The Effects of Message Sorting in the Diffusion of Information in Online Social Media

Diego F. M. Oliveira[1,2,3,4](✉), Kevin S. Chan[5], and Peter J. Mucha[6]

[1] Statistical and Applied Mathematical Sciences Institute, Durhan, NC, USA
diegofregolente@gmail.com
[2] Department of Mathematics, University of North Carolina, Chapel Hill, NC, USA
[3] Social Data Science Center, University of Maryland, College Park, MD, USA
[4] College of Information Studies, University of Maryland, College Park, MD, USA
[5] U.S Army Research Laboratory, 2800 Powder Mill Rd, Adelphi, MD, USA
[6] Department of Mathematics, Dartmouth College, Hanover, NH, USA

Abstract. In this work, we propose an agent-based model to study the effects of message sorting on the diffusion of low- and high-quality information in online social networks. We investigate the case in which each piece of information has a numerical proxy representing its quality, and the higher the quality, the greater are the chances of being transmitted further in the network. The model allows us to study how sorting information in the agent's attention list according to their quality, node's influence and popularity affect the overall system's quality, diversity and discriminative power. We compare the three scenarios with a baseline model where the information is organized in a first-in first-out manner. Our results indicate that such an approach intensifies the exposure of high-quality information increasing the overall system's quality while preserving its diversity. However, it significantly decreases the system's discriminative power.

Keywords: Networks · Competition · Limited attention · Information load

1 Introduction

The introduction of online social media platforms such as Twitter and Facebook have completely changed the ways the modern civilization consumes and shares information. Due to the low cost of information production and broadcasting, we are exposed to hundreds if not thousands of messages, or memes [1], every day which exceeds by far our capacity of content consumption [2], and each piece of information must compete for our limited attention. As a result, only a tiny fraction of the information created ends up going viral, while the vast majority will simply never be re-transmitted and quickly forgotten. If from one side online social media can facilitate the interaction between people from different

H. Cherifi et al. (Eds.): COMPLEX NETWORKS 2016 2022, SCI 1077, pp. 98–107, 2023.
https://doi.org/10.1007/978-3-031-21127-0_9

parts of the globe, they also provide the perfect ecosystem for the spreading of low-quality information such as fake news and misinformation (i.e., information that is misleading or inaccurate) that can be very harmful to our society. Examples of misuses of such platforms can be seen during events like terrorist attacks, natural disasters and even to negatively affect our economy [3–6]. Such an interesting behaviour and real life implications caught the attention of the scientific community and the field of information diffusion experienced a significant growth and every year new models are proposed to better understand the relationship between humans and the new universe of online social media [7–12].

