

Redes sociales basadas en ubicación en Buenos Aires (2009-2015)

Leonardo Salvador Rocco¹, Marcelo A. Soria²

¹ Maestría en Explotación de Datos y Descubrimiento de Conocimiento. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires

² Facultad de Agronomía. INBA-CONICET. Universidad de Buenos Aires
soria@agro.uba.ar

Resumen. El tipo de redes sociales que se basan en la ubicación de sus usuarios recibe el nombre de redes sociales basadas en localización (LBSN) y son un medio oportuno para el análisis del comportamiento espacio temporal de las personas. Utilizando datos registrados en Foursquare, aplicación para dispositivos móviles que provee a sus usuarios búsquedas personalizadas y localizadas, se construyó la red social de usuarios con actividades en lugares de la Ciudad Autónoma de Buenos Aires entre 2009 y 2015. En este trabajo se describen en primer lugar aspectos metodológicos para la recolección y procesamiento de datos de redes sociales utilizando información pública, y en segundo lugar se estudia las características estructurales de la red social que componen estos usuarios. Entre los hallazgos más importantes se encuentra una estructura fuertemente comunitaria, de mundo pequeño y con un grado neutral de afinidad selectiva, que no se enmarca en una distribución de grados de ley de potencias.

Abstract. The type of social networks that are based on the location of their users is called location-based social networks (LBSN). These networks are an adequate means for spatio-temporal users' behavior analysis. Using data from Foursquare, an application for mobile devices that provides its users with personalized and localized searches, the social network for users was built based on activities within Buenos Aires Federal District between 2009 and 2015. In this paper in the first place the methodological aspects for the collection and processing of social network data using public information is described. Secondly the structural characteristics of this social network are analyzed. Among the most relevant findings, we could see that the network has characteristics of a strong communitary, small-world and neutral degree of assortativity structure. It was also found that this network does not fit the power-law degree.

Keywords: LBSN, Foursquare, social networks, geolocation, Buenos Aires, social network analysis

1 Introducción

Las redes sociales basadas en localización (LSBN) se basan en la ubicación de sus usuarios y funcionan como un medio propicio para la investigación de la sociabilidad en el cruce de lo on-line y el contexto off-line, así como la movilidad y patrones espacio temporales.

El objetivo de este trabajo es describir las características topológicas de las redes sociales que conforman los usuarios de la plataforma Foursquare con actividades geolocalizadas dentro de la Ciudad Autónoma de Buenos Aires (CABA). Se toma como marco temporal para el análisis al período comprendido entre diciembre del 2009 y marzo del 2015, debido a que el acceso a los datos se volvió mucho más restrictivo desde ese entonces.

Este trabajo está parcialmente basado en los resultados de la tesis titulada “Redes sociales basadas en geolocalización en la Ciudad Autónoma de Buenos Aires (2009-2015)”, defendida en la Maestría en Explotación de Datos y Descubrimiento del Conocimiento de la Universidad de Buenos Aires.

1.1 Foursquare

La propuesta conceptual de la plataforma es la de recomendar *venues* (lugares) al usuario de acuerdo con su búsqueda, gustos, hábitos, necesidades, localización geográfica, ranking de los lugares y comportamiento de los usuarios amigos.

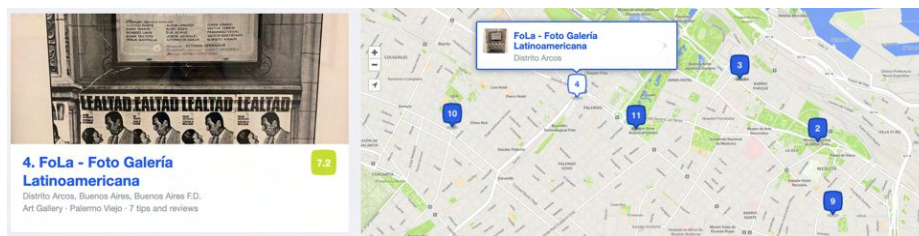


Fig 1. Ejemplo de lugares recomendados por la aplicación en base a una búsqueda y ubicación del usuario. En azul lugares recomendados y su ubicación.

En la mencionada LBSN los usuarios pueden registrarse y compartir su localización geográfica marcando su presencia (check-in) en un sitio específico (*venue*) que esté localizado físicamente cerca de donde se encuentra el usuario. Además de realizar check-in en *venues*, los usuarios pueden leer y publicar comentarios de sus experiencias, así como establecer vínculos en la plataforma con otros usuarios.

1.2 Extensión de las redes sociales localizadas

Para el periodo 2009-2015 considerados en este trabajo, se puede dar cuenta de algunos indicadores claves sobre el uso de dispositivos móviles e internet para comprender la extensión de esta aplicación diseñada principalmente para dispositivos móviles.

A nivel local, tomando en cuenta datos del Banco Mundial [1] podemos observar que la proporción de hogares que poseían telefonía móvil en Argentina y que hicieron uso de Internet desde el mismo fue del 48% para el año 2010. Si bien no existen datos respecto a la penetración de las redes sociales en la Argentina para el periodo analizado en el trabajo, se reconoce a Foursquare como la LBSN más extendida a nivel mundial durante el período considerado en este trabajo [2][3].

