

Estudio de la Red de Colectivos de la Ciudad de Buenos Aires

Fernando Cornes - Juan Herrera Mateos - Ignacio Sticco
Grupo 3

Redes Complejas con Aplicaciones a Sistemas Biológicos

12 de diciembre de 2018

1. Introducción

La movilidad brindada por el transporte urbano es uno de los problemas más importantes de una ciudad, e involucra muchos aspectos en lo concerniente a los ciudadanos, autoridades y el crecimiento económico de los países. El acceso al transporte es un factor fundamental para el desarrollo económico de un ciudadano [1].

En el marco de las redes complejas aplicadas a los sistemas de transporte del mundo real, estos grafos pueden consistir en millones de nodos vinculados por un conjunto complicado de enlaces que depende del tipo de red [4]. La teoría de redes, las técnicas de medición y el poder computacional han mejorado enormemente nuestra capacidad para analizar tales estructuras en los años recientes.

Basados en estadísticas realizadas por la Subsecretaría de Tránsito y Transporte de la Ciudad de Buenos Aires, las cuales indican que el colectivo es el medio de transporte predominante elegido entre los usuarios [3], centramos nuestra atención en la red de colectivos de esta ciudad.

Algunas de las cuestiones que intentaremos describir a lo largo de este trabajo son:

- ✓ Cómo construir una red compleja que describa el sistema de colectivos de la ciudad y permita determinar cuáles son las zonas de mayor conectividad.
- ✓ Determinar la importancia de distintas áreas de la ciudad en términos de medidas de centralidad vistas en la materia

2. Construcción de la red

Partimos un conjunto de datos publicados por el Ministerio de Modernización, Innovación y Tecnología de la Ciudad de Buenos Aires. Los datos se encuentran en formato CSV y cuentan con información de los 1060 recorridos de colectivos que pasan por la ciudad de Buenos Aires. Los campos son: coordenadas del recorrido, id, línea, tipo de servicio, ramal y sentido. Las coordenadas (latitud y longitud) vienen dadas en unidades de grados decimales. [2].

Si bien hay varias formas de obtener una red que permita visualizar el sistema de colectivos de la ciudad, puesto que nuestro principal objetivo es analizar la conectividad que hay entre distintas zonas de la Ciudad de Buenos Aires, el procedimiento implementado es el siguiente:

- ✓ Tomamos la menor superficie rectangular que contiene a CABA.

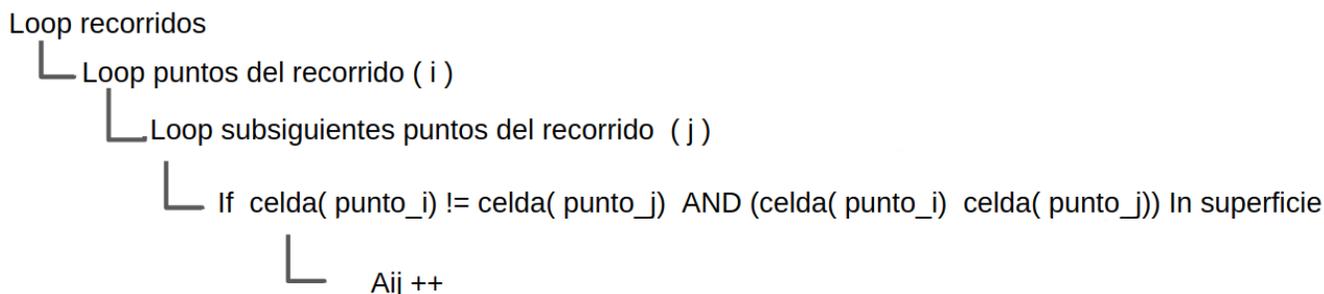
- ✓ Creamos una función tal que dado un $n \in \mathbb{N}$, crea una grilla de $n \times n$ sobre el rectángulo que contiene a CABA. Cada celda de la grilla es un potencial nodo de la red.
- ✓ Creamos una matriz de adyacencia (vacía) de tamaño $N \times N$ donde $N = n \times n$.
- ✓ Completamos la matriz de adyacencia con los valores correspondientes. Esta matriz en su elemento A_{ij} contiene la cantidad de recorridos que van desde la celda i hasta la celda j . El algoritmo utilizado para completar los valores de la matriz de adyacencia es el siguiente:
 - Primero tenemos un loop sobre recorridos. En cada recorrido hacemos un loop sobre los distintos puntos (es decir las distintas coordenadas geográficas del recorrido).
 - Luego tenemos otro loop que recorre los puntos subsiguientes del recorrido. Es decir, tomamos 2 puntos: (**punto_i** y **punto_j**). A cada punto le calculamos la celda de pertenencia y nos fijamos si la celda a la cual pertenece el **punto_i** es distinta a la celda a la cual pertenece el **punto_j** y además nos fijamos si ambos puntos se encuentran en la superficie de interés (rectángulo que contiene a CABA).
 - Si se cumplen las condiciones mencionadas, entonces incrementamos 1 al elemento A_{ij} de la matriz de adyacencia. En la [Figura 1](#) se muestra un esquema en pseudocódigo de este algoritmo.

