Network Analysis and Artificial Network Modeling

HPC Masters

SONIA MATEI

MORAR CRISTINA-MIRELA

TURCU CIPRIAN STELIAN

MURARIU TUDOR-CRISTIAN

PROJECT OBJECTIVES

- Analyze the structure of a real-world social network
- Create two artificial models based on the original network:
  - **Erdős–Rényi (ER)** random graph
  - **Barabási–Albert (BA)** scale-free graph
- Evaluate and compare network properties:
  - Visualization
  - Degree distribution
  - Clustering coefficient
  - Distances
- Determine which synthetic model best approximates the real-world network's behavior and topology.

DATASET OVERVIEW

- **Network Repository**: https://networkrepository.com/socfb-Reed98.php
- **Description**
  - A Facebook friendship network of students from **Reed College**.
  - Represents undirected, unweighted edges between users (friendship ties).
- **Key Statistics**
  - **Number of Nodes:** 962
  - **Number of Edges:** 18,812
  - **Average Degree:** 39
  - **Density:** 0.0407
  - **Clustering Coefficient:** 0.318
  - **Triangles:** ~291,400

VISUALIZATION

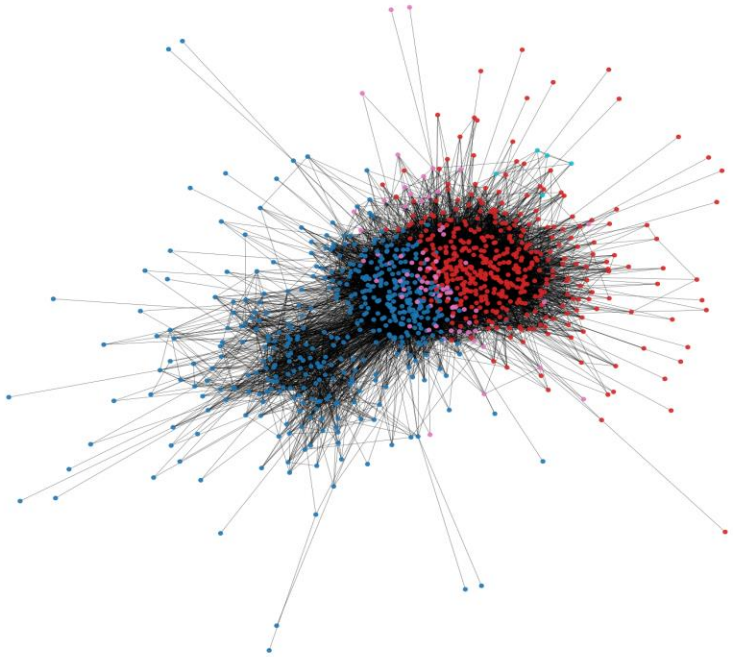
We'll go over 6 methods to visualize our 3 datasets:

1. 🌿 Spring Layout
2. 🔷 Fruchterman-Reingold Layout
3. 🔗 Kamada-Kawai Layout
4. 🧠 Spectral Layout
5. 🐚 Shell Layout
6. 🎲 Random Layout

SPRING LAYOUT

Spring layout simulates a force-directed layout where nodes repel each other like charged particles and edges act like springs.

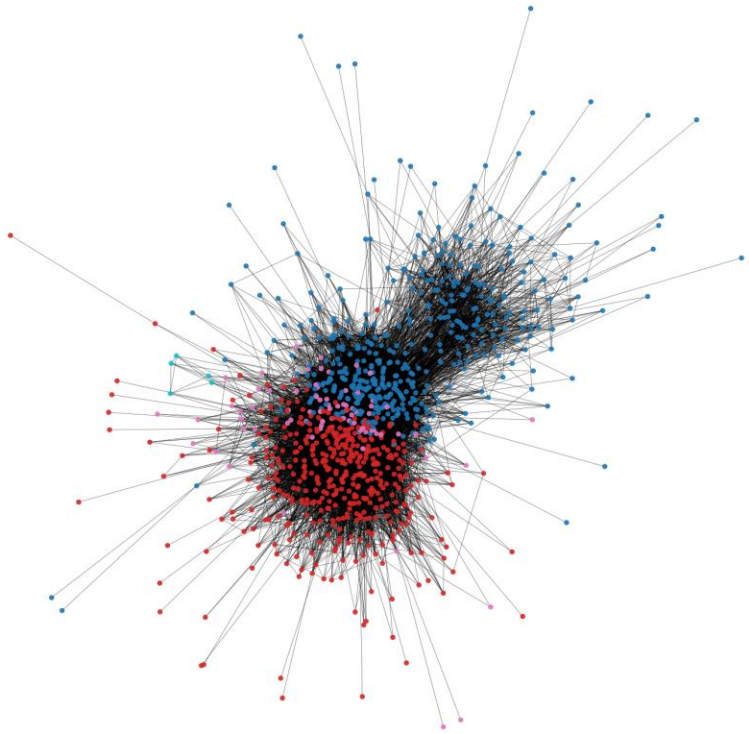
socfb-Reed98 - Spring Layout



FRUCHTERMAN-REINGOLD LAYOUT

Fruchterman-Reingold is a classical force-directed algorithm aiming for aesthetically pleasing layouts.

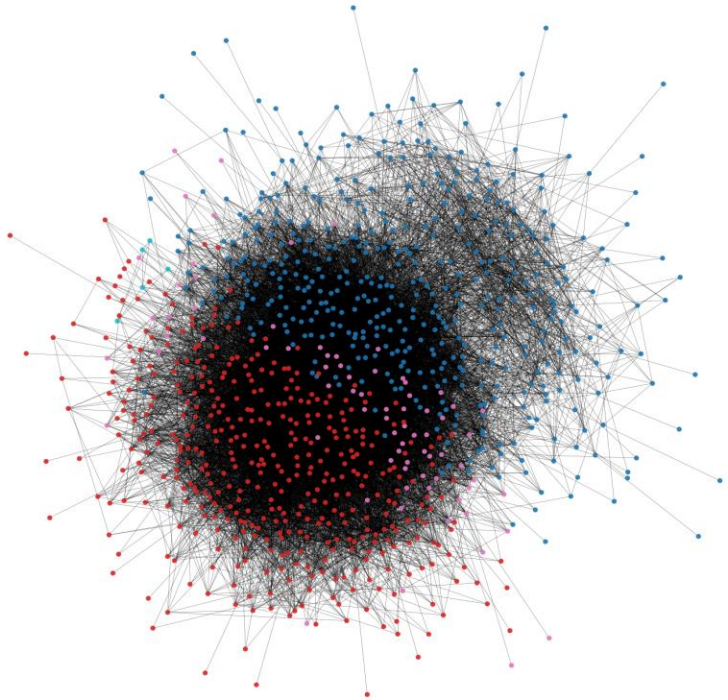
socfb-Reed98 - Fruchterman_reingold Layout



🔗 KAMADA-KAWAI LAYOUT

Kamada-Kawai layout is another force-directed algorithm that minimizes energy between node pairs based on graph-theoretic distances.

