Descubrimiento de Comunidades

Alondra Berzunza

Conceptos

Adyacencia

Sea $G = (V, E)$ un grafo. Un nodo $u \in V$ es **adyacente** o **vecino** a un nodo $v \in V$ si y sólo si hay un arista entre ellos; es decir, si y sólo si $(u, v) \in E$. El conjunto de vecinos de i es

$$N(u) = \{v \in V | (u, v) \in E\}$$

Si $e = \{u, v\}$, se dice que la arista e es **incidente** en los nodos u y v. También se dice que la arista e **conecta** u y v.

Peso

Existen casos en los cuales las redes no sólo representan conexiones simples de tipo booleano. Algunas veces es necesario modelar un peso, fuerza o valor asociado a la conexión.

Por ejemplo en un modelo de una red de comunicaciones, los nodos representan terminales y enrutadores, mientras que las aristas representan las conexiones y un valor asociado a cada arista puede representar el ancho de banda del que se dispone en esa conexión.

El peso de una arista representa un valor intrínseco a ella. Cuando en un grafo se modelan los pesos, este se denomina grafo ponderado.

Grado

El grado o conectividad de un nodo de un grafo es el número de aristas incidentes en él, cada arista contribuye con dos unidades al grado de los nodos que conecta y por lo general se denota ki o deg(i). Esta medida es de gran importancia en varios métodos de detección de comunidades, ya que permite comprender de una manera la relación entre los nodos y el número de vecinos adyacentes al nodo.

De manera formal el grado de un nodo se puede expresar como:

$$\deg(i) = |N(i)|$$

El conjunto es algunas veces llamado la frontera o límite de . El límite de junto con es llamado el cierre de v.

$$\mathrm{cierre}(v) = N(v) \cup \{v\}$$

Grado

A los nodos de grado cero, deg(i)=0, se les llama aislados. Si deg(i)=1, se les denomina nodos colgantes u hojas.

Existe una relación directa entre el número de aristas y el grado de cada uno de los nodos en un grafo . Esta propiedad se conoce como el teorema de los apretones de manos:

$$2e = \sum_{v=V} \deg(v)$$

Formalmente, el grado en un nodo en un grafo ponderado se define como:

$$\deg(i) = \sum_{j=1}^{n} w_{ij}$$

Grado

Donde w_{ij} es el peso de la arista del par de nodos (i, j).

La **matriz de grados** D se define como la matriz con los grados $\deg(1), \ldots, \deg(n)$ en la diagonal,

$$D = \begin{bmatrix} \deg(1) & & \\ & \ddots & \\ & & \deg(n) \end{bmatrix}$$

$$D(i,j) = \begin{cases} 0 \text{ si } i \neq j \\ \deg(v_i) \text{ si } i = j \end{cases} \Rightarrow D = \operatorname{diag}(\deg(1), \ldots, \deg(n))$$

Matriz de Adyacencia

La matriz de adyacencia permite representar, para un grafo no dirigido *G=(V,E)* con un conjunto de nodos *V={v1, ..., vn}*, las parejas de nodos que se encuentran interconectados. La matriz A es de tamaño nxn, donde n es el número total de nodos.

La matriz de adyacencia se define como

$$a_{ij} = \begin{cases} 1, & \text{si existe una arista entre el nodo } i \text{ y el nodo } j \\ 0, & \text{en caso contrario} \end{cases}$$

De la misma manera, un grafo ponderado se puede representar mediante una matriz de adyacencia

$$a_{ij} = \begin{cases} w, & \text{si existe una arista con peso } w \text{ entre el nodo } i \text{ y el nodo } j \\ 0, & \text{en caso contrario} \end{cases}$$

Matriz de incidencia

La **matriz de incidencia** permite representar un grafo no dirigido $G = (V, E)$ con un conjunto de nodos $V = \{v_1, \ldots, v_n\}$ y un conjunto de aristas $E = \{e_1, \ldots, e_m\}$. La matriz de incidencia es de tamaño $m \times n$ y se define como $M = [m_{ij}]$

$$m_{ij} = \begin{cases} 1, & \text{si la arista } e_j \text{ es incidente con } v_i \\ 0, & \text{en caso contrario} \end{cases}$$

Caminos

Si un par de nodos (i, j) no adyacentes pueden ser conectados por medio de una secuencia de m aristas $(i, l_1), (l_1, l_2), \ldots, (l_{m-1}, j)$; este conjunto de aristas describe un **camino** entre el nodo i y el nodo j, donde m es la **longitud del camino**

Se puede decir que dos nodos se encuentran **conectados** si existe al menos un camino entre ellos.

