How to Prevent Harmful Information Spreading in Social Networks Using Simulation Tools

Ivan Dmitriev[1(✉)] and Elena Zamyatina[2(✉)]

[1] Perm State University, Perm, Russian Federation
ivandmitriev5@gmail.com
[2] National Research University Higher School of Economics, Perm, Russian Federation
e_zamyatina@mail.ru

Abstract. The paper discusses the problems of preventing harmful information spreading in social Networks. Social networks are widespread nowadays and are used not only for managers and marketers propagation of advertising, promotion of goods, but also by attackers to spread harmful information. Thus, there is a need to counter the attackers. This paper presents simulation tools and several features that contribute to the successful application for modeling social networks and examine different strategies preventing rumors and harmful information spreading. The authors cite an example of a simulation model for identifying intruders in a social network, software tools and the results of simulation experiments.

Keywords: Social networks · Dynamic modeling · Static modeling · Random graphs · Information dissemination

1 Introduction

Social networks have become an integral part of human life. So we always have the opportunity to quickly connect with friends or other people, find out the news bulletin for today, share our opinion. Social networks are used by large companies, individual entrepreneurs or managers in order to promote products and brands. They are used to spread advertising, technology, knowledge, investment, etc. [1–3]. Social networks are also used by government agencies and various communities to control public opinion. However, rumors [4, 5], and "fake" news [6], and malicious information [7] are successfully spread on the social network. We should not forget that social networks are successfully used by criminals and terrorists for the purpose of criminal conspiracy [8–10].

Large numbers of studies related to monitoring, network analysis and prediction of online networks development and information management have appeared. The work [1] considers several tasks being popular in online social networks: a task of analyses, a task of forecasting, a task of management. So a task of analysis and monitoring implies the collection of statistical data, the identification of changes in the online social networks

The study was carried out with the financial support of the Russian Foundation for Basic Research in the framework of the research project No. 18-01-00359.

W. M. P. van der Aalst et al. (Eds.): AIST 2019, CCIS 1086, pp. 201–213, 2020.
https://doi.org/10.1007/978-3-030-39575-9_21

and an assessment of various indicators. This could be, for example, identifying the initiators of the discussion, monitoring the actual information objects being discussed, assessing the naturalness of the discussions, etc. In forecasting tasks based on data that were obtained during network analysis usually are trying to predict how the network will change in the future. An example is the problem of disseminating information in the network. Management tasks include all the above tasks for the reason that, first of all, it is necessary to perform an analysis of the current situation in the social network, make a forecast of its development, and, depending on the purpose of the research, consider various management strategies. The task of spreading public opinions is an excellent example.

There are various approaches to solving these problems, one of which is the modeling of social networks [2]. This approach, in contrast to the usual network analysis (Social Network Analyzes (SNA)) [1, 11] is more flexible: you can build a social network model that takes into account the characteristics of specific studies, set the required modeling conditions. It is known that the mathematical abstraction of a social network is a graph, and since from the point of view of an outside observer, connections are formed randomly, random graphs are used [2, 12].

The article is structured as follows: static and dynamic approaches to a modeling of social networks are examined. Static approach supposed an analysis of structural characteristics of the online social networks; dynamic approach examines changes in network characteristics depending on time, cause-and-effect relationships. Moreover special software tools are proposed. These tools can be successfully used by both approaches. The following is a description of the task of distributing prohibited information in a network implemented in the simulation system Triad.Net.

2 Online Social Networks Modeling

There are two basic approaches to the analysis of social networks: static and dynamic one (as it was mentioned above) [1].

The first approach involves the study of network structure (topology), its basic properties (contiguity, the degree of centrality, distance and others) [2]. This approach supposes the investigation of the current state of a "snapshot" of a social network. Main attention is paid to the geometric characteristics of the network (structure of network), as well as the different relations between the nodes (members of the social network).

Static (structural) approach allows one to characterize accurately the current state of the system, but does not make it possible to see one to-many pattern that become visible only in the study of the structure of the network in dynamic. Indeed the useful information about social network "can be achieved at points in time through the use of polling and survey data, but the most interesting questions typically lie in the space in between these snapshots in time" [3]. The causal mechanism of the changes in social networks may be obtained due to simulation (in time). The static approach allows to understand such complex adaptive system as society, assists the scientists and managers to take an appropriate decision, but only simulation (discrete event or agent-based) "provides a fully traceable implementation of these concepts that readily accommodates the varying timescales at which events unfold within society" [3].

