



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

75.30 - TEORÍA DE ALGORITMOS II

Trabajo práctico final

Manuel Longo Elia - Padrón 102425
Azul Zaietz - Padrón 102214

Índice

1	Objetivo	2
2	Exploración inicial	2
2.1	Grado promedio	2
2.2	Coeficiente de clustering	2
2.3	Diámetro de la red	2
3	Homofilia	3
3.1	Hipótesis	3
3.2	Resultados	4
4	Análisis de las comunidades	4
4.1	Introducción	4
4.2	Hipótesis	4
4.3	Resultados	4
5	Roles	6
6	Conclusiones	7
7	Anexo	9
8	Correcciones	10
8.1	Hipótesis	10
8.2	Resultados	11

1 Objetivo

El presente trabajo tiene como objetivo llevar a cabo un análisis de los desplazamientos efectuados por la compañía RideAustin en la ciudad de Austin, Texas, Estados Unidos, durante el periodo comprendido entre los años 2016 y 2017.

RideAustin es una plataforma de transporte compartido, semejante a empresas como Uber o Cabify. No obstante, se distingue de sus competidores debido a su enfoque sin ánimo de lucro, lo que implica que gran parte de los ingresos generados por las tarifas de transporte se destinaban a cubrir los gastos operativos y a apoyar a organizaciones benéficas locales. Los usuarios de *RideAustin* tenían la opción de seleccionar una organización benéfica de su preferencia, y automáticamente una pequeña porción de la tarifa de su viaje sería donada a dicha entidad.

Los datos de dichos viajes fueron descargados en formato *csv* aquí para ser leídos y trabajados desde distintos *Jupyter Notebook* que estarán detallados en el anexo.

En los *notebooks* se aplicaron diversas técnicas de transformación de datos, incluyendo la eliminación de valores nulos y la conversión de las coordenadas de inicio y fin de los viajes a direcciones físicas. Además, se procedió a filtrar el conjunto de datos para incluir únicamente las ubicaciones que sirvieron como punto de inicio o destino de al menos 150 viajes, resultando en un total de 165 ubicaciones distintas y un registro de 18,677 viajes que conectaban estas ubicaciones.

Posteriormente, para modelar la red, se construyó un grafo en el cual los nodos representan los puntos (direcciones) que fueron origen o destino de los viajes, mientras que las aristas simbolizan los trayectos entre dos puntos conectados por la misma. Con el objetivo de simplificar la representación de la red, cuando existían dos o más viajes entre dos puntos, se representaba por una única arista, cuyo peso indicaba la cantidad de viajes realizados en esa ruta específica.

2 Exploración inicial

En esta sección se realizarán los análisis básicos de una red. Estos son el grado promedio, el coeficiente de clustering y diámetro de una red.

2.1 Grado promedio

El grado promedio de la red es de 131.79, lo que indica que, en promedio, cada nodo representa el origen o destino de aproximadamente 131.79 viajes hacia ubicaciones distintas, dentro del conjunto de 165 ubicaciones analizadas.

La ubicación con el menor grado es 4015 Bergstrom Drive, que cuenta con un grado de 82. Esta ubicación corresponde a una de las calles de acceso secundario al aeropuerto internacional de Austin. En contraste, la ubicación de mayor grado es 709 West 6th Street, con un valor de 158. Esta dirección se encuentra en el centro de la ciudad y es un lugar muy concurrido debido a la presencia de diversas oficinas y bancos.

2.2 Coeficiente de clustering

El coeficiente clustering promedio de la red es de 0.88, lo cual refleja el nivel de interconexión entre los vecinos de cada nodo. Dado que el valor de clustering promedio se encuentra en el rango de 0 a 1, siendo 1 la máxima conectividad, se puede afirmar que esta red se encuentra altamente interconectada entre sus nodos debido a su valor de 0.88.

2.3 Diámetro de la red

El cálculo del diámetro de la red nos dio un valor de 2. Esto es indicativo de una red muy interconectada entre sí, como se confirmó antes con el coeficiente clustering. Lo que indica esto es

que para llegar de un nodo a otro del grafo como mucho usarás dos aristas.

3 Homofilia

3.1 Hipótesis

En esta sección, procederemos a realizar un análisis de la homofilia presente en la red. La homofilia se define como la tendencia de los nodos a conectarse con otros nodos que comparten características o atributos similares. Se representa mediante valores en el rango de 0 a 1, donde 0 representa la homofilia más alta posible y 1 la más baja.

