

Detección de patrones secuenciales generalizados de movilidad de ciclistas a partir de datos crowdsourcing

Quezada Adriana, Orellana Daniel

LlactaLAB, Departamento de
Espacio y Población, Universidad
de Cuenca
Av. 12 de Abril, Cuenca/Azuay/
Ecuador
adriana.quezada@ucuenca.edu.ec,
daniel.orellana@ucuenca.edu.ec

RESUMEN

La movilidad en bicicleta conlleva beneficios para el ambiente, la economía y la salud de la población. No obstante, la adopción de la bicicleta como medio de transporte, está limitada por la carencia de geoinformación detallada que describa el comportamiento de los ciclistas y permita la creación de estrategias adecuadas de planificación y promoción. En este estudio presentamos un método para obtener conocimiento sobre los patrones de movilidad de ciclistas urbanos utilizando datos crowdsourcing, minería de datos y técnicas de visualización de geo-información. Demostramos la viabilidad del método analizando un año de datos de la aplicación Strava METRO en la ciudad de Cuenca, y extrayendo y visualizando patrones secuenciales generalizados (GSP) de movimiento. Los resultados evidencian que la aplicación del método propuesto permite generar información relevante para el conocimiento sobre movilidad no motorizada utilizando conjuntos masivos de datos recolectados con técnicas de crowdsourcing.

Palabras clave: **movilidad no motorizada, bicicletas, geoinformación de crowdsourcing, patrones secuenciales generalizados, comportamiento colectivo espacial.**

ABSTRACT

Bicycle mobility lead to environmental, economic and health benefits to the population. Nonetheless, the implementation of bicycle-based transport systems is limited due to lack of detailed geodata describing the behavior of cyclists in order to build appropriate planning and promotion strategies. In this study, we present a method to extract knowledge about urban cyclist's mobility patterns using crowdsourced geodata, data mining and geodata visualization techniques. We demonstrate the viability of the method by analyzing a year-long data from the Strava METRO application in the city of Cuenca (Ecuador); and extracting and visualizing generalized sequential patterns (GSP) of movement. The results that data mining analysis technique applied to massive datasets allows the generation of relevant information about the non-motorized mobility knowledge.

Keywords: non-motorized mobility, bikes, crowdsourced geoinformation, generalized sequential patterns, collective spatial behavior.

I. INTRODUCCIÓN

Actualmente, la movilidad no motorizada, sea a pie o en bicicleta, se posiciona como una estrategia potente de mitigación ante el actual detrimento de la calidad ambiental, económica, y de la salud de la población. El predominio de la movilidad motorizada como medio de transporte y el incremento del parque automotor, son factores impulsores del aumento de las emisiones de gas de efecto invernadero, ante lo cual, grupos de expertos a nivel global sugieren a los gobiernos incentivar la inversión en transporte público, en la construcción de infraestructura para peatones y ciclistas, a fin de adoptar sistemas de transporte modales de baja emisión de CO₂. (Sims R., R. Schaeffer, F. Creutzig, X. Cruz-Núñez, M. D'Agosto, D. Dimitriu, M.J. Figueroa Meza, L. Fulton, S. Kobayashi, O. Lah, A. McKinnon, P. Newman, M. Ouyang, J.J. Schauer, D. Sperling, and G. Tiwari, 2014). En el campo de la salud, se ha definido a la bicicleta como un medio de transporte saludable. El uso de la bicicleta incrementa la condición física de las personas, mitigando los efectos nocivos del sedentarismo, los cuales se intensifican con el uso del auto privado (Fishman, 2016).

Ante los beneficios de este medio de transporte, planificadores e investigadores se cuestionan cómo incrementar la movilidad en bicicleta. Varias investigaciones a nivel mundial han encaminado sus esfuerzos al análisis de propuestas para el fomento del uso de la bicicleta. Los resultados evidencian que la dotación de infraestructura completa (vías, señalización, iluminación, parqueaderos, etc.) impacta más en la movilidad en bicicleta que la implementación parcial de esta (Pucher, Dill, & Handy, 2010). De acuerdo a Fishman (2016), aproximadamente al 60% de la población tiene interés en utilizar la bicicleta, pero al mismo tiempo les preocupa hacerlo por la falta de seguridad. Es clave que las estrategias para incrementar el número de usuarios incluyan la implementación de redes de ciclovías, complementadas con zonas de parqueo, señalización, iluminación, así como aspectos del como la disminución del tráfico vehicular, la accesibilidad a amenidades y servicios en las rutas, señalización, entre otros, a fin de mejorar la percepción de seguridad del usuario (Fishman, 2016). En efecto, los países que registran mayor actividad de ciclistas cuentan con una infraestructura extensa y segura (Pucher, Dill, & Handy, 2010). Otros aportes que la investigación ha hecho se enmarcan en la obtención de datos descriptivos y detallados de la movilidad en bicicleta, como son: el número de viajes, su origen y destino, el espacio geográfico y temporal, características demográficas y socioeconómicas de los usuarios, así como sus percepciones y motivaciones (Handy, van Wee, & Kroesen, 2014). No obstante, las ciudades afrontan el desafío de motivar el uso de la bicicleta como medio de transporte, con recursos económicos limitados. El entender el comportamiento espacial de los ciclistas, permite optimizar esos recursos e invertir de manera efectiva en la planificación de la movilidad no motorizada en bicicleta. Sin embargo, datos descriptivos que expliquen el comportamiento de los ciclistas generalmente son escasos, a veces inexistente, y su obtención es costosa (Iacono, Krizek, & El-Geneidy, 2010).

