



PROGRAMA DE POBLACIÓN
DOCUMENTOS DE TRABAJO

**Huellas digitales y oportunidades para el estudio de
la migración en América Latina y el Caribe**

Julieta Bengochea

Documento N°8

Julio 2022

ISSN 2393-7459

1. Demografía y Big Data..... 2

2. Big Data y el estudio de la migración internacional..... 8

3. Conclusiones.....15

Referencias 17

1. Demografía y Big Data

Si bien el uso de las nuevas fuentes de información derivadas de la expansión de tecnologías digitales para el estudio de los procesos demográficos es un área de investigación consolidada en la academia del norte global, es un área de incipiente desarrollo en la academia del sur global (Alexander et al., 2020; Billari & Zagheni, 2017; Cesare et al., 2018; Hilbert, 2013; Messias et al., 2016; Spyrtatos et al., 2018, 2019; Bodgan State et al., 2013; Zagheni et al., 2014; Zagheni & Weber, 2012, 2015). Estas nuevas fuentes de información surgen como consecuencia de la expansión de Internet y del uso de redes sociales web tales como: Facebook, Instagram, Twitter, Pinterest, LinkedIn, Yahoo! Web Services, Google Trends, IP¹ de teléfonos celulares o aplicaciones de citas, por mencionar algunas, y se caracterizan por: i) generarse de modo pasivo a través de su uso; ii) constituirse en grandes bases de datos que se archivan en espacios virtuales que se pueden recolectar en tiempo real; iii) poseer estructuras heterogéneas tal como imágenes, textos o números; y iv) representar muestras aleatorias y a un bajo costo (Cesare et al., 2018; Hilbert, 2013; Hitsch et al., 2010; Zagheni & Weber, 2015).

Este tipo de dato digital es comúnmente llamado “Big Data” o “huella digital”, y su uso en los estudios demográficos propone una lógica de tipo “bottom up”, ya que no ha sido diseñados a partir de proyectos financiados con preguntas de investigación concretas y cualquier persona puede hacer uso de ellos (Billari & Zagheni, 2017; Cesare et al., 2018). Por ejemplo, plataformas como Facebook y Twitter disponen de herramientas informáticas que permiten el acceso de modo gratuito, automatizado y agregado a datos sobre perfiles demográficos y de intereses de sus usuarios con diversos niveles de granularidad geográfica, o a tweets realizados y las IP desde dónde fueron efectuados (Ritterman et al., 2009; Zagheni et al., 2014, 2017).

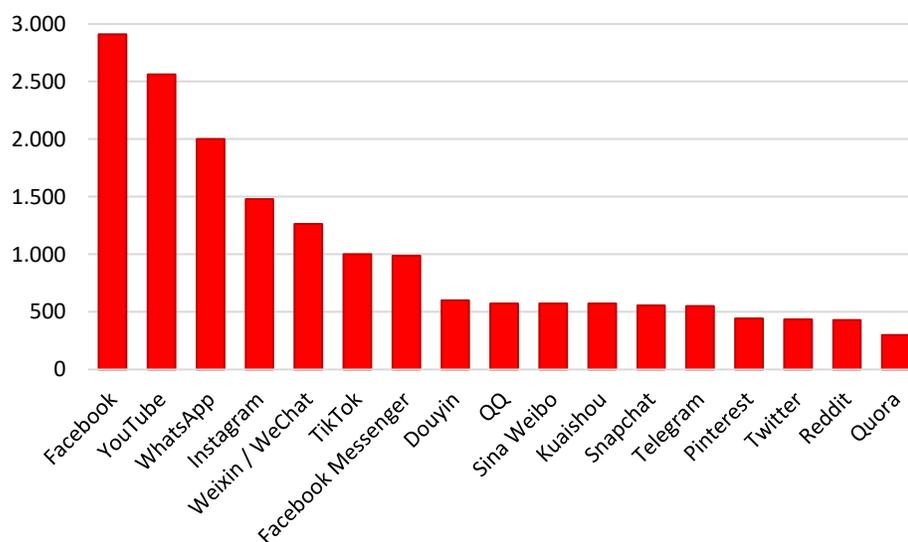
En particular, Facebook dispone de la herramienta *Facebook Ads Manager* (FAM) que contiene información sociodemográfica y de intereses de sus usuarios, desarrollada con el objetivo de segmentar y dirigir los anuncios publicitarios en esta plataforma (Alburez-Gutierrez et al., 2019; Cesare et al., 2018; Zagheni et al., 2017). Estos datos de usuarios son accesibles de modo automatizado y agregado mediante *Facebook Application Programming Interfaces* (API), herramienta también disponible en Twitter para acceder a los tweets de la red social (Twitter API). Asimismo, FAM es una herramienta útil para segmentar perfiles demográficos y realizar encuestas a poblaciones de difícil acceso mediante abordajes tradicionales de tipo cara a cara o telefónicos o mediante encuestas en línea autoadministradas (Cesare et al., 2018; Grow et al.,

¹ “Dirección de protocolo de Internet”: refiere a la dirección de coordenadas geográficas desde dónde un equipo o dispositivo se conecta a la red

2020, 2021; Perrotta et al., 2021; Pötzschke & Braun, 2017; Prieto Rosas et al., 2022; Rosenzweig et al., 2020).

En el universo de las redes sociales web, Facebook es la que cuenta con mayor número de usuarios, teniendo en promedio casi tres mil millones de usuarios mensuales en todo el mundo (Gráfico 1), lo que le ha hecho ganar el apelativo de “censo digital” (Cesare et al., 2018) y la ha convertido en la red social web de mayor uso en la investigación social en general y en el estudio demográfico de la migración en particular (Zagheni et al., 2017). Para el estudio de la migración internacional los datos de Facebook incluyen una categoría poblacional denominada Expat, que designa a personas cuyo país de residencia actual es diferente al de residencia anterior, por tanto, la definición no refiere a una persona migrante según su país de nacimiento sino según su residencia anterior. Lamentablemente, no hay documentación accesible al público en general sobre cómo es el algoritmo que Facebook utiliza para generar esta categoría, pero se estima que utiliza las variables auto reportadas sobre país de residencia, país de nacimiento y la estructura de amistades de la red social (Zagheni et al., 2017).

Gráfico 1. Redes sociales web más populares del mundo en enero de 2022, clasificadas por número de usuarios activos mensuales (en millones)



Fuente: elaboración propia con base en Statista (<https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>)

Por su parte, la plataforma *Google Trends* permite conocer el volumen de búsquedas de palabras claves realizadas en *Google Search*, según características geográficas y dimensiones temporales (Askitas & Zimmermann, 2009; Avramescu & Wiśniowski, 2021). Los datos

generados a través de esta herramienta han sido utilizados para estudiar comportamientos e intenciones reproductivas (Billari et al., 2016; Billari & Amuri, 2018; Ojala et al., 2017), analizar el impacto de políticas públicas sobre el aborto (Reis & Brownstein, 2010), monitorear brotes de influenza (Askitas & Zimmermann, 2015; Ginsberg et al., 2009; Lazer et al., 2014), proyectar niveles de desempleo (Askitas & Zimmermann, 2009; D'Amuri & Marcucci, 2017; Pavlicek & Kristoufek, 2015), predecir flujos migratorios (Avramescu & Wiśniowski, 2021), o identificar situaciones de violencia doméstica (Köksal et al., 2022). Algunos de estos resultados muestran que el volumen de búsquedas de Internet sobre aborto en Estados Unidos es mayor en aquellos estados donde las restricciones a la interrupción voluntaria del embarazo son mayores (Reis & Brownstein, 2010), que palabras como maternidad, embarazo y ovulación son útiles para estudiar intenciones reproductivas y proyectar tasas de fecundidad (Billari & Amuri, 2018) inclusive según características demográficas y socioeconómicas de los nacimientos (Ojala et al., 2017), y que las búsquedas de palabras compuestas relacionadas al empleo sirven para proyectar tasas de desempleo mensual (D'Amuri & Marcucci, 2017).

