Critical Points of Information Influence in Social Networks

Oleksandr Milov [1], Serhii Yevseiev [2], Stanislav Milevskyi [3], Krzysztof Kajstura [4] and Ruslana Ziubina [5]

[1,2,3] *Simon Kuznets Kharkiv National University of Economics, Nauki ave., 9a, Kharkiv, 61166, Ukraine*
[4,5] *University of Bielsko-Biala, Willowa str., 2, Bielsko-Biala, 43-309, Poland*

Abstract

Social networks are considered from the point of view of informational influences on network participants (agents). The dynamic processes of forming opinions and the dynamics of information influence on network agents are considered. Models and algorithms for identifying critical points of a social network (influencing agents) are presented, the impact on which allows manipulating the aggregate opinion of network participants that form a social network.

Keywords

social network, agent, informational influence, influence models

1. Introduction

An instrument of active influence on the actions of network users and a means of forming and disseminating opinions is undoubtedly a new type of resource - online social networks. Their role has grown significantly with the advent of Web 2.0 [1]. The target segments for using this tool can vary significantly and range from the formation of consumer demand to the formation of public opinion during elections at various levels (from state to district or city). All this allows to talk about the transformation of social networks into a tool for strategic management of the population [2].

A social network can be represented as a graph, the vertices of which are individuals (agents), and the edges are the various relationships between them. It is known that the opinion of an individual in a social network is largely determined by the opinion of his influential neighbors [3, 4]. Knowing this, it is possible, both outside the network and inside it, in order to achieve our goals, to try to change the opinions of a small set of key users in popular online social networks (such as Facebook, Twitter, LinkedIn), through which opinions will spread throughout the network.

The decisions of most agents can be based on the decisions of other agents they observe. This is especially typical in conditions of a lack of information or the impossibility for various reasons to process it and draw appropriate conclusions. At the same time, the structure of the network, which determines who trusts whom, can contribute to the emergence of large information cascade changes even with insignificant changes in the decisions of an insignificant part of agents [5,6].

In this paper, the formation and dynamics of opinions in a social network is considered, and an attempt is made to highlight those critical points of the social network (influencing agents), the impact on which allows manipulating the aggregate opinion of the network participants forming the social network, as well as the resulting game-theoretic problems information confrontation.

III International Scientific And Practical Conference "Information Security And Information Technologies", September 13–19, 2021, Odesa, Ukraine
EMAIL: Oleksandr.Milov@hneu.net (A. 1); Serhii.Yevseiev@hneu.net (A. 2); Stanislav.Milevskiy@hneu.net (A. 3); kkajstura@ath.bielsko.pl (A. 4); rziubina@ath.bielsko.pl (A. 5)
ORCID: 0000-0001-6135-2120 (A. 1); 0000-0003-1647-6444 (A. 2); 0000-0001-5087-7036 (A. 3); 0000-0001-8739-1224 (A. 4); 0000-0002-8654-6981 (A. 5)

2. Social network as a medium of information impact

A *social network* at a qualitative level is understood as a social structure consisting of a set of *agents* (subjects - individual or collective, for example, individuals, families, groups, organizations) and a set of relations defined on it (a set of *connections* between agents, for example, acquaintance, friendship, cooperation, communication). Formally, a social network is a *graph* $G(N, E)$, in which $N = \{1, 2, ..., n\}$ is a set of vertices (agents) and E is a set of edges reflecting the interaction of agents.

Social networks contribute, firstly, to the organization of *social communications* between people and, secondly, to the realization of their *basic social needs*. There are two intersecting interpretations of the social network - as a social structure and its specific Internet implementation.