Hasta aquí conseguimos la matriz de adyacencia correspondiente a todas las celdas del menor rectángulo que contiene a CABA. Lo que hicimos luego fue eliminar todas aquellas componentes de la matriz de adyacencia que están asociadas a zonas que no pertenecen a CABA (conurbano y Río de la Plata). Para esto fue necesario identificar el límite de la ciudad. Con el Google Earth Pro trazamos el límite de la ciudad de Buenos Aires identificando 18079 puntos (la separación entre un punto y otro corresponde a una distancia real de menos de 15 m). Una vez conseguidos los puntos del contorno, los vinculamos a las correspondientes celdas de la grilla $n \times n$ (rectángulo que contiene CABA). Usamos un algoritmo de *edge detection* para quitar todas las celdas que no corresponden a CABA (nos quedamos solo con las celdas del interior y el límite de la ciudad).

Una vez eliminadas las celdas que no pertenecen a CABA, nos quedamos con todos los elementos de la matriz de adyacencia que corresponden a las interacciones entre dos celdas de la ciudad.

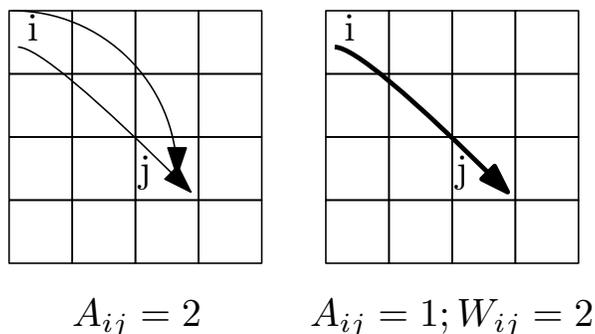
En términos teóricos, existe un enlace entre dos celdas cualesquiera por cada recorrido de colectivo que vincule geográficamente las respectivas celdas. Excluyendo auto-enlaces o *loops*, se tendrá un multigrafo dirigido cuya matriz de adyacencia no simétrica tiene elementos diagonales nulos y no diagonales $A_{ij} \geq 0$. Éste puede representarse equivalentemente (ver [Figura 2](#)) por medio de un grafo dirigido y pesado donde el peso de cada enlace entre dos nodos corresponde al número de vértices que los unen en el multigrafo. Esta representación es más conveniente de utilizar ante algunas funciones de Networkx. Hay que destacar que además de tener una matriz de adyacencia A_{ij} binaria se tiene una matriz de pesos W_{ij} .

Figura 1



Usamos la librería de Python Networkx para construir un grafo a partir de la matriz de adyacencia A (y la matriz de pesos W). Usando funciones implementadas de dicha librería, fuimos capaces de extraer varias propiedades de la red (reciprocidad, average shortest path, densidad, etc.).

Figura 2. Ejemplo de representación de un multigrafo dirigido por medio de un grafo pesado y dirigido



3. Resultados y discusión

3.1. Características de la red

Todos los resultados analizados corresponden a un grillado de 60×60 (equivale a una celda real de $300 \text{ m} \times 300 \text{ m}$). Usamos este grillado porque las celdas eran lo suficientemente chicas como para cubrir la Ciudad de Buenos Aires con una buena cantidad de nodos y eran suficientemente grandes como para que incluya una cantidad razonable de recorridos asociados a cada celda (de modo que la matriz de adyacencia no resulte muy esparza). En la [Tabla 1](#) se muestra un listado de las características básicas de la red.

