Assignment 2: Social Contagion with Malicious Nodes

Network Analysis

Seyedeh Aida Atarzade Hosseini
6936800

Delaram Doroudgarian
5881909

Academic Year 2024/2025

1 Introduction

In complex networks such as online social media or email communications, the spread of information follows patterns influenced by the network topology and the behavior of individual nodes. In this assignment, we simulate a realistic scenario where a secret message is propagated across a network. Some initial nodes, called **gossipers**, start the diffusion by sending the original message to their neighbors.

However, the network also contains **malicious nodes**, which tamper with the received message before forwarding it, thereby corrupting the content. The simulation models how information flows under such conditions and analyzes the impact of malicious behavior on the integrity of the message.

Each node in the network adopts a simple decision rule: it only accepts and forwards a message if a certain proportion of its neighbors have already accepted it. This threshold-based model captures realistic scenarios in which individuals require social confirmation before participating in spreading information.

This study investigates the dynamics of information diffusion under various conditions, including:

- Different numbers of initial gossipers and malicious nodes
- Variable acceptance thresholds
- Network structure and node roles

At the end of the simulation, we also analyze the **semantic consistency** of messages using *cosine similarity*, and visualize the results as a heatmap. This helps to understand how much the original message has been distorted by malicious actors.

2 Methodology

2.1 Dataset and Graph

The dataset used for this simulation is the **Email-Eu-core network**, which represents email interactions between members of a large European research institution. It includes directed email communications between 1,005 nodes (users), where an edge from node A to node B indicates that A sent at least one email to B.

2.2 Graph Preprocessing

We first converted the directed graph into an undirected version to simulate bidirectional communication. Then, we extracted the largest connected component (GCC) to ensure that all selected nodes belong to a single connected cluster capable of spreading information. Additionally, we cleaned the dataset by removing self-loops and isolated nodes, which could otherwise interfere with the diffusion process.

2.3 Node Roles: Gossipers and Malicious

- **Gossiper Nodes**: These nodes initiate the spread by holding the original message from the start. They actively forward the message to their neighbors and play a key role in how quickly and widely the information diffuses.
- **Malicious Nodes**: These nodes aim to disrupt the process. When they receive a message, they alter its content and spread the tampered version, reducing overall message integrity in the network.
- **Neutral Nodes**: Neutral nodes don't start with any message. They adopt and forward a message (original or tampered) once enough of their neighbors have shared it, based on a threshold.

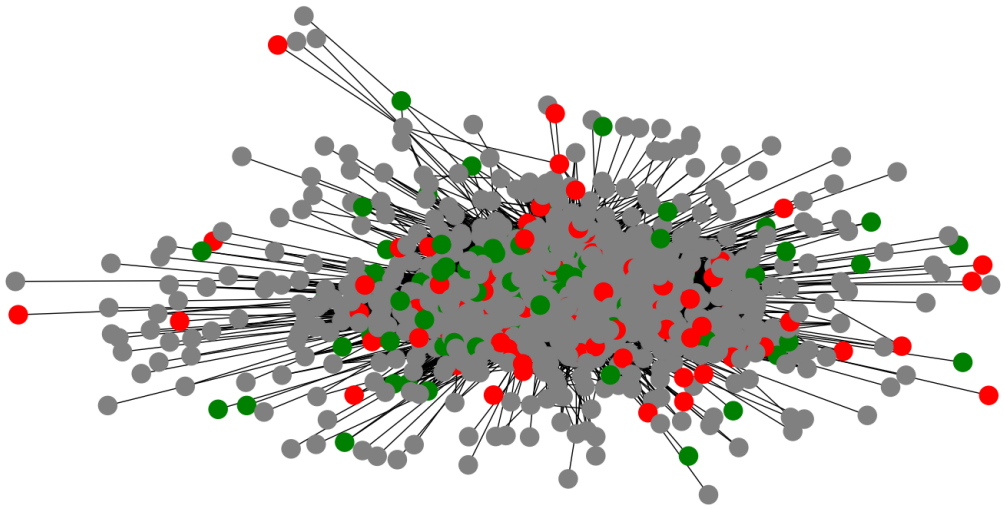


Figure 1: The giant component from the Email-Eu-core dataset. Gossipers are in green, malicious nodes are in red, and all others are in gray.

3 Message Diffusion Model

The message diffusion process models how information spreads in a social network under the influence of both cooperative and adversarial nodes. Each node holds a message, initially empty, and may decide to accept and forward a message based on the behavior of its neighbors.

3.1 Message Initialization

At the beginning of the simulation:

- Gossipers are initialized with the original message: `"this is my secret message"`.
- All other nodes start with an empty message.

3.2 Forwarding Rule and Threshold

Each node evaluates whether to accept a message in the following way:

- A node checks how many of its neighbors have already accepted a message.
- If the proportion of such neighbors exceeds a global threshold, the node accepts the message.

3.3 Malicious Node Behavior

When a node receives a message, one of two behaviors occurs:

- If the node is benign (normal), it forwards the message unchanged.
- If the node is malicious, it **tampers** with the message before forwarding it. Tampering is done by randomly replacing a character in the message with a string such as `"1111"`, `"2222"`, etc.

