

УДК 373.1

DOI: 10.25559/SITITO.16.202001.171-176

Сетевой анализ распространения информации в группе школьников

А. А. Алтухов, Д. С. Васянина, Е. Д. Ветрова, С. А. Тищенко*, О. А. Клачкова

Московский государственный университет имени М. В. Ломоносова, г. Москва, Россия
119991, Россия, г. Москва, ГСП-1, Ленинские горы, д. 1

* tichtch@mail.ru

Аннотация

В данной публикации представлены результаты исследования скорости распространения информации в группе школьников. Исследование выполнено методом сетевого анализа. Цель исследования – моделирование и анализ социальной структуры сообщества школьников и скорости распространения информации. Также ставились задачи получения ответов на следующие вопросы: насколько отличаются сетевые характеристики группы школьников от случайного графа; подчиняется ли устройство сети дружеских взаимосвязей внутри коллектива общим сетевым законам; каковы ключевые отличия распространения информации в графе, построенном на примере группы школьников, от распространения ее в случайном графе. Случайный граф с аналогичными базовыми параметрами был специально построен в целях данного исследования. Работа предваряется развернутым анализом исследований в области сетевого анализа, социальных моделей и поведения людей. Далее представлены эмпирические данные, которые собраны специально в целях проведения данного исследования. Объяснено, как и для чего нами получены данные сведения. В основной части исследования отражено то, что на базе собранных эмпирических данных, нами построены графы для изучения основных характеристик рассматриваемого сообщества школьников. Анализ включает в себя определение наличия кластеров, расчет и анализ центральности графа группы школьников и его сходства со случайным графом. На основании проведенного эмпирического исследования анализ результатов показал, что информация лучше распространяется (за меньшее число итераций) именно в условиях реальной группы школьников. Напротив, в случайной сети связи распределены более равномерно, что замедляет в ней скорость распространения информации. Исследование снабжено необходимыми пояснениями, а также иллюстрирующими графиками и таблицами.

Ключевые слова: взаимосвязи, графы, информация, моделирование, сетевой анализ, сеть, распространение, скорость, социальная структура, школьники.

Для цитирования: Алтухов, А. А. Сетевой анализ распространения информации в группе школьников / А. А. Алтухов, Д. С. Васянина, Е. Д. Ветрова, С. А. Тищенко, О. А. Клачкова. – DOI 10.25559/SITITO.16.202001.171-176 // Современные информационные технологии и ИТ-образование. – 2020. – Т. 16, № 1. – С. 171-176.

© Алтухов А. А., Васянина Д. С., Ветрова Е. Д., Тищенко С. А., Клачкова О. А., 2020



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.



Network Analysis of the Information's Dissemination in a Group of Schoolchildren

A. A. Altoukhov, D. S. Vasianina, E. D. Vetrova, S. A. Tichtchenko*, Klachkova O. A.

Lomonosov Moscow State University, Moscow, Russia

1, Leninskie gory, Moscow 119991, Russia

* tichtch@mail.ru

Abstract

This publication presents the results of a study about the speed of dissemination of information in a group of schoolchildren. The study was performed by network analysis. The purpose of the study is modeling and analysis of the social structure of the community of schoolchildren and the speed of dissemination of information. The objectives were also to obtain answers to the following questions: how different are the network characteristics of a group of students from a random graph; does the network device comply with the friendly relationships within the team with the general network laws; what are the key differences between the dissemination of information in a graph based on the example of a group of schoolchildren from its distribution in a random graph. A random graph with similar basic parameters was specially built for the purpose of this study. The work is preceded by a detailed analysis of researches in the fields of network analysis, social models and human behavior. The following are empirical data that are specifically collected for this study. It is explained how and why we received this information. The main part of the study reflects the fact that basing on the collected empirical data, we have constructed graphs for studying the main characteristics of the considered schoolchildren community. The analysis includes determining the presence of clusters, calculating and analyzing the centrality of the graph of a group of schoolchildren and its similarity to a random graph. Based on an empirical study analysis of the results showed that information is better disseminated (for fewer iterations) precisely in the context of a real group of schoolchildren. In contrast, in a random network, communications are more evenly distributed, which slows down the speed of information's dissemination. The study is provided with the necessary explanations, as well as illustrative graphs and tables.

Keywords: interconnections, graphs, information, modeling, network analysis, network, distribution, speed, social structure, schoolchildren.

