



La ciudad a vuelo de pájaro:

# Ciencia de datos para planificación urbana



**EDUARDO  
GRAELLS-GARRIDO**

Profesor Asistente del Departamento de Ciencias de la Computación de la Universidad de Chile e Investigador Joven en el Centro Nacional de Inteligencia Artificial (CENIA). Sus áreas de investigación principales son la informática urbana y la visualización de información. También es editor en Editorial Trazos de Aves y autor de la antología "GAME OVER".

✉ [egraells@dcc.uchile.cl](mailto:egraells@dcc.uchile.cl)



**DANIELA OPITZ**

Profesora Asistente de la Facultad de Ingeniería e Investigadora del Instituto Data Science (IDS) de la Universidad del Desarrollo. Sus áreas de investigación principales son informática urbana y desigualdad digital.

✉ [dopitz@udd.cl](mailto:dopitz@udd.cl)



**RESUMEN.** ¿Cómo medir lo que sucede en una ciudad hoy? Las problemáticas que afectan a las ciudades requieren un entendimiento actualizado y profundo de la movilidad de sus habitantes. En este artículo mostramos cómo las nuevas fuentes de datos, principalmente trazas de registros anonimizados de telefonía móvil, permiten observar una ciudad a vuelo de pájaro y obtener mediciones actualizadas y confiables sobre cómo se transportan sus habitantes. Ejemplificamos dos casos de estudio de proyectos de investigación con aplicaciones en planificación urbana real en Chile.

## Introducción

Las ciudades enfrentan problemas complejos en la actualidad. El crecimiento producido por la urbanización, la crisis climática, el cambio demográfico, la llegada de nuevas tecnologías, y otros fenómenos, han puesto a autoridades y ciudadanía bajo presiones que no tienen precedente. Las disciplinas urbanas han tradicionalmente enfrentado los problemas derivados de estos escenarios proponiendo e implementando soluciones basadas en la recolección de encuestas y en el uso de herramientas computacionales, sin embargo, los métodos tradicionales de recogida y análisis de datos no han dado abasto ante la complejidad que han alcanzado las problemáticas actuales. Lo anterior puede resumirse, en otras palabras, en que la ciudad del mañana requiere la solución de problemas hoy, pero las herramientas típicas son de un ayer que no contemplaba las dificultades actuales.

En la actualidad, el auge de la ciencia de datos plantea una oportunidad de creación de nuevas soluciones mediante la utilización de diversas fuentes de datos

## *El análisis de registros de uso de teléfonos móviles [...] permite realizar análisis a escalas temporales más finas que las disponibles en fuentes tradicionales, como las encuestas.*

y de enfoques computacionales [1]. En el contexto urbano, la ciencia de datos ha estudiado fenómenos relacionados con transporte, salud, comunicaciones, y otros, en base al análisis de registros de uso de teléfonos móviles [2]. Esta fuente de datos no tradicional permite realizar análisis a escalas temporales más finas que las disponibles en fuentes tradicionales, como las encuestas, lo que permite hacer estudios y evaluaciones de intervenciones ante sucesos disruptivos que podrían ocurrir en una ciudad [3]. Dichas soluciones requieren métodos computacionales avanzados, y su aplicabilidad plantea varios retos ya que un sistema basado en datos cuyo propósito es tener impacto en las políticas públicas debe ser transparente e interpretable, dos cualidades que no siempre son prioritarias en este tipo de proyectos [4, 5].

Aunque las disciplinas urbanas y la ciencia de datos trabajan en problemas similares, utilizan lenguajes diferentes, recurren a metodologías distintas y tienen prioridades divergentes. Las partes interesadas en soluciones, como las instituciones públicas y las autoridades/operadores de transporte, aun basan sus decisiones en los métodos y datos tradicionales y por lo general no tienen acceso a la escalabilidad y granularidad de los modelos provenientes de la ciencia de datos. Sin embargo, la situación actual y el crecimiento de las ciudades empuja con urgencia la aplicación de los métodos tradicionales complementada con métodos no tradicionales, ya que los primeros son sumamente caros y muchas veces el proceso de recolección de datos, y en consecuencia, los resultados de los análisis de estos datos, tardan demasiado. Ciertamente existen proyectos de ciencia de datos que han sido aplicados y adoptados con buenos resultados,

pero estos consisten principalmente en iniciativas individuales que son difíciles de replicar o de extender.

