Universidad de Oviedo

EL CONCEPTO DE VALOR SEMÁNTICO EN EL ANÁLISIS DE REDES SOCIALES

José Alejandro Ares Rodríguez

Dirigido por
Susana Montes Rodríguez y Noelia Rico Pachón

UNIVERSIDAD DE OVIEDO
Facultad de Ciencias
Grado en Matemáticas

Julio de 2022

Índice general

Capítulo 1

Introducción

1.1. Motivación

La teoría de grafos tiene aplicaciones en todas las ciencias y fuera de ellas, incluso está presente en nuestra vida diaria. Por ejemplo, un mapa de carreteras es, básicamente, un grafo. Un navegador que encuentra el camino más rápido entre origen y destino utiliza internamente una técnica que realiza una sucesión de aristas sobre el grafo. Son además una herramienta esencial en la extracción de información de las redes sociales y este sería uno de los usos más habituales en nuestros días. No obstante, en este trabajo se estudia una nueva aplicación que ha surgido recientemente, en la que se utilizan los grafos para analizar textos.

Se estudiará la manera de convertir un texto (ya sea un libro, un periódico, un mensaje de texto, etc.) en una red social. Para ello, se considerará que las palabras son los individuos que se relacionan entre ellos y se verá cómo esta metodología permitirá conocer los temas principales de un texto, sin necesidad de leerlo. Para ello, se introducirá el concepto de valor semántico, dividido en valor extrínseco e intrínseco. Estos valores están muy presentes en la rama de la filosofía. En ética, el valor intrínseco es una propiedad de todo lo que es valioso por sí mismo. El valor intrínseco contrasta con el valor instrumental (también conocido como valor extrínseco), que es una propiedad

de todo lo que deriva su valor de una relación con otra cosa intrínsecamente valiosa. El valor intrínseco es siempre algo que un objeto tiene "en sí mismo" o "por sí mismo", y es una propiedad intrínseca. Un objeto con valor intrínseco puede considerarse un fin o, en terminología kantiana, un fin en sí mismo. El objeto con valor intrínseco, el fin, puede ser tanto un objeto concreto como un objeto abstracto. El valor extrínseco es lo que hace valioso a un objeto como medio. Por ejemplo, el dinero por sí mismo no tiene valor, pero es valioso porque permite la adquisición de otros objetos.

Así pues, el objetivo fundamental de este trabajo es el análisis de los conceptos matemáticos necesarios para aplicar teoría de grafos en el análisis de textos. Los objetivos concretos del mismo aparecen detallados a continuación.

1.2. Objetivos

- Conocer la teoría de grafos; entender la diferencia entre un grafo orientado y no orientado y las características de cada uno de ellos.
- Introducir las redes de flujo y saber resolver un problema de flujo máximo.
- Entender cómo se modelan las redes sociales, qué son las medidas de centralidad, y las funciones de afinidad que relacionan los individuos de dicha red.
- Introducir el concepto de valor semántico de una palabra en un texto, formado por los valores extrínseco e intrínseco.
- Saber modelar una red a partir de un texto, en la que los individuos sean palabras, y crear un método para calcular las relaciones entre ellas.
- Crear una nueva función de afinidad que relacione palabras, basada en el valor semántico.
- Aplicar los resultados a un texto y analizar los resultados.

1.3. Estructura del trabajo

Derivado de todo lo anterior, este trabajo se estructura como sigue.

En el Capítulo 2 se introducen los conceptos básicos necesarios para comprender el resto del trabajo y se fija la notación utilizada. Se definirá el concepto de grafo y otros conceptos relacionados con la teoría de grafos. Se explicará también lo que es un problema de flujo y cómo resolverlo. Y, finalmente, se definirán los distintos tipos de medidas de centralidad y funciones de afinidad.

En el Capítulo 3 se desarrolla el concepto de valor semántico para el análisis de redes sociales, analizando la importancia del mismo. Se explicará como hallar el valor semántico y a partir de él, se creará una nueva función de afinidad; llamada afinidad semántica.

En el Capítulo 4 se presentará un ejemplo de aplicación. Se analizará un texto convirtiéndolo en una red. Se calcularán las distintas centralidades y afinidades explicadas en el Capítulo 2 y la afinidad semántica definida en el Capítulo 3, y, finalmente se analizarán los resultados numéricos.

Por último, en el Capítulo 5 aparecen las conclusiones finales del trabajo y las posibles futuras líneas de investigación relacionadas con él.

Capítulo 2

Conceptos básicos

2.1. Teoría de grafos

Gracias a la teoría de grafos se pueden resolver diversos problemas como por ejemplo la síntesis de circuitos secuenciales, contadores o sistemas de apertura. Se utiliza para diferentes áreas, por ejemplo, dibujo computacional, ingeniería, modelaje de trayectos como el de una línea de autobús a través de las calles de una ciudad, en el que podemos obtener caminos óptimos para el trayecto aplicando diversos algoritmos.

La teoría de grafos ha servido de inspiración para las ciencias sociales, en especial para desarrollar un concepto no metafórico de red social que sustituye los nodos por los actores sociales y verifica la posición, centralidad e importancia de cada actor dentro de la red. Esta medida permite cuantificar y abstraer relaciones complejas, de manera que la estructura social puede representarse gráficamente. Por ejemplo, una red social puede representar la estructura de poder dentro de una sociedad al identificar los vínculos, su dirección e intensidad y da idea de la manera en que el poder se transmite y a quiénes. Los grafos son importantes también en el estudio de la biología y el hábitat. El vértice representa un hábitat y las aristas representan los senderos de los animales o las migraciones. Con esta información, los científicos pueden entender cómo esto puede cambiar o afectar a las especies en su

hábitat. En matemáticas, la teoría de grafos estudia las propiedades fundamentales asociadas a los grafos y como obtener a partir de ellas algoritmos que permitan resolver problemas reales.

Introducimos a continuación los conceptos básicos sobre teoría de grafos que serán necesarios para comprender nuestro trabajo y permitirán además fijar la notación empleada en el mismo.

Definición 2.1. *Un* **grafo** *es un par $G = (V, T)$ donde V es un conjunto de puntos que representan elementos cualesquiera con $V \neq \emptyset$ y T es un conjunto de pares de elementos de V.*

Definición 2.2. *Los elementos de V se llaman* **vértices** *o* **nodos**.

Nota 2.1. *T es un multiconjunto, puesto que cualquiera de sus elementos puede estar más de una vez. V se puede interpretar como un conjunto de elementos que pueden o no estar relacionados entre sí, viniendo dicha relación expresada por T.*

Podemos distinguir entre dos tipos de grafos dependiendo de si los pares de elementos del conjunto T están ordenados o no. Si dicho orden es relevante en el estudio del grafo, diremos que trabajamos con un grafo orientado o dirigido. En caso contrario, diremos que nuestro grafo es no orientado o no dirigido.

Definición 2.3. *$G = (V, T)$ es un* ***grafo orientado*** *o* ***dirigido****, cuando influye el orden de los pares, es decir, si T es un conjunto de pares ordenados de elementos de V, es decir, $(i, j) \neq (j, i)\ \forall i, j \in V$. Con lo cual, $A \subseteq V \times V$.*

Un elemento cualquiera del conjunto T, $e = (i, j)$, se llama ***arco****, se representa gráficamente por $i \to j$ y los vértices i, j se llaman* ***vértice inicial*** *y* ***vértice final*** *del arco e, respectivamente.*

Algunos otros conceptos importantes en este trabajo, relacionados con los arcos son definidos a continuación.

Definición 2.4. *Sea $G = (V, T)$ un grafo orientado. Si $(i, j) \in T$, se dice que j es un* ***sucesor*** *de i y que i es un* ***antecesor*** *o* ***predecesor*** *de j.*

*Además, a los subconjuntos de V: $\Gamma(i) = \{j \in V \ : \ (i,j) \in T\}$ y $\Gamma^-(i) = \{j \in V \ : \ (j,i) \in T\}$ se les denomina **conjunto de sucesores y predecesores** de i, respectivamente, para cualquier vértice $i \in V$.*

Todos estos conceptos se van a describir en el siguiente ejemplo.

Ejemplo 2.2. *Si se considera el conjunto $V = \{1,2,3,4,5\}$ y el subconjunto de pares ordenados $T = \{(1,2),(3,1),(4,1),(2,4),(4,5)\} \subseteq V \times V$, se tiene que $G = (V,T)$ es un grafo orientado que puede representarse gráficamente tal como puede verse en la Figura 2.1.*

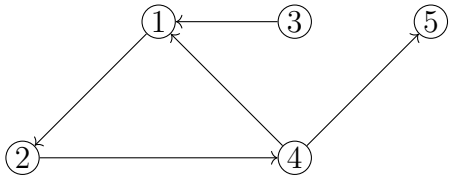


Figura 2.1: Ejemplo de grafo orientado.

Si consideramos un arco cualquiera, por ejemplo, $(3,1) \in T$, se tiene que 3 es el vértice inicial, y el predecesor del 1, que a su vez es el vértice final y el sucesor del 3. Además, $\Gamma(1) = \{2\}$ y $\Gamma^-(1) = \{3,4\}$.

De forma análoga, se define un grafo no orientado como aquel donde el orden de los pares de elementos en T no es importante. Así,

Definición 2.5. *$G = (V,T)$ es un **grafo no orientado** o **no dirigido** si T es un conjunto de pares no ordenados de elementos de V, con lo que $A \subseteq \mathcal{P}(V)$, en concreto de subconjuntos de V de cardinal 2.*

*Un elemento cualquiera del conjunto T, $e = \{i,j\}$. se llama **arista** o **eje**, se representa gráficamente por i — j y los vértices i, j se llaman **extremos**, **vértices** o **nodos** de la arista e.*

A partir de un grafo orientado, se puede definir un grafo no orientado asociado, sin más que eliminar la influencia del orden, tal como sigue.

Definición 2.6. *Dado un grafo orientado $G = (V,T)$, el **grafo subyacente** de G es un grafo no orientado $G^s = (V,T^s)$ donde T^s está formado por*

los pares no ordenados $\{i, j\}$ *tales que el par ordenado* $(i, j) \in T$ *o el par ordenado* $(j, i) \in T$*, es decir,*

$$T^s = \{\{i, j\} \in \mathcal{P}(V) \ : \ (i, j) \in T \ o \ (j, i) \in T\}$$

En ocasiones, si no hay ambigüedad, se denotan también a los elementos de T en los grafos no orientados como (i, j) en lugar de $\{i, j\}$, tal como se va a considerar en el siguiente ejemplo.

Ejemplo 2.3. *Dado el grafo orientado* $G = (V, T)$ *considerado en el ejemplo 2.2, el grafo subyacente asociado es* $G^s = (V, T^s)$ *con*

$$V = \{1, 2, 3, 4, 5\}$$

$$T^s = \{(1, 2), (1, 3), (1, 4), (2, 4), (4, 5)\}$$

cuya representación gráfica puede verse en la Figura 2.2.

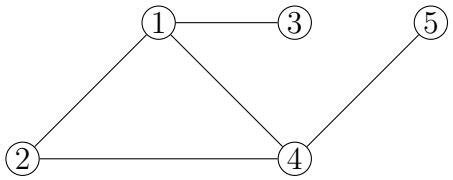


Figura 2.2: Ejemplo de grafo no orientado.

Este grafo no orientado tiene 5 aristas. Por ejemplo 1 *y* 3 *son los vértices o extremos de la arista* $(1, 3) \in T^s$*.*

En este trabajo se considerarán tanto grafos orientados como no orientados, por lo que es necesario conocer los conceptos relacionados con cada uno de ellos. Introduciremos a continuación algunas definiciones básicas comunes para ambos tipos de grafos.

Definición 2.7. *Sea* $G = (V, T)$ *un grafo (orientado o no orientado). Se tiene que:*

- *El cardinal de* V *se denota por* $o(G)$*, y se llama* **orden de** G*.*
- *El cardinal de* T *se denota por* $t(G)$ *y se llama* **tamaño de** G*.*

- *En el caso particular de que $o(G) = 1$ y $t(G) = 0$, se dice que $G = (V, T)$ es un* **grafo trivial**.

Ejemplo 2.4. *El orden y el tamaño en el grafo orientado del ejemplo 2.2 serían*

$$o(G) = 5, \quad t(G) = 5$$

También para el grafo no orientado asociado, visto en el ejemplo 2.3 se tiene el mismo orden y tamaño. En el caso del tamaño se tiene por definición que $t(G^s) \leq t(G)$. En el caso del orden la igualdad se verifica siempre, es decir, $o(G^s) = o(G)$.

Si consideramos $V = \{1\}$ y $T = \emptyset$, se tiene que (V, T) es un grafo trivial.

Otros conceptos importantes, que delimitan el tipo de grafos considerados en este trabajo, se muestran en la siguiente definición.

Definición 2.8. *Sea $G = (V, T)$ un grafo (orientado o no orientado). Se tiene que:*

- *Si V es un conjunto finito, se dice que G es un* **grafo finito**.
- *Si $(i, i) \in T$, se dice que el grafo tiene un* **bucle** *en i.*
- *Si T contiene a lo sumo p aristas con los mismos extremos i y j, se dice que G es un* **p-grafo**.
- *Un* ***grafo simple*** *es un 1-grafo sin bucles.*

En este trabajo vamos a trabajar siempre con grafos simples, denotando de forma genérica el orden y el tamaño por n y m, respectivamente, es decir, $o(G) = n$ y $t(G) = m$.

En el caso de que algunas conexiones sean más importantes que otras (por diferentes criterios), se les puede dar un valor a los arcos/aristas, normalmente conocido como peso, que cuantifique la fuerza de cierta conexión entre dos vértices. Varias métricas para grafos o redes sin aristas valoradas pueden ser adaptadas para grafos ponderados. Por ejemplo, las medidas de centralidad clásicas, que definiremos en este capítulo, también pueden considerar los pesos de las aristas.

Definición 2.9. *Un **grafo ponderado** o **red** (ordenado o no ordenado) se puede definir como una terna $G' = (V, T, p)$ donde (V, T) es un grafo (ordenado o no ordenado) y p es el conjunto de pesos asociados a cada arco/arista.*

Además de gráficamente, un grafo se puede representar de forma matricial, lo cual lo hace mucho más operativo, tanto desde el punto de vista matemático, como desde el computacional. Se crea una matriz cuyas columnas y filas representan los nodos del grafo. Por cada arista que une a dos nodos, se considera el valor 1 en la ubicación correspondiente de la matriz, siendo el resto de elementos ceros. Existe una matriz de adyacencia única para cada grafo (sin considerar las permutaciones de filas o columnas) y viceversa.

Definición 2.10. *La **matriz de adyacencia** A de un grafo G es una matriz cuadrada que representa las conexiones entre pares de vértices. Así, dados $i, j \in V$, se tiene que:*

$$a_{ij} = \begin{cases} 1 & si\ (i,j) \in T \\ 0 & si\ (i,j) \notin T \end{cases}$$

Es evidente que para un grafo no orientado la matriz de adyacencia es simétrica, pero que esta propiedad no es verifica, en general por los grafos ordenados.

Existe otra matriz asociada a un grafo, denominada matriz de distancias, que será de gran importancia a lo largo de este trabajo. Antes de introducirla es necesario recordar la definición de camino.

Definición 2.11. *Dado un grafo orientado/no orientado $G = (V, T)$ y dados $i_1, i_f \in V$. Un **camino** del vértice i_1 al vértice i_f es una sucesión alternada de vértices y arcos/aristas desde i_1 hasta i_f: $i_1, e_1, i_2, e_2, ..., e_{f-1}, i_f$, donde e_k denota el arco/arista $e_k = (i_k, i_{k+1}) \in T$.*

Al número de arcos/aristas de un camino se le llama **longitud del camino**.

La matriz de distancias se forma justo con las longitudes mínimas, es decir, con las longitudes de los caminos de menor longitud entre dos nodos. Para poder definirla es necesario la existencia de tales caminos, lo cual ocurre con un determinado tipo de grafos. En concreto:

Definición 2.12. *Un grafo no dirigido G se dice* ***conexo*** *si existe un camino entre cada par de vértices.*

Si el grafo es conexo, existe siempre al menos un camino entre cada par de vértices i y j. Al conjunto de todos ellos se le va a denotar por Π_{ij}.

Definición 2.13. *Un grafo orientado G es* ***débilmente conexo*** *si su grafo subyacente, G^s, es conexo.*

De todo lo anterior se deduce la definición de matriz de distancias para grafos conexos.

Definición 2.14. *Sea $G = (V, T)$ un grafo (ordenado o no ordenado) conexo. La* ***matriz de distancias*** *D del grafo G representa la longitud del camino de mínima longitud entre cada par de nodos. Así, dados dos vértices $i, j \in V$ cualesquiera, entonces el elemento (i, j) de la matriz será:*

$$d_{ij} = longitud(\pi_{ij})$$

siendo π_{ij} un camino de i a j tal que

$$longitud(\pi_{ij}) \leq longitud(\pi), \forall \pi \in \Pi_{ij}$$

Ejemplo 2.5. *Si se considera el grafo no orientado del ejemplo 2.3, se obtiene:*

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad D = \begin{pmatrix} 0 & 1 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 & 2 \\ 1 & 2 & 0 & 2 & 3 \\ 1 & 1 & 2 & 0 & 1 \\ 2 & 2 & 3 & 1 & 0 \end{pmatrix}$$

Para el grafo orientado del ejemplo 2.2, la matriz de adyacencia es:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

pero la de distancias no podría ser calculada al no ser el grafo conexo (por ejemplo no existe un camino de 1 a 3).