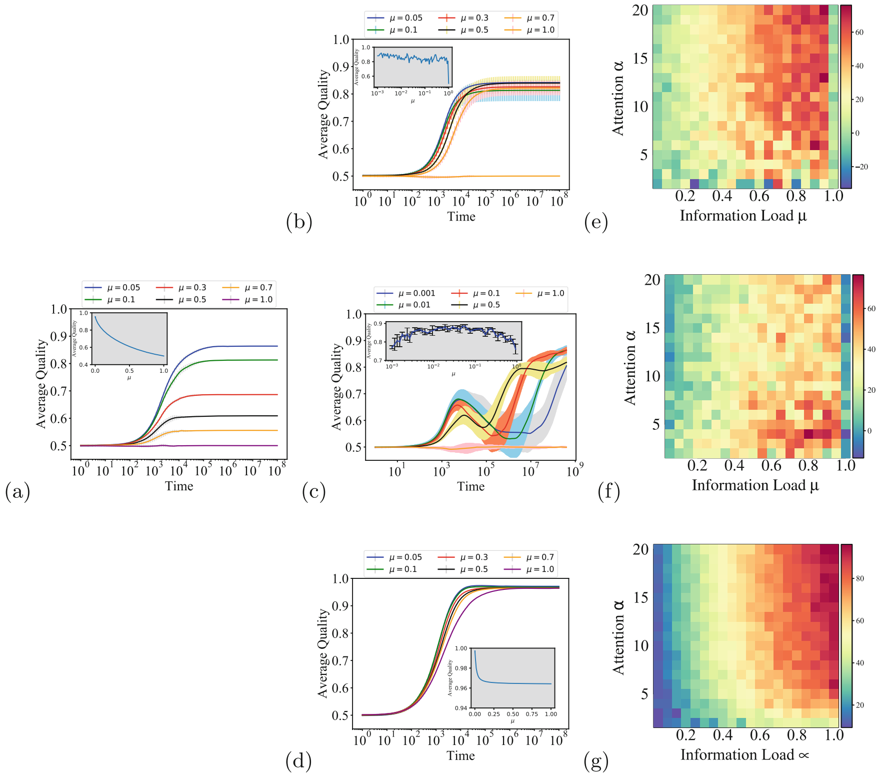

Fig. 1. Behaviour of the average system's quality as a function of time for **a** the baseline model and the model with sorted attention list according to **b** the meme's popularity, **c** node's influence and **d** meme's quality. The insets show the behaviour of the average system's quality at steady state for different values of the information load μ and with $\alpha = 14$. **e**–**g** Phase diagrams for different values of attention α and different values of the information load μ of the improvement in percent in the average system's quality at the steady state for the three models with respect to the baseline model. The error bars in (a–d) represent the standard deviation obtained from ten different simulations.

Traditionally, models of information diffusion are based on tools borrowed from theoretical epidemiology where susceptible agents became infected by interactions with infected agents and, in spite of their simplicity, they were able to reproduce several empirical observations [13–15]. For example, Weng [16], Gleeson [17] and Notarmuzi [18] have shown that a very simple model of information diffusion can produce a flat-tailed distribution for the popularity of a given meme. Such kinds of behaviors were commonly observed in a variety of systems that include citations [19–21], hashtags and URLs on Twitter [22], videos on YouTube [23], among many others [24–27]. Qiu et al. proposed a diffusion model that considers the user's limited attention and the quality of the information being transmitted. The authors showed that there exists a tradeoff between discriminative power and diversity. However, in realistic conditions, the model predicts that high-quality information has little advantage over low-quality information [28]. Simultaneously, Sreenivasan et al. [29] proposed a model of information cascades on feed-based networks also considering the finite attention, innovations and message diffusion. In such a case, the authors estimated the branching factor associated with the cascade process for different attention spans and different forwarding probabilities. They demonstrated that beyond a certain attention span, cascades tend to become viral. Ciampaglia et al. [30] proposed a model in which memes are selected based on their popularity or quality and the authors found that popularity bias hinders average quality when users are capable of exploring many items, as well as when they only consider very few top items due to scarce attention. They also found that an intermediate regime exists in which some popularity bias is good in distinguishing high-quality information, but too much can harm the system. More recently, Oliveira et al. investigated the impact of influential nodes on the spreading of information. The authors showed that a meme's quality does not guarantee virality, but there is a strong correlation between the meme's success and the influence of the agent who introduced it. Additionally, when trust is introduced into the model and the agents can decide whether or not to accept a meme, the authors observed that both lifetime and popularity distributions have broad power-law tails indicating that only a few memes spread virally through the population reproducing perfectly the broad distributions obtained from empirical data [31]. When considering situations where agents with heterogeneous criteria of quality, Cisneros-Velarde et al. proposed a simple method for enhancing the spread of high-quality information. Their results consist of strategically re-sorting the information feeds of users that share low-quality information. Under different settings of types of users, the authors showed that this policy has the best performance on homogeneous agents with a good criterion of what constitutes "good information". Moreover, they found that even in the case where agents are either purely malicious or have an opposite criterion of what constitutes high-quality information, the policy greatly reduces the spread of low quality information [32].