Para sistematizar la creación de soluciones modernas transdisciplinarias es necesaria la aplicación de metodologías de trabajo colaborativas que permitan cerrar brechas entre las disciplinas y partes interesadas. Existen propuestas de ese tipo de metodologías, incluyendo una definida y evaluada en Chile [6]. En este artículo describimos dos proyectos relacionados con ella. Primero, relatamos los resultados de un proyecto que definió maneras de medir lo que sucede en una ciudad utilizando datos de telefonía móvil, cuyos resultados incluyeron publicaciones científicas, el desarrollo de una patente, y su adopción para un proyecto real mediante licitación pública. Luego, describiremos un proyecto en curso que está observando y evaluando el transporte en los ejes más importantes del Gran Santiago, en particular en miras de una importante intervención urbana. Después, para finalizar el artículo, discutimos el presente y el futuro de esta línea de trabajo para mejorar la calidad de vida de las ciudades en Chile.

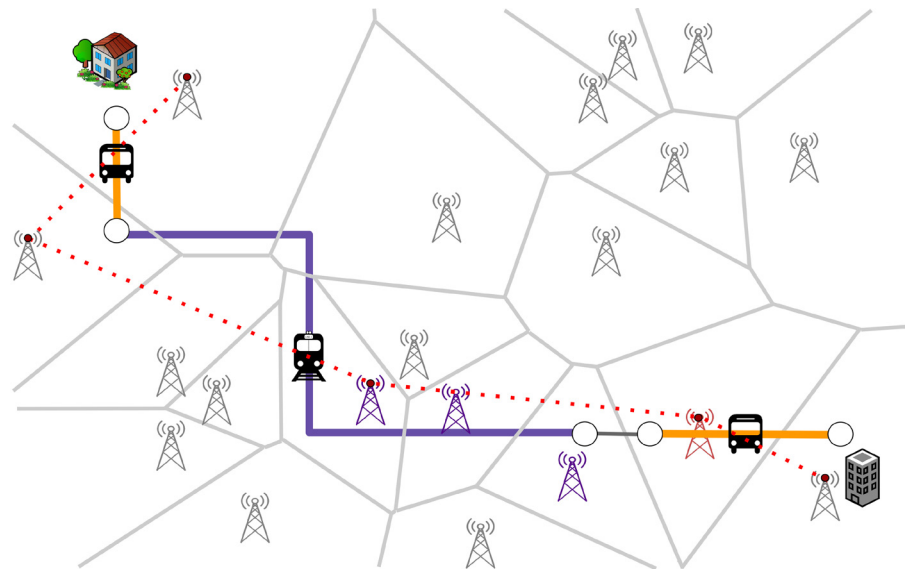
## Caso de estudio 1: Midiendo el uso actual de transporte en Santiago

Santiago tiene más de 7 millones de habitantes, con un sistema de transporte integrado multimodal [7]. Tradicionalmente, para medir el transporte en el área urbana de Santiago se han realizado encuestas de viaje cada 10 años, una tasa de frecuencia razonable en las décadas anteriores, pero insuficiente para captar



