А.Д. Яшунский

**Население, которое себя
переписывает: данные
из социальных сетей**

Рекомендуемая форма библиографической ссылки

Яшунский А.Д. Население, которое себя переписывает: данные из социальных сетей // Проектирование будущего. Проблемы цифровой реальности: труды 2-й Международной конференции (7-8 февраля 2019 г., Москва). — М.: ИПМ им. М.В.Келдыша, 2019. — С. 144-150. — URL: <https://keldysh.ru/future/2019/14.pdf> doi:[10.20948/future-2019-14](https://doi.org/10.20948/future-2019-14)

Население, которое себя переписывает: данные из социальных сетей

А.Д. Яшунский

ИПМ им. М.В.Келдыша РАН

Аннотация. В различных социологических, демографических и географических исследованиях недостаток статистической информации о населении приходится восполнять из альтернативных источников. В современном мире таким источником могут служить профили пользователей электронных социальных сетей. На примере данных, полученных для проекта «Виртуальное население России», показано, какого рода наблюдения можно сделать, используя подобную неофициальную статистику.

Ключевые слова: демография, география населения, перепись, электронные социальные сети

A population self-census: social network data

A.D. Yashunsky

Keldysh institute of applied mathematics, RAS

Abstract. Research in different areas of sociology, demography and geography often calls for alternative sources of information to compensate for the lack of official statistical data. A source that is available in the modern world is the collection of user profiles from on-line social networking services. We use the data collected for the project “Virtual population of Russia” to demonstrate what kind of observations can be made using such unofficial statistics.

Keywords: demography, population geography, census, social networking services

Всеобщая перепись, а также иные, менее глобальные инструменты учета населения используются, в первую очередь, для обеспечения информацией органов власти. Вместе с тем результаты этого учета являются первичными данными для исследователей в таких областях, как демография, экономика, социальная и экономическая география. Хотя качество и подробность собираемой статистики меняется от государства к государству и в каких-то странах исследователям доступны данные, о

4. Математические модели цифрового мира

которых их зарубежные коллеги могут только мечтать, ситуации, в которых интересы ученых выходят за рамки имеющихся данных, встречаются практически повсеместно. В случаях, когда государственной статистики оказывается недостаточно для целей научного исследования, ученым приходится изыскивать иные источники информации.

Так, например, еще в 1950-х годах И.М. Маергойз для исследования миграций собирал информацию из учетных карточек сотрудников крупных предприятий, получая к ним доступ через отделы кадров (см. [1]). В современных реалиях такой метод вряд ли реализуем, однако появились иные источники информации, которые если и не могут конкурировать с официальной государственной статистикой, то уж точно могут ее дополнить.

Первые электронные социальные сети появились в конце 1990-х годов, и их значимость для пользователей сети Интернет с тех пор неуклонно росла. Примечательно, что хотя сам термин «социальная сеть» существовал в социологии задолго до появления Интернета, сейчас это словосочетание используется, в первую очередь, для обозначения электронных (онлайновых) социальных сетей. Естественно, электронные социальные сети привлекли внимание тех социологов, которые ранее исследовали «оффлайновые» социальные сети, но вскоре стало понятно, что имеющаяся в электронных социальных сетях информация может быть полезна и в иных исследованиях. Так, например, в работе [2] отбор участников для социологического исследования осуществлялся путем просмотра их профилей в социальной сети, а в работе [3] подход И.М. Маергойза из работы [1] перенесен в современность с использованием профилей социальной сети «ВКонтакте» вместо учетных карточек из отдела кадров.

Постепенно использование данных из социальных сетей перестает быть чем-то экзотическим, а становится одним из методов сбора информации в узкоспециализированных исследованиях, для которых недостаточно «обычных» статистических демографических данных. Эта тенденция сохраняется даже при том, что изначально ничем не отличавшиеся от иной информации в Интернете данные из социальных сетей сейчас получают все больше и больше дополнительных «степеней защиты», и массовый сбор данных, находившихся в открытом доступе десять лет назад, на сегодняшний день зачастую невозможен. Альтернативой самостоятельному сбору данных в интернете оказывается непосредственное взаимодействие с компаниями, которым принадлежат электронные социальные сети. Как видно из различных публикаций, предоставление анонимизированных данных (из которых исключена информация, позволяющая идентифицировать какого-либо конкретного пользователя сети) рассматривается владельцами социальных сетей как вполне допустимая форма коммерции, а в некоторых случаях

анонимизированные данные могут предоставляться и на безвозмездной основе (см. [4,5]). Вероятно, к такому подходу рано или поздно прибегнут и владельцы российских социальных сетей, хотя сейчас все еще существенная часть профилей пользователей (в частности, пользователей сети «ВКонтакте» vk.com) находится в открытом доступе и может быть проанализирована без обращения к владельцам социальной сети.

Возможности глобального исследования массива профилей пользователей социальной сети были рассмотрены О.А. Чекмышевым и А.Д. Яшунским в работе [6]. Предложенные подходы для проведения переписи «онлайнового» населения были впоследствии воплощены в жизнь, а результаты сбора данных представлены на сайте «Виртуальное население России» webcensus.ru.

Собранный массив данных оказался настолько велик, что даже минимальная его обработка требовала значительных аппаратных средств (объемов памяти и процессорного времени). Тем более затруднительно было бы осуществлять такую обработку данных в режиме реального времени для обеспечения пользователям интерактивного доступа к данным. В связи с этим для публикации на сайте данные были предварительно агрегированы с учетом наиболее типичных исследовательских задач в области географии населения. На основе агрегированных наборов данных (доступных для скачивания на сайте проекта) был построен ряд интерактивных картосхем, позволивших сделать некоторые интересные наблюдения, часть из которых приведена ниже. Зачастую эти наблюдения не подвергались дополнительным проверкам и поэтому имеют фактически статус научных гипотез, но они дают представление о характере информации, которая может быть извлечена из профилей пользователей социальных сетей.