3.4 Simulation Loop

The diffusion process unfolds in discrete steps. At each step:

1. All nodes are scanned to determine if they meet the threshold condition.
2. Nodes that qualify accept and forward the message.
3. Colors are assigned to nodes for visualization:
    - Green: node received the **original** message.
    - Red: node received a **tampered** message.
    - Gray: node has not received any message.
4. The process continues until no more messages are changing.

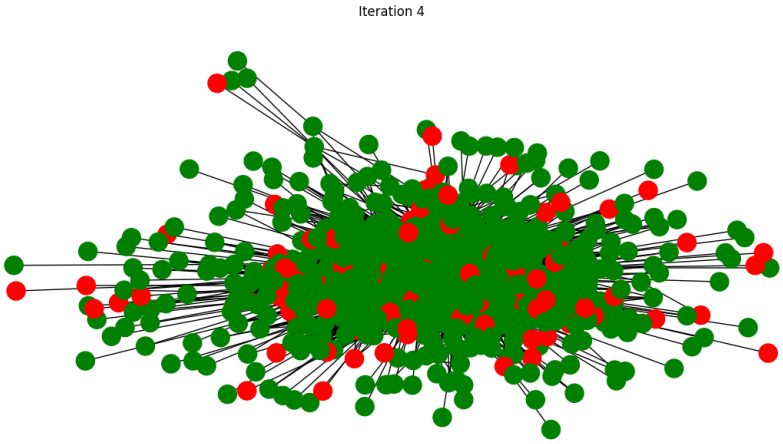
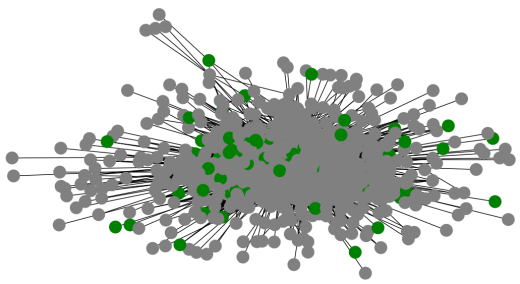
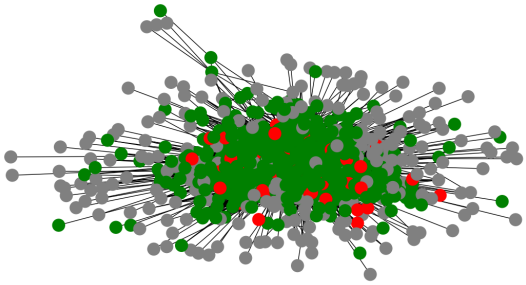


Figure 2: Final graph after diffusion. Nodes with the original message are green; those with tampered messages are red.

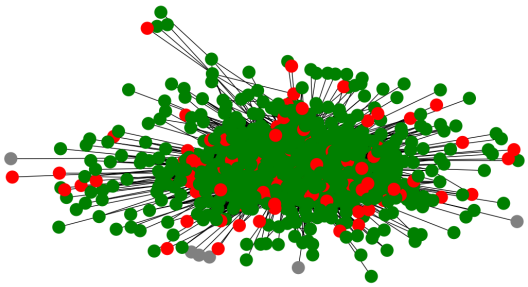
3.5 Iteration-wise Analysis



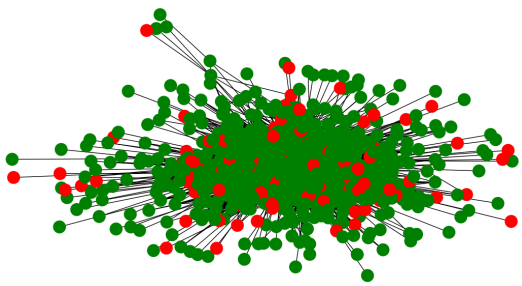
(a) Iteration 1



(b) Iteration 2



(a) Iteration 3



(b) Iteration 4

Threshold	Coverage	Original Message	Steps to Convergence
0.01	100%	99.5%	4
0.02	100%	99.1%	4
0.06	100%	99.5%	7
0.08	100%	99.1%	8
0.09	1.9%	1.9%	No convergence
0.1	1.4%	1.4%	No convergence

Table 1: Effect of threshold values on information spread dynamics

4 Results

4.1 Threshold Effects on Information Spread

- For **low thresholds (0.01–0.08)**, the network achieves **100% coverage** and retains over **99% original** messages. However, as the threshold increases, **convergence becomes slower** (from 5 to 9 steps), since nodes require more agreement from neighbors.
- When the threshold reaches **0.09 or 0.10**, diffusion **almost completely breaks down**. Very few nodes receive the message, and the process ends quickly due to lack of activation.
- This indicates a **critical threshold** between **0.08 and 0.09**, beyond which the network can no longer support information spread.
- Overall, thresholds in the range **0.02 to 0.06** offer the best balance between **coverage, integrity, and speed**.