In situations in which quality is not easily quantified, other metrics—such as ratings, number of views, likes, number of downloads, etc.—can be used to enhance the exposure of certain content to people. In principle, such approaches would allow high-quality information to prevail. However, this popularity-based approach can create bias since the systems can be easily manipulated by social bots, for example [12]. Another disadvantage was highlighted by Sunstein [33,34] and Pariser [35]. The authors argued that the reliance on personalization and social media can lead people to being exposed to a narrow set of point of views [36] and one's existing beliefs would be reinforced because they are locked inside so-called filter bubbles or echo chambers, which prevent the users from engaging with ideas different from their own. Such selective exposure could facilitate confirmation bias [37] and possibly create a fertile ground for polarization and misinformed opinions [38,39]. Although several other works [40–43] have been done trying to address to the crucial importance for the problem of competition for attention, there is still a lack of a better understanding of how memes behave in on-line social networks. In this work, we investigate how the adoption of different messages sorting mechanisms on the users' news feed will affect the system's quality, diversity and discriminative power. We assume that each piece of information carries a numerical proxy representing its quality, interestingness or truthfulness. We anticipate that by sorting the memes, will increase the exposure of high-quality information, therefore, increasing the overall system's quality. However, it is still unknown how it will affect other characteristics of the systems such as diversity of information and discriminative power.

2 The Model and Numerical Results

The model consists of a network with N agents, each of them equipped with a finite memory (or attention) containing α memes. Each meme is equipped with a numerical value drawn from an uniform distribution representing its quality or truthfulness. Furthermore, new memes are continuously introduced into the system in an exogenous way. The rate at which this happens determines the amount of diversity in the system in the sense that the higher information load μ, the higher the diversity and as a consequence, the harsher the competition. We assume that at time $t = t_0$ the system is in a state of higher diversity where each node has α unique memes. At every time step a node i is selected at random and with probability μ it introduces a new meme into the system by adding it to its attention list and sharing it with all its neighbors. On the other hand, with probability $1 - \mu$ the selected node chooses a meme from its attention list and, then, it transmits it to all of its neighbors. In this scenario, the exposed agent will accept the meme only if it is not already in his/her attention list. Furthermore, the probability that an agent selects a specific meme m from its list to transmit is proportional to the meme's quality $f(m)$. For example, if node i has a set of memes $m(i)$, the probability of meme m_k being selected is

$$P_i(k) = \frac{f(m_k)}{\sum_{j=1}^{\alpha} f_i(m_j)}, \tag{1}$$

therefore, the higher the quality, the higher the chances of being transmitted. Once all neighbors receive the meme, the memes at the bottom of the nodes'