DATOS CROWDSOURCING Y LA MOVILIDAD EN BICICLETA

La aparición de datos abiertos y Big Data, representa un instrumento idóneo tanto para planificadores urbanos y de transporte, así como para la comunidad científica que se enfoca en el estudio de la movilidad no motorizada en bicicleta, al ser información geolocalizada disponible, de rápida expansión y muchas veces, de acceso abierto. En el ámbito del estudio de la movilidad en bicicleta, estos recursos han sido ampliamente utilizados. Aspectos como el tráfico, los flujos,

el comportamiento, los modos de transporte, tipos de viaje, entre otros, han sido medidos con datos provenientes del sistema de geo-localización espacial. Estos han sido analizados y complementados con el uso de sistemas de información geográfica, plataformas virtuales, dispositivos móviles, teléfonos móviles inteligentes, y aplicaciones móviles de información geográfica voluntaria y crowdsourcing. Esta última, ha sido ampliamente usada por ciclistas, permitiendo la recolección masiva de datos. Entre las aplicaciones más populares están *Endomondo*, *MapMyRide* y *Strava*. Actualmente empresas como Strava METRO, comercializan estos datos para la planificación e investigación sobre movilidad no motorizada (Romanillos, Austwick, Ettema, & De Kruijf, 2015). De hecho, en los últimos, varios estudios los han utilizado para identificar: patrones de tiempo de los viajes, frecuencias, tipos de viaje, rutas origen-destino (Selala & Musakwa, 2016), volumen de tráfico de ciclistas y su asociación con factores del entorno (Jestico, Nelson, & Winters, 2016), evaluación de intervenciones en la infraestructura de la movilidad no motorizada (Heesch & Langdon, 2016), el rol de la movilidad en bicicleta en la actividad física (Handy et al., 2014), entre otros. El resultado de estas investigaciones evidencia que la información crowdsourcing de Strava es útil (Heesch & Langdon, 2016; Jestico et al., 2016; Romanillos et al., 2015; Selala & Musakwa, 2016). Comparada con otras metodologías de recolección de datos, como el conteo manual, la información proporcionada por Strava es útil (Jestico et al., 2016), no solo por la cantidad de datos, sino por la capacidad de abarcar simultáneamente escalas mayores, apropiadas para la planificación urbana (Griffin & Jiao, 2015). Sin embargo, se advierte no considerar los datos crowdsourcing como una muestra representativa de la población de ciclistas, ni una manera de identificar ciclistas potenciales (Griffin & Jiao, 2015; Romanillos et al., 2015). Quienes utilizan este tipo de aplicaciones generalmente son personas interesadas en conocer su condición física, sus rutas, la calidad de aire, etc. Esta falta de representatividad se traduce a términos de localización geográfica, aspectos demográficos y socioeconómicos (Romanillos et al., 2015). De hecho, estudios han demostrado que el número de viajes identificados con la aplicación se relaciona con el nivel de ingreso de sus usuarios (Selala & Musakwa, 2016).

Los datos crowdsourcing aportan principalmente a los planificadores y tomadores de decisiones; es información que les permite mejorar el entendimiento del comportamiento espacial de los ciclistas, y tener un acercamiento a las cualidades espaciales de sus rutas preferidas y de las que evitan, siendo insumos importantes para el diseño y planificación de infraestructura de movilidad en bicicleta, y para el monitoreo de la existente. Pocos son los estudios que se han enfocado en el campo de la planificación urbana, y en el desarrollo de metodologías para analizar las rutas registradas (Romanillos et al., 2015). En este sentido, es importante considerar que, si bien la información de estas bases de datos se caracteriza por su nivel de detalle, su utilidad sólo se manifiesta una vez que se detectan y se visualizan patrones que evidencian el comportamiento colectivo espacial de los ciclistas.

DETECCIÓN DE PATRONES DE MOVILIDAD EN BICICLETA

Los datos crowdsourcing de actividades en bicicleta, por si solos, no evidencian patrones que describen las rutas preferidas por el colectivo de ciclistas, sino únicamente las rutas individuales agregadas y anonimizadas. Al respecto, la minería de datos ofrece algoritmos enfocados al análisis de datos espaciales de trayectoria, que permiten extraer conocimiento de conjuntos de datos. La minería de datos de patrones de trayectoria facilita el análisis de patrones de movimiento de

