

Twitter Datasets Analysis

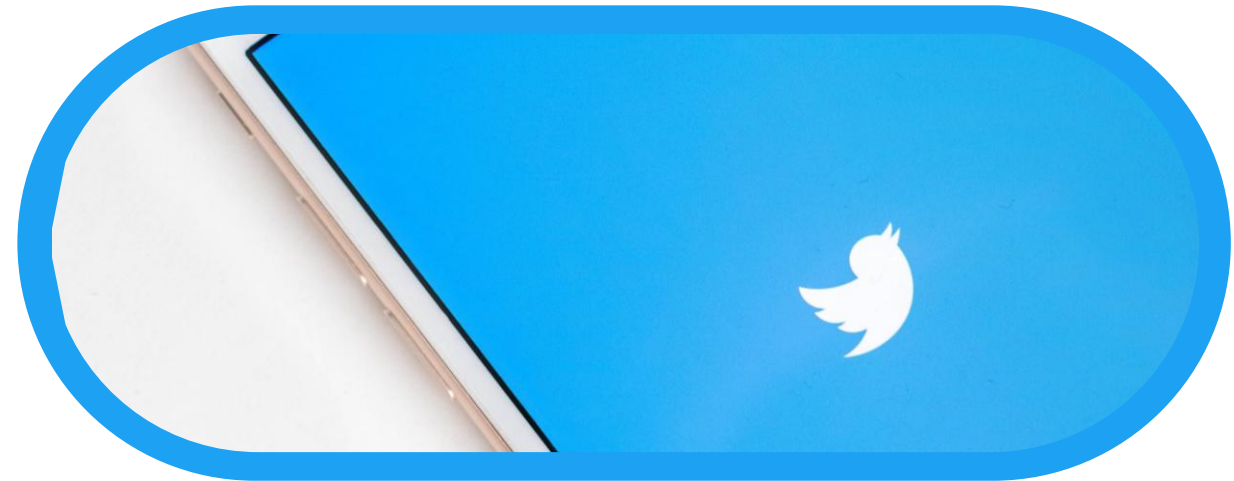


Social Network Computing

Identification

Name : Maryam Waheed Zamel

ID : 2205154





Goal :

1. **Analyzing two real twitter communities from WICO graph dataset. One from 5G-Conspiracy-Graph (misinformation clusters) and the other from Non-conspiracy-graphs (normal clusters) using methods used in step2**
2. **Performing full social network analysis for the two graphs**

WICO Dataset

The **WICO (Who Interacts with Whom) Graph Dataset** is a collection of twitter conversation graphs, specifically designed to study the structural differences between networks propagating misinformation and those engaging in ordinary discourse.



5G conspiracy Graph Analysis

Now I will investigate two subsets of the WICO dataset which the first one is 5G-conspiracy-Graph.

- ❖ The graphs represent malicious or anomalous social networks. Specifically, these graphs are reconstructed from twitter interactions surrounding debunked 5G conspiracy theories.
- ❖ The graphs are built from conversations that took place during the peak of the baseless theory that 5G technology was linked to the spread of the COVID-19 virus. This included claims that 5G waves suppressed the immune system or even that the virus was transmitted via the radio waves themselves.
- ❖ The edges in these graphs represent various forms of engagement on twitter such as:
 - **Retweets** and **Quote Tweets**
 - **Replies** and **Mentions**



Non-conspiracy-Graph Analysis

the second graph is Non-conspiracy-Graph.

- ❖ The graphs are representing normal twitter communities.
- ❖ These graphs are constructed from twitter interactions around neutral and factual topics. The exact topics can vary but are chosen specifically because they lack the emotional, and conspiratorial elements of the 5G narrative. Examples could include discussions about a new technology product launch, a popular TV show finale, or a local cultural event.
- ❖ The edges similarly represent **Retweets**, **Replies**, and **Mentions**, but the content being discussed is not misinformation. The interactions are more likely to reflect a diverse range of opinions, casual conversation, information sharing, and community building without a central, adversarial narrative.