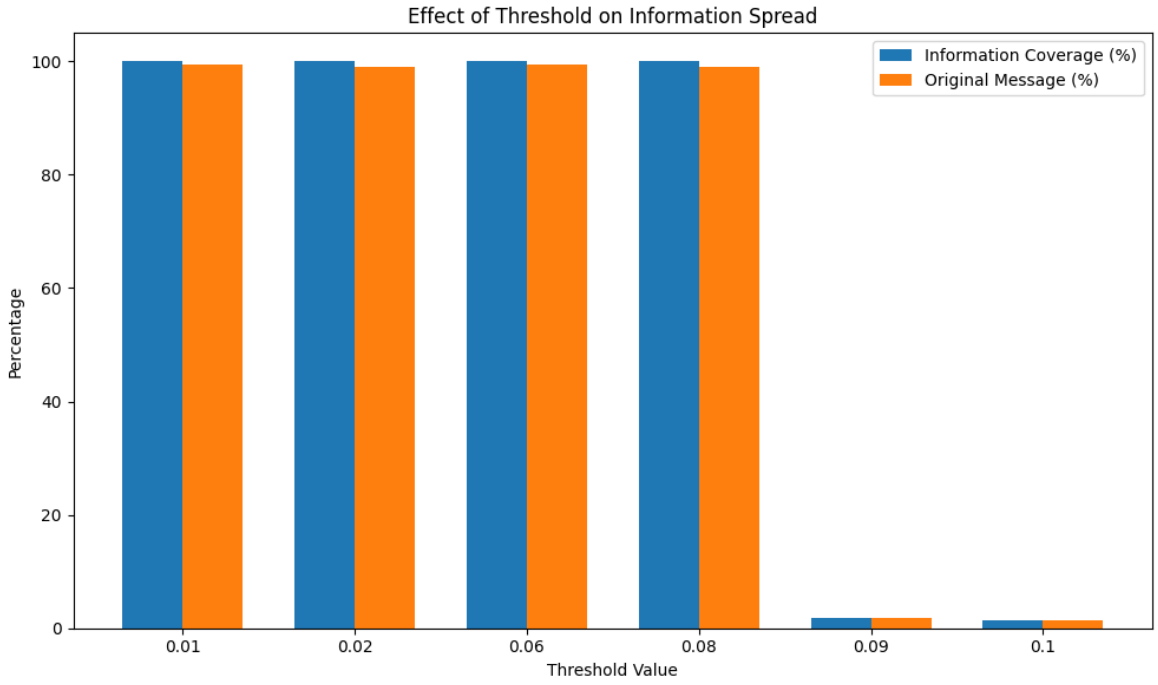


Figure 5: Effect of threshold on information spread

Gossipers

Gossipers	Coverage	Original Message	Steps to Convergence
1	100%	99.1%	7
4	100%	99.5%	7
10	100%	98.8%	6
30	100%	99.1%	4
95	100%	99.4%	3

Table 2: Effect of Gossipers on diffusion dynamics

- **Coverage remains 100%** in all cases, regardless of the number of gossipers. This shows that the network is well-connected and resilient — even a single gossiper can eventually reach all nodes.
- Increasing the number of gossipers **reduces the steps to convergence**, from 8 steps (with 1 or 4 gossipers) to just 4 steps (with 95 gossipers). More sources speed up the spread.
- **Message integrity fluctuates slightly**, with a minimum at 10 gossipers (98.8%) and a maximum at 4 gossipers (99.5%). These small differences are likely due to how close the gossipers are to malicious nodes in the network.

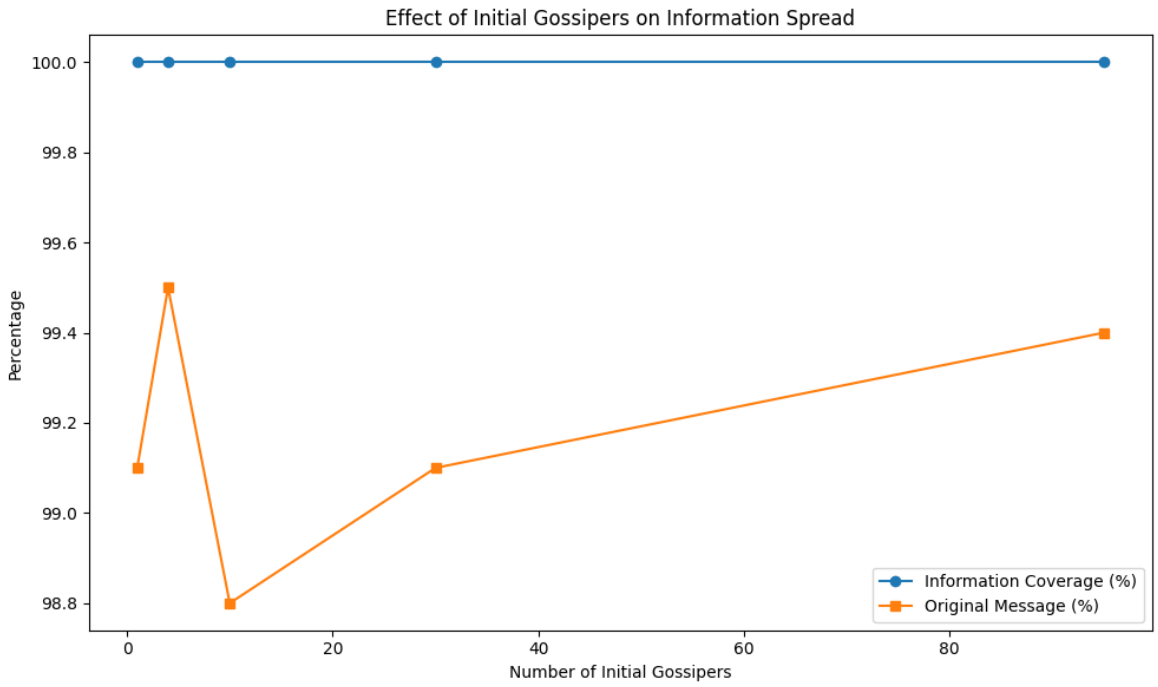


Figure 6: Effect of Gossipers on Information Spread

Malicious Nodes

Malicious Nodes	Coverage	Original Message	Unique Messages
0	100%	100%	1
1	100%	99.9%	2
8	100%	96.3%	9
20	100%	74.3%	21
100	100%	46.3%	95

