

UNIDAD 5: ANÁLISIS DE GRAFOS DE REDES SOCIALES

AGRUPAMIENTO EN GRAFOS DE REDES SOCIALES

Blanca Vázquez y Gibran Fuentes-Pineda

Noviembre 2020

¿Cómo identificar comunidades en las redes sociales?

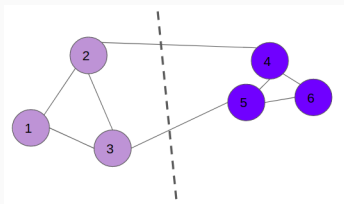
¿QUÉ ES UNA COMUNIDAD?

Una comunidad, con respecto a grafos, se puede definir como un subconjunto de nodos que están densamente conectados entre sí y débilmente conectados a los nodos de las otras comunidades en el mismo grafo.

Ejemplo: compañeros de la universidad, colegas, familia.

DETECCIÓN DE COMUNIDADES

- La detección de comunidades en una red es una de las tareas más importantes en el análisis de redes.
- Consiste en dividir una red en múltiples comunidades.



Existen principalmente 2 métodos para la detección de comunidades en grafos:

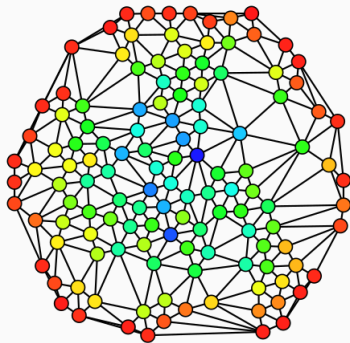
- Métodos aglomerativos
- Métodos de división

- En los métodos aglomerativos, se inicia con un grafo vacío que consta de nodos del grafo original pero sin bordes.
- A continuación, los bordes se agregan uno por uno al grafo, comenzando desde los bordes más fuertes a los más débiles.
- Para calcular la fuerza del borde se utilizan múltiples algoritmos, como el conteo de triángulos.

- En los métodos de división (vamos al revés) se inicia con el grafo completo y se eliminan los bordes de forma iterativa.
- Primero se elimina el borde con el mayor peso.
- En cada paso, se repite el cálculo del peso del borde, ya que el peso de los bordes restantes cambia después de que se quita un borde.
- Después de una cierta cantidad de pasos, se obtienen grupos de nodos densamente conectados.
- El algoritmo de Girvan-Newman (GN) es un algoritmo de métodos de división.

- En el algoritmo de Girvan-Newman, las comunidades en un grafo se detectan eliminando iterativamente los bordes del grafo, según el valor de centralidad de la intermediación del borde.

- La centralidad de intermediación de un borde (EBC) (*betweenness*) es una técnica especializada de agrupamiento.
- Se define como el número de caminos más cortos que pasan a través de un borde en una red.
- Todos y cada uno de los bordes reciben una puntuación EBC basada en las rutas más cortas entre todos los nodos del grafo.

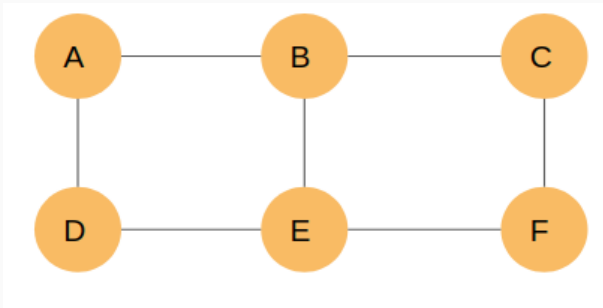


En las redes de telecomunicaciones, un nodo con alta intermediación se refiere a un nodo con mayor control en la red (debido a que mayor información pasa a través del nodo).

Imagen tomada de wikipedia.

EJEMPLO DE CENTRALIDAD DE INTERMEDIACIÓN

Con respecto a los grafos, la ruta más corta significa la ruta entre dos nodos que cubren la menor distancia. Consideremos el siguiente ejemplo para calcular EBC.



