Assignment No. 2

Sending information through a network

Network:

- Twitter Social Circles

Group Members:

- Parsa Moslem (5755015)
- Amir Mohammad Azimi (5795736)

Course:

- Network Analysis

The course is part of the below degree:

- Computer Science – Software & Security Engineering

Academic Year: 2023 - 2024

UNIVERSITÀ DEGLI STUDI
DI GENOVA

Dibris

Table of Contents

Introduction

Gossip protocols are a powerful tool for simulating the spread of information within networks. They allow us to examine the factors that influence how quickly (or slowly) data disseminates, helping us to identify which network topologies are more susceptible to various patterns of information diffusion. This can apply to the spread of data, messages, or even the propagation of viruses and diseases.

The presence of malicious nodes adds a layer of complexity to this process. These nodes can act as agents of misinformation, altering the original message and leading to the dissemination of false information throughout the network.

In this simulation, we delve into the dynamics of gossip protocols and explore how the presence of malicious actors can compromise the integrity of the message being spread. We aim to understand the conditions under which misinformation can flourish and the potential impact it can have on the overall information landscape of the network.

Simulation Setup

The simulation was conducted on a subgraph of the Twitter social circles network, a dataset inspired by research on identifying user-defined social circles within online platforms. This network, derived from Twitter's "circles" (or "lists") feature, represents connections between users who have categorized each other into specific groups. Due to the computational demands of the full dataset (81,306 nodes and 1,342,310 edges), a random sample of 2000 nodes were selected. The giant component of this subsampled network, representing the largest interconnected group, consisted of 355 nodes.

Within this giant component, 20 nodes were randomly designated as "gossipers," the initial spreaders of information. Another 20 nodes were designated as "malicious," tasked with altering the message as it passed through them. The remaining nodes were considered "regular," simply receiving and potentially forwarding the message based on the actions of their neighbors.

The simulation's acceptance threshold was set at 0.2, meaning a node would only adopt and potentially spread a message if at least 20% of its neighbors had already done so. The original message, a simple text string "Secret message," was the information disseminated by the gossipers. Malicious nodes, upon receiving the message, would tamper with it by randomly changing a single character, ensuring the message's content was altered before further transmission.

Simulation Result

Information Spread

The simulation demonstrates that the gossip protocol effectively disseminated information throughout the network, reaching 348 out of 355 nodes (98%). However, the presence of malicious nodes significantly impacted the integrity of the message. A total of 21 unique variations of the original message "super secret sentence" emerged due to tampering. The most prevalent version was the original message itself, received by 269 nodes. The remaining variations, each with minor alterations, were distributed among the other nodes, with some versions appearing more frequently than others. Notably, 7 nodes remained uninformed, highlighting potential limitations in the information spread due to network topology or the specific interactions between nodes.

This result underscores the vulnerability of gossip protocols to misinformation in the presence of malicious actors. Despite the high overall reach of the message, the