Para llevar a cabo el análisis, hemos dividido las ubicaciones de la ciudad en 5 zonas distintas: central, norte, sur, este y oeste. Nuestra hipótesis es que, según esta categorización, la homofilia en esta red será baja. Esto se debe a que estamos modelando viajes en automóvil, los cuales tienden a ser más largos que aquellos realizados mediante otros medios de transporte, como la bicicleta. En consecuencia, los viajes en automóvil tienen más probabilidad de atravesar diferentes zonas geográficas.

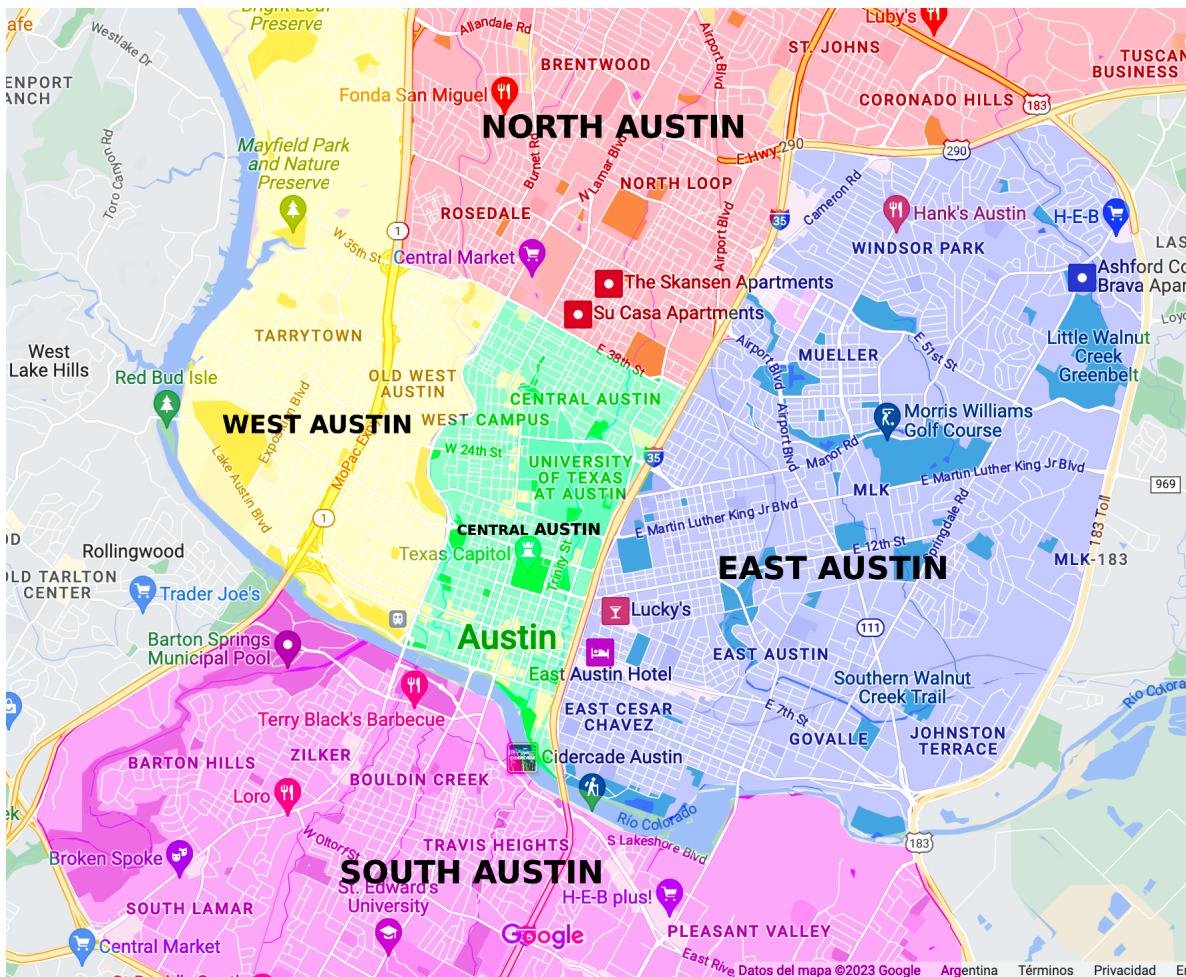


Figura 3.1: Zonas geográficas usadas para el análisis.

