

УДК 378:338.2

DOI: 10.28995/2073-6304-2018-3-9-22

Сравнительный анализ деятельности вузов на основании данных из социальных сетей

Юлия Ю. Шитова

*Российский государственный гуманитарный университет,
Москва, Россия, yu_shitova@mail.ru*

Юрий Ю. Шитов

*Объединенный институт ядерных исследований,
Россия, Московская обл., Дубна, shitov@jinr.ru*

Максим С. Кокорев

*Университет «Дубна»,
Россия, Московская обл., Дубна, kokms.11@gmail.com*

Аннотация. В работе проведен сравнительный анализ социально-экономической деятельности университета «Дубна» (7000 человек) и РГГУ (38 500 человек) на основании уникальных данных, извлеченных из сети ВКонтакте при помощи оригинального программного обеспечения. Исследованы структура и временная динамика ряда показателей: демографии обучающихся; географии приезда студентов, распределения по направлениям обучения в гендерном разрезе. Обсуждаются различия в развитии регионального и московского вузов, выявленные в ходе анализа. В заключении раскрыты перспективы развития методики.

Ключевые слова: социально-экономическая деятельность вуза, BigData, социальные сети, онлайн-данные, сравнительный анализ, демография студентов, география студентов

Для цитирования: Шитова Ю.Ю., Шитов Ю.Ю., Кокорев М.С. Сравнительный анализ деятельности вузов на основании данных из социальных сетей // Вестник РГГУ. Серия «Экономика. Управление. Право». 2018. № 3 (13). С. 9–22. DOI: 10.28995/2073-6304-2018-3-9-22

Comparative analysis of the activity of higher education institutions (HEI) on the basis of data from social networks

Yulia Yu. Shitova

Russian State University for the Humanities, Moscow, Russia, yu_shitova@mail.ru

Yuri Yu. Shitov

Joint Institute for Nuclear Research, Russia, Moscow Region, Dubna, shitov@jinr.ru

Maxim S. Kokorev

University of Dubna, Dubna, Moscow Region, Russia, kokms.11@gmail.com

Abstract. In this paper, a comparative analysis of the socio-economic activities of the University of Dubna (7000 people) and the Russian State University for the Humanities (38500) was carried out on the basis of unique data extracted from the VKontakte network with the help of original software. The structure and temporal dynamics of a number of indicators are studied: the students demographics; the geography of the students arrival, the distribution of studying directions in terms of gender. Differences in the development of regional and Moscow HEI revealed during the analysis are discussed. In conclusion, the prospects for the development of the methodology are disclosed.

Keywords: social and economic activity of the HEI Big Data, social networks, online data, comparative analysis, student demographics, student geography

For citation: Shitova YuYu., Shitov YuYu., Kokorev MS. Comparative analysis of the activity of higher education institutions (HEI) on the basis of data from social networks. *RSUH/RGGU Bulletin. "Economics. Management. Law" Series.* 2018;3(13):9-22. DOI: 10.28995/2073-6304-2018-3-9-22

Введение

Современное общество вступило в новую информационную эпоху развития, в которой все большее количество общественных коммуникаций осуществляется онлайн. Современные люди проводят все большую часть времени в Интернете, генерируя большое количество информации, так называемой BigData (Большие данные). Ценность этих данных трудно переоценить, они представляют главную ценность ведущих интернет-компаний с оборотами в миллиарды долларов (Гугл, Яндекс, Фейсбук, Амазон и др.). Анализ этой информации позволяет детально изучать поведение пользователей, их привычки, интересы, черты характера и т. п. Использование этой информации вооружает нас недоступными

