Medidas de Centralidad

Miguel Raggi

ENES Morelia

16 de mayo de 2018

Índice:

1 Introducción

2 Centralidad de Grado

3 Centralidad de cercanía

4 Centralidad de Intermediación
- Definición Formal
- Algoritmo

5 Centralidad de Eigenvector

6 Intermedio: Teoría Espectral
- Matriz de Adyacencia
- Matriz de Difusión
- Gráficas con Pesos

7 Centralidad de Katz

8 PageRank

Índice:

1 Introducción

2 Centralidad de Grado

3 Centralidad de cercanía

4 Centralidad de Intermediación
- Definición Formal
- Algoritmo

5 Centralidad de Eigenvector

6 Intermedio: Teoría Espectral
- Matriz de Adyacencia
- Matriz de Difusión
- Gráficas con Pesos

7 Centralidad de Katz

8 PageRank

Centralidad

- Considera un (di)grafo cualquiera (por ejemplo, amigos de facebook o seguidores de twitter).

Centralidad

- Considera un (di)grafo cualquiera (por ejemplo, amigos de facebook o seguidores de twitter).
- ¿Cómo puedes determinar a los “líderes” o las personas “más influenciales” en base a la estructura de la red?

Twitter

- Supón que tienes una red de usuarios de twitter en una comunidad (digamos, dentro de la ENES).

Twitter

- Supón que tienes una red de usuarios de twitter en una comunidad (digamos, dentro de la ENES).
- Pon una arista de $v \to w$ si w “sigue” a v. Es decir, si v “influencía” a w.

Twitter

- Supón que tienes una red de usuarios de twitter en una comunidad (digamos, dentro de la ENES).
- Pon una arista de $v \to w$ si w "sigue" a v. Es decir, si v "influencía" a w.
- Quizás una compañía quiere determinar qué personas son las más influenciales para hacer publicidad eficientemente.

Twitter

- Supón que tienes una red de usuarios de twitter en una comunidad (digamos, dentro de la ENES).
- Pon una arista de $v \to w$ si w "sigue" a v. Es decir, si v "influencía" a w.
- Quizás una compañía quiere determinar qué personas son las más influenciales para hacer publicidad eficientemente.
- Quizás si se ofrece un producto a los líderes, por medio de *retweets* puede llegar a muchos.

Centralidad

Hay muchas maneras de definir “líderes”:

Centralidad

Hay muchas maneras de definir "líderes":

- Centralidad de grado (degree)

Centralidad

Hay muchas maneras de definir "líderes":

- Centralidad de grado (degree)
- Centralidad de cercanía (closeness)

Centralidad

Hay muchas maneras de definir “líderes”:

- Centralidad de grado (degree)
- Centralidad de cercanía (closeness)
- Centralidad de intermediación (betweenness)

Centralidad

Hay muchas maneras de definir “líderes”:

- Centralidad de grado (degree)
- Centralidad de cercanía (closeness)
- Centralidad de intermediación (betweenness)
- Centrality de eigenvector

Centralidad

Hay muchas maneras de definir “líderes”:

- Centralidad de grado (degree)
- Centralidad de cercanía (closeness)
- Centralidad de intermediación (betweenness)
- Centrality de eigenvector
- Centralidad de Katz

Centralidad

Hay muchas maneras de definir “líderes”:

- Centralidad de grado (degree)
- Centralidad de cercanía (closeness)
- Centralidad de intermediación (betweenness)
- Centrality de eigenvector
- Centralidad de Katz
- PageRank

Centralidad

Hay muchas maneras de definir “líderes”:

- Centralidad de grado (degree)
- Centralidad de cercanía (closeness)
- Centralidad de intermediación (betweenness)
- Centrality de eigenvector
- Centralidad de Katz
- PageRank

Vamos a definir algunas de ellas.

Índice:

1 Introducción

2 Centralidad de Grado

3 Centralidad de cercanía

4 Centralidad de Intermediación
- Definición Formal
- Algoritmo

5 Centralidad de Eigenvector

6 Intermedio: Teoría Espectral
- Matriz de Adyacencia
- Matriz de Difusión
- Gráficas con Pesos

7 Centralidad de Katz

8 PageRank

Centralidad de grado

Definición

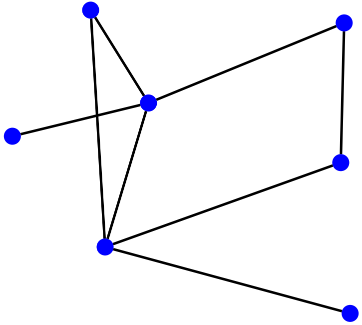
Dada una gráfica, la centralidad de grado de un nodo es simplemente el grado de ese nodo.

Centralidad de grado

Definición

Dada una gráfica, la centralidad de grado de un nodo es simplemente el grado de ese nodo.

Ejemplo: ¿Qué centralidad de grado tiene cada nodo?



Notas

- P: ¿Qué pasa si la gráfica tiene pesos?

Notas

- P: ¿Qué pasa si la gráfica tiene pesos?
- R: La centralidad será la suma de los pesos de sus aristas.

Notas

- P: ¿Qué pasa si la gráfica tiene pesos?
- R: La centralidad será la suma de los pesos de sus aristas.
- P: ¿Y si es dirigida?

Notas

- P: ¿Qué pasa si la gráfica tiene pesos?
- R: La centralidad será la suma de los pesos de sus aristas.
- P: ¿Y si es dirigida?
- R: Pues habrá dos medias, la "ex-centralidad" y la "in-centralidad".

Notas

- P: ¿Qué pasa si la gráfica tiene pesos?
- R: La centralidad será la suma de los pesos de sus aristas.
- P: ¿Y si es dirigida?
- R: Pues habrá dos medias, la “ex-centralidad” y la “in-centralidad”.
- P: ¿Cómo se calcula?

Notas

- P: ¿Qué pasa si la gráfica tiene pesos?
- R: La centralidad será la suma de los pesos de sus aristas.
- P: ¿Y si es dirigida?
- R: Pues habrá dos medias, la “ex-centralidad” y la “in-centralidad”.
- P: ¿Cómo se calcula?
  - Si la gráfica la guardamos como matriz de adyacencia, sumamos el renglón (o columna) correspondiente.

Notas

- P: ¿Qué pasa si la gráfica tiene pesos?
- R: La centralidad será la suma de los pesos de sus aristas.
- P: ¿Y si es dirigida?
- R: Pues habrá dos medias, la "ex-centralidad" y la "in-centralidad".
- P: ¿Cómo se calcula?
  - Si la gráfica la guardamos como matriz de adyacencia, sumamos el renglón (o columna) correspondiente.
  - Si la guardamos como listas de adyacencia, el tamaño de la lista.

Notas

- P: ¿Qué pasa si la gráfica tiene pesos?
- R: La centralidad será la suma de los pesos de sus aristas.
- P: ¿Y si es dirigida?
- R: Pues habrá dos medias, la “ex-centralidad” y la “in-centralidad”.
- P: ¿Cómo se calcula?
  - Si la gráfica la guardamos como matriz de adyacencia, sumamos el renglón (o columna) correspondiente.
  - Si la guardamos como listas de adyacencia, el tamaño de la lista.
- Nota: A veces normalizamos por el máximo grado.

Índice:

1 Introducción

2 Centralidad de Grado

3 Centralidad de cercanía

4 Centralidad de Intermediación
- Definición Formal
- Algoritmo

5 Centralidad de Eigenvector

6 Intermedio: Teoría Espectral
- Matriz de Adyacencia
- Matriz de Difusión
- Gráficas con Pesos

7 Centralidad de Katz

8 PageRank

Centralidad de cercanía

Definición

Dada una gráfica conexa, la centralidad de cercanía *de un nodo es el inverso de la suma de las distancias a los demás nodos. Es decir,*

$$C(x) = \frac{1}{\sum_{y \in V} d(x, y)}$$

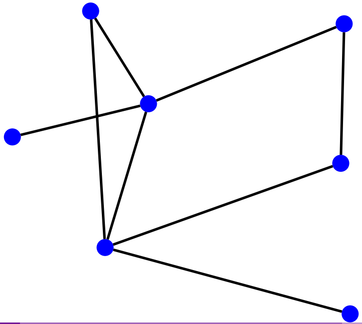
Centralidad de cercanía

Definición

Dada una gráfica conexa, la centralidad de cercanía de un nodo es el inverso de la suma de las distancias a los demás nodos. Es decir,

$$C(x) = \frac{1}{\sum_{y \in V} d(x,y)}$$

Ejemplo: ¿Qué centralidad de cercanía tiene cada nodo?



Notas

- P: ¿Qué pasa si la gráfica tiene pesos?

Notas

- P: ¿Qué pasa si la gráfica tiene pesos?
- R: La distancia será en la gráfica con pesos.

Notas

- P: ¿Qué pasa si la gráfica tiene pesos?
- R: La distancia será en la gráfica con pesos.
- P: ¿Y si es dirigida?

Notas

- P: ¿Qué pasa si la gráfica tiene pesos?
- R: La distancia será en la gráfica con pesos.
- P: ¿Y si es dirigida?
- R: Es más complicado: si es fuertemente conexa, pues está bien, pero si no...

Ejercicios

En una gráfica simple con n vértices,

Ejercicios

En una gráfica simple con n vértices,

- ¿Cuál es la mínima centralidad de cercanía posible y cuándo se da?

Ejercicios

En una gráfica simple con n vértices,

- ¿Cuál es la mínima centralidad de cercanía posible y cuándo se da?
- ¿Cuál es la máxima centralidad de cercanía y cuándo se da?

Ejercicios

En una gráfica simple con n vértices,

- ¿Cuál es la mínima centralidad de cercanía posible y cuándo se da?
- ¿Cuál es la máxima centralidad de cercanía y cuándo se da?
- ¿Cuál es la centralidad de cercanía normalizada?

Ejercicios

En una gráfica simple con n vértices,

- ¿Cuál es la mínima centralidad de cercanía posible y cuándo se da?
- ¿Cuál es la máxima centralidad de cercanía y cuándo se da?
- ¿Cuál es la centralidad de cercanía normalizada?

$$C(x) = \frac{n-1}{\sum_{y \in V} d(x,y)}$$

- ¿Cuál es la centralidad de cercanía en K_n? ¿Y C_n? ¿Y P_n?

Interludio: Floyd-Warshall

- Para lo anterior necesitamos calcular las distancias entre cada pareja de vértices.

