Comparing Malicious vs Benign Twitter Subgraphs

By Hager Abdal Hamid Essa 2205239

Introduction

This task was about WICO graph dataset that was used to analyze misinformation and conspiracy theory propagation on Twitter, particularly around the 5G and COVID-19 topics. The dataset provides graph-structured data representing Twitter conversations and interactions. Each graph represents a community or subgraph extracted from the wider Twitter discourse. I focused on 2 graphs: Non_Conspiracy_Graph_1323 and 5G_Conspiracy_Graph_323. This helped to compare malicious vs. benign networks, understand misinformation spread dynamics, identify influential users and echo-chamber behaviors, and analyze network structure such as centralization, modularity, and fragmentation.

Dataset Overview

The source is Twitter and type labeled graph dataset. Nodes represent Twitter users, edges represent interactions such as mentions, replies, and retweets. The labels are: malicious (misinformation clusters related to 5G-Corona conspiracy topics) and benign (regular, non-misinformation discussions). Each subgraph has features like node attributes, edge relationships, community labels, and graph-level metadata (topic, polarity, etc.).

Analysis Steps

For both graphs, I calculated the following using Gephi:

• Number of Nodes and Edges to detect size and engagement of the network.

• Top 3 Users by Degree Centrality to identify influential users or hubs.

• Top 3 Users by Betweenness Centrality to detect information brokers.

• Bridge Nodes between Communities to detect critical connectors.

• Average Clustering Coefficient to measure local cohesiveness and echo-chamber intensity.

• Modularity (Q) Value and Number of Communities to assess community separation.

• Network Density to see if the graph is tightly or loosely connected.

• Echo-Chamber or Isolated Patterns observation.

• Effect of Removing Top Bridge Node to simulate disruption of misinformation spread.

• Visual Description (open / closed / centralized / fragmented).

Comparative Analysis

Metric	5G Misinformation	Benign / Normal	Observation
Nodes / Edges	34 nodes, 42 edges	51 nodes, 127 edges	5G network is smaller, focused; benign network is larger, more diverse.
sDensity	0.037	0.05	5G network is sparse; benign network is more connected.
Average Path Length	4	7	5G is more connected with shorter paths; benign network longer paths → less centralization.
Average Clustering Coefficient	0.033	0.308	5G weakly clustered →less cohesive locally; benign moderately clustered →some local reinforcement.
Modularity (Q) & Communities	0.677, 7 communities	0.396, 5 communities	5G more segregated; benign more open with looser communities.
Top Influential Users (Degree Centrality)	121056... (15), 828444... (5), 61899869 (5), 40095521 (4), 123158... (4)	54156... (12), 63716... (34), 79494... (28), 51541... (16), 11329... (6)	5G has several active hubs; benign network has multiple moderately active users.

Brokers (Betweenness Centrality)	121056… (47.5), 61899869 (12.0), 40095521 (6.0), 828444… (4.0), 105141… (4.0)	79494… (769.75), 63716… (524.8), 14152… (420.35), 54156… (374.8), 19994… (54.0)	5G brokers are distributed; benign network relies on a few central mediators.
Bridge Nodes	121056…, 61899869	79494…, 63716…	Both networks rely on key connectors; removal would fragment communication.
Visual Structure	Directed, sparsely connected, semi-closed clusters	Directed, open, distributed	5G resembles star-like core → echo-chamber; benign more decentralized → smooth flow.
Echo-Chamber Behavior	More pronounced	Less pronounced	5G clusters recycle trust internally; benign network has freer information sharing.
Misinformation Spread Potential	Higher	Lower	Dense and repetitive 5G connections → faster spread; benign network slower spread.
Trust & Information Flow	Compartmentalized within closed groups	Open and organic	5G recycles trust internally; benign shares across groups.
Effect of Removing Top Bridge Node	Removing 121056… →fragmentation	Removing 79494… →isolates some users	Both networks fragile, dependent on central nodes.

Visual Description

For the 2 graphs, I imported nodes and edges data and ran ForceAtlas2, Fruchterman Reingold, and Yifan Hu graph layout algorithms for visualization. Malicious graphs often appear dense, centralized, and echo-chambered, with few nodes dominating information flow. Benign graphs are more open, distributed, and connected.

Conclusions

This analysis shows how the structure of online networks shapes information dissemination. Comparing the 5G misinformation subgraph with a benign one revealed clear behavioral and structural differences. The 5G conspiracy network was smaller but denser, forming strong echo chambers. Benign networks were more open, allowing freer information flow. Bridge nodes were crucial in both networks, and their removal caused fragmentation, highlighting the importance of central users in controlling misinformation spread.