A nivel internacional, para el 2011 más del 12% de los usuarios de teléfonos móviles de Estados Unidos habían usado su teléfono para LBSN tales como Foursquare o Gowalla [4].

2 Métodos

La obtención de datos se realizó a través de la API de la plataforma. Debido a la complejidad del proceso y las limitaciones del servicio, la recolección de datos fue realizada a lo largo del periodo comprendido entre febrero del 2014 y marzo del 2015.

Se recorrió exhaustivamente el espacio de la Ciudad de Buenos Aires, recolectando todos los lugares registrados en la plataforma, para lo que fue necesario recurrir a información geográfica oficial con la localización de las 12373 áreas geográficas (manzanas) de la ciudad. A tal efecto se trabajó con la base de datos Manzanero [5]. En la figura 2 se puede observar el resultado de este proceso.

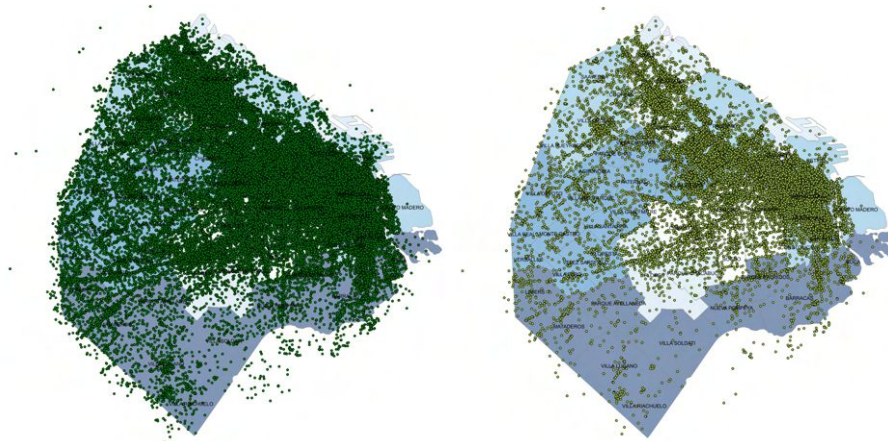


Fig. 2. Lugares de CABA recolectados inicialmente (izquierda) y distribución de los lugares considerados en el análisis (derecha).

Los puntos amarillos en la figura de la derecha son lugares que tienen comentarios y considerados en este trabajo por tal motivo. Como podemos ver, ambas distribuciones se asemejan espacialmente sobre las comunas de CABA, las cuales fueron segmentadas en 5 escalas de colores. Las escalas representan el nivel de densidad poblacional, en donde el tono más claro representa comunas de mayor densidad. Se observa que los lugares analizados se distribuyen en mayor medida en aquellos espacios con alta densidad poblacional.

2.2 Procesos y componentes de software

En la figura 3 podemos observar un resumen de los procesos a alto nivel y soportes de datos y software utilizados.

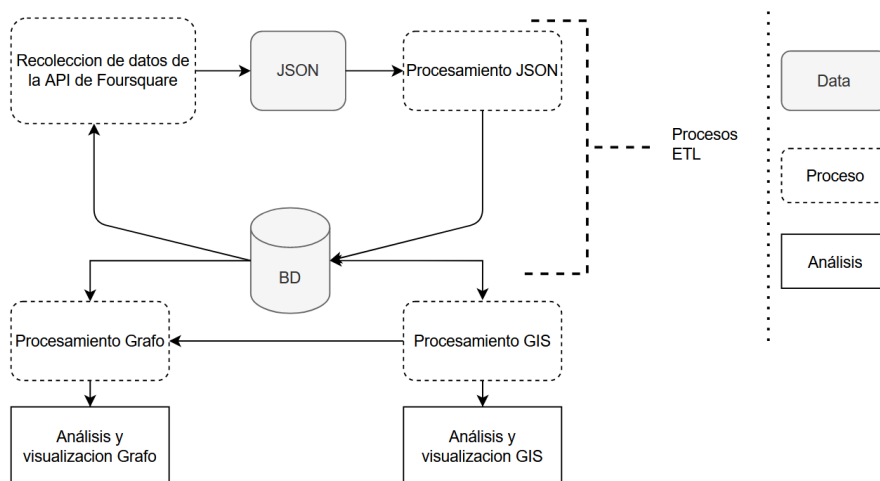


Fig. 3. Diagrama a alto nivel de procesos, información y análisis realizados.

En este trabajo se utilizó software libre en su totalidad tanto para la recolección, procesamiento como análisis de los datos. El uso conjunto de QGIS y PostGIS permitió manipular la base de datos de manzanas de la CABA, convertirla a coordenadas latitud-longitud y luego transformarla a una base de datos en PostgreSQL para que pueda hacerse cálculos espaciales necesarios para la interacción con la API de Foursquare. Para la integración de datos desde la API y orquestación de procesos se utilizó Talend Open Studio.