For citation: Altoukhov A.A., Vasianina D.S., Vetrova E.D., Tichtchenko S.A., Klachkova O.A. Network Analysis of the Information's Dissemination in a Group of Schoolchildren. *Sovremennye informacionnye tehnologii i IT-obrazovanie* = Modern Information Technologies and IT-Education. 2020; 16(1):171-176. DOI: <https://doi.org/10.25559/SITITO.16.202001.171-176>



Введение

Решения, которые люди принимают в жизни, и те действия, которые они совершают, во многом зависят от влияния и поступков их родственников, друзей и знакомых. Это верно в отношении как повседневных, так и принципиальных решений, определяющих жизненный путь человека. Поэтому научно-практический интерес представляет моделирование того, как социальная структура (окружение и общество) влияет на принятие экономических решений и на распространение информации, и того как все это влияет на поведение индивидов. В рамках нашего исследования проведен обзор актуальных работ, в которых изучалась роль социальной структуры в диффузии в различных сферах, и осуществлено моделирование скорости распространения информации на базе реальных данных – по результатам изучения социальных контактов школьников, которые принимали участие в Выездной конференции экономико-математической школы 2019 г.

Цель исследования

Целью исследования является моделирование и анализ социальной структуры сообщества школьников и скорости распространения информации в ней для получения ответов на следующие вопросы:

- насколько отличаются характеристики сети (социальной структуры) школьников из выбранного нами сообщества от случайного графа;
- подчиняется ли устройство сети дружеских взаимосвязей внутри выбранного нами коллектива общим сетевым законам;
- каковы отличия распространения информации в графе, построенном на примере выбранного нами коллектива школьников, от распространения ее в случайном графе.

Обзор литературы

В научной литературе существует большое количество работ, которые изучают и моделируют диффузию распространения различных материальных и нематериальных объектов. Изучалось, например, распространение принятия новых технологий и продуктов, и на примере исследования распространения новой технологии видео-сообщений в банке было установлено, что если новую технологию использует формальный начальник или неформальный лидер, то достигается наибольший эффект [1]. Другой пример – моделирование диффузии политических идей, которое показало большую вероятность того, что «соседи» по графу имеют схожее политическое мировоззрение [2]. А моделирование финансовой неустойчивости показало, что вероятность банкротства некоторой компании зависит от степени ее интеграции (того, какая доля активов компании находится в собственности других компаний) и от степени ее диверсификации (того, в собственности скольких других компаний находятся активы этой компании) [3]. Существуют исследования, которые показывают, как структура социальных взаимодействий помогает объяснить различия между городами США по уровню преступности [4]; или то, как сеть романтических связей в старшей школе связана, например, с распространением болезней, передаваемых половым путем [5]. Эти и другие аналогичные работы представляют

нам важные доказательства корреляции поведения индивидов с поведением связанных с ними людьми, а в случае заболеваний (учитывая природу заражения) предоставляют даже некоторые прямые доказательства диффузии.

Существуют классические исследования, которые показывают важность социальных связей для распространения информации, например, о вакансиях [6, 7] или показывают то, что даже случайные знакомства или непостоянные контакты («слабые связи») индивидов могут быть важны для распространения информации [8, 9, 10]. Некоторые другие работы идентифицируют местные корреляции в статусе занятости (в окрестностях Чикаго), изучая степень, в которой местные взаимодействия важны в сфере занятости [11, 12, 13]. Существует даже публикация, посвященная изучению связи между формированием сети дружеских отношений между сборщиками фруктов и эффективностью различных трудовых договоров [14].

Вышеперечисленные источники, которые опираются на значимый объем эмпирических исследований, указывают на важную роль социальных связей при передаче информации, на различия между типами социальных контактов, а также показывают, что даже слабые социальные связи могут быть важными.

Особенно значимой для нашего исследования является публикация, авторы которой отвечают на вопросы о том, как распространять полезную информацию в сети, чтобы наибольшее число людей смогло ей воспользоваться, и о том, как влияют на возможное распространение нового продукта сетевые позиции индивидов в обществе, которые первыми получают информацию об этом новом продукте [15]. Для этого авторами той работы была построена модель распространения информации, в которой использовались следующие обоснованные предположения:

- на скорость распространения информации влияют сетевые характеристики первого индивида, который получает информацию;
- индивиды должны знать о продукте, прежде чем смогут им воспользоваться, и они могут узнавать о продукте от своих друзей;
- индивид более вероятно узнает о продукте, если большее число его друзей может ему о нем рассказать;
- решение об использовании продукта информированным индивидом зависит от аналогичных решений его друзей.

