STEAM SOCIAL NETWORK ANALYSIS

Matei Sonia

Morar Cristina-Mirela

Turcu Ciprian

Murariu Tudor-Cristian

Steam Network Data Collection

- **Data Source -> Steam Web API endpoints (**`GetPlayerSummaries`, `GetFriendList`) as our sole source of user and friendship data.

- **Extraction Method -> Batched, multithreaded requests** with per-call caching and rate-limit delays to efficiently harvest profiles and friend lists.

- **Seed Selection Criteria ->** Fetch minimal profiles for initial IDs, sample x number of friends each, then score candidates by friend count, internal connectivity, country clustering, and recent activity.

- **Network Expansion Flow -> Perform a density-aware BFS:** in each batch, score friends with the same density function, enforce depth-based limits and a score threshold to add new nodes.

Steam Network Collected Data Overview

Basic Statistics

- Nodes: 1,224 users
- Edges: 2,081 friendships
- Type: Undirected, Unweighted
- Density: 0.002780 (0.28%)

Key Characteristics

- Gaming Social Network: Steam friendships
- Low Global Density: Sparse overall structure
- High Local Clustering: Dense community pockets
- Geographic Clustering: Regional friend groups

Significance

- Real Gaming Communities: Authentic social bonds
- Information Flow: Efficient within clusters
- Research Value: Game recommendations, social influence

VISUALIZATION

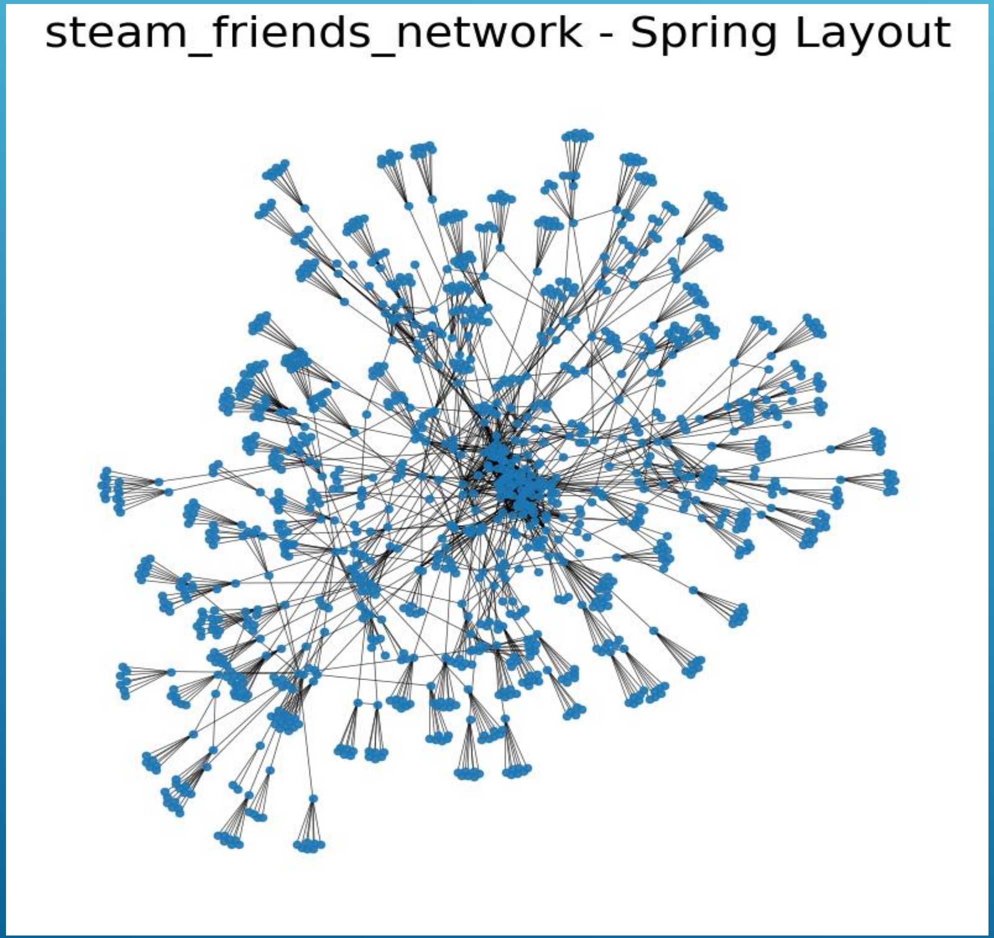
We'll go over 6 methods to visualize our dataset:

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

After that well use gephi to visialize the network by centrality, to get a better understanding of the relationships between users.

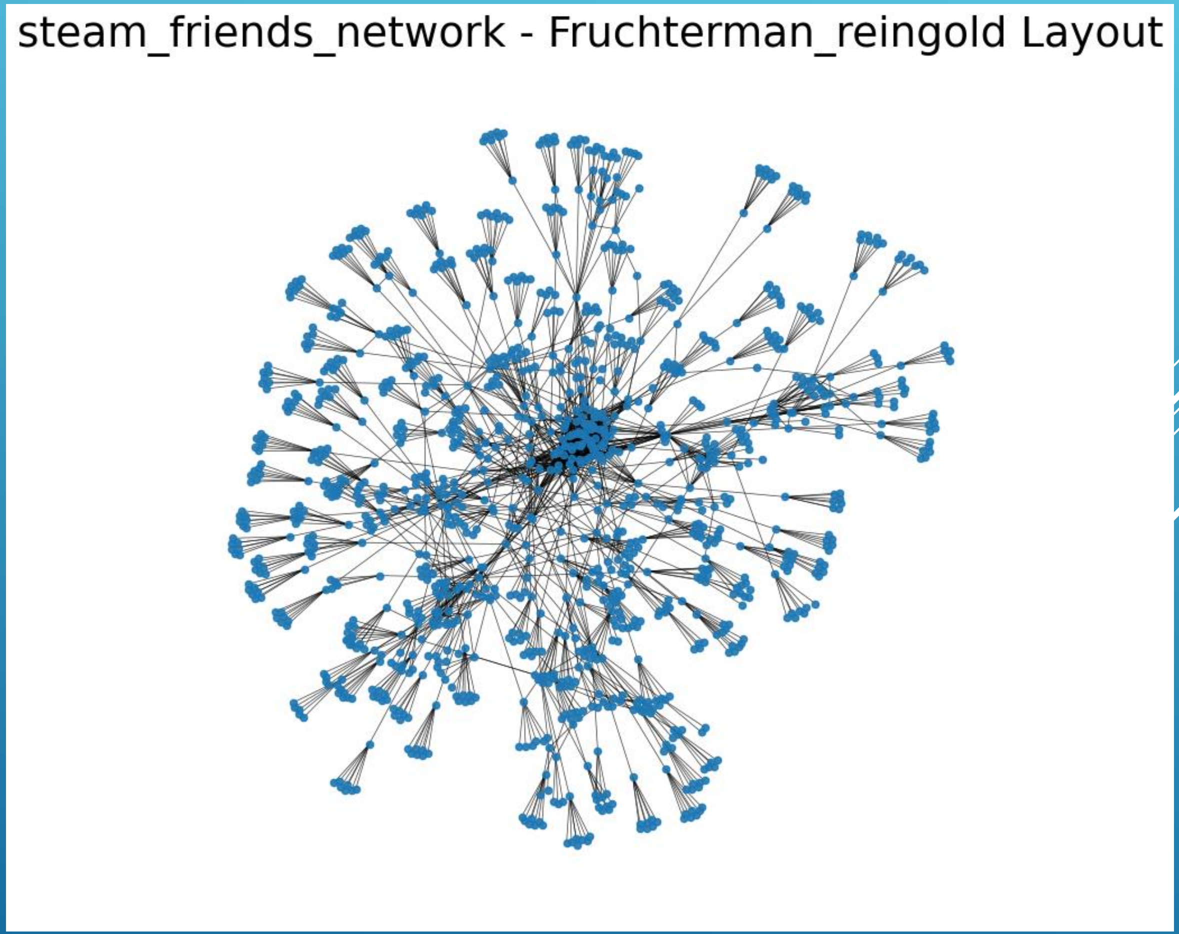
SPRING LAYOUT

The nodes behave like balls connected by elastic springs.



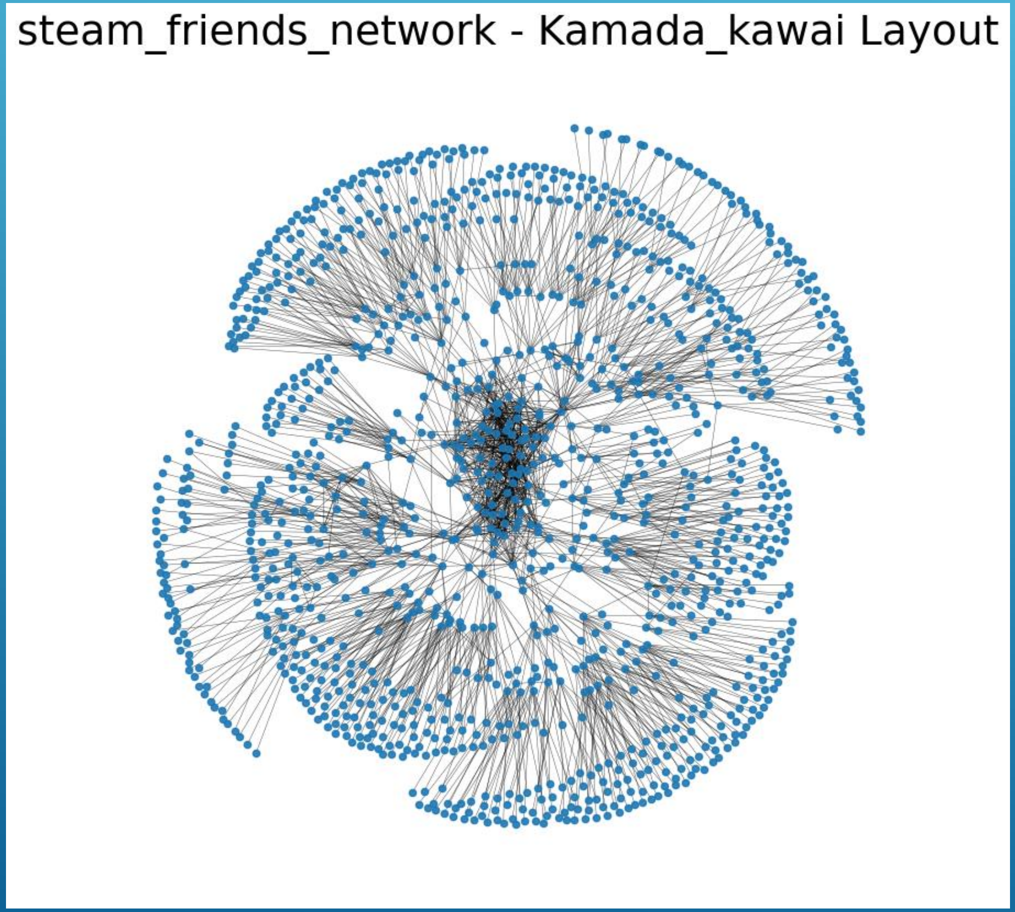
FRUCHTERMAN-REINGOLD LAYOUT

An optimized version of the spring model.



