

Análisis Estadístico de Redes Sociales: Taller #2

Valentina Cardona Saldaña

Ejercicio #1

Al reproducir los ejemplos 3.1, 3.2, 5.2, 5.3, 6.1 de Conceptos fundamentales de teoría de grafos en **Python** utilizando **igraph**, se encontró que todos los ejemplos pudieron replicarse exitosamente.

En el ejemplo 3.1 se analizaron los vecinos y el grado de una red no dirigida. Para el nodo 1, sus vecinos son los nodos 2 y 3, y el grado de cada vértice, ordenados de manera ascendente, es $[2, 3, 3, 4, 3, 3, 2]$. En el ejemplo 3.2 se examinaron los grados de entrada y salida de una red dirigida simple. Para los 3 nodos, los grados de entrada fueron $[0, 2, 2]$, y los grados de salida fueron $[2, 1, 1]$. En los ejemplos 5.2 y 5.3 se abordaron las componentes y conexiones de una red, respectivamente. En el 5.2, se exploró una red no conectada. Al extraer sus componentes se identificaron 3, de tamaños 7, 3 y 2. En el ejemplo 5.3, la red estaba conectada débilmente, ya que no era posible acceder desde cada vértice a todos los demás mediante una caminata dirigida.

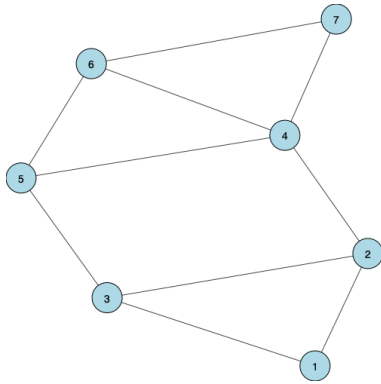


Figure 1: Ejemplo 3.1 y 6.1

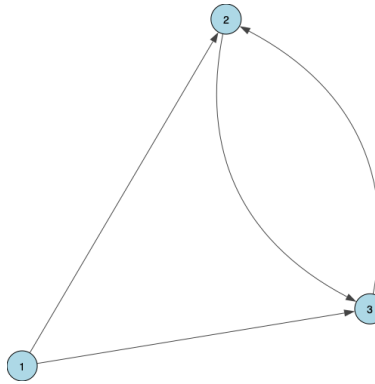


Figure 2: Ejemplo 3.2 y 5.3

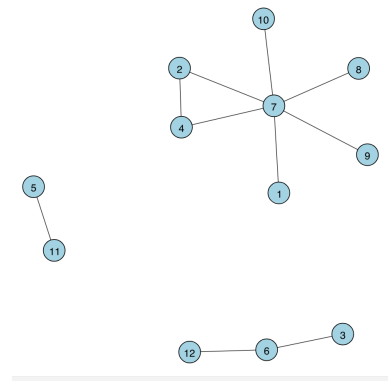


Figure 3: Ejemplo 5.2

Finalmente, en el ejemplo 6.1 se investigaron las distancias entre los nodos de la misma red que en el ejemplo 3.1. Esta tiene un diámetro de 3, que va desde el nodo 1, pasa por el 2 y 4, y llega al 6. Su distancia geodésica promedio es de 1.66, y en la figura 5 se muestra la distribución de estas distancias.

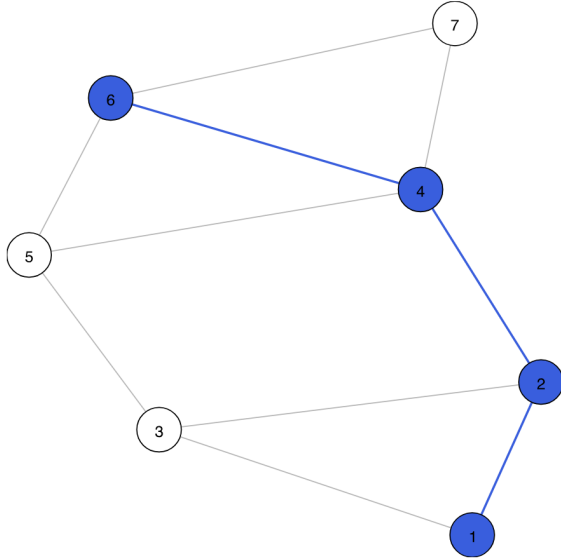


Figure 4: Ejemplo 6.1. Diámetro de Figure 1.

