How to Make the Information Dissemination Process More Effective: Studies based on Simulation Methods

Elena Strekalovskaya
HSE University
Perm, Russia
e-mail: e.strekalovskaja@yandex.ru

Ekaterina Syrvacheva
HSE University
Perm, Russia
e-mail: ket_prm@mail.rum

Elena Zamyatina
HSE University
Perm, Russia
e-mail: e_zamyatina@mail.ru

***Abstract*—The article deals with issues related to the definition of parameters that affect the process of information dissemination in social media. Directed or bidirectional links are considered as parameters. The study of the propagation process is carried out using simulation methods. AnyLogic was chosen as a tool for conducting simulation experiments.**

***Keywords*—Social media, directed and not directed links, the degree of influence.**

I. Introduction

Social media are web platforms and applications that allow to interact with the content of various types (text, photos, video and audio) and provide the user with the opportunity to interact with other users, rate and comment on content, and distribute it through likes, reposts and messages. In Russia, among the popular social media, the most popular is Vkontakte (VK). VK is a platform on which 9 out of 10 Russian users are registered. Another popular platform is the Odnoklassniki platform (7 out of 10 users). VK and Odnoklassniki allow users not only to communicate with each other, but also to comment on posts, organize events, etc., i.e. they are a channel for the dissemination of information, including the distribution of advertising products, services, materials for training, etc. [1]. The fast and reliable distribution of advertising is, without a doubt, beneficial for businesses. There is another side of this process - the dissemination of unwanted information. One way or another, the study of the process of dissemination of information, the parameters that influenced this process, or, conversely, the containment of information is relevant and important [2].

Simulation modeling methods were chosen for studying and conducting research, and the AnyLogic[3] simulation modeling system was used as a tool. Unlike the methods used in SNA (Social Network Analyzes), which mainly study the structural characteristics of social media, simulation methods allow us to consider the process in dynamics, to identify the causes that can cause certain events. There are various approaches to conducting simulation experiments. The object for these experiments can be either real or virtual social networks [4,5]. This article deals with virtual social networks, i.e., models of real social networks, and more precisely, the Barabasi-Alberta model [6,7,8].

Further, the article will be structured as follows: we will review similar works, consider a mathematical model of social media, information dissemination models, the main metrics that are used to assess the speed of information dissemination, and, finally, present the results of simulation experiments.

II. Related works

Simulation modeling methods have been successfully used to study the process of information dissemination in social media.

Thus, the authors of paper [9] explore the role of an influential person on the polarization of opinions in social networks. The authors built an agent-based simulation model and conducted a series of experiments. It is known that society polarizes if there is no agreement on some important issues. Experiments have shown that if an influential person has extremist views, then the rate of polarization of users in the network increases rapidly and depends on the degree of their activity and on the size of the network coverage. If the influential person has neutral views, then polarization depends on the tolerance of the society. The rate of polarization has high values if the society is conservative. One way or another, it is advisable to use the results of the experiments when conducting various campaigns that use the influential person social networks and the processes of disseminating information in them on public opinion.

In the next work [10], special attention is paid to the dependence of the activation of actions of social network users on external events. This article presents a modeling pipeline that presents stimulus/response models describing how social systems respond to external events that are relevant to them. Two case studies are presented to test the validity of the various models. One of them examines the online reaction to events related to the election crisis in Venezuela. The other examines online responses to the development of the China-Pakistan Economic Corridor (CPEC). These case studies indicate that simple stimulus/response models can predict aggregate online trends.

The authors of [11] discuss a new aspect of the social media model, more precisely, the social network increases in density, the new edges are formed over time. Researchers study the formation of common knowledge through local interactions and the characteristics of social network structures.

The authors of [12] studied various strategies for disseminating knowledge in the network of employees of the

https://doi.org/10.51408/csit2023_47

academic center. For this purpose, a dynamic model (using the Monte Carlo method) was developed. The knowledge dissemination strategy, in this study, implies choosing of agents who will initially disseminate knowledge. Four strategies were considered: (1) the first five agents selected by the degree of centrality, (2) 5 agents with a large number of published works, (3) the first five agents selected by intermediate centrality (4) 5 central agents in clusters. Results of the investigation: the scenario in which agents were selected on the basis of centrality in clusters had the greatest impact on the dissemination of knowledge.