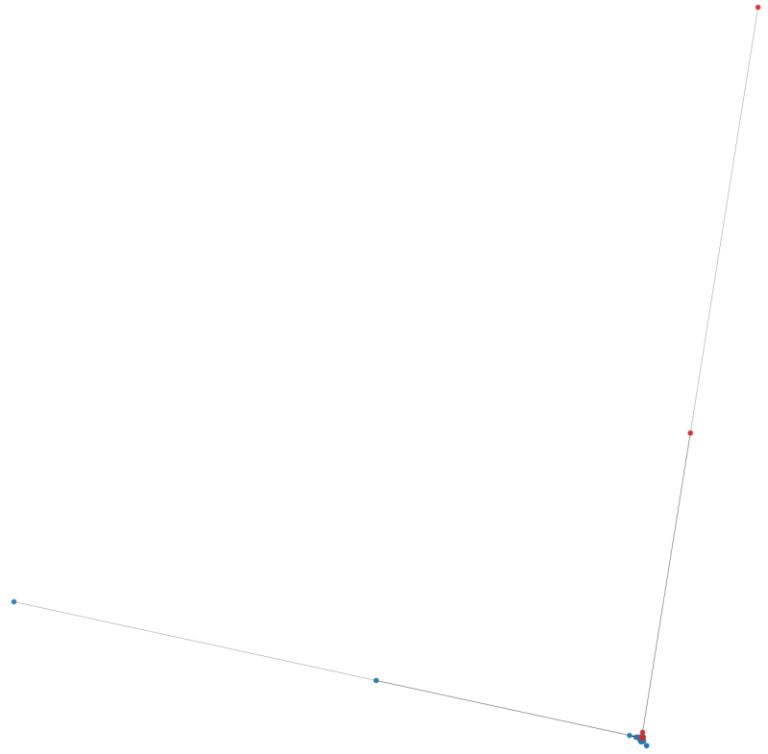
socfb-Reed98 - Kamada_kawai Layout



🧠 SPECTRAL LAYOUT

Spectral layout uses eigenvectors of the graph Laplacian to position nodes based on graph structure.

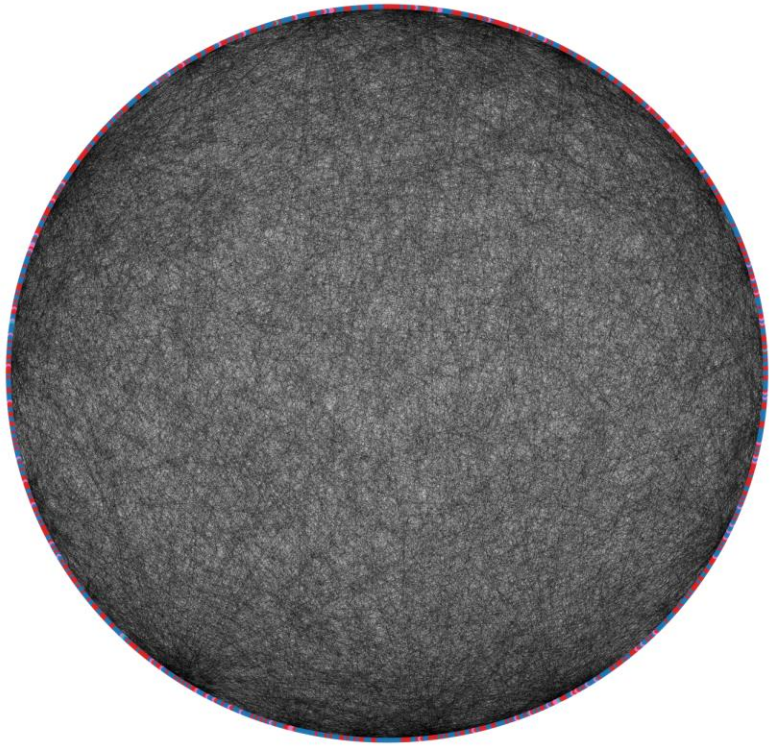
socfb-Reed98 - Spectral Layout



SHELL LAYOUT

Shell layout arranges nodes in concentric circles, good for layered or hierarchical structures.

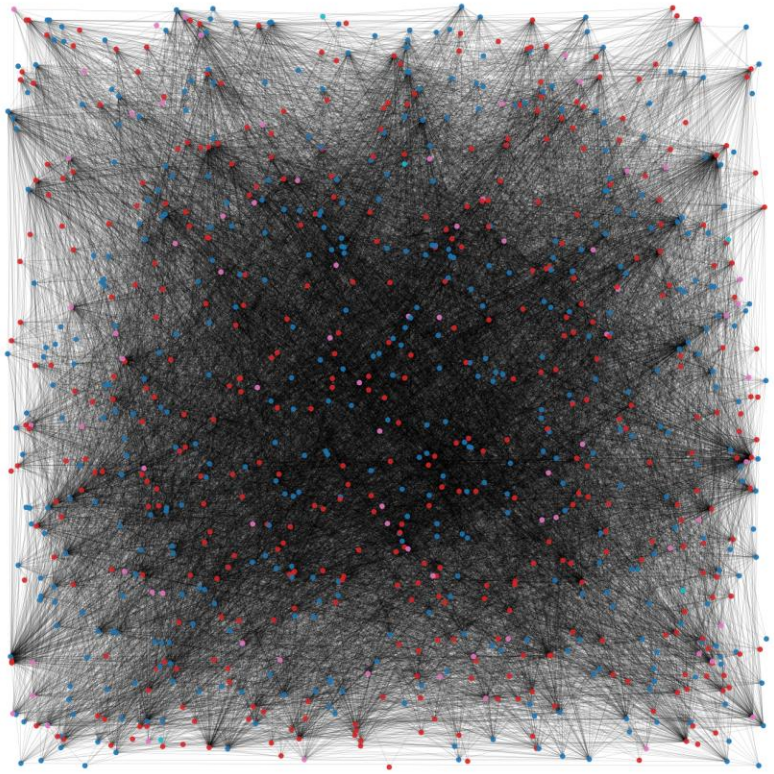
socfb-Reed98 - Shell Layout



RANDOM LAYOUT

Random layout places nodes arbitrarily in space, useful as a baseline or for testing layout effects.

