Topic Selectivity and Adaptivity Promote Spreading of Short Messages

Patryk Bojarski
Warsaw University of Technology
Krzysztof Suchecki
Warsaw University of Technology
Janusz Hołyst (janusz.holyst@pw.edu.pl)
Warsaw University of Technology

Research Article

Keywords:

Posted Date: January 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1232620/v1

Topic selectivity and adaptivity promote spreading of short messages

Patryk A. Bojarski[1], Krzysztof Suchecki[1], and Janusz A. Hołyst[1,2,*]

[1]Center of Excellence for Complex Systems Research, Faculty of Physics, Warsaw University of Technology, Koszykowa 75, 00-662, Warsaw, Poland
[2]ITMO University, Kronverkskiy Prospekt 49, St. Petersburg, 197101, Russia
[*]janusz.holyst@pw.edu.pl

ABSTRACT

We investigate how the spreading of messages in social networks is impacted by user selectivity for messages based on their content, competition for user's attention between different messages and message content adaptivity through user-introduced changes. An agent-based model of message spreading featuring these mechanisms is introduced and explored to asses the impact of each on statistical properties of resulting message cascades, in particular cascade size and mean length of the messages. We observe that selectivity changes the popularity distribution and is responsible for preference towards short or long messages – long messages for low and short messages for high selectivity. The competition drastically limits message spread if selectivity is low, but has only small impact for high selectivity. Message content adaptivity has small effect of increasing popularity of already popular messages, but only for medium levels of selectivity. We explore the spread process on random and scale-free synthetic networks and conclude that the existence of hubs in the network seems to have a marginal impact on cascade size distribution. We relate the distribution of mutated message variant popularity to observations in real social media.

1 Introduction

The spread of information through social networks is one of main modes of information transmission in modern society. Traditional media channels, such as news agencies, make efforts to report information thoroughly and accurately (or at least they should act in such a way), and its dissemination does not depend on end consumer reception directly. The casual character of communication between acquaintances in social media, however, implies that information passed on is chosen according to an individual's tastes and may be often incomplete, biased, or even deformed to mean something else than it originally did. In other words, individuals choose what they spread and may change the original message and, as a result, its topic is not fixed during the spread. The existence of confirmation bias is well known[1] and it has also been directly shown that it causes individuals not only to choose information sources that confirm what they know, but also selectively peruse stories and messages from the sources they already subscribe to[2,3]. Both selective news consumption and the choice of information sources may lead to creation of so-called echo chambers[3–6], where like-minded individuals form tight-knit communities that may consume almost no information from outside of their bubble, either by choice or ignorance. In recent times, due to the existence of automated recommendation systems, the selectivity of information consumption may not be completely a result of individual choice or preference[7–10]. Despite recommendation systems playing increasingly significant role in ecology of message processing, in this work we disregard their impact and focus on user behavior only. Several cases of message spreading with changing messages have been studied, such as changes in chain-letters[11], stories propagated through blogs[12] or online social networking services such as Facebook[13]. It has also been shown experimentally that the meaning of messages can change even if individuals are only trying to change its form, such as message shortening[14]. This behavior poses multiple problems for research on the subject of information spread and possible applications, most notably the fact that information spread is hard to track in absence of metadata such as URL links[15] or hashtags[13] that can be used as immutable markers. There was research on tracking changing information, such as specific chain letters[16] or meme-like phrases by phrase inclusion in another[12]. The online spreading of information itself has been extensively studied and many features of the real spreading process have been identified[17] and modeled[18,19], although adapted epidemic models have been used to represent this process[20]. Such models should be applied with care however, as, unlike new virus strains that partially bypass immunity developed for previous variant, users are likely to recognize the same information even if phrased differently or partially changed in meaning. Studies on propagation of changing information in social networks often show dynamics with similar characteristics to genetic mutations[11,21]. In fact, meme variant popularity shows characteristics of the Yule model designed to represent evolving populations[13]. However, unlike mutations in living organisms that are the result of random DNA transcription errors, changes in messages online are

intentional. It is usually quick and trivial to pass on a verbatim message and harder to do so when adding changes. It follows that, since humans have a tendency to be selective about what they read, they would also have the same tendency when sharing information, preferring to share messages that they believe in. This can motivate individuals to adapt or re-phrase messages, so that they more closely represent their beliefs. Regardless of user intent, messages are under evolutionary pressures that may promote certain characteristics. There have been studies investigating how different message features impact their spread — in particular, the communication on the Twitter platform[22,23]. Of specific interest here is the impact of message length, which has been found to be slightly positive[22] or inconclusively negative[23] depending on the particular study.