En este ejemplo se han obtenido los caminos de menor longitud de forma manual, puesto que el tamaño del grafo lo permite, pero evidentemente existen algoritmos para obtener dichos caminos (ver, por ejemplo, [11]).

Conociendo ya la definición de grafo y sus características principales, en la siguiente sección se va a profundizar en un problema clásico de grafos, el problema del flujo máximo. Posteriormente se definirán las medidas de centralidad y las funciones de afinidad, lo que completa la colección de conceptos básicos necesarios para este trabajo.

2.2. Redes de flujo

Las redes de flujo son modelos matemáticos aplicables a situaciones tales como: sistemas de tuberías (para fluidos como agua, petróleo o gas), redes de cableado eléctrico, sistemas de carreteras, sistemas de transporte de mercancías, etc. Un ejemplo de esto sería: "Un partido de baloncesto de gran interés tendrá lugar en Oviedo. Los aficionados del equipo visitante quieren ver el partido y animar a su equipo. Hay varias rutas para llegar a la ciudad, y el número máximo de coches en las autopistas es conocido. ¿Cómo determinar el máximo número de aficionados que pueden ir y las rutas que deben tomar?".

Las redes de flujo tienen muchas aplicaciones en la vida real, por ejemplo, en sistemas de vías públicas, en transporte de materiales a bodegas de almacenamiento, en redes de alumbrado público, etc. En nuestro caso, la aplicación no será ninguna de las mencionadas. Usaremos las redes de flujo para determinar la manera de llegar "de una palabra a otra" de la manera más eficiente posible. Esto se explicará en los capítulos sucesivos. Para entenderlos, será necesario comprender todos los conceptos que se dan en este apartado.

Sea un grafo ponderado orientado $G = (V, T, p)$. Supongamos ahora que los pesos de los arcos representan la capacidad máxima de los mismos (cantidad o flujo máximo que puede pasar por ese arco).

Definición 2.15. *Llamaremos* ***red estándar*** *a toda red* $R_{FS} = (V, T, p)$ *que*

satisface:

- *G=(V,T) es un grafo orientado débilmente conexo.*
- *Los pesos p_{ij}, para cada arco $(i,j) \in T$, son no-negativos y reciben el nombre de* ***capacidad*** *del arco (máxima capacidad del arco).*
- *Existe un único vértice $F \in V$ tal que $\Gamma^-(F) = \emptyset$ al que denominaremos* ***fuente*** *de la red.*
- *Existe un único vértice $S \in V$ tal que $\Gamma(S) = \emptyset$, al que denominaremos* ***salida***, ***destino*** *o* ***sumidero*** *de la red.*

Ejemplo 2.6. *Un ejemplo básico de una red estándar se puede ver en la Figura 2.3.*

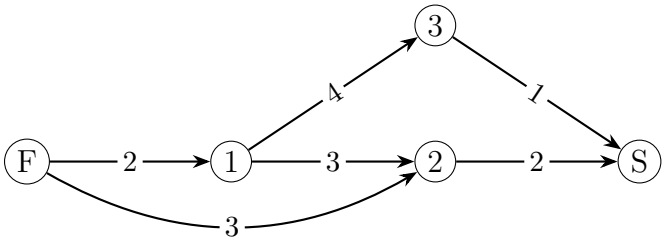


Figura 2.3: Ejemplo de red estándar.

En la misma se tiene que:

$$V = \{F, 1, 2, 3, S\}$$

$$T = \{(F,1), (F,2), (1,2), (1,3), (2,S), (3,S)\}$$

$$p_{F1} = 2, p_{F2} = 3, p_{12} = 3, p_{13} = 4, p_{2S} = 2, p_{3S} = 1$$

En una red estándar, para que exista un flujo, éste deberá cumplir ciertas restricciones:

- El flujo siempre será positivo.
- El flujo a través de un arco es menor o igual que la capacidad.
- El flujo que entra en un nodo es igual al que sale de él, para todos los nodos intermedios.

Así, se define el flujo como sigue:

Definición 2.16. *Partiendo de una red estándar $R_{FS} = (V, T, p)$, se dice que una colección de números reales $f = \{f_{ij} : (i,j) \in T\}$ es un **flujo** en R_{FS}, si se verifican las dos condiciones siguientes:*

- ***Acotación de flujo**: $0 \leq f_{ij} \leq p_{ij}, \forall (i,j) \in T$.*
- ***Conservación de flujo**: $\forall i \in V$ con $i \neq F$ e $i \neq S$,*

$$\sum_{j \in \Gamma(i)} f_{ij} - \sum_{k \in \Gamma^-(i)} f_{ki} = 0$$

Definición 2.17. *Se denomina **valor de flujo** a la cantidad*

$$v_f = \sum_{j \in \Gamma(F)} f_{Fj}$$

Si no hay ambigüedad posible, al valor del flujo f se le representa simplemente por v.

Ejemplo 2.7. *Imaginemos que queremos enviar un flujo de valor $v = 2$ de F a S en la red estándar del ejemplo 2.6. El valor del flujo f_{ij} asociado al arco (i,j) se representa en la Figura 2.4 delante de la capacidad p_{ij}.*

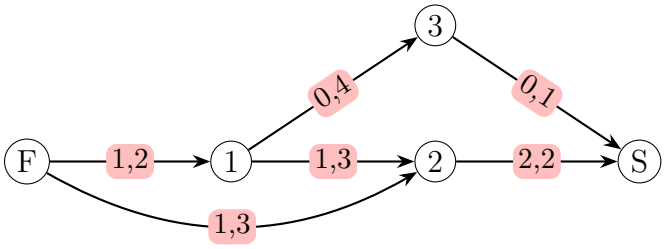


Figura 2.4: Ejemplo de flujo en una red estándar.

Es evidente que $0 \leq f_{ij} \leq p_{ij}, \forall (i,j) \in T$, con lo que se verifica la ley de acotación. En cuanto a la de conservación, se tiene en todos los nodos distintos de F y S:

$$\Gamma(1) = \{2,3\}, \Gamma^-(1) = \{F\}$$

$$\sum_{j \in \Gamma(1)} f_{1j} - \sum_{k \in \Gamma^-(1)} f_{k1} = (1+0) - (1) = 0$$

$$\Gamma(2) = \{S\}, \Gamma^-(2) = \{F, 1\}$$

$$\sum_{j\in\Gamma(2)} f_{2j} - \sum_{k\in\Gamma^-(2)} f_{k2} = (2) - (1+1) = 0$$

$$\Gamma(3) = \{S\}, \Gamma^-(3) = \{1\}$$

$$\sum_{j\in\Gamma(3)} f_{3j} - \sum_{k\in\Gamma^-(3)} f_{k3} = 0 - 0 = 0$$

Con lo cual los valores $\{f_{ij}\}$ *representan un flujo en esta red. El valor de dicho flujo es:*

$$v = \sum_{j\in\Gamma(F)} f_{Fj} = 1 + 1 = 2,$$

puesto que $\Gamma(F) = \{1, 2\}$.

Dada una red estándar cualquiera $R_{FS} = (V, T, p)$,

- El conjunto de números reales $f = \{f_{ij} : (i, j) \in T\}$ con $f_{ij} = 0$ es un flujo para dicha red, de valor 0. Con lo cual siempre existe, al menos, un flujo asociado a cada red.
- Dado un flujo f de R_{FS}, si $v_f > 0$, por la propiedad de conservación del flujo, se tiene que existe al menos un camino de F a S en R_{FS}.

2.2.1. Flujo máximo en una red

Aunque con una finalidad aparentemente distinta de para la que fue diseñado, en este trabajo se va a hacer uso del conocido problema de flujo máximo o maximal: ¿cuál es la tasa mayor a la cual el material puede ser transportado de la fuente al sumidero sin violar ninguna restricción de capacidad? En otras palabras, el problema consiste en determinar la máxima cantidad de flujo que puede ingresar a través de la fuente y salir por el nodo de destino. El procedimiento para obtener el flujo máximo de una red, consiste en seleccionar repetidas veces cualquier trayectoria de la fuente al destino y asignar el flujo máximo posible en esa trayectoria.

Así, el problema de flujo máximo consiste en enviar la cantidad máxima posible de flujo F a S de forma que se satisfagan las restricciones de capacidad sobre los arcos. Esto se traduce en encontrar un flujo f^* que maximice el valor del flujo, es decir, tal que

$$v_{f^*} = \max_{f \in \mathcal{F}} v_f$$

donde $\mathcal{F} = \{g : g \text{ es un flujo en } R_{FS}\}$.

Para resolver este problema es esencial conocer los siguientes conceptos: cuasicamino, holgura y capacidad. Se van a recordar todos ellos a continuación.

Definición 2.18. *Sea $R_{FS} = (V, A, p)$ una red estándar.*

- *Un **cuasicamino** de i_0 a i_r en R_{FS} es una sucesión alternada de vértices y arcos, $\pi_{i_0 i_r} = i_0 e_1 i_1 e_2 i_2 ... i_{k-1} e_k i_k ... i_{r-1} e_r i_r$, que es una cadena del grafo subyacente de (V, A).*
- *Un arco e_k de π_{ij} se llama **arco positivo** si está orientado de i_{k-1} a i_k, es decir, si $e_k = (i_{k-1}, i_k)$.*
- *Un arco e_k de π_{ij} se llama arco negativo si está orientado de i_k a i_{k-1}, es decir, si $e_k = (i_k, i_{k-1})$.*

Definición 2.19. *Sea $R_{FS} = (V, A, p)$ una red estándar. Un arco $(i, j) \in A$ tal que $f_{ij} = p_{ij}$ se llama **arco saturado** y un arco $(i, j) \in A$ tal que $f_{ij} = 0$ se llama **arco nulo**.*

Definición 2.20. *Sea $R_{FS} = (V, A, p)$ una red estándar. Sea f un flujo en R_{FS} y π_{ij} un cuasicamino de i a j en R_{FS}, diremos que π_{ij} es un **cuasicamino de flujo aumentable** de i a j si:*

$$\begin{cases} f_e < p_e \text{ para todo arco positivo } e \text{ de } \pi_{ij} \text{ (no saturado)} \\ f_e > 0 \text{ para todo arco negativo } e \text{ de } \pi_{ij} \text{ (no nulo)} \end{cases}$$

Los cuasicaminos de flujo aumentable son importantes porque, tal y como indica su nombre, permiten aumentar el flujo en una cantidad que va a depender de la holgura en cada arco.

Definición 2.21. *Sea $R_{FS} = (V, A, p)$ una red estándar. Para cada arco e de un cuasicamino π_{FS} de F a S se denomina **holgura del arco** e al valor positivo Δ_e definido por:*

$$\Delta_e = \begin{cases} p_e - f_e \text{ para todo arco positivo } e \text{ de } \pi_{FS} \\ f_e \text{ para todo arco negativo } e \text{ de } \pi_{FS} \end{cases}$$

Por la acotación del flujo, es evidente que $\Delta_e \geq 0, \forall e \in A$.

De la misma manera se define la ***holgura un cuasicamino*** π_{FS} *de* F *a* S *como:*

$$\Delta_{\pi_{FS}} = \min_{e \in \pi_{FS}} \{\Delta_e\}$$

La propiedad de conservación de flujo hace que el cambio en el flujo de los arcos de un cuasicamino tenga que ser de la misma magnitud, con lo que el cambio máximo permitido en el flujo de cada arco de un cuasicamino π_{FS} es $\Delta_{\pi_{FS}}$. Además, es evidente que si π_{FS} es un cuasicamino de flujo aumentable para un flujo f cualquiera, entonces $\Delta_{\pi_{FS}} > 0$.

Al aumentar el flujo que pasa por los cuasicaminos de flujo aumentable, conseguimos aumentar el valor del flujo en la red. En este hecho se va a basar el algoritmo de búsqueda de flujo máximo. Para determinar el criterio de parada del algoritmo, necesitamos definir el concepto de corte en una red y otros conceptos relacionados con él.

Definición 2.22. *Sea* $R_{FS} = (V, A, p)$ *una red estándar y* $\{B, B^c\}$ *una partición de* V *tal que* $F \in B$ *y* $S \in B^c$*, es decir, separa el vértice fuente del de salida. Al subconjunto de* A *formado por los arcos con origen en* B *y final en* B^c *se le denomina* ***corte*** *de la red y se denota por* $\delta^+(B)$*. Así,*

$$\delta^+(B) = \{(i, j) \in A : i \in B \text{ y } j \in B^c\}$$

Se denota por $\mathcal{D}^+$ *al* ***conjunto de cortes*** *de una red, es decir,*

$$\mathcal{D}^+ = \{\delta^+(B) : B \subset V, F \in B, S \notin B\}$$

Intercambiando los papeles de B y B^c se define el conjunto:

$$\delta^+(B^c) = \{(i, j) \in A : i \in B^c \text{ y } j \in B\}$$

que, evidentemente no es un corte.

Por otro lado, es evidente que toda red estándar R_{FS} tiene al menos un corte, puesto que, por ejemplo, si se consideran los subconjuntos de vértices $B_F = \{F\}$ y $B_S = V - \{S\}$, es evidente que $\delta^+(B_F), \delta^+(B_S) \in \mathcal{D}^+$.

Definición 2.23. *Dada una red estándar $R_{FS} = (V, A, p)$ y dados un flujo $f = \{f_{ij} : (i,j) \in A\}$ y un corte $\delta^+(B)$ de dicha red R_{FS}, se definen:*

- *La **capacidad** del corte $\delta^+(B)$, y se denota por $p(\delta^+(B))$, como la suma de las capacidades de los arcos del corte, es decir,*

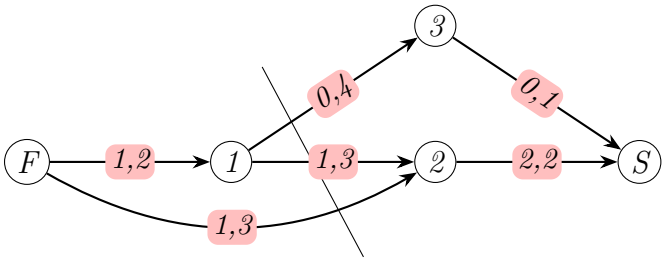
$$p(\delta^+(B)) = \sum_{(i,j)\in\delta^+(B)} p_{ij} = \sum_{i\in B, j\in B^c\cap\Gamma(i)} p_{ij}$$

- *El **flujo neto** a través del corte $\delta^+(B)$ como el valor:*

$$f(\delta^+(B)) = \sum_{(i,j)\in\delta^+(B)} f_{ij} - \sum_{(i,j)\in\delta^+(B^c)} f_{ij}$$

$$= \sum_{i\in B, j\in B^c\cap\Gamma(i)} f_{ij} - \sum_{i\in B, k\in B^c\cap\Gamma^-(i)} f_{ki}$$

Ejemplo 2.8. *Veamos un ejemplo de corte, capacidad y flujo neto, considerando como punto de partida la red y el flujo del ejemplo 2.7.*



Partición: $B = \{F, 1\}, B^c = \{2, 3, S\}$
Corte: $\delta^+(B) = \{(1,3), (1,2), (F,2)\}$
Capacidad: $p(\delta^+(B)) = 4 + 3 + 3 = 10$
Flujo neto: $f(\delta^+(B)) = (0 + 1 + 1) - 0 = 2$

Definición 2.24. *Un corte $\delta^+(B)$ de la red R_{FS} se llama **corte mínimo** si $p(\delta^+(B)) \leq p(\delta^+(B^*))$ para cualquier otro corte $\delta^+(B^*) \in \mathcal{D}^+$.*

Definición 2.25. *Un flujo f en la red R_{FS} se llama **flujo máximo** si $v_f \geq v_{f^*}$ para cualquier otro flujo $f^* \in \mathcal{F}$.*

Proposición 2.9. *Si $\delta^+(B)$ y f son, respectivamente, un corte y un flujo en una red estándar R_{FS}, tales que $v_f = p(\delta^+(B))$ entonces f es un flujo máximo y $\delta^+(B)$ es un corte mínimo.*

Teorema 2.10. *Sea f un flujo en una red estándar $R_{FS} = (V, A, p)$, entonces f es un flujo máximo en R_{FS} si, y solo si, no existe ningún cuasicamino de flujo aumentable de F a S en R_{FS}*

Según el teorema 2.10, el problema de encontrar el flujo máximo en una red se traduce a ir saturando todos los cuasicaminos de flujo aumentable hasta que no exista ninguno. Teniendo en cuenta esto, Ford y Fulkerson [4] desarrollaron en 1956 un algoritmo clásico de etiquetado que trata de buscar cuasicaminos de flujo aumentable. Este algoritmo simplemente buscaba la manera de llevar el flujo máximo sin tener en cuenta la eficiencia de ésta. Como en este trabajo se prefiere que el flujo sea llevado de la manera más rápida y eficiente posible, se considerará el algoritmo de Ford-Fulkerson-Edmonds-Karp [7]. Este algoritmo es idéntico al algoritmo de Ford-Fulkerson, excepto porque el orden para ir buscando los cuasicaminos aumentables viene fijado por el propio método. Edmonds y Karp solucionaron este problema en 1972, al elegir en la fase de etiquetado el cuasicamino de flujo aumentable con el menor número de arcos, en lugar de un cuasicamino de flujo aumentable cualquiera.

