

ELECTRONICS AND COMPUTER SCIENCE  
FACULTY OF PHYSICAL SCIENCES AND ENGINEERING  
UNIVERSITY OF SOUTHAMPTON

*Author*  
DANNY MARTINEZ HIBBERT

April 3, 2024

---

# **Final Report**

**Simulating social networks with user content bias to generate data for ml applications**

---

*Primary project Supervisor*  
Dr BOOJOONG KANG

*Secondary project Supervisor*  
Dr JIAN SHI

A final report submitted for the award of BSc Computer Science

## **Abstract**

This report presents a study aimed at advancing and improving OSN (online social networks) simulation methods to help in the analysis of botnet behaviour. The focus of the simulation method is through the use of a topic bias distribution to create targeted inauthentic node behaviour. The software project presented delves into the design and implementation of the base simulation model and its topic bias addition to shape user engagement within the network, similarly to a targeted spamming campaign.

Utilizing the SimSoM model as a foundational framework, this study introduces key modifications to incorporate topic bias, a novel feature that significantly enhances the model's realism and applicability in machine learning dataset generation.

The evaluation of the simulation tool demonstrates its effectiveness in replicating the complex interplay between authentic and inauthentic users within social networks, showcasing the impact of topic biases on information diffusion and user engagement. The findings highlight the tool's potential as a resource for researchers seeking to generate tailored datasets for studying the nuances of social media dynamics and bot influence.

### **Statement of Originality**

- I have read and understood the [ECS Academic Integrity](#) information and the University's [Academic Integrity Guidance for Students](#).
- I am aware that failure to act in accordance with the [Regulations Governing Academic Integrity](#) may lead to the imposition of penalties which, for the most serious cases, may include termination of programme.
- I consent to the University copying and distributing any or all of my work in any form and using third parties (who may be based outside the EU/EEA) to verify whether my work contains plagiarised material, and for quality assurance purposes.

***You must change the statements in the boxes if you do not agree with them.***

We expect you to acknowledge all sources of information (e.g. ideas, algorithms, data) using citations. You must also put quotation marks around any sections of text that you have copied without paraphrasing. If any figures or tables have been taken or modified from another source, you must explain this in the caption and cite the original source.

**I have acknowledged all sources, and identified any content taken from elsewhere.**

If you have used any code (e.g. open-source code), reference designs, or similar resources that have been produced by anyone else, you must list them in the box below. In the report, you must explain what was used and how it relates to the work you have done.

**I have not used any resources produced by anyone else.**

You can consult with module teaching staff/demonstrators, but you should not show anyone else your work (this includes uploading your work to publicly-accessible repositories e.g. Github, unless expressly permitted by the module leader), or help them to do theirs. For individual assignments, we expect you to work on your own. For group assignments, we expect that you work only with your allocated group. You must get permission in writing from the module teaching staff before you seek outside assistance, e.g. a proofreading service, and declare it here.

**I did all the work myself, or with my allocated group, and have not helped anyone else.**

We expect that you have not fabricated, modified or distorted any data, evidence, references, experimental results, or other material used or presented in the report. You must clearly describe your experiments and how the results were obtained, and include all data, source code and/or designs (either in the report, or submitted as a separate file) so that your results could be reproduced.

**The material in the report is genuine, and I have included all my data/code/designs.**

We expect that you have not previously submitted any part of this work for another assessment. You must get permission in writing from the module teaching staff before re-using any of your previously submitted work for this assessment.

**I have not submitted any part of this work for another assessment.**

If your work involved research/studies (including surveys) on human participants, their cells or data, or on animals, you must have been granted ethical approval before the work was carried out, and any experiments must have followed these requirements. You must give details of this in the report, and list the ethical approval reference number(s) in the box below.

**My work did not involve human participants, their cells or data, or animals.**

## **Acknowledgements**

Thank you to *Bao Tran Truong, Xiaodan Lou, Alessandro Flammini and Filippo Menczer* from the *Observatory on Social Media* for the great OSN model this project is based on[24].

# Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
1.1	Problem . . . . .	6
1.1.1	API access restrictions . . . . .	6
1.2	Goal . . . . .	6
1.3	Scope . . . . .	6
<b>2</b>	<b>Literature review</b>	<b>7</b>
2.1	Rogers' Diffusion of Innovations Theory . . . . .	7
2.2	Existing models, the SimSoM model . . . . .	7
2.2.1	SimSoM's key parameters . . . . .	7
2.3	Network generation and diffusion models . . . . .	8
2.3.1	Preferential attachment . . . . .	8
2.3.2	Triadic closure . . . . .	8
2.3.3	Homophily . . . . .	9
2.3.4	Reciprocity . . . . .	9
2.4	Agent generation . . . . .	10
2.4.1	Authentic agents . . . . .	10
2.4.2	Inauthentic users . . . . .	10
2.5	Topic exploitation for increase information diffusion . . . . .	10
<b>3</b>	<b>Planning and design</b>	<b>11</b>
3.1	Language and tools . . . . .	11
3.1.1	Github . . . . .	11
3.1.2	Lucid chart . . . . .	11
3.1.3	The code structure . . . . .	11
3.1.4	Module structure . . . . .	12
<b>4</b>	<b>Implementation</b>	<b>13</b>
4.1	Network . . . . .	13
4.2	Agents . . . . .	13
4.3	Messaging . . . . .	14
4.4	Topic bias system . . . . .	14
4.4.1	Topic distribution . . . . .	14
4.4.2	Topic similarity . . . . .	15
<b>5</b>	<b>Evaluation</b>	<b>15</b>
5.1	Message propagation . . . . .	15
5.2	Agent propagation . . . . .	15
5.3	Original base model without topic bias . . . . .	16
5.4	Base model with topic bias - Example 1 . . . . .	17
5.5	Base model with topic bias - Example 2 . . . . .	18
<b>6</b>	<b>Future work</b>	<b>21</b>
6.1	Improve existing model's metrics . . . . .	21
6.2	Improve topic bias layer . . . . .	21
6.3	Implement new complexity layer . . . . .	21
<b>7</b>	<b>Project management</b>	<b>22</b>
7.1	Implementation schedule report . . . . .	22
7.2	Software structure . . . . .	23
7.3	Gantt chart . . . . .	23
7.4	risk assessment . . . . .	24
<b>8</b>	<b>Bibliography</b>	<b>25</b>

# 1 Introduction

## 1.1 Problem

Online social media platforms are increasingly challenging to manage due to their expanding use and influence. The prevalence of fake profiles has escalated, allowing private interest groups to exploit these platforms [26]. Although there are some theoretical frameworks for simulating botnets within social networks, they fall short of addressing the complex interactions and dynamics among users.

There's a pressing need for researchers to delve into and understand the multifaceted aspects of social networks, particularly how their structures are manipulated by malicious actors using fake accounts. Yet, there is a noticeable absence of publicly accessible simulations of social platforms that examine the impact of content topics on user influence.

In addressing the challenge of bot detection, machine learning emerges as a cutting-edge solution with great results. Machine learning techniques analyse network and agent behaviour to detect inauthentic users and edge cases [25]. However, this approach is hampered by a dearth of datasets and significant issues with API access. A dedicated social network simulation tool could fill this gap by generating tailored datasets that meet specific research needs, thereby advancing our understanding of and ability to combat malicious activities on social media platforms.

### 1.1.1 API access restrictions

Over the recent years the most popular providers of OSN platforms such as Meta and twitter have severely restricted access to important data in regards to social trends on their platforms, instead favouring data that correlates to commercial purposes such as advertising reach [15].

Twitter is a very active area of research, their new API accessibility policies have changed, forcing many researchers to abandon their projects<sup>1</sup>. This decision was done to help the platforms sustainability<sup>2</sup>. Most researches spend a lot of time developing novel data extraction tools to circumvent the lack of API accessibility[7].

## 1.2 Goal

This project aims to alleviate this problem by designing a tool that simulates the effects of bots in a social network and how it shapes user engagement. The main purpose of the tool is to provide a more complex simulation that captures the underlying aspects of message re-sharing through the use of user topic interest.

The broader use case is to help researches in this field for the future by providing an accurate simulation and avoid relying on existing Private API's as well as avoid the time and financial bottleneck of sourcing data[1, 4, 8].

## 1.3 Scope

The key point of the simulation is to research the effects of inauthentic posts on user engagement through a simulation of topic bias, where each user in the network is biased towards engaging with certain topics.

The simulation is scaled realistically to avoid the dependencies on fast hardware, scalable solutions in OSN (online social networks) is a different problem all together [13]. A more in depth analysis of efficiency in the algorithms such as message propagation or topic comparison implementations is out of the scope of this project. The simulation will primarily use two key parameters to socialize the nodes with each other, engagement and quality of message/post.

There are primarily two types of social media networks, directional and bidirectional, this project focuses on directional networks, such as twitter based models. For example - when a user communicates with another user, it does not necessarily imply that the followed user will communicate back

---

<sup>1</sup><https://tinyurl.com/5a5spcyj>

<sup>2</sup><https://tinyurl.com/4x3my39t>

[14]. This unilateral nature of connections allows for a wide range of network structures, from highly reciprocal connections between users who follow each other to broad, non-reciprocal connections, like those between celebrities and their followers [21].