un gran volumen de datos espaciales, identificando patrones donde varios individuos se mueven en conjunto, patrones de rutas o espacios comunes (clusters) de objetos en movimiento, patrones de periodicidad, y patrones secuenciales de una trayectoria o de un conjunto de trayectorias (Zheng, 2015). La minería de patrones secuenciales es un proceso que extrae patrones secuenciales específicos cuyo valor de soporte excede un valor mínimo de soporte predeterminado. Los patrones extraídos con este proceso son secuencias que describen comportamientos comunes útiles para propósitos como: la detección de rutas apropiadas, predecir localizaciones, entender trayectos, etc., para diferentes propósitos en varios campos de las ciencias (Orellana, Bregt, Ligtenberg, & Wachowicz, 2012/6; Slimani & Lazzez, 2013; Zheng, 2015). La detección de aspectos comunes se ejecuta en base a un orden relativo y no absoluto (Orellana et al., 2012). Existen cinco algoritmos base para la extracción de los patrones secuenciales: *Apriori-like algorithm*, *Breadth First Search-based algorithm*, *Depth First Search-based algorithm*, *incremental-based algorithm* y *closed sequential patterns based algorithm*; a partir de estos se han desarrollado una serie de algoritmos que los han perfeccionado (Slimani & Lazzez, 2013). Es el caso del algoritmo BIDE, desarrollado para extraer patrones secuenciales generalizados (GSP), en base al esquema de control de cierre de secuencias denominado *BI Directional Extension*. Su eficacia radica en no necesitar un candidato (conjunto de patrones secuenciales generalizados) para identificar o validar nuevos patrones, lo que optimiza el uso de memoria en su proceso (Wang & Han, 2004). La relevancia de los GSPs se determina en base a un análisis exploratorio y a criterios subjetivos y objetivos. Los criterios son objetivos cuando su análisis se basa en datos crudos y en la aplicación de teorías de probabilidad, de estadística, o de información. Los criterios son subjetivos cuando en su análisis intervienen los datos y el usuario de los datos, ante el cual la novedad y la sorpresa queda a su juicio en base a conocimientos adquiridos. (Geng & Hamilton, 2006).

MOVILIDAD NO MOTORIZADA EN CUENCA

La ciudad de Cuenca, es la tercera cabecera cantonal más importante del Ecuador y se localiza en la zona montañosa de los Andes al sur del país. La ciudad crece de manera dispersa, continua y con baja densidad. Está formada por 15 parroquias urbanas y 21 parroquias rurales. Siguiendo las tendencias globales, en Cuenca, la movilidad motorizada es la que predomina como medio de transporte; el crecimiento del parque automotor ha incrementado al 11% anual, y consecuentemente, la disminución de la calidad del aire, considerando que el sector del transporte representa el 57% de emisiones de gases efecto invernadero local, llegando a índices de contaminación aún menores a la norma establecida por el Ministerio del Ambiente del Ecuador (BID, 2015). El panorama para la movilidad no motorizada en Cuenca no es alentador; el 24% de la población se desplaza a pie, y apenas el 1% en bicicleta (“Colectivo Cuenca Ciudad para Vivir,” 2016). El gobierno local ha desarrollado proyectos para motivar el uso de la bicicleta como transporte público, mediante la “Red Urbana de Ciclovías” y la “Bicicleta Pública para Cuenca” (BID, 2015), y desde el 2012 se incorpora parcialmente infraestructura para bicicletas en la ciudad. Los estudios realizados han favorecido a identificar el perfil del usuario (género, edad, ocupación, etc.) de diferentes medios de transporte, los viajes origen-destino, las características de los viajes (tiempo, motivo, medio de transporte) (GAD Municipal del Cantón Cuenca, 2015), entre otros aspectos; sin embargo, muy poco se conoce en relación a las trayectorias individuales y las rutas más frecuentadas por el colectivo de ciclistas en Cuenca, es decir, sobre su comportamiento espacial. Es así,

que en el marco del proyecto “Pies y Pedales”, desarrollado por el grupo de investigación LactaLAB - Ciudades Sostenibles de la Universidad de Cuenca, proponemos detectar, extraer y visualizar patrones secuenciales generalizados de movilidad de ciclistas de la zona urbana de Cuenca, utilizando datos crowdsourcing de la aplicación Strava METRO, algoritmos de minería de datos y técnicas de visualización, con la finalidad de brindar una metodología que permita construir conocimiento sobre la movilidad no motorizada en bicicleta.

II. METODOLOGÍA

La metodología utilizada en este artículo se basa en el enfoque de análisis exploratorio de datos espaciales (ESDA) para detectar e interpretar patrones de movimiento a fin de entender el comportamiento espacial de los usuarios de la bicicleta en el área urbana de Cuenca. El enfoque metodológico consistió en tres fases principales: La primera, la extracción y preprocesamiento de datos crowdsourcing de las trayectorias de ciclistas de la zona urbana de Cuenca. La segunda, la aplicación del algoritmo BIDE para extraer los patrones secuenciales generalizados (GSPs) de movimiento que representan rutas frecuentes de los lugares por los que los ciclistas cruzan durante sus viajes. Finalmente, el análisis de las propiedades principales de los patrones extraídos: tamaño, soporte, localización y relación con variables socioeconómicas.

