

Ю.Ю. Шитова, Ю.А. Шитов, М.С. Кокорев

Социально-экономический анализ
деятельности вуза
на основе данных из социальных сетей

В настоящей работе проведен анализ деятельности вуза на основании уникальных данных 7000 студентов университета «Дубна», полученных из социальной сети «ВКонтакте». Для полуавтоматического извлечения данных было разработано специальное программное обеспечение. Исследованы структура и временная динамика ряда показателей: демографии обучающихся; географии приезда студентов, включая ГИС-анализ (визуализацию); распределения по направлениям обучения. В заключение обсуждаются перспективы развития методики.

Ключевые слова: социально-экономическая деятельность вуза, социальные сети, анализ данных.

Проблемы полноты и качества исходной социально-экономической информации, используемой для научных исследований, всегда были и остаются актуальными в области социальных, экономических и гуманитарных наук. Недостаток информации делает проведение исследований невозможным, а некачественные данные ставят под сомнение ценность полученных результатов. В большинстве исследований используются стандартные источники информации: официальная статистика, панельные исследования, собственные опросы. Недостатки при использовании официальных данных широко известны: недостаточные объем, глубина, периодичность, широта охвата характеристик субъекта, проблема достоверности (в ходе агрегации данных выпускающий их агент может влиять на конечный результат в силу собственных интересов), высокая стоимость сбора качественных данных. В этой связи

появление и использование нового и независимого типа исходной статистики – онлайн-данных сети Интернет, при помощи которой можно решать широкий спектр задач социально-экономического профиля, является прорывом в области методологии науки. В рамках настоящей работы мы использовали данные из социальных сетей для экономико-социального анализа деятельности вуза.

Использование онлайн-данных в современных исследованиях

Научное осмысление феномена социальных медиа, революционно изменившего стиль жизнь людей в частности и трансформировавшего практически все сферы общественной деятельности, породило десятки тысяч публикаций. В подавляющем большинстве работ объектом исследования являются непосредственно сами социальные медиа. В том числе классификация, структуризация происходящих в них явлений и процессов, обзор и сравнение структуры и динамики различных социальных сетей¹, обсуждаются также возможности пользователей: объединения для совместной работы².

Большая часть публикаций нацелена, прежде всего, на взаимоотношения и влияние социальных медиа с конкретными областями деятельности. Интересны, прежде всего, возможности бизнеса в социальных сетях. В частности, способы вложения в сети и оценка полезного выхода и возвратности инвестиций³, возможности продвижения товаров и услуг через Интернет (опыт Румынии⁴), вопросы доверия и использования социальных сетей как в малом бизнесе⁵, так и поддержание лояльности больших брендов⁶.

Под влиянием «арабской весны» (революционных движений в арабских странах в первой половине 2013 г.) исследователи изучают влияние социальных медиа на политическую ситуацию⁷, корреляции активности политиков в сети и их результатов на выборах⁸ и другие вопросы.

Также исследуются: адаптация СМИ к социальным сетям⁹; использование социальных сетей в медицинских целях (пациентами и отдельными группами больных, обучающимися профессионалами, студентами); влияние социальных контактов на туристические потоки и другие темы.

В контексте настоящего исследования нас интересует извлечение и работа с данными социальных сетей. В этом аспекте научных работ представлено гораздо меньше. Отметим исследова-

ние скорости распространения слухов и новостей в социальных сетях – оценка временного отклика (по частоте и интенсивности упоминаний в твиттере) после «вброса» слуха или появления информации¹⁰. В данной методологии анализируется не контекст сообщений пользователей, а общий их поток во времени.

Для использования данных из социальных сетей необходимо применение «стандартизации» – определение стандартов, метрик (измеряемых показателей), методов анализа – данные вопросы исследуются в современный период¹¹. При этом основное внимание уделяется количественным оценкам контента, т. е. социальному и экономическому поведению пользователей, наиболее интересному политикам и производителям потребляемых товаров.

Подводя итоги, следует отметить актуальность обсуждаемого направления исследований. В следующем разделе будет рассмотрена методика сбора данных, применявшаяся в настоящей работе.

