Universidad Michoacana de San Nicolás de Hidalgo

INSTITUTO DE FÍSICA Y MATEMÁTICAS

"Multi-centralidad de intermediación "

Tesina Para Obtener el Título de:
MAESTRO EN CIENCIAS MATEMÁTICAS

PRESENTA:

Roberto Lara Sarmiento

DIRIGIDA POR:

Doctor en Ciencias Miguel Raggi Pérez

Morelia, Michoacán - Junio, 2017.

Índice general

Resumen

Una red consiste en un conjunto de actores y relaciones entre ellos. Un problema fundamental en redes es encontrar el actor, o conjunto de actores, más importantes dentro de la red. Existe una gran variedad de índices de centralidad para encontrar el actor más importante de la red en diversos contextos. Uno de estos índices de centralidad es la *centralidad de intermediación*, donde la importancia de un actor depende de la fracción de caminos mínimos entre actores que pasan por él. Ulrik Brandes proporcionó un algoritmo eficiente para calcular la centralidad de intermediación de todos los actores en una red.

En el presente documento, se estudia la generalización obvia de centralidad de intermediación a subconjuntos. Es decir, la importancia de un subconjunto de actores depende de la fracción de caminos mínimos entre actores que pasan por algún vértice del subconjunto. Se propone un algoritmo eficiente para calcular la centralidad de intermediación de un subconjunto y un algoritmo para encontrar el subconjunto con k elementos de mayor centralidad de intermediación.

Palabras clave: centralidad, intermediación, subconjuntos, redes, algoritmo.

Abstract

A network is a set of actors and relationships between them. A fundamental issue is to find the most important actor, or subset of actors, within the network. There exist several centrality indexes to find the most important actor of the network, according to several contexts. The *betweenness centrality* is one of these centralities, where the importance of an actor depends on the ratio of shortest paths between actors that go through that actor. Ulrik Brandes provided an efficient algorithm to compute the betweenness centrality of all the actors of a network.

In this document, we study the obvious generalization of betweenness centrality for subsets. This means that the importance of a subset of actors depends on the ratio of shortest paths between actors that go through an actor of the subset. We provide an efficient algorithm to compute the betweenness centrality of a subset and an algorithm to find the subset of k actor with the biggest betweenness centrality.

Keywords: centrality, betweenness, subsets, networks, algorithm.

Introducción

Una *red* consiste en un conjunto de actores y relaciones entre ellos. Por ejemplo, las redes sociales, donde los actores es un conjunto de personas y existe una relación entre dos actores siempre y cuando exista una relación social entre esas dos personas; una red de computadoras, donde los actores son las computadoras y las relaciones representan conexiones entre ellas o si son capaces de mandarse mensajes directamente; una red de ciudades y carreteras, etcétera.

Las redes pueden ser representadas y estudiadas con teoría de grafos, donde los vértices representan los actores y las aristas representan relaciones. Un problema fundamental en las redes es encontrar los actores más *importantes* de la red. Por supuesto, la definición de "importantes" depende completamente del contexto.

Existe una gran variedad de índices de centralidad en grafos que nos permiten asignarle un valor de *importancia* a cada actor. Por ejemplo:

- La *centralidad de grado* de un nodo se define simplemente como su grado. Así, los vértices más importantes serán aquellos con mayor grado.
- En la *centralidad de eigenvector* la importancia de un vértice es proporcional a la importancia de sus vecinos. Es decir, un vértice será importante si está conectado a otros vértices importantes. Existen variantes famosas de esta centralidad, como *centralidad de Katz* y *PageRank*.
- La *centralidad de cercanía* es el inverso de la suma de las distancias hacia los demás nodos. Es decir, un vértice será importante si se encuentra cerca de los demás vértices.
- En la *centralidad de intermediación* la importancia de un vértice está relacionada con la fracción de geodésicas (caminos de longitud mínima entre vértices) que pasan por él.

Existe una gran cantidad de referencias donde se estudia más a detalle distintas centralidades, como en [4]. En el presente documento nos concentraremos solamente en la centralidad de intermediación.

Podemos entender a la centralidad de intermediación con el siguiente ejemplo. Supongamos una red de comunicación donde todas las parejas de actores se envían mensajes con la misma frecuencia. Supondremos, además, que los mensajes se envían por alguna geodésica. La centralidad de intermediación responde a la pregunta: ¿cuál es el actor por el que pasan más mensajes?

La centralidad de intermediación (*betweenness centrality* es su nombre en inglés) ha sido ampliamente estudiada. Brandes ([2]) proporciona un algoritmo para calcularla en tiempo $O(n(n+m))$, con n y m el número de vértices y aristas, respectivamente. El algoritmo se basa en calcular de manera eficiente la contribución de un vértice en algún otro vértice.

Nosotros estudiamos una generalización natural de la centralidad de intermediación para subconjuntos. Es decir, dado k, queremos asignar un valor a cada subconjunto U de tamaño k de acuerdo a la fracción de geodésicas que tienen intersección no vacía con U. En particular, estamos interesados en encontrar un subconjunto que maximice este valor.

Una primera aproximación sería tomar los k vértices con más alto índice de centralidad de intermediación, pero este conjunto no siempre tiene la centralidad de intermediación más grande. Por ejemplo, en el grafo de la Figura 1, los vértices con más centralidad de intermediación son C y D. Sin embargo, el conjunto $\{B, E\}$ tiene una centralidad de intermediación mayor que $\{C, D\}$. Esto pasa porque $\{C, D\}$ tiene intersección vacía con dos geodésicas: de A a B y de E a F; por otro lado, $\{B, E\}$ solo tiene intersección vacía con una geodésica: de C a D.

Figura 1: Grafo de línea.