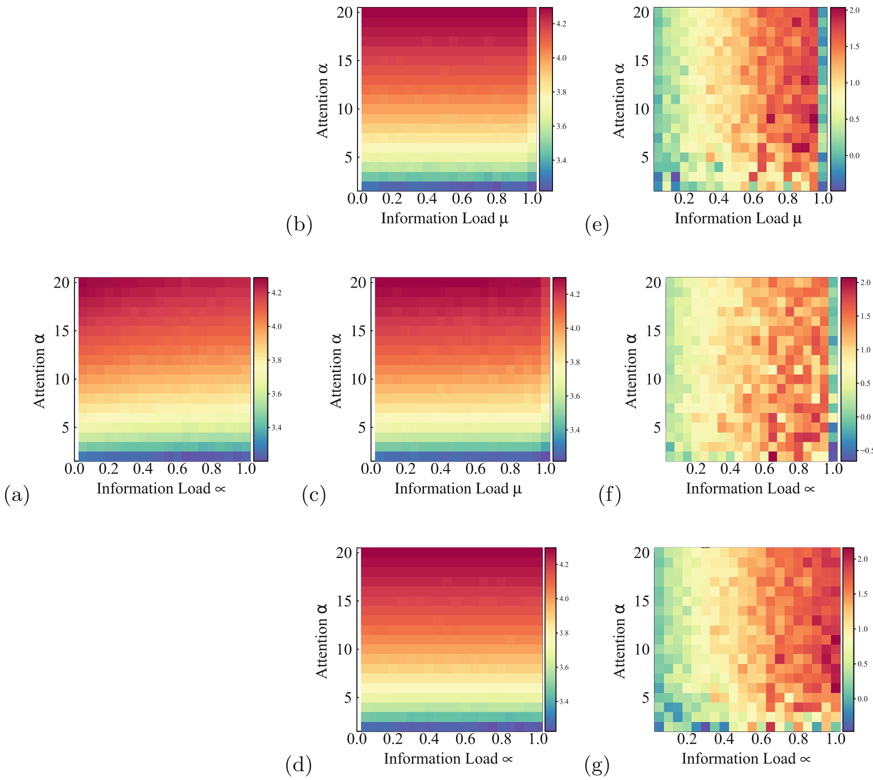

Fig. 2. The Diversity H (color scale bar) as a function of intensity of information load and attention for **a** the baseline model and the model with a sorted attention list according to **b** the meme's popularity, **c** node's influence and **d** meme's quality. **e**–**g** Difference in percentage between the baseline model and the models with different sorting mechanisms. The results indicate that sorting preserves the overall system's diversity.

attention lists are removed or forgotten to make space for the incoming message. Here, our goal is to understand how sorting the memes according to their quality, popularity and based on the source's influence (or degree) will affect the overall system's quality, diversity and discriminative power. We will compare these three scenarios with a baseline model where memes are organized in a first-in-first-out manner. The results were obtained from a scale-free network with $N = 1000$ and average degree $\overline{k} = 20$ for each of the scenarios considered. We run each simulation until the system reaches a steady state and once in such a state [44], we follow 4×10^6 memes for each combination of the control parameters considered from the moment they were first introduced into the model until they completely disappear from the network recording their quality, popularity, lifetime as well as some of the characteristics of the source of information. Here, we define popularity as the number of times a given meme is selected to be

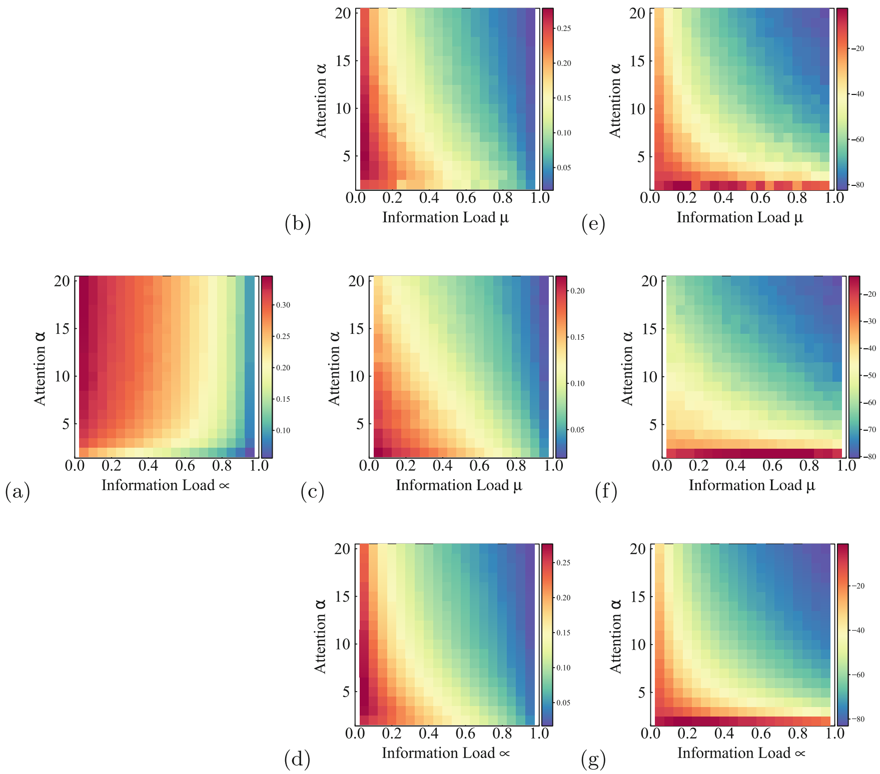

Fig. 3. The Kendall Tau (color scale bar) as a function of intensity of information load and attention for the **a** baseline model and the model with memes sorted according to **b** the meme's popularity, **c** node's influence and **d** meme's quality. **e**–**g** Difference in percentage between the baseline model and the models with sorted attention list. In this case, the introduction of sorting hinders the system's discriminative power with differences between models as high as 82.5%.