las dinámicas de la ciudad en la actualidad. De hecho, entre la última encuesta realizada, en 2012, y 2016, el año de nuestro primer estudio, la ciudad ya se había transformado de forma inesperada, y al día de hoy, 11 años después, estos cambios se han acentuado. Hay nuevas opciones de transporte, como las aplicaciones de viajes compartidos (por ejemplo, *Uber* y *Cabify*), los e-scooters personales y disponibles en aplicaciones (como *Lime*), y el sistema de bicicletas compartidas. Los servicios de planificación de rutas (como *Waze*) han cambiado la forma en que la gente toma decisiones durante un viaje, cambiando el tráfico en las calles. Además, también la población ha cambiado. En la última década se han producido olas de migración internacional sin precedentes en el país, y las personas migrantes se mueven de forma diferente por la ciudad en comparación con la gente local. Así pues, el potencial prescriptivo de la encuesta se vuelve limitado para la planificación y gestión de la red de transporte, ya que no se cuenta con información actualizada que permita medir la demanda de este. Una fuente de datos que permite medir la movilidad de las personas que usan el transporte público son los datos de tarjetas *bip!* Sin embargo, estos datos no permiten capturar todos los viajes que se realizan diariamente ya que un gran porcentaje de la población evade su pago. Por ejemplo, para la Región Metropolitana para el año 2022 se estimó una tasa de evasión de un 44%. Por otro lado, la demanda diaria de transporte privado es difícil de medir, ya que no existen fuentes de datos estándares ni públicas que se actualicen de manera oportuna.

Por tanto, las autoridades buscan activamente una solución que permita complementar la información de la encuesta de viajes con observaciones actualizadas de la demanda por cada modo de transporte, no sólo para satisfacer dicha demanda, sino también para mejorar la sostenibilidad de la ciudad y el uso del tiempo de sus habitantes. En vista de esta necesidad, el año 2016 se comenzó una



Fuente: Elaboración propia.

**Figura 1.** Representación de un viaje en datos de telefonía móvil y su comparación con la trayectoria real.

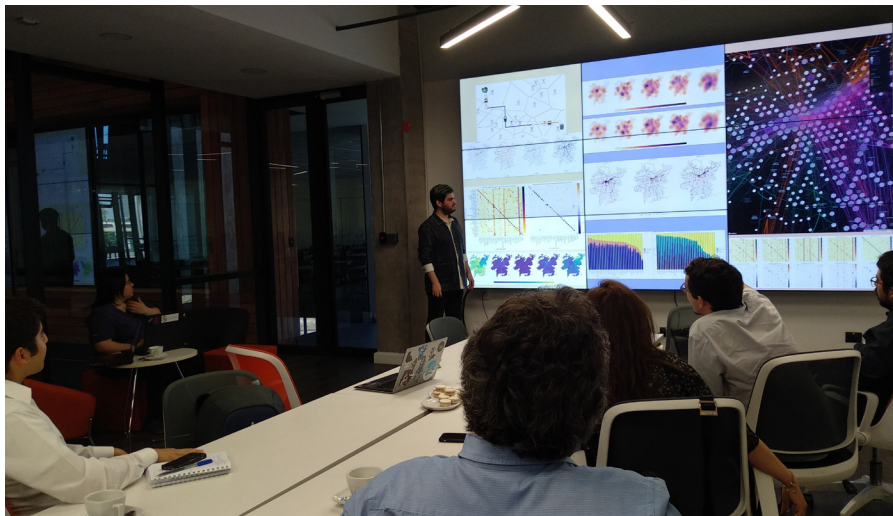
iniciativa colaborativa entre 1) Telefónica Chile, la mayor empresa de telecomunicaciones del país, que proporcionó registros de telefonía móvil anonimizados; 2) investigadores e investigadoras de diversas universidades, que aportaron conocimientos técnicos; y 3) profesionales de la Secretaría de Transporte (SECTRA) de Chile, que aportaron conocimiento del dominio y experiencia. El objetivo de esta iniciativa fue analizar y desarrollar nuevas metodologías basadas en datos de telefonía móvil, para resolver problemas relacionados a movilidad urbana como, por ejemplo, la inferencia del modo de viaje y la estimación de la partición modal de la ciudad (el cual detallaremos más adelante en esta sección). Además se desarrollaron metodologías para estimar el impacto en la ciudad del juego de realidad aumentada *Pokémon Go* [8], una metodología para estudiar la diversidad social en las visitas a centros comerciales [9] y un índice para medir la movilidad durante las restricciones en pandemia [10].