Es relevante destacar que las zonas del norte y sur de Austin son considerablemente más extensas de lo que se muestra en el mapa; ambas se extienden verticalmente. Se aplicó un zoom para presentar con mayor claridad las divisiones en el mapa.

Una característica fundamental de la ciudad es su perfil “angosto”, donde las zonas este y oeste son más reducidas en comparación. Por el contrario, la ciudad se extiende considerablemente más en dirección norte y sur, lo que contribuye a su forma característica.

3.2 Resultados

Como se observa en la Figura 3.2, la homofilia por grupos en la red es baja en general. La mayoría de las zonas muestran valores superiores a 0.72, excepto en South Austin, donde se registra un valor de 0.52. Esta disminución en el nivel de homofilia en la parte sur de la ciudad puede atribuirse a la presencia de lugares altamente frecuentados, como los parques metropolitanos Zilker y Butler, ambos situados junto al río Colorado, y el aeropuerto de la ciudad de Austin. Además, es importante destacar que la parte sur de la ciudad es más extensa, lo que aumenta la probabilidad de viajes intrazonales.

Por otro lado, los valores bajos en el este y oeste de la ciudad son razonables debido a que estas áreas son más pequeñas y principalmente residenciales, lo que lleva a que los viajes tiendan a cruzar de una zona a otra.

En cuanto a Central Austin, se observa la homofilia más baja, con un valor de 0.94. Esto se debe claramente a que en el centro de la ciudad se encuentran todas las oficinas y lugares importantes, lo que atrae a residentes de todos los puntos de la ciudad, generando una mayor mezcla de destinos.

En conclusión, los datos presentados confirman la hipótesis planteada en la sección anterior.

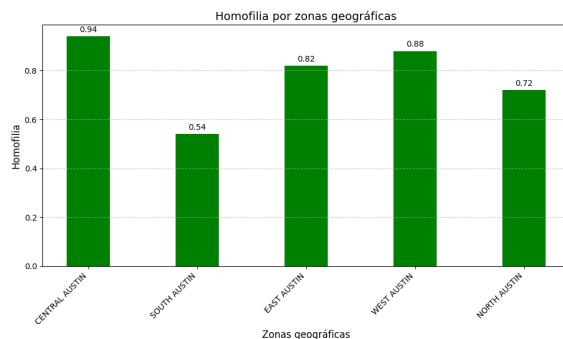


Figura 3.2: Homofilia por grupos

4 Análisis de las comunidades

4.1 Introducción

En esta sección, se hará un análisis de las visualizaciones de las comunidades que se obtuvieron mediante la realización del *Algoritmo de Louvain*.

4.2 Hipótesis

La hipótesis que se busca verificar es que existen comunidades, tal como fueron descritas en la sección anterior, que se organizan por zonas geográficas.

4.3 Resultados

En el anexo, en el apartado del *notebook*, se encuentra un detallado paso a paso de cómo se llevó a cabo el algoritmo previamente mencionado. Asimismo, en dicho *notebook*, se incluyen

las visualizaciones que serán presentadas a continuación y que pueden ser utilizadas de forma interactiva.

El algoritmo detectó inicialmente 4 comunidades, cada una con una cantidad diferente de direcciones.

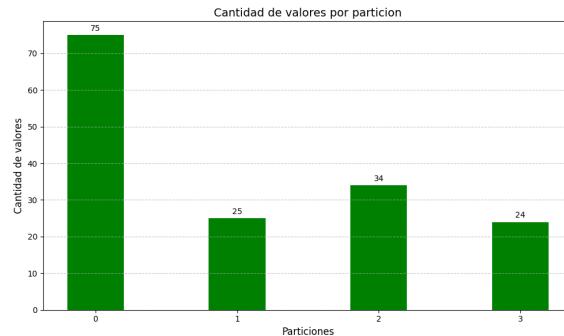


Figura 4.1: Cantidad de ubicaciones por comunidad

Como se aprecia en el gráfico, existe una comunidad que presenta significativamente más valores que las demás. Esta disparidad podría ser atribuible a la posibilidad de que el algoritmo haya identificado dos comunidades como una sola, fusionándolas en el proceso.