A. EXTRACCIÓN Y PREPROCESAMIENTO DE DATOS CROWDSOURCING

Strava METRO es un servicio de datos que ofrece información geográfica, temporal y descriptiva del movimiento de ciclistas y peatones de un área geográfica específica, compatible con diferentes ambientes de Sistemas de Información Geográfica (SIG) (Strava, 2015). La información fue recolectada por usuarios de la aplicación Strava, en teléfonos móviles inteligentes que permite monitorear las trayectorias a pie o en bicicleta. Los datos recolectados son pre-procesados, agregados y anonimizados por el proveedor de datos y ofertados como información para la planificación urbana. Se adquirió un conjunto de datos para la ciudad de Cuenca para el período Agosto 2014 - Julio 2015. El conjunto de datos consiste en tres productos principales: a) Streets, representa datos agregados de un minuto por cada Edge en una red de calles obtenidas de OpenStreetMap; b) Nodes, similares a los Edges pero agregados por intersección de calles con información acerca del tiempo de espera, y c) Origen/Destino (OD) el cual se deriva de las Streets y de los reportes de origen y destino, así como las localidades intermedias de cada viaje. Estas localidades son agregadas dentro de sectores censales para preservar el anonimato de los usuarios y para luego comparar con datos socioeconómicos. El análisis descrito en este artículo se enfoca en el producto OD, del cual extrajimos las secuencias correspondientes a la ciudad de Cuenca, para aplicar un análisis exploratorio de los viajes en bicicleta que inician en la zona urbana.

Debido al potencial sesgo socioeconómico de estos datos, analizamos la correlación de la frecuencia de orígenes de viajes con el estatus socioeconómico. Para esto se calculó para cada sector censal el Índice de Condiciones de Vida (ICV), el cual resume las condiciones de vida de la población y su estatus socioeconómico (Orellana & Osorio, 2014).

Todos los datos fueron procesados en R (R project, 2015) y QGIS (QGIS, 2017).

B. EXTRACCIÓN DE PATRONES SECUENCIALES GENERALIZADOS

Los patrones secuenciales generalizados (GSPs) pueden ser definidos como frecuencias de secuencias generalizadas que pueden ser encontradas en conjuntos de eventos ordenados temporalmente (Agrawal & Srikant, 1995). La palabra generalizado implica, por un lado, que los eventos ocurrieron en un orden relativo más que en uno absoluto, y por otro lado, que la detección de secuencias frecuentes podría tener eventos intermedios presentes únicamente en algunas de las secuencias originales. Un ejemplo ilustrativo de GSP es representado en la Figura 1. A la izquierda se encuentran 5 secuencias de números (s1...s5). A la derecha, cuatro GSPs fueron detectados. La secuencia generalizada más común (GSP1: 5, 6) se encuentra en todas las secuencias originales, mientras que en la segunda (GSP2: 5, 6, 1) se encuentra en cuatro de las cinco secuencias originales. El valor de soporte de un GSP indica la razón entre el número de secuencias en la que el GSP es encontrado y el número total de secuencias del conjunto de datos.

s1: 1,2,3,5,6,3,1	GSP1: 5,6 (support = 1)
s2: 1,2,4,5,6,1	GSP2: 5,6,1 (support = 0.8)
s3: 1,5,6,7,9	GSP3: 1,5,6 (support = 0.6)
s4: 2,5,3,6,1	GSP4: 2,5,6 (support = 0.6)
s5: 4,5,6,1	

Figura 1. Ejemplo de patrones secuenciales generalizados extraídos de un conjunto de secuencias de eventos.

En nuestro caso, cada secuencia es un viaje en bicicleta con todos los sectores censales, y en el orden en el que estos fueron recorridos. Por lo tanto, la misión de este análisis es encontrar rutas frecuentes seguidas por los ciclistas. La interpretación del ejemplo provisto en la Figura 1. en términos de comportamiento espacial es que todos los viajes fueron del sector censal 5 al sector censal 6, a pesar del lugar en el que estuvieron antes, después o en la mitad de estos tramos. Analizar los GPSs en lugar de las secuencias originales nos permite un mejor exploración y entendimiento del comportamiento de la movilidad colectiva.

Los GSPs fueron extraídos usando el algoritmo BIDE+, implementado en *Sequential Pattern Mining Framework* (SPMF) (Fournier-Viger et al. 2016). La ventaja del algoritmo BIDE+ es que evita la redundancia al extraer solo los patrones secuenciales "cerrados", i.e. secuencias que no están contenidas en otras secuencias que tiene el mismo valor de soporte. Un patrón cerrado produce una clase equivalente de patrones que comparten el mismo cierre, y aquellos patrones están parcialmente ordenados, es decir, de acuerdo a la relación de inclusión. Los elementos más pequeños en la clase equivalente son llamados generadores mínimos, y el único elemento máximo es llamado patrón cerrado (Fournier-Viger et al., 2008). Los GSPs cerrados, fueron extraídos con un valor mínimo de soporte igual a 0.14 (es decir, que aparece en al menos el catorce por ciento de los viajes). Los GPSs resultantes fueron exportados a un archivo separado por comas para un mayor análisis.

C. EXPLORACIÓN Y VISUALIZACIÓN DE LOS GSPS

Con el fin de interpretar los GSPs y obtener conocimiento sobre el comportamiento de la movilidad colectiva de los usuarios de bicicleta en la zona urbana de Cuenca, los patrones extraídos fueron analizados de acuerdo a sus propiedades intrínsecas: tamaño, soporte y frecuencia. Finalmente, los GPSs y sus propiedades destacadas fueron representadas usando la técnica de visualización *Flow maps*, que representan mo-

vimientos colectivos entre localidades, y son especialmente adecuado para visualizar patrones de movimientos secuenciales generalizados (Orellana et al., 2012).