ранее знаниями, с помощью которых можно революционно изменить жизнь общества как в лучшую сторону – улучшения стандарта жизни, так и в худшую – манипуляции массовым сознанием. Подтверждением второго положения является развивающийся сейчас скандал между Фейсбуком и «Кэмбридж Аналитикой», в котором данные о 87 млн пользователей ФБ были использованы в предвыборной кампании Дональда Трампа для влияния на результаты президентских выборов США в 2017 году [1]. Скандал стоил ФБ 60 млрд прямых потерь в капитализации, не считая возможных будущих штрафов от правительств США, Британии и ЕС. Уровень обсуждения и размер потерь еще раз ярко подчеркивает решающее значение BigData для современной цифровой экономики. Следует отметить, что скандалы подобного рода в целом негативно сказываются на свободе накопления и использования BigData для полезных и позитивных целей. А таких примеров также много. Прежде всего речь идет о контекстной рекламе, являющейся главным источником дохода многих интернет-компаний. В более широком смысле анализ BigData развился в целую индустрию DataScience, задачей которой является анализ и обработка BigData с целью обнаружения закономерностей, функционала, характерных особенностей, которые позволят фирмам различного профиля оптимизировать их бизнес-процессы.

Предметом нашего интереса является исследование активности студентов и выпускников вузов в социальных сетях. Анализ литературы показывает, что в последние два года появились десятки работ, посвященных этой тематике. Прежде всего вузы интересуют использование социальных сетей как инструмента для контакта и продвижения учебного заведения среди своей целевой аудитории – будущих студентов [2–4]. Другое перспективное направление – использование социальных сетей для учебных целей, исследование активности студенческих сообществ [5, 6]. Наконец, полезную для вуза информацию можно извлекать из активности выпускников, к примеру исследовать профильность их трудоустройства в статике и динамике [7]. Целый ряд работ посвящен специфике работы с данными социальных сетей: этическими проблемами автоматического сбора данных [8], актуально проявившимися в скандале с ФБ, вопросом достоверности данных [9], вносящим систематические погрешности.

Настоящая работа является логическим развитием предыдущего исследования, посвященного анализу деятельности вуза на основании уникальных данных 7000 студентов Университета «Дубна» (УД), полученных из социальной сети ВКонтакте (VK) [10]. Пользуясь той же методикой, из VK были извлечены и проанализи-

зированы данные 38 500 студентов и выпускников Российского государственного гуманитарного университета (РГГУ). В данной работе проведен сравнительный анализ двух вузов, результаты которого представлены далее.

Методика сбора данных

Методика сбора данных представлена в предыдущей работе [10], ограничимся здесь ее кратким изложением. Общая схема получения данных представлена на рис. 1. Для извлечения данных из сети ВК [11] использовалась специально написанная программа с графическим интерфейсом, через который пользователь в полуавтоматическом режиме мог загружать из сети ВК данные с определенным фильтром (в данном случае принадлежность к конкретному вузу).

Для каждого пользователя из ВК извлекались следующие данные: ФИО, пол, возраст, дата рождения, страна, город рождения, текущий город проживания, место обучения, год окончания обучения, факультет, кафедра/направление. Отдельной проблемой был обход ограничений на количество и скорость извлекаемой информации в автоматическом режиме, чтобы избежать блокирования программы со стороны ВК.



Рис. 1. Схема извлечения данных.

Пользователь с персонального компьютера (UserPC) через программу с графическим интерфейсом (GUI) извлекает данные из ВК, используя ее открытый программный интерфейс (OpenAPI)

В итоге нам удалось собрать данные об 7000/38 500 студентов и выпускников Университета «Дубна» и РГГУ соответственно. Результаты сравнительного анализа полученных данных представлены в следующем разделе.

Результаты анализа

Структура полученных данных позволяет провести социально-экономический анализ студентов в нескольких направлениях, каждое из которых представлено далее в отдельных параграфах.

География студентов

Мониторинг географии поступающих и ее временной динамики имеет важное значение для вуза при правильной оценке своей целевой аудитории, организации и координации рекламной и маркетинговой деятельности по привлечению будущих абитуриентов. Данные, извлекаемые из социальных сетей, позволяют проводить такой мониторинг. География студентов представлена в таблице. Из нее видно, что география студентов в Университете «Дубна» значительно более диверсифицирована.