socfb-Reed98 - Random Layout



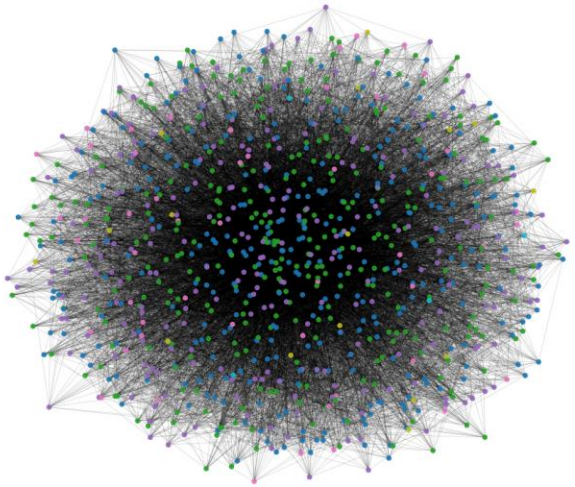
OBSERVATIONS

We can see that the first 3 ways to visualize the network matched very well with how the communities are distributed

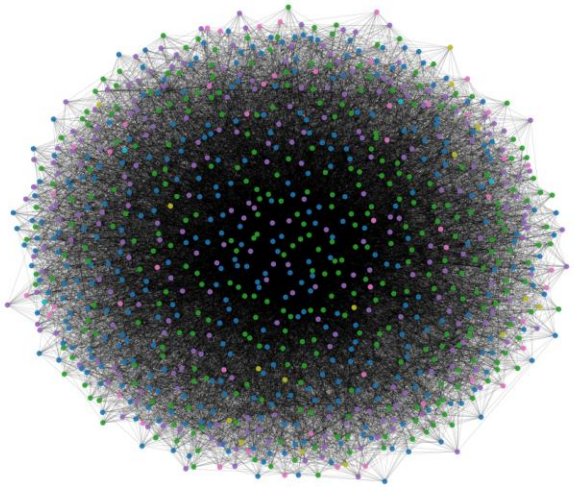
Our original network seems to be divided into two main clusters (comunities)