III. RESULTADOS

Del conjunto de datos, filtramos los viajes en bicicleta que inician únicamente en los sectores censales de la zona urbana de Cuenca, sin distinción del tipo de viaje. Obtuvimos 4939 secuencias (viajes), cuyas trayectorias se componen de 3 a 429 localidades; es decir, existen viajes en los que un ciclista cruza por 429 sectores censales. Se identificó que, a pesar que los viajes inician en la zona urbana, muchos de ellos realizan parte de su trayectoria, por la zona periurbana de Cuenca. Como se puede observar en la Figura 2, al visualizar en conjunto los viajes registrados, su análisis se dificulta, razón por lo cual fue necesario aplicar las técnicas de minería de datos para extraer información útil.

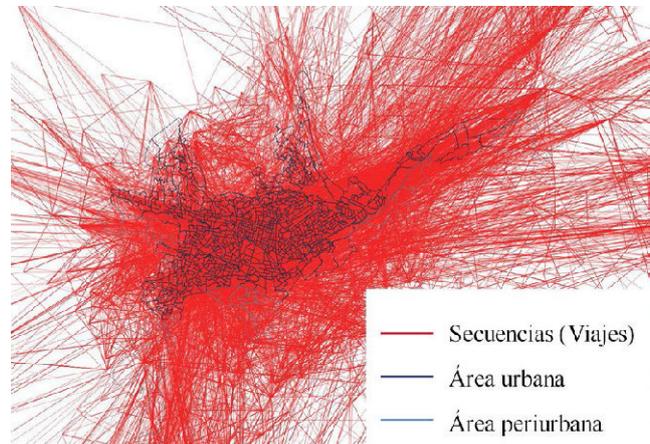


Figura 2. Mapa de viajes en bicicleta originados en la zona urbana de Cuenca y registrados por la aplicación móvil Strava.

En cuanto a la relación entre la frecuencia de origen de viajes y el Índice de Condiciones de Vida (ICV) en los sectores censales, se evidenció una correlación positiva leve pero estadísticamente significativa, lo cual confirmaría el sesgo socioeconómico del conjunto de datos (Figura 3). Los sectores con mayores condiciones de vida son también aquellas donde se originan el mayor número de viajes (Figura 4).

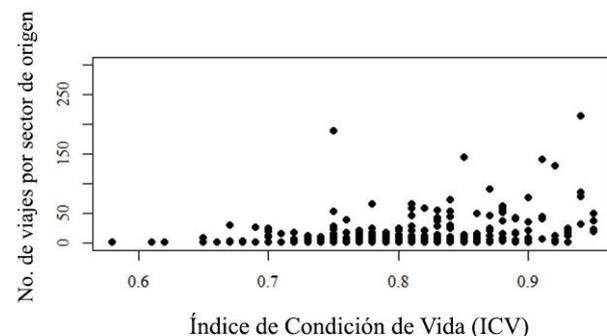


Figura 3. Relación entre la frecuencia de los sectores origen y su ICV ($r = 0,31$, p -valor $< 0,0001$).

El algoritmo BIDE+ extrajo 1365 patrones secuenciales generalizados (GSP) con un soporte mínimo de 0.10. (Figura 5), caracterizados por tener un tamaño de 8 localidades (Figura 6), y un valor de soporte inferior a 0,3 (Figura 7). En el mapa de

flujos, se evidencia que los GSPs se ubican al noreste (Sectores de Monay, Capulisipamba, Challuabamba,) y al sur (Sectores de El Ejido y Control Sur) de la zona urbana de Cuenca, extendiéndose hacia algunos sectores de su periferia. En relación a los patrones de movimiento de origen y destino, se identificó 316 rutas compuestas por dos localidades (correspondientes al origen y destino), de las cuales las más frecuentes también se concentran en la zona noreste de la ciudad, y se direccionan al norte y sur, y hacia sectores externos al límite urbano (Figura 8).

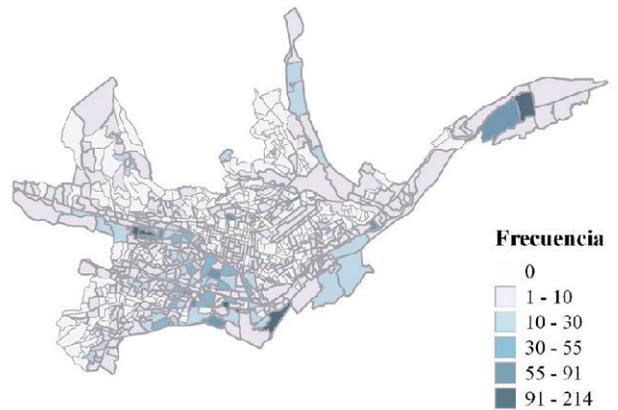


Figura 4. Mapa de los sectores de Cuenca donde se inician los viajes en bicicleta