The study of structural characteristics allows us to fairly accurately characterize the current state of the system, but it does not allow us to understand many of the patterns that are noticeable only when time changes in dynamics. The authors of [1] enumerate tasks in which the structural characteristics of a social network are investigated: (a) searching for exact algorithms for generating recommendations of friends and content based on a social graph [13]; (b) identifying the initiators of the discussion; (c) monitoring of the actual objects under discussion; (d) forecasting the further development of the network, etc.

Today, thanks to the study of the structural characteristics of social networks, a large number of models have been developed that have structural similarities with real networks (for example, the Buckley-Ostgus model [14], the Watts-Strogatz model [15], the copy model [16], etc.

Dynamic modeling studies social networks in dynamics. Over time, the number of nodes in the network, communication between nodes may change. In addition, the processes taking place in the network should be considered: the dissemination of information, opinions, rumors, knowledge, etc.

There are such models of activity in social networks [1].

1. Macro-level models consider the network as a whole, without taking into account connections between nodes;
2. Micro level models view the network, taking into account the links:
   a. models with thresholds are models in which there is a threshold value or a set of threshold values used when a state changes;
   b. models of independent cascades (in which each node gets at a certain step a chance to activate other nodes) [17];
   c. propagation models based on analogies with physics, medicine, and other branches of science [4, 5];
   d. Leakage [18] and contamination [19] models are a popular way of studying the dissemination of information and innovation in social systems.

Dynamic modeling, in contrast to static modeling, allows studying the reasons for why the network is in a certain state, which event is the cause of its transition to another state. Thus, it is possible to solve more complex problems: predict the dissemination of information in networks [20], form public opinion [1], form a discussion topic [13], etc.

3 Combining Dynamic and Static Modeling

So, a lot of research confirms that the combined use of static and dynamic modeling is relevant. Consider some of these researches.

The authors of [21] studied various strategies for disseminating knowledge in the network of employees of the academic center. For this purpose, the authors developed a dynamic model (using the Monte Carlo method). The network structure in this study is static (the number of agents and the connections between them do not change), but the process of knowledge dissemination is dynamic.

The simulation was carried out according to the following scenario: at each moment of time for each pair of related agents, it turned out whether they had been in contact during this period of time or not, after which, if a contact occurred, one of the agents passed some piece of information to another. Upon reaching a certain level of awareness, the agent joined the dissemination of knowledge.

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 degree of centrality, (2) 5 agents with the big number of published work, (3) the first five agents selected by intermediate centrality (4) 5 central agents in clusters.

To assess the effectiveness of strategies, two key indicators were considered: the proportion of aware agents at specific time intervals and the amount of time required to spread knowledge between specific portions of agents.

This model was tested on the network of cooperation of the research center. The model was built on the basis of information about collaboration, the number of publications, etc.

The authors examined the effects of various strategies on the dissemination of knowledge and obtained the following results: the scenario in which agents were selected on the basis of centrality in clusters had the greatest impact on the dissemination of knowledge.

The author of [22] attempted to analyze the distribution of information in an egocentric network that has a unique node as a source. The study is based on a stochastic multi-agent approach, where each agent is formed according to certain rules using data from the real social network Twitter. The modeling process consists of six phases and is iterative.

The first phase consists of uploading data from a real social network. After that, during the second phase, the topics and moods of the messages (posts) extracted from the sample data are classified. Then, in the third stage, sets of samples for each user are created from the previously classified data. Each set contains user messages and messages from his/her news feed (that is, messages from users who he/she is subscribed to). Each model of user behavior is built from these sets (fourth phase), and the models are used as input for a stochastic simulator (fifth phase). The model is executed. The loop is repeated several times until the most accurate model is found.

Experiments have shown that the proposed approach is promising for modeling user behavior in a social network.

A simulation model for the distribution of harmful information in the network was developed in [19]. An epidemiological model was used as the basis. The classical model of the spread of infection is based on the following cycle of the carrier disease: initially, a person is susceptible to infection. If this person contacts an infected person, he can with some probability become infected. Subsequently, the person over time either recovers, acquiring immunity, or dies; immunity decreases with time, and the person again becomes susceptible to infection.

A similar cycle was implemented in our paper. But we must notice that ordinary users who are susceptible to infection are attackers-agents.