Методика сбора данных

В настоящем исследовании для извлечения онлайн-данных была использована сеть «ВКонтакте» (ВК). Ее использование обусловлено двумя обстоятельствами. Во-первых, ВК является одной из самых популярных социальных сетей в русскоязычном секторе Интернета, поэтому представительство студентов и выпускников вузов в ней существенно. Во-вторых, ВК имеет развитый программный интерфейс API (Application Program Interface). Появление API у социальных медиа ознаменовало собой новую эру циркуляции информации. С этого момента практически любая информация, содержащаяся в социальных медиа, становится доступной потребителям в массовом, автоматическом режиме через стационарные, либо мобильные интернет-приложения. Тем самым стартовала эра получения и использования децентрализованной онлайн-информации.

В качестве объекта исследования были выбраны студенты и выпускники университета «Дубна». При помощи API ВК можно получить следующие данные о пользователе: ФИО, пол, возраст, дата рождения, страна, город рождения, текущий город проживания, место обучения, год окончания обучения, факультет, кафедра/направление.

Для извлечения данных нами было создано специальное приложение (на языке C#), при помощи которого в полуавтоматическом режиме нами из сети ВК были выгружены данные обо всех

студентах и выпускниках университета «Дубна» (указавших этот факт в своих аккаунтах) – всего около 7000 человек. Результаты 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 разделах.

Демография

Пол. Соотношение «мужчины/женщины» в выборке составило 37/63%. Динамика гендерной структуры в зависимости от года поступления показана на рис. 1, из которого видно, что доля обу-

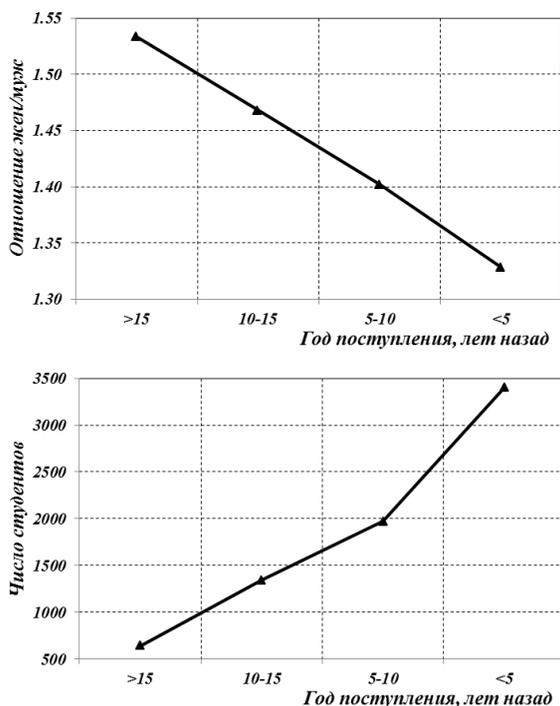


Рис. 1. Динамика гендерной структуры студентов и общего количества студентов в выборке в зависимости от года поступления

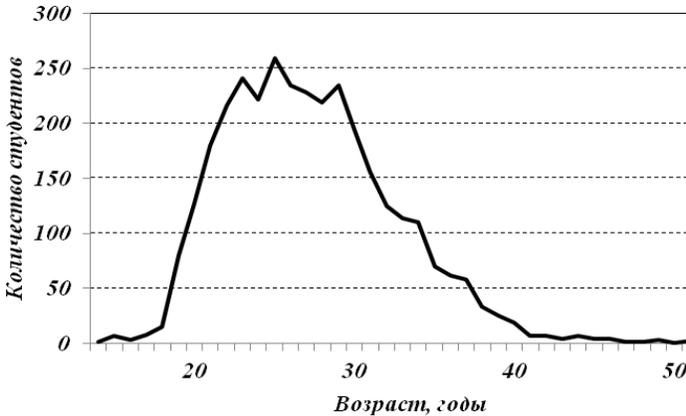


Рис. 2. Возрастная структура студентов и выпускников

чающихся мужчин увеличивается с течением времени. На рисунке показана также временная динамика студентов в выборке в зависимости от года поступления. Как и следовало ожидать, представительство более молодых студентов и выпускников в выборке существенно выше (разница более чем в 6 раз), что связано с общим трендом популярности социальных медиа, нацеленных, прежде всего, на молодых пользователей. Безусловно, рост также связан с развитием самого университета и увеличением числа студентов, однако этот вклад существенно ниже.