When modeling social networks, the mutual influence of their members (*agents*), the dynamics of their *opinions*, etc. there is a need to take into account the factors (effects) that take place in real social networks. In general, in real social networks, the following **effects and properties** can occur, due to both the characteristics and needs of agents (influencing and being influenced), the nature of their interaction, and the properties of the social network itself. Of the many effects and properties of the social network presented in [9,17-19], the following are of interest from the point of view of information impact:

1. the presence of agents' own opinions;
2. changing opinions under the influence of other members of the social network;
3. the different significance of the opinions (influence, trust) of some agents for other agents;
4. varying degrees of agents' susceptibility to influence (conformism, stability of opinions);
5. the existence of an indirect influence in the chain of social contacts. Decrease in indirect influence with increasing "distance";
6. the existence of "opinion leaders" (agents with the maximum "influence"), formalization of influence indices;
7. the impact of the structural properties of social networks on the dynamics of opinions;
8. the activity (purposeful behavior) of agents;
9. optimization of information impacts;
10. information management in social networks.

Should be noted the peculiarities of the impact of the structural properties of social networks on the opinions dynamics [7, 8]:

- the more connections an agent has, the more opportunities he has through his environment to influence the entire network, on the one hand, and, on the other, more vulnerability to someone else's influence;
- the effect of clustering (the higher the density of connections between active agents-neighbors, the greater the likelihood of activation of the agent associated with them; see below the related concept of “strong tie”);
- local intermediateness (the greater the intermediate value of the agent, the, on the one hand, the greater its value in the dissemination of opinion / information from one part of the network to another (the role of an information broker), and, on the other hand, the less its influence on the neighbor agent - see the related concept of “weak tie” below);
- the small diameter of the social network causes a short chain of dissemination of opinion in the network.

Influence is the process and result of an individual (subject of influence) changing the behavior of another subject (individual or collective object of influence), his attitudes, intentions, ideas and assessments (as well as actions based on them) in the course of interaction with it. *Influence* - the ability to influence someone's ideas or actions. Distinguish between directed and undirected influence. Directed (purposeful) influence - influence that uses persuasion and suggestion as mechanisms of influence on another subject. In this case, the subject of influence sets itself the task of achieving certain results (for example, choosing certain actions) from the object of influence. Non-directed (non-targeted) influence is an influence in which the individual does not set himself the task of achieving certain results from the object of influence.

In a social network, agents often do not have sufficient information for making decisions or cannot independently process it, so their decisions can be based on the decisions they observe or the perceptions of other agents (social influence). Social influence is realized in two processes: communication (in the course of communication, exchange of experience and information, discussion of certain issues with authoritative neighbors for the agent, he comes to certain ideas, attitudes, opinions) and comparison (in search of social identity and social approval, the agent accepts representations and actions expected from him by other agents in a given situation; the agent asks the question "what would the other agent (the standard for comparison) do if he were in my situation?" and, comparing himself with him, determines his adequacy and plays the corresponding role; can be explained by comparison and the search for strategic advantage: by comparing himself with other agents occupying the same positions in the social system, the agent can introduce or accept innovations that will make him more attractive as an object of relations). It should be noted that with a communicative approach to influence, agents may arrive at similar ideas, but not necessarily similar behavior. In comparison, the agent usually copies the behavior indirectly. Obviously, the behavior of an agent is determined not only by perceptions, but also by the constraints it faces. Therefore, agents with similar views can behave differently, and vice versa, agents with different views can behave in the same way.

The social network plays a large role in the dissemination of information, ideas and influence among its members. Influence in the social media literature is closely related to the term *diffusion of innovations*.

3. Identification of influential agents in the network

A social network can be viewed as a set of agents - potential voters who "vote" for a particular product, service, or candidate from a particular political party in the elections. In this case, the *value* (utility) of an *agent* in a social network depends not only on himself (for example, directly by the expected choice), but also on his influence on other agents. In other words, the configuration and state of the network is important - the totality of the opinions of potential voters regarding their choice. Therefore, there is a need to identify a small number of agents (*the problem of maximizing influence*) that contribute to the formation of the required opinion throughout the network.