socfb-Reed98_BA - All Layouts

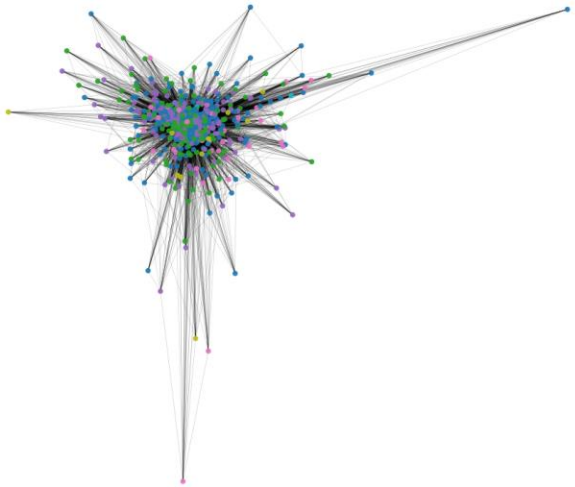
Spring Layout



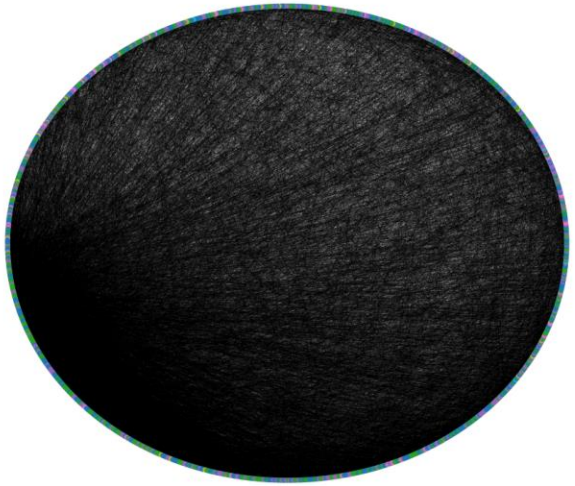
Kamada_kawai Layout



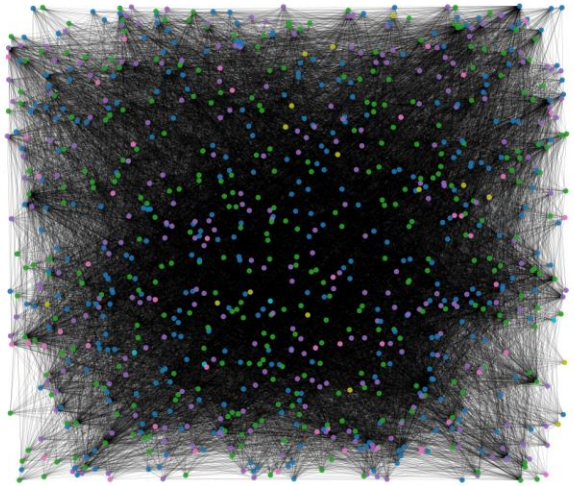
Spectral Layout



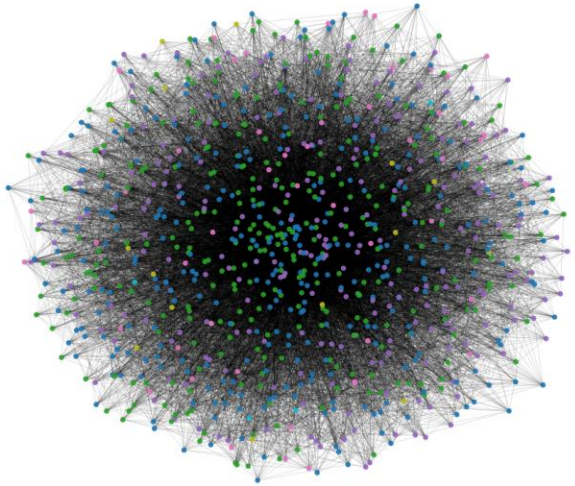
Shell Layout



Random Layout

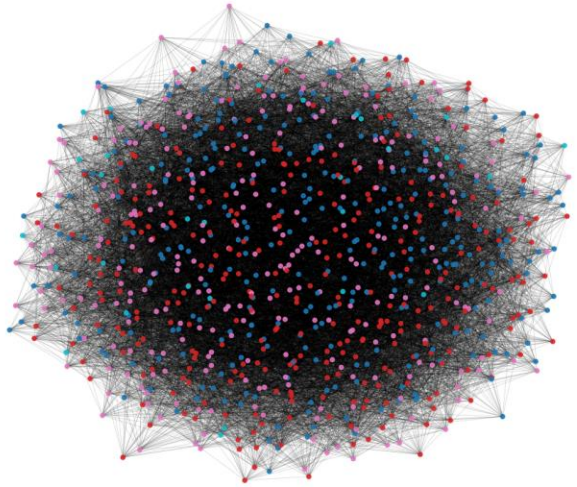


Fruchterman_reingold Layout

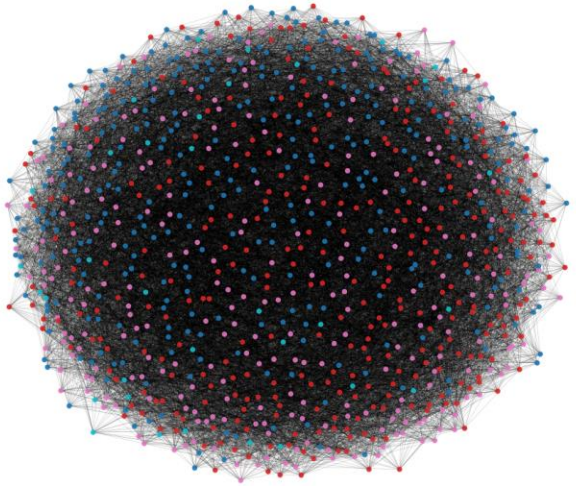


socfb-Reed98_ER - All Layouts

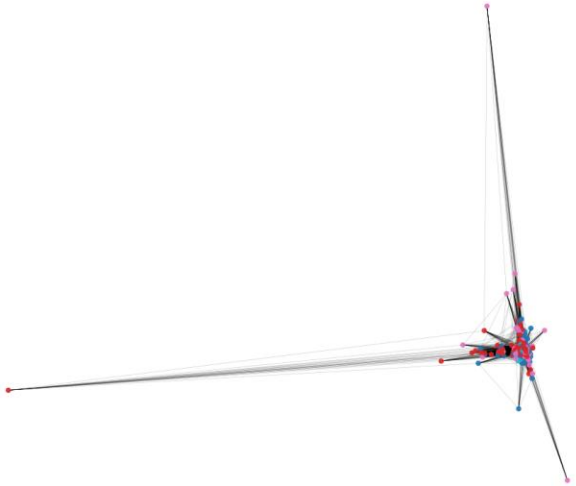
Spring Layout



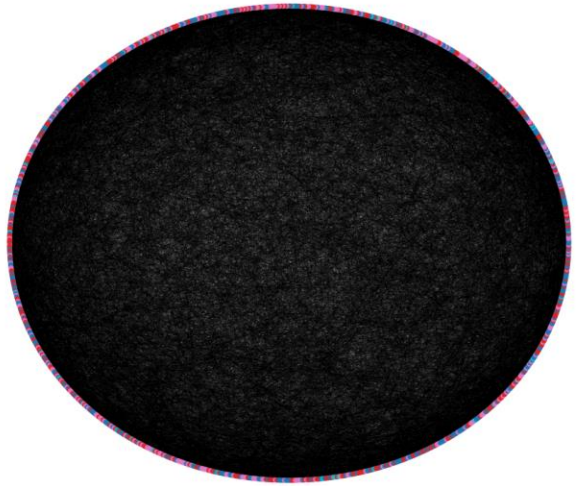