Respecto al software utilizado para el análisis de redes sociales, se escogió Python para el procesamiento, SQL y el software de exploración y visualización de redes

Gephi (<https://gephi.org/>). En términos de paquetes para trabajo con grafos y análisis de redes sociales, se utilizó principalmente NetworkX. En ambos casos los softwares fueron vinculados a la base de datos desarrollada en PostgreSQL para facilitar el procesamiento dinámico de los datos.

3 Procesamiento de datos

Una vez identificados todos los *venues* de la ciudad existentes en Foursquare, se procedió a la recolección de comentarios dejados en los mismos, las personas que habían dejado los comentarios y sus vínculos con otros usuarios.

Sobre el total de personas que dejaron comentarios en CABA, se procedió a excluir los outliers univariadamente en base a su puntuación transformado a una escala Z, descartando el 0,5% de los usuarios más alejado de la media de comentarios. Los bots y agentes de publicidad en Foursquare, así como otras redes sociales, han sido detectados y analizados en diferentes estudios [6], y se encuentran ampliamente extendidos [7].

Luego de la poda de usuarios con valores extremos, quedaron 31385 usuarios, con las cuales se conformó el grafo de relaciones entre usuarios comentaristas de *venues* de CABA. El proceso de depuración implicó también identificar a los usuarios únicos comentaristas de los lugares, deduplicar los lazos entre los mismos y finalmente identificar los componentes principales, por lo que se descartaron 5234 componentes no conectados al componente principal considerado en este estudio.

Si bien no fue posible atribuirle una ubicación geográfica a cada uno de los usuarios contemplados en el grafo, se pudo utilizar el contenido textual de los comentarios que estas personas dejaron en CABA para ser clasificados según el idioma de los mismos mediante un modelo de clasificación. A efectos prácticos y facilitar el análisis se decidió dividir a los usuarios en dos grupos: hispanos parlantes (78,2%) y no hispanos parlantes (21,8%).

4 Red social de comentaristas en CABA

4.1 Conformación del grafo

Debido a las características que habilita la relación entre usuarios en Foursquare, se utilizó un grafo no dirigido y sin peso en sus aristas. A continuación se puede observar un esquema del grafo de usuarios junto con sus relación a los *venues*, representando cuales usuarios y relaciones son contemplados en el grafo.

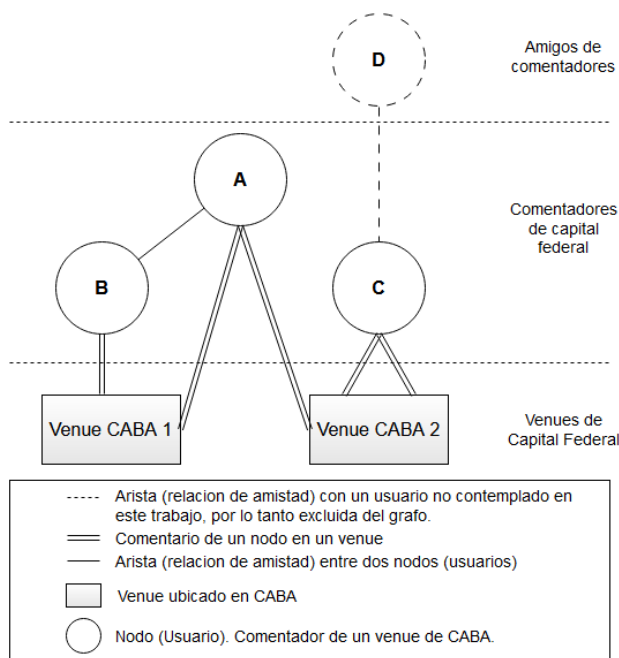


Fig. 4. Esquema de Grafo en relación a comentarios y venues de CABA

Podemos observar que una misma persona puede tener comentarios en diferentes lugares, así como un amigo de un comentarador puede ser: 1) comentarador de un mismo lugar, 2) comentarador de un diferente *venue* o 3) ser amigo directo de primer grado de una persona comentaradora aunque no ha comentado en ningún *venue* de CABA. El tercer caso ha sido omitido de este trabajo.

4.2 Características generales del grafo

Las redes sociales reales no son aleatorias, pero tampoco altamente regulares, poseen características únicas donde el orden y la estructura coexiste con el desorden en distintas medidas. En las siguientes secciones se analizan algunas de estas propiedades que nos permitirán identificar la estructura general, comunidades y características de esta red social. Si bien hay numerosas métricas de centralidades, en este trabajo se describe la red con los índices de centralidad comúnmente utilizados en el campo [8].

Se encuentra que muchas de estas métricas están correlacionadas entre sí y describen en parte las mismas propiedades, por lo que tan solo algunas de ellas son suficientes a motivos prácticos de análisis de la red social [9].

En la tabla 1 se puede observar un resumen del grafo y algunas de sus medidas.

