

Assignment-1

Programming Language: Python

Libraries/ Dependencies

1. Pandas
2. Matplotlib
3. Seaborn
4. NumPy
5. SciPy
6. Networkx

Input file: “BIOGRID-ORGANISM-Homo_sapiens-4.4.218.tab3.txt”

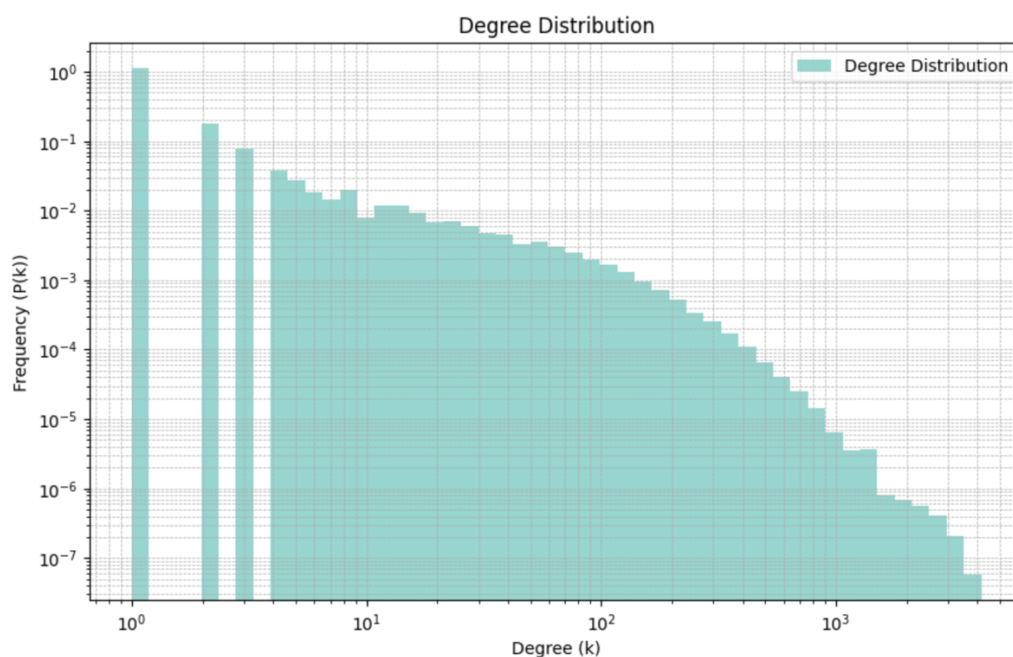
Question 1: For each node compute the degree and identify top 10 hub nodes

Total Nodes: 26771
Total Edges: 856635

Top 10 Hub Nodes by Degree:

ZRANB1: 4160
PARK2: 3426
EGFR: 3070
PLEKHA4: 2971
KIAA1429: 2932
KRAS: 2891
MYC: 2817
CCNF: 2798
CUL3: 2656
TRIM25: 2497

Question-2: Draw the graph of degree distribution to see if it is a scale free network or not?

