

# *Econometric models for evaluation of marketing activities` indicators of social network*

E. B. Gribanova, I. V. Shirenikov, A. V. Katasonova  
Tomsk State University of Control Systems and Radioelectronics  
Tomsk, Russia  
katag@yandex.ru

**Abstract**— In recent years, the analysis of social networks has received considerable attention in the literature because they have a global reach and the large number of data available. For organizations, there are the social networks attractive from the perspective of the promotion of their products and services. And the successful implementation of these activities must take into account a number of factors, existing dependencies in the system and laws for its development. This article is dedicated to study of marketing activities in the social network. The main factors that can influence the result of economic activities are identified. For the study, the authors used data from the groups in the social network VKontakte in the city of Tomsk. For this purpose, the authors calculated network analysis metrics (a clustering factor, a degree, a number of second-level elements) and specific indicators of social groups (activity, popularity). A Java-based program was implemented to collect data and perform group metric calculations. Two regression models have been developed. Also different types of dependencies were studied: linear, hyperbolic, progressive. The first regression model describes the change dependency on the number of members of the group as a result of the competition, based on the number of subscribers and the popularity of the group. The second regression model describes how the value of the advertisement in the group is dependent on the number of subscribers and the indicator of whether the group is the "leader". Both regressions are significant at 99% confidence level. The models can be used in planning and evaluating the results of marketing activities.

**Keywords**—social networks; econometric model; marketing activities; advertising

## I. INTRODUCTION

Social networks today include millions of users exchanging information with each other, publishing personal information in public, and are a promising site for marketing events. Among the benefits of using social networks, there is, first, the possibility to find target audiences, through the information, provided on the user's pages. Second, there are available thematic groups; simpler feedback implementation: participants can share their views on the activities of the organization, leaving comments, and thus they can indicate the development and improvement of products; availability of a wide range of tools to provide promotional information: video clips, audio recordings, and graphics.

Modern researchers [1] note that the use of a social network can significantly increase the sales of firms and the value of the effect depends on the content of the participants' interaction, thus exploring the mechanisms aimed at disseminating information. Identifying the laws of the network system remains an important area of research in the economy. Compared to classical patterns of trade in social networks, the role of other actors and their feedback in the choice of goods or services is increasing. For the organization, this can play both a positive and a negative role: negative information can also be disseminated very quickly. Thus, due to the large amount of information provided in the network and conditionality of the results of activities by the complex and often unpredictable behavior of people and their interaction, there is a need for tools and models to measure and develop strategies for decision-making by economic agents.

Due to the wide spread of social networks and their influence on social, economic and political processes, scientists in many countries are engaged in research in this area, and scientific laboratories have been established. The following main areas of research can be distinguished:

- The study of propaganda methods, information wars, identification of the information distribution centers (for example, the source of gossip [2]).
- Investigation of the causes and signs of dissemination of false information by network participants [3].
- Studying the influence of social networks on people's behavior, their health, economic and political processes [4].
- Development of classification systems for groups and network participants [5], identification of elements belonging to a particular group in the network [6].
- Monitoring of social processes through social networks, and etc.

Among the studies, related to marketing and activities of the organizations, the following can be noted:

One of the research topics focuses on viral marketing. However, there are different definitions of the term [7]. One of them defines viral marketing as the activities of companies to communicate with each other through a network for the

promotion and marketing of products. Also, viral marketing is understood as the transfer of advertising information by word of mouth through the network (online word-of-mouth). However, "leaders of public opinion" can play a significant role in the dissemination of information on goods (both positive and negative). The identification of such leaders in the network is also the subject of research.

