

МЕЖДИСЦИПЛИНАРНЫЕ ИССЛЕДОВАНИЯ

DOI: 10.14515/monitoring.2018.1.07

Правильная ссылка на статью:

Замятина Н. Ю., Яшунский А. Д. Виртуальная география виртуального населения // Мониторинг общественного мнения: Экономические и социальные перемены. 2018. № 1. С. 117—137. DOI: 10.14515/monitoring.2018.1.07.

For citation:

Zamyatina N. Yu., Yashunsky A. D. Virtual geography of virtual population. *Monitoring of Public Opinion: Economic and Social Changes*. 2018. № 1. P. 117—137. DOI: 10.14515/monitoring.2018.1.07.



Н. Ю. Замятина, А. Д. Яшунский ВИРТУАЛЬНАЯ ГЕОГРАФИЯ ВИРТУАЛЬНОГО НАСЕЛЕНИЯ

ВИРТУАЛЬНАЯ ГЕОГРАФИЯ ВИРТУАЛЬНОГО НАСЕЛЕНИЯ

ЗАМЯТИНА Надежда Юрьевна — кандидат географических наук, ведущий научный сотрудник географического факультета Московского государственного университета им. М. В. Ломоносова, Москва, Россия.

E-MAIL: nadezam@yandex.ru

ORCID: 0000-0002-4941-9027

ЯШУНСКИЙ Алексей Дмитриевич - кандидат физико-математических наук, заведующий сектором теоретической кибернетики, Институт прикладной математики имени М. В. Келдыша Российской академии наук, Москва, Россия.

E-MAIL: yashunsky@keldysh.ru

ORCID: 0000-0002-3297-7444

VIRTUAL GEOGRAPHY OF VIRTUAL POPULATION

Nadezhda Yu. ZAMYATINA¹ — Candidate of Geographic Sciences, Leading Research Fellow

E-MAIL: nadezam@yandex.ru

ORCID: 0000-0002-4941-9027

Alexey D. YASHUNSKY² — Candidate of Physico-Mathematical Sciences

E-MAIL: yashunsky@keldysh.ru

ORCID: 0000-0002-3297-7444

¹ Moscow State University, Moscow, Russia

² Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Moscow, Russia

Аннотация. В статье охарактеризованы перспективы практического использования результатов проекта «Виртуальное население России» — новой базы данных по структуре крупнейшей российской социальной сети «ВКонтакте». Сайт проекта представляет собой «полуфабрикат» для исследования структуры социальной сети: здесь собраны массивы отражающих ее данных, представлен интерактивный конструктор карт по этим данным. Конструктор позволяет показывать разные географические, временные, демографические параметры отображаемого набора данных.

На сайте предложен обширный материал для изучения структуры пользователей социальной сети по полу, возрасту, уровню и направлению образования, миграционному поведению. Практически не имеет аналогов массив данных о виртуальной дружбе пользователей — по-видимому, это первый в мире масштабный опыт изучения территориальной привязки дружеских связей в социальных сетях, открывающий возможности для развития нового направления исследований виртуальной среды.

Представленная база данных может быть применена как для изучения интернет-пространства, так и в качестве косвенных данных для изучения «офлайновых» социальных процессов — в частности, молодежных миграций — на микроуровне.

Ключевые слова: география, виртуальная перепись, вебометрика, «ВКонтакте»

Abstract. The paper focuses on the perspectives of the practical usage of the Russian Federation Virtual Population project data. The project represents a new database relative to Vkontakte social network. The project website is a customizable resource designed to study Vkontakte structure. Data set relative to the network structure can be found on the website, as well as an interactive map builder that summarizes the data in a map and provides different geographical, temporal and demographical filters.

The website disposes a large data set that allows studying Vkontakte audience's profile taking into account such parameters as sex, age, education and migratory behaviour. The website boasts its body of data on the users' virtual friendship that seems to be the first large-scale research on whether virtual friendship relations are territorially dependent. The date opens a prospect that a new way will develop in studying virtual environment.

The database seems to be useful for studying not only virtual space but also offline social processes like youth migration in microlevel, serving in this case as an implicit data source.

Keywords: geography, virtual population census, webometrics, VKontakte.

Введение

Словосочетание «социальная сеть», изначально возникшее в социологии, прочно закрепилось в быденном языке как обозначение электронных социальных сетей — веб-сервисов, позволяющих пользователям осуществлять социальные взаимодействия в виртуальном пространстве. С начала 2000-х годов социальные сети (далее понимаем под этим термином именно электронные социальные сети) постепенно проникали в повседневную жизнь людей, и к настоящему моменту регистрация в одной или нескольких из них уже распространена повсеместно. Многие из существующих социальных сетей (Facebook, «ВКонтакте», «Одноклассники» и т. д.) возникли как средство восстановления и поддержания связей между выпускниками (или учащимися) различных учебных заведений, поэтому регистрация в таких сетях предусматривает заполнение анкеты с упоминанием возраста, места жительства, учебных заведений, времени обучения в них и т. д. Массив таких анкет, хоть и с большими оговорками, оказывается подобен переписи населения.

Анкеты пользователей социальной сети «ВКонтакте» находятся частично в открытом доступе — информация скрывается только если пользователь выставил соответствующие настройки, — что позволяет использовать их в различных исследованиях. Вопросы о том, насколько достоверные выводы могут быть получены на основе таких данных, требуют специального исследования в каждом отдельном случае, однако сама совокупность данных — эта перепись части «виртуального населения» — представляет самостоятельный интерес. В отличие от «традиционных» статистических источников, где информация представлена в агрегированном виде, этот массив информации отличается высокой детальностью, в том числе и в пространственном измерении.

Опыт использования данных социальных сетей для исследований по географии населения побудил авторов к работе над веб-сайтом «Виртуальное население России»¹, где в виде виртуального атласа опубликованы массивы данных, полученные в результате обработки анкет пользователей социальной сети «ВКонтакте», находящихся в открытом доступе.

Настоящая статья призвана показать спектр возможностей использования данного ресурса в изучении миграций и других социальных процессов в российских городах и регионах. Картограммы, приведенные в качестве иллюстраций, сформированы непосредственно на описываемом сайте и демонстрируют возможности функционала интерактивного атласа.

Основных сфер применения виртуального атласа две.

1) Этот материал открывает почти незатронутое в России поле вебометрики — собственно виртуального населения России. Термин «вебометрика» (webometrics) активно используется в зарубежной науке с конца 1990-х годов, преимущественно для обозначения исследований количественных параметров информационных ресурсов интернета [Almind, Ingwersen, 1997; Thelwall, 2002, 2008a, 2008b, 2009; Kousha, 2005]. Несмотря на то, что понятие «вебометрика» используется наряду с целым рядом других — таких как «инфометрика», «библиометрика», «наукометри-

¹ Виртуальное население России [Электронный ресурс]. URL: <http://webcensus.ru> (дата обращения: 23.02.2018).

ка» и «киберметрика», — каждое из них обозначает свой сегмент исследований, и понятие «вебометрика» не тождественно ни одному из понятий-аналогов (соотношение названных сегментов исследования охарактеризовано в [Björneborn and Ingwersen, 2004]).

