



**Электронный научно-образовательный журнал «История».
2013-2024**

ISSN 2079-8784

URL - <http://history.jes.su>

Все права защищены

Выпуск 12 (134). Часть II Том 14 - Актуальные вопросы исторической науки и образования. 2023

Перспективы использования «больших данных» для анализа миграционных процессов

Гребенюк Александр Александрович

МГУ им. М. В. Ломоносова

Российская Федерация, Москва

Алешковский Иван Андреевич

МГУ им. М. В. Ломоносова

Российская Федерация, Москва

Максимова Анастасия Сергеевна

МГУ им. М. В. Ломоносова

Российская Федерация, Москва

Аннотация

В статье рассмотрены возможности и особенности применения методов обработки «больших данных» при исследовании миграции населения. Особое внимание уделено основным проблемам получения первичной информации из цифровых источников и ограничениям их использования. Оценены возможности использования «цифровых следов» мобильных операторов, приложений маршрутизации и открытых данных социальных сетей для анализа миграционных потоков. Проанализированы примеры подобных исследований. Сделаны выводы о перспективах использования «больших данных» цифровых сервисов как нового источника информации о миграции населения.

Ключевые слова: миграция населения, большие данные, источники данных, социальные сети

Дата публикации: 31.12.2023

Источник финансирования:

Настоящее исследование проведено в рамках программы развития МГУ им. М. В. Ломоносова и поддержано проектом № 23А-Ш-05-04 междисциплинарной научно-образовательной школы МГУ «Математические методы анализа сложных систем».

Ссылка для цитирования:

Алешковский И. А. , Гребенюк А. А. , Максимова А. С. Перспективы использования «больших данных» для анализа миграционных процессов // Электронный научно-образовательный журнал «История». – 2023. – Т. 14. – Выпуск 12 (134). Часть II. URL: <https://history.jes.su/s207987840029686-3-1/>. DOI: 10.18254/S207987840029686-3

1

Введение

Современная стадия развития общества характеризуется повсеместным проникновением сети интернет и цифровых сервисов, работающих на ее основе, в жизнедеятельность людей. По официальным данным, численность интернет-пользователей в России в 2023 г. составила около 100 млн человек¹. Доля ежедневных пользователей интернета среди опрошенных россиян достигла 74 %, при этом 35 % проводили в сети более 4 часов в день, не использовали сеть интернет только 16 % опрошенных². За последние пять лет число онлайн-покупателей в России выросло почти вдвое. В настоящее время примерно 60 % городских жителей 16—55 лет не реже двух раз в год делают покупки в интернете. Почти 85 % из них покупают минимум раз в месяц, почти 30 % — минимум раз в неделю.

² Также следует отметить возросшее влияние социальных сетей и различных мессенджеров, которые меняют практики социального взаимодействия, формат средств массовой информации, формируют новую сферу хозяйственной деятельности — «экономику внимания». По данным компании “Brand Analitics”, ежемесячное количество авторов в социальной сети «ВКонтакте» в 2023 г. превысило 24 млн человек, а общее количество созданных ими сообщений приближается к отметке 400 млн³.

³ Рост аудитории онлайн-сервисов, в свою очередь, приводит к увеличению «цифровых следов», к генерации все большего количества информации и, следовательно, к росту данных, доступных для анализа. Кроме того, внедрение информационных технологий в традиционные социальные и бытовые сферы дает возможность получения данных нового формата (например, геопро пространственная информация, фото- и видеоматериалы). Подобные тенденции трансформации скорости, количества и качества создаваемой информации требуют разработки новых методологических подходов и методического инструментария ее анализа, релевантных конкретным целям исследования. Потребителями полученных эмпирических данных могут выступать различные структуры, осуществляющие деятельность как в области ведения бизнеса и коммерческой деятельности, так и в сфере государственного управления.

⁴ Генерируемые сегодня «большие данные» (“Big Data”) представляют собой объемные наборы как структурированных, полуструктурированных, так и неструктурированных данных, при этом их интеллектуальный анализ требует очень много ресурсов. Технологии обработки «больших данных» работают с данными, создаваемыми как пользователями мобильных устройств, интернет-платформ, так и цифровыми датчиками и измерителями, например, спутниковыми устройствами⁴.

⁵ Использование технологий анализа информации в формате «больших данных» позволяет добиться совершенствования арсенала исследовательских методов и расширить круг исследовательских задач. К примеру, в социологии это позволило продвинуться в решении проблемы, связанной с определением природы социального явления⁵. Также в настоящее время вышло значительное количество публикаций, посвященных использованию технологий обработки «больших данных» социальных сетей для изучения различных видов миграционного движения (вынужденной, трудовой, учебной, интеллектуальной миграции) и процессов адаптации мигрантов, о которых будет сказано ниже.

⁶

Особенности применения технологий обработки «больших данных» для анализа социальных явлений

Безусловно, использование новых источников информации имеет существенный потенциал для развития социологических исследований. При этом анализ «больших данных» имеет свои особенности, а также методические проблемы. К преимуществам технологий обработки «больших данных» относятся:

- анализ большого объема информации, позволяющий провести наблюдение, приближенное к сплошному, а также детализировать данные по различным срезам генеральной совокупности;
- оперативность генерации данных: практически мгновенное их появление при совершении тех или иных социальных актов в реальности;
- оперативность поступления данных в центры обработки и дальнейшей работы с ними;
- непрерывность поступления информации: постоянное обновление баз данных; - уменьшение временных затрат и затрат ручного труда при анализе за счет автоматизации процесса.

