УДК 378:004.7

DOI: 10.28995/2073-6304-2019-1-56-65

Сравнительный анализ вузов Подмосквья на основе анализа активности студентов в социальных сетях

Юлия Ю. Шитова

*Российский государственный гуманитарный университет, Москва, Россия,
shitova.yu@rggu.ru*

Юрий А. Шитов

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

Максим С. Кокорев

*Университет Дубна, Дубна, Россия,
kokms.11@gmail.com*

Аннотация. Настоящая работа посвящена анализу активности студентов шести различных вузов Подмосквья в социальной сети ВКонтакте. На основании рассматриваемых данных, извлеченных при помощи специально написанного для этого программного обеспечения, проведен сравнительный анализ работы вузов по ряду направлений. Исследованы структура и временная динамика ряда показателей: демографии обучающихся; географии приезда студентов. В заключение 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.А., Кокорев М.С. Сравнительный анализ вузов Подмосквья на основе анализа активности студентов в социальных сетях // Вестник РГГУ. Серия «Экономика. Управление. Право». 2019. № 1. С. 56–65. DOI: 10.28995/2073-6304-2019-1-56-65

Comparative analysis of universities
in the Moscow region
on the basis of an analysis of student activity
in social networks

Yulia Yu. Shitova

*Russian State University for the Humanities, Moscow, Russia,
shitova.yu@rggu.ru*

Yuri A. Shitov

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

Maxim S. Kokorev

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

Abstract. The present work is devoted to the analysis of the activity of students from six different universities in the Moscow region using data of the social network VKontakte. A comparative analysis of the work of universities in a number of areas was carried out using the unique data, extracted with the help of specially written to that end software. The structure and dynamics of a number of indicators are studied: the demographics and geography of the students. In conclusion, the prospects for the development of the methodology are discussed.

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

For citation: Shitova YuYu., Shitov YuA., Kokorev MS. Comparative analysis of universities in the Moscow region on the basis of an analysis of student activity in social networks. *RSUH / RGGU Bulletin. "Economics. Management. Law" Series.* 2019;1:56-65. DOI: 10.28995/2073-6304-2019-1-56-65

Введение

В эпоху глобальной онлайн-революции в Интернет уходит все большая часть общественных отношений всех типов и уровней, от личных до государственных. Нормой стал тот факт, что важнейшие государственные новости политики высшего уровня анонсируют через социальные медиа (американский президент Д. Трамп предпочитает это делать через Твитер). Все большее количество человек в мире пользуется онлайн-новыми социальными сетями, где люди проводят все больше и больше времени, генерируя Большие

Данные (BigData). Эта информация является ключевой ценностью и ресурсом современного бизнеса ведущих инновационных технологических компаний с миллиардными оборотами (Гугл, Амазон, Фейсбук, Яндекс и др.). Анализ Биг Дата позволяет узнать все о поведении и желаниях людей, а использование информации в интересах общества и бизнеса революционно изменяет жизнь человечества.

Настоящее исследование продолжает цикл работ, посвященных анализу деятельности вуза на основании данных о студентах, извлекаемых из социальных медиа. Сначала была разработана, отлажена и реализована оригинальная авторская методика сбора и анализа информации на примере одного вуза – Университета «Дубна» [1], затем был проведен сравнительный анализ двух вузов: подмосковного (Университет «Дубна») и столичного (РГГУ) [2]. В указанных работах читатель найдет обзор современного состояния дел по данному направлению науки. В настоящей работе из социальной сети ВКонтакте (ВК)¹, при помощи той же методики были извлечены и проанализированы данные чуть менее 35 тыс. студентов и выпускников пяти вузов Московской области – государственных и негосударственных университетов, близких по направлению обучения:

- Государственный университет «Дубна» (Дубна);
- Государственный гуманитарно-технологический университет (Орехово-Зуево);
- Московский областной институт управления (Балашиха);
- Технологический университет (Королев);
- Государственный социально-гуманитарный университет (Коломна).

В работе проведен сравнительный анализ данных вузов, результаты которого представлены далее.

Методика сбора данных

Детали методики сбора данных представлены в предыдущих работах [1, 2], поэтому здесь опишем ее в общих чертах. Согласно схеме, показанной на рис. 1, данные из сети ВК² загружались в полуавтоматическом режиме через специальное ПО с графическим интерфейсом, работающее через API ВК. При этом обрабатывались профили пользователей, указавших принадлежность (факт обучения) к одному из изучаемых пяти вузов.