En la segunda sección de este documento describimos el algoritmo de Brandes para calcular la centralidad de intermediación de todos los vértices de un grafo. En la tercera sección mostraremos un algoritmo para calcular la centralidad de intermediación de un conjunto de vértices y el algoritmo para encontrar el subconjunto con mayor centralidad de intermediación. En la cuarta sección se presentan resultados y una serie de experimentos que se hicieron sobre algunos grafos particulares. El código en C++ que usamos se puede encontrar en:

https://github.com/rlara-sarmiento/multi-betweenness-centrality

Capítulo 1

Centralidad de intermediación

Pensemos en un grafo $G = (\mathrm{V}, \mathrm{E})$ conexo. Usaremos n y m para representar el número de vértices y el número de aristas, respectivamente. Por simplicidad, supondremos que G es un grafo simple y no dirigido, aunque los resultados pueden extenderse fácilmente a grafos dirigidos y con aristas múltiples. Así mismo, fijemos un número natural $0 < k \leq n$.

Un *camino* entre dos vértices $s, t \in \mathrm{V}$ es una sucesión alternante de vértices y aristas, comenzando en s y terminando en t, de tal manera que cada arista en el camino conecta al vértice anterior con el sucesor. La *longitud* de un camino será el número de aristas que contenga. Definimos la *distancia* entre s y t, denotada como $\mathrm{d}_G(s,t)$, como la mínima longitud de un camino entre s y t; por definición, $\mathrm{d}_G(s,s) = 0$. Notemos que, en general, pueden existir varios caminos de longitud mínima. Además, $\mathrm{d}_G(s,t) = \mathrm{d}_G(t,s)$. Finalmente, una *geodésica* será un camino de longitud mínima entre dos vértices.

Para cualesquiera $s, t \in \mathrm{V}$, definimos $\Upsilon(s,t)$ como el conjunto de geodésicas entre s y t. Para $U \subseteq \mathrm{V}$, $\Upsilon(s,t;U)$ será el subconjunto de caminos de $\Upsilon(s,t)$ que pasan por algún vértice de U. En caso de que $U = \{v\}$, escribiremos simplemente $\Upsilon(s,t;v)$ para denotar a las geodésicas de s a t que pasan por un vértice v.

Definición 1. *Para $s, t \in V$ y $U \subseteq V$, definimos*

$$\sigma(s,t) = |\Upsilon(s,t)|,$$

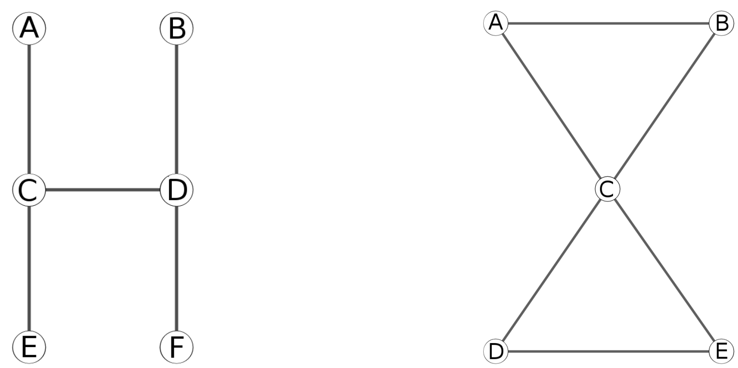
$$\sigma(s,t;U) = |\Upsilon(s,t;U)|.$$

Definiremos ahora la centralidad de intermediación para subconjuntos, denotada por $\mathrm{bc}(U)$ por sus siglas en inglés. También hablaremos de distintas definiciones y los valores mínimos y máximos que puede tomar la centralidad de intermediación.

Definición 2 (Centralidad de Intermediación)**.** *Dado $U \subseteq V$, definimos la centralidad de intermediación de U como*

$$bc(U) := \sum_{s,t \in V} \frac{\sigma(s,t;U)}{\sigma(s,t)}.$$

Cuando $U = \{v\}$, escribiremos simplemente $bc(v)$.



(a) Ejemplo 1 (b) Ejemplo 2

Figura 1.1: Ejemplos de grafos.

En el grafo del Ejemplo 1, los vértices que tienen máxima centralidad de intermediación son $\mathrm{bc}(C) = \mathrm{bc}(D) = 25$. Por otro lado, $\mathrm{bc}(A) = \mathrm{bc}(B) = \mathrm{bc}(E) = \mathrm{bc}(F) = 11$. Nótese que la centralidad de intermediación de estos últimos vértices se debe a que cada camino se cuenta dos veces: una vez de s a t y otra de t a s. Además, también contamos aquellos caminos en los que $s = t$. Más adelante se hablará de una definición más intuitiva donde se evitan estos casos y repeticiones. Para la centralidad de intermediación en subconjuntos, se tiene que $\mathrm{bc}(\{C, D\}) = 32$ es el mayor valor que se alcanza.

En el grafo del Ejemplo 2, encontramos que $\mathrm{bc}(A) = \mathrm{bc}(B) = \mathrm{bc}(D) = \mathrm{bc}(E) = 9$ y $\mathrm{bc}(C) = 17$, siendo este último el de mayor centralidad de intermediación. Para subconjuntos, se tiene que $\mathrm{bc}(\{C, D\}) = 20$ es el par con mayor centralidad de intermediación..

En el grafo del Ejemplo 3, los dos vértices con más centralidad de intermediación son el B y C, con $\mathrm{bc}(B) = 23$ y $\mathrm{bc}(C) = 29$. También se tiene que $\mathrm{bc}(D) = \mathrm{bc}(E) = 20$. Si calculamos centralidad de intermediación para subconjuntos se tiene que $\mathrm{bc}(\{B, C\}) = 32$ y $\mathrm{bc}(\{D, E\}) = 38$. Este es un ejemplo en el que el par con mayor centralidad de intermediación no contiene ninguno de los dos vértices con mayor centralidad de intermediación.