Si todos los nodos y aristas que componen un camino entre dos nodos i y j no se repiten, el camino se conoce comúnmente como **camino simple** o ***path***

Un **bucle** se define como un camino donde el nodo inicial y final es el mismo nodo y se pasa una única vez por cada uno de los $m - 1$ nodos intermedios

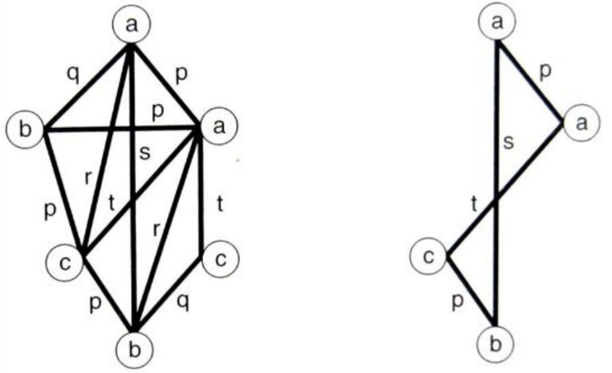
Subgrafos

Sea $G = (V, E)$ un grafo no dirigido. $G_X = (X, E_X)$ es un **subgrafo** de G (inducido por X) si y sólo si $X \subseteq V$ y $E_X \leq (X \times X) \cap E$; es decir, si contiene un subconjunto de nodos de G y las aristas correspondientes

Recíprocamente, el grafo G es un **supergrafo** de G_X si G_X es un **subgrafo** de G. Algunas veces, la condición de subgrafo hace referencia sólo a la condición de $E_X \subseteq E$ y un **subgrafo inducido** será aquel que cumpla la definición de subgrafo

Un grafo $G = (V, E)$ es un **grafo completo** si incluye todas las aristas posibles. Formalmente el grafo completo K_n de nodos n se define como:

$$K_n = V \times V - \{(v, v) | v \in V\}$$



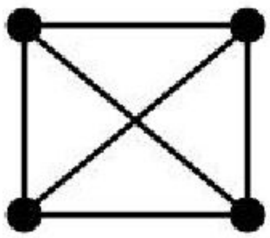
Clique

Un ***clique*** es un conjunto de nodos que se encuentran todos interconectados. Por lo tanto, cumplen con la definición de subgrafo completo. Un *clique* es llamado **máximo** si y sólo si no puede ampliarse mediante la inclusión de un vértice adyacente

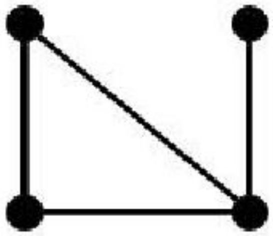
Un ***k-clique*** es un grupo de nodos conectados, en el que k es la máxima longitud del camino entre un par de nodos del *k-clique*. De manera más formal, se define como un subconjunto de nodos C tal que para todo $i, j \in C$, la distancia $d(i, j) \leq k$

Clique

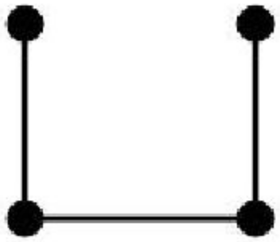
Por lo tanto, un *clique* es también un *1-clique,* ya que la distancia máxima entre cualquier par de nodos es directa. Un 2-*clique* es un grafo donde el camino que conecta cualquier par de nodos tiene una longitud $k \leq 2$. Este tipo de agrupamientos se puede asociar a los "amigos de amigos" en las redes sociales.