Métrica	Grafo
Nodos	31385
Aristas	229224
Componentes conectados	1
Grado promedio	14,6
Densidad	0,000465
Coef. Clustering Promedio	0,1304
Transitividad	0,0134

Tabla 1. Características de grafo

En la figura 5 se listan las correlaciones más significativas entre una matriz de correlaciones para las medidas más importantes. Estos altos valores de r son esperables y han sido reportados anteriormente y se deben a los altos niveles de asociación que ocurren en datos organizados en redes.

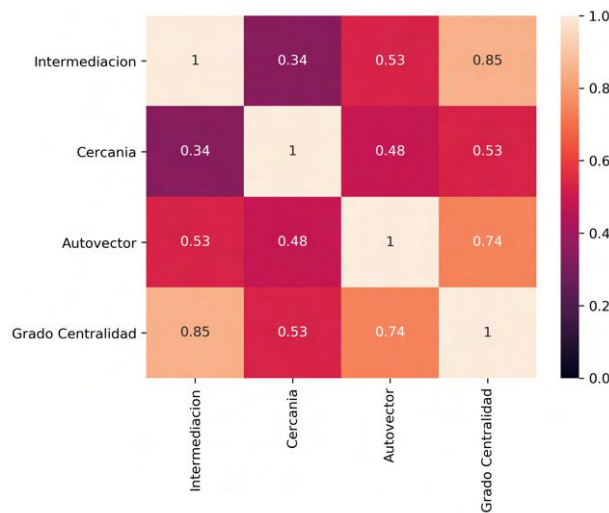


Fig. 5. Valores de correlación r de Pearson entre medidas de centralidad representadas en una matriz de calor.

Podemos identificar que el grado promedio es algo menor que el de otros estudios [10][11] en la misma plataforma, pero las diferencias se pueden deber a la manera de construcción del grafo analizado restringida solamente a usuarios con comentarios en CABA.

En tanto a otras medidas de estructura de la red, se ha encontrado que el coeficiente de clustering disminuye a medida que aumenta el grado de los nodos, indicando que los nodos con mayor grado pueden estar haciendo un uso menos orientado a relacionarse con verdaderas amistades, y en cambio establecen relaciones indiscriminadas.

4.3 Afinidad selectiva

Las redes sociales reales suelen presentar características de afinidad selectiva entre sus nodos basado en sus atributos (edad, género, etnia, etc) como también por atributos estructurales o propios de los nodos (grado de cercanía, etc). Esta afinidad selectiva, también suele llamarse homofilia en sociología y mide la tendencia de las personas a relacionarse con personas similares.

En primera instancia para analizar la afinidad selectiva de grados se analizó la relación entre el grado de los nodos y el promedio de grado de los vecinos de cada uno de estos (figura 6).

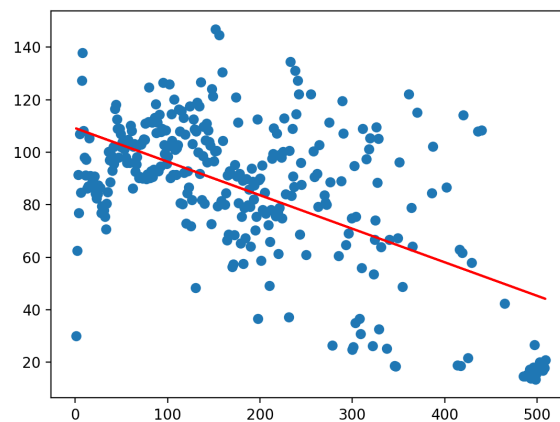


Fig 6. Asociación entre el grado del nodo (eje x) y el promedio de grado de vecinos (eje y)

Se puede observar a simple vista una tendencia de afinidad selectiva negativa con un valor de r de Pearson de 0,59, en donde nodos de mayor grado tienen vecinos con un grado menor en promedio. Un patrón similar emerge cuando se analiza los grados de los pares de nodos conectados, se observa que a medida que crece el grado de un nodo se relaciona en mayor medida con nodos de menor grado.

Finalmente este mismo patrón se comprueba con el coeficiente de Newman, el cual fue $-0,11$, lo que nos indica que los nodos de la red están en un grado intermedio o neutral de afinidad selectiva con una tendencia a la afinidad selectiva inversa.

Este valor está en concordancia con algunos estudios [12] donde proponen que las redes reales tienen un bajo grado de afinidad selectiva de grado y en un 58% de ellas se clasifican en un grado denominado “neutral” (entre $-0,19$ y $+0,19$). Este valor también es consistente con el grado encontrado en otras redes sociales en línea como Youtube [13].

Finalmente se estudió la afinidad selectiva no en base al grado sino a dos atributos que resultaron significativos en el análisis de la red: género declarado e idioma de los usuarios clasificado en base a sus comentarios.

Para el caso de idioma de los usuarios, observamos en la tabla 2 que el 76,9% de los nodos se relacionan con nodos de su mismo lenguaje, lo cual el coeficiente de afinidad confirma con un valor positivo de $0,11$. Mientras que el 73% de las aristas conectan hispanos-hispanos, solo el 3,9% lo hace entre no-hispanos parlantes.

