Social and Information Networks

Module 2 - Centrality Measures

Reference Book:

Wasserman Stanley, and Katherine Faust. (2009). Social Network Analysis: Methods and Applications, Structural Analysis in the Social Sciences.

Clustering Coefficient

- A clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together
- Tells how well connected the neighbourhood of the node is
- In social networks, nodes tend to create tightly knit groups characterised by a relatively high density of ties
- Two versions of this measure exist:
  - Global and Local
  - Global version gives an overall indication of the clustering in the network
  - Local version gives an indication of the embeddedness of single nodes

Clustering Coefficient (Local)

- Local clustering coefficient of a node in a graph quantifies how close its neighbours are to being a clique (complete sub graph)
- Given a graph G = (V , E)
  - The neighbourhood N_i for a vertex v_i is defined as its immediately connected neighbours as follows:

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

- Let k_i be the number of vertices, $|N_i|$ in the neighbourhood of a vertex i
- Local clustering coefficient C_i for a vertex v_i is given by the number of links between the vertices within its neighbourhood divided by the number of links that could possibly exist between them.

Clustering Coefficient (Local)

- For a directed graph, e_{ij} is distinct from e_{ji} and therefore for each neighbourhood N_i there are k_i (k_i -1) links that could exist among the vertices within the neighbourhood.
- Local clustering coefficient for directed graphs is

$$C_i = \frac{|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}$$

Clustering Coefficient (Local)

- In an undirected graph, e_{ij} and e_{ji} are identical
- If a vertex v_i has k_i neighbours, k_i (k_i -1) / 2 edges could exist among the vertices within the neighbourhood
- Local clustering coefficient for undirected graphs is

$$C_i = \frac{2|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}$$

- Clustering coefficient of a node is the ratio of number of connections in the neighbourhood of a node and the number of connections if the neighbourhood was fully connected
- **C_i=2 *(number of pairs of neighbours of i that are connected) / (number of pairs of neighbours of i)**

Clustering Coefficient (Local)

- Let $\lambda_G(v)$ be the number of triangles on $v \in V (G)$ for undirected graph G
  - $\lambda_G(v)$ is the number of subgraphs of G with 3 edges and 3 vertices, one of which is v
- Let $\tau_G(v)$ be the number of triples on $v \in G$
  - $\tau_G(v)$ is the number of subgraphs with 2 edges and 3 vertices, one of which is v and such that v is incident to both edges
- Alternatively, the clustering coefficient is $C_i = \frac{\lambda_G(v)}{\tau_G(v)}$
- C_i= number of triangles connected to node i / number of triples centered around node i , where a triple centered around node i is a set of two edges connected to node i

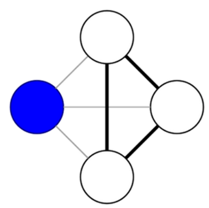
Clustering Coefficient (Local)

- These measures are
  - −> 1 if every neighbour connected to v_i is also connected to every other vertex within the neighbourhood
  - −> 0 if no vertex that is connected to v_i connects to any other vertex that is connected to v_i

Clustering Coefficient (Local)

- Example: local clustering coefficient = proportion of connections among its neighbours which are actually realised compared with the number of all possible connections
- Blue node has three neighbours, which can have a maximum of 3 connections among them
- All three possible connections are realised (thick black segments), giving a local clustering coefficient of 1

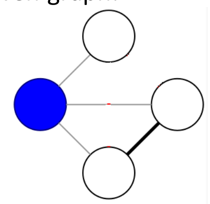
Example

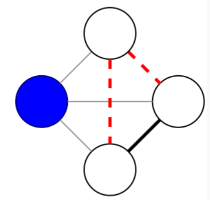


c = 1

Clustering Coefficient (Local)

- Example:
- Given graph:



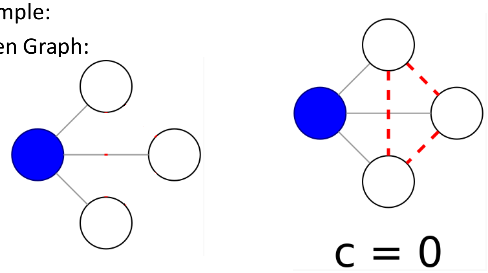


c = 1/3

- One connection is realised (thick black line) and 2 connections are missing (dotted red lines)
- Local cluster coefficient is 1/3

Clustering Coefficient (Local)

- Example:
- Given Graph:



- None of the possible connections among the neighbours of the blue node are realised, producing a local clustering coefficient value of 0

Clustering Coefficient (Local)

- If the neighbourhood is fully connected, the clustering coefficient is 1
- A value close to 0 means that there are hardly any connections in the neighbourhood

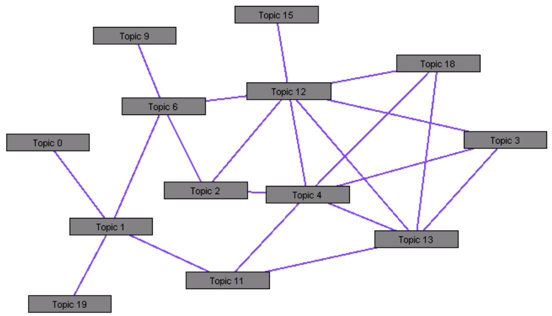
Example

- Consider a topic map that contains information about movies and actors. Each actor topic is connected to a movie topic if the actor stars in that movie. This kind of topic map would always have a clustering coefficient of 0 no matter how many topics it contained and how the movies and actors were connected. Reason for this is that the neighborhood of an actor topic will always consists of only movie topics and movie topics are never directly linked. Similarly neighborhood of a movie topic consists only of actor topics. Topics in topic maps are different in quality (or type) and direct connections between topics of same quality (or type) are rare.