Statistics applied

1. Avg degree
2. Graph density
3. Network diameter
4. Modularity
5. Connected Components
6. Avg Clustering Coefficient



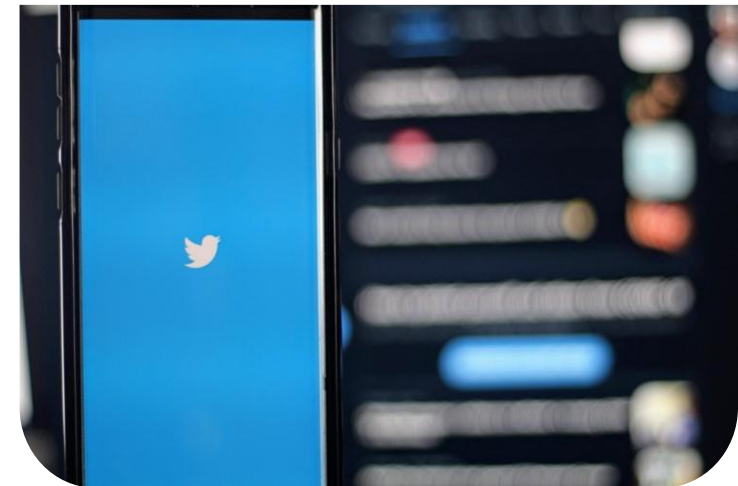
The chosen layout applied to the graph is :

- **ForceAtlas2** which is a state where the repelling and attracting forces are mostly balanced. The result is a visualization where nodes that are more interconnected are positioned closer together, forming visible clusters, while less connected nodes are pushed to the periphery.

Why used in 5G-conspiracy graph analysis :

- **It Perfectly discovers Community Structure (Echo Chambers):**
Misinformation networks often act as echo chambers or clusters where users primarily interact with others who share the same beliefs, forming communities. So, The edge attraction force will strongly pull these highly interconnected users together. The repulsion force will then push these different clusters apart from each other.
- **It Highlights Central Influencers (Super-spreaders):**
a node with a very high number of connections (a **high degree**) experiences a strong cumulative pull from all its connections. Because it's connected to so many nodes, it gets pulled into the center of a cluster, or even becomes a central hub that holds an entire community together.