En nuestro contexto, el uso de datos de telefonía móvil aplicado al transporte no

es directo y está sujeto a varios desafíos. Por ejemplo, las torres de telefonía tienen una cobertura media de conexión mayor a 1 kilómetro, y no están distribuidas uniformemente en la ciudad. La granularidad temporal de los registros tampoco es uniforme, con frecuencias variables a lo largo del día, con un promedio de 15 minutos y máximas de varias horas, de acuerdo al uso de cada dispositivo [11]. Sin embargo, a pesar de estas limitaciones, podemos construir una trayectoria aproximada para cada dispositivo con la secuencia de torres a las que se conecta, en sus respectivos tiempos (ver Figura 1), por lo que es posible trabajar sobre los datos para extraer información valiosa para el transporte. El desafío está en identificar los modos de transporte para estas trayectorias individuales, lo que contrasta con la información provista por la encuesta de viajes, la cual está disponible por cada viaje de cada persona encuestada.

La primera solución que se propuso para identificar los modos de transporte usando datos de telefonía móvil en Chile





Fuente: Espacio Atelier de INRIA Chile.

**Figura 2.** Sesión de trabajo en un entorno de visualización de datos colaborativa.

se desarrolló a partir de análisis grupal de muchas trayectorias e incorporando conocimiento relacionado con la infraestructura urbana que rodea a cada torre de telefonía. Por ejemplo, algunas torres están instaladas dentro de estaciones de metro o cerca de autopistas, por tanto, su ubicación contiene señales que se pueden utilizar para asociar a las personas a uno o más modos de transporte. Mediante experimentación con múltiples técnicas y modelos, descubrimos que la técnica de *soft clustering* (agrupamiento suave) llamada *Topic-Supervised Non-Negative Matrix Factorization* [12] permitía agrupar a las personas de acuerdo a las posibles combinaciones de modos de transporte que utilizaban, teniendo en cuenta sus trayectorias y la información adicional sobre las torres. En pocas palabras, lo que hace la técnica es asociar grupos de personas que transitan por torres similares (esto es lo que queremos identificar), y, simultáneamente, torres similares que tienden a aparecer en las mismas trayectorias (esto es lo que permite asociar el grupo a uno o más modos de transporte). La implementación eficiente de este modelo está disponible como código abierto en la página con resultados