Первое наблюдение заключается в том, что доля пользователей от населения несколько неожиданным образом ведет себя в окрестностях крупных населенных пунктов (см. рис. 1).

Эффект, наиболее ярко проявляющийся рядом с центрами крупных агломераций, но сохраняющийся вплоть до центров муниципальных районов, заключается в том, что пользователи предпочитают приписывать себя центру своей агломерации, нежели указывать свое истинное место жительства: большинство пользователей из Балашихи указывают местом жительства Москву. В результате в крупных населенных пунктах доля пользователей (не вполне оправданно) увеличивается, в их непосредственной окрестности резко падает, а по мере удаления – вновь возрастает. Подробнее об этой «виртуальной урбанизации» см. в работе [7].

В собранном массиве данных мало проявляются этнические особенности пользователей. Практически единственное, что позволяет как-то отследить этнокультурную дифференциацию пользователей, – это их

4. Математические модели цифрового мира

имена. Однако картосхема наиболее популярных имен в различных субъектах Российской Федерации несколько разочаровывает своим единообразием: везде, за исключением некоторых регионов Северного Кавказа, наиболее популярное мужское имя – Александр, а наиболее популярное женское – Елена. Различия проявляются, если рассмотреть долю имени-«лидера» от общего числа пользователей (см. рис. 2).

Рис. 1. Доля пользователей «ВКонтакте» от населения в муниципальных районах Московской области: изображение, получаемое с помощью ресурса <http://webcensus.ru>

Рис. 2. Доля пользователей с именем Александр в муниципальных районах южных регионов Российской Федерации: изображение, получаемое с помощью ресурса <http://webcensus.ru>

В республиках Башкирия, Калмыкия, Татарстан пользователи мужского пола с именем Александр хоть и преобладают, но их доля существенно меньше, чем в соседних регионах. Различие наиболее заметно на уровне муниципальных районов – фактически, даже не зная, какие районы входят в состав этих республик, это можно было бы восстановить по доле Александров среди пользователей. Примечательно, что соседняя с Башкирией и Татарстаном Удмуртия практически не выделяется по аналогичному показателю, позволяя предположить ее существенно меньшую этническую обособленность.

Информация о перемещениях пользователей формировалась на основе сопоставления их места жительства с местами учебы, в результате чего определялось направление «последнего переезда». Полученная таким образом информация несет несколько иной смысл, нежели официальная миграционная статистика, но позволяет сделать некоторые наблюдения о миграционных потоках в стране. Крупнейшие потоки, как можно было предположить, направлены из регионов в столицы – Москву и Санкт-Петербург. Однако, если исключить эти потоки, картина оказывается достаточно интересной (см. рис. 3).

Рис. 3. Тридцать крупнейших миграционных потоков (направления последнего переезда) пользователей за исключением потоков в Москву и Санкт-Петербург: изображение, получаемое с помощью ресурса <http://webcensus.ru>

Становится понятно, что «западный дрейф» населения (см. [8]) практически никак не проявляется, сменившись движением в сторону крупных агломераций, в первую очередь – столичных, но не только. Так, на Дальнем Востоке явным центром притяжения оказывается Хабаровск.

4. Математические модели цифрового мира

Рис. 4. Наиболее значимые дружеские связи между пользователями регионов, исключая Москву и Санкт-Петербург: изображение, получаемое с помощью ресурса <http://webcensus.ru>

Примером исследования дружеских связей пользователей социальных сетей может служить, например, работа [4] (данные о пользователях Facebook), однако в целом исследования в этой области еще только начинаются. Данные проекта webcensus.ru позволяют оценить интенсивность межрегиональной дружбы в России. Преобладание связей с Москвой и Петербургом проявляется и в этом случае, но картосхема, исключая эти преобладающие связи (см. рис. 4), позволяет выявить интересные нетривиальные закономерности. В частности, бросаются в глаза тесные связи между уральскими регионами. Этот феномен, как и многие другие, в том числе еще не выявленные, ждет своих исследователей.

Литература

1. Бухгольц Э.О., Маергойз И.М. Крупномасштабное изучение городов // Методы географических исследований. Сб. статей. – М., 1960. С.333–345.
2. Wadhwa V., Saxenian A., Freeman R., Gereffi G., Salkever A. America's Loss is the World's Gain: America's New Immigrant Entrepreneurs, Part 4 // SSRN, 2009, 2 марта. URL: <http://ssrn.com/abstract=1348616>
3. 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, 2(36), с.15–28. URL: [http://media.geogr.msu.ru/RI/RI_2012_02\(36\).pdf](http://media.geogr.msu.ru/RI/RI_2012_02(36).pdf)
4. Yearwood M.H., Cuddy A., Lamba N. et al. On wealth and the diversity of friendships: High social class people around the world have fewer

- international friends // Personality and Individual Differences. 2015. V.87, p.224-229. – DOI:10.1016/j.paid.2015.07.040
5. Dataverse : Social science one. URL: <https://socialscience.one/facebook-dataverse>
 6. 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>
 7. Ивлиева О.Д., Яшунский А.Д. Виртуальная урбанизация // Городские исследования и практики. 2016. Т.1(4). С.26–36. URL: https://usp.hse.ru/data/2017/05/31/1171921807/ГИиП_2016_04_compressed.pdf#page=26
 8. 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. URL: <https://www.econ.msu.ru/sys/raw.php?o=30120&p=attachment>