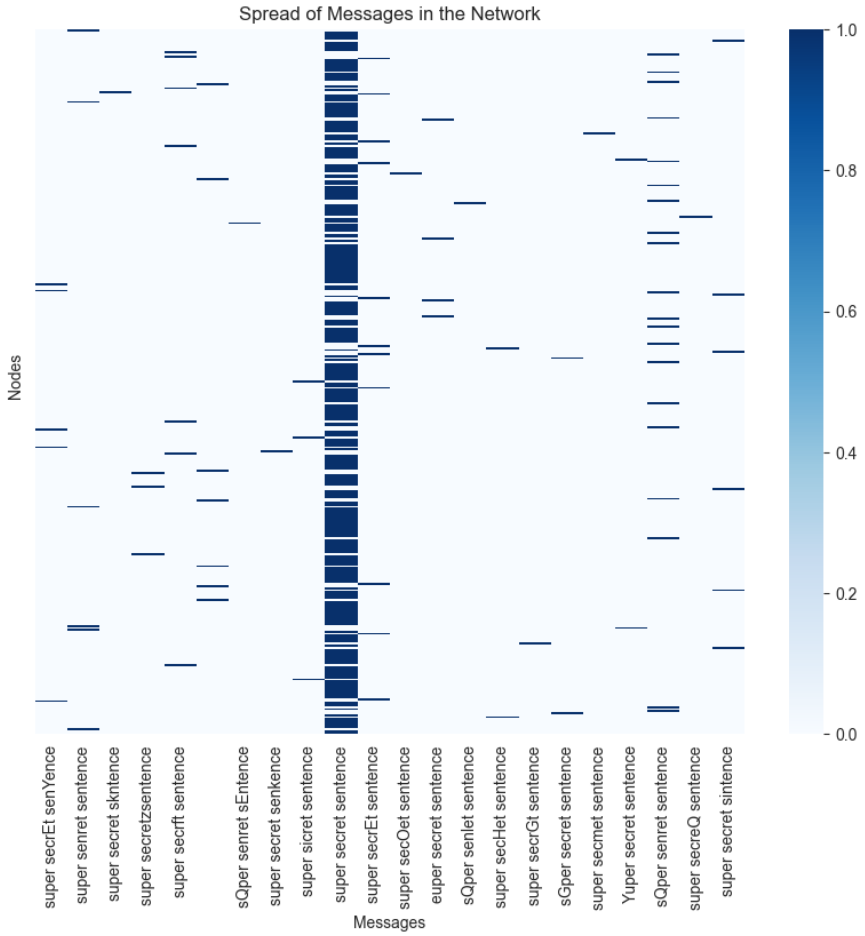

Figure 01

proliferation of altered versions raises concerns about the reliability and trustworthiness of information disseminated through such networks.

Figure 1 illustrates the spread of messages within the network. The heatmap reveals a concentration of the original message (dark blue) among a large portion of nodes, while the presence of lighter blue shades indicates the varying degrees of dissemination for the tampered messages.

The heatmap reveals that the original message, "super secret sentence," was the most widespread, reaching a significant portion of the network. However, the presence of multiple columns with varying shades of blue highlights the proliferation of tampered messages throughout the network. Some altered messages were received by a considerable number of nodes, while others were limited to a smaller subset. This visualization underscores the extent to which malicious actors can introduce and propagate misinformation within a network using a gossip protocol.

Additionally, the heatmap reveals several nodes that did not receive any messages, as indicated by entirely white rows. This suggests that certain nodes might be isolated or less connected within the network, hindering the complete spread of information.

Cosine Similarity Analysis

The cosine similarity heatmap in Figure 2 provides a visual representation of the similarity between message versions held by different nodes. Each cell in the heatmap represents the cosine similarity between the messages held by two nodes, with a value of 1 indicating identical messages and 0 indicating completely dissimilar messages.

The heatmap exhibits a complex pattern of similarity and dissimilarity, reflecting the impact of message tampering by malicious nodes. While a significant portion of the nodes share high cosine similarity, indicating the prevalence of the original or slightly altered messages, there are also numerous instances of low similarity, suggesting the presence of more significantly tampered messages.

The heatmap's complexity makes it difficult to visually identify the original gossipers or track the specific paths of message diffusion. However, the overall distribution of similarity values suggests a mixture of successful information dissemination and the propagation of misinformation due to the actions of malicious nodes.