Algoritmo de Ford-Fulkerson-Edmonds-Karp

Se considera el flujo f en la red estándar $R_{FS} = (V, A, p)$ con $f_e = 0, \forall e \in A$. Los pasos de este algoritmo son:

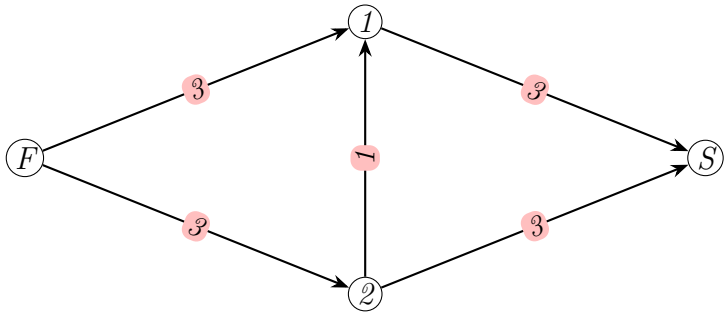
1. Etiquetado:
   - *a)* Hacer $d(F) = 0,\ d(i) = -1\ \forall i \neq F, k = 1$
   - *b)* $\forall u$ tal que d(u)=k-1, hacer d(w)=k si $d(w) = -1$ y $f_{uw} < p_{uw}$, o bien, d(w)=-1 y $f_{wu} > 0$.
     - Si $d(S) \neq -1$, considerar el cuasicamino de flujo aumentable de F a S que se forma yendo para atrás desde la etiqueta del S e ir al paso 3.
     - Si se han etiquetado nuevos vértices y $d(S) = -1$, hacer k=k+1 y repetir el paso 1.b).
     - Si no se puede asignar nuevas etiquetas y $d(S) = -1$, PARAR, el flujo actual es máximo.

2. Para el cuasicamino de flujo aumentable π_{FS} desde F a S, formado por los arcos entre vértices etiquetados, obtener $\Delta_{\pi_{FS}}$ y con ese valor obtener el nuevo flujo $f^* = f + \gamma_{\delta_{\pi_{FS}}}$.

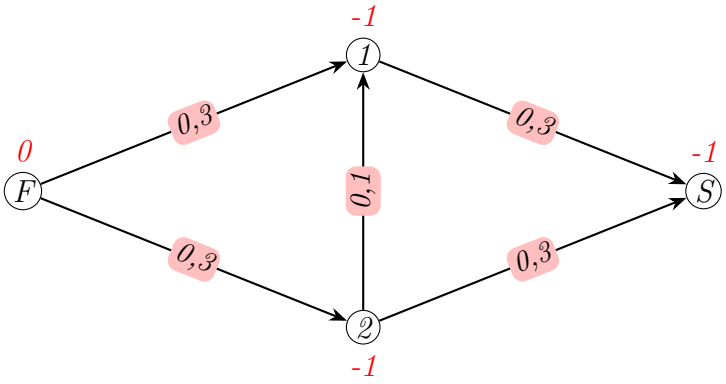
3. Volver al paso 1, borrando previamente todas las etiquetas.

A continuación se presenta un ejemplo de aplicación de este algoritmo.

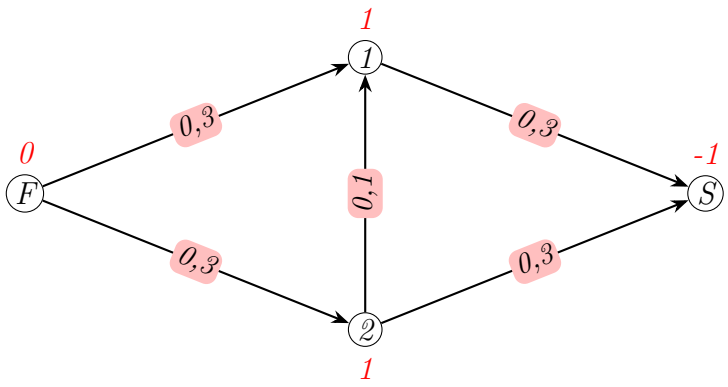
Ejemplo 2.11. *Se va a aplicar el algoritmo para obtener el flujo máximo en la red estándar siguiente:*



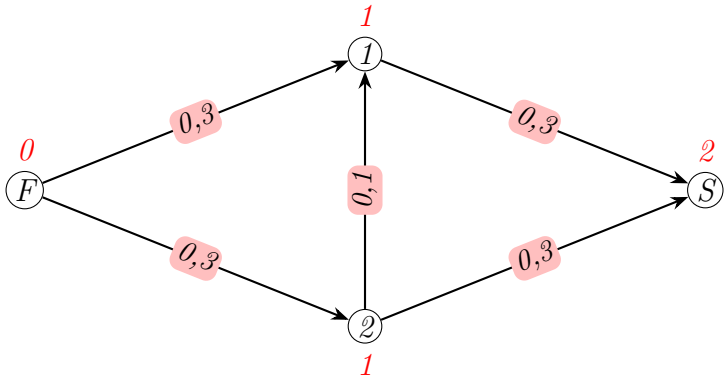
Primera iteración
Definimos un flujo inicial nulo y marcamos con -1 *a todos los nodos de la red excepto al nodo fuente, que lo marcamos con un* 0.
Flujo inicial: $f_e = 0, \forall e \in A$



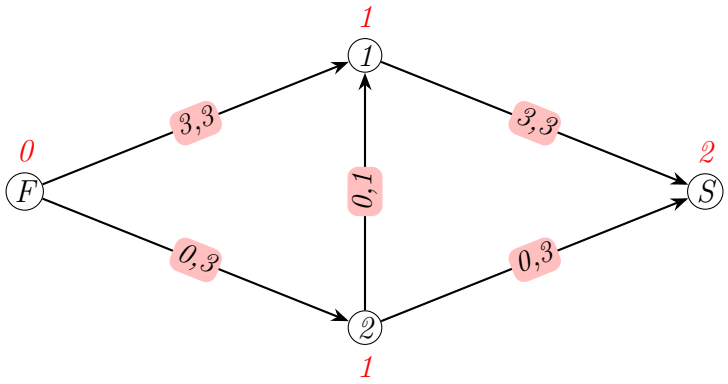
A los nodos sucesores del nodo fuente que forman cuasicaminos de flujo aumentable, se les cambia la etiqueta por 1. Como el sumidero sigue teniendo la etiqueta -1, proseguimos.
$k = 1, d(i) = 0, d(S) = -1$



Ahora, marcamos con $d(u) = k = 2$ *a los nodos sucesores de los marcados anteriormente con un* 1*, por los cuales pueda aumentarse el flujo.*
$k = 2, d(i) = 1, d(S) = 2$

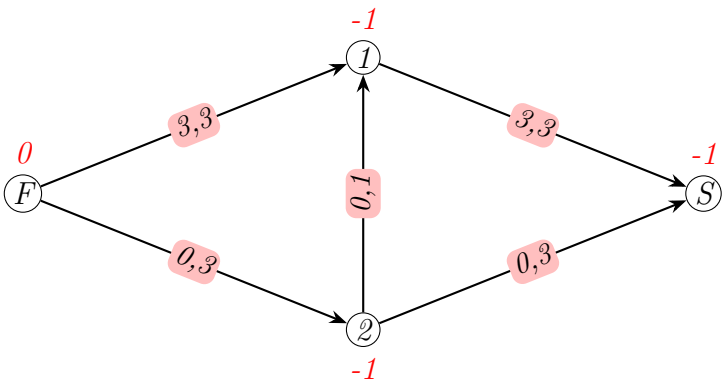


Como $d(S) = 2$*, consideremos el cuasicamino de flujo aumentable de* F *a* S *que se forma. En este caso, tenemos dos cuasicaminos posibles. Cogemos el camino que pasa por el nodo 1, y aumentamos el flujo.*
Aumento de flujo: $f^* = f + \gamma_{\Delta_{\pi_{FS}}} = 0 + 3 = 3.$



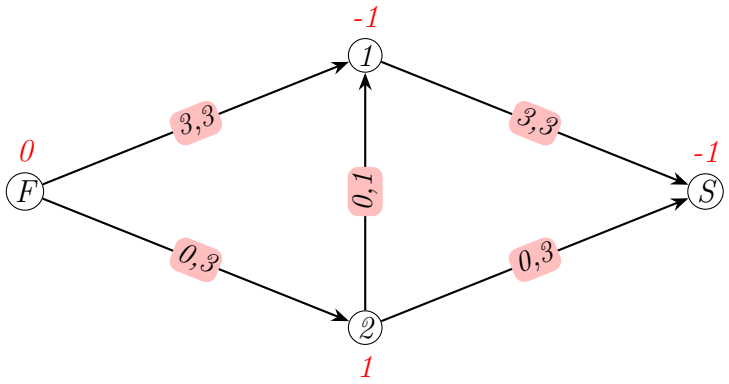
Segunda iteración

Borramos las etiquetas usadas previamente y etiquetamos de nuevo al nodo fuente con 0 *y al resto con* -1.

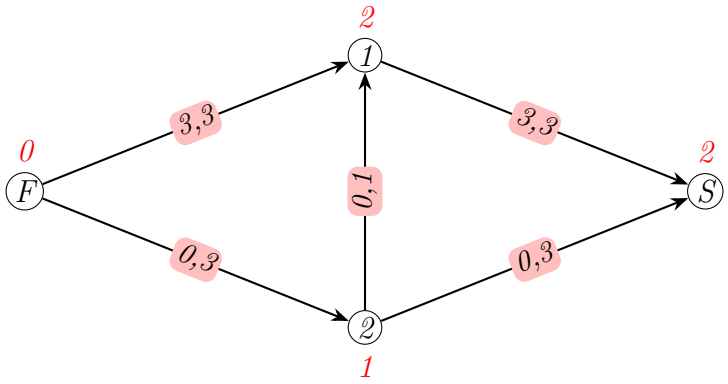


Los nodos sucesores de F *que forman cuasicaminos de flujo aumentable, se etiquetan con el número* 1. *Notamos que el algoritmo desprecia ya al nodo* 1 *porque por él no se puede aumentar el flujo.*

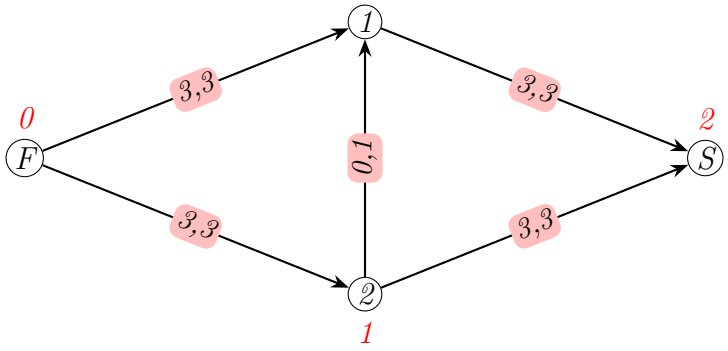
$k = 1, d(i) = 0, d(S) = -1$



Ahora etiquetamos los sucesores del nodo 2, *con la etiqueta* $k = 2$. *Como* $d(S) = 2$, *usamos el cuasicamino de flujo aumentable* $F - 2 - S$.
$k = 2, d(i) = 1, d(S) = 2$

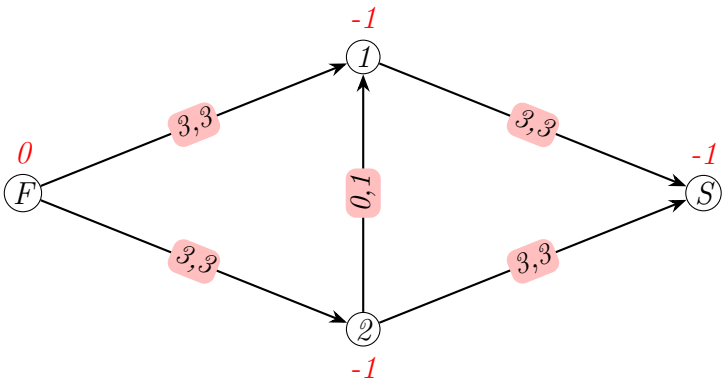


Aumento de flujo: $f^* = f + \gamma_{\Delta_{\pi_{FS}}} = 3 + 3 = 6$.



Tercera iteración

Borramos las etiquetas usadas previamente y etiquetamos de nuevo al nodo fuente con 0 *y al resto con* -1.



No se pueden asignar nuevas etiquetas, ya que no existen cuasicaminos de flujo aumentable, y además $d(S) = -1$*, con lo que se ha obtenido el flujo máximo en 2 iteraciones. El valor del flujo máximo es* 6.

Notar que el algoritmo encuentra satisfactoriamente la forma más "rápida" de enviar todo el flujo posible, evitando los caminos $F - 2 - 1 - S$ *y* $F - 1 - 2 - S$*, que requerirían de más pasos para alcanzar dicho flujo máximo.*

2.3. Medidas de centralidad

En teoría de grafos y redes, las medidas de centralidad indican la relevancia de cada nodo en una estructura [12], e identifican cuál es el nodo más importante en una red. El concepto de centralidad intenta cuantificar a la importancia o prominencia de los vértices (o nodos o actores) dentro de un grafo o red social. Existen cientos de medidas o índices de centralidad, para determinar y comparar cuantitativamente la importancia relativa de un actor dentro de la estructura definida por la red. Usualmente estas medidas se normalizan para que tomen valores en el intervalo $[0, 1]$, con el propósito de poder hacer comparaciones entre distintas redes y casos de estudio. La centralidad no es un atributo intrínseco de los nodos o actores de una red, como podrían serlo la autoestima, la temperatura, el ingreso monetario, etc. sino un atributo estructural, es decir, un valor asignado que depende de las relaciones del actor con los demás actores de la red. Intuitivamente (aunque dependerá de la medida de centralidad utilizada), en un grafo estrella o red egocéntrica el nodo central debería tener la mayor centralidad, mientras que los nodos periféricos compartirían todos un mismo valor de centralidad, inferior al del centro.

El concepto fue introducido inicialmente por Alex Bavelas [1] en 1948. Es uno de los conceptos más estudiados en el análisis de redes sociales, y muchos de los conceptos relacionados con las medidas de centralidad reflejan su origen sociológico.

En ocasiones, en el caso de relaciones dirigidas en lugar de hablar de "centralidad" se suele también hablar de medidas de prestigio o estatus, aunque dependiendo del tipo de relaciones consideradas, también se podría hablar de "rango", "deferencia" o "popularidad". En el análisis de redes sociales se suele distinguir también entre centralidad de actor y centralidad de grupo, siendo lo primero equivalente a la noción de centralidad utilizada en este trabajo, mientras que la centralidad de grupo se refiere al concepto asociado a la centralización.

A continuación se definen las medidas de centralidad más comunes, que son las que estudiaremos en los capítulos sucesivos. Las definiremos para grafos no orientados, ya que en nuestro caso trabajaremos la centralidad en redes en las que no influye el orden de los arcos.

2.3.1. Centralidad de grado

En análisis de redes sociales, la centralidad de grado es la primera y más simple de las medida de centralidad. Descrita inicialmente por Proctor y Loomis [13] en 1951, corresponde sencillamente al grado de un nodo o actor. Es decir, se mide el número de enlaces o conexiones que tiene un nodo con los demás nodos pertenecientes a un grafo.

Definición 2.26. *La **centralidad de grado** de un nodo corresponde al número de aristas del grafo tiene tienen como extremo a ese nodo. Así, para cada $i \in V$, su centralidad de grado $C_D(i)$ se define como*

$$C_D(i) = \sum_{jV} a_{ij}$$

Es usual normalizar las medidas de centralidad para la posterior comparación con otras redes. Al tratarse de grafos simples, esto se consigue dividiendo entre el número máximo de aristas que podrían tener al nodo como extremo, es decir, entre $|V| - 1 = n - 1$. Así:

$$C_D^{'}(i) = \frac{C_D(i)}{n-1}$$

Esta medida plasma la idea intuitiva de que un nodo será más importante (central) cuantas más conexiones tenga. Por tanto, esta medida adjudicará a cada nodo $i \in V$ un grado de centralidad de manera directamente proporcional al grado del mismo.

Ejemplo 2.12. *Veamos la centralidad de grado de cada nodo para el ejemplo 2.3 en la siguiente tabla.*

$Nodo$	1	2	3	4	5
C_D	3	2	1	3	1
$C_D^{'}$	0,75	0,5	0,25	0,75	0,25

Tabla 2.1: Valores de centralidad de grado de los nodos del ejemplo 2.3.