transmitted, lifetime is the time passed between the meme's creation and its extinction and the influence or degree of the node that generated the meme. In all situations, the selection depends purely on the meme's quality, however, in principle, sorting the user's attention list will enhance the exposure of high-quality information. In order to verify if our assumption is correct, Fig. 1 shows the behavior of the system's average quality as a function of time for different values of information load μ for (a) baseline model and the models with sorted attention list according to (b) meme's popularity (c) node's influence and (d) meme's quality. As one can see in Fig. 1 (a–d), at time $t = 0$, the average system's quality is 0.5 since all memes are drawn at random from a uniform distribution between 0 and 1. However, as the competition starts to take place, the average quality starts to increase in a way that highly depends on the information load μ and the sorting mechanism. The inset shows the dependence of the system's average quality $\overline{Q}$ on μ at the steady state. Observe that as $\mu \to 1$, $\overline{Q} \to 0.5$ which is equivalent to the case where there is no diffusion, only innovation with a new meme being introduced at every time step. The only exception is Fig. 1 (d). In such a case, the system is able to converge to an optimal state with only high quality information even when $\mu = 1$. Finally, Fig. 1 (e–g) shows the behavior of the attention α as a function of the information load μ. Color represents the difference in percent between the system's quality at the steady state of the baseline and the model with a sorted attention list. Observe that the overall system's quality improves significantly in all cases, reaching as high as 96% improvement for some combination of the control parameters. Confirming our hypotheses that by sorting the memes will enhance the exposure of high-quality information and as a consequence increase the overall system's quality.

Next, we investigate how the diversity of information and the system's discriminative power are affected by sorting mechanisms. To measure the amount of diversity in the system at the steady state, we start from the entropy $H = -\sum_m P(m) \log P(m)$ where $P(m)$ is the portion of attention received by meme m, i.e., the fraction of messages with m across all of the user feeds. The sum runs over all memes present at a given time and is averaged over a long period after stationarity has been achieved. Figure 2 (a–d) shows the behavior of the diversity (system's entropy) for (a) the baseline model and the models with sorted attention list according to the (b) meme's popularity (c) node's influence and (c) meme's quality for different values of α and μ. Observe that the information load does not significantly affect the system's diversity, however, as expected, it increases as the user's attention increases. Furthermore, Fig. 2 (e–g) shows the difference in percent between the baseline and the models with a sorted attention list. We observe that sorting does not affect the diversity of information in any significant way in any of these cases. On the other hand, as we will show next, it does considerably decrease the system's ability to distinguish between memes.

To measure the system's discriminative power, we employ the Kendall rank correlation [45] between popularity and quality, which is computed by ranking memes according to the two criteria and then counting the number of meme pairs for which the two rankings are concordant or discordant, properly accounting for ties. The extreme case $\tau = 1$ indicates a perfect correlation between quality and popularity and fitter memes are more likely to go viral. On the other hand, if $\tau = -1$, the two rankings are completely discordant. Figure 3(a-d) shows in color the Kendall correlation rank for the four models considered for different values of α and μ. We observed that in general the rank correlation decreases as the information load increases and a comparison between the models review that the introduction of the sorting mechanisms hinders the system's discriminative power with differences between models being as high as 82.5% as shown in Fig. 2(e-g).

3 Conclusions

We have considered an agent-based model to study the effects of message sorting on the diffusion of low and high-quality information in online social networks. We have considered three scenarios where memes are sorted according to their popularity, the influence of the node that posted the meme or the meme's quality. In order to understand the effects of such changes on the overall system's quality, diversity and discriminative power, we have compared these three situations with a baseline model where memes are organized in a first-in first-out manner. The results indicate that such approaches intensify the exposure of high-quality information, increasing the overall system's quality while preserving its diversity. On the other hand, a more significant change was observed when considering the system's discriminative power τ. In all cases, namely, the baseline model and the model with sorted attention lists, τ decreases as $\mu \to 1$, however when comparing them, sorting hinders the system's discriminative power with differences as high as 82.5% for some combination of the control parameters.