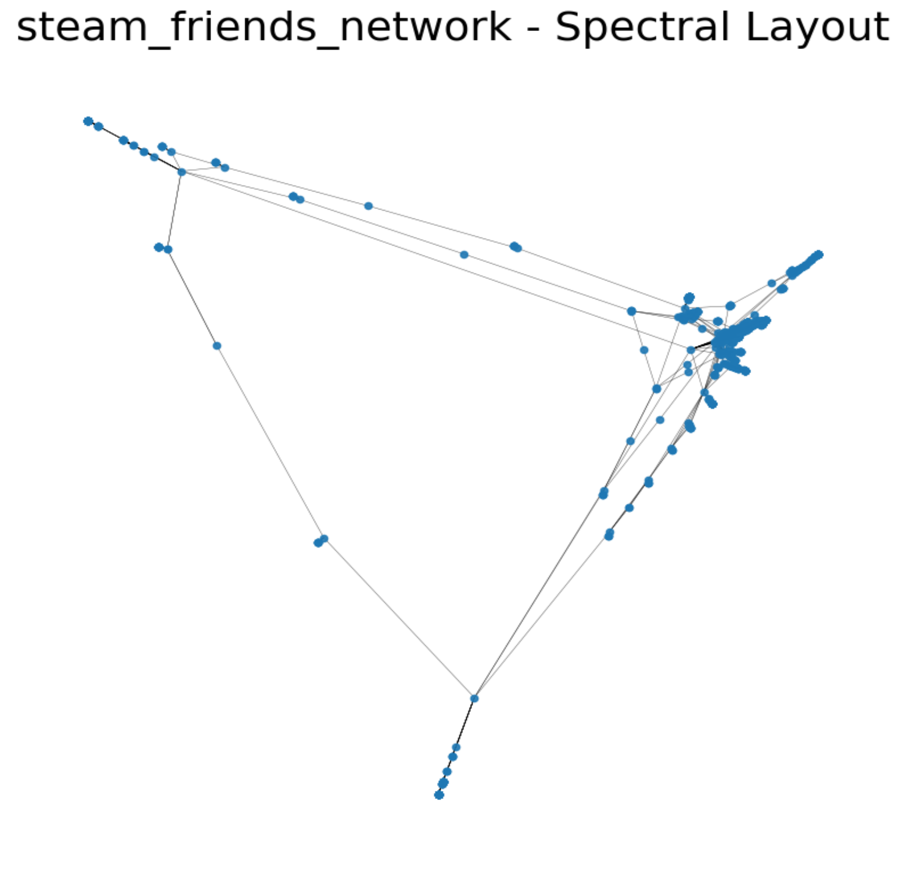
🔗 KAMADA-KAWAI LAYOUT

A clear image, where visual distances reflect network distances.



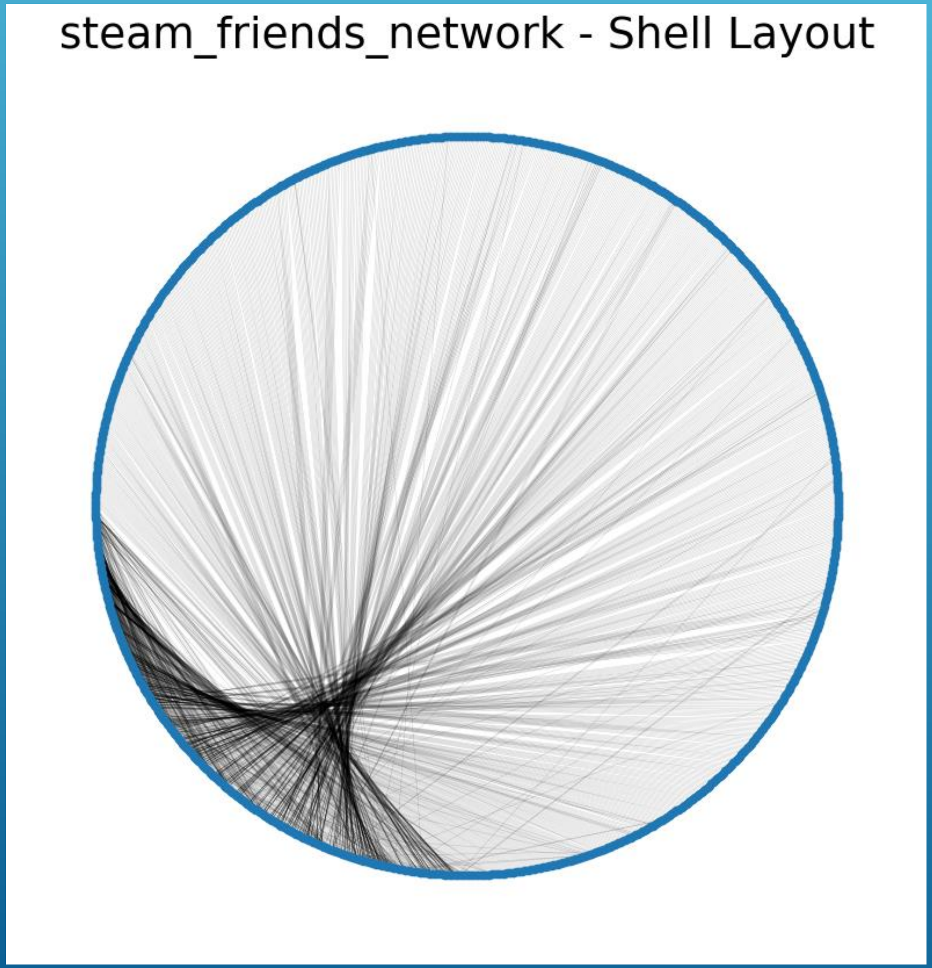
🧠 SPECTRAL LAYOUT

It uses mathematical methods to arrange the nodes.



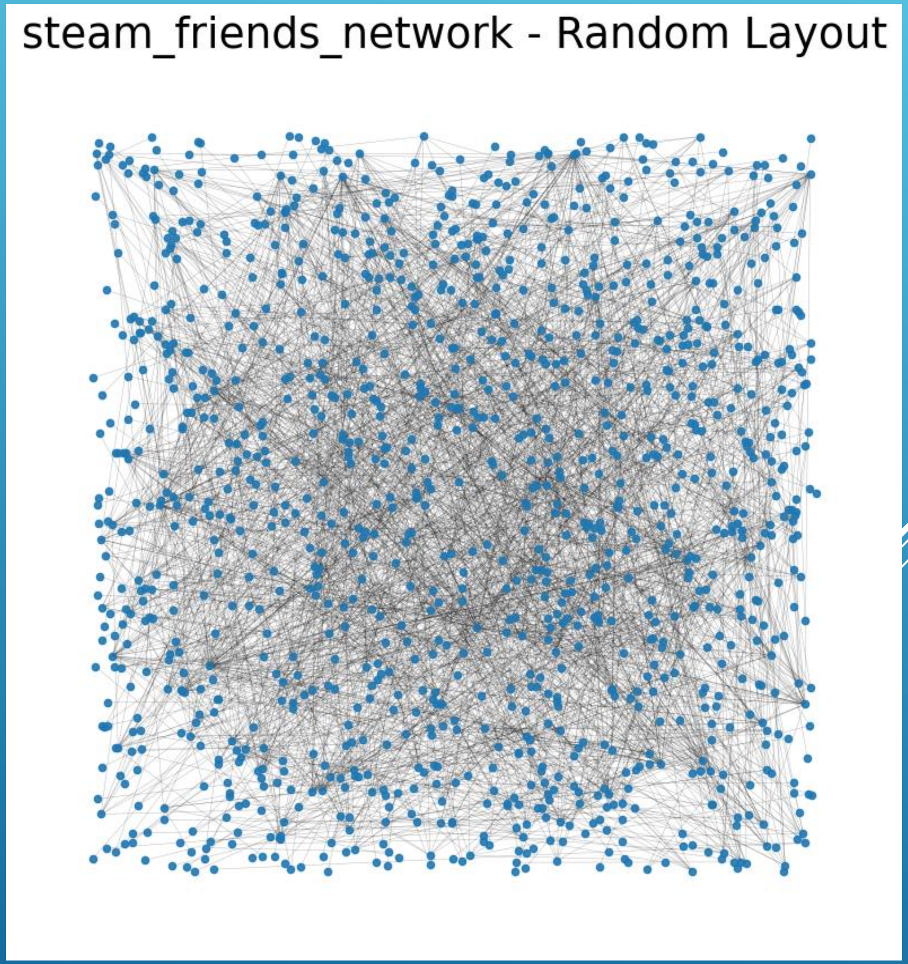
🐚 SHELL(CIRCLE) LAYOUT

The nodes are placed in concentric circles, like layers.

