

UNIDAD 5: ANÁLISIS DE GRAFOS DE REDES SOCIALES

SIMILITUD ENTRE NODOS

Blanca Vázquez y Gibran Fuentes-Pineda

Noviembre 2020

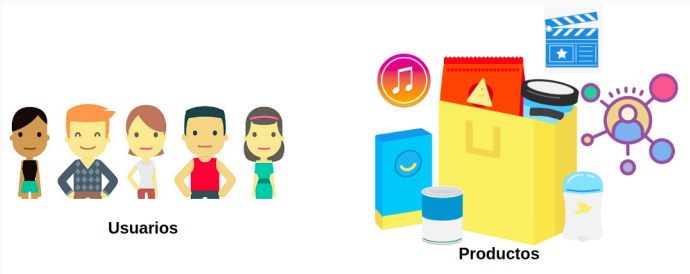
Recordemos que:

- Análisis de grafos sociales
- Detección de comunidades
- ¿Cómo detectar una comunidad a partir de grafos no dirigidos?
- ¿Cómo detectar nodos similares?

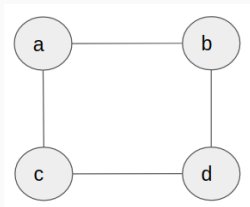
INTRODUCCIÓN

Muchas aplicaciones pueden requerir una medida de similitud entre objetos:

- Buscar documentos similares en la web.
- Agrupar objetivos (filtrado colaborativo)



En el 2002 Glen Jeh and Jennifer Widom ¹ propusieron **SimRank** un algoritmo para medir la similitud entre dos objetos.



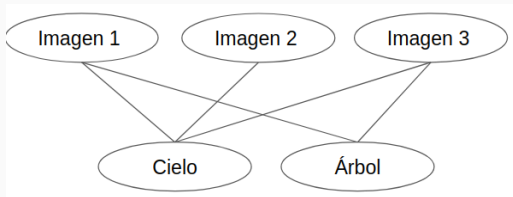
¹G. Jeh and J. Widom. “SimRank: a measure of structural-context similarity”, In KDD’02: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 538–543. ACM Press, 2002.

La idea básica de SimRank es:

Dos nodos son similares si están apuntados (tienen bordes entrantes) desde nodos similares.

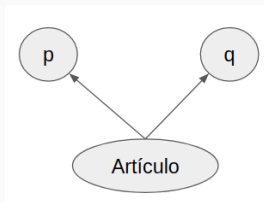
Esto se puede considerar como una agregación de similitudes basadas en la ruta entrante.

Ejemplo básico de SimRank:



- Es un algoritmo general que mide la similitud entre objetos y puede ser aplicado en cualquier dominio.
 - Similitud de páginas web
 - Sistemas de recomendación
 - Esquema de citas

En puntajes de citas, frecuentemente cualquier par de nodos son relacionados a partir únicamente de sus vecinos.



Sin embargo, en SimRank, se puede usar el grafo completo para determinar la similitud en cualquier par de nodos.

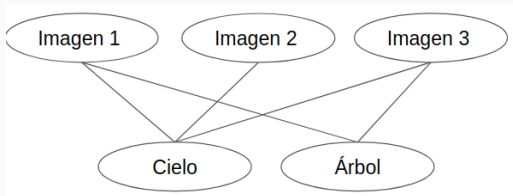
- El propósito de *SimRank* es medir la similitud entre nodos del mismo tipo.

¿Cómo lo hace?

- Hace uso de caminatas aleatorias, partiendo de cualquier nodo el algoritmo observa en qué nodo termina después de n iteraciones.

EJEMPLO DE SIMILITUD DE NODOS

Dado el siguiente grafo, calcular la similitud de la 'Imagen 1' al resto de los nodos.