2.3.2. Centralidad de proximidad

La centralidad de proximidades una medida de centralidad definida formalmente por Beauchamp [2] en 1965 y Sabidussi [15] en 1966, con aplicaciones en redes de comunicación. Es la más conocida y utilizada de las medidas radiales de longitud. Se basa en calcular la suma o bien el promedio de las distancias geodésicas (o longitudes de los caminos más cortos) desde un nodo hacia todos los demás. Notar que mientras mayor sea la distancia entre dos vértices, menor será la proximidad entre estos. Por lo tanto, la proximidad se define como el inverso de la "lejanía" entre dos vértices.

Definición 2.27. *La **centralidad de proximidad** de un nodo es la inversa de la suma de las longitudes de la ruta más corta entre ese nodo y el resto de nodos del grafo. Entonces, la centralidad de proximidad de un nodo $i \in V$, $C_P(i)$, se define como*

$$C_P(i) = \frac{1}{\sum_{j \in V} d_{i,j}}$$

Esta medida se normaliza de la siguiente manera

$$C'_P(i) = (n-1)C_P(i) = \frac{n-1}{\sum_{j \in V} d_{i,j}}$$

que para grafos con un orden muy elevado da un valor muy similar a la inversa de la longitud media.

Si consideramos una red de transportes, ya sea por tierra, mar o aire, es lógico pensar que una ciudad será más central cuanto antes se pueda llegar a otras desde ella. Esta forma de plantear la centralidad es importante si pensamos en flujo de información: cuanto más alejado esté un nodo del que envía la información, tardará más en llegarle y será más improbable que lo haga. La idea que subyace de la centralidad de proximidad es que un nodo será más central en la medida en que esté más cerca del resto de nodos de la red. Por tanto, para calcular la centralidad de cercanía de un nodo, primero se han de calcular todos los caminos más cortos desde ese nodo al resto de vértices de la red y posteriormente se hará la media de todas esas distancias.

Ejemplo 2.13. *Si consideramos de nuevo el grafo del ejemplo 2.3, los cami-*

nos de longitud mínima desde cada vértice a los demás son:

$1-2$	$2-1$	$3-1$	$4-1$	$5-4-1$
$1-3$	$2-1-3$	$3-1-2$	$4-2$	$5-4-2$
$1-4$	$2-4$	$3-1-4$	$4-1-3$	$5-4-1-3$
$1-4-5$	$2-4-5$	$3-1-4-5$	$4-5$	$5-4$

con lo que la suma de las longitud desde cada nodo es: $5, 6, 8, 5$ *y* 8*, respectivamente. Así, la centralidad de proximidad para cada nodo se calcula como la inversa de estas longitudes y tanto la original, como la normalizada, aparecen recogidas en la Tabla 2.2.*

$Nodo$	1	2	3	4	5
C_P	$1/5$	$1/6$	$1/8$	$1/5$	$1/8$
C'_P	$4/5$	$4/6$	$1/2$	$4/5$	$1/2$

Tabla 2.2: Valores de centralidad de proximidad de los nodos del ejemplo 2.3.

Es claro que los nodos que más cercanos están del resto, según este criterio, son el nodo 1 *y el* 4*.*

2.3.3. Centralidad del vector propio

La sencillez de la centralidad de grado y de la centralidad de proximidad es evidente, pero también lo es su falta de ajuste a la realidad en algunos casos. El hecho de que un usuario tenga muchos amigos en una red social no le hace inmediatamente importante. Sin embargo, tener un único amigo relevante en una red le puede otorgar mucha más visibilidad. La centralidad del vector propio soluciona este problema asignando una puntuación relativa a cada nodo en la red basada en la idea de que las conexiones a los nodos bien conectados deben ponderar más que las conexiones a nodos mal conectados. Así, la centralidad del vector propio, también llamado prestigio de rango o prestigio de estatus, es una medida de centralidad utilizada para cuantificar el nivel de influencia, prestigio o estatus de un nodo o actor en un grafo o red social. Fue propuesta por Phillip Bonacich [3] en 1972, y corresponde al principal vector propio de la matriz de adyacencia del grafo analizado. Esta centralidad intenta generalizar la centralidad de grado incorporando la importancia de sus nodos vecinos.

Definición 2.28. *La **centralidad del vector propio** de un nodo $i \in V$, $C_{VP}(i)$, depende de cómo de centrales sean los nodos a los que está conectado. Se define como*

$$C_{VP}(i) = \frac{1}{\lambda} \sum_{j=1}^{n} a_{ij} \cdot C_{VP}(j)$$

donde para calcular los valores de C_{VP} para los n nodos se construye un sistema de n ecuaciones con n incógnitas que se representa de manera matricial. Si $C = (C_{VP}(1), ..., C_{VP}(n))^T$ es el vector traspuesto que almacena los n valores de C_{VP} (C es un vector columna) y A es la matriz de adyacencia, entonces:

$$\lambda C = AC$$

De esta manera, se observa que λ es un autovalor de la matriz de adyacencia A. Tomaremos el mayor de los autovalores en módulo, dado que nos interesa que las medidas de centralidad sean positivas. Entonces,

$$\lambda := ||\lambda_{max}||$$

Ejemplo 2.14. *Para hallar la centralidad del vector propio para el grafo del ejemplo 2.3 lo primero es conocer los autovalores de la matriz de adyacencia. Para ello necesitamos calcular el determinante de $A - \lambda I$ (notar que I es la matriz identidad 5x5).*

$$\det(A - \lambda I) = \begin{vmatrix} -\lambda & 1 & 1 & 1 & 0 \\ 1 & -\lambda & 0 & 1 & 0 \\ 1 & 0 & -\lambda & 0 & 0 \\ 1 & 1 & 0 & -\lambda & 1 \\ 0 & 0 & 0 & 1 & -\lambda \end{vmatrix}$$

Obtenemos el polinomio característico:

$$det(A - \lambda I) = -\lambda^5 + 5\lambda^3 + 2\lambda^2 - 3\lambda$$

Ahora, calculamos los autovalores igualando el polinomio característico a 0. Puesto que $\lambda(-\lambda^4 + 5\lambda^2 + 2\lambda - 3) = 0$ si, y sólo si, $\lambda_1 = 0, \lambda_2 \approx 2{,}3, \lambda_3 \approx 0{,}62, \lambda_4 \approx -1{,}3, \lambda_5 \approx -1{,}62$, se tiene que: $\lambda_{max} = \text{máx}\{|0|, |2{,}3|, |0{,}62|, |-1{,}3|, |-1{,}62|\} = 2{,}3$. El autovector asociado a este autovalor sería el vector C que buscamos. Vamos a calcularlo resolviendo la ecuación matricial ($A -$

$\lambda_{\text{máx}}I)C = 0$.

$$\begin{pmatrix} -2{,}3 & 1 & 1 & 1 & 0 \\ 1 & -2{,}3 & 0 & 1 & 0 \\ 1 & 0 & -2{,}3 & 0 & 0 \\ 1 & 1 & 0 & -2{,}3 & 1 \\ 0 & 0 & 0 & 1 & -2{,}3 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Así, nos quedaría el siguiente sistema a resolver

$$\left.\begin{aligned} -2{,}3c_1 + c_2 + c_3 + c_4 &= 0 \\ c_1 - 2{,}3c_2 + c_4 &= 0 \\ c_1 - 2{,}3c_3 &= 0 \\ c_1 + c_2 - 2{,}3c_4 + c_5 &= 0 \\ c_4 - 2{,}3c_5 &= 0 \end{aligned}\right\}$$

Si tomamos $c_3 = c$ como parámetro, en la Tabla 2.3 podemos ver los valores para la centralidad del vector propio. Los nodos más centrales según esta centralidad son el 1 y el 4, y los que menos, el 3 y el 5.

Nodo	1	2	3	4	5
C_{VP}	2.3c	1.99c	c	2.3c	c

Tabla 2.3: Valores de la centralidad del vector propio para los nodos del ejemplo 2.3.

2.3.4. Centralidad de intermediación

La centralidad de intermediación, o simplemente intermediación es una medida de centralidad que cuantifica la frecuencia o el número de veces que un nodo se encuentra entre los caminos más cortos de otros actores. Esta medida fue formalizada por Freeman [18] en 1977. Si bien la medida fue inicialmente definida para grafos no dirigidos, una década más tarde, se demostró que también funcionaba para grafos dirigidos. La idea intuitiva es que si se eligen dos nodos al azar, y luego también al azar uno de los eventuales posibles caminos más cortos entre ellos, entonces los nodos con mayor intermediación serán aquellos que aparezcan con mayor frecuencia dentro de este camino.

Podemos pensar en las redes de transporte aéreo ya que son un ejemplo muy didáctico para esta medida. Una manera de plantear la centralidad de un aeropuerto, es decir, de un nodo de la red, es considerando el número de veces que se hace escala en él para viajar de un sitio a otro. En el caso de las redes sociales, esto equivaldría a tomar como usuario más importante aquel por el que pasa más información. La centralidad de intermediación considera que un nodo es más central cuantas más veces aparezca en el camino más corto que conecta a otros dos nodos distintos de la red.

Definición 2.29. *La **centralidad de intermediación** de un nodo es el número de veces que el nodo está en la ruta más corta de otros nodos. Así, la centralidad de intermediación de un nodo $i \in V$, $C_I(i)$, se define como*

$$C_I(i) = \sum_{j \neq k \neq i} \frac{b_{jik}}{b_{jk}}$$

donde b_{jk} es el número de caminos más cortos desde el nodo j hasta el nodo k, y b_{jik} es el número de caminos más cortos desde j hasta k que pasan a través del nodo i.

Para normalizar esta medida, se divide por el mayor número posible de pares de actores, excluyendo el nodo que se está midiendo. En este caso por $\frac{(n-1)(n-2)}{2}$. Así, se obtiene

$$C_I'(i) = \frac{2}{(n-1)(n-2)} C_I(i)$$

Ejemplo 2.15. *Para calcular la centralidad de intermediación de cada nodo del ejemplo 2.3, necesitamos la matriz de distancias del grafo. Además, para cada camino de longitud mínima k calculamos la matriz A^k, cuyas entradas son el número de caminos de longitud k entre cada par de vértices. Al observar la matriz D podemos apreciar que necesitamos el número de caminos mínimos de longitudes 1, 2 y 3, ya que $\max\{d_{ij}\} = 3$. Estos caminos mínimos los dan las matrices A, A^2 y A^3. La matriz A ya se había calculado en el ejemplo 2.5. Los valores de A^2 y A^3, se recogen en la Tabla 2.4 y en la Tabla 2.5, respectivamente.*

Nodos	1	2	3	4	5
1	3	1	0	0	1
2	1	2	1	1	1
3	0	1	1	1	0
4	1	1	1	3	0
5	1	1	0	0	1

Tabla 2.4: Número de caminos mínimos de longitud 2 entre cada par de vértices del ejemplo 2.3.

Nodos	1	2	3	4	5
1	2	4	3	5	1
2	4	2	1	4	1
3	3	1	0	1	1
4	5	4	1	2	3
5	1	1	1	3	0

Tabla 2.5: Número de caminos mínimos de longitud 3 entre cada par de vértices del ejemplo 2.3.

Para ver cuántos caminos mínimos existen, por ejemplo, del nodo 3 al 5, debemos consultar primero en la matriz de distancias D cual es esa distancia mínima. En este caso $d_{35} = 3$*, por lo tanto debemos ir a la posición* $(3, 5)$ *de la matriz de caminos de longitud 3 para saber cuantos caminos mínimos existen. Como se puede observar,* $a^3_{35} = 1$*, hay un único camino de longitud mínima que va desde el vértice 3 al vértice 5. Dicho camino es:*

$$3 - 1 - 4 - 5$$

Combinando la información de la matriz D con la matriz de adyacencia y las matrices de longitud 2 y longitud 3, se obtiene el número de caminos más cortos entre cada par de nodos, que se dan en la Tabla 2.6.

Nodos	1	2	3	4	5
1	0	1	1	1	1
2	1	0	1	1	1
3	1	1	0	1	1
4	1	1	1	0	1
5	1	1	1	1	0

Tabla 2.6: Número de caminos de longitud mínima entre cada par de vértices del ejemplo 2.3.

Ahora ya tenemos los elementos b_{jk} de la definición 2.29. El paso a seguir es, de todos estos caminos mínimos entre pares de nodos distintos de $V-\{i\}$, comprobar cuantos pasan por el nodo i. Este proceso hay que repetirlo para cada nodo $i \in V$. Notar que $b_{jk} = 1, \forall j,k \in V$ con $k \neq j$ nos facilita mucho el cálculo de la centralidad de intermediación para todos los nodos. Los valores para la centralidad de intermediación se dan en la Tabla 2.7.

Nodo	1	2	3	4	5
C_I	6	0	0	10	0
C_I'	1/2	0	0	0,830	0

Tabla 2.7: Medidas de centralidad de intermediación para los nodos del ejemplo 2.3.

Está claro que el nodo más intermediador es el 4 y los vértices 2, 3 y 5 no sirven de intermediario en ningún camino.

Notar que el hecho de estudiar un grafo con pocos nodos y arcos hace que al final lleguemos a que solo exista un único camino mínimo uniendo cada par de vértices. En el caso de grafos más complejos, los cálculos se complicarían considerablemente.

2.4. Funciones de afinidad

Las funciones de afinidad [8] sirven para medir la relación entre un par de individuos de una red, fijándose en la naturaleza de sus interacciones locales. Gracias a estas funciones podemos crear grafos ponderados según diferentes criterios.

Definición 2.30. *Una* ***función de afinidad*** F_A *se define como una función sobre el conjunto de aristas* T *de un grafo, que asigna a cada arista* (i, j) *un número en el intervalo* $[0, 1]$*, es decir,*

$$F_A : T \to [0, 1]$$

La afinidad entre dos individuos muestra cómo de fuerte es su relación usando diferentes criterios, dependiendo de los factores que estemos teniendo en cuenta en las relaciones en la red. Según estos factores, podemos dividir las funciones en dos grupos: funciones de afinidad personal y funciones de afinidad estructural.

2.4.1. Funciones de afinidad personal

La **afinidad personal** establece la fuerza de una conexión interpersonal entre i y j usando sus respectivas conexiones y relaciones comunes.

De la primera que hablaremos es de la afinidad del mejor amigo (MA), la cual mide la importancia de una relación de un individuo j para otro individuo i, teniendo en cuenta todas las otras relaciones que tiene i.

Definición 2.31. *La* ***función de afinidad del mejor amigo****,* F_A^{MA}*, se define como*

$$F_A^{MA}(i, j) = \frac{a_{ij}}{\sum_{k \in V} a_{ik}}$$

donde $A = (a_{ij})_{i,j \in V}$ *denota la matriz de adyacencia.*

Una variación de la anterior es la afinidad del mejor amigo común (MAC), que mide la importancia de una relación teniendo en cuenta cómo de impor-

tantes son las conexiones comunes entre los nodos conectados a i y a j, en relación a todas las otras relaciones que tiene i en la red.

Definición 2.32. *La* ***función de afinidad del mejor amigo común****,* F_A^{MAC}*, se define como*

$$F_A^{MAC}(i,j) = \frac{\text{máx}_{k \in V}\{\text{mín}\{a_{ik}, a_{jk}\}\}}{\sum_{k \in V} a_{ik}}$$

Ejemplo 2.16. *Veamos los valores para cada par de vértices resultantes al usar las funciones de afinidad MA y MAC para el grafo del ejemplo 2.3.*

	1	2	3	4	5
1	0	1/3	1/3	1/3	0
2	1/2	0	0	1/2	0
3	1	0	0	0	0
4	1/3	1/3	0	0	1/3
5	0	0	0	1	0

Tabla 2.8: Afinidad del Mejor Amigo entre pares de vértices del ejemplo 2.3.

	1	2	3	4	5
1	0	1/3	0	1/3	1/3
2	1/2	0	1/2	1/2	1/2
3	0	1	0	1	0
4	1/3	1/3	1/3	0	0
5	1	1	0	0	0

Tabla 2.9: Afinidad del Mejor Amigo Común entre pares de vértices del ejemplo 2.3.

2.4.2. Funciones de afinidad estructural

La **afinidad estructural** cuantifica la relación de un par de individuos basándose en las medidas de centralidad vistas anteriormente. Por ejemplo, en su centralidad de grado o de intermediación.

Se considera aquí la afinidad de Machiavelli, que calcula cómo de afines son dos individuos i y j basándose en las similitudes de las estructuras sociales que les rodean.

Definición 2.33. *La* ***función de afinidad de Machiavelli****,* F_A^{Mach}*, se define como*

$$F_A^{Mach}(i,j) = 1 - \frac{|W_i - W_j|}{\max\{W_i, W_j\}}$$

donde $W_i = \sum_{k \in Z} C_D(k)$*, siendo* $Z = \{z \in V : a_{iz} > 0\}$.