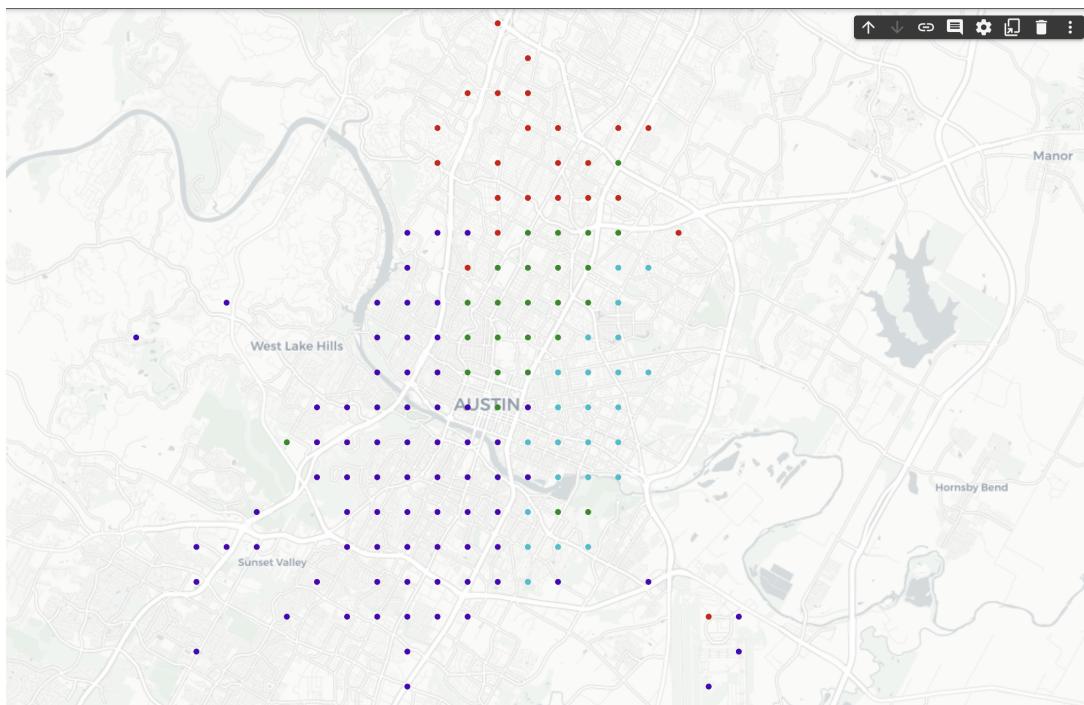


Figura 4.2: Visualización de las comunidades en el mapa de la ciudad de Austin

El gráfico reafirma lo mencionado previamente, evidenciando la unión de dos zonas geográficas, South Austin y West Austin. Por consiguiente, se puede considerar la hipótesis como confirmada. Sin embargo, se procederá a realizar un análisis más detallado de dicha comunidad fusionada para verificar de manera rigurosa la veracidad de la unión de esas dos zonas. Este análisis se enfocará en las subcomunidades de dicha agrupación con el propósito de asegurar que efectivamente las dos áreas mencionadas han sido combinadas en una sola entidad.