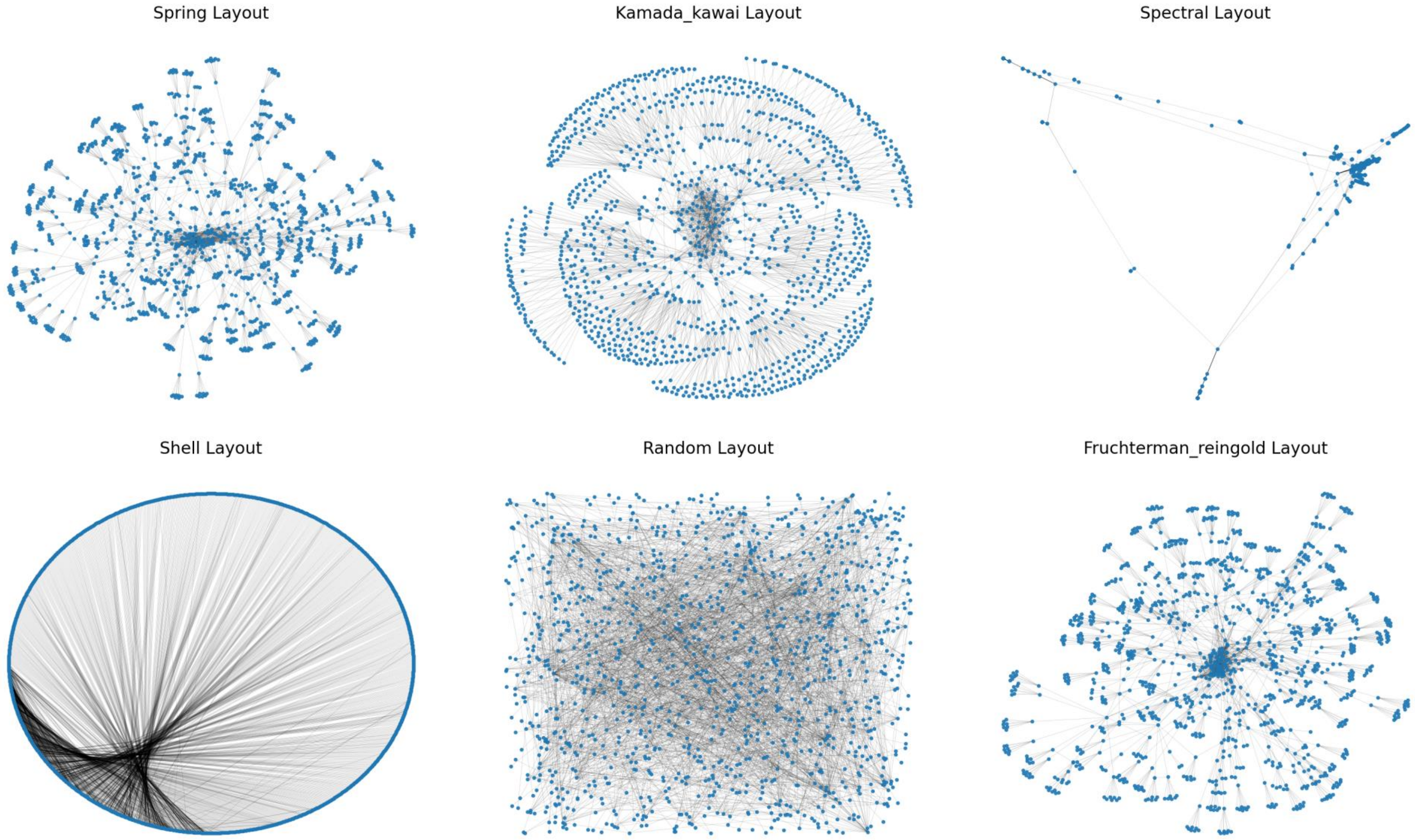


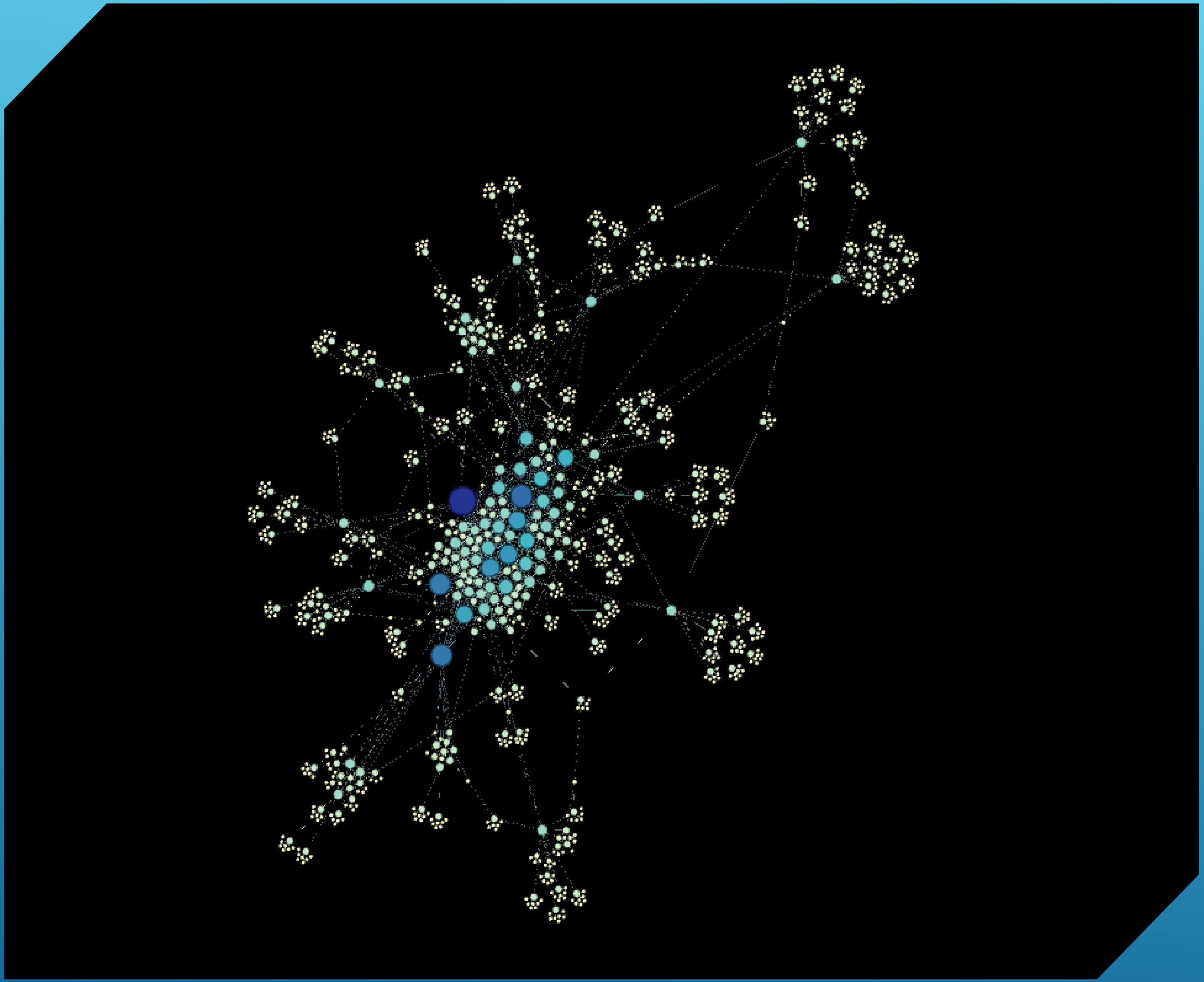
🎲 RANDOM LAYOUT

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



steam_friends_network - All Layouts



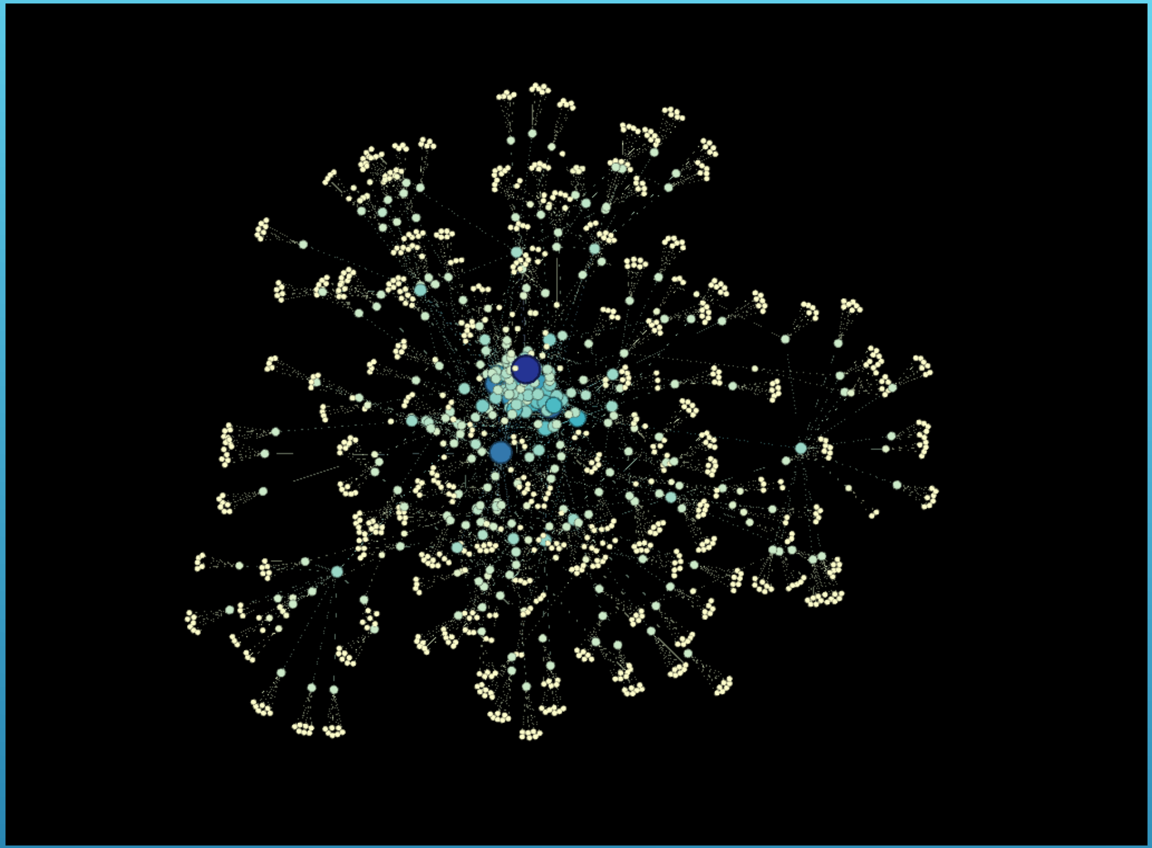


Force Atlas 2

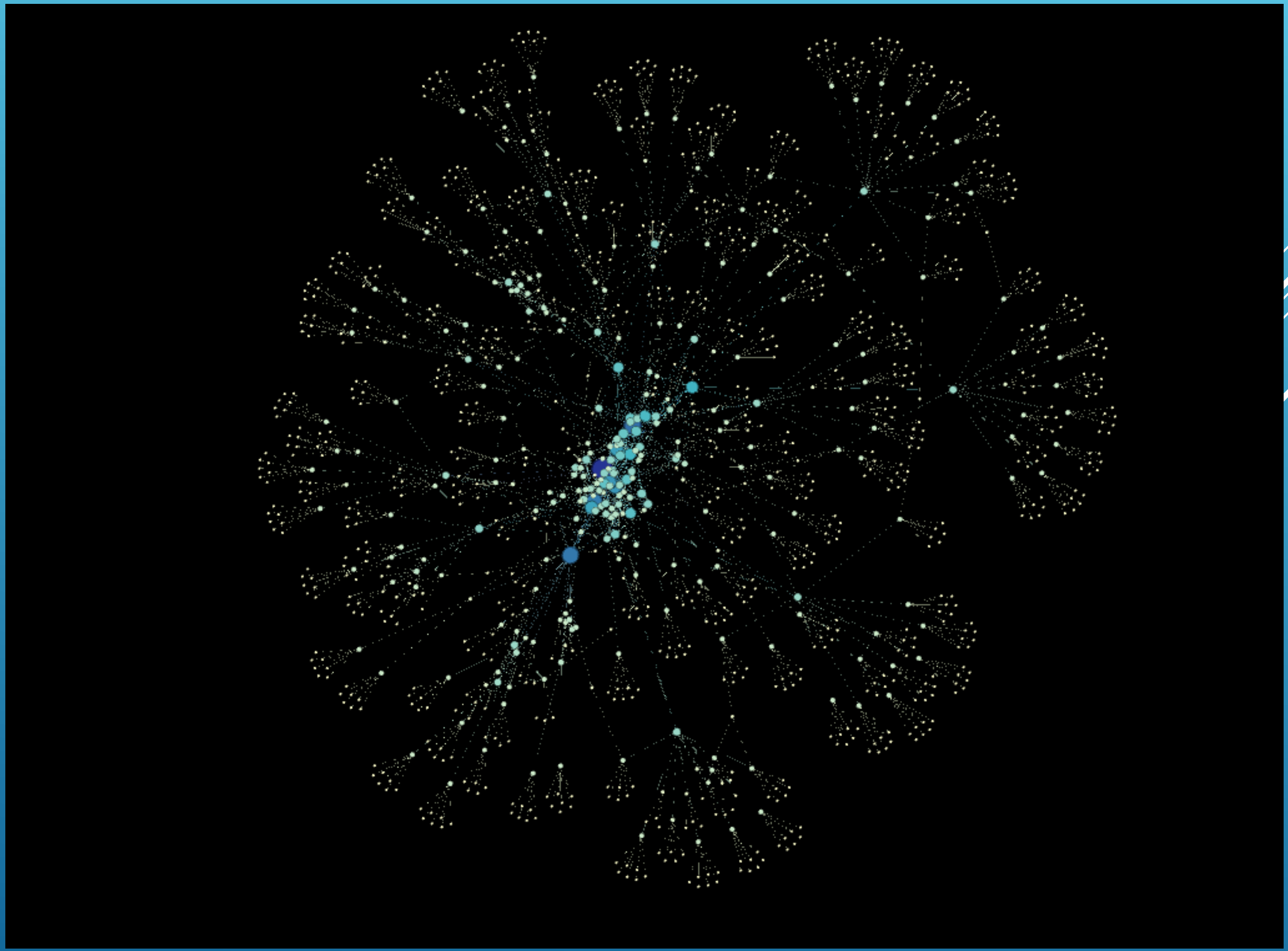
- simulates a physical system where:

- **Nodes repel each other** (like charged particles)
- **Edges pull connected nodes together** (like springs).