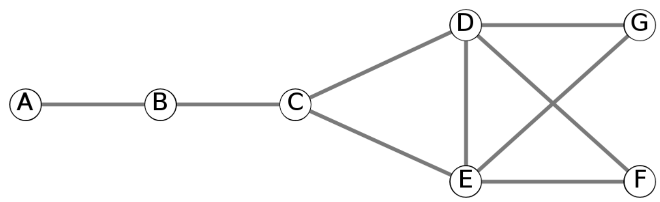


Figura 1.2: Ejemplo 3.

Para que la centralidad de intermediación tenga un valor entre 0 y 1, se puede normalizar sobre el número de parejas de V:

$$\mathrm{bc}(U) = \frac{1}{n^2} \sum_{s,t \in \mathrm{V}} \frac{\sigma(s,t;U)}{\sigma(s,t)}.$$

De esta manera, se puede comparar la importancia de vértices en distintos grafos. Por simplicidad, a lo largo de este documento, no normalizaremos la centralidad de intermediación.

Existe una definición alternativa de la centralidad de intermediación más natural:

$$\mathrm{bc}'(U) = \sum_{\{s,t\} \in \binom{V \setminus U}{2}} \frac{\sigma(s,t;U)}{\sigma(s,t)}.$$

En esta definición evitamos repeticiones y algunos casos triviales, como cuando $s = t$ o $s \in U$. Esta definición se vuelve más natural al momento de calcular la centralidad de intermediación *a mano*.

Ambas definiciones son equivalentes y se puede obtener una de la otra de la siguiente manera:

$$\mathrm{bc}'(U) = \frac{\mathrm{bc}(U) - k}{2} - \binom{k}{2} - k(n-k),$$

donde $k = |U|$. Claramente, dados $U, W \subseteq V$ con k elementos, $\mathrm{bc}(U) \geq \mathrm{bc}(W) \iff \mathrm{bc}'(U) \geq \mathrm{bc}'(W)$.

Trabajaremos con la primera definición pues hace que el algoritmo y su implementación sean más sencillos.

El valor mínimo que la centralidad de intermediación de un conjunto U con k elementos puede alcanzar es

$$\mathrm{bc}(U) = k^2 + 2k(n-k).$$

Este mínimo es 0, si consideramos la definición alternativa. Tal mínimo se alcanza cuando los vecinos de U, $N(U)$, forman un grafo completo.

Por otro lado, el valor máximo de la centralidad de intermediación de algún conjunto U con k elementos es n^2 (1 si consideramos la centralidad normalizada) y se alcanza cuando U se encuentra en el centro de una 'estrella'. Es decir, $\mathrm{bc}(U) = n^2$ si todo vértice $v \in V \setminus U$ solamente tiene como vecinos elementos de U.

En la presente sección solamente nos concentraremos en la centralidad de intermediación para un vértice, daremos una breve explicación del algoritmo propuesto en [2] y veremos algunas propiedades que serán de utilidad para calcular la centralidad de intermediación de subconjuntos.

Una observación fácilmente verificable que nos será de gran ayuda es la siguiente.

Lema 3 (Criterio de Bellman)**.** *Un vértice $v \in V$ está en un camino de longitud mínima entre $s, t \in V$, si y solo si $d_G(s,t) = d_G(s,v) + d_G(v,t)$. En este caso, $\sigma(s,t;v) = \sigma(s,v) \cdot \sigma(v,t)$.*

El algoritmo se basa en poder calcular de manera óptima la *dependencia* de un vértice $s \in \mathrm{V}$ en un vértice $v \in \mathrm{V}$, definida como

$$\delta(s;v) := \sum_{t \in \mathrm{V}} \frac{\sigma(s,t;v)}{\sigma(s,t)}.$$

De la definición, $\mathrm{bc}(v) = \sum_{s \in \mathrm{V}} \delta(s;v)$. Así pues, el siguiente teorema nos permitirá calcular de manera eficiente las dependencias, pero antes consideremos el conjunto de *sucesores*:

$$S(s;v) = \{u \in \mathrm{V} \; : \; \{u,v\} \in E, \mathrm{d}_G(s,u) = \mathrm{d}_G(s,v) + 1\}.$$

Es decir, $S(s;v)$ será el conjunto de vecinos de v para los cuales existe al menos un camino de longitud mínima desde s que pasa por v.

Teorema 4. *Dados $s, v \in V$, se tiene que*

$$\delta(s; v) = 1 + \sum_{w \in S(s;v)} \frac{\sigma(s,v)}{\sigma(s,w)} \cdot \delta(s; w).$$

Demostración. Para cada $w \in S(s; v)$ y para cualquier $t \in \mathrm{V} \setminus \{v\}$, definamos $\sigma(s, t; v, w)$ como el número de geodésicas de s a t que pasan por v y w. Además, notemos que si $\sigma(s, t; v, w) > 0$ entonces

$$\sigma(s, t; v, w) = \sigma(s, v) \cdot \sigma(v, w) \cdot \sigma(w, t) = \sigma(s, v) \cdot \sigma(w, t).$$

Afirmación 1.

$$\sigma(s, t; v, w) = \frac{\sigma(s,v)}{\sigma(s,w)} \cdot \sigma\left(s, t; w\right).$$

Demostración. Si $\sigma(s, t; w) = 0$, la afirmación se sigue trivialmente. En otro caso (véase Figura 1.3), se tiene que

$$\frac{\sigma(s,v)}{\sigma(s,w)} \cdot \sigma\left(s, t; w\right) = \frac{\sigma(s,v)}{\sigma(s,w)} \cdot \sigma(s, w) \cdot \sigma(w, t) = \sigma(s, v) \cdot \sigma(w, t) = \sigma(s, t; v, w).$$

□

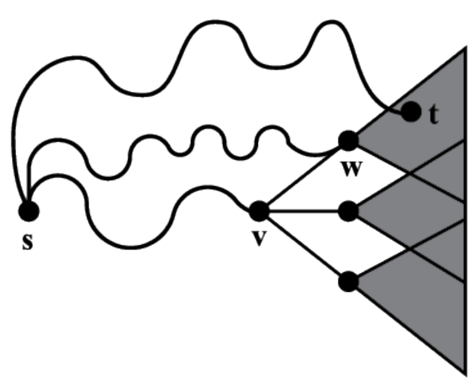


Figura 1.3: Todas las geodésicas pasan por un sucesor.