Одним из прикладных направлений вебометрики стало получившее широкую известность направление рейтингования университетов — между тем, в научном поле понятие «вебометрика» используется значительно шире. В частности, широко распространено изучение информационных связей как фактора регионального инновационного развития. С начала 2000-х годов, когда вслед за открытием методов изучения так называемых «следов» (paper trails) «переливов знания» («knowledge spill-overs», распространен также перевод «перетоки знания» [Пилясов, 2012]) наблюдался настоящий бум исследований аналогичных «следов» «перетоков знания» в виртуальной среде [Feldman, 1999; Feldman, Kogler, 2010; Anselin, Varga, Acs, 1997; Audretsch, Lehmann, Warning, 2004; Maggioni, Uberti, 2009]. В этом контексте вебометрика стала важным инструментом исследований регионального развития и географии инноваций в целом. В частности, за рубежом большую популярность в данном отношении получили исследования сетей гиперссылок между сайтами университетов, которые рассматриваются в контексте диффузии научных идей и инноваций — от первых исследований по развитым странам [Thelwall, 2002] до сетей африканских университетов [Adenkannbi, 2011].

В России первые попытки изучения по западному образцу научных связей в информационном пространстве дали предсказуемые результаты: подобно колоссальной централизации многих социальных и экономических процессов в нашей стране генерация инноваций также оказывается исключительно централизованной: подавляющее большинство ссылок с сайтов 40 крупнейших российских вузов приходится на ссылки на портал МГУ им. М. В. Ломоносова. Лишь если исключить МГУ из рассмотрения, удастся найти значимые межвузовские информационные связи [Утков, Замятина, 2012].

По-видимому, более перспективны другие направления: исследование социальных сетей и упоминаний географических объектов в новостях. При этом на смену первым попыткам изучения социально-экономических процессов по данным об упоминаниях географических названий [Mamatov, Samson, 2009; Agibetova, Samson, 2008; Замятина, Пилясов, 2013] пришли мощные инструменты изучения «больших данных». В сфере изучения социальных сетей развиваются похожие процессы, однако большая часть работ пока все же выполняется на сравнительно небольших «участках» данных [Koltsova, Koltcov, Sinyavskaya, 2017; Takhteyev, Gruz, Wellman, 2012].

Предлагаемый проект призван внести вклад в устранение технических барьеров широкого вовлечения отечественных ученых в изучение структуры виртуальных социальных сетей.

Проект «Виртуальное население России» предлагает исчерпывающий материал для исследования гендерных, возрастных и территориальных различий в использовании крупнейшей в стране социальной сети. Но самое захватывающее, пожалуй, это изучение виртуальной дружбы — по-видимому, это первый в мире масштабный опыт изучения территориальной привязки социальных сетей, откры-

вающий возможности для развития нового направления исследований виртуальной среды, big data и т. д.

2) Еще одна ключевая сфера применения результатов проекта — это сбор данных для косвенных оценок масштабов и характера происходящих офлайн процессов, в первую очередь, миграционных. Разумеется, сеть «ВКонтакте» охватывает преимущественно молодежную аудиторию, и представленные данные репрезентативны для косвенных оценок явлений, связанных с жизнью молодежи. Тем не менее в условиях отсутствия в России детальной статистики по миграционным потокам на микроуровне (между отдельными городами) именно данное направление получило первоначальный приоритет в изучении социальных сетей [Замятина, 2012; Замятина, Яшунский, 2012].

Теория

География населения оперирует в первую очередь данными официальной статистики, обращаясь к иным источникам в тех случаях, когда статистики не хватает или она вызывает сомнения. Помочь восполнить пробелы в статистике могут социологические методы — опросы и интервью, однако имеются и более неожиданные подходы. Еще более полувека назад для исследования географии населения на микроуровне в работе [Бухгольц, Маергойз, 1960] предлагалось использовать такие источники, как домовые книги и учетные карточки сотрудников на предприятиях. Подобные материалы, по своей сути в некоторых аспектах аналогичные анкетам переписи населения, могут содержать сведения, детализированность которых несопоставимо выше, чем у агрегированных статистических данных из официальных источников.

Методики прошлого века, трудно воспроизводимые в современных условиях, получают в некотором смысле новую жизнь, когда для исследований по географии населения привлекаются данные из анкет пользователей социальных сетей — они во многом воспроизводят те самые учетные карточки работников на предприятиях, по которым советские географы отслеживали миграции кадров между регионами.

Прецеденты использования анкет пользователей социальных сетей в исследовательских целях можно найти в зарубежных исследованиях: в работе [Wadhwa et al., 2009] профили пользователей сети LinkedIn использовались как база данных для формирования выборки респондентов в исследовании. Такое применение данных не вполне соответствует описанной выше парадигме дополнения официальной статистики, но все-таки демонстрирует, что подобные данные заслуживают определенного доверия.

Примечательно, что для исследования собственно аудитории социальных сетей используются преимущественно традиционные методы социологических опросов. Вместе с тем, например, в задачах отслеживания популярности брендов автоматическая обработка страниц пользователей социальных сетей и сбор соответствующей статистики вполне естественны.

Сужение выборки пользователей по возрасту и территориальному положению во многих случаях позволяет получить массивы анкет, покрывающие существенную часть исследуемой группы населения. Наиболее эффективно этот подход работает в отношении молодежи, которая ожидаемо преобладает среди пользователей

социальных сетей. В отдельных случаях, сужая выборку до выпускников отдельно взятой школы, можно получить результаты, сопоставимые с непосредственным анкетированием всех школьников определенных годов выпуска.

Подобные подходы использовались авторами при проведении нескольких тематических исследований. В работе [Замятина, 2012] рассматривалась образовательная миграция молодежи — анкеты пользователей (так же, как и в работе [Wadhwa et al., 2009]) обрабатывались вручную. Впоследствии переход к автоматической обработке позволил на основе более объемных массивов анкет детальнее исследовать образовательные миграции [Замятина, Яшунский, 2012] и миграции с Крайнего Севера России [Яшунский, Замятина, 2012].

Метод

Основные элементы технологии сбора данных описаны в работе [Чекмышев, Яшунский, 2014]. Для обработки анкет использовалось автоматизированное программное обеспечение, осуществляющее выборку из анкеты пользователя требуемой информации и занесение ее в базу данных. Здесь и далее, говоря об обработке анкет, мы почти всегда понимаем автоматическую обработку с использованием специально написанных для этого программ. Обработка многомиллионного массива анкет «вручную» представляется нам не только неосуществимой, но и бессмысленной.

Помимо самих анкет использовалась также вспомогательная информация, доступная в технических разделах сайта социальной сети «ВКонтакте», — список действующих (не удаленных) профилей пользователей, дополнительная информация о населенных пунктах, упоминаемых на страницах пользователей «ВКонтакте».

В отношении обрабатываемых анкет практически не применялось никаких процедур фильтрации — отсеивались только те, которые по техническим причинам не удалось разобрать (например, в результате ошибок передачи данных). Возможная недостоверность сведений в анкете ни при каких условиях не могла быть поводом для исключения анкеты из рассматриваемого массива данных.