Notar que la matriz de adyacencia resultante al usar la función de afinidad de Machiavelli es simétrica y con todos los elementos diagonales iguales a uno.

Ejemplo 2.17. *Para calcular los valores para cada* (i,j) *y hallar* F_A^{Mach} *para el ejemplo 2.3 necesitamos conocer primero* W_i*,* $\forall i \in V$*. Usando los valores de centralidad de grado de cada* $i \in V$ *sin normalizar, resultan*

$$\begin{aligned}
W_1 &= C_D(2) + C_D(3) + C_D(4) = 2 + 1 + 3 = 6 \\
W_2 &= C_D(1) + C_D(4) = 3 + 3 = 6 \\
W_3 &= C_D(1) = 3 \\
W_4 &= C_D(1) + C_D(2) + C_D(5) = 3 + 2 + 1 = 6 \\
W_5 &= C_D(4) = 3
\end{aligned}$$

Así, calculando cada $F_A^{Mach}(i,j)$ *nos quedan los siguientes valores de la afinidad de Machiavelli:*

	1	2	3	4	5
1	1	1	1/2	1	1/2
2	1	1	1/2	1	1/2
3	1/2	1/2	1	1/2	1
4	1	1	1/2	1	1/2
5	1/2	1/2	1	1/2	1

Tabla 2.10: Afinidad de Machiavelli entre pares de vértices del ejemplo 2.3.

La lógica difusa es una rama de la inteligencia artificial que le permite a una computadora analizar información del mundo real en una escala entre lo

falso y lo verdadero, manipula conceptos vagos, como çaliente.º "húmedo", y permite a los ingenieros construir dispositivos que juzgan la información difícil de definir. Las funciones de afinidad se han utilizado con diferentes operadores en lógica difusa para generar nuevas funciones de afinidad usando combinaciones de las previamente utilizadas.

En inteligencia artificial, la lógica difusa, también llamada lógica borrosa, se utiliza para la resolución de una variedad de problemas, principalmente los relacionados con control de procesos industriales complejos y sistemas de decisión en general, la resolución y la compresión de datos. Los sistemas de lógica difusa están también muy extendidos en la tecnología cotidiana, por ejemplo en cámaras digitales, sistemas de aire acondicionado, lavadoras, etc. Los sistemas basados en lógica difusa imitan la forma en que toman decisiones los humanos, con la ventaja de ser mucho más rápidos. Con la lógica convencional, las computadoras pueden manipular valores estrictamente duales, como verdadero/falso, sí/no o ligado/desligado. En la lógica difusa, se usan modelos matemáticos para representar nociones subjetivas, como caliente/tibio/frío, para valores concretos que puedan ser manipuladas por los ordenadores.

Existen muchos operadores efectivos para combinar diferentes funciones de afinidad. En este trabajo se implementarán combinaciones convexas de dos funciones de afinidad y las T-normas de funciones de $n-$afinidad. La idea de usar combinaciones convexas es que ponderan dos facetas diferentes de la relación, caracterizadas por las dos funciones de afinidad que queremos combinar. Las T-normas son de particular importancia porque la mayoría de las funciones de afinidad dan como resultado 0 en una red, pero algunas afinidades, como la de Machiavelli, resultan en valores mayores que 0 la mayoría de veces. Usando T-normas mantenemos una baja densidad en la afinidad resultante de la red, porque cuando una de las afinidades es 0 sabemos que el resultado de la agregación también será 0.

Capítulo 3

Valor semántico en el análisis de redes sociales

En esta sección presentamos nuestra formalización del valor semántico de un individuo en una red social. El objetivo de esta formalización es modelar el "Nóumeno" de la filosofía Kantiana [14] y otros trabajos filosóficos conocidos [16, 6], en términos y conceptos computacionales. Las cosas pueden tener un valor intrínseco, es decir, por sí mismas, o un valor extrínseco como, por ejemplo, el valor de cambio, es decir, el precio. Lo que vale un billete de banco no es lo que cuesta de fabricar, sino el valor que se quiere que signifique. El valor intrínseco es siempre algo que un objeto tiene "en sí mismo" o "por sí mismo" y es una propiedad intrínseca. Un objeto con valor intrínseco puede ser considerado como un "fin", o en terminología kantiana, como un "fin en sí mismo". Tradicionalmente, los filósofos sostenían que una entidad tiene valor intrínseco si es buena en o por sí misma. El valor intrínseco se contrapone al valor extrínseco, el cual se atribuye a las cosas que son valiosas solo como un medio para otra cosa.

Definición 3.1. *Definimos $\mp(x)$, el valor semántico de un individuo en una red social, como la unión del valor intrínseco y extrínseco:*

$$\mp(x) = \cup(\perp(x), \top(x)) \tag{3.1}$$

Este concepto depende a su vez del valor intrínseco y extrínseco. Primero definimos $\perp(x)$ para el valor intrínseco de un individuo x en una red, con

la propiedad de que sea único e inherente a él, no necesariamente deducible de la estructura o topología de la red. Si el individuo x se elimina de la red, entonces todos los valores intrínsecos $\perp(x)$ presentes en la red son también eliminados de todos los valores semánticos existentes en la red.

También definimos $\top(x)$ como el valor extrínseco de un individuo x, que representa la información de las interacciones locales de x, considerando las relaciones como transmisiones de información y recursos, como es usual en el análisis de redes sociales [17].

Definición 3.2. *Si denotamos por X al vector de nodos conectados a x (supuesto de dimensión a) y F_A es una función de afinidad, el valor extrínseco, $\top(x)$ es el conjunto dado por:*

$$\top(x) = \bigcup_{i=1}^{a} \{F_A(X_i, x) \mp (X_i) - \cup_{j \in J} \{\cap (F_A(X_i, x) \mp (X_i), F_A(X_j, x) \mp (X_j))\}\} \tag{3.2}$$

donde $J = \{j \in \{1, ..., a\}, i \neq j\}$.

Con esta expresión, establecemos que el valor extrínseco es la unión de los valores semánticos recibidos de los otros usuarios. Cada individuo en X envía su propio valor semántico a x, modulado en cada caso por la afinidad de dicha relación, y omitiendo las redundancias producidas por otras relaciones.

Para simplificar la expresión de la definición 3.2, podemos definir $V_x(b) = F_A(X_b, x) \mp (X_b)$. Así, la ecuación quedaría simplificada como:

$$\top(x) = \bigcup_{i=1}^{a} \{V_x(i) - \cup_{j \in J} \{\cap (V_x(i), V_x(j))\}$$

Pero esta definición tiene un problema importante: cuando calculamos el valor extrínseco obtenemos infinita recursividad. Esta recursividad es inevitable para esta definición y está en línea con la idea de que en los diccionarios, para definir una palabra, deben utilizar otras palabras.

3.1. Cálculo del valor semántico

Para dar una versión calculable del valor semántico para un nodo x, lo primero es dar una versión calculable de los valores intrínseco y extrínseco, $\perp(x)$ y $\top(x)$. Denotaremos la versión calculable del valor intrínseco como I, del valor extrínseco como E, y del valor semántico, $\mp(x)$, como S. Debido a la falta de claridad inherente al concepto de valor intrínseco, no podemos dar una fórmula matemática exacta para calcularlo en cualquier red. Sin embargo, dependiendo del contexto y de la aplicación, podemos usar una función para transformar esta idea abstracta en un número. Por ejemplo, si trabajamos con una red de *routers*, la localización en el "mundo real" de un router es una parte importante del valor intrínseco $\perp$ del mismo. Si consideramos la idoneidad de cada uno de los factores del "mundo real" para la tarea de transmisión de señales, podemos usar esta condición como I. Como este trabajo está centrado en el análisis de textos, aproximaremos el valor intrínseco de un nodo como la frecuencia absoluta (número de apariciones) del término asociado al nodo en los documentos.

Definición 3.3. *Se define el* ***valor intrínseco*** *de un nodo $x \in V$ como $I(x)$ siendo*

$$I(x) = freq(x)$$

El cálculo del valor extrínseco es más complicado. Primero, se escoge una función de afinidad para cuantificar las relaciones. Luego, como el valor extrínseco tiene el problema de la infinita recursividad (presente en la ecuación de la definición 3.2), para calcularlo, aproximaremos el valor de $\mp(X_i)$ en la fórmula original por $I(X_i)$. Así, eliminamos la recursividad. Una vez que tengamos números en vez de conjuntos, usaremos la suma en vez de la unión. La intersección de los valores semánticos recibidos por X_i y X_j para x la aproximamos como el resultado de propagar $I(X_i)$ a través de X_j a x.

Definición 3.4. *El* ***valor extrínseco*** *de un nodo $x \in V$ se define como E(x), siendo*

$$E(x) = \sum_{i=1}^{a} \text{máx}\{F_A(X_i, x)I(X_i) - \sum_{j \in J} F_A(X_i, X_j)I(X_i)F_A(X_j, x), 0\}$$

donde $J = \{j \in \{1, ..., a\}, i \neq j\}$.

Finalmente, podemos calcular $\mathcal{S}(x)$ de forma análoga a como se ha realizado en la definición 3.1.

Definición 3.5. *Se define el* ***valor semántico*** *de un nodo $x \in V$ como la suma del valor extrínseco E y el intrínseco I,*

$$\mathcal{S}(x) = I(x) + E(x)$$

3.1.1. El valor semántico como medida de centralidad

Después de haber calculado el valor semántico $\mathcal{S}$ para cada nodo de una red, podemos analizar los resultados con alguna otra medida de centralidad. Para que un individuo tenga un alto $\mathcal{S}$, deberá tener altos I y E en comparación con el resto de individuos de la red.

Para obtener un valor elevado de I, x debe ser importante en el dominio original desde el que construimos la red. El valor de función I en x, $I(x)$, no aporta mucha información directa sobre la estructura de la red pero puede reforzar la importancia de x si x tiene también valores elevados con otras medidas de centralidad. Si x no posee valores elevados en las medidas de centralidad restantes pero tiene un alto $I(x)$, esto revela que x era importante en el dominio original de una manera que no se está teniendo en cuenta para construir la nueva red.

Para tener un valor E elevado, el individuo x debe tener conexiones con otros nodos con un alto I que no estén conectados entre ellos. Un valor bajo de E significa que las relaciones del individuo no son relevantes en la dinámica de la red, forma parte de un grupo pequeño o sus conexiones son débiles o redundantes.

Por ejemplo, en el caso específico de la asociación de palabras de un texto en una red, no es probable que un término de dominio específico tenga un alto valor semántico. Como señala a un concepto estrecho, no tendrá un valor de I elevado. Además, si esos términos de domino específico se relacionan con otros términos específicos del mismo dominio, éstos no tendrán muchas conexiones con los otros usuarios, por lo que el posible valor extrínseco está muy limitado. Por otro lado, los valores semánticos elevados indicarán que

se trata de un concepto muy general. Por ejemplo, si tomamos la palabra "mano", podría significar literalmente una mano, pero también podría ser usada en diferentes contextos, y está muy relacionada con el concepto general de "utilidad". Podría referirse a ser la "mano derecha de alguien", o a algo que está cerca, es decir, está "a mano". Los significados relativos a la palabra "mano" pueden ser más generales: "lavarse las manos" puede significar quitarse responsabilidades en una situación determinada, o se puede interpretar literalmente, etc.

3.2. Afinidad semántica

La afinidad semántica de dos individuos x e y mide la afinidad entre ellos basándose en la idea de cuántos cambios debemos hacer para convertir $\mathcal{S}(x)$ en $\mathcal{S}(y)$. De esta manera, palabras que tengan significados parecidos deberían tener afinidad semántica alta y las palabras que no estén relacionadas, baja. Por ejemplo, la afinidad semántica entre "hielo" y "agua" debería ser alta porque son términos muy cercanos y en la vida real sólo se necesita congelar el agua para obtener hielo. Sin embargo, la afinidad semántica entre el "agua" y la "tierra" sería menor, ya que la diferencia entre ellos en la vida real es mayor.

Para calcular esta afinidad nos basaremos en cómo de eficiente es convertir $\mathcal{S}(x)$ en $\mathcal{S}(y)$, convirtiendo el grafo en un problema de flujo.

Podemos calcular la afinidad semántica basándonos en cuán eficiente es convertir $\mathcal{S}(x)$ en $\mathcal{S}(y)$. Modelaremos $\mathcal{S}(x)$ como si fuese un flujo que necesitamos transportar de x a y. Cada individuo x tiene una capacidad igual a su valor semántico $\mathcal{S}(x)$ y cada eje $x \to y$ puede llevar $FC(x, y) \cdot \mathcal{S}(y)$ a y. Es decir, cada eje se trata como si fuese un "tubo" que lleva el flujo y cada individuo es una bifurcación en el camino. Entonces, necesitamos llevar todo el flujo desde el individuo "fuente" hasta el individuo "salida" usando el camino más eficiente.

Existen diferentes posibilidades para definir "eficiencia" en este caso. Si consideramos que la eficiencia significa que no se requieren muchos indivi-

duos para el transporte, deberíamos denotar el camino más eficiente como aquel con menos intermediadores. También se podría definir eficiencia usando solo buenas conexiones, en este caso el mejor camino sería aquel con el mayor valor de afinidad. En este trabajo optaremos por esta última fórmula y denotaremos el camino más eficiente como aquel con la mayor media del valor de la afinidad.

Este problema es similar al "Shortest Capacitated Path Problem" [5], que consiste en buscar un conjunto de caminos por ejes desvinculados que conecte todos los nodos en un grafo, pero en nuestro caso tendremos solo en cuenta un camino, de x a y. El clásico "shortest path optimization problem" (problema de optimización del camino más corto) entre un par de individuos [10] está también muy relacionado con nuestra tarea. Pero este problema halla el camino más corto y nosotros necesitamos el más eficiente.

De esta manera, obtenemos el camino más corto posible usando todos los ejes que no están aún "completos". Normalmente, se requiere de más de un camino para llevar todo el valor semántico del origen a su destino. Entonces, necesitamos calcular un nuevo camino más corto con los nodos y ejes posibles cada vez que el camino actual llega a su límite de capacidad.

Para hallar este camino eficiente aplicaremos el Algoritmo de Ford-Fulkerson-Edmonds-Karp para cada problema de flujo entre cada par de palabras. Para usar este algoritmo, necesitaremos crear dos nodos imaginarios como "Fuente" y "Salida". A menudo, no será posible llevar todo el valor semántico de un nodo a otro, por diversas razones. Por ejemplo, si el camino es de x a y, y $S(x) > S(y)$ no se podría llevar todo el valor semántico necesario ya que y no puede recibirlo. En otros casos, aunque el destino pueda recibir ese valor semántico, puede suceder que en la red haya un corte con valor de flujo inferior al necesario. Imaginemos que queremos llevar el valor semántico de x a y. Por estas razones, crearemos un nodo ficticio F y un eje $F \to x$ con peso igual al valor semántico de x, $\mathcal{S}(x)$. Además, creamos un nodo S y un eje $y \to S$ con peso igual al valor semántico de y, $\mathcal{S}(y)$. Una vez llevemos todo el líquido de un nodo hasta otro, calculamos el resultado final teniendo en cuenta la diferencia original en sus valores semánticos $\mathcal{S}$ y el valor de afinidad medio en los ejes usados en el camino.

Algunos individuos seguramente tengan valores de afinidad bajos, por ejemplo si tienen muchas conexiones, los valores de afinidad medios de un

camino pueden ser decepcionantemente bajos. Con el fin de comparar las diferentes afinidades semánticas que origina un individuo x, reescalamos el resultado con el valor máximo de afinidad semántica que x emite.

Definición 3.6. *Dados dos vértices cualesquiera $x, y \in V$ tales que $x \neq y$, se define la* ***afinidad semántica*** *entre x e y, y se denota $A(x, y)$, como:*

$$A(x,y) = \left(1 - \frac{|\mathcal{S}(x) - \mathcal{S}(y)|}{\text{máx}(\mathcal{S}(x), \mathcal{S}(y))}\right) \frac{\sum P}{|P|} \cdot \frac{1}{\vee_{z \in Z} F_C(x,z)}$$

donde P es la lista de los valores de afinidad en los ejes usados en los caminos para llevar el valor semántico $\mathcal{S}(x)$ de x a y y Z es el conjunto de todos los individuos conectados a x.

Imaginemos la idea de calcular la afinidad semántica entre dos palabras iguales. "Convertir" esa palabra en ella misma sería el proceso más asequible, ya que sería inmediato. Por esta razón, la afinidad semántica entre un vértice y él mismo se ha tomado igual a 1, por convenio.

Para calcular la afinidad semántica en nuestro experimento, se ha usado una combinación de las afinidades del Mejor Amigo (MA) y de Macchiavelli (MAC) como funciones $F_C(x, y)$. Usando esta combinación de funciones afines podemos caracterizar cada eje basándonos en la importancia de la relación de pareja entre x e y, y también tendremos en cuenta la importancia relativa que tienen sus círculos sociales dentro de la red. Esto es importante por dos razones:

1. En individuos con un alto grado de centralidad, la afinidad del Mejor Amigo (MA) es necesariamente baja, lo que resulta en un valor semántico artificialmente bajo.