In this work, we explore a model of information propagation on social networks that includes selective sharing, competition between messages for user's attention, and intentional modification of the message content by the users (topic adaptivity). Our primary assumption is that online users want to propagate information they like or agree with, such as a conspiracy theory believer propagating information supporting the conspiracy he believes in, a scientist propagating only scientifically plausible content, or a politician tweeting only information supporting her agenda. If confronted with messages that users do not completely agree with but still think have some merit, they may decide to adjust the information, removing parts of message that they don't agree with, modify them, or even add something from themselves. Such a modification may be motivated by any of a need to shorten it, share their own thought on the matter, or push their own agenda. The key point here is that different users have different opinions or preferences. Users will forward information similar to their beliefs, while ignoring and effectively stifling information that they do not agree with. The spreading process may therefore bear some resemblance to opinion dynamics, as messages received and read may influence users' opinions[24]. Here, we opt to look at a short-term process, where we assume opinions and beliefs are constant. We do not consider the veracity of opinions or messages and represent both opinions and message contents by abstract vectors in multidimensional space.

The main question we try to answer is what is the impact of both user selectivity and content adaptivity on the outcome of the process, such as the popularity of messages or their lengths. Using numerical agent-based simulations we find that the most pronounced effect is the increased popularity of long messages at low selectivity and of short messages at high selectivity. The selectivity of individuals influences the characteristics of the message popularity distribution. For low selectivity, the competition for attention dominates and results in a power-law distribution, with more popular messages being on average longer. As selectivity increases, the trend reverses, with popular messages being the shortest possible. The popularity distribution features a peak at high popularities, for which short messages are responsible. The effect of message adaptivity is limited to medium values of selectivity, where it allows messages to spread even further, and high values, where it helps some messages to attain medium popularity rather than low. The competition for attention has fundamental influence on message spreading in the case of low selectivity, but it disappears if message spreading is highly selective. We find the effect of a scale-free connection topology to be minimal for the considered spread process.

2 Results

To investigate how selectivity, competition and message adaptivity can influence the spreading of messages, we have created an agent-based model and investigated its behavior for a set of parameter values. We assume agents, representing users online, possess their own personal opinions about a wide range of topics. For simplicity, an opinion of the agent i on the topic j is $x_{ij} \in \{-1,0,1\}$. Agents either post new messages, with probability of η, that reflect a few of their opinions, or read new messages posted recently by neighbors. The content of message covers only a few topics, each with an associated opinion $y_j \in \{-1,0,1\}$, with message *length* being the number of topics it contains. When agents read a message that is close enough to their own beliefs as defined by a selectivity parameter $\tau \in [-1,1]$, they share the message with their own neighbors, adapting it to fit their own opinions even more with a probability α. When a message is modified, we still consider it the same message, but a new *variant* of it. See the *Methods* section for the full description of our model.

We have performed a series of numerical simulations for our agent-based model and have tracked the popularity of messages, which is the total number of agents that shared it, and the length of its variants. In particular, the distribution of message popularity has been investigated – the fraction of the messages that attained a given total number of shares. The simulations have used synthetic Erdős–Rényi graphs (ER) and Barabási–Albert networks (BA) as connection topologies between users.

First, we checked the propagation of information on the ER graphs with the size $N = 600$ and an average degree $\langle k \rangle = 6$. We assumed $\eta = 0.1$ as the probability of creating a new message. The simulations have been performed for two different adaptivity probabilities $\alpha = 0.0$ (messages cannot be modified) and $\alpha = 0.2$ (agents can modify messages). For each value of the selectivity parameter τ, we made 10 independent realizations.

Behavior of the model was investigated for three selectivity parameters $\tau \in \{-0.4, 0.2, 0.8\}$. Note that since the cosine between vectors with possible positive and negative elements is between -1 and 1, the thresholds are also in this range. Main results of our work are shown in Fig. 1.