⁷ Перечисленные выше преимущества имеют особое значение для анализа миграционных процессов, в особенности тех, где миграционный статус индивида может неоднократно меняться в течение короткого периода (недели, месяца). В таких условиях классические инструменты учета мигрантов не позволяют оперативно и с необходимой полнотой получать первичную информацию.

⁸ Суть феномена «больших данных» была описана Д. Ленни в концепции 3V: Volume (Объем), Velocity (Скорость обновления), Variety (Разнообразие). Позднее она была доработана и расширена до 7V за счет включения Veracity (Достоверность), Variability (Изменчивость), Visualization (Визуализация), Value (Ценность)⁶. Благодаря высокопроизводительным технологиям — таким, как грид-

вычисления или аналитика в оперативной памяти — исследователи могут использовать практически любые объемы «больших данных» для анализа. Зачастую “Big Data” сначала подвергают предварительной обработке, отбирая только те, которые реально нужны для задач анализа. Все чаще «большие данные» применяют для задач в рамках расширенной аналитики, включая искусственный интеллект.

⁹ Однако работа с «большими данными» зачастую ставит ряд не только технических, но и этических проблем. В отношении использования данной информации ставятся два основных принципиальных вопроса:

¹⁰ 1. Проблема конфиденциальности используемых данных. Достаточно дискутируемой является тема правовых основ использования «больших данных» ввиду того, что она тесно касается вопроса защиты персональных данных. В случае изучения мобильности населения, вопрос о конфиденциальности персональных данных решается путем их сбора только из открытых источников, то есть анализируются исключительно те сведения, которые индивид сам о себе сообщает (в том числе невольно и неумышленно);

¹¹ 2. Проблема валидности данных, детерминируемая ограниченной сопоставимостью «виртуального» и «реального» мира. Проекция реального мира на виртуальное пространство зачастую носит фрагментарный характер. Если речь идет о социальных сетях, то наблюдается процесс конструирования цифровой идентичности, то есть создания желаемого образа путем сообщения о себе фактов, которые не являются подлинными, либо определяюще важными в данный момент для данного человека. Наиболее тщательное конструирование цифрового профиля характерно для цифровой гиперидентичности, которая подразумевает восприятие реального мира как части цифровой среды, в отличие от оптимальной цифровой идентичности и цифровой идентичности пониженного уровня, при которых глубина диффузии виртуального мира в реальный мир личности значительно ниже и может доходить до цифрового отчуждения⁷.

¹² Кроме того, наблюдается умышленное искажение качества и, как следствие, валидности информации в социальных сетях за счет ложных аккаунтов, спам-сообщений и другого. Идентификация личности в интернете может быть выражена отношением «один ко многим»: один пользователь создает несколько профилей в одной и той же социальной сети и использует их в зависимости от необходимости оперировать тем или иным информационным контентом. Другой вариант идентификации может быть выражен отношением «один к одному», когда один человек имеет только один профиль в каждой из социальных сетей. Также существует отношение «многие к одному», при котором несколько человек ведут один профиль в коммерческих или политических целях. Из-за указанных особенностей проекции реальной структуры населения в виртуальном пространстве возникают ограничения исследовательских возможностей получаемой информации⁸. В этой связи представляется необходимым говорить не о людях, а о цифровых идентичностях, виртуальных образах, информационном портрете, инфо-личности или информационных идентичностях;

¹³ 3. Существует также ряд проблем использования «больших данных» для анализа социальных явлений, касающихся применяемых методов получения, обработки и хранения информации, а также компетенций специалистов и наличия необходимых мощностей. Отметим, что особенности «больших данных» обуславливают и их институциональную смещенность. Академические круги менее ориентированы на имплементацию таких данных в анализируемые массивы информации, чем коммерческие структуры (прежде всего, ввиду существующих ограничений доступа и наличия низкого уровня доверия к таким данным).

¹⁴

«Большие данные» в исследованиях миграции населения

По мнению некоторых экспертов, «большие данные» пока не могут быть альтернативой традиционным источникам статистической информации о миграции населения⁹. Данная позиция обоснована, во-первых, разрозненностью существующих баз данных, в которых собираются цифровые следы мигрантов. Во-вторых, имеется незавершенность цифровой трансформации системы управления миграционными потоками в России, несмотря на указание этого направления в Концепции государственной миграционной политики Российской Федерации на 2019—2025 гг.¹⁰ Вместе с тем, источники информации в формате «больших данных» уже сегодня позволяют расширить возможности анализа отдельных видов миграции (таблица 1).

