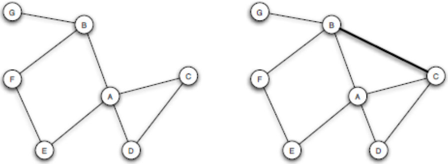
Web-graphs. Lecture 4.

Triadic Closure or Transitivity

- Basic idea of the *triadic closure*:
  If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future

From Easley & Kleinberg, Networks, crowds and markets



(a) *Before B-C edge forms.* (b) *After B-C edge forms.*

Figure 3.1: The formation of the edge between B and C illustrates the effects of triadic closure, since they have a common neighbor A.

- If nodes B and C have a friend A in common, then the formation of an edge between B and C produces a situation in which all three nodes A, B, and C have edges connecting each other – a structure we refer to as a *triangle* in the network.
- The basic role of triadic closure in social networks has motivated the formulation of simple social network measures to capture its prevalence.

Triadic Closure or Transitivity

- Why *transitivity*?
  In mathematics a relation "$\circ$" is said to be transitive if $a \circ b$ and $b \circ c$ together imply $a \circ c$. An example would be equality. If $a = b$ and $b = c$, then it follows that $a = c$ also, so "$=$" is a transitive relation. Other examples are "greater than", "less than", and "implies".
- In a network there are various relations between pairs of vertices, the simplest of which is "connected by an edge". If the "connected by an edge" relation were transitive it would mean that if vertex u is connected to vertex v, and v is connected to w, then u is also connected to w.
- Although this is only one possible kind of network transitivity – other network relations could be transitive too – it is the only one that is commonly considered, and networks showing this property are themselves said to be *transitive*.
- This definition of network transitivity could apply to either directed or undirected networks. We take the undirected case first, since it's simpler.

Global clustering coefficient

- Perfect transitivity only occurs in networks where each component is a fully connected subgraph or clique.
- Partial transitivity can be very useful. How to measure it formally?
- We can quantify the level of transitivity in a network as follows. If u knows v and v knows w, then we have a path uvw of two edges in the network. If u also knows w, we say that the path is *closed* – it forms a loop of length three, or a *triangle*, in the network. In the social network jargon, u, v, and w are said to form a *closed triad*. We define the **global clustering coefficient** $T(G)$:

$$
\begin{aligned}
T(G) &= \frac{\text{number of closed paths of length two}}{\text{number of paths of length two}} = \\
&= \frac{(\text{number of triangles}) \times 6}{\text{number of paths of length two}} = \\
&= \frac{(\text{number of triangles}) \times 3}{\text{number of connected triples}}
\end{aligned}
$$

Global clustering coefficient

Let

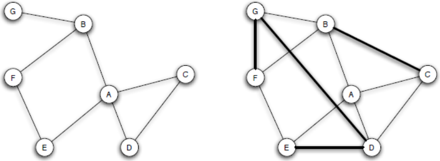
- $G = (V, E)$
- K_3 – triangle
- $\#(K_3, G)$ – number of triangles in G
- P_2 – connected triple
- $\#(P_2, G)$ – number of connected triples in G

$$T(G) = \frac{3\#(K_3, G)}{\#(P_2, G)}$$

Local clustering coefficient

- The *local* clustering coefficient of a node A is defined as the probability that two randomly selected friends of A are friends with each other. In other words, it is the fraction of pairs of A's friends that are connected to each other by edges.
- Example
  a) the clustering coefficient of node A in Figure 3.2(a) is 1/6 (because there is only the single C-D edge among the six pairs of friends B-C, B-D, B-E, C-D, C-E, and D-E)
  b) it has increased to 1/2 in the second snapshot of the network in Figure 3.2(b) (because there are now the three edges B-C, C-D, and D-E among the same six pairs).

From Easley & Kleinberg, Networks, crowds and markets



(a) *Before new edges form.* (b) *After new edges form.*

Figure 3.2: If we watch a network for a longer span of time, we can see multiple edges forming — some form through triadic closure while others (such as the *D-G* edge) form even though the two endpoints have no neighbors in common.

