

CÓMO APLICAR BIG DATA EN LA PLANIFICACIÓN DEL TRANSPORTE URBANO

EL USO DE DATOS DE TELEFONÍA MÓVIL EN EL ANÁLISIS DE LA MOVILIDAD

División de Transporte

NOTA TÉCNICA Nº
IDB-TN-1773

Autores:

Javier Gutiérrez Puebla
Carolina Benitez
Juan Manuel Leño
Juan Carlos García Palomares
Ana Condeço Melhorado
Carlos Mojica
Lynn Scholl
Veronica Adler
Felipe Vera
Borja Moya Gómez
Gustavo Romanillos Arroyo

Editora:

Carolina Benitez

CÓMO APLICAR BIG DATA EN LA PLANIFICACIÓN DEL TRANSPORTE URBANO

EL USO DE DATOS DE TELEFONÍA MÓVIL EN EL ANÁLISIS DE LA MOVILIDAD

Autores:

Javier Gutiérrez Puebla
Carolina Benitez
Juan Manuel Leño
Juan Carlos García Palomares
Ana Condeço Melhorado
Carlos Mojica
Lynn Scholl
Veronica Adler
Felipe Vera
Borja Moya Gómez
Gustavo Romanillos Arroyo

Editora:

Carolina Benitez

**Catalogación en la fuente proporcionada por la
Biblioteca Felipe Herrera del
Banco Interamericano de Desarrollo**

Cómo aplicar big data en la planificación del transporte urbano: el uso de datos de telefonía móvil en el análisis de la movilidad / Javier Gutiérrez Puebla, Carolina Benítez, Juan Manuel Leño, Juan Carlos García Palomares, Ana Condeço Melhorado, Carlos Mojica, Lynn Scholl, Veronica Adler, Felipe Vera, Borja Moya Gómez, Gustavo Romanillos Arroyo; editora, Carolina Benítez.

p. cm. — [Nota técnica del BID ; 1773]

Incluye referencias bibliográficas.

1. Urban transportation-Data processing. 2. Big data. 3. Urban transportation-Planning. 4. Cell phone systems. I. Gutiérrez Puebla, Javier. II. Benítez, Carolina. III. Leño, Juan Manuel. IV. García Palomares, Juan Carlos. V. Condeço-Melhorado, Ana. VI. Mojica, Carlos. VII. Scholl, Lynn. VIII. Adler, Veronica. IX. Vera, Felipe. X. Moya Gómez, Borja. XI. Romanillos Arroyo, Gustavo. XII. Banco Interamericano de Desarrollo. División de Transporte. XIII. Serie.

IDB-TN-1773

Diseño y Diagramación: **FILEZ**

Edición de estilo: **Ximena Abeledo**

Agradecimientos: **Mariano Greco y Martin Olmos, Dirección Implementación y Seguimiento SUBE, Ministerio de Transporte de la Nación Argentina.**

Contacto: **Carolina Benítez (cbenitez@iadb.org)**

<http://www.iadb.org>

Copyright © 2019 Banco Interamericano de Desarrollo. Esta obra se encuentra sujeta a una licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObrasDerivadas (CC-IGO 3.0 BY-NC-ND) y puede ser reproducida para cualquier uso no-comercial otorgando el reconocimiento respectivo al BID. No se permiten obras derivadas.

<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>

Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la CNUDMI (UNCITRAL). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID, no están autorizados por esta licencia CC-IGO y requieren de un acuerdo de licencia adicional.

Note que el enlace URL incluye términos y condiciones adicionales de esta licencia.

Las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.



Resumen

La presente nota técnica explica de qué manera el análisis de grandes volúmenes de datos de la telefonía móvil puede aplicarse a la planificación del transporte y la infraestructura urbana. La actividad de los usuarios permite conocer su huella digital y, por lo tanto, entender sus trayectorias espacio temporales de manera desagregada y extrapolada, estableciendo matrices de origen destino. La producción de grandes volúmenes de datos masivos, Big Data, abre interesantes posibilidades para entender los flujos de movilidad de nuestras ciudades de la región y su complementariedad con los métodos tradicionales de recolección de datos, como los son las encuestas de movilidad domiciliarias de origen destino de viajes, permitiendo disponer de información siempre actualizada y en menor tiempo. La nota técnica agrupa y sistematiza los conocimientos generados por especialistas e investigadores de distintos países del mundo; y analiza distintos casos de éxito realizados de implementación de dicha metodología en distintas ciudades latinoamericanas y en otros ámbitos internacionales, donde se indica cómo se han obtenido esas matrices de viajes y cuáles han sido sus aplicaciones.

Clasificaciones JEL: O2, O3, O18, O32, O35, Z18

Palabras clave : Transporte urbano, procesamiento de datos, big data, planeamiento del transporte urbano, sistemas de telefonía celular.



Cómo aplicar Big Data en la Planificación del Transporte Urbano

**El uso de datos de
telefonía móvil en el
análisis de la movilidad**

Índice

Resumen ejecutivo

1. LA MOVILIDAD URBANA Y LOS MÉTODOS TRADICIONALES DE RECOLECCIÓN DE DATOS DE MOVILIDAD

1.1 La movilidad y la planificación del transporte urbano

1.2 Fuentes de datos tradicionales: las encuestas de movilidad

1.3 Limitaciones de las encuestas de movilidad

2. DE LAS ENCUESTAS DE MOVILIDAD A LAS NUEVAS FUENTES DE DATOS: EL BIG DATA

2.1 La revolución de los datos: el Big Data

2.2 El análisis de la movilidad de la población a través del Big Data

3. ESTRUCTURA DE LAS REDES DE TELEFONÍA MÓVIL

3.1 Estructura de las redes

3.2 Características de las áreas de cobertura

3.3 Geolocalización de la actividad de los teléfonos móviles

4. DATOS DE TELEFONÍA MÓVIL

4.1 Tipos de información que aportan

4.2 Eventos activos: CDRs (Call Detail Records)

4.3 Eventos pasivos (Network-Driven Cell Phone Data)

4.4 Características de los datos de telefonía móvil

5. CÁLCULO DE MATRICES DE VIAJES ORIGEN-DESTINO (OD) A PARTIR DE DATOS DE TELEFONÍA MÓVIL

5.1 Limpieza y preparación de los datos

5.2 Selección de usuarios válidos

5.3 Identificación de estancias y secuencia de viajes

5.4 Localizaciones frecuentes y/o actividades: motivos de los desplazamientos

5.5 Enriquecimiento de datos

5.6 Cambios en la agregación espacial de los datos: de polígonos Voronoi a zonas de transporte datos

5.7 Expansión de la muestra

5.8 Obtención y visualización de las estadísticas de movilidad

5.9 Validación de los resultados

6. PRIVACIDAD: PASOS FUNDAMENTALES PARA ASEGURAR EL ANONIMATO DE LOS USUARIOS

7. EXPERIENCIAS INTERNACIONALES DE ESTUDIOS DE MOVILIDAD A PARTIR DE DATOS DE TELEFONÍA MÓVIL EN AMÉRICA LATINA Y EUROPA

7.1 Estudio sobre la movilidad en el área metropolitana de Neuquén

7.2 Pautas de la movilidad en el estudio de la segregación económica de ciudades de Suecia

7.3 Datos geolocalizados provenientes de telefonía móvil para la elaboración de estadísticas de flujos turísticos

7.4 Análisis dinámico de la accesibilidad a partir de datos de telefonía móvil

7.5 Estimación de la demanda a partir de datos móviles y su aplicación en modelos de transporte. El caso de la Región de Rotterdam

7.6 Propuesta de reestructuración de la red de líneas de autobuses urbanos basada en el cruce de datos de telefonía móvil y otras fuentes de datos

8. ANÁLISIS DE LA MOVILIDAD EN BARRIOS INFORMALES A PARTIR DE DATOS DE TELEFONÍA MÓVIL

8.1 Accesibilidad, movilidad y exclusión social en barrios desfavorecidos

8.2 Movilidad y exclusión social en el barrio informal que rodea a Costa Esperanza (Buenos Aires)

9. VENTAJAS Y LIMITACIONES. COMPARACIÓN CON LAS ENCUESTAS DE MOVILIDAD

10. CONSIDERACIONES FINALES

GLOSARIO DE ABREVIATURAS UTILIZADAS

BIBLIOGRAFÍA

ANEXO I: OTRAS EXPERIENCIAS INTERNACIONALES DE ESTUDIOS DE MOVILIDAD A PARTIR DE DATOS DE TELEFONÍA MÓVIL EN AMÉRICA LATINA Y EUROPA

Estudio de movilidad de los usuarios del Ferrocarril Línea Roca (Ramal Constitución – La Plata) & Línea C de Subterráneo. Una mirada de Género

Exposición de la población a gases contaminantes (NO₂) a partir de datos de telefonía móvil

ANEXO II: RECOMENDACIONES PARA LA CONTRATACIÓN DE ESTUDIOS DE MOVILIDAD BASADOS EN DATOS DE TELEFONÍA MÓVIL

Resumen Ejecutivo

La movilidad es uno de los grandes desafíos que tienen que enfrentar las ciudades de todo el mundo, particularmente las grandes metrópolis, muchas de ellas caracterizadas por unas tasas muy elevadas de crecimiento demográfico. Este desafío afecta particularmente a la región de América Latina y el Caribe, con una altísima tasa de población urbana y ciudades en continua expansión.

La producción de datos masivos, el denominado Big Data, abre nuevas oportunidades en el campo de la planificación del transporte en las ciudades de la región. Se trata de grandes volúmenes de datos producidos de forma continua por multitud de sensores y dispositivos, como los localizadores GPS de los smartphones, los registros de llamadas telefónicas, los lectores de las tarjetas inteligentes de transporte, las aplicaciones de los servicios de movilidad, las cámaras de tráfico, las espiras, etc. Además los usuarios de Internet generan una enorme cantidad de datos cuando usan la red: buscadores, páginas web visitadas, redes sociales, entre otros. Se habla de una auténtica revolución de los datos, ya que se calcula que en los últimos dos o tres años se han generado tantos datos como en toda la historia de la humanidad.

El Big Data constituye una valiosísima materia prima para el estudio de la movilidad urbana. De hecho, la mayor parte de los datos masivos que registran la actividad humana tienen localización espacial y temporal. Por lo tanto siguiendo la huella digital de un mismo usuario es posible conocer su localización en el espacio y en el tiempo e inferir sus patrones de movilidad.

Hasta ahora, los estudios de movilidad urbana y la planificación del transporte se han basado sobre todo en las encuestas domiciliarias de movilidad, que ofrecen información muy rica sobre las características de la movilidad en las ciudades, pero esa información se refiere a un momento en el tiempo y queda pronto desactualizada. En cambio el Big Data genera información de forma continua, por lo que puede contribuir a mantener actualizada la información de las encuestas, particularmente en lo que se refiere a las matrices de viajes, lo que resulta crítico en la evaluación de planes y proyectos de infraestructuras y servicios de transporte. Distintos actores del sector, necesitan disponer de información actualizada para mejorar el proceso a fin de tomar decisiones óptimas.

Los datos de telefonía móvil son actualmente la fuente de datos masivos más utilizada en la planificación del transporte. En la sociedad hiperconectada en que vivimos, casi todos llevamos con nosotros un teléfono móvil. Ese dispositivo informa sobre los lugares que visitamos y los viajes que hacemos. La actividad de cada teléfono móvil (llamadas, mensajes, sesiones de datos) es captada por la antena más próxima al dispositivo, informando sobre nuestra localización. La geolocalización se puede hacer a partir de las áreas de influencia de las torres de comunicaciones o por triangulación de señales, lo que permite aumentar la exactitud posicional de cada registro.

La geolocalización es más precisa en áreas urbanas, que cuentan con mayor densidad de antenas.

El análisis de los registros de actividad de cada usuario permite conocer su localización en el espacio y en el tiempo, diferenciando entre los periodos en los que permanece en un mismo lugar (secuencias de registros consecutivos en una misma localización) y los viajes entre lugares (registros consecutivos en los que cambia la localización del usuario). Por lo tanto, a partir de este tipo de datos es posible conocer los lugares en los que el usuario realiza sus actividades (domicilio, trabajo, otras actividades), analizar sus patrones de movilidad y generar matrices de viajes origen-destino, tanto totales como por franjas horarias.

La utilización de esta fuente de datos presenta sin duda innegables ventajas: los datos se recogen de forma pasiva sin que el usuario tenga que hacer nada para que se almacenen (se registra lo que el usuario hace, no lo que dice que hace); los datos se registran de forma continua, con lo que es posible disponer de información siempre actualizada, monitorizar procesos y analizar los patrones de movilidad en situaciones atípicas, como son periodos de vacaciones, eventos turísticos o catástrofes; las muestras telefonía móvil son enormes, mucho mayores que las que se obtienen en encuestas domiciliarias de movilidad, y su nivel de sesgo es bajo (aunque dependiendo de la penetración de la operadora en la ciudad a analizar); y los plazos y el costo de los estudios de movilidad basados en datos de telefonía móvil son mucho menores que los de los estudios basados en encuestas domiciliarias de movilidad.

Además, los datos de telefonía móvil registran también los viajes de la población no residente (particularmente commuters con residencia fuera del área de estudio y turistas). Adicionalmente, son capaces de penetrar en los barrios informales de forma efectiva. Son bien conocidas las dificultades que encuentran los encuestadores para hacer su trabajo en zonas de transporte de estas características, por lo las muestras obtenidas en las encuestas de movilidad suelen ser muy pobres o incluso inexistentes. El uso de datos de telefonía móvil puede ser una alternativa razonable para conocer los patrones de movilidad en barrios informales.

Junto a estas ventajas, aparecen también importantes limitaciones y dificultades en el uso de datos de telefonía móvil en los estudios de movilidad urbana: la naturaleza de los datos dificulta las tareas de almacenamiento y proceso de datos, y obliga a usar tecnologías Big Data; la información sobre los viajeros es limitada, sobre todo en comparación con la obtenida a partir de las encuestas de movilidad, si bien las tareas de enriquecimiento de datos contribuyen a mitigar esta debilidad; también es limitada la información sobre ciertas características de los viajes, como el modo o el motivo y el no acceso directo a las bases de datos, que se encuentran en dominio de las operadoras de telefonía móvil.

Así pues, tanto las encuestas de movilidad como los datos de telefonía móvil presentan ventajas y limitaciones. A priori no se puede afirmar de forma general que una fuente de datos sea mejor que la otra, ni que los datos de telefonía móvil vayan a sustituir a los de las encuestas de movilidad. Si una ciudad no dispone de medios económicos para hacer una encuesta domiciliaria de movilidad, la utilización de datos de telefonía móvil puede ser una buena alternativa. Si existen recursos económicos suficientes, y se dispone ya de una encuesta de movilidad, pero ha quedado desactualizada, la actualización de los datos de la encuesta con datos de telefonía móvil puede ser una buena estrategia. Para ciertos objetivos bastará con utilizar datos de telefonía móvil, por ejemplo para analizar el impacto de un evento en la movilidad de la ciudad o para estudiar los patrones de movilidad de los turistas.

A la hora de analizar la movilidad lo mejor sería poder conjugar las ventajas de los datos procedentes de las fuentes tradicionales con las de las nuevas fuentes de datos, buscando la complementariedad entre los resultados obtenidos a partir de registros de teléfonos móviles y los análisis que se pudiesen realizar con otras fuentes de datos, como encuestas origen destino y otras encuestas de transporte (flujos, preferencias declaradas de modos, etc.), encuestas de ingresos/pobreza, censos, encuesta de empleo, etc.

Aunque existen numerosos informes y artículos sobre el uso de datos de telefonía móvil en estudios de movilidad, se echa en falta una obra de síntesis, que agrupe y sistematice los conocimientos generados por técnicos e investigadores en los últimos años en distintos países del mundo, y que sirva de orientación para aquellos planificadores que quieren adentrarse en este campo. La nota técnica que aquí se presenta intenta cubrir este vacío.

“Cómo aplicar Big Data en la planificación del Transporte Urbano. El uso de datos de telefonía móvil en el análisis de la movilidad” pretende mostrar los elementos fundamentales para aplicar el uso de Big Data a partir de datos de telefonía móvil a la hora de entender la movilidad en ciudades o espacios mayores de la Región de América Latina y el Caribe. **Su principal objetivo es el de brindar toda la información específica necesaria a técnicos y gestores de cara a la contratación de estudios de movilidad basados en datos de telefonía móvil.**

Para ello, la presente nota técnica comienza aportando información sobre las características de las redes de telefonía móvil y de los registros activos y pasivos generados por estas redes. A continuación describe las metodologías y los pasos a seguir para derivar de los datos la información útil en la planificación del transporte en las ciudades latinoamericanas. Se describen de forma sucesiva las metodologías para la limpieza y preparación de datos, selección de usuarios válidos, identificación de estancias y secuencia de viajes, análisis de localizaciones frecuentes y motivos de los desplazamientos, operaciones para el enriquecimiento de datos, estrategias para adaptar la geolocalización de los datos de telefonía móvil a las zonificaciones de transporte, tareas para la expansión de la muestra, obtención y visualización de las estadísticas de movilidad y validación de los resultados. Todo ello respetando la privacidad de los usuarios.

Para enriquecer la descripción de la metodología se muestran distintos ejemplos de aplicaciones y casos de éxito en ciudades latinoamericanas y en otros ámbitos internacionales, indicando cómo se han obtenido las matrices de viajes, pero además cómo esa información de movilidad se ha usado para analizar temáticas como la segregación urbana, la movilidad turística, la accesibilidad, la alimentación de modelos de transporte o el estudio dinámico de la población expuesta a la contaminación. De particular interés para las ciudades latinoamericanas es el caso de estudio sobre la movilidad en un barrio informal de Buenos Aires, Argentina, en el que se demuestra el potencial de los datos de telefonía móvil en este tipo de áreas.

Al final del documento se presentan unas consideraciones generales a modo de recapitulación y se hacen unas recomendaciones de base que se deberían seguir para la contratación de este tipo de estudios, incluyendo los requerimientos a las empresas y las especificaciones que éstas deben incluir en sus informes.

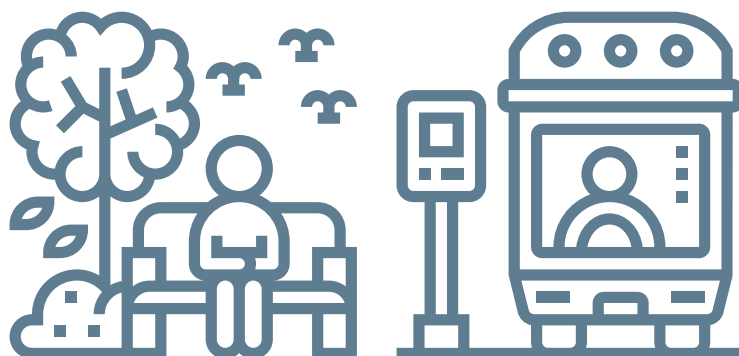
Una consideración práctica es que para obtener los mayores beneficios de las nuevas fuentes de datos será necesario modificar la composición de los equipos de planificación. Los planificadores del transporte no cuentan con la formación necesaria para trabajar con datos masivos y los analistas de datos masivos no tienen conocimientos en planificación del transporte. Parece claro que los equipos de planificación del transporte deberán incorporar expertos en análisis de Big Data y que cada perfil (el planificador y el analista de datos) deberán hacer un acercamiento al campo del otro, para poder llegar a alcanzar una colaboración fructífera. No se trata solo de datos, sino también de metodologías que nos permitan entender los patrones de movilidad. Ha quedado demostrado que los datos de telefonía móvil permiten que el planificador pueda disponer de información siempre actualizada para poder alimentar los modelos de transporte. Pero, además, las herramientas de inteligencia artificial ofrecerán al planificador nuevas oportunidades para dar solución a los viejos problemas de planificación.

En la región de América Latina y el Caribe se parte de una situación de cierto rezago con respecto a Europa en el uso de datos de telefonía móvil en el estudio de la movilidad urbana y la planificación del transporte, pero recientemente se están dando pasos muy esperanzadores tanto en las universidades como en las administraciones gubernamentales. La ventaja para la región es que las metodologías para el tratamiento de datos de telefonía móvil para analizar la movilidad urbana están bastante maduras y por lo tanto su aplicación en la región puede ser mucho más rápida y directa. Una de las claves para su difusión es sin duda que los planificadores de las ciudades de la región conozcan las nuevas fuentes de datos, sus ventajas y sus limitaciones. La movilidad del futuro requerirá y exigirá cuadros técnicos multidisciplinarios y especializados en el manejo de datos. Esta nota técnica es un intento de contribuir al objetivo de mejorar la capacitación de los cuadros técnicos de las ciudades de la región para aprovechar las ventajas de las nuevas fuentes de datos en la planificación del transporte.

The background of the entire page is a light blue-grey color with a white line pattern that resembles a stylized city street map or a network of roads. The lines are of varying thickness and form a complex, interconnected grid.

1

La movilidad urbana y los métodos tradicionales de recolección de datos de movilidad



1.1 La movilidad y la planificación del transporte urbano

La movilidad es uno de los principales retos a los que se enfrentan las ciudades contemporáneas. A medida que las ciudades han ido aumentando su tamaño, se han extendido en superficie y especializado internamente, con piezas dedicadas a distintas funciones. Los desplazamientos diarios de la población se han hecho más largos, complejos y numerosos, en gran parte mediante medios de transporte mecanizados. La movilidad urbana es un elemento clave en el funcionamiento de la ciudad y es causa de numerosas externalidades negativas, como contaminación atmosférica, ruido, consumo de espacio y consumo de energía, entre otros. El uso creciente del vehículo privado produce una congestión generalizada en las ciudades, especialmente en horas pico, y deteriora la calidad de vida de sus poblaciones.

La movilidad sostenible es uno de los grandes desafíos que tienen que enfrentar las ciudades de todo el mundo, particularmente las grandes metrópolis, muchas de ellas caracterizadas por muy elevadas tasas de crecimiento demográfico.

En este contexto se aboga por desarrollar políticas de movilidad basadas en los tres pilares de la sostenibilidad: ambiental, económica y social. La movilidad debe ser respetuosa con el medio ambiente, favorecer el desarrollo económico de las ciudades y promover la inclusión social. Así, se entiende que un sistema de transporte es sostenible cuando proporciona los medios para atender las necesidades económicas, medioambientales y sociales con equidad y eficiencia, evitando los impactos negativos y sus costes asociados, tanto en el tiempo como en el espacio. En un marco de movilidad sostenible, el sistema de transporte debe ser eficiente y fiable para garantizar que la actividad económica se desarrolle en las condiciones adecuadas; debe orientarse a asegurar el acceso de la población al empleo y los servicios, mejorando su calidad de vida y evitando situaciones

de exclusión social y desigualdad; y, por último, debe ser respetuoso con el medio ambiente, reduciendo los costos ambientales que produce: contaminación atmosférica y acústica, consumo de espacio, afecciones a fauna y flora y deterioro del paisaje, entre otros.

La planificación del transporte urbano se orienta a dar soluciones a los problemas de movilidad de las ciudades mediante la construcción de nuevas infraestructuras y la nueva oferta de servicios que permitan una movilidad en las mejores condiciones posibles. Para ello, es necesario disponer de modelos para estimar la demanda futura de transporte en diferentes escenarios, maximizar la rentabilidad económica y social de las actuaciones y priorizar las inversiones. Los modelos deben ser capaces de explicar satisfactoriamente la situación actual de la movilidad a partir de la relación entre el sistema territorial (actividades y usos del suelo) y el sistema de transporte. Al proyectar las principales variables explicativas (económicas, sociodemográficas y de usos del suelo) y tras considerar la futura oferta de transporte, es posible predecir las pautas de movilidad futuras y los efectos de las actuaciones previstas en materia de transporte para la toma de decisiones. Se puede simular la situación que se tendría en el escenario tendencial (sin actuaciones en materia de infraestructuras y servicios) para detectar déficits que sirvan para la propuesta de actuaciones. Asimismo, se puede comparar el escenario con proyecto (con la actuación o actuaciones seleccionadas) con el escenario sin proyecto (tendencial) para conocer así los efectos del proyecto en la movilidad futura.

Durante las últimas décadas, la planificación del transporte urbano ha estado dominada por el modelo de cuatro etapas. Este modelo debe obtener la matriz origen-destino total, es decir el número agregado de viajes entre zonas de transporte, previamente definidas y que agrupen sectores con comportamiento homogéneo contiguos. Este modelo también permite conocer el reparto modal entre cada zona origen y destino (matriz origen-destino por modo de transporte), cuáles son las rutas elegidas y, en consecuencia, el volumen de viajeros y vehículos en los elementos de la red de transporte.





1.2 Fuentes de datos tradicionales: las encuestas de movilidad

Los modelos de transporte se alimentan de datos de distintas fuentes, fundamentalmente de estadísticas oficiales, de usos del suelo, del funcionamiento del sistema de transporte (aforos de tráfico, frecuencias de paso del transporte público, etc.) y sobre todo de encuestas domiciliarias de movilidad. Con estas encuestas se pretende recoger información sobre los viajes realizados por la población residente en un día laborable medio y sobre determinadas variables que inciden en esos viajes. Normalmente los cuestionarios tienen tres partes:

- **Datos sobre el hogar:** dirección, características de la vivienda, número de personas que viven en el hogar, renta, número de vehículos, etc.
- **Datos sobre el individuo:** edad, género, nivel educativo, situación de empleo, disponibilidad de licencia de conducir, etc.
- **Datos sobre los viajes:** lugar de origen, destino, etapas del viaje, hora de salida y de llegada, motivo, modo, etc.

Los datos de los viajes se refieren al día anterior a la realización de la encuesta. La muestra debe ser representativa de la movilidad del conjunto de la población de la ciudad, pero también de cada una de sus zonas de transporte. Una vez realizada la encuesta, los datos deben ser depurados, para eliminar errores, y expandidos, así se obtiene una imagen de la totalidad de los movimientos que realizan los habitantes de la ciudad en día laborable medio.

Las matrices de viajes origen-destino (OD) se generan a partir del capítulo de viajes. Al seleccionar los viajes por franjas horarias y modos de transporte se calculan matrices que contienen el número de viajes entre cada par OD (zonas de transporte) por horas del día (por ejemplo, hora pico) y modo (por ejemplo, vehículo privado). Estas matrices OD constituyen una pieza clave para alimentar los modelos de transporte.

Las encuestas domiciliarias de movilidad son muy costosas y suponen un proceso largo y laborioso desde que se hace el diseño hasta que los datos están disponibles para el planificador. Especialmente crítico es el trabajo de campo, en el que un gran número de encuestadores realiza visitas a las personas seleccionadas en la muestra. Así, por ejemplo, en la encuesta de movilidad de Bogotá de 2015 se encuestó a algo más de 28.000 personas. Recientemente también realizaron encuestas telefónicas, donde los resultados obtenidos en estos casos no fueron exactamente comparables a los que se obtuvieron mediante visitas domiciliarias. El elevado costo y tiempo empleado en hacer y procesar las encuestas domiciliarias de movilidad explican que éstas se realicen en intervalos de tiempo relativamente dilatados, por ejemplo, cada 8 o 10 años.



1.3 Limitaciones de las encuestas de movilidad

Las encuestas domiciliarias de movilidad ofrecen una información de gran valor a los planificadores del transporte. Con estos datos se analizan las pautas de movilidad (número de viajes por persona, orígenes y destinos distribución temporal, etapas, modo, motivo...) y se alimentan los modelos de transporte. Sin embargo, esta fuente de datos presenta también algunas debilidades, como por ejemplo:

Sesgo: Cada vez es más difícil que las personas seleccionadas en la muestra respondan a los encuestadores, en un contexto de proliferación de encuestas de todo tipo que saturan a la población de las ciudades. Este hecho no solamente encarece el proceso, sino que además es una fuente de sesgo en la muestra. Aunque las personas que rehúsan contestar son sustituidas por otras con características sociodemográficas semejantes, el mero hecho de negarse a contestar puede reflejar características personales que influyan en la movilidad. Por otro lado, los encuestadores se enfrentan con graves dificultades para encuestar a determinados colectivos, como los de nivel de renta muy alto y los que habitan en barrios particularmente conflictivos. En el primer caso, porque estas personas suelen negarse a contestar para no suministrar información sobre sus rutinas de movilidad que pueda comprometer su seguridad personal; en el segundo caso, porque los encuestadores tienen serias dificultades para encuestar en barrios inseguros.

Omisión de viajes: Los encuestados tienden a declarar los viajes que consideran más importantes (por ejemplo, el viaje al trabajo), pero omiten algunos viajes que consideran menos relevantes, como los de corta distancia y los que no se realizan diariamente. Estas omisiones afectan particularmente a los desplazamientos en modos no motorizados, sobre todo peatonales, que deberían ser potenciados en una estrategia de movilidad sostenible. Por otro lado, estas encuestas tampoco recogen la movilidad de los turistas, un segmento de importancia creciente en la demanda de movilidad en numerosas ciudades, entre ellas muchas de América del Sur y el Caribe.

Desactualización de los datos: Dado el largo tiempo transcurrido entre una encuesta de movilidad y la siguiente, los planificadores de las ciudades con frecuencia se ven obligados a utilizar datos desactualizados, lo que resta fiabilidad a los resultados de los modelos en un contexto en el que la movilidad urbana cambia con gran rapidez. En realidad, nunca disponen de datos actualizados, ya que el proceso de elaboración de las encuestas de movilidad es bastante largo.

Falta de información de movilidad en días atípicos: Las encuestas de movilidad normalmente solo recogen la movilidad en un día laborable medio, de manera que las autoridades de transporte no suelen disponer de datos fiables sobre la movilidad durante los fines de semana o en vacaciones, y tampoco en situaciones excepcionales como la celebración de eventos masivos o las catástrofes naturales.

Las nuevas fuentes de datos del ámbito general del Big Data ofrecen información complementaria que permite superar o al menos mitigar estos problemas.

Dada la rápida evolución de las ciudades de Latinoamérica y el Caribe, disponer de matrices OD actualizadas y fiables obtenidas a partir de fuentes de datos Big Data repercutirá muy positivamente la evaluación de proyectos y la toma de decisiones en políticas de transporte en la región.

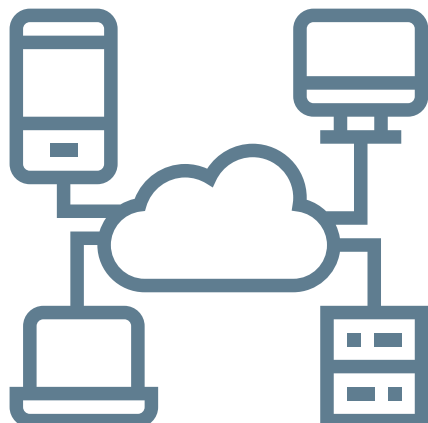
La posibilidad de identificar a los pasajeros con sus dispositivos permite la toma de datos de manera automática y en línea con la sola autorización de los mismos. Esta modalidad de recogida de datos nos permite minimizar las dificultades explicadas por la completitud y masividad que adquieren estos datos, pero conlleva otras dificultades que surgen por deficiencias técnicas en la exactitud de los datos recibidos, su verificación y la asociación con un pasajero unívoco que nos permita suponer una trazabilidad completa de sus conductas diarias. La ventaja del gran volumen de datos que tomaremos de esta forma es que nos permitirán ajustar nuestras mediciones a lo largo del tiempo, aumentando la exactitud y confianza de las mismas.



A stylized, light blue map of a city street grid is visible in the background. The lines represent streets and are of varying thicknesses, creating a complex network pattern.

2

De las encuestas de movilidad a las nuevas fuentes de datos: el Big Data



2.1 La revolución de los datos: el Big Data

Asistimos en la actualidad a una auténtica revolución de los datos.

Se calcula que en los últimos dos o tres años se han generado tantos datos como en toda la historia de la humanidad.

Estos grandes volúmenes de información reciben genéricamente el nombre de datos masivos o Big Data. Se trata tanto de datos producidos por multitud de sensores y dispositivos, como de contenidos generados por los usuarios cuando utilizan Internet:

- **Datos generados por sensores y dispositivos.** Multitud de sensores y dispositivos registran la actividad humana y aportan información sobre la movilidad: dispositivos GPS de los smartphones, registros de llamadas telefónicas, navegadores, tarjetas inteligentes de transporte, pagos con tarjetas de crédito, aplicaciones de servicios de movilidad, cámaras de tráfico, espiras, etc.
- **Datos generados por los usuarios en Internet.** La actividad de los usuarios en Internet queda registrada y almacenada en la nube. Esto ocurre con el uso de multitud de servicios, como el buscador de Google y también con las redes sociales, como Twitter o Flickr.

El Big Data se suele caracterizar por las denominadas Tres V (ver figura 2):

Volumen: Se trata de conjuntos de datos masivos, que ya no se miden en gigabytes, sino en terabytes (un billón de bytes), petabytes (mil billones de bytes), exabytes (un millón de billones de bytes), etc. Así, por ejemplo, las compañías de telefonía móvil almacenan millones de datos en un solo día en un solo país.

Velocidad: Los datos se producen de forma continua, a gran velocidad, lo que permite monitorizar procesos y tener datos siempre actualizados. Frente a las encuestas de movilidad, que ofrecen una instantánea de la movilidad en una ciudad en un momento dado, la velocidad asociada al Big Data aporta datos con

alta resolución temporal, lo que permite seguir la evolución de la movilidad de forma continua.

Variedad: La variedad hace referencia a la diversidad de tipos, formatos y fuentes de datos. Pueden tratarse de datos estructurados (los que pueden ser presentados en forma de tabla y manejados con sistemas gestores de bases de datos SQL), semi-estructurados (como los ficheros HTML de páginas web) y no estructurados (archivos de texto, fotografías, vídeos, etc).

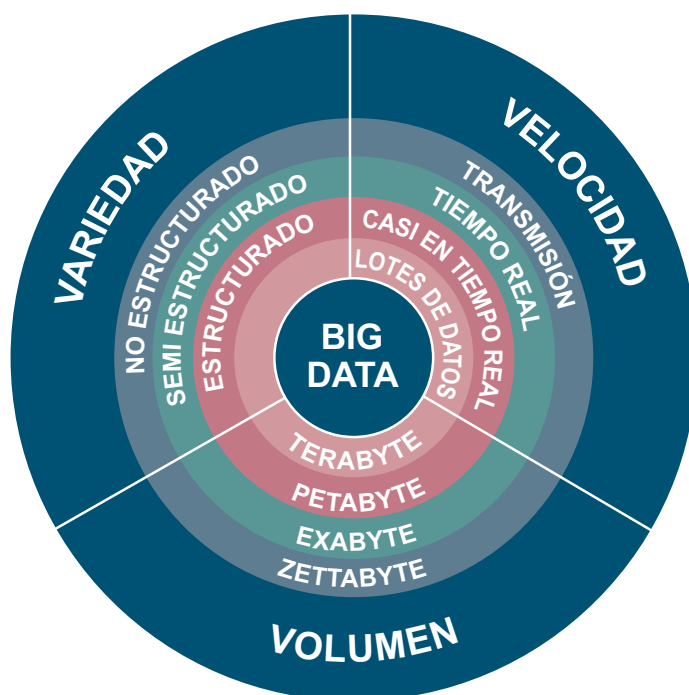


Figura 2: Las Tres Vs del Big Data: volumen, velocidad y variedad.

Los datos masivos, por sus características especiales, plantean nuevos retos para su tratamiento. Deben ser procesados mediante tecnologías específicas, las denominadas tecnologías Big Data. Normalmente se recurre a la computación paralela y distribuida:

- La computación distribuida permite aumentar la capacidad de los sistemas conectando entre sí tantos ordenadores como sean necesarios, formando clusters. Estos sistemas distribuidos tienen la ventaja de su escalabilidad, de forma que se puede adaptar la capacidad de computación de estos clusters a las necesidades de cada momento, añadiendo o retirando ordenadores.
- La computación paralela se basa en el principio de que un gran conjunto de datos se puede dividir en partes más pequeñas para realizar procesos de forma simultánea, es decir, en paralelo, utilizando varios ordenadores conectados en red, para hacer converger después los resultados obtenidos en un único ordenador. La paralelización de procesos se puede llevar a cabo en sistemas propios o mediante alquiler de servicios de computación en la nube (cloudcomputing).

Por otro lado, la variedad de datos inherente al Big Data puede aconsejar utilizar sistemas NoSQL más flexibles y rápidos que los sistemas gestores de bases de datos relacionales.

Computación Paralela

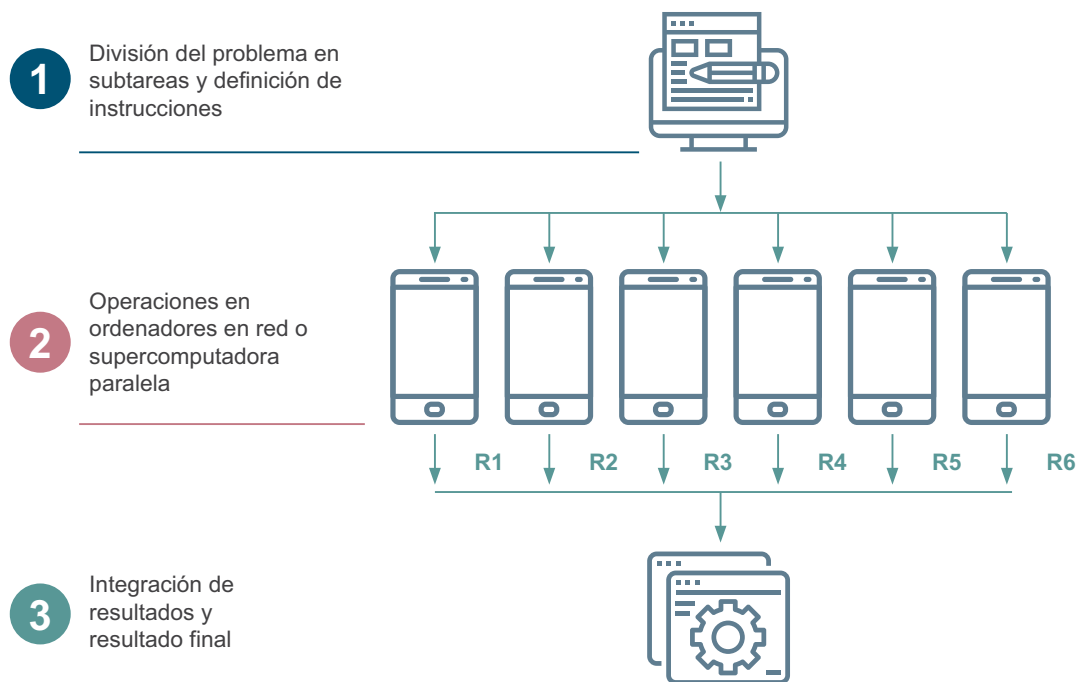


Figura 3: Computación paralela y distribuida para el tratamiento de grandes volúmenes de datos.