У указанного подхода есть почти очевидные недостатки, его основным достоинством является снижение трудоемкости обработки, которая и без того потребовала значительного времени.

Принимая возможные дополнительные искажения, вносимые в данные в результате такого подхода, перечислим некоторые аргументы, позволяющие примириться с этими искажениями. Во-первых, официальная перепись населения также содержит искажения, которые часто делают ее результаты не вполне достоверными, что, впрочем, не служит поводом, чтобы отказаться от использования ее результатов. Во-вторых, искажения, вносимые за счет анкет, не соответствующих реальным людям, а заводимые программами-роботами для различных целей, не так значительны — анкеты роботов зачастую не содержат практически никаких сведений, тем самым не влияя на массив содержательных анкет. Наконец, выявление содержимого анкет, которое может не соответствовать действительности (например, указание неверного года рождения), будучи возможно технически, создает риск потери интересной информации, которая может быть забракована только потому, что она слишком нестандартна.

Во многих случаях аномалии в имеющихся данных могут быть обнаружены и относительно легко объяснены. Так, например, пользователи, возраст которых превышает сто лет (такая аномалия встречается примерно в 0,1% анкет), по-видимому, в действительности младше 14 лет (минимальный возраст регистрации в сети «ВКонтакте») и указали вместо своего года рождения год веком ранее. Выявление таких аномалий в каждом отдельном случае лежит в зоне ответственности исследователя, использующего данные.

Данные

Использованные данные анкет из социальной сети «ВКонтакте» собирались в январе — марте 2015 г. Обработывались только анкеты, находящиеся в публичном доступе (не требующие для просмотра регистрации в социальной сети), из них могла быть извлечена только информация, в отношении которой пользователи не указали, что она должна быть скрыта. При обработке анкет фиксировались значения следующих полей (при наличии): имя, дата рождения, место проживания, места обучения и время обучения (средние и высшие учебные заведения), дружеские связи пользователя. Основной объем обработки информации пришелся именно на дружеские связи: их оказалось около 3,5 млрд на 200 млн обработанных анкет пользователей.

Упомянутые в анкетах населенные пункты — места проживания и места нахождения учебных заведений — были затем сгруппированы по административным единицам второго уровня Российской Федерации: муниципальным районам и городским округам. Для осуществления этой операции использовалась информация о населенных пунктах, предоставляемая в сети «ВКонтакте» в качестве дополнительной к названию (обычно название района и/или области расположения). Это позволило в некоторой степени решить проблемы локализации населенных пунктов, названия которых встречаются на территории России более чем один раз.

Согласно полученным результатам, далеко не все пользователи социальной сети географически связаны с Россией: лишь примерно в 88 млн анкет место проживания или последнее место обучения (при отсутствии информации о месте проживания) оказалось расположенным на территории Российской Федерации. Информация из прочих анкет также была использована для формирования некоторых специальных наборов данных, но основная масса собранной и обработанной информации относится к упомянутому массиву из примерно 88 млн анкет.

Собранный объем информации оказался чрезмерно большим для непосредственных манипуляций, поэтому он был подвергнут дополнительной обработке — агрегации по различным параметрам, и на сайте «Виртуальное население России» представлены именно агрегированные наборы данных. Даже их объемы весьма внушительны, поэтому потребовалась специальная настройка программного обеспечения для их отображения на сайте в режиме реального времени.

После обработки были сформированы следующие группы тематических наборов данных.

1. Основная статистика (пол, год рождения, наличие высшего образования, число мест проживания пользователей).
2. Имена пользователей.

3. Школьное и иное вневузовское образование.
4. Высшее образование.
5. «Последний переезд».
6. Дружба пользователей.

На основе этих данных были построены карты, некоторые из которых представлены ниже.

При использовании этих данных следует учитывать ограничения, которые накладывает характер используемого материала. Любые выводы, которые делаются по данным о пользователях «ВКонтакте», либо относятся именно к совокупности «пользователи сети «ВКонтакте»», либо свидетельствуют о некотором явлении — если что-то проявилось в данных, есть некоторые основания полагать, что оно существует на самом деле. Вместе с тем, соблюдая должную аккуратность, можно выводить из имеющихся данных и какие-то количественные суждения. Чаще всего это удастся сделать для более узких (территориальных или, например, возрастных) групп, нежели вся совокупность жителей России. Даже эти суждения будут ограничены в своей достоверности и точности, но часто это лучше, чем ничего.

Как можно показать репрезентативность выборки при использовании данного атласа? Очевидно, нет строгих доказательств того, что все люди из какой-то территориальной или возрастной группы имеют равные шансы быть зарегистрированными в сети «ВКонтакте». Поэтому нам доступны косвенные аргументы, например, сравнение распределения по полу в выборке и в генеральной совокупности. В таком случае на вопрос о репрезентативности мы получим либо громкое «нет», если соотношение полов различается кардинально, либо тихое «возможно», если колебания можно объяснить случайностью выборки.

Для оценки репрезентативности пользователей «ВКонтакте» относительно населения России данные атласа сопоставлены с результатами переписи населения России 2010 г. Сопоставление населения и пользователей с территориальной группировкой по субъектам федерации заведомо вносит погрешности для Москвы и Санкт-Петербурга, однако в большинстве случаев оно представляется допустимым, особенно для предварительных оценок. Методика детальной оценки репрезентативности данных для отдельных возрастных групп и отдельных регионов детально охарактеризована в подразделе «Вопросы репрезентативности» раздела «Данные» на сайте «Виртуальное население России».

Оперируя с данными сайта, также следует понимать, что доля «интересных» анкет пользователей в действительности не так уж велика: существенная часть анкет, которые получили территориальную привязку к России, не содержат каких-либо нетривиальных сведений. Так, в частности, свой возраст указали лишь чуть более 35 % пользователей, а (хотя бы косвенная) информация о миграциях имеется лишь в 7 % анкет.

Анализ

В собранных данных просматриваются разнообразные закономерности и тенденции. Некоторые из них удастся связать с какими-то известными явлениями, другие же так и остаются загадкой. При этом велик соблазн воспринимать эти данные как некоторую выборку, характеризующую население России или его

часть, но подобные утверждения даже при сужении массива данных (например, при ограничении возрастной группы и места проживания пользователей), как отмечалось выше, всегда требуют тщательной проверки.

С уверенностью можно сказать только, что массив собранных данных описывает анкеты пользователей социальной сети «ВКонтакте» по состоянию на начало 2015 г. Наблюдаемые явления относятся именно к этой совокупности «виртуальных жителей» России и могут служить лишь ориентиром для поиска каких-то закономерностей, касающихся «реального» населения.

Виртуальная урбанизация

Существенные отличия между реальным и виртуальным населением становятся видны уже при сравнении количества пользователей с численностью населения муниципальных районов и городских округов. Распределение пользователей оказывается неравномерным: более значимые населенные пункты концентрируют пользователей, в то время как менее населенные теряют. Этот эффект наиболее заметен в агломерациях городов-миллионников. Так, например, в муниципалитетах, непосредственно соседствующих с Москвой, доля пользователей среди населения составляет всего до 20 %, в то время как в остальных районах области она ближе к 40 %, а в крупных городах области достигает 50 % (см. рис. 1).

