

UNIDAD 5: ANÁLISIS DE GRAFOS DE REDES SOCIALES

CONTEO DE TRIÁNGULOS

Blanca Vázquez y Gibran Fuentes-Pineda

Noviembre 2020

- Consideremos un grafo no dirigido y no pesado, tal que $G = (V, E)$ donde G puede ser el grafo de *Facebook* donde los vértices V representan personas y los enlaces E representan relaciones de amistad.
- O dónde G es el grafo de *Twitter* con arcos (u, v) si u sigue a v .

- Coeficiente de agrupamiento
- Conteo de triángulos

DEFINICIÓN FORMAL DE UN GRAFO

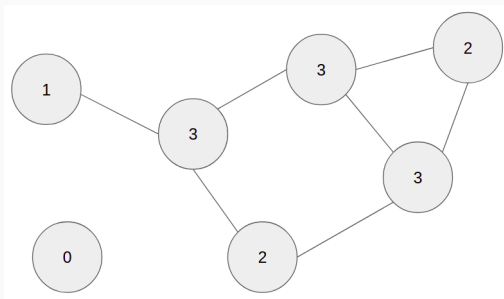
Un grafo $G = (V, E)$ formalmente consiste de un conjunto de vértices V y un conjunto de enlaces E entre ellos.

- Tal que un enlace e_{ij} conecta dos vértices i y j .
- La vecindad de vértices N para un vértice v_i se define como aquellos vértices inmediatamente conectados de tal forma que:

$$N_i = \{v_j\} : e_{ij} \in E \vee v_j \in V$$

GRADO DE UN VÉRTICE

El grado de un vértice K_v se define como el número de enlaces incidentes al vértice.



Grafo con vértices etiquetados según su grado.

El coeficiente de agrupamiento (*clustering coefficient*) de un vértice v en un grafo G cuantifica qué tanto el vértice está agrupado (o interconectado) con sus vecinos. Existen dos tipos:

- Para grafos dirigidos
- Para grafos no dirigidos

Para un grafo dirigido, e_{ij} es distinto de e_{ji} y por lo tanto para cada vecino N_i existen $k_i(k_i - 1)$ enlaces que podrían existir entre los vértices vecinos, por lo tanto:

$$CC_v = \frac{N_v}{K_v(K_v - 1)}$$

donde: v : es un nodo, N_v : es el número de enlaces entre los vecinos de v y K_v es el grado de v .

COEFICIENTE DE AGRUPAMIENTO PARA NO GRAFOS DIRIGIDOS

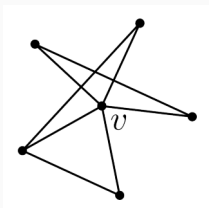
Para un grafo no dirigido, tienen la propiedad de que tanto los enlaces e_{ij} y e_{ji} son considerados idénticos, por lo tanto si un vértice v_i posee k_i vecinos, entonces existirán $\frac{k_i(k_i - 1)}{2}$ enlaces entre los vértices de su vecindad, por lo tanto:

$$CC_v = \frac{2Nv}{Kv(Kv - 1)}$$

donde: v : es un nodo, Nv : es el número de enlaces entre los vecinos de v y Kv es el grado de v .

EJEMPLO DE COEFICIENTE DE AGRUPAMIENTO

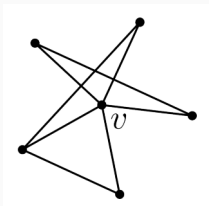
Dado el siguiente grafo no dirigido, calcular el coeficiente de agrupamiento del vértice v .



donde: N_v : es el número de enlaces entre los vecinos de v y K_v es el grado de v .

EJEMPLO DE COEFICIENTE DE AGRUPAMIENTO

Dado el siguiente grafo no dirigido, calcular el coeficiente de agrupamiento del vértice v .

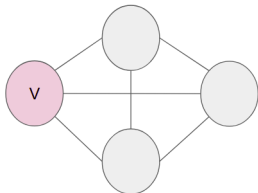


$$CC_v = \frac{2N_v}{K_v(K_v - 1)} = \frac{2 * 3}{5(5 - 1)} = \frac{6}{5 * 4} = \frac{3}{10}$$

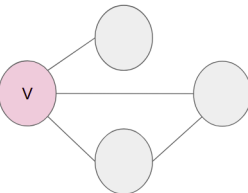
donde: $N_v=3$, $K_v = 5$

EJERCICIO

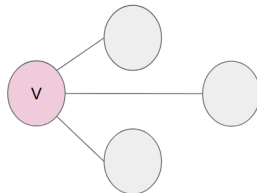
Calcular el CC del vértice v para los siguientes grafos.



Grafo a



Grafo b



Grafo c

¿Por qué en grafos sociales es importante el CC?

- Coeficientes de agrupamiento altos indican **comunidades muy unidas**.
 - Características: altos grados de confianza,
- Coeficientes de agrupamiento bajos indican **agujeros estructurales**
 - Características: un vértice conectado a diferentes comunidades, las cuales no están conectadas entre sí (peligro).

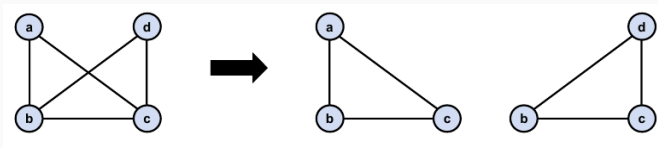
¿Qué relación existe en el coeficiente de agrupamiento y el conteo de triángulos?

- El algoritmo de coeficiente de agrupamiento se basa en el conteo de triángulos para la **detección de comunidades** y para medir la cohesión entre ellas.
- Dependiendo del tipo de aplicación, se puede utilizar para determinar la estabilidad de un grafo.