Sobre la totalidad de aristas que establecen las personas hispano parlantes, los hispanos se relacionan en un 86% entre sí. Mientras que considerando la totalidad de relaciones que establecen los no hispano parlantes, en un 74% lo hacen con hispanos.

% Aristas	no-hispano	hispano	Subtotal
no-hispano	3,9%	11,5%	15,4%
hispano	11,5%	73,0%	84,6%
Subtotal	15,4%	84,6%	100%

Tabla 2. Matriz de afinidad selectiva por clasificación de idioma

En cuanto a la fracción de nodos que se relacionan con el mismo género, esto es la proporción de aristas que conectan entre sí usuarias mujeres (10%) sumado a las aristas que conectan usuarios hombres entre sí (45,9%), suman apenas más que el 55% del total de aristas. Por tal motivo se puede decir que la red social no es homofílica desde el punto de vista de género, comprobándose al calcular el coeficiente de afinidad selectiva para el género, el cual es de apenas $-0,0078$.

4.4 Ley de potencias

Es posible clasificar las redes según su distribución de grados. En la premisa original de Barabási y Albert [14], afirmaron que una gran cantidad de redes reales presentan una distribución de ley de potencias, con una gran mayoría de nodos con bajo grado y una poca cantidad de ellos con un alto grado [15]. Esta distribución de grados representa una de las tres propiedades principales que habitualmente se encuentran en

las redes sociales, junto con las distancias cortas y el alto grado de clustering (descritas en la sección de redes de mundo pequeño). La ley de potencias no es solo la menos intuitiva de estas propiedades sino la más estudiada y debatida desde que se describió a fines del siglo pasado [16]. Cabe resaltar que, si bien era de común entendimiento que las redes sociales presentaban habitualmente esta propiedad, este presupuesto está bajo cuestionamiento en la actualidad [17].

Si bien en un primer análisis la distribución de grados en la red tiene una similitud con la ley de potencias, se probó también la bondad de ajuste de distribuciones alternativas, como la normal logística (log-normal), o exponencial, entre otras. A tal razón se compararon gráficamente las distribuciones de la ley de potencia contra las log-normal y exponencial, que podemos ver a continuación.

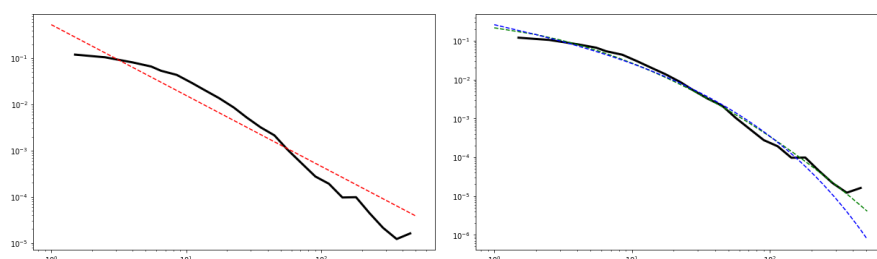


Fig. 7. Cuadro izquierdo: Ajuste de distribución para ley de potencias. Cuadro derecho: Ajuste a log-lineal (verde) y exponencial estrecha (azul) según distribución de grado (eje x)

Si bien las dificultades en distinguir las distribuciones log-normal de la ley de potencias son comunes [18], se evaluó comparativamente, utilizando la razón de verosimilitud para identificar cuál de las dos distribuciones se ajustaban mejor.

La distribución log-normal (también llamada log-lineal) se ajustó mejor. El valor de razón de verosimilitud entre la ley de potencia y la distribución log-lineal fue de -48.408. Este número es negativo ya que indica que es más posible que la distribución se ajuste a la log-lineal. El valor de significancia de p fue $< 0,001$, por lo que se descartó que la ley de potencias sea la mejor distribución teórica que ajuste a la distribución de grados del grafo analizado.

El ajuste a una ley de potencias en redes sociales es un presupuesto actualmente cuestionado dentro de la bibliografía citada, sumado a que estudios específicos de LBSN han también caracterizado que la distribución de grado ajusta parcialmente a la distribución log-lineal [19] mientras que otros lo describen como ley de potencias [20].

4.5 Un mundo pequeño

Para comprobar si esta red social tiene estructura de mundo pequeño, con distancias cortas y el alto grado de clustering, se procedió a realizar un experimento donde se utilizaron 7 probabilidades de recableado (p), utilizando un algoritmo generador de grafos basado en Watts-Strogatz. El algoritmo de Watts-Strogatz parte de una red en forma de anillo (la cual asegura un alto valor de coeficiente de clustering), y recablea aleatoriamente algunas de las aristas según una probabilidad p dada. Este recableo se hace entre nodos alejados entre sí para asegurar una creciente disminución del camino más corto en promedio, alcanzando su mayor aleatoriedad en el cableado de nodos cuando $p=1$.