2.2 El análisis de la movilidad de la población a través del Big Data

Las nuevas fuentes de datos conocidas de forma genérica como Big Data constituyen una valiosísima materia prima para el estudio de la movilidad de la población. La mayor parte de los datos masivos que registran la actividad humana tienen localización espacial y temporal. Por lo tanto, al seguir la huella digital de un mismo usuario es posible conocer su localización en el espacio y en el tiempo e inferir sus patrones de movilidad.

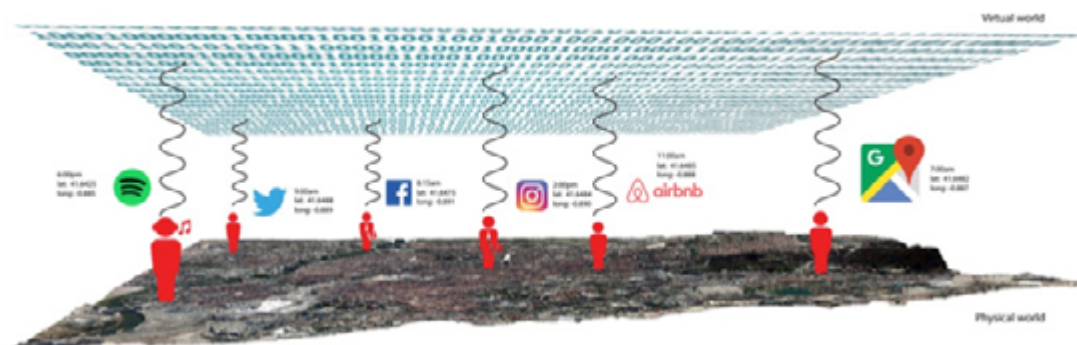


Figura 4: La huella digital y su localización en el espacio y en el tiempo. La secuencia temporal de la geolocalización de un individuo permite seguir sus desplazamientos en la ciudad.

Lógicamente este análisis será más preciso cuando mayor sea la resolución espacial y la granularidad temporal de los datos:

Resolución espacial: Hace referencia a la exactitud de la geolocalización, es decir el grado de precisión de la localización del dato en el espacio. La resolución espacial es muy alta en datos captados por sensores fijos o mediante GPS, para los que se dispone de coordenadas x y (latitud y longitud). Los datos de telefonía móvil tienen una precisión media-alta, con errores de centenares de metros, suficiente para analizar la movilidad en las ciudades. Datos con alta resolución espacial permiten conocer con precisión la localización de las actividades, los orígenes y destinos de los viajes y las rutas seguidas por los mismos.

Granularidad temporal: Se refiere a la frecuencia con que se registran los datos. Los tracks obtenidos a partir de dispositivos GPS tienen una resolución temporal muy alta (se registra la localización cada varios segundos o minutos, según la configuración del dispositivo), los registros de actividad de los teléfonos móviles tienen una granularidad temporal alta (por ejemplo, un registro cada veinte o treinta minutos) y los de transacciones con tarjetas bancarias suelen tener una frecuencia baja (por ejemplo, uno o dos registros al día). La granularidad temporal es clave para conocer las distintas localizaciones visitadas por un mismo individuo y para determinar las horas de comienzo y fin de los viajes.

Entre las nuevas fuentes de datos utilizadas en la planificación del transporte se destacan particularmente los registros de telefonía móvil

(ver Chen et al, 2016). El análisis de los registros de actividad de cada usuario (llamadas, mensajes, sesiones de datos) permite conocer sus trayectorias espacio-temporales, diferenciando entre las horas de permanencia en un mismo lugar (estancias) y los desplazamientos entre lugares (viajes) (Figura 5). Por lo tanto, a partir de este tipo de datos es posible generar matrices de viajes origen-destino según franjas horarias y, en general, analizar los patrones de movilidad de la población. La utilización de esta fuente de datos presenta sin duda innegables ventajas con respecto a las encuestas de movilidad tradicionales, pero también algunos puntos débiles que deben ser considerados (ver capítulo 8).

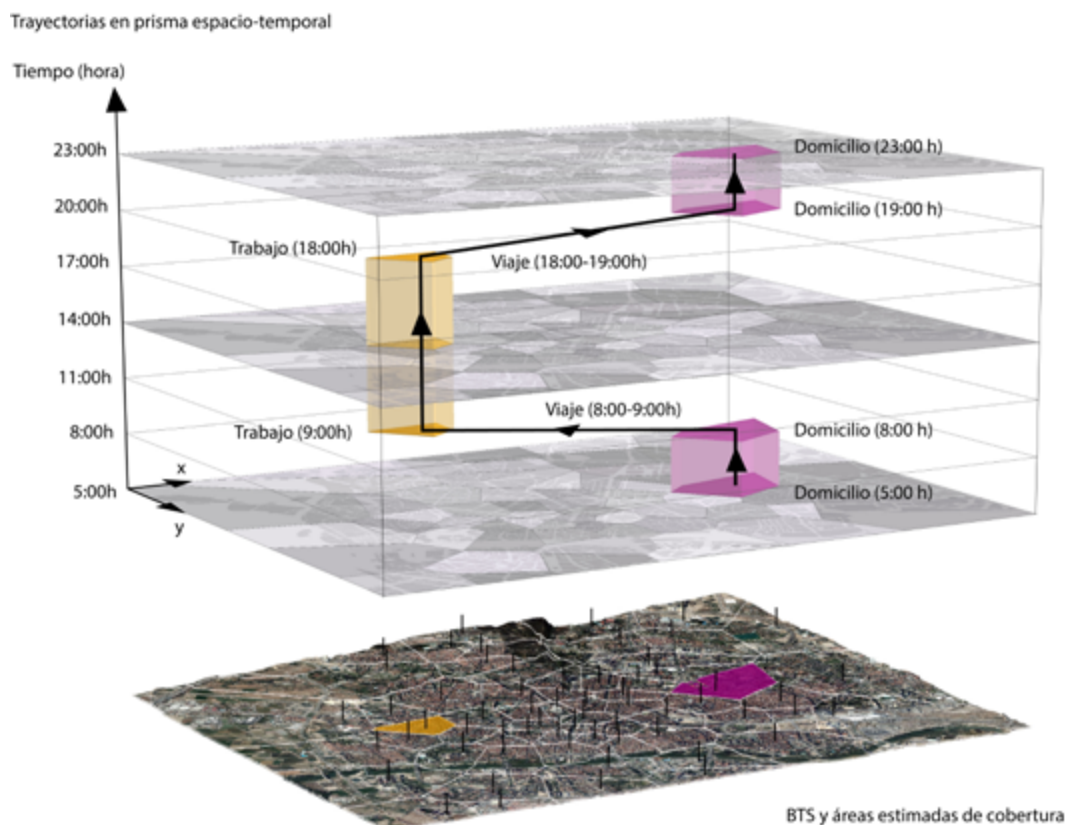


Figura 5: Trayectorias espacio-temporales. El eje vertical representa el tiempo (horas del día). El plano horizontal representa el espacio. Las líneas verticales indican estancias (permanencias en un mismo lugar asociadas a la realización de actividades). Las líneas diagonales se corresponden con viajes (desplazamientos entre lugares que llevan un cierto consumo de tiempo).

Otras nuevas fuentes de datos de gran utilidad en el análisis de la movilidad urbana son, por ejemplo, las siguientes:

Tarjetas inteligentes de transporte: Las tarjetas que se utilizan en muchas ciudades para viajar en transporte público generan una enorme cantidad de datos. Una ventaja de esta fuente de datos es que permite identificar inequívocamente el modo de transporte. Sin embargo, las tarjetas inteligentes de transporte en algunas ciudades no se utilizan en todos los viajes en transporte público. Además, en algunos casos solo se registra el lugar y el momento del día donde se accede al

sistema de transporte, pero no el lugar y momento de salida, que debe ser inferido, lo que supone una dificultad a la hora de construir matrices origen-destino.

Planificadores de rutas: Planificadores de rutas como Google Maps, TomTom MyDrive o Waze permiten a los usuarios calcular tiempos de viaje. Estos tiempos pueden ser obtenidos en forma de matrices según franjas horarias, ya sea a través de una API (Google Maps) o de la compra de datos (TomTom Speed Profiles). Waze ofrece además una información muy valiosa para conocer las incidencias del tráfico.

Registros GPS (tracks): Los autobuses urbanos (en muchas ciudades) y los vehículos de los nuevos sistemas de movilidad compartida (bicicletas, motos, patinetes, coches eléctricos) están equipados con dispositivos GPS que almacenan datos sobre los viajes realizados. Esta fuente de datos tiene una muy alta resolución espacial y temporal.

Redes sociales: Las redes sociales permiten la interacción con otras personas a través de diferentes medios, generando diariamente una gran cantidad de datos.

En muchos casos los datos de las redes sociales están geolocalizados y tiene un carácter dinámico debido a su constante uso, lo que permite conocer las pautas de distribución espacial de los usuarios y sus patrones de movilidad.

Una de las redes sociales más populares es Twitter, a partir de la cual se han estudiado los patrones de movilidad general en varias ciudades. En las redes sociales también se vierten opiniones sobre el funcionamiento de los sistemas de transporte, cuyo análisis pueden ser de utilidad para los gestores de la movilidad de las ciudades.

Tarjetas bancarias: Las tarjetas bancarias también permiten localizar a su tenedor en los dispositivos y cajeros donde hayan sido usadas, incluso en máquinas facturadoras de billetes de transporte público que aceptan pago por dicho medio. La localización en el espacio y en el tiempo de cada una de las transacciones bancarias realizadas por los usuarios de las tarjetas permite realizar un seguimiento de los individuos y conocer sus pautas de movilidad. Esta fuente tiene una resolución espacial alta, pero su resolución temporal es relativamente baja y está sesgada hacia la movilidad ligada al consumo. Refleja adecuadamente la movilidad interurbana, pero presenta muchas limitaciones en el análisis de la movilidad urbana.

Otros sensores y dispositivos: Existen diversos tipos de sensores y dispositivos que también aportan información útil para el análisis de la movilidad urbana. Por ejemplo, los detectores de bluetooth y las redes wifi registran la presencia de individuos a través de sus teléfonos móviles, las cámaras de tráfico permiten el conteo de vehículos mediante la identificación de sus matrículas y las espiras posibilitan conocer la velocidad de circulación..

Estas fuentes de datos, de carácter muy diverso, aportan información complementaria a las fuentes tradicionales. La telefonía móvil aporta información siempre actualizada sobre la movilidad de la población, incluso en días atípicos. Los tiempos de viaje y las horas de comienzo y fin del viaje, que en las encuestas

de movilidad son estimados por los entrevistados, se pueden conocer con precisión gracias a los registros GPS de los vehículos o sus ocupantes. Asimismo, la ruta seguida por el conductor, que no es recogida en las encuestas de movilidad, es almacenada con detalle en los registros GPS y de forma aproximada se puede obtener también a partir de los datos de telefonía móvil. Sin embargo, las encuestas domiciliarias de movilidad ofrecen una información conjunta mucho más rica en cuanto a características de los viajes y los individuos que cada una de las nuevas fuentes de datos. No se trata de debatir sobre qué fuentes son mejores, si las tradicionales o las Big Data. Lo que se busca es aprovecharlas que sean más útiles para cada objetivo y que el planificador pueda disponer de una información más variada y actualizada al cruzar fuentes tradicionales y Big Data (ver estudios de caso 7).

El presente documento se centrará exclusivamente en la primera fuente de datos mencionada: los registros de telefonía móvil.



A light gray background with a faint, stylized map pattern of streets and blocks.

3

Estructura de las redes de telefonía móvil



3.1 Estructura de las redes

La actividad de los teléfonos móviles se recoge a través de las redes de infraestructuras de las operadoras. Una red de telefonía móvil está compuesta por un conjunto de torres que dan cobertura a un área geográfica. Cada torre tiene su área de cobertura y éstas a su vez se agrupan en áreas de localización (LAs - Location Areas):

Torres de telefonía móvil. Los datos se recogen a nivel de torres, denominadas también Base Transceiver Station (BTS) o simplemente base station (Figura 6a). Una torre contiene varias antenas de telefonía, que cubren un ángulo de 120 grados, de forma tal que tres antenas cubren de forma conjunta los 360 grados.

Área de cobertura o celda. Cada torre da servicio a una porción limitada del territorio, llamada “área de cobertura celular”, o simplemente “celda” (Figura 6b). Solo los teléfonos móviles localizados dentro del área de cobertura pueden conectarse a la torre asociada. Una celda está dividida en tres sectores, uno por cada antena, cada uno de los cuales cubre 120 grados.

LA (Location Area). Para optimizar las señales de telefonía, las celdas se agrupan en zonas mayores para las que se guarda un código de Localización de Área Común (LAC), que identifica la zona de ubicación de la torre. Cada LA suele tener entre 10 y 100 torres (Figura 6d).

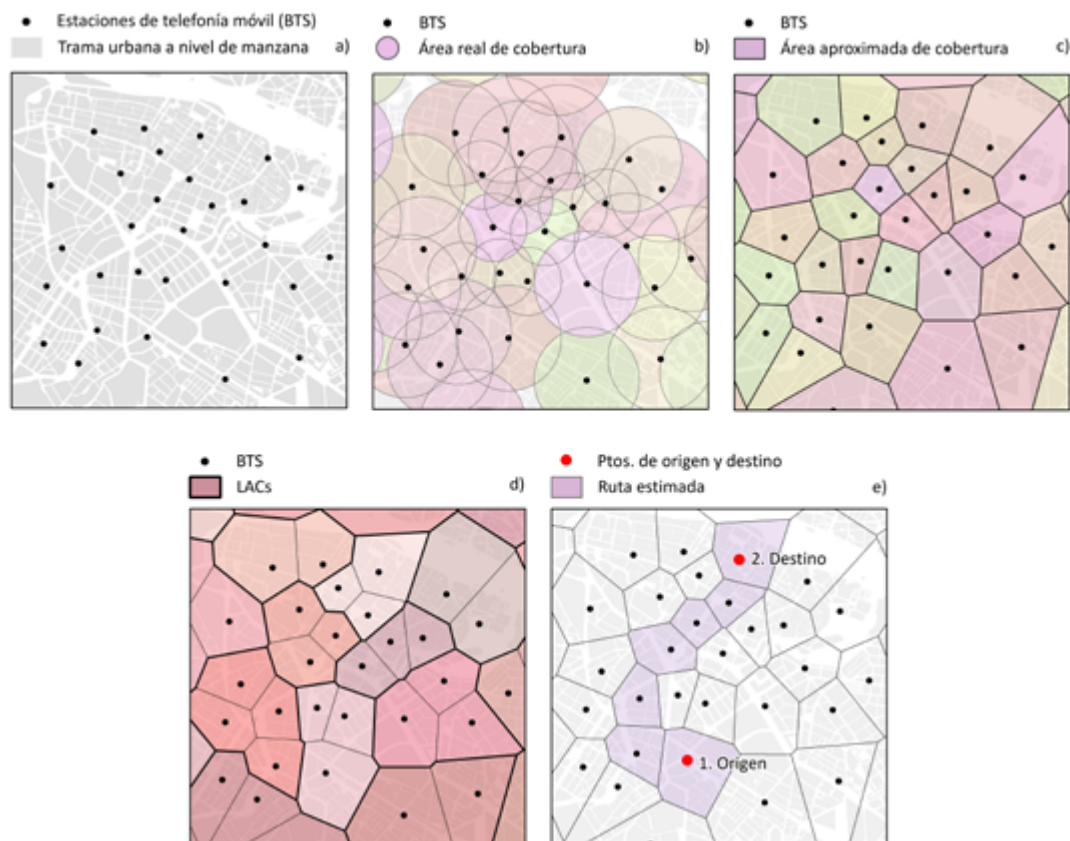


Figura 6: Áreas de cobertura de las antenas de telefonía móvil y encaminamiento de los viajes (centro de Zaragoza, España). a) Localización de las torres de telefonía móvil (BTS). b) Áreas de cobertura reales de las antenas. c) Áreas de cobertura aproximadas a través de polígonos Voronoi. d) Representación de los LAs como agrupación de polígonos Voronoi. e). Estimación del origen, destino y encaminamiento de un viaje.



3.2 Características de las áreas de cobertura

Las áreas de cobertura de las torres de telefonía móvil (BTS) presentan las siguientes características:

- Las áreas de cobertura reales de las torres se solapan parcialmente, de forma tal que la conexión de dos llamadas realizadas desde una misma localización puede establecerse a través de dos torres distintas, dependiendo de la distancia a las torres y del tráfico de la red. Para simplificar el problema, los estudios basados en datos de telefonía móvil delimitan las áreas de cobertura mediante polígonos Voronoi a partir de la localización de las torres, con lo que finalmente se obtiene un conjunto de polígonos (teselación) sin solapes. Estas son las áreas de cobertura simuladas o aproximadas (Figura 6d). Una persona que se desplaza atraviesa varios polígonos Voronoi, por lo que a partir de los registros de su teléfono móvil se puede seguir la trayectoria de su viaje de forma continua o discontinua (Figura 6f).
- Las áreas de cobertura tienen una extensión variable, dependiendo de la tecnología y de la demanda que soportan. La demanda introduce diferencias

significativas en el espaciado de las torres entre las zonas centrales y las periféricas en las propias ciudades y, sobre todo, entre las ciudades y las zonas rurales (Figura 7). La extensión de las áreas de cobertura normalmente oscila entre 1 km² en áreas urbanas densas y 4 km² en áreas rurales. Las celdas más pequeñas están en puntos específicos de alta densidad de población, tanto en exteriores (outdoor) como en interiores (indoor): centros comerciales, estaciones de tren, aeropuertos, etc. En el caso de las LAs el tamaño del área cubierta varía según operadores, pero en términos medios puede estar en unos 5-10 km de diámetro medio en las grandes ciudades, y de entre 20 km y 40 km de diámetro medio en las áreas rurales.

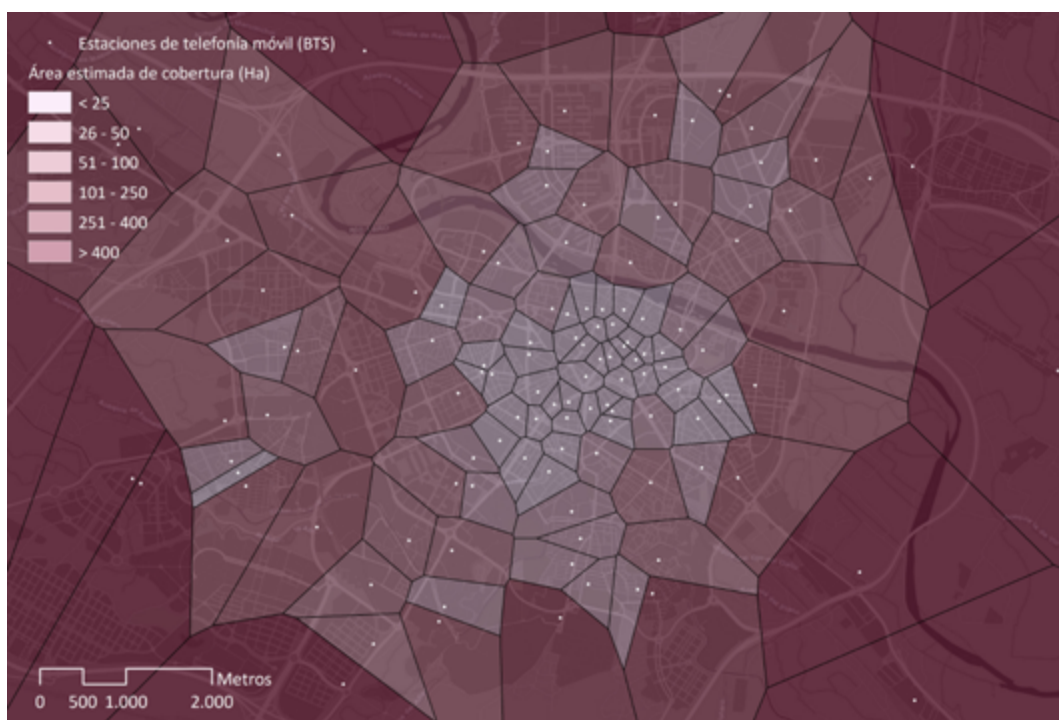


Figura 7: Tamaño de las áreas de cobertura estimadas (polígonos Voronoi) decrecientes del centro de la ciudad (con mayor densidad de torres de telefonía) a la periferia (con menor densidad). La precisión en la geolocalización de los usuarios aumenta del centro a la periferia (Zaragoza, España).



3.3 Geolocalización de la actividad de los teléfonos móviles

La geolocalización de los registros de actividad de los teléfonos móviles se puede hacer a través de dos procedimientos:

Por área de cobertura. La geolocalización de los registros de telefonía móvil se aproxima a la de la torre que transmite la llamada, para la que se conocen sus coordenadas x y, o directamente al polígono Voronoi de esa torre. Por lo tanto, lo que se conoce es que el teléfono móvil se encuentra cerca de la torre que transmite la llamada o, más concretamente, dentro del correspondiente polígono Voronoi, pero no se advierte con exactitud cuál es su localización. Dado que la densidad de antenas es mucho mayor en áreas urbanas que en ámbitos rurales, la exactitud

posicional de los registros de telefonía móvil es muy variable, lo que puede oscilar entre algunos centenares de metros en las primeras y varios kilómetros en las segundas. Una importante implicación de esta forma de geolocalización es que los viajes cortos que se desarrollan dentro del área de cobertura de una antena no pueden ser captados. Con esta forma de geolocalización no es posible diferenciar entre una persona que permanece en un mismo lugar y otra que se desplaza entre dos puntos dentro del área de cobertura de una antena. En ambos casos se interpretaría que la persona permanece fija.

Por triangulación (trilateración). Esa exactitud posicional puede ser mejorada utilizando técnicas de triangulación (trilateración) a partir de las antenas más próximas a la localización del teléfono móvil, de forma tal que es posible estimar las coordenadas x y con un cierto margen de error. En áreas urbanas, donde la densidad de torres de telefonía es mayor, puede situarse en torno a los 50 o 100 metros. La triangulación se hace a partir del cálculo de la distancia del dispositivo móvil con respecto a las antenas más próximas, considerando la atenuación de la señal con la distancia y el tiempo de viaje de la misma (Figura 8).

Los datos triangulados permiten conocer la posición del teléfono móvil de forma mucho más precisa, pero resultan muy costosos, por lo que no suelen estar disponibles para estudios de movilidad.

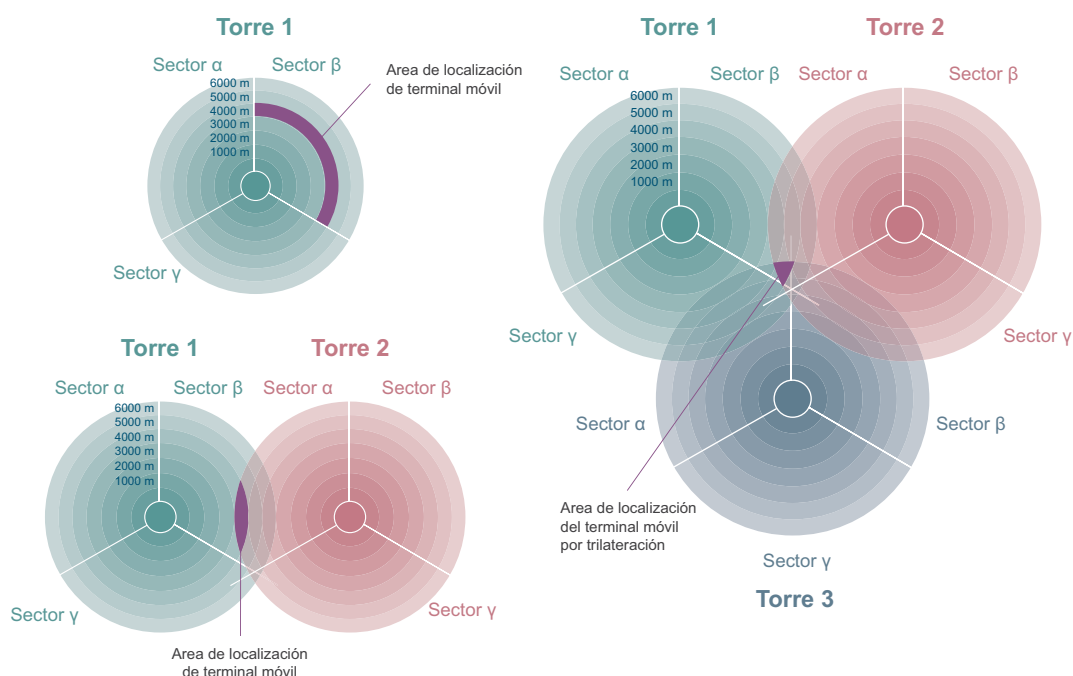


Figura 8: Localización de la llamada de un teléfono móvil cuando hay cobertura por una torre, dos torres o tres torres. En este último caso es cuando realmente se puede recurrir a la triangulación (o de forma más precisa, trilateración) y cuando la llamada se localiza con mayor exactitud posicional. Los círculos indican la distancia a la torre y los sectores se refieren a la antena con la que se conecta la llamada.

The background of the page is a light blue-grey color with a white line pattern that resembles a stylized map of city streets or a network of roads. The lines are of varying thickness and form a complex, interconnected grid.

4

Datos de telefonía móvil



4.1 Tipos de información que aportan

Los datos de telefonía móvil aportan tres grandes tipos de información:

Sobre los usuarios: Las operadoras disponen de información sobre el cliente asociada a los contratos, como la edad, género, domicilio y otros datos. Esta información normalmente no está disponible por razones de privacidad, aunque en algunos casos las propias operadoras pueden hacer uso de ella. Por otro lado, esta información no suele ser muy fiable. De hecho, muchas terminales no tienen contrato o los mismos figuran a nombre de empresas, aunque son utilizados por particulares; y además los clientes no suelen actualizar los datos de sus contratos (por ejemplo, cuando cambian de domicilio). De ahí que en los estudios de movilidad basados en registros de telefonía móvil se recurra a operaciones de enriquecimiento de los datos, ya sea de forma individual o agregada (ver apartado 5.5).

Sobre la infraestructura de la red: Datos sobre la localización (latitud y longitud) de las torres de comunicación y sobre la tecnología de las antenas (2G, 3G, 4G). Esta información permite la geolocalización de las llamadas.

Sobre la conexión del teléfono a la red: Son los datos asociados a los registros de comunicación del dispositivo con la red de telefonía móvil. Se les conoce como eventos y pueden ser de dos tipos:

- Eventos activos (Event-Driven Cell Phone Data), registros de las actividades realizadas por los usuarios de los teléfonos móviles. Incluyen los CDR (Call Detail Records) y las sesiones de datos.
- Eventos pasivos (Network-Driven Cell Phone Data), registros relacionados con la actualización periódica de la posición del dispositivo. Estos datos no suelen estar disponibles para estudios de movilidad.

A continuación, se describen las principales características de los datos activos y los datos pasivos.



4.2 Eventos activos: CDR (Call Detail Records) y sesiones de datos

Una de las fuentes de datos masivos más utilizadas actualmente en la planificación del transporte es sin duda la constituida por los registros de la actividad de los teléfonos móviles, que incluyen los denominados CDR (Call Detail Records) relativos a las llamadas y mensajes SMS y los registros de sesiones de datos (Internet Access logs), también denominados xDR. Estos registros, asociados a la facturación del servicio, se almacenan cada vez que:

- Se realiza o recibe una llamada,
- Se envía o recibe un mensaje (SMS),
- Se abre una sesión de datos (conexión a Internet o actualizaciones del sistema).

Los registros de llamadas (CDR) pueden presentar distintos formatos e información variada (según la tecnología empleada por el operador y según los operadores), pero en general comparten una serie de elementos comunes. En el caso de una llamada telefónica suelen incluir:

- Número de teléfono (encriptado) que realiza la llamada
- Número de teléfono (encriptado) que recibe la llamada
- Fecha y hora de inicio de la llamada
- Duración
- ID (identificador) de celda (antena) a la cual se conecta el dispositivo
- Tipo de llamada (entrante, saliente)
- Identificador del registro
- Etc.

Los registros de los CDR con localización por triangulación añaden tres campos más: latitud, longitud y error posicional estimado.

Los registros que crean las BTS cuando se abre una sesión de datos (Internet Access logs) tienen una estructura más simple que los CDR. Normalmente incluyen datos como los siguientes:

- Número de teléfono (encriptado)
- Fecha y hora de apertura de la sesión
- Página Web visitada
- Número de bytes transferidos

En el proceso de datos para identificar patrones de movilidad, el número de teléfono (anonimizado) permite seguir a un mismo usuario en el espacio y en el tiempo. El identificador de celda se asocia a otro fichero en el que se registra la localización de la torre (latitud y longitud) de cara a la geolocalización del usuario en el momento en el que hace la llamada. La fecha y hora de inicio de la llamada y la duración permiten conocer el intervalo temporal entre registros y estimar la

hora en que se realizan los viajes. Finalmente, los números de teléfono entrantes y salientes de los CDR permiten conocer la red social del usuario e identificar la posible convergencia espacio temporal de individuos de una misma red social (cuando ambos coinciden en el mismo tiempo en la misma área de cobertura) (ver tablas 1 y 2).

Tabla 1: Estructura básica de un CDR habitual en los estudios de movilidad (para un mismo usuario).

| ID_Usuario_Saliente | ID_Torre_Saliente | Día - Hora | Duración | ID_Usuario_Entrante | ID_Torre_Entrante |
|---------------------|-------------------|---------------------|----------|---------------------|-------------------|
| 52455562 | 12555 | 11/07/2012 15:36:05 | 38 | 55872462 | 2155 |
| 52455562 | 13476 | 12/07/2012 09:22:50 | 124 | 23547474 | 2741 |
| 52455562 | 578 | 12/07/2012 21:15:14 | 45 | 55872462 | 2155 |
| 52455562 | 578 | 12/07/2012 22:01:22 | 188 | 55872462 | 247 |
| 52455562 | 13476 | 13/07/2012 10:35:47 | 154 | 25241574 | 3325 |
| 52455562 | 1254 | 13/07/2012 16:21:12 | 132 | 55872462 | 2155 |
| 52455562 | 13476 | 14/07/2012 12:41:25 | 17 | 55872462 | 254 |
| 52455562 | 578 | 14/07/2012 21:00:32 | 35 | 92541869 | 1547 |

Tabla 2: Estructura tabla de localización de la torre.

| Torre_ID | Latitud | Longitud |
|----------|-----------|-----------|
| 12555 | 42,369639 | 12,332521 |
| 13476 | 42,378555 | 12,154822 |
| 578 | 43,005784 | 11,245878 |

Tabla 3: Estructura de un CRD con localización por triangulación.

| Señal_Usuario | Latitud | Longitud | Error | Día/Hora |
|--------------------------------|-----------|-----------|-------|---------------------|
| 3da524e5f6589dc94d555sfs54fdsf | 42,369639 | 12,332521 | 0,02 | 11/07/2012 15:36:05 |
| 3da524e5f6589dc94d555sfs54fdsf | 42,378555 | 12,154822 | 0,005 | 12/07/2012 09:22:50 |
| 3da524e5f6589dc94d555sfs54fdsf | 43,005784 | 11,245878 | 0,008 | 12/07/2012 21:15:14 |



4.3 Eventos pasivos (Network-Driven Cell Phone Data)

Se refiere a los datos generados a partir de la necesidad de que las operadoras conozcan dónde se ubican los dispositivos celulares de sus clientes para darles servicio (por ejemplo, para enrutar llamadas o entregar SMS, entre otros). Es lo que se conoce como monitorización pasiva. Normalmente estos datos no se almacenan, ya que las redes no suelen estar preparadas para ello. Solo son almacenados cuando la red cuenta con dispositivos específicos, lo que no es frecuente debido a su alto costo. Este tipo de datos no suele estar disponible para los análisis de movilidad.

Para este propósito, se generan eventos de red que actualizan la ubicación de los teléfonos. Los datos almacenados habitualmente son:

- Identificador del tipo de evento
- El número de teléfono
- El identificador de la torre (BTS) y conjunto de torres (LAC)
- La fecha y hora del evento
- Por lo general, hay varios tipos principales de eventos que la red captura:
- Cambio de LA, durante la realización de una llamada. Estos eventos reciben el nombre de handover (ver figura 9)
- Encendido y apagado del teléfono: en este caso, se registra la BTS a la que se conectó el teléfono por última vez
- Cambio en el tipo de tecnología que el teléfono está usando en el acceso a la red (por ej. de 3G a 4G).
- Solicitud de actualización de ubicación periódica (lo que técnicamente se conoce paging): cada cierto tiempo se emite una solicitud de ubicación que registra el identificador del BTS y el LAC al que está conectado el teléfono móvil. La frecuencia de esta actualización suele ser de 2 a 4 horas, es decir, si la red carece de información del teléfono durante las últimas 2-4 horas, se emite una solicitud de ubicación.

Además de los eventos anteriores, los registros de eventos pasivos también almacenan la actividad del teléfono móvil (llamadas y mensajes).

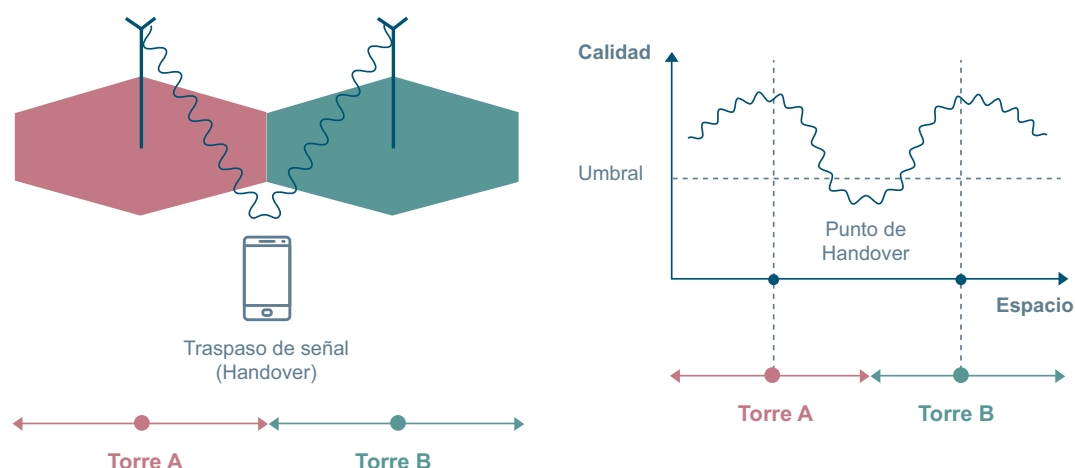


Figura 9: Handover: cambio de BTS y de LA durante la realización de una llamada en movimiento.



4.4 Características de los datos de telefonía móvil

Aunque los datos de telefonía móvil presentan diferencias según su tipo (activos o pasivos), operadoras, marco geográfico, etc., se pueden identificar las siguientes características principales de forma sintética:

Recogida pasiva de datos

Frente a las encuestas de movilidad, en las que la población tiene que hacer un esfuerzo por recordar los viajes que realizó el día anterior, los datos de telefonía móvil se recogen de forma pasiva, sin que el usuario tenga que hacer nada para que se almacenen: son las operadoras las que lo hacen automáticamente, tanto sobre los datos activos (CDRs) como los pasivos. Los registros de telefonía móvil recogen lo que la gente hace, no lo que dice que hace (encuestas). Esto lleva a un menor número de viajes omitidos, sobre todo viajes cortos y/o a pie.

Datos continuos, siempre actualizados

Los datos se registran de forma continua. Por ello es posible disponer de información siempre actualizada, de monitorizar procesos y de analizar los patrones de movilidad en situaciones atípicas, como son periodos de vacaciones, eventos turísticos o catástrofes.

Muestras de gran tamaño

Las muestras que se utilizan en estudios de telefonía móvil son enormes. Una operadora que cuente con una cuota de mercado del 30% tendría una muestra de casi un tercio de la población. Incluso después del proceso de selección de la muestra válida, las finales son mucho mayores que las que se obtienen en encuestas domiciliarias de movilidad.

Sesgo

Dado el tamaño de las muestras y la recogida pasiva de los datos, el nivel de sesgo es bajo siempre y cuando la operadora tenga una cuota de mercado alta en la ciudad a analizar y no exista un sesgo social o de otro tipo en sus clientes respecto de los de otras operadoras.

Resolución espacial media-alta

La localización del teléfono móvil se deriva de las coordenadas (latitud, longitud) de la torre a través de la cual ha establecido la conexión. La exactitud posicional de los registros de actividad es igual a la de los registros pasivos: se conoce que el dispositivo móvil se encuentra dentro del área de cobertura de una torre (aproximada mediante polígonos Voronoi), pero no dónde dentro de esa área de cobertura. La resolución espacial depende de la extensión de los polígonos Voronoi que a su vez está en función de la densidad de torres y oscila entre centenares de metros en áreas urbanas densas y varios kilómetros en áreas rurales. Por medio de técnicas de triangulación de señales se puede alcanzar una exactitud de unos 50-100 metros en áreas urbanas.

Alta granularidad temporal

Estudios empíricos sobre la granularidad temporal de los CDRs informan de un registro cada 5-8 horas, pero esa resolución es mucho mayor si se dispone también de sesiones de datos. En este caso se puede estimar en un registro cada 20 o 30 minutos para la muestra seleccionada. En los registros pasivos se pueden obtener promedios de 1-2 horas para el conjunto de los usuarios. La resolución temporal tiene una gran importancia en los estudios de movilidad. Si se dispone de datos activos que incluyan CDRs y sesiones de datos o registros pasivos se puede detectar un mayor número de lugares de actividad y de viajes y precisar mejor la hora en que se realiza cada viaje que si se dispone solo de CDRs.