El grafo tiene 6 nodos y 7 bordes.

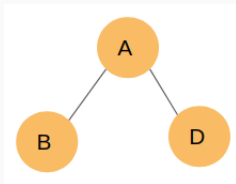
EJEMPLO DE CENTRALIDAD DE INTERMEDIACIÓN

Encontrar el EBC para todas las aristas en el grafo, el cual es un proceso iterativo:

- Se toma un nodo a la vez y se traza las rutas más cortas a los otros nodos desde el nodo seleccionado.
- Basándonos en los caminos más cortos, se calcula los puntajes EBC para todos los bordes.
- Se repite este proceso para cada nodo del grafo. Como observamos, tenemos 6 nodos en el grafo, por lo tanto, habrá 6 iteraciones de este proceso.
- Esto significa que cada borde obtendrá 6 puntuaciones. Estos puntajes se agregarán de forma escalonada
- Finalmente, la puntuación total de cada borde se dividirá por 2 para obtener la puntuación EBC.

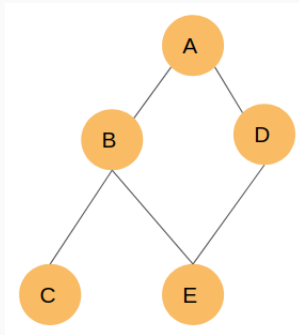
EJEMPLO DE CENTRALIDAD DE INTERMEDIACIÓN

Iniciamos con el nodo A. Los nodos conectados directamente al nodo A son los nodos B y D. Entonces, las rutas más cortas a B y D desde A son AB y AD respectivamente:



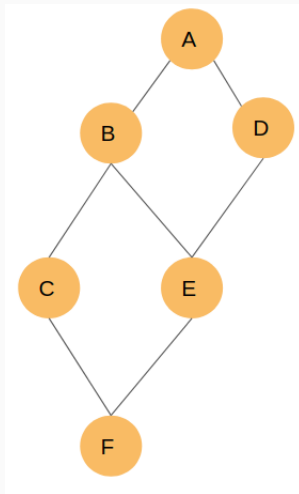
EJEMPLO DE CENTRALIDAD DE INTERMEDIACIÓN

Observemos que los caminos más cortos a los nodos C y E desde A pasan por B y D:



EJEMPLO DE CENTRALIDAD DE INTERMEDIACIÓN

Las rutas más cortas al último nodo F desde el nodo A pasan por los nodos B, D, C y E:

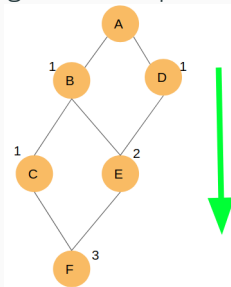


El proceso anterior muestra solo las rutas más cortas desde el nodo A a todos los demás nodos. Ahora veremos cómo se puntúan los bordes.

- Antes de dar puntajes a los bordes, asignaremos un puntaje a los nodos en el grafo de ruta más corta. Para asignar estos puntajes, tendremos que atravesar el grafo desde el nodo raíz, es decir, el nodo A hasta el último nodo (nodo F).

EJEMPLO DE CENTRALIDAD DE INTERMEDIACIÓN

Asignación de puntajes

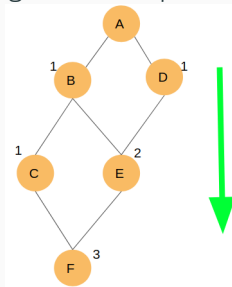


Observamos que los nodos B y D reciben una puntuación de 1 cada uno. Esto se debe a que la ruta más corta a cualquier nodo desde el nodo A es solo una.

Por la misma razón, al nodo C se le ha dado una puntuación de 1 ya que solo hay una ruta más corta desde el nodo A al nodo C.