В Университете «Дубна» (УД) представлены четыре большие группы студентов. Две наибольшие доли составляет Дубна (немногом более 40%) и Московская область (МО) (~37%). Две меньшие группы составляют Тверская область (ТО) и Москва (суммарно почти 20%). В РГГУ доминируют москвичи (70%) при доле немосквичей чуть более 20%. Отметим, что количество студентов, не указавших родной город, примерно одинаково и составляет ~10%.

Таблица

География студентов и выпускников исследуемых вузов

Университет Дубна			РГГУ		
Место	чел.	%	Место	чел.	%
Дубна	3228	31,6	Москва	26878	75
Московская обл.	26878	75			
Москва	891	8,7			
Тверская обл.	742	7,3			
Остальная Россия	601	5,9	Остальная Россия	8153	22,7
Нет данных	893	8,7	Нет данных	3502	9,1
<i>Итого</i>	10218	100,0	<i>Итого</i>	38533	100,0

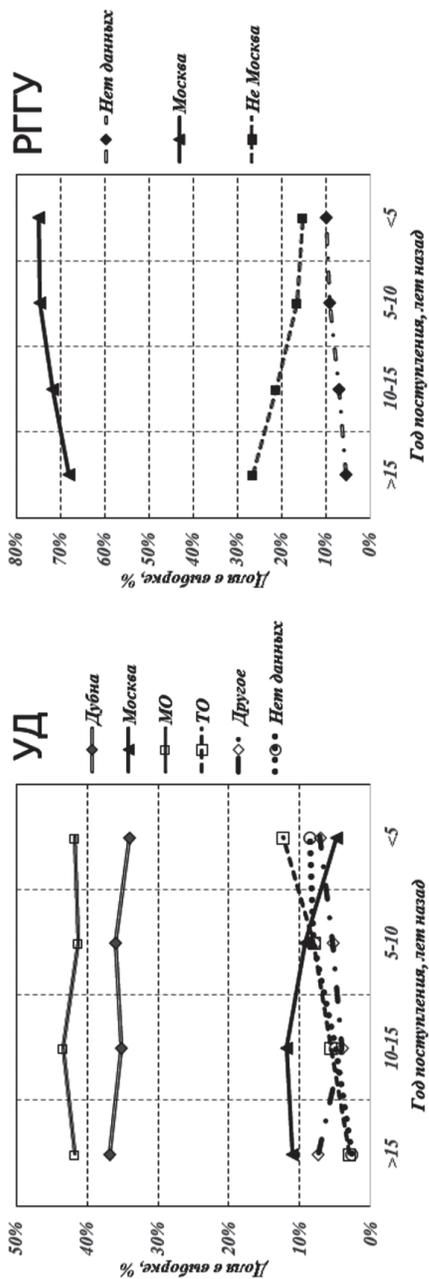


Рис. 2. Динамика географии поступающих Университета «Дубна» (слева) и РГГУ

Временная динамика географии студентов показана в таблице. Из нее видно, что доля крупнейших групп студентов в УД практически не меняется – УД остается учебным заведением жителей Дубны и МО. Однако остальные две группы показывают заметную динамику. Стабильно снижается доля москвичей (более чем в два раза, с 11% до 5%), которые замещаются студентами из ТО, доля которых выросла в 4 раза, с 3% до 12%. РГГУ становится все более московским вузом – доля москвичей увеличилась с 68% до 75%, а доля немосквичей упала почти в два раза: с 27% до 15%.

Тем самым налицо тенденция дальнейшей концентрации московских выпускников в московских вузах. В то время как абитуриенты не из Москвы выбирают регионально близкие сильные вузы, как это происходит с жителями ТО, обучающимися в УД.