The task of disseminating harmful information can be formulated as follows. The process of distribution of harmful information initiates any attacker agent by sending

messages with harmful information to his list of contacts. An attacker can start a single attacker or group of them. They send messages through each time unit.

Subscribers-recipients, having accepted the message, with probability β are included in the attack process (become intruders). It is assumed that the user either read a message, either ignored it or deleted it altogether.

In addition, in each unit of time, the attacked nodes can be protected because of defense mechanisms impact. Thus, they cease to send harmful information and become immune to further attacks.

The simulation results are numerical arrays of data describing the dynamic process of propagation of harmful information (the number of attacking, protected and potentially vulnerable nodes in each time unit).

The simulation experiments were carried out using this model. The next section presents the task of distributing harmful information in the network and illustrates its implementation in the simulation system Triad.Net.

So the popularity of social networks is growing every day. In this regard, there are more and more threats against which you need to protect network users. One of threats: the distribution of harmful information in networks [7, 17]. Recently, a large number of publications related to the study of the process of dissemination of harmful information, rumors, etc. have appeared.

So in [7] an algorithm for large-scale networks monitoring with dynamically changing cascades of harmful information is considered. Several ways to decrease the spread of harmful information through immunization of network nodes are considered in [17]. The search for candidates for immunization is performed dynamically during the process of information dissemination.

Let us consider the task of identifying intruders on the network.

Formulation of the problem:

Given: N - the number of nodes equal to the number of network users; I_0 - the number of malicious subscribers - the primary sources of threat; R_0 is the number of subscribers initially insensitive to attacking influences; β - parameter that reflects the strength of the threat, the likelihood of an attack; γ is a parameter reflecting the degree of resistance to the threat, the probability of subscriber protection (β and γ in this study are defined as constants, but can be expressed as functions depending on the profiles of social network subscribers); t is the process time (in arbitrary units of time).

The process of distribution of harmful information (Z_I) initiates any attacker agent by sending messages with harmful information to his list of contacts. An attacker can be launched by a single attacker or a group of attackers. They send messages through each time unit. Subscribers-recipients, having accepted the message, with probability β are included in the attack process (become intruders). It is assumed that the user either read a message; either ignored it or deleted it altogether.

In addition, in each unit of time, the attacked nodes can be protected due to the impact of defense mechanisms. Thus, they cease to send harmful information and become immune to further attacks.

We implement this task using the Triad.Net modeling system.

The Triad.Net modeling system was developed at Perm State University [23–25] and is intended for modeling computer systems.

The Triad simulation model is divided into three layers: *a structure layer* (a set of objects connected by communication lines, with the help of which objects exchange information with each other), *a routine layer* and a *message layer*. In our case, the set of objects is the set of network users who are in a relationship with each other. The layer of routines - software tools for the implementation of scenarios of behavior of network users. The message layer is intended to describe messages of a complex structure. These messages are shared by network users.

The structure layer is a procedure with parameters. So, the computer network model, which is a dedicated server and several client computing nodes, can be described in Triad language as follows: star (Server, Node [1..n]), where star is a graph constant corresponding to the star network topology. In the Triad language, other graph constants are also defined: cycle (cycle), rectan (lattice), tree (tree), etc.