В указанном исследовании [15] модель построена на примере распространения микрофинансовых займов в условиях, когда известно множество первоначальных потенциально знающих индивидов. За шесть пол года до того, как кредитное учреждение пришло в индийские деревни и стало предлагать сельским жителям микрофинансовые займы, авторы исследования [15] собрали подробные сетевые данные путем опроса домохозяйств о широком круге людей, с которыми у них существуют те или иные взаимоотношения. Внедрение микрофинансирования началось с приглашения «лидеров» (например, владельцев магазинов и учителей) на информационное собрание, а затем экспериментаторы попросили этих «лидеров» распространять информацию о кредитах. Используя данные местоположения в сети этих первых информированных жителей (или «точек инъекции») и данные о последующем участии жителей в использовании микрофинансовых займов, авторы оценили параметры построенной модели, используя метод моментов [15].



В результате были получены следующие важные выводы:

- участник микрофинансирования в семь раз чаще информирует другое лицо (домохозяйство), чем тот, кто не участвует в этом;
- однако информация, передаваемая теми, кто не участвует, также важна, так как обеспечивает около трети возможной информированности и участия (поскольку тех, кто не участвует, гораздо больше) [15].

Как только учтена передача информации, решение информированного субъекта (домохозяйства) об участии не зависит существенно от количества соседей. Централизованность коммуникации, когда она применяется к группе первых информированных индивидов (сельских жителей, домохозяйств), существенно превосходит в данном контексте другие стандартные сетевые показатели центральности в прогнозировании участия в микрофинансировании. Эмпирически, показатель «communication centrality» значимо и положительно влияет на долю домохозяйств, ставших участниками микрофинансирования. Корреляция с остальными мерами центральности не является значимой.

Методология исследования

Исследование проведено с помощью метода сетевого анализа, а также методов анализа и синтеза собранной информации и полученных данных.

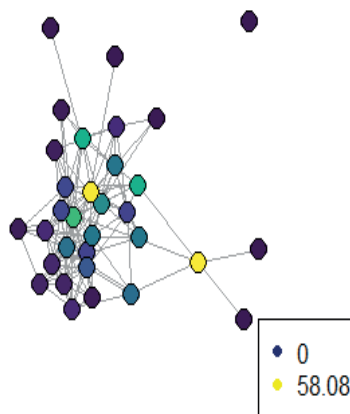
Эмпирические данные исследования

В целях проведения исследования нами были собраны данные о школьниках, которые принимали участие в Выездной школе ЭМШ весной 2019 г. (далее – Выездная школа). Участники мероприятия до этого посещали вечерние курсы, поэтому некоторые из них могли быть знакомы между собой, хотя знакомство не гарантировано, так как они могли посещать разные курсы или быть в разных группах. Данные собраны посредством заполнения анкеты в первый день выездного мероприятия, участникам которого было предложено из списка всех школьников выбрать тех, с кем они лично знакомы. По этим полученным данным была построена матрица смежности без весов, и в которой не учитывалось направление взаимодействия. Мы предположили, что если один школьник знает другого лично, то и второй с ним знаком, и таким образом мы получили неориентированный граф. Для дальнейшего сравнения нами был сконструирован случайный граф с таким же числом точек (32 участника) и тем же суммарным числом связей между ними. В качестве модели распространения информации использована модель Де Грута, с предположением того, что если один участник узнает от другого информацию, например, об отмене занятия, то он ей верит. Учитывая вышеизложенное, далее в публикации мы представляем итоги сравнительного анализа двух графов: основанного на реальных взаимодействиях школьников и случайного.

Основная часть исследования

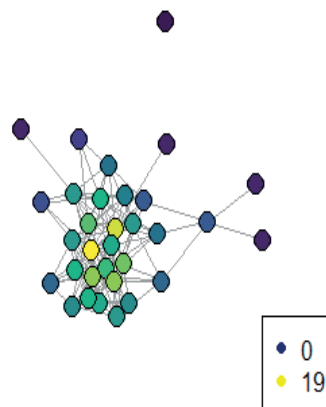
Используя собранные эмпирические данные, нами построено несколько графов с целью изучения основных характеристик рассматриваемого сообщества школьников, включая определение наличия кластеров, центральности графа и его сходство со случайным графом.

