Problem 1:

Social Computing

Assignment-02

Sharmin Sultana
ID : 01339552
Sharmin_Sultana@student.uml.edu.

Problem 1: Clustering Coefficient and Neighborhood OverlaP

Given that,

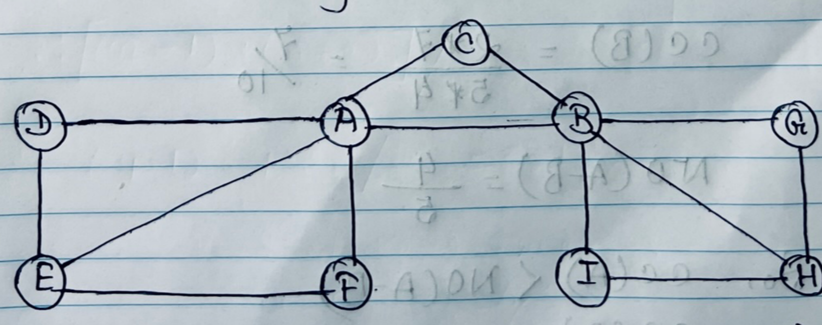
$$CC(V) = \frac{2 * N_v}{K_v * (K_v - 1)} \quad ; \quad \begin{array}{l} N_v = \# \text{ of nodes connected with } V \\ K_v = \# \text{ of links bet}^n \text{ neighbors of } V \end{array}$$

$$NO(A\text{-}B) = \frac{\# \text{ of neighbors connected with both } A \,\&\, B}{\# \text{ of neighbors who are connected at least one of } A \,\&\, B}$$

(i)

" both a, b have high CC but (a, b) has low NO "

Since, we need to build a graph where pair (a, b) will have higher clustering coefficiency, the number of connected nodes with a & b should be higher. Again, to make the pair with low neighborhood overlap, we need to have greater number of nodes that are at least connected with a & b than the common nodes shared by them.



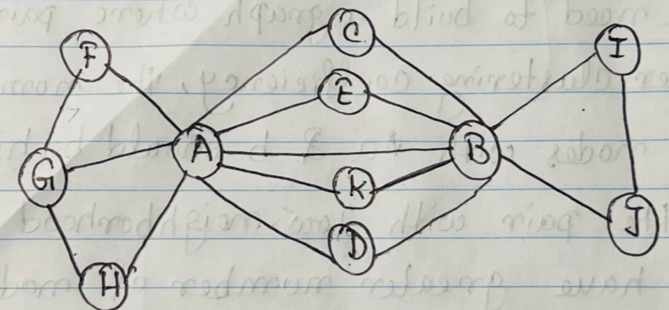
Here, $CC(A) = \frac{2 * \cancel{4}5}{3 * \cancel{4}2} = 5/3$ $\qquad CC(B) = \frac{2 * 5}{3 * 2} = 5/3$

$NO(A-B) = \frac{1}{\cancel{3}3+3} = 1/6$ So, $CC(A) > NO(A-B)$
$CC(B) > NO(A-B$

②

"both a and b have low clustering coefficient but (a,b) has high neighborhood overlap"

The resultant graph should have less num of neighbors for both a & b ~~B~~ but the the num. of common neighbors should be high.



Here, $CC(A) = \frac{2*8}{6*5} = 8/15$

$CC(B) = \frac{2*7}{5*4} = 7/10$

$NO(A-B) = \frac{4}{5}$

So, $CC(A) < NO(A-B)$
$CC(B) < NO(A-B)$

Problem 2:

1. The expected level for the edge between b and c will be W.

   According to Strong Triadic Closure, "we say that a node A violates the Strong Triadic Closure Property if it has strong ties to two other nodes B and C, and there is no edge at all (either a strong or weak tie) between B and C. We say that a node A satisfies the Strong Triadic Closure Property if it does not violate it."

   In our given example graph, if b-c is S then there must need to have another edge b-d with S ties to satisfy the above rule. Since there is no edge in between b and d, the level for b-c will be W.

2. In social network's perspective, strong ties refer to the people (nodes) with whom we are closely connected, and weak ties refer our acquaintances.

   Strength of weak ties explains the above comparison in following way:

   I. Weak ties connect us to new source of information.

   II. Weak ties don't need to be maintained continuously.

   III. With two or more weak ties, we can create more smaller clusters and together they form bigger community with verities of information.

Problem 3:

1. As an example, suppose we have a graph with 100,000 nodes, where 10,000 nodes are in the largest strongly connected component (SCC), and another 20,000 nodes are in the set of incoming nodes and 30,000 nodes are in the outgoing set. If we remove an edge between two nodes that are both in the SCC, and this edge was the only connection between this SCC and the rest of the graph, the SCC will be split into two or more smaller SCCs, each containing no more than 9,000 nodes, reducing the size of the largest SCC by at least 1,000 nodes.
2. Consider a directed graph with a large set of incoming nodes, a small strongly connected component, and a large set of outgoing nodes, with a total of 100,000 nodes. Let's assume that there are 10,000 nodes in the incoming set, 1,000 nodes in the strongly connected component, and 50,000 nodes in the outgoing set. Adding a directed edge from a node in the SCC to a node in the outgoing set that was previously unreachable from the SCC will

connect all nodes that are reachable from that node to the SCC. Consequently, the size of the outgoing set will be reduced by at least 1,000 nodes. It is possible that the size of the SCC increases as a result of this change, but it is not necessary for the size of the outgoing set to decrease.

Problem 4:

Let's make some assumptions for both cases:

Case 1: Consider that every student can choose any classes in UML. As the course registration portal shows the number of students that have already registered for a particular course, so students tend to take the classes that have more students because of the popularity.

Case 2: In case of elementary schools, most of them are location based. So most of the students of a certain neighborhood go to their neighborhood schools. Also assume that every elementary school have one 3rd grade class.

I would expect the first case, regarding the fraction of UML classes with k students enrolled, to follow a power-law distribution more closely as a function of k. In Power-law distributions a few nodes have many connections, while many nodes have only a few connections. UML classes may exhibit similar characteristics, with a few classes having a large number of students, while many classes have only a few students. In contrast, the number of students in 3rd-grade elementary school classrooms may be more evenly distributed, with fewer outliers, and may therefore not follow a power-law distribution as closely.

Problem 5:

In my opinion, we won't observe power laws and rich-get-richer phenomena since the model adds edges between nodes selected at random. This generates a graph with an approximately uniform degree distribution, unlike complex systems that show hierarchical structures with a few highly connected nodes. The model also lacks preferential attachment, a key feature of rich-get-richer phenomena. Therefore, the is not useful for generating random graphs that will have power laws or rich-get-richer effects.