- Identificación de características de un sitio web.
 - Elaboración de listas de contenido deseado y no deseado.
- Investigar la estructura de comunidades
 - Identificación de vecindarios en Facebook.
- Explorar la estructura de un sitio web
 - Detectar comunidades que se crean a partir de un conjunto de páginas

ALGORITMOS PARA EL CONTEO DE TRIÁNGULOS

Un triángulo se puede definir como un conjunto de tres vértices mutuamente adyacentes en un grafo.



El grafo de la izquierda consiste de 2 triángulos: $\{a,b,c\}$ y $\{b,c,d\}$.
Imagen tomada de GraphChallenge.org, 2020.

Los algoritmos para el conteo de triángulos se puede dividir en 2 categorías:

- Algoritmos basados en álgebra lineal que utilizan la matriz de adyacencia de un grafo.
- Algoritmos basados en grafos que utilizan la lista de enlaces y las vecindades del grafo.

- Tiempo de ejecución
- Espacio en memoria
- Tasa de enlaces procesados por segundo
- Cómputo paralelo

The MIT/Amazon/IEEE Graph Challenge¹

- Estimular la investigación y destacar las innovaciones en software, hardware, algoritmos y sistemas de análisis de gráficos.
 - Conteo de triángulos

¹<https://graphchallenge.mit.edu/champions>

The MIT/Amazon/IEEE Graph Challenge

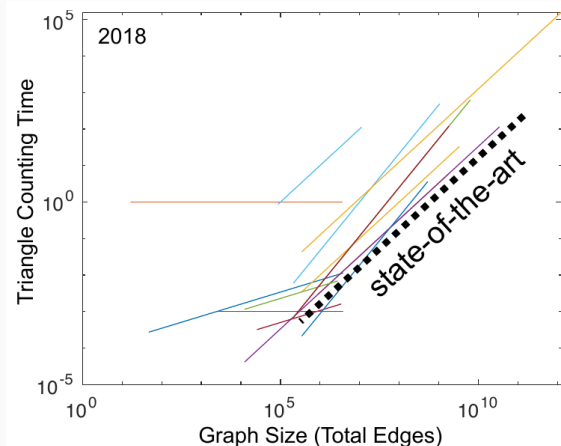


Imagem tomada de GraphChallenge.org, 2020.

The MIT/Amazon/IEEE Graph Challenge

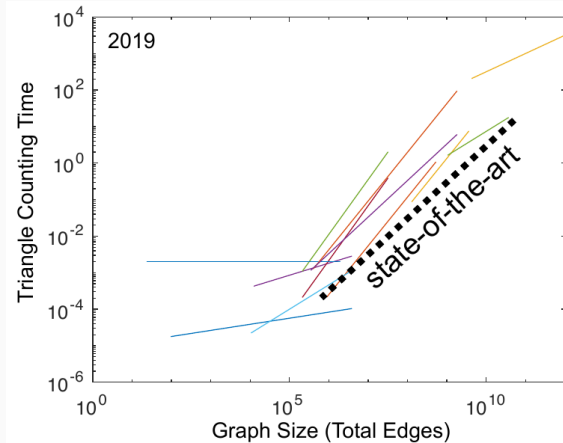


Imagen tomada de GraphChallenge.org, 2020.

The MIT/Amazon/IEEE Graph Challenge

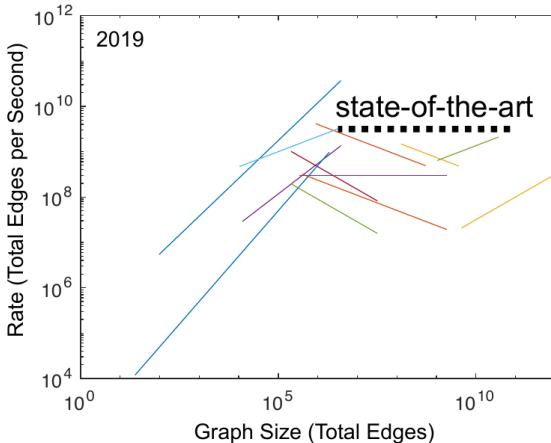


Imagen tomada de GraphChallenge.org, 2020.

- En el 2017, el tiempo de ejecución del conteo de triángulos Tri de acuerdo al número de enlaces N_e era:

$$Tri \approx (N_e/10^8)^{\frac{4}{3}}$$

- Mientras que el 2018 mejoró a:

$$Tri \approx (N_e/10^9)$$

CONTEO DE TRIÁNGULOS USANDO GPUS

En el 2018, el equipo de la Universidad George Washington y la Universidad de Massachusetts Lowell, definieron una arquitectura basada en GPUs.

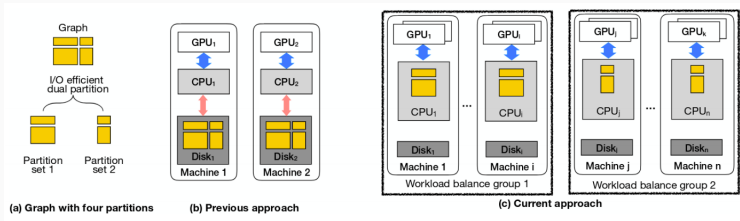


Imagen tomada de Graph Challenge: Counting Triangles and Enumerating Trusses in Graphs, 2019

- Tim Roughgarden, *CS167: Reading in Algorithms Counting Triangles*, 2014.
- Samsi, Siddharth and Kepner, Jeremy and Gadepally, Vijay and Hurley, Michael and Jones, Michael and Kao, Edward and Mohindra, Sanjeev and Reuther, Albert and Smith, Steven and Song, William and et al.; *GraphChallenge.org Triangle Counting Performance*, 2020 IEEE High Performance Extreme Computing Conference (HPEC).