The work [8] contains the results of a number of studies in the area of internet marketing. In particular, the following issues are being studied: "What are the best products to promote through social networks?", "At whom should advertising be directed?" etc. The article concludes that social networks are a good way to distribute information about books, restaurants, and movies (word of mouth application), it implies those products that a user can buy as soon as he/she gets information about them from even one participant, and about which people often talk. However, there are products that are selected only after they are accepted by the majority of the surroundings (adoption externalities application): new technologies, software and operating systems, a large number of linkages can be a deterrent and the use of the social network is not a good tool. In the latter case, it is offered to provide several units of production to individual people for free in the hope that this will launch a mechanism for further purchases by their friends. Also, the work concluded that in the case of promotion of the word of mouth application, advertising should be aimed at people, who have a small number of links, because this information is better to receive from a small number of people. And in the case of adoption externalities application, it is better when the information is shared by a person who has a significant number of connections.

A number of studies investigate the use of a targeted price strategy in social networks, when the cost of a product for a consumer depends on how many friends it purchased. Also the studies deal with a question of determining the optimal price. For example, in work [9], the price is proportional to the rate of centrality. Article [10] discusses the use of viral marketing strategies to maximize income. It provides two ways to promote the dissemination of information: returning a part of the purchase cost (cashback) in case a person has recommended the product to his or her friends; offer a discount to customers to encourage them to purchase the product after they have received the recommendations. The size of the discount and the availability of recommendations together with external factors (product quality, market conditions) form the probability of buying on the part of an individual participant in the proposed model. The problem of determining the size of a discount and cashback for a given network structure is being solved.

There are also studies devoted to the investigation of the labor market and the use of social networks and links to find work and transmit information about vacancies. So in work [11], the authors have revealed a high correlation between the employment of a person and the job placement of one's surroundings, and it is disclosed that the probability of finding a job by a person decreases as the period of unemployment

increases. Article [12] concluded that workers with a greater number of connections earn more, and firms engaging new employees earn more through their highly skilled workers.

Social networks can influence economic processes; in this case, it is important to determine the strength and direction of the relationship between phenomena, as well as the construction of predictive models. For example, in work [13], the authors studied the forecasting of behavior of the prices on a stock exchange with the help of the information in Twitter.

This work focuses on the factors influencing the performance of marketing activities, which are seen as an increase in the number of subscribers in the group and the value of advertising in groups. The study used the information of the VKontakte social network, which is currently the most numerous in Russia (its audience at the beginning of 2017 is 88-92 million a week). The data of the groups belonging to the city of Tomsk were received, and inactive groups, in which the administrator made an entry more than a month ago, were not considered, as well as groups where everyone can enter the page. The most significant factors are used to build regression models for estimating marketing activities. Two types of marketing activities are considered: advertising and holding competitions in groups. The developed models will allow estimation of the cost of advertising for a group with a given network structure, as well as forecast of an increase in the number of participants in the group due to the competition.

## II. GROUP AS PART OF THE SOCIAL NETWORK

If the social network is represented as an unfocused graph, the group will be a part bounded by edges with a number of neighboring vertices (Fig. 1), which are called group subscribers or participants. This grouping of participants is facilitated by the existence of similar interests related to the content of the information presented in the group. Vertices-members of a group can also be associated with an edge. This relationship is called friendship. The group administrator can share information with members by placing it on the group-page. Subscribers in turn can leave comments to posts, mark them with likes or post them on their page, doing a repost. There are two ways, in which users of the social network learn about the presence of a group and join it: through friends (for example, by viewing their page); by searching for a group using queries with the specified keywords (in case of interest in a specific subject).

According to the nature of the information provided by the group, they are divided into news groups (e.g. manifesting events in the city) and thematic groups (dealing with cars, handicrafts, etc.). From the point of view of advertising placement, the thematic group is more preferable in view of the possible greater interest of participants in the product / service. On the other hand, news groups provide, as a rule, a wider audience coverage.