## 2 Literature review

This research will focus on the current implementations of OSN simulation models, how they are constructed as well as their common aspects and elements.

The field of OSNs is lacking in depth for the study of information diffusion, most only feature macro level analysis and do no focus on explaining individual user behaviour[5].

### 2.1 Rogers' Diffusion of Innovations Theory

Most OSNs base their network diffusion on Roger's diffusion theory. For instance, the adoption of features like Stories, Reels, or network formation algorithms can be analyzed through the lens of knowledge, persuasion, decision, implementation, and confirmation stages<sup>3</sup>. "Diffusion is the process by which an innovation communicated through certain channels over time among the members of a social system. Diffusion is special type of communication in which the messages are about a new idea" [16, 5].

### 2.2 Existing models, the SimSoM model

This project will use the model from Truong, B. T., et al (2023). Vulnerabilities of the Online Public Square to Manipulation[24]. The paper discusses the SimSoM model, an agent-based simulation designed to explore information diffusion in social networks infiltrated by malicious actors.

The SimSoM model is the most sophisticated agent-based simulation designed to analyze the dynamics of information diffusion within social network, with a particular focus on the influence of inauthentic accounts. This model is perfect for the needs of this project due to its focus on agent behaviour through the application of parameters that change its behaviour for flexibility and control.

#### 2.2.1 SimSoM's key parameters

SimSoM has a few key parameters that are used to manipulate the output of the simulation by altering parts of its behaviour such as agent behaviour and social structure. These parameters will be fully adopted into our model and the topic distribution bias will be added on top.

**Information Load ( $\mu$ ):** Represents the frequency at which new content is introduced into the system versus re shared content.

**Agent Attention ( $\alpha$ ):** This models the limited attention span of social media users by defining the size of a user's news feed.

**Engagement of Messages ( $e_m$ ):** Influences the likelihood of the message being re shared.

**Quality of Messages ( $q_m$ ):** Reflecting the objective desirable properties of content, such as originality or truthfulness.

**Prevalence of Bots ( $\beta$ ):** Parameters control the ratio of bots to authentic accounts.

**Infiltration of Bots ( $\gamma$ ):** To what degree have inauthentic accounts over taken the network in regards to connections to other nodes.

**Deception Parameter ( $\phi$ ):** This parameter models the likelihood of bot-generated content being irresistibly engaging, regardless of its quality.

---

<sup>3</sup><https://files.eric.ed.gov/fulltext/ED501453.pdf>

**Flooding ( $\theta$ ):** This parameter represents the extent to which bots can spam the network with messages. Equivalent to a spam attack.

## 2.3 Network generation and diffusion models

The basis of any OSN simulation model needs an appropriate user distribution to reflect a realistic network. Real social media networks with the usual expected functionality (where a user in the network can follow anyone else in the network) will inter-connect users following a power-law distribution[6]. In fact the most popular social media networks such as Facebook and twitter feature an almost exclusive power law distribution for their entire network [2].

Social platform simulation tools are severely limited in regards to commercially or open source software. This area of research mainly involves research applicants from statistical and psychology backgrounds - this is where this project derives a lot of its background information in regards to existing implementations.

Most social media networks tend to follow a power law distribution, especially in terms of connections (followers, friends) and content distribution (likes, shares). A power law distribution in this context means that a small number of users have a disproportionately large number of connections or interactions, while the vast majority of users have relatively few.

This phenomenon is often referred to as the "80/20 rule" or Pareto principle<sup>4</sup>, where roughly 20% of the users hold 80% of the connections or influence.

### 2.3.1 Preferential attachment

This is the most common explanation for power law distributions in social media networks. "The mechanism of preferential attachment assumes that a vertex's probability of receiving a new edge is proportional to the number of edges it already has." [22]. This is an interesting algorithm that has each new node in the network establish an edge with nodes that have the most edges already through a probability bias.

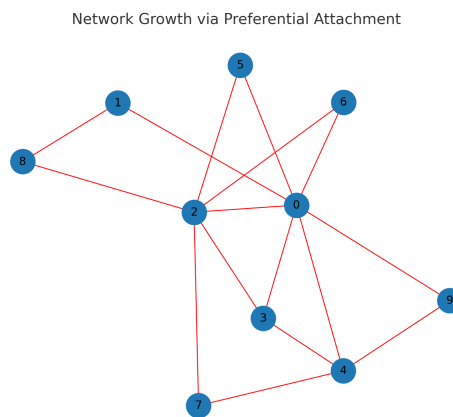


Figure 1: This graph was generated using the Barabási–Albert model, starting with a small number of interconnected nodes and then adding new nodes that preferentially attach to the more connected ones.