1-*clique*



2-*clique*



3-*clique*

k-plex

Un ***k-plex*** de *n* nodos es un subconjunto máximo de *n* nodos dentro de una red, de tal manera que cada nodo está conectado a, por lo menos, $n - k$ nodos de los demás nodos del *k-plex*. Si $k = 1$, conforma un *clique;* si $k = 2$, entonces cada nodo debe estar conectado a todos o a todos menos un nodo y así sucesivamente. Además, los *k-plexes* pueden solaparse si comparten nodos

k-core

Un ***k-core*** es un subconjunto máximo de n nodos de tal manera que cada nodo está conectado a por lo menos otros k nodos en el subconjunto. A diferencia de los *k-plexes*, los *k-cores* no pueden formar grupos solapados . Adicionalmente, un *k-core* tiene la propiedad de contener *k-cliques* y *k-plexes*, donde un *k-plex* de tamaño s forman parte de un $(s - k)$-*core*

Comunidades en Redes Sociales

Las comunidades de redes sociales son grupos de personas que se conectan e interactúan en plataformas digitales por tener intereses, objetivos o propósitos comunes. Fomentan la pertenencia, el apoyo mutuo y el intercambio de información a través de debates, contenido y redes. Ejemplos incluyen grupos en Facebook, comunidades en Reddit y servidores de Discord.

Comunidades en Redes Sociales

Son grupos de usuarios de internet que comparten intereses, como una marca, un hobby o una identidad, y se comunican a través de canales digitales.

Sirven para establecer vínculos y relaciones, ya sean personales, profesionales o de afición.

Permiten a los usuarios sentirse parte de algo y encontrar apoyo, información y experiencias compartidas.

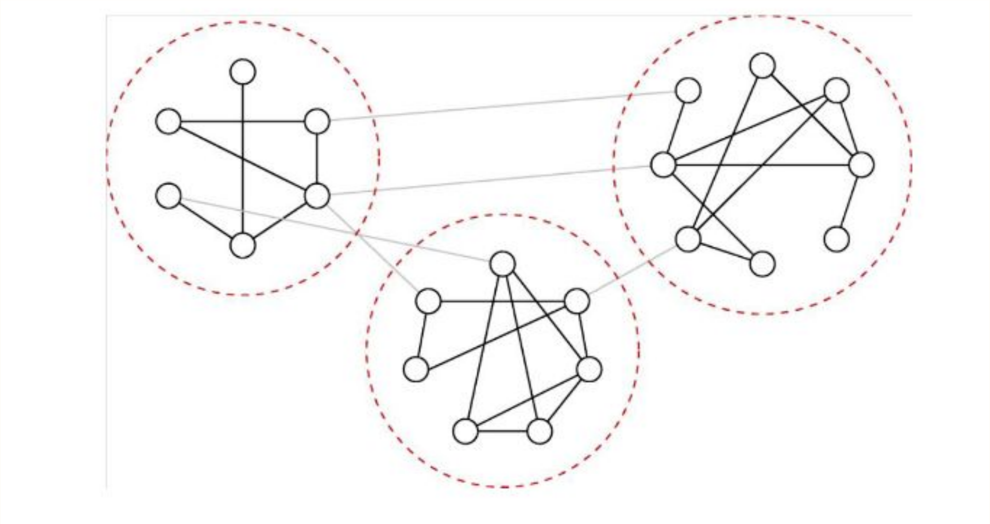
Importancia

- **Intercambian información y contenido:** Los miembros comparten y discuten sobre temas de su interés.
- **Generan sentido de pertenencia:** Crean un ambiente donde las personas se sienten conectadas y apoyadas por otros con experiencias similares.
- **Participan activamente:** A diferencia de ser un mero seguidor pasivo, los miembros de una comunidad se involucran en debates y actividades.
- **Sirven para marcas y empresas:** Permiten a las marcas crear un vínculo más fuerte con sus clientes potenciales y reales al ofrecerles contenido interesante y fomentar la lealtad.