EJEMPLO DE CENTRALIDAD DE INTERMEDIACIÓN

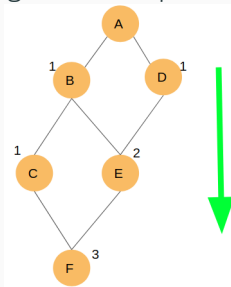
Asignación de puntajes



Pasando al nodo E. Está conectado al nodo A a través de dos rutas más cortas, ABE y ADE. Por tanto, obtiene una puntuación de 2.

EJEMPLO DE CENTRALIDAD DE INTERMEDIACIÓN

Asignación de puntajes

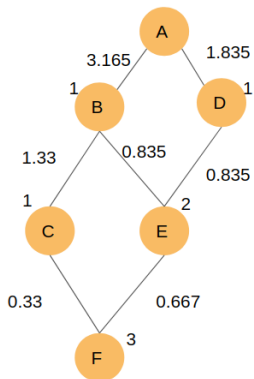


El último nodo F está conectado a A a través de tres rutas más cortas: ABCF, ABEF y ADEF. Entonces, obtiene una puntuación de 3.

EJEMPLO DE CENTRALIDAD DE INTERMEDIACIÓN

A continuación, procederemos a calcular las puntuaciones de los bordes. Aquí nos moveremos hacia atrás, del nodo F al nodo A:

Calculando los puntajes por aristas



$$FC = \frac{1}{3} = 0.33$$

$$FE = \frac{2}{3} = 0.667$$

$$CB = 1 + 0.33 = 1.33$$

$$EB = (1 + 0.667)/2 = 0.835$$

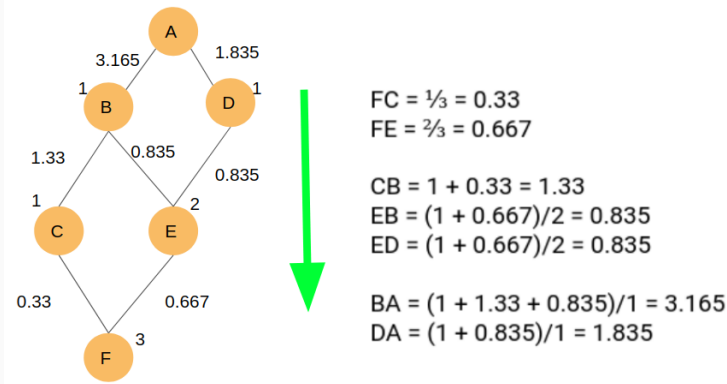
$$ED = (1 + 0.667)/2 = 0.835$$

$$BA = (1 + 1.33 + 0.835)/1 = 3.165$$

$$DA = (1 + 0.835)/1 = 1.835$$

EJEMPLO DE CENTRALIDAD DE INTERMEDIACIÓN

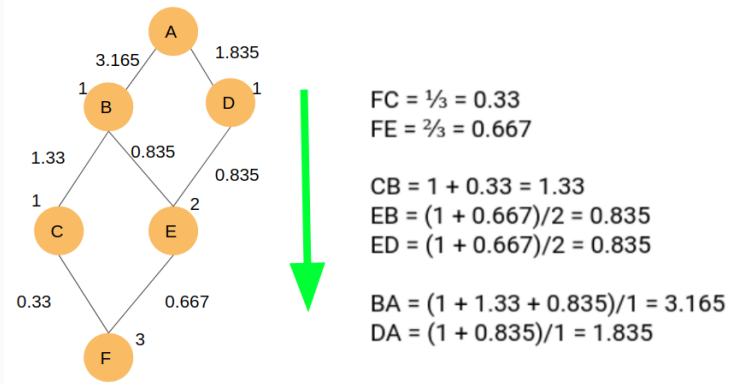
Calculando los puntajes por aristas



Primero calculamos la puntuación de los bordes FC y FE. Como se observa, la puntuación de borde para el borde FC es la relación de las puntuaciones de nodo de C y F, es decir, $1/3$ o 0.33 . De manera similar, para FE, la puntuación de borde es $2/3$.