Trivialmente, para cualquier geodésica p de s a t que pase por v, debe existir un $w \in S(s;v)$ tal que $w \in p$. Es decir,

$$\delta(s;v) = \frac{\sigma(s,v;v)}{\sigma(s,v)} + \sum_{t \in \mathrm{V} \setminus \{v\}} \frac{\sigma\,(s,t;v)}{\sigma(s,t)} = 1 + \sum_{t \in \mathrm{V} \setminus \{v\}} \sum_{w \in S(s;v)} \frac{\sigma(s,t;v,w)}{\sigma(s,t)} =$$

$$= 1 + \sum_{w \in S(s;v)} \sum_{t \in \mathrm{V} \setminus \{v\}} \frac{\sigma(s,t;v,w)}{\sigma(s,t)}.$$

Por lo tanto,

$$\delta(s;v) = 1 + \sum_{w \in S(s;v)} \sum_{t \in \mathrm{V} \setminus \{v\}} \frac{\sigma(s,v)}{\sigma(s,w)} \cdot \frac{\sigma\,(s,t;w)}{\sigma(s,t)} =$$

$$= 1 + \sum_{w \in S(s;v)} \frac{\sigma(s,v)}{\sigma(s,w)} \cdot \sum_{t \in \mathrm{V} \setminus \{v\}} \frac{\sigma\,(s,t;w)}{\sigma(s,t)} = 1 + \sum_{w \in S(s;v)} \frac{\sigma(s,v)}{\sigma(s,w)} \cdot \delta(s;w). \quad \square$$

Corolario 5. *Dado $s \in V$, podemos calcular $\delta(s;v)$ para todo $v \in V$ en tiempo $O(n+m)$ y espacio $O(n+m)$.*

Demostración. Primero, haciendo una búsqueda en anchura, para cada $v \in \mathrm{V}$ calculamos la distancia a s y $\sigma(s,v)$. También, para cada $v \in \mathrm{V}$, almacenaremos una lista con sus predecesores, es decir, los vértices $w \in \mathrm{V}$ tales que $v \in S(s;w)$. Tal búsqueda se puede hacer en tiempo $O(n+m)$.

Haciendo un recorrido en orden decreciente de acuerdo a la distancia con s, podemos calcular $\delta(s;v)$ para cada $v \in \mathrm{V}$ de la siguiente manera: para un vértice $v \in \mathrm{V}$ con distancia máxima tendremos $\delta(s;v) = 1$; para el resto de los vértices, hacemos uso del Teorema 4. Tal proceso se puede hacer en tiempo $O(n+m)$, pues habrá que recorrerse todos los vértices y sus predecesores para utilizar el Teorema 4.

$\square$

Corolario 6. *La centralidad de intermediación de cada veŕtice de un grafo se puede calcular en tiempo $O(n(n+m))$.*

Se implementa el Algoritmo 1, donde se utiliza el Corolario 5 para cada vértice $s \in \mathrm{V}$.

Algoritmo 1 Algoritmo para calcular bc(v).

Para cada $v \in$ V, $BC[v] \leftarrow 0$
para todo $s \in$ V **hacer**
$S \leftarrow$ pila vacía
para cada $w \in$ V, $P[w] \leftarrow$ lista vacía
para cada $t \in$ V, $\sigma[t] \leftarrow 0$; $\sigma[s] \leftarrow 1$
para cada $t \in$ V, $d[t] \leftarrow -1$; $d[s] \leftarrow 0$
$Q \leftarrow$ cola vacía
insertar $s \rightarrow Q$
mientras Q no sea vacía **hacer**
sacar $v \leftarrow Q$
insertar $v \rightarrow S$
para todo w vecino de v **hacer**
si $d[w] < 0$ **entonces**
insertar $w \rightarrow Q$
$d[w] \leftarrow d[v] + 1$
fin si
si $d[w] = d[v] + 1$ **entonces**
$\sigma[w] \leftarrow \sigma[w] + \sigma[v]$
insertar $v \rightarrow P[w]$
fin si
fin para
fin mientras
para cada $v \in$ V, $\delta[v] \leftarrow 1$
mientras S no sea vacía **hacer**
sacar $w \leftarrow S$
para $v \in P[w]$ **hacer**
$\delta[v] \leftarrow \delta[v] + \frac{\sigma[v]}{\sigma[w]} \cdot \delta[w]$
fin para
$BC[w] \leftarrow BC[w] + \delta[w]$
fin mientras
fin para

Capítulo 2

Un algoritmo eficiente

En la presente sección introduciremos un algoritmo para encontrar el subconjunto de tamaño k con mayor centralidad de intermediación. Un algoritmo de fuerza bruta podría calcular la centralidad de intermediación de un subconjunto U en tiempo $O(n^2(n+m))$ y tomaría tiempo $O(n^{k+2}(n+m))$ encontrar el subconjunto con mayor centralidad de intermediación, pues deberíamos calcular la centralidad de intermediación de cada subconjunto.