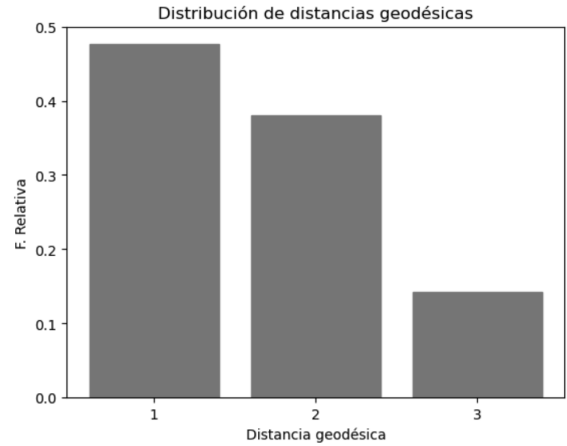


Figure 5: Distribución distancias geodésicas

Ejercicio #2

El grado de un nodo en una red tanto dirigida como no dirigida se puede calcular fácilmente a partir de la matriz de adyacencia $Y = [y_{i,j}]$. El *out-degree* d_i^{out} y el *in-degree* d_i^{in} del nodo i se pueden calcular respectivamente como:

$$d_i^{out} = \sum_{j:j \neq i} y_{i,j} \quad \text{y} \quad d_i^{in} = \sum_{j:j \neq i} y_{j,i}$$

Muestre que si la red es no dirigida entonces $d_i^{out} = d_i^{in}$

Respuesta 2.1:

Una red no dirigida es aquella en que cada arista $\{i, j\} \in E$ es tal que $\{i, j\} \equiv \{j, i\}$, para todo $i, j \in V$ [4]. Al ser la matriz de adyacencia Y simétrica, $y_{i,j} = y_{j,i}$, sumar todas las filas o todas las columnas corresponde al mismo resultado:

$$\sum_{j:j \neq i} y_{i,j} = \sum_{j:j \neq i} y_{j,i}$$

Por lo tanto, si la red es no dirigida entonces

$$d_i^{out} = d_i^{in}$$

Ejercicio #3

Tanto para redes dirigidas como no dirigidas, se define la media global de las interacciones como

$$\tilde{y} = \frac{1}{n(n-1)} \sum_{i,j:i \neq j} y_{i,j}$$

donde $Y = [y_{i,j}]$ es la matriz de adyacencia de la red correspondiente. Tal estadístico describe de manera rudimentaria la posibilidad de observar una relación entre dos nodos cualesquiera, dado que no tiene en cuenta la heterogeneidad nodal (algunos nodos son más propensos a enviar/recibir más relaciones).

1. Muestre que para relaciones no dirigidas la media global \tilde{y} es igual a la media tanto de la parte triangular superior de Y como de la parte triangular inferior de Y .

Respuesta 3.1:

Sea una matriz de adyacencia $\mathbf{Y} = [y_{i,j}]$, se tiene que la media de su triangular inferior L y su triangular superior U se pueden calcular respectivamente como:

$$\begin{aligned}\tilde{u} &= \frac{1}{n(n-1)/2} \sum_{i=1}^{n-1} \sum_{j=i+1}^n y_{i,j} & y & \quad \tilde{l} = \frac{1}{n(n-1)/2} \sum_{i=2}^n \sum_{j=1}^{i-1} y_{i,j} \\ \tilde{u} &= \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n y_{i,j} & y & \quad \tilde{l} = \frac{2}{n(n-1)} \sum_{i=2}^n \sum_{j=1}^{i-1} y_{i,j}\end{aligned}$$