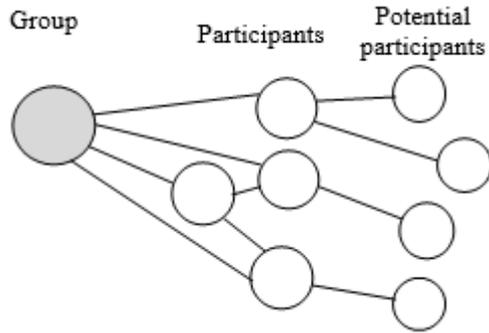


Fig. 1. Group of the social network

As the characteristics of the group, both standard network analysis metrics and specific estimates can be used. Let us consider the basic network analysis metrics that can be used to evaluate groups in terms of marketing activities.

The degree (or number of neighboring vertices) is the number of community members. The more team members there are, the greater audiences one can deliver information to. Also this number makes it possible to judge how many people are interested in this community and its products / services. In Fig. 1, the degree of the group is 4. This indicator can be determined directly by going to the group page. Besides, in VK API there is a method that makes it possible to get this value.

The path from  $i_1$  to  $i_2$  is a series of vertices  $(i_1, i_2, \dots, i_k)$  and ribs  $(i_1i_2, i_2i_3 \dots, i_{k-1}i_k)$ . The length of the path is determined by the number of edges that link the vertices  $i_1$  and  $i_2$ . For group subscribers, the shortest path to the group will be 1 (neighboring vertices, top-level vertices), for friends of non-group subscribers-2 (second-level vertices). The vertices with the shortest path equal to 2 are the potential members of the group who, seeing the information from their friends, can enter it. The manual determination of this indicator is a time-consuming task and, therefore, has been obtained through the programme. The following algorithm was used to calculate the number of potential participants: the group members' id was placed in the object-collection, and then a check was made in the group's subscribers' enumeration cycle. If the participant's id is not in the collection, then the number of vertices of the second level is increased by 1. Fig. 2 shows the dependency of the number of potential members on the group's subscribers. The values were calculated for 93 groups of the city of Tomsk, and groups with a number of participants from 100 to 20000 were selected.

The  $\beta$  clustering factor is the percentage of neighboring vertices, associated with each other. The coefficient is calculated by the formula:

$$\beta = \kappa_1 / \kappa,$$

where  $\kappa$  is the number of possible pairs of neighboring nodes;

$\kappa_1$  is number of pairs of neighboring nodes, connected to each other.

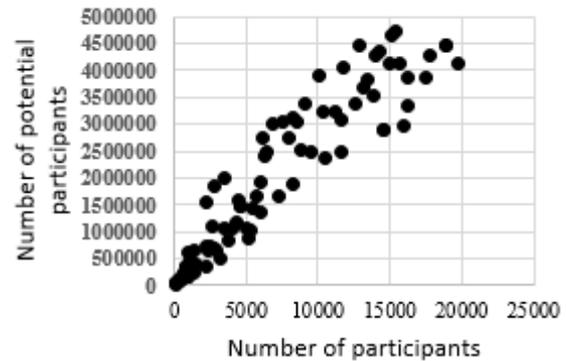


Fig. 2. The dependency of the number of potential members on the group's subscribers

For example, for the structure, presented in Fig. 1, the clustering factor is 0.33.

The clustering factor shows the percentage of team members, who are friends with each other. As the number of subscribers to the group increases, the value of the clustering factor decreases. The highest value of the clustering coefficient (0.3-1), as a rule, occurs in small communities (up to 100 people), which have not only a virtual but also a real expression: student and school groups, groups of employees of firms, religious organizations, events, related to meetings, for example, joint trips, etc. Also, the high value of the clustering factor occurs in the initial stages of the creation of a group, when the participants are primarily the friends of the administrator. Fig. 3 shows the dependency of the number of participants on the clustering factor. It is possible to see that as the number of subscribers increases, the average value of the clustering coefficient decreases, and the spread of values reduces.