Local clustering coefficient

Let

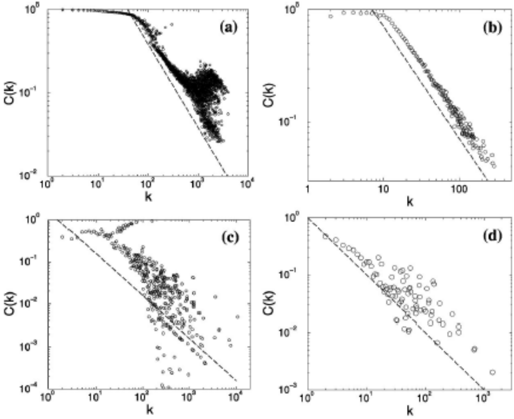
- $G = (V, E)$
- N_v be the set of node v's neighbors in G
- $n_v = |N_v|$

If $n_v \geq 2$

$$C_v = \frac{\text{number of pairs of neighbors of } v \text{ that are connected}}{\text{number of pairs of neighbors of } v} =$$
$$= \frac{|\{(x, y) \in E : x, y \in N_v\}|}{C_{n_v}^2}$$

Local clustering coefficient: real networks properties

In many real networks local clustering coefficient is found empirically to have a rough dependence on degree. Vertices with higher degree having a lower local clustering coefficient on average.

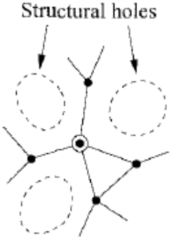


From Ravasz, E., & Barabási, A. L. (2003). Hierarchical organization in complex networks. Physical review E, 67(2), 026112.

FIG. 3. The scaling of $C(k)$ with k for four large networks: (a) Actor network, two actors being connected if they acted in the same movie according to the www.IMDB.com database. (b) The semantic web, connecting two English words if they are listed as synonyms in the Merriam Webster dictionary [27]. (c) The World Wide Web, based on the data collected in Ref. [7]. (d) Internet at the autonomous system level, each node representing a domain, connected if there is a communication link between them. The dashed line in each figure has slope -1, following Eq. (1).

Local clustering coefficient: real networks properties

- Local clustering can be used as a probe for the existence of so-called "structural holes" in a network.
- While it is common in many networks, especially social networks, for the neighbors of a vertex to be connected among themselves, it happens sometimes that these expected connections between neighbors are missing.
- If we are interested in efficient spread of information or other traffic around a network, then structural holes are a bad thing-they reduce the number of alternative routes information can take through the network.
- Central node's power: If two friends of i are not connected directly and their information about one another comes instead via their mutual connection with i then i can control the flow of that information.



When the neighbors of a node are not connected to one another we say the network contains "structural holes."

From M.Newman, Networks. An Introduction

Two overall clustering coefficients

- There are two characteristics of the overall network clustering:
  - global clustering coefficient,
  - mean local clustering coefficient:

$$C(G) = \frac{1}{|V|} \sum_{v \in V} C_v \tag{1}$$

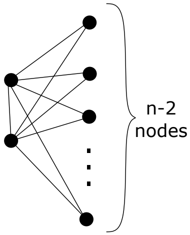
- Expression for the global clustering coefficient can be rewritten as follows:

$$T(G) = \frac{3\#(K_3, G)}{\#(P_2, G)} = \frac{\sum_{v \in V} C^2_{n_v} C_v}{\sum_{v \in V} C^2_{n_v}}$$

- The higher nodes degree the higher its weight in $T(G)$. $T(G)$ mostly depends on hubs local clustering properties, while $C(G)$ takes all nodes equivalently.