Otras fuentes de información son los datos georreferenciados provenientes de *Yahoo! Web Services*, los cuales han sido utilizados para estimar tasas de migración por sexo y edad, ilustrando cómo los datos obtenidos de correos electrónicos pueden complementar fuentes tradicionales (Zagheni & Weber, 2012). Además, información extraída de LinkedIn permite analizar tendencias de la migración internacional de personas calificadas a través de datos geolocalizados de historias académicas (Bogdan State et al., 2014), y el análisis de datos bibliométricos provenientes de plataformas como Web of Science o Scopus identificar trayectorias migratorias de migrantes calificados (Alburez-Gutierrez et al., 2019).

El uso de la información de conexiones de IP de teléfonos celulares móviles ha sido utilizado para estudios de movilidad urbana a través de la trayectoria anonimizada de teléfonos celulares (Csáji et al., 2012, 2013; González et al., 2008; Pei et al., 2014; Phithakkitnukoon et al., 2012), o para el análisis de tendencias de la migración interna (Bodgan State et al., 2013; Zagheni et al., 2014). Asimismo, Twitter dispone de información sobre la ubicación de los usuarios al momento de twittear y permite acceder a las publicaciones, mostrando que esta información es útil para monitorear enfermedades, desastres naturales, discursos de opinión pública, resultados electorales y dinámicas migratorias (Alburez-Gutierrez et al., 2019; Diaz et al., 2016; Gendronneau et al., 2019; Hawelka et al., 2014; Lamb et al., 2013; Righi, 2019; Ritterman et al., 2009; Zagheni et al., 2014).

Esta “revolución digital” (Alburez-Gutierrez et al., 2019; Billari & Zagheni, 2017) que vuelve más democrático el proceso de investigación científica, requiere sin embargo de ciertas

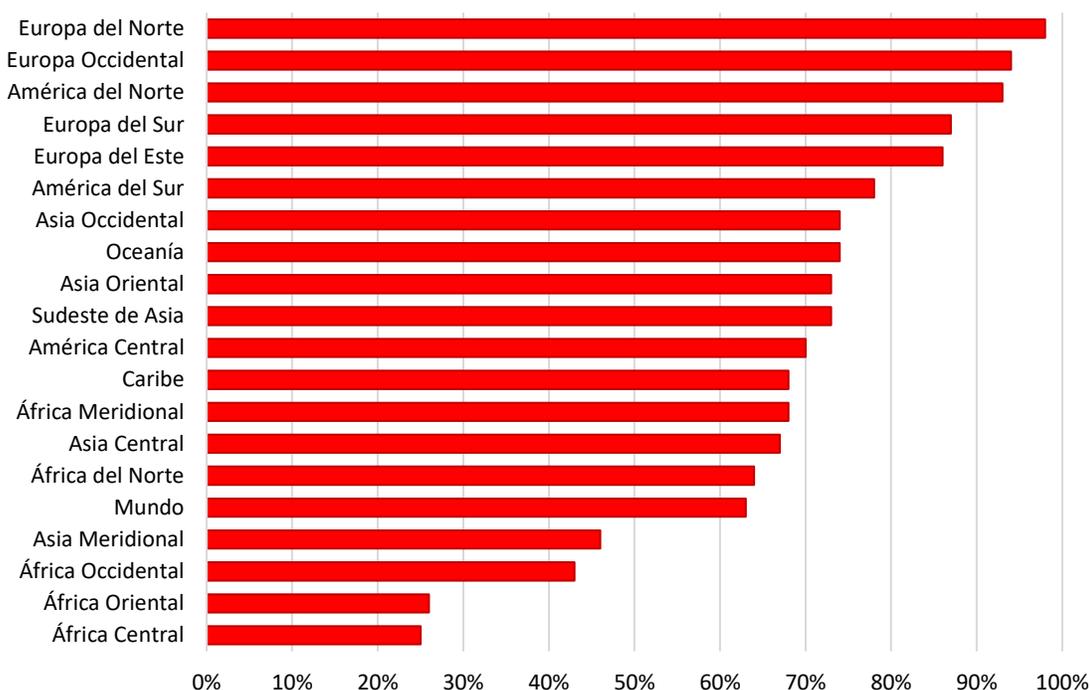
habilidades (Alburez-Gutierrez et al., 2019; Cesare et al., 2018) que pueden ser más complejas de adquirir entre las generaciones de demógrafos más antiguas (Billari & Zagheni, 2017). En este sentido, es reciente el vínculo entre las instituciones académicas, demógrafos y *computer scientists* para promover su uso (Cesare et al., 2018). Tanto así que la *International Union for the Scientific Study of Population* (IUSSP) desarrolla desde 2014 un conjunto de seminarios para la discusión y capacitación de demógrafos sobre el uso de estas nuevas fuentes de información² (Cesare et al., 2018), y el portal de información migratoria de la Organización Internacional para las Migraciones (OIM) está integrando datos derivados de estas fuentes³.

Estos nuevos datos digitales presentan una serie de desafíos del tipo éticos y metodológicos que aún están en discusión y resolución. Éticamente, los desafíos están asociados a la privacidad y grados de voluntariedad de los consentimientos de uso de las plataformas web ya que la información no es recolectada en una primera instancia con fines de investigación (Alburez-Gutierrez et al., 2019; Askitas & Zimmermann, 2015; Cesare et al., 2018; Hilbert, 2013; Letouzé, 2015; Spyrtatos et al., 2019, 2020; Zagheni & Weber, 2012, 2015). Metodológicamente, el desafío concierne a que son datos con alto sesgo y que no permiten generalizaciones ya que no todas las personas tienen acceso a Internet y redes sociales web, lo que se observa en las diferentes tasas de penetración según características demográficas y socioeconómicas (Alburez-Gutierrez et al., 2019; Cesare et al., 2018; Gil-Clavel & Zagheni, 2019; Grow et al., 2020; Kashyap et al., 2022; Zagheni & Weber, 2015). En el Gráfico 2 se observa la heterogeneidad actual en el acceso a Internet a nivel mundial. En promedio, la tasa de penetración mundial de Internet es de 63% mientras que en los países del Europa del Norte la tasa de penetración es cercana al 100% y en entre los países de África Central es de 26%.

² <https://iussp.org/en/demography-and-data-revolution>

³ <https://migrationdataportal.org/themes/big-data>

Gráfico 2. Tasa de penetración mundial de Internet en abril de 2022 por región



Fuente: elaboración propia con base en Statista (<https://www.statista.com/>)

En este sentido, el trabajo de Hargittai (2020) analiza sobre los perfiles sociodemográficos que quedan excluidos cuando se utilizan huellas digitales en la investigación social. Sus resultados muestran que la raza, la etnia y el nivel socioeconómico son variables que segmentan no solo el uso de Internet sino también el conocimiento que se tiene sobre las diferentes redes sociales, y que estas diferencias persisten en el tiempo a pesar de la expansión temporal de Internet. Por ejemplo, da cuenta de que Facebook es la red social no solo con mayor penetración de uso sino la más conocida, y que en promedio sus usuarios tienen un mayor nivel educativo en comparación con aquellos que únicamente utilizan Google, mientras que Twitter es una red social utilizada por una minoría poblacional. En este sentido, al utilizar las publicaciones de Twitter como datos secundarios para la investigación social no solo se está incurriendo en un sesgo de selección poblacional, sino que la extracción de ciertos *hashtags* sobre otros es en sí parcial.