Возраст. Структура выборки студентов согласно заявленному ими возрасту показана на рис. 2.

Следует отметить, что у существенной доли студентов возраст либо скрыт, либо не указан год рождения. Интересно отметить, что доля девушек, скрывающих год рождения, составляет 43%, в то время как у юношей таковых всего 30%. Тем самым подтверждается гипотеза, что в социальных сетях девушки чаще склонны скрывать свой возраст, чем юноши. При этом возможен второй вариант – пользователи могут указать неверный год рождения, прежде всего, в сторону увеличения для соответствующего занижения возраста. Логично предположить, что такое поведение также должно быть более вероятным для женщин. Проверкой величины данного возможного эффекта является сравнение среднего возраста в выборке в гендерном разрезе. Если, к примеру, женщины в массовом порядке занижают возраст в ВК, то это приведет к занижению сред-

него возраста женщин в выборке по сравнению со средним возрастом мужчин. Однако расчеты показывают, что средние значения возраста мужчин и женщин практически одинаковы (27 лет) и не имеют значимого отличия. Это означает, что занижение возраста не носит массового характера. Вместо этого студентки, желающие скрыть возраст, просто не указывают год рождения.

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 обучающихся по специальностям представлена на рис. 3, из которого видно, что все направления развиваются более-менее равномерно.

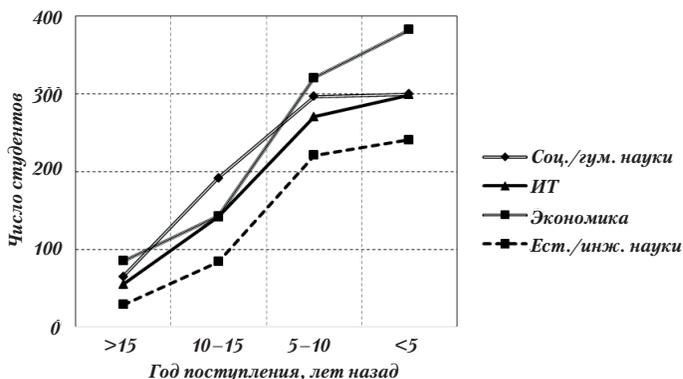


Рис. 3. Временная динамика студентов по направлениям обучения в абсолютных цифрах

На рис. 4 представлена структура и динамика обучающихся в гендерном разрезе. Из него видны ожидаемые результаты по гендерной специфике популярности направлений.

Мужчины предпочитают информационные технологии (ИТ) и естественно-инженерные специальности, и этот тренд усиливается по времени. Соответственно меньший интерес мужчины проявляют к социально-гуманитарным и экономическим дисциплинам,

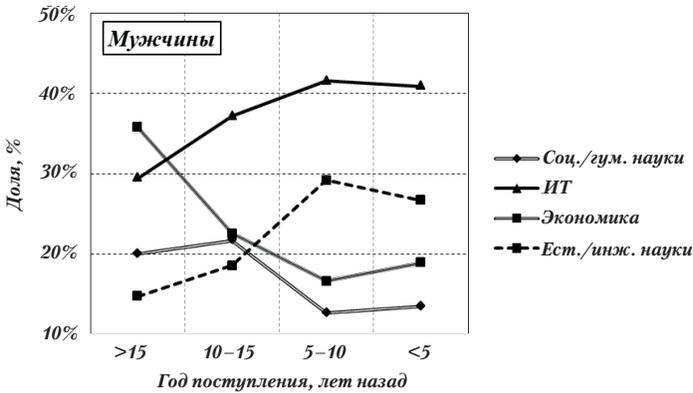


Рис. 4. Структура и динамика студентов по направлениям в гендерном разрезе

и падение интереса усиливается во времени. И практически противоположное поведение наблюдается у женщин. Они предпочитают экономические, социальные и гуманитарные дисциплины при меньшем внимании к остальным дисциплинам. Временным трендом является снижение интереса женщин к ИТ и небольшой рост интереса к естественно-инженерным направлениям, хотя абсолютная доля женщин тут остается относительно малой. Отметим общее снижение интереса обеих полов к социальным и гуманитарным дисциплинам. Скорее всего, это связано с меньшей востребованностью и низкими зарплатами выпускников этих направлений на современном рынке труда.