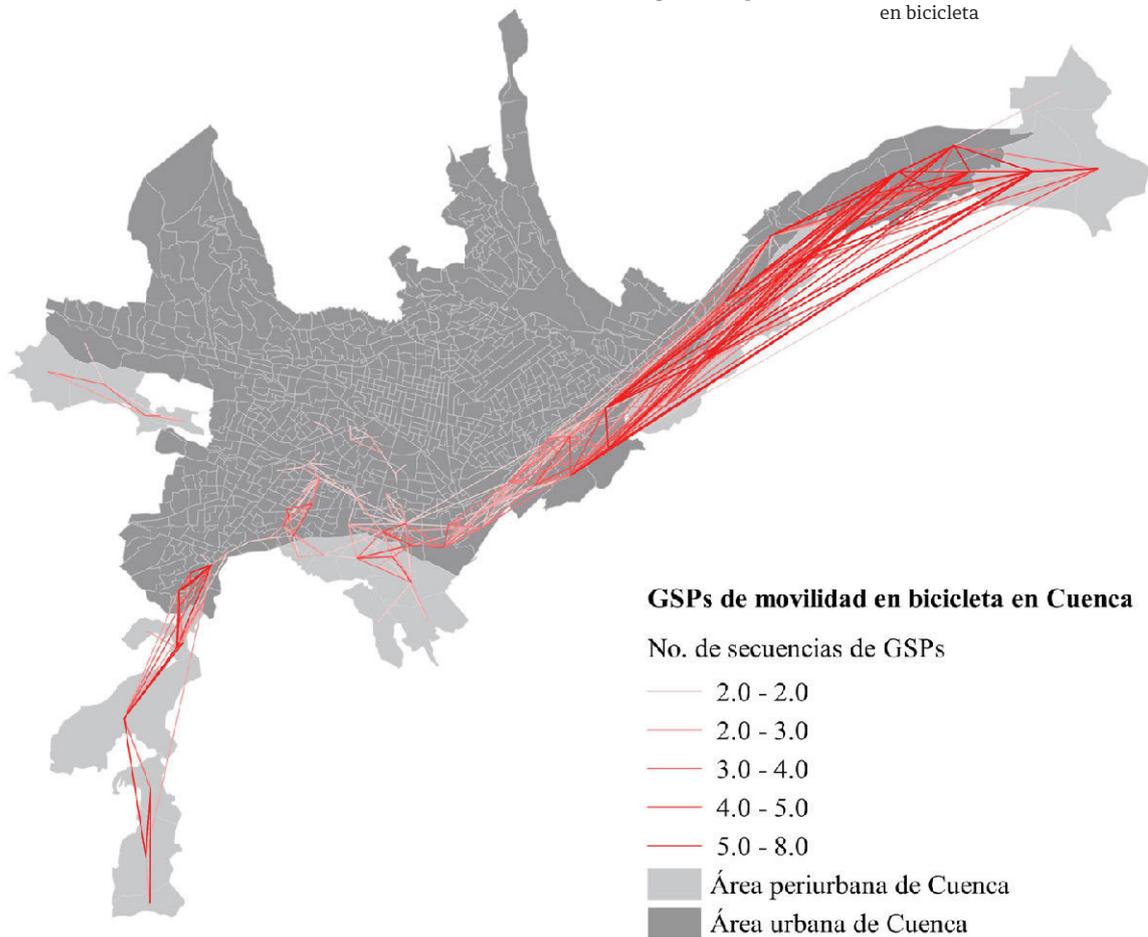


Figura 5. Mapa de Patrones Secuenciales Generalizados de la movilidad en bicicleta en Cuenca.

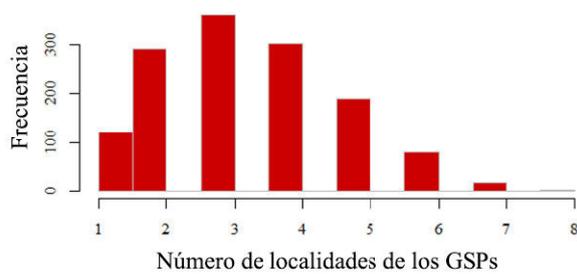


Figura 6. Número de localidades de los Patrones Secuenciales Generalizados y su frecuencia

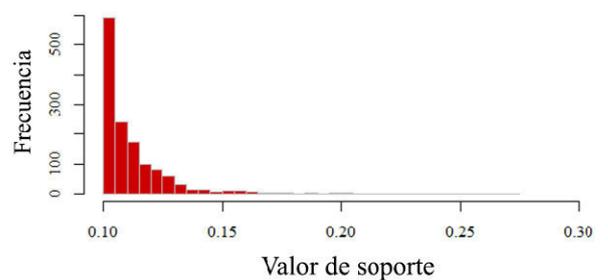


Figura 7. Frecuencia del valor de soporte de los Patrones Secuenciales Generalizados.

