

Análisis de Redes Sociales

Medidas descriptivas

Alejandro Espinosa-Rada
Social Networks Lab
ETH Zurich

10 de julio de 2023

Tabla de Contenidos

- 1 Centralidades
- 2 Grados Nodales
- 3 Densidad
- 4 Distancias
- 5 Centralidad de Cercanía
- 6 Centralidad de Intermediación
- 7 Eigenvector
- 8 Índices de centralidad

Tabla de Contenidos

- 1 Centralidades
- 2 Grados Nodales
- 3 Densidad
- 4 Distancias
- 5 Centralidad de Cercanía
- 6 Centralidad de Intermediación
- 7 Eigenvector
- 8 Índices de centralidad

¿Cuál es la preeminencia de un actor en la red? (Wasserman y Faust, 1994)

- Puede ser debido al número de conexiones que goza un actor, su cercanía con otros o quizás por sus niveles de intermediación (Freeman, 1979)
- Tal vez los actores que a su vez están bien conectados con otros nodos igualmente relevantes (Bonacich, 1972).
- Entre muchísimas otras opciones...

Algunas implicancias de posiciones aventajadas en la red (Borgatti y Everett, 2006):

- Organizaciones influyentes (Laumann and Pappi, 1976; Marsden and Laumann, 1977; Galaskiewicz, 1979)
- Poder (Burt, 1982; Knoke and Burt, 1983)
- Ventajas de intercambios en la red (Cook et al., 1983; Marsden, 1982)
- Competencia en organizaciones formales (Blau, 1963)
- Oportunidades para encontrar trabajo (Granovetter, 1974)
- Adopción de innovaciones (Coleman et al., 1966)
- Relaciones corporativas (Mariolis, 1975; Mintz and Schwartz, 1985; Mizruchi, 1982)

Centralidad:

- La noción de punto central como representación de grafo se originó en el concepto sociométrico de “estrella”
- El estudio experimental de Alex Bavelas (velocidad y eficiencia en la difusión de información en un grupo) dió origen al estudio y formalización sobre las posiciones relativas de los actores.
- La centralidad refiere a estar “en el centro” de un número de conexiones.
- Centralidad local: un nodo focal en el vecindario.
- Centralidad global: preeminencia en la red global.

Grafo estrella: uno de sus nodos es completamente superior en centralidad al resto de los nodos y, además, los otros nodos son intercambiables entre sí.

Figura: Grafo Estrella



Grafo circular: todos los nodos son intercambiables y, por consiguiente, deberían tener la misma centralidad.

Figura: Grafo Circular

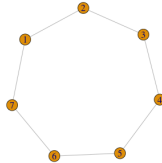
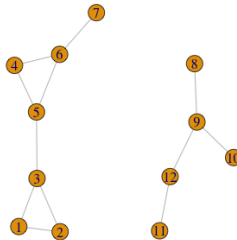


Tabla de Contenidos

- 1 Centralidades
- 2 **Grados Nodales**
- 3 Densidad
- 4 Distancias
- 5 Centralidad de Cercanía
- 6 Centralidad de Intermediación
- 7 Eigenvector
- 8 Índices de centralidad

Grado nodal (*node degree*):

- Indica “actividad” del actor que representa.
- Número de líneas que hay incidentes a un nodo.
- Mínimo: 0 (aislado).
- Máximo: $(g - 1)$.
- $d(n_i)$



Ejemplos de grados nodales en Grafo 1:

- $d(n_1) = 2$
- $d(n_3) = 3$
- $d(n_{10}) = 1$

Grado nodal medio:

$$\bar{d} = \frac{\sum_{i=1}^g (d(n_i))}{g} = \frac{2L}{g} \quad (1)$$

- Grado promedio de los nodos del grafo

Varianza de los grados:

$$S_D^2 = \frac{\sum_{i=1}^g (d(n_i) - \bar{d})^2}{g} \quad (2)$$

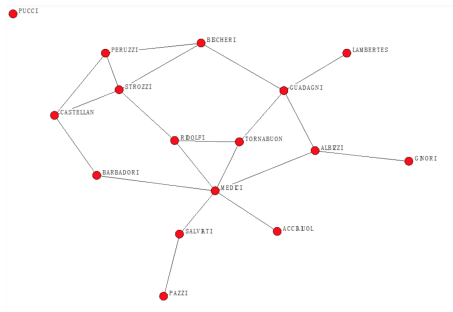
- Si $S_D^2 = 0$, grafo es *d-regular*.