2. La afinidad de Macchiavelli da alta afinidad a individuos que tienen un rol similar en la red. En los textos que estamos estudiando, esto nos da valores altos en afinidad entre conceptos que tienen un rol similar en los textos.

De esta manera, es natural pensar que estos individuos se transmiten información entre ellos mismos, a pesar de que la afinidad del Mejor Amigo entre ellos no es alta. Hemos combinado ambas funciones de afinidad utilizando una función convexa, así el valor de cada eje es el 90 % la afinidad MA

y el 10 % restante la afinidad de Machiavelli, pero cambiamos a 0 todos los valores de afinidad en los ejes donde el valor original del MA era 0.

Capítulo 4

Análisis de un texto

Para entender cómo se aplica el método anteriormente explicado, utilizaremos un texto corto como ejemplo.

El texto que se da a continuación es una sinopsis de las dos últimas películas de Harry Potter. El primer párrafo es el resumen de la primera parte de Harry Potter y las Reliquias de la Muerte y el segundo de la parte dos. Se ha decidido usar este ejemplo ya que se repiten los personajes y los conceptos más relevantes de las películas.

Harry, Ron y Hermione se lanzan a la búsqueda de los horrocruxes que guardan una parte del alma de Voldemort y los necesita para sobrevivir. Sólo cuando los hayan destruido, podrán vencer a Voldemort. Pero en su búsqueda, deberán hacer frente a muchos peligros.
Harry, Ron y Hermione continuan la búsqueda de los horrocruxes para vencer a Voldemort. Harry deberá usar todos los conocimientos adquiridos gracias a Dumbledore sobre Voldemort para poder sobrevivir y derrotarlo.

A simple vista se entiende que Harry y Voldemort son los protagonistas principales y luego hay conceptos y personajes secundarios. Podremos ver cómo después de haber convertido los párrafos en una red, el análisis

semántico nos dará estos resultados de manera matemática.

En este caso se trata de un pequeño texto que podría analizarse directamente, pero este ejemplo servirá para ver el potencial de este método en el análisis automático de grandes textos.

4.1. Construcción de la red

Definimos la red $G = (V, T)$ donde necesitaremos escoger criterios para crear el conjunto de vértices V y el conjunto de arcos T.

Las palabras más relevantes de un texto, son los nombres propios e impropios. Por lo tanto, no pondremos atención al resto de palabras como verbos, preposiciones, adverbios, etc. De esta manera, cada sustantivo será un nodo de la red. Es decir, el orden de nuestro grafo será igual al número de sustantivos diferentes que haya en el texto.

Así, en nuestro caso el conjunto V será el siguiente:

$$V = \{Harry, Ron, Hermione, búsqueda, horrocruxes, alma,$$

$$Voldemort, peligros, conocimientos, Dumbledore\}$$

El orden de nuestra red será el cardinal del conjunto V,

$$o(G) = |V| = 10$$

La siguiente cuestión es cómo crear las relaciones de la red. Se podrían seguir diferentes criterios para conectar las palabras. En nuestro caso, seguiremos el criterio de la distancia k. Al ser un texto corto, escogemos $k = 7$. Entonces, dos sustantivos estarán conectados si están a menos de siete palabras uno del otro.

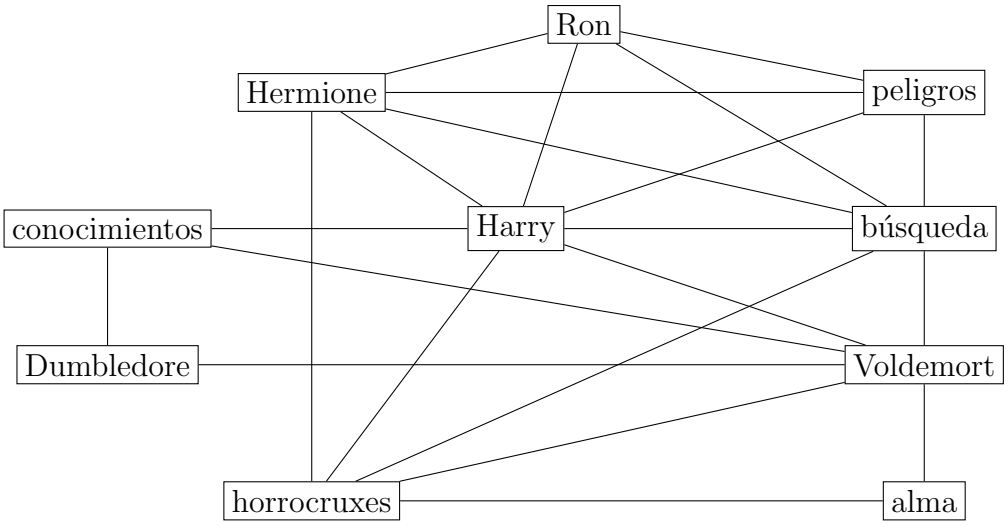
G será un grafo no dirigido ya que simplemente miraremos si dos palabras están a cierta distancia sin importar cuál de ellas aparece primero. Es decir, no importará el orden de los pares del conjunto T.

La matriz de adyacencia A asociada al grafo G se da en la Tabla 4.1.

Nodos	Harry	Ron	Hermione	búsqueda	horrocruxes	alma	Voldemort	peligros	conocimientos	Dumbledore
Harry	0	1	1	1	1	0	1	1	1	0
Ron	1	0	1	1	0	0	0	1	0	0
Hermione	1	1	0	1	1	0	0	1	0	0
búsqueda	1	1	1	0	1	0	1	1	0	0
horrocruxes	1	0	1	1	0	1	1	0	0	0
alma	0	0	0	0	1	0	1	0	0	0
Voldemort	1	0	0	1	1	1	0	0	1	1
peligros	1	1	1	1	0	0	0	0	0	0
conocimientos	1	0	0	0	0	0	1	0	0	1
Dumbledore	0	0	0	0	0	0	1	0	1	0

Tabla 4.1: Matriz de adyacencia A.

Teniendo la matriz de adyacencia, es fácil representar la red:



4.2. Cálculo de centralidades

Ahora que tenemos el texto convertido en una red, podemos proceder a calcular cómo de central es cada palabra, es decir, hallar la centralidad de cada nodo.

4.2.1. Centralidad de grado

Podemos hallar esta centralidad de manera sencilla calculando cuántas conexiones directas tiene cada palabra. De esta manera, la centralidad de grado asociada a cada nombre del texto queda reflejada en la Tabla 4.2.

	Harry	Ron	Hermione	búsqueda	horrocruxes
C_D	7	4	5	6	5
C'_D	7/9	4/9	5/9	2/3	5/9

	alma	Voldemort	peligros	conocimientos	Dumbledore
C_D	2	6	4	3	2
C'_D	2/9	2/3	4/9	1/3	2/9

Tabla 4.2: Centralidad de grado para cada palabra.

Podemos apreciar como Harry obtiene el mayor valor de esta centralidad, cosa que siendo el protagonista de las películas es lógico. Detrás de él, Voldemort y la palabra búsqueda son las más importantes. Voldemort es el coprotagonista de las películas, y éstas se resumen en la búsqueda de los horrocruxes, que también tienen un valor alto de centralidad. Dumbledore y la palabra alma son las menos relevantes de acuerdo a la centralidad de grado.

4.2.2. Centralidad de proximidad

Para el cálculo de esta centralidad, necesitamos la matriz de distancias D asociada al grafo G, que se da a continuación:

Nodos	Harry	Ron	Hermione	búsqueda	horrocruxes	alma	Voldemort	peligros	conocimientos	Dumbledore
Harry	0	1	1	1	1	2	1	1	1	2
Ron	1	0	1	1	2	3	2	1	2	3
Hermione	1	1	0	1	1	2	2	1	2	3
búsqueda	1	1	1	0	1	2	1	1	2	2
horrocruxes	1	2	1	1	0	1	1	2	2	2
alma	2	3	2	2	1	0	1	3	2	2
Voldemort	1	2	2	1	1	1	0	2	1	1
peligros	1	1	1	1	2	3	2	0	2	3
conocimientos	1	2	2	2	2	2	1	2	0	1
Dumbledore	2	3	3	2	2	2	1	3	1	0

Tabla 4.3: Matriz de distancias D.

Usando la matriz de la Tabla 4.3 se obtiene fácilmente la centralidad de proximidad de cada palabra. Dichos valores se recogen en la Tabla 4.4 que recoge los datos para cada nombre.

	Harry	Ron	Hermione	búsqueda	horrocruxes
C_P	0,091	0,063	0,071	0,083	0,077
C'_P	0,818	0,563	0,643	0,750	0,692

	alma	Voldemort	peligros	conocimientos	Dumbledore
C_P	0,056	0,083	0,063	0,067	0,053
C'_P	0,500	0,750	0,563	0,600	0,474

Tabla 4.4: Centralidad de proximidad para cada palabra.

Harry es el nodo más próximo al resto de nodos de la red. Además, Voldemort y la palabra búsqueda obtienen valores altos en esta centralidad.

4.2.3. Centralidad del vector propio

Para hallar esta centralidad necesitamos hallar los autovalores de la matriz A.

Con la ayuda de Matlab hayamos el siguiente determinante que nos dará el polinomio característico, y al igualarlo a 0 tendremos los autovalores.

$$det(A-\lambda I) = \lambda^{10}-22\lambda^8-38\lambda^7+69\lambda^6+222\lambda^5+123\lambda^4-134\lambda^3-191\lambda^2-82\lambda-12$$

$$det(A-\lambda I) = 0 \Leftrightarrow \lambda_1 = -2{,}23, \lambda_2 = -1{,}85, \lambda_3 = -1{,}59, \lambda_4 = -1,$$

$$\lambda_5 = -0{,}7, \lambda_6 = -0{,}56, \lambda_7 = -0{,}39, \lambda_8 = 1{,}07, \lambda_9 = 2{,}25, \lambda_{10} = 4{,}99$$

Escogemos el autovalor $\lambda = \text{máx}\{|\lambda_i| \ , \ i \in [1, 10]\} = 4{,}99$. El autovector asociado a este autovalor será el vector C que buscamos. Lo calculamos resolviendo la ecuación matricial $(A - 4{,}99I)\, C = 0$.

A es una matriz de dimensión 10, por lo que nos quedará un sistema de 10 ecuaciones y 10 incógnitas. Esto sería complicado de resolver de forma manual, con lo que se ha calculado la centralidad del vector propio con ayuda del programa GEPHI. Los valores para esta centralidad se recogen en la Tabla 4.5.

	Harry	Ron	Hermione	búsqueda	horrocruxes
C_{VP}	1	0,684	0,810	0,933	0,757

	alma	Voldemort	peligros	conocimientos	Dumbledore
C_{VP}	0,300	0,728	0,684	0,396	0,229

Tabla 4.5: Centralidad del vector propio de cada palabra.

Notemos que Harry obtiene una centralidad del vector propio del 100 %, sigue siendo el nodo más importante. En cambio, los horrocruxes en este caso obtienen un valor mayor que Voldemort.

4.2.4. Centralidad de intermediación

Como máx$\{d_{ij}\} = 3$, necesitamos las matrices A, A^2 y A^3 que se recogen en las tablas 4.1, 4.6 y 4.7 respectivamente.

Nodos	Harry	Ron	Hermione	búsqueda	horrocruxes	alma	Voldemort	peligros	conocimientos	Dumbledore
Harry	7	3	4	5	3	2	3	3	1	2
Ron	3	4	3	3	3	0	2	3	1	0
Hermione	4	3	5	4	2	1	3	3	1	0
búsqueda	5	3	4	6	3	2	2	3	2	1
horrocruxes	3	3	2	3	5	1	3	3	2	1
alma	2	0	1	2	1	2	1	0	1	1
Voldemort	3	2	3	2	3	1	6	2	2	1
peligros	3	3	3	3	3	0	2	4	1	0
conocimientos	1	1	1	2	2	1	2	1	3	1
Dumbledore	2	0	0	1	1	1	1	0	1	2

Tabla 4.6: Número de caminos mínimos de longitud 2 entre pares de palabras.

Nodos	Harry	Ron	Hermione	búsqueda	horrocruxes	alma	Voldemort	peligros	conocimientos	Dumbledore
Harry	22	19	21	23	21	6	20	19	12	4
Ron	19	12	16	18	11	5	10	13	5	3
Hermione	21	16	16	20	17	5	12	16	7	4
búsqueda	23	18	20	20	19	5	19	18	8	4
horrocruxes	21	11	17	19	12	8	15	11	7	5
alma	6	5	5	5	8	2	9	5	4	2
Voldemort	20	10	12	19	15	9	12	10	10	8
peligros	19	13	16	18	11	5	10	12	5	3
conocimientos	12	5	7	8	7	4	10	5	4	5
Dumbledore	4	3	4	4	5	2	8	3	5	2

Tabla 4.7: Número de caminos mínimos de longitud 3 entre pares de palabras.

Combinando los datos de la Tabla 4.3 con las matrices A, A^2 y A^3, tenemos el número de caminos mínimos entre cada par de nombres, que se recoge en la Tabla 4.8.

Nodos	Harry	Ron	Hermione	búsqueda	horrocruxes	alma	Voldemort	peligros	conocimientos	Dumbledore
Harry	0	1	1	1	1	2	1	1	1	2
Ron	1	0	1	1	3	5	2	1	1	3
Hermione	1	1	0	1	1	1	3	1	1	4
búsqueda	1	1	1	0	1	2	1	1	2	1
horrocruxes	1	3	1	1	0	1	1	3	2	1
alma	2	5	1	2	1	0	1	5	1	1
Voldemort	1	2	3	1	1	1	0	2	1	1
peligros	1	1	1	1	3	5	2	0	1	3
conocimientos	1	1	1	2	2	1	1	1	0	1
Dumbledore	2	3	4	1	1	1	1	3	1	0

Tabla 4.8: Número de caminos de longitud mínima entre cada par de vértices.

Entonces, para calcular cada valor de intermediación necesitamos fijarnos en los caminos mínimos que tengan intermediarios, es decir, de longitud 2 o más.

En la Tabla 4.9 se presentan los valores de la centralidad de intermediación de la red.

	Harry	Ron	Hermione	búsqueda	horrocruxes
C_I	17,270	0	2,130	7,430	7,570
C'_I	0,240	0	0,030	0,103	0,105

	alma	Voldemort	peligros	conocimientos	Dumbledore
C_I	0	9,383	0	1,417	0
C'_I	0	0,261	0	0,020	0

Tabla 4.9: Centralidad de intermediación de cada palabra.

Harry, que era el más central teniendo en cuenta las centralidades anteriores, ahora es el segundo. Voldemort es el nodo que más aparece en caminos

mínimos entre pares vértices. Además, Ron, Dumbledore y las palabras alma y peligros no sirven de intermediario en ningún caso.

4.3. Afinidad entre pares de vértices

Calculamos ahora los valores para cada tipo de afinidad entre todos los pares de individuos de nuestra red. Para las afinidades MA, MAC Y Mach se dan simplemente los valores de dichas afinidades. Los cálculos son asequibles aplicando las definiciones explicadas en el Capítulo 2. Por lo que, no se detallan explícitamente. Nos centraremos, en cambio, en el procedimiento para el cálculo de la afinidad semántica.

Los valores de afinidad y de centralidad se analizarán en conjunto al final del trabajo en el apartado Análisis de resultados.

4.3.1. Afinidad del Mejor Amigo

Nodos	Harry	Ron	Hermione	búsqueda	horrocruxes	alma	Voldemort	peligros	conocimientos	Dumbledore
Harry	0	1/7	1/7	1/7	1/7	0	1/7	1/7	1/7	0
Ron	1/4	0	1/4	1/4	0	0	0	1/4	0	0
Hermione	1/5	1/5	0	1/5	1/5	0	0	1/5	0	0
búsqueda	1/6	1/6	1/6	0	1/6	0	1/6	1/6	0	0
horrocruxes	1/5	0	1/5	1/5	0	1/5	1/5	0	0	0
alma	0	0	0	0	1/2	0	1/2	0	0	0
Voldemort	1/6	0	0	1/6	1/6	1/6	0	0	1/6	1/6
peligros	1/4	1/4	1/4	1/4	0	0	0	0	0	0
conocimientos	1/3	0	0	0	0	0	1/3	0	0	1/3
Dumbledore	0	0	0	0	0	0	1/2	0	1/2	0

Tabla 4.10: Afinidad del Mejor Amigo entre pares de vértices.