Демография

Возраст. Возрастное распределение студентов УД и РГГУ согласно заявленному ими возрасту показано на рис. 4 слева.

Из него видно, что основной возрастной контингент, представленный в Сети, – 20–40 лет. При этом студенты и выпускники УД гораздо моложе (повышенное представительство в диапазоне 20–25 лет), что связано с молодостью самого УД, который существует всего лишь 23 года. Ядро РГГУ с его гораздо большей историей составляют студенты и выпускники в возрасте почти 30 лет, при этом виден длинный правый хвост больших возрастов. Еще один показатель, демонстрирующий «возраст» вуза, – представительство (доля) студентов и выпускников в Сети в зависимости от года поступления, показанное на рис. 5. УД – молодой и растущий вуз, поэтому его онлайн-аудитория представлена преимущественно молодежью, выпускников с большим стажем очень мало (растущая сплошная кривая). В противовес ему возрастной РГГУ, обладая большим пулом выпускников среднего возраста, показывает пик в этом диапазоне (пунктирная кривая).

Пол. На рис. 3 справа показана возрастная структура студентов РГГУ в гендерном разрезе. Между гистограммами наблюдается заметный сдвиг: женщины «моложе» мужчин. Это видно и для респондентов 20–30 лет, и возрастное правостороннее асимптотическое смещение у мужчин заметно больше. На наш взгляд, этот эффект связан с тем обстоятельством, что девушки указывают в профиле ВК меньший возраст, чем есть на самом деле. Другой эффект той же природы связан с указанием или не указанием возраста в профиле. Распределение этого показателя показано

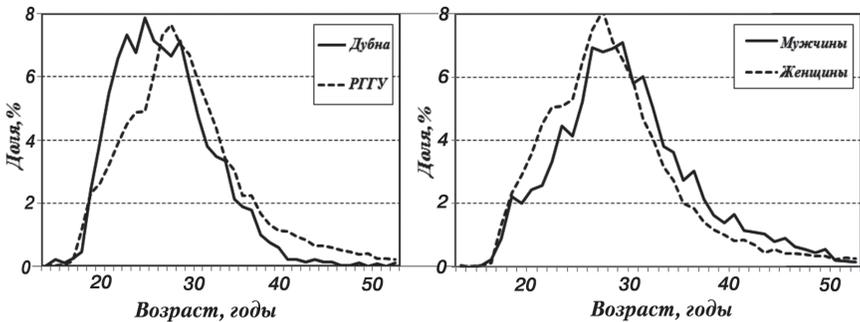


Рис. 3. Распределение студентов и выпускников по возрасту
 Слева: Университет «Дубна» против РГГУ
 Справа: состав РГГУ в гендерном разрезе

на рис. 5. Из него видно, что доля девушек, не указывающих возраст в профиле, выше, чем у мужчин (60% против 50%). Причем это соотношение одинаково для обоих вузов, что говорит в пользу общей гипотезы: девушки склонны занижать или не указывать возраст в профиле социальных сетей. Эти эффекты важны для учета систематических погрешностей при использовании персональных данных, извлеченных из социальных сетей.

Гендерная динамика студентов показана на рис. 6. Из него очевидно, что тенденции в исследуемых вузах кардинально противоположны. РГГУ становится все более женским: доля женщин за последнюю декаду выросла в два раза. В противоположность ему в УД немного, но стабильно увеличивается доля мужчин.

