

ВОЗМОЖНОСТИ И ОГРАНИЧЕНИЯ ДАННЫХ СОТОВЫХ ОПЕРАТОРОВ В ИЗУЧЕНИИ МИГРАЦИИ НАСЕЛЕНИЯ

Юмагузин В.В., к.соц.н., научный сотрудник Института демографии НИУ ВШЭ, e-mail: vyumaguzin@hse.ru

***Аннотация.** Данные сотовых операторов позволяют оценить потоки и контингенты мигрантов, центры притяжения и спальные районы. Однако данные не предоставляют информации о гражданстве, уровне образования, семейном положении, занятости мигрантов, цели въезда/переезда, кроме того они дают информацию о передвижениях не всегда человека, а его телефона/сим-карты. В связи с этим данные должны быть использованы совместно с традиционными источниками.*

***Ключевые слова:** миграция, цифровые данные, данные сотовых операторов.*

OPPORTUNITIES AND LIMITATIONS OF MOBILE OPERATORS ' DATA IN STUDYING POPULATION MIGRATION

Yumaguzin V.V., PhD of social Sciences, Institute of demography, HSE, research associate, e-mail: vyumaguzin@hse.ru

***Abstract.** Cell phone operator data allows us to assess the flows and contingents of migrants, centers of attraction and residential areas. However, the data does not provide information about the citizenship, education level, marital status, employment of migrants, purpose of entry/relocation. In addition, they provide information about the movements of person's cellphone/SIM-card. Therefore, the data should be used together with traditional sources.*

***Keywords:** migration, digital data, cell phone data.*

С развитием информационных технологий возникают новые цифровые источники данных, которые все больше начинают вовлекаться в изучение российской миграции: ГИС-анализ [0, 2], анализ социальных сетей [3] и анализ данных сотовых операторов [4, 5]. Мировой опыт показывает, что даже информация об аффилиации авторов научных журналов также может служить источником данных о миграции [6].

Информационная революция, которая привела к переходу от аналоговых к цифровым технологиям позволила накопить огромный объем данных на индивидуальном уровне. Развитие Интернета и устройств «Интернета вещей» ускорили этот процесс, производя беспрецедентный объем данных об обществе и поведении человека, которые изначально не были предназначены для нужд социологов [7]. К новым цифровым данным относятся также данные сотовых операторов, которые могут быть использованы, в том числе в изучении миграции населения. Распространение сотовой связи среди населения во всем мире постоянно увеличивается в связи с ростом ее доступности: совершенствуются существующие технологии и появляются новые, позволяющие снижать стоимость услуг и оборудования при повышении качества предоставления [8]. В 2018 году число используемых сим-карт в России составило более 255 млн, уровень проникновения услуг сотовой связи (количество сим-карт на человека) по данным сотовых операторов составил 179%. Наибольший уровень проникновения отмечен в Москве (249%) и Санкт-Петербурге (248%). В регионах средний уровень проникновения составил в прошлом году 164% [8]. По этому показателю Россия находится в ряду наиболее развитых телекоммуникационных рынков мира. В то же время фактический уровень проникновения услуг сотовой связи в середине 2010-х годов оценивался в 75%-79% [9]. Для сравнения уровень проникновения социальных сетей в России в этот же период составил всего 33% [10]. То есть результаты анализа данных сотовых операторов скорее будут более достоверны ввиду меньшей селективности пользователей услуг.

По данным отчета «Ericsson Mobility Report, 2018» общее количество сим-карт в мире достигло 7,9 млрд, при этом количество активных абонентов составляет 5,6 млрд [11]. К 2024 г. аналитики компании Ericsson ожидают рост активных абонентов до 6,2 млрд человек. Ожидается также рост и устройств «Интернета вещей», в т.ч. с использованием сотовых технологий.

Данные сотовых операторов уже сейчас используются для оценки наличного и постоянного населения (в т.ч. планируется привлечение операторов мобильной связи к переписи населения [12]), плотности населения, выявления центров притяжения и границ спальных районов, оценки миграционных потоков и доли маятниковых мигрантов в общем числе населения региона. Дальнейшее развитие могут получить оценки среднего расстояния передвижения маятниковых мигрантов, определение режима работы и в целом формирование социально-демографического портрета мигрантов. Потенциально с помощью новых цифровых данных возможно оценить коммуникационную связность регионов, которая может отражать миграционные сети и межрегиональную деловую активность [13]. Помимо изучения миграции, данные сотовых операторов могут быть полезными при анализе здоровья, транспорта, городского планирования, энергетики, электорального поведения и качества государственной статистики [10]. Говоря о репрезентативности выборок, полученных на основе анализа новых данных, следует отметить, что сотовыми телефонами пользуются начиная со школьного возраста и заканчивая пенсионными возрастами, хотя уровень проникновения мобильной связи в старших возрастах может быть еще недостаточно высоким ввиду поколенческих различий. В целом исследователи отмечают, «что нет ни одной группы потребителей, которые не являются пользователями мобильной связи. Группы могут различаться по активности пользования, но пользователями являются практически все» [14]. Наряду с интернет-технологиями мобильная связь лучше охватывает меньшинства и подвижные группы населения – жителей удаленных регионов, недокументированных мигрантов, временных и маятниковых мигрантов и др.