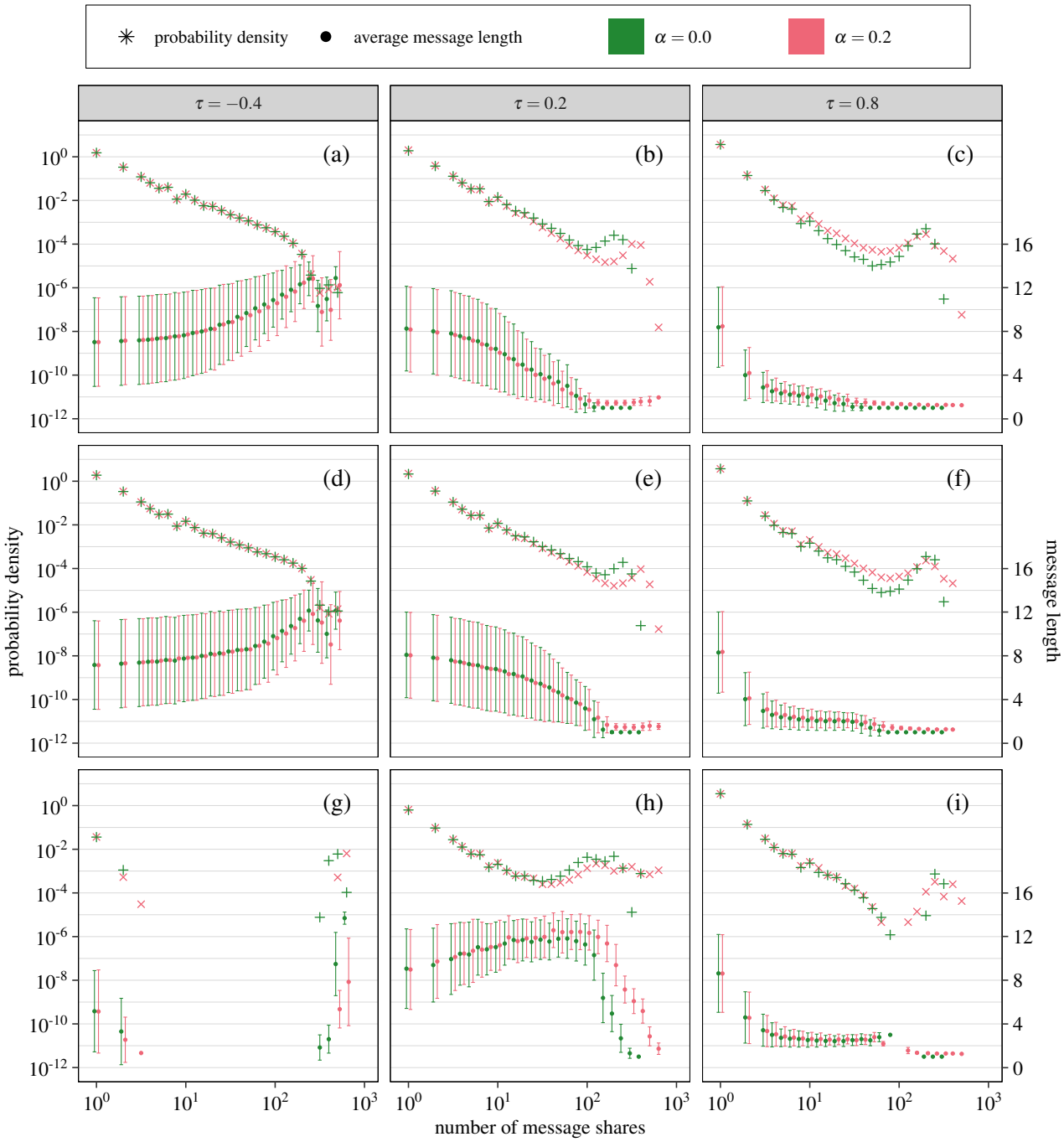


Figure 1. The influence of user selectivity τ and message adaptivity α on the popularity and average length of messages propagating in synthetic networks. Panels a-f show spreading with competition between messages (agents can share only the most recent message he likes), while panels g-i show spreading without this constraint. Panels a-c show popularity for ER networks and 500 000 time steps, d-f for BA networks with 500 000 time steps and g-h for ER network, but with a shorter time horizon – 50 000 time steps. The simulations were carried out for networks with size $N = 600$ and an average node degree $\langle k \rangle = 6$. The agents' opinion vectors were randomly selected. The probability of creating a new message equals $\eta = 0.1$. The probability density of the number of shared messages (left vertical axis) is marked with crosses and pluses. The average length of the message (right vertical axis) with a given popularity is marked with dots, and its standard deviation with error bars.

With a low value of the selectivity parameter, $\tau = -0.4$, the majority of messages will be passed on, and the main limiting factor to their spread is competition for user's attention (i.e., which of the message has arrived as the last one). The distribution of popularity is mostly a power-law with a cut-off for high popularity values (Fig. 1a). The average length of a message increases with its popularity, which is in opposition to the behavior for higher selectivity values (Fig. 1b,c). This happens because the longer the message is, the closer to zero the cosine similarity between message content and the agent's opinion vector will be if the agent's interests are essentially random (Central Limit Theorem). If the selectivity parameter is below zero, that means that longer messages have a greater chance to not fall below the selectivity threshold and to be shared. As result, there is competition over not being exceptionally disagreeable, that is won by messages long enough that nobody can truly completely disagree with them. For low selectivity, we find no significant impact of information adaptivity (red and green symbols are very close one to another). The reason behind this is that users share almost any message they see, meaning the deciding factor is how fresh the message is and not its content. Changed information does not gain much of an advantage in the dynamics. It is still enough to increase the mean length of the popular messages, but not enough to alter popularity distribution in a significant way.