III. ONLINE SOCIAL NETWORKS

A. Mathematical model of online social networks

So, an online social network is an abstraction that defines the interaction between people in the Internet infrastructure. Typically, an online social network is represented as a graph. Nodes and edges represent users and relationships between them, respectively. In this article, the social graph is denoted by G = (V,E), where V = {v_1, v_2,…, v_N}, N= ｜V｜ and E ⊆ V × V represent the sets of vertices and edges of the graph. If e_{ij} ∈ E, this means that there is a connection between the nodes v_i and v_j, and then these nodes are called neighbors. The neighbors form the set of $Neigb_i$. neighbors of the node v_i. The cardinality of this set indicates the degree of the node v_i, i.e., d_i = ｜$Neigb_i$｜. A weight w_{ij} can be associated with each edge e_{ij} indicating the influence (expansion) of the probability of the node v_i on v_j, i.e., how likely (weight values closer to 1) or unlikely (weight values closer to 0) that node v_i can influence node v_j. In some studies, the network is considered as a directed graph. In a directed graph, $e_{ij} \in E$ means that v_i is a neighbor of v_j and v_j is an outer neighbor of v_i. This means that influence is not bidirectional: if one node influences another, then this does not mean that the reverse is always true. Random graphs [7] are often used as graph models of online social networks, namely: Erdesy-Renyi graph, Barabasi-Alberta graph, etc.

To study a random graph, it is useful to determine: (1)the difference in distribution - indicates a large number of bonds for some agents and minimal for others; the important phenomenon in this case is the "rich becomes richer" phenomenon, which leads to a high dispersion of vertices; (2)mutual orientation — the property indicates whether the relationship between the vertices is binary (whether the connection is bidirectional); (3) transitivity of bonds — an increase in the likelihood of bonds between agents that have bonds with the same peaks; (4)homogeneity - indicates the degree of appearance of bonds between similar agents (by gender, age, interests); (5)centrality - a metric that allows you to determine the significance or influence of a particular node or group in a network; (6) assortativeness - a tendency to form bonds between peaks of a large degree [8].

B. Diffusion Models

Various diffusion models have been proposed to model the process of information dissemination and to determine the influence of the initial set of nodes. Diffusion models are designed to describe the propagation process based on some network observations. There are three main classes of widely used diffusion models: threshold models, cascade models, and epidemic models.

The most popular threshold model is the Linear Threshold (LT) model. Each node v_i has an activation threshold Θ_i and can be active or inactive during the propagation of information in the model. At each time t, an inactive node can change its state if the activation threshold is exceeded (it depends on the number of active neighbors at time t-1). At the end of the simulation, the number of active nodes indicates the influence of the initial set of propagator nodes.

The most popular cascading model - the Independent Cascade (IC) model. Each node can be in either active or inactive state (as in the previous model). Initial spreader nodes are set as active in t = 0. In each timestamp t > 0, each node v_i activated in $t-1$ has one chance to activate each of its neighbor v_j with probability α. Then these nodes move to inactive state. This process continues until no node is activated in a timestamp t. The number of nodes activated during the process indicates the influence of the initial spreader nodes.

The Susceptible–Infected–Recovered (SIR) model is a widely used epidemic model in the literature. In this model, each node can be in either susceptible (SU), infected (IN), or recovered (RE) state. In timestamp t = 0, the initial spreader nodes are set to IN and all other nodes are set to SU. In each timestamp t > 0, each infected node v_i moves to recovered state with probability β after its attempt to infect each of its susceptible neighbors with probability α. The infection process continues until no infected nodes remain in the graph. At the end of the process, the number of recovered nodes represents the influence of the initial spreader set.

In this paper, the Barabashi-Albert model is chosen as a graph model, as it most closely matches the social network. The epidemic models SIR and SEIR were chosen as propagation models. To evaluate the efficiency of the distribution process, the centrality metrics are considered.

C. Metrics

Degree centrality CD of vertex i is defined as the number of vertices adjacent to i[13]:

$$C_D(i) = \sum_{j=1}^{n} a(i, j)$$

Here a(i,j) = 1 if and only if the vertices are connected by an edge.