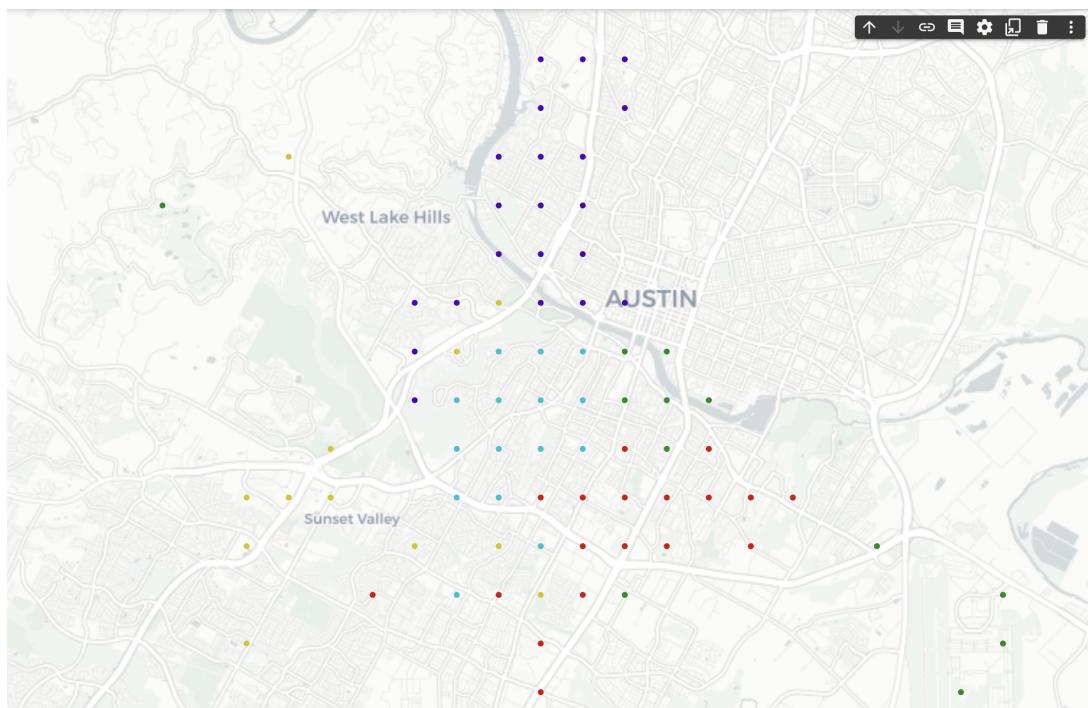


Figura 4.3: Visualización de las subcomunidades

A partir de la visualización, se puede observar que las zonas geográficas se encuentran en gran medida respetadas. Las ubicaciones pertenecientes a West Austin se encuentran representadas en violeta, mientras que las de South Austin están identificadas por los demás colores, zona que abarca la mayor extensión territorial.

Considerando toda la información presentada, se confirma la hipótesis planteada.

5 Roles

En esta sección, analizaremos los roles desempeñados por diferentes ubicaciones en la red. Mediante la aplicación del *algoritmo RolX*, se han identificado tres roles distintos: Rol 1, Rol 2 y Rol 3, con un número de nodos de 20, 14 y 124, respectivamente. Los resultados de estos roles se encuentran visualizados en la imagen siguiente y en el *notebook* del anexo con la opción de interactividad.

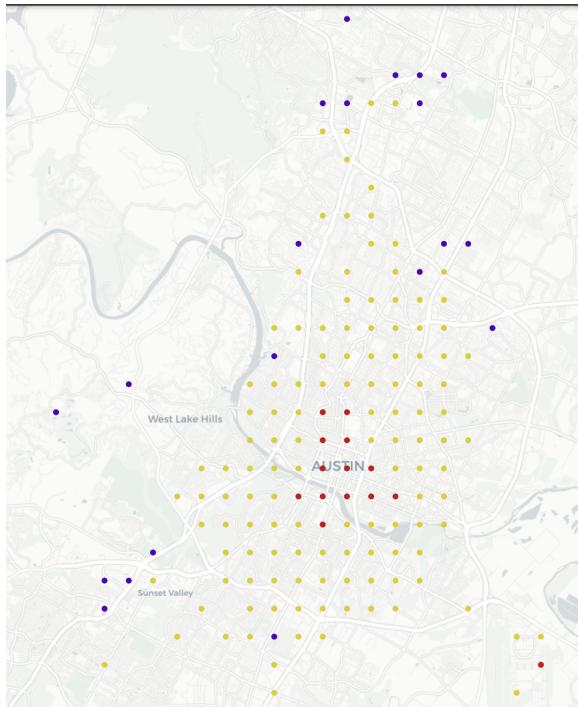


Figura 5.1: Visualización de los tres tipos de roles

Las ubicaciones correspondientes al Rol 1 se representan mediante el color violeta, destacando que se trata de nodos periféricos situados en las zonas más distantes del centro de la ciudad de Austin. La mayoría de ellos se encuentran en la parte oeste de la ciudad, con la excepción de un caso aislado en el este, y algunos en la parte norte.

Por otro lado, las ubicaciones del Rol 2 se marcan en rojo. Estas localizaciones se identifican como nodos centrales, ya que ocupan posiciones en el centro de la ciudad, que es el área con mayor concentración de personas. También, se le suma a este set de ubicaciones la entrada principal del aeropuerto internacional de Austin.

Finalmente, se destaca el Rol 3, que se representa con un mayor número de nodos. Estas ubicaciones se consideran nodos intermedios, ya que no se encuentran ni muy alejados del centro ni son particularmente centrales.