Durante el experimento se mantuvieron las características topológicas en cuanto a cantidad de nodos (31385) y grado promedio para los mismos (14) que presenta el grafo analizado en este trabajo. Se pueden observar resultados del experimento en la tabla 3.

p	Clustering promedio	Camino más corto promedio
0,000001	0,6923	1066,26
0,00001	0,6922	955,34
0,0001	0,6921	210,59
0,001	0,6902	40,03
0,01	0,6721	11,48
0,1	0,5078	5,81
1	0,0004	4,23

Tabla 3. Resultado del experimento para grafos aleatorizados usando Watts-Strogatz

Podemos observar que a medida que se incrementa la probabilidad de recableado p , el coeficiente de clustering disminuye mucho más lentamente que el camino más corto promedio del grafo, siendo que la caída del valor de clustering solo ocurre con grafos con p muy alto.

En la figura 8 se puede observar con más claridad el comportamiento de ambas medidas utilizando una escala logarítmica para p .

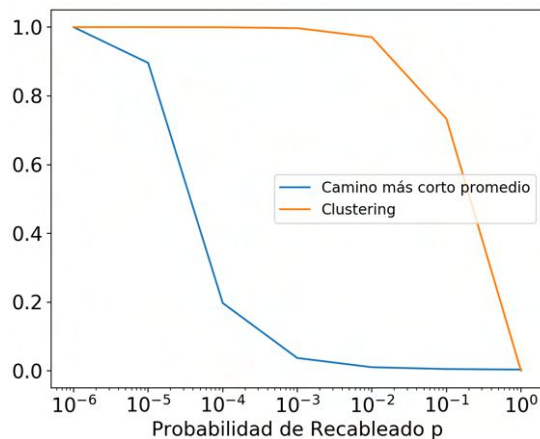


Fig. 8. Coeficiente de clustering C y camino más corto L según p para experimento de grafos aleatorizados usando Watts-Strogatz

Si tomamos en cuenta que el grafo analizado tiene un coeficiente de clustering promedio de 0,148 y un valor de camino más corto promedio de 4,23, podemos concluir que no se encuentra en el extremo en que el recableado es totalmente aleatorio, por lo que podemos clasificar esta red como una estructura espacial de mundo pequeño.

Al mismo tiempo se ha encontrado valores similares para redes sociales como Facebook [21], en donde la distancia promedio es de 4,7 y otras redes sociales en línea [22] o LBSN como Foursquare, en donde la distancia promedio del camino más corto es menor a 6 y el coeficiente de clustering promedio es entre 0,18 y 0,26 [23]. Estas propiedades van también en línea con la naturaleza de mundo pequeño de las LBSN [24] estudiadas.

4.6 Estructura comunitaria

La gran mayoría de las redes sociales tienen una estructura comunitaria [25] compuesta por grupos, también llamados clusters, lo que significa que hay grupos de nodos densamente conectados entre sí y poco conectados con otros grupos.

Para la identificación y partición de comunidades se utilizó la heurística de Louvain, basado en la optimización de la modularidad del grafo. Si bien el resultado de la heurística utilizada no es determinista, el mismo presentó resultados similares en diferentes ejecuciones para segmentar en comunidades el grafo.

La modularidad resultado de la partición obtenida del grafo fue de 0,52. Este valor indica una estructura de comunidad alta, en donde las aristas tienden a ser incidentes entre nodos de la misma comunidad.

Sobre un total de 37 comunidades identificadas, las mismas tienen un promedio de 848 nodos, con un máximo de 4229 y un mínimo de 3 nodos. Podemos observar en la figura 9 el diagrama de densidad Kernel de la distribución del tamaño de las comunidades.

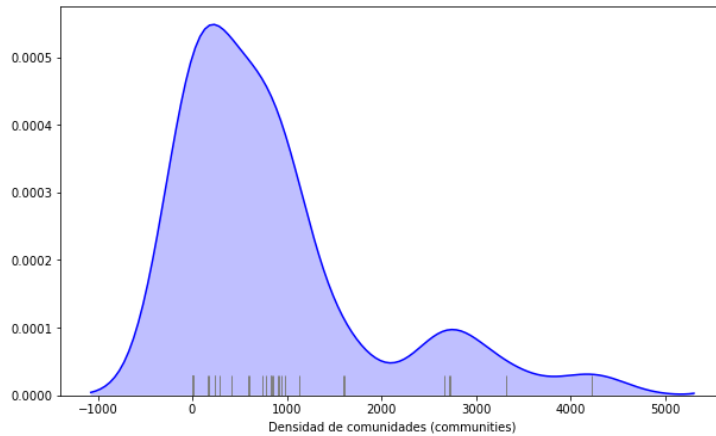


Fig. 9. Estimación de densidad Kernel tamaño de comunidades

A modo de ejemplo, se observa en la figura 10 el grafo de una de las comunidades, de menos de 200 nodos, la cual tiene la característica de tener un 55% de usuarios clasificados como no hispano parlantes (en color rosa), alejándose de la distribución promedio de idiomas encontradas en otras comunidades.