Usando información previamente calculada por el Algoritmo 1, podemos calcular la centralidad de intermediación de un subconjunto en tiempo $O(n)$. Aunque para encontrar el subconjunto con mayor centralidad de intermediación no podemos evitar una búsqueda sobre todos los subconjuntos, siendo el algoritmo en tiempo $O(n^{k+1})$, al final de la sección presentamos una poda que reduce la búsqueda de manera significativa.

Ahora, denotemos como $\overline{\Upsilon}(s,t;U)$ al conjunto de caminos de s a t que pasan por todos los elementos de U. Análogamente, definimos $\overline{\sigma}\,(s,t;U) = \left|\overline{\Upsilon}(s,t;U)\right|$.

Teorema 7. *Dado $U \subseteq V$ con k elementos, se tiene que*

$$\sigma\,(s,t;U) = \sum_{i=1}^{k} (-1)^{i+1} \cdot \sum_{W \in \binom{U}{i}} \overline{\sigma}\,(s,t;W)\,,$$

para cualesquiera $s,t \in V$.

Demostración. Primero, notemos que dados $U \subseteq \mathrm{V}$ y $u \in \mathrm{V}$, se tiene que

$$\overline{\Upsilon}(s,t;U) \cap \Upsilon(s,t;u) = \overline{\Upsilon}(s,t;U \cup \{u\}).$$

Ademas, se tiene que $\Upsilon(s,t;U) = \bigcup_{u\in U} \Upsilon(s,t;u)$.

Por el principio de inclusión-exclusión, se tiene que

$$|\Upsilon(s,t;U)| = \sum_{i=1}^{k} (-1)^{i+1} \cdot \sum_{W\in\binom{U}{i}} \left| \bigcap_{u\in W} \Upsilon(s,t;u) \right|,$$

es decir,

$$\sigma(s,t;U) = \sum_{i=1}^{k} (-1)^{i+1} \cdot \sum_{W\in\binom{U}{i}} \overline{\sigma}(s,t;W). \quad \square$$

Definimos la *dependencia* de s en un subconjunto U como $\delta(s;U) = \sum_{t\in \mathrm{V}} \frac{\sigma(s,t;U)}{\sigma(s,t)}$, para $U \subseteq \mathrm{V}$ y $s \in \mathrm{V}$. Entonces podemos escribir la centralidad de intermediación como

$$\mathrm{bc}(U) = \sum_{s\in \mathrm{V}} \delta(s;U).$$

Por otro lado, usando el Teorema 7 e intercambiando sumas, podemos calcular la dependencia $\delta(s;U)$ de la siguiente manera

$$\delta(s;U) = \sum_{i=1}^{k} (-1)^{i+1} \cdot \sum_{W\in\binom{U}{i}} \left(\sum_{t\in \mathrm{V}} \frac{\overline{\sigma}(s,t;W)}{\sigma(s,t)} \right).$$

Definamos $\overline{\delta}(s;U) := \sum_{t\in \mathrm{V}} \frac{\overline{\sigma}(s,t;W)}{\sigma(s,t)}$. Ahora, nos concentraremos en calcular $\overline{\delta}(s;U)$ de manera rápida.

Fijemos $s,t \in \mathrm{V}$ y $U = \{v_0, ..., v_{k-1}\} \subseteq \mathrm{V}$. Si existe un camino $p \in \overline{\Upsilon}(s,t;U)$ no pueden existir dos vértices de U con la misma distancia a s. Además, tal camino debe recorrer a U en el orden dado por la distancia de s a cada $u \in U$. Es decir, si definimos $\overrightarrow{U_s} = (v'_0, ..., v'_{k-1})$ como el vector de vértices tal que

$$\{v_0, ..., v_{k-1}\} = \{v'_0, ..., v'_{k-1}\} \text{ y}$$
$$0 \leq i < j < k \Rightarrow \mathrm{d}_G(s, v'_i) < \mathrm{d}_G(s, v'_j),$$

entonces p debe recorrer a U en el orden dado por $\overrightarrow{U_s}$.

En particular, si $w \in U$ es tal que, para todo $u \in U$, $\mathrm{d}_G(s,u) \leq \mathrm{d}_G(s,w)$, entonces w debe ser el último vértice de U en aparecer en p. Aún más, si definimos $U^{-} = U \setminus \{w\}$, usando el Criterio de Bellman podemos verificar la siguiente ecuación fácilmente (véase Figura 2.1):

$$\overline{\sigma}\,(s,t;U) = \begin{cases} \overline{\sigma}\,(s,w;U^{-}) \cdot \sigma(w,t) & \text{si } \mathrm{d}_G(s,t) = \mathrm{d}_G(s,w) + \mathrm{d}_G(w,t), \\ 0 & \text{de lo contrario.} \end{cases} \tag{2.1}$$

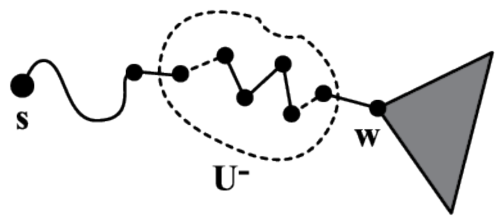


Figura 2.1: U se debe recorrer en orden.

Por lo tanto, una vez calculado $\mathrm{d}_G(s,t)$ y $\sigma(s,t)$, para todo $s,t \in \mathrm{V}$, podremos calcular de forma recursiva $\overline{\sigma}\,(s,t;U)$ haciendo uso de la ecuación (2.1). Es decir, una vez calculado $\overline{\sigma}\,(s,w;U^{-})$, podremos calcular $\overline{\sigma}\,(s,t;U)$ efectuando solamente una operación.

El siguiente teorema nos dice que, una vez calculado $\overline{\sigma}\,(s,w;U^{-})$ y $\delta(s;w)$, podemos calcular $\overline{\delta}(s;U)$ efectuando solamente una operación.