Size and Color: Centrality

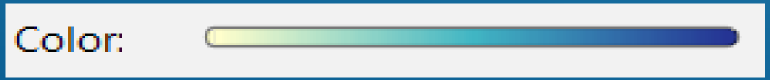


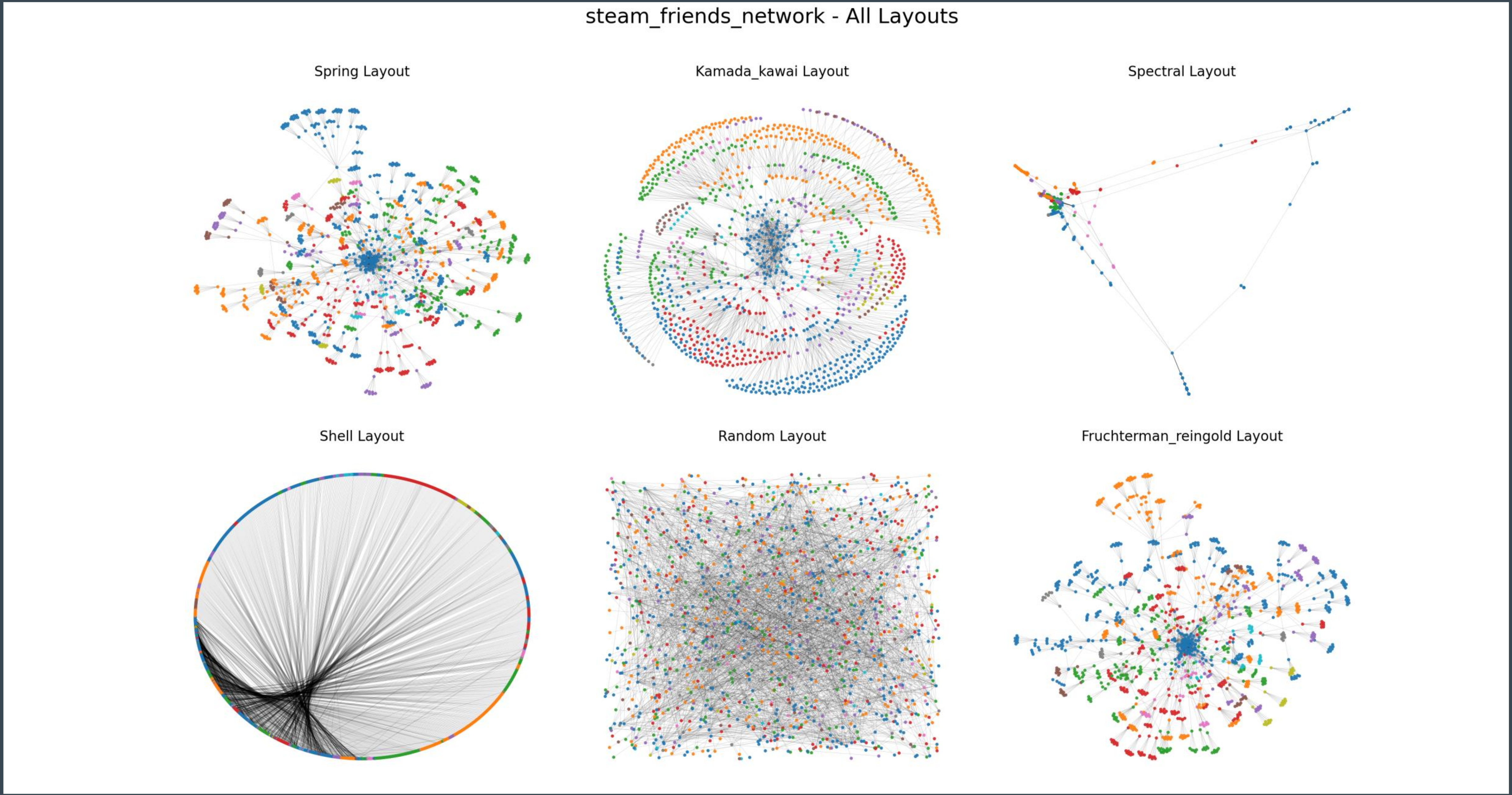
FRUCHTERMAN-REINGOLD

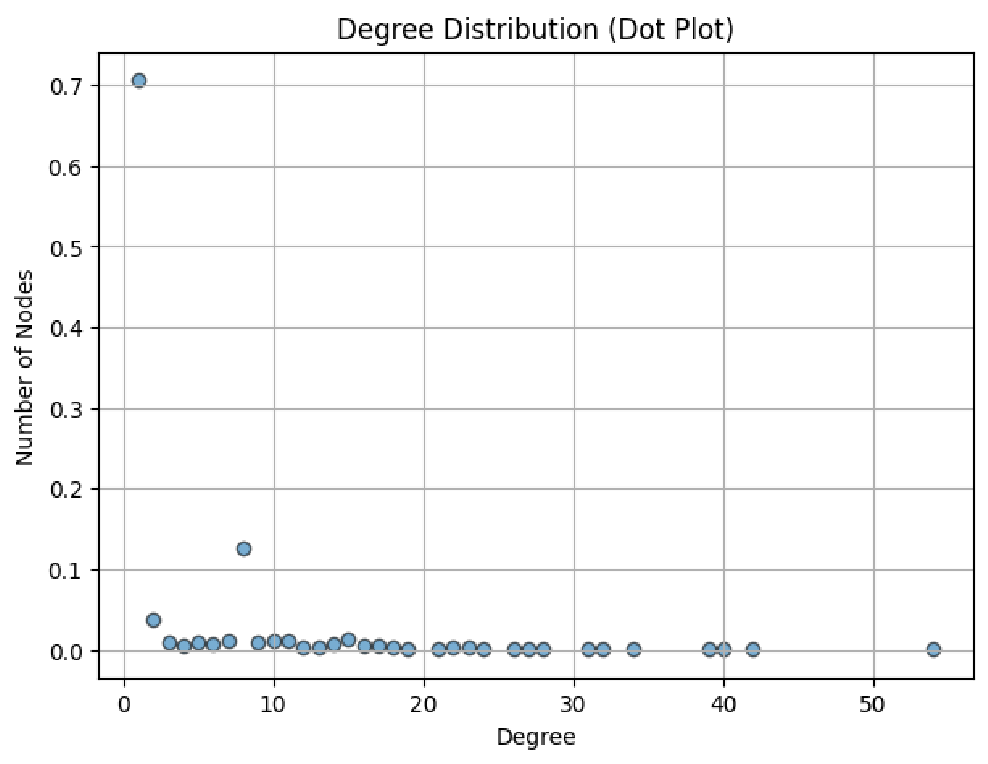


YIFAN HU

Size and Color: Centrality

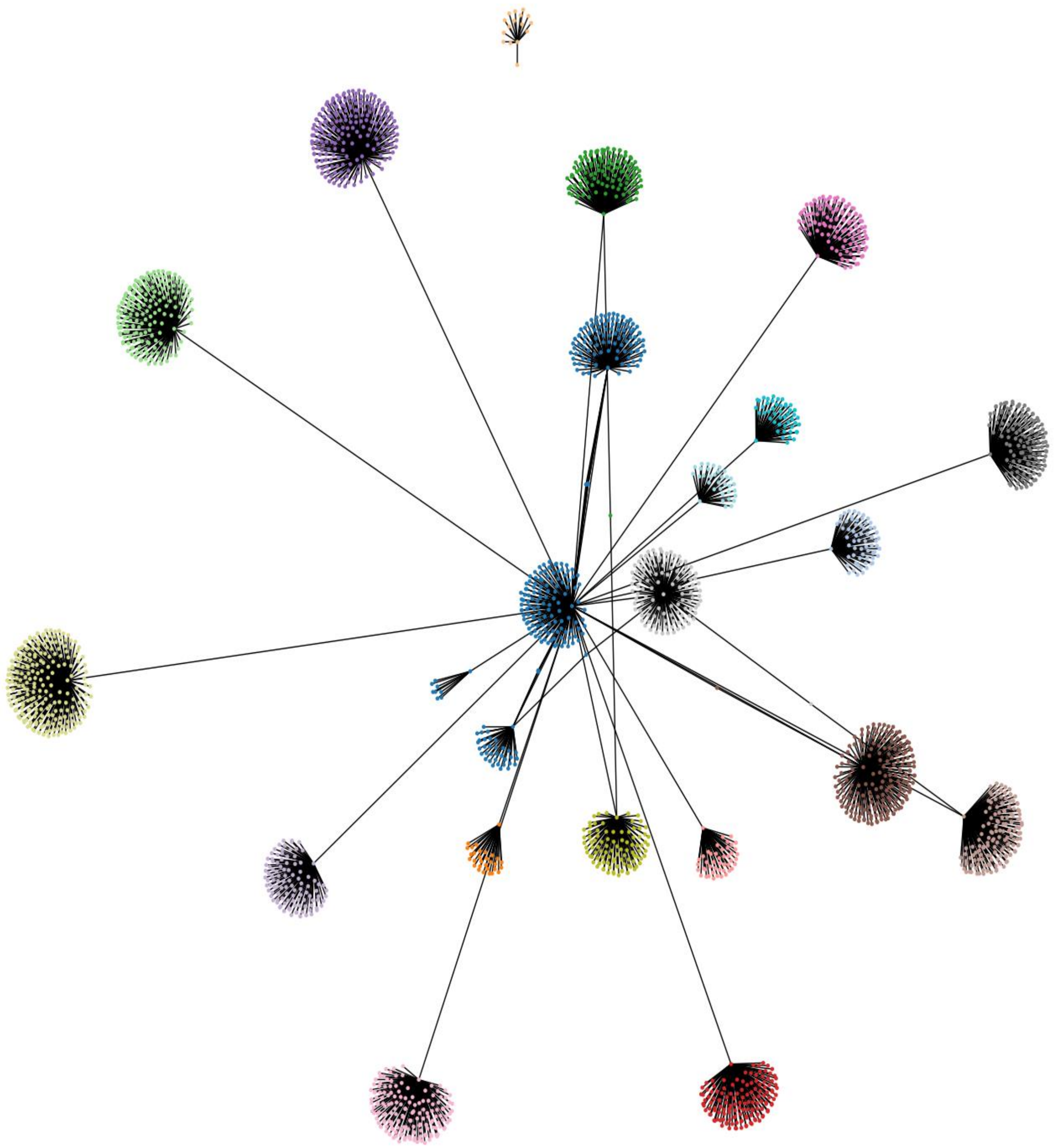


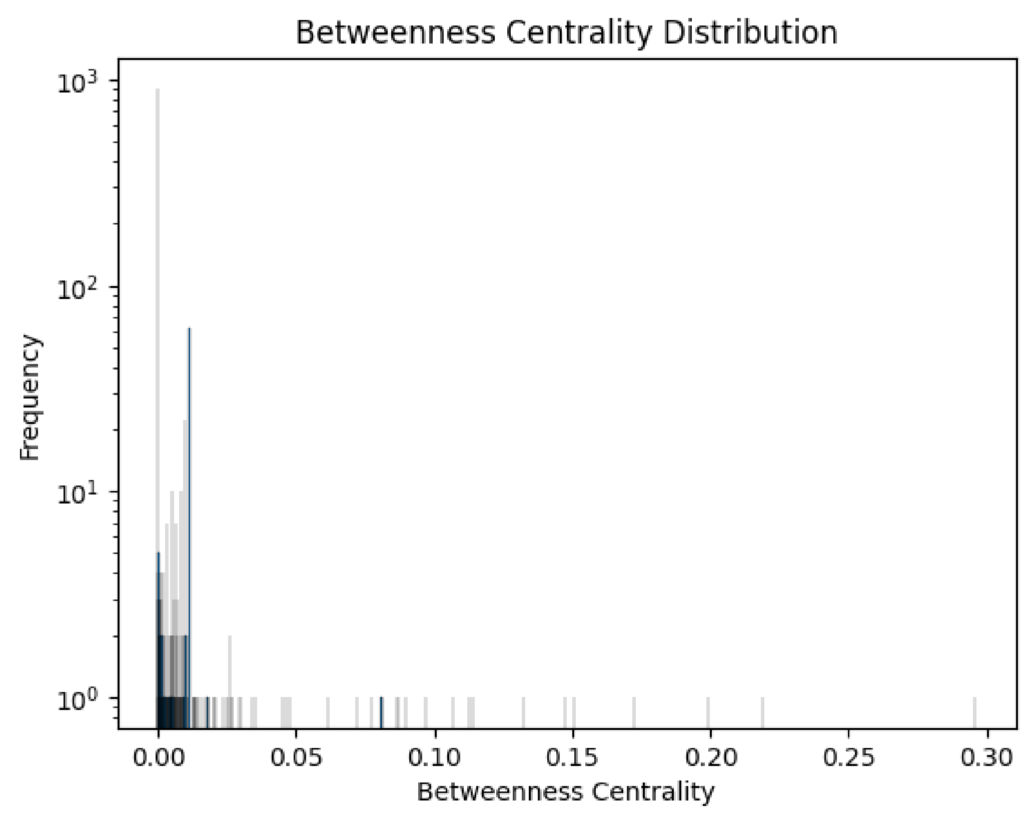

steam_friends_network - All Layouts
Spring Layout
Kamada_kawai Layout
Spectral Layout
Shell Layout
Random Layout
Fruchterman_reingold Layout