Interludio: Floyd-Warshall

- Para lo anterior necesitamos calcular las distancias entre cada pareja de vértices.
- Podríamos para cada vértice correr Dijkstra para cada vértice, lo cual toma tiempo $O(VE\log(E))$.

Interludio: Floyd-Warshall

- Para lo anterior necesitamos calcular las distancias entre cada pareja de vértices.
- Podríamos para cada vértice correr Dijkstra para cada vértice, lo cual toma tiempo $O(VE\log(E))$.
- Hay un algoritmo que lo hace en tiempo $O(V^3)$, lo cual es malo pero es muy sencillo de programar, mucho más que Dijkstra: Floyd-Warshall.

Interludio: Floyd-Warshall

- Para lo anterior necesitamos calcular las distancias entre cada pareja de vértices.
- Podríamos para cada vértice correr Dijkstra para cada vértice, lo cual toma tiempo $O(VE\log(E))$.
- Hay un algoritmo que lo hace en tiempo $O(V^3)$, lo cual es malo pero es muy sencillo de programar, mucho más que Dijkstra: Floyd-Warshall.
- La idea es programación dinámica e ir "relajando" aristas.

Interludio: Floyd-Warshall

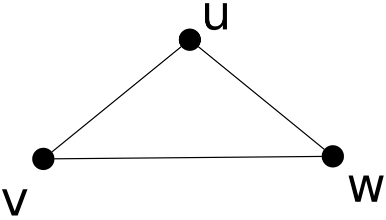
- Creamos una matrix de $V \times V$ llamada D (por "distancia"), que empieza con puros ∞.

Interludio: Floyd-Warshall

- Creamos una matrix de $V \times V$ llamada D (por “distancia”), que empieza con puros ∞.
- $D(v, v) = 0$ y $D(u, v) = w(u, v)$ si $u \to v$ es arista.

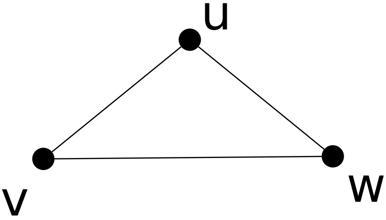
Interludio: Floyd-Warshall

- Creamos una matrix de $V \times V$ llamada D (por “distancia”), que empieza con puros ∞.
- $D(v,v) = 0$ y $D(u,v) = w(u,v)$ si $u \to v$ es arista.
- Para cada tercia (u,v,w), si $D(v,w) > D(v,u) + D(u,w)$, la hacemos menor:



Interludio: Floyd-Warshall

- Creamos una matrix de $V \times V$ llamada D (por "distancia"), que empieza con puros ∞.
- $D(v, v) = 0$ y $D(u, v) = w(u, v)$ si $u \to v$ es arista.
- Para cada tercia (u, v, w), si $D(v, w) > D(v, u) + D(u, w)$, la hacemos menor:



- Regresamos D.

Índice:

1 Introducción

2 Centralidad de Grado

3 Centralidad de cercanía

4 Centralidad de Intermediación
- Definición Formal
- Algoritmo

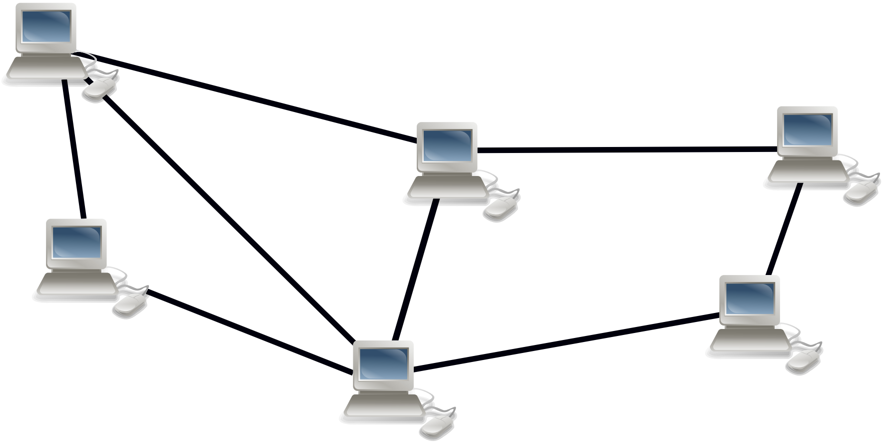
5 Centralidad de Eigenvector

6 Intermedio: Teoría Espectral
- Matriz de Adyacencia
- Matriz de Difusión
- Gráficas con Pesos

7 Centralidad de Katz

8 PageRank

¿Cuál hackear para maximizar la probabilidad de interceptar el mensaje?



Supongamos ciertas cosas...

1. Tú no sabes cuáles son las computadoras de los malos (o *buenos*, según sea el caso).

Supongamos ciertas cosas...

1. Tú no sabes cuáles son las computadoras de los malos (o *buenos*, según sea el caso). Pensemos que cada pareja de computadoras tienen la misma probabilidad.

Supongamos ciertas cosas...

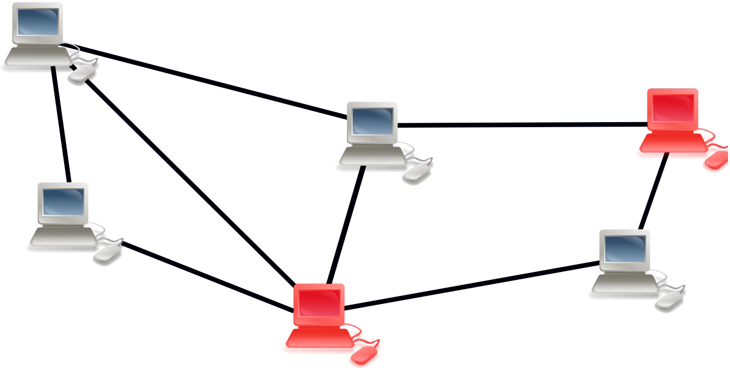
1. Tú no sabes cuáles son las computadoras de los malos (o *buenos*, según sea el caso). Pensemos que cada pareja de computadoras tienen la misma probabilidad.
2. Si el mensaje pasa por la que hackeamos, lo interceptamos.

Supongamos ciertas cosas...

1. Tú no sabes cuáles son las computadoras de los malos (o *buenos*, según sea el caso). Pensemos que cada pareja de computadoras tienen la misma probabilidad.
2. Si el mensaje pasa por la que hackeamos, lo interceptamos.
3. Los mensajes viajan por geodésicas (caminos más cortos).

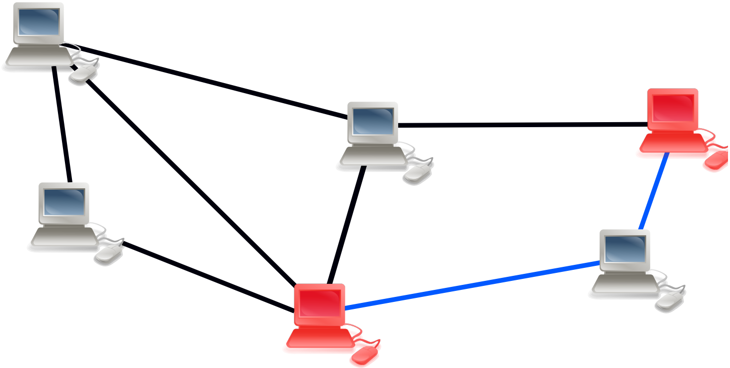
Supongamos ciertas cosas...

1. Tú no sabes cuáles son las computadoras de los malos (o *buenos*, según sea el caso). Pensemos que cada pareja de computadoras tienen la misma probabilidad.
2. Si el mensaje pasa por la que hackeamos, lo interceptamos.
3. Los mensajes viajan por geodésicas (caminos más cortos).



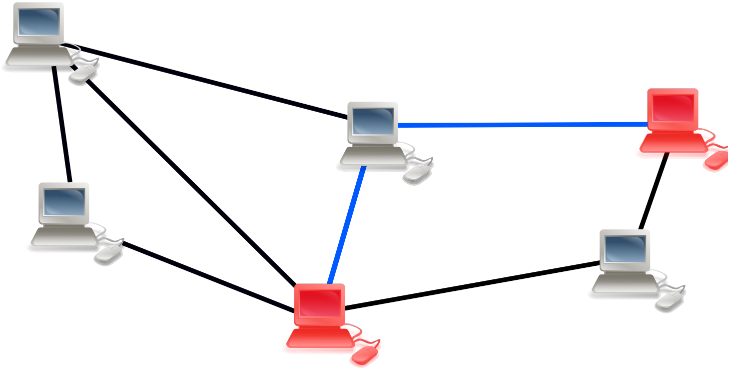
Supongamos ciertas cosas...

1. Tú no sabes cuáles son las computadoras de los malos (o *buenos*, según sea el caso). Pensemos que cada pareja de computadoras tienen la misma probabilidad.
2. Si el mensaje pasa por la que hackeamos, lo interceptamos.
3. Los mensajes viajan por geodésicas (caminos más cortos).



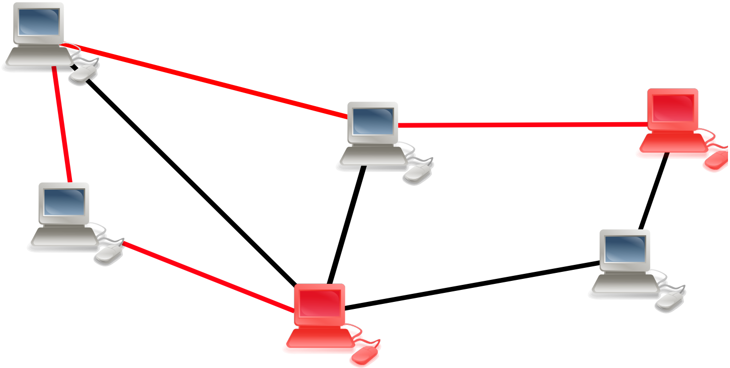
Supongamos ciertas cosas...

1. Tú no sabes cuáles son las computadoras de los malos (o *buenos*, según sea el caso). Pensemos que cada pareja de computadoras tienen la misma probabilidad.
2. Si el mensaje pasa por la que hackeamos, lo interceptamos.
3. Los mensajes viajan por geodésicas (caminos más cortos).



Supongamos ciertas cosas...

1. Tú no sabes cuáles son las computadoras de los malos (o *buenos*, según sea el caso). Pensemos que cada pareja de computadoras tienen la misma probabilidad.
2. Si el mensaje pasa por la que hackeamos, lo interceptamos.
3. Los mensajes viajan por geodésicas (caminos más cortos).