Tabla 1. Características de la red utilizada

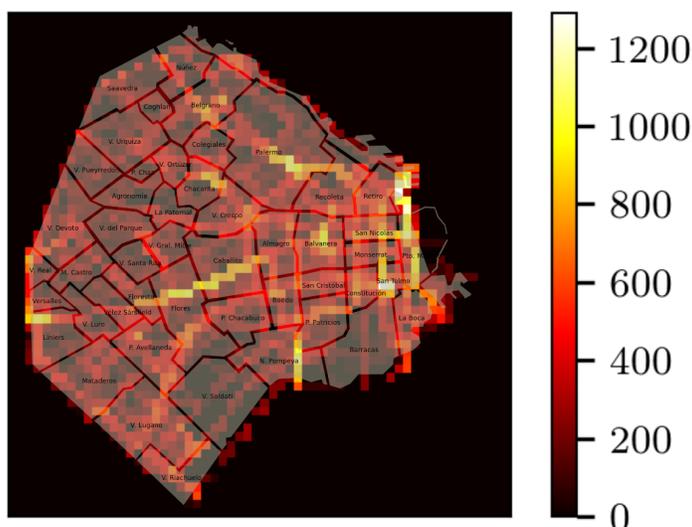
Red	60×60
N (total de nodos)	2066
Componente gigante	1787
Average shortest path	1,84
Reciprocidad	67%
Diámetro	3
Densidad	0,1

3.2. Representación de la red

En primer lugar, se realizó un mapa de grado de la red de transporte con el fin de establecer una correspondencia entre lo que uno conoce “a priori” respecto de las zonas de la Ciudad de Buenos Aires por donde pasan más recorridos y el resultado obtenido al analizar la red. En la [Figura 3](#) se muestra el mapa de la Ciudad de Buenos Aires donde se indican los barrios superpuesto al mapa de grado.

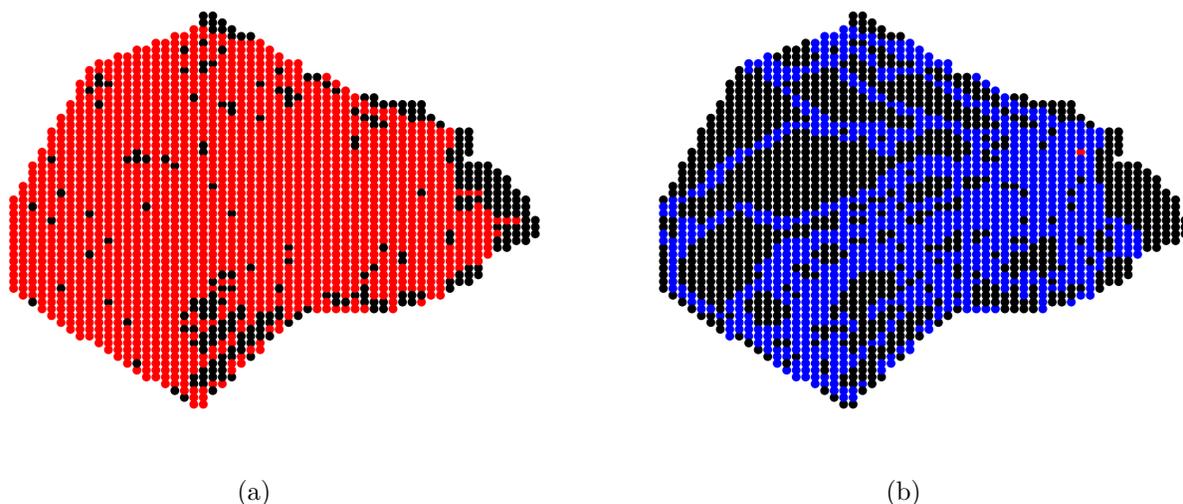
Como puede observarse, las zonas de Retiro, Constitución, Liniers, Plaza Italia y Av. Rivadavia son algunas de las zonas de mayor grado. Es decir, zonas por donde pasan muchos recorridos. Por otro lado, Soldati es una de las zonas de la capital con menos recorridos de colectivos. Cabe destacar, que en dicha zona hay muchos espacios verdes como por ejemplo el parque Indoamericano.

Figura 3. Mapa de grado de la red de transporte de colectivos superpuesto con el mapa de la Ciudad de Buenos Aires. La intensidad del color de cada nodo es proporcional a su grado (ver escala de colores). La red utilizada fue de 60×60 nodos, lo que equivale a un tamaño de celda de $300 \text{ m} \times 300 \text{ m}$.