After increasing the selectivity parameter to $\tau = 0.2$, the popularity distribution changes somewhat, with an added peak at high popularity values (Fig. 1b). This essentially means that aside from the spectrum of "regular" messages, a population of very popular viral messages appear that did not exist when selectivity was low. This population consists of very short messages that have a better chance to be attractive for some users. The trend of message length visible for low selectivity reverses, and short messages have the advantage (compared to lower selectivity Fig. 1a). It is now a competition for being agreeable, which is won this time by short messages that have a relatively good chance to fit a user's view perfectly. In fact, the peak consists almost solely of messages of the shortest possible length: 1. For medium selectivity, the effects of content adaptivity become visible. The ability to change message content is conducive to shortening the message and thus improving its popularity, shifting the peak to somewhat higher popularity values. On the other hand, the short, viral information may mutate producing some longer variants that disappear quickly, but still bump the average length of the most popular messages.

When the selectivity parameter is even greater, close to one ($\tau = 0.8$), the spreading is driven mostly by selection. The advantage of short messages is more pronounced, with a much clearer peak. (Fig. 1c). Again, it is the short messages that dominate there, but while changing a message's content may bridge the gap between popular and unpopular a bit, it does not affect the top. This is mainly due to the fact that only the shortest messages ever get a chance, and they either fit a user's beliefs ideally, or do not, in which case they are never read and do not get any chance to evolve. Aside from investigating the dynamics on a random ER graphs, we wanted to see if the dynamics will differ for the scale-free BA networks, where hubs (nodes with very high degree) exist. The investigated BA network has the same basic network measures (size N, density $\langle k \rangle$ as the ER network. We observe (Fig. 1a-c vs Fig. 1d-f) no significant statistical differences for messages spreading on both these networks, aside from the fact that the influence of message adaptivity for the medium selectivity parameter ($\tau = 0.2$) was a bit smaller for the BA network than for the ER graph.

What, at first, may have seemed counter-intuitive is that when we increase the selectivity parameter τ from -0.4 to 0.2, some messages reach more agents than when the selectivity was lower. We considered this to be the influence of the competition for user attention, which we verified. For the same network and dynamics parameters as before, we performed simulations of the model variant without the competition, where agents can share all messages that fulfill the selectivity constraint. (Fig. 1g-i). In the situation where the selectivity parameter is $\tau = -0.4$, a large fraction of messages spread over almost the entire network (Fig. 1g). This is in contrast to the previous case, where most messages did not spread widely. Therefore, competition plays a significant role in message spread for this value of τ.

When increasing the selectivity parameter to $\tau = 0.2$, the importance of competition also decreases (Fig. 1h). Nevertheless, it still plays a big role in spreading the message – more information reaches each agent than with competition. Interestingly, in the absence of competition, there is slight increasing trend of average message length. This may be because all medium-popularity short messages become more popular, leaving only long messages in the mid-popularity range.

When selectivity reaches a value close to one, $\tau = 0.8$, the role of competition becomes even smaller (Fig. 1i). The distribution of message popularity without competition is similar to the distribution with competition, except for a more pronounced peak, practically separated from the rest of the distribution. Since users are so picky that very few of the messages they see each time step count as interesting for them, it becomes more common for users to have only one or even no new shared messages, thus making competition much less common and thus less influential on overall behavior. The average message length is very slightly longer than in dynamics with competition. Nevertheless, when messages are viral, this number is still close to one topic only. It can be clearly seen that the peak is exclusively (without adaptivity) and almost exclusively (with adaptivity) a result of messages of length 1, that from the epidemic point of view would have the spreading rate over 1 and thus encompass the whole network, unlike longer messages that can form only local cascades.

Along with the distribution of message popularity, we have also investigated the distribution of relative popularity of message variants. In particular, this distribution has been calculated for a single message, in order to see this distribution for

single realization, as one would see distribution of popularity of variants for real messages. Similarly to the work of Adamic et al.[13], we chose only the message with the most variants across all model simulations. This is done intentionally, so as to make comparisons between our results and the results of Adamic et al.[13], which feature real cascades. As seen in Fig. 2, the distribution appears to be somewhat similar to a power-law, meaning that there usually is just one or two main dominant variants, along with a plethora of niche adaptations.