На рисунках 1-3 ниже представлены разные варианты визуализации графа знакомств рассматриваемой группы школьников. Показано, какие точки стали более центральными в зависимости от способа измерения центральности: по кратчайшему пути (см. рис.1), по степени вершины (см. рис.2), по собственному значению (см. рис.3). Можно заметить, что в случае измерения по кратчайшему пути, есть только две точки, которые значительно центральнее остальных, но при измерении центральности другими способами выделяется группа школьников, которые занимают относительно центральные позиции. Это значит, что у этих школьников довольно много знакомств, которые не критичны с точки зрения скорости передачи информации. Центральность по кратчайшему пути составляет 0,09, что говорит о том, что в среднем через каждого школьника проходит 0,09 кратчайших путей знакомств двух других участников рассматриваемого коллектива. Центральность графа по степени вершины 0,33 означает, что средняя степень вершины для графа составляет обратную величину – 3,03.



Р и с. 1. Граф связей школьников на Выездной школе, центральность по кратчайшему пути («betweenness centrality»). Составлено авторами.

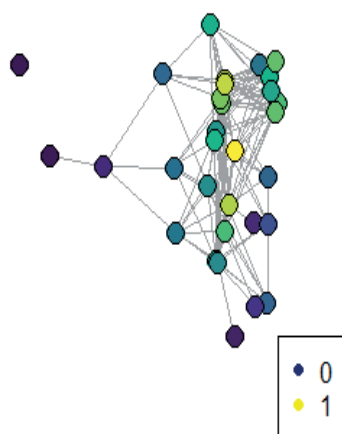
F i g. 1. The graph of connections of schoolchildren at the traveling school, the centrality along the shortest path ("betweenness centrality"). Compiled by the authors.



Р и с. 2. Граф связей школьников на Выездной школе, центральность по степени вершины («closeness centrality»). Составлено авторами.

F i g. 2. Graph of connections of schoolchildren at the traveling school, "closeness centrality". Compiled by the authors.





Р и с. 3. Граф связей школьников на Выездной школе, центральность по собственному значению («eigenvector centrality»). Составлено авторами.

Fig. 3. The graph of connections of schoolchildren at the traveling school, eigenvector centrality. Compiled by the authors.

Чтобы определить то, насколько случаен граф, построенный по связям школьников, его следует сравнить с аналогом: для этого нами построен случайный граф с таким же количеством вершин и ребер. Если бы школьников случайным образом отобрали для участия в Выездной школе, то их знакомства подчинялись бы закону нормального распределения. Результаты сравнения двух графов представлены ниже (см. табл. 1).

Т а б л и ц а 1. Сравнительные характеристики графа связей школьников и аналогичного случайного графа

Table 1. Comparative characteristics of the graph of connections between schoolchildren and a similar random graph

| | Граф связей школьников | Случайный граф |
|------------------------|------------------------|----------------|
| Число вершин | 32 | 32 |
| Число связей | 141 | 141 |
| Closeness centrality | 0.33 | 0.19 |
| Betweenness centrality | 0.09 | 0.06 |
| Eigenvector centrality | 0.56 | 0.41 |

Составлено авторами по данным исследования.

По всем трем показателям центральности граф, полученный на данных о рассматриваемой группе школьников, более централен, чем случайный граф с тем же количеством вершин и связей. Значит, каждый школьник из группы Выездной школы в среднем «ближе» к остальным, чем вершины в случайном графе, и в среднем через каждого проходит больше кратчайших путей. Это происходит из-за неравномерного распределения знакомств: так, например, в случайном графе не возникает вершин, не связанных с остальными, а в графе, который основан на собранных нами данных, такая вершина есть, равно как и вершины с единственным знакомством. Максимальная степень вершины случайного графа – 15, в то время как среди реальных школьников максимальное число связей – 20. Проведенные нами тесты показывают, что распределение степеней вершин графа по данным о школьникам статистически значимо отличается от нормального. Таким образом, на небольшой выборке школьников видно, что знакомства между ними происходят не в произвольном порядке. Различие в плотности связей реальных школьников и точек

случайного графа можно увидеть, выполнив кластеризацию. В графе, построенном по данным выбранной группы школьников, выделяются пять кластеров – тесных кругов общения – один из которых состоит из единственного участника. А в случайном графе существует только один единый кластер, при этом все его вершины связаны между собой и равномерно соединены. Далее мы проверили то, насколько отсутствие случайности в знакомствах школьников влияет на распространение информации внутри их коллектива. Иными словами – насколько быстро распространится та или иная информация, если сообщить ее школьнику с максимальным числом связей или в случайном графе. Исследование показало, что если сообщить некоторую значимую новость самому общительному школьнику, то через три шага эту информацию узнают все 32 человека из группы; а в случайном графе потребуется на один шаг больше.