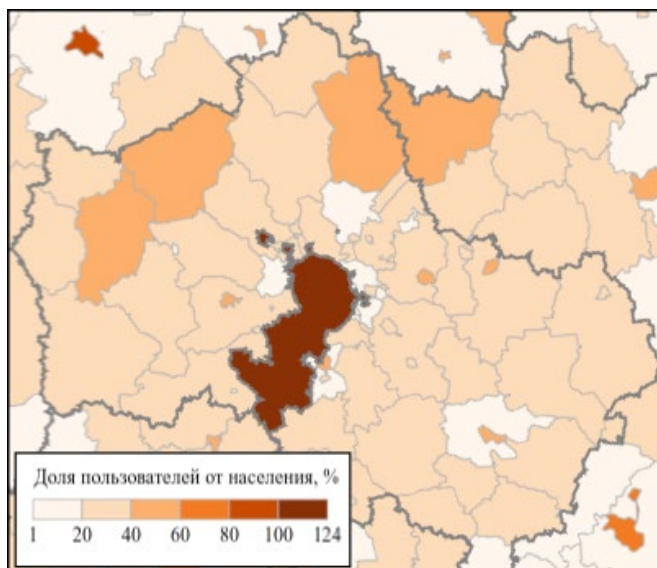


Рисунок 1. Соотношение числа пользователей и населения в Москве и Московской области²

При этом в Москве количество анкет пользователей превышает число официально зарегистрированных жителей. Причин для подобного эффекта, по-видимому, несколько. Низкая доля пользователей в соседних муниципалитетах позволяет

² Здесь и далее на рисунках представлены изображения, получаемые с помощью ресурса «Виртуальное население России».

предположить, что проживающие на периферии агломерации пользователи «приписывают» себя к ее центру. Тот же эффект наблюдается и для меньших городов: на рисунке 1 можно видеть характерные «провалы» вокруг Коломны, Твери, Рязани и рост доли пользователей непосредственно в городских округах (подробнее о явлении виртуальной урбанизации см. [Ивлиева, Яшунский, 2016]).

Возраст пользователей

Помимо локальных пространственных неоднородностей в распределении пользователей сети «ВКонтакте» имеются и некоторые глобальные неравномерности по различным параметрам. Так, например, в аудитории социальной сети, как и многих других интернет-сервисов, преобладают пользователи, которых можно по возрасту условно отнести к «молодежи», однако эта категория определяется по-разному в зависимости от региона. На европейской части России (см. рис. 2) имеется отчетливый тренд в сторону повышения медианного возраста пользователей «ВКонтакте» по мере продвижения с юга на север. Если в Дагестане медианный возраст пользователей равен 23 годам, то в Карелии он достигает 31 года. Эта диспропорция (возможно, чуть менее явно) сохраняется на всей территории России, и далеко не везде ее можно объяснить возрастной структурой населения. В частности, северные регионы, где доля молодежи достаточно высока (см., например, [Фаузер и др., 2016: 21]), тем не менее имеют более высокий медианный возраст пользователей, нежели регионы средней полосы. Это может рассматриваться как свидетельство того, что использование сети «ВКонтакте» представителями старших групп населения — более распространенная практика в северных регионах.

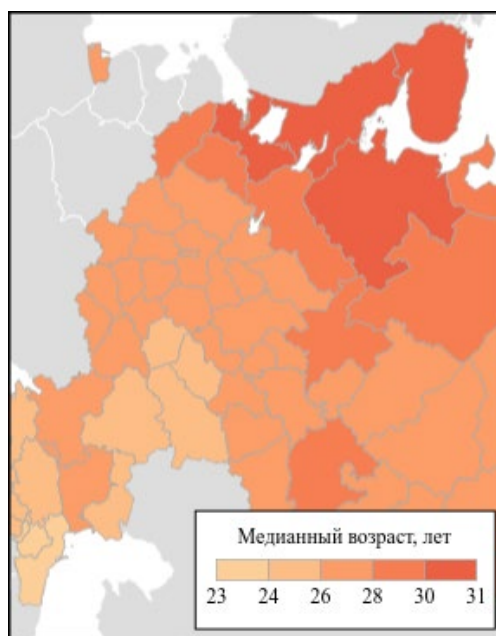


Рисунок 2. Медианный возраст пользователей по регионам европейской части России

С возрастом пользователей связано еще одно почти курьезное наблюдение. Год рождения, как и любое другое поле в анкете пользователя, может быть скрыт. Людей, предпочитающих не афишировать свой возраст, достаточно много — в большинстве регионов России более 50% от общего числа пользователей. Вместе с тем, существует ярко выраженная неравномерность доли скрывающих возраст при разделении пользователей по полу.

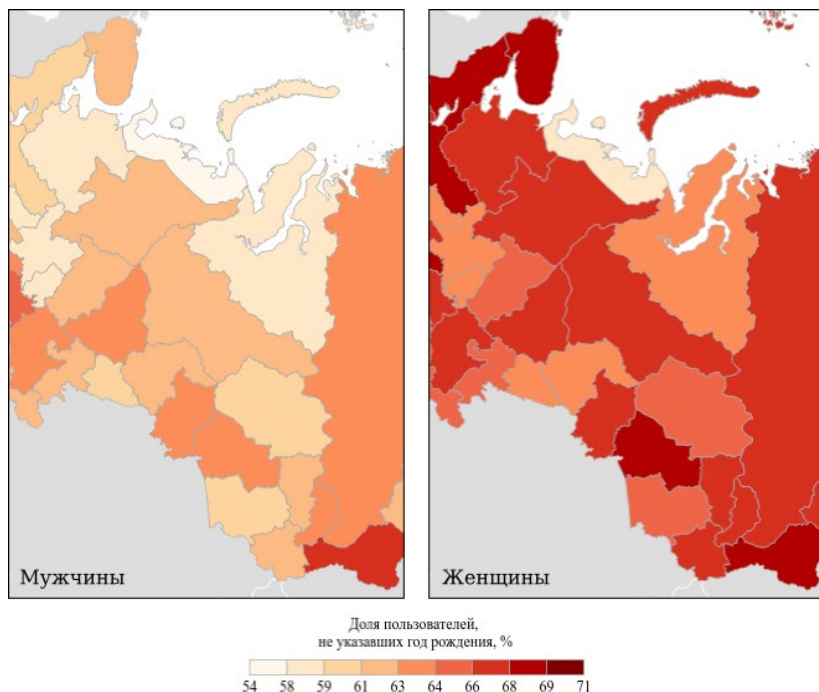


Рисунок 3. «Безвозрастные» пользователи по регионам Урала и Западной Сибири

Как видно из рисунка 3, в абсолютно подавляющем числе регионов доля женщин, скрывающих свой возраст, превышает соответствующую долю мужчин. Картина наблюдается повсеместно, регионы на рисунке 3 выбраны просто для примера.

