

Assign 6 ReadMe.

(Akshaj Patil) MT19111.

DataSet: CollegeMsg Temperol network.

This dataset contains private messages sent on online social network at University of California.

Dataset statistics	
Nodes	1899
Temporal Edges	59835
Edges in static graph	20296
Time span	193 days

Dataset contains 1899 nodes they are from 1 to 1899.

PreProcessing DataSet:

Here dataset is in text format. We load the file and one by one read lines of file. First parameter means node 'u' second parameter is node 'v' and this signifies edge from (u to v). This network is directed graph.

We then design adjacency list and adjacency matrix for the above network (Directed Graph).

Assumptions:

Total edges contained in network are 59835 edges. Since it is communication network it contains many edge from 'u' to 'v'. But here we only consider one edge. So out total edges become 20296.

Methodology:

Q1: Formulas used.

Avg In-Degree=(sum of in-degree of all nodes)/(total nodes).

Avg out-degree=(sum of out-degree of all nodes)/(total nodes).

Density of Network= (total edges in network)/(total_nodes*(total_nodes -1))

Degree Distribution of network:

Number of nodes for each degree is calculated. And degree distribution graph is plotted between degree as X-axis and fraction of node of that degree as Y-axis.

Above method is followed both for In-Degree and Out-Degree.

Clustering Coefficient:

Neighbors Ni of each node is calculated. Node 'v' is neighbor of particular node 'u' if there is edge from 'u' to 'v' OR edge from 'v' to 'u' OR both.

Clustering coefficient is calculated using following formula.

coefficient for directed graphs is given as

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

Where 'Ki' is number of vertices in the neighborhoods of that particular node.

Centrality Measure:

Here we have used Degree centrality measure.

Therefore Centrality Measure of each node is Degree of that Node.

Here we have plotted graph for both In-Degree Centrality measure and Out-Degree centrality measure.

Q2]

NetworkX library is used to calculate pagerank, hub and authority score.

1:Page Rank: networkx.pagerank(G) returns dictionary of every node with its pagerank value. Here 'G' is network graph.

2:Hubs and Authority score: network.hits(G) returns two dictionary one of nodes with hub score and second one is nodes with authority score.

How to run program:

Load dataset with proper path.

Then run all blocks one after another.