Полученные результаты

Проведенное нами эмпирическое исследование показало, что информация лучше распространяется в реальных условиях (в реальной группе школьников), чем в случайном графе. При сообщении новости наиболее общительному школьнику, за три итерации даже наименее общительные участники группы, которые имеют всего одно знакомство, получают эту информацию.

Заключение

По результатам проведенных нами изысканий с использованием метода сетевого анализа показано, что центральность графа, который построен на реальных данных о группе школьников, оказалась выше, чем у случайного графа, по всем трем использованным метрикам. Из-за того, что в реальной жизни не все школьники одинаково общительны, возникает смещение, которое способствует более быстрому распространению информации в этом коллективе. Благодаря этому, если сообщить некоторую значимую информацию одному наиболее общительному школьнику из рассматриваемой группы (который знаком с наибольшим числом других школьников), то уже за три коммуникации эту информацию получают даже наименее общительные члены группы, которые имеют всего одно знакомство.

В противовес реальной группе, в случайной сети связи распределены более равномерно, т.е. они статистически отличаются от распределения реальных связей, и именно это замедляет скорость распространения информации, даже в том случае, если она попала к самому «общительному» агенту.

В целом же можно констатировать, что по результатам проведенных изысканий и анализа была достигнута цель исследования и найдены ответы на сформулированные в начале вопросы.

Исследование может быть продолжено на примере более крупных, различных по составу групп. Это позволит либо укрепить полученные выводы, либо обоснованно скорректировать их.

References

- [1] Tucker C.E. Identifying Formal and Informal Influence in Technology Adoption with Network Externalities. *Management Science*. 2008; 54(12):2024-2038. (In Eng.) DOI: <https://doi.org/10.1287/mnsc.1080.0897>



- [2] Huckfeldt R.R., Sprague J. Citizens, Politics and Social Communication: Information and Influence in an Election Campaign. Cambridge University Press; 1995. (In Eng.)
- [3] Elliott M., Golub B., Jackson M.O. Financial Networks and Contagion. *American Economic Review*. 2014; 104(10):3115-53. (In Eng.) DOI: <https://doi.org/10.1257/aer.104.10.3115>
- [4] Glaeser E.L., Sacerdote B., Scheinkman J.A. Crime and Social Interactions. *The Quarterly Journal of Economics*. 1996; 111(2):507-548. (In Eng.) DOI: <https://doi.org/10.2307/2946686>
- [5] Bearman P.S., Moody J., Stovel K., Thalji L. Social and sexual networks: The national longitudinal study of adolescent health. In: Morris M. (ed.) *Network Epidemiology: A Handbook for Survey Design and Data Collection*. New York, NY: Oxford University Press; 2004. p. 201-237. (In Eng.)
- [6] Rees A. Information Networks in Labor Markets. *American Economic Review*. 1966; 56(1/2):559-566. Available at: <https://www.jstor.org/stable/pdf/1821319.pdf> (accessed 27.01.2020). (In Eng.)
- [7] Rees A., Schultz G.P. Workers and Wages in an Urban Labor Market. Chicago: University of Chicago Press; 1970. (In Eng.)
- [8] Granovetter M.S. The Strength of Weak Ties. *American Journal of Sociology*. 1973; 78(6):1360-1380. Available at: https://sociology.stanford.edu/sites/g/files/sbiybj9501/f/publications/the_strength_of_weak_ties_and_exch_w-gans.pdf (accessed 27.01.2020). (In Eng.)
- [9] Granovetter M.S. Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology*. 1985; 91(3):481-510. Available at: https://www.jstor.org/stable/2780199?seq=1#metadata_info_tab_contents (accessed 27.01.2020). (In Eng.)
- [10] Granovetter M.S. Getting a Job: A Study of Contract and Careers. Chicago: University of Chicago Press; 1995. (In Eng.)
- [11] Topa G. Social Interactions, Local Spillovers and Unemployment. *Review of Economic Studies*. 2001; 68(2):261-295. (In Eng.) DOI: <https://doi.org/10.1111/1467-937X.00169>
- [12] Conley T.G., Topa G. Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Econometrics*. 2002; 17(4):303-327. (In Eng.) DOI: <https://doi.org/10.1002/jae.670>
- [13] Bayer P., Ross S.L. Identifying Individual and Group Effects in the Presence of Sorting: A Neighborhood Effects Application. *NBER Working Paper*. 2006; 12211. (In Eng.) DOI: <https://doi.org/10.3386/w12211>
- [14] Bandiera O., Barankay I., Rasul I. The Evolution of Cooperative Norms: Evidence from a Natural Field Experiment. *The B.E. Journal of Economic Analysis & Policy*. 2006; 6(2). (In Eng.) DOI: <https://doi.org/10.2202/1538-0637.1484>
- [15] Banerjee A., Chandrasekhar A.G., Duflo E., Jackson M.O. The Diffusion of Microfinance. *Science*. 2013; 341(6144):1236498. (In Eng.) DOI: <https://doi.org/10.1126/science.1236498>