Этнические особенности

Среди обработанных полей анкеты пользователя только поле «имя» могло бы позволить каким-то образом дифференцировать регионы России этническому принципу. На первый взгляд такой инструмент кажется достаточно грубым, и картосхема наиболее популярных имен по субъектам Российской Федерации подтверждает это опасение. За исключением республик Северного Кавказа повсеместно на территории России наиболее популярное мужское имя — Александр, а женское — Елена. При рассмотрении административно-территориальных единиц второго уровня картина становится более пестрой, но почти повсеместное пре-

обладание Александров и Елен сохраняется: вкрапления зон популярности иных имен выглядят почти бессистемными, хотя особенности национальных республик начинают проявляться не только на Северном Кавказе.

Несколько неожиданно на роль индикатора этнического своеобразия региона может претендовать доля имени Елена (или Александр) в административно-территориальной единице. Как видно из рисунка 4, Александр, оставаясь наиболее популярным именем в Башкирии и Татарстане (по-видимому, в первую очередь за счет жителей Уфы и Казани), не только «проигрывает» на уровне муниципальных образований другим более популярным в этих республиках именам, но и существенно теряет в популярности вообще. В некоторых районах его доля падает ниже 2% от общего числа пользователей мужского пола. Примечательно, что «провал» в популярности Александра почти точно повторяет границы субъектов федерации — именно это заставляет обратить внимание на этот показатель как на потенциальный индикатор этнических особенностей.

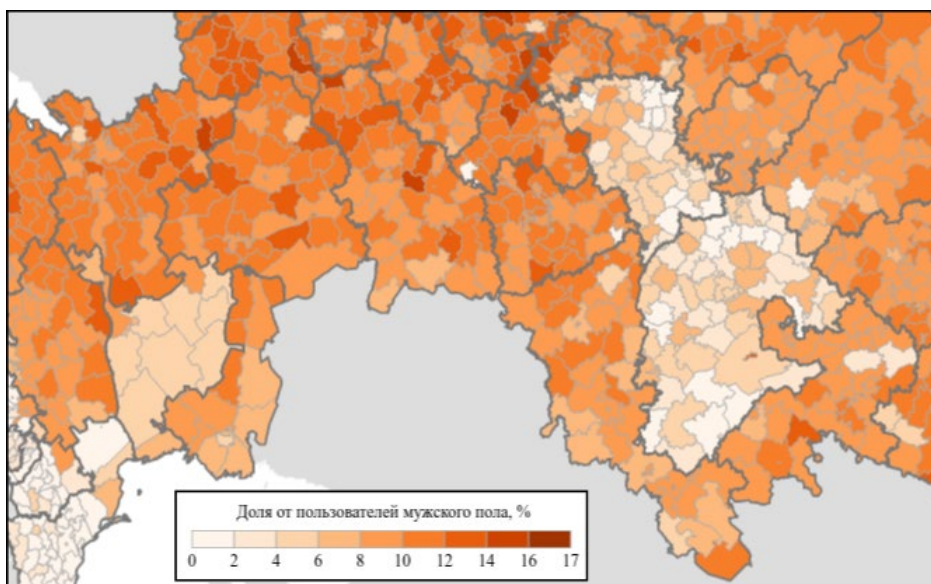


Рисунок 4. Доля пользователей с именем Александр по муниципальным образованиям Поволжья и Урала

Картина с падением популярности имен-лидеров федерального уровня воспроизводится во многих республиках Российской Федерации. В частности, на рисунке 4 так же отчетливо, как Татарстан и Башкирия, видна и Калмыкия. Примечательно, что в соседней с Татарстаном Удмуртской республике имя Александр достаточно популярно. Этот контраст можно рассматривать как косвенный индикатор уровня ассимиляции³.

³ Благодарим Анну Герасимову за идею интерпретации.

Привлекательность вузов

Исследование образовательных миграций, в особенности переездов для учебы в вузе непосредственно после окончания школы, по-видимому, наиболее очевидное применение данных из анкет пользователей «ВКонтакте». Именно эти миграции отображаются в анкете непосредственно, в то время как об иных перемещениях можно судить по косвенным данным, сопоставляя места учебы и текущее место проживания.

Вопрос о том, из каких регионов поступают студенты в вузы города, имеет значение для характеристики экономико-географического положения города. Примечательно, что для ответа на него в работе [Маергойз, 1956] предлагается использовать студенческие анкеты, хранящиеся в архиве вуза. Анкеты пользователей социальных сетей можно рассматривать как их практически полноценные современные аналоги.

Массив данных, представленный на сайте, позволяет анализировать образовательные миграции весьма детально. Более того, для использования в дальнейших исследованиях на сайте также представлены более детальные наборы данных, чем те, что используются при формировании интерактивных картосхем.

Приведенный ниже пример иллюстрирует возможности ресурса для анализа привлекательности вузов, географические зоны преимущественного набора абитуриентов, причем по разным специальностям. Для примера приведены две картосхемы, показывающие соотношение количества абитуриентов, поступивших в вузы Новосибирской области (преимущественно непосредственно Новосибирска) на специальности технической и гуманитарной направленности (см. рис. 5). С помощью настроек интерактивного атласа получена пара карт географического распределения всех, кто когда-либо учился в вузах Новосибирской области на технических и гуманитарных специальностях, по регионам окончания школы. Рисунок 5 представляет собой увеличенные фрагменты данных карт. При их сопоставлении бросается в глаза гендерная дифференциация абитуриентов (юноши преобладают на технических специальностях, девушки — на гуманитарных) — на всех диаграммах карты по распределению мест рождения студентов технических специальностей преобладает синий цвет (юноши), гуманитарных — красный (девушки). Однако сопоставление данной пары карт позволяет сделать и более интересные выводы — например, о том, что роль Новосибирска как межрегионального образовательного центра по-разному проявляется для разных отраслей знаний. На факультеты технической направленности поступили сотни студентов из различных, и не только соседних, регионов Сибири и Дальнего Востока (внушительные кружки на фоне многих регионов свидетельствуют о значительной численности уроженцев многих регионов Сибири, поступивших в новосибирские вузы); атлас позволяет при необходимости дать количественные оценки (см. примечание о репрезентативности данных). На гуманитарных специальностях студенты из других областей встречаются существенно реже (круговые значки, отражающие число уроженцев конкретных регионов, поступивших в Новосибирск, почти незаметны). Сопоставление двух картосхем позволяет предположить, что факультеты гуманитарной направленности имеют локальное значение, «обслуживая» лишь саму Новосибирскую область, в то время как

технические факультеты формируют славу Новосибирска как образовательного центра федерального уровня.