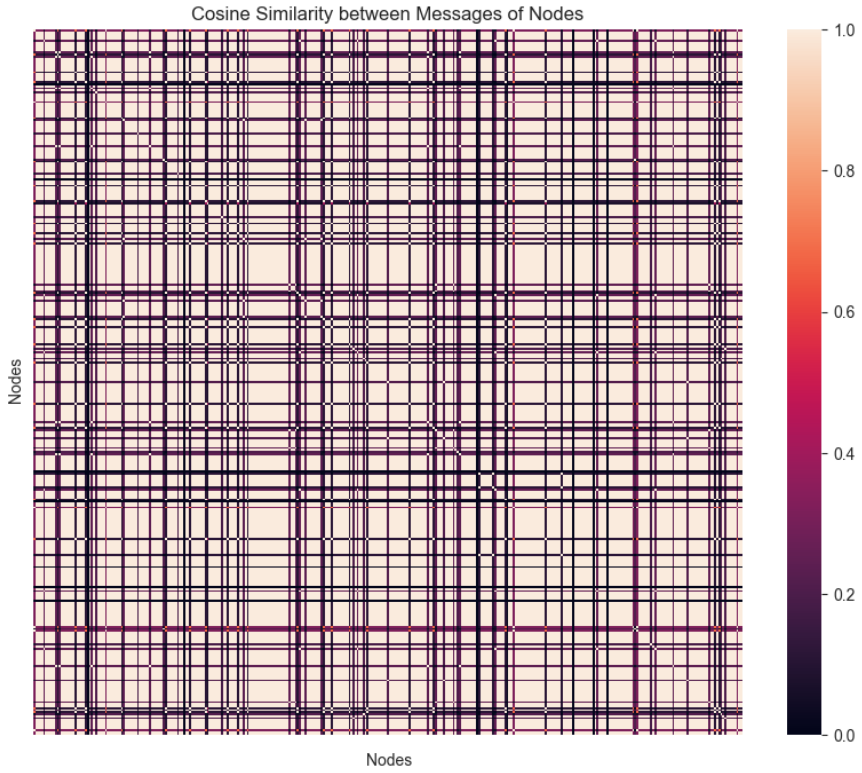

Figure 2

The presence of both high and low similarity clusters within the network indicates that while the gossip protocol can effectively spread information, it is susceptible to the creation and dissemination of diverse message variants in the presence of malicious actors. This diversity of messages can lead to confusion and uncertainty among nodes, hindering the establishment of a shared understanding within the network.

Experimenting with different parameters

To further explore the impact of varying numbers of initial gossipers, malicious nodes, and acceptance thresholds, additional experiments were conducted. The table below summarizes the key findings from these experiments:

Gossiper Count	Malicious Count	Threshold	Original Messages	Tampered Messages	Uninformed Nodes
10	20	0.1	337	36	0
10	20	0.2	79	3	291
10	20	0.3	38	4	331
10	40	0.1	241	132	0
10	40	0.2	99	189	85
10	40	0.3	18	0	355
10	60	0.1	154	219	0

10	60	0.2	46	38	289
10	60	0.3	13	1	359
20	20	0.1	345	28	0
20	20	0.2	268	83	22
20	20	0.3	68	0	305
20	40	0.1	289	84	0
20	40	0.2	252	63	58
20	40	0.3	86	28	259
20	60	0.1	260	113	0
20	60	0.2	192	150	31
20	60	0.3	65	23	285
30	20	0.1	295	78	0
30	20	0.2	291	52	30
30	20	0.3	241	33	99
30	40	0.1	264	109	0
30	40	0.2	250	93	30
30	40	0.3	130	9	234
30	60	0.1	269	104	0
30	60	0.2	241	93	39
30	60	0.3	114	7	252

Analysis

- **Impact of Gossiper Count**: Increasing the number of gossipers generally improves the spread of the original message.
- **Impact of Malicious Nodes**: The presence of more malicious nodes increases the spread of tampered messages.
- **Threshold Effect**: Higher thresholds reduce the overall spread of information.

These additional experiments align with the initial findings and provide a more comprehensive understanding of how different parameters influence information dissemination in the network.

Visualizations

To gain a deeper understanding of the information diffusion process, let's visualize the spread of messages across the network over time.

The initial state of the network is shown in Figure 3 at the onset of the gossip simulation. Green nodes represent the original gossipers, the sources of the "super secret sentence," while red nodes represent malicious actors prepared to tamper with the message. The majority of nodes, depicted in blue, are regular nodes who may receive and spread the message depending on their neighbors' actions.