Example (cont'd)

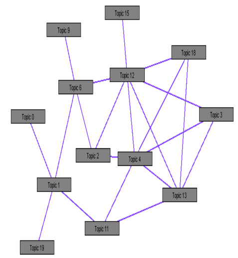
- Consider a topic map with only movie topics which are connected if they have at least one same actor

Example (cont'd)



Example (cont'd)

- Neighborhood of topic 6 consists of topics 9, 12, 2 and 1.
- Between these topics there is only one connection, from topic 2 to topic 12.
- If the four topics were fully connected, there would be 4*3/2=6 connections.
- Clustering coefficient of topic 6 is 1/6=0.17
- Clustering coefficient of topic 1 is 0 because there is no connection at all between topics 0, 6, 11 and 19
- Clustering coefficient of topic 3 is 1 because the neighborhood consisting of topics 12, 4 and 13 is fully connected



Clustering Coefficient (Global)

- Global clustering coefficient is based on triplets of nodes
- A triplet consists of three connected nodes
- A triangle includes three triplets, one centered on each of the nodes
- This measure gives an indication of the clustering in the whole network (global) and can be applied to both undirected and directed networks

Clustering Coefficient (Global)

- Global clustering coefficient is the number of closed triplets (or 3 x number of triangles) over the total number of triplets
- Global clustering coefficient is defined as:

  C = (3 × number of triangles) / number of connected triplets of vertices

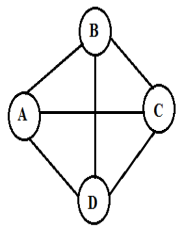
  = number of closed triplets / number of connected triplets of vertices

  - A connected triplet is defined to be a connected subgraph consisting of three vertices and two edges
  - Each triangle forms three connected triplets, leading to the factor of three in the formula

Example

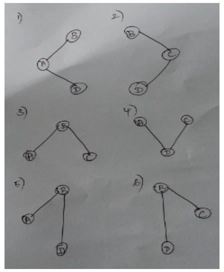
- Example: Global clustering coefficient C = (3 × number of triangles) / number of connected triplets of vertices
- C = (3 * 4)/ 12 = 1
- In the example graph, there are 12 connected triplets

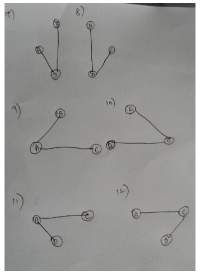
Example



Example (cont'd)

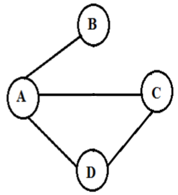
The 12 connected triplets are given below





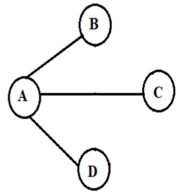
Example

- Given graph: Compute Global clustering coefficient



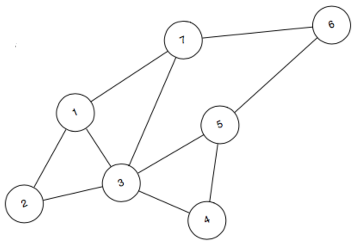
Example

- Given Graph: Compute Global clustering coefficient



Example

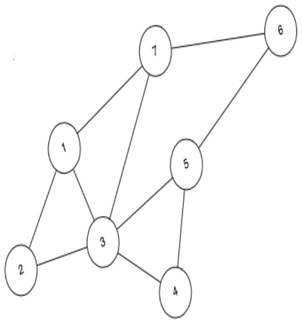
- Compute Global Clustering Coefficient for the given graph



Example (cont'd)

Triplets:

1-7-6	4-3-7
1-3-5	4-5-6
1-3-4	5-3-7
2-1-7	5-6-7
2-3-7	+ 3 *No. Of. Triangles
2-3-5	
2-3-4	
3-5-6	
3-7-6	



Transitivity of a graph

- The transitivity T of a graph is based on the **relative number of triangles in the graph, compared to total number of connected triples of nodes**.
  - T=3×number of triangles in the network/number of connected triples of nodes in the network
- The factor of three in the number accounts for the fact that each triangle contributes to three different connected triples in the graph, one centered at each node of the triangle. With this definition, $0 \le T \le 1$, and T=1 if the network contains all possible edges.
- The transitivity of a graph is closely related to the clustering coefficient of a graph, as both measure the relative frequency of triangles.