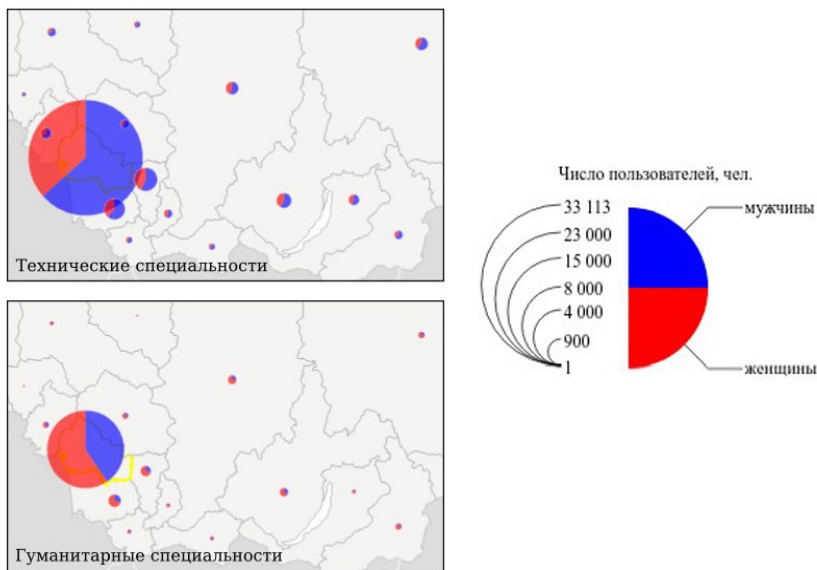


Рисунок 5. «Зоны сбора» студентов в вузы Новосибирской области

Миграционные потоки

Данные, полученные в результате обработки анкет пользователей сети «ВКонтакте», не аналогичны данным о миграциях в официальной статистике. В анкетах пользователей содержится информация о месте жительства, а также о местах и времени обучения в различных учебных заведениях: на основании этой информации можно частично восстановить последовательность пространственных перемещений пользователя. Располагая места учебы хронологически и добавляя в конце место жительства (если оно указано), получаем «траекторию» перемещений пользователя, в которой можно выбрать последнее по времени перемещение — «последний переезд». Рассматриваемые далее данные о миграционных потоках представляют собой общее число переехавших из одного региона в другой в рамках последнего (хронологически) переезда, просуммированное по всем анкетам пользователей. Таким образом, данные, представленные на сайте, позволяют выявить относительное преобладание одних направлений миграционных потоков над другими, но не дают возможности оценить собственно объем миграции. Картина, получающаяся в результате наложения множества факторов, тем не менее позволяет сделать некоторые наблюдения.

Общемировая тенденция концентрации населения в крупных агломерациях, особо сильно проявляющаяся за пределами развитых стран [Towards a more urban world..., 2015], а также, по-видимому, сверхцентрализация Российской Федерации

находят отражение в миграционных потоках, которые в первую очередь направлены в центральные мегаполисы — Москву и Санкт-Петербург.

Преобладание столь сильное, что на его фоне практически незаметны прочие перемещения населения. Однако и в этих «первых среди последних» миграционных потоках просматриваются любопытные закономерности.

Потоки, не ведущие в Москву или Санкт-Петербург, распределены по территории России нерегулярно: можно выявить «притягивающие» регионы, а также сети регионов с интенсивным миграционным обменом.

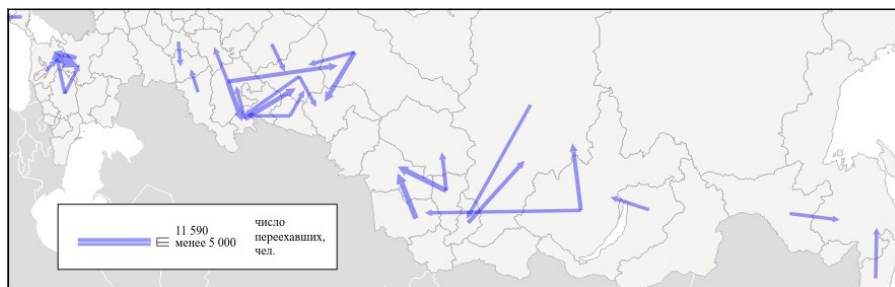


Рисунок 6. Тридцать крупнейших миграционных потоков между регионами России, исключая Москву и Санкт-Петербург

При рассмотрении тридцати крупнейших интегральных миграционных потоков можно выявить следующие области притяжения: Хабаровский край собирает приезжих из соседних Приморского края и Амурской области, в Сибири потоки направлены в сторону Красноярского края и Новосибирской области, разнонаправленная сеть связывает регионы Уральского федерального округа — Башкирию, Челябинскую область, Пермский край, Свердловскую область, Тюменскую область и входящие в нее автономные округа. На юге России центрами притяжения оказываются Краснодарский край и Ростовская область.

В представленной картине достаточно слабо заметен так называемый западный дрейф [Мкртчян, 2015]: в миграционных потоках, направленных не в столичные города, преимущественно проявляется локальное перераспределение населения между регионами, нежели общее движение с восточных окраин в сторону центрального региона. Примечательно, что среди рассмотренных крупных потоков ни один не оказался направлен в Московскую и Ленинградскую области. По-видимому, переезжая в агломерацию, пользователи предпочитают указывать в качестве места своего жительства ее центр.

Данные о миграционных потоках между административно-территориальными единицами второго уровня (муниципальными районами и городскими округами) не имеют аналогов в официальной статистике и позволяют с высокой точностью определять преобладающие направления переездов жителей отдельно взятых городов. Подобные сведения использовались при исследовании миграционного поведения выходцев из Норильска [Замятина, Яшунский, 2015; Zamyatina, Yashunsky, 2017]. В таблице 1 представлены десять крупнейших миграционных потоков из Норильска, вычисленные по данным анкет пользователей «ВКонтакте».

Таблица 1. **Миграция из Норильска**

Направление миграции	Число мигрантов
Санкт-Петербург	1673
Москва	1658
Красноярск	1281
Новосибирск	348
Нижний Новгород	290
Белгород	259
Краснодар	251
Екатеринбург	226
Ростов-на-Дону	178
Воронеж	163

Не позволяя точно оценить число переехавших (тем более за определенный период времени), эти данные заставляют обратить внимание на особую роль Санкт-Петербурга (поток превышает поток в Москву), Белгорода и Краснодара (поток превышает более многолюдный и более близкий Екатеринбург) для жителей Норильска.

Межрегиональные социальные связи

Как и в случае миграционных потоков, основная масса межрегиональных дружеских связей приходится на Москву и Санкт-Петербург. Все прочие связи меркнут на их фоне: в первой сотне наиболее значимых межрегиональных связей только в восьмом десятке появляются связи, в которых не участвует один из столичных городов. Вместе с тем, картина этих «внестоличных» связей достаточно интересна (см. рис. 7). В качестве «центров дружбы» явно выделяются Краснодар и Новосибирск, а регионы Урала и частично Западной Сибири оказываются связаны мощной сетью. Возникновение именно такой конфигурации межрегиональных связей, по-видимому, связано не в последнюю очередь с направлениями миграционных потоков: собственно, дружеские связи часто повторяют картину преобладающих направлений миграций. Подобное наблюдение представляется вполне ожидаемым, но этот вопрос, вероятно, требует более детального исследования.