Comunidades en Redes Sociales

Una definición es la propuesta por Newman y Girvan es:

“una comunidad es la división de los nodos de una red en grupos dentro de los cuales las conexiones son densas, pero entre ellos son escasas”



Comunidades en Redes Sociales

Trabajos realizados por Reichardt y Bornholdt a partir de las definiciones de comunidad plantadas por Newman , definen de manera formal una comunidad como un conjunto de nodos para el que se verifican las siguientes desigualdades:

$$\frac{2m}{n(n-1)} > \frac{2M}{N(N-1)} > \frac{m_{nN}}{n(N-n)}$$

donde m es el número de aristas entre los n nodos de la comunidad y m_{nN} es el número de aristas que conectan a los restantes $N-n$ nodos de la red. En la cual, de manera general, una comunidad debe presentar una densidad mayor con respecto a la densidad promedio del grafo, y esta a su vez debe ser superior a la densidad externa a la comunidad

Comunidades explícitas

Las comunidades explícitas son aquellas que se construyen a partir de la intervención humana, es el humano quien decide los enlaces entre los nodos de la red. Ejemplos de este tipo de comunidades son Facebook, Google+, LinkedIn o Flick

Comunidades implícitas

Las comunidades implícitas son aquellas donde se supone la existencia de una comunidad en un conjunto de datos y el sistema es el encargado de “descubrir” la existencia de esas comunidades. La investigación en el desarrollo de técnicas que permitan descubrir de manera automática las comunidades ofrece ventajas al no requerir la intervención humana en su elaboración.

A partir de la idea descrita anteriormente, donde una comunidad se caracteriza por una mayor cantidad de conexiones entre sus nodos con respecto a los de fuera, varias definiciones se han ideado desde el campo de las ciencias humanas y, de manera más reciente, desde la física y las ciencias de la computación, distinguiéndose tres tipos de definición, local, global y basados en la similitud de nodos.

Medidas en redes y comunidades

Una de las áreas importantes en el estudio de las redes y comunidades es determinar que características se deben resaltar de los grupos a encontrar. Estas características pueden ser de centralidad, transitividad, entre otras. Muchas de estas medidas se originan en los primeros estudios sobre redes sociales, elaborados por sociólogos y demás profesiones afines a las ciencias humanas, que hoy en día se han complementado con otros enfoques como la física, las ciencias de la computación o la biología.

Medidas en redes y comunidades

1. **DISTANCIA.**

En un grafo $G = (V, E)$ la **distancia** entre dos nodos i y j se define como el camino más corto existente entre ellos. Si $i = j$, entonces $d(i,j) = 0$. Si no existe un camino posible entre i y j, entonces $d(i,j) = \infty$ [9].

Si el grafo es ponderado la distancia se calcula como la suma de los pesos asociados a las aristas del camino [9]:

$$d(i,j) = \sum w_{i,j}$$

Medidas en redes y comunidades

1. **DISTANCIA.**

Un **camino geodésico** es el camino más corto que puede existir entre un par de nodos que se encuentren conectados por medio de un camino [9].

La **distancia media o camino medio** mide la distancia promedio que existe entre un par de nodos i y j que se comunican por el camino geodésico, de manera formal se define como [9]:

$$d_m(G) = \frac{1}{n(n-1)} \sum_{i,j} d(i,j)$$

Medidas en redes y comunidades

1. **DISTANCIA.**

El **diámetro** en una red es la máxima distancia que puede existir entre dos nodos i y j de entre todos los nodos del grafo que estén unidos por un camino geodésico. Esta medida es de gran importancia para establecer el tamaño de la red [9].

$$\text{Diámetro}(G) = \max\{d(i,j)\}\ \forall\ i,j \in V$$

Para calcular su valor, se deben calcular todos los caminos geodésicos para cada par de nodos existentes en la red y, de ellos, escoger el de mayor longitud.

Medidas en redes y comunidades

2. DENSIDAD.

La **densidad** $\delta(G)$ es la relación entre el número de aristas presentes en el grafo (m) y el número máximo de aristas que puede tener el grafo, cuyo valor se calcula a partir del número de nodos presentes (n) [55].