Definición Intuitiva

Definición

Dada una gráfica (simple, conexa), la centralidad de intermediación (betweenness centrality) de un vértice es la probabilidad de que una geodésica arbitraria pase por el vértice.

Definición Intuitiva

Definición

Dada una gráfica (simple, conexa), la centralidad de intermediación (betweenness centrality) de un vértice es la probabilidad de que una geodésica arbitraria pase por el vértice.

Nota: La distribución de las geodésicas se forma tomando un par de vértices aleatorios y luego una geodésica aleatoria.

Definición Formal

Definición

Sea G una gráfica y v un vértice.

Definición Formal

Definición

Sea G una gráfica y v un vértice. Para cada par de vértices $s, t \in V(G)$, sea

$$\sigma(s,t) = \#\ \textit{geodésicas de } s \textit{ a } t$$

Definición Formal

Definición

Sea G una gráfica y v un vértice. Para cada par de vértices $s, t \in V(G)$, sea

$$\sigma(s,t) = \#\ \textit{geodésicas de } s \textit{ a } t$$

y sea

$$\sigma_v(s,t) = \#\ \textit{geodésicas de } s \textit{ a } t \textit{ que pasan por } v$$

Definición Formal

Definición

Sea G una gráfica y v un vértice. Para cada par de vértices $s, t \in V(G)$, sea

$$\sigma(s,t) = \#\ \textit{geodésicas de } s \textit{ a } t$$

y sea

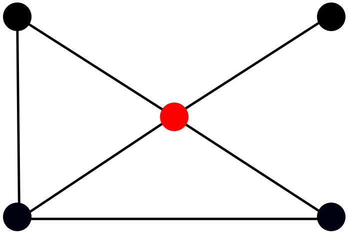
$$\sigma_v(s,t) = \#\ \textit{geodésicas de } s \textit{ a } t \textit{ que pasan por } v$$

Definimos la centralidad de intermediación (normalizada) de un vértice v como:

$$bc(v) = \frac{1}{\binom{n-1}{2}} \sum_{\{s,t\} \in \binom{V(G) \setminus \{v\}}{2}} \frac{\sigma_v(s,t)}{\sigma(s,t)}$$

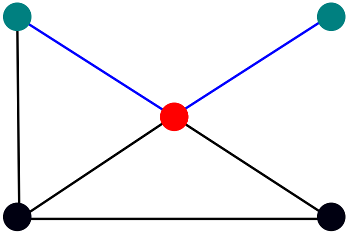
Nota: A veces también se define sin quitar a v de $V(G)$, y que $(s,t) \in V(G) \times V(G)$, etc.

Ejemplo



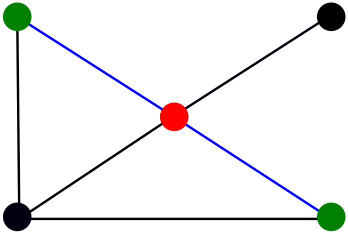
bc(●) =

Ejemplo



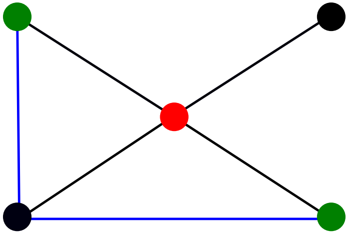
bc(●) = 1

Ejemplo



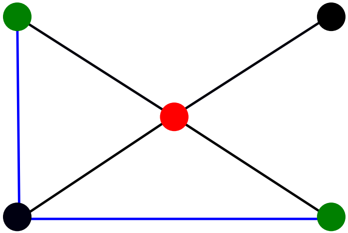
bc(●) = 1

Ejemplo



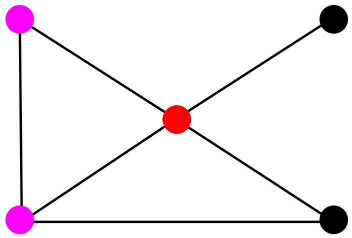
bc(●) = 1

Ejemplo



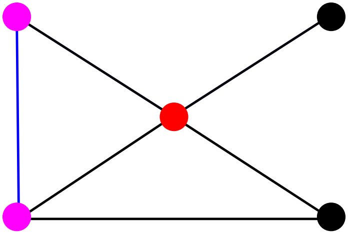
bc(●) = 1+0.5

Ejemplo



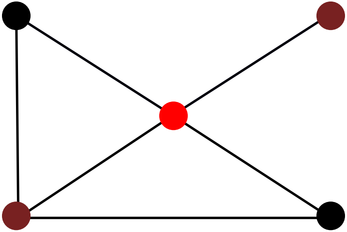
bc(●) = 1+0.5

Ejemplo



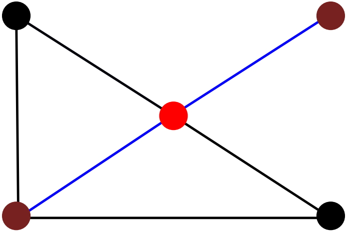
bc(●) = 1+0.5+0

Ejemplo



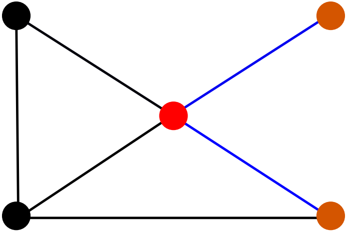
bc(●) = 1+0.5+0

Ejemplo



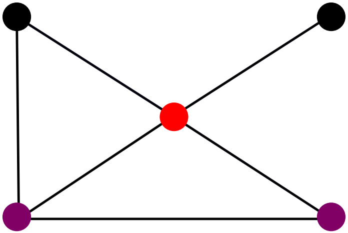
bc(●) = 1+0.5+0+1

Ejemplo



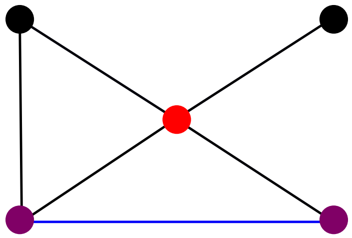
bc(●) = 1+0.5+0+1+1

Ejemplo



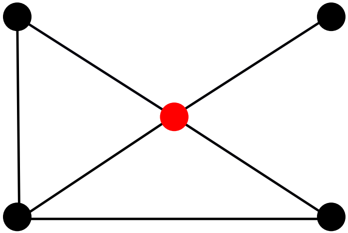
$\mathrm{bc}(\bullet) = 1+0.5+0+1+1$

Ejemplo



bc(●) = 1+0.5+0+1+1+0

Ejemplo



$$\mathrm{bc}(\bullet) = [1 + 0.5 + 0 + 1 + 1 + 0]/\binom{4}{2} = 0.58$$

Máxima y Mínima posible BC

- P: ¿Cuál es la máxima posible centralidad de intermediación? ¿Cuándo ocurre?

Máxima y Mínima posible BC

- P: ¿Cuál es la máxima posible centralidad de intermediación? ¿Cuándo ocurre?
- R: Máxima: 1, sólo en la estrella $K_{1,n}$.

Máxima y Mínima posible BC

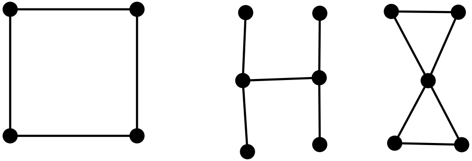
- P: ¿Cuál es la máxima posible centralidad de intermediación? ¿Cuándo ocurre?
- R: Máxima: 1, sólo en la estrella $K_{1,n}$.
- P: ¿Cuál es la mínima posible centralidad de intermediación? ¿Cuándo ocurre?

Máxima y Mínima posible BC

- P: ¿Cuál es la máxima posible centralidad de intermediación? ¿Cuándo ocurre?
- R: Máxima: 1, sólo en la estrella $K_{1,n}$.
- P: ¿Cuál es la mínima posible centralidad de intermediación? ¿Cuándo ocurre?
- R: La mínima es 0, y ocurre siempre que todos los vecinos de un vértice están conectados entre sí.

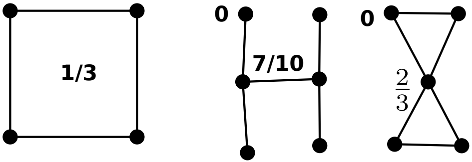
Ejemplos

Calcula la betweenness centrality en las siguientes:



Ejemplos

Calcula la betweenness centrality en las siguientes:



Preguntas

- P: ¿Cuál es la máxima y mínima posible betweenness centrality? ¿Cuándo ocurren?

Preguntas

- P: ¿Cuál es la máxima y mínima posible betweenness centrality? ¿Cuándo ocurren?
- R: Máxima: 1, sólo en la estrella *.

Preguntas

- P: ¿Cuál es la máxima y mínima posible betweenness centrality? ¿Cuándo ocurren?
- R: Máxima: 1, sólo en la estrella *.
- R: La mínima es 0, y ocurre siempre que todos los vecinos de un vértice están conectados entre sí.

Naturalidad vs Elegancia

- La definición anterior de betweenness centrality es la que a mí se me hace más natural.

Naturalidad vs Elegancia

- La definición anterior de betweenness centrality es la que a mí se me hace más natural.
- Sin embargo, el algoritmo queda bastante más sencillo si la definimos de una manera un poquito diferente.

Naturalidad vs Elegancia

- La definición anterior de betweenness centrality es la que a mí se me hace más natural.
- Sin embargo, el algoritmo queda bastante más sencillo si la definimos de una manera un poquito diferente.
- Entonces vamos a volver a definir betweenness centrality, vamos a ver cómo se comparan las dos definiciones y luego explicaré el algoritmo para la nueva definición, y cómo cambiarlo para la otra definición.

D. Formal

Definición

Sea G una gráfica.