Table 3: Effect of malicious nodes on information integrity

- **Coverage remains 100%** in all cases, indicating that every node in the network receives a message regardless of the number of malicious nodes. However, this does **not guarantee integrity**—the content of the messages may be altered.
- **Original Message % drops significantly** as the number of malicious nodes increases: From **100%** when no malicious nodes exist, down to **46.3%** with 100 malicious nodes. This shows that a higher number of malicious nodes dramatically reduces the number of nodes that receive the correct message.
- **Unique Messages grow steadily**, from just **1** (when all nodes receive the same original message) to **95** (indicating widespread tampering and diversity in received messages). This metric effectively captures **information pollution** in the network. A high number of unique messages means that different parts of the network have received conflicting versions of the message, causing confusion and inconsistency.

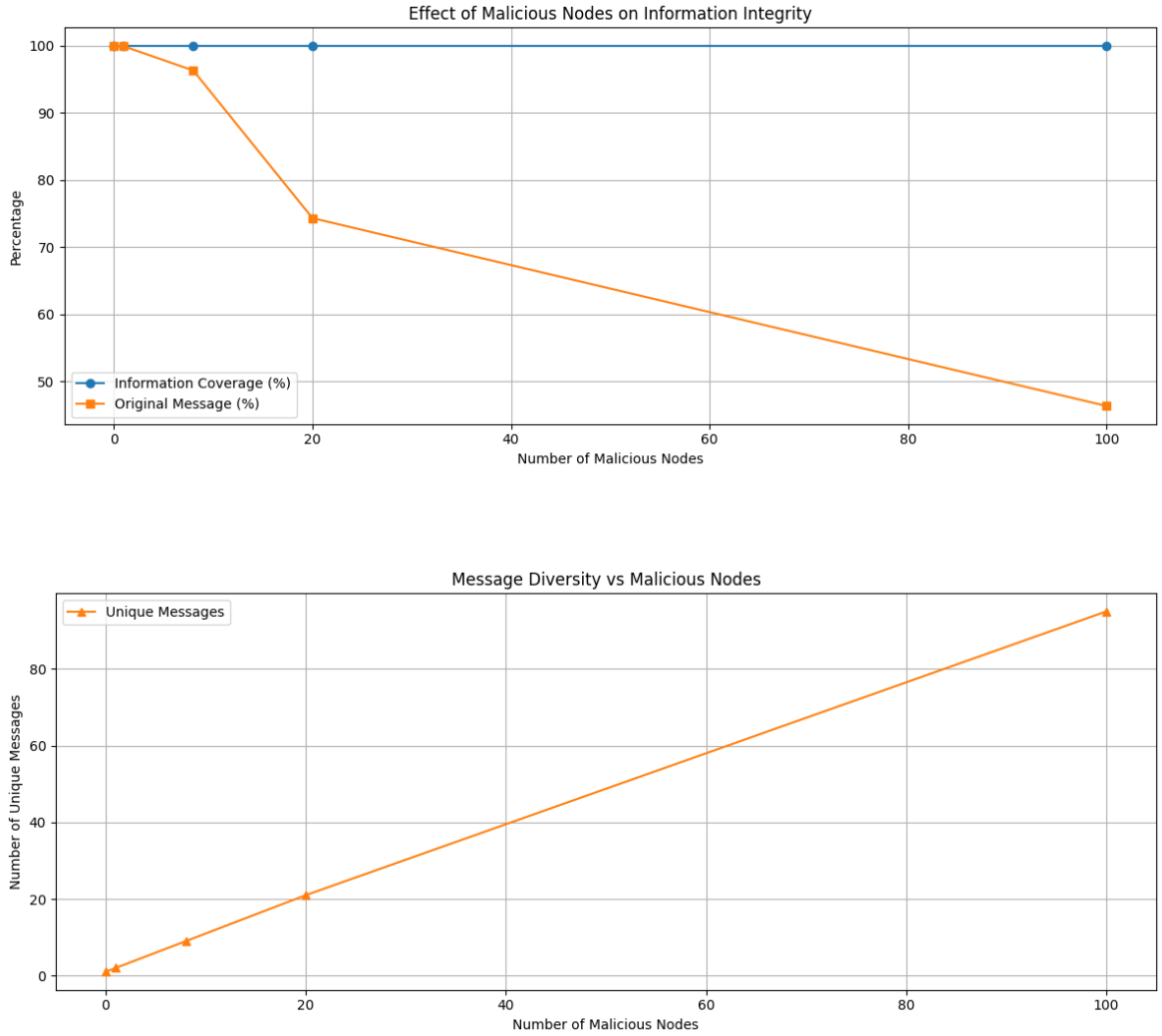


Figure 7: Effect of malicious nodes on information integrity and diversity

Targeting Strategy Analysis

Strategy	Coverage	Original Message
Random	100%	99.5%
Highest Degree	100%	89%
Betweewnness Centrality	100%	88.6%
Closeness Centrality	100%	90%

Table 4: Effectiveness of malicious node targeting strategies

- All strategies result in 100% **coverage**, meaning every node receives a message. However, **message integrity** (i.e., receiving the original message) varies greatly.
- **Random** targeting causes the **least damage** (99.5% original messages), as it often selects low-impact nodes.