De manera formal, se expresa como:

$$\delta(G) = \frac{m}{n(n-1)}$$

En un grafo completo, la densidad presente es máxima, $\delta(G) = 1$.

Medidas en redes y comunidades

2. DENSIDAD.

La **densidad intra-cluster** $\delta_{int}(C)$ es la relación entre en número de aristas internas y el número aristas que podrían existir dentro de un cluster. [19].

$$\delta_{int}(C) = \frac{\#\text{aristas internas de } C}{\frac{n_c(n_c - 1)}{2}}$$

Recíprocamente, se define la **densidad inter-cluster** $\delta_{ext}(C)$ como la relación entre el número de aristas que conectan el grupo con el resto del grafo y el número de todas las posibles aristas. [19].

$$\delta_{ext}(C) = \frac{\#\text{ aristas externas de } C}{n_c(n - n_c)}$$

Medidas en redes y comunidades

3. CENTRALIDAD

Las medidas de centralidad ayudan a evaluar los nodos en busca de aquéllos que sean más importantes o céntricos en una agrupación [51]. El concepto de la centralidad se Inicia en los estudios de comportamiento humano y sus relaciones; basándose en el uso de conceptos matemáticos, estadísticos y de teoría de grafos. Hay diferentes tipos de medidas para la centralidad, a continuación se explicarán algunos de ellos: centralidad por grado (*degree centrality*), centralidad por vector propio *eigenvector centrality*), centralidad por cercanía (*closeness centrality*) y centralidad por intermediación (*betweenness centrality*).

Medidas en redes y comunidades

3. CENTRALIDAD

3.1. Centralidad por grado *(Degree centrality)*

Como se vio anteriormente, el **grado** de un nodo en un grafo es el número de aristas conectadas a él. Se denota como k_i donde i representa el nodo. De manera formal, se expresa de la siguiente forma a partir de la matriz de adyacencia del grafo:

$$k_i = \sum_{j=1}^{n} a_{ij}$$

El **grado medio** de un grafo se determina como:

$$c = \frac{1}{n}\sum_{i=1}^{n} k_i$$

Medidas en redes y comunidades

3. CENTRALIDAD

3.1. Centralidad por grado *(Degree centrality)*

Sea un subgrafo C de un grafo G, donde $|C| = n_c$ y $|G| = n$ nodos, respectivamente. El **grado interno** de un nodo $v \in C$ es k_v^{int}, donde el numero de aristas conectados en v con otros nodos de C. Recíprocamente, se aplica este concepto al **grado externo** k_v^{ext}como el número de aristas que conectan nodos de C con nodos del resto del grafo ($v \notin C$) [19].

Si $k_v^{ext} = 0$, los nodos son vecinos únicamente de otros nodos de C, indicando que el nodo v pertenece al *cluster* C. Si $k_v^{int} = 0$, el nodo no pertenecerá a C, si bien puede pertenecer a otro *cluster* [19].

Medidas en redes y comunidades

3. CENTRALIDAD

3.1. Centralidad por grado *(Degree centrality)*

El **grado interno del subgrupo** C, k_C^{int} es la suma de los grados internos de los nodos pertenecientes al clúster, de la misma manera el **grado externo del subgrupo** C, k_C^{ext} es la suma de los grados externos de los vértices [19].

El **grado total del subgrupo** C es la suma de los grados de los nodos de C [19].

$$k_C = k_C^{int} + k_C^{ext}$$

Medidas en redes y comunidades

3. CENTRALIDAD

3.1. Centralidad por vector propio *(Eigenvector Centrality)*

En la centralidad por grado se considera la importancia de un nodo con respecto a su entorno por el número de conexiones. En este caso, se considera la importancia de que el nodo tenga conexiones con otros nodos de igual importancia, siendo esta la base para la centralidad por vector propio (eigenvector centrality).
Esta medida inspiró el desarrollo del algoritmo de PageRank el cual sirve de base para el motor de búsqueda de Google.