### 2.3.2 Triadic closure

This method of social network formation is also very popular for formulating simulated networks. It relies on the principle of association. Newly added nodes to the network are likely to establish edges with nodes if they have other nodes (friends) in common with each other [22, 3].

---

<sup>4</sup>[https://en.wikipedia.org/wiki/Pareto\\_principle](https://en.wikipedia.org/wiki/Pareto_principle)



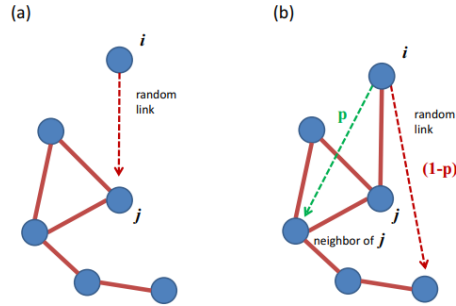


Figure 2: "Basic model. One link associated with a new node  $i$  is attached to a randomly chosen node  $j$ , the other links are attached to neighbors of  $j$  with probability  $p$ , closing triangles, or to other randomly chosen nodes with probability  $1 - p$ " [3]

### 2.3.3 Homophily

In this case, new nodes added to the network are more likely to form edges with nodes that share similar features. This is likely to cause clusters, for example reddit asks new users to join groups upon joining the network, creating clusters is what they are looking do to as the network operates on 'subreddit' clusters or groups[22, 18].

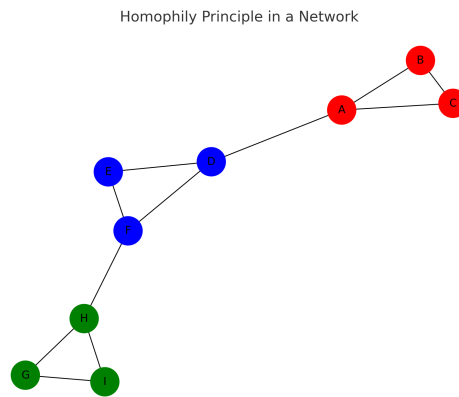


Figure 3: In this graph, nodes are more likely to connect with others within the same group, which is denoted by their color.

### 2.3.4 Reciprocity

Another phenomena that might explain the power law distribution in social networks is the theory of users reciprocating an edge formation in directed networks. If node A follows B, then B is more likely to be following A[22, 11].

In terms of popularity and current relevance, Preferential Attachment and Homophily are particularly prominent in explaining the dynamics of most large-scale social media networks, due to their focus on network growth and content sharing. Triadic Closure is also a significant model, especially in networks with a strong emphasis on personal relationships. Reciprocity is more specific to networks where mutual connections are a core feature.

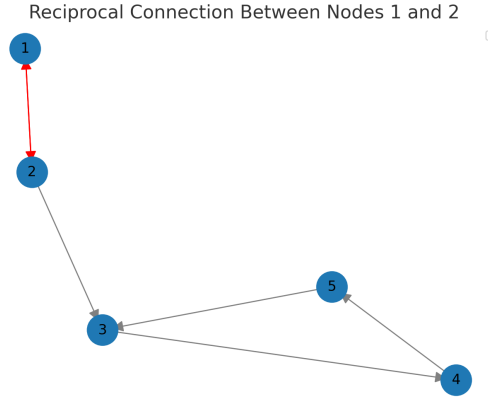


Figure 4: The red connection shows a reciprocal connection, node 1 is friends with node 2 through mutual interest, hence the bidirectional edge.

## 2.4 Agent generation

User agents within the OSN can vary greatly depending on the application of the simulation and type of data required. User interactions can be very specific to certain platforms that offer unique features as part of their service. Agents usually should be generic in nature in order to reflect users in different social media networks such as Facebook, twitter, tiktok, etc. Some of the agent behaviours of interest are based on a mix of active, rest, and sleep states, generation of novel messages and the re sharing of existing ones [9].

### 2.4.1 Authentic agents

A general assumption to make in regards to authentic users within the network is that they naturally join and interact based on their interests, forming connections and sharing content that reflects their real-life experiences[19]. This type of behaviour is fundamentally different from a user that is trying to subvert the network. Real users in the network engaging with the community and contributing to it naturally, no malicious activity or harm is intended[23]. The output of these nodes is generally considered to be of high quality.

### 2.4.2 Inauthentic users

These types of agents are strategically created and deployed to manipulate social media platforms. They often operate in networks, using both real and fake accounts to amplify specific narratives or silence opposition. Their behavior includes the coordination of content sharing, leveraging platform algorithms for greater visibility, and engaging in harassment campaigns against individuals or groups[19].