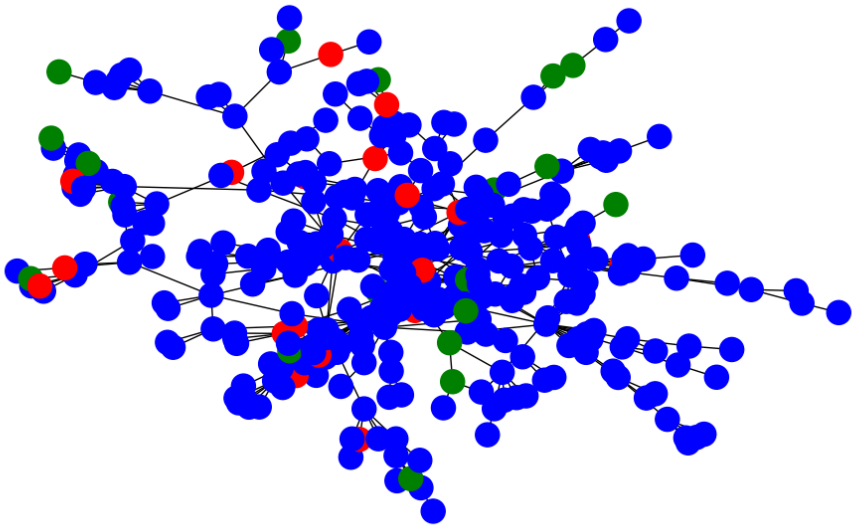

Figure 3

As the simulation progresses, Figure 4 showcases an intermediate stage where the message has begun to spread throughout the network. The growing number of green nodes indicates that many regular nodes have received and adopted either the original or a tampered version of the message, now acting as secondary gossipers.

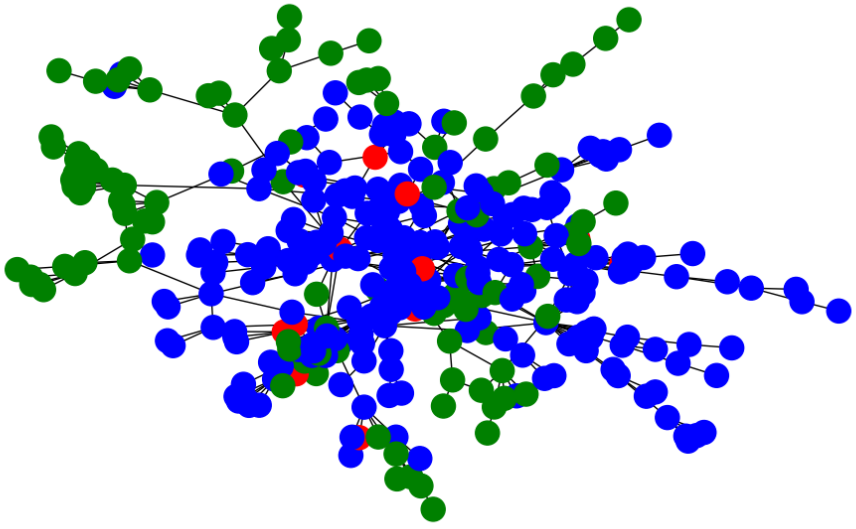

Figure 4

In Figure 5, the diffusion process is further along, with the message reaching a wider portion of the network. The abundance of green nodes demonstrates the successful

spread of information, although the specific version of the message held by each node may vary due to the actions of malicious actors.

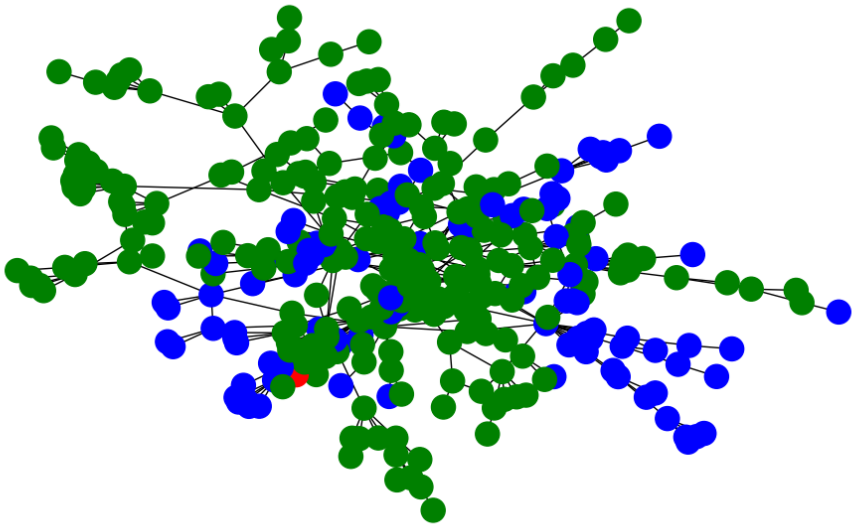

Figure 5

Finally, Figure 6 depicts the network near the end of the simulation. Most nodes have now received some version of the message, as shown by the predominance of green nodes. However, a small number of blue nodes remain, indicating that a few nodes did not receive any message due to their position in the network or the specific interactions during the simulation.