Medidas en redes y comunidades

3. CENTRALIDAD

3.1. Centralidad por vector propio *(Eigenvector Centrality)*

Formalmente la **centralidad por vector propio** se define de la siguiente manera [51]:

$$C_e(i) = \frac{1}{K_1} \sum_j A_{ij} x_j$$

Donde A_{ij} es la matriz de adyacencia, K_1 es una constante [51].

Con esta medida, un nodo adquiere "importancia" por medio de dos caminos: tener muchas conexiones con otros miembros de la comunidad o tener conexiones con otros nodos "importantes", aunque sean pocos.
Una de las limitantes de esta medida es el análisis sobre redes dirigidas.

Medidas en redes y comunidades

3. CENTRALIDAD

3.1. Centralidad por cercania *(Closeness Centrality)*

La centralidad por cercanía consiste en calcular las distancias que existen entre el nodo analizado y los demás nodos. Cuanto más grande es la distancia, esta representa menor importancia y, recíprocamente, las distancias menores indican mayores distancias [51].

De manera formal, se define como [51]:

$$C_c(i) = \frac{n}{\sum_{j=1}^{n} d_{ij}}$$

Donde $C_c(i)$ es la centralidad por cercanía del nodo i, n es el número de nodos y d_{ij} la distancia que hay desde el nodo i hasta el nodo j. [51]

Medidas en redes y comunidades

3. CENTRALIDAD

3.1. Centralidad por cercania *(Closeness Centrality)*

Uno de los inconvenientes que presenta este tipo de centralidad es que solo puede ser utilizado en redes donde exista un camino entre cada par de nodos existentes en la red. Otro inconveniente consiste en su variable complejidad computacional, que puede variar de $O(mn)$ hasta $O(n^4)$ [51].

Medidas en redes y comunidades

3. CENTRALIDAD

3.1. Centralidad por intermediación *(Betweenness Centrality)*

La medida de intermediación o de ***betweenness*** se define como una medida para identificar las aristas que conectan comunidades, otorgando valores altos para aquellas aristas que conectan nodos fuera de la comunidad y penalizando aquellas aristas que se encuentran dentro de la comunidad.

Para calcular la medida de ***betweenness***, Girvan y Newman exponen 3 métodos.

Medidas en redes y comunidades

3. CENTRALIDAD

   3.1. Centralidad por intermediación *(Betweenness Centrality)*

-Camino más corto (Shortest - path)

Se basa en calcular el camino más corto entre un par de nodos, para todo el conjunto de nodos existentes en la red, y contar cuántos “caminos” pasan por cada arista, tomando como criterio de elección las aristas con mayor importancia.

Medidas en redes y comunidades

3. CENTRALIDAD

3.1. Centralidad por intermediación *(Betweenness Centrality)*

-Camino más corto (Shortest - path)

De manera formal, la medida de *betweenness* se define para cada arista $e \in E$ como[52]:

$$C_B(e) = \sum_{u,v \in V} \frac{g_e(u,v)}{g(u,v)},$$

Donde $g(u,v)$ es el número total de caminos mínimos entre los nodos u y v, y $g_e(u,v)$ es el número de caminos mínimos entre los nodos u y v que pasan por el nodo e. Valores pequeños de *betweenness* indican que las aristas pertenecen a una misma comunidad y aquellas aristas que conectan nodos de diferentes comunidades tendrán valores más altos [52].