- **Highest Degree** and **Closeness Centrality** reduce integrity more (89–90%), as these nodes influence many others directly or quickly.
- **Betweenness Centrality** causes the **most damage** (only 88.6% original messages).
- Nodes with the most betweenness centrality act as bridges between different parts of the network. Most message paths go through them. If they're malicious, they tamper with many messages traveling between communities.

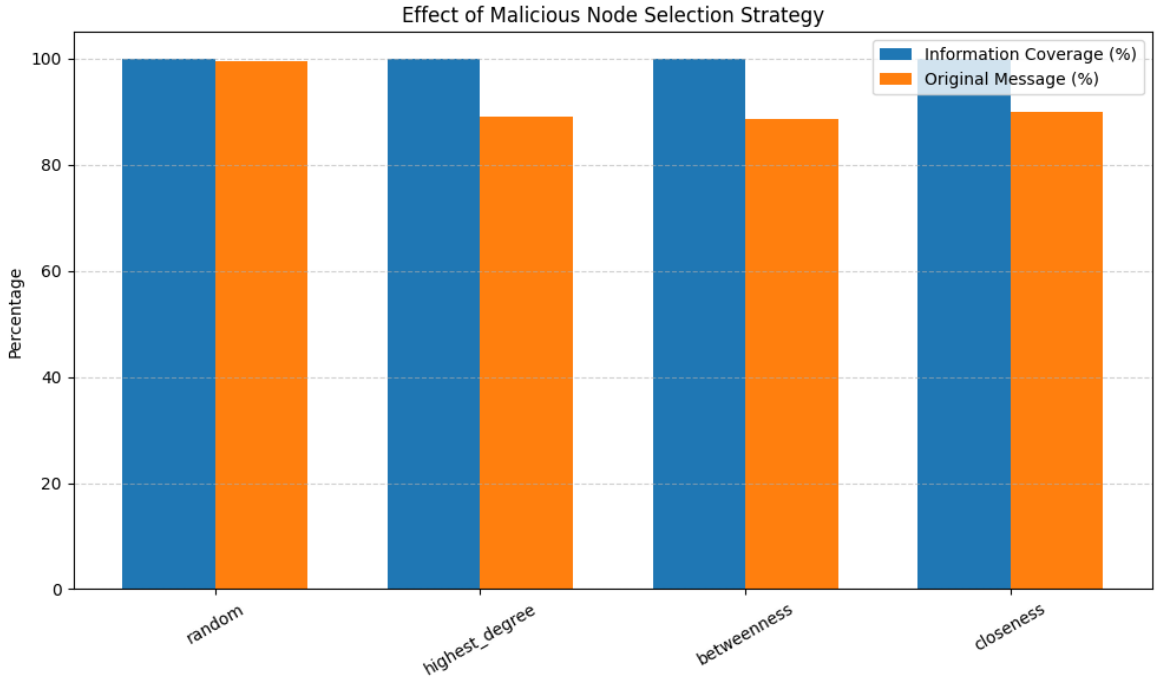


Figure 8: Comparison of malicious node targeting strategies

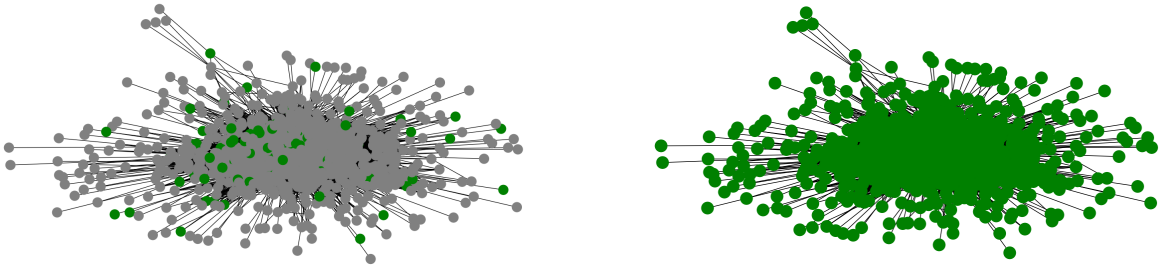
Combined Scenario Analysis

Scenario	Threshold	Gossipers	Malicious Nodes	Coverage	Original Message
Best Case	0.01	95	0	100%	100%
Worst Case	0.06	2	100	100%	50.3%
Realistic Case	0.07	10	15	100%	97.1%