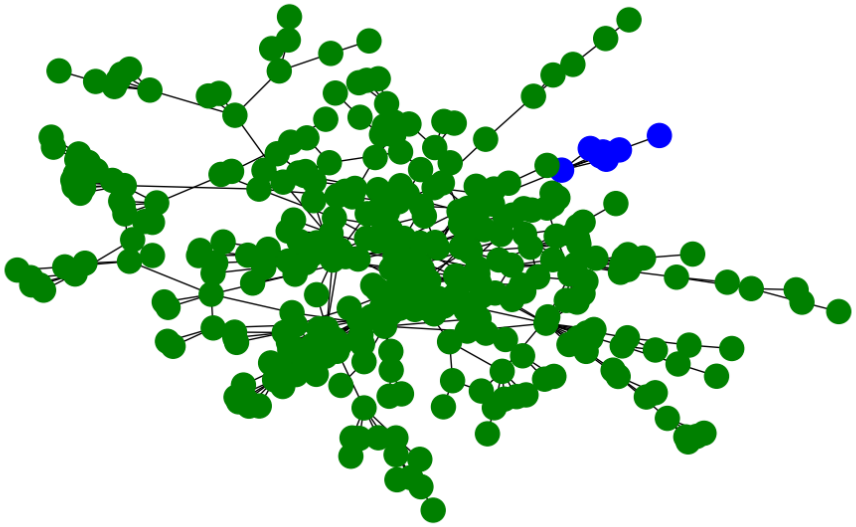

Figure 6

These visualizations collectively demonstrate the dynamic nature of information diffusion in a gossip protocol. They highlight the role of gossipers in initiating the spread, the influence of malicious nodes in introducing variations of the message, and the varying degrees of message penetration throughout the network.

Discussion

The simulation effectively demonstrated the rapid dissemination of information within a network, even with a relatively modest acceptance threshold of 0.2. In just eight iterations, the message reached 98% of the nodes, underscoring the efficiency of gossip protocols in propagating information.

However, the presence of malicious nodes significantly impacted the integrity of the message. With 20 malicious nodes, approximately 20.3% of the received messages were tampered with. This highlights the vulnerability of such protocols to misinformation and the potential for significant distortion of the original message as it spreads through the network.

The findings of this simulation resonate with real-world scenarios, particularly in the context of social networks like Twitter, which was the inspiration for the dataset used in this simulation. The rapid spread of information observed in the simulation mirrors the swift dissemination of news and messages on platforms like Twitter. Moreover, the impact of malicious nodes underscores the real-world issue of misinformation, where intentional or unintentional spread of false information can significantly distort the information landscape.

This simulation serves as a cautionary tale, emphasizing the need for robust mechanisms to verify information and mitigate the influence of malicious actors in online communication networks.

Conclustion

In conclusion, this simulation of a gossip protocol on a subset of the Twitter social circles network has revealed key insights into the dynamics of information dissemination and the impact of malicious actors. The gossip protocol proved to be highly efficient in spreading information, reaching almost all nodes within a few iterations. However, the presence of

malicious nodes significantly compromised the integrity of the message, with a substantial portion of nodes receiving altered versions.

This study underscores the inherent trade-off between the efficiency of gossip protocols and their vulnerability to misinformation. While these protocols excel at rapidly disseminating information, they are susceptible to manipulation by malicious actors, leading to the spread of diverse and potentially inaccurate message variants. This highlights the importance of developing robust mechanisms for verifying information and mitigating the influence of malicious actors in real-world communication networks.