Teorema 8. *Sean $s \in V$ y $U \subseteq V$. Tomemos $w \in U$ tal que, para todo $u \in U$, $d_G(s,u) \leq d_G(s,w)$ y definamos $U^{-} = U \setminus \{w\}$, entonces*

$$\overline{\delta}(s;U) = \delta(s;w) \cdot \frac{\overline{\sigma}\,(s,w;U^{-})}{\sigma(s,w)}.$$

Demostración. Definamos $V^0 = \{t \in \mathrm{V} \mid \mathrm{d}_G(s,t) = \mathrm{d}_G(s,w) + \mathrm{d}_G(w,t)\}$ y $V^1 = V \setminus V^0$.

Para $t \in V^1$, por el Criterio de Bellman tenemos que $\overline{\sigma}(s,t;U) = 0$ y $\sigma(s,t;w) = 0$. En este caso trivialmente se tiene que

$$\overline{\sigma}(s,t;U) = \sigma(s,t;w) \cdot \frac{\overline{\sigma}(s,w;U^-)}{\sigma(s,w)}.$$

Para $t \in V^0$, multiplicamos por $\frac{\sigma(s,w)}{\sigma(s,w)}$ y usando la ecuación (2.1), se tiene que

$$\overline{\sigma}(s,t;U) = \frac{\sigma(s,w)}{\sigma(s,w)} \cdot \overline{\sigma}\left(s,w;U^-\right) \cdot \sigma(w,t) =$$

$$= \sigma(s,w) \cdot \sigma(w,t) \cdot \frac{\overline{\sigma}(s,w;U^-)}{\sigma(s,w)} = \sigma(s,t;w) \cdot \frac{\overline{\sigma}(s,w;U^-)}{\sigma(s,w)}.$$

Por definición de $\delta(s;w)$, se sigue el teorema. □

Recordemos que, por el Teorema 5, el algoritmo de centralidad de intermediación para un vértice encuentra $\delta(s;v)$ para cualesquiera $s,v \in$ V. Una vez ejecutado este algoritmo, podemos almacenar cada $\delta(s;v)$ y $\sigma(s,v)$. Usando la ecuación (2.1) y el Teorema anterior, podemos aplicar el Algoritmo 2 (donde ultimo$(s;W)$ es un vértice $w_0 \in W$ tal que $\mathrm{d}_G(s,w) \leq \mathrm{d}_G(s,w_0)$ para todo $w \in W$) y entonces tenemos el siguiente corolario.

Corolario 9. *Dado $s \in V$ y $U \subseteq V$, si para cualesquier par de vértices $s,t \in V$ ya hemos encontrado $\delta(s;v)$ y $\sigma(s,t)$, entonces podemos encontrar $\delta(s;U)$ en tiempo $O(1)$.*

Notemos que el Algoritmo 2 corre en tiempo $O(1)$, pues k es constante. Por otro lado, para calcular $\mathrm{bc}(U)$, hay que calcular $\delta(s;U)$ para cada $s \in$ V. Usando el Algoritmo 3 se sigue el siguiente resultado.

Corolario 10. *Sea $U \subseteq V$ y supongamos que para cualesquiera $s,t,v \in V$ ya hemos encontrado bc$(s;v)$ y $\sigma(s,t)$, entonces podemos encontrar bc(U) en tiempo $O(n)$.*

Finalmente, para encontrar el subconjunto con k elementos con mayor centralidad de intermediación, tendríamos que calcular la centralidad de intermediación para todos los posibles subconjuntos. Sin embargo, podemos hacer una poda a la búsqueda que funciona muy bien en la práctica. Notemos que $\sigma(s,t;U) \leq \sum_{u \in U} \sigma(s,t;u)$. En efecto, cualquier camino que contemos en $\sigma(s,t;U)$ debe pasar por algún elemento $u \in U$, tal camino lo contaremos en $\sigma(s,t;u)$. De aquí es inmediato el siguiente teorema.

Algoritmo 2 Algoritmo para calcular $\delta(s;U)$.

Entrada: Un vértice s y un subconjunto $U \subseteq \mathrm{V}$

Salida: $\delta(s;U)$

$\delta \leftarrow 0$
para todo $i \in \{1,2,..,k\}$ **hacer**
  para todo $W \in \binom{U}{i}$ **hacer**
    $w_0 \leftarrow \mathrm{ultimo}(s;W)$
    $w_1 \leftarrow \mathrm{ultimo}(s;W \setminus \{w_0\})$
    $\overline{\sigma}(s,w_0;W \setminus \{w_0\}) \leftarrow 0$
    si $\mathrm{d}_G(s,w_0) = \mathrm{d}_G(s,w_1) + \mathrm{d}_G(w_0,w_1)$ **entonces**
      $\overline{\sigma}(s,w_0;W \setminus \{w_0\}) \leftarrow \overline{\sigma}(s,w_1;W \setminus \{w_0,w_1\}) \cdot \sigma(w_1,w_0)$
    fin si
    $\delta \leftarrow \delta + \delta(s;w_0) \cdot \frac{\overline{\sigma}(s,w_0;W \setminus \{w_0\})}{\sigma(s,w_0)}$
  fin para
fin para
devolver δ

Algoritmo 3 Algoritmo para calcular $\mathrm{bc}(U)$.

Entrada: Un subconjunto $U \subseteq V$

Salida: $\mathrm{bc}(U)$

$sum \leftarrow 0$
para s **in** V **hacer**
  $sum \leftarrow sum + \delta(s;U)$
fin para
devolver sum

Teorema 11. *Para todo $U \subseteq V$, se tiene que*

$$bc(U) \leq \sum_{u \in U} bc(u).$$

Una vez ejecutado el Algoritmo 1, deberemos almacenar $\mathrm{d}_G(s,v), \mathrm{bc}(s;v)$ y $\sigma(s,v)$ para cada $s, v \in$ V. Luego, ordenamos V de mayor a menor de acuerdo a su centralidad de intermediación. Una vez que empecemos a calcular la centralidad de intermediación de subconjuntos, usando el teorema anterior, podemos ir eliminando aquellos subconjuntos U tales que la suma de la centralidad de intermediación de sus elementos es menor que la mayor centralidad de intermediación calculada hasta el momento.