4.3.2. Afinidad del Mejor Amigo Común

Nodos	Harry	Ron	Hermione	búsqueda	horrocruxes	alma	Voldemort	peligros	conocimientos	Dumbledore
Harry	0	1/7	1/7	1/7	1/7	1/7	1/7	1/7	1/7	1/7
Ron	1/4	0	1/4	1/4	1/4	0	1/4	1/4	1/4	0
Hermione	1/5	1/5	0	1/5	1/5	1/5	1/5	1/5	1/5	0
búsqueda	1/6	1/6	1/6	0	1/6	1/6	1/6	1/6	1/6	1/6
horrocruxes	1/5	1/5	1/5	1/5	0	1/5	1/5	1/5	1/5	1/5
alma	1/2	0	1/2	1/2	1/2	0	1/2	0	1/2	1/2
Voldemort	1/6	1/6	1/6	1/6	1/6	1/6	0	1/6	1/6	1/6
peligros	1/4	1/4	1/4	1/4	1/4	0	1/4	0	1/4	0
conocimientos	1/3	1/3	1/3	1/3	1/3	1/3	1/3	1/3	0	1/3
Dumbledore	1/2	0	0	1/2	1/2	1/2	1/2	0	1/2	0

Tabla 4.11: Afinidad del Mejor Amigo Común entre pares de vértices.

4.3.3. Afinidad de Machiavelli

Para el cálculo de esta centralidad primero necesitamos calcular el valor de W para cada palabra, que se dan a continuación:

$$\begin{array}{l|l} W_{Harry} = 33 & W_{alma} = 11 \\ W_{Ron} = 22 & W_{Voldemort} = 25 \\ W_{Hermione} = 26 & W_{peligros} = 22 \\ W_{búsqueda} = 31 & W_{conocimientos} = 15 \\ W_{horrocruxes} = 26 & W_{Dumbledore} = 9 \end{array}$$

Con la ayuda de estos valores, ya se puede hallar la afinidad de Machiavelli para cada par de vértices del grafo. Dichos valores se presentan en la Tabla 4.12.

Nodos	Harry	Ron	Hermione	búsqueda	horrocruxes	alma	Voldemort	peligros	conocimientos	Dumbledore
Harry	1	0,670	0,780	0,930	0,780	0,300	0,750	0,670	0,450	0,270
Ron	0,670	1	0,846	0,710	0,846	0,500	0,880	1	0,681	0,409
Hermione	0,780	0,846	1	0,839	1	0,423	0,962	0,846	0,577	0,346
búsqueda	0,930	0,710	0,839	1	0,839	0,356	0,806	0,710	0,484	0,290
horrocruxes	0,780	0,846	1	0,839	1	0,423	0,962	0,846	0,577	0,346
alma	0,330	0,500	0,423	0,356	0,423	1	0,440	0,500	0,730	0,810
Voldemort	0,760	0,880	0,962	0,806	0,962	0,440	1	0,880	0,600	0,360
peligros	0,670	1	0,846	0,710	0,846	0,500	0,880	1	0,681	0,409
conocimientos	0,450	0,681	0,577	0,484	0,577	0,730	0,600	0,681	1	0,600
Dumbledore	0,270	0,409	0,346	0,290	0,346	0,810	0,360	0,409	0,600	1

Tabla 4.12: Afinidad de Machiavelli entre pares de vértices.

4.3.4. Afinidad semántica

Para calcular la afinidad semántica entre cada par de palabras, primero tenemos que hallar el valor semántico de cada una de ellas. Recordemos que el valor semántico $\mathcal{S}$ era la suma del valor extrínseco E y el intrínseco I.

Valor Intrínseco

El valor intrínseco de cada palabra se ha aproximado por la frecuencia con la que aparece en el texto. Dichos valores se dan en la Tabla 4.13.

	Harry	Ron	Hermione	búsqueda	horrocruxes
I	3	2	2	3	2

	alma	Voldemort	peligros	conocimientos	Dumbledore
I	1	4	1	1	1

Tabla 4.13: Valor Intrínseco de cada palabra.

Valor Extrínseco

Para hallar el valor extrínseco necesitamos usar una función de afinidad convexa, que refleje los valores de la afinidad del Mejor Amigo en un 90 % y de Machiavelli en un 10 %, para tener en cuenta la afinidad estructural y la personal. Estos valores se recogen en la Tabla 4.14.

Nodos	Harry	Ron	Hermione	búsqueda	horrocruxes	alma	Voldemort	peligros	conocimientos	Dumbledore
Harry	0	0,195	0,207	0,223	0,207	0	0,204	0,195	0,174	0
Ron	0,292	0	0,310	0,296	0	0	0	0,325	0	0
Hermione	0,259	0,265	0	0,264	0,28	0	0	0,265	0	0
búsqueda	0,244	0,221	0,234	0	0,234	0	0,231	0,221	0	0
horrocruxes	0,259	0	0,280	0,264	0	0,222	0,276	0	0	0
alma	0	0	0	0	0,492	0	0,546	0	0	0
Voldemort	0,225	0	0	0,231	0,246	0,194	0	0	0,210	0,186
peligros	0,292	0,325	0,310	0,296	0	0	0	0	0	0
conocimientos	0,345	0	0	0	0	0	0,360	0	0	0,600
Dumbledore	0	0	0	0	0	0	0,486	0	0,510	0

Tabla 4.14: Afinidad entre pares de vértices para el valor extrínseco.

Teniendo ya la función de afinidad que vamos a utilizar, veamos un ejemplo de cómo hallar el valor extrínseco de Ron.

Primero, definimos el conjunto X que contiene las conexiones del nodo, y procedemos a calcular E(Ron).

$$X_{Ron} = \{Harry, Hermione, búsqueda, peligros\}$$

$$\textbf{E(Ron)} = \text{máx}\{0, I(Harry)[F_E(Harry, Ron) - F_E(Harry, Hermione) F_E(Hermione, Ron) - F_E(Harry, búsqueda)F_E(búsqueda, Ron) - F_E(Harry, peligros)F_E(peligros, Ron)]\} + \text{máx}\{0, I(Hermione)[F_E(Hermione, Ron) - F_E(Hermione, Harry)F_E(Harry, Ron) - F_E(Hermione, búsqueda)F_E(búsqueda, Ron) - F_E(Hermione, peligros) F_E(peligros, Ron)]\} + \text{máx}\{0, I(búsqueda)[F_E(búsqueda, Ron) - F_E(búsqueda, Harry)F_E(Harry, Ron) - F_E(búsqueda, Hermione) F_E(Hermione, Ron) - F_E(búsqueda, peligros)F_E(peligros, Ron)]\} + \text{máx}\{0, I(peligros)[F_E(peligros, Ron) - F_E(peligros, Harry) F_E(Harry, Ron)F_E(peligros, Hermione)F_E(Hermione, Ron) - F_E(peligros, búsqueda)F_E(búsqueda, Ron)]\} = 0{,}082461 + 0{,}140052 + 0{,}118755 + 0{,}120494 = 0{,}461762$$

Siguiendo el mismo procedimiento para el resto de palabras, tenemos en la Tabla 4.15 recogidos todos los valores extrínsecos, redondeados a las milésimas.

	Harry	Ron	Hermione	búsqueda	horrocruxes
E	0,648	0,462	0,568	0,759	1,210

	alma	Voldemort	peligros	conocimientos	Dumbledore
E	0,894	1,243	0,582	1,105	0,773

Tabla 4.15: Valor extrínseco de cada palabra.

Valor Semántico

Sumando el valor extrínseco y el valor intrínseco de cada palabra se halla el valor semántico. Los valores semánticos se recogen en la Tabla 4.16.

	Harry	Ron	Hermione	búsqueda	horrocruxes
$\mathcal{S}$	3,648	2,462	2,568	3,759	3,210

	alma	Voldemort	peligros	conocimientos	Dumbledore
$\mathcal{S}$	1,894	5,243	1,582	2,105	1,773

Tabla 4.16: Valor semántico de cada palabra.

Analizando los valores semánticos en la Tabla 4.16, se observa que Voldemort es con diferencia el que más valor semántico obtiene. Podríamos pensar que es por aparecer cuatro veces en el texto y tener un valor intrínseco muy alto, pero si nos fijamos en la Tabla 4.15 también supera en valor extrínseco al resto. En Harry Potter y las Reliquias de la Muerte se revela la verdadera preocupación de Voldemort, que no se sabía en las anteriores películas. Voldemort necesita hacerse con el poder y además conocemos por primera vez lo que es un horrocrux, un objeto donde se puede guardar una parte del alma y así conseguir ser inmortal. Los horrocruxes también tienen un valor semántico alto gracias al valor extrínseco, siendo éste cercano al de Voldemort. Aquí podemos apreciar la importancia en un texto de tener en cuenta el valor extrínseco de las palabras.

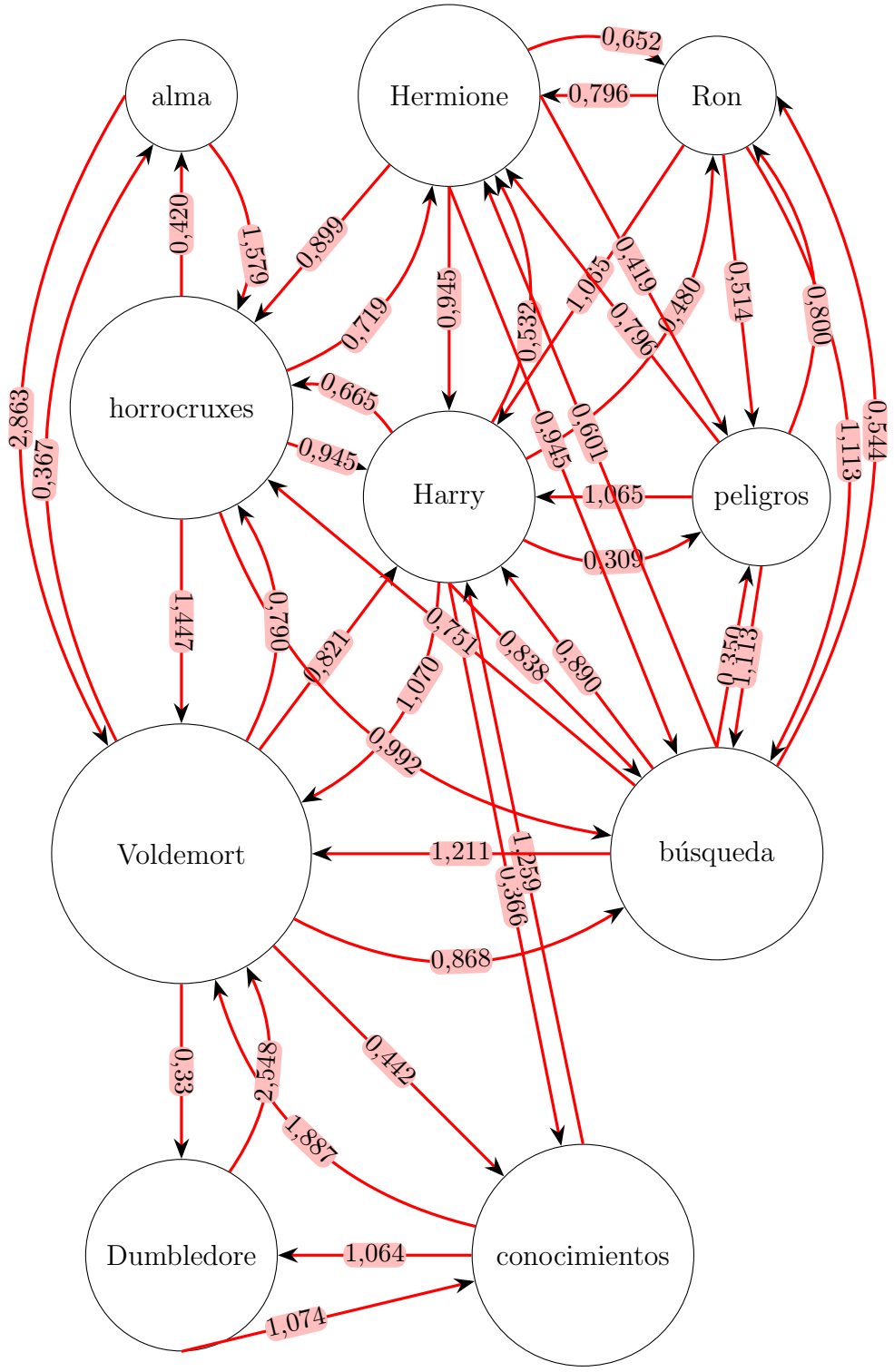
Afinidad semántica

Ahora que tenemos todos los valores semánticos podemos proceder a calcular la afinidad semántica entre cada par de palabras. Para ello necesitamos calcular primero la cantidad de valor semántico que puede llevar cada eje en la red. Cada origen $x \in V$ puede llevar a $y \in V$ $F(x,y)\mathcal{S}(y)$ de valor semántico. Recogemos en la Tabla 4.17 los valores de cada eje.

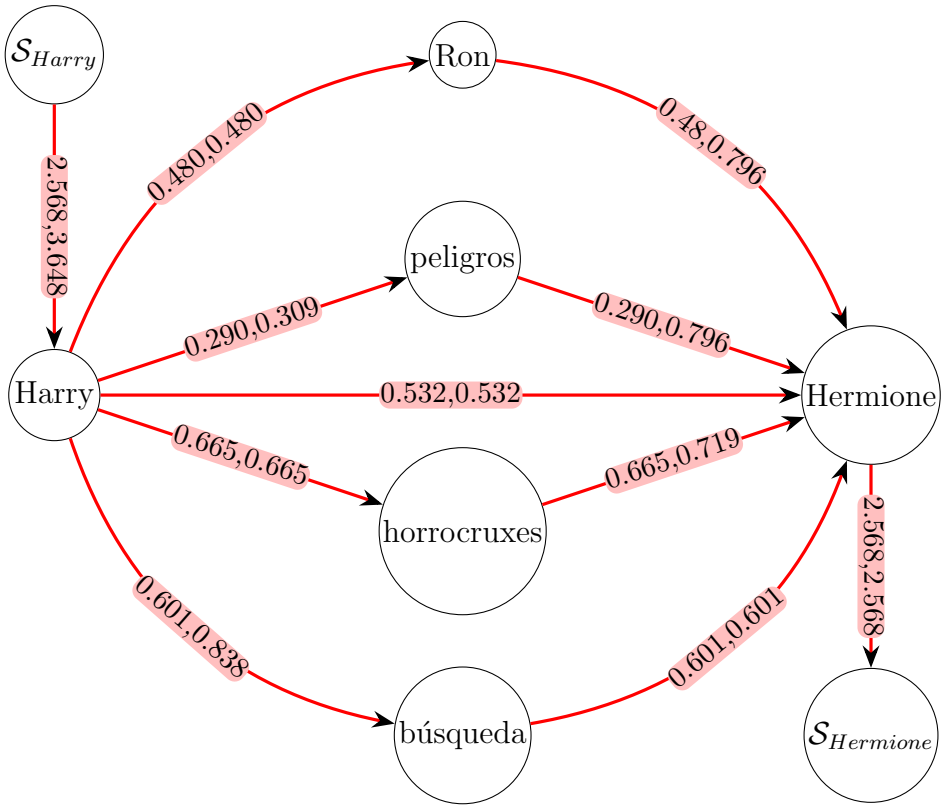
Nodos	Harry	Ron	Hermione	búsqueda	horrocruxes	alma	Voldemort	peligros	conocimientos	Dumbledore
Harry	0	0,480	0,532	0,838	0,665	0	1,070	0,309	0,366	0
Ron	1,065	0	0,796	1,113	0	0	0	0,514	0	0
Hermione	0,945	0,652	0	0,992	0,899	0	0	0,419	0	0
búsqueda	0,890	0,544	0,601	0	0,751	0	1,211	0,350	0	0
horrocruxes	0,945	0	0,719	0,992	0	0,420	1,447	0	0	0
alma	0	0	0	0	1,579	0	2,863	0	0	0
Voldemort	0,821	0	0	0,868	0,790	0,367	0	0	0,442	0,330
peligros	1,065	0,800	0,796	1,113	0	0	0	0	0	0
conocimientos	1,259	0	0	0	0	0	1,887	0	0	1,064
Dumbledore	0	0	0	0	0	0	2,548	0	1,074	0

Tabla 4.17: Máximo valor semántico posible que puede llevar cada eje.

La red en la que estudiaremos los problemas de flujo máximo queda reflejada en la página siguiente. No se han representado los arcos para el valor semántico de cada palabra, ya que sería de un tamaño demasiado grande para dibujarlo en las dimensiones que trabajamos.


alma
Hermione
Ron
horrocruxes
Harry
peligros
Voldemort
búsqueda
Dumbledore
conocimientos
0,652
0,796
0,420
1,579
0,899
0,719
0,945
0,532
1,063
0,419
0,480
0,514
0,800
2,863
0,367
0,665
0,945
0,945
0,601
0,544
1,113
1,065
0,309
1,447
0,790
0,821
0,751
0,838
0,890
1,070
0,992
0,350
1,113
1,211
1,259
0,366
0,868
0,33
2,548
0,442
1,887
1,064
1,074

Veamos un ejemplo de cómo calcular la afinidad semántica entre Harry y Hermione. Como $\mathcal{S}(Harry) > \mathcal{S}(Hermione)$ existe la posibilidad de llevar el máximo valor semántico que Hermione puede recibir.