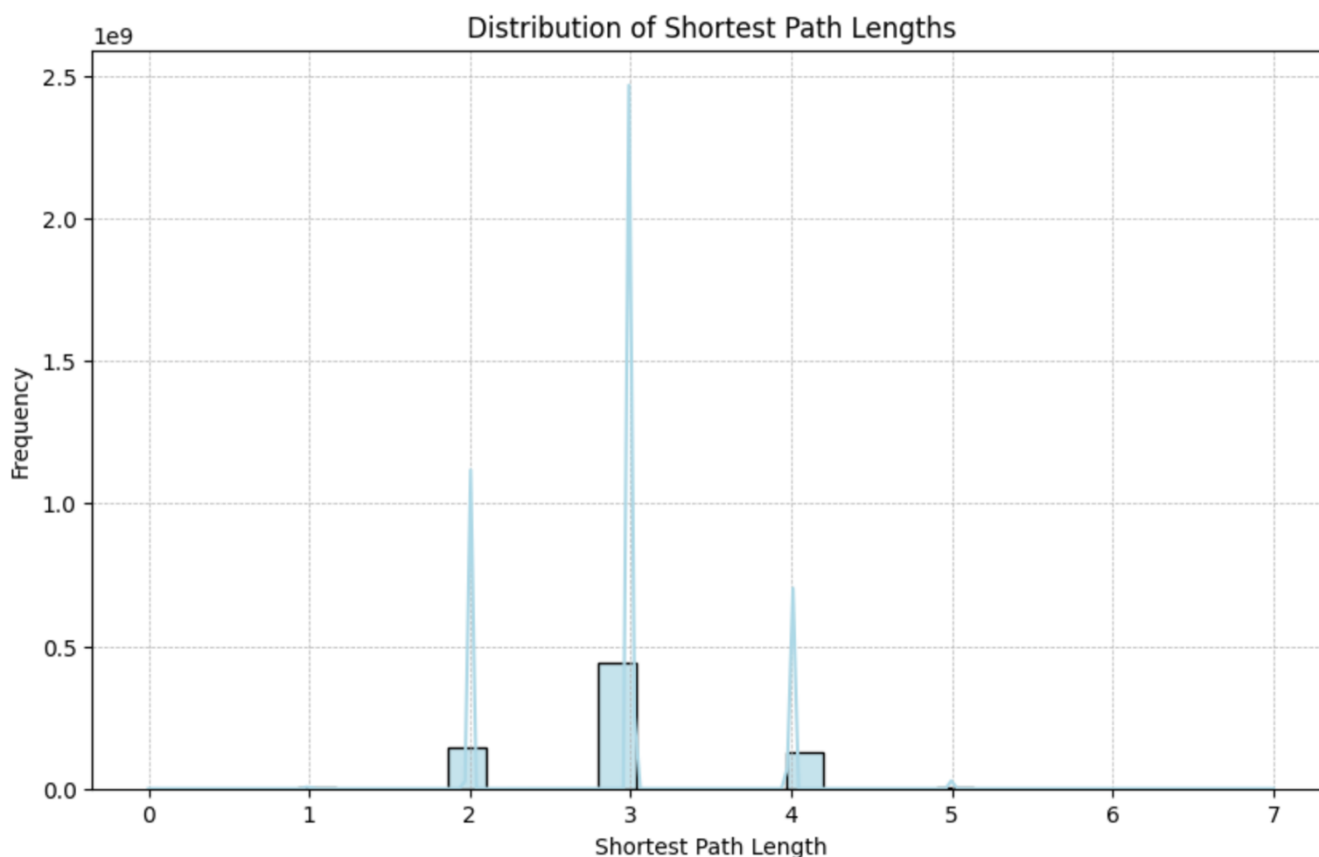


On a log-log scale, the degree distribution graph displays a heavy-tailed pattern, meaning that certain nodes function as highly linked hubs, whereas most nodes have modest degrees. Given that the degree roughly follows a power-law distribution, which is a typical feature of biological networks, this distribution points to a possible scale-free network. Although these networks can withstand sporadic outages, they are susceptible

to focused hub attacks. In keeping with the anticipated characteristics of scale-free networks, the existence of a small number of high-degree nodes validates the prior hub identification.

Question – 3: Compute the shortest path for all node pairs and plot the distribution of shortest path length.

Total shortest path lengths computed: 716525827



Question – 4: For each node, compute its betweenness and closeness, and identify top 10 nodes with highest betweenness and closeness respectively.

Top 10 Nodes by Betweenness Centrality:

UBC: 0.045594

NR3C1: 0.036977

CCNF: 0.036762

APP: 0.029426

KIAA1429: 0.028654

TRIM25: 0.027188

PARK2: 0.024566

ZRANB1: 0.023217

TP53: 0.022526

EGFR: 0.022525

Computing closeness centrality...

Top 10 Nodes by Harmonic Closeness Centrality:

ZRANB1: 0.549418

PARK2: 0.536138

EGFR: 0.529757

PLEKHA4: 0.524511

KIAA1429: 0.522796

CUL3: 0.520676

MYC: 0.516455

CCNF: 0.515839

HSPA8: 0.514877

TP53: 0.514198