

1. Parte 1, 2, 3 Overleaf, 4 en Google Colab.

1. Para un grafo dirigido considera las siguientes medidas de centralidad: Eigencentrality, PageRank, Katz-Bonacich

a) Definir la medida de centralidad de Eigencentrality para un grafo dirigido.

**Sol:** Para un grafo con matriz de adyacencia  $\mathbf{A}$  tenemos que la centralidad  $x_v$  de un vértice  $v$  está dada en términos de la centralidad  $x_u$  de todos los vértices  $u$  a los que está conectado  $v$ , y a una constante  $\lambda$ , como

$$x_v = \frac{1}{\lambda} \sum_u A_{v,u} x_u,$$

lo cual puede escribirse como el problema de valores propios

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x},$$

que aunque en general tendrá varias soluciones, el requerimiento de que todas las medidas de centralidad sean no-negativas implica que debe considerarse la solución con valor propio máximo (por Perron-Frobenius).

Cuando tenemos que el grafo es dirigido, esta medida de centralidad será para aristas entrantes, por lo que si se quiere la centralidad para aristas salientes debe primero reversarse el grafo, i.e., considerar  $\mathbf{A}^t$ .

b) Calcular la expansión a primer orden del radio espectral para perturbaciones de la matriz.

**Sol:** En el problema de valores propios anterior, el valor propio que consideramos es el máximo, que corresponde con el radio espectral. Siendo así, el radio espectral está dado como solución a la ecuación

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

con todas las entradas de  $\mathbf{x}$  no-negativas. Considérese también el vector propio a izquierda asociado al mismo valor propio  $\lambda$  dado por

$$\mathbf{y}^t \mathbf{A} = \lambda \mathbf{y}^t,$$

así como una pequeña perturbación  $\mathbf{A} + \Delta\mathbf{A}$  de la matriz.

El problema de valores propios perturbado es ahora

$$\begin{aligned} (\mathbf{A} + \Delta\mathbf{A})(\mathbf{x} + \Delta\mathbf{x}) &= (\lambda + \Delta\lambda)(\mathbf{x} + \Delta\mathbf{x}) \\ \mathbf{A}\mathbf{x} + \mathbf{A}\Delta\mathbf{x} + \Delta\mathbf{A}\mathbf{x} + \Delta\mathbf{A}\Delta\mathbf{x} &= \lambda\mathbf{x} + \lambda\Delta\mathbf{x} + \Delta\lambda\mathbf{x} + \Delta\lambda\Delta\mathbf{x} \\ \mathbf{A}\Delta\mathbf{x} + \Delta\mathbf{A}\mathbf{x} + \Delta\mathbf{A}\Delta\mathbf{x} &= \lambda\Delta\mathbf{x} + \Delta\lambda\mathbf{x} + \Delta\lambda\Delta\mathbf{x} \\ \mathbf{A}\Delta\mathbf{x} + \Delta\mathbf{A}\mathbf{x} &= \lambda\Delta\mathbf{x} + \Delta\lambda\mathbf{x}, \end{aligned}$$

donde en la última línea se usó que la perturbación es pequeña. Multiplicando ahora por  $\mathbf{y}^t$  a izquierda tenemos

$$\begin{aligned}
\mathbf{y}^t \mathbf{A} \Delta \mathbf{x} + \mathbf{y}^t \Delta \mathbf{A} \mathbf{x} &= \mathbf{y}^t \lambda \Delta \mathbf{x} + \mathbf{y}^t \Delta \lambda \mathbf{x} \\
\mathbf{y}^t \Delta \mathbf{A} \mathbf{x} &= \mathbf{y}^t \Delta \lambda \mathbf{x} \\
\Delta \lambda &= \frac{\mathbf{y}^t \Delta \mathbf{A} \mathbf{x}}{\mathbf{y}^t \mathbf{x}},
\end{aligned}$$

donde vemos que a primer orden la perturbación del valor propio (el radio espectral) está dada en términos de los vectores propios a izquierda y derecha asociados al mismo y a la perturbación de la matriz.