Al tener del ejercicio 2 que \mathbf{Y} es una matriz simétrica en redes no dirigidas, tal que $y_{i,j} = y_{j,i}$, así como la transpuesta de sí misma $\mathbf{Y}^t = \mathbf{Y}$, la parte triangular superior es igual a la parte triangular inferior. Entonces, sus medias son iguales

$$\tilde{u} = \tilde{l}$$

Para probar que la media global \tilde{y} es igual a la media de la parte triangular superior \tilde{u} :

$$\begin{aligned}\tilde{y} &= \frac{1}{n(n-1)} \sum_{i,j:i \neq j} y_{i,j} \\ &= \frac{1}{n(n-1)} \left(\sum_{i=1}^{n-1} \sum_{j=i+1}^n y_{i,j} + \sum_{i=2}^n \sum_{j=1}^{i-1} y_{i,j} \right) \quad (\text{sustitución}) \\ &= \frac{1}{n(n-1)} \left(\sum_{i=1}^{n-1} \sum_{j=i+1}^n y_{i,j} + \sum_{i=1}^{n-1} \sum_{j=i+1}^n y_{i,j} \right) \quad (\text{debido a la simetría}) \\ &= \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n (y_{i,j} + y_{i,j}) \\ &= \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2y_{i,j} \\ &= \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n y_{i,j} \\ \tilde{y} &= \tilde{u}\end{aligned}$$

2. Muestre que tanto para relaciones dirigidas como no dirigidas la media global corresponde a la densidad de la red.

Respuesta 3.2:

La densidad de un grafo se calcula como

$$\text{den}(\mathbf{G}) = \frac{|E|}{|V|(|V|-1)/2}$$

En el caso de un digrafo el denominador debe ser $|V|(|V| - 1)$ [2]. Sea $|E|$ la cantidad de aristas y $|V|$ el número de vértices, entonces:

$$|E| = \frac{1}{2} \sum_{i,j:i \neq j} y_{i,j} \quad \text{para una red no dirigida, } y_{i,j} = y_{j,i}$$

$$|E| = \sum_{i,j:i \neq j} y_{i,j} \quad \text{para una red dirigida}$$

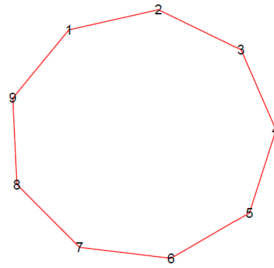
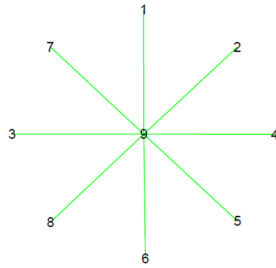
Para probar que la media global corresponde a la densidad de una red no dirigida:

$$\begin{aligned} \tilde{y} &= \frac{1}{n(n-1)} \sum_{i,j:i \neq j} y_{i,j} \\ &= \frac{1}{n(n-1)} \cdot 2|E| \quad (\text{para una red no dirigida, } y_{i,j} = y_{j,i}) \\ &= \frac{2|E|}{n(n-1)} \\ &= \frac{|E|}{n(n-1)/2} \\ &= \frac{|E|}{|V|(|V|-1)/2} \quad (\text{porque } |V| = n) \\ &= \text{den}(\mathbf{G}) \end{aligned}$$

Para una red dirigida, la demostración sigue un razonamiento similar, solo que al sustituir $\sum_{i,j:i \neq j} y_{i,j}$, se cambia por $|E|$ solo.

Ejercicio #4

Considere un grafo estrella de orden n y un grafo círculo de orden n . A continuación se representan ambos grafos para $n = 9$. Estos grafos tienen aproximadamente la misma densidad, pero su estructura es muy diferente. Recuerde que la densidad de un grafo se puede calcular como el grado promedio dividido por $n - 1$.



1. Muestre que para el grafo círculo de orden n el grado promedio es $\tilde{d} = 2$.
2. Muestre que para el grafo estrella de orden n el grado promedio es $\tilde{d} = \frac{2n-1}{n} \rightarrow 2$ cuando $n \rightarrow \infty$.