6 Conclusiones

En este informe se han analizado las características principales de la red presentada, lo que nos permitió identificar las zonas principales de la ciudad, su división y los lugares más frecuentados por su población, sin necesidad de haber visitado físicamente el lugar.

Podemos constatar que, a pesar de su extenso territorio (709 km^2 , más del doble de la Ciudad Autónoma de Buenos Aires), la mayor actividad se concentra en el centro de la ciudad y en la zona sur cercana al Río Colorado.

Al igual que otras ciudades de Estados Unidos, Austin se encuentra zonificada según los puntos cardinales. Esta conclusión fue respaldada al observar la homofilia por zonas y verificar que los resultados coincidían con el análisis de las comunidades.

Por último, como resultado destacado, identificamos los tres lugares más concurridos por los usuarios de la aplicación:

- Palacio de Justicia Homer Thornberry (221 East 9th Street)

- Banco Americano del Comercio (610 West 5th Street)
- Edward Rendon Park (Lady Bird Lake Hike and Bike Trail)

Los dos primeros lugares revisten una gran importancia tanto para el sistema judicial como para la economía del estado de Texas, siendo que Austin ostenta el título de capital. Por otro lado, el último lugar es altamente concurrido debido a sus pasarelas para bicicletas que atraviesan el Río Colorado y las rutas de senderismo del bosque circundante.

7 Anexo

- Notebook transformación de datos: colab
- Notebook exploración de datos: colab
- Notebook Homofilia: colab
- Notebook Comunidades: colab
- Notebook roles: colab
- Link al dataset: link

8 Correcciones

Homofilia

8.1 Hipótesis

Luego de analizar las comunidades y los roles, quedó en evidencia que nuestra hipótesis sobre la homofilia en la red era incorrecta. Para corregir el planteo anterior, generamos una nueva división del territorio:

- En primer lugar, dejamos de considerar a la zona “West Austin” individualmente. La integramos a la región “South Austin” creando “Southwest Austin”. Como habíamos aclarado en la versión anterior del trabajo, el mapa no reflejaba el tamaño real de la zona por el zoom que se había aplicado para generar la visualización. La región no era significativa en kilómetros cuadrados y tampoco con respecto a la cantidad de puntos del dataset dentro de la misma.
- En segundo lugar, agrandamos las regiones norte y centro. Nuevamente, buscando reflejar mejor la realidad. La universidad y los campus, que ocupan una parte importante del centro de la ciudad, habían quedado afuera en los límites que habíamos marcado. Con la nueva división quedaron incorporados al centro de la ciudad. Al mismo tiempo, el norte de la ciudad era verticalmente más extenso de lo que habíamos delimitado. Las modificaciones del norte y del centro generaron la disminución de la zona este del territorio.

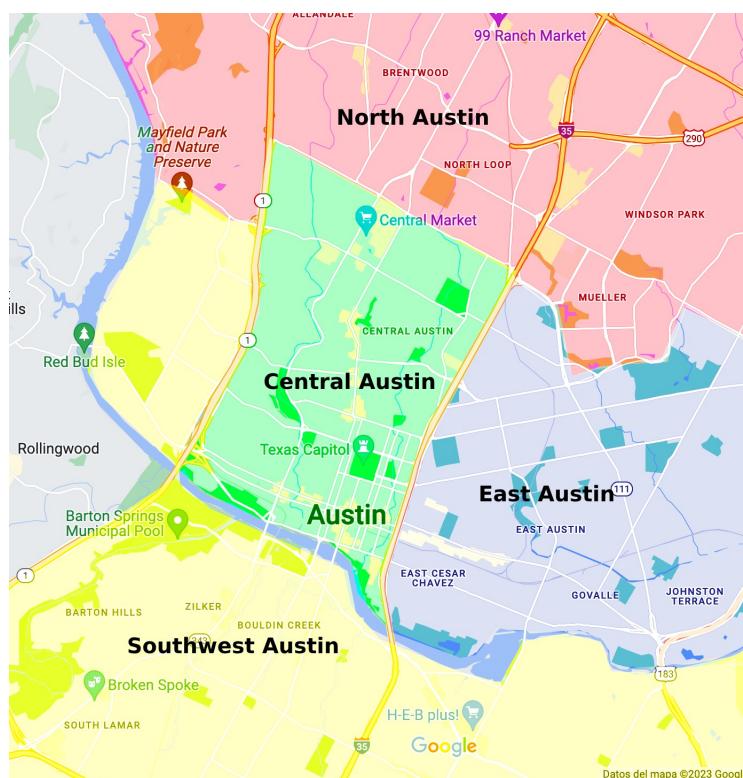


Figura 8.1: Zonas geográficas usadas para el análisis.