¹ Сеть ВКонтакте [Электронный ресурс]. URL: <http://vk.com>

² Там же.

Рис. 1. Схема сбора данных

Для каждого настоящего или бывшего студента из профиля извлекалась и записывалась следующая информация: ФИО (никнейм), возраст, пол, дата рождения, город и страна рождения, текущий город проживания, место и год окончания обучения, факультет, кафедра/направление. Извлечение существенной (десятки тысяч записей) объема информации из социальной сети является проблемой, поскольку возможно блокирование (бан) программы сбора информации со стороны сервера ВК, который может посчитать большое количество запросов DDoS-атакой³ на сервис. Поэтому требуется искусственное замедление скорости запросов и работа с нескольких IP-адресов. Несмотря на проблемы, в итоге нам удалось собрать данные об 33 342 студентах и выпускниках пяти вузов Московской области.

Результаты анализа

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

³ Иванов О. Что собой представляет DDoS-атака? [Электронный ресурс]. URL: https://www.anti-malware.ru/analytics/Threats_Analysis/what-is-a-ddos-attack (дата обращения 12 фев. 2019).

Статистика выборки и география студентов

Количественные данные в разрезе вузов и географии поступающих представлены в табл. 1.

Таблица 1

Общая собранная статистика и география студентов

Показатели	Вуз				
	Дубна	Орехово-Зуево	Коломна	Балашиха	Королев
Количество данных, чел.	6365	8030	13869	1146	3932
Доля от общей выборки, %	19,09%	24,08%	41,60%	3,44%	11,79%
<i>География студентов, в процентах от выборки своего вуза</i>					
Из того же города, что и вуз	43,5	36,0	48,1	52,0	44,4
Московская область	10,0	28,0	26,1	13,4	18,7
Иной регион	34,9	19,3	13,2	24,7	24,5
Нет данных	11,6	16,7	12,6	9,9	12,4

Источник: расчеты авторов

Контроль структуры и динамики географии обучающихся важен вузу для точного выявления целевой аудитории, правильной координации и организации рекламных и маркетинговых услуг с целью привлечения абитуриентов. Полученные нами данные из социальных сетей позволяют проводить такой мониторинг.

Год поступления в вуз в профиле ВК не указывается, поэтому определялся расчетным путем по возрасту пользователей. При этом мы исходили из предположения, что большинство студентов поступают учиться сразу после школы в возрасте ~17 лет. Исходя из этого, далее производилась группировка студентов по пятилетним группам.

Динамика географии учащихся и выпускников представлена на рис. 2. Из нее видно, что в целом структура обучающихся за 20 лет не претерпела существенных изменений за исключением Королева. Технологический университет Королева образован в 1998 г., естественным образом большинство студентов в первые годы работы составили жители города. Поэтому в первой группе

(поступление 15 и более лет назад) наблюдается большая доля учащихся из своего города. Однако уже через пять лет уровень опустился до сравнимого с другими за счет притока студентов из Московской области (МО).

Мы видим, что доля обучающихся из города, где расположен университет, сравнима по всем пяти вузам, разброс здесь небольшой (35–50%). Наибольшее количество жителей Московской области – каждый третий студент – привлекают Коломна и Орехово-Зуево, далее эта доля снижается последовательно в Королеве, Балашихе и Дубне. А по количеству студентов, привлеченных из других регионов (не родной город и не МО), безусловным лидером является Дубна (более 35% обучающихся). Доля студентов, не указавших родной город, колеблется на уровне ~6–14%. Интересно, что динамика ее изменений полностью повторяется для четырех из пяти вузов, что позволяет предположить, что доля студентов с незаполненными профилями коррелирует с их возрастом.

Рис. 2. Динамика обучающихся по географии студентов

Релевантность статистики и скрытые данные

Одна из проблем информации, извлекаемой из социальных сетей – проверка ее качества. Часть пользователей сетей: а) не заполняют все поля профиля, б) заполняют его неверными данными. В предыдущей работе [2] было показано, что девушки чаще скрывают свой возраст или указывают заниженный возраст. Анализ настоящих данных также подтверждает эти выводы. В частности, доля мужчин, скрывающих свой возраст, составила 27%, в то время как 42% женщин предпочли не называть свой возраст.