Данные о геолокации собираются при любой абонентской транзакции (звонок, смс, интернет-сессия), а также в целом в течение суток для абонентов, которые ежедневно совершают звонки [15], при этом различные приложения в смартфоне также содержат функции геопозиционирования либо непосредственно указывая GPS координаты (Facebook, Twitter, Instagram, VK, LinkedIn, YouTube, Foursquare, Tumblr, Flickr, Reddit и др.), либо оставляя данные об уникальных IP адресах (почтовые сервисы), поэтому анализ социальных сетей и почтовых сервисов может еще в большей степени расширить источниковую базу миграциологии.

Доступ исследователей к персонифицированным данным сотовых операторов чаще всего осуществляется посредством коллаборации с Департаментом информационных технологий г. Москва, который закупает данные у операторов «большой тройки». Однако не исключено, что российские телекоммуникационные компании, взяв пример с Sonatel и Orange Group, самостоятельно создадут персонифицированную базу данных телефонных звонков, которая будет доступна для исследователей. Исследователи также могут сформировать уникальную базу данных, содержащую информацию о передвижении человека, направлении и длительности звонков, договорившись с участниками эксперимента об установлении специального программного обеспечения или приложения на их телефон/смартфон. По такому пути пошли исследователи из Массачусетского технологического университета, США, организовав в 2009 г. девятимесячный эксперимент по сбору данных, в котором приняли участие сто человек [16].

В отличие от традиционных источников данных о миграции, результаты которых часто публикуются в годовом исчислении и в разрезе субъектов Российской Федерации, данные сотовых операторов достаточно оперативно отражают события с большой степенью детализации. Однако, как любой источник данных о населении, он имеет свои ограничения [17]. Важно понимать, что учитывается активность не конкретных людей, а сим-карт. При этом на одного человека может приходиться более одной сим-карты: это может быть телефонный аппарат, поддерживающий 2 сим-карты, это может быть также и второй телефонный аппарат или другое мобильное устройство, используемое в повседневной жизни. Еще одно уточнение касается геолокации, которая может быть определена двумя путями. В первом случае может использоваться замер удаленности абонента от трех станций по мощности сигнала от его сотового те-

лефона. Во втором случае оператор сотовой связи получает координаты не абонента, а определенной сотовой ячейки, которую обслуживает базовая станция. Данное ограничение особенно важно учитывать при включении в исследование сельских территорий, где сеть базовых станций может быть редкой и соответственно радиус их обслуживания будет большим (до нескольких километров), что приведет к меньшей точности геолокации. В городах плотность базовых станций достаточно высокая и соответственно зона покрытий каждой из них является более детализированной, чтобы позиция абонента была определена с погрешностью в несколько десятков метров. Многие исследователи миграции опираются на данные голосового трафика мобильных операторов, однако при дальнейшем увеличении доступности интернета, коммуникация будет переходить в область IP-телефонии (голосовые сообщения и звонки, а также видеозвонки в WhatsApp и пр.), ввиду чего оценки миграции по голосовому трафику могут стать недостаточно полными. В этом случае фокус исследователей будет смещаться на геолокацию интернет-сессий, которые, однако, не позволят учесть направление и длительность коммуникации.

Данные сотовых операторов также не представляют информации о гражданстве, уровне образования, семейном положении, занятии мигрантов [17], цели въезда/переезда. Хотя исследователи работают с неперсонифицированными данными абонентов и используют их в агрегированном виде, это не всегда позволяет решить все вопросы, связанные с информированным согласием, защитой и неприкосновенностью личной жизни [10]. Обществу еще предстоит прийти к консенсусу в вопросе о пользе используемых данных в научных и гуманитарных целях и сохранением приватности.