D. Formal

Definición

Sea G una gráfica. Para cada par de vértices $s, t \in V(G)$, sea

$$\sigma(s,t) = \# \text{ geodésicas de } s \text{ a } t$$

D. Formal

Definición

Sea G una gráfica. Para cada par de vértices $s, t \in V(G)$, sea

$$\sigma(s,t) = \#\ \textit{geodésicas de } s \textit{ a } t$$

y sea

$$\sigma_v(s,t) = \#\ \textit{geodésicas de } s \textit{ a } t \textit{ que pasan por } v$$

D. Formal

Definición

Sea G una gráfica. Para cada par de vértices $s, t \in V(G)$, sea

$$\sigma(s,t) = \#\ \textit{geodésicas de } s \textit{ a } t$$

y sea

$$\sigma_v(s,t) = \#\ \textit{geodésicas de } s \textit{ a } t \textit{ que pasan por } v$$

Definimos la centralidad de intermediación (normalizada) de un vértice v como:

$$bc(v) = \frac{1}{n^2} \sum_{s,t \in V(G)} \frac{\sigma_v(s,t)}{\sigma(s,t)}$$

D. Formal

Definición

Sea G una gráfica. Para cada par de vértices $s, t \in V(G)$, sea

$$\sigma(s,t) = \#\ \textit{geodésicas de } s \textit{ a } t$$

y sea

$$\sigma_v(s,t) = \#\ \textit{geodésicas de } s \textit{ a } t \textit{ que pasan por } v$$

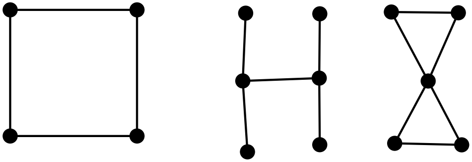
Definimos la centralidad de intermediación (normalizada) de un vértice v como:

$$bc(v) = \frac{1}{n^2} \sum_{s,t \in V(G)} \frac{\sigma_v(s,t)}{\sigma(s,t)}$$

Note: s, t y v no son necesariamente distintos, así que cada pareja se cuenta dos veces cuando $s \neq t$.

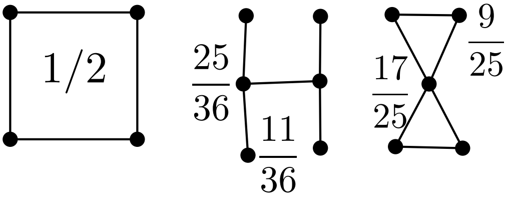
Ejemplos

Calcula bc:



Ejemplos

Calcula bc:



Preguntas

- P: ¿Máxima y mínima bc?

Preguntas

- P: ¿Máxima y mínima bc?
- R: Ahora la máxima es $\frac{n^2-n+1}{n^2}$.

Preguntas

- P: ¿Máxima y mínima bc?
- R: Ahora la máxima es $\frac{n^2-n+1}{n^2}$.
- R: La mínima ahora es $\frac{2n-1}{n^2}$.

Preguntas

- P: ¿Máxima y mínima bc?
- R: Ahora la máxima es $\frac{n^2-n+1}{n^2}$.
- R: La mínima ahora es $\frac{2n-1}{n^2}$.
- ¿Cómo se comparan la definición anterior y esta? ¿Cómo se calcula una a partir de la otra?

Algoritmo

- ¿Cómo calculamos la betweenness centrality de un vértices en una gráfica? (o, mejor aún, calcularlo para *todos* los vértices)

Algoritmo

- ¿Cómo calculamos la betweenness centrality de un vértices en una gráfica? (o, mejor aún, calcularlo para *todos* los vértices)
- Podríamos hacerlo por "fuerza bruta": calcular todas las posibles geodésicas entre pares de vértices utilizando los algoritmos estándar que vimos.

Algoritmo

- ¿Cómo calculamos la betweenness centrality de un vértices en una gráfica? (o, mejor aún, calcularlo para *todos* los vértices)
- Podríamos hacerlo por “fuerza bruta”: calcular todas las posibles geodésicas entre pares de vértices utilizando los algoritmos estándar que vimos.
- Claro, esto es ligeramente más complicado si hay más de un camino más corto.

Algoritmo

- ¿Cómo calculamos la betweenness centrality de un vértices en una gráfica? (o, mejor aún, calcularlo para *todos* los vértices)
- Podríamos hacerlo por "fuerza bruta": calcular todas las posibles geodésicas entre pares de vértices utilizando los algoritmos estándar que vimos.
- Claro, esto es ligeramente más complicado si hay más de un camino más corto.
- Búsqueda a lo ancho toma tiempo $O(n+m)$, si hay n vértices y m aristas.

Algoritmo

- ¿Cómo calculamos la betweenness centrality de un vértices en una gráfica? (o, mejor aún, calcularlo para *todos* los vértices)
- Podríamos hacerlo por "fuerza bruta": calcular todas las posibles geodésicas entre pares de vértices utilizando los algoritmos estándar que vimos.
- Claro, esto es ligeramente más complicado si hay más de un camino más corto.
- Búsqueda a lo ancho toma tiempo $O(n+m)$, si hay n vértices y m aristas.
- Hay $\binom{n-1}{2} \in O(n^2)$ pares de vértices.

Algoritmo

- ¿Cómo calculamos la betweenness centrality de un vértices en una gráfica? (o, mejor aún, calcularlo para *todos* los vértices)
- Podríamos hacerlo por "fuerza bruta": calcular todas las posibles geodésicas entre pares de vértices utilizando los algoritmos estándar que vimos.
- Claro, esto es ligeramente más complicado si hay más de un camino más corto.
- Búsqueda a lo ancho toma tiempo $O(n+m)$, si hay n vértices y m aristas.
- Hay $\binom{n-1}{2} \in O(n^2)$ pares de vértices.
- Tendríamos que checar, para cada geodésica, si v está en la geodésica. El tamaño promedio de una geodésica es $\log(n)$ y el peor caso es $n/2$.

Algoritmo

- ¿Cómo calculamos la betweenness centrality de un vértices en una gráfica? (o, mejor aún, calcularlo para *todos* los vértices)
- Podríamos hacerlo por "fuerza bruta": calcular todas las posibles geodésicas entre pares de vértices utilizando los algoritmos estándar que vimos.
- Claro, esto es ligeramente más complicado si hay más de un camino más corto.
- Búsqueda a lo ancho toma tiempo $O(n+m)$, si hay n vértices y m aristas.
- Hay $\binom{n-1}{2} \in O(n^2)$ pares de vértices.
- Tendríamos que checar, para cada geodésica, si v está en la geodésica. El tamaño promedio de una geodésica es $\log(n)$ y el peor caso es $n/2$.
- Así que en una gráfica típica, tomaría tiempo $O((n^3+n^2m)\log(n))$.

Algoritmo

- ¿Cómo calculamos la betweenness centrality de un vértices en una gráfica? (o, mejor aún, calcularlo para *todos* los vértices)
- Podríamos hacerlo por “fuerza bruta”: calcular todas las posibles geodésicas entre pares de vértices utilizando los algoritmos estándar que vimos.
- Claro, esto es ligeramente más complicado si hay más de un camino más corto.
- Búsqueda a lo ancho toma tiempo $O(n+m)$, si hay n vértices y m aristas.
- Hay $\binom{n-1}{2} \in O(n^2)$ pares de vértices.
- Tendríamos que checar, para cada geodésica, si v está en la geodésica. El tamaño promedio de una geodésica es $\log(n)$ y el peor caso es $n/2$.
- Así que en una gráfica típica, tomaría tiempo $O((n^3 + n^2 m)\log(n))$.
- Terrible!

Podemos mejorar?

- Cómo podríamos mejorar? Qué estamos haciendo “de más”?

Podemos mejorar?

- Cómo podríamos mejorar? Qué estamos haciendo “de más”?
- La idea es utilizar el hecho de que muchos caminos más cortos utilizan las mismas aristas.

Podemos mejorar?

- Cómo podríamos mejorar? Qué estamos haciendo "de más"?
- La idea es utilizar el hecho de que muchos caminos más cortos utilizan las mismas aristas.
- El siguiente algoritmo corre en tiempo $O(n(n+m))$.

Podemos mejorar?

- Cómo podríamos mejorar? Qué estamos haciendo “de más”?
- La idea es utilizar el hecho de que muchos caminos más cortos utilizan las mismas aristas.
- El siguiente algoritmo corre en tiempo $O(n(n+m))$.
- Empezaremos en un árbol, para entenderlo mejor, y luego generalizamos.

Caso: Árbol

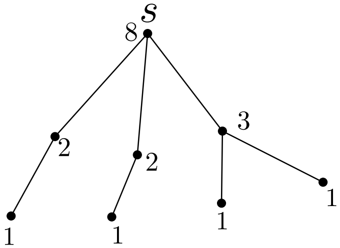
- Escoge un nodo raíz s.

Caso: Árbol

- Escoge un nodo raíz s.
- Cuenta el número de trayectorias que comienzan en s y pasan por un nodo dado.

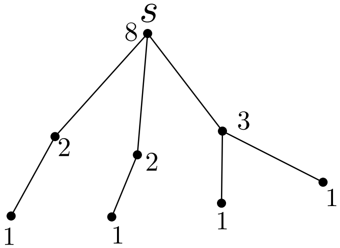
Caso: Árbol

- Escoge un nodo raíz s.
- Cuenta el número de trayectorias que comienzan en s y pasan por un nodo dado.
- En un árbol, es bien fácil: El número de trayectorias es el número de vértices “abajo” del nodo dado.



Caso: Árbol

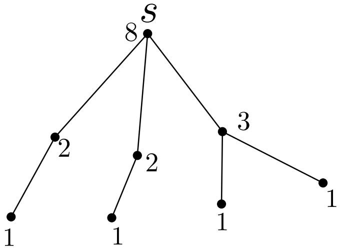
- Escoge un nodo raíz s.
- Cuenta el número de trayectorias que comienzan en s y pasan por un nodo dado.
- En un árbol, es bien fácil: El número de trayectorias es el número de vértices "abajo" del nodo dado.



- Empieza desde abajo y trabaja hacia arriba, sumando el peso de los "hijo" más uno.

Caso: Árbol

- Escoge un nodo raíz s.
- Cuenta el número de trayectorias que comienzan en s y pasan por un nodo dado.
- En un árbol, es bien fácil: El número de trayectorias es el número de vértices "abajo" del nodo dado.



- Empieza desde abajo y trabaja hacia arriba, sumando el peso de los "hijo" más uno.
- Después, haz lo mismo para cada posible "raíz" y suma: sólo hay una geodésica para cada par.

Caso general

- En general, podría haber más de una geodésica por pareja.

Caso general

- En general, podría haber más de una geodésica por pareja.
- Cómo detectamos el número de geodésicas que van de s a algún vértices u que pasan por v?

Caso general

- En general, podría haber más de una geodésica por pareja.
- Cómo detectamos el número de geodésicas que van de s a algún vértices u que pasan por v?
- Pero podemos calcular, en una sóla búsqueda a lo ancho, el número total de geodésicas de s a u, y el número total de geodésicas de s a v y de v a u;

Caso general

- En general, podría haber más de una geodésica por pareja.
- Cómo detectamos el número de geodésicas que van de s a algún vértices u que pasan por v?
- Pero podemos calcular, en una sóla búsqueda a lo ancho, el número total de geodésicas de s a u, y el número total de geodésicas de s a v y de v a u;
- En una sóla búsqueda a lo ancho (para cada s) podemos obtener toooda la información necesaria: simplemente hay que pegar las piezas del rompecabezas. déjalos pensar y absorber un rato, está difícil

Algoritmo

- Para cada nodo s, haz lo siguiente:

Algoritmo

- Para cada nodo s, haz lo siguiente:
- Para cada vértice v, vamos a guardar 3 números:

Algoritmo

- Para cada nodo s, haz lo siguiente:
- Para cada vértice v, vamos a guardar 3 números:
  - d_v, que representará la distancia de s a v.