Con el fin de continuar caracterizando la red de transporte, en la [Figura 4\(a\)](#) se representan en diferentes colores los nodos con y sin grado. Es decir, los lugares (celdas) por donde pasa al menos un recorrido se indican en color rojo mientras que en color negro se representan aquellas celdas por donde no pasa ningún recorrido. Como se mencionó anteriormente, la zona de Soldati (parte inferior de la [Figura 4\(a\)](#)) posee muchas zonas verdes (parques) y eso se refleja en la gran cantidad de nodos de color negro. Por otro lado, al este de la Capital Federal puede observarse una gran cantidad de nodos de grado nulo debido a que la misma corresponde a la Reserva Ecológica. Finalmente, cabe destacar que la mayor parte de los nodos de la Ciudad de Buenos Aires posee grado diferente de cero. Es decir, existen pocas zonas por donde no pasan colectivos.

Figura 4. Esquema de la red de transporte de acuerdo al grado de cada celda: (a) Los nodos con cero grado se indican en color negro, mientras que los nodos con grado no nulo se representan en color rojo. (b) En rojo se indica el nodo correspondiente a la zona de Retiro. Los nodos de color azul son aquellos que se encuentran conectados con Retiro por al menos un recorrido. En negro se indican todos aquellos nodos desde los cuales no es posible llegar desde Retiro. Ambos gráficos corresponden a una red de 60×60 nodos, lo que equivale a un tamaño de celda de $300 \text{ m} \times 300 \text{ m}$.



Como se mencionó anteriormente y se puede observar en la [Figura 3](#), la zona de Retiro es uno de los lugares de la Capital Federal por donde pasan la mayor cantidad de recorridos. Por lo tanto, se analizó el grado de conexión de Retiro con los demás lugares de la capital. En la [Figura 4\(b\)](#) se indican en color azul los lugares a los que es posible llegar desde Retiro, el cual se encuentra representado en color rojo. Como puede observarse, gran parte de la Ciudad de Buenos Aires se encuentra conectada con Retiro. En particular, la mayor cantidad de recorridos que salen desde Retiro pasan por el sur de la capital, mientras que pocas zonas al oeste de la misma son accesibles desde Retiro.

3.3. Caracterización de la red en términos de centralidad

A continuación se detallan los análisis realizados sobre la red de transporte en términos de centralidad. Centralidad por grado (saliente), por autovector, por cercanía y por intermediación corresponden a las medidas de centralidad estudiadas. Para poder comparar las diferentes medidas, los mapas de centralidad se normalizaron usando el máximo valor de la centralidad correspondiente. Por lo tanto, cabe destacar que los análisis que se realizan a continuación tienen carácter cualitativo y no cuantitativo.

3.3.1. Centralidad por grado

En la [Figura 5\(a\)](#) se muestra el mapa de centralidad por grado. En este caso, se graficó el grado saliente de cada nodo. Cabe recordar que la red de transporte es una red dirigida, por lo tanto cada nodo posee grado entrante y saliente. Sin embargo, en este caso se graficó el grado saliente. No se incluyó el gráfico correspondiente al grado entrante debido a su similitud con la [Figura 5\(a\)](#). La semejanza entre ambos gráficos se debe a que los recorridos de “ida y vuelta” en general pasan por los mismos lugares. Esto se refleja en el valor de

la reciprocidad de la red (67% según la [Tabla 1](#)).

Como se mencionó anteriormente en la [Figura 3](#), las zonas de Retiro, Constitución, Liniers, parte de Av. Rivadavia, etc. poseen alto grado. Por otro lado, gran parte del sur de la Capital Federal (Barracas y La Boca) son algunas de las zonas de menor grado. Finalmente, puede observarse como por la zona del Macrocentro de la ciudad pasan muchos recorridos.

3.3.2. Centralidad por autovector

La centralidad por autovector considera que un nodo es importante si tiene vecinos importantes. Por lo tanto, esta medida de centralidad se corresponde con la idea de que “la importancia se propaga por la red”. En la [Figura 5\(b\)](#) se muestra el mapa de centralidad por autovector. En la misma puede verse que la zona correspondiente a Plaza Italia es la región de la Ciudad de Buenos Aires de mayor centralidad. Al verificar este resultado mediante la aplicación “Como llego” del Gobierno de la Ciudad de Buenos Aires, pudimos corroborar que muchos de los colectivos que pasan por Plaza Italia llegan tanto a Constitución como a Retiro. Es decir, Plaza Italia conecta con nodos importantes, por lo que su centralidad por autovector es alta.