- Most nodes have low degree
- Central node with the highest degree

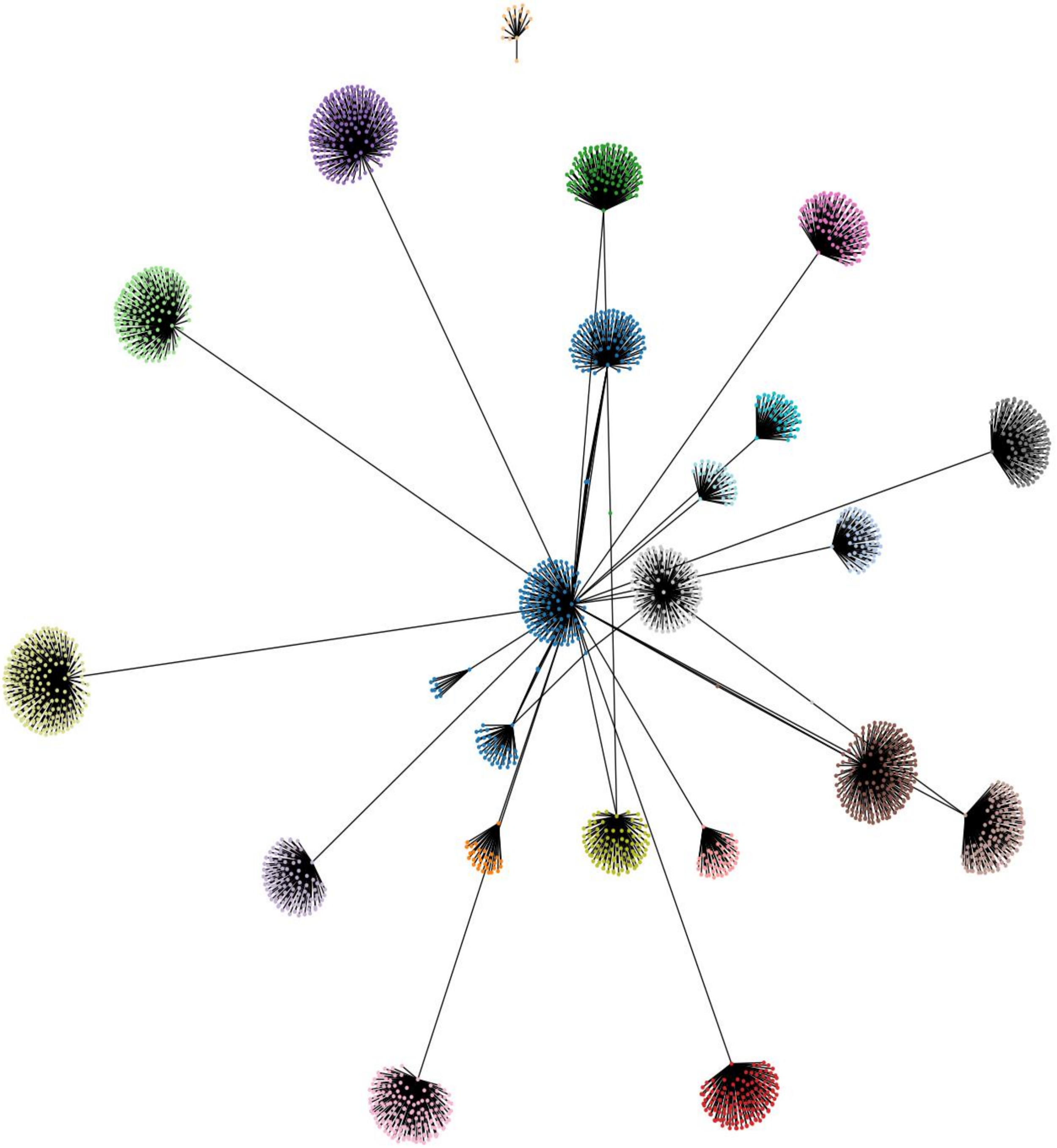
DEGREE DISTRIBUTION

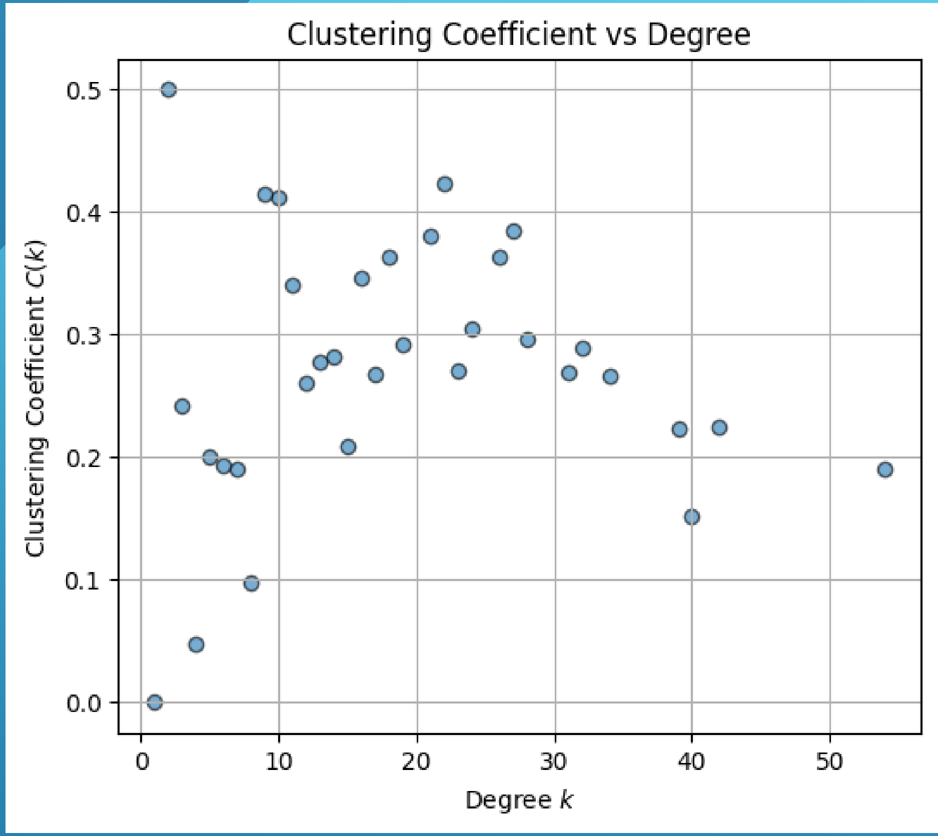


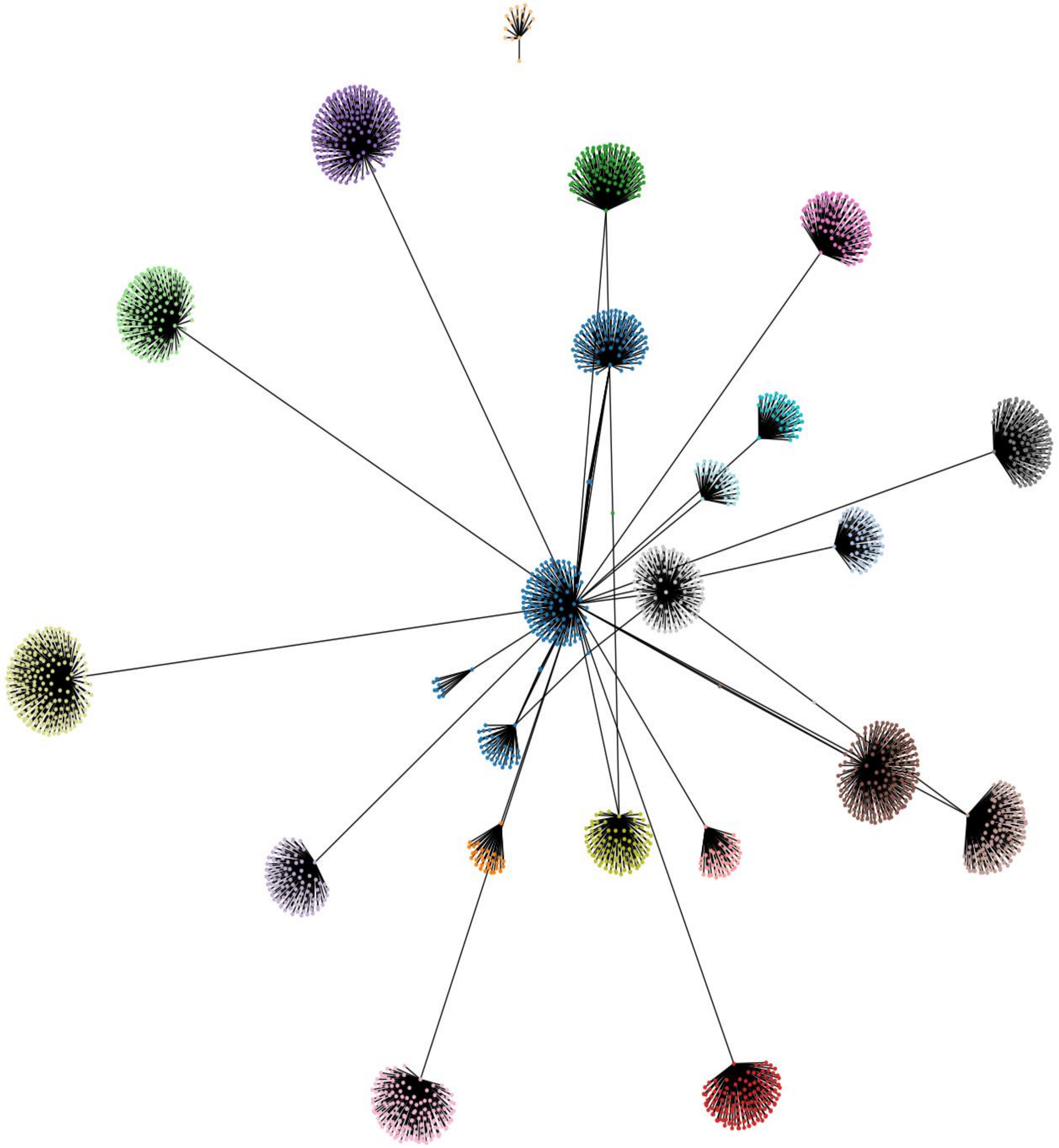


- Few nodes connect other nodes, most don't
- The rest connect the network

BETWEENESS DISTRIBUTION

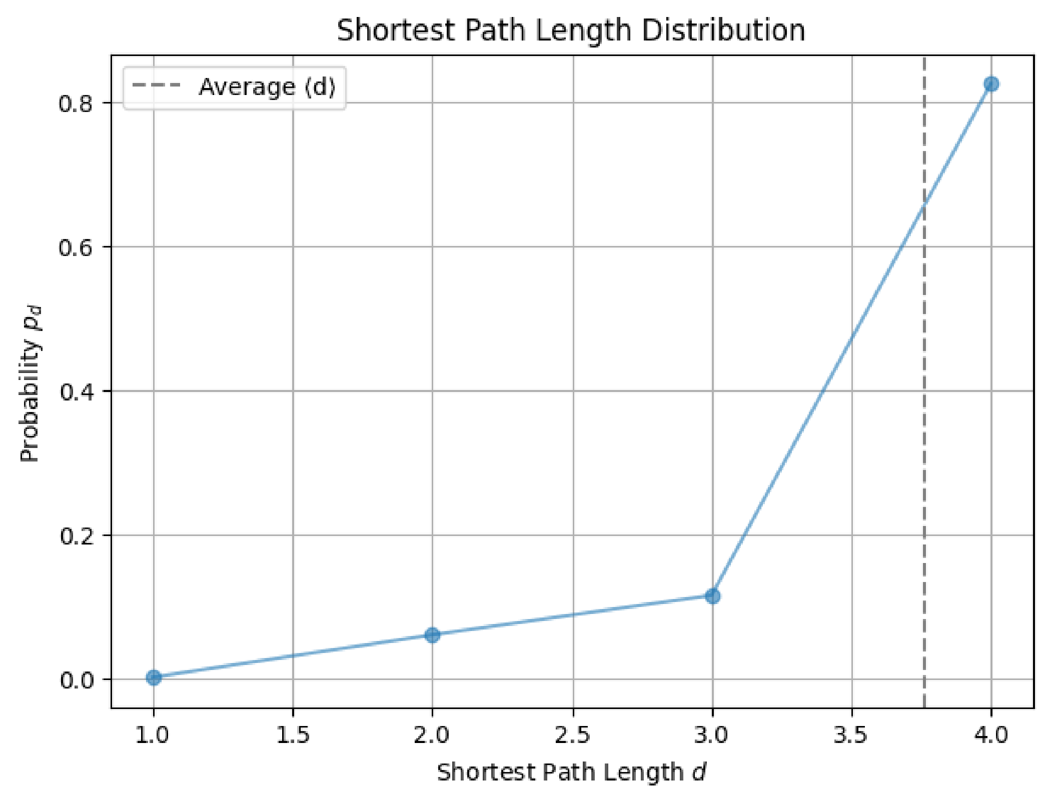


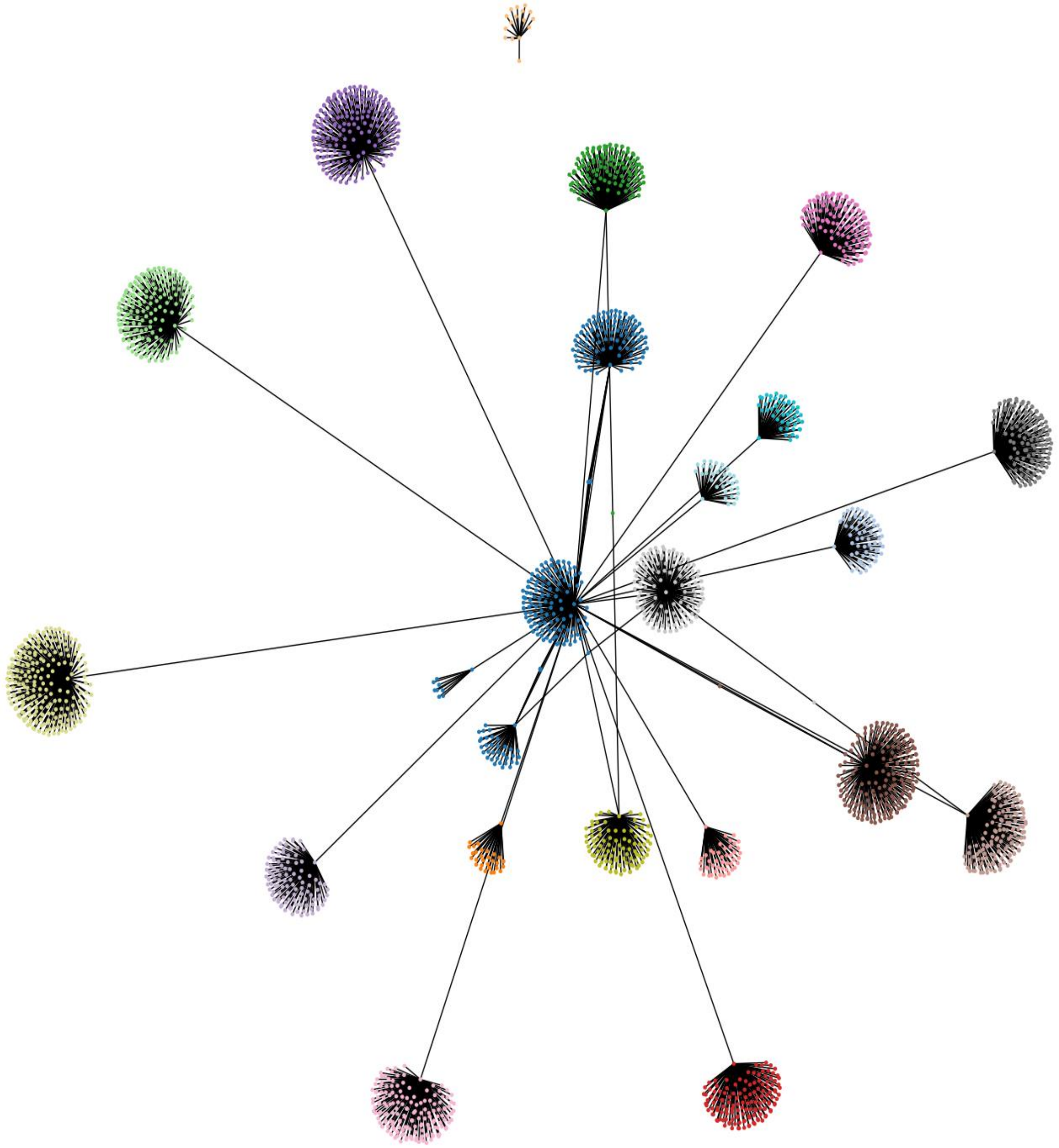




- Leaves don't cluster
- Low degree nodes form clusters
- High and medium degree modes keep the network together (interconnected between each other only)

CLUSTERING COEFFICIENT





- "Small-world" graph, everyone is connected with just a few hubs