EJEMPLO DE CENTRALIDAD DE INTERMEDIACIÓN

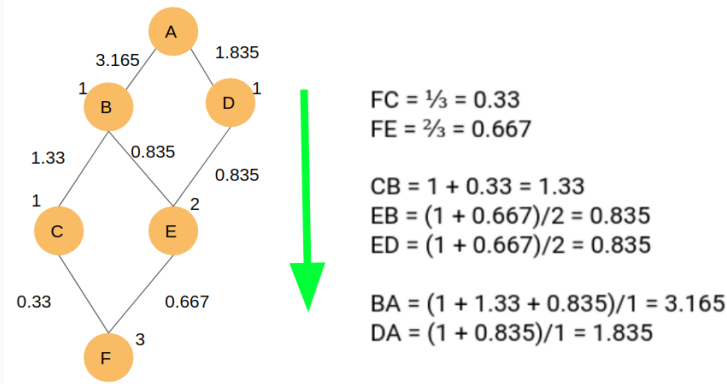
Calculando los puntajes por aristas



Ahora tenemos que calcular la puntuación del borde para los bordes CB, EB y ED. Según el algoritmo de GN, a partir de este nivel, cada nodo tendrá un valor predeterminado de 1 y los puntajes de borde calculados en el paso anterior se sumarán a este valor.

EJEMPLO DE CENTRALIDAD DE INTERMEDIACIÓN

Calculando los puntajes por aristas



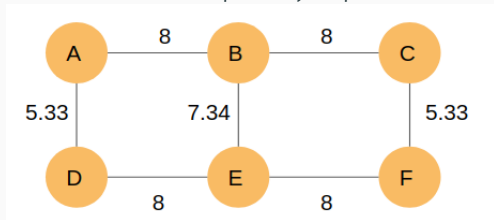
Entonces, la puntuación de borde de CB es $(1 + 0.33) / 1$. De manera similar, la puntuación de borde EB o ED es $(1 + 0.667) / 2$. Luego pasamos al siguiente nivel para calcular los puntajes de borde para BA y DA.

Calculando los puntajes por aristas

- Hasta ahora, hemos calculado las puntuaciones de los bordes de las rutas más cortas con respecto al nodo A.
- Tenemos que repetir los mismos pasos nuevamente desde los otros cinco nodos restantes.
- Al final, obtendremos un conjunto de seis puntuaciones para todos los bordes de la red.
- Agregaremos estos puntajes y los asignaremos al grafo original como se muestra a continuación.

EJEMPLO DE CENTRALIDAD DE INTERMEDIACIÓN

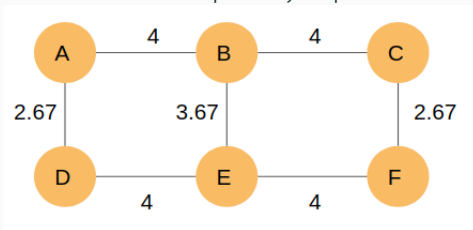
Calculando los puntajes por aristas



En la imagen se muestra la suma para cada nodo y borde.

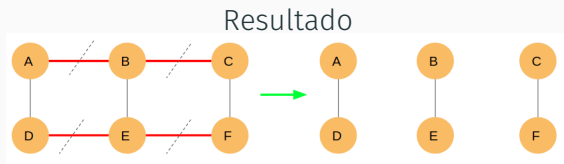
EJEMPLO DE CENTRALIDAD DE INTERMEDIACIÓN

Calculando los puntajes por aristas



Dado que es un grafo no dirigido, dividiremos estos puntajes entre dos y finalmente obtendremos los puntajes EBC.

EJEMPLO DE CENTRALIDAD DE INTERMEDIACIÓN



Según el algoritmo de GN, después de calcular los puntajes EBC, los bordes con los puntajes más altos se quitarán hasta el punto en que el grafo se divide en dos.

Entonces, en el grafo anterior, podemos ver que los bordes AB, BC, DE y EF tienen la puntuación más alta, es decir, 4. Tacharemos estos bordes y nos da 3 subgrafos que podemos llamar comunidades.

- Prateek Joshi, Getting Started with Community Detection in Graphs and Networks, 2020.