Tiempo y costo empleado

Los plazos y el costo de los estudios de movilidad basados en datos de telefonía móvil son mucho menores que los correspondientes a estudios basados en encuestas domiciliarias de movilidad. Aunque depende de cada caso, el tiempo empleado puede ser 5 veces menor y el costo 10 veces menor. Por lo tanto, se puede disponer de estudios con un menor desfase temporal y sobre todo a un costo mucho más bajo, si bien el final dependerá de distintos factores, como el menor o mayor nivel de desagregación de las zonas objeto de estudio.

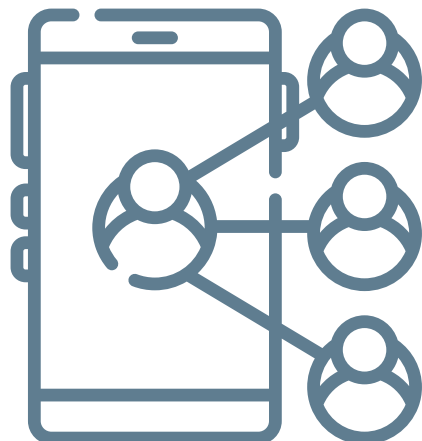
Información sobre las características sociodemográficas de los usuarios y de los viajes

Sin embargo, en los estudios con datos de telefonía móvil la información sobre las características de los usuarios y de los viajes es en general pobre. Pero esta información puede enriquecerse posteriormente tanto de forma individual (por ejemplo, mediante encuestas enviadas a través de mensajes SMS) como agregada (por ejemplo, utilizando cartografía con estratos sociales o usos del suelo), aunque difícilmente se puede llegar a alcanzar la cantidad de variables que suministra una encuesta domiciliaria.

A faint, light blue background pattern resembling a city street map or a network of roads, with various lines and intersections.

5

**Cálculo de matrices
de viajes origen-
destino (od) y de
indicadores de
movilidad a partir de
datos de telefonía
móvil**



El cálculo de matrices de viajes origen-destino al utilizar datos de telefonía móvil requiere seguir una serie de pasos, que se enumeran y describen a continuación (ver figura 10):

- 1. Limpieza y preparación de los datos
- 2. Selección de usuarios válidos
- 3. Identificación de estancias y secuencia de viajes
- 4. Localizaciones frecuentes y/o actividades: motivos de los desplazamientos
- 5. Enriquecimiento de los datos
- 6. Cambios en la agregación espacial de los datos: de polígonos Voronoi a zonas de transporte
- 7. Expansión de la muestra
- 8. Obtención y visualización de las estadísticas de movilidad
- 9. Validación de los resultados





Figura 10: Metodología para la obtención de matrices OD de viajes e información de movilidad a partir de datos de telefonía móvil.



1. Limpieza y preparación de los datos

Los datos procedentes del Big Data se generan habitualmente con un fin diferente del que le damos posteriormente en los estudios socio-territoriales o del transporte. Es el caso de los datos de telefonía con los que trabajamos en movilidad, cuyo objeto inicial es la facturación de la actividad realizada por los usuarios. Por eso los primeros pasos a realizar son tareas de pre-proceso y limpieza de errores con el fin de adaptar la base de datos de la mejor manera posible al objetivo final del estudio.

Limpieza de registros

No toda la información almacenada en un CDR o en los registros de un evento pasivo es necesaria en los estudios de movilidad.

Para evitar consumir recursos a la hora de almacenar las bases de datos o esfuerzos y tiempos en los procesos de análisis, es importante seleccionar solo aquellos campos de información necesarios para el objeto de estudio.

En el caso de la movilidad, la información básica con la que vamos a trabajar es:

- un identificador (anonimizado) del usuario que realiza el evento;
- un identificador de la BTS o el LA asignado o, en su caso, la localización del evento obtenida por triangulación;
- la referencia al momento temporal en el que se realiza el evento (fecha y hora del mismo).

A esta información básica pueden añadirse otros campos que tengan que ver con los datos del evento o con las características sociodemográficas del usuario. En relación con el evento, datos que pueden ser de interés son, por ejemplo, otras referencias a BTS o LA usados durante el evento en el caso de handover. Para éste también interesa la duración de la llamada, o el identificador anonimizado del usuario con el que se interacciona en el caso de llamadas o SMS. Mientras que entre las características del usuario, son útiles si están disponibles, el género y/o la edad para estudiar diferencias entre grupos.

Una vez definidos los campos de información temática que va a contener la base de datos a utilizar, debemos realizar un proceso de validación de los registros incluidos en ella. En el caso de las localizaciones, cuando trabajamos con datos triangulados debemos comprobar que las coordenadas asociadas tienen los rangos correctos, localizándose en todos los casos en nuestra área de estudio. Si trabajamos con datos georreferenciados a nivel de BTS, debemos asegurarnos que las relaciones entre los ID de cada BTS asignado al evento y los ID asignados a las BTS en la base de datos con sus coordenadas geográficas son las correctas. Con frecuencia aparecen eventos que tienen asignado un ID de BTS que no aparece en la base de datos con las localizaciones de cada BTS, normalmente porque esa BTS se ha sustituido o se ha reasignado, de manera que perdemos la referencia espacial de esos eventos. En estos casos, podemos reasignar los

eventos a la nueva BTS en el caso de conocer qué BTS está cubriendo el área de cobertura de la que ha sido eliminada, o simplemente eliminando esos eventos de la base final de datos que usaremos en el análisis.

Además de las localizaciones debe validarse el resto de la información a utilizar, asegurando por ejemplo que todos los registros tienen en sus campos de referencia temporal datos de fecha u hora en los rangos adecuados (por ejemplo, nunca una fecha posterior al momento final de recogida de datos o anteriores a una fecha dada), rangos adecuados en la edad de los usuarios, etc.

Problemas con la oscilación en la BTS asignada

Los datos georreferenciados a nivel de BTS presentan habitualmente errores de localización que tienen que ver con las asignaciones de los eventos a las BTS. Con frecuencia se producen saltos en las conexiones de los usuarios de una BTS a otra. Así, usuarios que no se han movido y se mantienen en una misma localización pueden ser asignados a diferentes torres en varios eventos sucesivos. Este fenómeno se conoce como oscilaciones o saltos y se produce en usuarios que se encuentran en un espacio cubierto por áreas de cobertura de dos o más BTS. Los operadores de telefonía tienen un sistema en sus redes que optimizan la asignación de las BTS según la demanda y el nivel de servicio en la red en cada uno de los momentos. De manera que, en el caso de encontrarse en un espacio cubierto por varias torres, la asignación a una torre puede variar de un momento a otro según la demanda de las mismas. Estas oscilaciones, en el caso de no ser eliminadas, introducen errores importantes en los estudios de movilidad al generar viajes inexistentes o introducir orígenes, puntos de paso o destinos nuevos en los desplazamientos (ver figura 11).

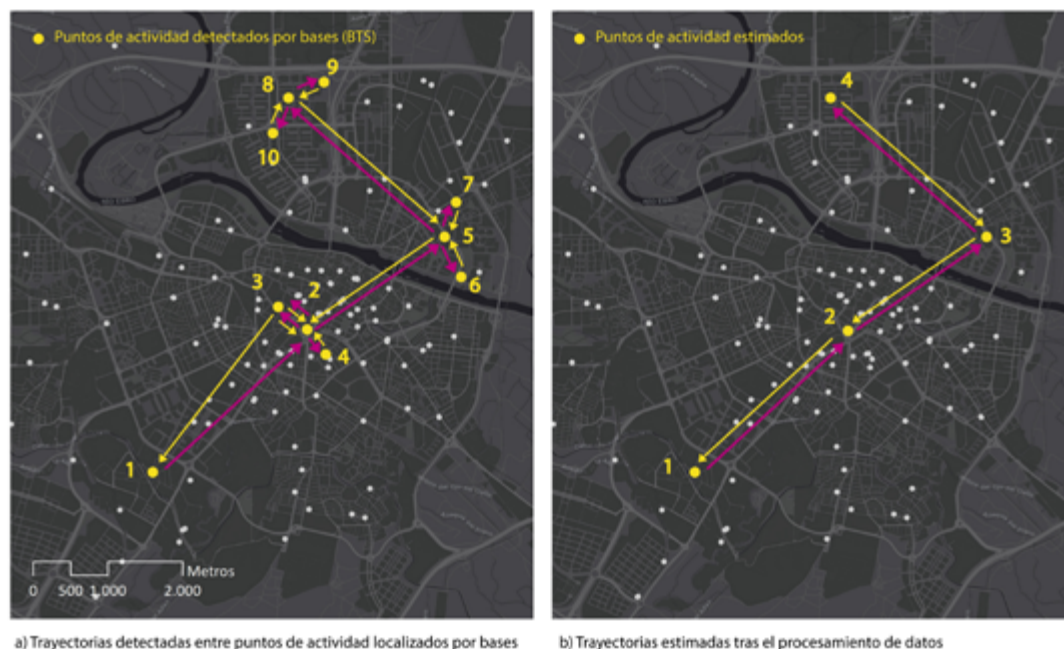
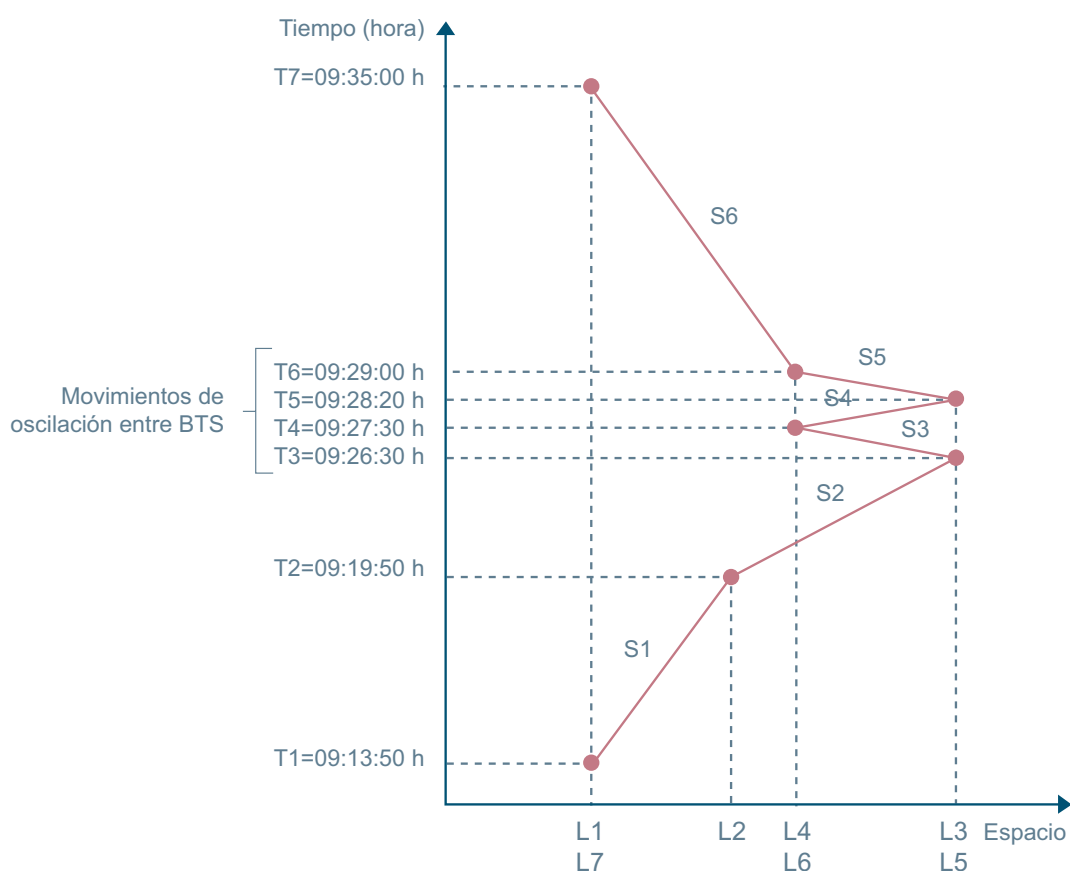


Figura 11: Trayectorias de un usuario de muestra (a) antes del procesamiento de datos y (b) después del procesamiento de datos.

Para identificar y posteriormente eliminar estas oscilaciones o saltos en la asignación de las BTS se suelen utilizar dos procedimientos:

- Como los saltos en la red suelen producirse en periodos cortos de tiempo, es posible identificar esos saltos cuando la velocidad a la que habría que desplazarse entre BTS es mayor a un cierto umbral (figura 12).
- También es posible identificar estos saltos cuando el número de cambios entre antenas es muy alto en cortos periodos de tiempo.

Una vez que se han identificado los saltos entre antenas se opta por tomar como BTS la que aparece como más frecuente, eliminando las referencias del resto de BTS asociadas a los saltos.



| Segmento | Hora Inicio | Hora Final | Tiempo (sec) | Distancia (m) | Velocidad (Km/h) |
|--------------|-------------|------------|--------------|---------------|------------------|
| L1 - L2 (S1) | 9:13:50 | 9:19:50 | 360 | 3,000 | 30 |
| L2 - L3 (S2) | 9:19:50 | 9:26:30 | 400 | 4,000 | 36 |
| L3 - L4 (S3) | 9:26:30 | 9:27:30 | 60 | 2,500 | 150 |
| L4 - L5 (S4) | 9:27:30 | 9:28:20 | 50 | 2,500 | 180 |
| L5 - L6 (S5) | 9:28:20 | 9:29:00 | 40 | 2,500 | 225 |
| L6 - L7 (S6) | 9:29:00 | 9:35:00 | 360 | 3,200 | 32 |

Figura 12: Ejemplo de efectos de la oscilación y método de identificación por velocidad en el cambio de BTS.

Errores de localización para datos georreferenciados por triangulación

Cuando los datos se georreferencian por triangulación es necesario corregir lo que se conoce como incertidumbre en la localización. Como vimos en el apartado 3.3, en la localización de los eventos por esta técnica aparece un error que varía en función del número de torres utilizadas y de la distancia a éstas. De esta manera, en las localizaciones frecuentes, los eventos aparecen en forma de nube de puntos que debemos asociar a una localización dada (ver figura 13)



Figura 13: Ejemplo de incertidumbre en los datos georreferenciados por triangulación.

Los métodos habituales para eliminar los errores asociados a la incertidumbre en la localización por triangulación trabajan con técnicas de clusterización. Se trata de asociar los eventos que están próximos en el espacio a una única localización. Para ello se fija un umbral mínimo de distancia, en algunos trabajos se ha usado por ejemplo los 300 metros, y las localizaciones observadas en ese menor umbral de distancia se ubican todas en el centroide de todas ellas.

Mejoras en la calidad de la localización en datos asociados a BTS

La estructura de las redes presenta, como se ha visto, superposiciones de las áreas de cobertura de las BTS, que complican la localización de cada uno de los eventos. La figura 14 muestra un ejemplo de la localización de las torres y la posible superposición de las áreas de cobertura de sus antenas.

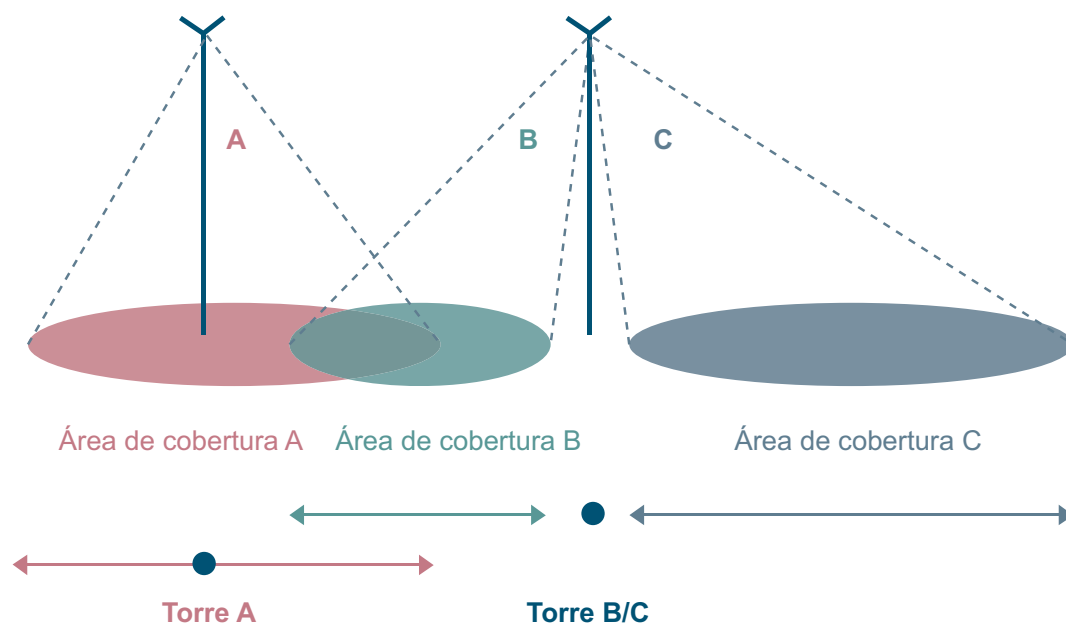


Figura 14: Ejemplo de localización de dos torres y superposición de las áreas de cobertura de sus antenas.

Como hemos visto, cuando los eventos están localizados por la BTS asignada, lo habitual es trabajar con polígonos de Thiessen o Voronoi y suponer que los eventos se localizan en el polígono asociado a esa BTS. Sin embargo, es posible utilizar otras aproximaciones diferentes a la hora de localizar los eventos:

- En algunos casos, el operador puede proporcionar un mapa con una cuadrícula que identifica para cada celda de esa cuadrícula cuál es la torre que cubre mejor esa localización. De esta manera, los eventos asignados a una torre serán aquellos que están en las cuadrículas mejor cubiertas por ella.
- De no existir esa cartografía es posible asignar al usuario la probabilidad de estar en un polígono Voronoi determinado, a partir de la torre a la que está asignado y de las distancias de las torres vecinas. La principal ventaja de esta solución es que toma en consideración el hecho de que varias torres podrían estar cubriendo la misma ubicación.
- Es posible desagregar la información de la localización dentro del polígono de Voronoi, mediante métodos de desagregación espacial (ver apartado 5.6).



2. Selección de usuarios válidos

Una vez seleccionados los campos de información con los que vamos a trabajar y limpiada la base de datos de posibles errores, debemos proceder a realizar una muestra de los usuarios que sean válidos para el análisis de movilidad. No todos son válidos para obtener información de movilidad de sus eventos.

Para realizar la selección de los usuarios válidos vamos a trabajar con la información sobre la resolución temporal de la base de datos inicial. Los valores promedio sobre la resolución temporal de los registros de actividad están muy influidos por los valores extremos. De hecho, los registros de algunos usuarios presentan una granularidad temporal muy inferior a la media y no son apropiados para derivar de ellos patrones de movilidad. Por eso una práctica habitual en los estudios de movilidad basados en registros de actividad, y particularmente en los basados solo en CDR de llamadas y SMS, es seleccionar aquellos usuarios que tienen una alta resolución temporal, de forma tal que sea posible conocer en detalle su posición en el espacio y tiempo. Esta selección supone una reducción del tamaño de la muestra final, pero las muestras iniciales son tan grandes cuando las operadoras tienen una cuota de mercado elevada que aun así resultan muestras finales de mucho mayor tamaño que las de las encuestas domiciliarias de movilidad. La selección de usuarios también puede introducir un cierto sesgo en la muestra final, por lo que es necesario emplear elementos de validación (comparar algunos resultados, por ejemplo, el número promedio de viajes o el porcentaje de personas no móviles, con los de una encuesta de movilidad anterior).

Existen varios métodos para seleccionar los usuarios con los que vamos a realizar los análisis y obtener las matrices de viajes finales:

- Como hemos señalado, lo habitual es utilizar una muestra que incluya a los usuarios con mayor resolución temporal. Pueden ser seleccionados los que tengan un número de registros por encima de un determinado umbral o quedándonos con un percentil dado (por ejemplo, eliminando el cuartil de usuarios con menor número de eventos).
- Algunos autores consideran también que usuarios excesivamente activos y móviles, que pueden asociarse a transportistas, repartidores o comerciales, pueden incorporar ruido en los análisis, de forma tal que optan por selecciones de usuarios que tengan valores medios en su frecuencia de uso.
- En otros casos se propone dividir el día en intervalos de tiempo, por ejemplo, cada 30 minutos, y descartar aquellos usuarios que no tengan registros en un número mínimo de esos intervalos. Se pueden tomar como válidos solo usuarios con registros en al menos 8 intervalos diferentes de tiempo de 30 minutos.
- En esa misma línea, otros trabajos proponen descartar usuarios con poca información asociada a los lugares de localización más frecuentes, que como veremos posteriormente permiten identificar el lugar de residencia (casa) y/o de trabajo. Por ejemplo, en el trabajo de Alexander (et al 2015) eliminan a los usuarios con menos de 8 estancias en el hogar, para una base de datos de un periodo de 2 meses, lo que supone tener al menos una estancia por semana basada en el hogar.



3. Identificación de estancias y secuencia de viajes

La tercera de las fases a la hora de obtener datos que alimenten una matriz de viajes a partir de telefonía móvil es conocer la secuencia de viajes, para lo que necesitamos identificar lo que se conoce como estancias y, a partir de ellas, los viajes entre éstas. Así, vamos a considerar:

- **Estancias:** como aquellos lugares en los que las personas permanecen un cierto tiempo, es decir los lugares donde se realizan actividades.
- **Viajes:** serán los desplazamientos entre dos estancias consecutivas.
- **Puntos de paso:** en aquellas localizaciones donde los usuarios han realizado algún evento, pero donde únicamente aparece un evento o el periodo entre eventos sucesivos es inferior a un determinado umbral de tiempo, se considera un punto de paso en el desplazamiento entre dos estancias.

De esta forma, para identificar una estancia se usa el tiempo que una persona permanece en una misma localización. La mayor parte de los trabajos identifican estancias a partir de la permanencia de más de 30 minutos en una misma localización, aunque según la resolución temporal de la base de datos en otros casos se ha usado un tiempo de 10 minutos (Jiang. et al. 2013, Çolak et. al 2015) o de más de 20 minutos (Holleczek et al. 2014).



Figura 15: Definición de estancias y puntos de paso a partir de registros telefónicos. En este ejemplo, para un usuario anónimo y ficticio a partir de los registros de su actividad telefónica se definen 7 ubicaciones de estancias distintas (triángulos rojos) en un determinado período de tiempo. El resto de los puntos se denominan puntos de paso, ya que en los eventos realizados en esas localizaciones no se han observado estancias superiores a 30 minutos.

Una vez identificadas las estancias y los puntos de paso de cada usuario, es posible definir los viajes realizados entre las estancias. En el caso de los viajes existe la incertidumbre de establecer tanto la hora de inicio como la de final del viaje. Esta incertidumbre dependerá nuevamente de la resolución temporal de la base de datos con la que estemos trabajando. Así, los eventos nos proporcionan la última hora detectada en la estancia de origen del viaje y la primera detectada en la estancia de destino, lo que determina la mínima hora de inicio del viaje y la máxima hora de llegada respectivamente. Con estos datos, existen varias opciones para establecer las horas de inicio y final del viaje:

- A partir del momento del último evento en origen y el primero en destino, algunos trabajos utilizan una función de probabilidad, basada en estadísticas oficiales sobre distribución horaria de viajes para establecer el momento temporal del viaje (Alexander et al. 2015).
- En otros casos se propone realizar estimaciones de tiempos de viaje entre esas dos estancias, para reducir así el intervalo de incertidumbre y después asignar como hora del viaje la intermedia dentro del nuevo intervalo de horas posibles (Widhalm et al. 2015). En estos casos, los tiempos de viaje entre las dos estancias suelen calcularse como una aproximación obtenida a partir de la distancia en línea recta entre las dos estancias y una velocidad media de desplazamiento entre ellas. Esta velocidad media puede obtenerse a partir de otras fuentes, ya sean encuestas domiciliarias previas o, para distintos modos de transporte, en función de mapas de velocidades medias de circulación medidos a través de vehículo flotante, nuevas fuentes de datos como redes de navegadores, ficheros GTFS de transporte público o aplicaciones como Google Transit.

Se trata de tener un tiempo de viaje estimado entre las estancias con el que aproximar las horas de salida y llegada del viaje.

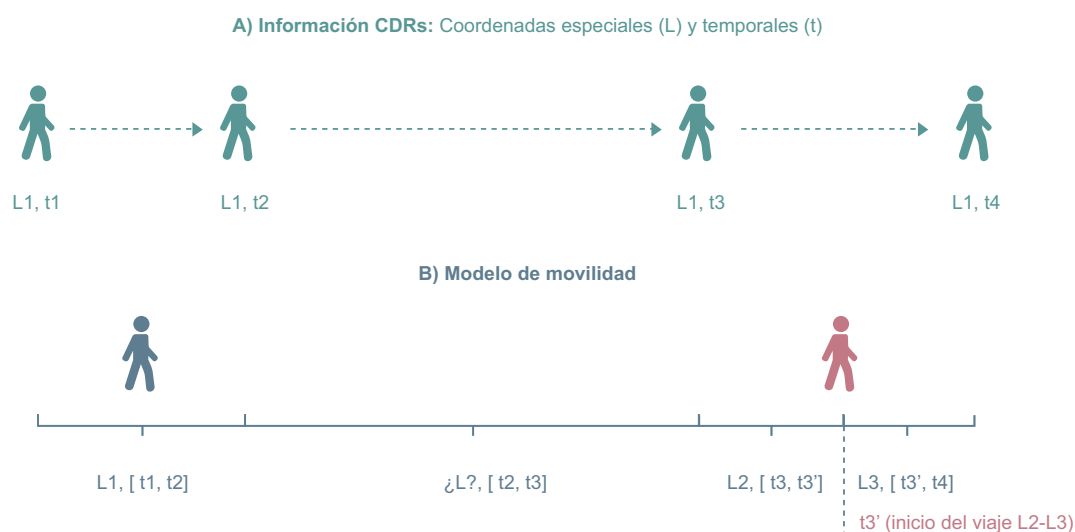


Figura 16: Estimación de la movilidad a partir de los datos de CDRs. Los dos primeros CDRs (L1, t1 y L1,t2) tienen la misma localización. Si t1 y t2 superan el umbral de tiempo establecidos se define una estancia en L1. El CDR(L2, t3) se realiza en una localización diferente (L2), al igual que el siguiente CDR (L3, t4). La incertidumbre surge a la hora de determinar las salidas y llegadas entre cambios de localización. Fuente: Basado en Picornell et al. (2015).

La resolución espacial de los datos es fundamental a la hora de definir la localización de las estancias y los viajes. En la mayor parte de los trabajos realizados se utilizan datos procedentes de eventos activos localizados a nivel de BTS. Sin embargo, es posible trabajar también con eventos pasivos localizados a nivel de LA o con datos georreferenciados por triangulación.

En el caso de trabajar con datos de eventos pasivos a nivel de LA se incrementa la resolución temporal, pero se reduce la resolución espacial. De esta manera, podemos conocer el LA en los que se localizan los usuarios a lo largo del día, pero es muy difícil determinar los momentos en los que se producen los viajes, dado que únicamente conocemos el momento temporal en el que se ha cambiado de LA. La figura 17 muestra las diferencias entre trabajar con localizaciones a nivel de BTS o de LA.

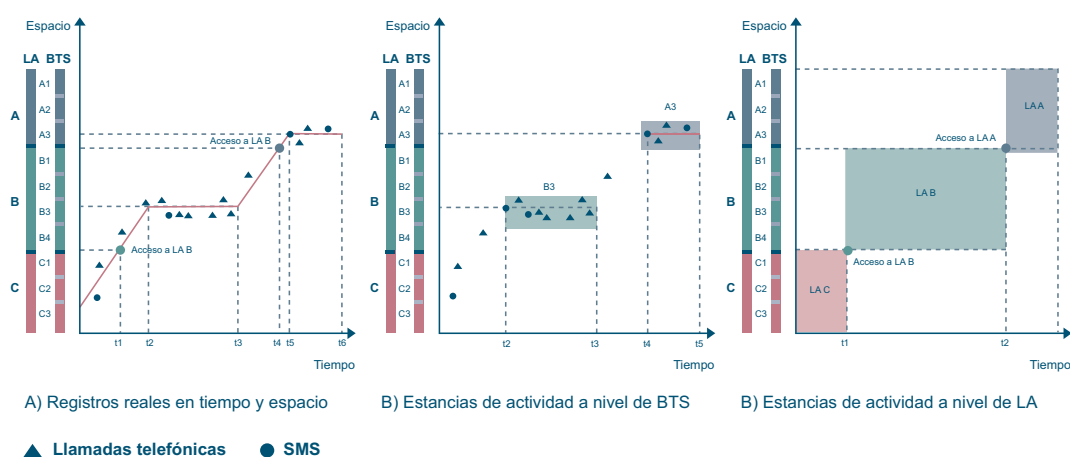


Figura 17: Definición de estancias si se trabaja con datos georreferenciados a nivel de BTS o de LA.

Al contrario de la información georreferenciada a nivel de LA, cuando trabajamos con datos georreferenciados por triangulación la localización de la estancia es mucho más fina. En este caso, para la definición de las estancias se utilizan métodos de clusterización, agrupando los eventos localizados a una determinada distancia a partir de su centroide. Una vez agrupados se utiliza un umbral temporal, similar a lo usado anteriormente (10, 20, 30 minutos), para definir estancias donde aparecen eventos durante periodos de tiempo superiores a dicho umbral. En el caso del viaje puede usarse una metodología similar a la anterior, determinando el viaje como el desplazamiento entre dos estancias sucesivas, o utilizando nuevamente métodos de clusterización para identificar eventos que tengan una secuencia lineal (ver la figura 19 en el apartado siguiente).

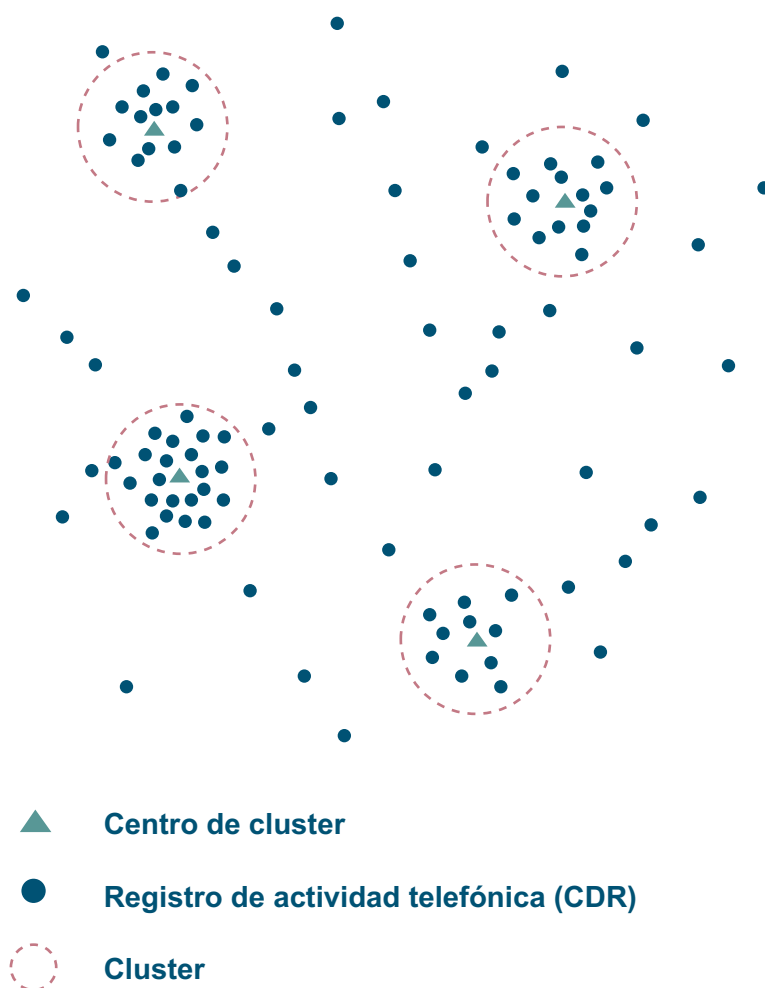


Figura 18: Definición de estancias mediante procesos de clusterización para datos georreferenciados por triangulación.



4. Localizaciones frecuentes y/o actividades: motivos de los desplazamientos

Una vez definidas, las estancias pueden ser clasificadas según el tipo de actividad que se realiza en la mismas, lo que permite aproximarnos al análisis de los motivos de los desplazamientos. Para conocer el tipo de actividad, lo más habitual y sencillo es tratar de identificar inicialmente el lugar de residencia y trabajo de los usuarios de la muestra. Esta metodología permite obtener información de la movilidad recurrente de la población.

La forma de localizar tanto el lugar de residencia como de trabajo es identificar los lugares donde se realizan eventos con una mayor frecuencia y considerar como el lugar de residencia la localización más visitada (donde se han realizado más eventos) en el horario de tarde-noche y como lugar de trabajo la localización más visitada en el horario por la mañana. Los valores horarios asignados a residencia

y trabajo varían según países y zonas, lo que debe adaptarse a los patrones generales de comportamiento de la población de cada área de estudio.

La tabla 1 muestra algunos ejemplos de horarios utilizados para establecer el lugar de residencia y de trabajo:

| Autores | Horario de residencia | Horario de trabajo |
|-------------------------|-----------------------|--------------------|
| Calabrese et. al (2013) | 18:00 – 08:00 | - |
| Lenormand et al. (2014) | 20:00 – 07:00 | 09:00 – 17:00 |
| Çolak et. al (2015) | 19:00 – 08:00 | 09:00 – 18:00 |
| Picornell et al (2015) | 20:00 – 07:00 | 08:00 – 17:00 |

Para asegurar una mejor definición tanto del lugar de residencia como de trabajo se suele utilizar una base de datos que incluya solo los días laborables, normalmente excluyendo los viernes, al tratarse de un día bastante atípico. No obstante, algunos autores han trabajado en el caso de la residencia también incluyeron los fines de semana, mientras que para el caso del lugar de trabajo siempre se utilizan únicamente días laborables.

Utilizar las localizaciones más frecuentes y los horarios para identificar el hogar y el lugar de trabajo es el método más usado, pero hay otras opciones que también se utilizan:

- Algunos autores incluyen adicionalmente la distancia entre los lugares más frecuentes en el horario de mañana, para los casos en los que aparecen varias localizaciones frecuentes y la residencia para determinar el lugar de trabajo (Alexander et al. 2015). Así, definen el lugar de trabajo como la localización más visitada dentro del periodo temporal de mañana que se encuentra a mayor distancia del lugar de residencia.
- Otros modelos incorporan en el método anterior un modelo de regresión para obtener la probabilidad de esa localización. Estos modelos se usan cuando la resolución temporal de la base de datos es baja (por ejemplo, si solo se trabaja con eventos de llamadas y SMS georreferenciados a nivel de BTS). Usan como variables el número de días con actividad, la duración de la actividad en el área de estudio (para eliminar posibles turistas o no residentes), el número de torres y días con actividad, los eventos en horario de trabajo y los eventos en horario de residencia (Isaacman y otros 2011).

Una vez definidos el lugar de residencia y de trabajo se procede a identificar otras localizaciones visitadas y sus posibles motivos. Picornell et al (2015) trabajando con CDR localizados a nivel de polígono Voronoi diferencian, junto a la residencia y el trabajo otras actividades que denominan “otras actividades frecuentes” y a las restantes las clasifican como “actividades no frecuentes”. De esta manera tendríamos cuatro tipos de actividades:

- Residencia
- Trabajo
- Otras localizaciones frecuentes. Serían estancias que se repiten con una cierta frecuencia, por ejemplo, actividades relacionadas con motivos

deportivos (como ir al gimnasio 3 días a la semana), actividades de ocio (como ir al cine habitual) o compras (ir a un centro comercial determinados días).

- Actividades no frecuentes. Serían estancias que no tienden a repetirse.

Para trabajar con estas clasificaciones de las estancias, el primer paso es definir como actividades frecuentes aquellas estancias en la que el usuario aparece un mínimo de días respecto del total de días analizados. Para ello, es posible trabajar respecto del total de días de un mismo tipo (por ejemplo, martes), del total de días laborables (considerando días laborables los días de lunes a jueves) o del total de fines de semana. Como ya hemos señalado, el viernes suele ser excluido al tratarse de un día de transición al fin de semana. En cualquier caso, es posible realizar otras clasificaciones de días laborables y festivos según la idiosincrasia del área de estudio.

El número mínimo de días (frecuencia mínima) para considerar una actividad como frecuente viene determinado por la siguiente expresión:

$$\text{Frecuencia_mínima} = \alpha * \text{"muestra de días analizados"}$$

donde 'α' es un coeficiente de reducción y la "muestra de días analizados" es el total de días de un mismo tipo presentes en la muestra (días específicos de la semana, días laborables y/o días festivos).

El coeficiente de reducción a utilizar deberá ajustarse a la granularidad temporal de los datos. Algunos valores de 'α' considerados en varios trabajos son de 0.2 para la residencia y 0.3 para el trabajo, considerando así las menores horas efectivas en el caso del lugar de trabajo. Otros valores de referencia para el conjunto de lugares frecuentes se mueven en una horquilla de 0,15 a 0,35.

Cuando los datos están georreferenciados por triangulación es posible utilizar la metodología anterior a partir de la frecuencia de los eventos y las distribuciones



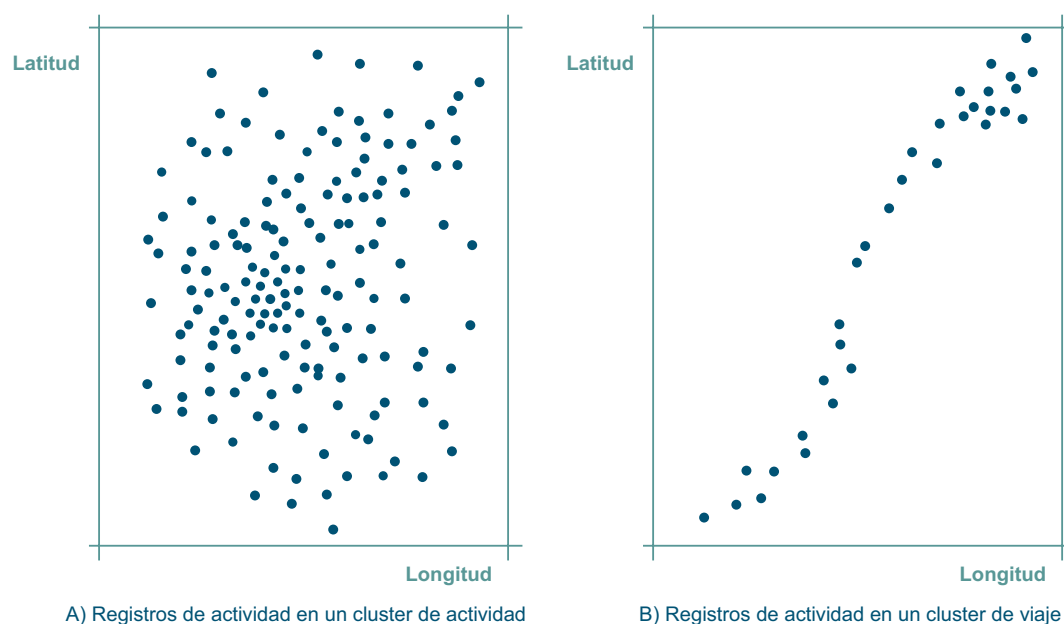


Figura 19: Definición de actividades mediante procesos de clusterización para datos georreferenciados por triangulación.

horarias, pero utilizando métodos de clusterización para la agrupación de los datos. Pero además, en el caso de los desplazamientos, estos pueden identificarse a partir de la huella del cluster, clasificando como viaje a aquellos que tienen una secuencia espacial de carácter lineal (ver figura 19).