Поступила 27.01.2020; принята к публикации 20.04.2020;
опубликована онлайн 25.05.2020.
Submitted 27.01.2020; revised 20.04.2020;
published online 25.05.2020.

Об авторах:

Алтухов Алексей Валерьевич, инженер, аспирант кафедры экономики инноваций, экономического факультет, Московский государственный университет имени М. В. Ломоносова (119991, Россия, г. Москва, ГСП-1, Ленинские горы, д. 1), ORCID: <http://orcid.org/0000-0002-4757-0401>, alexei.altoukhov@gmail.com

Васянина Дарья Сергеевна, магистрант экономического факультета, Московский государственный университет имени М. В. Ломоносова (119991, Россия, г. Москва, ГСП-1, Ленинские горы, д. 1), ORCID: <http://orcid.org/0000-0001-8148-2563>, vasyanina@econ.msu.ru

Ветрова Екатерина Дмитриевна, магистрант экономического факультета, Московский государственный университет имени М. В. Ломоносова (119991, Россия, г. Москва, ГСП-1, Ленинские горы, д. 1), ORCID: <http://orcid.org/0000-0001-5903-2099>, vetrova@econ.msu.ru

Тищенко Сергей Александрович, доцент кафедры экономической информатики, экономического факультет, Московский государственный университет имени М. В. Ломоносова (119991, Россия, г. Москва, ГСП-1, Ленинские горы, д. 1), кандидат физико-математических наук, ORCID: <http://orcid.org/0000-0002-6410-068X>, tichtch@mail.ru

Клачкова Ольга Александровна, ассистент кафедры математических методов анализа экономики, экономического факультет, Московский государственный университет имени М. В. Ломоносова (119991, Россия, г. Москва, ГСП-1, Ленинские горы, д. 1), кандидат экономических наук, ORCID: <http://orcid.org/0000-0002-5300-7930>, sparrow889@gmail.com

Все авторы прочитали и одобрили окончательный вариант рукописи.

About the authors:

Alexei V. Altoukhov, Engineer and PhD student of "Innovation Economics" Department, Faculty of Economics, Lomonosov Moscow State University (1, Leninskie gory, Moscow 119991, Russia), ORCID: <http://orcid.org/0000-0002-4757-0401>, alexei.altoukhov@gmail.com

Darya S. Vasianina, Master of the Faculty of Economics, Lomonosov Moscow State University (1, Leninskie gory, Moscow 119991, Russia), ORCID: <http://orcid.org/0000-0001-8148-2563>, vasyanina@econ.msu.ru

Ekaterina D. Vetrova, Master of the Faculty of Economics, Lomonosov Moscow State University (1, Leninskie gory, Moscow 119991, Russia), ORCID: <http://orcid.org/0000-0001-5903-2099>, vetrova@econ.msu.ru

Sergey A. Tichtchenko, Associate Professor of the Department of Economic Informatics, Faculty of Economics, Lomonosov Moscow State University (1, Leninskie gory, Moscow 119991, Russia), Ph.D. (Phys.-Math.), ORCID: <http://orcid.org/0000-0002-6410-068X>, tichtch@mail.ru

Olga A. Klachkova, Assistant of the Department of Mathematical Methods of Economic Analysis, Faculty of Economics, Lomonosov Moscow State University (1, Leninskie gory, Moscow 119991, Russia), Ph.D. (Economy), ORCID: <http://orcid.org/0000-0002-5300-7930>, sparrow889@gmail.com

All authors have read and approved the final manuscript.