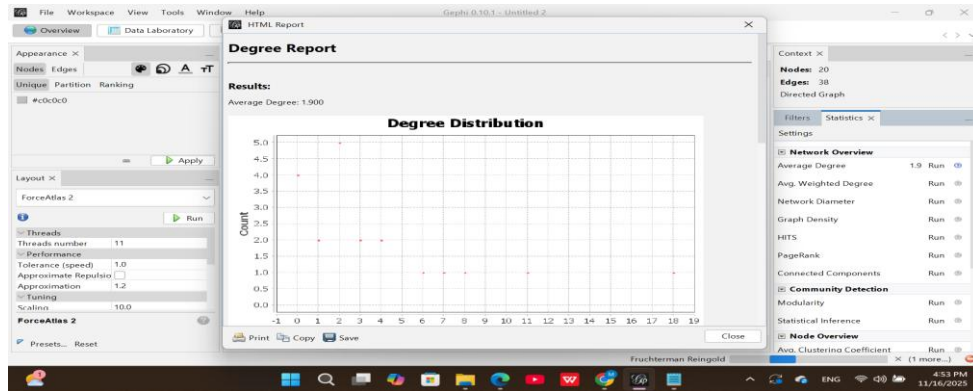
5G-Conspiracy Graph Analysis



5G-Conspiracy Graph Analysis

Beginning statistics applied to analyze the graph

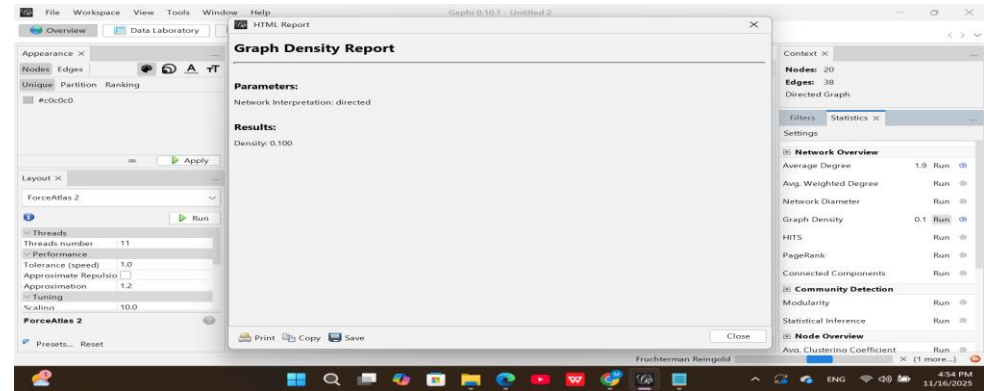
1. **Avg degree** In a directed graph, this counts both incoming and outgoing connections.



What it means: It measures the general level of connectedness. A higher average degree suggests a more densely interconnected network.

Result = 1.9 which means on average, each user in this conspiracy graph has about 2 connections. This is a low to moderate average degree. It suggests that while there is activity, the network is not extremely dense. Many users might be connected to just one or two central figures like influencers rather than everyone talking to everyone else. It hints at a broadcast-style structure rather than a tight-knit discussion group.

2. **Graph Density:** The ratio of the actual number of edges to the total possible number of edges. It ranges from 0 (no connections) to 1 (everyone is connected to everyone else).

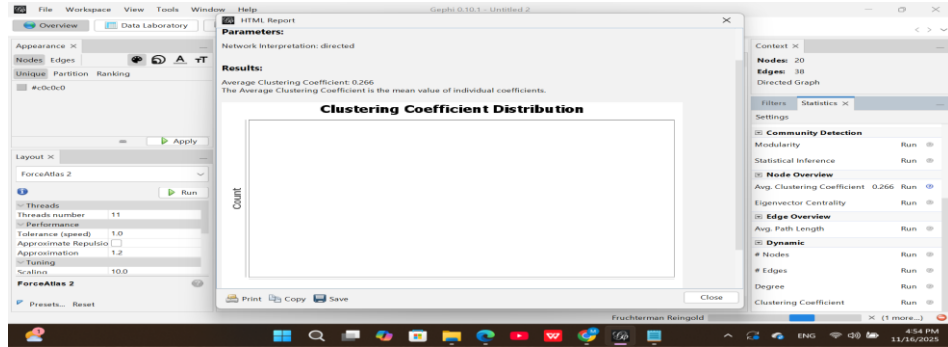


What it means: It measures how 'complete' the network is. High density indicates strong community. Low density indicates a sparse network.

Result = 0.1 so the density of 0.1 (or 10%) is very low. It means only 10% of all possible connections in this network actually exist. This is typical for large, real-world networks and supports the idea of an information-spreading network: information flows through specific pathways (from influencers to followers) rather than bubbling up from everyone talking to everyone.

5G-Conspiracy Graph Analysis

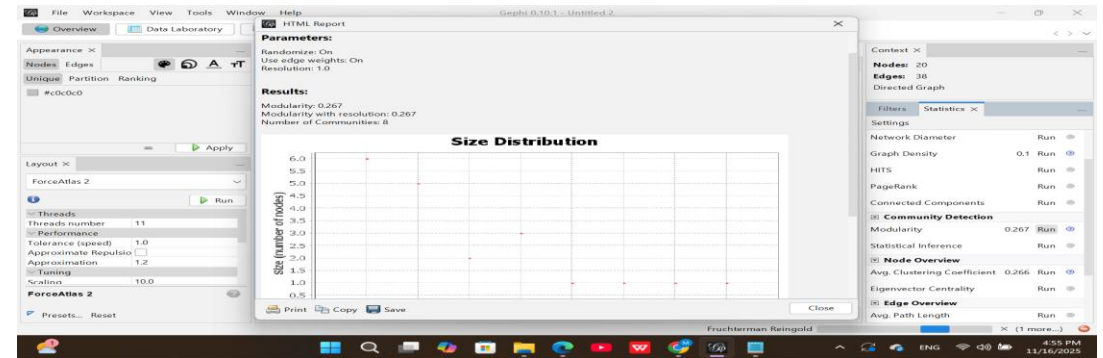
3. Average Clustering Coefficient: This measures the degree to which nodes in a network tend to cluster together



What it means: A high coefficient (close to 1) indicates strong behavior forming tight friend groups. A low coefficient (close to 0) indicates a more open, broadcast-oriented network.