В заключение следует отметить, что появление новых цифровых источников данных (данные сотовых операторов, социальных сетей, сайтов агрегаторов по продаже авиа- и железнодорожных билетов и др.) не отменяют важность традиционных источников данных (выборочные обследования, переписи населения, текущий учет, создание регистров), наоборот, последние могут служить «золотым стандартом» при введении новых данных в научный оборот и помочь выявить систематические смещения, чтобы впоследствии выполнить необходимые поправки [7]. В любом случае генерация новых цифровых данных о населении требует совершенствования методов и инструментов их анализа (в т.ч. визуализации и интерпретации), навыков разработки дизайна исследования, а также более глубокого понимания преимуществ и ограничений цифровых данных.

СПИСОК ЛИТЕРАТУРЫ И ИСТОЧНИКОВ

1. Махрова А.Г., Медведев А.А., Нефедова Т.Г. Садово-дачные поселки горожан в системе сельского расселения // Вестник Московского университета. Серия 5. География. 2016. №2.
2. Шитова Ю., Шитов Ю. Маятниковая трудовая миграция в Московском регионе // Демоскоп Weekly №569-570. 2013. URL: <http://www.demoscope.ru/weekly/2013/0569/tema01.php>
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. 2, 2012, с. 15 -29.
4. Юмагузин В. В. Возможности изучения маятниковой миграции по данным сотовых операторов (на примере г. Уфы) // В кн.: Тезисы VI международной социологической Грушинской конференции "Жизнь исследования после исследования: как сделать результаты понятными и полезными", 16-17 марта 2016. Материалы конференции. М.: ВЦИОМ, 2016. С. 732-737;
5. Махрова А. Г., Бабкин Р. А. Методические подходы к делимитации границ Московской агломерации на основе данных сотовых операторов // Региональные исследования. – 2019. – № 2(64). – С. 48–57.
6. Aref, S.; Zagheni, E.; West, J. The demography of the peripatetic researcher: evidence on highly mobile scholars from the Web of Science. In: Flöck, F. (Ed.): Social Informatics: 11th International Conference, SocInfo 2019, Doha, Qatar, November 18-21, 2019, Proceedings. Cham: Springer (2019). pp. 50-65.

7. Alburez-Gutierrez, D., Aref, S., Gil-Clavel, B. S., Grow, A., Negraia, D. V., Zagheneh, E. Demography in the digital era: new data sources for population research. In: Arbia, G., Peluso, S., Pini, A., Rivellini, G. (Eds.): Smart statistics for smart applications: book of short papers SIS2019, 23-30 (2019). Pearson.
8. Фомина Т.А. Анализ рынка операторов сотовой связи // Молодой ученый. – 2014. – №18. – С. 466-468.
9. Зубарев А. Е., Перевозникова М. В. Обзор рынка услуг сотовой связи России // Вестник Тихоокеанского государственного университета. 2015. - №3(36)
10. Inferring migrations, traditional methods and new approaches based on mobile phone, social media, and other big data. Feasibility study on inferring (labour) mobility and migration in the European Union from big data and social media data. Wiśniewski, Arkadiusz; Tatem, Andrew J; Abel, Guy J; Zagheneh, Emilio; Weber, Ingmar; Sorichetta, Alessandro; Hughes, Christina. 2016. 40 p.
11. "Ericsson Mobility Report, 2018" - текущее состояние и перспективы рынка мобильной связи. Технологии связи, 2019. URL: <https://itechinfo.ru/content/ericsson-mobility-report-2018>
12. Посчитают каждого. Российская газета, Союз. Беларусь-Россия - № 40(908). 30.10.2019. URL: <https://rg.ru/2019/10/30/glava-rosstata-pavel-malkov-o-rossijskoj-i-belorusskoj-perepisi-naselenia.html>
13. Blumenstock, Joshua. Using mobile phone data to measure the ties between nations. ACM International Conference Proceeding Series. 2011. 195-202.
14. Иванов М.Ю. Потребительская свобода пользователей мобильной связи в условиях глобализации // Модернизация экономики и глобализация: В 3 кн. Кн. 2 Под ред: Е. Г. Ясин Кн. 2. М.: Издательский дом ГУ-ВШЭ, 2009, с. 177-188.
15. Махрова А. Г., Кириллов П. Л., Бочкарев А. Н. Маятниковые трудовые миграции населения в Московской агломерации: опыт оценок потоков с использованием данных сотовых операторов. Региональные исследования, 2016. 3(53), 71-82
16. Bayir, Murat & Demirbas, Murat & Eagle, Nathan. Discovering SpatioTemporal mobility profiles of cellphone users. 2009 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks and Workshops, WOWMOM 2009. 2009, 1 - 9.
17. Чудиновских О. С. Большие данные и статистика миграции // Вопросы статистики. 2018. Т. 25, № 2. С. 48–56.