Estandarización de grado nodal:

$$\frac{\sum C_D(n_i)}{g(g-1)} \quad (3)$$

- Matemáticamente, la densidad es también el grado nodal estandarizado.
- Se estandariza para hacer comparables las mediciones.
- El índice tiene un rango entre 0 y 1.

Familias de la Elite Renacentista del siglo XV (Padgett y Ansell, 1993)

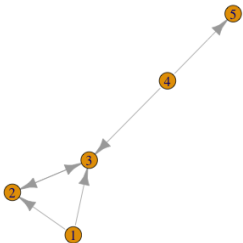


¿Cuál es la familia que posee más conexiones?

Grado de nodos en un grafo dirigido

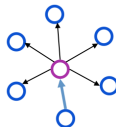
- El grado de entrada (*in*) de un nodo $d_I(n_i)$ es el número de nodos que son adyacentes hacia n_i .
- El grado de salida (*out*) de un nodo $d_O(n_i)$ es el número de nodos que son adyacentes desde n_i .

Figura: Grafo 2 - Grados nodales direccionados



Ejemplos (hipótesis):

- Un documento muy citado tendrá mucho prestigio (*in degree centrality*).
- Una persona que emite mucha información tendrá más incidencia en la población (*out degree centrality*).



Grado de entrada y salida medio:

$$\bar{d}_I = \frac{\sum_{i=1}^g (d_I(n_i))}{g} \quad (4)$$

$$\bar{d}_O = \frac{\sum_{i=1}^g (d_O(n_i))}{g} \quad (5)$$

Varianza de los grados de entrada y salida:

$$S_{D_I}^2 = \frac{\sum_{i=1}^g (d_I(n_i) - \bar{d}_I)^2}{g} \quad (6)$$

$$S_{D_O}^2 = \frac{\sum_{i=1}^g (d_O(n_i) - \bar{d}_O)^2}{g} \quad (7)$$

Nodo es $a(n)$:

- *Aislado*: $d_I(n_i) = d_O(n_i) = 0$
- *Transmisor*: $d_I = 0$ y $d_O(n_i) > 0$
- *Receptor*: $d_I(n_i) > 0$ y $d_O(n_i) = 0$
- *Portador y ordinario*: $d_I(n_i) > 0$ y $d_O(n_i) > 0$

Cálculo de grados nodales a través de matrices:

$$d(n_i) = \sum_{j=1}^g x_{ij} = \sum_{j=1}^g x_{ji} = x_{i+} = x_{+j} \quad (8)$$

- Equivale a las sumas de las filas o las sumas de las columnas.

Cálculo de grados nodales a través de matrices:

$$d_O(n_i) = \sum_{j=1}^g x_{ij} = x_{i+} \quad (9)$$

$$d_I(n_i) = \sum_{j=1}^g x_{ji} = x_{+j} \quad (10)$$

- Totales filas: igual a los grados nodales de salida.
- Totales columnas: igual a los grados nodales de entrada.

Tabla de Contenidos

- 1 Centralidades
- 2 Grados Nodales
- 3 **Densidad**
- 4 Distancias
- 5 Centralidad de Cercanía
- 6 Centralidad de Intermediación
- 7 Eigenvector
- 8 Índices de centralidad

Densidad:

- Es una medida simple de cohesión, interpretada como la probabilidad de que exista una relación entre un par de nodos aleatorios.
- Suele usarse la densidad de un modo comparativo.
- La densidad en general depende del tamaño de la red:
 - Redes pequeñas la densidad suele ser alta
 - Redes grandes la densidad suele ser baja
- Una alternativa para comparar redes de distintos tamaños son los promedios de grados nodales.
- En caso de redes ego-céntricas: se suele extraer ego para observar los vínculos existentes entre sus contactos.

Densidad de los grafos:

$$\Delta = \frac{L}{g(g-1)/2} = \frac{2L}{g(g-1)} \quad (11)$$

- Proporción de líneas en el grafo total.
- Mínimo es 0 cuando $L = 0$.
- Máximo es 1 cuando todas las líneas se encuentran presentes.
- Donde L corresponde al total de líneas posibles.
- Existen $g(g-1)/2$ posibles líneas que pueden estar presentes en el grafo.

Densidad de subgrafo:

$$\Delta_s = \frac{2L_s}{g_s(g_s - 1)} \quad (12)$$

- Proporción de lazos que están presentes entre un subconjunto de actores en la red.
- Medida que permite evaluar la cohesión de los subgrupos

Densidad de los grafos dirigidos:

$$\Delta = \frac{L}{g(g-1)} \quad (13)$$

- Par ordenado
- Proporción de líneas en el grafo total.
- Mínimo es 0 cuando $L = 0$.
- Máximo es 1 cuando todas las líneas se encuentran presentes.