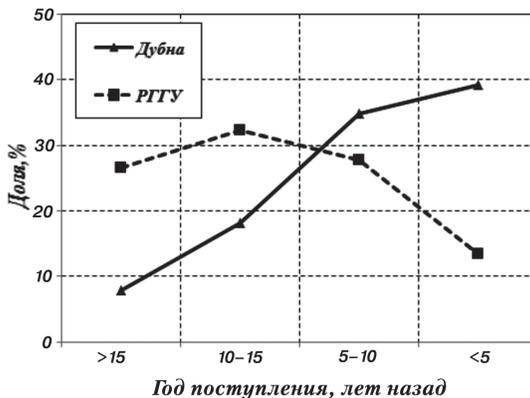


Рис. 4. Представительство в Сети по времени поступления

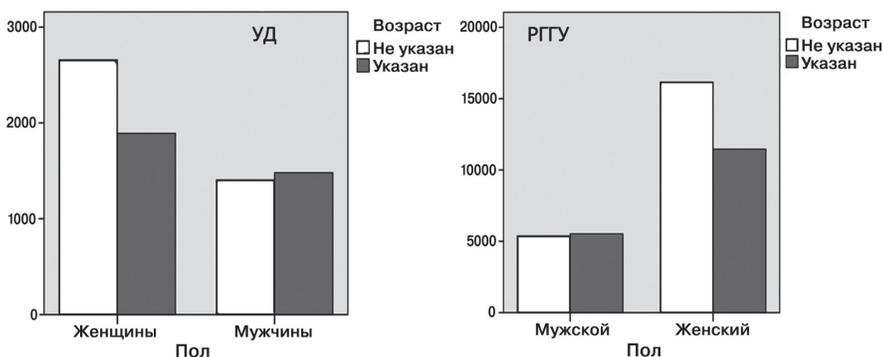


Рис. 5. Открытость данных по возрасту. Слева – УД, справа – РГГУ

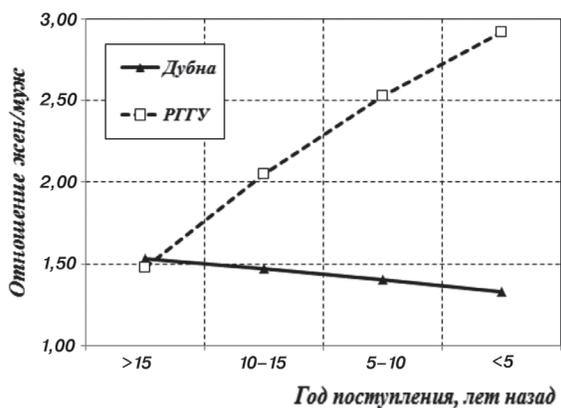


Рис. 6. Динамика гендерной структуры студентов

Направление обучения

Еще один важный показатель деятельности вуза – востребованность тех или иных направлений (специализаций) обучения, структура и динамика обучающихся по специализациям с учетом дополнительных параметров, к примеру пола. Структура и динамика студентов в разрезе профиля обучения и пола представлены на рис. 7. Из него видны ожидаемые результаты по гендерной специфике популярности направлений. Мужчины УД предпочитают информационные технологии (ИТ) и естественно-