- c) Demostrar que aproximadamente a primer orden, el nodo que más contribuye al radio espectral es el que tiene mayor eigencentality

**Sol:** En el caso de un grafo dirigido no tenemos que la matriz de adyacencia sea simétrica, por lo que los vectores propios a izquierda y derecha no tienen por qué ser iguales. Sin embargo de manera análoga al caso de un grafo no dirigido podemos calcular el cambio relativo en el radio espectral al remover la fila y columna  $i$ -ésima, i.e., al remover el nodo  $i$  del grafo, como

$$\begin{aligned}
\Delta \rho(\mathbf{A}) &= \frac{\mathbf{y}^t (-(\mathbf{A} \mathbf{e}_i) \mathbf{e}_i^t - \mathbf{e}_i (\mathbf{A} \mathbf{e}_i)^t) \mathbf{x}}{\mathbf{y}^t \mathbf{x}} \\
&= \frac{-\mathbf{y}^t (\mathbf{A} \mathbf{e}_i) \mathbf{e}_i^t \mathbf{x} - \mathbf{y}^t \mathbf{e}_i (\mathbf{A} \mathbf{e}_i)^t \mathbf{x}}{\mathbf{y}^t \mathbf{x}} \\
&= \frac{-\mathbf{y}^t (\mathbf{A} \mathbf{e}_i) x_i - y_i (\mathbf{A} \mathbf{e}_i)^t \mathbf{x}}{\mathbf{y}^t \mathbf{x}} \\
&= -\mathbf{x}^t (\mathbf{A} \mathbf{e}_i) x_i - y_i (\mathbf{A} \mathbf{e}_i)^t \mathbf{y} \\
&= -\rho(\mathbf{A}) \mathbf{x}^t \mathbf{e}_i x_i - \rho(\mathbf{A}) y_i \mathbf{e}_i^t \mathbf{y} \\
\frac{\Delta \rho(\mathbf{A})}{\rho(\mathbf{A})} &= -(x_i^2 + y_i^2),
\end{aligned}$$

es decir que a primer orden el nodo que más contribuye al radio espectral es aquel con mayor eigencentality (tomando en este caso la suma de los cuadrados de la centralidad para aristas entrantes y salientes).

- d) Determinar una medida de centralidad para aristas para encontrar la arista que contribuye mas (en la aproximacion a primer orden) al radio espectral de un grafo

**Sol:**

- e) Para un grafo dirigido definir PageRank y PageRank personalizado

**Sol:** La medida de PageRank da la distribución de probabilidad de llegar a un vértice del grafo siguiendo aristas aleatoriamente. En su forma más simple, el PageRank  $r_v$  para un vértice  $v$  es sencillamente la suma del PageRank de todos los vértices  $u$  que tienen aristas de  $u$  hacia  $v$ , dividido cada uno por el número total  $L_u$  de aristas salientes que tiene  $u$ , i.e.,

$$r_v = \sum_u A_{v,u} \frac{r_u}{L_u}.$$

Por otra parte, la teoría incorpora un factor  $d$  llamado *damping parameter* que tiene en cuenta que en algún punto se dejan de seguir aristas y el recorrido se detiene, con lo cual el PageRank se define ahora como

$$r_v = \frac{1-d}{N} + d \sum_u A_{v,u} \frac{r_u}{L_u},$$

donde  $N$  es el número total de vértices.

El PageRank personalizado se refiere a esta misma medida pero con sesgo hacia un conjunto específico de vértices como vértices iniciales para realizar el recorrido.

- f) Demostrar que PageRank existe y es único y es una función racional del "damping parameter", porque es la solución a un sistema lineal con matriz asociada invertible.