Entonces, para llevar a cabo el nuevo análisis, trabajamos con la clasificación de las ubicaciones de la ciudad en 4 zonas distintas: central, norte, suroeste y este.

Antes de avanzar con el análisis, vamos a definir a p como la proporción de ubicaciones pertenecientes a una determinada zona z de la red para el caso de distribución completamente aleatoria

de las aristas en el grafo. Si tomamos una arista insertada aleatoriamente en la red cuyo origen está en la zona z, la probabilidad de que el destino también esté en la zona z es igual a p. Entonces, la probabilidad de que no esté en z será de 1-p.

Habiendo dicho esto, podemos entender que para detectar la homofilia por zona debemos verificar si las conexiones entre las ubicaciones dentro de la misma zona son más frecuentes en nuestra red que en un grafo aleatorio. Por eso, para determinar si se cumple y en qué medida, vamos a calcular para cada zona qué proporción de los viajes que inician en esa zona también terminan dentro de ella.

Para poder obtener un valor más preciso de esa proporción, modificamos la función que devuelve la proporción real que se obtiene de cruces respecto del total de aristas de un tipo en particular. Hicimos que la función tenga en cuenta los pesos de las aristas, es decir, la cantidad de viajes que se hicieron con esos nodos como origen y destino respectivamente.

Planteamos como nueva hipótesis que, según esta nueva categorización y teniendo en cuenta la cantidad de viajes hechos por cada camino, es más probable que encontremos una mayor cantidad de viajes internos por zona. Por lo tanto, deberíamos obtener valores más altos de homofilia, por lo menos para alguna de las zonas.

8.2 Resultados

Graficamos la proporción de nodos (ubicaciones) por zona sobre el total de zonas (calculada con la función: proporción por tipo) versus la proporción de aristas que unen dos nodos del grafo pertenecientes a la misma zona. La última proporción la calculamos como uno menos la proporción calculada con la función: proporción cruzan campo de tipo, actualizada para que tenga en cuenta el peso de las aristas.

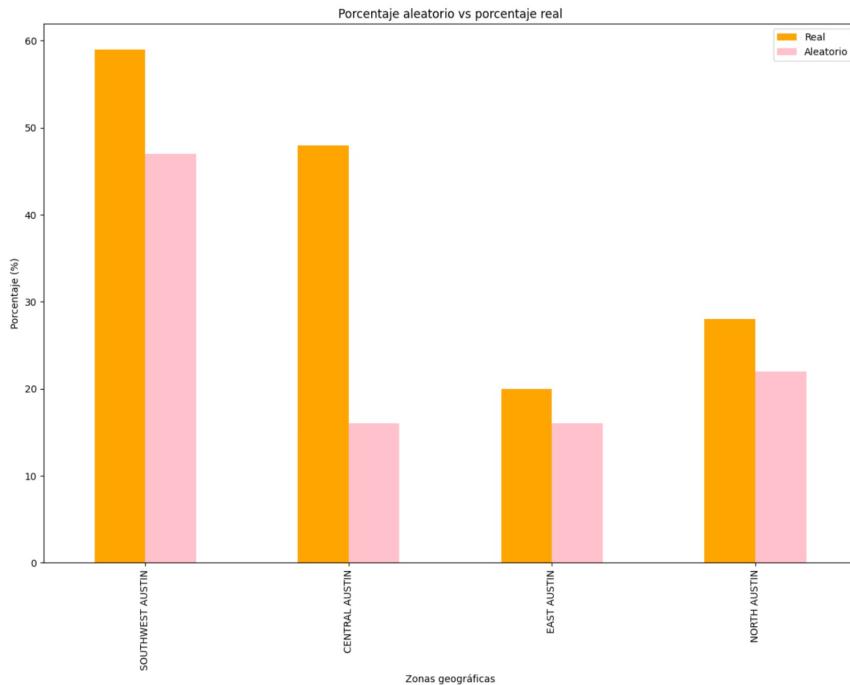


Figura 8.2: Proporción real versus proporción aleatoria

La información que podemos extraer de este gráfico es que para el caso de la zona central la cantidad de aristas entre dos nodos pertenecientes a la misma en una red aleatoria es menor que en la red que estamos analizando. Esto significa que, cuando inicia un viaje en una ubicación localizada en la zona central hay una probabilidad alta (superior a la aleatoria) de que la ubicación final también esté dentro de la zona central. Esto refleja homofilia por la zona central.