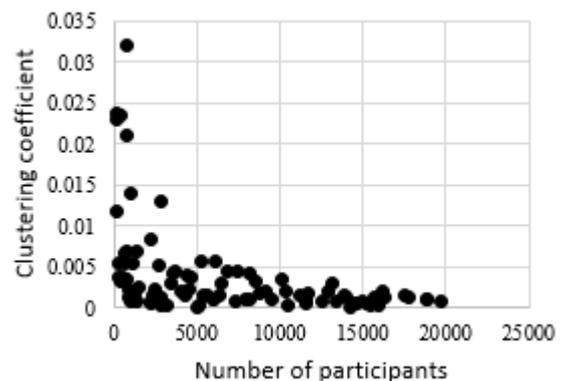


Fig. 3. The dependency of the number of participants on the clustering factor

For programmatic calculation of the clustering coefficient, adjacency matrix  $A$  was formed, with the number of columns and rows equal to the number of team members. The matrix

element is zero ( $a_{ij}=0$ ), if the members that match the row and column of the matrix are not friends with each other and are equal to the unit ( $a_{ij}=1$ ), if they are friends. To calculate the clustering factor, there is the sum of the units below the main diagonal of the matrix (the number of related vertices), which is then divided by the total number of elements under the main diagonal of the matrix  $A$  (the number of possible participant relationships). To generate such matrix, the collection object was first placed with group subscriber information (id and sequence number). Then, for each participant, a search for his friends and checking for their id in the object-collection were performed. In the case the friend is in the group, the corresponding element of the adjacency matrix was assigned the value 1.

The betweenness indicator reflects the connecting role of the group. It shows, how many members are interconnected through a given community without being friends. The formula for calculating the indicator for a group of social networks is:

$$cl = \sum l(j,k),$$

where  $l(j, k)$  is the value that accepts 1, if the vertices of  $j$   $k$  are linked, and 0 - otherwise.

The summation is performed on all pairs of members that are participants of the group. Since this indicator has a high correlation with the clustering factor, it has not been calculated in this study.

The specific characteristics of the social network group include:

- the group activity: the average number of records on the group's wall per day
- the popularity of the group: the sum of likes, reposts and comments divided by the number of records.

The calculation of these indicators was done using the program; this calculations were used in the last 100 records.

In the study of factors, influencing the behavior of participants in the group, endogenous, exogenous and correlated effects are distinguished [14].

Endogenous effects: human behavior is determined by the behavior of people in a group, so the network has the effect of other vertex outcomes on the outcome of the connected with them vertex  $i$ . Such effects are called, depending on the context, as "social norms", "imitation", "epidemic", "influence", etc. For example, the attitude to the information presented in the group may change along with the opinion of other participants.

Exogenous (contextual) effects: the behavior of a group member depends on the group's exogenous characteristics. For example, team members' activity (commenting on records) is determined by the nature of the group itself: whether it was for information only or it implies active discussion.

Correlated effects: some individuals of the group behave in the same way because of similar individual characteristics or

because they belong to the same environment. So, participants-players can actively participate in the group's competitions.

Econometric models are used to represent the dependence, existing in the network between the selected indicators in the form of an equation for the purpose of forecasting or estimating the resultant indicator. The following models of networks are distinguished:

- Network-wide regression is a model, for the construction of which the characteristics of the group are used, i.e. the collection of the necessary data (outcomes and factors) for each community is carried out.
- Node-wide regression is a model, for the construction of which network members' characteristics are used. In this case, data (outcomes and factors) for each participant are used. An example of such regression for solving the problem of network marketing is given in work [15].
- Node-level regression is a model, for the construction of which the data of group members are used, and the influence of these characteristics on some outcome of the group is studied. The classical linear model of the impact of network structure on the community of  $N$  participants is proposed in article [14].

This work will address the network-wide regression.

### III. ECONOMETRIC MODEL OF THE RESULT 'S EVALUATION OF THE COMPETITIVE ACTION

Both deterministic and stochastic methods are used to evaluate the effectiveness of the activities carried out. The essence of them is to measure the amount of a certain value before the event (at the period  $t-t_1$ ) and after it. The same time interval is used. The difference between the measured values provides insight into how much the magnitude of the indicator studied has changed as a result of the event. The growth for a specified period can also be calculated in the formula:

$$\varepsilon = (b-a)/a,$$

where  $b$  is the change in the indicator for the period following the event;

$a$  is a change in the figure for a similar period prior to the event.