Acknowledgment. The research was supported partially by NSF-DMS Grant 1929298 and ARL through ARO Grant W911NF-16-1-0524. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

1. Dawkins, R.: The Selfish Gene, no. 199. Oxford University Press (2006)
2. Dunbar, R.: The social brain hypothesis. Brain **9**(10), 178–190 (1998)
3. Starbird, K., Maddock, J., Orand, M., Achterman, P., Mason, R.M.: Rumors, false flags, and digital vigilantes: misinformation on Twitter after the 2013 Boston marathon bombing. In: IConference 2014 Proceedings (2014)
4. Mendoza, M., Poblete, B., Castillo, C.: Twitter under crisis: can we trust what we RT? In: Proceedings of the First Workshop on Social Media Analytics, pp. 71–79. ACM (2010)
5. Gupta, A., Lamba, H., Kumaraguru, P., Joshi, A.: Faking Sandy: characterizing and identifying fake images on Twitter during hurricane Sandy. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 729–736. ACM (2013)
6. Domm, P.: False rumor of explosion at white house causes stocks to briefly plunge; AP confirms its Twitter feed was hacked. http://cnb.cx/2thGiq5, 2013. Online. Accessed 05-Oct-2020
7. Rosen, L.D., Carrier, L.M., Cheever, N.A.: Facebook and texting made me do it: media-induced task-switching while studying. Comput. Human Behav. **29**(3), 948–958 (2013)
8. Junco, R.: Too much face and not enough books: the relationship between multiple indices of Facebook use and academic performance. Comput. Human Behav. **28**(1), 187–198 (2012)

9. World Economic Forum: Digital Wildfires in a Hyperconnected World. http://bit.ly/2sY6djW, 2013. Online. Accessed 30-Jan-2019
10. Tambuscio, M., Oliveira, D.F., Ciampaglia, G.L., Ruffo, G.: Network segregation in a model of misinformation and fact-checking. J. Comput. Social Sci. **1**(2), 261–275 (2018)
11. Lazer, D.M., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., et al.: The science of fake news. Science **359**(6380), 1094–1096 (2018)
12. Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A.: The rise of social bots. Commun. ACM **59**(7), 96–104 (2016)
13. Newman, M.: Networks: An Introduction. Oxford University Press (2010)
14. Easley, D., Kleinberg, J.: Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press (2010)
15. Goffman, W., Newill, V.: Generalization of epidemic theory: an application to the transmission of ideas. Nature **204**(4955), 225–228 (1964)
16. Weng, L., Flammini, A., Vespignani, A., Menczer, F.: Competition among memes in a world with limited attention. Sci. Rep. **2** (2012)
17. Gleeson, J.P., Ward, J.A., O'Sullivan, K.P., Lee, W.T.: Competition-induced criticality in a model of meme popularity. Phys. Rev. Lett. **112**(4), 048701 (2014)
18. Notarmuzi, D., Castellano, C.: Analytical study of quality-biased competition dynamics for memes in social media. EPL (Europhys. Lett.) **122**(2), 28002 (2018)
19. Stringer, M.J., Sales-Pardo, M., Amaral, L.A.N.: Statistical validation of a global model for the distribution of the ultimate number of citations accrued by papers published in a scientific journal. J. Am. Soc. Inf. Sci. Technol. **61**(7), 1377–1385 (2010)
20. Wang, D., Song, C., Barabási, A.-L.: Quantifying long-term scientific impact. Science **342**(6154), 127–132 (2013)
21. Penner, O., Pan, R.K., Petersen, A.M., Kaski, K., Fortunato, S.: On the predictability of future impact in science. Sci. Rep. **3** (2013)
22. Lerman, K., Ghosh, R., Surachawala, T.: Social contagion: an empirical study of information spread on Digg and Twitter follower graphs. arXiv:1202.3162 (2012)
23. Crane, R., Sornette, D.: Robust dynamic classes revealed by measuring the response function of a social system. Proc. Nat. Acad. Sci. **105**(41), 15649–15653 (2008)
24. Leskovec, J., Adamic, L.A., Huberman, B.A.: The dynamics of viral marketing. ACM Trans. Web (TWEB) **1**(1), 5 (2007)
25. Aral, S., Walker, D.: Creating social contagion through viral product design: a randomized trial of peer influence in networks. Manage. Sci. **57**(9), 1623–1639 (2011)
26. Rogers, E.M.: Diffusion of Innovations. Simon and Schuster (2010)
27. Jamali, S.: Comment mining, popularity prediction, and social network analysis. PhD thesis, George Mason University (2010)
28. Qiu, X., Oliveira, D.F., Shirazi, A.S., Flammini, A., Menczer, F.: Limited individual attention and online virality of low-quality information. Nat. Human Behav. **1**(7), s41562-017 (2017)
29. Sreenivasan, S., Chan, K.S., Swami, A., Korniss, G., Szymanski, B.: Information cascades in feed-based networks of users with limited attention. IEEE Trans. Netw. Sci. Eng. (2017)
30. Ciampaglia, G.L., Nematzadeh, A., Menczer, F., Flammini, A.: How algorithmic popularity bias hinders or promotes quality. Sci. Rep. **8**(1), 15951 (2018)