Algoritmo

- Para cada nodo s, haz lo siguiente:
- Para cada vértice v, vamos a guardar 3 números:
  - d_v, que representará la distancia de s a v.
  - w_v, que representará el número de geodésicas que comienzan en s y terminan en v

Algoritmo

- Para cada nodo s, haz lo siguiente:
- Para cada vértice v, vamos a guardar 3 números:
  - d_v, que representará la distancia de s a v.
  - w_v, que representará el número de geodésicas que comienzan en s y terminan en v
  - x_v, que representará la suma acumulada para la betweenness centrality que las geodésicas que comienzan en s contribuyen.

Algoritmo

- Para cada nodo s, haz lo siguiente:
- Para cada vértice v, vamos a guardar 3 números:
  - d_v, que representará la distancia de s a v.
  - w_v, que representará el número de geodésicas que comienzan en s y terminan en v
  - x_v, que representará la suma acumulada para la betweenness centrality que las geodésicas que comienzan en s contribuyen.
- Paso W: Usa búsqueda a lo ancho comenzando en s para calcular las w's

Algoritmo

- Para cada nodo s, haz lo siguiente:
- Para cada vértice v, vamos a guardar 3 números:
  - d_v, que representará la distancia de s a v.
  - w_v, que representará el número de geodésicas que comienzan en s y terminan en v
  - x_v, que representará la suma acumulada para la betweenness centrality que las geodésicas que comienzan en s contribuyen.
- Paso W: Usa búsqueda a lo ancho comenzando en s para calcular las w's
- Paso X: Ahora pasa por la gráfica en reversa para calcular los x's, utilizando los pesos calculados.

Paso W

1. Pon $d_s = 0$ y $w_s = 1$.

Paso W

1. Pon $d_s = 0$ y $w_s = 1$.
2. Para cada vértice v a distancia $d \in \{0, 1, 2, ...\}$ (en orden) de s itera sobre sus vecinos u:

Paso W

1. Pon $d_s = 0$ y $w_s = 1$.
2. Para cada vértice v a distancia $d \in \{0, 1, 2, ...\}$ (en orden) de s itera sobre sus vecinos u:
   - Si a u no le hemos asignado una distancia aún, ponle distancia $d + 1$ y peso w_v (el del padre).

Paso W

1. Pon $d_s = 0$ y $w_s = 1$.
2. Para cada vértice v a distancia $d \in \{0, 1, 2, ...\}$ (en orden) de s itera sobre sus vecinos u:
   - Si a u no le hemos asignado una distancia aún, ponle distancia $d + 1$ y peso w_v (el del padre).
   - Si a u ya le asignamos la distancia $d + 1$, entonces aumenta el peso del vértice por w_v.

Paso W

1. Pon $d_s = 0$ y $w_s = 1$.
2. Para cada vértice v a distancia $d \in \{0, 1, 2, ...\}$ (en orden) de s itera sobre sus vecinos u:
   - Si a u no le hemos asignado una distancia aún, ponle distancia $d + 1$ y peso w_v (el del padre).
   - Si a u ya le asignamos la distancia $d + 1$, entonces aumenta el peso del vértice por w_v.
   - Si u tiene distancia menor, no hagas nada.

Step X

Mientras recorremos la gráfica en el paso anterior, podemos llevar una lista de todos los vértices en orden de distancia a s (es decir, ponemos primero a s, luego todos los que están a distancia 1, luego los que están a distancia 2, etc.).

Step X

Mientras recorremos la gráfica en el paso anterior, podemos llevar una lista de todos los vértices en orden de distancia a s (es decir, ponemos primero a s, luego todos los que están a distancia 1, luego los que están a distancia 2, etc.). Vamos a recorrer esa lista en orden inverso.

Step X

Mientras recorremos la gráfica en el paso anterior, podemos llevar una lista de todos los vértices en orden de distancia a s (es decir, ponemos primero a s, luego todos los que están a distancia 1, luego los que están a distancia 2, etc.). Vamos a recorrer esa lista en orden inverso.

- A los últimos vértices les ponemos $x_v = 1$ (pues no tienen nada abajo, así que cualquier geodésica, o termina en ellos, o no los usa para nada)

Step X

Mientras recorremos la gráfica en el paso anterior, podemos llevar una lista de todos los vértices en orden de distancia a s (es decir, ponemos primero a s, luego todos los que están a distancia 1, luego los que están a distancia 2, etc.). Vamos a recorrer esa lista en orden inverso.

- A los últimos vértices les ponemos $x_v = 1$ (pues no tienen nada abajo, así que cualquier geodésica, o termina en ellos, o no los usa para nada)
- A cada vértice v le asignamos un score de $x_v = 1 + \sum_u x_u w_v / w_u$, donde sumamos sobre los hijos u de v.

Step X

Mientras recorremos la gráfica en el paso anterior, podemos llevar una lista de todos los vértices en orden de distancia a s (es decir, ponemos primero a s, luego todos los que están a distancia 1, luego los que están a distancia 2, etc.). Vamos a recorrer esa lista en orden inverso.

- A los últimos vértices les ponemos $x_v = 1$ (pues no tienen nada abajo, así que cualquier geodésica, o termina en ellos, o no los usa para nada)
- A cada vértice v le asignamos un score de $x_v = 1 + \sum_u x_u w_v / w_u$, donde sumamos sobre los hijos u de v.
- Repite.

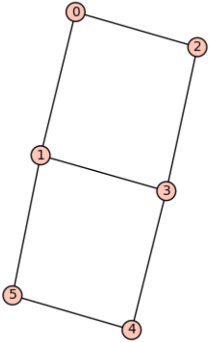
Step X

Mientras recorremos la gráfica en el paso anterior, podemos llevar una lista de todos los vértices en orden de distancia a s (es decir, ponemos primero a s, luego todos los que están a distancia 1, luego los que están a distancia 2, etc.). Vamos a recorrer esa lista en orden inverso.

- A los últimos vértices les ponemos $x_v = 1$ (pues no tienen nada abajo, así que cualquier geodésica, o termina en ellos, o no los usa para nada)
- A cada vértice v le asignamos un score de $x_v = 1 + \sum_u x_u w_v / w_u$, donde sumamos sobre los hijos u de v.
- Repite.

Tenemos que hacer el paso W y el paso X para cada nodo s, añadir todo y normalizar al final.

Ejemplo



$$W = [1, 1, 1, 2, 3, 1]$$

$$X = [6, 19/6, 11/6, 5/3, 1, 4/3]$$

$$bc = [19/54, 53/108, 19/54, 53/108, 19/54, 19/54]$$

```python
from collections import deque
# Los vertices los llamamos de 0 a n-1.
def betweenness_centrality_fea(G):
    n = G.num_verts()
    B = [0]*n
    for s in range(n): # para cada vertice s
        orderedvertices = []
        D = [-1]*n # -1 significa "todavia no se su distancia a s
        W = [0]*n
        D[s] = 0
        W[s] = 1
        X = [1]*n
        d = 0
        frontier = deque([s])  #estamos haciendo busqueda a lo ancho, asi que frontera sera una cola
        orderedvertices = [s]
```

betweennesscentrality.py

```
    #Paso X
    while (frontier):
      lastnode = frontier.popleft()
      d = D[lastnode]
      for v in G.neighbor_iterator(lastnode):
      # Caso 1: Si nunca hemos visto el nodo
        if (D[v] == -1):
          D[v] =  d+1
          W[v] = W[lastnode]
          # Si no hemos visto el vertice antes, aniadelo a la frontera
          frontier.append(v)
          orderedvertices.append(v)
        elif (D[v] == D[lastnode]+1):
          W[v] += W[lastnode]
```

betweennesscentrality.py

```
    # Paso W
    for v in reversed(orderedvertices):
      for u in G.neighbors(v):
        if (D[u] != D[v]+1):
          continue
        X[v] += X[u]*W[v]/W[u]
    for i in range(n):
      B[i] += X[i]
  for i in range(n):
    B[i] /= n^2
  return B
```

betweennesscentrality.py

Índice:

1 Introducción

2 Centralidad de Grado

3 Centralidad de cercanía

4 Centralidad de Intermediación
- Definición Formal
- Algoritmo

5 Centralidad de Eigenvector

6 Intermedio: Teoría Espectral
- Matriz de Adyacencia
- Matriz de Difusión
- Gráficas con Pesos

7 Centralidad de Katz

8 PageRank

Ecuaciones

- Vamos ahora a pensar otra manera de definir la importancia de un vértice.

Ecuaciones

- Vamos ahora a pensar otra manera de definir la importancia de un vértice.
- Idea: Mi "importancia" depende de la "importancia" de mis vecinos.

Ecuaciones

- Vamos ahora a pensar otra manera de definir la importancia de un vértice.
- Idea: Mi “importancia” depende de la “importancia” de mis vecinos.
- Específicamente, podría definir que mi importancia es proporcional a la importancia de mis vecinos.

Ecuaciones

- Vamos ahora a pensar otra manera de definir la importancia de un vértice.
- Idea: Mi "importancia" depende de la "importancia" de mis vecinos.
- Específicamente, podría definir que mi importancia es proporcional a la importancia de mis vecinos.
- Pero a su vez, su importancia depende entonces de mí, pero también de sus demás vecinos.

Ecuaciones

- Sea G una gráfica.

Ecuaciones

- Sea G una gráfica.
- ¿Cómo se verían las ecuaciones que dicen que la importancia x_v de cada vértice es proporcional a la suma de sus vecinos?

Ecuaciones

- Sea G una gráfica.
- ¿Cómo se verían las ecuaciones que dicen que la importancia x_v de cada vértice es proporcional a la suma de sus vecinos?

$$x_v = \lambda \left(\sum_{u \in N(v)} x_u \right)$$

Ecuaciones

- Sea G una gráfica.
- ¿Cómo se verían las ecuaciones que dicen que la importancia x_v de cada vértice es proporcional a la suma de sus vecinos?

$$x_v = \lambda \left(\sum_{u \in N(v)} x_u \right)$$

- Ahora, pongámoslo en forma vectorial.