Result = 0.266 It suggests that the "friends of my friends are not necessarily my friends." In the context of a conspiracy graph, this means the structure is more like a star or a tree: followers of a central account don't necessarily interact with each other. They are connected through the hub, not among themselves

4. Modularity and Number of Communities: measures the strength of division of a network into modules (communities or clusters)

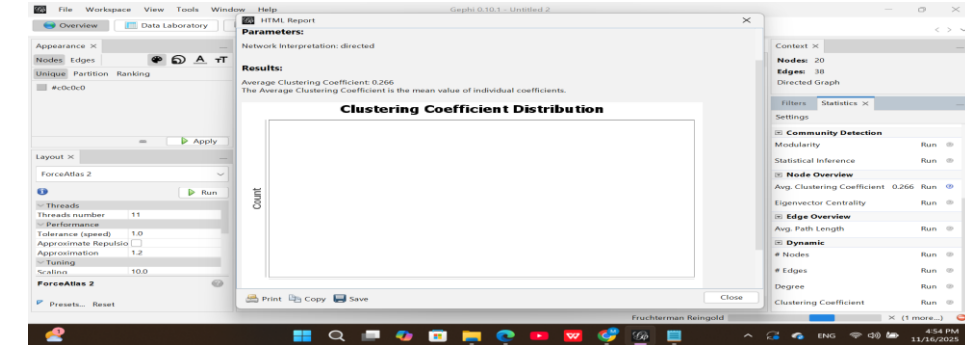


What it means: It identifies "echo chambers" or distinct subgroups within the larger network. **As high modularity** (closer to 1) means the network has strong, separate communities. The algorithm also tells you how many communities it found.

Result = 0.267, Number of Communities = 8
Modularity (0.267): This is a **moderate value**. network has some community structure. However, it's not extremely strong. This could mean there are a few distinct sub-groups discussing different facets of the 5G conspiracy, or different influencer circles. And no.of communities = 8 so with only **20 total nodes**, having **8 communities** means the network is **highly sparse**

5G-Conspiracy Graph Analysis

5. Betweenness Centrality & closeness centrality: can be got from network diameter which is the longest path of all the shortest paths between any two nodes in the network.



What it means: Betweenness Centrality of a node measures how often it acts as a "bridge" along the shortest path between two other nodes. A node with high betweenness has control over the flow of information in the network. **Closeness Centrality** measures how close a node is to all other nodes in the network.

- **High Betweenness Node:** the **bridge** that controls flow between different parts of the network.
- **High Closeness Node:** An **efficient broadcaster** located at the heart of the network.

Result of diameter = 5 means in the graph of 20 nodes, the two most distant users are only **5 steps** apart

6. Connected components: subgraph where any two nodes are connected to each other by paths



What it means: In a directed graph, we look at:

Weakly Connected: Ignores direction. Are they connected in any way?

Strongly Connected: Respects direction. Can every node reach every other node following the direction of edges?

Result Weakly Connected Components = 5, Strongly Connected Components = 8

- There are **5 separate, disconnected segments** in this network when we ignore direction. This means there are 5 groups that have no connection whatsoever to the other groups.
- When we respect direction, this fragments further into **8 components**. This is a sign of a lack of reciprocal communication. For example, many users might be retweeting an influencer (a one-way link), but the influencer doesn't retweet them back, preventing the formation of a strongly connected group.

The chosen layout applied to the graph is :

Fruchterman–Reingold is a force-directed layout algorithm used to visualize graphs. which means:

1. Nodes repel each other

this prevents nodes from overlapping and spreads the graph naturally.

2. Edges pull connected nodes together

strong edges pull tightly connected nodes closer.

3. The layout tries to minimize overall energy

It keeps the entire graph balanced and readable.

Result :A visually clean graph where communities form clusters, hubs move to the center, and outliers move to the edge.