The problem of determining the k most influential agents in a social network arose in the context of the so-called *viral marketing* [13,20]. To solve the problem, the market is modeled as a social network of agents (Markov network), the value of each of which is determined not only by the immediate expected profit from the sale (*intrinsic value of customer*), but also by the expected profit from sales to other agents that will be affected by this, from sales to agents which they can influence, etc. (*network value of customer*).

To identify the most valuable (authoritative, influential) agents, the task can be formulated as follows. Let us define the optimal informational influences $IA = \{IA_1, \ldots, IA_n\}$ (IA_i can be both a Boolean variable: 1 - the presence of informational influence, 0 - its absence for the i-th agent; and continuous - the level of influence) for a set of n agents with a predicate $X_i = 1$ if agent i made the required choice and $X_i = 0$ otherwise. Suppose that the choice is described by the following set of attributes: $Y = \{Y_1, \ldots, Y_m\}$. Each agent i has a set of neighbors N_i that directly affect X_i, thereby defining a network of agents. In turn, the i-th agent influences its neighbors.

Let the cost c of the implementation of the information influence per one agent be given, the utility rv_1 from the adoption of the required decision, if the corresponding information influence was exerted on it, and the utility rv_0 from the adoption of the required decision, if the information influence was not carried out. For simplicity, let IA be a Boolean vector.

Let $f_i^1(IA)$ will be the set-result of setting IA_i to 1 (all other values are unchanged), similarly defined for $f_i^0(IA)$. Then the expected increase in utility from the information impact for the agent without taking into account its impact on other agents, i.e., the expected utility from the successful implementation of the information impact (*intrinsic value of customer*) is determined by the formula

$$ELP_i\left(X^k, Y, IA\right) = rv_1 P\left(X_i = 1 \mid X^k, Y, f_i^1(IA)\right) - ,$$
$$- rv_0 P\left(X_i = 1 \mid X^k, Y, f_i^0(IA)\right) - c$$

where X^k – the set of agents whose decisions are known (about whom it is known that they made the required decisions), $P\left(X_i \mid X^k, Y, IA\right)$ – conditional probability of making the required decision by the i-th agent.

Then the expected increase in utility from the information campaign for the selected agents will be

$$ELP\left(X^k, Y, IA\right) = \sum_{i=1}^{n} rv_1 P\left(X_i = 1 \mid X^k, Y, IA\right) - ,$$
$$- \sum_{i=1}^{n} rv_0 P\left(X_i = 1 \mid X^k, Y, IA_0\right) - |IA| c$$

where IA_0 – zero vector; $rv_i = rv_1$, if $IA_i = 1$ (else $rv_i = rv_0$); $|IA|$ – number of selected agents.

The overall value of an agent on the network (*total value of customer = network value of customer + intrinsic value of customer*) will be $ELP\left(X^k, Y, f_i^1(IA)\right) - ELP\left(X^k, Y, f_i^0(IA)\right)$, (i.e., the value of *IA* will change for other agents and may affect their probability of making a decision). Then the agent's *network value* (*network value of customer*) is the difference between his general and personal value (*network value of customer = total value of customer - - intrinsic value of customer*). As can be seen, the value depends on whether the promotions were held for other agents and whether other agents made the required decision.

Let's return to the problem of determining the *k* most influential nodes in a social network. Obviously, in order to find them in this case, you need to find an *IA* that maximizes *ELP*. In the general case, finding the optimal *IA* requires an enumeration of all its possible combinations. The following approximating procedures are possible, giving an approximate solution:

1) A single bypass. For the *i*-th agent there is a special offer

$IA_i = 1$, if $ELP\left(X^k, Y, f_i^1(IA_0)\right) > 0$;

2) Greedy algorithm. Set $IA = IA_0$. It is necessary to bypass IA_i in the loop, setting the value to one, if

$ELP\left(X^k, Y, f_i^1(IA)\right) \geq ELP\left(X^k, Y, IA\right)$;

Hill-climbing search. Set $IA = IA_0$, $IA_{i1} = 1$, where $i_1 = \arg\max_i \left(ELP\left(X^k, Y, f_i^1(IA)\right)\right)$. Repeat as long as the *i*-th agent exists, setting for which $IA_i = 1$ leads to an increase in *ELP*.