El siguiente punto de la Capital Federal de mayor centralidad corresponde a Constitución. El mismo conecta con Retiro y Plaza Italia a través de varios recorridos. Otras zonas de gran centralidad por autovector son la mayor parte de Av. Rivadavia, Puente Saavedra, el Obelisco y la Av. 9 de Julio.

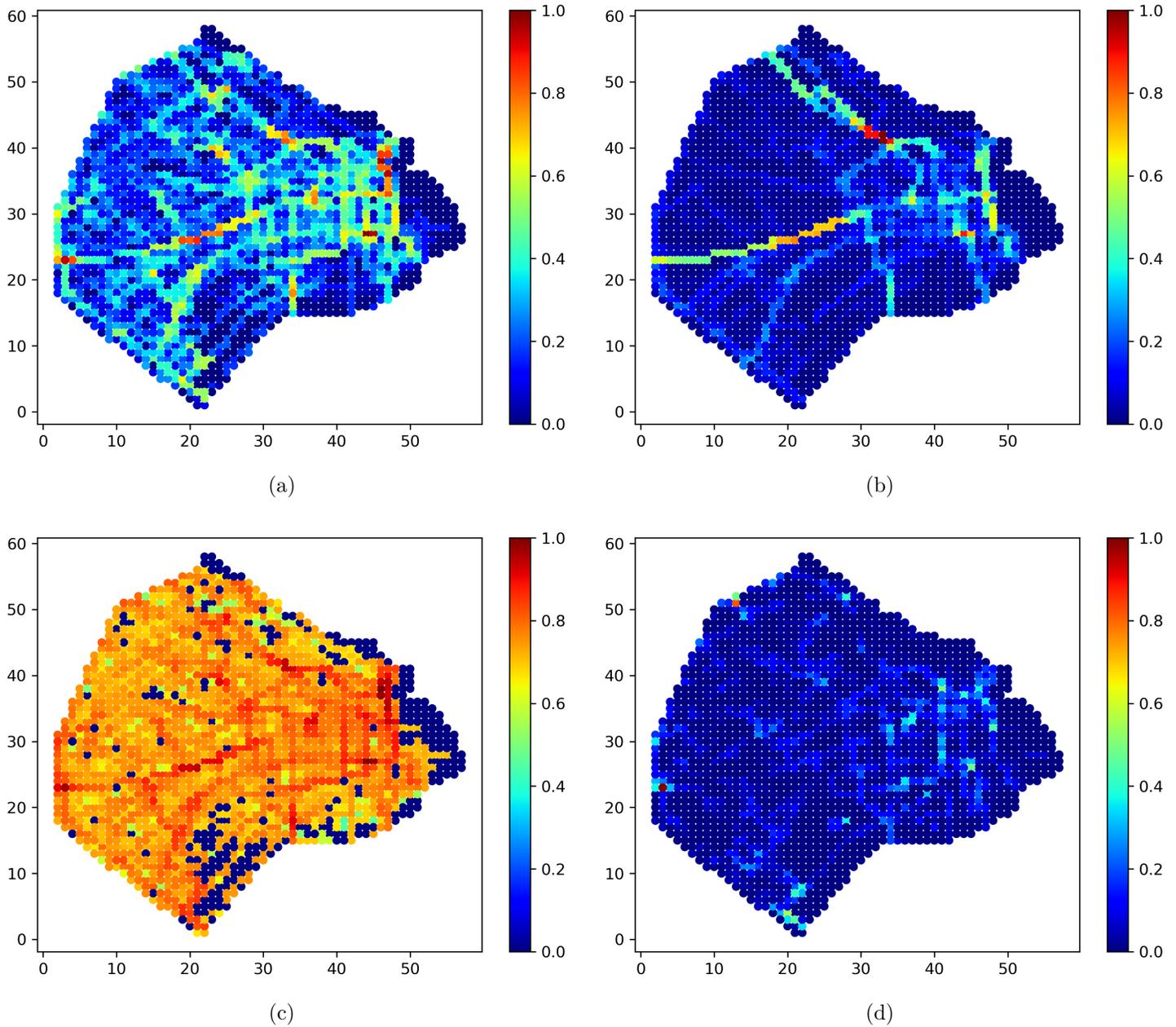
3.3.3. Centralidad por cercanía

La [Figura 5\(c\)](#) corresponde al mapa de centralidad por cercanía (closeness). En esta medida de centralidad un nodo es importante si es fácil acceder a él. Es decir, si hay muchos caminos posibles que conectan cualquier nodo con este nodo. Como puede observarse en la figura, la centralidad en la mayor parte de la ciudad de Buenos Aires es (aproximadamente) homogénea. Es decir, cualquier par de puntos de la Capital Federal puede ser alcanzados con la misma “dificultad” que cualquier otro.