**Sol:** Para esto notamos que la ecuación presentada anteriormente para el PageRank de un vértice puede escribirse para todos los vértices en términos matriciales, tomando  $\mathbf{r}$  como el vector de todos los PageRanks tenemos la ecuación

$$\mathbf{r} = \frac{1-d}{N} \mathbf{1} + d \tilde{\mathbf{A}} \mathbf{r},$$

donde  $\mathbf{1}$  es el vector columna con  $N$  entradas todas iguales a 1 y  $\tilde{\mathbf{A}}$  es la matriz de adyacencia modificada tal que sea una matriz estocástica (toda columna suma a 1). Por supuesto esto tiene solución pues así como la matriz de adyacencia es invertible, también lo es  $\tilde{\mathbf{A}}$ .

- g) Demostrar que PageRank es el vector propio de una matriz positiva e irreducible y por lo tanto puede ser calculado usando el método de potencia.

**Sol:** Para esto notamos ahora que como por construcción  $\mathbf{r}$  es una distribución de probabilidad,  $\mathbf{J}\mathbf{r} = \mathbf{1}$ , donde  $\mathbf{J}$  es una matriz  $N \times N$  con todas las entradas iguales a 1, con lo cual podemos reescribir la ecuación matricial como

$$\mathbf{r} = \left( \frac{1-d}{N} \mathbf{J} + d \tilde{\mathbf{A}} \right) \mathbf{r},$$

lo cual identificamos inmediatamente como un problema de valor propio donde  $\mathbf{r}$  es el vector propio asociado al valor propio  $\lambda = 1$  (dado que estamos tratando con una matriz estocástica éste es el valor propio máximo y tenemos por Perron-Frobenius que éste es el vector propio principal).

Que la matriz es positiva es evidente pues es la suma de un múltiplo de  $\tilde{\mathbf{A}}$ , que es una matriz estocástica (y por lo tanto positiva), y la matriz con todas las entradas iguales a  $\frac{1-d}{N}$ , que es igualmente positiva. Que la matriz es irreducible es consecuencia de que no se trate únicamente de la matriz de adyacencia sino que contenga el término  $\frac{1-d}{N} \mathbf{J}$  con el cual todas las entradas de la matriz son  $\neq 0$ . Con esto tenemos que el PageRank puede calcularse como el vector propio de una matriz positiva e irreducible, y por lo tanto puede calcularse usando el método de potencia.

- h) Usando formulas relacionadas con la formula de Sherman-Morrison calcular el orden de convergencia del método potencia aplicado al calculo de PageRank

**Sol:**

- i) Definir la centralidad de Katz-Bonacich y la total Katz-Bonacich para un grafo dirigido.

**Sol:** Para un grafo dirigido, la matriz de adyacencia  $\mathbf{A}$  tiene entradas  $A_{j,i} \neq 0$  donde se tenga una arista que conecte el vértice  $i$  al vértice  $j$ . Una extensión natural de esto es tomar productos de esta matriz con sí misma  $\mathbf{A}^k$ , donde tendremos que  $(A^k)_{j,i} \neq 0$  donde se tenga un camino con  $k$  aristas que conecte el vértice  $i$  al vértice  $j$ . Consideramos entonces como medida de centralidad  $c_v$  del vértice  $v$  la suma de todos los caminos hacia  $v$  desde cualquier otro vértice  $u$  y de cualquier longitud, i.e.,

$$c_v = \sum_k \sum_u (A^k)_{v,u}.$$

Sin embargo debería incluirse un cierto factor de atenuación  $\lambda$  tal que se “penalicen” caminos más largos, en otras palabras, un camino desde un vértice  $u$  contribuye más a la centralidad del vértice  $v$  entre más corto sea, por lo que finalmente escribimos la centralidad como

$$c_v = \sum_k \sum_u \lambda^k (A^k)_{v,u},$$

o en términos matriciales (notando que se tiene algo como una serie geométrica)

$$\mathbf{c} = (\mathbf{I} - \lambda \mathbf{A})^{-1} \mathbf{1}.$$

La centralidad así definida, nuevamente, es para aristas entrantes, en caso de querer calcularse para aristas salientes sencillamente se usa la matriz de adyacencia transpuesta. Por último se puede definir una total Katz-Bonacich, como el nombre lo sugiere, como la suma de las centralidades para todos los vértices, en términos matriciales

$$\mathbf{1}^t \mathbf{c}.$$

- j) Demostrar que la centralidad de Katz-Bonacich y la total Katz-Bonacich se puede interpretar en términos de conteo de caminos