Por otro lado, los niveles de audiencia de las redes sociales web (observado para el caso de Facebook por Rampazzo (2021), Gendronneau (2019) y Alexander (2019)) así cómo también los algoritmos que generan los datos (Kashyap et al., 2022) varían con el tiempo. Sin embargo, estos sesgos de audiencia, penetración y algoritmo se pueden suponer como sistemáticos y conocer, por ejemplo mediante la modelización de la relación entre datos web y datos tradicionales (Billari &

Zagheni, 2017; Zagheni & Weber, 2015). Asimismo, a medida que aumente la penetración de Internet los sesgos de selección irán desapareciendo ya que la población general y la usuaria de Internet serán similares (Askitas & Zimmermann, 2015).

Sin embargo, como fue mencionado en párrafos anteriores Hargittai (2020) da cuenta de que los sesgos persisten a lo largo del tiempo. Frente a estos desafíos de tipo metodológico, y bajo la hipótesis de la regularidad de los sesgos surge en la demografía una línea dedicada a combinar los datos tradicionales con datos digitales para mejorar los modelos de análisis demográficos (Alexander et al., 2020; Avramescu & Wiśniowski, 2021; Rampazzo et al., 2021; Spyrtatos et al., 2019; Wiśniowski, 2017; Zagheni et al., 2017, 2018; Zagheni & Weber, 2015). Esta línea de análisis ha sido mayormente desarrollada para la estimación de flujos migratorios mediante la utilización de modelos bayesianos que serán desarrollados con mayor profundidad en el próximo apartado.

A pesar de los límites metodológicos y los desafíos éticos, hay consenso dentro de la comunidad demográfica en cuanto a que los datos provenientes de huellas digitales representan una oportunidad para el estudio de los procesos poblacionales siempre y cuando sean complementarios con los datos tradicionales (Alburez-Gutierrez et al., 2019) y se reconozcan los sesgos inherentes (Hargittai, 2020). Esto es gracias a que estos datos: 1) se obtienen en tiempo real y permiten el estudio de los fenómenos demográficos mientras están sucediendo; 2) habilitan comparaciones entre un gran número de países a costos significativamente menores que los que requieren los abordajes tradicionales; y 3) facilitan el acceso a población ocultas (Cesare et al., 2018; Gendronneau et al., 2019; Rampazzo & Weber, 2020). En este sentido, varios autores mencionan que su uso es una verdadera oportunidad para estudiar la dinámica poblacional de los países del sur global, que se caracterizan por tener menor calidad y periodicidad de datos demográficos tradicionales (Kashyap et al., 2022; Rampazzo & Weber, 2020)

2. Big Data y el estudio de la migración internacional

El estudio de la migración internacional con base en fuentes estadísticas tradicionales (ya sea censos de población, encuestas continuas de actividad, o encuestas específicas sobre migración) presenta una serie de problemas y desafíos que no son nuevos para este campo de conocimiento. Entre los principales se destacan: i) disponibilidad y accesibilidad del dato; ii) problemas de comparabilidad del dato; y iii) fiabilidad y cobertura (Abel, 2013; Bell et al., 2002; Lomax et al., 2013; Poulain & Perin, 2009; Rees et al., 2010; Torrealba, 2018; Zlotnik, 1987).

La identificación de un migrante se realiza mediante un conjunto de preguntas y dado que generalmente estas no suelen ser similares entre fuentes ni a lo largo del tiempo, se genera un problema de comparabilidad del dato migratorio. Si bien la pregunta sobre país de nacimiento del individuo (que contrastada con el país donde se realiza el censo o encuesta permite identificar a un migrante internacional) suele estar presente, no siempre lo están las que identifican la dimensión temporal del evento migratorio.

Los censos de población son una de las principales fuentes de información para el estudio de la migración internacional dada la universalidad de la muestra. Sin embargo, el periodo intercensal sugerido por Naciones Unidas en 1976 de diez años implica un periodo de tiempo suficientemente extenso para que sucedan cambios sustantivos de la dinámica migratoria de un país (Rampazzo et al., 2021). Otra fuente de subregistro de los migrantes está dada por aquellos que ingresan a un país en condición de indocumentación, o que habiendo realizado su movimiento migratorio entre censos no dejan rastro estadístico. Estos elementos dan cuenta de cómo las fuentes tradicionales para el estudio de la migración internacional (censos, encuestas de hogares y registros de movilidad) presentan un problema de subrepresentación de los migrantes limitando el estudio de dinámicas migratorias recientes. Esto ha sido evidente en la crisis migratoria de venezolanos en el continente latinoamericano: la ausencia de datos sobre su magnitud, perfiles demográficos y direccionalidad ha sido un impedimento para la gestión humanitaria de la misma.

Considerando las limitaciones mencionadas, el campo demográfico de los estudios de migración se ha volcado a desarrollar modelos estadísticos de regresión que permitan combinar diferentes fuentes de información tradicionales para estimar flujos de migración entre países de origen y destino (Abel, 2018; Abel & Cohen, 2019; Azose & Raftery, 2019; Bijak et al., 2019; Raymer, 2017; Raymer et al., 2012, 2013; Wiśniowski, 2017). Más recientemente, y partiendo de los avances desarrollados principalmente por Raymer (2017) y Azose y Raftery (2019) el interés está puesto en explorar la aplicabilidad de utilización de datos provenientes del uso de redes sociales web para el estudio de comportamientos migratorios (Cesare et al., 2018; Hilbert, 2013;

Zagheni & Weber, 2012, 2015), principalmente mediante datos de usuarios mensuales de Facebook (Alexander et al., 2020; Palotti et al., 2020; Rampazzo et al., 2021; Spyrtatos et al., 2018, 2019, 2020; Zagheni et al., 2017, 2018).

El denominador común de estos trabajos radica en cómo controlar los sesgos inherentes de los datos provenientes de Internet y como poder hacer inferencias estadísticas a partir de estos. En este sentido, el trabajo de Zagheni y Weber (2015) se propone desarrollar un marco analítico para controlar estos sesgos que sea útil para la investigación en Ciencias Sociales. Sus resultados dan cuenta de que mediante la aplicación de ciertas técnicas estadísticas se pueden extraer regularidades de los datos no representativos de Internet que pueden ser contrastados con datos y registros poblaciones completos, disponibles en algunos países.

A continuación, se presenta una revisión de trabajos relevantes para el uso de datos web sobre migración internacional, centrada en tres inquietudes: 1) cómo utilizar datos de redes sociales web para estimar stocks de migrantes abordando los dilemas metodológicos asociados a los sesgos de sesgos de audiencia, penetración y algoritmo; 2) cómo, luego de controlar los sesgos, se pueden combinar los datos web con los datos tradicionales para lograr estimaciones en tiempo real y proyecciones a futuro de stocks y flujos de migrantes; y en la más reciente línea de investigación, 3) cómo realizar encuestas específicas a migrantes de difícil acceso mediante FAM, y utilizar métodos de post-estratificación para garantizar niveles de generalización estadística.