Кроме того, выяснилось, что доля пользователей с заполненным профилем (полная информация) коррелирует с возрастом пользователей (см. рис. 3). Из него следует, что наиболее молодые и возрастные пользователи меньше склонны представлять информацию о себе, поэтому об этих пользователях можно получить меньше данных.

Рис. 3. Количество зарегистрированных пользователей в соответствии с возрастом для общей выборки

Демография

Возрастное распределение студентов выбранных вузов согласно заявленному ими возрасту показано на рис. 4.

Из него видно, что контингенты вузов заметно отличаются по возрастному составу. Орехово-Зуево и Коломна имеют сходное распределение и являются самыми возрастными, судя по длинным правым хвостам распределения. В противоположность им Дубна и Королев – молодые вузы в силу относительно недолгого периода работы, поэтому они еще не успели обзавестись выпускниками в преклонном возрасте. И особняком стоит Балашиха – она имеет компактный пик в районе 25–40 лет и малое количество студентов других возрастов. Особенно удивительно малое количество молодежи (17–25 лет) по сравнению со всеми другими вузами. Природу такого отличия необходимо изучать отдельно (данный вопрос выходит за рамки настоящего исследования).

Рис. 4. Распределение студентов и выпускников по возрасту

Еще один показатель, демонстрирующий «возраст» вуза, — представительство (доля) студентов и выпускников в сети в зависимости от года поступления, показанное на рис. 5.

Рис. 5. Представительство в сети по времени поступления

Исходя из данной динамики трудно предположить, что за последние пять лет снижается доля поступающих. Как было показано выше, молодежь склонна к анонимности в сети, поэтому среди них велика доля людей, не указывающих свои данные. Анонимность в сети – современный тренд, понимание которого связано с текущей социально-политической ситуацией в стране. В последние годы было ужесточено законодательство в отношении действий, осуществляемых в сетевых медиа, особенно среди молодежи. В ответ на это существенная доля пользователей, опять же преимущественно молодых, стала убирать из сетей свои персональные данные с целью затруднения их идентификации и трекинга сетевой активности. Полученные нами результаты демонстрируют этот процесс.

Также на рис. 5 видно различие между парами молодых вузов Дубны, Королева и возрастных Коломны, Орехово-Зуева, подтверждающее выводы на основании анализа возрастных распределений (рис. 4), обсуждавшихся выше.

Заключение

В настоящем исследовании мы продолжили развивать оригинальную методику социально-экономического анализа деятельности вуза на основании данных, извлеченных из сети ВКонтакте при помощи специально разработанного программного обеспечения. Проведен сравнительный анализ 33 тысяч студентов и выпускников пяти вузов Московской области. Исследованы структура и временная динамика демографии студентов, географии их приезда, релевантности данных. Выявлены различия в формировании и динамики развития региональных вузов.

Предлагаемая методика сбора и анализа онлайн-данных имеет широкие перспективы дальнейшего развития. Во-первых, она может быть легко масштабирована на больший масштаб, охват любых 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. 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.
2. Шитова Ю.Ю., Шитов Ю.А., Кокорев М.С. Сравнительный анализ деятельности вузов на основании данных из социальных сетей // Вестник РГГУ. Серия: «Экономика. Управление. Право». 2018. № 3. С. 9–23.

References

1. Shitova YuYu., Shitov YuA., Kokorev MS. Socio-economic analysis of the activities of the university based on data from social networks. *RSUH/RGGU Bulletin. "Economics. Management. Law" Series*. 2017;3:50-60. (In Russ.)
2. Shitova YuYu., Shitov YuA., Kokorev MS. Comparative analysis of the activities of universities on the basis of data from social networks. *RSUH/RGGU Bulletin. "Economics. Management. Law" Series*. 2018;3:9-23. (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; shitova.yu@rggu.ru

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

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

Information about the authors

Yulia Yu. Shitova, Dr. of Sci. (Economics), professor, Russian State University for the Humanities, Moscow, Russia; bld. 6, Miusskaya Sq., Moscow, Russia, 125993; shitova.yu@rggu.ru

Yuri A. Shitov, Cand. of Sci. (Phys. and math.), Joint Institute for Nuclear Research, Moscow, Russia; bld. 6, Joliot Curie Str., Dubna, Russia, 141980; shitov@jinr.ru

Maxim S. Kokorev, University of Dubna, Dubna, Russia; bld. 9, Universitetskaya Str., Dubna, Russia, 141982; kokms.11@gmail.com