Como se mostrará posteriormente, una de las fases en las que se pueden aprovechar los procesos de enriquecimiento de datos es a la hora de afinar en la definición del tipo de actividad, por ejemplo cruzando los datos de las localizaciones de las estancias identificadas con capas de usos del suelo para poder inferir así el motivo del viaje.



5. Enriquecimiento de datos

Se ha visto en anteriores apartados cómo es posible mejorar nuestro conocimiento sobre individuos y viajes analizando los patrones espacio-temporales de los usuarios. Dado que muchos teléfonos móviles no tienen contrato y en los que sí lo tienen el lugar de residencia del usuario que figura en el mismo no es un dato fiable, (teléfonos móviles de empresas, varios teléfonos de una familia registrados a un mismo nombre, etc.) el domicilio del usuario se infiere a partir de los patrones espacio-temporales derivados del uso del teléfono, al igual que el lugar de trabajo (ver apartado 4).

Pero además, los datos sobre movilidad extraídos de los registros de telefonía móvil pueden enriquecerse utilizando fuentes externas. Es lo que se conoce como enriquecimiento de datos. De hecho, la geolocalización de los datos permite cruzarlos con otros de distintas fuentes cartográficas utilizando un SIG, lo que permite añadir información de gran valor para los estudios de movilidad en cuestiones como las siguientes:

Características de los viajeros. Se ha comentado anteriormente (subapartado 4.1) que los datos que disponen las operadoras sobre sus clientes a partir de los contratos son limitados y poco fiables. En los móviles con contrato, según el plan que tenga el usuario se infiere el estrato socio-económico y se permite conocer género y edad, por ejemplo. Pero muchos usuarios utilizan dispositivos móviles sin contrato. Sin embargo, en estos casos es posible inferir las características de los usuarios a partir de las secciones censales o zonas de transporte en las que viven.

Así, por ejemplo, al conocer la localización del domicilio del usuario del teléfono móvil a partir de sus patrones espacio-temporales, se pueden cruzar esos datos de domicilios con otros sobre estratos sociales.

Si la zona de transporte en que se sitúa un domicilio tiene un bajo nivel medio de ingresos, se debe inferir que ese usuario pertenecerá al estrato de ingresos bajo. Esto es especialmente válido en ciudades muy segregadas socialmente, como la mayoría de las ciudades latinoamericanas. Variables propiamente demográficas, como el género o la edad, no pueden ser inferidas por este procedimiento.

Motivo de viaje. A partir de las características de la zona de destino del viaje (usos del suelo) y de los patrones temporales del viaje se puede llegar a inferir el motivo para aquellos viajes que sobre la base del análisis espacio-temporal fueron calificados como no recurrentes (subapartado 5.4). Así, un viaje con una estancia de corta duración en un centro comercial puede interpretarse como un viaje por motivo de compras. Es evidente que este procedimiento es especialmente útil en zonas de transporte altamente especializadas (campus universitarios, complejos hospitalarios y grandes superficies comerciales, entre otros.), pero más difícil de aplicar en áreas de la ciudad con gran mezcla de usos del suelo. Para superar esta dificultad se puede utilizar un enfoque probabilístico basado en las características de los lugares de atracción (capacidad de atracción de viajes, horarios de apertura). Por otro lado, la existencia de antenas indoor y redes WIFI o sensores de Bluetooth en grandes equipamientos o infraestructuras (como centros comerciales) también es de gran utilidad para inferir el motivo del viaje, al aportar información con mayor exactitud posicional sobre la presencia de un usuario en un determinado lugar.

Estimación de la ruta. Para estimar la ruta se parte de la localización de los puntos de origen y destino del viaje y de los intermedios. Estos puntos se superponen sobre capas de redes de transporte (transporte privado y transporte público) para determinar la ruta más probable usando modelos de asignación. Cuando los datos de localizaciones intermedias son suficientes y las áreas de cobertura de las torres son pequeñas se pueden usar algoritmos que obtengan la ruta minimizando las distancias de cada una de las rutas más probables a todos los centroides de celda.

Identificación del modo.

La identificación del modo de transporte es uno de los grandes retos en el estudio de la movilidad urbana a partir de datos de telefonía móvil.

En viajes interurbanos la diferencia de tiempos de viaje entre modos (avión, ferrocarril, carretera) es un buen indicador, pero en los viajes urbanos las diferencias de tiempos entre modos no resultan particularmente indicativas, ya que

la congestión prácticamente puede igualar las velocidades en transporte público y privado. Sin embargo, a través de las rutas seguidas en ocasiones es posible conocer el modo de transporte. Esto ocurre, por ejemplo, cuando en la periferia de las ciudades el ferrocarril y la carretera no corren paralelos. Asimismo, es posible identificar viajes en transporte público a partir de las antenas indoor y redes WIFI o sensores de Bluetooth de algunas estaciones de ferrocarril. Así, por ejemplo, en el estudio de Telefónica sobre el Ferrocarril Línea Roca (Ramal Constitución-La Plata) & Línea C de Subterráneo de Buenos Aires (ver estudio de caso) se consideraron las antenas que cubren las estaciones de la línea de ferrocarril y en el caso del metro las microantenas que dan servicio exclusivo dentro de cada estación.

Algunas operadoras realizan encuestas en sus estudios de movilidad a través de mensajes SMS para solicitar a sus clientes información sobre sus características sociodemográficas y sus viajes. Así, por ejemplo, en el estudio de movilidad de Neuquén (ver estudios de caso) Movistar+ realizó una encuesta a sus clientes mediante su servicio de Encuestas por SMS, donde preguntó sobre las siguientes cuestiones:

- Consulta sobre el nivel educativo de la población para caracterizar la matriz de viajes.
- Consulta sobre el motivo del primer desplazamiento realizado por las personas en el día de la encuesta.
- Consulta el modo de transporte utilizado en el primer desplazamiento realizado en el día de la encuesta.
- Consulta sobre tiempo de viaje en el primer desplazamiento realizado en el día de la encuesta.

El cruce de los datos de la encuesta con los de movilidad de cada encuestado obtenidos a partir de los registros de teléfonos móviles aporta una valiosa información para conocer mejor las características de los individuos y los viajes. Sin embargo, este procedimiento puede introducir un cierto sesgo, ya que determinados grupos sociales pueden ser más proclives que otros a contestar la encuesta.





6. Cambios en la agregación espacial de los datos: de polígonos Voronoi a zonas de transporte

Dado que la localización de los teléfonos móviles normalmente se hace por áreas de cobertura (polígonos Voronoi) a partir de una relación entre los datos de eventos y los datos de red basada en el identificador de la torre (ver capítulo 4), las matrices de viajes se calculan inicialmente para las áreas de cobertura. Sin embargo, esta geolocalización de los datos no es útil en la planificación del transporte, que requiere datos referidos a zonas de transporte y no a áreas de cobertura. El problema estriba en que ambas zonificaciones (áreas de cobertura de las torres y zonas de transporte) no coinciden.

En este apartado se describen varios procedimientos para hacer este cambio de unidades espaciales. Este paso no sería necesario en el caso que la operadora proporcione datos triangulados, con coordenadas x y, que por lo tanto pueden ser directamente agregados por zonas de transporte. Una vez que los datos de la muestra están referidos a zonas de transporte es posible elevarlos a partir de datos oficiales de población residente o por otros procedimientos (ver apartado 5.7).

Los pasos que hay que dar para convertir la información de polígonos de Voronoi a zonas de transporte son básicamente los siguientes:

- **1-** Se debe disponer de dos capas de información en un Sistema de Información Geográfica (SIG): una capa de polígonos Voronoi y otra de zonas de transporte. Una capa de información en el SIG consta de un mapa digital (información cartográfica) y una base de datos asociada (información temática). El mapa digital registra la localización de los polígonos y la tabla asociada registra las características de esos polígonos. Así, en la tabla de la capa de polígonos Voronoi se almacenan los datos de telefonía móvil tratados anteriormente y en la de zonas de transporte las características de las mismas (población, usos del suelo, tipología de la edificación, etc.).
- **2-** Las dos capas deben ser superpuestas en un SIG. La capa resultante de la superposición contiene la información cartográfica y temática de las dos capas originales. Así, cada polígono Voronoi puede quedar partido por varias zonas de transporte y sus datos deben ser desagregados (repartidos) entre los polígonos en que ha sido partido como consecuencia de la superposición.
- **3-** Posteriormente los polígonos pertenecientes a la misma zona de transporte deben ser agregados para construir una nueva capa de zonas de transporte que contenga al mismo tiempo información de las características de las zonas de transporte y de los viajes.

El reparto de los datos de los polígonos Voronoi al que se alude en el punto 2 se puede realizar por varios procedimientos.

Reparto proporcional entre zonas de transporte (areal weighting).

Los datos de telefonía móvil asociados a cada polígono Voronoi se reparten entre las diferentes zonas de transporte que lo cubren de forma proporcional a la superficie que ocupa cada una de ellas en el correspondiente polígono Voronoi. Así, por ejemplo, como se muestra en la Figura 20, si tenemos una relación entre dos polígonos Voronoi (A y B) de 100 viajes. El polígono de Voronoi de origen (A)

queda cubierto en un 20% por la zona de transporte A1, en un 50% por la A2 y un 30% por la A3. De esta manera, el 20% de los viajes con origen en ese polígono serán asignados a la zona de transporte A1 (en este caso 20 viajes), el 50% a la A2 (50 viajes) y el 30% a la A3 (30 viajes). Esos viajes se tienen que repartir entre las zonas de transporte que cubren el polígono Voronoi de destino (B). Para ello se realiza un proceso similar. Así, en el caso del ejemplo, el Voronoi de destino queda cubierto en un 40% por la zona de transporte B1 y un 60% por la B2. De esta manera, los 20 viajes con origen en la zona de transporte A1 se repartirán en un 40% con destino en la zona de transporte B1 (esto es, 8 viajes) y en un 60% con destino en la zona de transporte B2 (12 viajes). Y así sucesivamente. Este procedimiento es el más simple, pero da resultados poco satisfactorios en los bordes de las ciudades y en las municipalidades vecinas en las que la población se concentra solo en una parte de las mismas.

Estimación de viajes a nivel de zona de transporte (A1, A2... Xn) a partir de los viajes obtenidos a nivel de áreas de cobertura (A y B)

$$\text{Viajes (A1 - B1)} = A1/A \cdot V_{AB} \cdot B1/B$$

| Origen | Destino | Nº viajes |
|--------|---------|-----------|
| A | B | 100 |
| A1 | B1 | 8 |
| A2 | B1 | 20 |
| A3 | B1 | 12 |
| A1 | B2 | 12 |
| A2 | B2 | 30 |
| A3 | B2 | 18 |

* Superficies de zona de transporte en % sobre superficies de área de cobertura
(Superficie Zona transporte A1 = 20% Área de cobertura A)

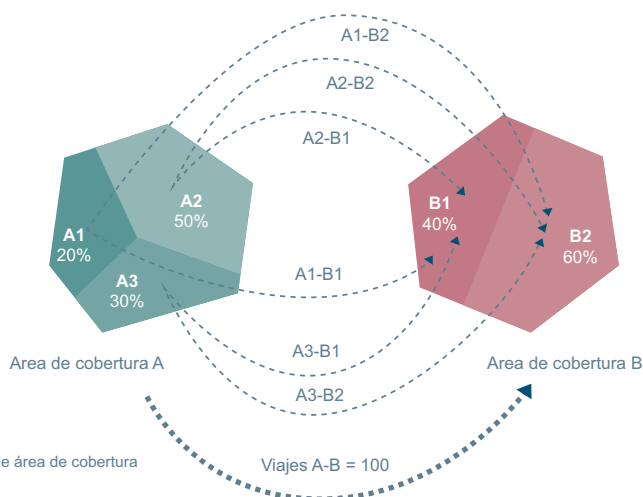


Figura 20: Superposición de áreas de cobertura y zonas de transporte.

Reparto proporcional entre zonas de transporte (are al weighting)

apoyado en análisis dasimétrico. Este método es semejante al anterior, solo que previamente se utiliza una cartografía de usos del suelo para eliminar de los polígonos Voronoi las áreas no cubiertas por el suelo urbano, excluyendo así las áreas sin desarrollar, no frecuentadas por la población. Este proceso se realiza de forma automática utilizando un SIG. El término de análisis dasimétrico alude a la utilización de una capa auxiliar (en este caso, una capa de usos del suelo) para mejorar la localización espacial de una variable (en este caso, los datos de telefonía móvil).

Reparto proporcional utilizando un indicador de actividad referido a las zonas de transporte.

Los métodos anteriores no consideran que la distribución espacial de la población de una ciudad cambia de localización a lo largo del día y por lo tanto también cambia la distribución de los viajes. Para superar esta debilidad se puede considerar algún indicador de actividad de las zonas de transporte según franjas horarias para repartir los datos de los viajes. Así, por ejemplo, la transferencia de datos de viajes desde los polígonos Voronoi hasta las zonas de transporte puede hacerse considerando la actividad de las

redes wifi en cada zona de transporte como proxy de la actividad que registra esa zona a esa hora del día. En otros casos se han utilizado encuestas de uso del tiempo combinadas con capas de usos del suelo para inferir el nivel de actividad de las distintas zonas según franjas horarias.



7. Expansión de la muestra

El objetivo de esta fase es expandir la información obtenida a partir de los eventos de telefonía móvil con el fin de obtener después estadísticas de movilidad que representen al total de la población en el área de estudio. No debemos olvidar que con los datos de telefonía estamos recogiendo una proporción (muestra) de la población total, cuyo volumen dependerá de la cuota de mercado de la operadora que proporciona los datos y de la selección de usuarios válidos que se realiza al inicio del proceso de cálculo de matrices de viajes. En relación con la expansión de la muestra es importante considerar, por lo tanto:

- Si la cuota de mercado de la operadora con la que estamos trabajando está segmentada socio-demográficamente en nuestra área de estudio. Así, existen operadoras en determinadas regiones que tienen un mercado enfocado hacia un cierto nicho de población, por ejemplo, los jóvenes y los grupos de rentas bajas o altas, entre otros. En estos casos, tendremos que considerar este hecho a la hora de realizar la expansión, con diferentes factores de expansión para esos grupos y valorando si la muestra de datos es representativa en el caso de los grupos sociodemográficos en los que la operadora no se está centrando.
- Si el mercado está segregado espacialmente. Así, puede ocurrirnos que la cuota de penetración de la operadora con la que trabajamos presente distribuciones territoriales diferentes, de manera tal que su penetración sea elevada en unos espacios y baja en otros. Este hecho está directamente relacionado con el punto anterior, debido a que determinados grupos sociales tienden a concentrarse más en unas zonas que en otras. Además puede haber razones técnicas: es posible que aparezcan pequeñas áreas de la ciudad que, por una peor cobertura o peor nivel de servicio, tengan una cuota de penetración menor de la operadora en cuestión. En estudios regionales o estatales puede ocurrir que la penetración de la operadora sea diferente entre unas áreas y otras, lo que deberemos considerar a la hora de realizar la expansión de los datos.



Existen varios procedimientos para la expansión de los datos. El más habitual es realizar la expansión a partir de la información obtenida del número de residentes por cada unidad espacial de referencia y utilizando datos oficiales de distribución de la población, por ejemplo, datos censales. En este caso, los pasos a seguir para la expansión son:

- Calcular el factor de expansión para cada zona de referencia como la proporción entre el número de residentes identificados a partir de los datos de telefonía en esa zona y su población total según los datos oficiales de referencia (censos, etc.).
- Realizar un análisis de los factores de expansión obtenidos, que nos permita identificar la calidad de los datos finales y los espacios donde éstos puedan presentar problemas de representación. Algunos autores establecen un factor de expansión de 0 en las zonas donde los datos de telefonía han recogido un número bajo de residentes (por ejemplo, menos de 10 residentes) o donde los factores de expansión superan un determinado umbral (por ejemplo, un factor de expansión de más de 100). De esta manera se evitan sobrerrepresentaciones de usuarios que no son representativos de una determinada zona.

En otros casos la expansión de los datos obtenidos puede realizarse considerando:

- Para matrices de viajes recurrentes (residencia-trabajo), si tenemos datos oficiales tanto de residentes como de trabajadores, podemos realizar una expansión que considere tanto la distribución de la población residente en las zonas de origen de los viajes como la distribución del empleo en las zonas de destino. Se trata de expansiones denominadas de doble constricción.
- En otros casos, se han usado como datos para la expansión información de Intensidades Medias Diarias de tráfico (IMD). Así, por ejemplo Iqbal et. al. (2014) generaron una matriz de viajes origen destino a nivel de torre, que asignaron después a una matriz de nodos en la red y la expandieron con los datos de IMD de 14 puntos de aforo obtenidos por cámaras de video.



8. Obtención y visualización de las estadísticas de movilidad

Una vez agregados los datos a las unidades espaciales objeto de estudio y expandidas las matrices de viajes es posible obtener estadísticas que permitan describir las características de la movilidad de la población en el área de estudio. Los principales indicadores de movilidad que pueden obtenerse a partir de las matrices con los datos de telefonía son:

Número de viajes con origen en cada zona y sus destinos. Para cada una de las unidades espaciales de referencia podremos conocer el número total de viajes que tienen origen en la misma (normalmente trabajando con viajes diarios). Esos datos pueden ser desagregados según los destinos de los viajes, de manera tal que podemos conocer la distribución según destinos para cada unidad espacial.

Viajes con destino en cada zona y sus orígenes. Al igual que con los viajes generados podemos conocer cuántos viajes son atraídos por cada unidad

espacial y cuál es la distribución según zonas de origen. Estos datos pueden visualizarse a través de tablas de doble entrada (matrices OD) donde los orígenes se representan en las filas y los destinos se visualizan en las columnas. La diagonal de la matriz representa los viajes internos en cada una de las unidades espaciales. Para su representación gráfica suele usarse tanto mapas de coropletas que representan la cantidad de viajes con origen en una zona según sus destinos (o viceversa) como los mapas denominados de arañas o “líneas de deseo”, donde cada par de relaciones OD se representa con una línea, cuyo grosor varía según la cantidad de viajes.

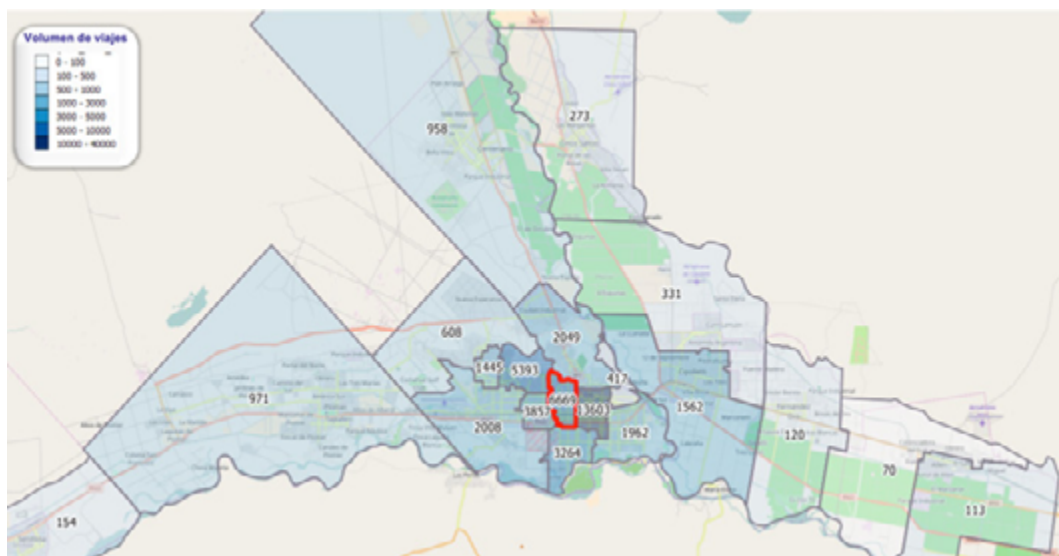


Figura 21: Ejemplo de mapa de coropletas para representar el destino de los viajes generados en una zona en concreto (en este caso la zona resaltada en rojo). Tomado del trabajo realizado por Telefónica para el Banco Iberoamericano de Desarrollo en Neuquén.

Número de viajes por persona. Para cada unidad espacial podemos obtener el total de viajes diarios por persona, simplemente dividiendo el total de viajes por los residentes en esa unidad espacial.

Distancias y tiempos medios de viajes: Los datos de los viajes individuales pueden ser agregados según unidades espaciales a partir de las distancias y tiempos en sus viajes, de manera que podemos obtener las estadísticas descriptivas de distancias y tiempos de viajes agregados para cada zona de origen y destino.

Viajes según motivos-actividades: A partir de la clasificación de las estancias según tipos de actividades, los resultados pueden diferenciarse según se trate de viajes “basados en casa”, cuando una de las puntas (ya sea el origen o el destino) es el lugar de residencia, o “viajes no basados en casa”, cuando ninguna de las puntas es el lugar de residencia. Si la información ha sido enriquecida (ver apartado 5.5) pueden diferenciarse otro tipo de motivos de viajes.

Distribución temporal de los viajes realizados: Los datos de matrices de viajes pueden ser desagregados según franjas horarias, lo que nos permite conocer cómo es la distribución de los viajes en el área de estudio en cada momento del día, conociendo cuáles son los principales orígenes (ejemplo en la Figura 22-izquierda) y destinos en cada franja horaria (ejemplo en la figura 22-derecha)

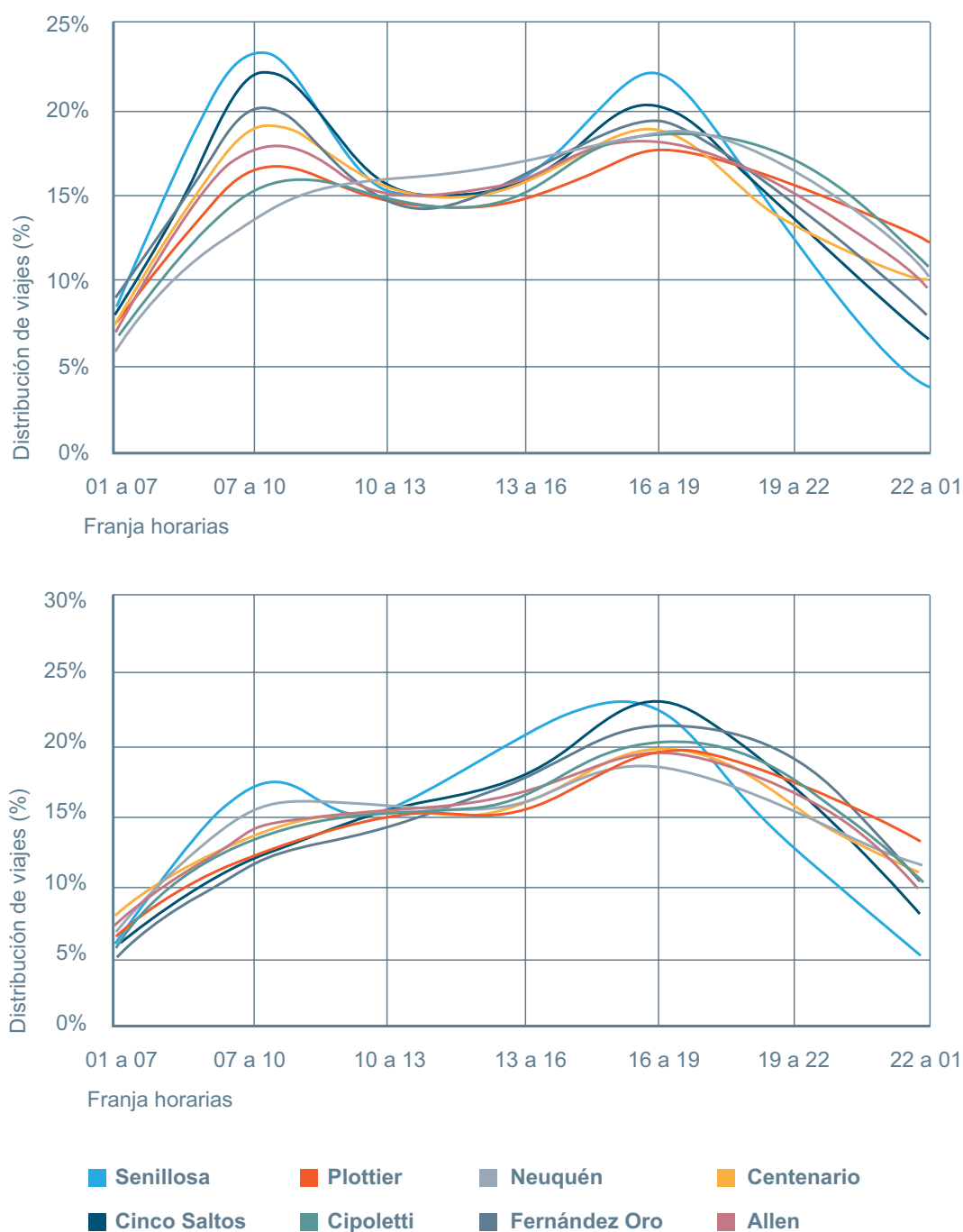


Figura 22: Ejemplo de distribución horaria de los viajes según municipios de origen (izquierda) y según municipio de destino (derecha). Tomado del trabajo realizado por Telefónica para el Banco Interamericano de Desarrollo en el Área Metropolitana de Neuquén, Argentina.



9. Validación de los resultados

La validación de los resultados se realiza a partir de la comparación de los resultados obtenidos con los datos de telefonía y los de otras fuentes. Según los datos disponibles es posible realizar la validación por:

- Las distribuciones de residencia y del lugar de trabajo obtenidas con los datos de telefonía pueden ser comparadas con las distribuciones de la población y del empleo de las fuentes oficiales (por ejemplo, por comparación con datos censales, registros de actividad económica, etc.). Obviamente esta validación debe realizarse antes de la expansión de los datos. Lo habitual es realizar una correlación bivariada entre las distribuciones de la distribución de la residencia en las muestras obtenidas y la de las fuentes oficiales (censos), esperando que los ajustes entre ambas sean muy altos.
- Los resultados de los análisis de movilidad pueden validarse también por comparación con los obtenidos en encuestas domiciliarias de movilidad si se ha realizado alguna previamente, con datos de tráfico (por ejemplo IMD), e incluso con datos de encuestas de uso del tiempo. En este caso se trata contrastar que los resultados obtenidos sean coherentes con los análisis previos.
- Otra forma de validar los resultados obtenidos es utilizando un grupo de control para el que podamos conocer su movilidad real y compararla con la que hemos obtenido a través de los eventos que han realizado con sus teléfonos móviles. Una posibilidad para realizar este procedimiento de validación es tomar un grupo de personas de los que se haya obtenido permiso para realizar un seguimiento de su actividad del teléfono móvil y obtener sus patrones de movilidad y, posteriormente, realizarles una encuesta sobre los viajes que han realizado.
- El procedimiento de validación mediante un grupo de control es muy raro, debido a los problemas de privacidad y las dificultades para acceder a dichos grupos. Por eso algunos autores han optado por generar este tipo de grupos de control (ver Chen et al., 2014). En este caso se opta por generar una base de datos simulada de viajes por eventos a partir de los datos obtenidos en una encuesta de movilidad. Esta base de datos generada se usa como grupo de control o verdad terreno (ground truth) con los que validar la información obtenida de los datos de telefonía.
- Finalmente, algunos estudios han validado los viajes detectados en transporte público con los registrados por las tarjetas de transporte. Este es el caso de uno de los estudios que aparecen en este documento (Estudio de movilidad de los usuarios del Ferrocarril Línea Roca) en el que se utilizaron los datos de la tarjeta SUBE para la validación de los resultados.



The background of the entire page is a light blue-grey color with a white line pattern that resembles a stylized city street map or a network of paths. The lines are of varying thickness and form a complex, interconnected grid.

6

**Privacidad: Pasos
fundamentales
para asegurar el
anonimato de los
usuarios**



Cada vez aparecen con más frecuencia reglamentos de carácter nacional o supranacional sobre protección de datos que tratan de evitar los riesgos que puedan derivarse del intercambio o mal uso de datos personales y obligan a las empresas a tomar medidas al respecto. Entre éstas, es habitual que, por un lado, se solicite a empresas y administraciones el consentimiento de los usuarios cuando sus datos van a ser utilizados en fines diferentes de aquellos para los que fueron creados y, por otro, que se garantice en todo momento la privacidad de esos usuarios. En todo caso, en cada país existen unas regulaciones de privacidad específicas, por lo que no es posible generalizar sobre los aspectos legales de privacidad en el tratamiento de los datos en el ámbito de la región de Latinoamérica.

En los estudios de movilidad a partir de datos de telefonía móvil es obligatorio mantener la privacidad de los usuarios de la compañía que genera los datos que vamos a utilizar. Los usuarios o clientes tienen legítimamente la expectativa de que su privacidad individual será preservada.

Por ello, en el desarrollo de cualquier trabajo es muy importante asegurar que tanto en las fases de la elaboración como de la presentación y distribución de los resultados no se infrinjan aspectos relacionados con la privacidad de los usuarios.

Sin embargo, en el estudio de la movilidad, si el foco se pone en la movilidad de los individuos esa privacidad se pone en riesgo. Algunos trabajos han mostrado cómo con conocer cuatro puntos espacio-temporales de la actividad generada por un usuario son suficientes para identificar de manera única al 95% de los individuos. En general, puede establecerse una relación entre la temática de los estudios realizados con datos de telefonía móvil, el volumen de datos a utilizar y la privacidad de los usuarios. Estudios sobre distribuciones de densidades de usuarios cambiantes en el tiempo a partir de sus registros telefónicos usan grandes volúmenes de datos y de forma agregada, sin que aparezcan riesgos de privacidad. Por el contrario, si el estudio es sobre el comportamiento en la movilidad de un grupo específico de población, el volumen de datos válidos se reduce y el detalle de la información que contienen aumenta, incrementado la

posibilidad de vulnerar la privacidad de esos usuarios. En el caso de la generación de matrices origen destino de viajes lo habitual es trabajar con un volumen alto de datos y agregado en unidades espaciales que aseguran el anonimato de los individuos.

En todo caso, como indicábamos, en los estudios de movilidad con datos de telefonía deben considerarse una serie de pasos que permitan proteger la privacidad de los individuos:

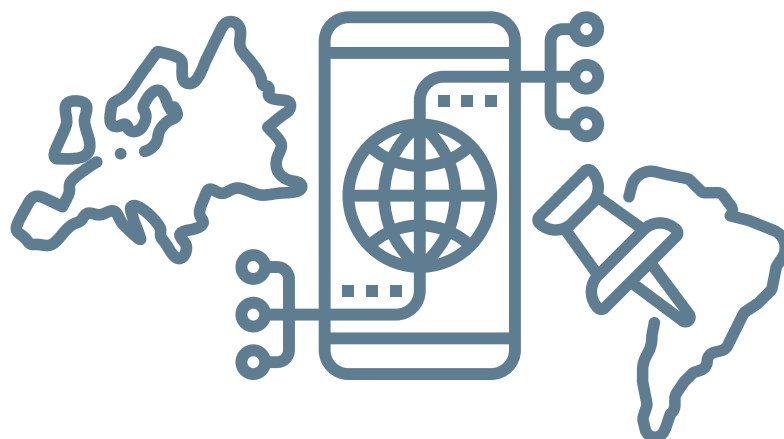
- **Anonimización.** Los datos de partida (los CDRs) deben ser en todo momento anónimos. Además, es obligatorio que esa anonimización sea realizada por alguien que no esté involucrado en el análisis de datos. Para ello, las compañías tienen equipos independientes que realizan este proceso y proporcionan los datos ya anonimizados.
- **Información mínima.** Debe utilizarse la mínima información necesaria para el estudio. En los estudios de movilidad se trabaja con datos de CDRs simplificados, que incluye solo los datos a utilizar en el trabajo. Así, normalmente es suficiente con: el identificador anónimo del teléfono, la fecha y hora del evento y las antenas celulares involucradas en el evento. En algunos casos pueden utilizarse datos como el tiempo transcurrido de una llamada y raramente se incluyen información demográfica del suscriptor, como el código postal de facturación del teléfono ni ninguna información sobre la otra parte en la comunicación.
- **Resultados agregados.** La presentación de los resultados se debe realizar a partir de agregaciones de la información en unidades espaciales. No se debe enfocar el análisis en teléfonos individuales, salvo los casos en los que se usen voluntarios que tengamos un permiso (“consciente”) para revisar sus registros. En todos los casos, la presentación de los resultados debe ser agregada, asegurando la confidencialidad. Las unidades espaciales sobre las que se van a agregar los datos deben tener un número mínimo de usuarios que asegure nuevamente la privacidad de los mismos.



The background of the entire page is a light blue-grey color with a faint, stylized map pattern. The map consists of a network of thin, light grey lines representing streets and boundaries, creating a grid-like but irregular pattern across the surface.

7

**Experiencias de
estudios de movilidad
a partir de datos de
telefonía móvil en
América Latina y
Europa**



Actualmente disponemos de datos provenientes de multitud de sensores que registran los movimientos de las personas de forma masiva y con una gran precisión espacial y temporal. Se destacan los provenientes de la telefonía móvil debido a la penetración de mercado que tienen este tipo de dispositivos. Gracias a estos datos podemos conocer la localización de los individuos en cualquier momento del día, lo que supone disponer de datos totalmente actualizados, de gran precisión y a un menor costo, comparados con otro tipo de datos como los obtenidos a través de las tradicionales encuestas de movilidad. Todo ello amplía la capacidad de realizar estudios capaces de captar las dinámicas espacio-temporales de los patrones de movilidad de los ciudadanos, de una forma más desagregada, lo que permite extraer conclusiones que hasta ahora eran únicamente posibles a escalas micro.

En este documento se presenta un conjunto de 6 casos de estudios que ofrecen una visión sobre las posibilidades de los datos provenientes de la telefonía móvil para los estudios de transporte. Para seleccionarlos, se ha tenido en cuenta la diversidad de los temas y su aplicación a diversos contextos espaciales tanto en América Latina como en Europa. En su conjunto, los casos muestran la aplicación de los datos provenientes de los teléfonos móviles al 1) análisis de patrones de movilidad a través del cálculo de matrices origen destino, 2) el estudio de la segregación socioeconómica, 3) los flujos turísticos, 4) la calibración de modelos de transporte, o 5) el rediseño de la red de transporte público. Asimismo, se presentan dos casos de estudio que utilizan estos datos para analizar 6) la concentración la accesibilidad a lo largo del día. La tabla siguiente presenta los casos de estudio, su temática y el contexto espacial donde se han aplicado.

Adicionalmente, en el anexo I se han incluido otros dos casos de estudio, el primero sobre cuestiones de género relativas al uso del transporte público y el segundo sobre la exposición de la población a los gases contaminantes.

| Casos | Temática | Contexto espacial |
|---|--|--|
| Capítulo 7 | | |
| Telefónica & BID, 2018 | Patrones de movilidad y obtención de matrices OD | Neuquén (Argentina) |
| Östh, J.; Shuttleworth, I.; Niedomysl, T. (2018) | Segregación socioeconómica | Estocolmo, Gotemburgo y Malmö (Suecia) |
| Ahas, R.; Aasa, A.; Roose, A.; Mark, U.; Silm, S. (2008) | Flujos turísticos nacionales | Estonia |
| García-Albertos, P.; Picornell, M.; Salas-Olmedo M.H.; Gutiérrez J. (2019) | Análisis de accesibilidad | Madrid (España) |
| Wismans, L.; Friso, K.; Rijdsdijk, J.; de Graaf, S.; Keij, J. (2018) | Calibración de modelos de transporte | Rotterdam (Holanda) |
| Positium (2017) | Rediseño de la red de buses | Tartu (Estonia) |
| Anexo I | | |
| Telefónica & BID, 2018 | Flujos de viajes en el ferrocarril Roca y línea de Subterráneo C y perspectiva de género | Provincia de Buenos Aires (Argentina) |
| Picornell, M.; Ruiz, T.; Rafael, B.; García-Albertos, P.; de la Paz, D.; Lumbreras, J. (2018) | Exposición a la contaminación | Madrid (España) |

El primer estudio (Telefónica & BID, 2018) analiza la movilidad en el área metropolitana de Neuquén (Argentina), utiliza los datos de telefonía móvil para estimar matrices origen destino y desagrega entre día laborable y fin de semana, motivo del viaje, hora, género, edad y nivel socioeconómico. También en esta línea se destacan otros estudios realizados por Telefónica y encargados por el BID como el llamado “Estudio de la zona que rodea el barrio Costa Esperanza como polo generador y receptor de viajes”, o el estudio “Aplicación de telefonía móvil para planificación y gestión de movilidad en el Distrito Metropolitano de Quito”. Ambos no se han incluido en el conjunto de los casos aquí presentados por su similitud en términos metodológicos al estudio de Neuquén y al estudio que se presenta a continuación.

A continuación, se presenta un caso (Östh, J.; Shuttleworth, I.; Niedomysl, T., 2018) que supone una crítica a los estudios de segregación social que caracterizan los individuos según su lugar de residencia y propone una metodología que introduce la movilidad de los residentes como un factor más en esa ecuación de segregación.