AVERAGE LENGTH PATH

COMMUNITIES

Hierarchical agglomerative clustering

This is a **star** graph -> many leaves from one node.

Each colored blob is a tightly-knit group of players who **mostly interact among themselves.**

One or two nodes with many spikes -> those are your **hub players** (popular group leaders who friend tons of people).

Their friends themselves **aren’t generally friends with one another**.

Not many nodes with edges into two different color-blobs. That means people tend to friend one “hub” rather than multiple hubs.

The **most central community** could be linked to the person who manages the CS community.

Dense, clique-like communities
- Each colored blob is a tightly interconnected set of players with many internal “friend-of-friend” triangles.

Community hubs over spokes
- High-degree nodes act as facilitators within each blob, their neighbors also linked to one another.

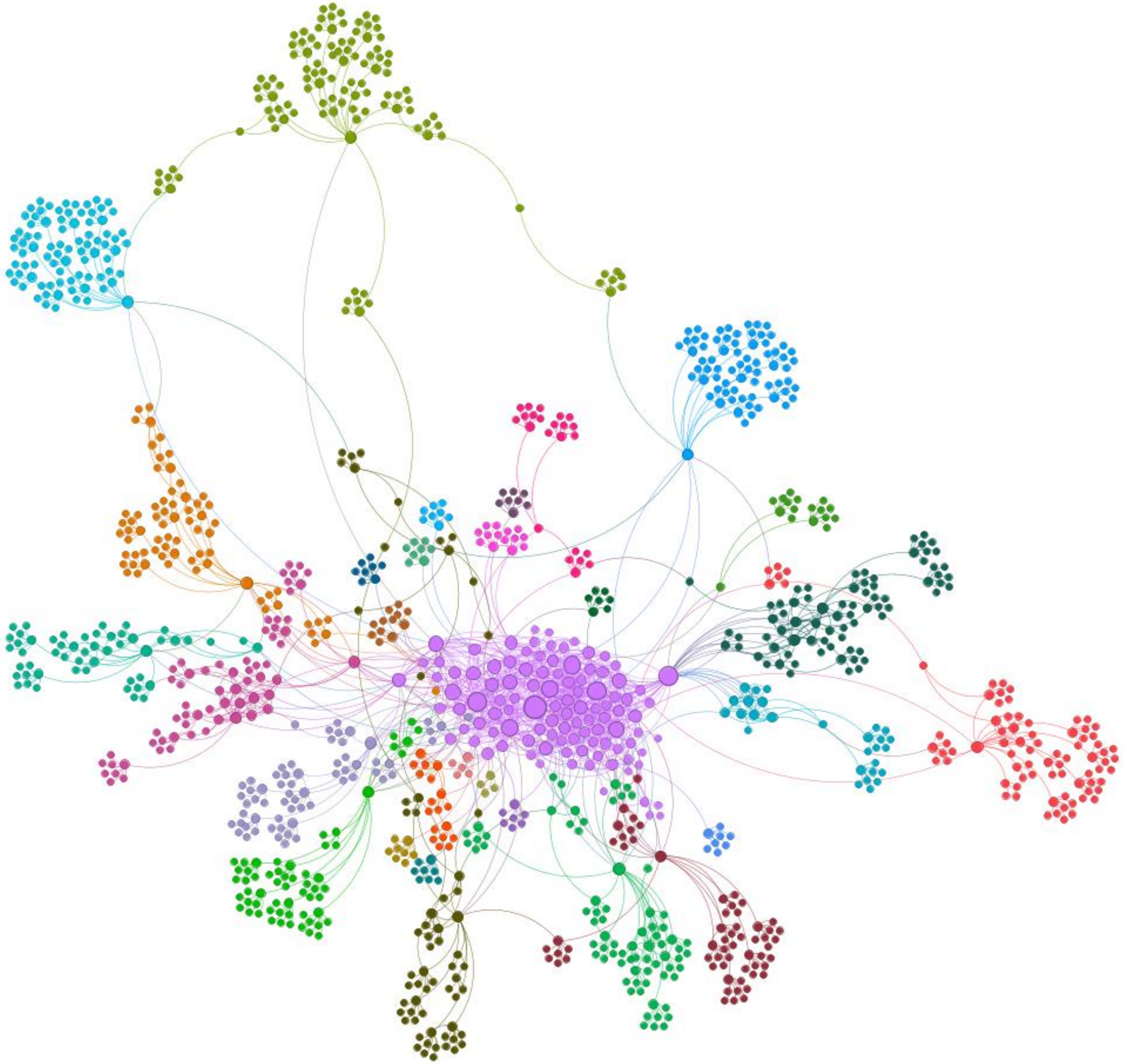
Isolated interaction islands
- Scant edges between blobs show that modularity optimization yields largely self-contained groups.

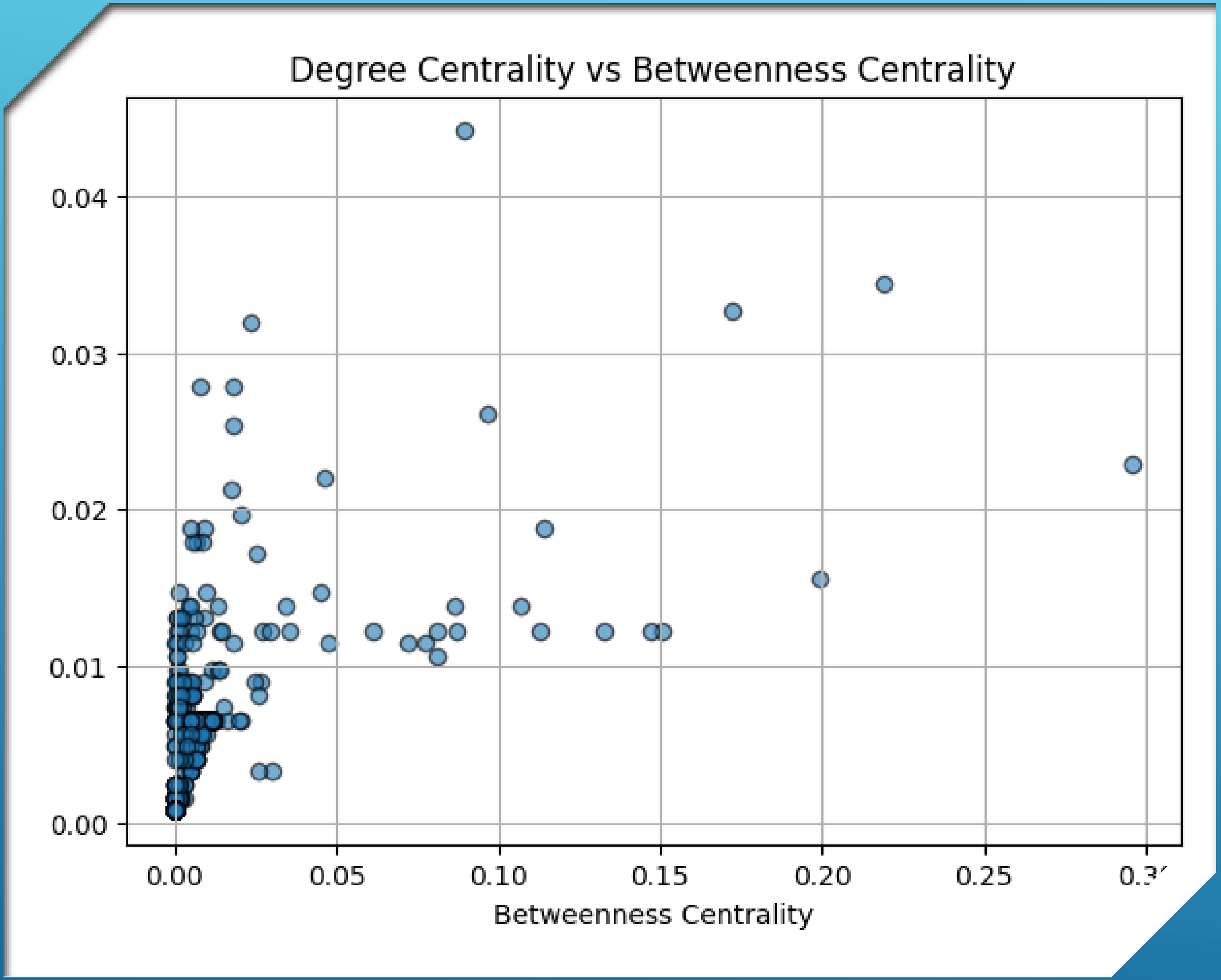
Varied group sizes
- One large community (~115 nodes), several mid-sized clusters (60–90), and many small ones (< 10).