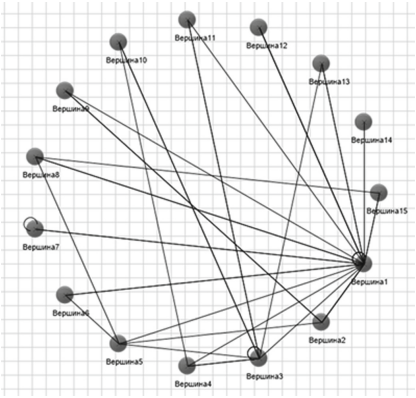

Fig. 1. Bollobás-Riordan model

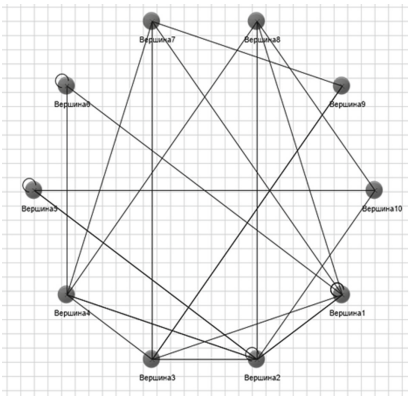

Fig. 2. Buckley-Osthus model

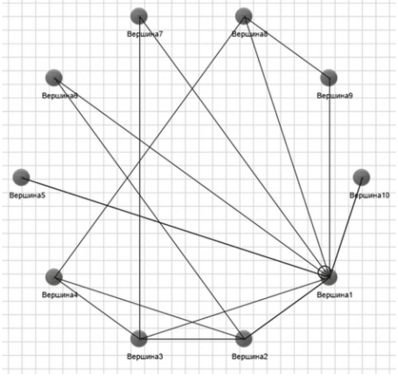

Fig. 3. Copying model

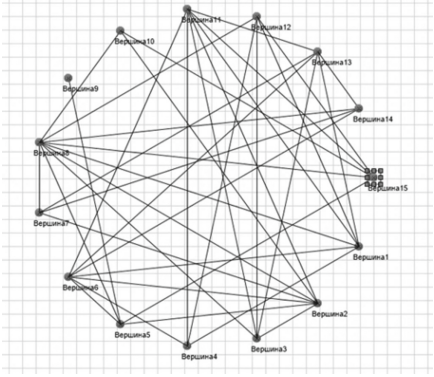

Fig. 4. Erdos- Renyi model

A distributed system consisting of 3 servers (client nodes are connected to each of the servers) can be described as follows: star (Server, Serv [a..c]) + star (Serv [a], Node [1..k]) + star (Serv [b], Node [k + 1..m]) + star (Serv [c], Node [m + 1..n]). Here operations (union of graphs) are used. It should be noted that the following operations with graphs are defined in the structure layer: add/delete vertices, arcs, edges, union, intersection of graphs, etc. The description above defines a whole family of structures. Also there are procedures for constructing random graphs, used as models of social networks (Figs. 1, 2, 3 and 4).

In addition, graph analysis procedures (diameter, number of vertices, edges, etc.), as well as: clustering coefficient, centrality, transitivity, mutual orientation, etc., are implemented in the layer of structures.

Routine is a specific scenario in which agents act. A routine consists of a sequence of events that plan each other. The execution of an event is accompanied by a change in the state of the object. Routine has input and output poles, with their help, agents interact with each other.

A description of routine using the simulation language Triad [25] may be represented as:

```
Routine <Name> { <A section of parameters> | <A defini-
tion of poles> }
[<A section of initialization of routine> ] { <A descrip-
tion of events>}
EndRout
```

The following is the code for the routine used in the task.

```
routine Rout[boolean Defence; boolean bad; real beta; re-
al gama](InOut pol[50])
// initialization
initial
  integer i;
  real Seed:=0;
  if (bad) then
    schedule SendMessage in 0;
  endif;
endi
//Events
// Event send messages
event SendMessage;
  // The command to send a message to all poles
  out " ";
  // With probability gama, an agent turns from an at-
  tacker into a protected one,
  // and stops sending messages.
  if (RandomReal()<gama) then
```

```
    Defence:=true;
    bad:=false;
  endif;
  if (Defence=false) then
    schedule SendMessage in 1;
  endif;
ende

// Input Processing Event
event;
  // With the probability of the beta, the agent joins
  the attack.
  //(Defence=false) check that the agent is not protected
  //(bad=false) check that the agent is not attacking
  if (Defence=false)&(bad=false)&(RandomReal()<beta) then
    schedule SendMessage in 1;
    bad:=true;
    Seed:=Seed+1;
  endif;
ende
endrout
```

Information procedures are used to collect statistical information during simulation experiments. Information procedures in the process of modeling observe the model elements (events, variables, poles). When the state of the observed object changes (i.e. when a variable value changes, when an event is executed, after a message arrives at the input pole or a message is transmitted from the output pole) the information procedure is connected to a specific model element and the data is processed according to the algorithm specified in the information procedure.