Closeness centrality

$$C_c(x) = \frac{1}{\sum_{y}^{N} d(x, y)},$$

where d(x,y) is the distance between vertices x and y.

Betweenness centrality shows the number of shortest paths between all network nodes passing through a particular node [9]:

$$C_B(i) = \sum_{j \neq k} \frac{g_{jk}(i)}{g_{jk}},$$

where g_{jk} is the total number of the shortest paths from node j to node k, and $g_{jk}(i)$ is the number of paths that pass hrough i.

D. Simulation Tools

For simulation experiments, the AnyLogic simulation system was chosen. AnyLogic was developed in Russia. The simulation system is interesting in that it supports three approaches to simulation modeling: system dynamics, process-oriented and agent-based approaches. The agent approach is characterized by the fact that the object being modeled is a set of interacting agents acting according to certain scenarios. To study the factors influencing the process of dissemination of information in social media, in this work, an agent model is built in the AnyLogic environment, a model of a social network in which users interact with each other by sending messages to each other, putting likes and signatures.

IV. The Results of Simulation Experiments

Let's take a closer look at an agent-based simulation model designed to study the influence of directed and symmetrical bidirectional links on the process of information dissemination in social media.

A. Simulation model for determining the influence of directional and bidirectional links

First of all, let's define what is meant by directed and bidirectional connections in this study: (1) In directed connections there is a certain direction of information transfer. So, if some network user has a subscription to a blog, then the connection is considered to be directed, information is received (directed) from the blogger to the user, it is considered that the blogger "influences" the subscriber, he is the initiator; (2) Relationships that exist between friends will be considered bidirectional. Information is equally likely to be sent from one user to another and vice versa; (3) The number of users for whom this information is intended may be limited if the connections are directed. If the connections are bidirectional, then the information can be sent by different routes, it can reach the goal faster.

B. Epidemic models SIR and SEIR

To simulate the content transfer algorithm between network users, the SIR and SEIR models were created. Network users are represented by agents, agents are characterized by states and parameters, such as the probability of infection, recovery time, etc. For the epidemic SIR model, agents have 3 states that reflect the process of information dissemination: "susceptible", "infected", "recovered". Initially, all agents are in the "susceptible" state, which means they are susceptible to infection. Upon receiving the "Infection" message, the agent proceeds with the "sendcontenttofriends" transition. I select the user-initiator of information dissemination based on the value of the "getConnectionsNumber" parameter.

When an agent in the "susceptible" state receives an "Infected" message, it means that it has received information from another infected agent and changes its state from "susceptible" to "infective". Now he is "infected" and can spread information to his friends.An infected agent in the "infective" state can continue to interact with other agents and transmit information received by them. This can lead to further spread of the infection in the network.

Model parameter: (1) Number of users "userssize". This setting determines the total number of users (agents) on the network. It indicates the number of nodes that will participate in the simulation and interact with each other.(2)Infection duration "infectionduration": This parameter defines the length of time that the user remains infected (it is in the "infected" state before being "recovered").(3)The degree of infection "infectivity": the parameter determines the probability of transmission of infection from an infected agent to a susceptible agent at the time when the agents are in contact.(4)"contactrate" frequency: This parameter defines the speed or frequency with which agents in the network communicate with each other. Parameter values can be customized and changed to explore different scenarios. When the time interval associated with the duration of the disease (the agent is infected) ends, the infected agent enters the "recovered" state.

Based on the epidemiological SEIR model, 4 states were created to display the spread of information: "susceptible", "exposed", "infected", "recovered", once more parameter "latency" was added, the model extends the SIR model by adding a state of readiness for infection.