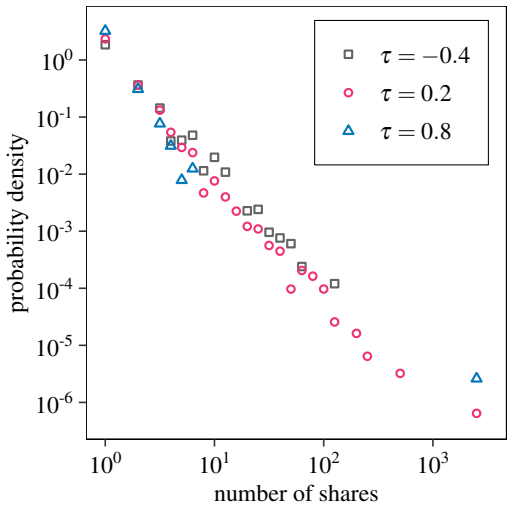


Figure 2. Probability distribution of the number of shares of message variants for different selectivity parameters τ. Simulations were conducted on ER networks with $N = 10\,000$, $\langle k \rangle = 6$, time steps: 500 000. Please note that the number of agents is much larger than in previous plots.

3 Discussion

We have investigated the impact of selectivity of message sharing, competition for user's attention, and message adaptivity in a model of information spread on artificial network topologies. We have observed that a high selectivity promotes short messages and, in the presence of competition for user's time, is also partially responsible for the emergence of popular, viral messages.

Using numerical simulations, we find that the total cascade size distributions change significantly depending on how selective users are, while the existence of message adaptivity has a smaller impact. For a low selectivity, the distribution is largely a power-law, with an additional peak of highly popular messages appearing as the selectivity grows stronger. The content adaptivity has the largest impact for medium selectivity, when it significantly increases the popularity of the most popular messages and lesser for high selectivity, where it boosts some messages from low to medium popularity. An interesting observation is that either short or long messages are more prevalent depending on selectivity. If users reject only messages clearly in opposition to their beliefs, as is the case for low selectivity, then longer messages are more popular. On the other hand, simple messages dominate in the presence of strict selection. The length here means how many different opinions or beliefs the messages are relevant to. A short message, for example, conveys only one fact that can be in agreement, disagreement, or neutral for an individual's belief, such as a message stating that moon landing was fake, without any other content that could be perceived true or false according to individual beliefs. A long message, on the other hand, relates to many facts or beliefs, such as political opinion, conspiracy theories and scientific facts all in one message, so that different people may agree or disagree on different parts of that message.

The above conclusions assume that there is competition for a user's attention between messages, so that users cannot read all messages they see and only pay attention to the newest ones, disregarding older ones even if they may find them interesting. If we relax this assumption and make all messages effectively spread independently, the behavior changes significantly. For a low selectivity, most messages end up viral, which could be expected since the spreading rate is usually above the epidemic threshold. For a higher selectivity, however, the situation changes, as the limited acceptance by neighbors means that users end up with much fewer messages in their queue, up to the point where competition becomes irrelevant when people are very selective.

Interestingly, the spread of information on ER random graphs and BA networks looks very similar in the investigated model. There are slight differences between them, but they are not very noticeable. We expected that hubs in BA networks, coupled with the possibility of changing the content of messages, would allow information to spread more broadly. Perhaps this was not

visible because the agents' opinions were purely random. However, real-life experience has shown that people tend to keep in touch with like-minded people. Perhaps in such a situation the hubs would play a more significant role.

The preference for short messages for positive selectivity may be a potential explanation for the Twitter phenomenon. Artificial limits on message length may have contributed to its popularity as it may have become preferred over longer internet forum or blog posts because it facilitated broader information spread. The research of factors contributing to tweet spreading may not seem to support this claim, showing small positive influence[22] or statistically insignificant negative[23] of length on popularity. It must be noted however, that what we mean by "short" messages here is number of topics and related opinions expressed in the message. With Tweets being very short, many of them may express just a single opinion and it may be not directly related to message length counted in characters. In addition, while our model features uncorrelated, static opinions of agents, interests of interacting users, such as followers on Twitter, do change in time to be more uniform[25]. This difference may alter the overall behavior, especially for the small cascades that are most frequent in real systems and may change the perceived influence of factors such as length.

Our results are in agreement with other findings regarding the spreading of messages with changing content. In particular, according to Adamic et al.[13], the memes on Facebook behave according to a Yule model that describes evolving populations. The popularity distribution of variants of a message (Fig. 2) is qualitatively the same as distributions found by Adamic et al.[13], regardless of the assumed user selectivity. Our finding that the most popular messages are short is in agreement with another conclusion of Adamic et al.[13], where the maximum of the popularity occurs for mean lengths significantly below the average (unlike in Twitter research[22,23] mentioned above). While this is not direct comparison, it still shows that the model shows behavior similar to real social network communication.

Our work lays a foundation for possible further research. The first avenue is to make the spreading model richer, featuring correlated user opinions, asymmetric opinion distributions, or the inclusion of heterogeneity in agents' selectivity or adaptivity. The second avenue is testing the underlying principles of the spreading model in real social networks and adapting the model to better reflect real social information processing. The third potential research avenue is exploiting the obtained results, for example by including a message recommendation system to the model and testing whether its influence could be detected from message spread statistics, similar to how it was possible to do in internet forum discussions[26]. If so, it may be possible to develop a method to detect the presence and assess the influence of recommendation systems in real online social networking services.