A description of procedure using the simulation language Triad [26] may be represented as:

```
Infprocedure<Name>(< A section of parameters >)]
initial< initial conditions >endi
handling// executed when one of the parameters changes
<Information Procedure Operator> { ";" <Information Pro-
cedure Operator> }endh
processing// executed when requesting a result
<Information Procedure Operator> { ";" <Information Pro-
cedure Operator> }endp Endinf
```

Below is the code of the information procedure, which calculated the number of attacking, vulnerable and protected agents.

```
infprocedure ip(
    in array[50] of boolean bad;
    in array[50] of boolean Defence)
    // initialization
    initial
      integer cbad,cdef,cfree,i;
      real prevtime:=0;
      print "time\t"+"bad\t"+"def\t"+"\tfree";
    endi
    handling
      if prevtime!=SystemTime then
        print prevtime+"\t"cbad+"\t"+cdef+"\t"+cfree;
      endif;
      cbad:=0;cdef:=0;cfree:=0;
      for i:=0 to 49 do
        if Defence[i] then
          cdef:=cdef+1;
        endif
        if bad[i] then
          cbad:=cbad+1;
        else
          cfree:=cfree+1;
        endif
      endf;
      prevtime:=SystemTime;
    endh
    endinf
```

During the experiments, several runs of a simulation model with different input parameters were carried out.

According to the results of the experiments, the following conclusions can be drawn:

- the I (t) attack process has an exponential dependence (Fig. 5);
- as β increases, the rate of infection of the nodes increases (attack intensity) (Figs. 5 and 6);
- with an increase in the probability of an attack β and a low probability of protection γ, the time of the defense process increases (Fig. 6);
- the number of protected agents grows slower due to the fact that the agents transition to this state from the attack state (Fig. 6);
- with a low probability of attack β and a high probability of protection γ, the attack process has a non-exponential form (Fig. 7).

The authors conducted a number of other experiments and confirmed the studies [19], which showed that the rate of infection of agents depends not on the network topology, but on the number of connections between agents.