El estudio tiene lugar en tres grandes áreas metropolitanas de Suecia (Estocolmo, Gotemburgo y Malmö) y utiliza los datos móviles para ahondar en cómo las pautas de movilidad pueden agudizar o, por el contrario, contrarrestar situaciones de segregación socioeconómica de los ciudadanos.

Pero los datos de telefonía móvil no solo permiten analizar el comportamiento de los residentes, sino que también son capaces de caracterizar los flujos turísticos de una determinada región. Estonia es uno de los países que más ha avanzado en estudios de este tipo y por ello aquí se presenta un caso que evalúa la aplicabilidad de los sistemas pasivos de posicionamiento de los teléfonos móviles para el análisis de los flujos turísticos en este país (Ahas, R. et al., 2006). El estudio ofrece una información útil para la planificación turística de los recursos y aporta una fuente complementaria a las fuentes de datos oficiales.

Una de las grandes ventajas de los datos de telefonía móvil es su capacidad de captar la dinámica de los flujos de movilidad.

El número de personas en las zonas residenciales, de empleo, o de ocio cambia según va avanzando el día y con ello la demanda localizada en las diferentes zonas de la ciudad.

Este aspecto cambiante la movilidad no se ha tenido en cuenta en la mayoría de los estudios de accesibilidad, llevados a cabo en la planificación del transporte y los estudios urbanos. Sin embargo, es fundamental para la mejora de la efectividad de las políticas que incidan sobre el aumento de la accesibilidad. El cuarto estudio (García-Albertos, P. et al., 2019) presenta una metodología que aplica datos móviles obtenidos para la ciudad de Madrid, introduciendo el dinamismo de los flujos de movilidad en los análisis de accesibilidad.

Los datos de telefonía móvil permiten mejorar los resultados de los modelos de transporte y la eficacia de las políticas de planificación. El siguiente estudio (Wismans, L. et al., 2018) tiene lugar en Rotterdam, una de las principales ciudades holandesas en el sector logístico y financiero. Los datos de telefonía móvil se usan para calibrar la matriz origen-destino inicial del modelo de transporte y por consiguiente obtener unas mejores predicciones del modelo.

El último caso de estudio de este capítulo (Positium, 2017) utiliza los datos móviles para estimar los flujos de viajes a diferentes horas del día e investigar las necesidades de movilidad de la población de la ciudad de Tartu (Estonia). Gracias a estos datos, el estudio propone un rediseño de la red de líneas de autobuses, eliminando ciertas líneas/paradas y reestructurando algunas de las líneas existentes. Todo ello ha supuesto un aumento de la demanda de usuarios por parada y una disminución de los tiempos de espera.

En cuanto a los casos de estudio del anexo I, primero se incluye un estudio (Telefónica & BID, 2018) que aporta una mirada sobre cuestiones de género a la vez que analiza los patrones de movilidad de los usuarios de una línea de ferrocarril (Ramal Constitución-Plata) y la línea C del Subterráneo de Buenos Aires. Una de las grandes aportaciones del caso es el foco en la problemática de inseguridad relacionada con el acoso hacia las usuarias, durante el viaje en transporte público o en su acceso al mismo. También en esta línea relacionada con la evaluación de corredores de transporte, cabe señalar el estudio realizado por Telefónica a pedido del BID sobre la “Aplicación de tecnología móvil para análisis de la movilidad poblacional para la evaluación de impacto del corredor vial Atalaya- Mariatoquebró-Flores y Ramar a Varadero, Panamá”, no incluido en este documento.

Este anexo se cierra con un estudio de caso (Picornell, M., 2018) que muestra cómo los datos provenientes de la telefonía móvil mejoran los análisis sobre la exposición de los ciudadanos a los gases contaminantes. Hasta ahora este tipo de análisis utilizaba información sobre el lugar de residencia de los individuos para estimar su nivel de exposición. Sin embargo, las personas se exponen a diversos niveles de contaminación a lo largo del día en función de cómo varían sus pautas de movilidad. El estudio se ha llevado a cabo en el área metropolitana de Madrid.



7.1 CASO 1: Estudio sobre la movilidad en el área metropolitana de Neuquén¹

Sobre el caso de estudio

Se centra en el análisis de la movilidad en el área metropolitana de Neuquén, que incluye municipios del departamento de Confluencia (provincia de Neuquén) y del departamento de General Roca (provincia de Río Negro) (Figura 23). Su economía se basa en las actividades terciarias, particularmente funciones comerciales, de servicios y de administración (salud, educación superior y servicios sociales). En este sistema metropolitano Neuquén es el principal polo de atracción de los flujos urbanos, seguido en menor medida por las ciudades de Plottier y Cipolletti.

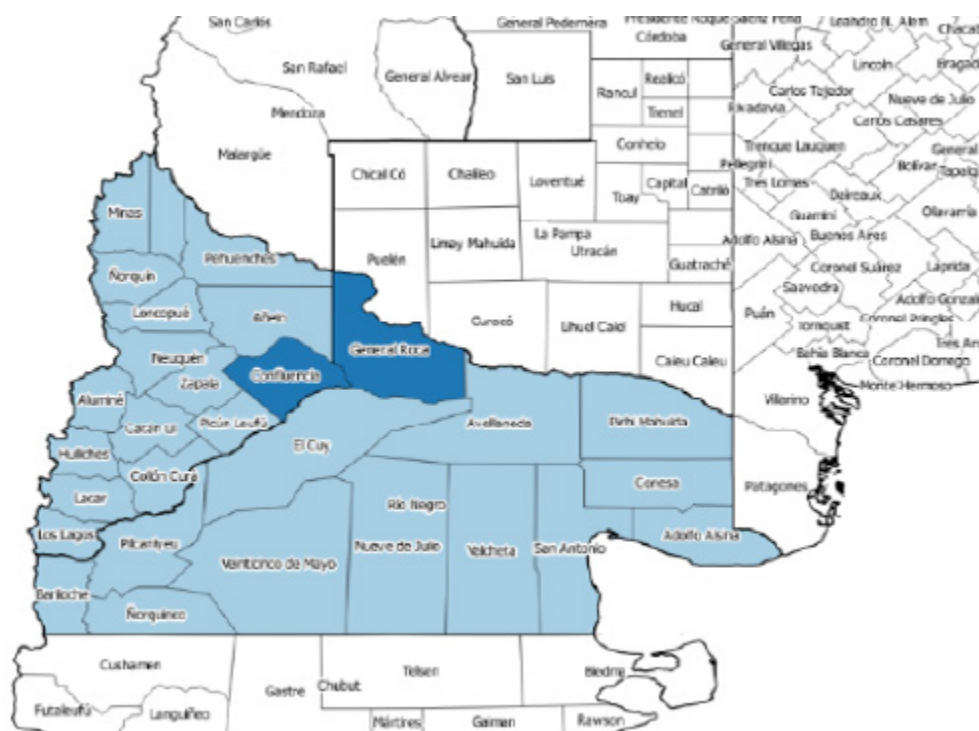


Figura 23: Encuadre del área de estudio.

El desafío

Este estudio pretende dar respuesta a tres objetivos principales, que son fundamentales en la planificación de transporte y del medio urbano:

- Actualizar y ampliar la encuesta de movilidad domiciliar de Neuquén-Cipolletti de 2010, estimando y caracterizando el flujo de viajes en el área metropolitana de Neuquén, incluyendo no solo residentes sino también otras personas que se mueven en la zona.

¹ Estudio de Movilidad utilizando datos de telefonía móvil, realizado por el Banco Interamericano de Desarrollo y Ministerio de Transporte de Nación Argentina.

- Determinar el dinamismo de los flujos de viaje y su evolución a lo largo del día, comparando días hábiles y fines de semana.
- Entender las diferentes dimensiones de la movilidad, considerando movilidad interna (dentro de cada localidad) y externa (entre las diferentes localidades), así como las principales características socioeconómicas de las personas que viajan en la zona.

La solución

A partir de la herramienta Smart Steps de Telefónica, basada en tecnologías Big Data para telefonía móvil, se obtienen datos de movilidad en un periodo acotado de tiempo (4 semanas de mayo 2016). Smart Steps utiliza datos de telefonía móvil anonimizados provenientes de un conjunto de terminales adscritos a la red de telecomunicaciones de Telefónica. La muestra de los datos la constituyen personas de más de 15 años. Los datos son agregados a nivel espacial para un conjunto de zonas situadas en el área de estudio y extrapolados para representar la población total de la zona.

Se define como zonas de interés ocho municipios: Neuquén, Plottier, Senillosa y Centenario (provincia de Neuquén); Cipolletti, Cinco Saltos, General Fernández Oro y Allen (provincia de Río Negro). En el caso de Neuquén, Cipolletti y Allen, municipios que concentran gran parte de la población del área de estudio, se utilizan zonas más pequeñas provenientes de la agregación de las localidades censales.



Figura 24: Municipios del área de estudio.

Se estima el número de viajes en un día típico y uno de fin de semana, desagregando por localidades y presentando los resultados en forma de matrices origen – destino. Esta información se desagrega según el motivo de los viajes (casa, trabajo, otros), franjas horarias, género, edad y estrato económico, a partir de inferencias respecto del uso del teléfono móvil (ver cuadro 1) y una encuesta con mensajes SMS.

A partir de las matrices origen-destino se estiman los principales centros productores y receptores de viajes, se calcula el número de personas en cada zona identificando los períodos de mayor demanda y se establecen los principales flujos de interacción entre las diferentes zonas analizadas.

Los resultados provenientes de los datos móviles se complementan con una encuesta realizada a diferentes usuarios a través del servicio Movistar+ sobre: 1) las actividades que motivan los viajes (estudios, trabajo, salud, ocio, otros, sin desplazamiento); 2) el modo de transporte (a pie, bicicleta, moto, automóvil particular, taxi / remis, colectivo); 3) el nivel de educación.

Cuadro 1

La segmentación por motivo del viaje considera que la casa de un usuario se sitúa en la localidad en la que se encuentra más frecuentemente entre las 20h y las 6h del día siguiente. Mientras que el lugar de trabajo de un usuario determinado es la localidad en la que se encuentra más frecuentemente entre las 10h y 16h del día siguiente. Respecto a la segmentación socioeconómica, se utiliza la facturación de cada usuario como proxy a la clase social.

Beneficios

El estudio ha permitido entender las diferentes dimensiones de la movilidad con el fin de mejorar la planificación urbana, del transporte y la calidad de vida de las personas.

Para ello se profundizó en el conocimiento sobre la articulación funcional del área de estudio, lo que permite proponer mejoras más efectivas en la oferta de transporte de la zona. En este sentido, se identificaron los principales centros productores y atractores de viajes.



Viajes con origen en la zona resalta-da

Viajes con destino en la zona resalta-da

Figura 25: Ejemplo de mapas de viajes con origen / destino en una zona de transporte.

La identificación de los principales flujos de viaje entre las zonas estudiadas facilita la efectividad de las políticas de gestión del transporte público y privado, mejorando las conexiones entre las zonas de mayor demanda, considerando además su dinámica a lo largo del día.

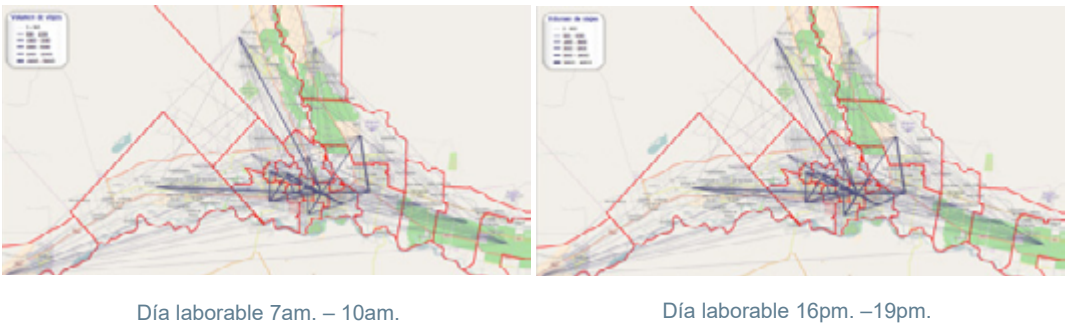


Figura 26: Ejemplo de mapas de viajes entre zonas de transporte según franjas horarias.

El estudio de los patrones espaciales de las personas y su dinámica a lo largo del día se presentan en forma de mapas que contabilizan el número de personas en cada zona. El análisis permite determinar las áreas de mayor concentración de usuarios y optimizar la asignación de recursos.



Figura 27: Densidad de población según franjas horarias.



7.2 CASO 2: Pautas de la movilidad en el estudio de la segregación económica de ciudades de Suecia

Sobre el caso de estudio

La segregación social es uno de los principales problemas de las ciudades contemporáneas y afecta de forma muy especial a las urbes latinoamericanas. Los estudios de segregación social se basan en el lugar de residencia de la población. Pero la población urbana invierte su tiempo no solo en el lugar de residencia, sino también en otros sitios donde realiza actividades. Algunos habitantes de las zonas de bajo nivel de ingresos pueden permanecer a lo largo del día en su barrio, mientras que otros tienen un comportamiento menos localista y aprovechan mejor las oportunidades que les brinda la ciudad. Por lo tanto, un estudio integral de segregación social debería considerar también las pautas de movilidad de la población.

El estudio se centra en las tres grandes áreas metropolitanas de Suecia: Estocolmo, Gotemburgo y Malmö. Según los autores la segregación económica ha aumentado en los países nórdicos en los últimos años, por lo que su estudio es de interés tanto desde el punto de vista académico como político. Hay además un especial interés por los barrios de la periferia de las grandes ciudades, debido a la concentración de estratos económicos menos favorecidos, con problemas sociales y la presencia de minorías étnicas.

El desafío

La mayoría de los estudios de segregación social y económica ponen la atención en los lugares de residencia, sin embargo, las personas que residen en un mismo barrio pueden estar expuestas a contextos sociales y económicos muy diferentes dependiendo de sus pautas de movilidad a lo largo del día. En materia de integración, unos de los prerrequisitos fundamentales es que miembros de diferentes grupos usen los mismos espacios en al menos algunas de las actividades de su vida cotidiana.

La principal cuestión planteada en este estudio es conocer en qué medida la movilidad de los individuos reduce o aumenta la segregación cuando ésta es comparada con el nivel de segregación asociado a su lugar de residencia.

Para ello, el primer desafío fue incorporar la movilidad y la dimensión temporal en las metodologías de segregación, con el fin de determinar hasta qué punto la movilidad diaria en las ciudades cambia las experiencias de segregación de las personas y los niveles de segregación de los lugares.

Otro de los desafíos de este estudio es la combinación de diferentes fuentes de datos, en particular utilizando las trayectorias de los usuarios estimadas a partir de datos de teléfonos móviles, que son cruzados con datos socioeconómicos detallados sobre los lugares de residencia.

La solución

Se utilizan datos anonimizados provenientes de telefonía móvil y otros sobre el nivel de ingresos en las tres áreas metropolitanas. Los datos de telefonía móvil proceden de uno de los principales operadores de telecomunicaciones en Suecia y cubren las 24 horas de un día laborable (martes). Se estima que representan un 15% de la población de Suecia. Los datos socioeconómicos provienen de PLACE, una base de datos sueca que contiene registros longitudinales sobre aspectos de la demografía, socioeconomía y la geografía de todos los residentes suecos.

Los datos socioeconómicos se utilizan para estimar la proporción de ricos y pobres, según una definición aceptada en la Unión Europea (ver cuadro 1). Estos datos se utilizan para estimar el nivel de riqueza y pobreza de cada área de cobertura de las torres (BTS), según las características de los 800 vecinos más cercanos y usando la metodología k-nearest, implementada en el software EquiPop (la cifra de 800 corresponde al número medio de teléfonos conectados a cada antena en las tres áreas metropolitanas).

Cuadro 1. Definición de pobreza y riqueza relativa

La pobreza relativa se atribuye a los individuos con el 60% o menos de los ingresos medios del país. La riqueza relativa se asigna a los individuos con el 140% o más de los ingresos medios del país.

El lugar de residencia se estima como la localización más frecuente en el intervalo de tiempo entre las 00:00 y las 07:20 y se asocia con el nivel de riqueza previamente definida para esa zona. De este modo, todos los móviles que comparten localización nocturna comparten además el mismo estatus socioeconómico.

A partir de estos datos se desarrolla una metodología espacio-temporal (ver cuadro 2) que permite realizar un seguimiento de la movilidad de los individuos a lo largo del día, con observaciones cada 5 minutos, para poder analizar la presencia simultánea de individuos provenientes de áreas con diferentes niveles de riqueza y pobreza. La exposición de los usuarios a diferentes estatus socioeconómicos supone una influencia de otros usuarios y una reducción de su grado de segregación.

Con esta información construyen dos conjuntos de datos comparables que llaman TG y Home, con la única diferencia que uno asume que la población no se mueve a lo largo de las 24h (Home) mientras que el TG asume que la población se mueve de acuerdo con los patrones de movilidad recogidos por sus teléfonos móviles. La comparación de ambos conjuntos de datos se realiza a través de métodos estadísticos globales como el Coeficiente de Gini que permite comparar entre áreas metropolitanas y entre ambos conjuntos de datos y a través de mapas que permiten visualizar la variación local de los valores TG frente a Home.

2 Östh J., Shuttleworth I., Niedomysl T. (2018) Spatial and temporal patterns of economic segregation in Sweden's metropolitan areas: A mobility approach. *Environment and Planning A: Economy and Space*. Vol. 50 (4) 809-825.

Cuadro 2. Metodología para el análisis espacio-temporal de la exposición a diferentes niveles de renta

La exposición de los usuarios a la riqueza y la pobreza es el producto de las exposiciones medias a lo largo de las 24h, según las ecuaciones 1 y 2 (con la pobreza como ejemplo):

1)

$$Pobreza_{jt} = \frac{\sum_i Pobreza_{nocturna_i}}{N_{Co\ presencia_{jt}}}$$

El valor de *Pobreza* es específico de cada espacio – tiempo, donde *j* es el área de cobertura de la torre, *t* la unidad temporal (intervalo de 5 minutos) e *i* el valor atribuible a los teléfonos móviles. El valor calculado es compartido por todos los usuarios que coinciden en ese espacio - tiempo. El valor de *Pobreza nocturna_i* representa la proporción de individuos pobres que coinciden en *j t*, frente al número total de usuarios en *j t* (*Co presencia*)

2)

$$PobrezaTG_t = \frac{\sum_i Pobreza_{jt}}{UnidadesTiempoObservadas_i}$$

En la ecuación 2 los valores de *Pobreza_{jt}* se agregan a nivel de individuo reflejando su grado de exposición a la pobreza a lo largo de las 24h. El valor *Pobreza TG_i* representa la exposición media a la pobreza experimentada por cada usuario fruto de sus pautas de movilidad (conjunto de datos TG) y que se utiliza para comparar con los valores socioeconómicos asociados a su lugar de residencia (conjunto de datos *Home*). *Unidades Tiempo Observadas_i* representa el número de intervalos de 5 minutos registrado por cada teléfono a lo largo de las 24h.

El conjunto de datos *Home* es similar al conjunto TG con la excepción de que el primero considera la población como estacionaria a lo largo del día. Esto significa que los valores usados para crear la *Pobreza nocturna_i* y *Riqueza nocturna_i* se mantienen invariables y que la proporción se calcula sobre la base de la población residente en el área de cobertura de la torre

Beneficios

Las estadísticas tradicionales, como los censos, no ofrecen datos sobre las trayectorias diarias y la movilidad diaria, por ello la gran mayoría de los estudios de segregación se han centrado en los lugares de residencia sin tener en cuenta las pautas de movilidad de los individuos. Los datos provenientes de telefonía móvil ofrecen esa oportunidad, a la vez que son suficientemente representativos de todos los grupos sociales debido a su fuerte penetración en todos los estratos sociales y demográficos.

La segregación económica es un fenómeno que ha ido en aumento en muchos países y que cobra una relevancia especial a nivel político.

La metodología propuesta permite comparar la segregación económica entre áreas metropolitanas y a lo largo del tiempo. Su aplicación es especialmente interesante en países fuertemente segregados social y económicamente como ocurre en muchos países de América Latina.

La metodología desarrollada permite saber hasta qué punto los patrones de segregación residencial en entornos urbanos se traducen en una separación física en el espacio-tiempo y en qué medida grupos con distintos niveles de renta

En este estudio los resultados muestran que la movilidad ayuda a reducir las diferencias en la exposición a la riqueza y a la pobreza en las tres áreas metropolitanas analizadas.

coexisten en el mismo espacio y tiempo fruto de sus pautas de movilidad, algo determinante en materia de integración social.

Los mapas de exposición a la riqueza y pobreza permiten ver dentro de cada área metropolitana como la segregación se ve afectada positiva o negativamente por la movilidad de sus habitantes.

A nivel local los resultados muestran que la movilidad permite reducir la segregación para algunos individuos, pero la mayor parte de la población permanece altamente segregada, independientemente de sus pautas de movilidad. En general las zonas del centro presentan menores niveles de segregación que la periferia. Los residentes en las zonas centrales están más expuestos a individuos residentes en otras zonas, debido a la fuerte atracción que ejerce el centro urbano. Por otro lado, los individuos que residen en la periferia, pero frecuentan el centro, también ven reducidos sus niveles de segregación debido a sus patrones de movilidad a lo largo del día. En cambio, los que residen en la periferia y apenas se mueven no ven alterados sus niveles de segregación diurna y nocturna.

Stockholm Metropolitan Area-Poverty

Stockholm Metropolitan Area-Wealth

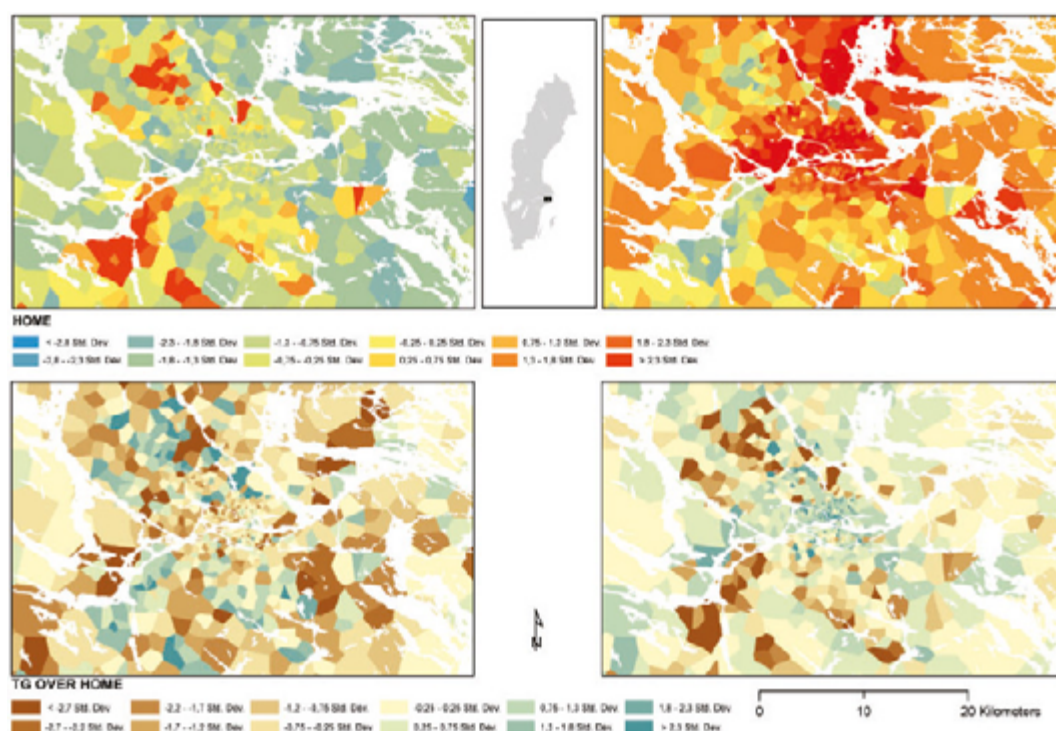


Figura 33: Los mapas muestran la distribución de la pobreza (izquierda) y riqueza (derecha) en el área metropolitana de Estocolmo. Colores rojos indican altas concentraciones (de pobreza o riqueza) y los azules bajas.

Los mapas inferiores representan la ratio de TG sobre Home. Cuando TG es mayor que Home (tonos azules) indican que la movilidad (TG) ha ayudado a exponer al grupo a otros contextos sociales y por lo tanto a reducir su grado de segregación, cuando Home es mayor que TG (tonos marrones) la movilidad ha debilitado dicha exposición.



7.3 CASO 3: Datos geolocalizados provenientes de telefonía móvil para la elaboración de estadísticas de flujos turísticos en Estonia

Sobre el caso de estudio

La movilidad turística tiene una importancia creciente en las ciudades latinoamericanas. Son numerosas las ciudades que reciben un número muy elevado de turistas. La población turística representa una proporción importante de la demanda de transporte no solo en los destinos típicamente vacacionales (como Punta del Este, Mar del Plata o Viña del Mar) o en ciudades con un muy destacado patrimonio histórico (Cartagena de Indias, Quito, La Antigua y Colonia Sacramento, entre otras.) sino también en muchas ciudades de Latinoamérica que combinan diferentes atractivos turísticos (Buenos Aires, Río de Janeiro, Santiago de Chile y México DF). Sin embargo, la movilidad turística es una gran desconocida, ya que las fuentes tradicionales (particularmente las encuestas domiciliarias origen-destino) solo recogen los viajes de los residentes.

Este estudio de caso presenta cómo se pueden utilizar datos de telefonía móvil para conocer la movilidad de los turistas. Se toma como ejemplo Estonia, un destino turístico emergente en el contexto europeo, conocida por su patrimonio histórico de la ciudad antigua de Tallin, por el turismo de naturaleza y por su clúster de spas. El incremento del turismo evidenciado en los últimos años en el país supone un reto para una correcta planificación de este sector. La existencia de fuentes de datos y estudios sobre los flujos turísticos cobra así una especial relevancia.

El desafío

Las estadísticas tradicionales sobre turismo incluyen la recogida de datos en los pasos de aduana, el transporte, las pernoctaciones, cuestionarios y diferentes conjuntos de datos modelizados. Sin embargo, las estadísticas existentes son limitadas en el espacio y en el tiempo y no permiten el análisis de cuestiones más complejas como la elección de destinos, la valoración de lugares de interés o los puntos de atracción turística visitados. El presente estudio evalúa la aplicabilidad de los sistemas pasivos de posicionamiento de los teléfonos móviles para el análisis de los flujos turísticos. Con ello se pretende construir una serie de indicadores cuantitativos que ofrezcan una información útil para la planificación turística y sean complementarios a las fuentes de datos oficiales.

La solución

Se parte de una base de datos pasivos que contiene todos los registros de llamadas roaming (teléfonos extranjeros) en la red de Estonian Mobile Telephone (EMT). Esta es la mayor compañía telefónica de Estonia con un 60% de cuota de mercado. Los datos, que cubren 17 meses (abril 2004 a agosto 2005), incluyen 12.8 millones de

llamadas y 1.2 millones de visitantes de 96 países. La nacionalidad se determina por el país donde está registrado el teléfono móvil y únicamente se trabaja con el turismo internacional.

La red incluye 690 torres y sus respectivas áreas de cobertura, calculadas a partir de polígonos de Voronoi. Posteriormente estas áreas son agregadas a 224 municipios y 15 condados en base a métodos de interpolación para facilitar comparaciones con las estadísticas oficiales (pernoctaciones).

Los autores utilizan un sistema gestor de bases de datos espacial (PostgreSQL 8.1 con la extensión espacial de PostGIS) y un sistema de información geográfica (ArcGIS) para realizar análisis de interpolación (IDW) y Densidad (Kernel), además de las salidas cartográficas de los resultados. Los datos han permitido analizar la distribución de los flujos turísticos en Estonia en los diferentes condados, evaluar la densidad de turistas, su variación temporal y determinar su correlación con los datos oficiales sobre pernoctaciones.

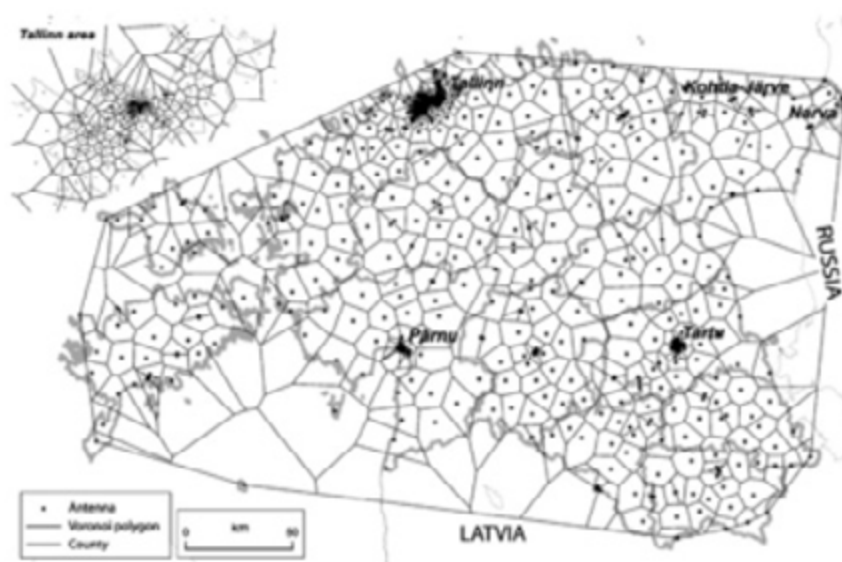


Figura 34: Áreas de coberturas de las torres

Beneficios

Los resultados prueban que los datos de telefonía móvil tienen un gran valor para los estudios de la movilidad urbana, particularmente en las ciudades. Se pueden obtener estadísticas tanto para grandes áreas como para zonas menos visitadas donde apenas hay datos. Además, hay que resaltar la mayor precisión espacio-temporal de los datos de los teléfonos móviles, respecto de lo que ofrecen hoy las estadísticas oficiales existentes.

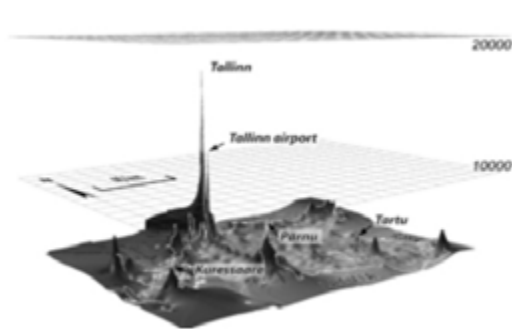


Figura 35: Distribución espacial de las llamadas roaming en Estonia.

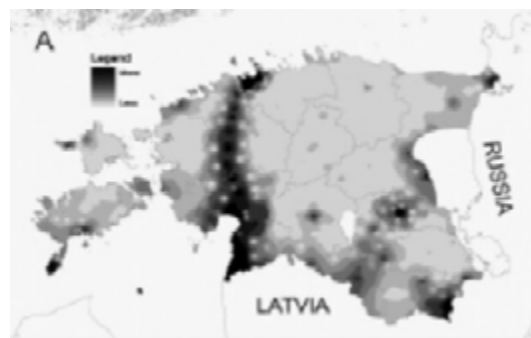


Figura 36: Localización de la primera llamada en Estonia (A) de Letonios y (B) Rusos.

La figura 35 muestra los puntos de mayor densidad turística a nivel nacional, evidenciando el peso de las principales ciudades, pero también de ciertos corredores de transporte. Por otro lado, gracias a los Identificadores anonimizados de los teléfonos fue posible hacer un seguimiento de las rutas de los turistas durante su estancia en Estonia. Esto ha permitido por ejemplo determinar un punto de entrada al país a partir del lugar donde se establece la “primera llamada” en Estonia (Figura 36), diferenciando según nacionalidades.

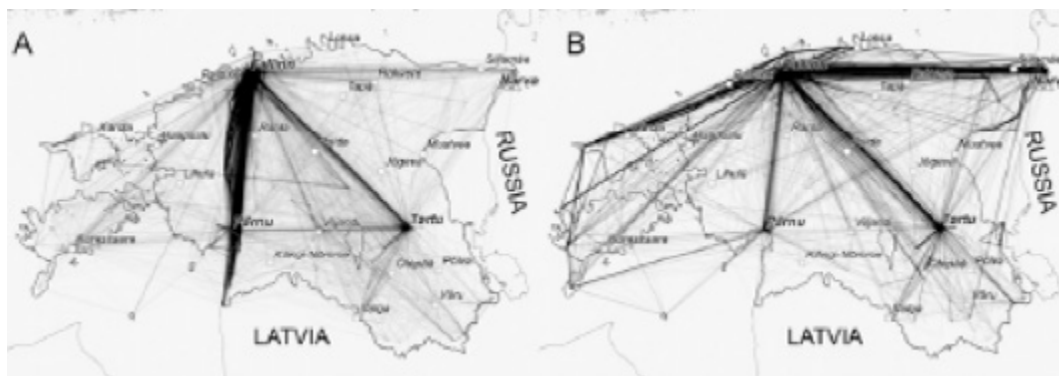


Figura 37: Corredores de movilidad de (A) Letonios y (B) Rusos en Estonia durante un periodo festivo (Junio 22-25, 2004).

Los registros de los datos móviles fueron modelados a lo largo de las principales carreteras del país para determinar las principales rutas turísticas (Figura 37), una información crucial para conocer los destinos más populares y mejorar las estrategias turísticas. Asimismo, es posible conocer la actividad turística en otras zonas del país menos frecuentadas y para las cuales hay menos datos, como ocurre con determinados parques, localizaciones aisladas y ciudades pequeñas.

Gracias a la componente temporal de los datos fue posible conocer la variación estacional de los flujos turísticos viendo por ejemplo la variación entre los días laborables y de fin de semana o el efecto de las vacaciones o días festivos a lo largo del año (Figura 38).

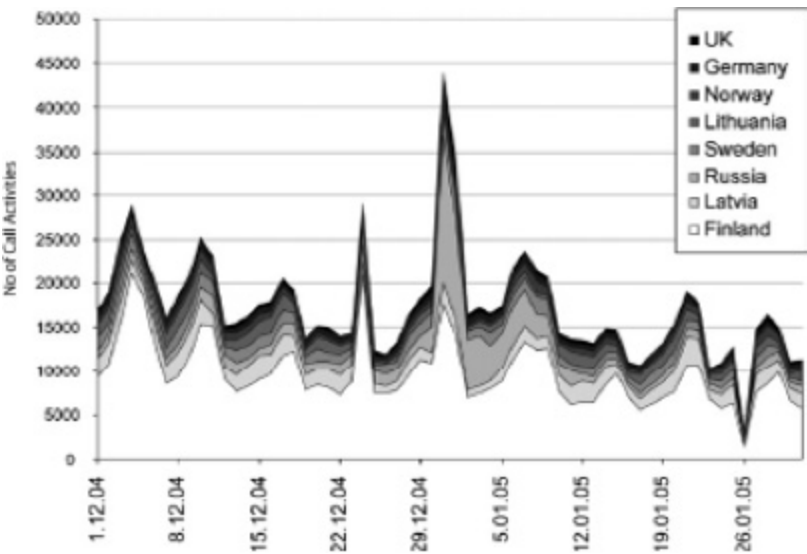


Figura 38: Dinámica temporal de los flujos turísticos.

Finalmente, destacar la fiabilidad de los datos de telefonía móvil cuando son comparados con las estadísticas tradicionales de pernoctaciones. En el caso de Estonia el mayor volumen de datos roaming se alcanza en verano, siendo considerablemente superiores los datos de los países vecinos (Finlandia, Letonia, Rusia y Lituania). En las regiones más turísticas de Estonia la correlación entre los datos móviles y las estadísticas convencionales es del 0.99 ($P < 0.005$), siendo algo menor en regiones de paso o en las regiones menos visitadas (Figura 39).

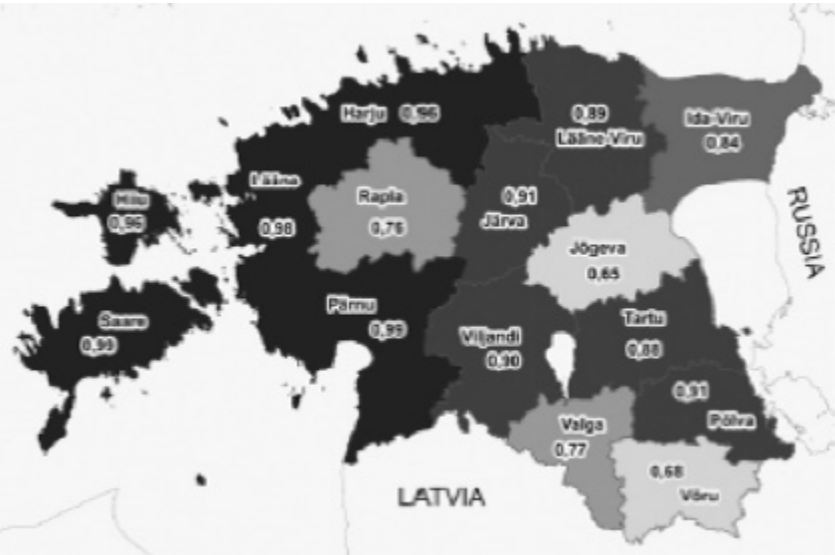


Figura 39: Correlación entre las estadísticas de pernoctaciones y la actividad de llamadas.



7.4 CASO 4: Análisis dinámico de la accesibilidad a partir de datos de telefonía móvil en Madrid

Sobre el caso de estudio

Numerosas ciudades latinoamericanas sufren graves problemas de accesibilidad, debidos a la congestión circulatoria, la insuficiente cobertura de la red de transporte público o las bajas frecuencias. Las autoridades de las ciudades deben tomar medidas para mejorar la accesibilidad, ya sean referentes al transporte (mejora de las infraestructuras y servicios) o a los usos del suelo (construcción de nuevas instalaciones de servicios y áreas de expansión de la ciudad).

Los análisis de accesibilidad han adoptado siempre una perspectiva estática, cuando ésta en realidad varía a lo largo del día en función de los cambios en las redes de transporte (congestión, frecuencias del transporte público) o de la elección de los destinos de viaje por parte de los usuarios. El estudio que se presenta aquí analiza la accesibilidad en transporte privado en la ciudad de Madrid (España) de forma dinámica, considerando la variabilidad temporal de los tiempos de acceso y de los destinos elegidos. El municipio de Madrid cuenta con cerca de 3 millones de habitantes y se compone por áreas muy diversas en términos de densidad de población y usos del suelo. Sin embargo, la mayor concentración de actividades se da en el centro y en el norte de la ciudad, mientras que el sur es predominantemente residencial. Esta configuración provoca importantes flujos de movilidad desde las periferias hacia las zonas del centro y del norte por la mañana, y flujos en la dirección opuesta por la tarde. Conocer sus dinámicas es esencial para una correcta planificación urbana y del transporte.