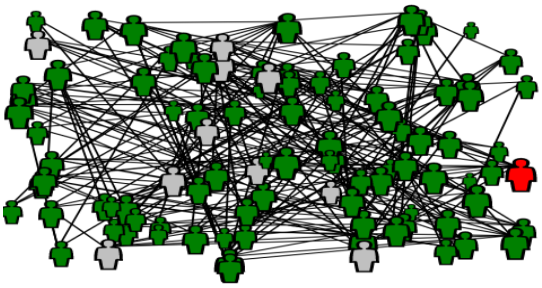
C. The results of experiments

Experiment 1 (SIR). Let's set the model parameters: (a) Population size (usersize): 100 users; (b) Probability of receiving information from an infected contact (probabilityofreceivinginfo): 0.4.; (c) Duration of stay in the infected state (infectionduration): 7 days (an infected user can spread the infection within 7 days); (d) infected user - selected user with the most connections: 88. As a result of the first run of the model, all users subscribed to the "infected" became infected. There were 90 uninfected, since the "infected" had 10 connections. Full recovery time - 69 days. The simulation results are presented in Fig.1.

Experiment 2 (SIR). Let's change the values of the model parameters: (a) Population size (usersize): 200 users; (b) Probability of receiving information from an infected contact (probabilityofreceivinginfo): 0.4.; (c) Duration of stay in the infected state (infectionduration) 7 days: infected user - the user with the most connections (16) was selected. As a result of the second run of the model, the infection spread in 20 days, 166 users remained uninfected, since the number of connections for the selected agents is 16 and 18. The results of the experiment are shown in Fig.2. Next, consider the epidemiological model of SEIR

Experiment 1 (SEIR). Set the model parameters: (a) Population size (usersize): 50 users; (b) The latency duration is: 2 days; (c)Duration of stay in the infected state infectionduration: 7 days (an infected user can spread the infection within 7 days); (d) Infected user – 4. As a result of the first run of the model, all users subscribed to the "infected" became infected. There were 28 uninfected, since the "infected" had 22 conditional subscribers. Full recovery time - 30 days.

Experiment 2 (SEIR). Set the model parameters: (a) Population size (usersize): 50 users; (b) The latency duration is: 2 days; (c) Duration of stay in the infected state infectionduration: 7 days (an infected user can spread the infection within 7 days); (d) Infected user (selected user with the most connections): 4. As a result of the second experiment, all network users who had a subscription to already two infected users were infected, the time for complete recovery was 17 days. The results show that the number of bloggers and the number of their connections can have a significant impact on the speed of information dissemination. A similar series of experiments was carried out with a model in which the links are bidirectional. Two epidemic models SIR and SEIR were also used.



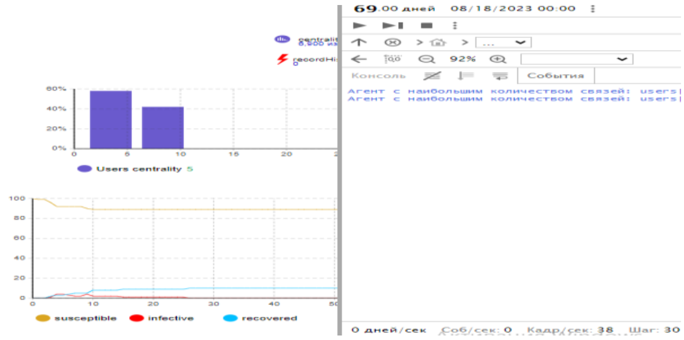
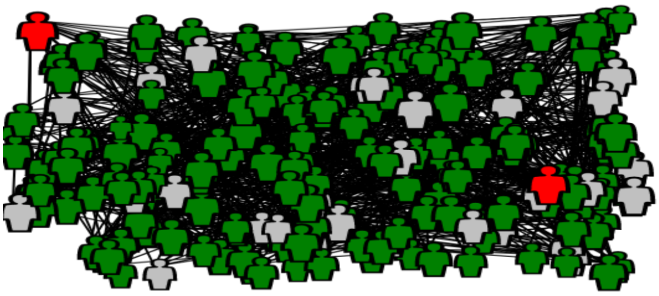


Fig. 1. The results of the first experiment (SIR)



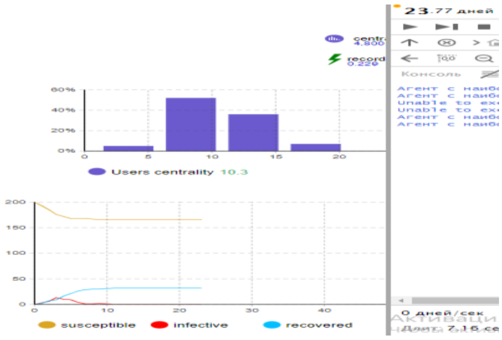


Fig. 1. The results of the second experiment (SIR)

V. CONCLUSION

So, the paper presents simulation models that allow you to determine what factors affect the reliability and speed of information transfer in a social network. This is important, for example, for the distribution of advertising in solving marketing problems. Thus, in a particular situation, it is possible to influence the process of information dissemination.