Table 5: Combined scenario effects under different conditions

5 Detailed Scenario Visualization

5.1 Best Case Scenario



(a) Initial state with 95 gossipers

(b) Final state with 95 gossipers

Figure 9: Best case scenario network states

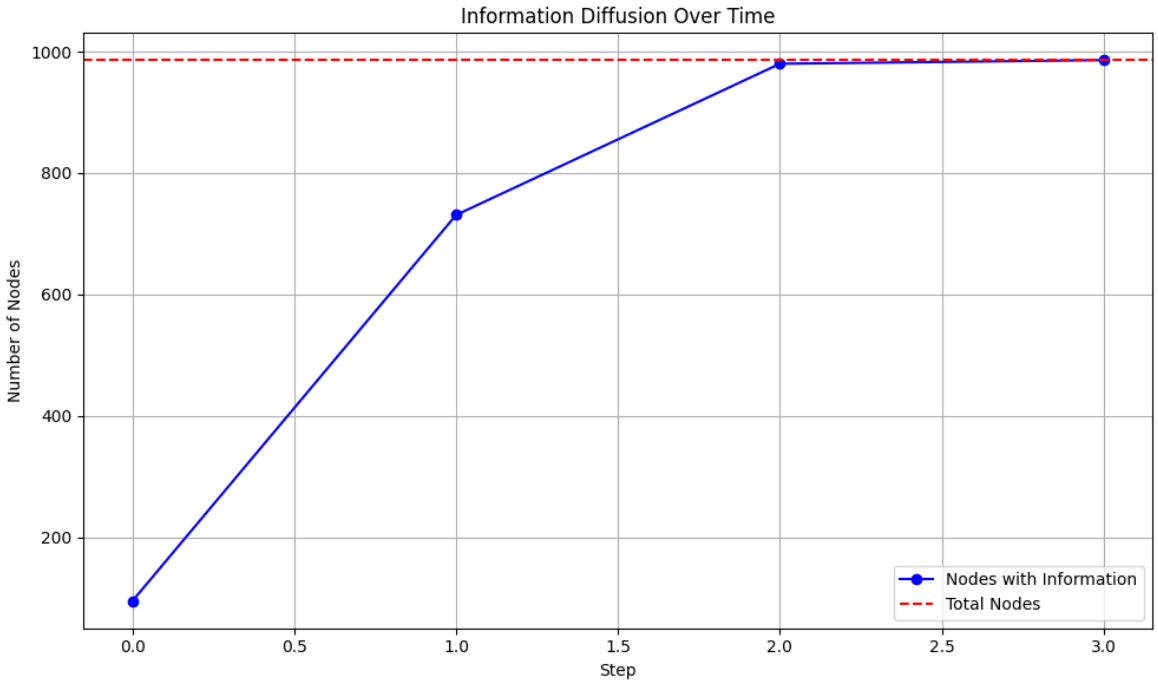


Figure 10: Best case diffusion curve showing rapid exponential growth (3 steps)

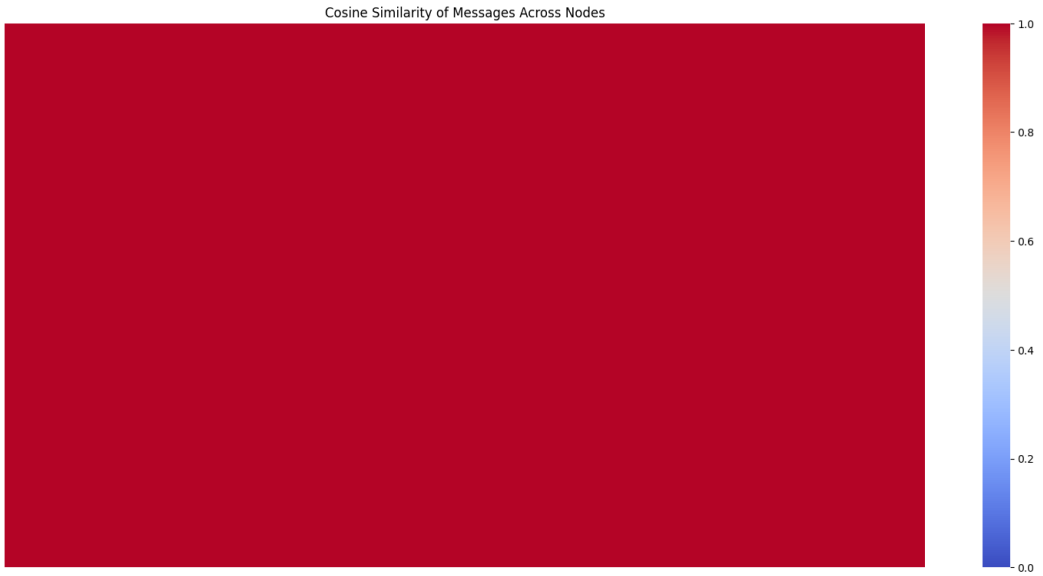
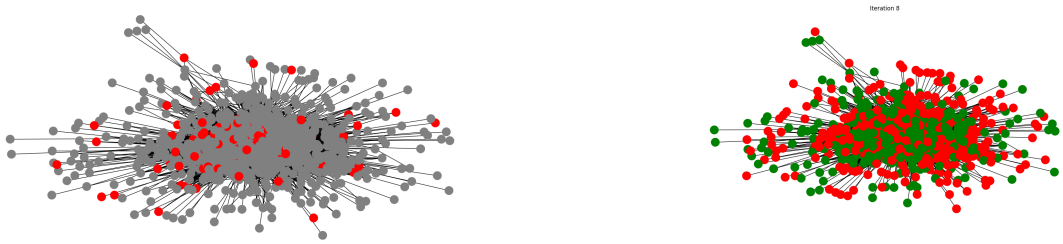