## 2.5 Topic exploitation for increase information diffusion

More sophisticated botnets within an OSN have used topic bias to increase the influence of their campaigns by tailoring produced content to complement current trends. These types of agents are defined by the Department of Social Sciences, University of Naples Federico II, Naples, Italy as CIB's (Coordinated Inauthentic Behavior). They strategically spread misinformation and manipulate public opinion by amplifying divisive narratives, particularly around COVID-19 vaccines. This exploitation of topic bias enhances information diffusion by engaging users more likely to share content that aligns with their pre-existing beliefs, thereby increasing the spread of inauthentic content across networks[20].

It has been observed that there's a higher activity of bot behaviour in OSN when scanning popular platforms such as Facebook and twitter [27] showing a tendency for bots to target topic bias. This is more visible on platforms such as Twitter that have trending topics through the hashtag system. Research presented in the Journal of Global Security Studies by Oxford Academic focused on the strategic use of social media networks by states for disinformation campaigns. It specifically exam-

ined how entities like RT and Sputnik, linked to the Kremlin, targeted networks of social media users by mimicking cultural cues to camouflage their intentions and attract sympathizers [17].

### 3 Planning and design

The main goal of this project is to implement the SimSoM[24] model and improve upon it by adding a layer of abstraction for topic bias to capture more realistic metrics for analysis purposes such as training machine learning models. Therefor this paper will implement the following key goals:

- **1.** Network generation method
- **2.** Agent generation for authentic and inauthentic nodes
- **3.** Inauthentic agent infiltration method
- **4.** Network manipulation through parameters
- **5.** Agent behaviour manipulation through parameters
- **6.** Message generation and re sharing method
- **7.** Topic bias algorithm
- **8.** Dashboard for user interaction

#### 3.1 Language and tools

The backend of the project is implemented using python3. The reason for choosing python is to take advantage of the rich graph and visual libraries - allowing the project to progress faster and preventing the need to write common algorithms from scratch. Another main reason for using python is to leverage its support for web interface integration. The frontend is used to display network information for the user to easily see any changes and developments in the network.

##### 3.1.1 Github

All code and utilities are kept in a github repository, this allows the code to be version controlled and managed appropriately. It also allows other developers to join the project and review code changes.

##### 3.1.2 Lucid chart

The Lucid chart service is used to provide a detailed UML diagram to visualise sections of the application - detailing the different constituent parts and their sub elements.

##### 3.1.3 The code structure

Python is a dynamic interpreted language that allows the use of object oriented programming as well as procedural programming, the project features both paradigms where appropriate.

The project is split into separate modules for increased flexibility:

#### Files

- **App.py** The entry point for the application, contains all globally defined parameters
- **data\_extraction.py** Returns dataset of entire network with its feature set
- **interface.py** Frontend code for displaying all visual representations

- **messaging.py** All logic to do with the way nodes send messages to eachother within the network
- **network.py** The logic behind the formulation of the network structure
- **plotter\_\*.py** All logic for generating plotted graph
- **topic\_distribution.py** Logic for generating topic biases for each node in the network

### Libraries

- **networkx** For generating graph data
- **dash + visdcc** For providing the frontend interface
- **numpy** For general math applications
- **plotly** For plotting data

#### 3.1.4 Module structure

The project is split into modules for better separation of concerns, their dependencies are shown as arrows. The **network** module is built first as its the base for all other data. The **messages** module is next as its dependent on the **network** module by generating messages for each timeline. The **plotter\_\*** library is very important for making sense of all the data in the network, from individual nodes to distributed messages. Finally **data\_extraction** and **interface** can be finalised independently. The very last module is the **topic\_distribution** which can be added to the system at the very end as an extra abstraction layer.

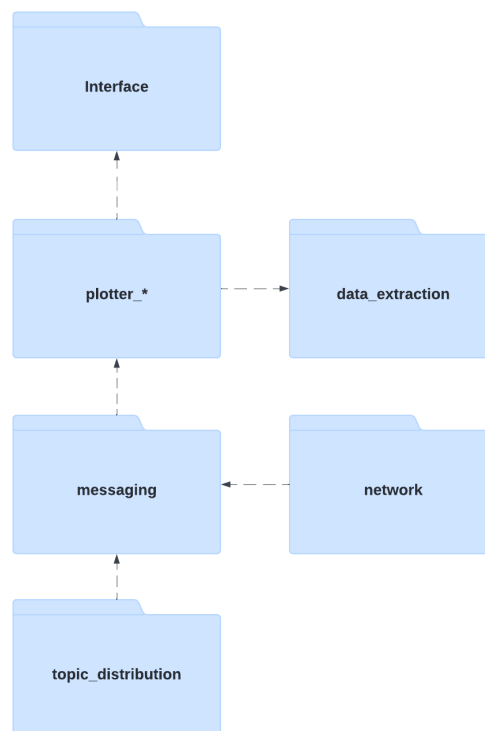


Figure 5: Gantt chart

## 4 Implementation

### 4.1 Network

The random-walk growth model leads to a network structure that exhibits properties of networks following a power-law distribution. In social networks, this means a small number of nodes (hubs) have many connections, while most nodes have few. This is characteristic of scale-free networks.