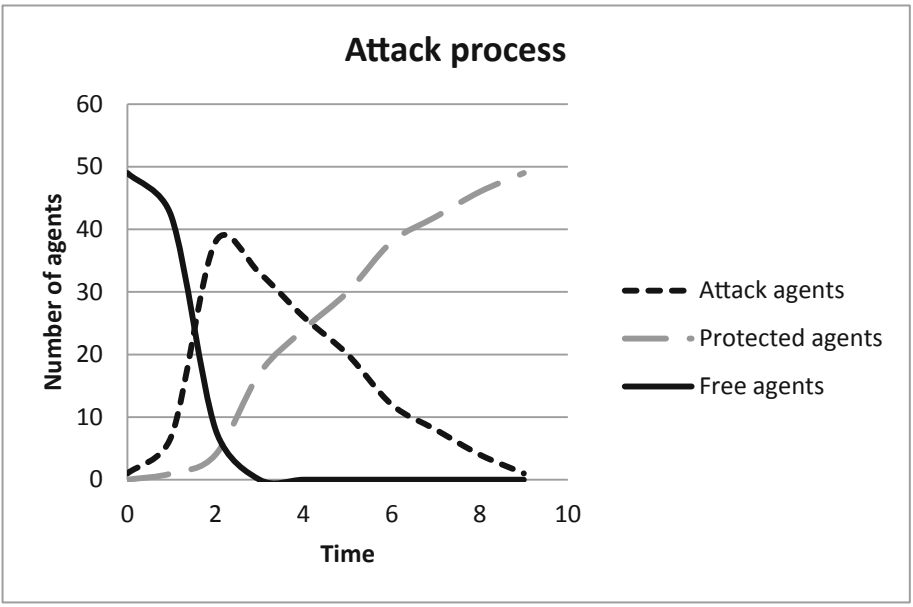

Fig. 5. Attack process, n = 50, β = 0.6, γ = 0.3

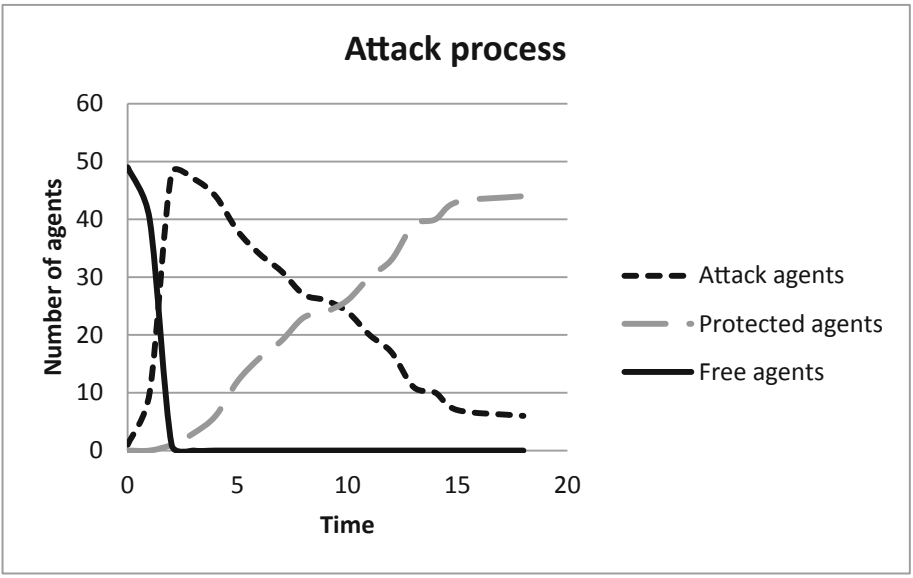

Fig. 6. Attack process, n = 50, β = 0.9, γ = 0.1

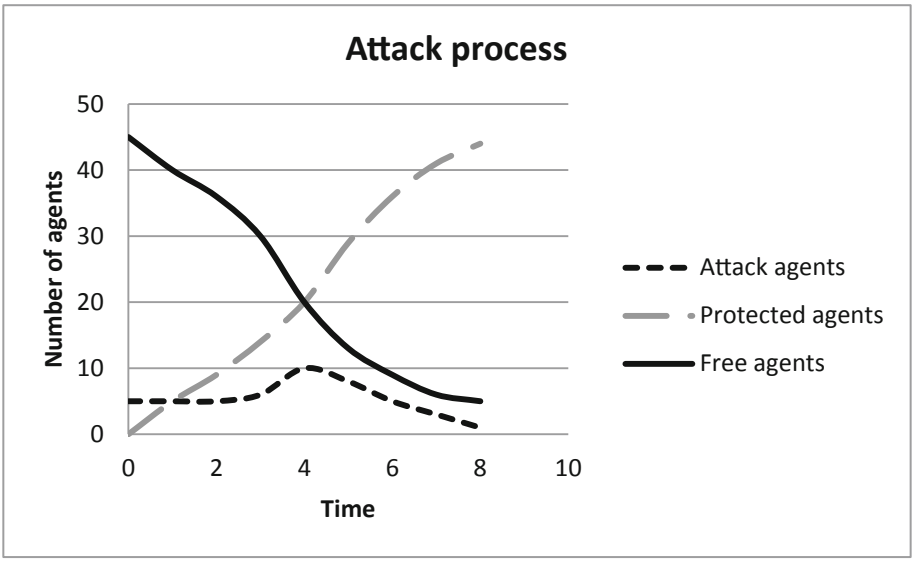

Fig. 7. Attack process, n = 50, β = 0.1, γ = 0.9

4 Conclusion

The paper considers an example of solving a problem of disseminating malicious information in a social network. To study the dissemination of information in the network, the simulation method and the simulation system Triad.Net were used.

Triad.Net software has a number of characteristics that make it a convenient tool for modeling social networks. Simulation tool Triad.Net includes procedures for constructing graphs, including random ones, used as models of social networks. The number of graph vertices is given as input parameters, so it is possible to construct a graph with a large number of vertices and with a complex structure, while the structure description in the Triad language remains concise. The procedures for analyzing graphs of a layer of structures make it possible to determine the structural characteristics of a graph, which are usually obtained when performing network analysis procedures. In addition, Triad.Net allows to explore and dynamic processes occurring in a social network.

So this paper demonstrated the viability of the Triad.Net for solving problems related to the research of social networks, including the tasks of disseminating of harmful information.

Acknowledgements. The study was carried out with the financial support of the Russian Foundation for Basic Research in the framework of the research project No. 18-01-00359.

References

1. Gubanov, D., Chkhartishvili, A.: A conceptual approach to the analysis of online social networks. Large-Scale Syst. Control (45), 222–236 (2013)