¹⁵ *Таблица 1. Источники информации и анализируемые виды миграции*

Источники	Способ идентификации миграционного перемещения	Вид миграции
1. Данные активности аккаунта в социальных медиа (социальные сети, блоги и так далее)	Сообщения о переезде, о работе в регионе, отличающемся от места постоянного проживания, информация о месте рождения, месте окончания учебного заведения, месте проживания	Постоянная, временная (в том числе образовательная, трудовая)
2. Данные сотовых операторов о передвижении устройств	Регистрация подключений к вышкам мобильных операторов	Маятниковая, временная
3. Данные маршрутизации мобильных устройств	Данные о запросах маршрутов	Маятниковая
4. Данные о банковских операциях	Геолокация совершения банковских операций	Маятниковая, постоянная, трудовая
5. Данные о научной публикационной активности	Аффилиация публикаций автора в научных изданиях	Постоянная миграция (утечка умов), временная (трудовая, циркуляция умов)

¹⁶ Остановимся на рассмотрении аналитических возможностей обработки «больших данных» из трех источников: операторов сотовой связи, запросов маршрутизации мобильных устройств и социальных сетей.

¹⁷ 1. Данные запросов маршрутизации мобильных устройств — данные, генерируемые самими пользователями. Их принципиальное отличие от сведений, получаемых с помощью Wi-Fi сетей и GPS-навигации, состоит в том, что именно пользователи на своем устройстве устанавливают возможность доступа к такой информации для анализа, в то время как данные о подключении к сотовым

вышкам или Wi-Fi узлам собираются вне зависимости от желания пользователей сделать их доступными.

¹⁸ С точки зрения изучения миграции населения, эти данные можно использовать для характеристики динамики подвижности населения. Чем больше обращений пользователей с запросами о построении маршрута к приложению, тем большее количество поездок или пеших прогулок они совершают, и, соответственно, имеется более высокая подвижность населения. Следует оговориться, что подобный вывод справедливо делать лишь для отдельных территориальных локализаций, устанавливаемых исследователями. Кроме того, размер населенного пункта прямо пропорционален количеству запросов маршрутизации не только из-за разницы в количестве жителей, а значит и потенциальных пользователей, но и из-за территориальной удаленности тех или иных объектов социальной инфраструктуры. Также определенное значение имеет генезисная структура населения: чем больше в населенном пункте вновь прибывших переселенцев, тем больше количество запросов о маршрутизации. Однако оно снижается по мере повышения продолжительности проживания, поскольку возрастает информированность новоселов об адресах расположения тех или иных объектов. Также очевидно, что запросы маршрутов во вновь построенный район или ко вновь построенным объектам социальной инфраструктуры будут более частыми, поэтому следует принимать во внимание наличие новых инфраструктурных объектов. С точки зрения изучения передвижения населения, справедливо анализировать динамику запросов в границах одной территориальной единицы, без сравнения населенных пунктов между собой.

¹⁹ Аналитические возможности этого источника информации позволяют выявлять маятниковую миграцию за счет сравнения различий запросов одного аккаунта в динамике, в частности, по дням недели за длительный интервал времени. Если несколько дней в неделю запросы локализованы по точке отправления в одном населенном пункте, а оставшиеся — в другом, можно предполагать, что два населенных пункта являются местом постоянного проживания и местом ведения трудовой деятельности данного индивида.

²⁰ Рассмотрим пример двух сервисов открытых данных, созданных в целях расширения информационного пространства во время борьбы с пандемией.

²¹ Сервис “Mobility Trends Reports” от компании Apple до 14 апреля 2022 г. аккумулировал анонимизированные данные о запросах маршрутизации, которые доступны пользователям в виде динамических базисных показателей ежедневного прироста по странам и отдельным регионам¹¹. Сервис от Google “Covid-19 Community Mobility Reports” агрегировал данные до 17 октября 2022 г. и охватывал 135 государств, в том числе Российскую Федерацию¹². Для России доступны данные по 16 городским агломерациям, в то время как для США — по 52. Данные в сервисах генерировались на основе сведений пользователей, открывших доступ к истории своих местоположений. Конфиденциальность соблюдается путем отсутствия информации в неагрегированном виде, что не позволяет установить местоположение, траекторию движения и количество обращений к приложению о построении маршрута отдельного пользователя. В

отчетах представлена активность обращений с запросами о построении маршрута, пунктами назначения в которых являются общественные места, в том числе в разрезе их типов: аптеки, точки розничной торговли, точки общественного питания, музеи, парки, пляжи и др. Обновление информации происходило непрерывно и занимало 2—3 дня с момента ее генерации пользователями.

²² Оба сервиса имели схожий интерфейс представления информации потребителям, где за базу расчета и базовую линию был взят уровень мобильности населения за определенный период, и динамика запросов маршрутизации строилась относительно указанной базы сравнения. Для сервиса от Google базовая линия — это уровень медианной мобильности населения в соответствующий день недели, в период 3 января — 6 февраля 2020 г. (5 недель). Для сервиса Apple базовая линия — один день, 13 января 2020 г., при этом отмечено, что во многих странах (регионах) и городах относительный объем перемещений с 13 января увеличился, что соответствует обычной регулярно наблюдаемой сезонности. С точки зрения базы сравнения, использование медианного значения за период позволяет получить более надежные данные.