Experiments have shown that the parameters of simulation models built on the basis of epidemic models (SIR,SEIR), such as population size, frequency of getting contact with another user, probability of infection, etc., significantly affect the course of dissemination of information in the social network. Increasing the size of the population and the frequency of obtaining contact contributes to faster infection, and, consequently, more rapid receipt of information by a large number of network users. The length of the time interval and the period when the user is in a latent state also leads to the fact that the information dissemination process continues as long as possible, which increases the reliability of information transmission to more users. Consideration should also be given to the selection of specific users, namely users with a large number of connections (Experiment 2, SIR).

REFERENCES

[1] A. Y. Dombrovskaya. “Repesentation of the Russian Youth Civil Activity in Social Media. Monitoring of Public Opinion: Economic and Social Changes. No. 6., pp. 203--225, 2021.

[2] D. Gubanov and A.Chkhartishvili. “A conceptual approach to the analysis of online social networks”. Large-Scale Systems Control, pp.222--236, 2013.

[3] (2023) The AnyLogic website. [Online]. Available:http://www.anylogic.ru/

[4] N.Zhao, X.Cheng, X.Guo: Impact of information spread and investment behavior on the diffusion of internet investment products. Physica A: Statistical Mechanics and its Applications. Elsevier, Vol. 517, pp.427--436, 2019.

[5] M. Gatti. Large-Scale Multi-agent-Based Modeling and Simulation of Microblogging-Based Online Social Network. Multi-Agent-Based Simulation, MABS, pp.17--33, 2014.

[6] A.Usanin, I.Zimin and E.Zamyatina. “Study of Strategies for Disseminating Information in Social Networks Using Simulation Tools”, *Proceedings of .in: Analysis of Images, Social Networks and Texts: 9th International Conference, AIST 2020,* Skolkovo, Moscow, Russia, Vol. 12602, Springer, pp.303—315, 2021.

[7] A. M. Raigorodskii. “Random graph models and their applications”. Random graph models and their application”, *Proceedings of Moscow Institute of Physics and Technology (State University)*. Vol.2, No 4, pp.130–140, 2010.

[8] I.Dmitriev and E.Zamyatina. “How to Prevent Harmful Information Spreading in Social Networks Using Simulation Tools”, *Proceedings of 8th International Conference “Analysis of Images, Social Networks and Texts”, AIST 2019*, Kazan, Russia, 2019, Revised Selected Papers “Communications in Computer and Information Science”, Vol. 1086, Springer, pp. 201--213, 2020.

[9] J. M. Betts and A. – M. Bliuc .”Effect of Influencers on Societal Polarization‖”, *Proceedings of the 2022 Winter Simulation Conference B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.*, Singapore, pp. 370--381, 2022.

[10] H.Shao et al. Simulating Online Social Response: a Stimulus/Response Perspective. *Proceedings of the 2021 Winter Simulation Conference S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.,*Phoenix, USA, pp. 1--12, 2021.

[11] C.J. Kuhlman, G. Korkmaz, S. S. Ravi and F.Vega-Redondo. An Agent-Based Model of Common Knowledge and Collective Action Dynamics on Social Networks. *Proceedings of the 2020 Winter Simulation Conference K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.*, Virtual Conference, USA, pp. 218—229, 2020

[12] H.Kang and D.Munoz. “A dynamic network analysis approach for evaluating knowledge dissemination in a multi-disciplinary collaboration network in obesity research”, *Proceedings of the 2015 Winter Simulation Conference, L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds*. Huntington Beach, USA, pp.1319—1330, 2015.

[13] M.E.J. Newman A measure of betweenness centrality based on random walks. URL: http://aps.arxiv.org/pdf/cond-mat/0309045.pdf (дата обращения: 10.06.2023).