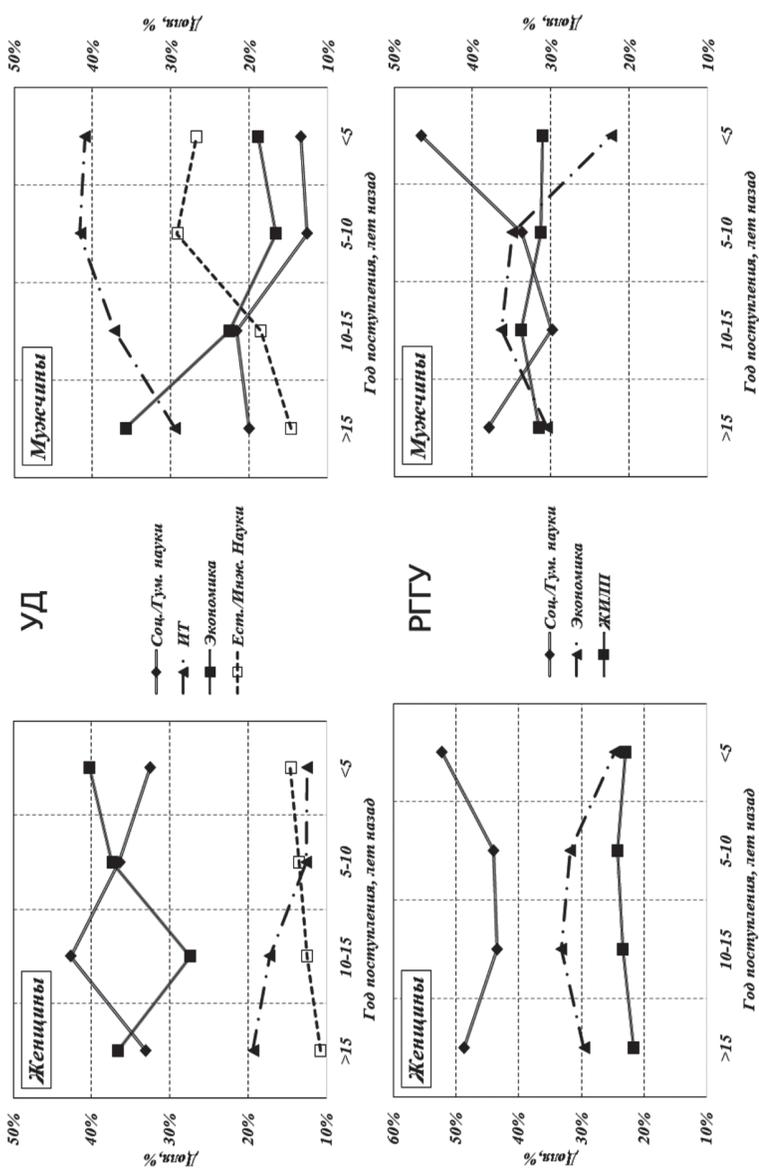


Рис. 7. Динамика выбора направления обучения студентами Университета «Дубна» (сверху) и РГГУ (снизу): женщины – слева, мужчины – справа

инженерные специальности, и этот тренд усиливается по времени при соответствующем уменьшении интереса к социально-гуманитарным и экономическим дисциплинам, усиливающимся во времени. И практически противоположное поведение наблюдается у женщин УД. Они предпочитают экономические, социальные и гуманитарные дисциплины при меньшем внимании к остальным дисциплинам. Временным трендом является снижение интереса к ИТ и небольшой рост интереса к естественно-инженерным направлениям, хотя абсолютная доля женщин тут остается относительно малой.

Отметим общее снижение интереса обоих полов УД к социальным и гуманитарным дисциплинам.

В РГГУ мы видим противоположную тенденцию: стабильно растет доля социально-гуманитарного направления, причем за счет уменьшения обучающихся экономическим специальностям. Доля остальных направлений ЖИЛП (журналисты, ИТ, лингвисты, право) остается стабильной и не зависит от пола.

Столь различное поведение студентов в отношении социально-гуманитарных наук, на наш взгляд, объясняется двумя противоречивыми факторами воздействия. С одной стороны, выпускники этого направления в целом менее востребованы и получают меньшие зарплаты по сравнению с выпускниками других специальностей. С другой стороны, социально-гуманитарный профиль традиционно является главным направлением в РГГУ, куда вкладываются основные деньги и где наработан наибольший опыт. Поэтому на фоне кризиса финансирования можно предположить, что РГГУ пытается сконцентрироваться на обучении профильных направлений.

Заключение

В настоящем исследовании мы продолжили развивать оригинальную методiku социально-экономического анализа деятельности вуза на основании данных, извлеченных из сети ВКонтакте при помощи специально разработанного программного обеспечения. Проведен сравнительный анализ 7000 и 38 500 студентов и выпускников Университета «Дубна» и РГГУ соответственно. Исследованы структура и временная динамика демографии студентов, географии их приезда, направления обучения в гендерном разрезе. Выявлены различия в формировании и динамики развития регионального и столичного вузов.