Respuesta 4.1:

En un grafo círculo de orden n , cada vértice está conectado exactamente con dos vértices adyacentes. Por lo tanto, cada vértice tiene grado 2. La fórmula para el grado promedio \tilde{d} del grafo se define como:

$$\tilde{d} = \frac{2|E|}{|V|}$$

En un grafo círculo de orden n , hay n vértices y n aristas (cada vértice está conectado a dos aristas):

$$\tilde{d} = \frac{2n}{n} = 2$$

Por lo tanto, el grado promedio para un grafo círculo de orden n es $\tilde{d} = 2$.

Respuesta 4.2:

Para un grafo estrella de orden n , el vértice central tiene grado $n - 1$ ya que está conectado a todos los otros $n - 1$ vértices. Los vértices periféricos tienen grado 1 ya que están conectados únicamente al vértice central. Utilizando la anterior fórmula para el grado promedio \tilde{d} de un grafo:

$$\tilde{d} = \frac{2|E|}{|V|}$$

Para un grafo estrella de orden n , hay n vértices y $n - 1$ aristas. Entonces,

$$\tilde{d} = \frac{2(n - 1)}{n}$$

$$\tilde{d} = 2 \frac{n - 1}{n}$$

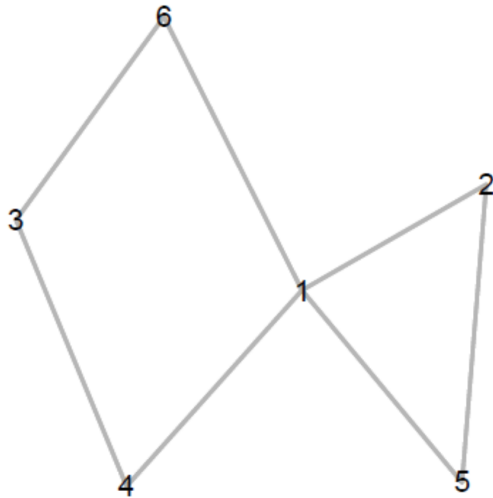
Cuando n tiende a infinito, $\frac{n-1}{n}$ tiende a 1. Por lo tanto,

$$\lim_{n \rightarrow \infty} \tilde{d} = 2$$

Se concluye que cuando n tiende a infinito, el grado promedio para un grafo estrella de orden n tiende a 2.

Ejercicio #5

A continuación se especifican las secuencias que son caminatas, senderos, circuitos y ciclos del grafo.