El trabajo seminal de Zagheni y Weber (2012) ahonda en la oportunidad de utilizar fuentes no tradicionales para complementar las estadísticas migratorias y resolver el problema de comparabilidad de los datos migratorios entre los países. Para esto, los autores estiman tasas de emigración por sexo y edad a partir de datos anonimizados provenientes del correo electrónico de Yahoo!. Puntualmente, a partir de las direcciones de IP de correo enviados entre setiembre de 2009 y junio de 2011, y la información demográfica autorreportada, identifican a migrantes cuando hay dos países desde donde se envían con mayor frecuencia los correos y el país de residencia habitual es el que concentra un mayor nivel de confianza en la frecuencia de envío. También estiman un modelo para corregir los sesgos de penetración de Internet por sexo y edad, y validan los resultados obtenidos con datos de EUROSTAT. Sus resultados muestran que hay una regularidad entre sus estimaciones y aquellas obtenidas a partir de estadísticas migratorias tradicionales, y que el método utilizado fue eficaz para los diferentes países. Igualmente, concluyen que persisten desafíos, en tanto una persona puede tener varias casillas de correo o incluso que una casilla de correo sea utilizada por varias personas, y que estas sean utilizadas de modo aleatorio por sus usuarios.

Usando *Yahoo! Web Services State*, Zagheni y Weber (2013) extraen los datos de las direcciones de IP de inicio de sesión de más de 100 millones de usuarios anónimos durante julio de 2011 y julio de 2012, lo que permitió identificar el país de conexión. A partir de esta información, identifican como persona migrante a aquel usuario de la red social que pasó más de tres meses en un país diferente de aquel de residencia, y como turista a quien pasó menos de un mes. A partir de los flujos estiman las probabilidades condicionadas de migración y de turismo a nivel global y por continentes, y a partir de un modelo lineal de efectos mixtos los factores asociados a la migración. Las conclusiones muestran que Estados Unidos es el principal país de destino, seguido por Inglaterra, Francia, India y Australia, y que estas estimaciones convergen con las de Abel (2013). Además, que los patrones de migración internacional se encuentran dominados por factores de tipo geográficos, idiomáticos y económicos.

Posteriormente, Zagheni et al. (2014) utilizan datos georreferenciados de 500.000 tuits para evaluar tendencias recientes de la migración en países de OCDE, con el objetivo de generar herramientas metodológicas que permitan monitorear en tiempo real los flujos migratorios. Los datos requirieron de un proceso previo de estructuración para su uso, principalmente para poder organizarlos según usuarios y según país o países desde dónde los usuarios realizaron sus tuits. Luego, los autores seleccionan a los usuarios móviles considerando diferentes ponderaciones según el nivel de migración del país de procedencia, y la muestra de estudio final incluye a los usuarios para los cuales se disponía de información geolocalizada desde los primeros meses de 2011. Para controlar el sesgo de selección de los datos utilizados, implementan un enfoque de Diferencias en las Diferencias (DED) entre las tasas de emigración estimadas con base en datos de Twitter y en datos de la OCDE. Los resultados obtenidos muestran convergencia en las tendencias de la migración entre los datos de Twitter y los datos provenientes de estimaciones del Pew Research Center.

En State et al. (2014), con base en datos geolocalizados de historias académicas extraídas de los perfiles de la red social LinkedIn, se analizan las tendencias mundiales de migración internacional de profesionales durante el periodo 1990 – 2012. Este trabajo muestra que, si bien Estados Unidos es el principal país receptor de personas calificadas, su capacidad de atracción descendió durante el periodo y aumentó la de Asia. Para controlar el sesgo de los datos provenientes de esta red social (ya que no es representativa de todas las personas profesionales a nivel global) comparan sus estimaciones con datos de la encuesta permanente *American Community Survey* (ACS). Sus resultados muestran que LinkedIn sobreestima la migración reciente de personas calificadas hacia Estados Unidos, pero que la tendencia es similar a la obtenida con datos de ACS. Sin embargo, los datos provenientes de esta red social para el estudio

de la migración internacional tienen limitaciones como la ausencia de información sobre la ciudadanía o país de nacimiento de las personas, así como falta de certidumbre sobre tendencias de migración circular ya que más del 90% de los resultados refieren a personas que solo realizaron un movimiento migratorio.

Por primera vez, en Zagheni, Weber y Gummadi (2017) se utilizan datos de Facebook para el estudio de la migración internacional (extraídos a través de la API de Facebook). Utilizando la categoría de Expat, desarrollan un método para estimar stocks de migrantes y eludir algunos de los problemas de trabajar con datos no representativos. Para ahondar en las características de estos sesgos, los autores estiman un modelo de regresión lineal donde la variable dependiente es el número de migrantes según país de origen, sexo y edad residiendo en Estados Unidos según la ACS, fuente que es tomada como el estándar de oro sobre el cual medir el nivel de precisión de los datos provenientes de Facebook. Sus resultados muestran que a pesar de los sesgos de selección de Facebook es posible generar estimadores demográficos de migración robustos utilizando los datos de esta red social (Zagheni et al., 2017).

A partir de este trabajo, se comienza a observar en la investigación en demografía digital una tendencia al uso de los datos de Facebook para la estimación de stocks y flujos de migrantes, intentando dar cuenta de los sesgos de selección de Facebook.

En este sentido, los trabajos liderados por Spyrtatos (Spyrtatos et al., 2018, 2019, 2020) se proponen utilizar la información de usuarios mensuales de Facebook para complementar vacíos en las estadísticas migratorias tradicionales. En general, estos trabajos se caracterizan por el abordaje metodológico sobre cómo controlar los sesgos de selección de los datos provenientes de esta red social. Por un lado, desarrollan una corrección de sesgos para los datos crudos de Facebook que considera la tasa de penetración de la red social⁴ tanto en el país de origen como en el país de destino del usuario de Facebook catalogado como Expat, y utilizan como línea de base comparativa las estimaciones poblacionales de Naciones Unidas. Por otro lado, mediante un modelo de regresión estiman la probabilidad de que una persona sea usuaria de Facebook y la correlación de estas estimaciones con estadísticas migratorias de la OCDE y ACS. En particular, Spyrtatos et al. (2019) estiman el stock de migrantes según lugar de residencia previa para 119 países de destino e incluyen países de la región latinoamericana y caribeña, validando sus resultados con datos provenientes de Naciones Unidas, ACS, OCDE y Eurostat. Los resultados muestran una correlación significativa entre los datos de Facebook y los de las fuentes

⁴ La tasa de penetración de Facebook refiere al resultante de la división de la medición total mensual (MAU) de usuarios de Facebook de un país x en año t sobre el total de población del país x en el año t de Naciones Unidas, multiplicado por 100.

tradiciones en general, y algunas diferencias en algunos países latinoamericanos de residencia previa.

Siguiendo esta línea de trabajo, Palotti (2020) utiliza los datos de Facebook como fuente alternativa para monitorear la crisis migratoria venezolana en los países de América del Sur. En este trabajo, las estimaciones obtenidas son comparadas con los datos de la Plataforma de Coordinación Interagencial para Refugiados y Migrantes de Venezuela (R4V) coordinada por OIM y ACNUR. Para corregir el sesgo de selección de Facebook, realiza el ajuste de corrección de dos pasos propuesto por Spyrtos (2018). Primero, genera un valor de ajuste de la audiencia mensual de Facebook en los países de destino, dividiendo dicho valor sobre el total de población estimada por Naciones Unidas. Segundo, divide el total de Expats de Facebook venezolanos en el país de destino sobre el valor de ajuste estimado en el paso anterior. Los resultados obtenidos muestran que los valores de R4V entran dentro del intervalo de valor comprendido entre el valor crudo y el valor ajustado de Expats de Facebook en los países de América del Sur.