²³ Следует заметить, что данные маршрутизации помогают сепарировать маятниковую миграцию и разовые поездки, что немаловажно при изучении мобильности. Вместе с тем проанализировать потоки долгосрочных переселенцев с помощью данного инструмента сложно.

²⁴ 2. Данные сотовых операторов. Данные о перемещениях можно также получить с помощью регистрации подключений пользователей к вышкам мобильных операторов. Среди недостатков таких данных следует отметить то, что регистрация мобильных устройств позволяет получить информацию о местоположении только тогда, когда человек пользуется сотовым телефоном. При этом пользователи мобильных устройств, как правило, совершают большую часть своих звонков за короткие интервалы времени, за которыми следуют длительные периоды без звонков. В течение них отсутствует какая-либо информация о местоположении пользователя.

²⁵ Ряд исследований, проведенных с использованием эмпирических данных о подключении к вышкам операторов и о звонках, позволили сделать выводы о мобильности населения. Например, исследование предиктивности мобильности населения, результаты которого получены на основе изучения траекторий движения пользователей мобильных телефонов, показывает, что потенциальная предсказуемость мобильности составляет 93 %¹³. Цель исследования состояла в том, чтобы количественно оценить взаимодействие между регулярным (и, следовательно, предсказуемым) и случайным (и, следовательно, непредсказуемым) перемещением. Выборка составила 50 тысяч человек. Траектории движения пользователей, совершавших звонки с мобильных телефонов, были описаны в виде траекторий перемещения между вышками с учетом различной продолжительности времени, которое пользователь провел в непосредственной близости от конкретной вышки. Анализ показал, что люди, как правило, проводят большую часть своего времени в нескольких местах. Также, благодаря фиксации последовательности посещаемых пользователем вышек, было установлено, что передвижение имеет определенную временную структуру, то есть удалось

установить пространственно-временную траекторию перемещения пользователей¹⁴. Одним из значимых результатов исследования стало отсутствие гендерных и возрастных различий в регулярности и предсказуемости перемещений, а также существенных изменений в регулярности передвижений в выходные дни по сравнению с буднями, из чего следует, что регулярность не обосновывается рабочим графиком, а сама по себе присуща человеческой жизнедеятельности. Наиболее существенным выводом, по мнению авторов исследования, является обнаружение отсутствия вариативности предсказуемости в различных по мобильности группах: казалось бы, что для людей, которые мало путешествуют, прогноз из-за меньшего количества предиктивных перемещений должен быть более точным, в то время как передвижения тех, кто регулярно совершает длительные поездки, должны иметь низкую предсказуемость. Однако оказалось, что это не так, и повседневная подвижность населения фактически характеризуется глубоко укоренившейся регулярностью¹⁵.

²⁶ Другое исследование моделей мобильности, проведенное на основе анализа схем путешествий 500 тысяч человек в Кот-д'Ивуаре с использованием данных мобильных операторов, демонстрирует, что теоретическая максимальная предсказуемость передвижения населения достигает 88 %. На основе реализации математического построения серии моделей Маркова для прогнозирования мест, посещаемых пользователем, было установлено, что точность прогнозирования при использовании данного метода составляет 87 % для стационарных траекторий и 95 % — для нестационарных траекторий¹⁶.

²⁷ Реализуемый в настоящее время проект “Italy (Istat) — Persons and Places: Mobility Estimates based on Mobile Phone Data” изначально был посвящен анализу маятниковой трудовой и учебной миграции. В настоящее время на его основе осуществляются различные исследования, в том числе в интересах туристической отрасли¹⁷. В задачи проекта входит сравнение данных административных источников информации и данных, полученных с помощью мобильных операторов. Предварительные результаты продемонстрировали прочную взаимную корреляцию двух величин — показателей численности маятниковых мигрантов, полученных из двух вышеупомянутых источников информации.

²⁸ 3. Данные социальных сетей. Говоря о социальных сетях, следует различать значение данного термина в концепции классической социологии и современной коммуникационной веб-среды. Термин «сетевой анализ» используется в социологической теории в значении особого аналитического подхода, который используется при изучении таких явлений, как организационное поведение и межорганизационные отношения, распространение инфекционных заболеваний, взаимная поддержка индивидов и домохозяйств, распространение информации, политические, культурные и научные связи и так далее. Сети представляют собой «особый тип связей между позициями индивидов, объектов или событий, которые отбираются в зависимости от целей построения сети»¹⁸.

²⁹ В то же время сегодня термин «социальные сети» в подавляющем большинстве случаев используют для обозначения цифровой технологии построения коммуникаций и распространения информации в виртуальном пространстве. Отличие электронных социальных сетей как источника информации

о миграции от первых двух рассмотренных источников состоит в том, что эти данные обладают гораздо более широким спектром параметров. Если сведения о мобильных подключениях и звонках с мобильных телефонов имеют набор характеристик, которые известны в полном объеме (дата, время, точка подключения, продолжительность подключения, точка отправления, точка назначения, протяженность маршрута и другие), то набор характеристик страниц социальных сетей гораздо более вариативен. Одни страницы позволяют получить информацию о полном наборе социально-демографических характеристик индивида, в том числе о семейном положении, возрасте, количестве детей, месте рождения, месте работы и так далее, другие не позволяют получить информацию ни по одному из вышеперечисленных пунктов. Но даже в случае наличия нужной информации на личной странице пользователя следует поставить вопрос о ее достоверности, поскольку, как уже было упомянуто, зачастую цифровая идентификация личности весьма далека от идентификации в реальности. Вместе с тем, социальные сети дают возможность специалистам выявлять новые социальные практики, в частности, миграционные.

