Social Network Analysis (SNA)

Clustering: Real vs. Random

Objective: To calculate a particular graph parameter on a real social network dataset and compare it to the value of the same parameter on a randomly generated graph.

Tools used: NetworkX

Dataset: SNAP dataset

Procedure: The following are the subtasks to approach this graph problem.

1. First task is to calculate the edge probability. We know, there is an explicit formula that determines the number of expected edges in a graph.

   E = p(n(n-1)/2), where p = edge probability, n = number of nodes.

   I get the number of edges from the code:

   ```
   >>>edges = G_fb.number_of_edges()
   ```

   Then I get the number of nodes in the graph from the code below:

   ```
   >>>nodes = G_fb.number_of_nodes()
   ```

   Then, edge probability = no. of edges / [n*(n-1)/2]

2. Now,for the second task assigned, the graph parameter to be used, here, is the *average clustering coefficients (ACC)* which is computed as the mean value of the *local clustering coefficients (LCC)* for all vertices in the graph.

   To calculate ACC for graph 1(G-fb), I use the following code:

   ```
   >>>av_clust_coeff = nx.average_clustering(G_fb)
   ```

3. Third task needs to generate an Erdos Renyl random graph, whose ACC value needs to be compared with the ACC value of graph 1.

   The following syntax is used for the same:

   ```
   >>>G_rand= nx.erdos_renyi_graph(n,p)
   ```

   Here, n = no. of vertices and p = edge probability

Computing the ACC of this graph:

```
>>>av_clust_coeff_2= nx.average_clustering(G_rand)
```

Calculation:

Task 1:

Since an edge is a combination of 2 vertices from a pool of n vertices, the number of these is

Edges (E) = `n!/((n-2)!*2)`
= `n*(n-1)*(n-2)!/((n-2)!*2)`
= `n*(n-1)/2`... the (n-2)! cancels out.

Expected Edges = `[n*(n-1)/2]*p`

or, `p` = Expected Edges / `[n*(n-1)/2]`
`p = 88234 / 8154741`
`p = 0.0108199635`

Task 2:

ACC for G_fb = 0.6055467186200876

Task 3:

ACC for G_rand = 0.010682206700308675

By: Subhodeep Sinha
Dated: 23-10-2020