4 Methods

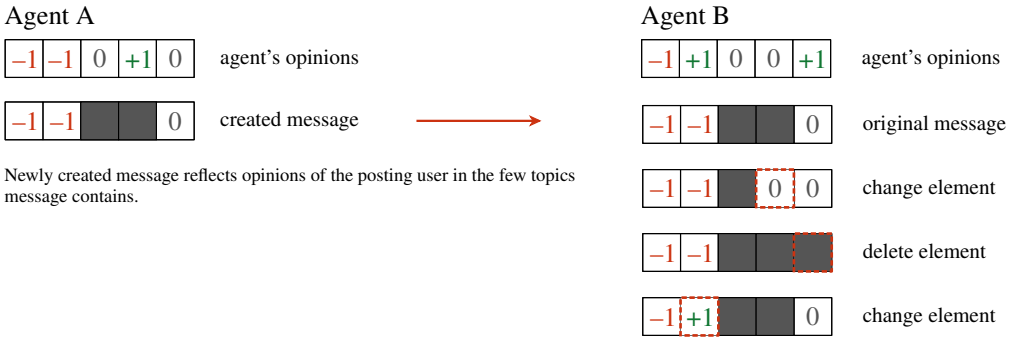
4.1 Model

We assume every agent (user) i is connected to k_i other agents, creating a network consisting of N users, that represents contacts in the online service along which messages can spread. Each agent has its own opinions on D different, independent topics, represented as a vector $\vec{x}$ of length D, which we call an opinion vector. Each element has a value $x_{ij} \in \{-1,0,1\}$. These values represent the agent's attitude or opinion towards a given subject – positive ($+1$), neutral (0), or negative (-1). Each element of the vector corresponds to the same topic for all agents. For example, the third element of any agent's vector opinion may represent opinion about global warming – whether they believe and care about it ($x_{i3} = 1$), deny it ($x_{i3} = -1$) or simply don't care ($x_{i3} = 0$).

We use the asynchronous dynamical rule, picking agent that will act at random. With a probability η, an agent will create a new *message* and *share it* with his closest neighbors. Each message made in this way gets a unique ID that does not change during the dynamics. We assume that a given message will concern a certain issue (an event, politician's speech, scientific discovery, etc.) and while its later variants may contain different opinions or paint facts in a different manner, the message will be essentially about the same issue. The content of every new message reflects a part of the author's opinion. We randomly pick $d \in [1, 0.15D]$ topics and assign values to them from the opinion vector of the selected agent (as shown on Fig. 3a). Thus, the message contains values $y_j \in \{-1,0,1\}$ for a few of the topics and contains no mention about other topics at all. The number of topics included in the message is the message's *length*.

With a probability $1-\eta$, instead of creating a new message, the user examines messages shared by his closest neighbors. It does so in order from the newest to the oldest received. Users do not know when original information was created and only see the time when a message was shared with them. Then the cosine similarity between the agent's opinion vector and the first message is calculated to consider the *selectivity criterion*, taking into account only the components of the opinion vector that are also present in the message itself, which is why 0 in a message is not the same as the absence of a topic. In the case of a vector having cartesian length 0, we assume cosine similarity equal to 0. If the similarity exceeds a certain threshold τ (the dynamics parameter called selectivity), the agent decides he likes it and will share this message with his own neighbors. If, on the other hand, the similarity is equal or less than τ, the user reads increasingly older messages he received. He does this until he finds the first attractive information. If after looking through the whole set of messages received there are none to its liking,

the agent simply does not share anything. When the euclidean length of either an agent opinion vector or a message is equal to zero, then we assume that the cosine similarity is equal to zero.



(a) Newly created message reflects opinions of the posting user in the few topics the message contains.

(b) When sharing a message user may adapt it by adjusting one of opinons, deleting a topic or adding a topic.

Figure 3. The creation of new messages and possible adaptation of messages when sharing them.

Before sharing a message, an agent can modify it and create a new *message variant*. With a probability of α, it can add a new random topic to the message, remove one, or replace an opinion on one topic present in the message with its own (as shown on Fig. 3b). If it decides to modify the message, it chooses each of these actions at random with the equal probability of $1/3$. If it adds a new topic, the value on the topic is the same as the agent's opinion. If it deletes one subject, it is always the one that it disagrees with. And if it changes one topic, it is replacing the disagreeable part with its own opinion.

Users never share the same message (with the same ID) more than once, even if the message has been modified, since it is assumed that it will relate to the same event, even if it paints it in different colors. Users completely ignore any incoming messages with an ID they have already read and considered before. The dynamics ends after a given number of time steps. The dynamics can be represented by the diagram as in Fig. 4.