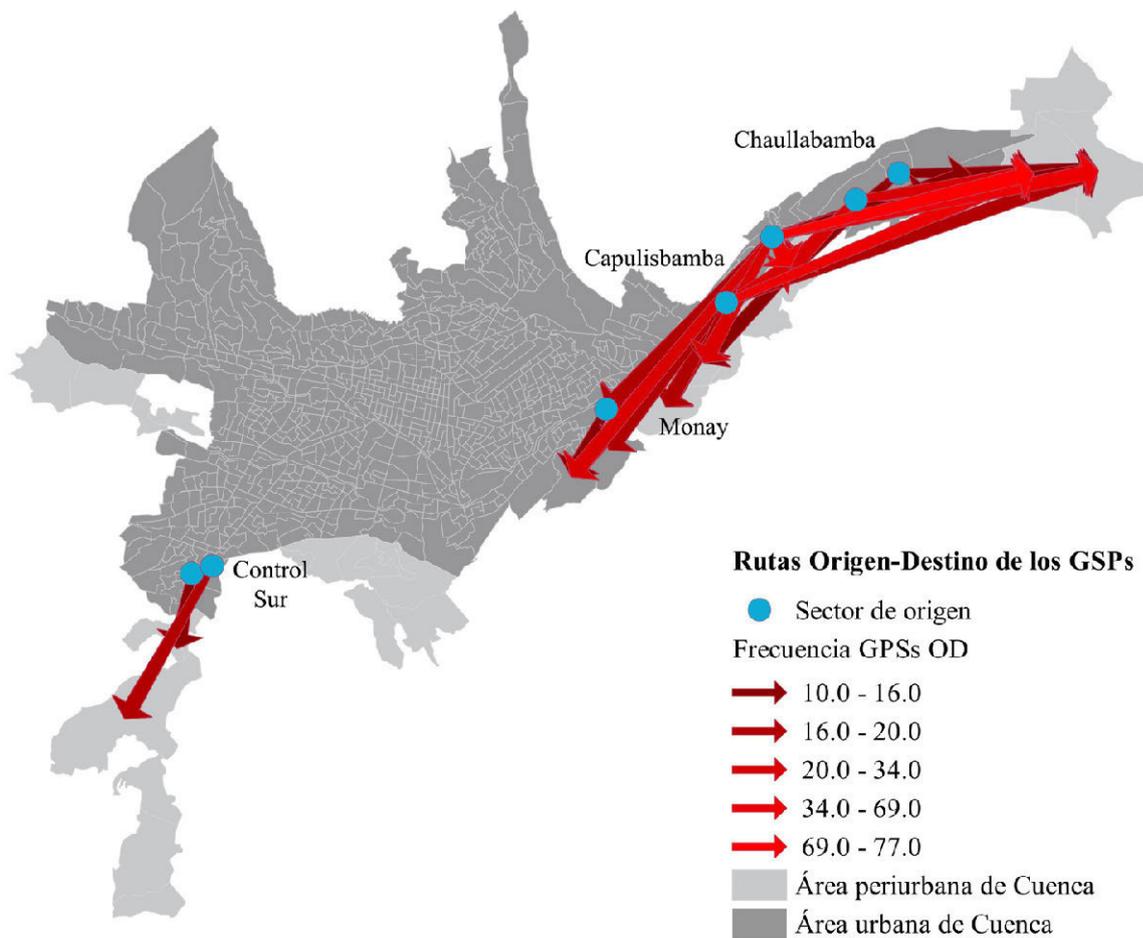


Figura 8. Rutas origen-destino más frecuentes identificadas en los patrones secuenciales generalizados de la movilidad en bicicleta en Cuenca.

IV. CONCLUSIONES

En este artículo presentamos una metodología para detectar patrones secuenciales generalizados de movilidad de ciclistas a partir de datos crowdsourcing. La extracción y pre-procesamiento de estos datos produjo información descriptiva de la movilidad en bicicleta, tal como el número de viajes en la zona urbana de Cuenca que se han registrado en la aplicación Strava durante el periodo de un año, y el tamaño de los trayectos de acuerdo al número de sectores censales que han cruzado en cada viaje. Al explorar la relación del ICV de los sectores censales con el número de actividades en bicicleta hemos encontrado una correlación estadísticamente significativa, lo cual confirmaría un sesgo socioeconómico en los datos que se debe tener en cuenta para la interpretación de los patrones. Aún así, los patrones detectados en los datos de Strava podrían ser indicadores de la cantidad de actividad ciclística en general (Jestico et al., 2016).

Por otro lado, los patrones secuenciales generalizados (GSPs) detectados, describen aspectos comunes de las trayectorias registradas por el colectivo de ciclistas en la aplicación Strava. De acuerdo a los valores de soporte, tamaño y frecuencia de los GSPs, las rutas más comunes de la movilidad de ciclistas se caracterizan por ser trayectos cortos, que carecen de

localidades intermedias destacadas. El valor de soporte bajo, también indica que la preferencia de rutas de los ciclistas es diversa.

A diferencia de las rutas descritas por los datos crowdsourcing, los GSPs permiten visualizar y localizar en el área de estudio, las rutas comunes de la movilidad de ciclistas, lo que ofrece un acercamiento a las zonas urbanas de Cuenca claves para la planificación de infraestructura para la movilidad no motorizada en bicicleta. La metodología presentada permite obtener insumos relevantes del comportamiento espacial de los ciclistas en base a datos crowdsourcing, minería de datos y visualización, la cual es aplicable a conjuntos de datos provenientes de diferentes fuentes de recolección de información georeferenciada. Para aprovechar el potencial de esta metodología, y generar conocimiento sobre el comportamiento espacial colectivo de los ciclistas de una población específica, es necesario orientar esfuerzos para generar información georeferenciada de aquellos ciclistas que no tienen acceso a tecnologías móviles. En cuanto al sesgo que representan los datos de Strava, es importante profundizar en el estudio de los perfiles de usuarios potenciales de bicicleta en un contexto específico, para relacionar y definir el papel que cumplen los usuarios de Strava en ese contexto y el conocimiento que aportan al mismo.