География студентов

Мониторинг географии поступающих и ее временной динамики имеет важное значение для вуза для правильной оценки своей целевой аудитории, организации и координации деятельности по привлечению будущих абитуриентов. Данные, извлекаемые из социальных сетей, позволяют проводить такой мониторинг.

На рис. 5 показана временная динамика структуры географии поступающих в университет, из которой можно сделать следующие выводы.

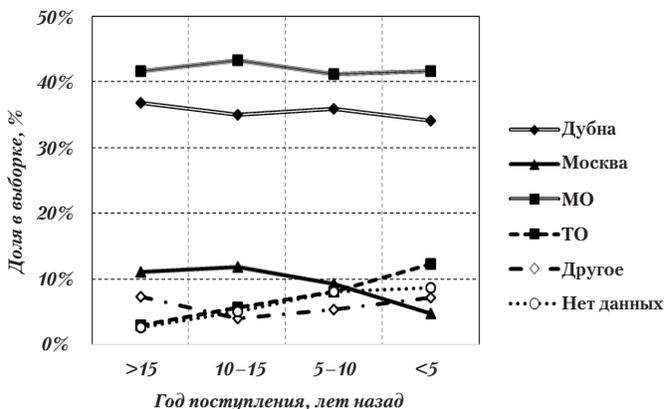


Рис. 5. Динамика структуры обучающихся по географии приезда

В целом структура поступающих по географии остается стабильной в течение всего времени существования университета, кардинальных сдвигов не зафиксировано, все наблюдаемые изменения – плавные и небольшие. Основную долю студентов (чуть более 40%) обеспечивает Московская область (МО) и Дубна (около 35%), доля которой немного снижается. Отметим снижение доли абитуриентов из Москвы более чем два раза, с 11 до 5%. При этом замещение москвичей происходит за счет жителей Тверской области (ТО), доля которых выросла в 4 раза, с 3% до 12%. Таким образом, университет имеет устойчивое и стабильное ядро потенциальных абитуриентов из МО при уменьшающемся интересе москвичей, но расширяющейся аудитории за счет прилегающей ТО.

Анализ динамики структуры студентов в гендерном разрезе не выявил значимых различий: география и ее динамика студентов мужского и женского полов практически идентичны.

ГИС-визуализация

Собранные данные позволяют нам анализировать и визуализировать географию студентов с использованием средств современных ГИС-технологий (геоинформационные системы). Такой подход позволяет более наглядно представить «ареал обитания» студентов вуза. Для настоящего анализа была использована выборка



Рис. 6. География проживания поступающих во временном разрезе

студентов, указавших родной город, географические координаты которого были получены при помощи онлайн-сервиса Яндекс-карты¹². На рис. 6–7 представлены географии поступающих в университет во временном разрезе и с учетом направления обучения.

Из рис. 6 можно заметить, что география студентов, приезжающих учиться в Дубну, со временем становится шире, с заметным представительством Южного, Северо-Западного и Сибирского регионов. Распределения прибывающих по различным направлениям обучения (рис. 7) достаточно близки.

В настоящем исследовании нами была предложена оригинальная методика социально-экономического анализа деятельности вуза на основании уникальных онлайн-данных студентов и выпускников университета «Дубна», полученных из социальной сети «ВКонтакте» при помощи специально разработанного программного обеспечения. На основании выборки из 7000 человек нами были проанализированы структура и временная динамика демографии студентов, география их приезда, направление обучения.