La densidad de un grafo (no dirigido, dirigido o valorado):

$$\Delta = \frac{\sum_{i=1}^g \sum_{j=1}^g x_{ij}}{g(g-1)} \quad (14)$$

Tabla de Contenidos

- 1 Centralidades
- 2 Grados Nodales
- 3 Densidad
- 4 Distancias**
- 5 Centralidad de Cercanía
- 6 Centralidad de Intermediación
- 7 Eigenvector
- 8 Índices de centralidad

Distancia geodésica (1/2)

- Distancia más pequeña entre dos actores.
- “Óptimo” o “Más eficiente”.
- Redes muy densas: camino geodésico es generalmente pequeño.
- Redes conectadas: actores son “accesibles” por todos los actores.
- Redes no conectadas: no podemos calcular todas las distancias geodésicas entre pares de actores.
- Mayor distancia geodésica: *excentricidad*.

Distancia geodésica (2/2)

- Es posible calcular la media (o mediana) de la distancia geodésica.
- Es posible calcular la desviación estándar de la matriz.
- Un concepto similar es *diámetro*: Distancia más larga existente en la red.
- Redundancia de la conexión: puede generarse porque hay muchos caminos geodésicos entre dos actores.

Tabla de Contenidos

- 1 Centralidades
- 2 Grados Nodales
- 3 Densidad
- 4 Distancias
- 5 Centralidad de Cercanía
- 6 Centralidad de Intermediación
- 7 Eigenvector
- 8 Índices de centralidad

Cercanía (1/2):

- Qué tan *cerca* está un actor de todos los demás.
- Un actor está cerca si puede interactuar rápidamente con todos los actores.
- Suelen ser actores muy productivos para comunicar información.
- Medida basada en las geodésicas de los grafos.
- Interpretación de Freeman (1979) como dualidad conceptual (Brandes et al., 2016):
 - Independencia del control de los otros
 - Medida de eficiencia en los accesos

Cercanía (2/2):

- Si geodésicas aumentan, la centralidad de los actores debería decrecer.
- Distancias se calculan de modo inverso para calcular el índice de cercanía (de Sabidussi).
- Sin embargo, el índice solo tiene sentido para un grafo conexo.

$$C_C(n_i) = \left[\sum_{j=1}^g d(n_i, n_j) \right]^{-1} \quad (15)$$

Normalización de centralidad de cercanía:

$$C'_C(n_i) = \frac{g - 1}{\left[\sum_{j=1}^g d(n_i, n_j) \right]} \quad (16)$$

- Rango que va de 0 a 1.

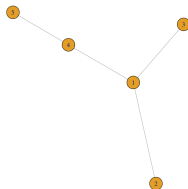
Familias de la Elite Renacentista del siglo XV (Padgett y Ansell, 1993)

MEDICI	56%
RIDOLFI	50%
ALBIZZI	48%
TORNABUON	48%
GUADAGNI	47%
BARBADORI	44%
STROZZI	44%
BISCHERI	40%
CASTELLAN	39%
SALVIATI	39%
ACCIAIUOL	37%
PERUZZI	37%
GINORI	33%
LAMBERTES	33%
PAZZI	29%

¿Cuál es la familia que posee una mayor cercanía?

Se asume que la familia Medici posee una cercanía normalizada del 56 % del total de la red, dado que la suma de las distancias geodésicas con los otros actores es la mayor (y su lejanía, por tanto, es la menor). En otras palabras, éste actor tiene la capacidad del 56 % del total de geodésicas de llegar a los demás actores. Estos actores son capaces de alcanzar a otros en longitudes de caminos más cortos o son más accesibles, teniendo así posiciones favorables en la red.

Cercanía:



Sin normalizar:

$$C_c(n1) = 1/(1 + 1 + 1 + 2) = 0,2$$

Normalizado:

$$C_c(n1) = (5 - 1)/(1 + 1 + 1 + 2) = 0,8$$

Tabla de Contenidos

- 1 Centralidades
- 2 Grados Nodales
- 3 Densidad
- 4 Distancias
- 5 Centralidad de Cercanía
- 6 Centralidad de Intermediación**
- 7 Eigenvector
- 8 Índices de centralidad

Intermediación:

- Interacciones pueden depender de actores que se encuentran en los caminos que están entre dos actores.
- Actores intermediarios controlan en parte de las interacciones (caminos) del grafo.
- Conocidos en la literatura como *gatekeeper*.