El resto de las zonas también presentan una diferencia con respecto al porcentaje aleatorio pero en menor medida. Es decir, la probabilidad de que el viaje inicie y termine dentro de la misma zona es menor (sigue siendo superior a la aleatoriedad).

Graficamos los nuevos valores obtenidos de homofilia:

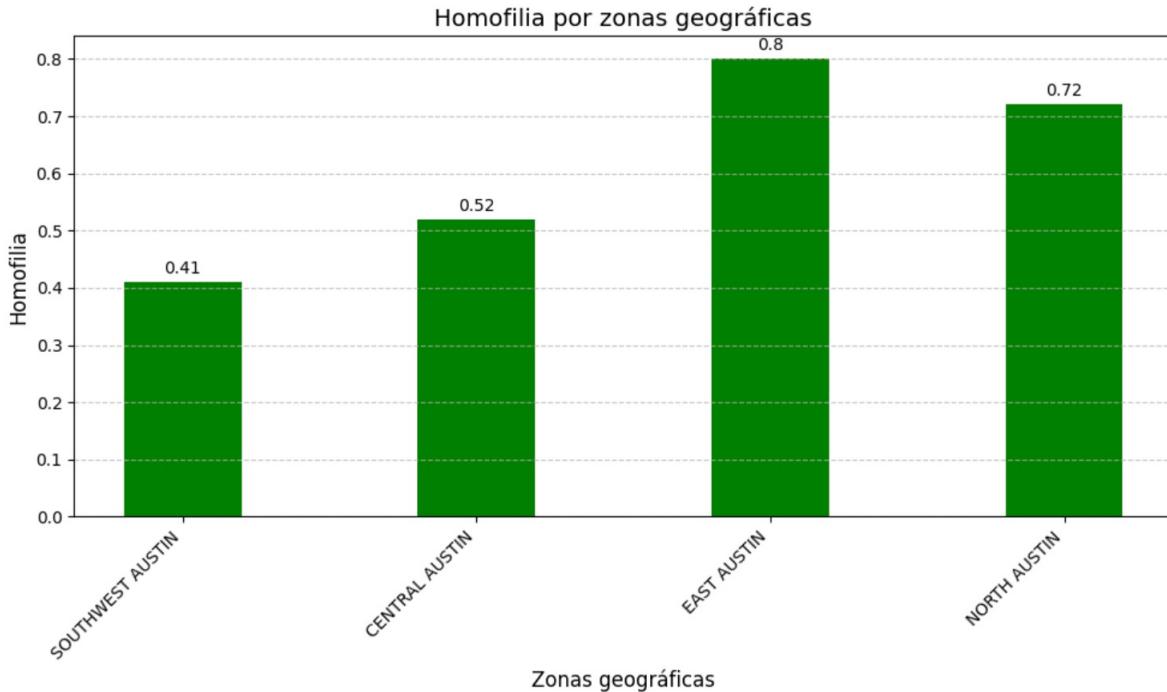


Figura 8.3: Homofilia por grupos

Podemos ver que los nuevos valores obtenidos reflejan un valor más alto de homofilia para las zonas sur y central.

Podemos pensar que el sur, al ser una zona muy extensa y la parte más turística de Austin, tiene muchos viajes internos. En la zona se encuentran espacios verdes, galerías de arte, restaurantes y bares. La zona ofrece desde actividades como kayak, paddleboarding y senderismo hasta espacios para disfrutar de la vida nocturna de la ciudad.

Al mismo tiempo, tiene sentido que haya muchos viajes concentrados en el centro de la ciudad, por ser una zona concurrida tanto por trabajadores como por estudiantes. La zona se caracteriza por ser distrito de negocios, tener una gran concentración de empresas, oficinas y centros de innovación. Además, dentro de la zona se encuentra ubicada la sede de la Universidad de Texas en Austin (una de las sedes más importantes de la Universidad de Texas) y el campus de la misma.

El este es la zona más pequeña en comparación con las otras, además de ser una zona residencial, por eso se justifica que presente la homofilia más baja. Tiene sentido que la gente haga viajes para desplazarse hacia otras zonas.