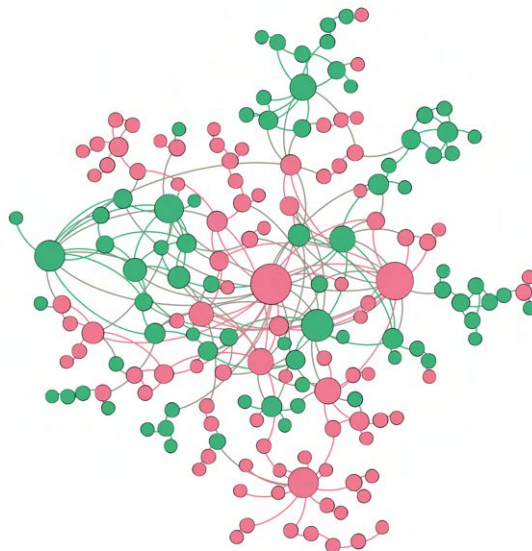


Fig. 10. Grafo de comunidad seleccionada. Hispano parlantes en verde. No hispano parlantes en rosa. Tamaño por centralidad de grado.

Esta comunidad, nos da un indicio de que las personas parecen relacionarse entre sí de acuerdo a su idioma, pero esto no parece ser extensible en gran medida para la totalidad del grafo, y fue abordado en más detalle en la sección de afinidad selectiva. Sin embargo al ser graficada sobre el mapa de la ciudad en la figura 11, se encontró un indicio espacial de que las personas no hispano parlantes frecuentan sectores más concentrados de la ciudad que personas hispano parlantes, hipótesis que será abordada con más profundidad en otro trabajo.

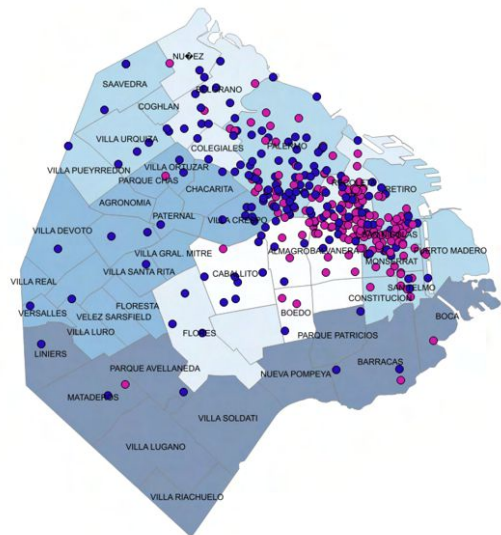


Fig. 11. Centroides geográficos de los comentarios de cada integrante de la comunidad azul y comunidad rosa. La composición de hispano parlantes para la primera es del 55% mientras que para la segunda es del 15%.

Finalmente, si bien la identificación de comunidades es útil para descubrir propiedades subyacentes en la red social que puedan ser exclusivas de esa comunidad o luego extenderse al conjunto de la red, también es posible identificar cómo se relaciona cada una de estas comunidades entre sí, y entender la centralidad de cada una de ellas para distintos intereses como difusión de mensajes o segmentación de usuarios.

5 Conclusiones y trabajos futuros

En este trabajo se hace uso del análisis de redes sociales (SNA) para caracterizar los aspectos estructurales de una red social basada en geo localización del ámbito de la Ciudad Autónoma de Buenos Aires. Adicionalmente se describe la metodología utilizada para la recolección, procesamiento y análisis utilizando software de código abierto y fuentes de datos disponibles públicamente en el estado y plataformas online.

En tanto características de la red social analizada, se ha encontrado los siguientes aspectos estructurales y propiedades:

- Se descarta que la ley de potencias sea la mejor distribución teórica que ajuste a la distribución de grados del grafo analizado, encontrando la distribución log-lineal como más adecuada, en consonancia con la literatura académica más reciente al respecto.

- Se comprueba que la red presenta un bajo grado de afinidad selectiva negativa, dentro de los parámetros que presentan otras redes sociales similares.

- De igual manera, se comprueba la estructura de mundo pequeño, coincidiendo con otros estudios.

- Por último, se encuentra una estructura de comunidad alta y se describen las características estructurales distintas comunidades presentes.

Desde un punto de vista aplicado este estudio muestra la utilidad del enfoque de grafos para analizar los comportamientos de los turistas en CABA. Y así encontrar, por ejemplo, zonas de la ciudad que están adquiriendo interés turístico, segmentado por su región de origen. También, el grado de conectividad hallado sugiere que mediante un direccionamiento adecuado de las acciones de publicidad y marketing es posible alcanzar niveles de difusión comparables a los que se obtendrían con campañas más costosas que busquen una alta saturación.

En futuros trabajos se analizará el componente espacial de la red social, poniendo en foco los vínculos entre los usuarios en relación a la distancia espacial y estructura de su red social. Si bien en la bibliografía se encuentra un acuerdo sobre que la probabilidad de un vínculo entre dos usuarios decrece a medida que aumenta la distancia geográfica de los mismos [26][27], las características e importancia de tal relación todavía siguen siendo debatidas según diferentes tipos de redes, plataformas y países, por lo que reviste de importancia analizar qué características adquiere en el ámbito local.