Why used in non-conspiracy graph analysis :

- It gives a natural, balanced shape (similar to real social graphs)
- It visually separates communities clearly
- It avoids distortion (good for graphs with moderate edge density)

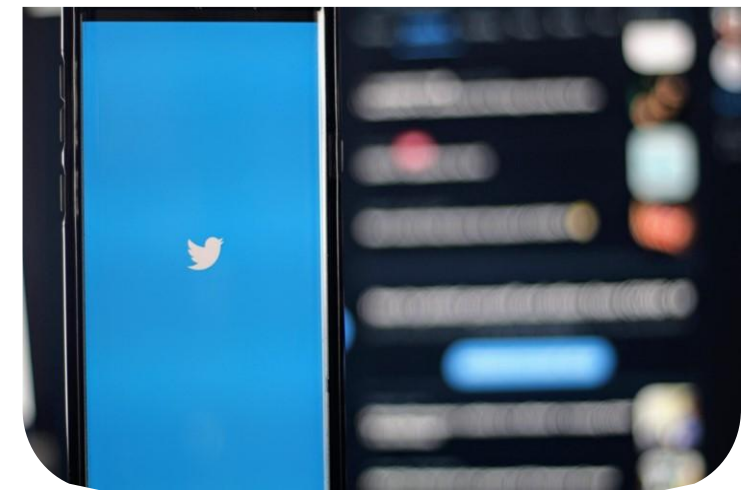
Non-conspiracy graphs are not extremely dense and not extremely sparse. they sit in the middle.

So this layout works best in this middle zone because:

it doesn't collapse the graph into one big blob (like ForceAtlas2 sometimes does)



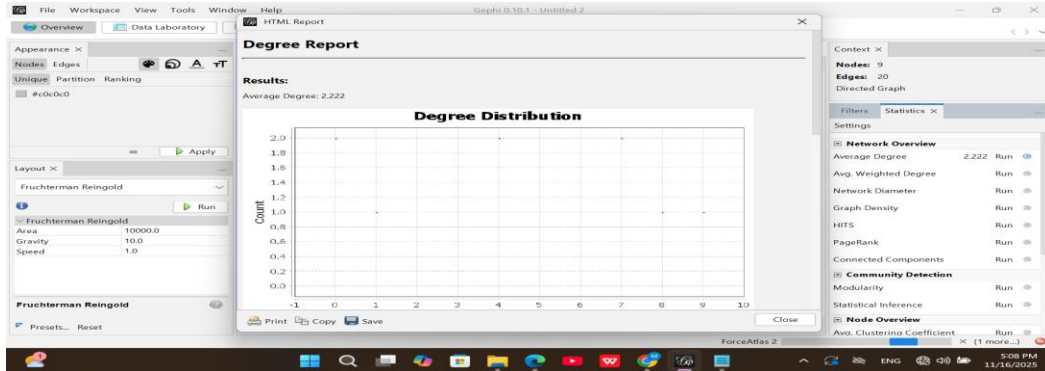
Non- Conspiracy Graph Analysis



Non-Consipracy Graph Analysis

Beginning statistics applied to analyze the graph

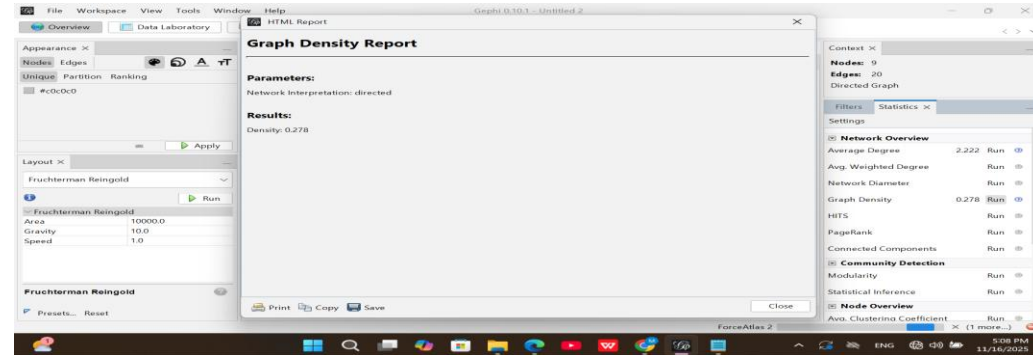
1. Avg degree



Result = 2.222

This is **higher** than the conspiracy graph which its avg degree was 1.9, despite having fewer total nodes(9 nodes) but the conspiracy graph was 20 nodes . This means that, on average, each user in this normal community is connected to more people. It suggests a more actively engaging community where users interact with multiple others.