El desafío

La accesibilidad se puede definir como la facilidad de poder alcanzar las actividades deseadas a partir de una determinada localización. Representa una potencialidad de un territorio, que depende por un lado de la proximidad a los recursos deseados y por otro, del sistema de transporte que facilite ese desplazamiento.

Depende además de como varíen estos factores a lo largo del día, una vez que las actividades suelen tener un horario y un nivel de afluencia determinado y que existen diferentes niveles de congestión que repercuten en la facilidad de acceso.

El presente estudio tiene como principal desafío medir la accesibilidad de una forma dinámica, incorporando nuevas fuentes de datos como son los datos de telefonía móvil y la información de tráfico obtenida a partir de la API de Google. La presencia de los teléfonos móviles en la ciudad permite representar la capacidad de atracción de la demanda a diferentes horas del día, mientras que los datos de tráfico ofrecen la posibilidad de analizar la dinámica de la congestión. Las

diferentes áreas de la ciudad podrán ser clasificadas respecto de su accesibilidad y los factores que la caractericen (congestión y concentración de usuarios) en cada momento del día.

La solución

En este estudio los autores usan la API de Google Maps para calcular los tiempos de viaje en coche entre cada par origen-destino y su variación a lo largo del día. Para analizar la atracción de las diferentes zonas de la ciudad se obtienen matrices de viajes origen-destino a partir de los datos de telefonía móvil, asumiendo que el número de viajes que llegan a una zona es representativo de su nivel de atracción.

Los datos de los teléfonos móviles representan el comportamiento de la población en un día laborable promedio de noviembre 2014, para una población residente de más de 16 años. Proviene de uno de los principales operadores en España, con una cuota de mercado de un tercio de los consumidores. Se han utilizado las 584 zonas de transporte definidas por la Autoridad de transporte de Madrid (Consortio Regional de Transportes).

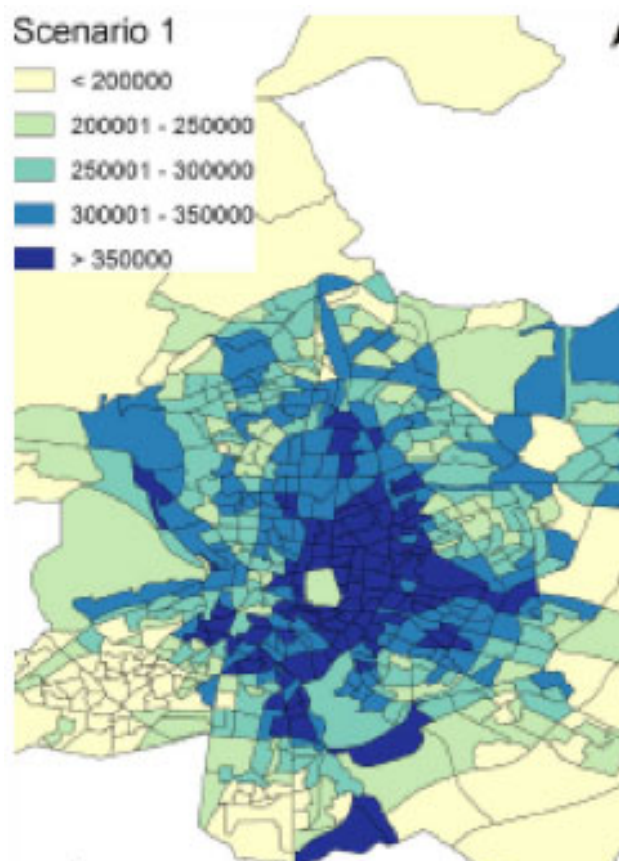


Figura 40: Zonas de transporte utilizadas.

4 García-Albertos P., Picornell M., Salas-Olmedo M.H., Gutiérrez J. (2018) Exploring the potential of mobile phone records and online route planners for dynamic accessibility analysis. *Transportation Research Part A: Policy and Practice*. <https://doi.org/10.1016/j.tra.2018.02.008>.

La metodología aplicada se divide en tres pasos. El primero consiste en la estimación de la matriz de viajes a partir de los teléfonos móviles. Para ello en primer lugar se eliminan los usuarios con pocos datos sobre su movilidad, se estiman los lugares de actividad de los usuarios utilizando un método de clustering desarrollado por García-Perez et al.⁵ (2010) así como su lugar de residencia en base a la metodología propuesta por Picornell et al.⁶ (2015)

En segundo lugar, se obtienen las matrices de tiempo de viaje a partir de la API de Google Maps. Los tiempos de viaje se calculan para cada hora del día, lo que permite medir el impacto de la congestión en la hora punta de la mañana y de la tarde. Se eligen los centroides de las zonas de transporte como puntos de origen y destino.

Por último, se calculan los valores de accesibilidad para tres momentos del día: hora pico de la mañana (entre las 9:00 y las 10:00), hora valle (entre las 12:00 y las 13:00) y hora pico de la tarde (entre las 19:00 y las 20:00). Se utiliza un indicador de oportunidades acumuladas (A_i) que contabiliza el número de viajes en cada zona de destino (w_j), que pueden ser alcanzadas dentro de un límite de tiempo (t_{ij}) desde una determinada localización de origen (i). Al tiempo de viaje calculado con la API de Google Maps se suman 10 minutos para simular el tiempo de acceso andando desde el lugar de origen hasta donde se ha aparcado el coche, así como el tiempo de encontrar un lugar de aparcamiento y caminar hasta el lugar de destino. El límite de tiempo (t_{ij}) para considerar las actividades acumuladas se establece en 30 minutos.

Beneficios

El estudio ha permitido medir la accesibilidad recogiendo la influencia de dos factores con efectos opuestos. La congestión aumenta el tiempo de viaje y disminuye la accesibilidad, mientras que la concentración de usuarios en un determinado punto aumenta la atracción de esa zona y ejerce un efecto positivo sobre la accesibilidad. La metodología utilizada en este estudio permite ver el efecto conjunto de la congestión y la atracción de los destinos, analizados de forma dinámica para recoger su variación a lo largo del día.

5 Gatica-Perez, D., Montoliu, R., Raul Montoliu, 2010. Discovering human places of interest from multimodal mobile phone data. In: Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia, Limassol, Cyprus, 2010.

6 Picornell, M., Ruiz, T., Lenormand, M., Ramasco, J., Dubernet, T., Frias-Martinez, E., 2015. Exploring the potential of phone call data to characterize the relationship between social network and travel behavior. *Transportation* 42, 647–668.

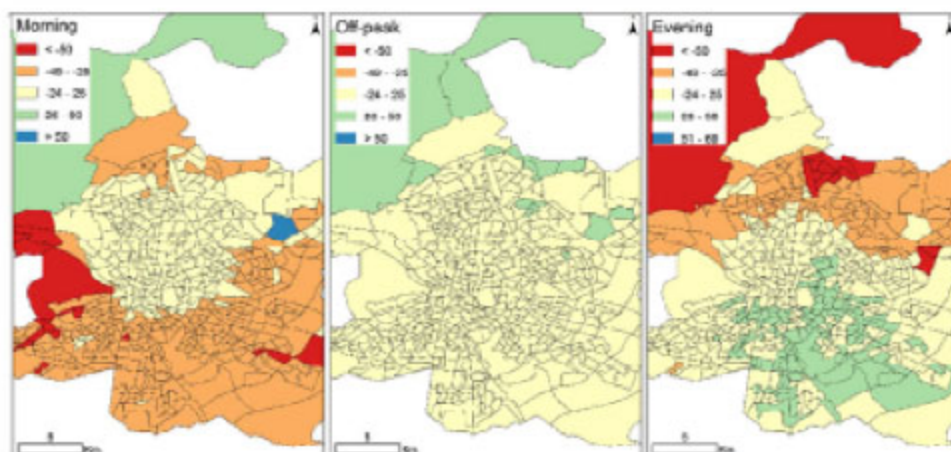


Figura 41: Accesibilidad dinámica (variando congestión y la atracción de los destinos)

Los mayores niveles de accesibilidad ocurren como sería de esperar en la hora valle (Figura 3), sin congestión, mientras que la hora punta de la mañana es la que más penaliza la accesibilidad de la ciudad en su conjunto. Las zonas de la periferia sur pierden accesibilidad sobre todo por la mañana debido al efecto conjunto de la congestión y los usuarios que salen para trabajar en el centro y norte de la ciudad. En la hora valle los niveles de accesibilidad se asemejan a la accesibilidad media diaria y en la hora pico de la tarde la periferia sur aumenta de nuevo su accesibilidad gracias al retorno de los residentes en esta área.

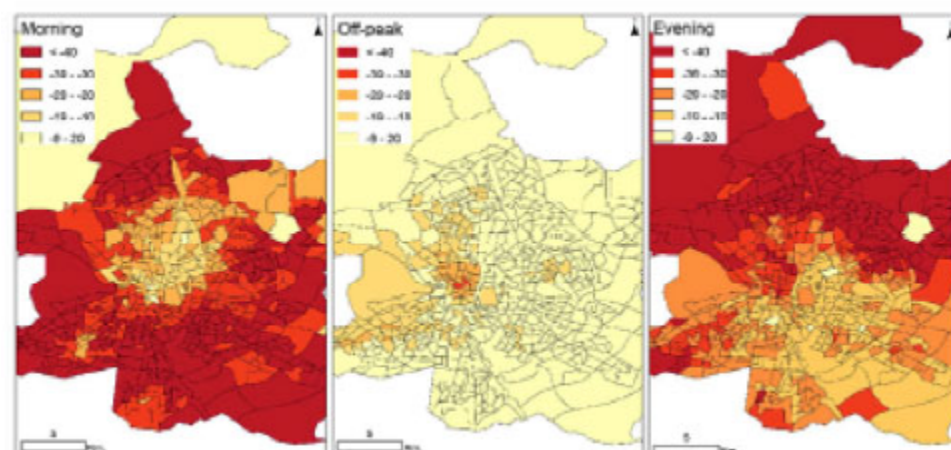


Figura 42: Efecto de la congestión sobre la accesibilidad

El estudio analiza también el efecto atribuible exclusivamente a la congestión. Para ello crea un escenario que mantiene fijo el factor de la masa de los destinos (w) para medir el efecto de los cambios en los niveles de congestión (Figura 43). La congestión tiene efectos negativos sobre la accesibilidad, pero especialmente en la hora pico de la mañana y en la periferia, debido a la enorme cantidad de viajes hacia el centro y norte de la ciudad, donde se concentran una gran parte de las empresas en Madrid.

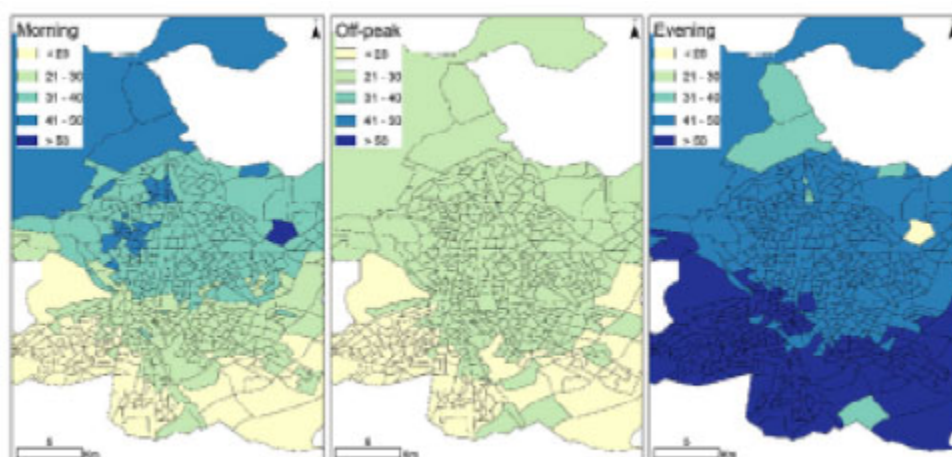


Figura 43: Efecto de la atracción de los destinos sobre la accesibilidad.

Para estimar el efecto de la variación temporal del factor de atracción de los destinos sobre la accesibilidad, se elaboró un nuevo escenario fijando los tiempos de viaje (t) en los valores medios diarios. En este caso, se pueden comprobar las ganancias en accesibilidad en la hora pico tanto en el norte como en centro, debido a que estas zonas concentran muchos empleos y atraen muchos viajes (Figura 5).

Finalmente, los autores realizan una clasificación de las zonas de la ciudad en base a sus niveles de accesibilidad, los cambios debidos a la congestión y los cambios en la componente de atracción de los destinos, estimando un total de 6 tipos de zona:

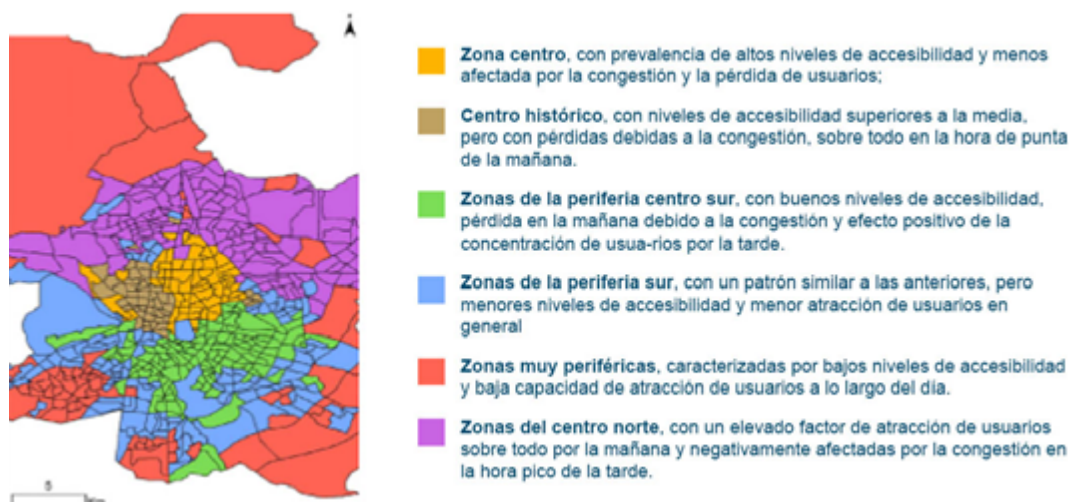


Figura 44: Tipos de zonas de transporte (análisis cluster).

Los resultados de este estudio contribuyen a un mayor conocimiento sobre las causas de una baja accesibilidad en medio urbano, diferenciando entre problemas relacionados con las prestaciones de la red de carreteras y aquellos relacionados con los usos del suelo.



7.5 CASO 5: Estimación de la demanda a partir de datos móviles y su aplicación en modelos de transporte en Rotterdam⁷

Sobre el caso de estudio

Los modelos de transporte siguen siendo cruciales para su planificación y la evaluación de la efectividad de las inversiones en infraestructuras. Su utilización se basa normalmente en datos provenientes de distintas fuentes sobre el sistema territorial (censos, usos del suelo...) y el sistema de transporte (aforos, capacidad de las vías, frecuencias de paso del transporte público...), así como en encuestas domiciliarias de movilidad. Sin embargo, los datos tradicionales tienen varias deficiencias, como la limitación del área a la que hacen referencia, que no siempre incluye todos los flujos de movilidad, o la falta de actualización de los datos. En este sentido, los datos de telefonía móvil tienen un gran potencial para mejorar los resultados de los modelos y la toma de decisiones en las políticas de transporte. El presente caso de estudio tiene lugar en la ciudad de Rotterdam que ha hecho la transición de una economía industrial a una economía de servicios, sobre todo en lo que se refiere a logística, transporte y finanzas. Los flujos de viaje entre Rotterdam y las principales ciudades de Holanda son muy importantes y deben ser incluidos en los modelos de transporte, motivo por el cual se utilizarán los datos provenientes de telefonía móvil.

El desafío

Un modelo de transporte estima la demanda y la asigna a la infraestructura generalmente en 4 etapas: producción/atracción, distribución, elección modal y asignación (elección de la ruta). Los primeros tres pasos se corresponden al modelo de demanda que genera una matriz de viajes inicial por modos de transporte, mientras que el último asigna la demanda a los arcos a través de un modelo de red (ver capítulo 1). La matriz OD inicial se obtiene normalmente a partir de encuestas domiciliarias de movilidad. Sin embargo, los datos de las encuestas son caros, representan una reducida muestra (en usuarios y periodo) y generalmente se encuentran anticuados, debido a que pasan varios años hasta que se puedan repetir las encuestas. También se reconoce que las encuestas tienen un sesgo respecto al número de viajes de cada hogar, porque se suele reportar menos viajes de las realizadas en realidad.

Una correcta calibración de la matriz OD inicial determina la exactitud de la matriz de viajes generada a posteriori y finalmente la calidad del modelo de transporte. Este estudio propone introducir los datos de los teléfonos móviles para mejorar la estimación de la matriz OD, mejorando la estructura de la matriz inicial y por ende la calidad de la matriz OD generada a posteriori, así como las predicciones del modelo de transporte.

⁷ Wismans L.J.J., Friso K., Rijdsdijk J., de Graaf S.W., Keij J. (2018) Improving a priori demand estimates transport models using mobile phone data: A Rotterdam-Region case. *Journal of Urban Technology*, 25, 2, 63-83. <https://doi.org/10.1080/10630732.2018.1442075>.

La solución

Se utiliza un modelo de transporte multimodal (coche, camión, transporte público y modos activos) que se aplica a la región de Rotterdam, para un día laborable, incluyendo las horas pico de la mañana y la tarde. La región se divide en cerca de 6.000 zonas de transporte y se incluyen además la red y otras regiones, pero con un menor nivel de detalle. En base a los datos oficiales de la encuesta domiciliaria de viajes de Holanda, se construye la matriz OD inicial. La asignación de viajes se basa en un modelo de equilibrio utilizando una función de tiempo de viaje en los arcos y retrasos en los nodos.

El estudio aplica los datos móviles (CDR) para derivar una matriz inicial de viajes para Holanda a nivel de ciudades y para el mes de noviembre de 2014. Esto se hace primero eliminando los usuarios con poca actividad, luego determinando su localización en base a la torre asociada y agregando la información para 1259 zonas, que corresponden a las ciudades y, en el caso de las ciudades grandes, a los distritos.



Figura 50: Flujos de viaje calculados a partir de los CDR.

La localización se traduce en información origen-destino, ordenando las estancias (localizaciones donde el usuario ha permanecido más de 30 minutos) en una secuencia temporal, mientras que los viajes corresponden al tiempo transcurrido entre dos estancias consecutivas.

Los viajes internos (dentro de las ciudades) solo se tienen en cuenta en el caso de las grandes ciudades, lo que supone una discrepancia entre las 6.000 zonas de transporte del modelo y las 1.259 zonas de la matriz de viaje calculada a partir de los CDR. En la práctica esto significa que el uso de la matriz calculada a partir de datos móviles, solo se puede utilizar para enriquecer la información sobre los viajes externos.

Finalmente se propone y evalúa la combinación de los métodos tradicionales y los datos móviles. Para ello se combinan la matriz OD obtenida con los datos de la encuesta (basada en 6.000 zonas de transporte) con la matriz calculada a partir de los CDR (1.259 distritos y ciudades), utilizando el procedimiento mostrado en el cuadro 1.

Cuadro 1. Procedimiento para enriquecer la matriz de demanda inicial

1. Cálculo de la matriz de los CDR en día laborable medio.
2. Conversión de la zonificación del modelo a la zonificación de los datos móviles.
3. Agregación de la matriz inicial de viajes por modo y motivo para obtener una matriz OD general para un día laborable medio
4. Escalar las relaciones OD del modelo inicial (filas y columnas) a las relaciones OD medidas con los datos móviles.
5. Expandir la matriz OD enriquecida al sistema de zonas del modelo de Rotterdam, así como para los diferentes modos, utilizando las tasas de reparto modal del modelo del modelo inicial, resultando así una matriz OD por modos enriquecida.

La diagonal de la matriz inicial no se tiene en cuenta en el proceso de enriquecimiento de datos por falta de información en los datos móviles.

El proceso de enriquecimiento se llevó a cabo solo para las relaciones con distancias comprendidas entre los 8 y los 60 km debido al error observado entre los datos móviles y los resultados de la encuesta.

Beneficios

El análisis de los CDRs permite extraer el movimiento de los usuarios y determinar los viajes realizados. Ofrece la oportunidad de entender las dinámicas diarias y actualizar los modelos de transporte más frecuentemente. Además, permite la toma de decisiones más informadas y de una forma más rápida, a la vez que posibilita contestar nuevas preguntas. En este estudio, una matriz OD calculada a partir de datos móviles se utiliza para enriquecer una matriz estimada a partir de la encuesta domiciliaria de viajes, mejorando su estructura, es decir la distribución de los viajes entre las diferentes zonas de transporte.

Los viajes cortos en la matriz obtenida con los datos móviles aparecen infravalorados y lo opuesto ocurre en los viajes largos. Esta desviación se ha tenido en cuenta a la hora de enriquecer la matriz original y mejorar los resultados del modelo de transporte.

Los resultados de la asignación de viajes obtenidos con la matriz original se comparan con los obtenidos con la matriz enriquecida, utilizando los aforos de tráfico para su validación.



Figura 51: Comparación de la asignación de coches obtenida con la matriz enriquecida y la matriz original.

Como resultado se observa que la matriz enriquecida produce un mejor ajuste y un menor error que la matriz original. Todo ello redundará en una mayor calidad de los resultados, así como en mejores previsiones de tráfico.



7.6 CASO 6: Propuesta de reestructuración de la red de líneas de autobuses urbanos basada en el cruce de datos de telefonía móvil en Tartu⁸

Sobre el caso de estudio

El diseño de los sistemas de transporte público debe facilitar el acceso de la población al empleo y los servicios. Sin embargo, algunas rutas quedan obsoletas con el paso de los años debido a los cambios en la ubicación de los hogares, lugares de trabajo y servicios de la ciudad.

Una red que no atiende de manera óptima las necesidades de movilidad de los residentes limita la posibilidad de cambiar la distribución modal de vehículo privado al transporte público, dificultando así la transición hacia una movilidad sostenible y una mayor calidad de vida.

La ciudad de Tartu inició el proceso de rediseño para una nueva red de líneas de autobuses en 2016 que pretendía mejorar la eficiencia de la red de acuerdo con la actual demanda de viajeros. La red existente fue diseñada al principio de los 90' y cuenta con 27 líneas, que seguían un esquema de rutas que conectaban muchos lugares con cada línea. La limitación de este tipo de líneas es su baja velocidad, lo que unido a las bajas frecuencias, en algunos casos repercutía en una baja demanda. Otros aspectos de la red de autobuses habían sido mejorados anteriormente, como la flota de buses, el sistema de billetes electrónicos, sistema de información al usuario en los buses y las paradas, información disponible en la página web y las aplicaciones de móviles. El estudio se lleva a cabo por las empresas Positium y WSP Finlandia que se encargan de desarrollar y validar los resultados de una metodología basada en el análisis de datos masivos de telefonía móvil y la simulación del sistema de transporte.

El desafío

El proyecto de rediseño de la red de autobuses se inició en 2017 con la elaboración de una nueva visión del sistema público de autobuses de cara a 2030, que incluía los siguientes aspectos:

- Imagen atractiva del transporte público: identidad visual común y comunicación clara.
- Sistema inteligente integrado: conexión de las redes de autobuses, trenes y bicicletas de alquiler.
- Uso simple y lógico: planificación de viajes cómoda y fácil, información en tiempo real y sistemas de emisión de billetes.

Pilsas, Gerttu: (2018): Public Transportation Planning with Data Driven Approach. Tartu use case. Positium

- Conexiones fáciles y rápidas: provistas por menos líneas de buses que son directas, rápidas y con mayor frecuencia.
- El servicio de transporte público tiene prioridad sobre el transporte privado, lo que garantiza la sostenibilidad.
- El servicio de transporte público es asequible: el servicio efectivo tiene costes más bajos y ofrece una opción de calidad asequible para la movilidad diaria.

El diseño de la red se planificó a partir de tres principios fundamentales:

- Menos líneas de autobuses. El diseño de líneas rectas con mejor interconexión permite una mayor frecuencia.
- Los tiempos de espera constantes permiten una mejor conexión entre las líneas y evitan la acumulación de autobuses en las paradas. El tiempo de espera constante elimina también la necesidad de memorizar horarios.
- La reducción del número de líneas de ruta circular y el uso principalmente de rutas radiales crea una red que es más fácil de entender.

La solución

El proceso de diseño se inició con el análisis de la demanda de viajes realizada por Positium. El análisis se basó en varios pasos y capas de datos:

- Análisis cualitativo de los documentos estratégicos de la ciudad de Tartu y la antigua red de líneas de autobuses. Éstos dieron la información necesaria para crear futuros escenarios de desarrollo urbano, la auditoría del servicio de transporte público existente brindó la experiencia del usuario existente.
- Creación de la base de datos PostgreSQL para analizar datos espaciales. Constaba de 57 tablas de datos.
- Importación de capas de datos tradicionales a la base de datos espacial: datos de mapas base de la base de datos Estonian Land Board; demográficos, de densidad de población, de lugar de residencia y lugar de trabajo de la base de datos del censo; domicilios estudiantiles del Sistema de Información Educativo de Estonia; datos actuales y futuros de usos del suelo del Plan Integral de Tartu; datos de GTFS de líneas de buses existentes; volumen de tráfico de automóviles.
- Cálculo y visualización del uso del transporte público a partir de datos de venta de billetes electrónicos. Este paso creó una matriz de origen-destino (O-D) del uso del transporte público. Los datos electrónicos fueron validados con datos sobre el conteo manual en buses.

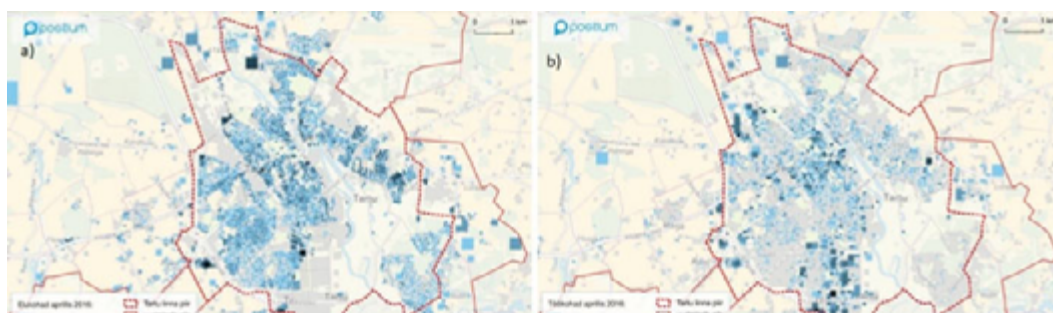


Figura 52: Lugar de residencia (a) y trabajo (b) estimados a partir de los datos de telefonía móvil.

- Cálculo del lugar de residencia y trabajo a partir de la muestra de 1 año de Mobile Positioning Data (MPD). Estos datos permitieron categorizar la población en trabajadores, viajeros frecuentes, usuarios en ruta, etc.
- Estimación de las necesidades de movilidad a partir de las matrices MPD: O-D para las horas del día, los días de semana y las estaciones.
- Cálculo de áreas de servicio de parada de autobús y ubicación óptima. Las mismas se calcularon a partir del principio de distancia mínima y se compararon con los datos de ubicación de los edificios y su funcionalidad.
- Creación de escenarios futuros de uso del suelo y cambio de población para la ciudad de Tartu.
- Seguidamente se realizaron talleres con diferentes partes interesadas para integrar en el diseño de la red las diversas opiniones y necesidades de los usuarios, incluyendo:
 - Un taller con representantes de grupos de usuarios para hacer un mapa de experiencias de usuario, deseos y sueños positivos y negativos.
 - Un taller con políticos, funcionarios del ayuntamiento, expertos en movilidad y empresas privadas comparando el servicio de transporte público existente con la información del taller del grupo de usuarios.

El siguiente paso fue el establecimiento de objetivos sobre el nivel de servicio para la ciudad de Tartu y las áreas circundantes. Este paso previo fue crucial para la planificación de la red de transporte público: todas las decisiones de diseño de la red se basan en los niveles de servicio necesarios, que se fijaron para los próximos 5 años.

WSP Finlandia llevó a cabo el modelo de simulación de la red de autobuses. Se modelaron 3 escenarios utilizando datos generados en los pasos del proceso de diseño anterior. El escenario 1 es uno de referencia que refleja las líneas de buses existente, con datos de uso actuales. Además, se crearon 2 escenarios alternativos de la red de buses (escenarios 2 y 3). La comparación de los 3 escenarios de red se basa en los parámetros de costos de operación, número de buses necesarios, velocidad, tiempos de viaje, accesibilidad y emisiones.

Beneficios

Como resultado del proceso de simulación, se propuso una nueva red de líneas de autobuses (figura 53) que consistía en un menor número de líneas, optimizando

las paradas y rediseñando algunas rutas de modo a dar respuesta a la actual localización de la demanda de viajeros. De este modo se asegura una eficiente cobertura de los lugares de residencia, de trabajo y otras actividades (figura 54). La reducción de la complejidad en el diseño de las rutas y del número de paradas permitió además aumentar la frecuencia de autobuses en las horas punta.



Figura 53: Propuesta de la nueva red de líneas de autobuses.



Figura 53: Propuesta de la nueva red de líneas de autobuses.

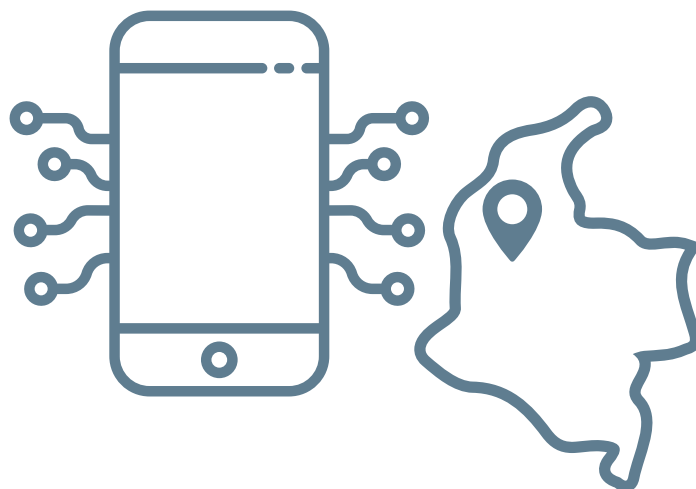
Esta nueva red se validó a través de una campaña de comunicación que recogió las opiniones de usuarios y los principales actores. Además, una aplicación web que permitió comparar las rutas calculadas con la red existente y la nueva red propuesta, además de diversos encuentros con grupos de especialistas. Algunas de las sugerencias fueron incorporadas a la propuesta final de la red de autobuses que se entregó a la empresa proveedora del servicio.

La metodología desarrollada trabaja con una muestra muy grande de datos, gracias al uso de datos móviles y cruce con otras fuentes de información complementarias. Tiene una gran cobertura tanto espacial, pudiendo recopilar potencialmente todos los flujos de movilidad, y temporal lo que permite analizar variaciones a lo largo del día, semana o meses. Todo ello repercute en un mayor conocimiento de la movilidad urbana y una mejor ordenación del sistema de transporte en nuestras ciudades.

The background of the entire page is a stylized, light blue map of an informal settlement. It features a complex, irregular grid of streets and blocks, with some areas appearing more densely packed than others. The lines representing the streets are of varying thicknesses, creating a sense of depth and movement.

8

Análisis de la movilidad en barrios informales a partir de datos de telefonía móvil



8.1 Accesibilidad, movilidad y exclusión social en barrios desfavorecidos

Los barrios informales, así como los asentamientos no planificados, no son consecuencia de procesos inmobiliarios, sino de la acción social. Se suelen desarrollar en localizaciones periféricas y se caracterizan por tener graves carencias de urbanización y déficits de servicios e infraestructuras (calles estrechas y generalmente no asfaltadas, ausencia de redes de alcantarillado, déficits de servicios sanitarios y educativos, entre otros.) y estar habitados por grupos sociales con bajos niveles de ingresos y de educación. La baja accesibilidad suele ser también una nota distintiva de estos barrios, ya que a su localización generalmente periférica se une una oferta de transporte muchas veces inadecuada e insuficiente.

Las deficientes condiciones de accesibilidad y de servicios de transporte se convierten en un factor que refuerza los procesos de exclusión social, conduciendo a procesos de guetificación. Por el contrario, la disponibilidad de servicios de transporte adecuados puede promover procesos inclusivos, al favorecer el acceso de los habitantes de estos barrios a otras partes de la ciudad, donde pueden interactuar con otros grupos de población, mejorar sus oportunidades de empleo y acceder a los servicios de los que carecen en sus propios barrios.

El análisis de la movilidad de los habitantes de los barrios informales permite detectar situaciones de baja integración social relacionadas con el transporte y la accesibilidad y, por lo tanto, aportar información útil a los planificadores para que se puedan adoptar acciones inclusivas en el campo del transporte o la dotación de servicios.

Latinoamérica cuenta con experiencias muy ilustrativas al respecto. Así, por ejemplo, el Metrocable de Medellín ha sido fundamental en la estrategia de integración social de la ciudad, al permitir que la población que vive en barrios informales de la periferia tenga un acceso mucho más fácil al centro de la ciudad,

incrementando así sus oportunidades de empleo y de acceso a los servicios. Al mismo tiempo, los barrios servidos por Metrocable han recibido inversiones para la provisión y mejora de infraestructuras y servicios públicos. El resultado final es que estos barrios informales están más integrados en la ciudad y se ha reducido el riesgo de exclusión social de sus habitantes.

Sin embargo, las características de la movilidad de los habitantes de los barrios informales son relativamente poco conocida. Las encuestas domiciliarias de movilidad no suelen cubrir adecuadamente estos barrios debido a las dificultades a las que se enfrentan los entrevistadores, tanto por las condiciones de inseguridad de muchos de estos barrios como por la desconfianza que muestran sus habitantes cuando se trata de abrir las puertas de sus hogares a los entrevistadores.

En cambio, la telefonía móvil sí es capaz de penetrar en estos barrios de forma efectiva. En contra de lo que podría esperarse a priori, la gran mayoría de los habitantes de los asentamientos informales utiliza teléfonos móviles, por lo que los registros almacenados por las compañías permiten analizar sus patrones de movilidad.

Por lo tanto, los datos de telefonía móvil constituyen una clara oportunidad para mejorar nuestro conocimiento sobre las pautas de movilidad de los habitantes de los barrios informales.





8.2 Movilidad y exclusión social en el barrio informal que rodea a Costa Esperanza (Buenos Aires)⁹

Un interesante ejemplo de estudio de movilidad a partir de datos de telefonía móvil en barrios informales es el llevado a cabo en 2018 por Telefónica Movistar y el BID en el asentamiento informal que rodea el Barrio Costa Esperanza, en el partido de San Martín, en el AMBA (Área Metropolitana de Buenos Aires). Se trata de un barrio de muy reciente desarrollo, sometido a un rápido proceso de crecimiento. Estructurado a través de un conjunto de calles angostas no asfaltadas, carece de redes de alcantarillado, agua corriente, electricidad y gas natural, así como de equipamientos públicos (escuelas, puestos sanitarios, policiales, de bomberos). Más de la mitad de su población percibe la Asignación Universal por Hijo o la asistencia del estado para niños con madre o padre desocupados. Los niveles de desocupación y empleo informal se sitúan en torno al 30%. La población total es de 21.903 habitantes en 2018, de los cuales 13.635 son mayores de 18 años.

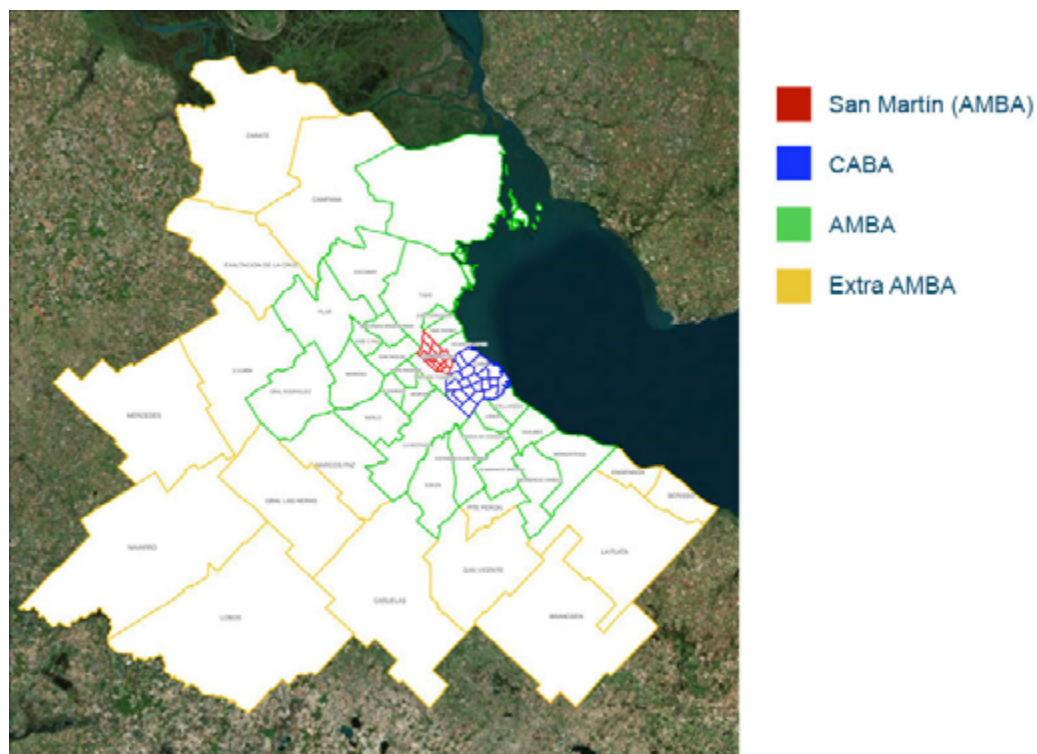


Figura 55: Propuesta de la nueva red de líneas de autobuses.