Example 1: Bollobas example

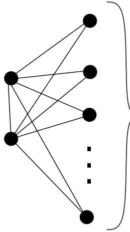
Let G be complete bipartite graph $K_{2,n-2}$



- **Local clustering coefficient**
  - For left part nodes:
$$C_v = \frac{n-2}{C_{n-1}^2} = \frac{2}{n-1}$$
  - For right part nodes:
$$C_v = 1$$
  - Mean local clustering coefficient:
$$C(G) = \frac{1}{n}\left(n - 2 + \frac{4}{n-1}\right) \to 1,\ n \to \infty$$

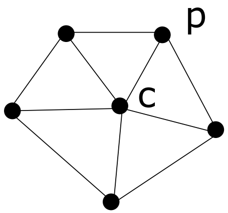
Example 1: Bollobas example

Let G be complete bipartite graph $K_{2,n-2}$



- **Global clustering coefficient**
  - Number of triangles $\#(K_3, K_{2,n-2}) = n - 2$
  - Number of connected triples:
    - with center in left nodes: $2C^2_{n-1}$
    - with center in right nodes: $n - 2$
    - total: $2C^2_{n-2} + n - 2 = (n-2)n \sim n^2$
  - $T(G) = \Theta\left(\frac{1}{n}\right)$

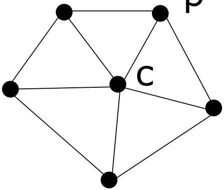
Example 2



Case $n = 5$ peripheral

- $C_c = \frac{5}{C_5^2} = \frac{1}{2}$
- $C_p = \frac{2}{C_3^2} = \frac{2}{3}$
- $C(G) = \frac{1/2+5\cdot 2/3}{6} = \frac{23}{36} \approx 0.64$
- $T(G) = \frac{\sum_{v\in V} C_{n_v}^2 C_v}{\sum_{v\in V} C_{n_v}^2} = \frac{C_5^2\cdot 1/2+5C_3^2\cdot 2/3}{C_5^2+C_3^2} = \frac{3}{5} \approx 0.6$

Example 2



General case: n peripheral nodes

- $C_c(n) = \frac{n}{C_n^2} = \frac{2}{n-1}$
- $C_p(n) = \frac{2}{C_3^2} = \frac{2}{3}$
- $C(G) = \frac{2/(n-1)+n\cdot 2/3}{n+1} \to \frac{2}{3},\ n \to \infty$
- $T(G) = \frac{\sum_{v\in V} C_{n_v}^2 C_v}{\sum_{v\in V} C_{n_v}^2} = \frac{C_n^2\cdot 2/(n-1)+nC_3^2\cdot 2/3}{C_n^2+C_3^2} \to 0,\ n \to \infty$

Theoretical results

In real web-graph $C(G) = \Theta(1)$. We have no reliable data on $T(G)$.

Theorem (Ostroumova, Samosvat). Let $\{G_n\}$ be a sequence of graphs with growing number of nodes n and power-law degree distribution with exponent $\gamma \in (2, 3)$, then

$$T(G_n) \to 0, \quad n \to \infty$$

Theorem (Ostroumova). There exists sequence of multi-graphs $\{G_n\}$ with loops, growing number of nodes n and power-law degree distribution with exponent $\gamma \in (2, 3)$ in which

$$T(G_n) \geq \text{const}, \quad n \to \infty$$

Theorem (Bollobas, Riordan). Let $m \geq 1, \;\; G_m^n$ be the random graph in the Bollobas-Riordan model. Then

$$\mathbb{E}(T(G_m^n)) \asymp \frac{\ln^2 n}{n}$$

Theoretical results

Theorem (Ryabchenko, Samosvat). Let $m \geq 1$, G_m^n be the random graph in the Bollobas-Riordan model and H be some fixed graph. Then

$$\mathbb{E}\left(\#(H, G_m^n)\right) \asymp n^{\#(d_i=0)} (\sqrt{n})^{\#(d_i=1)} (\ln n)^{\#(d_i=2)},$$

where $\#(d_i = k)$ is the number of nodes with degree k in H.

- Asymptotically, there are no graphs in which $d_i > 2,\ \forall i$
- $C(G_m^n) \to 0$ in the Bollobas-Riordan model.