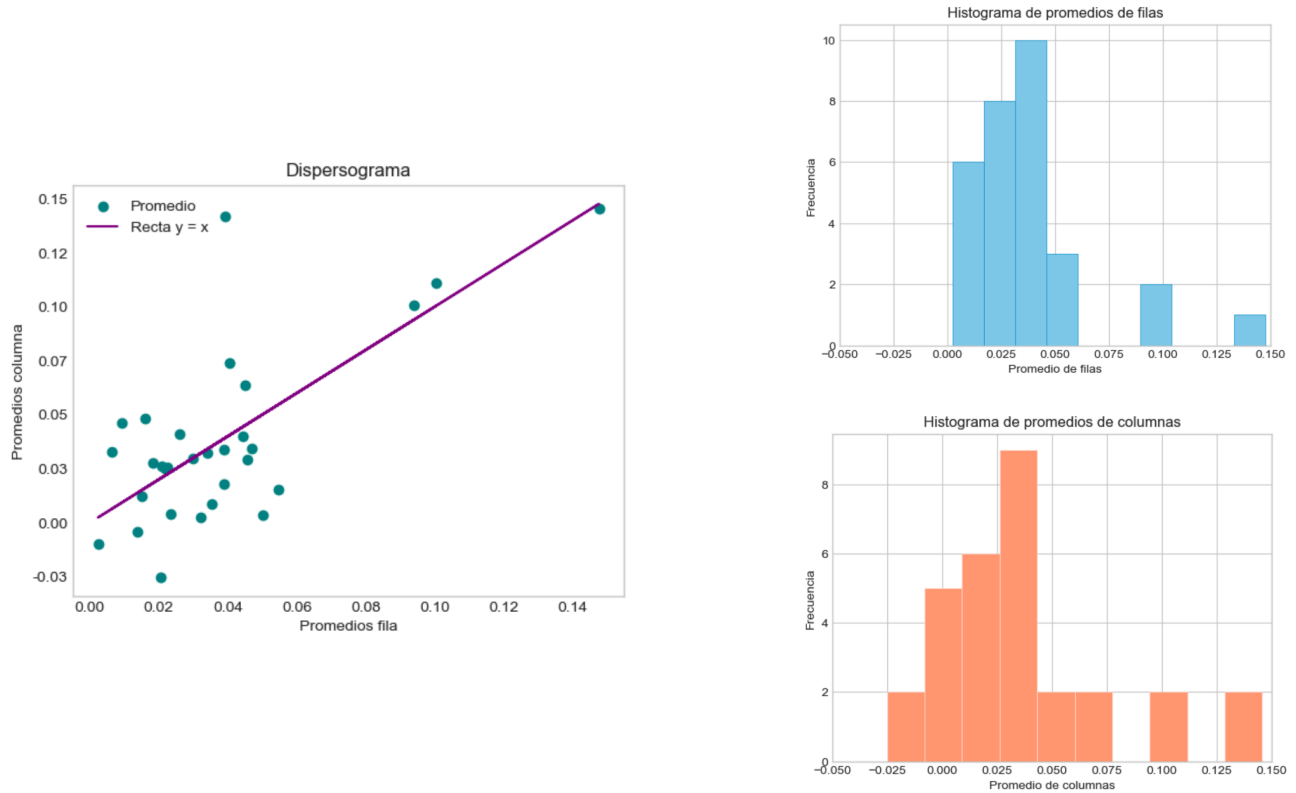


- (a) 2 - 1 - 6 - 3 - 4.
- (b) 2 - 1 - 6 - 3 - 4 - 1 - 5.
- (c) 2 - 1 - 2 - 5 - 1 - 4.
- (d) 2 - 1 - 6 - 3 - 4.
- (e) 2 - 1 - 6 - 3 - 4 - 1 - 5.
- (f) 2 - 1 - 2 - 5 - 1 - 4
- (g) 2 - 1 - 6 - 3 - 4.
- (h) 2 - 1 - 6 - 3 - 4 - 1 - 5.
- (i) 2 - 1 - 2 - 5 - 1 - 4

1. **Caminatas:** Todas las secuencias son *caminatas abiertas*, no hay alguna *caminata cerrada*.
2. **Senderos:** (a), (b), (d), (e), (g), y (h).
3. **Circuitos:** Dado que no hay caminatas cerradas, tampoco hay circuitos (caminatas cerradas sin aristas repetidas) [3].
4. **Ciclos:** Dado que no hay caminatas cerradas, tampoco hay ciclos (caminatas cerradas con al menos tres aristas no repetidas y vértices intermedios distintos) [3].

Ejercicio #6

Considerando los datos presentes en el archivo **comtrade.RData**, el comercio de bienes manufacturados ha experimentado, en promedio, un crecimiento anual del 3.8% ($\tilde{y} = 0.038$) en términos logarítmicos durante la década analizada con respecto al año base 2000.