³⁰ Использование социальных сетей позволяет решить проблему отсутствия переменных, характеризующих структуру миграционных потоков и контингентов. При этом эксперты отмечают, что «социальные сети могут стать источником предварительных оценок и снизить необходимость проведения статистических обследований, поскольку оценки на основе данных из социальных сетей могут быть получены с большей частотностью и с меньшими затратами»¹⁹.

³¹ Анализ информации, извлеченной из социальных сетей, может проводиться как автоматически — посредством использования лингвистических моделей, так и вручную — путем детального анализа аккаунта специалистом. Соотношение использования указанных подходов и их последовательность зависят от исследовательской задачи. Кроме традиционных социальных сетей, существуют также геолокационные социальные сети. Одной из них является “Foursquare”, эта сеть дает возможность отмечать пользователями свое местонахождение вблизи различных объектов. С технической точки зрения “Foursquare” — это технологическая платформа определения местоположения, служащая маркетинговым целям, анализирующая перемещения пользователей. С точки зрения получения информации о перемещениях, данное решение и аналогичные ему можно использовать в качестве идентификатора маятниковой миграции.

³² Появление возможности использования «больших данных» электронных социальных сетей для анализа перемещений населения дало старт целому ряду исследований различных видов миграции, проводимых как национальными научными центрами, так и международными организациями. Так, например, одним из конкретных объектов подобного анализа стала вынужденная миграция, которой в последние годы был посвящен целый ряд исследований²⁰. В 2017 г. вышел доклад “Social Media and Forced Displacement: Big Data Analytics and Machine Learning: White Paper”, в котором были использованы данные из социальных медиа для мониторинга проблем беженцев в Европе, а также ксенофобских настроений принимающих сообществ по отношению к мигрантам. В частности, исследование было сосредоточено на том, как европейцы

воспринимают беженцев в контексте серии терактов, произошедших в Европе в 2016 г. По итогам исследования был разработан стандартизированный информационный продукт, позволяющий УВКБ ООН отслеживать и анализировать соответствующие каналы социальных сетей в режиме реального времени²¹.

³³ Еще одно направление использования данных социальных сетей связано с изучением интеграции и ассимиляции мигрантов. А. Дубуа и соавторы на основании анализа «лайков» в социальных сетях разработали методику анализа ассимиляции говорящих по-арабски мигрантов, находящихся в Германии²². И. Стюард проанализировал интеграцию мигрантов из Мексики в США на основе анализа прослушиваемых в социальных сетях музыкальных композиций²³.

³⁴ Российскими исследователями Н. Ю. Замятиной и А. Д. Яшунским в 2017 г. представлены результаты реализации проекта «Виртуальное население России» — базы данных по структуре крупнейшей российской социальной сети «ВКонтакте». Предложенную авторами методику можно использовать для анализа внутрироссийских перемещений молодежи²⁴. В работе К. А. Чернышева на основе анализа цифровых данных социальной сети «ВКонтакте» осуществлены расчеты коэффициентов интенсивности миграционных связей населения Крыма²⁵.

³⁵ В настоящее время социальные сети дают нам возможность агрегировать данные о миграции из двух источников. Во-первых, широко используется информация аккаунта, которую человек самостоятельно указывает при регистрации и заполнении своей страницы в социальной сети. Как уже говорилось, у этих сведений есть ограничения для использования, связанные с верификацией данных. Вместе с тем на больших выборках мы можем оперативно отслеживать определенные тенденции существующих или зарождение новых миграционных процессов.

³⁶ Во-вторых, большое количество данных генерируется социальной активностью аккаунта в сети. Для анализа социальных процессов это наиболее интересная информация, так как позволяет получать субъективное отношение индивида посредством не классического социологического опроса, а, по сути, включенного наблюдения. Конечно, у этого метода также имеется достаточно много ограничений, которые находятся в плоскости социальной психологии и сферы контроля содержания контента, и которые социальному знанию только предстоит преодолеть.

³⁷

Заключение

По итогам проведенного исследования можно сделать следующие выводы:

1. Разнообразие источников наполнения «больших данных» обуславливает перспективность их использования для анализа различных миграционных процессов. При этом важно отметить, что качество получаемой информации будет достижимо не за счет дальнейшего улучшения определенного метода, а за счет синхронизации всех источников в целях формирования так называемого «озера данных» о миграции. Например, значительным потенциалом обладает

синхронизация данных сотовых операторов, финансовых организаций и транспортных компаний для учета иностранных граждан или внутренних мигрантов;

³⁸ 2. Данные маршрутизаторов, в первую очередь, интересны для анализа миграционных процессов и передвижения населения в крупных городских агломерациях. Уже сегодня эти данные позволяют регулировать потоки людей, снижать нагрузку на социальную инфраструктуру, бороться с пробками на дорогах и другое. Также большое значение имеет использование этого инструмента для развития туристической отрасли, поскольку позволяет выявить места притяжения мигрантов, которые не учитывались ранее;

³⁹ 3. Изучение электронных социальных сетей, в отличие от классических статистических методов, позволяет получить богатый социальный контекст миграционного процесса, в особенности, если анализ проводится на микроуровне и затрагивает малые социальные группы или конкретных индивидов. Благодаря определению коннотации сообщения, анализу «лайков» и «репостов», исследователь имеет возможность оценить, насколько мигрант доволен своим переселением, насколько успешна адаптация той или иной социальной группы, какова вероятная миграционная стратегия индивида: закрепление, миграция в третью страну или возвращение в страну (регион) выезда.