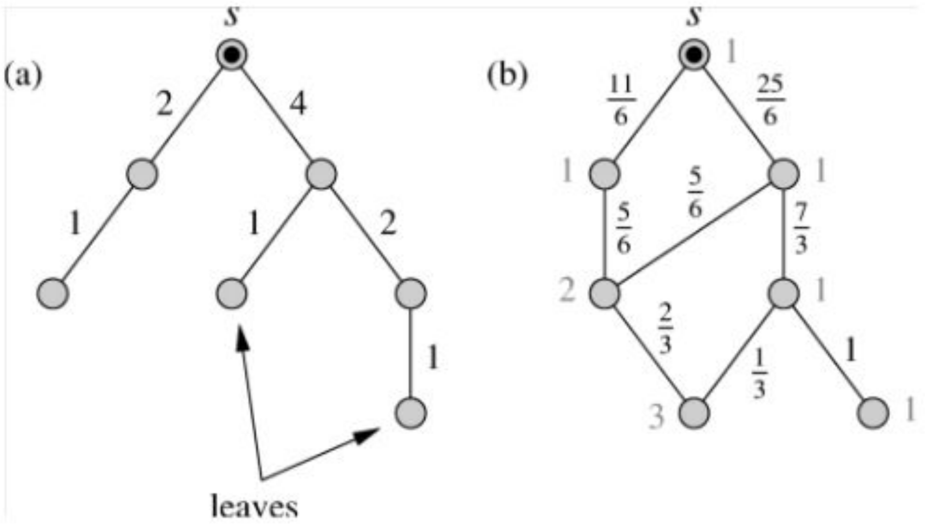
3. CENTRALIDAD

3.1. Centralidad por intermediación *(Betweenness Centrality)*

-Camino más corto (Shortest - path)

En la **Figura 2-6** se puede apreciar el cálculo del *betweenness* para un árbol (a), donde sólo hay un posible camino desde cualquier nodo al nodo S. En este tipo de red es muy simple el procedimiento. Para el caso (b), donde existe más de un camino desde cualquier nodo al nodo S, el cálculo es más complejo.

Figura 2-6: Cálculo de *betweenness* por el camino más corto [52].



3. CENTRALIDAD

3.1. Centralidad por intermediación *(Betweenness Centrality)*

-Flujo de corriente (Current - flow)

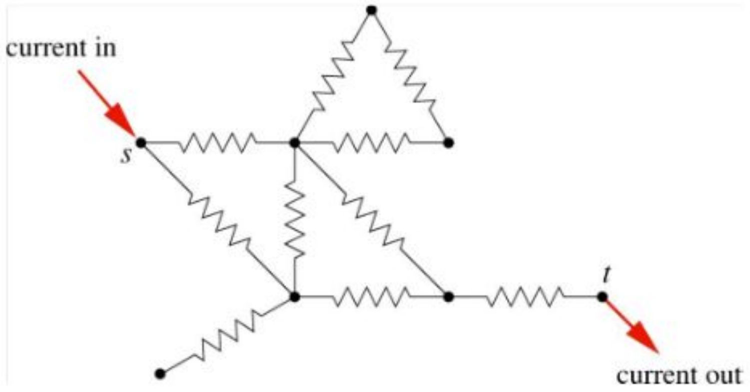
Tanto por el procedimiento por ruta más corta como por el de caminos aleatorios, siempre se está determinando un flujo o una importancia de la arista. Sobre esta premisa se inspira el método del flujo de corriente, en el cual la red se modela como un circuito eléctrico, donde las aristas se consideran resistencias eléctricas, (ver Figura 2-7) y donde se determina la corriente que fluye desde un nodo *s* hasta un nodo *i.*

3. CENTRALIDAD

3.1. Centralidad por intermediación *(Betweenness Centrality)*

-Flujo de corriente (Current - flow)

Figura 2-7: ejemplo de una red con resistencias[52].



Si se aplica una diferencia de voltaje entre los dos nodos (s, t), cada arista lleva una cierta cantidad de corriente, la cual se puede calcular mediante el uso de las leyes de Kirchhoff, repitiéndose este procedimiento en todos los nodos; como resultado se obtiene es el promedio de la corriente que se transporta entre el par de nodos. Este resultado es homologo a los otros métodos realizados, (ruta mínima, camino aleatorio)

3. CENTRALIDAD

3.1. Centralidad por intermediación *(Betweenness Centrality)*

-Camino aleatorio (Random - walk)

Otro enfoque para calcular la medida de ***betweenness*** se basa en considerar caminos aleatorios. En este caso, la medida de ***betweenness*** de una arista está dada de forma aproximada por la frecuencia con la que se pueden generar recorridos entre un par de nodos *(s,t)*.
Se escogen al azar un par de nodos *(s,t)*, el “caminante” empieza en *s* y se mantiene en movimiento hasta que llega a *t*, donde se detiene. Se calcula la probabilidad de cada arista por la cual se cruza y se realiza el promedio de todos los recorridos posibles entre *s* y *t*.