Kamada_kawai Layout



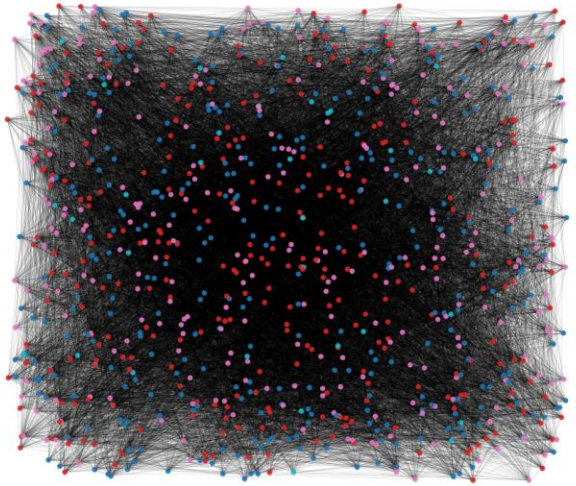
Spectral Layout



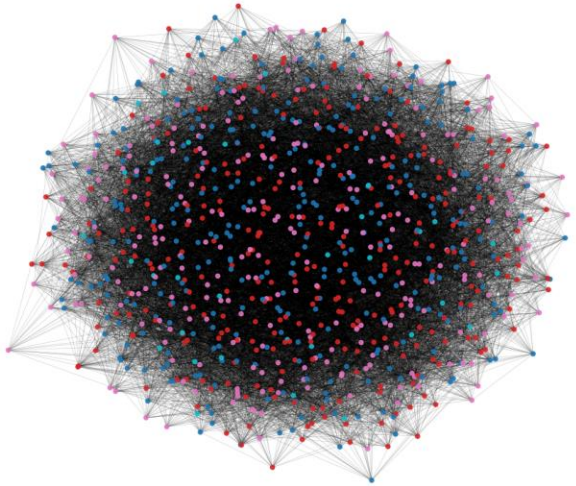
Shell Layout



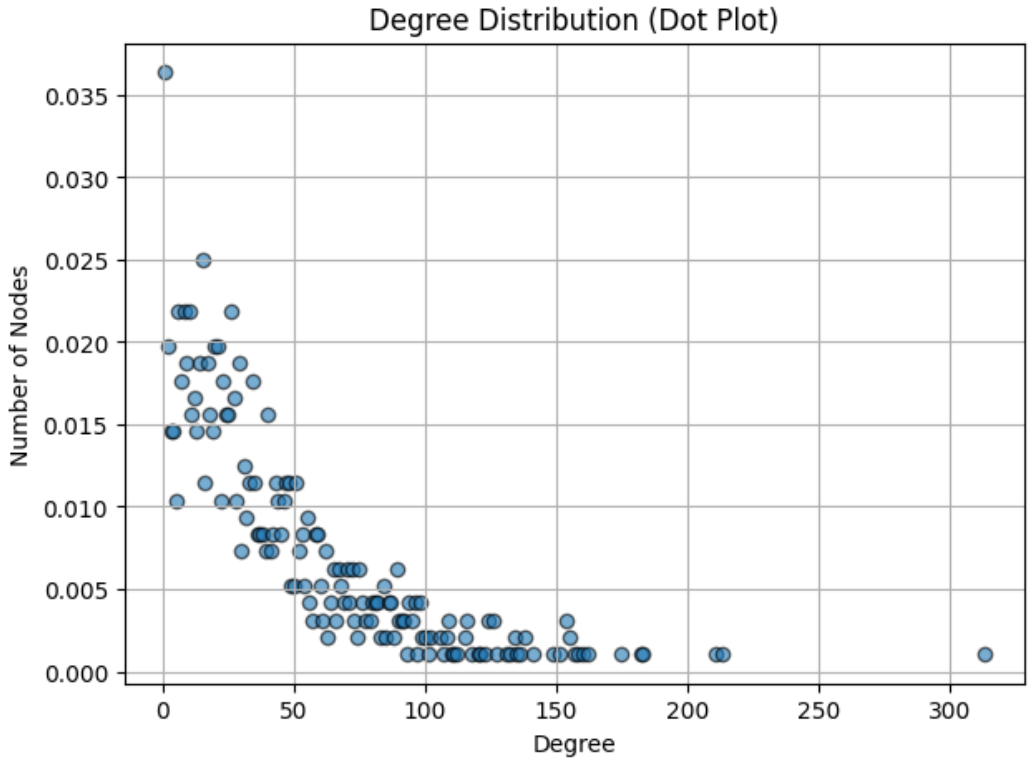
Random Layout



Fruchterman_reingold Layout

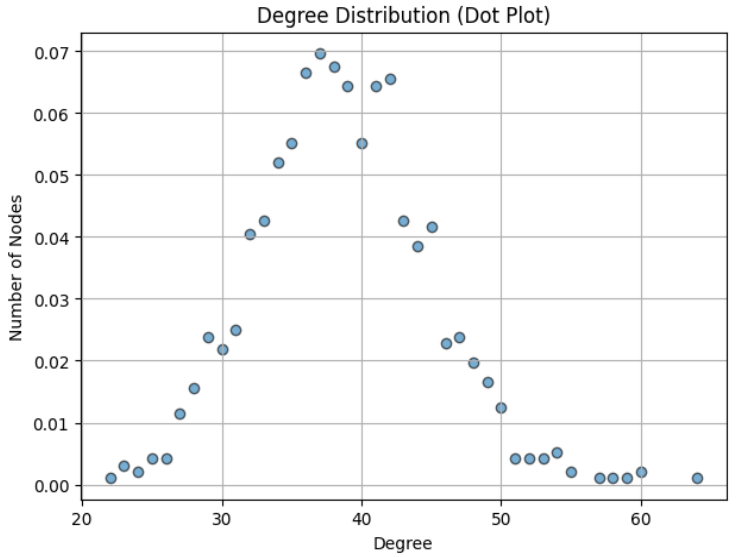


DEGREE DISTRIBUTION



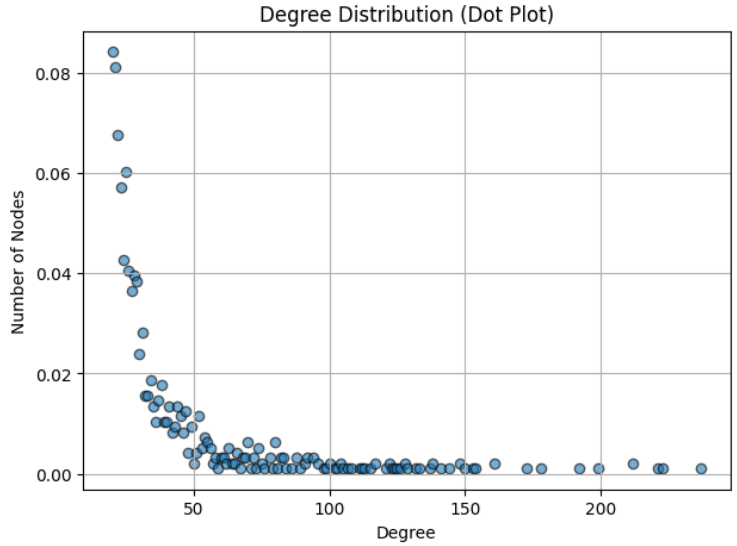
Original Network

Min Degree: 1
Max Degree: 313
Average Degree: 39.110187110187110187



Random Network

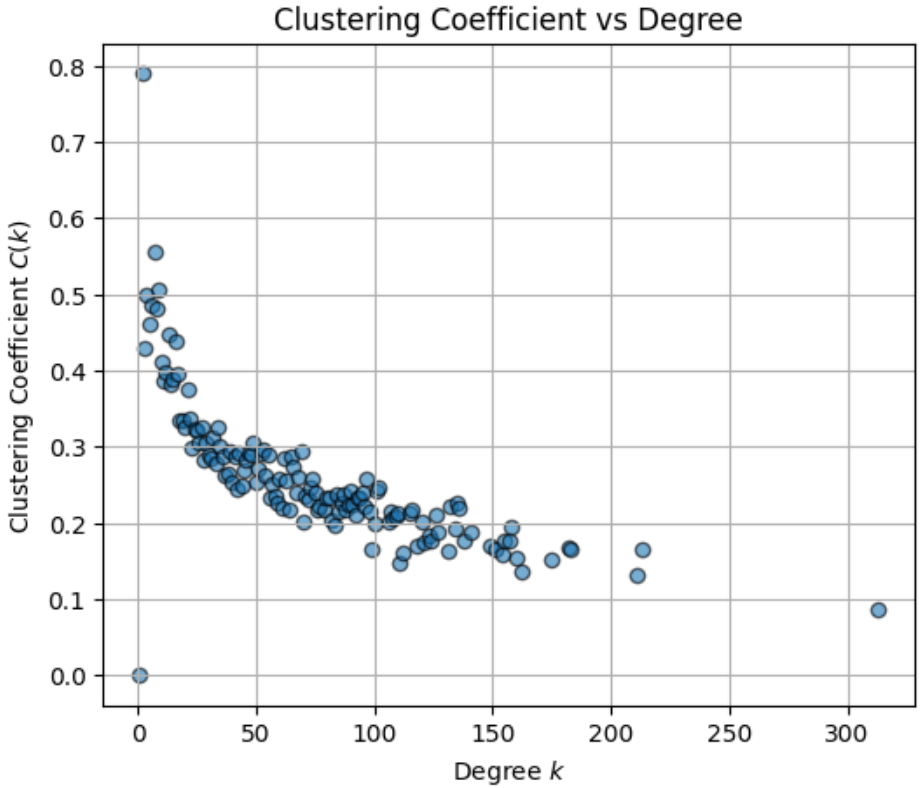
Min Degree: 22
Max Degree: 64
Average Degree: 38.6