Referencias

1. World Bank. Information, Communication Technologies, & infoDev (Program). Information and communications for development 2012: Maximizing mobile. World Bank Publications. (2012)
2. PreoŃiu-Pietro, D., Trevor, C. Mining user behaviours: a study of check-in patterns in location based social networks. In Proceedings of the 5th annual ACM web science conference. (2013).
3. Zhang, K., Konstantinos, P., and Theodoros, L. Effects of promotions on location-based social media: evidence from foursquare. In International Journal of Electronic Commerce 22.1. (2018).
4. Zickuhr, K. Three-quarters of smartphone owners use location-based services. Pew Internet & American Life Project. (2012)
5. Gobierno de la Ciudad de Buenos Aires."Manzanas de la Ciudad de Buenos Aires". <http://data.buenosaires.gob.ar/dataset/manzanas>, last accessed 12, Jan. 2015.
6. Hussain, A., & Keshavamurthy, B. N. Analyzing Online Location-Based Social Networks for Malicious User Detection. In Recent Findings in Intelligent Computing Techniques. Springer, Singapore. pp. 463-471. (2019)
7. Vasconcelos, M. A., Ricci, S., Almeida, J., Benevenuto, F., & Almeida, V. Tips, dones and todos: uncovering user profiles in foursquare. In Proceedings of the fifth ACM international conference on Web search and data mining, pp. 653-662). (2012)
8. Zhang, J., & Luo, Y. Degree centrality, betweenness centrality, and closeness centrality in social network. In 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017). Atlantis Press. (2017).
9. Zinoviev, D. Complex network analysis in Python: Recognize-construct-visualize-analyze-interpret. Pragmatic Bookshelf. (2018)
10. Agryzkov, T., Martí, P., Tortosa, L., & Vicent, J. F. Measuring urban activities using Foursquare data and network analysis: a case study of Murcia (Spain). International Journal of Geographical Information Science, 31(1), pp. 100-121. (2017)
11. Ferreira, A. P. G., Silva, T. H., & Loureiro, A. A. F. Beyond sights: Large scale study of tourists' behavior using foursquare data. In 2015 IEEE International Conference on Data Mining Workshop (ICDMW), pp. 1117-1124). (2015)
12. Meghanathan, N. Assortativity Analysis of Real-World Network Graphs based on Centrality Metrics. Computer and Information Science, 9(3), pp 7-25. (2016)
13. Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. Measurement and analysis of online social networks. In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, pp. 29-42. (2007)
14. Barabási, A. L., & Albert, R. Emergence of scaling in random networks. science, 286(5439), pp. 509-512. (1999)
15. Oliveira, M., & Gama, J. An overview of social network analysis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. (2012)
16. Muchnik, L., Pei, S., Parra, L. C., Reis, S. D., Andrade Jr, J. S., Havlin, S., & Makse, H. A. Origins of power-law degree distribution in the heterogeneity of human activity in social networks. Scientific reports, 3(1), pp 1-8. (2013)
17. Broido, A. D., & Clauset, A. Scale-free networks are rare. Nature communications, 10(1), pp 1-10. (2019)
18. Malevergne, Y., Pisarenko, V., & Sornette, D. Empirical distributions of stock returns: between the stretched exponential and the power law?. Quantitative Finance, 5(4), pp. 379-401. (2005)

19. Scellato, S., & Mascolo, C. Measuring user activity on an online location-based social network. In 2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPs), pp. 918-923. (2011)
20. Wei, W., Zhu, X., & Li, Q. LBSNSim: Analyzing and modeling location-based social networks. In IEEE INFOCOM 2014-IEEE Conference on Computer Communications. pp. 1680-1688. (2014)
21. Ugander, J., Karrer, B., Backstrom, L., & Marlow, C. The anatomy of the facebook social graph. arXiv preprint arXiv:1111.4503. (2011)
22. Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. Measurement and analysis of online social networks. In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. (2007)
23. Scellato, S., Noulas, A., Lambiotte, R., & Mascolo, C. Socio-spatial properties of online location-based social networks. ICWSM, 11, pp. 329-336. (2011)
24. Leskovec, J., & Horvitz, E. Planetary-scale views on a large instant-messaging network. In Proceedings of the 17th international conference on World Wide Web, pp. 915-924. (2008)
25. Girvan, M., & Newman, M. E. Community structure in social and biological networks. Proceedings of the national academy of sciences, 99(12), 7821-7826. (2002)
26. Levy, M., & Goldenberg, J. The gravitational law of social interaction. Physica A: Statistical Mechanics and its Applications, 393, 418-426. (2014)
27. Backstrom, L., Sun, E., & Marlow, C. Find me if you can: improving geographical prediction with social and spatial proximity. In Proceedings of the 19th international conference on World wide web (pp. 61-70). (2010)