Medidas en redes y comunidades

4. COEFICIENTE DE AGRUPAMIENTO O *Clustering*

Otro tipo de métricas empleadas en la detección de comunidades evalúan el nivel de agrupamiento de los nodos, si los nodos están estrechamente relacionados. El coeficiente de agrupamiento está relacionado con el número de triángulos.

Los triángulos en un grafo son los grupos conformados por tres nodos interconectados de manera total entre ellos, que conforman una triada.

El coeficiente de agrupamiento o clustering permite determinar el nivel de presencia de triángulos en una red. Este coeficiente se puede definir de dos formas, de manera global o de manera local.

Medidas en redes y comunidades

4. COEFICIENTE DE AGRUPAMIENTO O *Clustering*

El coeficiente de agrupamiento global también es conocido como **coeficiente de transitividad.** De manera formal, se define como [48]:

$$C_g = \frac{(\#\text{ de triángulos en la red}) \times 3}{\#\text{ de caminos de longitud 2 en la red}}$$

Medidas en redes y comunidades

4. COEFICIENTE DE AGRUPAMIENTO O *Clustering*

Para encontrar el número de triángulos en la red (N_Δ) y el número de caminos de longitud 2 (N_3) se pueden utilizar las siguientes expresiones [9]:

$$N_\Delta = \sum_{k<j<i} a_{ij} a_{ik} a_{jk}$$

$$N_3 = \sum_{k<j<i} a_{ij} a_{ik} + a_{ji} a_{jk} + a_{ki} a_{kj}$$

Donde a_{ij} son los elementos de la matriz de adyacencia A.

4. COEFICIENTE DE AGRUPAMIENTO O *Clustering*

De manera similar, el **coeficiente de agrupamiento local** se calcula sobre un nodo en particular. De manera formal, se define como [48]:

$$C_i = \frac{\#\ \text{de triángulos conectados a}\ i}{\#\ \text{de caminos de longitud 2 conectando a}\ i}$$

También se puede encontrar este coeficiente en formulaciones mas amigables como [64]:

$$C_i = \frac{2m_i}{k_i(k_i - 1)}$$

Donde m_i es el número de conexiones entre k_i vecinos del nodo i.

Medidas en redes y comunidades

4. COEFICIENTE DE AGRUPAMIENTO O *Clustering*

También hay otra formulación para calcular el coeficiente de agrupamiento global a partir de la medida local [48]:

$$C_g = \frac{1}{n} \sum_{i \in V} C_i$$

Medidas en redes y comunidades

4. MODULARIDAD

El índice de modularidad fue propuesto por Girvan y Newman, para medir la presencia de clusters en una red no aleatoria. Se fundamenta en determinar las densidades dentro y fuera de cada cluster y comparándolas con la densidad general que tendrían si la red fuera construida de manera aleatoria.

Medidas en redes y comunidades

4. MODULARIDAD

De manera formal se define como:

$$Q = \frac{1}{2m}\sum_{ij}\left(a_{ij} - \frac{k_i k_j}{2m}\right)\delta(c_i, c_j)$$

Este Q es llamado modularidad, m es el número de aristas, k_i es el grado del nodo i, a_{ij} es el elemento de la matriz de adyacencia de los nodos i y j, $\delta(m, n)$ es la función delta de Kronecker [2] y c_i es el *cluster* i. [51]

Medidas en redes y comunidades

4. MODULARIDAD

Otra forma en la cual se presenta el índice de modularidad es la siguiente:

$$Q = \sum_r (e_{rr} - a_r^2)$$

Donde e_{rr} son el conjunto de aristas que se conectan entre sí en un *cluster* denominado r y a_r representa al conjunto de nodos que forman parte del *cluster* r [51].