Giant component + periphery
- Nearly all nodes join the main component, yet tiny, tightly bound micro-communities still sit on the outskirts.

COMMUNITIES

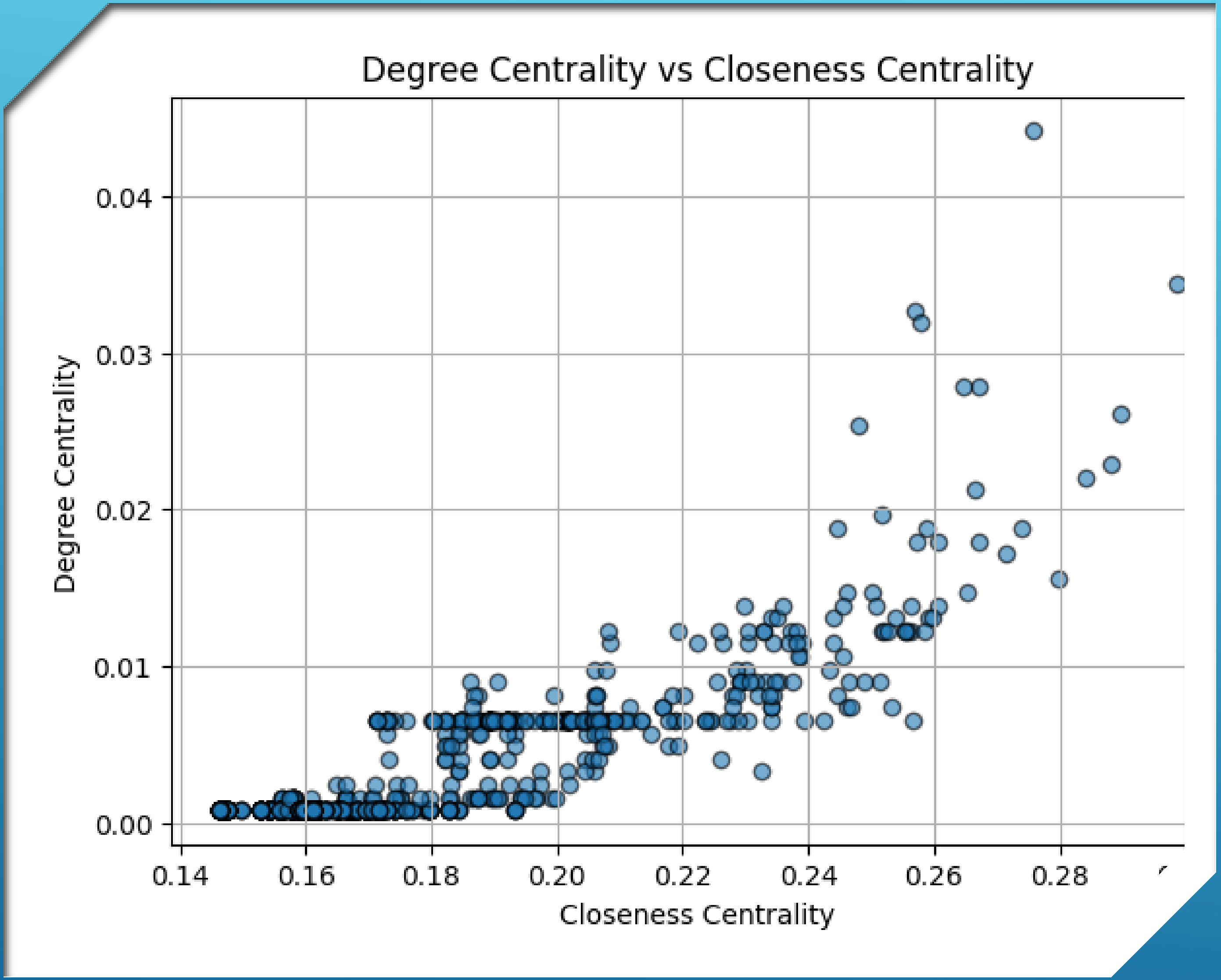
Gephi Louvain approach



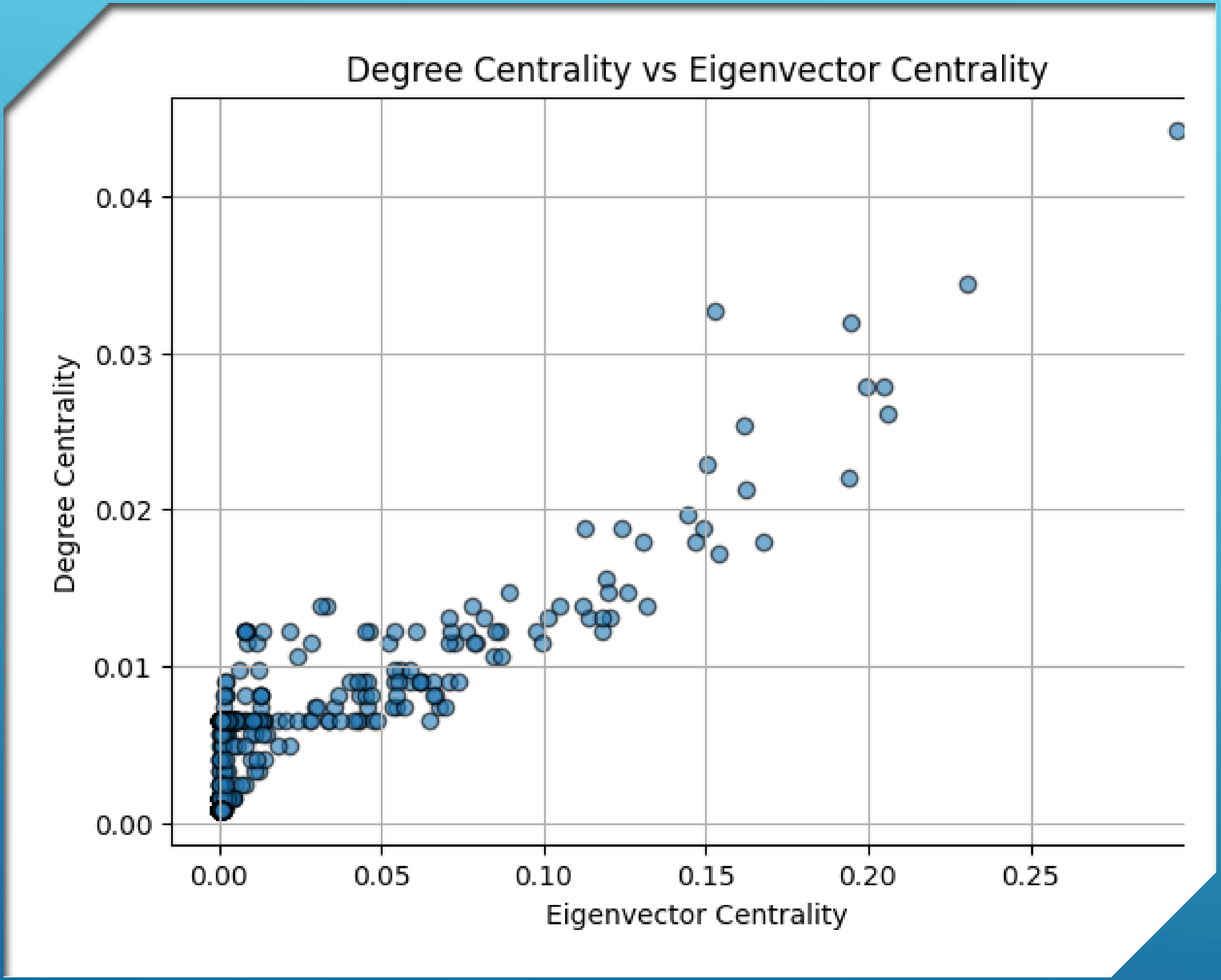


IMPORTANT NODES

- Most nodes exhibit low values for both metrics, forming a dense cluster near the origin.
- A few nodes stand out with relatively high betweenness and moderate-to-high degree centrality.
- The outliers with high betweenness centrality likely act as critical intermediaries or bottlenecks, even if they're not the most connected.



- Most nodes have low to moderate values on both centrality measures.
- There is a visible positive correlation: nodes with higher degree centrality tend to also have higher closeness centrality.
- A few outliers exhibit both high degree and high closeness, indicating well-connected nodes that are also close to others in the network.



- Most nodes cluster near the origin, with low values for both centrality measures.
- A clear positive correlation is observed: nodes with higher degree centrality often have higher eigenvector centrality.
- A few prominent outliers dominate both measures, suggesting their influence and connectivity to other influential nodes.

REAL NETWORK VS. ERDŐS–RÉNYI RANDOM MODEL

Degree Distribution:

- Real: Heavy-tailed, hubs present
- Random: Poisson-like, uniform connectivity

Clustering Coefficient:

- Real: High, strong local groups
- Random: Very low, no local cohesion

Shortest Path Length:

- Both are small, but real network shows structured connectivity

Conclusion:

- The random model fails to explain the clustering or degree variability
- **Real network is not random**

REAL NETWORK VS. WATTS–STROGATZ SMALL-WORLD MODEL

Degree Distribution:

- Real: Heavy-tailed
- WS: Uniform, lacks hubs

Clustering Coefficient:

- Both high — WS preserves local cliques

Shortest Path Length:

- Both short — WS uses shortcuts, real has hubs

Conclusion:

- Small-world model captures clustering, but not degree variability
- **Partially similar, but lacks scale-free behavior**

REAL NETWORK VS. BARABÁSI–ALBERT SCALE-FREE MODEL

Degree Distribution:

- Real & BA: Heavy-tailed, hub nodes dominate

Clustering Coefficient:

- Real: High
- BA: Moderate (lower than real network)

Shortest Path Length:

- Both exhibit short path lengths (efficient communication)

Conclusion:

- BA model explains hub structure and connectivity efficiency
- **Closest structural match to real network**

THANK YOU!