This indicator is relative and shows the percentage of change in the property to be examined in relation to the initial value.

An example of a stochastic method is the difference of the difference method, which consists in constructing two regression models. The first model is based on data for the period, preceding the event. In the simplest case, a linear trend can be used:

$$y_1 = \Theta_0 + \Theta_1 x, \quad (1)$$

where  $x$  is a period;

$y_1$  is a forecasting value;

$\Theta_0, \Theta_1$  are parameters of the regression model.

The second model is based on the data for the period, following the event:

$$y_2 = \Theta_2 + \Theta_3 x, \quad (2)$$

Next, the difference between model values, obtained by model (1) and model (2), is defined, and the sum of the received quantities is calculated. The calculated indicator reflects the absolute value of the increase in the property being investigated.

As the object of the study, competitions in social networks were chosen, and the number of subscribers is an investigated value. To participate in such competitions, members of the social network should make a repost of the record by placing it on their page, and after a specified period, random choices are made between such users. The purpose of these events for organizations is to increase the loyalty to the product or service presented, as well as to disseminate information about them. The following indicators can be used to estimate the output of the events: an increase the group's activity and number of users.

Fig. 4 shows the calculation of the effectiveness of one of the competitions; as an explored characteristic, an increase in the number of subscribers of the group is considered.

Here,  $t$  is the day of the competition, and  $d$  is the difference between the model values of the two regressions ( $y_2 - y_1$ ). For the submitted competition, the sum of differences of model values is equal to 1160. This value shows the increase in subscribers as a result of the competition.

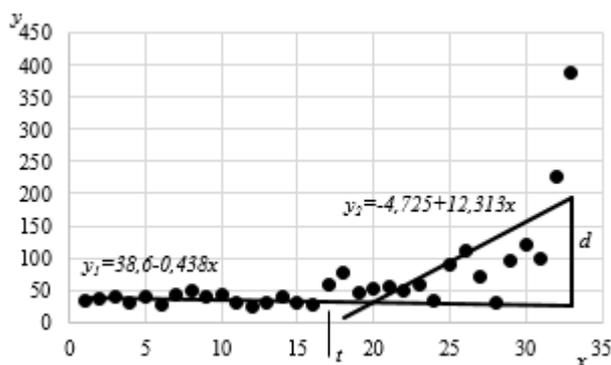


Fig. 4. The calculation of the effectiveness of one of the competition

Models (1) and (2) can be used in cases where historical data already exist. If such data are not available and a new contest is planned, then to assess the increase in subscribers, it is necessary to determine the factors that influence the change in the resulting indicator.

### A. Correlation analysis

Because only a small part of the groups provides access to their statistics, the observation of organized tenders was carried out with data collection at the beginning of the competition and at its end. The following indicators were identified (for 41 competitions):

- number of subscribers at the time of the competition;
- number of subscribers after the competition;
- number of winning competitive positions;
- period of competition;
- prize type: money prize or not;
- popularity of the group;
- group activity;
- clustering factor;
- number of potential participants.

The calculation of the last four indicators was carried out by the programme.

The highest value of the Pearson correlation coefficient was obtained between the following indicators:

- number of subscribers and group activity (0.956);
- number of subscribers and the number of potential participants (0.939);
- number of subscribers and the group's popularity (0.72).

### B. The regression model

To build the regression model as the result variable, the value of the change in the number of subscribers per unit of time (one day) was chosen:

$$\text{change} = (\text{num\_p1} - \text{num\_p0}) / t,$$

where *change* is the change in the number of subscribers per unit of time;

num\_p1 is the number of subscribers after the competition;

num\_p0 is the number of subscribers before the competition;

t is the period of the competition.