En una línea similar, con el objetivo de conocer el impacto del Huracán María en 2017 en la migración de personas desde Puerto Rico, Alexander, Polimis y Zagheni (2019) utilizan datos de usuarios de Facebook recogidos de modo trimestral desde enero de 2017 hasta marzo de 2018. En particular, utilizan información del total usuarios mensuales de Facebook en Estados Unidos que anteriormente residían en Puerto Rico, según sexo, edad y Estado de residencia. Para controlar el sesgo de selección y de algoritmo de Facebook, aplican el método de DED para poder comparar con los datos de la ACS. Este método permite obtener una estimación del porcentaje de cambio en el número de migrantes de los diferentes grupos migrantes que componen la muestra e identificar en qué medida los cambios se producen en los diferentes grupos o no. Luego, estos porcentajes se aplican a los datos de la ACS para estimar poblaciones migrantes según atributos demográficos. Registraron un incremento del 17% en el número de personas migrantes de Puerto Rico en Estados Unidos durante octubre 2017 y enero 2018, lo que corresponde a un total de 185.200 personas, quienes residían principalmente en los Estados de Florida, Nueva York y Pensilvania, y se encontraban en edades laborales y eran mayormente varones. También identificaron un proceso de retorno, ya que luego del impacto del huracán (durante la ola 3 y 4 de sus extracciones) el total de Expat puertorriqueños en Estados Unidos disminuyó en un 1,8%.

Dando un paso más en el estudio de datos no representativos para el análisis de los flujos migratorios, una serie de trabajos se proponen estimar stocks y flujos migratorios mediante métodos estadísticos que combinen fuentes de datos tradicionales y no tradicionales. En esta línea, el trabajo de Rampazzo et al. (2021) se propone estimar un “valor real” del stock de migrantes en Reino Unido mediante un método que complementa los datos de la Encuesta de

Fuerza Laboral (LFS) de Reino Unido con datos de audiencia total y de Expats de Facebook. Para esto, parten del modelo desarrollado por Raymer et al. (2013) para estimar flujos de migración a partir de muestras sesgadas, y desde un enfoque bayesiano estiman un modelo en dos partes: por un lado, un modelo para los datos de LFS y por otro lado, un modelo para los datos de Facebook. Asimismo, los datos de Facebook son corregidos considerando tres puntos: primero, que alrededor del 4% de las cuentas son falsas o duplicadas; segundo, que hay que considerar las diferentes tasas de penetración de Facebook con datos poblacionales provenientes de Eurostat; y tercero, que durante el periodo de recolección de datos Facebook realizó un cambio en su algoritmo que resultó en una caída de sus usuarios (Rampazzo et al., 2021). Los resultados a los que llegan son de tipo conceptual, dado que discuten sobre la importancia de considerar sesgos en la información tanto en los datos web como en los datos tradicionales y sobre la falta de transparencia de la información y de algoritmos para la construcción de los perfiles de usuarios de Facebook.

Por su parte, Alexander, Polimis y Zgaheni (2020) estiman un modelo estadístico jerárquico bayesiano para combinar datos provenientes de Facebook con datos tradicionales de la ACS, con el objetivo de hacer estimaciones de flujos migratorios en tiempo real. Para esto, en un primer modelo de regresión relacionan el total de migrantes según la ACS con el total de migrantes de Facebook en un mismo periodo de tiempo. Luego, estiman un modelo de series de tiempo con datos de ACS para dar cuenta de las tendencias históricas y la estructura de edad de los migrantes. Así, el modelo resultante produce estimaciones y proyecciones de migrantes según estado de destino en Estados Unidos y país de origen.

Asimismo, Avramescu y Wisniowski (2021) estiman tasas de inmigración por sexo y edad de personas provenientes de Rumania en Inglaterra, a partir de: a) datos tradicionales provenientes de International Passenger Survey (IPS); y b) datos no tradicionales provenientes de *Google Trends*. Puntualmente, generan una serie de indicadores compuestos que incluyen los principales motivos de la migración entre las personas migrantes provenientes de Rumania, tal como educación, empleo y vivienda, con información proveniente del buscador de Google. Para validar estos datos, estiman modelos de serie de tiempo autorregresivos (ARIMAX) con base en los datos de IPS y las estimaciones obtenidas a través de Google. Los datos obtenidos muestran que (para el caso de estudio)- los datos provenientes de las búsquedas de Google sirven para ostrar tendencias de la migración en tiempo real.

Una línea emergente de investigación en el campo de la demografía digital se concentra en la realización de encuestas específicas a usuarios de Facebook a través de FAM, la cual fue diseñada para segmentar perfiles demográficos para los anunciantes de la red social (Cesare et

al., 2018; Grow et al., 2020, 2021; Perrotta et al., 2021; Pham et al., 2019; Pöttschke & Braun, 2017; Prieto Rosas et al., 2022; Rosenzweig et al., 2020). Algunas de sus ventajas de esta herramienta incluyen que i) permite acceder a poblaciones de difícil acceso (tal como algunos colectivos de migrantes); ii) hacer estudios de casos y encuestas comparativas entre países en un periodo de tiempo corto; y iii) a costos significativamente inferiores a los de las encuestas tradicionales cara a cara (Rosenzweig et al., 2020). Por ejemplo, el trabajo realizado por Pham et al. (2019) logró reclutar a 957 personas de origen keniano en un plazo de tres días, con un costo total de 1,130 dólares. Por su parte, Perrotta et al. (2021) recolectaron 71.612 cuestionarios entre marzo y abril de 2020 en un total de ocho países, y Grow et al. (2020) un total de 137,224 cuestionarios entre marzo y agosto de 2020, ambos con el objetivo de tener información sobre el impacto y desarrollo de la pandemia por COVID – 19. En este sentido, esta herramienta permitió obtener datos oportunos para el seguimiento de los efectos de la pandemia en la salud y calidad de vida de las poblaciones. Trabajos como los de Grow et al. (2020, 2021), Perrotta et al. (2021), y Prieto Rosas et al. (2022), son ejemplo de la utilidad de realizar encuestas a través de FAM para abordar problemas reales en el momento en que se están produciendo.

Los desafíos actuales que tiene el uso de FAM para la investigación social refieren a los ya nombrados sesgos de selección y categorización demográfica de las poblaciones por parte de Facebook. En este sentido, los trabajos analizados concentran parte de su desarrollo metodológico en evaluar la fiabilidad y precisión de las muestras de Facebook mediante el nivel de acierto que tienen las etiquetas, por ejemplo, en cuanto al sexo, edad o país de residencia anterior (Grow et al., 2020, 2021). Los resultados obtenidos sobre este punto muestran que hay una importante correlación entre los perfiles sociodemográficos de las muestras generadas mediante FAM y los perfiles sociodemográficos de encuestas tradicionales (Grow et al., 2021; Perrotta et al., 2021; Prieto Rosas et al., 2022; Rosenzweig et al., 2020).

Asimismo, para evaluar la calidad de los datos obtenidos se realizan comparaciones entre las respuestas obtenidas por FAM y encuestas tradicionales, siempre y cuando se realicen en momentos similares del tiempo y se efectúen las mismas preguntas. Esta técnica ha sido utilizada en los trabajos de Pham et al. (2019) y Perrotta et al. (2021) con resultados alentadores. Por último, es emergente la utilización de técnicas estadísticas de pos-estratificación para mejorar la representatividad de los datos obtenidos a través de muestras FAM desbalanceadas (Grow et al., 2020, 2021; Perrotta et al., 2021; Pham et al., 2019; Rosenzweig et al., 2020).