3.3.4. Centralidad por intermediación

En la [Figura 5\(d\)](#) se presenta el mapa de centralidad por intermediación (betweenness). Recordemos que un nodo que posee un valor alto de centralidad está asociado a la idea de ser un “puente” que conecta dos regiones. En el gráfico puede verse que salvo Liniers y Puente Saavedra, la mayor parte de la ciudad de Buenos Aires posee baja centralidad. Esto significa que existen varios caminos alternativos para ir desde un punto dado a otro. En cambio, Liniers y en menor medida Puente Saavedra son “pasos obligados” para entrar o salir de Capital Federal. Por lo tanto, su valor de centralidad por intermediación es alto comparado con los demás nodos.

Figura 5. Caracterización de la red de transporte en términos de centralidad: (a) Centralidad por grado (saliente); (b) Centralidad por autovector; (c) Centralidad por cercanía; y, (d) Centralidad por intermediación. Todos los gráficos están normalizados por el máximo valor correspondiente con cada centralidad. Todos los gráficos corresponden a una red de 60×60 nodos, lo que equivale a un tamaño de celda de $300 \text{ m} \times 300 \text{ m}$.



3.3.5. Comparación centralidad por grado y por autovector

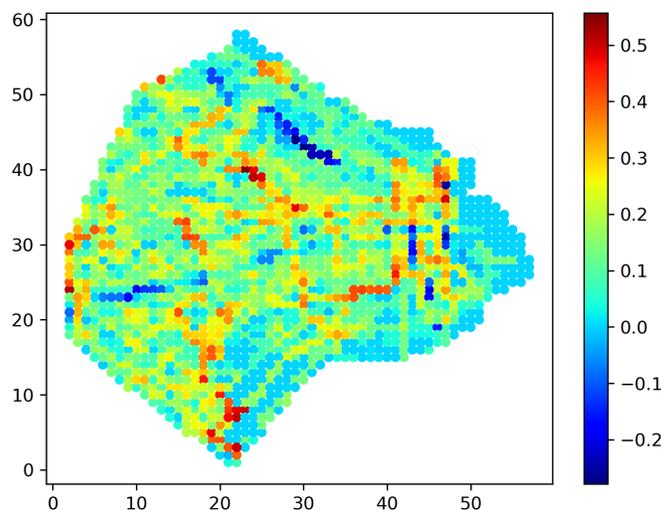
Finalmente y con el propósito de realizar una comparación más exhaustiva entre dos medidas de centralidad, se analizó si un nodo es más central por grado que por autovector o viceversa. Para ello, se realizó la resta entre los mapas de centralidad por grado (Figura 5(a)) y por autovector (Figura 5(b)). El resultado obtenido se muestra en la Figura 6. Los valores positivos corresponden a una mayor centralidad por grado que por autovector, mientras que los valores negativos corresponden a lo opuesto (alta centralidad por autovector

en comparación con la de grado).

Como puede observarse, Av. Cabildo posee mayor centralidad por autovector que por grado. Esto significa (y se pudo comprobar mediante la aplicación “Como llego”) que gran parte de los colectivos que pasan por Av. Cabildo conectan con Retiro y Constitución y eso “pesa” más que la cantidad de recorridos que pasan por allí (a pesar de que son bastantes). Esto también ocurre en la parte más al oeste de Av. Rivadavia y en la zona de Constitución.

Otro resultado interesante es la zona de Chacarita en donde ocurre el caso contrario. El grado pesa mas que el autovector. Esto ocurre debido a que esta zona de la ciudad conecta muy poco con zonas importantes (Ej. Retiro y Constitución) pero pasan muchos colectivos por ahí. Cabe destacar que esto se pudo corroborar con la app del Gobierno de la Ciudad. Es decir, por ejemplo hay muy pocas opciones para ir desde Chacarita hasta Retiro, a pesar de que por esa zona pasan muchos colectivos.

Figura 6. Comparación entre la centralidad por grado y por autovector. Los valores positivos indican que la centralidad por grado es mayor a la de autovector. En cambio, los valores negativos de la escala corresponden a una mayor centralidad por autovector que por grado. El gráfico corresponde a una red de 60×60 nodos, lo que equivale a un tamaño de celda de $300 \text{ m} \times 300 \text{ m}$.