⁴⁰ Таким образом, возможность обогащения первичной информации текущего учета миграции, переписей населения данными цифровых следов может значительно улучшить качество результатов исследований. Вместе с тем, как показал проведенный нами анализ осуществленных эмпирических исследований, на текущий момент «больших данных» все еще недостаточно для всеобъемлющего изучения миграционных процессов.

Примечания:

1. Официальный сайт Правительства России. [Электронный ресурс]. URL: www.gov.ru (дата обращения: 30.11.2023).
2. ВЦИОМ Исследование: Цифровой детокс — 2023: о пользовании интернетом и отдыхе от него. Июнь 2023 г. [Электронный ресурс]. URL: www.vciom.ru (дата обращения: 30.11.2023).
3. Официальный портал компании “Brand Analytics” [Электронный ресурс]. URL: brandanalytics.com (дата обращения: 30.10.2023).
4. Big data, migration and human mobility. [Электронный ресурс]. URL: www.bigmigration.com (дата обращения: 15.11.2023).
5. Губа К. Большие данные в социологии: новые данные, новая социология? // Russian Sociological Review. 2018. Vol. 17. No. 1. P. 213—236. DOI: 10.17323/1728-192X-2018-1-213-236.
6. Wasif & Arshad, Iram & Alsamhi, Saeed. Big Data Testing Techniques: Taxonomy, Challenges and Future Trends // Computers, Materials & Continua. 2022. No. 74. DOI: 10.32604/cmc.2023.030266.
7. Шнейдер Л. Б., Сыманюк В. В. Пользователь в информационной среде: цифровая идентичность сегодня // Психологические исследования. 2017. Т. 10. № 52. DOI: 10.1017/v10n52/1406.
8. Шульц В. Л., Гребенюк А. А., Ашманов И. С. Теоретико-методологические проблемы цифровой социологии // Вестник Московского университета. Серия 18. Социология и политология. 2022. Т. 28. № 1. С. 126—144. DOI 10.24290/1029-3736-2022-28-1-126-144.
9. Чудиновских О. С. Большие данные и статистика миграции // Вопросы статистики. 2018. № 25 (2). С. 48.
10. Указ Президента Российской Федерации от 31 октября 2018 г. № 622 «О Концепции государственной

миграционной политики Российской Федерации на 2019—2025 гг.» (с изменениями и дополнениями от 12 мая 2023 г.). [Электронный ресурс]. URL: [>>>](#) (дата обращения: 30.11.2023).

11. COVID-19-Mobility Trends Reports — Apple. [Электронный ресурс]. URL: [>>>](#) (дата обращения: 30.11.2023).

12. COVID-19 Community Mobility Report. [Электронный ресурс]. URL: [>>>](#) (дата обращения: 30.11.2023).

13. Chaoming Song, Zehui Qu, Nicholas Blumm, Albert-László Barabási Limits of Predictability in Human Mobility // Science 19 Feb 2010. Vol. 327. Issue 5968. P. 1018—1021. DOI: 10.1126/science.1177170.

14. Там же. P. 1019.

15. Там же. P. 1021.

16. Lu X., Wetter E., Bharti N., Tatem A. J., Bengtsson L. Approaching the limit of predictability in human mobility // Sci Rep. 2013. Oct 11.3. Issue 2923. DOI: 10.1038/srep02923.

17. Bertocchi D. Exploring mobile network data for tourism statistics: the collaboration between Istat and Vodafone Business Italia // Rivista di Statistica Ufficiale. 2022. № 3. P. 43—76. DOI: 10.1481/ISTATRIVISTASTATISTICAUFFICIALE_3.2022.02.

18. Градосельская Г. В. Сетевые измерения в социологии: учебное пособие. М.: Феникс, 2004.

19. Мусин У. Р., Нусратуллин И. В. Применение «больших данных» в оценке миграционных процессов // Вестник университета. 2017. № 7—8. С. 192.