31. Oliveira, D.F., Chan, K.S.: The effects of trust and influence on the spreading of low and high quality information. Phys. A Stat. Mech. Appl. **525**, 657–663 (2019)
32. Cisneros-Velarde, P., Oliveira, D.F., Chan, K.S.: Spread and control of misinformation with heterogeneous agents. In: International Workshop on Complex Networks, pp. 75–83 (2019)
33. Sunstein, C.R.: The law of group polarization. J. Polit. Philos. **10**(2), 175–195 (2002)
34. Sunstien, C.R.: Republic 2.0. Princeton University Press, Princeton (2009)
35. Pariser, E.: The filter bubble: How the new personalized web is changing what we read and how we think. Penguin (2011)
36. Nikolov, D., Oliveira, D.F., Flammini, A., Menczer, F.: Measuring online social bubbles. PeerJ Comput. Sci. **1**, e38 (2015)
37. Baron, J.: Thinking and Deciding. Cambridge University Press (2000)
38. Nyhan, B., Reifler, J.: When corrections fail: the persistence of political misperceptions. Polit. Behav. **32**(2), 303–330 (2010)
39. Stanovich, K.E., West, R.F., Toplak, M.E.: Myside bias, rational thinking, and intelligence. Curr. Directions Psychol. Sci. **22**(4), 259–264 (2013)
40. Ratkiewicz, J., Fortunato, S., Flammini, A., Menczer, F., Vespignani, A.: Characterizing and modeling the dynamics of online popularity. Phys. Rev. letters **105**(15), 158701 (2010)
41. Lerman, K., Ghosh, R.: Information contagion: an empirical study of the spread of news on Digg and Twitter social networks. ICWSM-Int. Conf. Weblogs Social Media **10**, 90–97 (2010)
42. González-Bailón, S., Borge-Holthoefer, J., Rivero, A., Moreno, Y.: The dynamics of protest recruitment through an online network. Sci. Rep. **1**, 197 (2011)
43. Baños, R.A., Borge-Holthoefer, J., Moreno, Y.: The role of hidden influentials in the diffusion of online information cascades. EPJ Data Sci. **2**(1), 6 (2013)
44. Oliveira, D.F., Chan, K.S., Leonel, E.D.: Scaling invariance in a social network with limited attention and innovation. Phys. Lett. A **382**(47), 3376–3380 (2018)
45. Kendall, M.G.: A new measure of rank correlation. Biometrika **30**(1/2), 81–93 (1938)