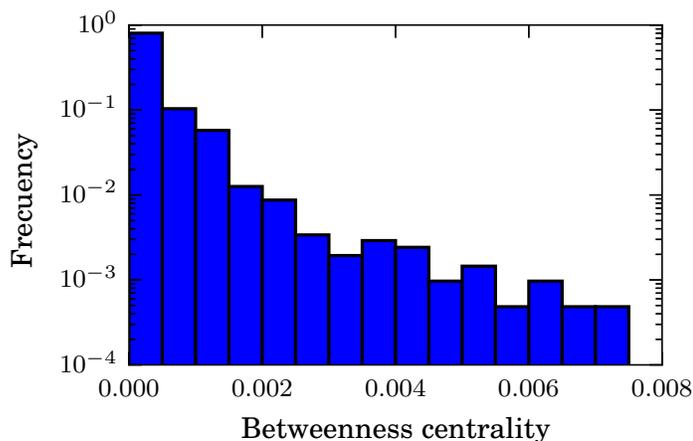
3.4. Distribuciones

En la [Figura 7](#) mostramos las distribuciones de centralidad. El gráfico [7\(a\)](#) muestra la distribución de la centralidad por betweenness. Análogamente, la figura [7\(b\)](#) muestra la distribución de los nodos según centralidad de autovector. Las figuras [7\(c\)](#) y [7\(d\)](#) muestran la distribución de grado. Todos los histogramas tienen su eje de frecuencia normalizado y en escala logarítmica. La figura [7\(d\)](#) cuenta también con el eje del grado (eje x) en escala y binned logarítmico.

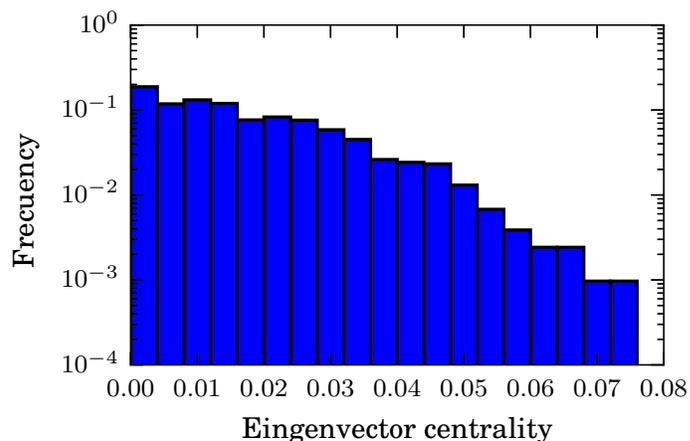
Los nodos de la red poseen baja centralidad por betweenness. Esto ocurre porque para ir de un lugar a otro en la ciudad existen varios caminos alternativos (no hay “pasos obligados”). La distribución por autovector muestra menores variaciones que la distribución por grado. En la centralidad por autovector, un nodo es importante si conecta con nodos importantes. En el caso de la red analizada, los nodos con mayor centralidad por autovector conectan con nodos importantes pero también conectan con nodos poco importantes. Esto

hace que no haya grandes diferencias en los valores de centralidad por autovector en la red. En cuanto a la distribución por grado, la figura 7(d) se encuentra en escala logarítmica y con binnedo logarítmico. Podemos ver que la distribución no sigue una ley de potencias. La distribución parece tener una media y varianza bien definidas.

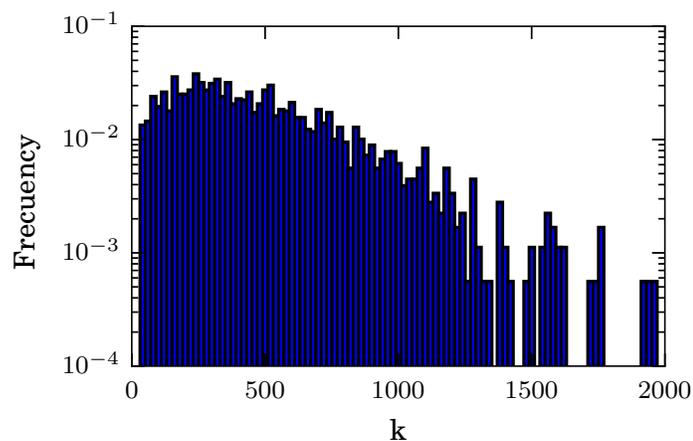
Figura 7. Distribución de centralidad: (a) Distribución por betweenness; (b) Distribución por autovector; (c) Distribución por grado; y, (c) Distribución por grado en escala logarítmica