20. Alexander M., Polimis K., Zagheni E. The impact of Hurricane Maria on out-migration from Puerto Rico: Evidence from Facebook data // Population and Development Review. 2019. V. 45. No. 3. P. 617—630. DOI: 10.1111/padr.12289; Witteborn S. The digital gift and aspirational mobility // International Journal of Cultural Studies. 2019. V. 22. No. 6. P. 754—769. DOI: 10.1177/1367877919831020; Gualda E., Rebollo C. The refugee crisis on Twitter: A diversity of discourses at a European crossroads // Journal of Spatial and Organizational Dynamics. 2016. V. 4. No. 3. P. 199—212; Marquez N., Garimella K., Toomet O., Weber I. G., Zagheni E. Segregation and sentiment: Estimating refugee segregation and its effects using digital trace data // Guide to Mobile Data Analytics in Refugee Scenarios: The Data for Refugees Challenge Study / ed. A. Salah, A. Pentland, B. Lepri, E. Letouzé. Cham: Springer, 2019. P. 265—282. DOI: 10.1007/978-3-030-12554-7_14; Righi A. Assessing migration through social media: a review // Mathematical Population Studies. 2019. V. 26. No. 2. P. 80—91. DOI: 10.1080/08898480.2019.1565271.

21. Social Media and Forced Displacement: Big Data Analytics and Machine Learning. UN Global Pulse and UNHCR Innovation Service. White Paper. September 2017. [Электронный ресурс]. URL: [>>>](#) (дата обращения: 23.11.2023).

22. Dubois A., Zagheni E., Garimella K., Weber I. Studying migrant assimilation through Facebook interests // Social Informatics. SocInfo. Lecture Notes in Computer Science, V. 11186 / ed. S. Staab, O. Koltsova, D. Ignatov. Cham: Springer, 2018. P. 51—60. DOI: 10.1007/978-3-030-01159-8_5.

23. Stewart I., Flores R., Riffe T., Weber I., Zagheni E. Rock, Rap, or Reggaeton? Assessing Mexican immigrants' cultural assimilation using Facebook data // Proceedings of the World Wide Web Conference (WWW '19). San Francisco, CA, USA, May 13—17, 2019 / ed. Liu L., White R. N. Y.: Association for Computing Machinery, 2019. P. 3258—3264. DOI: 10.1145/3308558.3313409.

24. 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.

25. Чернышев К. А. Межрегиональные связи населения Крыма: исследование на основе цифровых и статистических данных о местах рождения мигрантов // Геополитика и экогеодинамика регионов. 2022. Т. 8. № 3. С. 253—264.

Библиография:

1. Алешковский И. А., Гребенюк А. А., Сидоров И. Н. Социальные риски и негативные последствия распространения технологий искусственного интеллекта // Электронный научно-образовательный журнал «История». 2022. Т. 13. Выпуск 4 (114). URL: <https://history.jes.su/s207987840019849-2-1/> DOI: 10.18254/S207987840019849-2

2. ВЦИОМ Исследование: Цифровой детокс — 2023: о пользовании интернетом и отдыхе от него. Июнь 2023 г. [Электронный ресурс]. URL: <https://wciom.ru/analytical-reviews/analiticheskii-obzor/cifrovoy-detoks-2023-o-polzovanii-internetom-i-otdykhe-ot-nego> (дата обращения: 30.11.2023).

3. Градосельская Г. В. Сетевые измерения в социологии: учебное пособие. М.: Феникс, 2004.
4. Губа К. Большие данные в социологии: новые данные, новая социология? // Russian Sociological Review. 2018. Vol. 17. No. 1. P. 213—236. DOI: 10.17323/1728-192X-2018-1-213-236.
5. Дмитриев А. С. Big data, 4v: volume, velocity, variety, value. // Мониторинг общественного мнения. № 2 (126). Март — апрель 2015. С. 156—159.
6. 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.
7. Мусин У. Р., Нусратуллин И. В. Применение «больших данных» в оценке миграционных процессов. // Вестник Московского университета. 2017. № 7—8.
8. Официальный портал компании “Brand Analytics” [Электронный ресурс]. URL: <https://brandanalytics.ru/statistics/author> (дата обращения: 30.11.2023).
9. Официальный сайт Правительства России [Электронный ресурс]. URL: <http://government.ru/news/49647/> (дата обращения: 30.11.2023).
10. Указ Президента Российской Федерации от 31 октября 2018 г. № 622 «О Концепции государственной миграционной политики Российской Федерации на 2019—2025 гг.» (с изменениями и дополнениями от 12 мая 2023 г.) [Электронный ресурс]. URL: <https://base.garant.ru/72092260/> (дата обращения: 30.11.2023).
11. Чернышев К. А. Межрегиональные связи населения Крыма: исследование на основе цифровых и статистических данных о местах рождения мигрантов // Геополитика и экогеодинамика регионов. 2022. Т. 8. № 3. С. 253—264.
12. Чудиновских О. С. Большие данные и статистика миграции // Вопросы статистики. 2018. № 25 (2).
13. Шнейдер Л. Б., Сыманюк В. В. Пользователь в информационной среде: цифровая идентичность сегодня // Психологические исследования. 2017. Т. 10. № 52. DOI: 2017v10n52/1406.
14. Шульц В. Л., Гребенюк А. А., Ашманов И. С. Теоретико-методологические проблемы цифровой социологии // Вестник Московского университета. Серия 18. Социология и политология. 2022. Т. 28. № 1. С. 126—144. DOI 10.24290/1029-3736-2022-28-1-126-144.
15. Alexander M., Polimis K., Zagheni E. The impact of Hurricane Maria on out-migration from Puerto Rico: Evidence from Facebook data // Population and Development Review. 2019. V. 45. No. 3. P. 617—630. DOI: 10.1111/padr.12289.