3. Conclusiones

La revisión bibliográfica ha dejado ver que los problemas de calidad, comparabilidad y periodicidad de los datos tradicionales para el estudio de las tendencias de la migración internacional ocupa un gran porcentaje de la agenda de investigación en demografía. La necesidad e importancia de disponer de datos actualizados y en tiempo real sobre el comportamiento de los flujos migratorios ha sido evidente para la gestión de las crisis de movilidad causadas en años recientes por eventos climáticos, políticos y de salud, tales como la crisis migratoria y de refugio de personas provenientes de Venezuela, el impacto del huracán María en Puerto Rico, o, la pandemia por COVID-19.

En este sentido, el orden cronológico de las publicaciones revisadas para la producción de este texto identifica que la problematización y desarrollo de soluciones a los problemas de los datos migratorios surge desde la academia del norte global desde tres líneas de investigación complementarias. Una primera línea donde el foco está puesto en la aplicación de modelos estadísticos (principalmente bayesianos) para la estimación de flujos y/o stocks de poblaciones migrantes, a partir de datos provenientes de fuentes tradicionales principalmente censos de población. Con el surgimiento del Big Data y las oportunidades que las huellas digitales abren para el análisis de la migración internacional, una segunda línea de investigación se posiciona en la agenda de investigación que se concentra en conocer los sesgos de selección, algoritmo y penetración de los datos provenientes de Internet. Esto con los objetivos de corregir estos sesgos y mejorar el nivel de generalización de los datos no representativos provenientes de huellas digitales tal como Facebook o Twitter. Una tercera línea de investigación en este campo se concentra en la combinación de datos tradicionales y datos provenientes de huellas digitales tanto para la estimación de flujos y stocks de migración en tiempo real como para la proyección a futuro de su comportamiento.

Junto a estos retos metodológicos se encuentran retos de tipo éticos, principalmente por el nivel de conocimiento y consenso que tienen los usuarios de redes sociales web sobre los datos que van dejando de modo pasivo en su uso y la posterior utilización de éstos para fines de investigación. Sin bien estos retos deben seguir problematizándose, los beneficios que brinda el uso de huellas digitales para el estudio de los comportamientos demográficos en general y de las tendencias migratorias en particular es indiscutible. Más aún, entre países que se caracterizan por disponer de estadísticas poblacionales y migratorias de baja calidad y escasa periodicidad.

Particularmente, los países de América Latina se caracterizan por tener buena información censal pero no tan buenos registros administrativos sobre ingresos y salidas de personas y/o

encuestas específicas sobre movilidad. Más allá de los desafíos metodológicos y éticos, y la heterogénea cobertura de Internet entre los países latinoamericanos y caribeños, la extracción y generación de datos provenientes de huellas digitales resultan en un bajo costo de investigación, una gran capacidad comparativa entre países y unidades administrativas menores, y en la posibilidad de obtener datos en tiempo real.

En este sentido, el uso de huellas digitales resulta una oportunidad de oro para el estudio de los procesos migratorios Sur – Sur en la región, principalmente entre periodos intercensales o ante el surgimiento de crisis migratorias. La revisión hecha ha mostrado que Facebook reúne una serie de excelentes condiciones para el análisis de la migración no solo a través de la información sobre usuarios de la plataforma sino también como plataforma para aplicar encuestas específicas y acceder a población de difícil acceso a través de la FAM. Este tipo de encuestas han sido fundamentales para poder conocer el impacto de la pandemia por COVID-19 en ciertos comportamientos demográficos sin tener que esperar el transcurso de largos periodos de tiempo que llevan las encuestas tradicionales (Prieto Rosas et al., 2022).

Sin embargo, la revisión hecha muestra una brecha en el uso de huellas digitales entre la academia europea y estadounidense y la academia latinoamericana y caribeña, que se materializa en que sólo cuatro del total de los trabajos analizados refieren al análisis de problemáticas migratorias regionales, y solo dos fueron conducidos en su totalidad por un equipo de académicos latinoamericanos. Esta problemática es planteada por Letouzé (2015), quien advierte que a medida que aumenta la atención y el uso de Big Data para la investigación y generación de políticas públicas se genera una “nueva división digital” que lejos de disminuir brechas entre países del Norte y Sur Global las vuelve más profundas.

Esta brecha académica en la utilización de huellas digitales en la investigación demográfica debe de llamar nuestra atención en la región latinoamericana y caribeña. Primero, por la oportunidad que representa en términos de investigación, y segundo, por la urgencia de desarrollar espacios de especialización que formen personas especializadas en demografía capaces de manejar los datos provenientes del Big Data.

Referencias

- Abel, G. (2013). Estimating global migration flow tables using place of birth data. *Demographic Research*, 28(March), 505–546. <https://doi.org/10.4054/DemRes.2013.28.18>
- Abel, G. (2018). Estimates of Global Bilateral Migration Flows by Gender between 1960 and 2015. *International Migration Review*, 52(3), 809–852. <https://doi.org/10.1111/imre.12327>
- Abel, G., & Cohen, J. E. (2019). Bilateral international migration flow estimates for 200 countries. *Scientific Data*. <https://doi.org/10.1038/s41597-019-0089-3>
- Alburez-Gutierrez, D., Zagheni, E., Aref, S., Gil-Clavel, S., Grow, A., & Negraia, D. V. (2019). Demography in the Digital Era: New Data Sources for Population Research. In G. Arbia, S. Peluso, A. Pini, & G. Rivellini (Eds.), *Book of Short Papers SIS2019* (pp. 1–8). Pearson. <https://doi.org/10.31235/osf.io/24jp7>
- Alexander, M., Polimis, K., & Zagheni, E. (2019). The Impact of Hurricane Maria on Out-migration from Puerto Rico: Evidence from Facebook Data. *Population and Development Review*, 45(3), 617–630. <https://doi.org/10.1111/padr.12289>
- Alexander, M., Polimis, K., & Zagheni, E. (2020). Combining Social Media and Survey Data to Nowcast Migrant Stocks in the United States. *Population Research and Policy Review*, 0123456789. <https://doi.org/10.1007/s11113-020-09599-3>
- Askatas, N., & Zimmermann, K. F. (2009). Google Econometrics and Unemployment Forecasting. *Applied Economics Quarterly*, 55(2), 107–120. <https://doi.org/10.3790/aeq.55.2.107>
- Askatas, N., & Zimmermann, K. F. (2015). The internet as a data source for advancement in social sciences. In *Discussion Paper Series* (No. 8899; Vol. 36, Issue 1). <https://doi.org/10.1108/IJM-02-2015-0029>
- Avramescu, A., & Wiśniowski, A. (2021). Now-casting Romanian migration into the United Kingdom by using Google Search engine data. *Demographic Research*, 45, 1219–1254. <https://doi.org/10.4054/DEMRES.2021.45.40>
- Azose, J., & Raftery, A. (2019). Estimation of emigration, return migration, and transit migration between all pairs of countries. *Proceedings of the National Academy of Sciences of the United States of America*, 116(1). <https://doi.org/10.1073/pnas.1722334116>