⁹ Estudio de Movilidad utilizando datos de telefonía móvil, realizado por el Banco Interamericano de Desarrollo y Ministerio de Transporte de Nación Argentina.



Figura 56: Detalle de la zonificación en el partido de San Martín. La zona 1 es el área de estudio.

El estudio se centró en la población mayor de 18 años. Los datos sobre los viajes y sus características fueron obtenidos a partir del análisis de los registros de telefonía móvil de agosto de 2018 completados con encuestas SMS. Se consideró residente a todo usuario que durante los seis meses anteriores al envío de la encuesta pasó la mayor parte de su tiempo en la zona de estudio, entre las 22hs y las 5hs. En cuanto a las características sociodemográficas de los usuarios, el nivel socioeconómico se infirió a partir de datos de facturación. El grupo de edad se obtuvo a partir de la información que registra la operadora sobre la fecha de nacimiento del cliente y el número de su DNI, y el género se pudo conocer a partir del nombre que figura en el contrato utilizando un algoritmo de clasificación automática.

La muestra fue extrapolada para ofrecer resultados representativos del conjunto de la población analizada. Los resultados fueron contrastados y completados con los de la tarjeta de transporte SUBE. Lo primero, para comprobar la validación de los resultados obtenidos; lo segundo, para añadir información complementaria, como por ejemplo las líneas de colectivos más utilizadas por la población del área de estudio o el grado de la utilización de la tarifa social, destinada a grupos desfavorecidos (jubilados, excombatientes, asignación universal por hijo o por embarazo, becas y pensiones no contributivas, entre otros.). El 41% de los viajes en colectivos se hicieron con tarifa social.

Los resultados obtenidos son reveladores de las especificidades de los patrones de movilidad de la población de los barrios informales y desvelan rasgos de una preocupante problemática social. La encuesta refleja que sólo un 45% de las personas hizo un viaje el día anterior y que la gran mayoría de los viajes declarados tienen como motivo el trabajo (61%), lo que es coherente con la distribución temporal de los viajes según los resultados obtenidos a partir de los registros de telefonía móvil, que presentan una punta muy marcada en la mañana para los viajes de ida y en la tarde para los viajes de vuelta.

Por lo tanto, una gran parte de la población permanece durante todo el día en la zona de estudio, o, lo que es lo mismo, no aprovecha las oportunidades de acceso al trabajo y a los servicios derivadas de la localización del barrio en una gran área metropolitana como la de Buenos Aires.

Por otra parte, la distribución de los viajes por motivos, con un gran predominio de los viajes al trabajo, también sugiere que la población en general se beneficia poco de los servicios localizados fuera del barrio.

Propósito del viaje

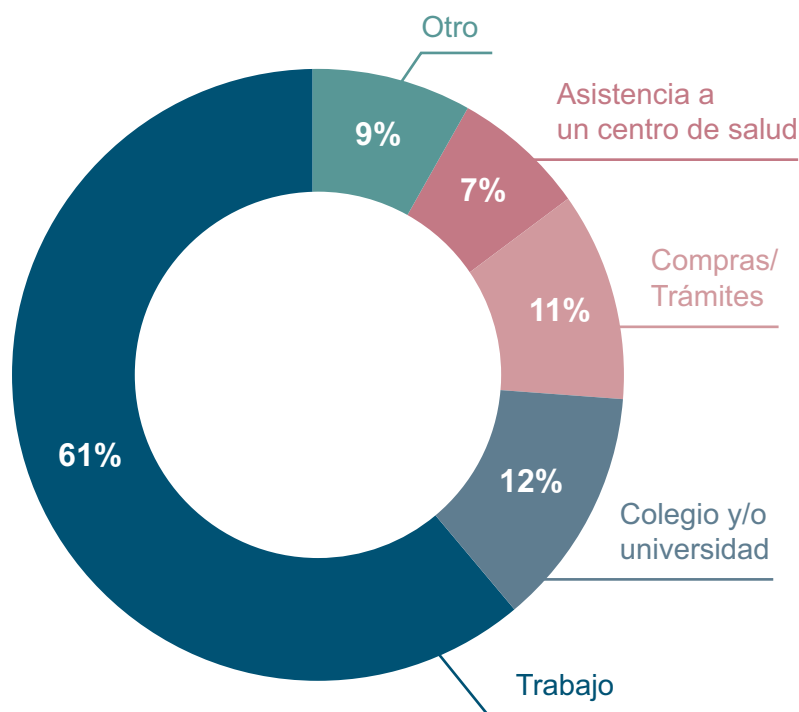


Figura 57: Motivos de viaje según los resultados de la encuesta.

Los encuestados que declararon no haber realizado ningún viaje el día anterior esgrimieron como razones más frecuentes que no tenían la necesidad, no tenían dinero o que ello fue debido a la presencia de barro e inundaciones en las calles (recuérdese que las calles no están asfaltadas y que el barrio no cuenta con red de alcantarillado).

La alta proporción de personas que no viajan (en una población en general muy joven) y las razones aducidas (no necesidad o falta de dinero) es consistente con el alto índice de desempleo observado en la zona de estudio y parece reflejar una clara situación de riesgo de exclusión social.

En coherencia con lo anterior, en la zona de estudio se observa un bajo número de viajes por persona, muy por debajo de la media del AMBA: los días laborables se realizan en promedio 1.88 viajes por persona y los fines de semana baja a 1.54, cuando en el resto de AMBA el promedio de viajes por persona en día laborable es de 2.64 y durante los fines de semana llega a 1.79. Estas circunstancias de baja movilidad afectan particularmente a las mujeres. De hecho, los hombres viajan en promedio un 6% más que las mujeres y esta diferencia se hace mayor durante los horarios pico (de 7 a 9hs y de 18 a 20hs) en los viajes largos, donde los hombres realizan hasta un 20% más viajes que las mujeres.

Como era de esperar, la mayor parte de los viajes de ida tiene como destino zonas próximas al barrio (mayoritariamente dentro del distrito de San Martín) y en el centro del área metropolitana (es decir, en la Ciudad Autónoma de Buenos Aires). En cuanto a las distancias recorridas, las personas con mayor poder adquisitivo y los hombres tienden a realizar viajes a destinos más lejanos; en cambio, los de nivel socioeconómico bajo y las mujeres se mueven más frecuentemente dentro de la zona en la que viven o en las zonas vecinas.



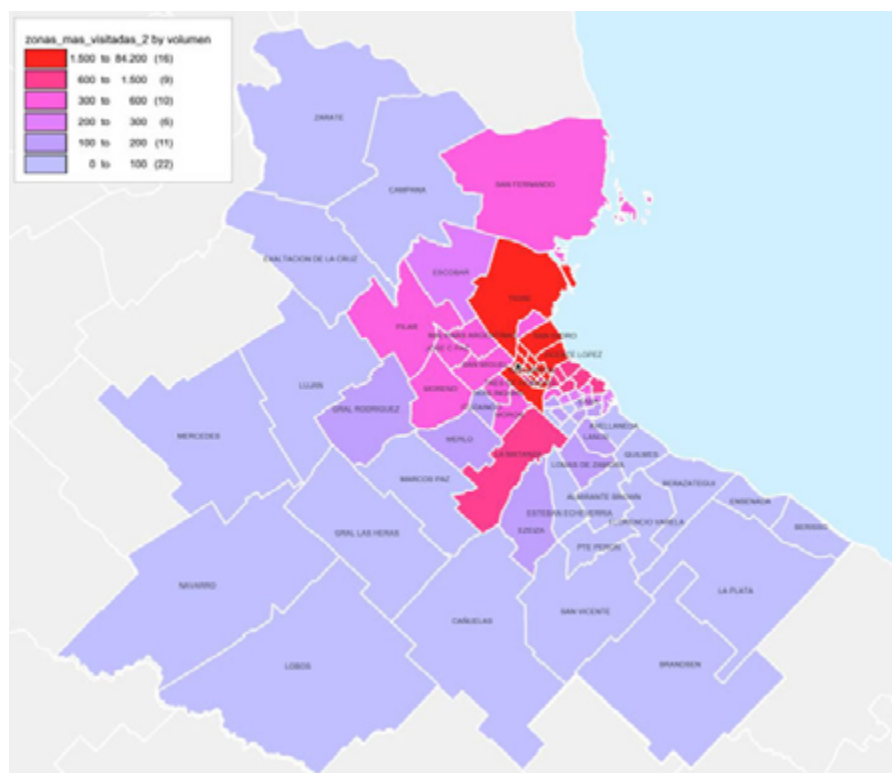


Figura 58: Destinos de los viajes según zonas de transporte.

Como era de esperar en un barrio con bajo nivel de ingresos, el modo de transporte más utilizado por los que realizan viajes con frecuencia es el colectivo (71%), mientras que el auto privado solo alcanza al 11% de los viajes y el ferrocarril el 7%. La seguridad es una de las principales preocupaciones para los que viajan. Un 33% afirmaron sentirse inseguros durante el recorrido, porcentaje que asciende a un 62% para los que viajan a pie. Durante la espera del transporte público, en la parada, el 46% de los encuestados respondió que se siente inseguro y un 24%, muy inseguro. En cuanto a la calidad de los servicios de transporte público, merece destacarse que mientras el recorrido y la ubicación de las paradas reciben una calificación neutra, los horarios, la calidad de las paradas y la ubicación de la recarga de la tarjeta SUBE reciben valoraciones claramente negativas. Esto muestra a los gestores de transporte un conjunto de puntos en los que se debería actuar para mejorar las condiciones de movilidad y de calidad de vida de la población residente en el barrio.

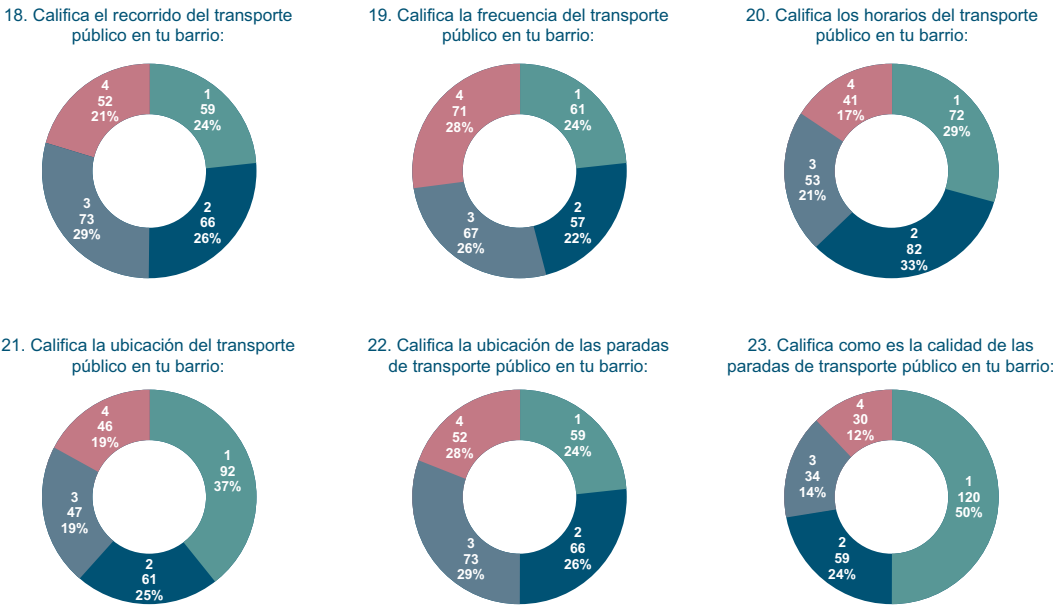


Figura 59: Calidad de servicio del transporte público según la encuesta, usando una escala donde “1” es muy malo y “4” es muy bueno.

En conclusión, el estudio sobre el asentamiento que rodea el Barrio Costa Esperanza, en el partido de San Martín, llevado a cabo con datos de telefonía móvil y encuestas SMS, junto con datos de la tarjeta de transporte SUBE, muestra las potencialidades de estas fuentes de datos, cuando utilizadas de forma combinada, para el análisis de los patrones de movilidad en barrios informales y la producción de información relevante para que las autoridades públicas puedan tomar de decisiones orientadas a mejorar la movilidad de la población.

A light blue background with a faint, stylized map pattern of streets and blocks.

9

**Ventajas y limitaciones.
Comparación con las encuestas de movilidad**



En este apartado se extraen las conclusiones principales sobre el uso de datos de telefonía móvil para el análisis de la movilidad urbana y, particularmente, para la obtención de matrices origen-destino de viajes. Como hemos visto, las oportunidades que proporcionan estas nuevas fuentes son muchas. A continuación, se resumen las principales ventajas que estos datos tienen para el estudio de la movilidad en la ciudad, ya sea como fuente independiente para el estudio de la movilidad o como complemento a los datos usados tradicionalmente procedentes de las encuestas domiciliarias. Sin embargo, frente a las encuestas y otros datos tradicionales en planificación del transporte, las nuevas fuentes de datos tienen también limitaciones. Al contrario que los tradicionales, este tipo de datos no han sido creados ni recogidos para el estudio de la movilidad; su naturaleza y su fin son otro (en el caso de los CDR es la facturación del servicio) y eso condiciona las posibilidades que nos ofrecen. Por lo tanto, tenemos que adaptar los nuevos datos a nuestro objetivo de estudio y realizar tareas y procesos que nos ayuden a superar o mitigar las limitaciones que presentan.

En cualquier caso, son muchas las ventajas y oportunidades que nos ofrece el Big Data en general y los datos de telefonía móvil en particular para el estudio de la movilidad urbana en nuestras ciudades. Veamos las más significativas:

Despliegue de la solución: Una de las ventajas de los datos de telefonía es que la recogida de los mismos no necesita de ningún tipo de despliegue de sensores adicionales a los ya instalados para la gestión de la red de la compañía telefónica que suministra los datos. Del mismo modo, tampoco es necesario que el usuario instale ningún nuevo software en su dispositivo móvil, o que active ciertas configuraciones de su dispositivo para la recogida de información. Frente a los costos derivados del despliegue de toda una serie de encuestadores que supone la recogida de datos tradicional o al contrario que en la información procedente de otro tipo de sensores (como los usados para el conteo de aforos), con los datos de telefonía no es necesario ningún tipo de actuación o infraestructura nueva. Además, a diferencia de las aplicaciones móviles, por ejemplo app para la recogida de datos con el GPS del móvil, con los datos de telefonía evitamos posibles reticencias de los usuarios a instalar dichas apps o problemas de consumo excesivo de la batería del teléfono.

Gran alcance espacial: Las redes de telefonía móvil están normalmente desplegadas por todo el territorio. Esto permite analizar la movilidad de los usuarios

tanto dentro como fuera del área de estudio definida, pudiendo caracterizar con bastante detalle el origen y destino final de los viajes.

Recogida pasiva de la información:

Uno de los elementos más importantes de este tipo de fuentes, en contraposición con las encuestas, es que en muchos de los casos, como lo es en la telefonía móvil, los datos se recogen sin necesidad de que el usuario tenga que responder un cuestionario o activar algún tipo de aplicación. Al recogerse de esta forma pasiva, los datos muestran lo que la gente hace, y no lo que la gente dice que hace.

Es lo que se conoce con el concepto de verosimilitud. En las encuestas, es frecuente que los encuestados tiendan a redondear ciertos datos (la edad, el salario, horas de salida o llegada de los viajes.) o a obviar cierto tipo de información (es el caso de ciertos viajes peatonales que en las encuestas de movilidad suelen estar infrarrepresentados). Respecto de las apps, con la recogida pasiva evitamos las posibles pérdidas de información de usuarios que olvidan entrar en la aplicación o activarla en su momento.

Grandes volúmenes de datos: Con algunas fuentes de Big Data, como las tarjetas inteligentes de transporte o de los sistemas de bicicletas públicas, podemos llegar a tener datos de todos los usuarios del sistema (compleción). En el caso de la telefonía, el tamaño de la muestra dependerá de la cuota de mercado de la operadora. En todo caso, las muestras de usuarios que se obtienen con los datos de telefonía son de un tamaño mucho mayor que en las encuestas de movilidad. Particularmente, los operadores con red propia suelen tener una cuota de mercado elevada. En algunos países de la región latinoamericana, las operadoras tienen cuotas de mercado muy altas. Es el caso por ejemplo de Telcel, con más del 60% de la cuota de mercado de telefonía móvil en México, de ANTEL en Uruguay, que concentra más del 50% de los usuarios del país, o Telefónica en Argentina o Brasil con cerca de un 35% de los usuarios de telefonía móvil.

Resolución temporal alta:

Los datos de telefonía móvil se recogen de manera continua (datos continuos), lo que permite tener una nuestra con información constantemente actualizada, realizar un análisis o monitorización de la movilidad en tiempo real o casi real, y estudiar secuencias temporales en la movilidad.

Así, es posible conocer los cambios en la movilidad a lo largo del día, de la semana, del mes o del año. Podemos incluso analizar la movilidad en periodos atípicos (como festivos o vacaciones) o durante determinados eventos urbanos.

Resolución espacial alta – media/alta: La información del nivel de detalle en la geolocalización de los datos varía, como hemos visto a lo largo del documento, llegando a ser muy precisa cuando trabajamos con datos de GPS o con registros telefónicos que han sido georreferenciados por triangulación. Cuando los datos de telefonía se refieren a la antena, la resolución es menor y presenta variaciones entre las zonas centrales de la ciudad, donde es alta, y las de la periferia urbana, con una resolución media. Pero en general, el nivel de detalle es alto y el número de unidades espaciales con las que trabajamos mayor a las zonificaciones de transporte habituales.

Costo y tiempo: Los costos que supone el uso del Big Data son mucho menores que los empleados en la realización de las encuestas de movilidad, habitualmente extremadamente costosas. También es mucho menor el periodo necesario para la elaboración de los análisis y la obtención de la información.

Como señalábamos, junto a estas ventajas, aparecen limitaciones y dificultades en el uso del Big Data en los estudios de movilidad urbana, en particular de los datos de telefonía móvil. Así debemos tener en cuenta elementos como:

Sesgos: Una de las críticas que se hace con frecuencia a las nuevas fuentes de datos asociadas al Big Data es el problema del sesgo, entendido como el posible error en los resultados de los análisis a partir de una muestra donde uno determinado grupo de población está infra o sobrerrepresentado. Sin embargo, el sesgo varía mucho según las fuentes de Big Data. Así:

- En algunas fuentes de datos no existe ningún tipo de sesgo, ya que podemos trabajar con información de toda la población a estudiar (compleción). Es el caso del estudio de la movilidad de los usuarios de transporte público o de sistemas de bicicletas públicas a partir de tarjetas inteligentes de transporte o de los registros de esos sistemas de bicicletas públicas.
- En otras fuentes el sesgo es bajo. Este es el caso de los registros de telefonía. En la actualidad, casi el ciento por ciento de la población tiene móvil y el sesgo únicamente dependerá, como hemos visto previamente, de las tasas de penetración de la compañía, las características de su mercado y las tareas de filtrado que realicemos en el pre-proceso de la información.
- Los sesgos aumentan (sesgo medio) en los datos procedentes de redes sociales, que tienen un uso de partida más reducido en el conjunto de la población y en muchos casos está más concentrado en determinados grupos sociodemográficos (por ejemplo entre la población más joven).

Almacenamiento y proceso de datos: La naturaleza del Big Data y su carácter de datos masivos condiciona las tareas de almacenamiento y proceso de datos, y obliga a usar tecnologías Big Data y personal técnico formado en el uso de estas tecnologías. Son necesarias plataformas y equipos que permitan almacenar los datos y hacerlo de forma continua. También es necesario el uso de técnicas Big Data para el procesamiento y análisis de los datos. Como hemos visto previamente, el uso de almacenamiento y computación en la nube o el trabajo mediante computación distribuida o paralelización de los procesos de cálculo son elementos fundamentales en el trabajo con este tipo de fuentes.

Dificultad para la limpieza y filtrado de datos: Cuando se trabaja con Big Data y con datos de telefonía hemos visto que es fundamental una primera etapa de limpieza y filtrado de los datos, que permita depurar los errores que aparecen en los mismos.

La limpieza de los rebotes entre torres en el caso de los datos de telefonía es, por ejemplo, un paso fundamental antes de cualquier análisis.

Información limitada sobre los usuarios: En algunos casos se puede disponer de información sociodemográfica básica del usuario a través de los contratos, como la edad o el género, pero por lo general esta información está limitada por motivos de privacidad. También se pueden inferir las características socioeconómicas de los usuarios a partir del consumo, tipo de terminal y/o el plan contratado. Otras veces se recurre, como hemos visto a procesos de enriquecimiento de datos ya sea de forma indirecta a través de los propios registros de los datos, por patrones de comportamiento (domicilio-trabajo) o por el uso de técnicas de análisis espacial en un Sistema de Información Geográfica. Asimismo, algunas operadoras realizan encuestas mediante mensajes SMS para obtener información sobre características de los usuarios.

Información limitada sobre ciertas características de los viajes: Si la información sobre el usuario es baja, también lo es la información que podemos obtener de forma directa sobre sus patrones de movilidad. Nuevamente, a través de procesos de enriquecimiento de datos es posible inferir información sobre la hora del viaje, el motivo, modo, ruta u otras características de los viajes. Será necesario un proceso de validación de los resultados, que permita conocer la calidad del proceso realizado. De forma general se puede afirmar que el motivo de viaje se recoge mejor en las encuestas de movilidad que en los estudios basados en CDRs, pero en cambio estos últimos registran la ruta seguida por el usuario, mientras que en los segundos la ruta en vehículo privado es desconocida y debe ser modelizada. En cuanto a la hora de salida y llegada del viaje y la duración del mismo, se obtiene una mejor aproximación en las encuestas (a pesar de los redondeos en que incurrir los encuestados). Por lo que se refiere al tiempo de viaje, el tiempo de paso entre torres de telefonía (puntos intermedios de la ruta) es preciso, pero la dificultad de estimar la hora de salida y de llegada resta fiabilidad al tiempo total del viaje. Algunas operadoras realizan encuestas mediante mensajes SMS para obtener información adicional sobre las características de los viajes.

Accesibilidad a los datos: Cuando trabajamos con Big Data el primero de los elementos a considerar es el acceso a los datos. En la mayor parte de los casos, como es el de la telefonía, dependemos de empresas propietarias de los mismos. Será necesario conocer las políticas de las operadoras sobre la forma de acceso a los datos. Algunas operadoras no permiten el acceso a sus datos a terceros y ellas mismas realizan los servicios de consultoría para estudios de movilidad. Otras operadoras, en cambio, no participan directamente en el negocio de consultoría, pero tienen acuerdos con empresas consultoras especializadas mediante los cuales éstas pueden entrar en sus servidores y procesar los datos.

En resumen, tanto las encuestas de movilidad como los datos de telefonía móvil presentan ventajas e inconvenientes (Cuadro 1).

A priori no se puede afirmar de forma general que los datos de telefonía móvil vayan a sustituir a los de las encuestas de movilidad. Si una ciudad no dispone de medios

económicos para hacer una encuesta domiciliaria de movilidad, la utilización de CDRs puede ser una buena alternativa. Si existen recursos económicos suficientes, y se dispone ya de una encuesta de movilidad, la actualización de los datos de la encuesta con datos de telefonía móvil puede ser una buena estrategia. En ciertos casos bastará con utilizar datos de telefonía móvil, por ejemplo para conocer el impacto de un evento en la movilidad de la ciudad.

Cuadro 1: Comparación sintética entre las encuestas domiciliares de movilidad y los datos de telefonía móvil (CDRs) en estudios de movilidad urbana

| | ENCUESTA DOMICILIARIA DE MOVILIDAD | DATOS DE TELEFONÍA MÓVIL (CDRs) |
|---|--|---|
| Tamaño de la muestra | Del orden de varias decenas de miles de encuestados en ciudades de varios millones de habitantes | Unas diez veces mayor que el de las encuestas |
| Sesgo | El inherente a las encuestas Es difícil acceder a determinados colectivos | Dependiente de la cuota de mercado de la operadora, su implantación según grupos de población y el proceso de selección de la muestra Es posible llegar a colectivos a los que no se accede con la encuesta (barrios informales y áreas de muy alto nivel de ingresos) |
| Coste | 10 veces mayor (aprox.) | 10 veces menor (aprox.) |
| Tiempo | 5 veces mayor (aprox.) | 5 veces menor (aprox.) |
| Actualización de datos | Cada 7-8 años (aprox.) | Constante |
| Días o periodos atípicos | No. Solo movilidad en día laborable medio | Sí. Movilidad en día laborable medio y en días y periodos atípicos |
| Cobertura geográfica y poblacional | El área de estudio (área urbana/metropolitana) Se registran solo los viajes de la población residente en el área de estudio | Todo el país Se registran también los viajes de la población no residente, particularmente conmuters con residencia fuera del área de estudio y turistas |
| Datos sobre viajeros | Información rica Diferentes variables sociodemográficas (según cuestionario) | Información inicialmente limitada. Posibilidad de enriquecerla mediante: - Tipo de contrato y consumo - Encuestas SMS - Cruce con datos sociodemográficos de las secciones censales o ZT |
| Número de viajes | Los que declarara el encuestado Viajes peatonales e infrecuentes están infrarrepresentados Solo viajes de residentes | Los que se infieren a partir de los CDR Viajes muy cortos (dentro del mismo polígono Voronoi) o con actividades en destino muy breves no son registrados. Viajes de residentes y no residentes |
| Localización del origen y el destino | Localización más precisa A nivel de zona de transporte A nivel de dirección postal Coordenadas xy | Localización menos precisa A nivel de polígonos Voronoi (área de influencia de la torre) No es posible ajustarla a nivel de dirección postal Coordenadas xy (con cierto margen de error) solo en datos triangulados |
| Motivo | Información detallada del motivo, según las categorías del cuestionario | Información poco detallada: domicilio, trabajo, otros. Inferencia de otros motivos en algunos casos. Posibilidad de obtener Información adicional con encuestas SMS. |

| | | |
|---------------------------------|---|--|
| Modo | Información detallada del modo, según las categorías del cuestionario | Dificultad para inferir el modo en áreas urbanas. Se puede inferir a partir del tipo de antena, la ruta, el tiempo de viaje... Posibilidad de obtener Información adicional con encuestas SMS |
| Ruta | Transporte privado. Ruta no registrada en la encuesta Transporte público. Ruta reproducible a partir de las líneas utilizadas (cuando esta información se recoge en la encuesta) | Queda registrada de forma aproximada a partir de los polígonos Voronoi atravesados en el viaje. Se puede ajustar a partir de cálculos de caminos mínimos. |
| Hora de salida y llegada | Las declaradas por el encuestado Errores de estimación del encuestado y de redondeo | Se infieren a partir de la hora final y la hora de inicio de estancias sucesivas y del tiempo estimado del viaje Errores mayores que en la encuesta |
| Tiempo de viaje | Diferencia entre las horas de salida y de llegada declaradas por el encuestado Errores de estimación del encuestado y de redondeo | Diferencia entre las horas de salida y de llegada inferidas Error mayor que en la encuesta |
| Etapas de viaje | Las declaradas por el encuestado Normalmente algo infrarrepresentadas | Las permanencias de menos de 30 minutos en una misma localización pueden ser interpretadas como paradas entre etapas Información menos fiable y más infrarrepresentada que en las encuestas |
| Motivaciones del viajero | Se registran mediante preguntas específicas (por ejemplo, ¿por qué no utilizó el auto en este viaje?) | No es posible conocerlas |



The background of the page is a light blue-grey color with a white line pattern that resembles a stylized map or a network of streets and paths. The lines are of varying thickness and form a complex, interconnected web across the entire page.

10

Consideraciones finales

Es evidente que a la hora de estudiar la movilidad lo mejor sería poder conjugar las ventajas de los datos procedentes de las fuentes tradicionales con las nuevas fuentes de datos, buscando la complementariedad entre los resultados obtenidos a partir de registros de teléfonos móviles y los análisis que se pudiesen realizar con información complementaria, como encuestas origen-destino y otras de transporte (flujos, preferencias declaradas de modos, etc.), encuestas de ingresos/pobreza, censos y de empleo, entre otras.

Para obtener los mayores beneficios de las nuevas fuentes de datos será necesario modificar la composición de los equipos de planificación. Los planificadores del transporte no cuentan con la formación necesaria para trabajar con datos masivos y los analistas de datos masivos no tienen conocimientos en planificación del transporte. Parece claro que estos equipos futuros deberán incorporar expertos en análisis de Big Data y que cada perfil (el planificador y el analista de datos) debe hacer un acercamiento al campo del otro, para poder llegar a alcanzar una colaboración fructífera. El planificador debe tener conocimientos básicos sobre las fuentes de datos Big Data y los procedimientos de análisis de datos masivos, conociendo sus ventajas y limitaciones, para poder obtener el mayor provecho de las oportunidades que brindan las nuevas fuentes de datos. Y el analista de datos masivos debe entender de forma básica el proceso de planificación del transporte y los modelos para poder interactuar adecuadamente con el planificador. No se trata solo de datos, sino también de metodologías.

Está claro que los datos de telefonía móvil permiten que el planificador pueda disponer de información siempre actualizada para poder alimentar los modelos de transporte.

Pero, además, las herramientas de inteligencia artificial ofrecerán al planificador nuevas oportunidades para dar solución a los viejos problemas de planificación.

En la región de América Latina y el Caribe se parte de una situación de cierto retraso respecto de Europa en el uso de datos de telefonía móvil en el estudio de la movilidad urbana y la planificación del transporte, pero recientemente se han dado pasos muy esperanzadores tanto en las universidades como en las administraciones gubernamentales. La ventaja para la región es que las metodologías para el tratamiento de datos de telefonía móvil para analizar la movilidad urbana están bastante maduras y por lo tanto su aplicación en la región puede ser mucho más rápida y directa. Una de las claves para su difusión es, sin duda, que los planificadores de las ciudades de la región conozcan las nuevas fuentes de datos, sus ventajas y sus limitaciones.

La movilidad del futuro requerirá y exigirá cuadros técnicos multidisciplinarios y especializados en el manejo de datos. Este documento intenta contribuir al objetivo de mejorar la capacitación de los cuadros técnicos para aprovechar las ventajas de las nuevas fuentes de datos en la planificación del transporte. En el anexo II se enumeran las acciones que deben seguirse para implementar con éxito análisis de la movilidad en las ciudades latinoamericanas utilizando los datos de telefonía móvil, se identifican los requerimientos a las empresas y consultoras a las que se va a contratar, y se detallan las especificaciones que las mismas deben incluir en los informes finales sobre sus estudios de movilidad.

Glosario de abreviaturas utilizadas

API - Application Programming Interface. Una interfaz de programación de aplicaciones es un conjunto de rutinas que provee acceso a un determinado software. En la web las APIs permiten la descarga de datos de forma relativamente fácil.

BTS - Acrónimo de Base Transceiver Station. Son las torres de telefonía móvil. Una torre contiene varias antenas de telefonía, que cubren un ángulo de 120 grados, de forma que tres antenas cubren de forma conjunta los 360 grados.

CDR - Call Detail Records. Son los registros de las actividades realizadas por los usuarios de los teléfonos móviles. Incluyen llamadas y mensajes. De forma genérica también se incluyen las sesiones de datos bajo esta categoría.

GPS - Global Positioning System. Es un sistema que permite determinar la localización sobre la superficie terrestre utilizando un dispositivo específico (receptor GPS) o simplemente el dispositivo utilizado para este fin en los teléfonos inteligentes.

ID - Identificador. Los registros activos (CDR) registran el identificador del usuario saliente y entrante y de la torre entrante y saliente.

IMD - Intensidad Media Diaria de Tráfico. Expresa el número de vehículos que transitan como media al día por un determinado tramo.

LA - Acrónimo de Location Area. Para optimizar las señales de telefonía, las celdas se agrupan en zonas mayores para las que se guarda un código de Localización de Área Común (LAC), que identifica la zona de ubicación de la torre. Cada LA suele tener entre 10 y 100 torres.

LAC - Location Area Code. Es el código de identificación de cada LA (ver LA).

OD - Origen-Destino. Se utiliza este acrónimo para aludir a las matrices que registran el número viajes para cada par origen y destino.

SIG - Sistema de Información Geográfica.

Sistema informáticos diseñado para la captura, almacenamiento, manipulación, análisis, modelización y presentación de datos geolocalizados para la resolución de problemas planificación y gestión.

SMS - Short Message Service. Servicio disponible en los teléfonos móviles que permite el envío de mensajes cortos

xDR - Son los registros de las sesiones de datos

2G, 3G, 4G - Segunda, tercera y cuarta generación de tecnologías de telefonía móvil. El despliegue de las redes de distintas generaciones da lugar a diferentes características y funcionalidades.

Bibliografía

- Alexander, L., Jiang, S., Murga, M., & González, M. C. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation research part c: emerging technologies*, 58, 240-250.
- Bahoken, F., & Olteanu-Raimond, A. M. (2013, August). Designing origin-destination flow matrices from individual mobile phone paths: The effect of spatiotemporal filtering on flow measurement. In *ICC'13-26th International Cartographic Conference* (p. 15p).
- Bar-Gera, H. (2007). Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel. *Transportation Research Part C: Emerging Technologies*, 15(6), 380-391.
- Bonnel, P., Hombourger, E., Olteanu-Raimond, A. M., & Smoreda, Z. (2015). Passive mobile phone dataset to construct origin-destination matrix: potentials and limitations. *Transportation Research Procedia*, 11, 381-398.
- Calabrese, F., Di Lorenzo, G., Liu, L., & Ratti, C. (2011). Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area. *IEEE Pervasive Computing* 10, no. 4 (April 2011): 36–44. <http://dx.doi.org/10.1109/mprv.2011.41>
- Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira Jr, J., & Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation research part C: emerging technologies*, 26, 301-313.
- Calabrese, F., Ferrari, L., & Blondel, V. D. (2015). Urban sensing using mobile phone network data: a survey of research. *Acm computing surveys (csur)*, 47(2), 25.
- Çolak, S., Alexander, L. P., Alvim, B. G., Mehndiratta, S. R., & González, M. C. (2015). Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities. *Transportation Research Record*, 2526(1), 126-135.
- Chen, C., Bian, L., & Ma, J. (2014). From traces to trajectories: How well can we guess activity locations from mobile phone traces?. *Transportation Research Part C: Emerging Technologies*, 46, 326-337.
- Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation research part C: emerging technologies*, 68, 285-299.
- Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation research part C: emerging technologies*, 68, 285-299.
- Frias-Martinez, V., Soguero, C., & Frias-Martinez, E. (2012, August). Estimation of urban commuting patterns using cellphone network data. In *Proceedings of the ACM SIGKDD international workshop on urban computing* (pp. 9-16). ACM.
- Gundlegård, D. (2018). *Transport Analytics Based on Cellular Network Signalling Data* (Vol. 1965). Linköping University Electronic Press. https://books.google.es/books?hl=es&lr=&id=IdV7DwAAQBAJ&oi=fnd&pg=PP4&ots=qT7f-cFoVe&sig=_dUHrNiPVVtli77KQ9FkIBrFM2A#v=onepage&q&f=false
- Hadachi, A., & Lind, A. (2019). Exploring a New Model for Mobile Positioning Based on CDR Data of The Cellular Networks. *arXiv preprint arXiv:1902.09399*.
- Holleczeck, T., Yu, L., Lee, J. K., Senn, O., Ratti, C., & Jaillet, P. (2014). Detecting weak public transport connections from cellphone and public transport data. In *Proceedings of the 2014 International Conference on Big Data Science and Computing* (p. 9). ACM.

- Iqbal, M. S., Choudhury, C. F., Wang, P., & González, M. C. (2014). Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, 40, 63-74.
- Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., & Varshavsky, A. (2011). Identifying important places in people's lives from cellular network data. In *International Conference on Pervasive Computing* (pp. 133-151). Springer, Berlin, Heidelberg.
- Järv, O., Tenkanen, H., & Toivonen, T. (2017). Enhancing spatial accuracy of mobile phone data using multi-temporal dasymetric interpolation. *International Journal of Geographical Information Science*, 31(8), 1630-1651.
- Jiang, S., Fiore, G. A., Yang, Y., Ferreira Jr, J., Frazzoli, E., & González, M. C. (2013). A review of urban computing for mobile phone traces: current methods, challenges and opportunities. In *Proceedings of the 2nd ACM SIGKDD international workshop on Urban Computing* (p. 2). ACM.
- Lenormand, M., Picornell, M., Cantú-Ros, O. G., Tugores, A., Louail, T., Herranz, R., ...& Ramasco, J. J. (2014). Cross-checking different sources of mobility information. *PloS one*, 9(8), e105184.
- Milne, D., & Watling, D. (2019). Big data and understanding change in the context of planning transport systems. *Journal of Transport Geography*, 76, 235-244.
- Picornell, M., Ruiz, T., Lenormand, M., Ramasco, J. J., Dubernet, T., & Frías-Martínez, E. (2015). Exploring the potential of phone call data to characterize the relationship between social network and travel behavior. *Transportation*, 42(4), 647-668.
- Ricciato, F., Widhalm, P., Craglia, M. y Pantisano, F. (2015): Estimating population density distribution from network-based mobile phone data. ISPR, European Commission, Joint Research Centre, Institute for Environment and Sustainability
- Tettamanti, T., Demeter, H., & Varga, I. (2012). Route choice estimation based on cellular signaling data. *Acta Polytechnica Hungarica*, 9(4), 207-220.
- Tolouei, R., Psarras, S., & Prince, R. (2017). Origin-destination trip matrix development: Conventional methods versus mobile phone data. *Transportation research procedia*, 26, 39-52.
- Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A., & González, M. C. (2015). The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*, 58, 162-177.
- Wang, Z., He, S. Y., & Leung, Y. (2018). Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society*, 11, 141-155.
- Widhalm, P., Yang, Y., Ulm, M., Athavale, S., & González, M. C. (2015). Discovering urban activity patterns in cell phone data. *Transportation*, 42(4), 597-623.
- Wismans, L. J. J., Friso, K., Rijdsdijk, J., de Graaf, S. W., & Keij, J. (2018). Improving A Priori Demand Estimates Transport Models using Mobile Phone Data: A Rotterdam-Region Case. *Journal of Urban Technology*, 25(2), 63-83.

The background of the entire page is a light blue-grey color with a faint, stylized map pattern. The map consists of a network of thin, light grey lines representing streets and roads, forming a complex grid-like structure with various rectangular and irregular shapes.