Помимо дружеских связей на уровне регионов Российской Федерации на сайте также представлена информация о дружеских связях на уровне муниципальных районов и городских округов. Этот массив информации, до сих пор практически неисследованный, вероятно, кроме очевидных наблюдений (преобладание дружеских связей со столицами) позволит выявить какие-то более интересные скрытые закономерности.

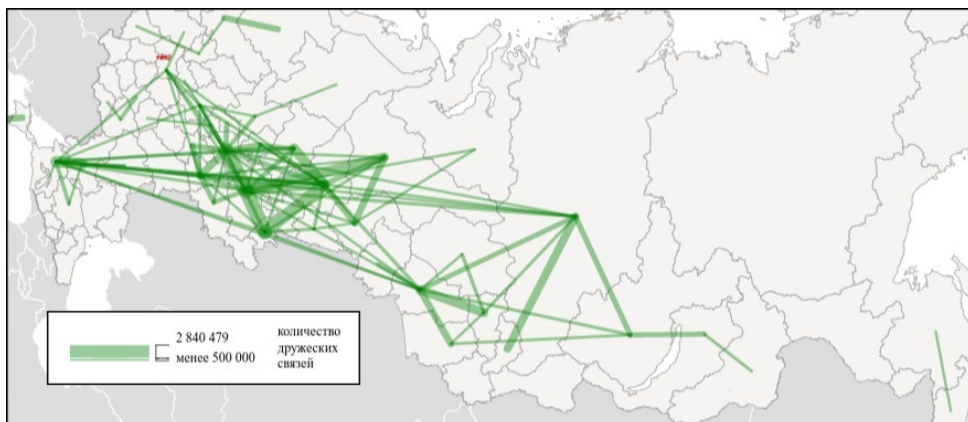


Рисунок 7. Наиболее значимые дружеские связи между регионами России, исключая Москву и Санкт-Петербург

Заключение

Опубликованные массивы данных во многих аспектах не имеют аналогов и предназначены для широкого круга исследователей. Представленные в данной работе наблюдения не претендуют на полноту характеристики спектра возможных путей использования данных социальных сетей для анализа социальных явлений. В атласе на сайте «Виртуальное население России» каждый из возможных к построению типов картосхем сопровождается пояснениями и рекомендациями по возможным направлениям использования сформированного массива данных.

Однако десятки возможных вариантов наблюдений группируются в два крупных магистральных направления. Первое — это исследование собственно структуры социальных сетей, их распространения, особенностей наполнения и функционирования в разных регионах и городах страны, в разных типах населенных пунктов (например, в крупных и малых городах, на периферии и т. д., в национальных республиках и др.). Второе направление — это оценка миграционных потоков и других социальных явлений, для которой материалы социальных сетей, собранные на сайте «Виртуальное население России», являются косвенным источником данных. Парадоксально, но до настоящего времени машинным образом собранные большие данные по структуре социальных сетей использовались именно в этом (в общем-то, побочном) направлении, см. [Замятина, 2012; Замятина, Яшунский, 2012] и др. Это связано с катастрофической нехваткой оперативной официальной статистики: например, матрицы миграций «регион — регион», не говоря уже о матрицах «город — город», давно доступных многим нашим зарубежным коллегам, — именно поэтому данные социальных сетей вынужденно стали источниками данных для хотя бы оценочных суждений о локальных миграционных потоках.

Мы надеемся, что публикация интерактивного атласа социальных сетей станет отправной точкой исследований виртуальной географии в целом, откроет дорогу к исследованию региональной специфики функционирования социальных сетей

в целом — задача, ранее недоступная из-за трудоемкости и ресурсоемкости сбора больших данных. К сожалению, действовать нужно быстро: данные по тем же причинам обновить практически невозможно, и представленный ресурс, отражающий полный срез крупнейшей российской социальной сети «ВКонтакте» по состоянию на 2015 г., взывает к срочному исследованию.

Он может быть полезен, как уже говорилось, и для косвенных наблюдений социальных явлений. Здесь задача ресурса — показать общую картину на уровне детальности, пока недоступной в официальных статистических источниках. Мы надеемся, что представленная информация не только послужит источником дополнительных сведений по различным вопросам, но и позволит ставить и решать новые исследовательские задачи в различных областях знаний.

Список литературы (References)

Бухгольц Э. О., Маергойз И. М. Крупномасштабное изучение городов // Методы географических исследований. Сб. статей. М., 1960. С. 333—345.

Bukhgoltz E. O., Maergoyz I. M. (1960) Studying cities in large scale. In: Geographical research methods. Collection of articles. P. 333—345. (In Russ.)

Замятина Н. Ю. Метод изучения миграций молодежи по данным социальных Интернет-сетей: Томский государственный университет как «центр производства и распределения» человеческого капитала (по данным социальной Интернет-сети «ВКонтакте») // Региональные исследования. 2012. Вып. 2 (36). С. 15—28.

Zamyatina N. Yu. (2012) Studying youth migrations through social networks data: Tomsk State University as a 'centre that produces and distributes' human capital (based on Vkontakte social network data). *Regional research*. Vol. 36. No. 2. P. 15—28. (In Russ.)

Замятина Н. Ю., Пилясов А. Н. Россия, которую мы обрели: исследуя пространство на микроуровне. М.: Хронограф. 2013.

Zamyatina N. U., Pelyasov A. N. (2013) The Russia we have acquired: microexamination of space. *Chronograph*. (In Russ.)

Замятина Н. Ю., Яшунский А. Д. Межрегиональные центры образования // Отечественные записки. 2012. Вып. 3 (48). С. 74—84.

Zamyatina N. U., Yashunsky A. D. (2012) Interregional education centres. *Otechestvennye zapiski*. Vol. 48. No. 3. P. 74—84. (in Russ.)

Замятина Н. Ю., Яшунский А. Д. Миграции с Севера: социальные сети и ментальная «близость» // Внеэкономические факторы пространственного развития. Сб. статей / отв. ред. Стрелецкий В. Н. М.: Эслан, 2015. С. 147—173.

Zamyatina N. U., Yashunsky A. D. (2015) Migrations from the North: social networks and mental 'proximity'. In: Non-economic factors of spatial development. Collection of articles. P. 147—173. (in Russ.)

Ивлиева О. Д., Яшунский А. Д. Виртуальная урбанизация // Городские исследования и практики. 2016. Т. 1(4). С. 26—36.

Ivlieva O. D., Yashunsky A. D. (2016) Virtual urbanization. *Urban studies and practices*. Vol. 1. No. 4. P. 26—36. (in Russ.)

Маергойз И. М. К экономико-географическому изучению городов // Вопросы географии. Сб. 38: География городов. М. 1956. С. 5—26.

Maergoyz I. M. (1956) Studying cities in economical and geographical aspects. *Geographical issues*. Vol. 38. City geography. M. P. 5—26. (in Russ.)

Мкртчян Н. В. Пространственные особенности внутрироссийской миграции в постсоветский период // Современные исследования миграции населения. М. : Экономический факультет МГУ им. М. В. Ломоносова. 2015. С. 94—111.

Mkrtychyan N. V. (2015) Spatial peculiarities of internal migration in Russia in the Post-Soviet era. In: Contemporary research of migration: Collection of articles. P. 94—111. (in Russ.)