- Bell, M., Blake, M., Boyle, P., Duke-Williams, O., Rees, P., Stillwell, J., & Hugo, G. (2002). Cross-national comparison of internal migration: issues and measures. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 165(3), 435–464. <https://doi.org/10.1111/1467-985X.T01-1-00247>
- Bijak, J., Disney, G., Findlay, A. M., Forster, J. J., Smith, P. W. F., & Wiśniowski, A. (2019). Assessing time series models for forecasting international migration: Lessons from the United Kingdom. *Journal of Forecasting*, 38(5), 470–487. <https://doi.org/10.1002/for.2576>
- Billari, F., & Amuri, F. D. (2018). *Forecasting US Birth Rates with Google Trends Data and Determinants for US fertility* (Issue March).
- Billari, F., D'Amuri, F., & Marcucci, J. (2016). Forecasting Births Using Google. *First International Conference on Advanced Research Methods and Analytics*. <https://doi.org/10.4995/carma2016.2016.4301>
- Billari, F., & Zagheni, E. (2017). Big Data and Population Processes: A Revolution? In A. Petrucci & R. Verde (Eds.), *Conference of the Italian Statistical Society* (Issue Statistics and Data Science: new challenges, new generations, pp. 167–178). Firenze University Press. <https://doi.org/10.31235/osf.io/f9vzp>
- Cesare, N., Lee, H., Spiro, E., McCormick, T., & Zagheni, E. (2018). Promises and Pitfalls of Using Digital Traces for Demographic Research. *Demography*, 55(5). <https://doi.org/10.1007/s13524-018-0715-2>
- Csáji, B. C., Browet, A., Traag, V. A., Delvenne, J. C., Huens, E., Van Dooren, P., Smoreda, Z., & Blondel, V. D. (2013). Exploring the mobility of mobile phone users. *Physica A: Statistical Mechanics and Its Applications*, 392(6), 1459–1473. <https://doi.org/10.1016/j.physa.2012.11.040>
- Csáji, B. C., Browet, A., Traag, V., Delvenne, J.-C., Huens, E., Van Dooren, P., Smoreda, Z., & Blondel, V. D. (2012). Exploring the Mobility of Mobile Phone Users. *Physica A*.
- D'Amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, 33(4), 801–816. <https://doi.org/10.1016/j.ijforecast.2017.03.004>
- Diaz, F., Gamon, M., Hofman, J. M., Kiciman, E., & Rothschild, D. (2016). Online and social media data as an imperfect continuous panel survey. *PLoS ONE*, 11(1), 1–22. <https://doi.org/10.1371/journal.pone.0145406>

- Gendronneau, C., Wiśniowski, A., Yildiz, D., Zagheni, E., Fiorio, L., Hsiao, Y., Stepanek, M., Weber, I., Abel, G., & Hoorens, S. (2019). Measuring Labour Mobility and Migration Using Big Data: Exploring the potential of social-media data for measuring EU mobility flows and stocks of EU movers. In *Ep68038*. <https://www.rand.org/randeurope/research/projects/measuring-migration-using-big-data.html>https://www.rand.org/pubs/external_publications/EP68038.html<https://ec.europa.eu/social/main.jsp?catId=738&langId=en&pubId=8264>
- Gil-Clavel, S., & Zagheni, E. (2019). View of Demographic Differentials in Facebook Usage around the World. *Proceedings of the Thirteenth International AAAI Conference on Web and Social Media*. <https://ojs.aaai.org/index.php/ICWSM/article/view/3263/3131>
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012–1014. <https://doi.org/10.1038/nature07634>
- González, M. C., Hidalgo, C. A., & Barabási, A. L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782. <https://doi.org/10.1038/nature06958>
- Grow, A., Perrota, D., Fava, E. Del, Cimentada, J., Gil-Clavel, S., Zagheni, E., Flores, R. D., Ventura, L., & Weber, I. (2021). *How Reliable is Facebook's Advertising Data for Use in Social Science Research? Insights from a Cross-National Online Survey* (No. 006; MPIDR Working Paper). <https://doi.org/10.4054/MPIDR-WP-2021-006>
- Grow, A., Perrotta, D., Del Fava, E., Cimentada, J., Rampazzo, F., Gil-Clavel, S., & Zagheni, E. (2020). Addressing Public Health Emergencies via Facebook Surveys: Advantages, Challenges, and Practical Considerations. *Journal of Medical Internet Research*, *22*(12), e20653. <https://doi.org/10.2196/20653>
- Hargittai, E. (2020). Potential Biases in Big Data: Omitted Voices on Social Media. *Social Science Computer Review*, *38*(1), 10–24. <https://doi.org/10.1177/0894439318788322>
- Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2014). Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, *41*. <https://doi.org/10.1080/15230406.2014.890072>
- Hilbert, M. (2013). Big Data for Development: From Information - to Knowledge Societies. *SSRN Electronic Journal*, 1–39. <https://doi.org/10.2139/ssrn.2205145>
- Hitsch, G. J., Hortacsu, A., & Ariely, D. (2010). What makes you click?-mate preferences in online dating. *Quantitative Marketing and Economics*, *8*(4), 393–427. <https://doi.org/10.1007/s11129-010-9088-6>

- Kashyap, R., Rinderknecht, R. G., Akbaritabar, A., Alburez-gutierrez, D., Grow, A., Kim, J., Leasure, D. R., Lohmann, S., Negraia, D. V., Perrotta, D., Rampazzo, F., Tsai, C., Verhagen, M. D., & Zagheni, E. (2022). Digital and Computational Demography. *Research Handbook of Digital Sociology*, 1–50.
- Köksal, S., Pesando, L. M., Rotondi, V., & Şanlıtürk, E. (2022). Harnessing the Potential of Google Searches for Understanding Dynamics of Intimate Partner Violence Before and After the COVID-19 Outbreak. *European Journal of Population*. <https://doi.org/10.1007/S10680-022-09619-2>
- Lamb, A., Paul, M. J., & Dredze, M. (2013). Separating fact from fear: Tracking flu infections on twitter. *NAACL HLT 2013 - 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Main Conference, June*, 789–795.
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of google flu: Traps in big data analysis. *Science*, 343(6176), 1203–1205. <https://doi.org/10.1126/science.1248506>
- Letouzé, E. (2015). *Big Data and Development: General Overview Primer* (Data-Pop Alliance White Paper Series, Issue March).
- Lomax, N., Norman, P., Rees, P., & Stillwell, J. (2013). *Subnational migration in the United Kingdom: producing a consistent time series using a combination of available data and estimates*. <https://doi.org/10.1007/s12546-013-9115-z>
- Messias, J., Benevenuto, F., Weber, I., & Zagheni, E. (2016). *From migration corridors to clusters: The value of Google+ data for migration studies*. <https://doi.org/10.1109/ASONAM.2016.7752269>
- Ojala, J., Zagheni, E., Billari, F. C., & Weber, I. (2017). Fertility and its meaning: Evidence from search behavior. *Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017, Icwsm*, 640–643.
- Palotti, J., Adler, N., Morales-Guzman, A., Villaveces, J., Sekara, V., Herranz, M. G., Al-Asad, M., & Weber, I. (2020). Monitoring of the Venezuelan exodus through Facebook's advertising platform. *PLoS ONE*, 15(2). <https://doi.org/10.1371/JOURNAL.PONE.0229175>
- Pavlicek, J., & Kristoufek, L. (2015). Nowcasting unemployment rates with google searches: Evidence from the Visegrad Group countries. *PLoS ONE*, 10(5), 1–12. <https://doi.org/10.1371/journal.pone.0127084>
- Pei, T., Sobolevsky, S., Ratti, C., Shaw, S. L., Li, T., & Zhou, C. (2014). A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science*, 28(9), 1988–2007. <https://doi.org/10.1080/13658816.2014.913794>