4. Maximizing influence in the basic models of the diffusion of innovations

In [10], the problem of influence maximization is considered on the example of the following two basic models of the propagation of innovations: *a linear threshold model and a model of independent cascades*, in which there is an initial set of active agents A_0 and at some moment in time a new active agent gets a chance to activate its neighbors with probability p_{vw}, and the latter, if successful, are activated at the next step, and so on until new activations are possible.

The problem of maximizing influence can be formulated as follows. The influence $\sigma(A)$ of the set of agents *A* is defined as the expected number of active agents upon completion of the process of propagation of information actions initiated by agents from the set *A*. For both models (linear threshold and independent cascades), an *NP*-hard problem arises: for a given parameter *k*, find *k*-elements set *A* maximizing $\sigma(A)$. Since the problem of maximizing the influence is similar to the problem of maximizing submodular functions, then for the appropriate application of the algorithm it is only necessary to prove that $\sigma(A)$ is a submodular function. The submodular function f maps a finite set U to non-negative real numbers and satisfies the natural property of "diminishing returns" (the marginal revenue from adding an element to a set S is at least as high as the marginal revenue from adding the same element to any set including S).

Generalized Threshold Model. An agent's decision to activate is determined by a monotonic threshold function $f_v : S \subseteq N_v \to [0,1]$, where N_v is the set of neighbors *v* and $f_v(\varnothing) = 0$. Each agent initially chooses a threshold θ_v uniformly randomly and becomes active if $f_v(S) \geq 0$.

Generalized cascade model. The probability $p_v(u, S)$ that agent *u* activates agent *v* depends on the set *S* of agents that have already unsuccessfully tried to activate agent *v*. A restriction is imposed on the model: if neighbors $u_1, ..., u_l$ try to activate *v*, then the probability that *v* will become active after *l* attempts does not depend on the order of activation attempts.

Generalized information impact strategies. Let there be *m* different ways of informational influence $I_1, \ldots, I_m$, each of which can affect a certain subset of agents of the social network, increasing their probability of activation. That is, the initial set of active agents A_0 is not defined. The amount of investments x_i in each marketing action is selected, which is limited in aggregate by the budget. Marketing strategy - vector $\boldsymbol{x} = \{x_1, ..., x_m\}$. The probability $h_v(x)$ of agent *v* becoming active is determined by strategy *x*. The function $h_v(\cdot)$ is non-decreasing and has the property of "*diminishing incomes*", that is

$$\forall x \geq y \; \forall a \geq 0 \; h_v(x+a) - h_v(x) \leq h_v(y+a) - h_v(y)$$

The resulting expected number of active agents in this case (taking into account direct marketing and subsequent influence) is equal to

$$EG(x) = \sum_{A \subseteq V} \sigma(A) \prod_{u \in A} h_u(x) \prod_{v \notin A} \left[1 - h_v(x)\right]$$

In order to approximately maximize this functional, it is assumed that can be estimated

$EG(x)$ at each point $\boldsymbol{x}$ and can be found the direction i with an approximately maximum gradient. Let e_i be the unit vector of the i-axis and δ a constant. It is assumed that there exists $\gamma \leq 1$ such that can be found i for which $EG(\boldsymbol{x} + \delta e_i) - EG(\boldsymbol{x}) \geq \gamma(EG(\boldsymbol{x} + \delta e_i) - EG(\boldsymbol{x}))$ for any j. Then, dividing the budget k into parts of size δ, at each step (all of these parts k/δ), we can invest δ funds from the budget into I_i, which maximizes the gradient $EG(\cdot)$.