Capítulo 3

Implicaciones prácticas

Se hizo un análisis sobre todos los grafos conexos de 5 a 9 vértices. Para cada grafo obtuvimos la pareja con más centralidad de intermediación, digamos Max, y también obtuvimos los dos vértices con más centralidad de intermediación, digamos v_0, v_1, donde $\mathrm{bc}(v_0) \geq \mathrm{bc}(v_1)$. Para los grafos tales que $\mathrm{bc}(Max) > \mathrm{bc}(\{v_0, v_1\})$ definimos tres tipos distintos de grafos:

1. **Tipo 1:** Son aquellos grafos tales que $Max \neq \{v_0, v_1\}$,

2. **Tipo 2:** Son aquellos grafos tales que $v_0 \notin Max$,

3. **Tipo 3:** Son los grafos tales que $Max \cap \{v_0, v_1\} = \emptyset$.

Claramente, **Tipo 3**$\subseteq$**Tipo 2**$\subseteq$**Tipo 1**. En la siguiente tabla están los resultados que obtuvimos, donde se muestra el número total de grafos y el número de grafos de cada tipo.

# vértices	Total de grafos	**Tipo 1**	**Tipo 2**	**Tipo 3**
5	21	2	1	0
6	112	25	8	0
7	853	283	71	4
8	11117	4123	826	42
9	261080	101078	17238	639

Análogamente, se hizo el mismo experimento para subconjuntos de tamaño 3. Para cada grafo, sea Max el subconjunto con 3 vértices con mayor centralidad de intermediación. Así mismo, sean v_0, v_1, v_2 los vértices con más centralidad de intermediación de tal modo que $\mathrm{bc}(v_0) \geq \mathrm{bc}(v_1) \geq \mathrm{bc}(v_2)$. Para los grafos tales que $\mathrm{bc}(Max) > \mathrm{bc}(\{v_0, v_1, v_2\})$ definimos los tres mismos tipos de grafos:

1. **Tipo 1:** Son aquellos grafos tales que $Max \neq \{v_0, v_1, v_2\}$,
2. **Tipo 2:** Son aquellos grafos tales que $v_0 \notin Max$,
3. **Tipo 3:** Son los grafos tales que $Max \cap \{v_0, v_1, v_2\} = \emptyset$.

En la siguiente tabla se resumen los resultados encontrados.

# vértices	Total de grafos	**Tipo 1**	**Tipo 2**	**Tipo 3**
6	112	21	3	0
7	853	313	39	0
8	11117	5575	693	3
9	261080	149247	16686	62

Notemos que para los grafos con 9 vértices cerca del 57 % son de **Tipo 1**. Es decir, aunque son muy pocos los grafos para los cuales $Max \cap \{v_0, v_1, v_2\} = \emptyset$, hay muchos grafos en los que el subconjunto con más centralidad de intermediación es distinto a tomar los vértices con más centralidad de intermediación.

Los grafos anteriores se obtuvieron de una recopilación hecha por Brendan McKay, que puede encontrarse en:

`http://users.cecs.anu.edu.au/~bdm/data/graphs.html`

En las siguientes secciones analizamos los grafos de distintas redes. El primer experimento fue sobre las avenidas de la ciudad de Morelia, Michoacán, México. Los siguientes dos experimentos se realizaron sobre las redes sociales de los personajes de una novela y de una serie animada. Usando los algoritmos de centralidad de intermediación, encontramos los actores más importantes, así como los subconjuntos de actores más importantes.

3.1. Análisis de una ciudad: Morelia

El primer experimento que realizamos se hizo sobre el grafo creado por las calles de Morelia, Michoacán, México, una ciudad de aproximadamente 100 km^2 y 600 mil habitantes.

Para crear el grafo, se uso la información proporcionada por:

https://www.openstreetmap.org

Descargamos el mapa de Morelia de la página anterior y seleccionamos solamente las calles etiquetadas como "primary", "trunk", "secondary" y "tertiary". Es decir, solamente aquellas de más flujo e importancia. Los vértices de nuestro grafo son las intersecciones de dichas calles. El resultado fue un grafo con 1370 vértices y 2829 aristas.

En la Figura 3.1 se muestra un mapa de Morelia con los 6 puntos más importantes según la centralidad de intermediación.

Figura 3.1: Mapa de Morelia, Michoacán con los puntos con más centralidad de intermediación. Los vértices tienen los siguientes valores: $\mathrm{bc}(1) = 67300{,}6, \mathrm{bc}(2) = 64322{,}4, \mathrm{bc}(3) = 48964, \mathrm{bc}(4) = 48953, \mathrm{bc}(5) = 48136{,}5, \mathrm{bc}(6) = 45743{,}4$.

El vértice 2 parece ser una inconsistencia. La razón se debe a que consideramos como vértices las intersecciones de calles con mayor flujo y usamos aristas sin peso. Como puede verse en la imagen, se puede llegar al vértice dos a través de una o dos aristas desde Avenida Camelinas, una de las avenidas más importantes y que le da la vuelta a Morelia. Por esta razón, el vértice 2 es central en nuestro grafo, aunque en la práctica este punto no sea central.