Further, a procedure was performed to select factors with the calculation of the adjusted determination index. Different types of dependencies were considered: linear, gradual, logarithmic. For the linear model, the highest value of the determination index was defined. Finally, the resulting regression model is as follows:

$$\text{change} = 7.2956 + 0.001 \text{num\_p0} + 0.0352 \text{popular},$$

where *popular* is a popularity of the group.

Adjusted R-squared presented model is: 0.703, F-statistics 37.89, model is significant at the 99% confidence level, t-value for factors: 8.097; 1.18.

Regression model allows us to determine how current factors can influence on changes of dependent variables. In this case, one can say that the change in the number of subscribers per unit of time depends on: the number of subscribers at the time of the competition, popularity of the group. Fig. 5 shows the fragment of reference sample.

For example, a group that includes 15000 subscribers with a popularity of 60 will increase the number of subscribers per day to 24.636. If the competition period is 20 days, then the total change in the number of subscribers will be 493 people.

This value can be compared to what we receive, if the competition is not held (1).

For some groups, with an equal number of subscribers, there was a significant spread in the values of the increase in participants. This make it possible to conclude, that other factors, for example, characterizing the subject matter of the competition (its value and exclusivity) or the quality of information presentation (humor, spectacular feature) can influence the result. The study of such characteristics will be the subject of further research. In addition, it is planned to take into account a group of users, who only enter the group for the duration of the competition, leaving it out, so that there is an outflow of participants.

Name	Number of participants	Popularity	Increasing ▲
Sneaker place	1802	67,25	12,27
DNS Tomsk	19352	25,39	29,40
Youthful Tomsk	12495	35,30	8,25
I'm from Tomsk	105371	142,44	86,36
News. Tomsk	4293	6,97	9,00
Tomsk.ru	80501	96,75	97,86

Fig. 5. Fragment of reference sample of data

#### IV. ECONOMETRIC MODEL FOR ESTIMATING THE VALUE OF ADVERTISEMENT

Firms may advertise in existing groups to promote their products and sales growth [16]. As a rule, the placement of such information is a paid service. The study examines factors influencing the cost of advertising. The effects of the following factors on the resulting indicator were investigated: the number of subscribers of the group; availability of topics (1-topic group, 0-news group); group activity; popularity of the group; the time, when the advertisement is in the first place; the total time of placement of the post in the group. To collect data on the cost of advertising in groups in the city of Tomsk, the site sociate.ru was used; requests to the administrators of the groups were made. In total, 45 groups were selected that satisfy the specified requirements.

As a result of the procedure for sequential selection of factors and consideration of various types of dependencies, the obtained regression model is (Fig. 6):

$$\text{cost} = -776.144 + 110.213 \cdot \ln(\text{num\_p}),$$

where *cost* is the cost of advertising;

*num\_p* is the number of subscribers.

Adjusted R-squared presented model: 0.519, F-statistics 46.35, model is significant at the 99% confidence level, t-value for factor: 6.8.

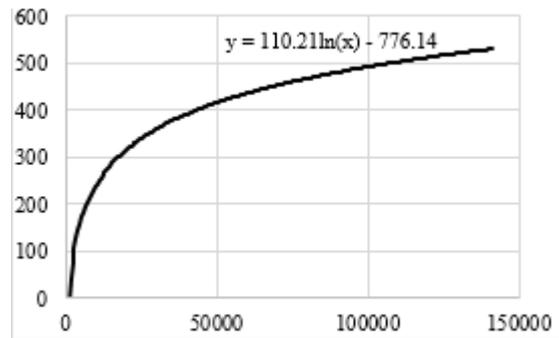


Fig. 6. The logarithmic regression model

The status of the group, i.e. whether it is a leader, was also studied (1=yes, 0=no). It was determined that the group is a leader in the following cases: there are no other groups dedicated to this topic; if there are several groups of competitors, then at the search query it appears among the first.