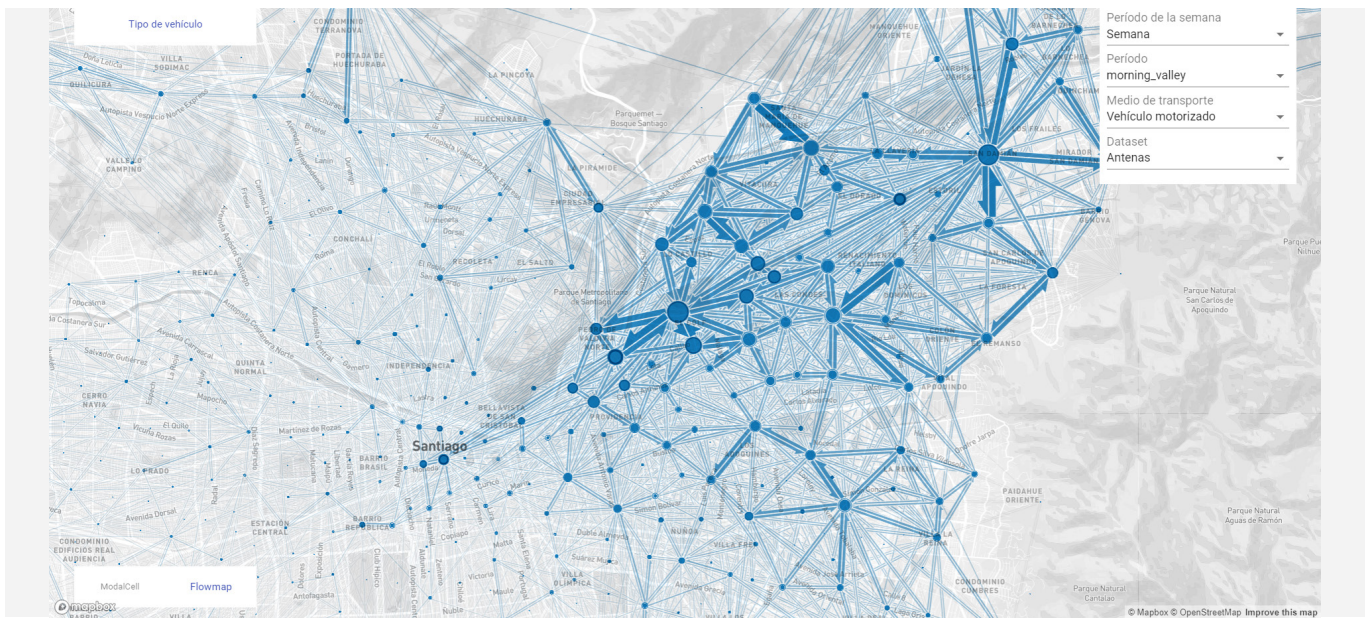
históricos del proyecto: <http://datagramas.cl/proyectos/runrun/>.

Durante el desarrollo del proyecto validamos todas las decisiones de procesamiento e interpretación de datos con personas expertas en gestión y planificación de transporte. Esto incluyó sesiones de trabajo con representantes de la Secretaría de Transporte, centros de investigación de disciplinas urbanas, e ingenieros de planificación de Metro de Santiago (ver Figura 2). La colaboración con personas expertas fue más allá de la validación de nuestras propuestas. Por ejemplo, ellas veían valor en el sistema, pero se preguntaron cómo integrar los resultados en sus propias herramientas, tanto desde el punto de vista operativo como metodológico. En términos operativos, ya se cuenta con herramientas de análisis de transporte público [13] que aportan valor y son utilizadas día a día en la gestión y planificación de la ciudad, por lo que no necesitan un software nuevo que haga todo desde cero, sino una aplicación que trabaje junto a lo existente. En términos metodológicos, necesitan estimaciones del sesgo de los datos y de características adicionales de los viajes que no habían sido consideradas.

Por ejemplo, declararon la importancia de incluir información sociodemográfica. Sin embargo, el carácter anónimo de los datos dificulta la estimación de la representatividad y de las características de estos, pero la existencia de fuentes actualizadas permite realizar cálculos, como el Censo de 2017 y la encuesta CASEN de 2020, junto a las proyecciones del Instituto Nacional de Estadísticas. Esta es una de las líneas de investigación que estamos explorando en la actualidad y cuyos resultados son útiles en el siguiente proyecto descrito en este artículo.

El proyecto generó software y publicaciones [14], sin embargo, un proyecto de ciencia de datos urbana sólo se puede considerar exitoso si tiene impacto real en la planificación de la ciudad. Esto también impone desafíos. Una de las preocupaciones del proveedor de datos era lograr que una implementación de este proyecto fuese sostenible en términos económicos. Para ello se necesita la protección de cualquier invención generada durante el proyecto mediante una patente. En nuestro caso, se solicitó una patente en Chile de manera conjunta entre centro de investigación y operadora de telecomunicaciones [15].

Hoy, en 2023, podemos comentar que la solución sí fue adoptada y puesta en operación. Un proyecto que incluye la solución descrita en este artículo ganó una licitación pública para complementar la Encuesta Origen-Destino 2019 en la conurbación de Rancagua-Machalí. Esto refuerza la idea de que un proyecto de ciencia de datos se debe integrar en los procesos y herramientas existentes, para aumentar el conocimiento disponible a partir del valor extraído de fuentes de datos no tradicionales. Así, los resultados de este proyecto pueden cumplir un rol relevante en mejorar la calidad de vida de los cientos de miles de habitantes de Rancagua-Machalí. Tenemos la esperanza de que esfuerzos como ese se repliquen para otras ciudades de Chile.