The preferential attachment mechanism, which is key to the emergence of scale-free networks, implies that the probability  $P(k)$  that a node in the network has  $k$  connections follows a power-law distribution:

$$P(k) \sim k^{-\gamma}$$

where  $\gamma$  is a parameter typically in the range  $2 < \gamma < 3$  for many real-world networks.

The growth model, through its iterative process where new nodes may attach to nodes based on existing connections (with probability  $p$ ) or link to any node in the network (with probability  $1 - p$ ), encourages a network structure with hubs and clustering. These features are indicative of the scale-free and small-world properties, respectively. Both properties can be described by power-law or similar heavy-tailed distributions in their degree distributions:

$$\text{For each edge } e_i : \begin{cases} \text{Select a friend of } j \text{ with probability } p, \\ \text{Select any node from the network with probability } 1 - p. \end{cases}$$

Here,  $j$  is a node already in the network that has been randomly selected as a starting point for the new edge. This process, while not explicitly preferential attachment, leads to a similar emergent behavior where the network develops characteristics of scale-free networks.

### 4.2 Agents

#### Bot Generation and Infiltration Parameters:

$\beta$ : Ratio of bots to authentic accounts in the network.

$\gamma$ : Probability that an authentic account follows a bot, modeling bot infiltration.

#### Engagement and Quality of Content for each agent:

For authentic accounts, engagement equals quality:

$$q = e$$

With high-quality information assumed to be rare. Quality and engagement are drawn from a distribution where,

$$P(e) = 2(1 - e)$$

For bots, all content is of low quality:

$$q_m = 0$$

but can have deceptively high engagement. This is modeled by the deception parameter  $\phi$ , where:

$$e_m = 1$$

with probability  $\phi$ , making some bot-generated content highly engaging.

### Spamming Behavior:

Parameter  $\theta$  indicates how much more content a bot generates compared to an authentic account, representing the bot's spamming behavior.

## 4.3 Messaging

### Message Selection and Sharing:

Probability of a message  $m$  being selected from a user's feed for sharing is proportional to its engagement:

$$P(m) = \frac{e_m}{\sum_{j \in M_i} e_j}$$

where  $M_i$  is the set of messages in user  $i$ 's feed.

### Effect on Information Quality:

The overall quality of the system at time  $t$  is given by:

$$Q_t = \frac{1}{\alpha N} \sum_{i=1}^N \sum_{m \in M_i} q_m^t$$

where  $q_m^t$ , item is the quality of message  $m$  in user  $i$ 's feed at time  $t$ ,  $N$  is the number of authentic accounts, and  $\alpha$  is the size of a user's feed.

### Cascade Sizes:

Reshare and exposure cascade sizes are analyzed to understand the spread and visibility of content, with reshare cascades initiated by posting a new message and exposure cascades measuring the potential views of a message.

## 4.4 Topic bias system

The topic bias feature in the network simulation is implemented by assigning topics to messages and nodes, then adjusting the likelihood of message creation, sharing, or engagement based on the alignment between a node's topics and the message's topic. This involves weighting topics differently to reflect their popularity/relevance to the simulated community, influencing how likely nodes are to engage with or spread messages related to those topics.

### 4.4.1 Topic distribution

- Let  $G = (V, E)$  represent the social network, where  $V$  is the set of nodes (users), and  $E$  is the set of edges (connections between users).
- Let  $T = \{t_1, t_2, \dots, t_k\}$  be the set of topics, with each topic  $t_i$  having an associated weight  $w_i$  that represents its popularity. The weights satisfy the condition  $\sum_{i=1}^k w_i = 1$  to form a probability distribution.
- For each node  $n \in V$ , the process of assigning a topic is modeled as a random selection based on the weights, mathematically given by:

$$P(\text{node } n \text{ is assigned topic } t_i) = w_i$$

- This means the probability that a node  $n$  is assigned topic  $t_i$  is equal to the weight  $w_i$ , ensuring that the distribution of topics among nodes reflects their relative popularity.