**Sol:** Como ya se mencionó en la definición de la centralidad y la total de Katz-Bonacich, las potencias de la matriz de adyacencia dan los distintos caminos que existen entre vértices de una cierta longitud, y la forma en que se calcula la centralidad de un vértice no es más que el número de caminos que llevan a él con la atenuación dada por  $\lambda^k$  para cada camino de longitud  $k$  (si además el grafo es pesado entonces esta suma tendrá en cuenta los pesos de las aristas que conforman cada camino).

Inmediatamente con esto, y con la definición de la total como la suma de las centralidades, podemos interpretar la total como un conteo (potencialmente pesado) del número total de caminos que pueden seguirse en el grafo con la atenuación dada por  $\lambda^k$  para cada camino de longitud  $k$ .

- k) Usar el teorema de Perron-Frobenius para demostrar que no es posible definir la centralidad de Katz-Bonacich para un valor mayor que el radio espectral que siempre sea positiva.

**Sol:**

El teorema de Perron-Frobenius nos dice que el único valor propio con vector propio asociado estrictamente positivo es  $\rho(\mathbf{A})$ , que además es el valor propio máximo, dada una matriz no negativa  $\mathbf{A}$  (como lo es la matriz de adyacencia). Ahora si tenemos  $\lambda > \rho(\mathbf{A})$  y buscamos  $\mathbf{c}$  positivo, debe ser que éste es un múltiplo positivo del vector propio asociado a  $\rho(\mathbf{A})$  y es en sí un vector propio asociado a un valor propio  $< \rho(\mathbf{A})$ , una contradicción.

- l) Demostrar que la centralidad de Katz-Bonacich es una función racional del parametro  $\lambda$  y que el polo en  $\frac{1}{\rho(\mathbf{A})}$  se puede calcular en términos de eigencentality.

**Sol:** Considérese la expresión matricial en términos de la descomposición espectral de la matriz de adyacencia

$$\begin{aligned}\mathbf{c} &= (\mathbf{I} - \lambda \mathbf{A})^{-1} \mathbf{1} \\ &= \sum_{n=1}^N \frac{\mathbf{x}_n \mathbf{x}_n^t}{1 - \lambda \lambda_n} \mathbf{1} \\ &= \sum_{n=1}^N \frac{(\mathbf{1}^t \mathbf{x}_n) \mathbf{x}_n}{1 - \lambda \lambda_n},\end{aligned}$$

con lo cual la centralidad de Katz-Bonacich es una función racional de  $\lambda$  con polos en  $\lambda = \frac{1}{\lambda_n}$  para cada valor propio  $\lambda_n$ . En particular considerando el polo en  $\lambda = \frac{1}{\rho(\mathbf{A})}$ , con el radio espectral  $\rho(\mathbf{A})$  el valor propio dominante, tenemos

$$(\mathbf{1}^t \mathbf{x}) \mathbf{x} = \mathbf{x} \sum_i x_i,$$

donde como en el ejercicio (a)  $x_i$  es el eigencentality del vértice  $i$  y  $\mathbf{x}$  es el vector de eigencentalityes.

2. Para un modelo determinístico Multi-SIR en un conjunto de poblaciones todas del mismo tamaño sin dinámica vital (no hay nacimientos ni muertes)
- a) Determine la relación que deben cumplir el número final de infectados como función de  $R_0$  (curiosamente la dependencia con respecto a los infectados iniciales es binaria).