ANEXO I

Otras experiencias de estudios de movilidad a partir de datos de telefonía móvil en América Latina y Europa



Estudio de movilidad de los usuarios del Ferrocarril Línea Roca (Ramal Constitución – La Plata) & Línea C de Subterráneo. Una mirada de Género¹⁰

Sobre el caso de estudio

Entre los principales problemas a los que se enfrentan los estudios de movilidad a partir de datos de telefonía móvil figuran la identificación del modo de transporte y la captura de información relativa a las condiciones en que los usuarios realizan sus viajes. El presente estudio analiza la movilidad en la línea de Ferrocarril Roca, en el tramo que va desde la Plaza de la Constitución (Buenos Aires) hasta la ciudad de La Plata. En esta línea el tiempo de viaje es de 70 minutos a través de 19 estaciones. Además, se analiza la Línea C del subte de Buenos Aires, que cuenta con una extensión de 4,4 km y un total de 9 estaciones.



Figura 1: Área de estudio.

En la línea de ferrocarril se incidirá sobre las tres estaciones de transferencia importantes, señaladas en la Figura 1 con un círculo verde, que son La Plata, Berazategui y Constitución, donde se permite hacer el trasbordo a la línea C. Mientras que en la línea C de Subte, se analizan todas las estaciones.

¹⁰ Estudio de Movilidad utilizando datos de telefonía móvil, realizado por el Banco Interamericano de Desarrollo y Ministerio de Transporte de Nación Argentina.

El desafío

Se pretende conocer la demanda de pasajeros en días hábiles en diferentes sistemas de transporte, incluyendo el ferrocarril Roca, ramal Constitución – La Plata y la línea C del Subte. Asimismo, se propone una segmentación de la demanda por perfil de género, nivel socioeconómico y etario, diferenciando además por franjas horarias. Los resultados del estudio servirán para actualizar y complementar las estadísticas existentes sobre el flujo de viajeros en estas líneas. Los datos de telefonía móvil se utilizan para determinar una serie de métricas relacionadas con la movilidad. Cuestiones como el volumen medio de personas en la región; su variación según franjas horarias y según el día de la semana; segmentación del flujo por edad, género y nivel económico; lugares de origen y destino de los viajes; o la diferencia entre flujos internos y externos, son algunos de los datos que permitirán obtener un mayor conocimiento de la demanda de estas líneas.

Uno de los aspectos novedosos del estudio es su perspectiva de género. Para ello se propone un estudio comparativo de las pautas de movilidad entre hombre y mujeres, así como un análisis centrado en la identificación de posibles situaciones de acoso sufridas por las mujeres en el transporte público.

La solución

La región de análisis se divide en 12 zonas localizadas alrededor de las estaciones señaladas anteriormente. La base de datos utilizada corresponde a registros de eventos de red generados por teléfonos conectados a la red de Telefónica en Argentina, durante agosto de 2018. Estos datos se cruzan con otras fuentes como son los Censos, los datos geoespaciales de las zonas de interés o los provenientes de la tarjeta de transporte SUBE.

En lo que se refiere a la zonificación, para el caso de las estaciones de ferrocarril, las zonas se definen como las áreas de cobertura de las antenas que dan servicio a la estación (Figura 2). En el caso del Subte, donde existen microantenas que dan servicio exclusivo dentro de cada estación, las zonas respetan el límite interno de cada estación.

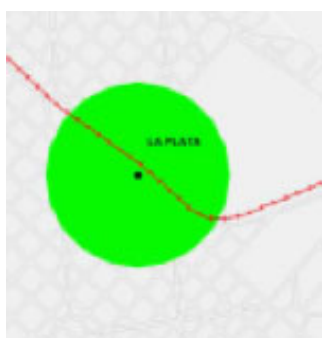


Figura 2: Zona La Plata.

Los registros son anonimizados, sustituyendo el número del teléfono por un identificador único, y se agregan a nivel de zona. Una vez agregados, los datos son extrapolados hasta alcanzar la población total mayor de 15 años en la zona en cuestión sobre la base de los datos del Censo. Finalmente, se segmenta la población por género, rangos de edad y nivel socioeconómico. Se estima su lugar

de origen y destino, siendo el origen la localidad donde residen y el destino la zona donde hacen la primera parada mayor a 30 minutos. Se identifican todos los viajes de la muestra con origen y destino en las zonas de estudio, incluyendo aquellos desplazamientos con origen o destino fuera de las zonas de interés. De este modo, es posible conocer los viajes internos, como los que ocurren dentro de una misma zona, y los viajes externos, que equivalen a todos los demás. También se han diferenciado los viajes que hacen trasbordo de la línea de subterráneo a la línea de tren o al revés.

Para ahondar en las cuestiones relaciones con las posibles agresiones sufridas por las mujeres en el transporte público, o bien mientras accedían a él, el análisis se complementa a través de la solución Movistar+ que permite realizar una encuesta por SMS a costo 0 para los usuarios. Las encuestas se lanzaron entre el 15 de noviembre de 2018 y el 22 de enero de 2019. Los SMS con la encuesta se enviaron a usuarios de Movistar de sexo femenino que, a través de los datos de las antenas, recibieron la clasificación de usuarios frecuentes del Ferrocarril Roca (ramal La Plata-Buenos Aires) y de la línea C de subterráneos. El desarrollo y diseño de la encuesta estuvo a cargo del BID en conjunto con el Ministerio de Transporte.

Beneficios

El estudio mide un volumen medio de 200.000 viajes en los días laborables y 100.000 en los fines de semana. El Subte registra un mayor número de usuarios respecto del tren. En ambos casos, los mayores picos de usuarios en días laborables suelen ocurrir entre las 5 y las 6 de la tarde y las 7 y las 9 de la mañana. En términos de distancia, la mitad de los usuarios recorre de media más de 1.400 m desde su casa hasta la estación de tren o Subte más cercana.

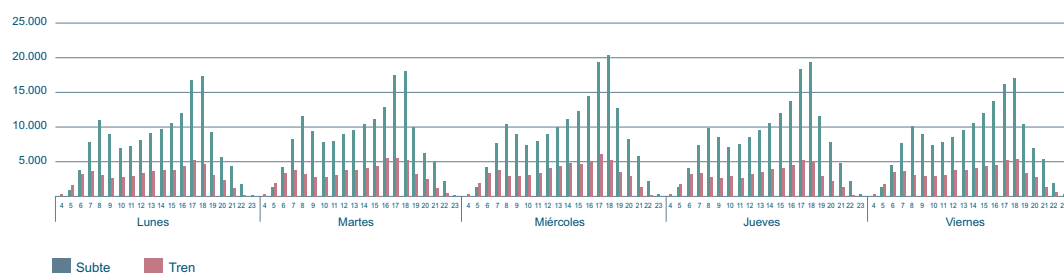


Figura 3: Volumen de usuarios por día, hora y modo de transporte.

Las estaciones de Diagonal Norte, Plaza de la Constitución, Av. De Mayo, La Plata e Independencia son las más demandadas respecto de los lugares de origen de los viajes, mientras que las estaciones de Constitución, Av. de Mayo, Diagonal Norte son las más demandadas respecto de los sitios de destino.

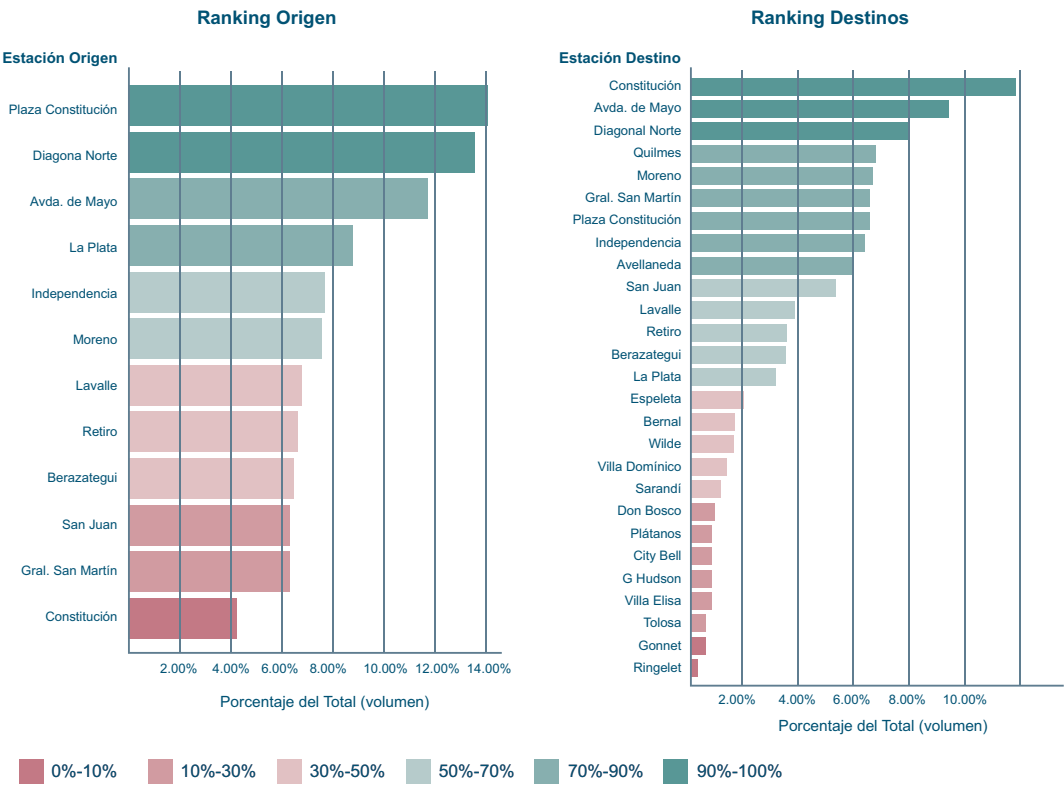


Figura 4: Ranking de estaciones de origen y destino.

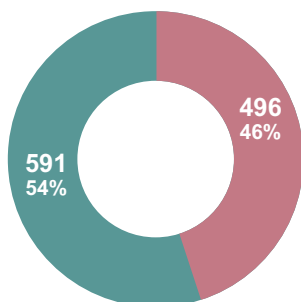
Para las tres estaciones seleccionadas del Ramal Roca (Berazategui, La Plata y Plaza de la Constitución), se ha analizado el flujo de viajeros a lo largo del día y se encontraron importantes diferencias entre ellas. En el caso de Berazategui el mayor flujo de viajes con origen en esta estación ocurre entre las 5 y las 9 de la mañana, mientras que en las otras dos estaciones ese flujo se distribuye de un modo más uniforme, incrementando sobre todo en las horas de la tarde. El estudio permite determinar para cada una de estas estaciones, sus estaciones de destino más frecuentes.

Los datos de la tarjeta de transporte SUBE se han utilizado para validar los resultados obtenidos previamente a través de los datos de telefonía móvil. Para ello, se compararon los volúmenes de viajeros en las diferentes estaciones del Subte y del ferrocarril, flujos de viajes origen-destino, así como volumen total según día y horario. Globalmente ambas fuentes obtienen pautas de movilidad muy similares, con algunas pequeñas variaciones debidas sobre todo a diferencias metodológicas en la captación de los datos.

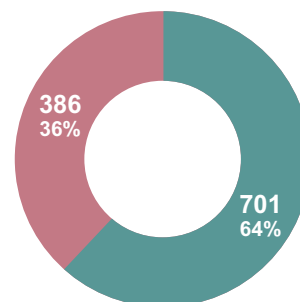
Los usuarios masculinos de las líneas estudiadas superan en cerca del 10% a los femeninos. En cambio, las mujeres suelen recorrer más distancia para llegar a la estación de origen de su viaje. En términos de estaciones de destinos preferidas, hombres y mujeres presentan patrones similares.

Asimismo, se ha realizado una encuesta de acoso, utilizando la metodología de Smart Steps, mediante SMS, a 1200 mujeres mayores de edad usuarias del ferrocarril. Se detalla a continuación los siguientes resultados:

Mujeres que sufrieron acoso sexual **mientras esperaban** el transporte público **y/o caminaban hacia/desde** una parada



Mujeres que sufrieron acoso sexual **cuando caminaban hacia/desde** una parada



■ Sufrieron acoso ■ No sufrieron acoso

Figura 5: Casos de acoso entre las encuestadas.

Los resultados de la encuesta Movistar+ referente a posibles situaciones de acoso a mujeres indican que el 46% de las encuestadas ha sufrido acoso en el transporte público, mientras un 36% lo ha sufrido mientras caminada hacia/desde alguna parada.

El mayor número de casos de acoso ocurre en las horas pico y el periodo nocturno. Diferenciando entre medios de transporte, el tren registra el mayor número de casos, seguido del subte y del colectivo. En relación al tipo de acoso, la mayor parte de los casos (74%) se refiere a agresiones sin contacto físico, mientras que el 26% restante son con contacto físico. Estas últimas ocurren sobre todo por la mañana mientras que las primeras predominan en horario de tarde. La principal respuesta de las mujeres ante a una situación de acoso es cambiar de recorrido o de calle. A pesar del alto porcentaje de mujeres que han sido víctimas de acoso, solo un 6% de ellas dice haber denunciado. Este resultado puede indicar la necesidad de afianzar mecanismos formales para el registro y atención de las usuarias.





Exposición de la población a gases contaminantes (NO₂) a partir de datos de telefonía móvil¹¹

Sobre el caso de estudio

Actualmente, la contaminación atmosférica en general y la concentración de NO₂ es una de las principales preocupaciones de numerosas ciudades en todo el mundo, entre ellas muchas latinoamericanas. México DF y Santiago de Chile son buenos ejemplos de ciudades latinoamericanas particularmente afectadas por la contaminación atmosférica.

El transporte es el principal responsable de la concentración de NO₂ en las ciudades y por lo tanto las principales medidas tomadas por las autoridades están relacionadas con la gestión de la movilidad.

En días de alta contaminación se toman medidas del tipo de prohibición de circulación a los vehículos más contaminantes, a las matrículas pares o impares, limitación de la velocidad de circulación, etc.

El estudio analiza la exposición de la población a la contaminación en el área metropolitana de Madrid, en un día típico laborable. Incluye además las provincias colindantes a esta región (Segovia, Ávila, Toledo, Guadalajara y Cuenca) para tener en cuenta los flujos de viajes desde estas provincias hacia la capital.



Figura 6: Casos de acoso entre las encuestadas.

¹¹ Picornell M., Ruiz T., Borge R., García-Albertos P., de la Paz D., Lumbreras J. (2018) Population dynamics based on mobile phone data to improve air pollution exposure assessments. Journal of exposure science & environmental epidemiology. <https://doi.org/10.1038/s41370-018-0058-5>.

El desafío

El análisis de la cantidad de población expuesta a diferentes niveles de contaminación se ha realizado tradicionalmente con datos provenientes de fuentes estáticas con pueden ser los censos. Sin embargo, la movilidad diaria de la población determina una diferente exposición a los niveles de contaminación según horas del día que es necesario cuantificar para una correcta mitigación de sus efectos negativos. El estudio pretende estimar la dinámica diaria de la población sobre la base de datos provenientes de telefonía móvil y cruzar esa información con datos de concentración de NO₂, también dinámicos, en la ciudad de Madrid. Finalmente se pretende comparar los resultados de la metodología desarrollada con los obtenidos a partir de datos del censo, para así determinar el sesgo incurrido cuando se utilizan fuentes de datos estáticas.

A nivel metodológico el estudio afronta dos retos principales, por un lado, se pretende mejorar la estimación de la localización de los usuarios y superar el enfoque que asocia su localización a la torre a la que está conectado el usuario. Por otro, se profundiza en las técnicas de expansión de la muestra de datos mejorando la estimación de la población expuesta a los gases contaminantes.

La solución

El estudio desarrolla una metodología para estimar la dinámica de la población en lo que se refiere a las actividades y viajes a lo largo del día. El proceso lo dividen en cuatro pasos: 1) identificación del lugar de residencia; 2) determinación de las actividades y los viajes; 3) extrapolación de toda la población; 4) extracción de la dinámica de la población.

- 1) El lugar de residencia se identifica como la localización más frecuente entre las 8 p.m. y las 7 a.m. de los días laborables de octubre y noviembre de 2014.
- 2) Los lugares de actividad se definen como aquellos donde el usuario pasa algún tiempo mientras que los viajes se definen como los desplazamientos entre dos actividades. Los datos utilizados se refieren al 17 de noviembre de 2014, como proxy a un típico día laborable en términos de movilidad y concentraciones de NO₂.

Una vez eliminados los usuarios con periodos largos de tiempo con poca información (8h para el periodo nocturno de 8 p.m. a 7 p.m. y 4h para el periodo diurno de 7 a.m. a 8 p.m.), se identifican todas las actividades llevadas a cabo por los usuarios. Se parte de un límite de tiempo previamente establecido (30 min), a partir del cual se considera que el usuario realiza una actividad. Luego se determina el viaje, estimando el tiempo de origen del viaje (cuadro 1). La figura 7 muestra un ejemplo de cómo los datos móviles se transforman en lugares de actividad y viajes.

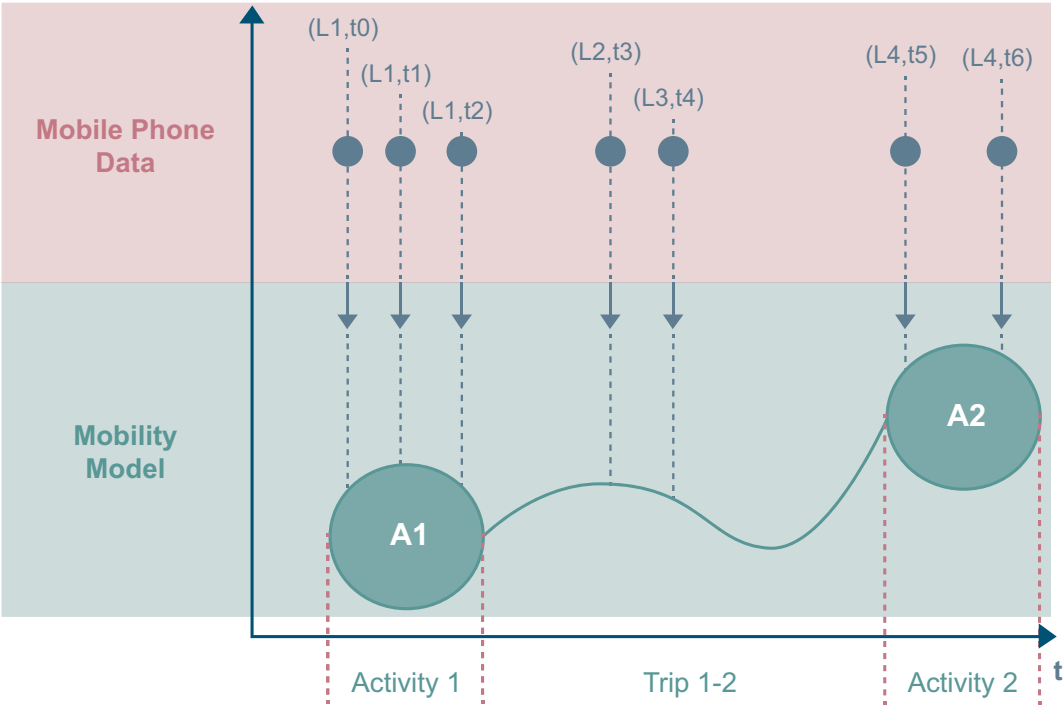


Figura 7: Teléfonos móviles a información de actividades y viajes.

Cuadro 1 – Estimación del tiempo de origen del viaje

Tiempo de origen del viaje = [**origen último registro** , **destino primer registro** - **duración del viaje**]

Donde **origen-último-registro** es el tiempo del último registro del teléfono móvil en origen, **destino-primer-registro** es el tiempo del primer registro observado en la actividad de destino y la **duración del viaje** es el tiempo de viaje entre el origen y el destino, considerando una distancia

- 3) La expansión de la muestra se realiza comparando la muestra de datos provenientes de telefonía móvil, con los datos del censo y diferenciando por zona, edad y género, obteniendo un factor de expansión para cada combinación de estos tres factores.
- 4) Finalmente se extrae la dinámica de la población contabilizando el número de personas en zonas de 1 km² y para cada hora del día

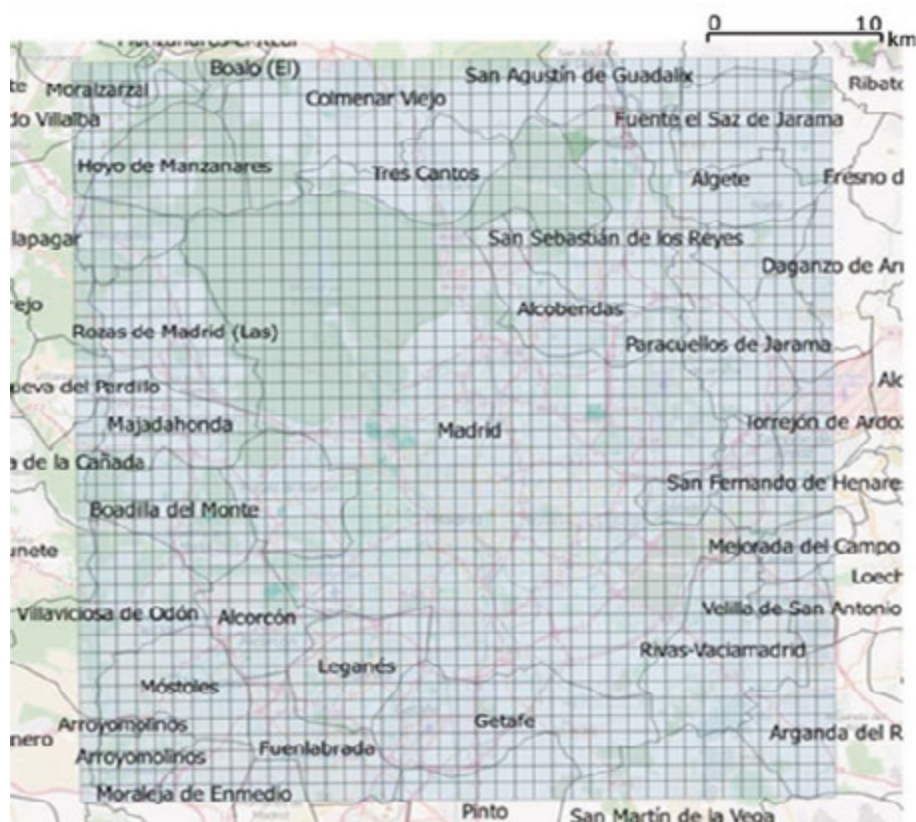


Figura 8: Malla de celdas de 1 km².

Se utilizan diferentes conjuntos de datos: a) datos de población del censo a nivel de secciones censales; b) malla de celdas de 1km² con datos sobre contaminación NO₂ en intervalos de tiempo de 1h, provenientes de modelos de simulación; c) datos de telefonía móvil provenientes de uno de los principales operadores en España con cerca del 30% de cuota de mercado, recolectados durante octubre y noviembre de 2014.

El indicador de exposición de la población al NO₂ para una determinada zona se calcula como el producto de la concentración de NO₂ y el número de personas por hora en la zona ($\mu\text{g}/\text{m}^3 \times \text{personas} \times \text{hora}$). Este indicador se calcula de dos formas: 1) estática, donde el número de personas x hora es constante a lo largo del día, utilizando los datos del censo; 2) dinámica donde el número de personas x hora en cada zona se estima a partir de los datos de telefonía móvil.

Beneficios

La exposición de la población se calcula para cada celda de 1km² y para cada hora del día. Los resultados muestran que a primeras horas del día (4 a.m.) los niveles de NO₂ son menores y la población está todavía en sus residencias. A las 9 a.m. mucha población se ha movido hacia el centro de la ciudad y también aumenta la concentración de NO₂ por lo que se incrementan de forma considerable los niveles de exposición. A mediodía decrecen los valores de exposición debido a la disminución de los niveles de NO₂ mientras que la población permanece en el centro.

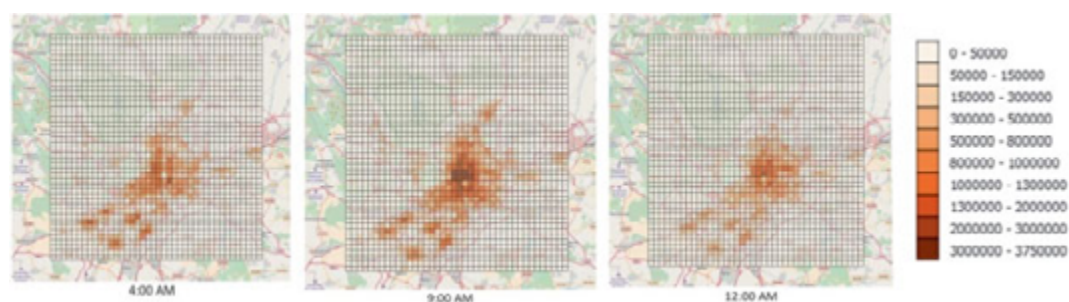


Figura 9: Indicador de exposición a diferentes horas.

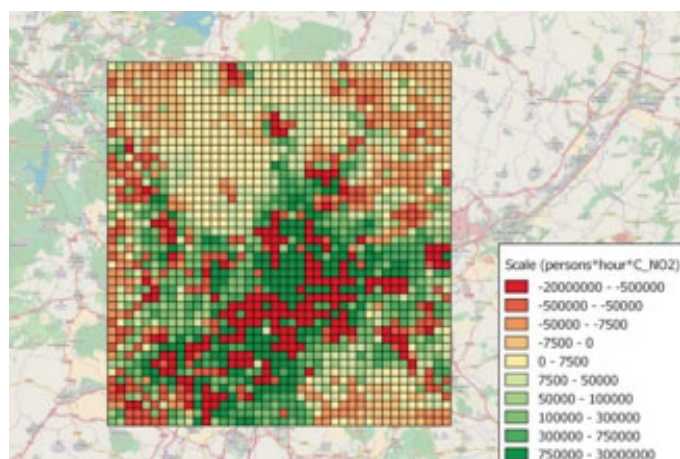
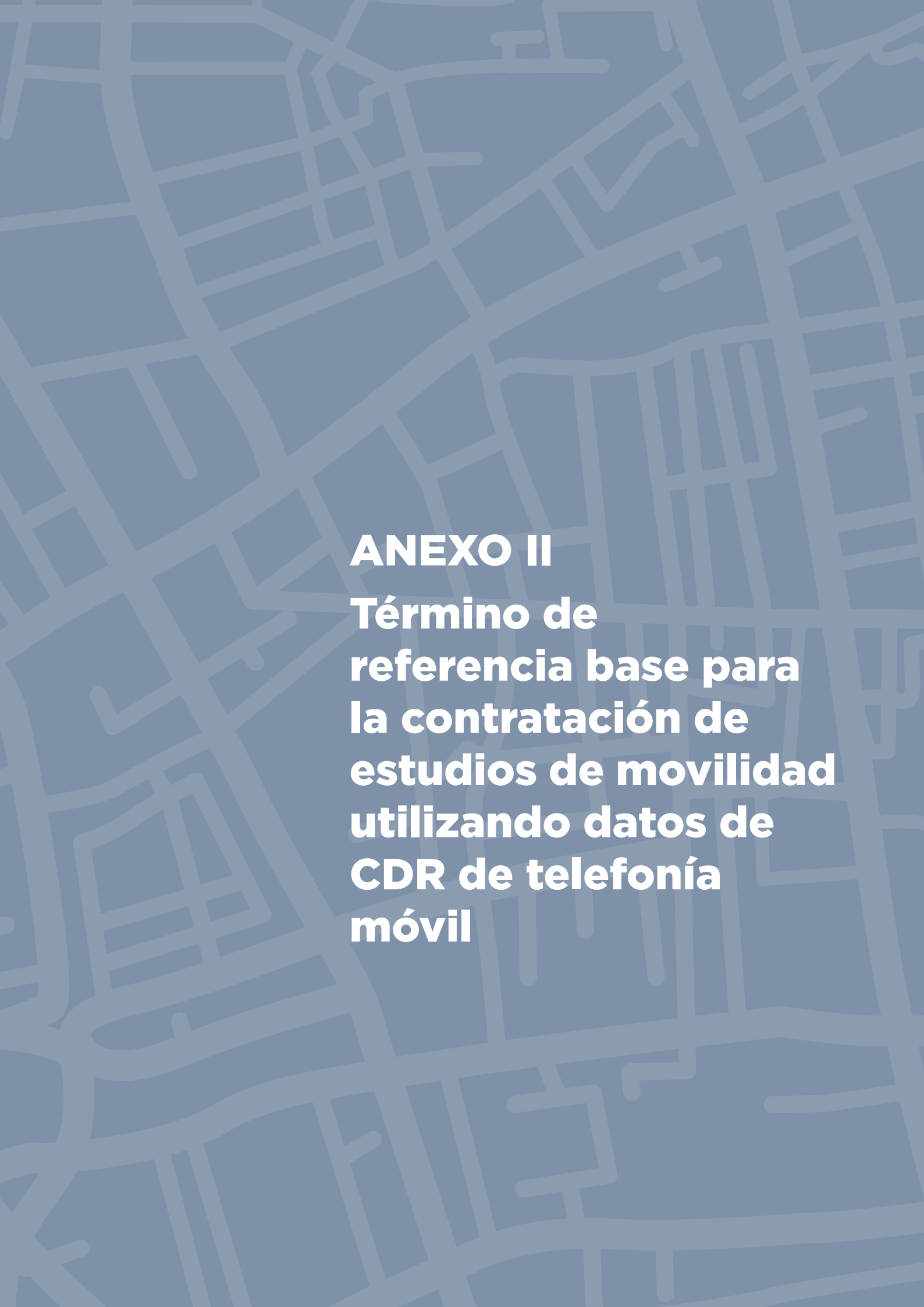


Figura 10: Indicador de exposición a diferentes horas.

La comparación del indicador de exposición de la población calculado con los datos de telefonía móvil (dinámico) y los datos del censo (estático) revela diferencias significativas en el centro de Madrid, en las zonas industriales del este de la ciudad (incluyendo el aeropuerto), la zona sur de Madrid y la periferia sur. Estas zonas aparecen infravaloradas respecto a sus niveles de exposición a la contaminación.

Una de las fortalezas de este estudio es que el uso de un indicador de población dinámica permite identificar las zonas que aparecen infravaloradas cuando se utilizan datos del censo para realizar análisis similares. Es necesario considerar no solo la población residente en la ciudad, como también la población que vive en los alrededores y se desplaza diariamente a la ciudad. Medir correctamente los niveles de exposición de la población es fundamental si se quiere por ejemplo analizar las causas de una baja calidad del aire y proponer políticas de prevención y mitigación de emisiones.

The background of the page is a light blue-grey color with a faint, stylized map pattern. The map consists of a network of lines representing streets or roads, with some lines forming rectangular blocks and others branching out. The lines are a slightly darker shade of the background color.

ANEXO II

Término de referencia base para la contratación de estudios de movilidad utilizando datos de CDR de telefonía móvil

En este apartado se enumeran las acciones que deben seguirse para implementar con éxito el análisis de la movilidad en las ciudades latinoamericanas utilizando los datos de telefonía móvil. Posteriormente, se identifican también los requerimientos a las empresas y consultoras a las que se va a contratar y se detallan las especificaciones que las mismas deben incluir en los informes finales sobre sus estudios de movilidad.

La contratación de un estudio de movilidad utilizando CDRs de telefonía móvil requiere una serie de acciones necesarias de considerar:

Definición de los objetivos del estudio. Aunque es algo que puede resultar evidente, la definición clara de los objetivos es un primer paso y fundamental a realizar. A partir de los objetivos vamos a poder determinar si los datos de telefonía móvil serán una fuente adecuada para el estudio o, si por el contrario, será necesario recurrir a encuestas. A la vez, si los datos de telefonía son la fuente óptima, una buena definición de los objetivos del estudio va a ser fundamental para establecer los requerimientos de la contratación.

Definición del área de estudio y nivel de detalle de los análisis. Junto a los objetivos, otra de las acciones de inicio es la definición del área de estudio y las escalas de análisis, tanto la escala espacial (tamaño de las unidades espaciales de referencia, zonificación) como la escala temporal. Como veremos posteriormente, la definición del área de estudio y su adecuación a los objetivos planteados inicialmente nos permitirá establecer los requerimientos en relación a la cobertura de la red de telefonía móvil necesaria para cubrir cumplir con esos objetivos en ese espacio. También, la escala de trabajo y el tamaño de las zonas espaciales con las que vamos a trabajar van a determinar si la geolocalización de los datos de telefonía móvil a utilizar debe ser por triangulación, por referencia a BTS o si por el contrario se podrían usar datos procedentes de handover o cambios de LA. Finalmente, establecer la escala temporal del estudio permite fijar los requerimientos temporales de la base de datos a utilizar por la compañía que realiza el estudio, de manera tal que se cubra por un lado todo el periodo de estudio establecido y por otro se puedan obtener los datos de movilidad con la resolución temporal establecida en los objetivos (ya sea distribución diaria de los viajes, semana, mensual, etc).



Definición de los plazos y entregables. Los plazos para la realización del estudio deben ser acordes con los objetivos fijados y las necesidades del tratamiento de la base de datos para las escalas y el área de estudio establecida. Es conveniente realizar un seguimiento y trabajo activo del proyecto junto a la empresa contratada, para lo cual es bueno establecer un calendario de entregables y reuniones de trabajo adecuado a dicho seguimiento.

Una vez definidos los objetivos del estudio y su adecuación a las posibilidades que nos ofrecen los datos de telefonía móvil, a continuación se enumeran un conjunto de requerimientos a considerar a la hora de la contratación de estudios de movilidad basados en telefonía móvil:

1. Información necesaria a considerar sobre la compañía telefónica de la que proceden los datos. En relación a la compañía es necesario conocer su cuota de mercado en el área de estudio. Este dato es una proxy fundamental para conocer el tamaño de la muestra y, por tanto, la fiabilidad de los resultados. Normalmente se considera que una cuota de mercado del 20% o más es suficiente. Es muy recomendable que se proporcione la cuota de mercado específica para el área de estudio y no solo para el conjunto del país. Además es necesario que se informe de las características de esa cuota de mercado, señalando posibles sesgos como consecuencia de un mercado segmentado, ya sea a nivel sociodemográfico (presencia de grupos socio-demográficos con baja penetración) o diferencias espaciales a nivel local (espacios con baja penetración de mercado).

2. Información sobre los datos utilizados. Es necesario que la empresa informe del total de registros que incluye la base de datos utilizada, sus características y la estructura de atributos que contienen la base de datos. Debe señalarse si se ha trabajado con datos de CDRs procedentes de eventos activos (y si estos incluyen sesiones de datos) o eventos pasivos. La empresa informará también del periodo de tiempo cubierto en la base de datos y de la resolución (granularidad) temporal de los datos, mostrando las estadísticas descriptivas de los tiempos entre eventos consecutivos. Se recomienda que se muestre un histograma de frecuencias de eventos según tiempos de eventos sucesivos.

Además de la resolución temporal, la empresa deberá informar sobre el tipo de georreferenciación de los datos que se ha usado, señalado si se ha trabajado a nivel de Localización de Área (LA), antena o con coordenadas por triangulación, o si se han combinado datos y en qué casos. Cuando se trabaje a nivel de torre (lo más habitual) es conveniente que se describan bien las características de la red (número de torres, tamaños medios de los polígonos de Voronoi y diferencias espaciales, etc.). Sería conveniente disponer de una cartografía de la localización de las torres y la definición de los polígonos de Voronoi y de los LA.

En relación con la información sobre el usuario, la empresa debe informar si tiene la posibilidad de trabajar con información sociodemográfica sobre los usuarios, como el género o el tipo de contrato. En caso afirmativo es necesario conocer también las políticas de privacidad de la empresa y las medidas a utilizar para mantener el anonimato de los usuarios.

3. Información sobre pre-procesos y limpieza de los datos. La empresa contratada deberá señalar qué proceso y pasos ha seguido para la selección de la muestra de usuarios válidos y la limpieza de errores en los datos.

Debe señalar el total de usuarios válidos con los que ha trabajado. Es importante conocer también el proceso seguido para la limpieza de rebotes entre torres.

4. Información sobre la metodología de cálculo de las matrices de viajes OD. Es importante que los técnicos que contratan el trabajo conozcan muy bien los pasos que se ha seguido para el cálculo de las matrices de viajes y las estadísticas de movilidad obtenidas a partir de ellas. Para ello, la empresa deberá describir los pasos fundamentales:

- a) Qué proceso ha seguido para determinar la localización del domicilio del usuario
- b) Qué proceso se han seguido para la definición de estancias y puntos de pasos de los viajes, indicando los parámetros usados, como el tiempo mínimo para definir la estancia, franjas horarias para determinar el lugar de trabajo u otro tipo de motivos de viaje, así como el error estimado para la hora de inicio y de finalización del mismo.
- c) Qué proceso se ha seguido para la conversión de los datos de polígonos de Voronoi a zonas de transporte.
- d) Cómo se ha realizado la expansión de la muestra. Es necesario conocer los datos usados para la expansión y unas estadísticas básicas descriptivas de los factores de expansión utilizados. Sería recomendable presentar una cartografía de los factores de expansión usados.
- e) Cuáles han sido los procesos sobre la validación de los resultados utilizados. Se deberían indicar las fuentes y los métodos utilizados para la validación y mostrar los resultados de la misma, tanto a nivel del conjunto del área de estudio como a nivel local.

5. Procesos de enriquecimiento de datos e información complementaria (rutas, modos, etc.). En los casos en los que se haya realizado operaciones de enriquecimiento de datos, la empresa debe indicar claramente la información que ha sido obtenida mediante estos procesos de enriquecimiento de datos y no directamente de los datos originales. A la vez, se deberá informar de los procesos y los métodos empleados para obtener la nueva información y de los metadatos de las fuentes secundarias que hayan podido utilizarse. En el caso de que se realicen encuestas por SMS para mejorar la información de los individuos y los viajes, deberá indicarse el número de encuestas obtenidas y la tasa de respuesta, así como la distribución espacial de la muestra resultante.

6. La empresa contratada deberá realizar una presentación y visualización de los resultados acorde a los objetivos del proyecto y cumpliendo con las especificaciones del pliego, ya sea en formato informe en pdf o en visualizaciones o proyectos web.

7. Finalmente, la empresa deberá entregar en soporte digital los datos ya agregados por zonas de transporte, preferiblemente en formato csv y shp, para que puedan ser visualizados y analizados en un SIG y utilizados fácilmente para alimentar los modelos de los planificadores.