#### 4.4.2 Topic similarity

- The similarity between two topic distributions  $d_1$  and  $d_2$ , with a small constant  $\epsilon = 1e-10$  to ensure numerical stability, is calculated as:

$$\text{similarity} = \frac{d_1 \cdot d_2}{\|d_1\| \times \|d_2\| + \epsilon}$$

- where  $d_1 \cdot d_2$  is the dot product of the two distributions,  $\|d_1\|$  and  $\|d_2\|$  are the norms of the distributions, respectively, and  $\epsilon$  is a small constant to avoid division by zero.

## 5 Evaluation

The implementation of the base model in this project is successful and recreates the network power distribution, agent propagation and infiltration as well as message sharing and generation. The addition of a topic distribution layer is also successful and captures the case when botnets in a OSN focus on specific topics.

### 5.1 Message propagation

First the implemented model confirms the expected behaviour of authentic and inauthentic nodes. The base model produces a series of messages and over time the engagement of content increases as more spam populates the network and overruns the overall message quality of the network.



Figure 6: The average quality and engagement of all messages in the network

### 5.2 Agent propagation

The agent propagation implementation is also successful in the model. The distribution of nodes follow a power law distribution and the ratio of authentic to inauthentic nodes is controlled by beta and the amount of connections by gamma.

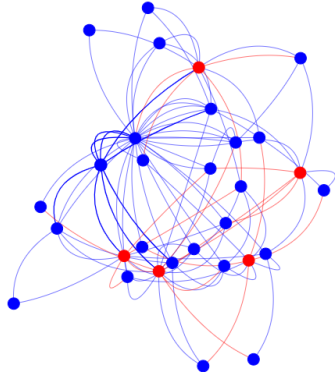


Figure 7:  $n = 25$ ,  $\beta = 0.7$ ,  $\gamma = 0.25$ ,  
flood factor = 1

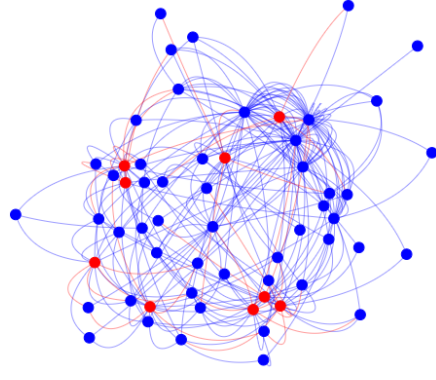


Figure 8:  $n=50$ ,  $\beta = 0.7$ ,  $\gamma = 0.25$ ,  
flood factor = 1

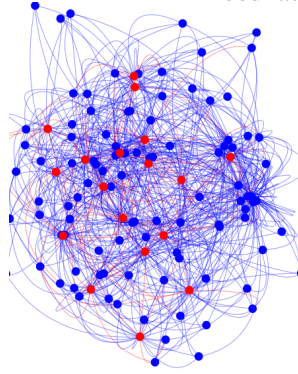


Figure 9:  $n=100$ ,  $\beta = 0.7$ ,  $\gamma = 0.25$ ,  
flood factor = 1

### 5.3 Original base model without topic bias

We can see the simulation doesn't produce messages based on topic type, users have a distribution of topics of interest similarly to real user behaviour but messages generated by bad actors are not targeting them. This is the base model behaviour.





Figure 10: Average engagement by topic with no bias - message topics are randomly generated

#### 5.4 Base model with topic bias - Example 1

In this case for the purpose of modeling a distribution of example topics have been propagated throughout the network for each node and newly generated message. The messages are generated and re-shared based on a distribution, topic featured are "tech", "planes", "right wing", "left wing" and "charity".

**Topic distribution** = {left\_wing: 0.4, right\_wing: 0.3, cars: 0.05, charity: 0.05, planes: 0.1, travel: 0.05, tech: 0.05}

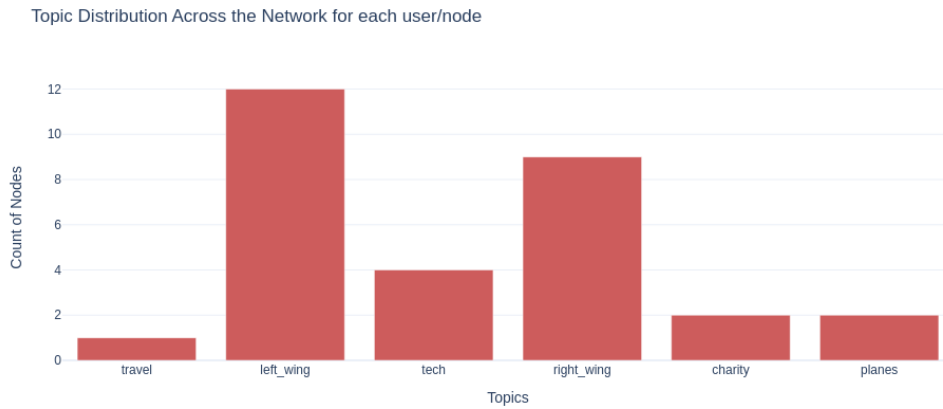


Figure 11: topic distribution for nodes in the network

The added topic bias mechanism generates multiple topics and assigns each one a probability of its appearance in the network for each node. Every node in the network is assigned a topic bias, which was not present previously.