2. Graph density : The ratio of actual connections to possible connections.

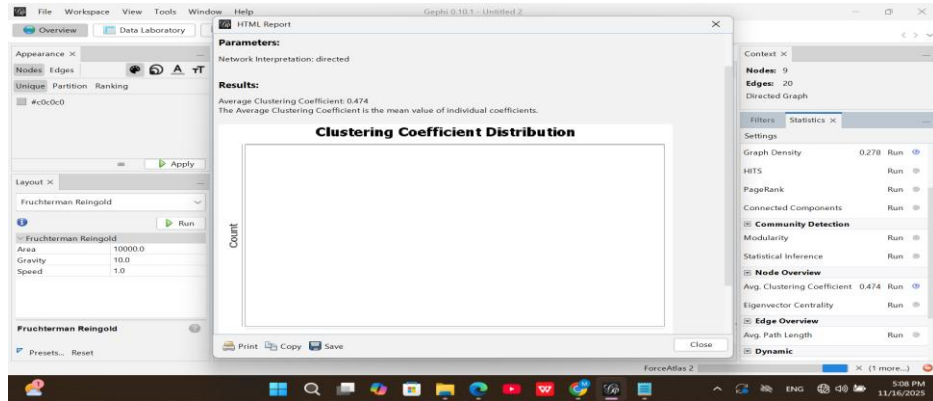


Result = 0.278

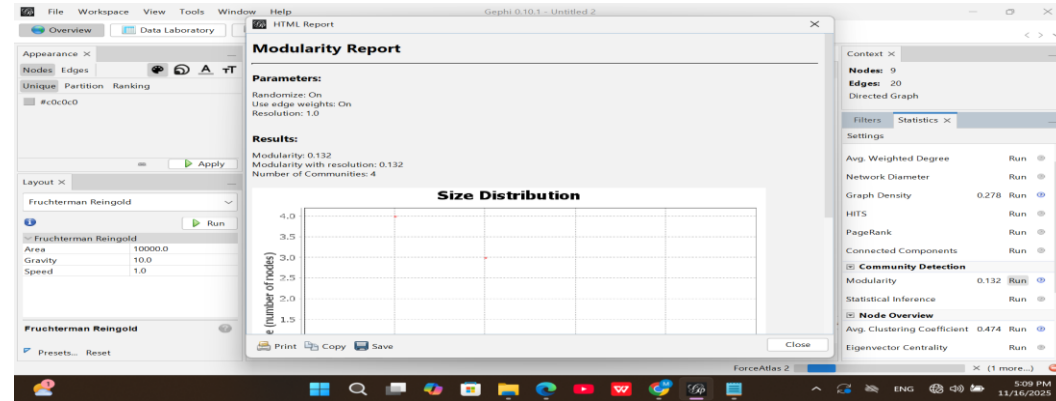
This is **significantly higher** than the conspiracy graph's density of 0.1. A density of 27.8% indicates a much more tightly-knit and well-connected community. It's not just a broadcast network; it's a web of conversations. A lot more of the possible connections between users actually exist.