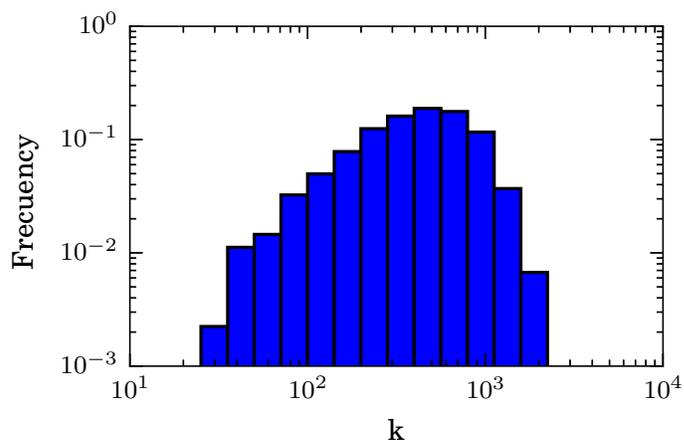
(a)



(b)



(c)



(d)

4. Conclusión

En este trabajo estudiamos la conectividad de distintas zonas de la ciudad de Buenos Aires en términos de transporte. En particular, se analizó el sistema de colectivos. Pudimos crear una red compleja que lo represente, de la cual se pudo determinar sus principales características. Entre ellas destacamos: una componente gigante de 1787 nodos conectados sobre el total de 2066 nodos, un *average shortest path* de 1,84 el cual indica que en promedio para llegar a dos nodos cualesquiera de la red, deben abordarse casi dos colectivos. La red tiene un diámetro de valor 3, esto implica que hay dos celdas a las que se llega por medio de tres colectivos. Por otro lado, determinamos que la red tiene una reciprocidad del 67% lo que implica que la matriz de adyacencia que representa la red es muy simétrica. Esto se debe a que los recorridos ida y vuelta hacen caminos similares

(en términos geográficos). Determinamos que si bien la mayor cantidad de los nodos de la ciudad tienen grado no nulo, existen algunas zonas urbanizadas en el sur de la capital por donde no pasan recorridos.

Por otro lado, determinamos la importancia de distintas áreas según medidas de centralidad aprendidas a lo largo de la materia. En ese sentido encontramos que la centralidad por cercanía es aproximadamente homogénea, lo cual refleja que las distintas zonas de la ciudad de Buenos Aires son fácilmente accesibles. La centralidad por intermediación de distintos nodos resultaron bajas a excepción de los correspondientes a Liniers y en menor medida a Puente Saavedra, los cuales son accesos o egresos a CABA (“pasos obligatorios”). Finalmente comparamos dos medidas de centralidad diferentes (grado y autovector). Encontramos que en algunas zonas de capital predomina la centralidad por grado sobre la centralidad por autovector (Chacarita). También encontramos que ocurre lo contrario; es decir, existen lugares de la capital que poseen bajo grado (no conectan con demasiados) puntos de la ciudad pero que conectan con zonas importantes (Av. Cabildo).

Referencias

- [1] Raj Chetty y Nathaniel Hendren. *The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects*. Inf. téc. Dic. de 2016. DOI: [10.3386/w23001](https://doi.org/10.3386/w23001). URL: <https://doi.org/10.3386/w23001>.
- [2] Ministerio de Modernización Innovación y Tecnología - SS de Innovación y Ciudad Inteligente - DG de Gestión Digital - Unidad de Sistemas de Información Geográfica (USIG). *Paradas de Colectivo*. 2015. URL: <https://data.buenosaires.gob.ar/dataset/colectivos>.
- [3] La Nación. *La mitad de los viajes en la Capital y el conurbano se hace en transporte público*. 2015. URL: <https://www.lanacion.com.ar/1821690-la-mitad-de-los-viajes-en-la-capital-y-el-conurbano-se-hace-en-transporte-publico>.
- [4] Harold Soh y col. “Weighted complex network analysis of travel routes on the Singapore public transportation system”. En: *Physica A: Statistical Mechanics and its Applications* 389.24 (dic. de 2010), págs. 5852-5863. DOI: [10.1016/j.physa.2010.08.015](https://doi.org/10.1016/j.physa.2010.08.015). URL: <https://doi.org/10.1016/j.physa.2010.08.015>.