Competing information influences. In [11], the problem of influence maximization is considered for the case of two competing influences A and B (there are player A and player B) for the model of independent cascades. Accordingly, an agent in the network represented by the graph *G(N, E)* can be in three states: A (reaction to informational action A), B (reaction to informational action B) and C (no decision has been made yet - no response). An agent can move from state C to any other and nothing more. The initial disjoint active sets of nodes are I_A and I_B, respectively ($I_A \cup I_B = I$). The problem of influence maximization is considered for player A. Formally, it is necessary to maximize $f(I_A| I_B)$ - the expected number of agents that will be affected by A for a given I_B by choosing I_A.

Two models extended in relation to the model of independent cascades are proposed:

1) *A model based on distance (distance-based),* in which the agent receives the corresponding innovation from the "closest" activated agent from *I*.

2) *The wave model.* The innovation is spreading step by step. An agent that is not active at the previous step is activated at the current step by uniformly randomly choosing one of the neighbors located at a distance proportional to the number of the step.

For these conditions, it is promising to calculate the Nash equilibrium and consider the Stackelberg game.

Voting model. In [12,14-16], the problem of maximizing influence is considered on the example of a probabilistic *voting model*. In the voting model (belonging to the class of *Interacting Particle Systems models*), at each step, each agent can change his mind, randomly choosing one of the neighbors and accepting his opinion. This model is similar to the threshold model in the sense that the agent is more likely to change his mind to the one supported by the majority of his neighbors. However, in the voting model, in contrast to the threshold model, the agent can become inactive.

The social network is represented by an undirected graph with loops *G(N, E)*. Each node v has many neighbors *N(v)* and is randomly initialized (assigned a value of 1 or 0). At each moment in time, each node randomly chooses one of its neighbors (the probability of choosing each neighbor is the same) and accepts his opinion:

$$f_{i+1}(v) = \begin{cases} 1, \text{ with probability } \dfrac{\left|\{u \in N(v) : f_t(u) = 1\}\right|}{|N(v)|} \\ 0, \text{ with probability } \dfrac{\left|\{u \in N(v) : f_t(u) = 0\}\right|}{|N(v)|} \end{cases}$$

The budget is bounded from above by a constant B, the cost of the initial “persuasion” $f_0(v)=1$ of agent v is c_v. Thus, the problem of maximizing influence is formulated as follows: f_0: $N \rightarrow 0; 1\}$ maximizing the mathematical expectation $E[\sum_{v \in N} f_t(v)]$ for a given budget constraint $\sum_{\{v | f0(v)=1\}} c_v \leq B$.

5. Conclusions

The paper considers the dynamic processes of forming opinions in a social network, and also presents models and algorithms for identifying critical points of a social network (influencing agents), the impact on which allows manipulating the aggregate opinion of network participants forming a social network, as well as the resulting game-theoretic problems information confrontation.

6. References

[1] *Watts D., Dodds P.* Influentials, Networks, and Public Opinion Formation // Journal of Consumer Research. — December, 2007. — P. 123—134.

[2] *Granovetter M.* Threshold Models of Collective Behavior // The American Journal of Sociology. — 1978. — Vol. 83, N 6. — P. 1420—1443.

[3] *Kempe D., Kleinberg J., Tardos E.* Maximizing the Spread of Influence through a Social Network / Proc. of the 9-th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining, Washington, DC, 2003. — P. 137—146.

[4] *Morris S.* Contagion // The Review of Economic Studies. — 2000. — Vol. 67, N 1.

— P. 57—78.

[5] *Goldenberg J., Libai B., Muller E.* Talk of the network: A complex systems look at the underlying process of word-of-mouth // Marketing Letters. — 2001. — N 2. — P. 11—34.

[6] D. A. Gubanov, D. A. Novikov, A. G. Chkhartishvili, "Models of influence in social networks", UBS, 27 (2009), 205–281 (Д. А. Губанов, Д. А. Новиков, А. Г. Чхартишвили, "Модели влияния в социальных сетях", *УБС*, **27** (2009), 205–281).