Ecuaciones

- Sea G una gráfica.
- ¿Cómo se verían las ecuaciones que dicen que la importancia x_v de cada vértice es proporcional a la suma de sus vecinos?

$$x_v = \lambda \left(\sum_{u \in N(v)} x_u \right)$$

- Ahora, pongámoslo en forma vectorial.
- Si x es el vector de importancia y A es la matriz de adyacencia de G,

Ecuaciones

- Sea G una gráfica.
- ¿Cómo se verían las ecuaciones que dicen que la importancia x_v de cada vértice es proporcional a la suma de sus vecinos?

$$x_v = \lambda \left(\sum_{u \in N(v)} x_u \right)$$

- Ahora, pongámoslo en forma vectorial.
- Si x es el vector de importancia y A es la matriz de adyacencia de G,

$$x = \lambda A x \implies A x = \frac{1}{\lambda} x$$

Ecuaciones

- Sea G una gráfica.
- ¿Cómo se verían las ecuaciones que dicen que la importancia x_v de cada vértice es proporcional a la suma de sus vecinos?

$$x_v = \lambda \left(\sum_{u \in N(v)} x_u \right)$$

- Ahora, pongámoslo en forma vectorial.
- Si x es el vector de importancia y A es la matriz de adyacencia de G,

$$x = \lambda A x \implies A x = \frac{1}{\lambda} x$$

- En otras palabras, $1/\lambda$ es un eigenvalor de la matriz de adyacencia, y x es un eigenvector.

Ecuaciones

- Sea G una gráfica.
- ¿Cómo se verían las ecuaciones que dicen que la importancia x_v de cada vértice es proporcional a la suma de sus vecinos?

$$x_v = \lambda \left(\sum_{u \in N(v)} x_u \right)$$

- Ahora, pongámoslo en forma vectorial.
- Si x es el vector de importancia y A es la matriz de adyacencia de G,

$$x = \lambda A x \implies Ax = \frac{1}{\lambda} x$$

- En otras palabras, $1/\lambda$ es un eigenvalor de la matriz de adyacencia, y x es un eigenvector.

¿Cuál eigenvector debo escoger?

- Queremos que todas las centralidades sean positivas.

¿Cuál eigenvector debo escoger?

- Queremos que todas las centralidades sean positivas.
- Hay un teorema de álgebra lineal, el teorema de Perrón-Frobenius, que dice que si tomas el máximo eigenvalor, entonces el eigenvector correspondiente tendrá todas las componentes positivas!

¿Cuál eigenvector debo escoger?

- Queremos que todas las centralidades sean positivas.
- Hay un teorema de álgebra lineal, el teorema de Perrón-Frobenius, que dice que si tomas el máximo eigenvalor, entonces el eigenvector correspondiente tendrá todas las componentes positivas!
- El teorema está feo, y no es necesario enunciarlo ni demostrarlo para entender la centralidad.

¿Cuál eigenvector debo escoger?

- Queremos que todas las centralidades sean positivas.
- Hay un teorema de álgebra lineal, el teorema de Perrón-Frobenius, que dice que si tomas el máximo eigenvalor, entonces el eigenvector correspondiente tendrá todas las componentes positivas!
- El teorema está feo, y no es necesario enunciarlo ni demostrarlo para entender la centralidad.
- A grandes rasgos: Si A es una matriz en donde todas sus entradas son positivas:

¿Cuál eigenvector debo escoger?

- Queremos que todas las centralidades sean positivas.
- Hay un teorema de álgebra lineal, el teorema de Perrón-Frobenius, que dice que si tomas el máximo eigenvalor, entonces el eigenvector correspondiente tendrá todas las componentes positivas!
- El teorema está feo, y no es necesario enunciarlo ni demostrarlo para entender la centralidad.
- A grandes rasgos: Si A es una matriz en donde todas sus entradas son positivas:
  - Hay un eigenvalor real que es más grande (en norma) que todos los demás eigenvalores.

¿Cuál eigenvector debo escoger?

- Queremos que todas las centralidades sean positivas.
- Hay un teorema de álgebra lineal, el teorema de Perrón-Frobenius, que dice que si tomas el máximo eigenvalor, entonces el eigenvector correspondiente tendrá todas las componentes positivas!
- El teorema está feo, y no es necesario enunciarlo ni demostrarlo para entender la centralidad.
- A grandes rasgos: Si A es una matriz en donde todas sus entradas son positivas:
  - Hay un eigenvalor real que es más grande (en norma) que todos los demás eigenvalores.
  - Para este eigenvalor, el eigenespacio asociado es de dimensión 1.

¿Cuál eigenvector debo escoger?

- Queremos que todas las centralidades sean positivas.
- Hay un teorema de álgebra lineal, el teorema de Perrón-Frobenius, que dice que si tomas el máximo eigenvalor, entonces el eigenvector correspondiente tendrá todas las componentes positivas!
- El teorema está feo, y no es necesario enunciarlo ni demostrarlo para entender la centralidad.
- A grandes rasgos: Si A es una matriz en donde todas sus entradas son positivas:
    - Hay un eigenvalor real que es más grande (en norma) que todos los demás eigenvalores.
    - Para este eigenvalor, el eigenespacio asociado es de dimensión 1.
    - Además, tiene un eigenvector en el "primer cuadrante"; es decir, hay un eigenvector para el cual todas sus entradas son positivas.

¿Cuál eigenvector debo escoger?

- Queremos que todas las centralidades sean positivas.
- Hay un teorema de álgebra lineal, el teorema de Perrón-Frobenius, que dice que si tomas el máximo eigenvalor, entonces el eigenvector correspondiente tendrá todas las componentes positivas!
- El teorema está feo, y no es necesario enunciarlo ni demostrarlo para entender la centralidad.
- A grandes rasgos: Si A es una matriz en donde todas sus entradas son positivas:
  - Hay un eigenvalor real que es más grande (en norma) que todos los demás eigenvalores.
  - Para este eigenvalor, el eigenespacio asociado es de dimensión 1.
  - Además, tiene un eigenvector en el "primer cuadrante"; es decir, hay un eigenvector para el cual todas sus entradas son positivas.
- Después hay una generalización a matrices no negativas, pero se pone medio complicado.

¿Cómo calcularlo?

- Hay métodos para calcular el eigenvector específicamente en matrices así.

¿Cómo calcularlo?

- Hay métodos para calcular el eigenvector específicamente en matrices así.
- Pero esos métodos tienen que ver más con álgebra lineal numérica.

¿Cómo calcularlo?

- Hay métodos para calcular el eigenvector específicamente en matrices así.
- Pero esos métodos tienen que ver más con álgebra lineal numérica.
- Básicamente hay que tomar el límite cuando k tiende a infinito de W^k

¿Cómo calcularlo?

- Hay métodos para calcular el eigenvector específicamente en matrices así.
- Pero esos métodos tienen que ver más con álgebra lineal numérica.
- Básicamente hay que tomar el límite cuando k tiende a infinito de W^k
- Mejor utilizamos una biblioteca que ya lo haga.

Índice:

1 Introducción

2 Centralidad de Grado

3 Centralidad de cercanía

4 Centralidad de Intermediación
- Definición Formal
- Algoritmo

5 Centralidad de Eigenvector

6 Intermedio: Teoría Espectral
- Matriz de Adyacencia
- Matriz de Difusión
- Gráficas con Pesos

7 Centralidad de Katz

8 PageRank

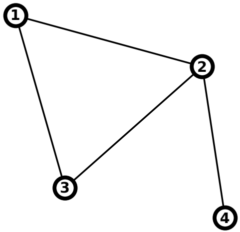
Matriz de Adyacencia

Definición

Dada una gráfica G, la matriz de adyacencia A_G de G está definida por:

$$A_G[i,j] = \begin{cases} 1 & \text{si } ij \in E(G) \\ 0 & \text{si } ij \notin E(G) \end{cases}$$

Ejemplo:



$$\begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Versión no dirigida, o con pesos

- P: ¿Qué pasa si tomamos un grafo dirigido?

Versión no dirigida, o con pesos

- P: ¿Qué pasa si tomamos un grafo dirigido?
- R: Pues nada, simplemente la matriz no será simétrica.

Versión no dirigida, o con pesos

- P: ¿Qué pasa si tomamos un grafo dirigido?
- R: Pues nada, simplemente la matriz no será simétrica.
- P: ¿Qué pasa si tomamos un grafo con pesos?

Versión no dirigida, o con pesos

- P: ¿Qué pasa si tomamos un grafo dirigido?
- R: Pues nada, simplemente la matriz no será simétrica.
- P: ¿Qué pasa si tomamos un grafo con pesos?
- R: Pues nada, simplemente la matriz tendrá números diferentes de 0 y 1.

¿Para qué?

- Pensamos un vector en $\mathbb{R}^n$ como una función de los vértices en los reales.

¿Para qué?

- Pensamos un vector en $\mathbb{R}^n$ como una función de los vértices en los reales.
- Aplicar el operador de la matriz de adyacencia (i.e. multiplicar la matriz por el vector) es "darle" a cada uno de mis vecinos el número que yo tengo.

¿Para qué?

- Pensamos un vector en $\mathbb{R}^n$ como una función de los vértices en los reales.
- Aplicar el operador de la matriz de adyacencia (i.e. multiplicar la matriz por el vector) es “darle” a cada uno de mis vecinos el número que yo tengo.

$$(A_G x)(a) = \sum_{b:ab\in E(G)} x(b)$$

¿Para qué?

- Pensamos un vector en $\mathbb{R}^n$ como una función de los vértices en los reales.
- Aplicar el operador de la matriz de adyacencia (i.e. multiplicar la matriz por el vector) es "darle" a cada uno de mis vecinos el número que yo tengo.

$$(A_G x)(a) = \sum_{b:ab \in E(G)} x(b)$$

- En este contexto, un eigenvector es una función de los vértices para la cual si aplico el operador de adyacencia, multiplica todo por una constante.

¿Para qué?

- Pensamos un vector en $\mathbb{R}^n$ como una función de los vértices en los reales.
- Aplicar el operador de la matriz de adyacencia (i.e. multiplicar la matriz por el vector) es "darle" a cada uno de mis vecinos el número que yo tengo.

$$(A_G x)(a) = \sum_{b: ab \in E(G)} x(b)$$

- En este contexto, un eigenvector es una función de los vértices para la cual si aplico el operador de adyacencia, multiplica todo por una constante.
- Podemos hacer un ejemplo: ¿qué pasa si multiplicamos la matriz de la gráfica anterior por el vector $(1, 1, 1, 1)$?