- Perrotta, D., Grow, A., Rampazzo, F., Cimentada, J., Del Fava, E., Gil-Clavel, S., & Zagheni, E. (2021). *Behaviours and attitudes in response to the COVID-19 pandemic: insights from a cross-national Facebook survey*. <https://doi.org/10.1140/epjds/s13688-021-00270-1>
- Pham, K. H., Rampazzo, F., & Rosenzweig, L. R. (2019). Online Surveys and Digital Demography in the Developing World: Facebook Users in Kenya. *MIT Conference on Digital Experimentation*.
- Phithakkitnukoon, S., Smoreda, Z., & Olivier, P. (2012). Socio-Geography of Human Mobility: A Study Using Longitudinal Mobile Phone Data. *PLoS ONE*, 7(6), e39253. <https://doi.org/10.1371/journal.pone.0039253>
- PöTzschke, S., & Braun, M. (2017). Migrant Sampling Using Facebook Advertisements: A Case Study of Polish Migrants in Four European Countries. *Social Science Computer Review*, 35(5), 633–653. <https://doi.org/10.1177/0894439316666262>
- Poulain, M., & Perin, N. (2009). Measuring International Migration: A challenge for demographers. In J. Surkyn, P. Deboerck, & J. Van Bavel (Eds.), *Demographic challenges for the 21st century : a state of the art in demography*. Brussels University Press. <https://books.google.com.uy/books?hl=en&lr=&id=w2L3D4KWgIC&oi=fnd&pg=PA143&dq=Measuring+International+Migration:+A+challenge+for+demographers&ots=12jCDJ1CUD&sig=UBNMukB7sn7hoHLZcQ6FGOURLiI#v=onepage&q=Measuring+International+Migration%3A+A+challenge+for>
- Prieto Rosas, V., Montiel, C., Barasch, G., & Bengochea, J. (2022). Validating Facebook Tagging to Examine the Impacts of COVID-19 pandemia on a South - South migration Setting. *Conference of the Population Association of America*.
- Rampazzo, F., Bijak, J., Vitali, A., Weber, I., & Zagheni, E. (2021). A Framework for Estimating Migrant Stocks Using Digital Traces and Survey Data: An Application in the United Kingdom. *Demography*, 58(6), 2193–2218. <https://doi.org/10.1215/00703370-9578562>
- Rampazzo, F., & Weber, I. (2020). Facebook advertising data in Africa. In P. Fargues, M. Rango, E. Borgnäs, & I. Schöfberger (Eds.), *Migration in West and North Africa and across the Mediterranean. Trends, risks, development and governance* (pp. 32–37). International Organization for Migration.
- Raymer, J. (2017). Measuring flows of international migration. *IZA World of Labor*, April, 1–10. <https://doi.org/10.15185/izawol.354>
- Raymer, J., Forster, J. J., Smith, P. W. F., Bijak, J., & Wisniowski, A. (2012). Integrated Modelling of European Migration: Background, specification and results. *Norface Migration Discussion Paper*, 4, 17.

- Raymer, J., Wiśniowski, A., Forster, J. J., Smith, P. W. F., & Bijak, J. (2013). Integrated modeling of European migration. *Journal of the American Statistical Association*, 108(503), 801–819. <https://doi.org/10.1080/01621459.2013.789435>
- Rees, P., Bell, M., Duke-Williams, O., & Blake, M. (2010). Problems and solutions in the measurement of migration intensities: Australia and Britain compared. *Http://Dx.Doi.Org/10.1080/713779082*, 54(2), 207–222. <https://doi.org/10.1080/713779082>
- Reis, B. Y., & Brownstein, J. S. (2010). Measuring the impact of health policies using Internet search patterns : the case of abortion. *BMC Public Health*.
- Righi, A. (2019). Assessing migration through social media: a review. *Mathematical Population Studies*, 26(2), 80–91. <https://doi.org/10.1080/08898480.2019.1565271>
- Ritterman, J., Osborne, M., & Klein, E. (2009). *Using Prediction Markets and Twitter to Predict a Swine Flu Pandemic. 2004*.
- Rosenzweig, L., Bergquist, P., Pham, K. H., Rampazzo, F., & Mildenberger, M. (2020). *Survey sampling in the Global South using Facebook advertisements*. SocArXiv. <https://doi.org/10.31235/OSF.IO/DKA8F>
- Spyratos, S., Vespe, M., Natale, F., Iacus, S. M., & Santamaria, C. (2020). Explaining the travelling behaviour of migrants using facebook audience estimates. *PLoS ONE*, 15(9 September), 1–17. <https://doi.org/10.1371/journal.pone.0238947>
- Spyratos, S., Vespe, M., Natale, F., Weber, I., Zagheni, E., & Rango, M. (2018). *Migration Data using Social Media: a European Perspective*. European Union. <https://doi.org/10.2760/964282>
- Spyratos, S., Vespe, M., Natale, F., Weber, I., Zagheni, E., & Rango, M. (2019). Quantifying international human mobility patterns using Facebook Network data. *PLOS ONE*, 14(10), 22. <https://doi.org/10.1371/journal.pone.0224134>
- State, Bodgan, Weber, I., & Zagheni, E. (2013). *Studying Inter-National Mobility through IP Geolocation*. 265–274.
- State, Bogdan, Rodriguez, M., Helbing, D., & Zagheni, E. (2014). Migration of professionals to the U.S. Evidence from linkedin data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8851, 531–543. https://doi.org/10.1007/978-3-319-13734-6_37

- Torrealba, R. (2018). International Migration Data: Their Problems and Usefulness in Venezuela: *Https://Doi.Org/10.1177/019791838702100418*, 21(4), 1270–1276. <https://doi.org/10.1177/019791838702100418>
- Wiśniowski, A. (2017). Combining Labour Force Survey data to estimate migration flows : the case of migration from Poland to the UK. *Royal Statistical Society*, 180(1), 185–202. <https://doi.org/10.1111/rssa.12189>
- Zagheni, E., Garimella, V. R. K. R. K., Weber, I., & State, B. (2014). Inferring international and internal migration patterns from twitter data. *WWW 2014 Companion - Proceedings of the 23rd International Conference on World Wide Web*, 439–444. <https://doi.org/10.1145/2567948.2576930>
- Zagheni, E., Polimis, K., Alexander, M., Weber, I., & Billari, F. C. (2018). Combining Social Media Data and Traditional Surveys to Nowcast Migration Stocks. *PAA*, 1–17.
- Zagheni, E., & Weber, I. (2012). You are where you E-mail: Using E-mail data to estimate international migration rates. *Proceedings of the 4th Annual ACM Web Science Conference, WebSci'12, volume*, 348–357. <https://doi.org/10.1145/2380718.2380764>
- Zagheni, E., & Weber, I. (2015). Demographic research with non-representative internet data. *International Journal of Manpower*, 36(1), 13–25. <https://doi.org/10.1108/IJM-12-2014-0261>
- Zagheni, E., Weber, I., & Gummadi, K. (2017). Leveraging Facebook's Advertising Platform to Monitor Stocks of Migrants. *Population and Development Review*, 43(4), 721–734. <https://doi.org/10.1111/padr.12102>
- Zlotnik, H. (1987). Measuring international migration: theory and practice. *The International Migration Review*, 21(4). <https://doi.org/10.1177/019791838702100401>