Figure 11: Best case cosine similarity showing perfect message integrity

5.2 Worst Case Scenario



(a) Initial state with 1 gossiper and 100 malicious nodes

(b) Final state with 1 gossiper and 100 malicious nodes

Figure 12: Worst case scenario network states

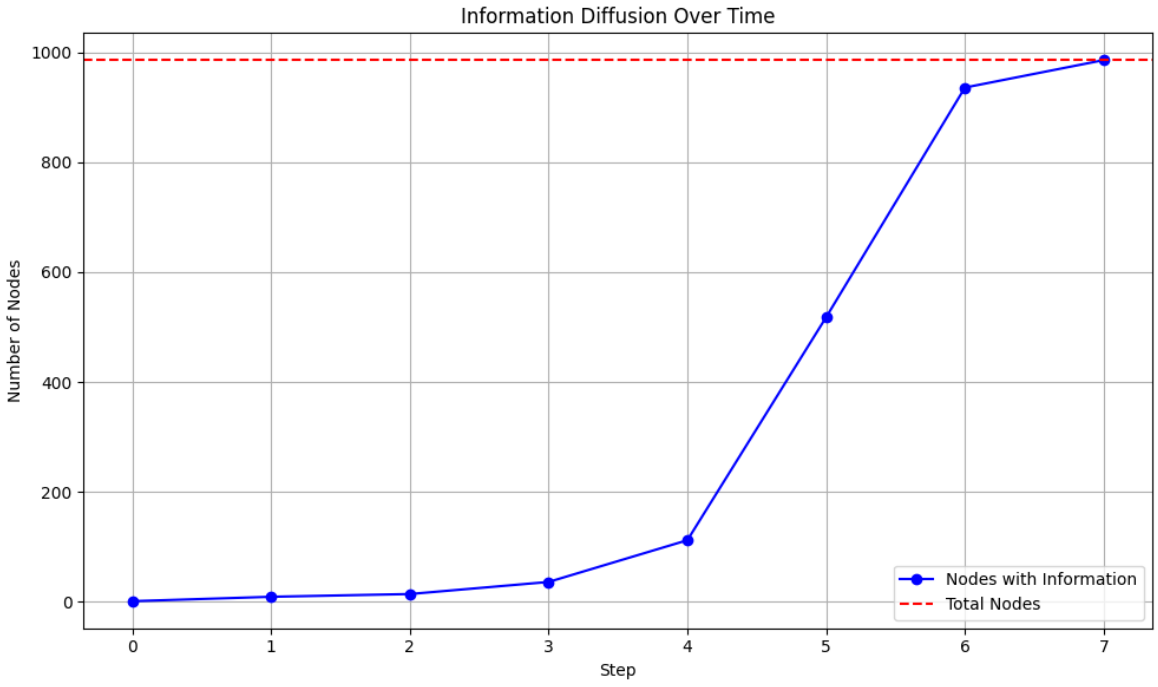


Figure 13: Worst case diffusion showing slower growth and incomplete coverage

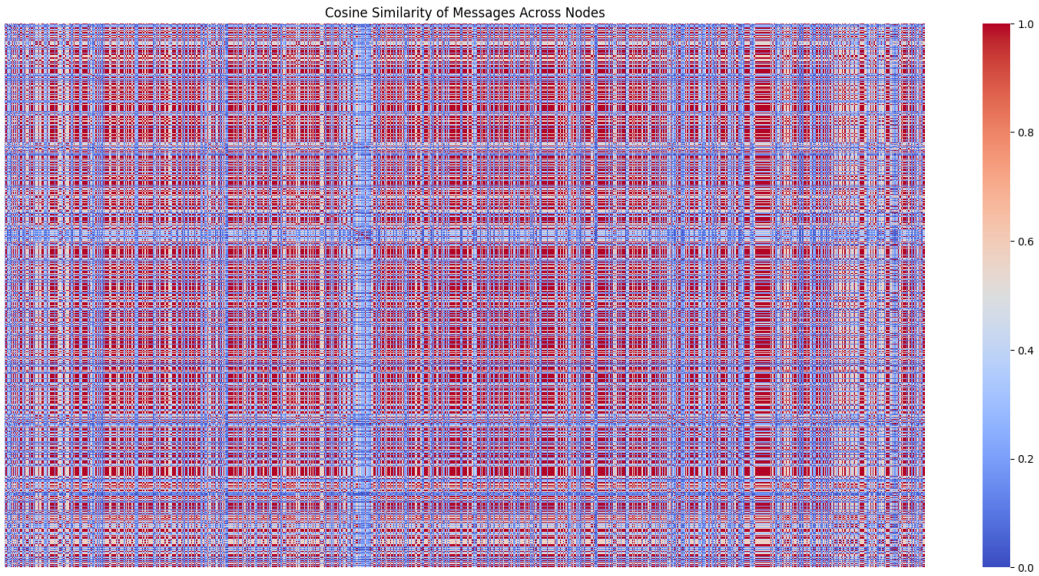
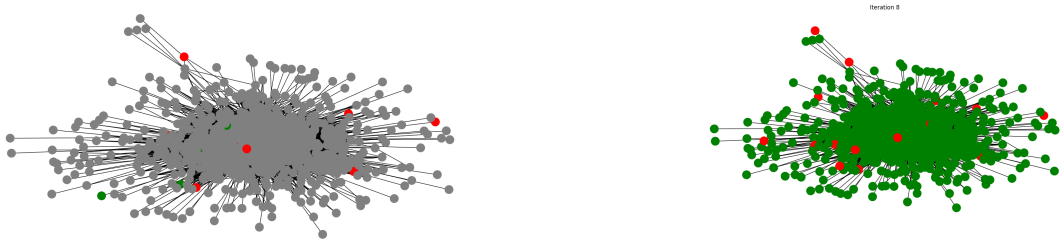