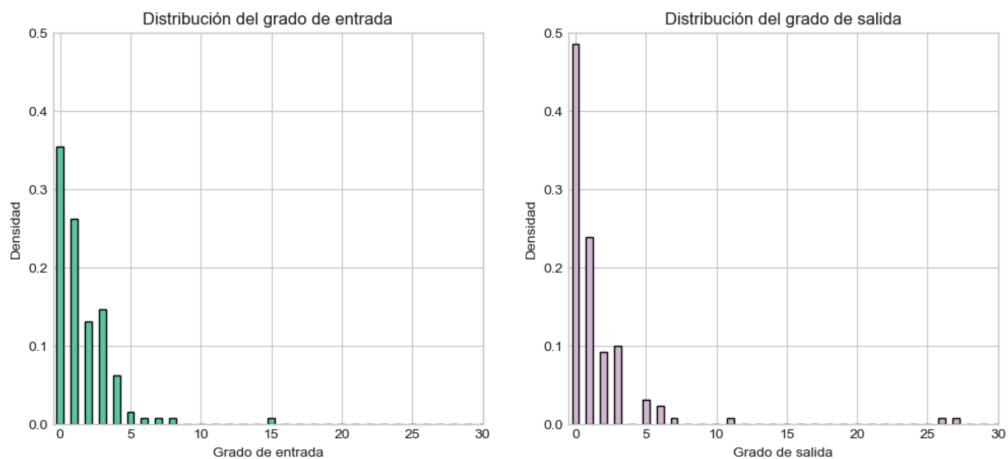
Se calcularon los promedios fila (exportaciones) y columna (importaciones) para cada país. Los resultados muestran que tanto los promedios de exportaciones como de importaciones tienen una media de 0.0378. Sin embargo, la desviación estándar es ligeramente mayor para las importaciones, siendo de 0.041 en comparación con 0.03 para las exportaciones. Esto sugiere una mayor variabilidad en los datos de importaciones en comparación con las exportaciones. Por último, los promedios de filas y los promedios de columnas obtuvieron un coeficiente de correlación de 0.7, indicando una relación positiva fuerte entre las exportaciones e importaciones de los países en el conjunto de datos. Esto significa que, en general, los países que tienden a exportar más también tienden a importar más, y viceversa.

Ejercicio #7

La siguiente figura exhibe una representación visual de la red extraída del archivo **conflict.RData**, presentando diferentes diseños. Mientras tanto, la matriz de adyacencia proporciona información sobre el número de conflictos iniciados por el país i hacia el país j . Una media global de 0.018 indica que, en promedio, cada par de países tiene una probabilidad baja de estar implicado en conflictos durante este período.



Sin embargo, es crucial examinar los grados de entrada y salida. La siguiente figura ilustra la distribución del grado de entrada y salida en la red de conflictos. Ambas distribuciones muestran un sesgo positivo o hacia la derecha, lo que sugiere que la mayoría de los países tienen un bajo número de conflictos tanto iniciados como recibidos, pero hay una pequeña proporción de países que han iniciado más de 25 conflictos y otros que han recibido al menos 15.



En promedio, los nodos en la red tienen un grado de entrada y salida similar, con una media de 1.5615. Sin embargo, la desviación estándar es considerablemente mayor para los grados de salida (3.5756) en comparación con los grados de entrada (1.9768), lo que sugiere una mayor variabilidad en los grados de salida; reflejando que algunos nodos tienen una cantidad considerablemente mayor de conexiones de salida en comparación con otros nodos en la red.

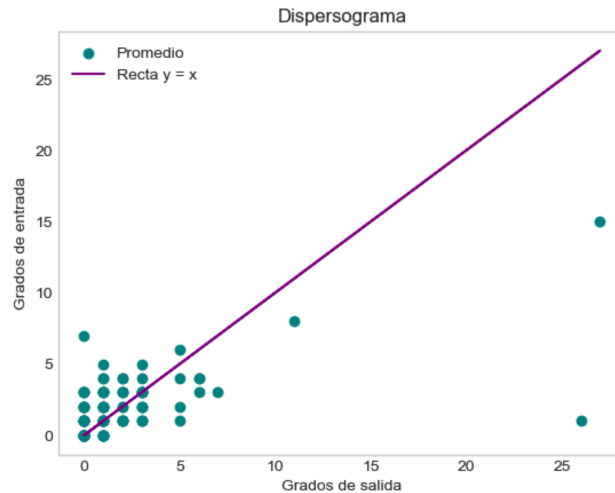


Figure 6: Dispersograma de los grados de entrada y salida

Un coeficiente de correlación de 0.64 entre los valores del out-degree y el in-degree indica una correlación positiva moderadamente fuerte entre estas dos medidas en la red. Esto sugiere que los nodos que tienen un mayor número de conexiones de salida tienden a tener también un mayor número de conexiones de entrada, y viceversa. Por último, los cinco países con los mayores grados de entrada son Irak (IRQ), Estados Unidos (USA), Haití (HAI), Turquía (TUR) y Japón (JPN). Por otro lado, los cinco países con los mayores grados de salida son Irak (IRQ), Jordania (JOR), Estados Unidos (USA), Uganda (UGA) y la República Democrática del Congo (DRC). Esto sugiere que estos países tienen una influencia significativa en las interacciones de la red, ya sea como origen o destino de conexiones.