Предлагаемая методика сбора и анализа онлайн-данных имеет широкие перспективы дальнейшего развития. Во-первых, она может быть легко масштабирована на больший масштаб, охват всех российских вузов на основании разработанного технического инструментария. Во-вторых, анализ может быть существенно углублен за счет анализа не только профилей студентов и выпускников, но и анализа их активности в сети (публикаций, связей в сообществе и т. д.). Эти методики и инструментарий для них уже появляются и используются в научных исследованиях.

В целом следует отметить тот факт, что анализ BigData является приоритетным и востребованным направлением развития науки и бизнеса, а настоящая работа следует этому современному тренду.

Литература

1. *Meredith S.* Facebook-Cambridge Analytica: A timeline of the data hijacking scandal [Электронный ресурс] CNBC online, 10 April 2018. URL: <https://www.cnbc.com/2018/04/10/facebook-cambridge-analytica-a-timeline-of-the-data-hijacking-scandal.html> (дата обращения 15 августа 2018).
2. *Ромм М.В., Заякина Р.А.* Сетевые сообщества с участием вуза: сложившиеся практики социального взаимодействия // Высшее образование в России. 2006. № 11. С. 28–36.
3. *Wolski K.* How to Effectively Promote Universities and Research Institutes in the Network? Psychological Mechanisms of E-Marketing Effectiveness // MINIB [Электронный ресурс]. 2015. Vol. 16. Issue 2. P. 67–90. URL: <https://doi.org/10.14611/minib.16.02.2015.07> (дата обращения 15 августа 2018).
4. *Rutter R., Roper S., Lettice F.* Social media interaction, the university brand and recruitment performance // Journal of Business Research [Электронный ресурс]. 2016. Vol. 69. Issue 8. P. 3096-3104. URL: <https://doi.org/10.1016/j.jbusres.2016.01.025> (дата обращения 15 августа 2018).
5. *Разгон А.В., Винникова А.Б.* Сравнительная оценка активности сообществ в социальных сетях вузов г. Барнаула по экономическим направлениям подготовки // Вестник Алтайской академии экономики и права. 2016. № 1 (43). С. 100–104.
6. *Peruta A., Alison B.* Shields Social media in higher education: understanding how colleges and universities use Facebook // Journal of Marketing for Higher Education [Электронный ресурс]. 2017. Vol. 27. Issue 1. P. 131–143. URL: <http://dx.doi.org/10.1080/08841241.2016.1212451> (дата обращения 15 августа 2018).
7. *Павлова О.Н., Казин Ф.А., Бутаков Н.А.* Профильность трудоустройства выпускников вузов: анализ данных социальных сетей // Университетское управление: практика и анализ. 2017. Т. 21. № 3 (109). С. 38–56.
8. *Alim S.* Automated Data Extraction from Online Social Network Profiles: Unique Ethical Challenges for Researchers // International Journal of Virtual Communities and Social Networking. 2013. № 5 (4). Article 2. 19 p.

9. *Смирнов И.Б., Сивак Е.В., Козьмина Я.Я.* В поисках утраченных профилей: достоверность данных «ВКонтакте» и их значение для исследований образования // *Вопросы образования*. 2016. № 4. С. 106–122.
10. *Шитова Ю.Ю., Шитов Ю.А., Кокорев М.С.* Социально-экономический анализ деятельности вуза на основе данных из социальных сетей // *Вестник РГГУ. Серия «Экономика. Управление. Право»*. 2017. № 3 (9). С. 50–60.
11. Сеть ВКонтакте [Электронный ресурс]. URL: <http://vk.com> (дата обращения 15 августа 2018).