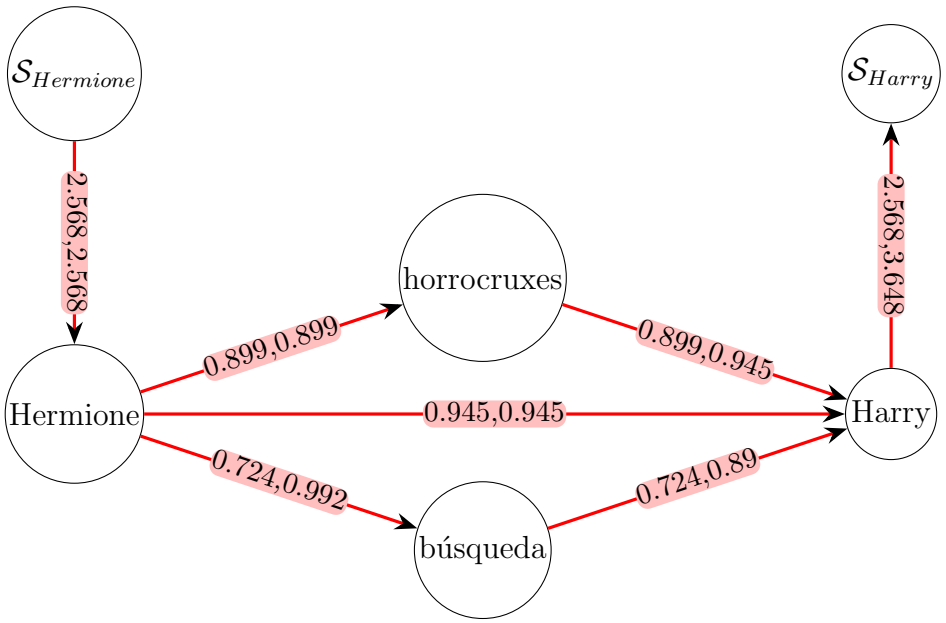


Así, Hermione obtiene todo su valor semántico a partir de Harry, por medio de Ron, peligros, horrocruxes y búsqueda. Además, Harry no se ha "vaciado" completamente, ya que tenía mayor valor semántico que Hermione. Para calcular su afinidad semántica debemos mirar el valor de afinidad en todos los ejes usados para llevar el valor semántico. Entonces, utilizando la definición de la afinidad semántica:

$$A(Harry, Hermione) = \left(1 - \frac{|3{,}648 - 2{,}568|}{3{,}648}\right) \frac{2{,}161}{9} \frac{1}{0{,}223} = 0{,}758$$

Ahora calculemos la afinidad en el otro sentido, es decir, el valor de $A(Hermione, Harry)$. Como $\mathcal{S}(Hermione) < \mathcal{S}(Harry)$ no podremos "lle-

nar" a Harry pero si existe la posibilidad de llevar todo el valor semántico de Hermione hasta él. El problema de flujo resuelto sería de la siguiente manera:



$$A(Hermione, Harry) = \left(1 - \frac{|2{,}568 - 3{,}648|}{3{,}648}\right) \frac{1{,}306}{5} \frac{1}{0{,}28} = 0{,}657$$

Después de hacer todos los problemas de flujo entre cada par de vértices, recogemos los datos en la Tabla 4.18.

Nodos	Harry	Ron	Hermione	búsqueda	horrocruxes	alma	Voldemort	peligros	conocimientos	Dumbledore
Harry	1	0,712	0,758	0,998	1	0,481	0,739	0,450	0,664	0,599
Ron	0,582	1	0,809	0,577	0,618	0,530	0,390	0,529	0,714	0,630
Hermione	0,657	0,841	1	0,651	0,755	0,640	0,441	0,567	0,793	0,675
búsqueda	0,981	0,646	0,717	1	0,899	0,454	0,807	0,420	0,704	0,530
horrocruxes	0,811	0,678	0,704	0,783	1	0,486	0,540	0,447	0,630	0,561
alma	0,329	0,423	0,409	0,319	0,463	1	0,361	0,414	0,598	0,661
Voldemort	0,711	0,469	0,489	0,785	0,644	0,324	1	0,290	0,426	0,384
peligros	0,404	0,560	0,597	0,396	0,409	0,595	0,237	1	0,627	0,780
conocimientos	0,276	0,366	0,366	0,241	0,274	0,383	0,323	0,327	1	0,536
Dumbledore	0,370	0,381	0,369	0,258	0,340	0,527	0,322	0,530	0,530	1

Tabla 4.18: Afinidad semántica entre cada par de palabras.

4.4. Análisis de resultados

Recordemos que la afinidad semántica entre dos palabras se basaba en la idea de cómo de eficiente sería "convertir" una en la otra. Según esto, una palabra con pocas conexiones y un valor semántico bajo, debería tener una afinidad alta con otras de las mismas características. Con esta idea, analizamos todos los resultados anteriores para cada una de las diez palabras.

- **Harry** era el nodo más central según todas las centralidades excepto la de intermediación, en la que era el segundo. Sólo por esto debería ser una palabra muy importante en el texto y, efectivamente es así ya que es el protagonista de las películas. Harry tiene conexión directa con 7 de los 10 individuos que hay en el texto, por lo que la afinidad del mejor amigo se dividía entre 7. Sin embargo, si tenemos en cuenta la afinidad del mejor amigo común, está conectado con todos los individuos (excepto él mismo). Esto suele pasar para nodos muy centrales en redes de un tamaño pequeño, ya que al tener muchas conexiones es muy probable que alguna de ellas esté conectada a los nodos restantes. Si nos fijamos en la Tabla 4.12 su mayor afinidad es con la palabra búsqueda. La afinidad de Machiavelli es la que más se parece a la afinidad semántica ya

que se fija en la estructura y no sólo en las conexiones. Centrándonos en el valor semántico, Harry tiene un valor intrínseco alto y sorprendentemente resulta en un valor extrínseco relativamente bajo respecto a otras palabras. Concretamente, hay otras 6 palabras con mayor valor extrínseco. Sin embargo, al calcular el valor semántico ya obtiene un valor mayor y se refleja que es muy importante en nuestro texto. En la Tabla 4.18 vemos que su mayor afinidad es con búsqueda (al igual que en la de Machiavelli) y con horrocruxes. Además, con una diferencia significativa, ya que obtiene un 99,8 % y un 100 %, respectivamente.

- **Ron y Hermione** tienen un rol similar en las películas (son los mejores amigos de Harry). Si se diferencian en algo en nuestra red, es porque Hermione tiene una conexión más que Ron. Esto no es demasiado importante y no altera los resultados de forma significativa. En un texto más grande como un libro, también sucedería que por muy similar que fuese el rol que tuviesen en la historia, llegarían a tener distintas conexiones. Lo importante es que en su mayoría, son muy parecidos. En las centralidades siempre han resultado tener valores similares y siempre Hermione sale más central que Ron, gracias a tener una conexión más. Respecto al valor semántico, tienen igual valor intrínseco y en el valor extrínseco se diferencian en una décima. Es obvio que sus valores semánticos resultan prácticamente iguales. Centrándonos en la Tabla 4.18, apreciamos que la columna 2 y la columna 3, tienen valores siempre muy cercanos. Esto es porque "convertir" a otras palabras en "Ron" o en "Hermione" sería un trabajo similar. También sucede con las filas 2 y 3, por la razón inversa, para llevar el valor semántico desde Ron o desde Hermione a otros nodos, el camino sería parecido en la mayoría de casos. Además, la mayor afinidad semántica de Ron es Hermione, con un 80,9 % y de Hermione, Ron con un 84,1 % de afinidad.

- **Voldemort** es el antagonista de la historia y justo en estas últimas películas cobra más importancia que nunca. Obtiene centralidades muy altas en todos los casos y en el caso de la centralidad de intermediación es el que mayor valor obtiene. Es el nodo más intermediador de la red ya que está conectado a palabras muy aisladas del resto, como por ejemplo, alma y Dumbledore. Además, al ser un nodo muy central tiene afinidad del mejor amigo común con todas las otras palabras. Estructuralmente, tiene más afinidad con Hermione y con horrocruxes, concretamente del 96,2 %, que es muy elevada. Respecto al valor semántico, es el que mayor resultado obtiene con diferencia, ya que tiene el mayor número de valor intrínseco y extrínseco. Voldemort es muy intermediador, por lo que

aparecerá muchas veces en los caminos de flujo máximo para calcular la afinidad semántica. Por último, su mayor afinidad semántica resulta con la palabra búsqueda (78,5 %) y con Harry (71,1 %).

- **Dumbledore** es de los individuos que menos centralidad obtiene con todas las medidas de centralidad. No sirve de intermediario en ningún camino de longitud mínima. Dumbledore muere en la primera parte de Harry Potter y las Reliquias de la Muerte, por lo que tiene sentido que no sea muy central en la historia. Aún así, es un personaje muy importante en general. Volvemos a notar la importancia del valor extrínseco de una palabra en un texto, Dumbledore tiene el quinto valor extrínseco más alto, por encima incluso de Harry y búsqueda, que son de las palabras más centrales del texto. Así, gracias al valor extrínseco, Dumbledore no obtiene el valor semántico más bajo. Dumbledore no obtiene demasiada afinidad semántica con ninguna otra palabra, la más alta es con las palabras peligros y conocimientos, que obtiene una afinidad del 53 %. Con la palabra alma también obtiene una afinidad alta, del 52,7 %.

- La trama principal de la historia es la **búsqueda** de los horrocruxes, esta palabra obtiene un valor de centralidad muy elevado en todos los casos. Según la centralidad del vector propio es la segunda más importante, con un 93,3 % , después de Harry que obtenía un 100 %. Vuelve a ocurrir que su valor extrínseco es relativamente bajo, como sucedía con Harry. Aún así, gracias a su importancia en el texto, obtiene el segundo valor semántico más alto, de 3,759. La mayor afinidad semántica, al igual que en la afinidad de Machiavelli, la obtiene con Harry con un 98,1 %. También con horrocruxes tienen una alta afinidad semántica (89,9 %).

- Los **horrocruxes** son parte del alma de Voldemort, por lo que para acabar con él, primero se debe encontrar y destruir cada uno de ellos. Además descubrimos al final de la historia que el propio protagonista, Harry, es un horrocrux. Esta palabra resulta en centralidades altas. En la que más destaca es en la de intermediación, ya que mientras que otros individuos obtenían un valor nulo, en este caso es la tercera palabra más intermediaria de la red. Además, obtiene afinidad MAC con todos los nodos, lo que quiere decir que está bien conectada. Estructuralmente, la afinidad de Machiavelli nos aporta que su mayor coincidencia es con Voldemort, concretamente del 96,2 %. Esto no debería sorprendernos cuando sabemos que los horrocruxes son parte de él. Los horrocruxes en valor intrínseco no sobresalían demasiado, pero en valor extrínseco,

sabiendo la importancia que de verdad se les da en la historia, obtienen el segundo valor más alto, después de Voldemort y con una diferencia no significativa (0,033). Respecto a la afinidad semántica, el mayor valor lo obtiene con Harry, un 81,1 %.

- Los **peligros** a los que se enfrentarán Harry, Hermione y Ron, el **alma** de Voldemort y los **conocimientos** de Harry adquiridos gracias a Dumbledore, son los conceptos menos relevantes en nuestro ejemplo. Son las palabras que menos centralidad obtienen en general (junto a Dumbledore). La palabra conocimientos sirve de intermediaria en algún camino de longitud mínima, ya que obtiene un valor mayor que cero en la centralidad de intermediación. Esto último, a diferencia de alma y peligros, que obtienen una intermediación nula. Los peligros a los que se someten los protagonistas son más importantes incluso que Dumbledore, por ello obtienen más centralidad que él en algunas centralidades como, por ejemplo, en la del vector propio. Los conocimientos que Dumbledore ofreció a Harry obtienen un valor extríseco alto comparado al valor que obtenía en las centralidades. Así, el valor semántico de las palabras alma y conocimientos es superior al de Dumbledore. Respecto a la afinidad semántica, las tres palabras obtienen su mayor valor con Dumbledore. Esto se da, ya que ninguno de los cuatro individuos son muy relevantes en nuestra red, y es relativamente fácil "convertir" a unos en otros, ya que tienen pocas conexiones.

Capítulo 5

Conclusiones

En este trabajo se han usado diferentes metodologías de teoría de grafos, como las medidas de centralidad y las funciones de afinidad, relacionadas con problemas de flujo máximo, con el objetivo de analizar desde un punto de vista matemático las palabras y la idea filosófica del valor semántico asociado a ellas.

Hemos unido dos disciplinas tan aparentemente opuestas como son la lengua y las matemáticas. Con nuestro método, es posible convertir cualquier texto en una red social. Se ha procedido con un ejemplo sencillo, que permita detallar paso a paso la metodología propuesta. La afinidad semántica es un problema complejo y largo de calcular. Por lo que, en **líneas futuras de investigación,** existiría la posibilidad de crear un algoritmo que calcule la afinidad semántica recurriendo al uso del Algoritmo de Ford-Fulkeron-Edmonds-Karp, para automatizar el procedimiento. Existe ya una solución realizada por el autor del archivo original [9], pero no es mediante una red de flujo, ya que implementa el algoritmo del tubo.

Además, se podría adaptar para usarlo en algo muy relevante a día de hoy como es la inteligencia artificial. Nuestros móviles tienen acceso a todo lo que decimos o escuchamos, ya que sus micrófonos, con previa autorización, recogen las palabras para luego procesarlas y saber qué publicidad puede encajar más con nuestros gustos o preferencias. Así, convirtiendo las conversaciones

que obtienen en redes, y calculando qué palabras son más centrales o tienen más afinidad semántica entre ellas, el ordenador podría saber qué anuncios mostrarnos. Por ejemplo, imaginemos que leemos el texto analizado a un teléfono. Aplicando los resultados, los "anuncios" que nos debería mostrar serían los que contengan la palabra Harry o Voldemort, ya que teniendo en cuenta las centralidades y el valor semántico, son las más relevantes. Aunque en la realidad no leeríamos un texto tan bien organizado. Al hablar, puede que digamos: "estoy interesado en comprarme un ordenador" a un amigo, luego a otra persona: "he encontrado un ordenador que me gusta mucho y es barato". Entonces, el texto a convertir podría ser simplemente el que tiene todas estas frases juntas, tal como las va recogiendo el micrófono. Y, con el valor semántico, gracias a las redes de flujo y a la teoría de grafos, podríamos obtener mediante esa inteligencia artificial buenos y eficientes resultados.

Bibliografía

[1] A. Bavelas, "A mathematical model for group structures", Human Organization 7, pp. 16–30, 1948.

[2] M. A. Beauchamp, "An improved index of centrality", Systems Research and Behavioral Science 10(2), pp. 161–163, 1965.

[3] P. Bonacich, "Factoring and weighting approaches to clique identification", Journal of Mathematical Sociology 2, pp. 113–120, 1972.

[4] T. H. Cormen, "Introduction to Algorithms", MIT Press, 2009.

[5] M.C. Costa, A. Hertz y M. Mittaz, "Bounds and heuristics for the shortest capacitated paths problem", Journal of Heuristics 8(4), pp. 449–465, 2002.

[6] R. C. Cross, "Logos and forms in Plato", Mind 63(252), pp. 433–450, 1954.

[7] J. Edmonds y R. M. Karp, "Theoretical improvements in algorithmic efficiency for network flow problems", Journal of the ACM 19(2), pp. 248–264, 1972.

[8] J. Fumanal-Idocin, A. Alonso-Betanzos, O. Cordon, H. Bustince y M. Minárová, "Community detection and social network analysis based on the italian wars of the 15th century", Future Generation Computer Systems 113, pp. 25–40, 2020.

[9] J. Fumanal-Idocin, O. Cordón Fellow, G. Dimuro, M. Minarová y H. Bustince, "The Concept of Semantic Value in Social Network Analysis: an Application to Comparative Mythology", ArXiv abs/2109.08023, 2021.

[10] A. V. Goldberg y C. Harrelson, “Computing the shortest path: A* search meets graph theory”, in Proc. 16th ACM-SIAM Symposium on Discrete Algorithms 5, pp. 156–165, 2005.

[11] J.L. Gross y J. Yellen, “Graph theory and its applications, Chapman & Hall/CRC, 2006.

[12] A. Landherr, B. Friedl y J. Heidemann, “A critical review of centrality measures in social networks”, Business & Information Systems Engineering 2(6), pp. 371–385, 2010.

[13] C. H. Proctor y C. P. Loomis, “Analysis of sociometric data” in Marie Johada, M. Deutsch y S.W. Cook (eds), Research Methods in Social Relations, Part 2 (New York: Holt, Renehart& Winston), 1951.

[14] N. Rescher, “On the status of “things in themselves” in Kant”, Synthese 47(2), pp. 289–299, 1981.

[15] G. Sabidussi, “The centrality index of a graph”, Psychometrika 31(4), pp. 581–603, 1966.

[16] D. J. Schopenhauer, “Philosophy and the arts”, Cambridge University Press, 2007.

[17] S. Wasserman y K. Faust, “Social network analysis: Methods and applications”, Cambridge University Press, vol. 8, 1994.

[18] S. Wasserman y K. Faust, “Centralidad y prestigio”, Cambridge University Press, pp. 191–240, 2008.