**Sol:** Consideramos un sistema

$$\begin{aligned}\frac{d\mathbf{S}}{dt} &= -\beta \text{diag}(\mathbf{S}) \mathbf{A} \mathbf{I}, \\ \frac{d\mathbf{I}}{dt} &= \beta \text{diag}(\mathbf{I}) \mathbf{A} \mathbf{S} - \gamma \mathbf{I}, \\ \frac{d\mathbf{R}}{dt} &= \gamma \mathbf{I},\end{aligned}$$

donde tratamos con *vectores* de números de susceptibles, infectados, y muertos, y modelamos las  $N$  poblaciones por un grafo no dirigido sin loops con matriz de adyacencia  $\mathbf{A}$

(en esta primera aproximación asumimos por simplicidad que los parámetros de infección y recuperación,  $\beta$  y  $\gamma$ , son los mismos para todas las poblaciones y es por esto que pueden representarte sencillamente como escalares). Para una población  $k$  dada tenemos entonces

$$\begin{aligned}\frac{dS_k}{dt} &= -\beta S_k \sum_l A_{kl} I_l, \\ \frac{dI_k}{dt} &= \beta I_k \sum_l A_{kl} S_l - \gamma I_k, \\ \frac{dR_k}{dt} &= \gamma I_k,\end{aligned}$$

con lo cual podemos hallar una relación similar a la que se tiene en el caso con una sola población, i.e.,

$$\sum_l A_{kl} R'_l = -\frac{\gamma}{\beta} (\log S_k)',$$

y así tener las  $N$  cantidades conservadas

$$\sum_l A_{kl} (S_l + I_l - \frac{\gamma}{\beta} \log S_l).$$

Consideramos condiciones iniciales  $S_k(0) = S_k^0, I_k(0) = I_k^0, R_k(0) = 0, \forall k$  y el límite cuando  $t \rightarrow \infty$  con  $S_k \rightarrow S_k^\infty, I_k \rightarrow 0, R_k \rightarrow R_k^\infty$  con esta cantidad conservada y definimos el número básico de reproducción  $R_0 = \frac{\beta}{\gamma}$  para escribir

$$\begin{aligned}\sum_l A_{kl} (S_l^\infty - \frac{\gamma}{\beta} \log S_l^\infty) &= \sum_l A_{kl} (S_l^0 + I_l^0 - \frac{\gamma}{\beta} \log S_l^0) \\ \sum_l A_{kl} \log \frac{1 - R_l^\infty}{S_l^0} &= -R_0 \sum_l A_{kl} R_l^\infty \\ \prod_l \left( \frac{1 - R_l^\infty}{S_l^0} \right)^{A_{kl}} &= e^{-R_0 \sum_l A_{kl} R_l^\infty} \\ R_k^\infty &= 1 - S_k^0 e^{-R_0 \sum_l A_{kl} R_l^\infty}\end{aligned}$$

- b) Considere el numero final de infectados como función de  $R_0$  y calcule la derivada con respecto a  $R_0$  (implícitamente).

**Sol:** Derivamos directamente

$$\begin{aligned}\frac{d}{dR_0} R_k^\infty(R_0) &= \frac{d}{dR_0} \left( 1 - S_k^0 e^{-R_0 \sum_l A_{kl} R_l^\infty(R_0)} \right) \\ &= S_k^0 \left( \sum_l A_{kl} R_l^\infty(R_0) + \sum_l A_{kl} (R_l^\infty)'(R_0) \right) e^{-R_0 \sum_l A_{kl} R_l^\infty(R_0)},\end{aligned}$$

y si luego evaluamos la derivada en  $R_0 = 0$ , considerando que  $R_l^\infty(0) = 0$ , tendremos