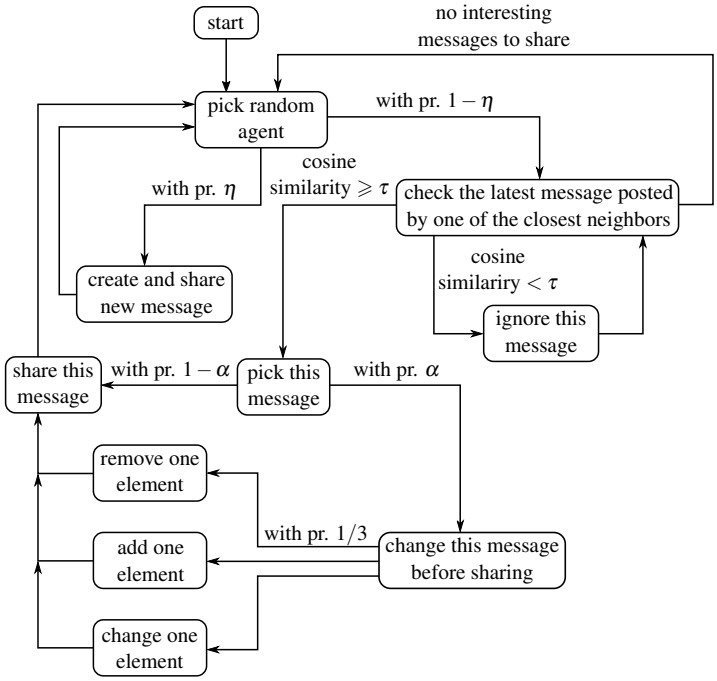


Figure 4. Flowchart of the algorithm for creation, sharing and modification of messages in our model.

At each time step, the agent can only create or share (possibly modifying) one message, to represent a limited capacity to consume and process information. We also introduced a variant of the model, allowing each agent to examine all received messages in one time step. They can share (with or without modifications) all messages that fulfill the selectivity criterion. We

call this variant a model without competition, as messages no longer need to compete for user's attention and effectively spread independently from each other.

4.2 Measures

We use the distribution of message popularity and average message length measures (Fig 1). The *popularity distribution* is the distribution of number of times S_m a message m has been shared throughout the whole simulation, including all variants of the message. The distribution shows the probability density of the number of shares being a certain value S. *Average length of message* is an unweighted average of the individual message lengths over all messages within a given popularity bin. $\langle L \rangle(S) = \sum_{m:S_m=S} L_m / N_S$, where S is number of shares (argument), m are different messages, L_m is length of individual message, S_m is number of shares for message m and N_S is total number of messages that have number of shares S. Individual message length is an average number of topics over all its variants, weighted by number of times given variant was shared. $L_m = \sum_v l_v s_v / \sum_v s_v$ where v are variants of message m, l_v is length of particular variant and s_v is number of shares of given variant. The error bars on Fig. 1 represent the standard deviation of the distribution of message lengths L_m, and do not include variance of lengths between variants of each message, so it does not depend on distribution of l_v for any given message (other than through L_m itself).