2. Davydenko, V.A., Romashkina, G.F., Chukanov, S.N.: Modelirovanie sotsial'nkh setei, pp. 68–79. Vesntik TSU (2005)
3. Zhao, N., Cheng, X., Guo, X.: Impact of information spread and investment behavior on the diffusion of internet investment products. Phys. A **512**, 427–436 (2018)
4. Zhang, Y., Zhu, J.: Stability analysis of I2S2R rumor spreading model in complex networks. Phys. A **503**, 862–881 (2018)
5. Zan, Y., Wu, J., Li, P., Yu, Q.: SICR rumor spreading model in complex networks: Counterattack and self-resistance. Phys. A **405**, 159–170 (2014)
6. Ilieva, D.: Fake news, telecommunications and information security. Int. J. "Inf. Theor. Appl." **25**(2), 174–181 (2018)
7. Yang, D., Liao, X., Shen, H., Cheng, X., Chen, G.: Dynamic node immunization for restraint of harmful information diffusion in social networks. Phys. A **503**, 640–649 (2018)
8. Bindu, P.V., Thilagam, P.S., Ahuja, D.: Discovering suspicious behavior in multilayer social networks. Comput. Hum. Behav. **73**, 568–582 (2017)
9. Tumbinskaya, M.V.: Protection of information in social networks from social engineering attacks of the attacker. J. Appl. Inform. **12**(3(69)), 88–102 (2017)
10. Filippov, P.B.: Use and implementation of personal data protection in social networks of the Internet. J. Appl. Inform. (2(38)), 71–77 (2012)
11. Dang-Pham, D., Pittayachawan, S., Bruno, V.: Applications of social network analysis in behavioural information security research: concepts and empirical analysis. Comput. Secur. **68**, 1–15 (2017)
12. Raigorodskii, A.M.: Proceedings of Moscow Institute of Physics and Technology (State University). In: Random Graph Models and Their Application, pp. 130–140 (2010)
13. Roth, M., et al.: Suggesting friends using the implicit social graph. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, pp. 233–242 (2010)
14. Buckley, P., Osthus, D.: Popularity based random graph models leading to a scale-free degree sequence. Discrete Math. **282**(1–3), 53–68 (2004)
15. Watts, D., Strogatz, S.: Collective dynamics of 'small-world' networks. Nature **393**, 440–442 (1998)
16. Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A., Upfal, E.: Stochastic models for the web graph. In: Proceedings of the 41st Symposium on Foundations of Computer Science, p. 57 (2000)
17. Zhou, C., Lu, W.-X., Zhang, J., Li, L., Hu, Y., Guo, L.: Early detection of dynamic harmful cascades in large-scale networks. J. Comput. Sci. **28**, 304–317 (2018)
18. Zhukov, D., Khvatova, T., Lesko, S., Zaltcman, A.: Managing social networks: applying the percolation theory methodology to understand individuals' attitudes and moods. Technol. Forecast. Soc. Chang. **129**, 297–307 (2018)
19. Abramov, K.G.: Modeli ugrozy rasprostraneniya zapreshchennoi informatsii v informatsionno-telekommunikatsionnykh setyakh., Vladimir (2014)
20. Newman, M.E.: A measure of betweenness centrality based on random walks. http://aps.arxiv.org/pdf/cond-mat/0309045.pdf
21. Kang, H., Munoz, D.: A dynamic network analysis approach for evaluating knowledge dissemination in a multi-disciplinary collaboration network in obesity research. In: Proceedings of the 2015 Winter Simulation Conference, Huntington Beach, pp. 1319–1330 (2015)
22. Gatti, M., et al.: Large-scale multi-agent-based modeling and simulation of microblogging-based online social network. In: Alam, S.J., Parunak, H. (eds.) MABS 2013. LNCS (LNAI), vol. 8235, pp. 17–33. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-54783-6_2
23. Zamyatina, E.B., Mikov, A.I., Mikheev, R.A.: TRIADNS computer networks simulator linguistic and intelligent tools. Int. J. "Inf. theor. Appl." (IJ ITA) **19**(4), 355–368 (2012)

24. Zamyatina, E.B., Mikov, A.I.: Programmnye sredstva sistemy imitatsii Triad.Net dlya obespecheniya ee adaptiruemosti i otkrytosti. Informatizatsiya i svyaz (5), 130–133 (2012)
25. Mikov, A.I.: Formal method for design of dynamic objects and its implementation in CAD systems. In: Gero, J.S., Sudweeks F. (eds.) Advances in Formal Design Methods for CAD, Preprints of the IFIP WG 5.2 Workshop on Formal Design Methods for Computer-Aided Design, Mexico, pp. 105–127 (1995)
26. Mikov, A.I.: Avtomatizatsiya sinteza mikroprotsessornykh upravlyayushchikh system. Irkutsk University Publ., Irkutsk (1987)