$$\begin{aligned}
(R_k^\infty)'(0) &= S_k^0 \left( \sum_l A_{kl} R_l^\infty(0) + \sum_l A_{kl} (R_l^\infty)'(0) \right) \\
&= S_k^0 \sum_l A_{kl} (R_l^\infty(0) + (R_l^\infty)'(0)) \\
&= S_k^0 \sum_l A_{kl} (R_l^\infty)'(0),
\end{aligned}$$

que es el eigencentality  $x_k$  con  $\lambda = \frac{1}{S_k^0}$ . Luego podremos aproximar el número final de infectados como

$$\begin{aligned}
R^\infty(R_0) &= \sum_k R_k^\infty(R_0) \\
&\approx \sum_k (R_k^\infty)'(0) R_0 \\
&= R_0 \mathbf{1}^t \mathbf{x}.
\end{aligned}$$

3. Para un modelo determinístico Multi-SIR en un conjunto de poblaciones todas del mismo tamaño sin dinámica vital (no hay nacimientos ni muertes) considere la relación que debe cumplir el número final de infectados como función de la fracción inicial de infectados en cada nodo (la misma para cada población) y calcule la derivada de esa función (implícitamente).

**Sol:** Consideramos una fracción inicial de infectados  $\alpha$ , en principio igual para toda población. Derivamos ahora directamente

$$\begin{aligned}
\frac{d}{d\alpha} R_k^\infty(\alpha) &= \frac{d}{d\alpha} \left( 1 - (1 - \alpha) e^{-R_0 \sum_l A_{kl} R_l^\infty(\alpha)} \right) \\
&= e^{-R_0 \sum_l A_{kl} R_l^\infty(\alpha)} + (1 - \alpha) R_0 \sum_l A_{kl} (R_l^\infty)'(\alpha) e^{-R_0 \sum_l A_{kl} R_l^\infty(\alpha)},
\end{aligned}$$

y si luego evaluamos la derivada en  $\alpha = 0$ , considerando que  $R_l^\infty(0) = 0$  tendremos

$$\begin{aligned}
(R_k^\infty)'(0) &= 1 + R_0 \sum_l A_{kl} (R_l^\infty)'(0), \\
\mathbf{R}^{\infty'}(0) &= (\mathbf{I} - R_0 \mathbf{A})^{-1} \mathbf{1}
\end{aligned}$$

que es la centralidad de Katz-Bonacich  $\mathbf{c}$  con  $\lambda = R_0$ . Luego podremos aproximar el número final de infectados como

$$\begin{aligned}
R^\infty(\alpha) &= \sum_k R_k^\infty(\alpha) \\
&\approx \sum_k (R_k^\infty)'(0) \alpha \\
&= \alpha \mathbf{1}^t \mathbf{c}.
\end{aligned}$$

4. Considere grafos aleatorios Erdős-Rényi, Watts-Strogatz y Barabási-Albert con el mismo número de nodos y aristas (o valor esperado). Para varios valores de los parámetros. (Ejemplos de módulos de python de grafos que pueden usar: Networkx, Igraph, Networkit, Graph-tools) compare el uso de las medidas de centralidad como PageRank, Eigencentrality y Katz-Bonacich para controlar epidemias: Por ejemplo vaya removiendo nodos y mire cómo disminuye el número total de infectados.

**Sol:** Basado en los resultados anteriores, tenemos que el número total de infectados puede ser aproximado como algo proporcional a la suma de las centralidades de todos los vértices (lo vimos para eigencentrality y para Katz-Bonacich pero lo asumiremos también para PageRank), luego debería poderse comparar el efecto en cada caso tomando el cambio relativo de las sumas de las centralidades en cada caso.

Calculos en notebook.

Notamos que dados los grafos aleatorios que se usaron, el caso usando centralidad de Katz-Bonacich resultó en la reducción más grande, seguido del caso usando PageRank, y por último el caso usando Eigencentrality. Sin embargo, cabe resaltar que las diferencias no fueron muy grandes y en general con cualquier medida de centralidad tenemos que la eliminación del nodo con mayor centralidad resultó en un final size aproximadamente 92% tan grande como el original.