Members of Group 57:
1. **Prakhar Kumar**
2. **Meenal Jain**

Dataset Link:
https://snap.stanford.edu/data/p2p-Gnutella06.html

Preprocessing -

Downloaded graph dataset has some extra information that is removed. We read line by line and remove unwanted lines and then create the graph.

Methodology -

1. Process line by line.
2. Split the line on \t to get an edge from u to v.
3. In the adjacency matrix mark a 1 for the edge.
4. In the edge list add this pair as an edge.
5. For indegree and outdegree calculation the edge list is used.
6. For clustering coefficients an adjacency matrix is used. We check for neighbors of current nodes and check how many of the neighbors are connected with each other and then divide by total possible connections between the neighbors to get local clustering coefficient for a node.
7. For pageRank we first create a networkx graph from the adjacency matrix.
8. For authority and hub score the same networkx graph is used.
9. The algorithm for the above two is shown in the code file.

Formulas:

Local clustering Coefficient:
(number of connections between all neighbors of a node) **/** (total possible connections that can be made between the neighbors of that node)

Total connections = n * (n-1) / 2

PageRank:
Formula mentioned in this video explaining pagerank -
https://www.youtube.com/watch?v=F5fcEtqysGs

Hubs and Authority:
Formula mentioned in this video explaining hits algorithm -
https://www.youtube.com/watch?v=-kiKUYM9Qq8&t=627s