$$C_B(n_i) = \sum_{j < k} g_{jk}(n_i) / g_{jk} \quad (17)$$

- Sea g_{jk} número de geodésicas que unen a dos actores.
- Sea $g_{jk}(i)$ número de geodésicas que unen a dos actores en donde actor i se encuentra presente.
- Si todos los caminos tienen la misma probabilidad: $1/g_{jk}$.
- Si un actor distinto i está implicado en la geodésica, la probabilidad será equivalente a $g_{jk}(n_i)/g_{jk}$

Estandarización:

$$C'_B(n_i) = C_B(n_i) / [(g - 1)(g - 2)/2] \quad (18)$$

- Toma valores de 0 a 1
- Permite ser comparado fácilmente con otros índices, redes y relaciones.
- Es posible calcular este índice pese a que el grafo no esté conectado.

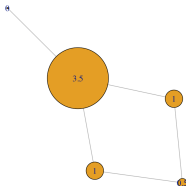
Familias de la Elite Renacentista del siglo XV (Padgett y Ansell, 1993)

MEDICI	52%
GUADAGNI	25%
ALBIZZI	21%
SALVIATI	14%
RIDOLFI	11%
BISCHERI	10%
STROZZI	10%
BARBADORI	9%
TORNABUON	9%
CASTELLAN	5%
PERUZZI	2%
ACCIAIUOL	0%
GINORI	0%
LAMBERTES	0%
PAZZI	0%

¿Cuál es la familia que posee una mayor intermediación?

Se aprecia que la familia Medici tienen una gran relevancia en el traspaso de la información, intermediando un 52 % del total de los posibles vínculos entre dos nodos de esta red.

Intermediación:



Ejemplo, ¿Cuál es el nivel de intermediación del nodo 2?:

$$\begin{aligned}
 C_b(n_2) = & ((1/1)\{e = n_1, t = n_3\} \\
 & + (1/1)\{e = n_1, t = n_4\} \\
 & + (2/2)\{e = n_1, t = n_5\} \\
 & + (1/2)\{e = n_3, t = n_4\} \\
 & + 0\{e = n_3, t = n_5\} \\
 & + 0\{e = n_4, t = n_5\}) = 3,5
 \end{aligned}$$

Tabla de Contenidos

- 1 Centralidades
- 2 Grados Nodales
- 3 Densidad
- 4 Distancias
- 5 Centralidad de Cercanía
- 6 Centralidad de Intermediación
- 7 Eigenvector
- 8 Índices de centralidad

Centralidad de *eigenvector*:

- Relevancia de actores centrales en términos de estructura “global” o “general”, prestando menor atención a los patrones locales.
- El análisis de factores busca “dimensiones” de la distancia entre actores.
- La primera “dimensión” captura los aspectos más “globales” de distancia entre actores.
- Las segundas y demás dimensiones, capturan subestructuras específicas.
- Puntuaciones más altas indican que actores son “más centrales”.
- Puntuaciones más bajas indican que actores son “más periféricos”.

- La centralidad es proporcional a la suma de la centralidad de los vecinos.
- Adquiere valor por su conexión con otros, pero es proporcional a éste.
- Algunas medidas similares son:
 - Katz
 - Poder de Bonacich
 - Page Rank (conocido como “el algoritmo de google”)

Eigenvector:

$$x_i = \lambda^{-1} \sum_j A_{ij} x_j \quad (19)$$

En lenguaje matricial (no direccionado):

$$Ax = \lambda x \quad (20)$$

En lenguaje matricial (direccionado):

$$A'x = \lambda x \quad (21)$$

En donde A' corresponde a la matrix de adyacencia traspuesta.

* Se utiliza el eigenvalue dominante para cálculo del eigenvector (Bonacich, 1972)

Tabla de Contenidos

- 1 Centralidades
- 2 Grados Nodales
- 3 Densidad
- 4 Distancias
- 5 Centralidad de Cercanía
- 6 Centralidad de Intermediación
- 7 Eigenvector
- 8 Índices de centralidad

Usos interesantes pueden ser: ¿Hay nodos que poseen más poder que otros dadas sus conexiones con otros nodos igualmente influyentes? ¿Existe variabilidad en la tenencia de poder? (equilibrio en la estructura) ¿El poder de influencia incide en otras variables?

Las centralidades se han convertido en un campo fértil de desarrollo científico y hoy existe un conjunto importante de posibilidades....

- Actualmente, de acuerdo al proyecto Centi Server (<http://www.centiserver.org/>) existen más de 400 índices!
- En este curso solo revisamos las medidas más clásicas.