Scale-Free Network

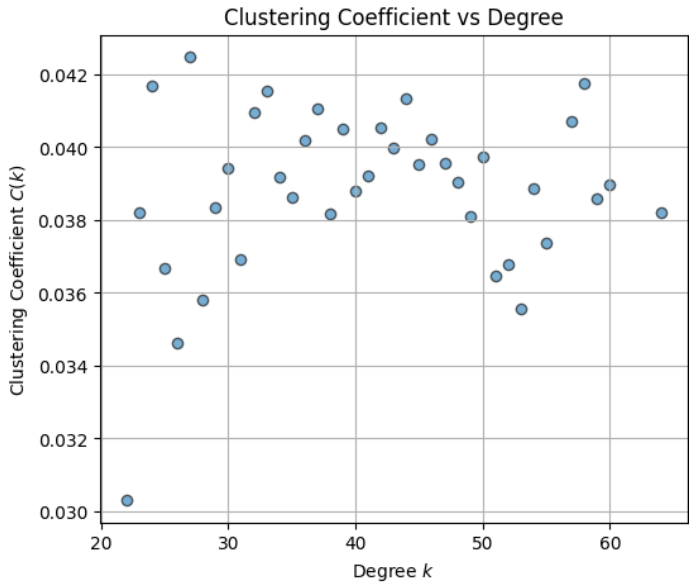
Min Degree: 20
Max Degree: 237
Average Degree: 39.1

CLUSTERING COEFFICIENT



Original Network

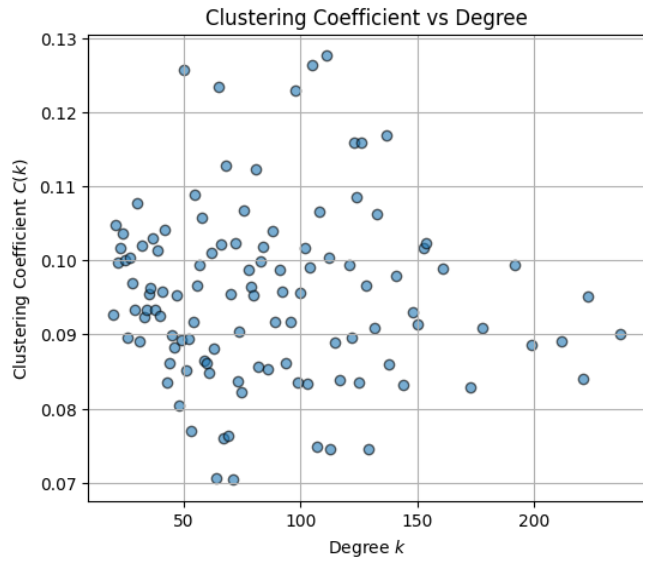
Average Clustering Coefficient: 0.318
Edge Density: 0.040
Clustering Coefficient is high compared to the edge density.



Random Network

Average Clustering Coefficient: 0.039

Clustering Coefficient is low compared to the edge density.

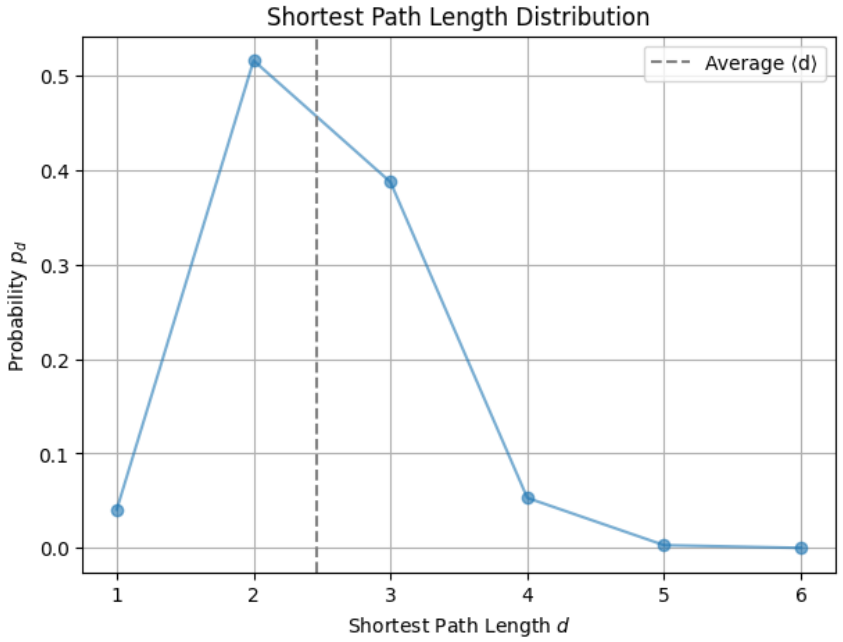


Scale-Free Network

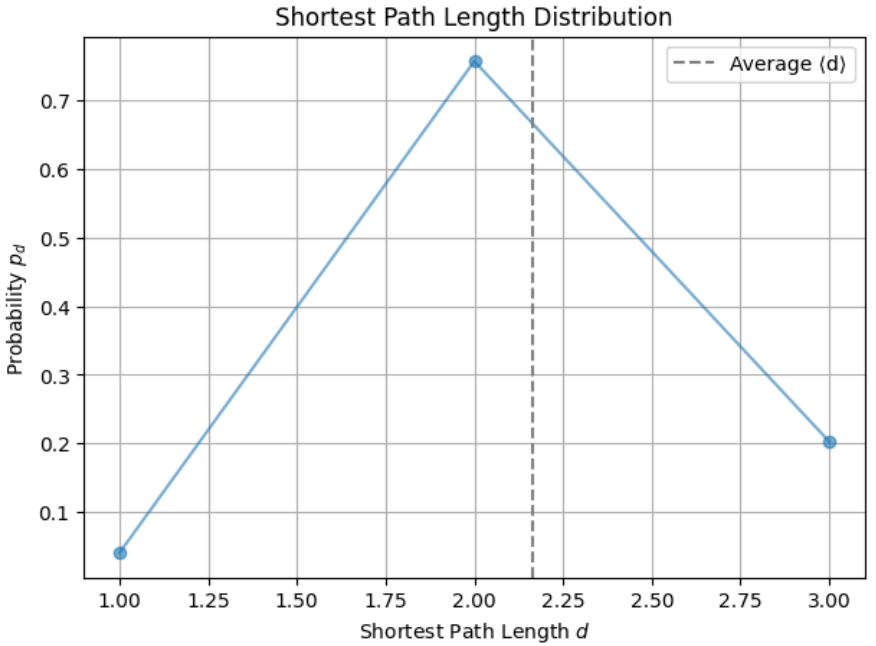
Average Clustering Coefficient: 0.096

Clustering Coefficient is high compared to the edge density.