Пилясов А. Н. (ред.) Синергия пространства: региональные инновационные системы, кластеры и перетоки знания / отв. Ред. А. Н. Пилясов. Смоленск: Ойкумена. 2012.

Pelyasov A. N. (2012) Synergy in space: regional innovation systems, clusters and knowledge spillovers. Ed. by A. N. Pelyasov. Smolensk: Oecumene. (in Russ.)

Утков П., Замятина Н. Ю. Ссылки между сайтами университетов: территориальная структура виртуального пространства // Российская глубинка — модели и методы изучения. Сборник статей. М. : Эслан, 2012. С. 451—458.

Utkov P., Zamyatina N. U. (2012) Web links between universities' websites: territorial structure of virtual space. In: Russian province: models and ways of studying it. Collection of articles. P. 451—458. (In Russ.)

Фаузер В. В., Лыткина Т. С., Фаузер Г. Н., Залевский В. А. Население северных регионов: от количественных показателей к качественному измерению. Сыктывкар : Изд-во СГУ им. Питирима Сорокина, 2016. URL: http://vvfauzer.ru/pub/mon/m_2016_1.pdf (дата обращения 25.11.2017).

Fauzer V. V., Lytkina T. S., Fauzer G. N., Zalevsky V. A. (2016) Northern regions: from quantity to quality measuring. Syktyvkar State University. http://vvfauzer.ru/pub/mon/m_2016_1.pdf (accessed 25.11.2017). (in Russ.)

Чекмышев О. А., Яшунский А. Д. Извлечение и использование данных из электронных социальных сетей // Препринты ИПМ им. М. В. Келдыша. 2014. Вып. 62. URL: <http://library.keldysh.ru/preprint.asp?id=2014-62> (дата обращения: 25.11.2017).

Chekmyshev O. A., Yashunsky A. D. (2014) Extraction and usage of online social network data. Preprint of the Keldish institute of applied mathematics. No. 62. URL: <http://library.keldysh.ru/preprint.asp?id=2014-62> (accessed 25.11.2017). (in Russ.)

Яшунский А. Д., Замятина Н. Ю. Севера как зона роста российской провинции // Отечественные записки. 2012. Вып. 5 (50). С. 227—239.

Yashunsky A. D., Zamyatina N. U. (2012) Northern region as Russian province zone of development. *Otechestvennyye zapiski*. Vol. 50. No. 5. P. 227—239. (in Russ.)

Adekannbi J. (2011) Web link analysis of interrelationship between top ten African universities and world universities. *Annals of Library and Information Studies*. Vol. 58. No. 2. P. 128—138.

Agibetova U., Samson I. (2008) The Metropolization of the FSU Temptative Measurement via the Method of Hyperlinks Notoriety. In: Eurasian Integration Yearbook, Almaty: Eurasian Development Bank. P. 115—135.

Almind T. C., Ingwersen P. (1997) «Informetric analyses on the World Wide Web: methodological approaches to 'webometrics'». *Journal of Documentation*. Vol. 53. No. 4. P. 404—426. <http://doi.org/10.1108/EUM0000000007205>.

Anselin L., Varga A., Acs Z. (1997) Local Geographic Spillovers between University Research and High Technology Innovations. *Journal of Urban Economics*. No. 42. P. 422—448. <https://doi.org/10.1006/juec.1997.2032>.

Audretsch D., Lehmann E., Warning S. (2004) University spillovers: does the kind of science matter? *Industry and Innovation*. Vol. 11. No. 3. P. 193—205. <https://doi.org/10.1080/1366271042000265375>.

Björneborn L., Ingwersen P. (2004) Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*. Vol. 55. No. 14. P. 1216—1227. <https://doi.org/10.1002/asi.20077>.

Feldman M., Kogler D. (2010) Stylized facts in the geography of innovation. In: Handbook of the Economics of Innovation. Vol. 1. P. 381—410.

Feldman M. (1999) The New economics of innovation, spillovers and agglomeration: a review of empirical studies. *Economics of Innovation and New Technology*. Vol. 8. No. 1—2. P. 5—25. <https://doi.org/10.1080/10438599900000002>.

Koltsova O., Koltcov S., Sinyavskaya Y. (2017) When Internet Really Connects Across Space: Communities of Software Developers in Vkontakte Social Networking Site. In: Ciampaglia G., Mashhadi A., Yasseri T. (eds.) Social Informatics. Lecture Notes in Computer Science. Vol. 10540. P. 431—442. https://doi.org/10.1007/978-3-319-67256-4_34.

Kousha K. (2005) Webometrics and Scholarly Communication: An Overview. *Quarterly of Library and information Science*. Vol. 14. No. 4. P. 7—16.

Maggioni, M.A., Uberti, T.E. (2009) Knowledge networks across Europe: which distance matters? *Annals of Region Science*. Vol. 43. No. 3. P. 691—720. <https://doi.org/10.1007/s00168-008-0254-7>.

Mamatov T., Samson I. (2009) Metropolisation and enlargement of European Union: an analysis by the hyperlinks with filters. In: Territorial Cohension of Europe and Integrative Planning. 49-th European Congress of the Regional Science Association International 25th-29th Agust 2009. Book of Abstracts. P. 299—300.

Takhteyev Yu., Gruzd A., Wellman B. (2012) Geography of Twitter networks. *Social Networks*. No. 34. P. 73—81. <https://doi.org/10.1016/j.socnet.2011.05.006>.

Theilwall M. (2002) A research and institutional size-based model for national university Web site interlinking. *Journal of Documentation*. Vol. 58. No. 6. P. 683—694. <https://doi.org/10.1108/00220410210448219>.

Theilwall M. (2008a), Social networks, gender, and friending: An analysis of MySpace member profiles. *Journal of the American Society for Information Science and Technology*. Vol. 59 No. 8. P. 1321—1330. <https://doi.org/10.1002/asi.20835>.

Theilwall M. (2008b) Bibliometrics to Webometrics. *Journal of Information Science*. Vol. 34. No. 4. P. 605—621. <https://doi.org/10.1177/0165551507087238>.

Theilwall M. (2009) Introduction to webometrics. Quantitative web research for the social science. San New York: Morgan & Claypool. <https://doi.org/10.2200/S00176ED1V01Y200903ICR004>.

Towards a more urban world (GMT 2). (2015) SOER. 2015. URL: <https://www.eea.europa.eu/soer-2015/global/urban-world> (accessed 25.11.2017).

Wadhwa V., Saxenian A., Freeman R., Gereffi G., Salkever A. (2009) America's Loss is the World's Gain: America's New Immigrant Entrepreneurs, Part 4, SSRN. URL: <http://ssrn.com/abstract=1348616> (accessed 25.11.2017). <https://doi.org/10.2139/ssrn.1348616>.

Zamyatina N., Yashunsky A. (2017) Migration cycles, social capital, and networks. A new way to look at Arctic mobility. In: *New mobilities and social changes in Russia's Arctic regions*. Ed. Laruelle M. Routledge research in Polar Regions. P. 59—84.