Figure 14: Worst case similarity matrix showing 94 message variants with 49.9% similarity maintained

5.3 Realistic Scenario



(a) Initial state with 10 gossipers and 15 malicious nodes

(b) Final state with 10 gossipers and 15 malicious nodes

Figure 15: Realistic scenario network states

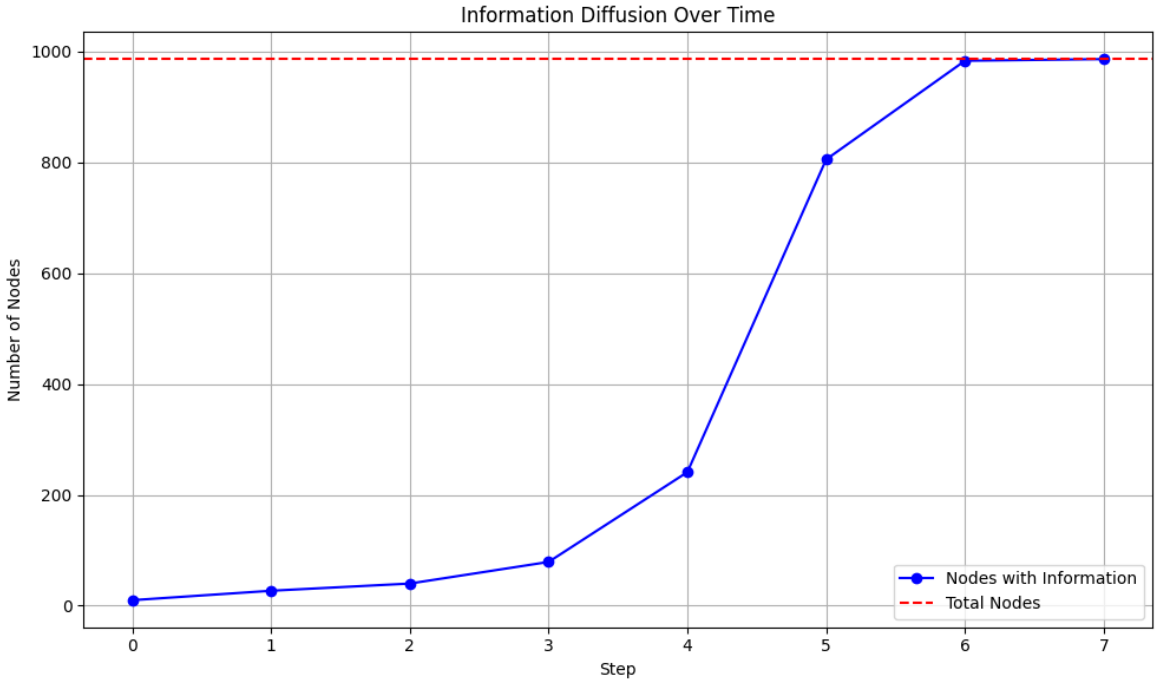


Figure 16: Realistic scenario showing balanced growth (complete coverage in 7 steps)

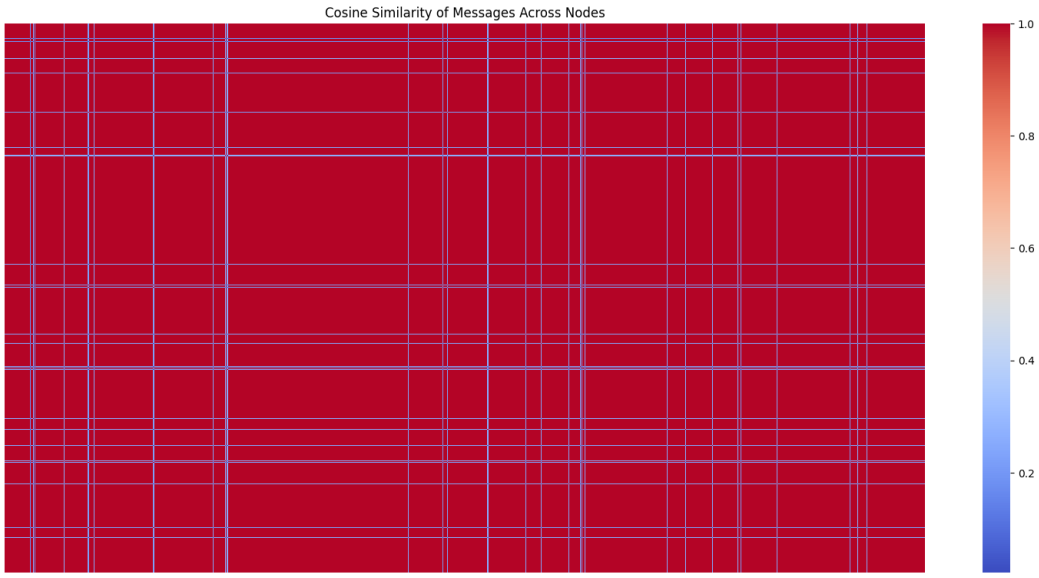


Figure 17: Realistic similarity matrix with 97.1% similarity maintained