Respecto a la centralidad de subconjuntos se obtuvieron los siguientes resultados. La pareja con más centralidad de intermediación fue $\{1, 2\}$ con 131536. Para $k = 3$, existe una sorpresa, pues el conjunto con más centralidad de intermediación es $\{1, 2, 5\}$. Es decir, se deja fuera a los puntos 3 y 4. Finalmente, para $k = 4$, el conjunto más importante es $\{1, 2, 3, 5\}$, dejando fuera al punto 4.

3.2. Análisis de personajes en una novela

Game of Thrones es una serie de televisión de fantasía medieval basada en la serie de novelas *Canción de hielo y fuego*. La historia detrás de las novelas se torna muy compleja y tiene una gran cantidad de personajes, volviéndose difícil decidir quién o quiénes son los verdaderos protagonistas de la serie. En [1], los autores crean un grafo para descubrir a los protagonistas de la serie usando varias medidas de centralidad, entre ellas, centralidad de intermediación.

El grafo creado por los A. Beveridge y Jie Shan está basado en un análisis de la novela *Tormenta de espadas*, de la serie Canción de hielo y fuego. La siguiente tabla muestra los diez personajes más importantes, según la centralidad de intermediación.

	Nombre	$bc(v)$
1	Jon	2772.51
2	Robert	2544.21
3	Tyrion	2415.77
4	Daenerys	1962.67
5	Robb	1626.11
6	Sansa	1623.4
7	Stannis	1356.05
8	Jaime	1325.37
9	Arya	1099.03
10	Tywin	942.442

Utilizando nuestro algoritmo, evaluamos el grafo anterior para descubrir a los subconjuntos de personajes más importantes.

Para $k = 2$, obtuvimos que la pareja más importante es Jon y Tyrion, es decir, se deja fuera al segundo lugar, Robert. Encontramos que bc({Jon,Tyrion}) = 5026,5, contra bc({Jon,Robert}) = 4574,92.

Por otro lado, para $k = 3$, el conjunto más importante es el conformado por los tres personajes más importantes: Jon, Robert y Tyrion. Su centralidad de intermediación es 6569.43.

Existe otra sorpresa para $k = 4$, donde el subconjunto más importante es el conformado por Jon, Robert, Tyrion y Jaime, dejando fuera a Daenerys, Robb, Sansa y Stannis. Es decir, lo conforman los lugares $\{1, 2, 3, 7\}$.

Para $k = 5$, el subconjunto más influyente lo conforma Jon, Robert, Tyrion, Robb y Jaime. Para $k = 6$, {Jon, Robert, Tyrion, Robb, Sansa, Jaime} es el subconjunto con más centralidad de intermediación. Al parecer, Daenerys y Stannis no aportan mucha centralidad al conjunto conformado por Jon, Robert y Tyrion.

3.3. Análisis de una serie animada

One Piece es un manga japonés publicado desde 1997 y después llevado a una versión anime. One Piece es uno de los mangas más extensos de la historia con una gran cantidad de personajes. Cuenta con 82 volúmenes de manga y más de 750 capítulos en su versión de Anime.

Realizamos un experimento para descubrir a los personajes más importantes de la serie según la centralidad de intermediación y la centralidad de intermediación para subconjuntos.

Usamos la información de la página: `http://onepiece.wikia.com/` para construir el grafo. Los vértices del grafo son los personajes 'In Canon' que se encuentran en la página anterior, siendo un total de 837 personajes. Estos personajes son aquellos que aparecen o han sido mencionados en el manga. Es decir, se excluyen los personajes que aparecen en el anime, películas o especiales, pero no han aparecido en el manga.

Para crear las relaciones hicimos un análisis de la descripción de cada personaje. Usando Python, hicimos la extracción del código fuente de la descripciones y creamos una arista entre dos personajes si uno era mencionado en la descripción del otro en cualquier parte antes de la sección 'References'. El resultado final fue un grafo con 837 vértices y 7941 aristas.

Los resultados no tuvieron ninguna sorpresa, pues los personajes con más centralidad de intermediación son el protagonista y aquellos muy cercanos a él.

	Nombre	$bc(v)$
1	Monkey D. Luffy	252667
2	Roronoa Zoro	45543.5
3	Nami	42511.5
4	Sanji	40395.6
5	Portgas D. Ace	39687.1
6	Usopp	36554.2
7	Nico Robin	36516.2
8	Donquixote Doflamingo	28963.7
9	Edward Newgate	23811.9
10	Franky	21999.9

Respecto a la centralidad de intermediación de subconjuntos, la siguiente tabla muestra los subconjuntos con mayor centralidad de intermediación para $2 \leq k \leq 5$. Habiendo numerado los personajes de acuerdo a su centralidad de intermediación, en Subconjunto se muestran los subconjuntos de personajes con más centralidad de intermediación.

k	Subconjunto	$bc(U)$
2	{1,2}	296198
3	{1,2,4}	335588
4	{1,2,3,4}	373195
5	{1,2,3,4,6}	408272

Cabe resaltar que el subconjunto de tres personajes con más centralidad de intermediación es {Luffy, Zoro, Sanji}, dejando fuera a Nami. Así mismo, para $k = 5$, el subconjunto con más centralidad de intermediación es {Luffy, Zoro, Nami, Sanji, Usopp} dejando fuera al número 5, Portgas D. Ace.

Bibliografía

[1] A. Beveridge, J. Shan, *Network of Thrones*, Math Horizons Magazine 23(4):18-22, 2016.

[2] U. Brandes, *A Faster Algorithm for Betweenness Centrality*, Journal of Mathematical Sociology 25(2):163-177, 2001.

[3] D. Easley, J. Kleinberg, *Networks, Crowds and Markets*, Cambridge University Press, Cambridge, 2010.

[4] M. Newman, *Networks; An Introduction*, Oxford University Press, Oxford, 2010.

[5] J. Scott, *Social Network Analysis: A Handbook*, Sage Publications, 1991.

[6] S. Wasserman, K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, 1994.