The resulting model is as follows:

$$\text{cost} = -724.704 + 174.892 \cdot \text{leader} + 92.874 \cdot \ln(\text{num\_p}),$$

Adjusted R-squared presented model: 0.705, F-statistics 50.09, model is significant at the 99% confidence level, t-value for factors: 5.14; 6.99.

#### V. CONCLUSION

The research of marketing activities in the social network VKontakte in the groups of the city of Tomsk was carried out. The selection of groups and activities meeting the specified requirements was carried out; the metrics of the network analysis were calculated and the factors influencing the resulting indicators were identified. The research work presents two regression models: to assess the change in the number of subscribers as a result of a competitive event and to assess the cost of placing advertisements in a group. The most important indicator in two cases was the number of subscribers. The presented models can be used to make the decision by the economic agents, by planning the marketing activities.

## References

- [1] S. Goyal, J. Gagnon, "Social networks and the firm", *Revista de Administracao*, vol. 51, pp. 240-243, 2016.
- [2] A. Banerjee, A. Chandrasekhar, E. Duflo, M. Jackson, "Gossip: Identifying Central Individuals in Social Network", *The National bureau of economic research*, vol. 20422, pp. 1-30, 2014.
- [3] M. Kleek, D. Murray-Rust, A. Guy, D. Smith, K. O'Hara, N. Shadbolt, "Self curatio, social partitioning, escaping from prejudice and harassment: the many dimensions of lying online", *Proceedings of the ACM web science conference*, pp. 1-9, June 2015.
- [4] T. Althoff, P. Jindal, J. Leskovec, "Online actions with offline impact: how online social networks influence online and offline user behavior", *Proc. of the tenth ACM international conference on web search and data mining*, pp. 537-546, February 2017.
- [5] I. Hemalatha, G. Saradhi Varma, A. Govardhan, "Social networks analysis and mining using machine learning techniques", *AETS*, pp. 603-608, 2013.
- [6] S. Kumar, J. Cheng, J. Leskovec, V. Subrahmanian, "An army of me: Sockpuppets in online discussion communities", *26<sup>th</sup> International world wide web conference*, pp. 1-10, April 2017.
- [7] M. Petrescu, P. Korgaonkar, "Viral advertising: definitional review and synthesis", *Journal of internet commerce*, vol. 10, pp. 208-226, 2011.
- [8] A. Galeotti, S. Goyal, "Influencing the influencers: a theory of strategic diffusion", *Rand Journal of economics*, vol. 40, pp. 509-532, 2009.
- [9] F. Bloch, N. Querou, "Pricing in social network", *Games and Economic behavior*, vol. 80, pp. 263-281, 2013.
- [10] A. David, M. Rajeev, S. Aneesh, X. Ying, "Pricing strategies for viral marketing on social networks", *Internet and Network economic*, vol. 5929, pp. 101-112, 2009.
- [11] A. Calvo-Armengol, M. Jackson, "The effects of social networks on employment and inequality", *The American economic review*, vol. 94, pp. 426-454, 2004.
- [12] J. Montgomery, "Social networks and and labor-market outcomes: toward an economic analysis", *The American conomic review*, vol. 81, pp. 1408-1418, 1991.
- [13] J. Bollen and H. Mao, "Twitter mood as a stock market predictor", *Computer*, vol. 44, pp. 91-94, 2011.
- [14] C. Manski, "Identification of endogenous social effects: the reflection problem", *The review of economic studies*, vol. 60, pp. 531-542, 1993.
- [15] M. Grigoryeva, E. Griбанова, T. Kust, "Research activities of students as a method to support interdisciplinary teaching in training process of technical university", *III International scientific conference on information technologies in science, management, social sphere and medicine*, pp. 270-275, May 2016.
- [16] S. Goyal, "Advertising on social media", *Scientific journal of pure and applied sciences*, vol. 5, pp. 220-223, 2013.