[7] GRANOVETTER M. *The Strength of Weak Ties* // American Journal of Psychology. – 1973. – №78(6). – P. 1360-1380.

[8] BURT R. S. *Brokerage and Closure*. – Oxford: Oxford University Press, 2005.

[9] DOMINGOS P., RICHARDSON M. *Mining the Network Value of Customers* / Proceedings of the Seventh International Confer- ence on Knowledge Discovery and Data Mining. 2002. P. 57-66.

[10] Oleksandr Laptiev, Volodymyr Tkachev, Oleksii Maystrov, Oleksandr Krasikov, Pavlo Open'ko, Volodimir Khoroshko, Lubomir Parkhuts. The method of spectral analysis of the determination of random digital signals. International Journal of Communication Networks and Information Security (IJCNIS). Vol 13, No 2, August 2021 pp.271-277. ISSN: 2073-607X (Online).

[11] KEMPE D., KLEINBERG J., TARDOS E. *Maximizing the Spread of Influence through a Social Network* / Proceedings of the 9-th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – 2003. – P. 137-146.

[12] Oleg Barabash, Oleksandr Laptiev, Valentyn Sobchuk, Ivanna Salanda, Yulia Melnychuk, Valerii Lishchyna. Comprehensive Methods of Evaluation of Distance Learning System Functioning. International Journal of Computer Network and Information Security (IJCNIS). Vol. 13, No. 3, Jun. 2021. pp.62-71, DOI: 10.5815/ijcnis.2021.03.06.

[13] Laptiev O., Savchenko V., Kotenko A., Akhramovych V., Samosyuk V., Shuklin G., Biehun A. Method of Determining Trust and Protection of Personal Data in Social Networks. International Journal of Communication Networks and Information Security (IJCNIS), Vol. 13, No. 1, 2021. pp.15-21.

[14] Serhii Yevseiev, Oleksandr Laptiev, Sergii Lazarenko, Anna Korchenko, Iryna Manzhul. Modeling the protection of personal data from trust and the amount of information on social networks. Number 1 (2021), «EUREKA: Physics and Engineering» pp.24–31. DOI:10.21303/2461-4262.2021.001615

[15] S. Toliupa, N. Lukova-Chuiko, O. Oksiuk. Choice of Reasonable Variant of Signal and Code Constructions for Multirays Radio Channels. Second International Scientific-Practical Conference Problems of Infocommunications. Science and Technology. IEEE PIC S&T 2015. pp. 269 – 271.

[16] N. Lukova-Chuiko, V. Saiko, V. Nakonechnyi, T. Narytnyk, M. Brailovskyi. Terahertz Range Interconnecting Line For LEO-System. 2020 IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), Lviv-Slavske, Ukraine. 2020. pp. 425– 429.

[17] I. Ruban, V. Martovytskyi & N. Lukova-Chuiko. Approach to Classifying the State of a Network Based on Statistical Parameters for Detecting Anomalies in the Information Structure of a Computing System. Cybernetics and Systems Analysis. Vol. 54, 2018.pp.302–309.

[18] Barabash O., Kopiika O., Zamrii I., Sobchuk V., Musienko A. Fraktal and Differential Properties of the Inversor of Digits of Qs-Representation of Real Number. Modern Mathematics and Mechanics. Springer, Cham, 2019. P. 79 – 95.

[19] CARNES T., NAGARAJAN C., WILD S. M., ZUYLEN A. *Maximizing Influence in a Competitive Social Network: A Follower's Perspective* / Proceedings of the Ninth International Conference on Electronic Commerce. – 2007. – P. 351-360.

[20] EVENDAR E., SHAPIRA A. *A Note on Maximizing the Spread of Influence in Social Networks* / Internet and Network Economics. – 2007. – P. 281-286