AGRADECIMIENTOS

Este trabajo es parte del proyecto “Pies y Pedales, estudio de los patrones de movilidad de ciclistas y peatones para la movilidad sustentable”, financiado por la Dirección de Investigación DIUC de la Universidad de Cuenca. Los autores desean agradecer a los miembros del grupo de investigación LlactaLAB - Ciudades Sustentables del Departamento de Espacio y Población de la Universidad de Cuenca por su contribución.

REFERENCIAS

- Agrawal, R., & Srikant, R. (1995). Mining sequential patterns (pp. 3-14).
- BID. (n.d.). Cuenca, Ciudad Sostenible / Plan de Acción. Retrieved from <https://dl.dropboxusercontent.com/u/13114503/Cuenca%20Ciudad%20Sostenible.pdf>
- Colectivo Cuenca Ciudad para Vivir. (n.d.). Retrieved May 31, 2017, from [http://www.cccv.ec/download/20170316_EncuestaCCCV2016\(Sitio\).pdf](http://www.cccv.ec/download/20170316_EncuestaCCCV2016(Sitio).pdf)
- Fishman, E. (2016). Cyclist as transport. *Transports Reviews*, 36(1), 1-8.
- Fournier-Viger, P., Lin, J. C.-W., Gomariz, A., Gueniche, T., Soltani, A., Deng, Z., & Lam, H. T. (2016). The SPMF Open-Source Data Mining Library Version 2. In *Lecture Notes in Computer Science* (pp. 36-40).
- GAD Municipal, del C. C. (2015). Plan de Movilidad y Espacios Públicos: Plan Operativo de Bicicletas.
- Geng, L., & Hamilton, H. J. (2006). Interestingness measures for data mining. *ACM Computing Surveys*, 38(3), 9-es.
- Griffin, G. P., & Jiao, J. (2015). Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus. *Journal of Transport & Health*, 2(2), 238-247.
- Handy, S., van Wee, B., & Kroesen, M. (2014). Promoting Cycling for Transport: Research Needs and Challenges. *Transport Reviews*, 34(1), 4-24.
- Heesch, K. C., & Langdon, M. (2016). The usefulness of GPS bicycle tracking data for evaluating the impact of infrastructure change on cycling behaviour. *Health Promotion Journal of Australia: Official Journal of Australian Association of Health Promotion Professionals*. <https://doi.org/10.1071/HE16032>
- Iacono, M., Krizek, K. J., & El-Geneidy, A. (2010). Measuring non-motorized accessibility: issues, alternatives, and execution. *Journal of Transport Geography*, 18(1), 133-140.
- Jestico, B., Nelson, T., & Winters, M. (2016). Mapping ridership using crowdsourced cycling data. *Journal of Transport Geography*, 52, 90-97.
- Orellana, D., Bregt, A. K., Ligtenberg, A., & Wachowicz, M. (2012/6). Exploring visitor movement patterns in natural recreational areas. *Tourism Management*, 33(3), 672-682.
- Orellana, D., & Osorio, P. (2014). Segregación socio-espacial urbana en Cuenca, Ecuador. *Analitika*, 8(2), 27-38.
- Pucher, J., Dill, J., & Handy, S. (2010). Infrastructure, programs, and policies to increase bicycling: an international review. *Preventive Medicine*, 50 Suppl 1, S106-25.
- Romanillos, G., Austwick, M. Z., Ettema, D., & De Kruijff, J. (2015). Big Data and Cycling. *Transport Reviews*, 36(1), 114-133.
- Selala, M. K., & Musakwa, W. (2016). The Potential of Strava Data to Contribute in Non-Motorised Transport (NMT) Planning in Johannesburg. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B2, 587-594.
- Sims R., R. Schaeffer, F. Creutzig, X. Cruz-Núñez, M. D'Agosto, D. Dimitriu, M.J. Figueroa Meza, L. Fulton, S. Kobayashi, O. Lah, A. McKinnon, P. Newman, M. Ouyang, J.J. Schauer, D. Sperling, and G. Tiwari. (2014). Transport. In *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- Slimani, T., & Lazzez, A. (2013, November 2). Sequential Mining: Patterns and Algorithms Analysis. *arXiv [cs.DB]*. Retrieved from <http://arxiv.org/abs/1311.0350>
- Strava. (2015). Strava METRO. Comprehensive User Guide (Version Version 2.0 for 2015). Retrieved from http://ubdc.ac.uk/media/1323/stravametro_200_user_guide_withoutpics.pdf
- Wang, J., & Han, J. (2004). BIDE: efficient mining of frequent closed sequences. In *Proceedings. 20th International Conference on Data Engineering* (pp. 79-90).
- Zheng, Y. (2015). Trajectory Data Mining: An Overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3), 29.