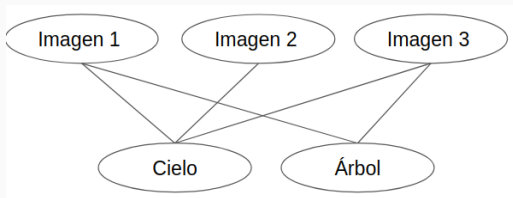
$$v' = \beta Mv + (1 - \beta)e_N$$

dónde:

- β es la probabilidad de elegir un nodo de forma aleatoria
- M es la matriz de adyacencia
- v es un vector columna que refleja la probabilidad de estar en un nodo en cada iteración
- e_N es un vector columna que tiene 1s en la fila del nodo N y 0s en el resto.
- v' es la probabilidad de que estar en cada nodo en la siguiente iteración.

EJEMPLO DE SIMILITUD DE NODOS

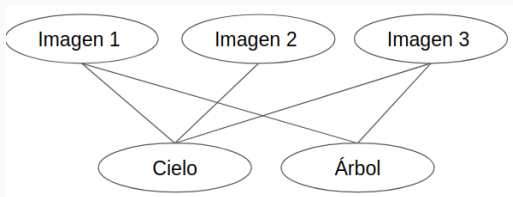
Dado el siguiente grafo, calcular la similitud de la 'Figura 1' al resto de los nodos, donde $B = 0.8$



$$v' = \beta Mv + (1 - \beta)e_N$$

EJEMPLO DE SIMILITUD DE NODOS

Paso 1: definir de la matriz de adyacencia.



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

Sustituir cada valor en la ecuación:

$$v' = \beta Mv + (1 - \beta)e_N$$

$$v' = 0.8 * \begin{bmatrix} 0 & 0 & 0 & 1/3 & 1/2 \\ 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/2 \\ 1/2 & 1 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + (1 - 0.8) \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

EJEMPLO DE SIMILITUD DE NODOS

Paso 2: resolver la parte izquierda de la ecuación:

$$BM_v = 0.8 * \begin{bmatrix} 0 & 0 & 0 & 1/3 & 1/2 \\ 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/2 \\ 1/2 & 1 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$BM_v = \begin{bmatrix} 0 & 0 & 0 & 4/15 & 2/5 \\ 0 & 0 & 0 & 4/15 & 0 \\ 0 & 0 & 0 & 4/15 & 2/5 \\ 2/5 & 4/5 & 2/5 & 0 & 0 \\ 2/5 & 0 & 2/5 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

EJEMPLO DE SIMILITUD DE NODOS

Paso 2: resolver la parte izquierda de la ecuación:

$$BM_v = \begin{bmatrix} 0 & 0 & 0 & 4/15 & 2/5 \\ 0 & 0 & 0 & 4/15 & 0 \\ 0 & 0 & 0 & 4/15 & 2/5 \\ 2/5 & 4/5 & 2/5 & 0 & 0 \\ 2/5 & 0 & 2/5 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$BM_v = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2/5 \\ 2/5 \end{bmatrix}$$

Paso 2: resolver el resto de la ecuación

$$v' = \beta Mv + (1 - \beta)e_N$$

$$v' = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2/5 \\ 2/5 \end{bmatrix} + (1/5) \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Paso 2: resolver el resto de la ecuación

$$v' = \beta Mv + (1 - \beta)e_N$$

$$v' = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2/5 \\ 2/5 \end{bmatrix} + \begin{bmatrix} 1/5 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/5 \\ 0 \\ 0 \\ 2/5 \\ 2/5 \end{bmatrix}$$

EJEMPLO

Iteraciones

$$t_0 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, t_1 \begin{bmatrix} 1/5 \\ 0 \\ 0 \\ 2/5 \\ 2/5 \end{bmatrix}, t_2 \begin{bmatrix} 35/75 \\ 8/75 \\ 20/75 \\ 6/75 \\ 6/75 \end{bmatrix}, t_3 \begin{bmatrix} 95/375 \\ 8/375 \\ 20/375 \\ 142/375 \\ 110/375 \end{bmatrix}, t_4 \begin{bmatrix} 2353/5625 \\ 568/5625 \\ 1228/5625 \\ 786/5625 \\ 690/5625 \end{bmatrix} \dots \begin{bmatrix} .345 \\ .066 \\ .145 \\ .249 \\ .196 \end{bmatrix}$$