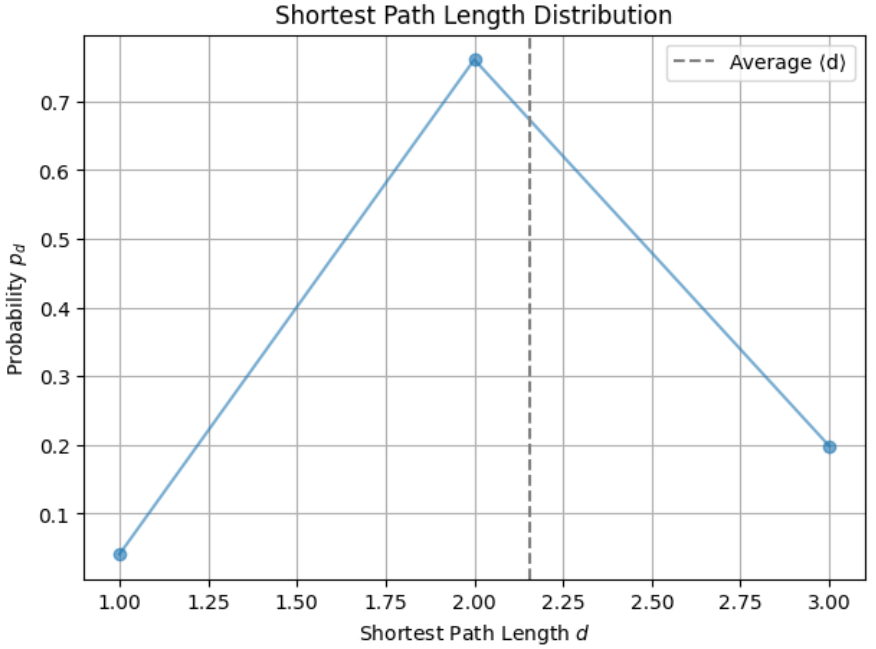
DISTANCES



Original Network

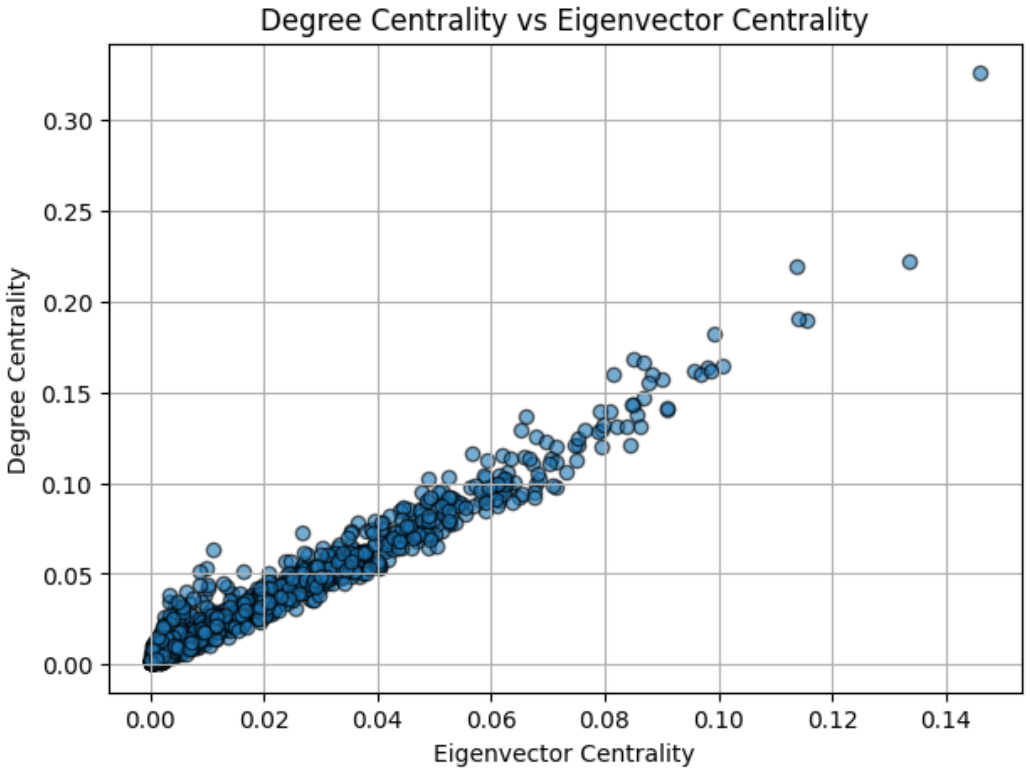


Random Network

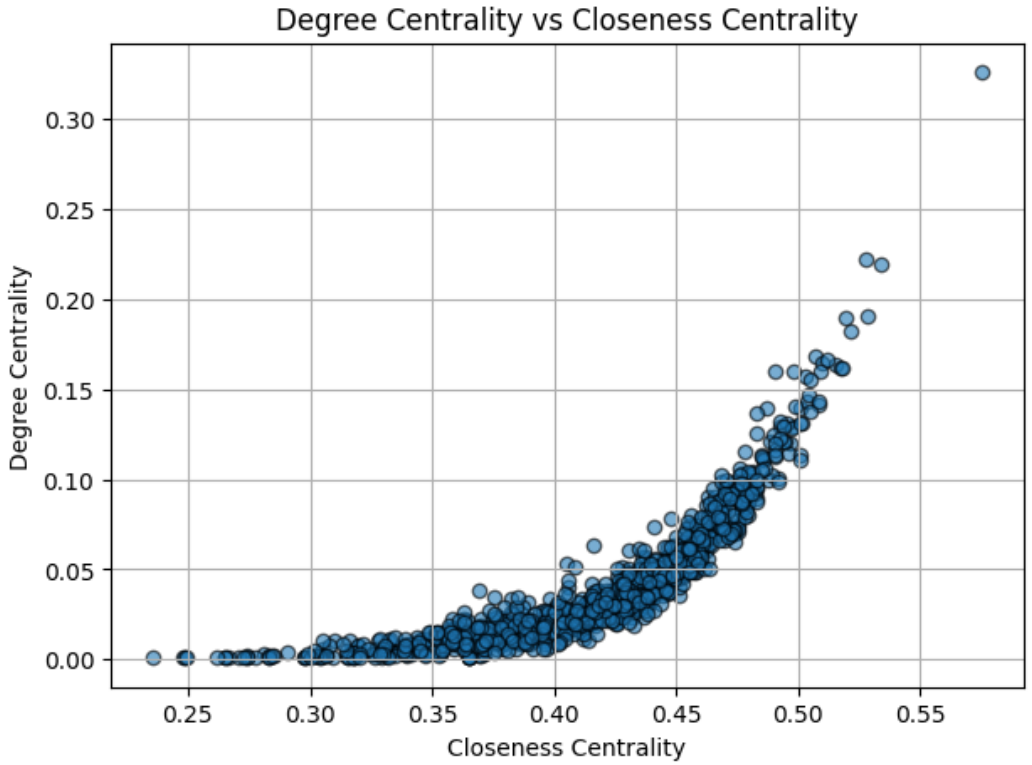


Scale-Free Network

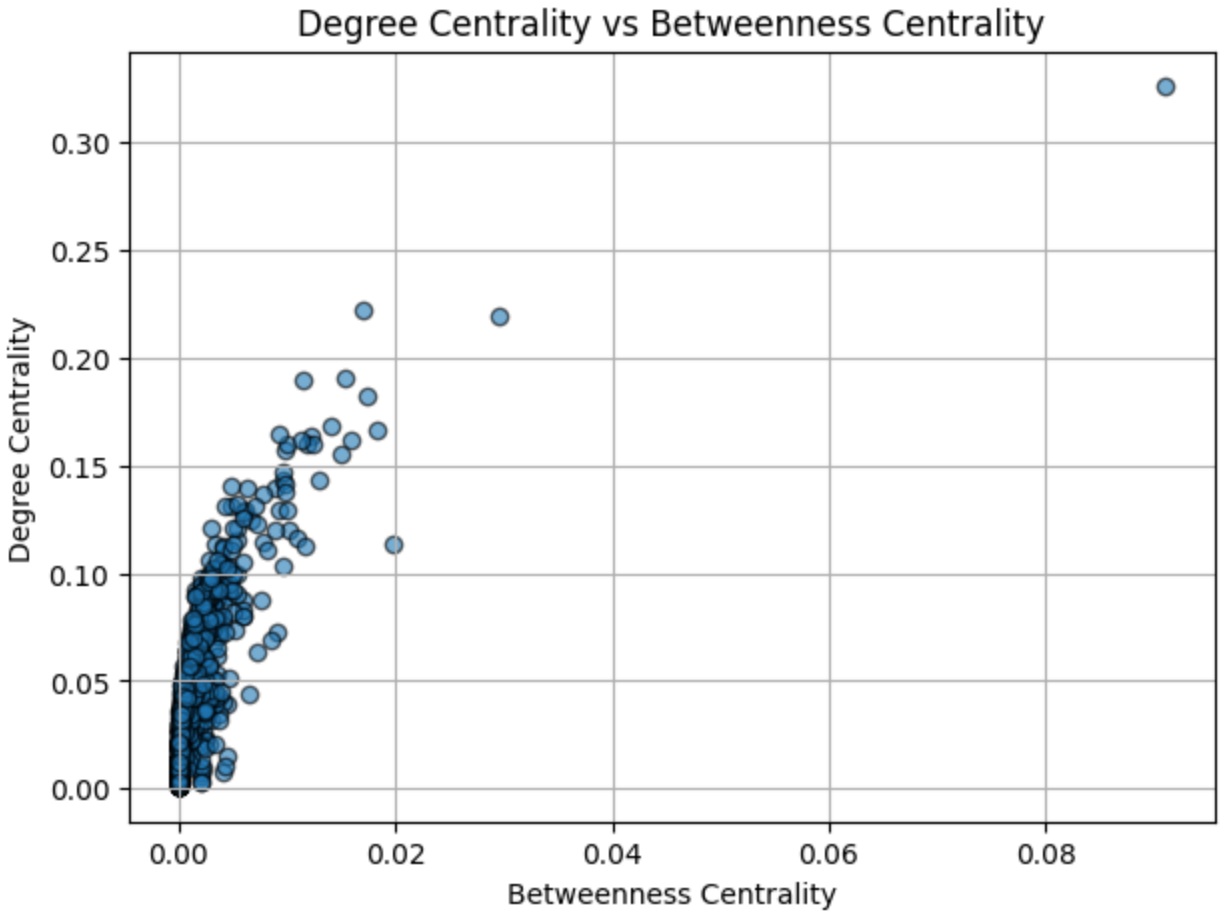
IMPORTANT NODES



- There's a very strong linear-looking correlation between degree and eigenvector centrality.
- Nodes with high degree also have high influence — meaning they are likely connected to other important nodes.



- There is a non-linear but strong correlation: as degree increases, closeness tends to increase too.
- Most nodes cluster at the lower range of both metrics, but a few stand out with high values.



- Most nodes have very low betweenness, even if their degree centrality is moderate.
- A few outliers have significantly higher betweenness – these are likely bridge nodes connecting communities.
- The correlation between degree and betweenness is weak.

TYPE OF NETWORK - DEGREE DISTRIBUTION & VISUALIZATION

Degree Distribution:

- Heavy-tailed pattern with a few nodes acting as hubs.

Visualization:

- Multiple layout algorithms (e.g., spring, Kamada-Kawai) reveal distinct clusters and hub formations.

Key Insight:

- These patterns are characteristic of a scale-free network, not of a random (Erdős–Rényi) model.

TYPE OF NETWORK - CLUSTERING COEFFICIENT & DISTANCES

Clustering Coefficient:

- Significantly high, indicating strong local cohesion and tightly knit groups.

Average Shortest Path (Distances):

- Low average path lengths demonstrate efficient, small-world connectivity.

Key Insight:

- The combination of high clustering and short distances further supports a scale-free topology over a random network model.

REAL NETWORK - OVERVIEW

- Steam Platform as Data Source
- Nodes = Steam Users
- Edges = Friendship Links
- Aim for N > 100 (Minimum)

DATA COLLECTION METHOD

- Steam Web API / Public Profiles
- Gather User IDs, Friend Lists
- Check Privacy Settings
- Data Cleaning & Validation

RESEARCH QUESTIONS & MOTIVATION

- Which users form hubs or bridges?
- Are there distinct game-based communities?
- Do certain friendship patterns predict activity levels?
- Why study this?
  - Understand social dynamics
  - Identify influential gamers & communities