Ejercicio #8

La siguiente figura muestra una visualización de la red del punto anterior después de eliminar los puntos aislados. Se identificaron 5 componentes con 2 nodos, 2 componentes con 3 nodos y una componente con 32 nodos conectados. Por lo tanto, el tamaño de la componente gigante sería de 32, lo que representa el 67% del grafo inicial. Esta componente gigante comprendería un total de 34 aristas y una conectividad de los vértices de 1.



Posteriormente, se calculó la matriz de distancias geodésicas de la componente gigante y se utilizó un escalamiento multidimensional para incrustar estas distancias en un espacio de dos dimensiones. La figura 8 muestra una relación no lineal entre las coordenadas en este espacio bidimensional.

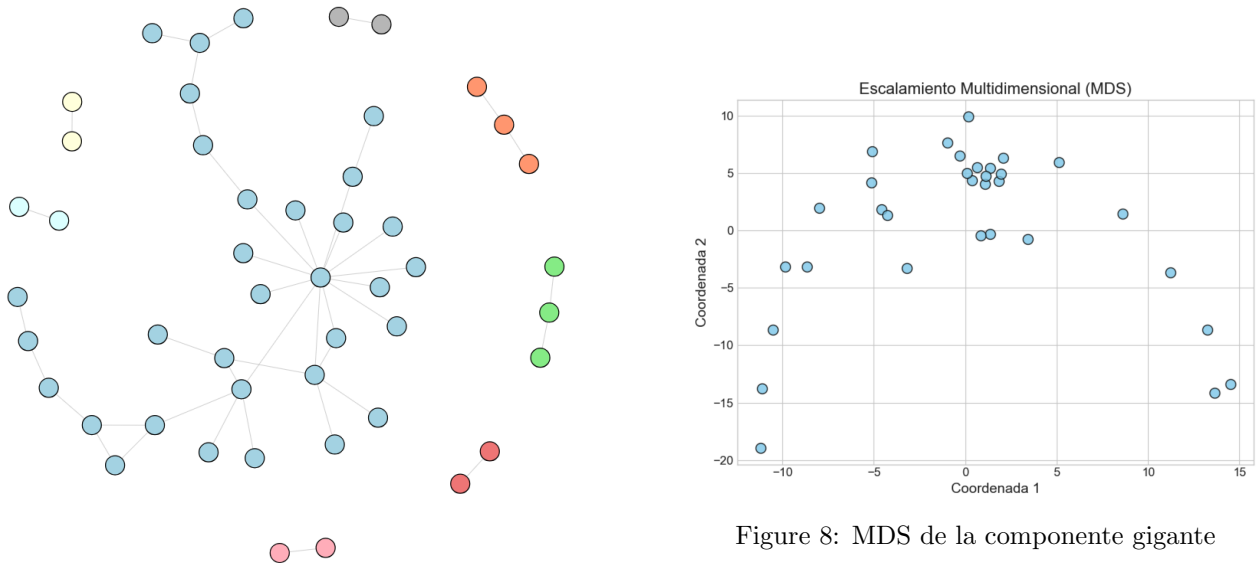


Figure 8: MDS de la componente gigante

Figure 7: Componentes

Ejercicio #9

Congress network representa la red de interacción de Twitter, utilizado en el artículo '*A centrality measure for quantifying spread on weighted, directed networks*' [1], para el 117^o Congreso de los Estados Unidos, tanto de la Cámara de Representantes como del Senado. Los datos base se recopilieron a través de la API de Twitter, y están disponibles en <https://snap.stanford.edu/data/congress-twitter.html>. Corresponde a una *red dirigida* con 475 *vértices* y 13.289 *aristas*. Los vértices son los usuarios de Twitter de los congresistas y las aristas contienen los pesos de conexión, es decir, las probabilidades de transmisión empírica que se cuantificaron de acuerdo con la fracción de veces que un miembro retuiteó, citó, respondió o mencionó el tweet de otro miembro.