Matriz de Difusión

- El operador de difusión representa un proceso en el cual “masa” o alguna otra cosa se mueve de los vértices a sus vecinos.

Matriz de Difusión

- El operador de difusión representa un proceso en el cual "masa" o alguna otra cosa se mueve de los vértices a sus vecinos.
- Como la "masa" debe conservarse, vamos a pensar que se reparte a sus vecinos en partes iguales.

Matriz de Difusión

- El operador de difusión representa un proceso en el cual "masa" o alguna otra cosa se mueve de los vértices a sus vecinos.
- Como la "masa" debe conservarse, vamos a pensar que se reparte a sus vecinos en partes iguales.
- La matriz de grado D_G es la que en la diagonal tiene el grado de cada vértice y en lo demás tiene 0's.
- Por ejemplo, para la gráfica anterior:

$$D_G = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Matriz de Difusión

Definición

La matriz de difusión W_G *(también llamada la "walk matrix") está dada por:*

$$W_G := A_G D_G^{-1}$$

Matriz de Difusión

Definición

La matriz de difusión W_G *(también llamada la "walk matrix") está dada por:*

$$W_G := A_G D_G^{-1}$$

$$\begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 1 \\ \frac{1}{2} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \end{bmatrix}$$

- ¿Cómo actúa esta matriz en un vector x?

Matriz de Difusión

Definición

La matriz de difusión W_G *(también llamada la "walk matrix") está dada por:*

$$W_G := A_G D_G^{-1}$$

$$\begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 1 \\ \frac{1}{2} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \end{bmatrix}$$

- ¿Cómo actúa esta matriz en un vector x?
- Básicamente, cada quien "reparte" lo que tenga entre sus vecinos.

Matriz de Difusión

Definición

La matriz de difusión W_G *(también llamada la "walk matrix") está dada por:*

$$W_G := A_G D_G^{-1}$$

$$\begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 1 \\ \frac{1}{2} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \end{bmatrix}$$

- ¿Cómo actúa esta matriz en un vector x?
- Básicamente, cada quien "reparte" lo que tenga entre sus vecinos.
- Es decir:

$$(W_G x)(a) = \sum_{b:ab \in E} x(b)/d(b)$$

Diles qué pasa cuando no es invertible D.

Caminatas Aleatorias

- Esta matriz sirve para estudiar la dinámica de una caminata aleatoria en G.

Caminatas Aleatorias

- Esta matriz sirve para estudiar la dinámica de una caminata aleatoria en G.
- Una caminata aleatoria es simplemente lo siguiente: Estás parado en un vértice, y en cada momento escoges un vecino y caminas.

Caminatas Aleatorias

- Esta matriz sirve para estudiar la dinámica de una caminata aleatoria en G.
- Una caminata aleatoria es simplemente lo siguiente: Estás parado en un vértice, y en cada momento escoges un vecino y caminas.
- Entonces si das k pasos, la probabilidad de terminar en el vértice b si empezaste en el vértice a se puede calcular así: $W_G^k e_a(b)$, donde e_a es el vector característico de a.

Caminatas Aleatorias

- Esta matriz sirve para estudiar la dinámica de una caminata aleatoria en G.
- Una caminata aleatoria es simplemente lo siguiente: Estás parado en un vértice, y en cada momento escoges un vecino y caminas.
- Entonces si das k pasos, la probabilidad de terminar en el vértice b si empezaste en el vértice a se puede calcular así: $W_G^k e_a(b)$, donde e_a es el vector característico de a.
- En realidad se puede hacer con una distribución de probabilidad, simplemente multiplicando por W_G la distribución original k veces para obtener la distribución después de caminar k pasos.

En el caso con pesos

- Una gráfica con pesos en las aristas simplemente es una gráfica en donde cada arista tiene asociado un número real (su “peso”).

En el caso con pesos

- Una gráfica con pesos en las aristas simplemente es una gráfica en donde cada arista tiene asociado un número real (su “peso”).
- Distinguimos entre el grado combinatorio y el grado pesado de un vértice. Usamos d para el grado pesado:

$$d(a) = \sum_{b:ab \in E} w(a,b)$$

En el caso con pesos

- Una gráfica con pesos en las aristas simplemente es una gráfica en donde cada arista tiene asociado un número real (su “peso”).
- Distinguimos entre el grado combinatorio y el grado pesado de un vértice. Usamos d para el grado pesado:

$$d(a) = \sum_{b:ab \in E} w(a,b)$$

- Es decir, en vez de tomar el número de aristas, tomas la suma de las aristas de los pesos que le salen.

Caminata aleatoria

- Una caminata aleatoria en una gráfica con pesos se mueve de un vértice a a un vecino b con probabilidad proporcional a $w(a, b)$.

Caminata aleatoria

- Una caminata aleatoria en una gráfica con pesos se mueve de un vértice a a un vecino b con probabilidad proporcional a $w(a,b)$.
- Así que sigue funcionando la matriz W_G que habíamos definido antes como

$$W_G = A_G D_G^{-1}$$

Caminata aleatoria

- Una caminata aleatoria en una gráfica con pesos se mueve de un vértice a a un vecino b con probabilidad proporcional a $w(a, b)$.
- Así que sigue funcionando la matriz W_G que habíamos definido antes como

$$W_G = A_G D_G^{-1}$$

- Simplemente ahora A_G y D_G toman en cuenta los pesos.

Índice:

1 Introducción

2 Centralidad de Grado

3 Centralidad de cercanía

4 Centralidad de Intermediación
- Definición Formal
- Algoritmo

5 Centralidad de Eigenvector

6 Intermedio: Teoría Espectral
- Matriz de Adyacencia
- Matriz de Difusión
- Gráficas con Pesos

7 Centralidad de Katz

8 PageRank

Centralidad de Katz

- La centralidad de Katz es una medida de centralidad que es una generalización de la centralidad de grado y una variante de la centralidad de eigenvector.

Centralidad de Katz

- La centralidad de Katz es una medida de centralidad que es una generalización de la centralidad de grado y una variante de la centralidad de eigenvector.
- En la centralidad de grado, cada vértice depende de sus vecinos.

Centralidad de Katz

- La centralidad de Katz es una medida de centralidad que es una generalización de la centralidad de grado y una variante de la centralidad de eigenvector.
- En la centralidad de grado, cada vértice depende de sus vecinos.
- En la centralidad de Katz, la importancia de un vértice depende de cuántos caminos de longitud k salgan de ese vértice, para $k \in \{1, 2, ..., n\}$.

Centralidad de Katz

- La centralidad de Katz es una medida de centralidad que es una generalización de la centralidad de grado y una variante de la centralidad de eigenvector.
- En la centralidad de grado, cada vértice depende de sus vecinos.
- En la centralidad de Katz, la importancia de un vértice depende de cuántos caminos de longitud k salgan de ese vértice, para $k \in \{1, 2, ..., n\}$.
- Pero bueno, vamos a "penalizar" a los caminos más largos:

$$x_i = \sum_{k=1}^{\infty} \sum_{j=1}^{n} \alpha^k (A^k)[i,j]$$

Centralidad de Katz

- La centralidad de Katz es una medida de centralidad que es una generalización de la centralidad de grado y una variante de la centralidad de eigenvector.
- En la centralidad de grado, cada vértice depende de sus vecinos.
- En la centralidad de Katz, la importancia de un vértice depende de cuántos caminos de longitud k salgan de ese vértice, para $k \in \{1, 2, ..., n\}$.
- Pero bueno, vamos a "penalizar" a los caminos más largos:

$$x_i = \sum_{k=1}^{\infty} \sum_{j=1}^{n} \alpha^k (A^k)[i,j]$$

- Claro, estamos pensando que $0 \leq \alpha < 1$.

Centralidad de Katz

- La centralidad de Katz es una medida de centralidad que es una generalización de la centralidad de grado y una variante de la centralidad de eigenvector.
- En la centralidad de grado, cada vértice depende de sus vecinos.
- En la centralidad de Katz, la importancia de un vértice depende de cuántos caminos de longitud k salgan de ese vértice, para $k \in \{1, 2, ..., n\}$.
- Pero bueno, vamos a "penalizar" a los caminos más largos:

$$x_i = \sum_{k=1}^{\infty} \sum_{j=1}^{n} \alpha^k (A^k)[i,j]$$

- Claro, estamos pensando que $0 \leq \alpha < 1$.
- Si dividimos entre α y tomamos el límite cuando α tiende a 0, nos da la centralidad de grado.

Katz

- Ejercicio: ¿Cuánto vale la centralidad de Katz en C_n?

Katz

- Ejercicio: ¿Cuánto vale la centralidad de Katz en C_n?
  - $2\alpha + 4\alpha^2 + 8\alpha^3 + 16\alpha^3 + ...$

Katz

- Ejercicio: ¿Cuánto vale la centralidad de Katz en C_n?
  - $2\alpha + 4\alpha^2 + 8\alpha^3 + 16\alpha^3 + ...$
- ¿Y en K_n?