Non-Consipracy Graph Analysis

3. Average clustering coefficient



4. Modularity and number of communities



Result = 0.474

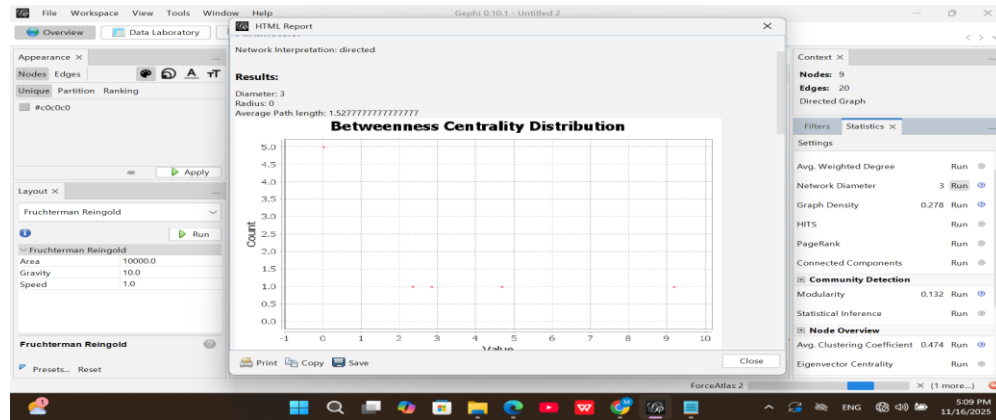
This is **much higher** than the conspiracy graph's 0.266. This is a critical finding. It confirms that this network is composed of real social clusters. Users who talk to the same person are very likely to also talk to each other, forming triangles of interaction. This is the collaborative community, not a one-way broadcast tree.

Result: Modularity = 0.132, Number of Communities = 4

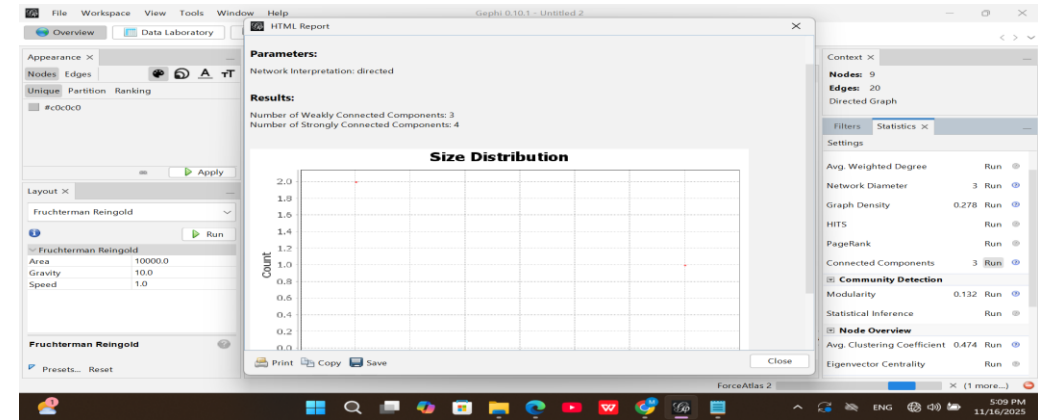
- **Modularity (0.132):** This is a **low value**, especially compared to the conspiracy graph's 0.267. A low modularity means the network is not clearly split into separate, isolated communities. there is more interconnection across the entire network.
- **Number of Communities (4):** With only **9 total nodes**, having 4 communities means the network is still somewhat grouped, but these groups are likely interconnected. It's a mixed structure, not the extreme fragmentation we saw in the conspiracy graph.

Non-Consspiracy Graph Analysis

5. Betweenness and closeness centrality



6. Connected Components



Result of Network Diameter: 3

The shortest path between the two most distant nodes is only 3. This is a very small world, just like the conspiracy graph, showing efficient internal communication.

Betweenness Centrality (from distribution):

The distribution appears less extreme than in the conspiracy graph. The values are more spread out, suggesting that control over information flow is **less concentrated** in one or two super-influencers. It's a more decentralized forming triangles of interaction. This is the collaborative community, not a one-way broadcast tree.

Result: Weakly Connected Components = 3, Strongly Connected Components = 4

There is **difference** from the conspiracy graph.

The network is less fragmented at the macro level (3 weak components VS 5 in the conspiracy graph).

the number of **Strongly Connected Components is 4** is very close to the number of **Weakly Connected Components is 3**. This means that within the weakly connected groups, there is a high degree of reciprocal communication. The connections are not just one-way. This allows for the formation of proper, strongly connected groups where users can actually reach each other through mutual interactions.

Thank You 😊