References

1. Nickerson, R. Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* **2**, 175–220, DOI: 10.1037/1089-2680.2.2.175 (1998).
2. Newman, T., Nisbet, E. & Nisbet, M. Climate change, cultural cognition, and media effects: Worldviews drive news selectivity, biased processing, and polarized attitudes. *Public Underst. Sci.* **27**, 985–1002, DOI: 10.1177/0963662518801170 (2018). Cited By 34.
3. Cinelli, M. *et al.* Selective exposure shapes the facebook news diet. *PLoS ONE* **15**, DOI: 10.1371/journal.pone.0229129 (2020).
4. Terren, L. & Borge, R. Echo chambers on social media: A systematic review of the literature. *Rev. Commun. Res.* **9**, 1–39, DOI: 10.12840/ISSN.2255-4165.028 (2021).
5. Cinelli, M., de Francisci Morales, G., Galeazzi, A., Quattrociocchi, W. & Starnini, M. The echo chamber effect on social media. *Proc. Natl. Acad. Sci. United States Am.* **118**, DOI: 10.1073/pnas.2023301118 (2021).
6. Choi, D., Chun, S., Oh, H., Han, J. & Kwon, T. T. Rumor Propagation is Amplified by Echo Chambers in Social Media. *Sci. Reports* **10**, 310, DOI: 10.1038/s41598-019-57272-3 (2020).
7. Pariser, E. *The Filter Bubble: What the Internet Is Hiding from You* (Penguin Group , The, 2011).
8. Bakshy, E., Messing, S. & Adamic, L. Exposure to ideologically diverse news and opinion on facebook. *Science* **348**, 1130–1132, DOI: 10.1126/science.aaa1160 (2015).
9. Anspach, N. M. The new personal influence: How our facebook friends influence the news we read. *Polit. Commun.* **34**, 590–606 (2017).
10. Möller, J., Trilling, D., Helberger, N. & van Es, B. Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Commun. & Soc.* **21**, 959–977 (2018).
11. Bennett, C., Li, M. & Ma, B. Chain letters & evolutionary histories. *Sci. Am.* **288**, 76–81, DOI: 10.1038/scientificamerican0603-76 (2003).
12. Leskovec, J., Backstrom, L. & Kleinberg, J. Meme-tracking and the dynamics of the news cycle. *Proc. ACM SIGKDD Int. Conf. on Knowl. Discov. Data Min.* 497–505, DOI: 10.1145/1557019.1557077 (2009).
13. Adamic, L., Lento, T., Adar, E. & Ng, P. Information evolution in social networks. *WSDM 2016 - Proc. 9th ACM Int. Conf. on Web Search Data Min.* 473–482, DOI: 10.1145/2835776.2835827 (2016).
14. Ribeiro, M., Gligorić, K. & West, R. Message distortion in information cascades. *The Web Conf. 2019 - Proc. World Wide Web Conf. WWW 2019* 681–692, DOI: 10.1145/3308558.3313531 (2019).
15. Adar, E. & Adamic, L. Tracking information epidemics in blogspace. *Proc. - 2005 IEEE/WIC/ACM Int. on Web Intell. WI 2005* **2005**, 207–214, DOI: 10.1109/WI.2005.151 (2005).
16. Liben-Nowell, D. & Kleinberg, J. Tracing information flow on a global scale using internet chain-letter data. *Proc. Natl. Acad. Sci. United States Am.* **105**, 4633–4638, DOI: 10.1073/pnas.0708471105 (2008).

17. Kim, M., Paini, D. & Jurdak, R. Real-world diffusion dynamics based on point process approaches: a review. *Artif. Intell. Rev.* **53**, 321–350, DOI: 10.1007/s10462-018-9656-9 (2020).

18. Guille, A. Information diffusion in online social networks. *Proc. ACM SIGMOD Int. Conf. on Manag. Data* 31–35, DOI: 10.1145/2483574.2483575 (2013).

19. Huang, J. *et al.* Temporal scaling in information propagation. *Sci. Reports* **4**, 5334, DOI: 10.1038/srep05334 (2015).

20. Girvan, M., Callaway, D., Newman, M. & Strogatz, S. Simple model of epidemics with pathogen mutation. *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* **65**, DOI: 10.1103/PhysRevE.65.031915 (2002).

21. Cheng, M., Li, Y., Nazarian, S. & Bogdan, P. From rumor to genetic mutation detection with explanations: a GAN approach. *Sci. Reports* **11**, 5861, DOI: 10.1038/s41598-021-84993-1 (2021).

22. Lahuerta-Otero, E., Cordero-Gutiérrez, R. & De la Prieta-Pintado, F. Retweet or like? that is the question. *Online Inf. Rev.* **42**, 562–578, DOI: 10.1108/OIR-04-2017-0135 (2018).

23. Han, X., Gu, X. & Peng, S. Analysis of tweet form's effect on users' engagement on twitter. *Cogent Bus. Manag.* **6**, 1–15, DOI: 10.1080/23311975.2018.1564168 (2019).

24. Sîrbu, A., Loreto, V., Servedio, V. & Tria, F. Opinion dynamics: Models, extensions and external effects. *Underst. Complex Syst.* 363–401, DOI: 10.1007/978-3-319-25658-0_17 (2017).

25. Myers, S. & Leskovec, J. The bursty dynamics of the twitter information network. *WWW 2014 - Proc. 23rd Int. Conf. on World Wide Web* 913–923, DOI: 10.1145/2566486.2568043 (2014).

26. Kujawski, B., Hołyst, J. & Rodgers, G. Growing trees in internet news groups and forums. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **76**, DOI: 10.1103/PhysRevE.76.036103 (2007).

Code availability

The code for this project is available on GitHub at:
`https://github.com/PatBoj/mutated-information-spread`.

Acknowledgements

We are thankful to Dr. Curtis Atkisson for useful comments to this article. The project was funded by POB Research Centre Cybersecurity and Data Science of Warsaw University of Technology within the Excellence Initiative Program - Research University (ID-UB). J.A.H. was partially supported by the Russian Science Foundation, Agreement No 17-71-30029 with co-financing of Bank Saint Petersburg, Russia.

Author contributions

P.A.B performed simulations, created plots and participated in interpreting results, K.S. has participated in interpreting results, J.A.H. forming conclusions. All authors have participated in defining the model and reviewed the manuscript.

Competing interests

The authors declare no competing interests.