Fuente: Elaboración propia.

**Figura 3.** Imagen de prototipo del sistema de visualización *SuperFly* para análisis de movimientos entre orígenes y destinos.

## Caso de estudio 2: Evaluando intervenciones de transporte a través de la integración de datos

Actualmente el Gobierno Regional (GORE) Metropolitano está ejecutando una serie de proyectos que buscan mejorar la calidad de vida de los habitantes de la región. Uno de ellos es la construcción de la Ciclovía Metropolitana que irá desde Pajaritos hasta Plaza Baquedano por el eje Alameda-Providencia.

Para la ejecución de este proyecto el GORE ha iniciado una colaboración activa con investigadores del Instituto de Data Science de la Universidad del Desarrollo, la Universidad de Chile, la consultora especializada en transporte Cityplanning y Fundación País Digital. El objetivo de esta colaboración es estimar la potencial demanda de la ciclovía y evaluar el impacto que tendrá su construcción mediante la integración

de distintas fuentes de datos, incluyendo datos de telefonía móvil. Parte del proceso de análisis incluye el modelo descrito en el caso de estudio anterior, y también una extensión que permite su integración con fuentes de datos oficiales y otras trazas digitales mediante técnicas de fusión de datos [16]. Los datos oficiales incluyen fuentes como la encuesta CASEN y estadísticas disponibles en el Instituto Nacional de Estadísticas. El proyecto también innova al incorporar datos de conteo de vehículos, ciclistas y peatones en intersecciones relevantes para el proyecto, obtenidos a través de cámaras e inteligencia artificial. Además se realizó un estudio de preferencias declaradas y se desarrollará un estudio etnográfico en varios puntos del eje para entender cómo distintos actores conviven en el entorno de una ciclovía. Los resultados de los análisis se transfieren al GORE a través de visualizaciones interactivas de la movilidad de la ciudad basadas en los resultados del proyecto anterior (el sistema extiende el prototipo expuesto en la Figura 3).

Esta iniciativa ilustra las potencialidades de colaboración entre el gobierno, la academia y la industria y cómo las metodología de evaluación de intervenciones urbanas tradicionales, se pueden nutrir y complementar con fuentes de datos no tradicionales —como las trazas digitales de los celulares y los datos de conteo de cámaras— y se pueden desarrollar metodologías integradoras que capturen lo mejor de ambos mundos.

## Presente y futuro

Ante el auge de las herramientas de inteligencia artificial en los últimos años, es inevitable preguntarse cómo las nuevas tecnologías impactarán la planificación urbana. A decir verdad, en dichas problemáticas no es necesario que una solución de ciencia de datos presente un desarrollo completamente nuevo, más bien, el mayor reto es integrar una solución novedosa para un problema real en el contexto que la requiere, y que esta solución sea adoptada y puesta en operación. En



## Para medir el transporte en [...] Santiago [tradicionalmente] se han realizado encuestas de viaje cada 10 años, una tasa de frecuencia razonable en las décadas anteriores, pero insuficiente [...] en la actualidad.

este texto hemos presentado dos proyectos, uno terminado y otro en ejecución, donde el uso de fuentes de datos no tradicionales y los métodos computacionales permiten la creación de nuevos conocimientos accionables. Ahora bien, dejamos claro que este conocimiento

no es automatizado por máquinas o por el software que hemos creado, sino que la tecnología es mediadora en el trabajo que realizan personas especialistas en múltiples disciplinas, tanto ciencia de datos como disciplinas urbanas y políticas. Como hemos descrito, parte importante del resultado positivo de estos proyectos se debe al conocimiento disponible en fuentes de datos tradicionales y a la experiencia de las personas que diseñan o ejecutan políticas públicas. El valor que aporta la ciencia de datos se sustenta en la sinergia creada entre las distintas personas y organizaciones que extraen valor y conocimiento de fuentes tradicionales y digitales de datos.