References

1. *Meredith S.* Facebook-Cambridge Analytica. A timeline of the data hijacking scandal [Internet]. CNBC online. 10 April 2018. URL: <https://www.cnn.com/2018/04/10/facebook-cambridge-analytica-a-timeline-of-the-data-hijacking-scandal.html> (data obrashcheniya 15 Aug. 2018).
2. Romm MV., Zayakina RA. Network communities with the participation of the university. Established practices of social interaction. *Vyssee obrazovanie v Rossii*. 2006;11:20-36. (In Russ.)
3. Wolski K. How to effectively promote Universities and research Institutes in the Network? Psychological mechanisms of E-Marketing Effectiveness. *MINIB* [Internet]. 2015;16(2):67-90. URL: <https://doi.org/10.14611/minib.16.02.2015.07> (data obrashcheniya 15 Aug. 2018).
4. Rutter R., Roper S., Lettice F. Social media interaction, the university brand and recruitment performance. *Journal of Business Research* [Internet]. 2016;69(8):3096-3104. URL: <https://doi.org/10.1016/j.jbusres.2016.01.025> (data obrashcheniya 15 Aug. 2018).
5. Razgon AV., Vinnikova AB. Comparative assessment of the communities activity in social networks of universities in Barnaul in the economic training areas. *Vestnik Altaiskoi akademii ekonomiki i prava*. 2016;1(43):100-104. (In Russ.)
6. Peruta A., Alison B. Shields Social media in higher education: understanding how colleges and universities use Facebook. *Journal of Marketing for Higher Education* [Internet]. 2017;27(1):131-43. URL: <http://dx.doi.org/10.1080/08841241.2016.1212451> (data obrashcheniya 15 Aug. 2018).
7. Pavlova ON., Kazin FA., Butakov NA. The graduates employment profile. Analysis of social networks data. *Universitetskoe upravlenie: praktika i analiz*. 2017;3(109):38-56. (In Russ.)
8. Alim S. Automated Data Extraction from Online Social Network Profiles: Unique Ethical Challenges for Researchers. *International Journal of Virtual Communities and Social Networking*. 2013;5(4). Article 2. 19 p.
9. Smirnov IB., Sivak EV., Koz'mina YaYa. Looking for the lost profiles. The accuracy of the "Vkontakte" data and its significance for educational studies. *Voprosy obrazovaniya*. 2016;4:106-22. (In Russ.)
10. Shitova YuYu., Shitov YuA., Kokorev MS. The university activity socio-economic analysis based on the data from social networks. *RSUH/RGGU Bulletin. "Economics. Management. Law" Series*. 2017;3(9):50-60. (In Russ.)
11. VKontakte network [Internet]. URL: <http://vk.com> (data obrashcheniya 15 Aug. 2018). (In Russ.)

Информация об авторах

Юлия Ю. Шитова, доктор экономических наук, профессор, Российский государственный гуманитарный университет, Москва, Россия; 125993, Россия, Москва, Миусская пл., д. 6; yu_shitova@mail.ru

Юрий Ю. Шитов, кандидат физико-математических наук, Объединенный институт ядерных исследований, Московская область, Дубна, Россия; 141980, Россия, Московская обл., Дубна, ул. Жолио Кюри, 6; shitov@jinr.ru

Максим С. Кокорев, аспирант, Университет «Дубна», Московская область, Дубна, Россия; 141982, Россия, Московская обл., Дубна, ул. Университетская, д. 9; kokms.11@gmail.com

Information about the authors

Yulia Yu. Shitova, Dr. in Economics, professor, Russian State University for the Humanities, Moscow, Russia; bld. 6, Miusskaya Square, Moscow, Russia, 125993; yu_shitova@mail.ru

Yuri Yu. Shitov, PhD in Physics and Mathematics, Joint Institute for Nuclear Research, Dubna, Moscow Region, Russia; bld. 6, Joliot Curie Street, Dubna, Moscow Region, Russia, 141980, shitov@jinr.ru

Maxim S. Kokorev, postgraduate student, University of Dubna, Dubna, Moscow Region, Russia; bld. 9, Universitetskaya Street, Dubna, Moscow Region, Russia, 141982; kokms.11@gmail.com