Предложенные в настоящем исследовании методики сбора и анализа данных имеют многообещающие перспективы, поскольку анализ онлайн-данных (Биг-дата) уже стал приоритетным направлением развития науки и бизнеса. В настоящей работе мы продемонстрировали анализ данных студентов одного вуза, но данная методика, с учетом уже разработанного технического инструментария, может быть легко расширена на практически любой масштаб.



Рис. 7. География проживания поступающих с учетом направления обучения

К примеру, это может быть сравнительный анализ разных вузов, как российских, так и зарубежных, по их студентам. Сейчас большое внимание уделяется вопросам повышения эффективности и качества образовательного процесса в вузе, предложенный инструментарий позволяет разрабатывать количественные методики и критерии оценки для решения данной задачи.

Примечания

- 1 Kietzmann Jan H., Hermkens Kristopher, McCarthy Ian P., and Silvestre Bruno S. Social media? Get serious! Understanding the functional building blocks of social media // Business Horizons. 2011. № 54. P. 241–251. [Электронный ресурс] URL: <http://www.sciencedirect.com/science/article/pii/S0007681311000061?via%3Dihub> (дата обращения: 23.08.2017); Ganea M., Constantin A., and Comşa Aurel D. Quick Overview of the Social Media // Revista Română de Statistică Trim. III/2012 Supplement. P. 113–115; То же. [Электронный ресурс] URL: http://www.revistadestatistica.ro/suplimente/2012/3/srrs3_2012a14.pdf (дата обращения: 23.08.2017).
- 2 Kaplan Andreas M., Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media // Business Horizons. 2010. № 53. P. 59–68.
- 3 Weinberg Bruce D., Pehlivan Ekin. Social spending: Managing the social media mix // Business Horizons. 2011. № 54. P. 275–282.
- 4 Buzilă Nicoleta. Promoting through the internet with social-media networking // Annals. Economic Science Series. 2012. Vol. 18. P. 658–663. [Электронный ресурс]

URL: http://fse.tibiscus.ro/images/anale/lucrari/Lucrari2012/kssue2012_105.pdf
(дата обращения: 23.08.2017).

- ⁵ *Rohayah Kahar, Faizal Yamimi, Ghazali Bunari, Hadina Habil.* Trusting the Social Media in Small Business // *Procedia: Social and Behavioral Sciences*. 2012. Vol. 66. P. 564–570.
- ⁶ İrem Eren Erdoğan, Mesut Çiçek, The Impact of Social Media Marketing on Brand Loyalty // *Procedia – Social and Behavioral Sciences*/ 2012. Vol. 58. P. 1353–1360.
- ⁷ *Sorin Suciu, Dalia Petcu.* Social Media and Political Unrest, Comşa // *Annals. Economic Science Series*. 2012. Vol. 18. P. 847–850; То же. [Электронный ресурс] URL: http://fse.tibiscus.ro/images/anale/lucrari/Lucrari2012/kssue2012_138.pdf (дата обращения: 23.08.2017).
- ⁸ *Cameron Michael P, Barrett Patrick & Stewardson Bob.* Can Social Media Predict Election Results? Evidence From New Zealand // *Journal of Political Marketing*. 2016. № 15. P. 416–432; То же. [Электронный ресурс] URL: <http://dx.doi.org/10.1080/15377857.2014.959690> (дата обращения: 23.08.2017).
- ⁹ *Sounman Hong.* Online news on Twitter: Newspapers' Social Media Adoption and their Online Readership // *Information Economics and Policy*. 2012. № 24. P. 69–74.
- ¹⁰ *Situngkir Hokky.* Spread of Hoax in Social Media. BFI Working Paper No. WP-4-2011. [Электронный ресурс] URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1831202 (дата обращения: 23.08.2017).
- ¹¹ *Drula G.* Social and online media research – data, metrics and methods // *Review of Applied Socio-Economic Research*. 2012. № 3. P. 77–86; То же. [Электронный ресурс] URL: http://reaser.eu/RePec/rse/wpaper/10_Rev3_Drula_DigitalMedia.pdf (дата обращения: 23.08.2017).
- ¹² Сервис Яндекс.карты, <https://tech.yandex.ru/maps/> (дата обращения: 23.08.2017).