Nuestro grupo de investigación PLUMAS (Plataforma de Urbanismo, Movilidad,

Análisis y Simulación) está trabajando en esta línea actualmente, en una colaboración entre el Instituto de Data Science (IDS) de la Universidad del Desarrollo, el Departamento de Ciencias de Computación de la Universidad de Chile (DCC), y el Centro Nacional de Inteligencia Artificial (GENIA). En estas temáticas mantendremos colaboraciones con investigadores e investigadoras como Loreto Bravo (U. del Desarrollo), Francisco Rowe (U. de Liverpool), Jacqueline Arriagada (U. de Leeds), Patricio Reyes (Barcelona Supercomputing Center), Horacio Samaniego (U. Austral) y Rossano Schifanella (U. de Turín), y guiamos proyectos de título y tesis de estudiantes en computación y otras disciplinas vinculadas a la ciencia de datos, con aplicaciones en temáticas urbanas, geográficas y sociales. ■

### REFERENCIAS

- [1] Cao, L. (2017). Data science: Challenges and directions. *Communications of the ACM*, 60, 59–68.
- [2] Blondel, V. D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4, 10.
- [3] Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography*, 3, 274–279.
- [4] Castelvocchi, D. (2016). Can we open the black box of AI? *Nature News*, 538, 20.
- [5] Krause, J., Perer, A., & Ng, K. (2016). Interacting with predictions: Visual inspection of black-box machine learning models. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5686–5697. ACM.
- [6] Graells-Garrido, E., Peña-Araya, V., & Bravo, L. (2020). Adoption-driven data science for transportation planning: Methodology, case study, and lessons learned. *Sustainability*, 12, 6001.
- [7] Muñoz, J. C., & Gschwender, A. (2008). Transantiago: A tale of two cities. *Research in Transportation Economics*, 22, 45–53.
- [8] Graells-Garrido, E., Ferres, L., Caro, D., & Bravo, L. (2017). The effect of Pokémon Go on the pulse of the city: A natural experiment. *EPJ Data Science*, 6, 1–19.
- [9] Beiró, M. G., Bravo, L., Caro, D., Cattuto, C., Ferres, L., & Graells-Garrido, E. (2018). Shopping mall attraction and social mixing at a city scale. *EPJ Data Science*, 7, 1–21.
- [10] Pappalardo, L., Cornacchia, G., Navarro, V., Bravo, L., & Ferres, L. (2023). A dataset to assess mobility changes in Chile following local quarantines. *Scientific Data*, 10, 6.
- [11] Graells-Garrido, E., Peredo, O., & García, J. (2016). Sensing urban patterns with antenna mappings: The case of Santiago, Chile. *Sensors*, 16, 1098.
- [12] MacMillan, K., & Wilson, J. D. (2017). Topic supervised non-negative matrix factorization. *arXiv Preprint arXiv:1706.05084*.
- [13] Gschwender, A., Munizaga, M., & Simonetti, C. (2016). Using smart card and GPS data for policy and planning: The case of Transantiago. *Competition and Ownership in Land Passenger Transport (Selected Papers from the Thredbo 14 Conference)*, 59 IS -, 242–249.
- [14] Graells-Garrido, E., Caro, D., & Parra, D. (2018). Inferring modes of transportation using mobile phone data. *EPJ Data Science*, 7, 49.
- [15] Graells-Garrido, E., & Caro Alarcón, D. (2020). *Patent no. WO2020093181A1*.
- [16] Graells-Garrido, E., Opitz, D., Rowe, F., & Arriagada, J. (2023). Feel old yet? Updating mode of transportation distributions from travel surveys using data fusion with mobile phone data. *arXiv Preprint arXiv:2204.09482v2*.