16. Bertocchi D. Exploring mobile network data for tourism statistics: the collaboration between Istat and Vodafone Business Italia // *Rivista di Statistica Ufficiale*. 2022. № 3. P. 43—76. DOI: 10.1481/ISTATRIVISTASTATISTICAUFFICIALE_3.2022.02.
17. Big data, migration and human mobility [Электронный ресурс]. URL: <https://www.migrationdataportal.org/themes/big-data-migration-and-human-mobility> (дата обращения: 15.11.2023).
18. Chaoming Song, Zehui Qu, Nicholas Blumm, Albert-László Barabási Limits of Predictability in Human Mobility. // *Science*. 19 February 2010. Vol. 327. Issue 5968. P. 1018—1021. DOI: 10.1126/science.1177170.
19. COVID-19-Mobility Trends Reports — Apple [Электронный ресурс]. URL: <https://covid19.apple.com/mobility> (дата обращения: 30.11.2023).
20. COVID-19 Community Mobility Report [Электронный ресурс]. URL: <https://www.google.com/covid19/mobility/> (дата обращения: 30.11.2023).
21. Dubois A., Zagheni E., Garimella K., Weber I. Studying migrant assimilation through Facebook interests // *Social Informatics. SocInfo. Lecture Notes in Computer Science*, V. 11186 / ed. S. Staab, O. Koltsova, D. Ignatov. Cham: Springer, 2018. P. 51—60. DOI: 10.1007/978-3-030-01159-8_5.
22. Gualda E., Rebollo C. The refugee crisis on Twitter: A diversity of discourses at a European crossroads // *Journal of Spatial and Organizational Dynamics*. 2016. V. 4. No. 3. P. 199—212.
23. Lu X., Wetter E., Bharti N., Tatem A.J., Bengtsson L. Approaching the limit of predictability in human mobility // *Sci Rep*. 2013. October 11.3. Issue 2923. DOI: 10.1038/srep02923.
24. Marquez N., Garimella K., Toomet O., Weber I. G., Zagheni E. Segregation and sentiment: Estimating refugee segregation and its effects using digital trace data // *Guide to Mobile Data Analytics in Refugee Scenarios: The Data for Refugees Challenge Study* / ed. A. Salah, A. Pentland, B. Lepri, E. Letouzé. Cham: Springer, 2019. P. 265—282. DOI: 10.1007/978-3-030-12554-7_14.
25. Righi A. Assessing migration through social media: a review. // *Mathematical Population Studies*. 2019. V. 26. No. 2. P. 80—91. DOI: 10.1080/08898480.2019.1565271.
26. Social Media and Forced Displacement: Big Data Analytics and Machine Learning. UN Global Pulse and UNHCR Innovation Service. White Paper. September 2017 [Электронный ресурс]. URL: <https://www.unhcr.org/innovation/wp-content/uploads/2017/09/FINAL-White-Paper.pdf> (дата обращения: 23.11.2023).
27. Stewart I., Flores R., Riffe T., Weber I., Zagheni E. Rock, Rap, or Reggaeton? Assessing Mexican immigrants' cultural assimilation using Facebook data // *Proceedings of the World Wide Web Conference (WWW '19)*. San Francisco, CA, USA,

May 13—17, 2019 / ed. L. Liu, R. White. N. Y.: Association for Computing Machinery. 2019. P. 3258—3264. DOI: 10.1145/3308558.3313409.

28. Wasif & Arshad, Iram & Alsamhi, Saeed. Big Data Testing Techniques: Taxonomy, Challenges and Future Trends // Computers, Materials & Continua. 2022. No. 74. DOI: 10.32604/cmc.2023.030266.

29. Witteborn S. The digital gift and aspirational mobility // International Journal of Cultural Studies. 2019. V. 22. No. 6. P. 754—769. DOI: 10.1177/1367877919831020.

Big Data for Studying Population Migration: Analytical Capabilities

Alexandr Grebenuk

Lomonosov Moscow State University

Russian Federation, Moscow

Ivan Aleshkovski

Lomonosov Moscow State University

Russian Federation, Moscow

Anastasiya Maksimova

Lomonosov Moscow State University

Russian Federation, Moscow

Abstract

The article discusses the possibilities and features of usage the “big data” processing methods for studying population migration. Particular attention is paid to the main problems of obtaining primary information from digital sources and the limitations of their use. The article assesses the possibilities of using “digital traces” of mobile operators, application routing and open data from social networks to analyze migration flows. It also provides specific examples of such research and draws conclusions about the prospects for using digital “big data” services to enrich the information of classical migration registration tools.

Keywords: population migration, big data, information sources, social networks

Publication date: 31.12.2023

Citation link:

Aleshkovski I., Grebenuk A., Maksimova A. Big Data for Studying Population Migration: Analytical Capabilities // ISTORIYA. – 2023. – V. 14. – Issue 12 (134). Part 2. URL: <https://history.jes.su/s207987840029686-3-1/>. DOI: 10.18254/S207987840029686-3