

Fig. 2 Evaluation of model by comparison of simulations with empirical distributions of meme lifetime (left) and meme popularity (right). We simulate the model on the real social network with different levels of competition; posts are removed from screen and memory after t_w time units. We compare the standard model ($t_w = 1$) against versions with less and more competition ($t_w = 5$ and $t_w = 0.1$, respectively).

Our aim is to determine a minimal set of individual-level assumptions necessary to interpret these collective patterns. To evaluate the role of the competition among memes for limited user attention, we simulated variations of the model with *stronger* or *weaker* competition 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 leads to less attention and thus strong competition, while a longer time window allows for weaker competition. As shown in simulation results, stronger competition fails to reproduce the large observed number of long-lived memes (Fig. 2(A)), while weaker competition cannot generate extremely popular memes (Fig. 2(B)).

To gauge the role of underlying network structure in shaping the diffusion process, we simulated the model on both the real social network and a random network. The model is able to reproduce the main empirical features on the social network, while the observed heterogeneity is largely reduced on a random network [16]. The structure of the network is thus another key ingredient of the system dynamics.

3 Conclusion

In this chapter we discussed agent-based modeling as a powerful computational tool to analyze system dynamics, and in particular to study global patterns by simulating individual behaviors and interactions between agents [1].

As an example, we presented a case study where agent-based modeling was used to shed light on how memes compete for limited attention on social networks [16]. The computational approach allows us to demonstrate that, surprisingly, a combination of social network structure and competition for finite user attention is a sufficient condition for the emergence of broad diversity in meme popularity and lifetime, without having to assume exogenous factors.

In general, agent-based models can be very useful in identifying minimal hypotheses consistent with collective patterns generated by human computation systems.

References

1. Bonabeau, E.: Agent-based modeling: Methods and techniques for simulating human systems. Proc. Natl. Acad. Sci., **99**(Suppl 3), 7280–7287 (2002).
2. Castellano, C., Fortunato, S., Loreto, V..: Statistical physics of social dynamics. Reviews of modern physics, **81**(2), 591 (2009).
3. Axelrod, R.: The complexity of cooperation: Agent-based models of competition and collaboration. Princeton University Press (1997).
4. Holme, P., Newman, M. E. J.: Nonequilibrium phase transition in the coevolution of networks and opinions. Phys. Rev. E, **74**(5), 2006.
5. Sun, X., Kaur, J., Milojevic, S., Flammini, A., Menczer, F.: Social dynamics of science. Sci. Rep., **3**(1069) (2013).

6. Erdös, P., Rényi, A.: On the evolution of random graphs. Magyar Tud. Akad. Mat. Kutató Int. Közl, **5**, 17–61 (1960).
7. Watts, D. J., Strogatz, S. H.: Collective dynamics of 'small-world' networks. Nature, **393**, 440–442 (1998).
8. Barabási, A.-L., Albert, R.: Emergence of scaling in random graphs. Science, **286**, 509–511 (1999).
9. McPherson, M., Lovin, L., Cook, J.: Birds of a feather: Homophily in social networks. Annual Review of Sociology, **27**(1), 415–444 (2001).
10. Granovetter, M.: The strength of weak ties. American Journal of Sociology, **78**, 1360–1380 (1973).
11. Shi, X., Adamic, L. A., Strauss, M. J.: Networks of strong ties. Physica A, **378**, 3347 (2007).
12. Leskovec, J., Backstrom, L., Kumar, R., Tomkins, A.: Microscopic evolution of social networks. In: Proc. SIGKDD Intl. ACM Conf. on Knowledge discovery and data mining, 462–470 (2008).
13. Watts, D. J.: A simple model of global cascades on random networks. Proc. Natl. Acad. Sci., **99**(9), 5766–5771 (2002).
14. Pastor-Satorras, R., Vespignani, A.: Epidemic spreading in scale-free networks. Phys. Rev. Lett., **86**, 3200–3203 (2001).
15. Goetz, M., Leskovec, J., McGlohon, M., Faloutsos, C.: Modeling blog dynamics. In: Proc. Intl. AAAI Conf. on Weblogs and Social Media (2009).
16. Weng, L., Flammini A., Vespignani A., Menczer, F.: Competition among memes in a world with limited attention. Scientific Reports, **2**(335) (2012).
17. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Alstyne, M.V.: Computational social science. Science, **323**(5915), 721–723 (2009).
18. Vespignani, A.: Predicting the behavior of techno-social systems. Science, **325**(5939), 425–428 (2009).
19. Dawkins, R.: The selfish gene. Oxford University Press (1989).
20. Simon, H.: Designing organizations for an information-rich world. In: Greenberger, M. (eds.) Computers, Communication, and the Public Interest, pp. 37–52. The Johns Hopkins Press, Baltimore (1971).
21. Crane, R., Sornette, D.: Robust dynamic classes revealed by measuring the response function of a social system. Proc. Natl. Acad. Sci., **105**(41), 15649–15653 (2008).
22. Lerman, K., Ghosh R.: Information contagion: an empirical study of the spread of news on digg and twitter social networks. In: Proc. Intl. AAAI Conf. on Weblogs and Social Media (2010).
23. Wu, F., Huberman, B. A.: Novelty and collective attention. Proc. Natl. Acad. Sci., **104**(45),17599–17601 (2007).
24. Moussaid, M., Helbing, D., Theraulaz, G.: An individual-based model of collective attention. In: Proc. European Conf. on Complex Systems (2009).
25. Dunbar, R.I. M.: The social brain hypothesis. Evolutionary Anthropology, **6**, 178–190 (1998).
26. Gonçalves, B., Perra, N., Vespignani, A.: Validation of dunbar's number in twitter conversations. PLOS One, 6:e22656 (2011).
27. Asur, S., Huberman B. A., Szabo G., Wang C.: Trends in social media: Persistence and decay. In: Proceedings Intl. AAAI Conf. on Weblogs and Social Media (2011).
28. Bakshy, E., Mason W. A., Hofman, J. M., Watts, D. J.: Everyone's an influencer: Quantifying influence on twitter. In: Proc. ACM Intl. Conf. on Web Search and Data Mining (2011).
29. Ienco, D., Bonchi, F., Castillo, C.: The meme ranking problem: Maximizing microblogging virality. In: Proc. IEEE Intl. Conf. on Data Mining Workshop, 328–335 (2010).
30. Yang, L., Sun, T., Mei, Q.: We know what @you #tag: Does the dual role affect hashtag adoption? In: Proc. Intl. ACM World Wide Web Conf. (2012).