Katz

- Ejercicio: ¿Cuánto vale la centralidad de Katz en C_n?
  - $2\alpha + 4\alpha^2 + 8\alpha^3 + 16\alpha^3 + ...$
- ¿Y en K_n?
  - $(n-1)\alpha + (n-1)^2\alpha^2 + (n-1)^3\alpha^3 + ...$

Katz

- Ejercicio: ¿Cuánto vale la centralidad de Katz en C_n?
  - $2\alpha + 4\alpha^2 + 8\alpha^3 + 16\alpha^3 + ...$
- ¿Y en K_n?
  - $(n-1)\alpha + (n-1)^2\alpha^2 + (n-1)^3\alpha^3 + ...$
- ¿Y en P_n para el primer vértice? (difícil)

Katz

- Ejercicio: ¿Cuánto vale la centralidad de Katz en C_n?
  - $2\alpha + 4\alpha^2 + 8\alpha^3 + 16\alpha^3 + ...$
- ¿Y en K_n?
  - $(n-1)\alpha + (n-1)^2\alpha^2 + (n-1)^3\alpha^3 + ...$
- ¿Y en P_n para el primer vértice? (difícil)
- Respuesta:
  - $\alpha + 2\alpha^2 + 3\alpha^3 + 6\alpha^3 + 10\alpha^4 + ...+?.$

Katz

- Ejercicio: ¿Cuánto vale la centralidad de Katz en C_n?
  - $2\alpha + 4\alpha^2 + 8\alpha^3 + 16\alpha^3 + ...$
- ¿Y en K_n?
  - $(n-1)\alpha + (n-1)^2\alpha^2 + (n-1)^3\alpha^3 + ...$
- ¿Y en P_n para el primer vértice? (difícil)
- Respuesta:
  - $\alpha + 2\alpha^2 + 3\alpha^3 + 6\alpha^3 + 10\alpha^4 + ... + ?.$

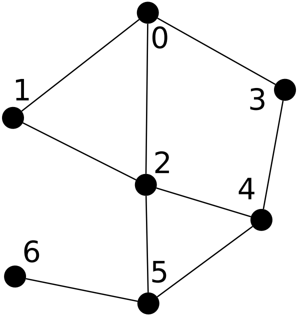
Calcular Katz numéricamente

```
def katz(G, alpha, v):
    A = G.adjacency_matrix()
    n = G.num_verts()
    return sum([sum([alpha^k*(A^k)[v,j] for j in range(n)])
    for k in range(1000)])
```

katz.py

Centralidad de Katz: Ejemplo

Ejemplo: Calculemos la centralidad de Katz para $\alpha = 0.1$:



- Katz del vertice 0 es 1.41354466858790
- Katz del vertice 1 es 1.29682997118156
- Katz del vertice 2 es 1.55475504322767
- Katz del vertice 3 es 1.28386167146974
- Katz del vertice 4 es 1.42507204610951
- Katz del vertice 5 es 1.41210374639769

Índice:

1 Introducción

2 Centralidad de Grado

3 Centralidad de cercanía

4 Centralidad de Intermediación
- Definición Formal
- Algoritmo

5 Centralidad de Eigenvector

6 Intermedio: Teoría Espectral
- Matriz de Adyacencia
- Matriz de Difusión
- Gráficas con Pesos

7 Centralidad de Katz

8 PageRank

PageRank Simplificado

- Imaginemos que tenemos una digráfica.

PageRank Simplificado

- Imaginemos que tenemos una digráfica.
- La idea de pagerank es sencilla: Imagínate que comienzas en un vértice aleatorio y que repetidamente sigues una arista.

PageRank Simplificado

- Imaginemos que tenemos una digráfica.
- La idea de pagerank es sencilla: Imagínate que comienzas en un vértice aleatorio y que repetidamente sigues una arista.
- Entonces, básicamente comenzamos con el vector $(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$, que representa la distribución de probabilidad inicial.

PageRank Simplificado

- Imaginemos que tenemos una digráfica.
- La idea de pagerank es sencilla: Imagínate que comienzas en un vértice aleatorio y que repetidamente sigues una arista.
- Entonces, básicamente comenzamos con el vector $(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$, que representa la distribución de probabilidad inicial.
- Después, repetidamente escogemos una arista aleatoria que sale, y la seguimos. Es decir, multiplicamos por la "walk matrix".

PageRank Simplificado

- Imaginemos que tenemos una digráfica.
- La idea de pagerank es sencilla: Imagínate que comienzas en un vértice aleatorio y que repetidamente sigues una arista.
- Entonces, básicamente comenzamos con el vector $(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$, que representa la distribución de probabilidad inicial.
- Después, repetidamente escogemos una arista aleatoria que sale, y la seguimos. Es decir, multiplicamos por la "walk matrix".
- Etc.

PageRank Menos simplificado

- Ahora pensemos que no nos "vamos a infinito", sino que en cada momento hay cierta probabilidad $0 < d < 1$ de detenerse. (Google, dicen, utiliza $d = 0.85$).

PageRank Menos simplificado

- Ahora pensemos que no nos "vamos a infinito", sino que en cada momento hay cierta probabilidad $0 < d < 1$ de detenerse. (Google, dicen, utiliza $d = 0.85$).
- Digamos que **v** es el vector $(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$.

PageRank Menos simplificado

- Ahora pensemos que no nos "vamos a infinito", sino que en cada momento hay cierta probabilidad $0 < d < 1$ de detenerse. (Google, dicen, utiliza $d = 0.85$).
- Digamos que **v** es el vector $(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$.
- ¿Cuánto vale entonces PageRank?:

PageRank Menos simplificado

- Ahora pensemos que no nos “vamos a infinito”, sino que en cada momento hay cierta probabilidad $0 < d < 1$ de detenerse. (Google, dicen, utiliza $d = 0.85$).
- Digamos que **v** es el vector $(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$.
- ¿Cuánto vale entonces PageRank?:

$$\begin{aligned} PR &= \left[dI + (1-d)dW + d(1-d)^2W^2 + ... \right] \mathbf{v} \\ &= dI\mathbf{v} + (1-d)W \cdot PR \end{aligned}$$

- Entonces, si a es un vértice,

$$PR(a) = \frac{d}{n} + (1-d) \left(\sum_{b \to a} \frac{PR(u)}{\delta_{ex}(a)} \right)$$

PageRank Menos simplificado

- Ahora pensemos que no nos "vamos a infinito", sino que en cada momento hay cierta probabilidad $0 < d < 1$ de detenerse. (Google, dicen, utiliza $d = 0.85$).
- Digamos que **v** es el vector $(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$.
- ¿Cuánto vale entonces PageRank?:

$$\begin{aligned} PR &= \left[dI + (1-d)dW + d(1-d)^2 W^2 + ... \right] \mathbf{v} \\ &= dI\mathbf{v} + (1-d)W \cdot PR \end{aligned}$$

- Entonces, si a es un vértice,

$$PR(a) = \frac{d}{n} + (1-d) \left(\sum_{b \to a} \frac{PR(u)}{\delta_{ex}(a)} \right)$$

- Se puede calcular en cualquier programa de álgebra lineal también:

$$PR = d \left[I - (1-d)W \right]^{-1} \mathbf{v}$$

Grafos pesados: detalle técnico

- Para grafos con pesos en las aristas, todo funciona igual, simplemente hay que tomar las matrices correspondientes pesadas.

Grafos pesados: detalle técnico

- Para grafos con pesos en las aristas, todo funciona igual, simplemente hay que tomar las matrices correspondientes pesadas.
- Detalle técnico: Hay que considerar una cosa para grafos pesados:

Grafos pesados: detalle técnico

- Para grafos con pesos en las aristas, todo funciona igual, simplemente hay que tomar las matrices correspondientes pesadas.
- Detalle técnico: Hay que considerar una cosa para grafos pesados:
- ¿Qué significa peso de 0 o casi 0?

Grafos pesados: detalle técnico

- Para grafos con pesos en las aristas, todo funciona igual, simplemente hay que tomar las matrices correspondientes pesadas.
- Detalle técnico: Hay que considerar una cosa para grafos pesados:
- ¿Qué significa peso de 0 o casi 0?
- Dependiendo del contexto, a veces significa que la arista "no está" (o "casi no está").

Grafos pesados: detalle técnico

- Para grafos con pesos en las aristas, todo funciona igual, simplemente hay que tomar las matrices correspondientes pesadas.
- Detalle técnico: Hay que considerar una cosa para grafos pesados:
- ¿Qué significa peso de 0 o casi 0?
- Dependiendo del contexto, a veces significa que la arista "no está" (o "casi no está").
    - Por ejemplo, si el peso de la arista es "cuántas conversaciones de facebook tienen dos amigos", entonces entre más conversaciones tengan, más "fuerte" es la arista, y si no han tenido ninguna, es que no son amigos.

Grafos pesados: detalle técnico

- Para grafos con pesos en las aristas, todo funciona igual, simplemente hay que tomar las matrices correspondientes pesadas.
- Detalle técnico: Hay que considerar una cosa para grafos pesados:
- ¿Qué significa peso de 0 o casi 0?
- Dependiendo del contexto, a veces significa que la arista "no está" (o "casi no está").
    - Por ejemplo, si el peso de la arista es "cuántas conversaciones de facebook tienen dos amigos", entonces entre más conversaciones tengan, más "fuerte" es la arista, y si no han tenido ninguna, es que no son amigos.
- Y a veces significa que los dos vértices están "casi pegados".

Grafos pesados: detalle técnico

- Para grafos con pesos en las aristas, todo funciona igual, simplemente hay que tomar las matrices correspondientes pesadas.
- Detalle técnico: Hay que considerar una cosa para grafos pesados:
- ¿Qué significa peso de 0 o casi 0?
- Dependiendo del contexto, a veces significa que la arista “no está” (o “casi no está”).
  - Por ejemplo, si el peso de la arista es “cuántas conversaciones de facebook tienen dos amigos”, entonces entre más conversaciones tengan, más “fuerte” es la arista, y si no han tenido ninguna, es que no son amigos.
- Y a veces significa que los dos vértices están “casi pegados”.
  - Por ejemplo, si el peso de una arista es “la longitud de la carretera que las une”, entonces peso 0 significa que son la misma ciudad.

Grafos pesados: detalle técnico

- Para grafos con pesos en las aristas, todo funciona igual, simplemente hay que tomar las matrices correspondientes pesadas.
- Detalle técnico: Hay que considerar una cosa para grafos pesados:
- ¿Qué significa peso de 0 o casi 0?
- Dependiendo del contexto, a veces significa que la arista "no está" (o "casi no está").
    - Por ejemplo, si el peso de la arista es "cuántas conversaciones de facebook tienen dos amigos", entonces entre más conversaciones tengan, más "fuerte" es la arista, y si no han tenido ninguna, es que no son amigos.
- Y a veces significa que los dos vértices están "casi pegados".
    - Por ejemplo, si el peso de una arista es "la longitud de la carretera que las une", entonces peso 0 significa que son la misma ciudad.
- Es decir, a veces pensamos que las aristas que "no están" tienen peso 0, y a veces pensamos que tienen peso infinito.

Grafos pesados: detalle técnico

- Para grafos con pesos en las aristas, todo funciona igual, simplemente hay que tomar las matrices correspondientes pesadas.
- Detalle técnico: Hay que considerar una cosa para grafos pesados:
- ¿Qué significa peso de 0 o casi 0?
- Dependiendo del contexto, a veces significa que la arista "no está" (o "casi no está").
  - Por ejemplo, si el peso de la arista es "cuántas conversaciones de facebook tienen dos amigos", entonces entre más conversaciones tengan, más "fuerte" es la arista, y si no han tenido ninguna, es que no son amigos.
- Y a veces significa que los dos vértices están "casi pegados".
  - Por ejemplo, si el peso de una arista es "la longitud de la carretera que las une", entonces peso 0 significa que son la misma ciudad.
- Es decir, a veces pensamos que las aristas que "no están" tienen peso 0, y a veces pensamos que tienen peso infinito.
- Simplemente hay que tomar $peso = 1/costo$, para que 0 signifique "no hay arista".