Estudiar la red de interacción de Twitter de congresistas puede proporcionar una comprensión más profunda de cómo los políticos y legisladores se comunican y colaboran en plataformas de redes sociales, así como la influencia que tienen dichos usuarios y la formación o identificación de comunidades y/o posturas políticas. Lamentablemente no encontré una red de interacción similar para Colombia, dado que resulta interesante abordar en este tipo de redes la evidencia de polarización política, el uso de la información (o la desinformación) por parte de partidos políticos, etc.

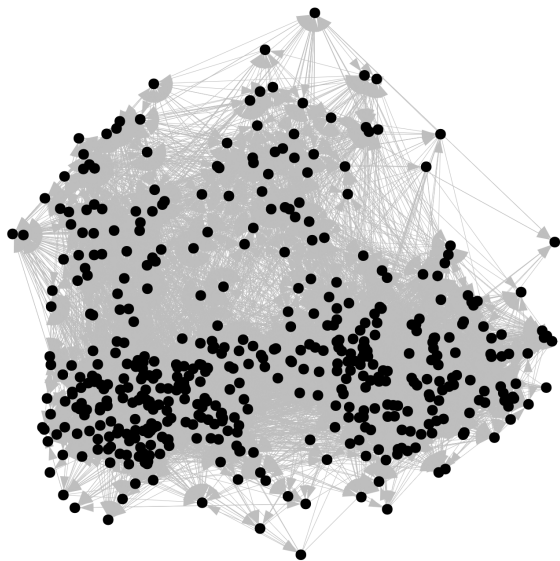


Figure 9: Visualización simple

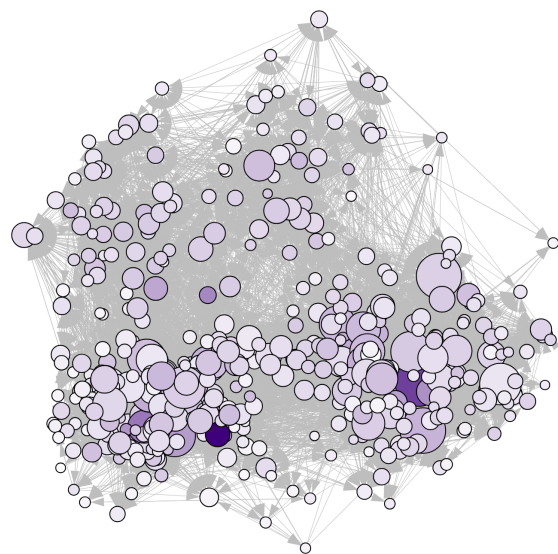


Figure 10: Visualización decorada

La Figura 9 y Figura 10 corresponden a visualizaciones de la base de datos **Congress network**. La Figura 10 se decoró de tal forma que la intensidad del color representa los grados de salida (interacción en Twitter hacia otro miembro del congreso) y, por otro lado, los grados de entrada se ven representados por el tamaño de los vértices que, corresponderían a los miembros del congreso más populares en esta red social. De acuerdo con lo anterior, se observa lo siguiente: 1) Se percibe una segmentación en dos grupos distintos; 2) Los congresistas que más envían interacciones no necesariamente son los que reciben más; 3) Un número reducido de miembros del congreso envían un volumen significativo de interacciones, mientras que hay varios nodos populares que las reciben.

References

- [1] Christian G Fink, Kelly Fullin, Guillermo Gutierrez, Nathan Omodt, Sydney Zinnecker, Gina Sprint, and Sean McCulloch. A centrality measure for quantifying spread on weighted, directed networks. *Physica A*, 2023.
- [2] Juan Camilo Sosa. Caracterización de conectividad.
- [3] Juan Camilo Sosa. Conceptos fundamentales de teoría de grafos.
- [4] Juan Camilo Sosa. Gestión de datos relacionales.