Figure 12: Topic distribution across all messages in the network

Across all messages generated, we can see a topic majority following the defined distribution. In this case the the most popular topics are 'left wing' and 'right wing'. The least popular topics in this case 'travel' and 'planes' aren't discussed are prominently and do not receive a focus when generating messages.

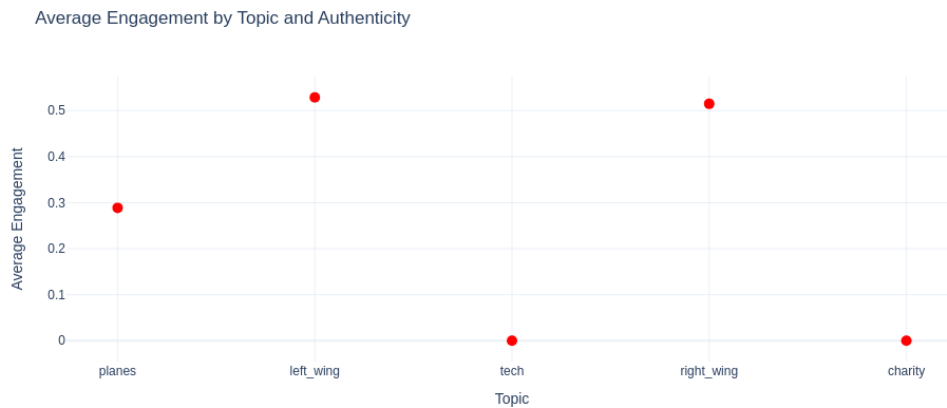


Figure 13: Average engagement by topic - with bias

With this system, the inauthentic nodes are targeting engagement based on certain topics. Nodes assigned the most popular topics are targeted the most - receiving messages with higher engagement after analysing topic similarity.

## 5.5 Base model with topic bias - Example 2

Another example of inauthentic nodes targeting produced messages towards authentic accounts, this time there's a larger topic distribution.

**Topic distribution** = { tv\_shows: 0.06, architecture: 0.08, jewelry: 0.012, weather: 0.10, politics: 0.22, education: 0.012, hacking: 0.16, property: 0.08, dating: 0.06 }

Figure below shows all the topics and their distribution across nodes in the network.

Topic Distribution Across the Network for each user/node

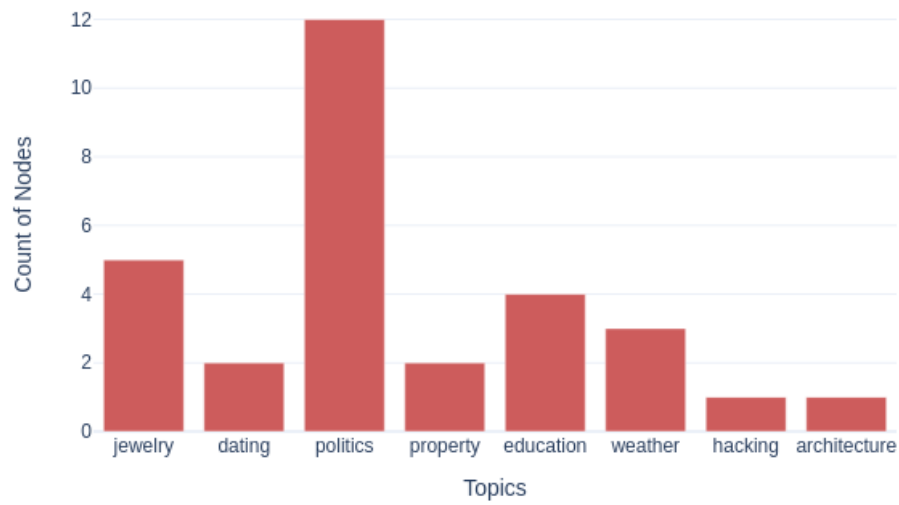


Figure 14: Topic distribution for nodes in the network - example 2

The following figure shows messages produced by inauthentic nodes targeting the most popular topics within the network. Due to the probability distribution, the most uncommon topics are dropped entirely - this behaviour is also a by product of simulating a smaller network (less nodes)

Distribution of Topics Across All Messages

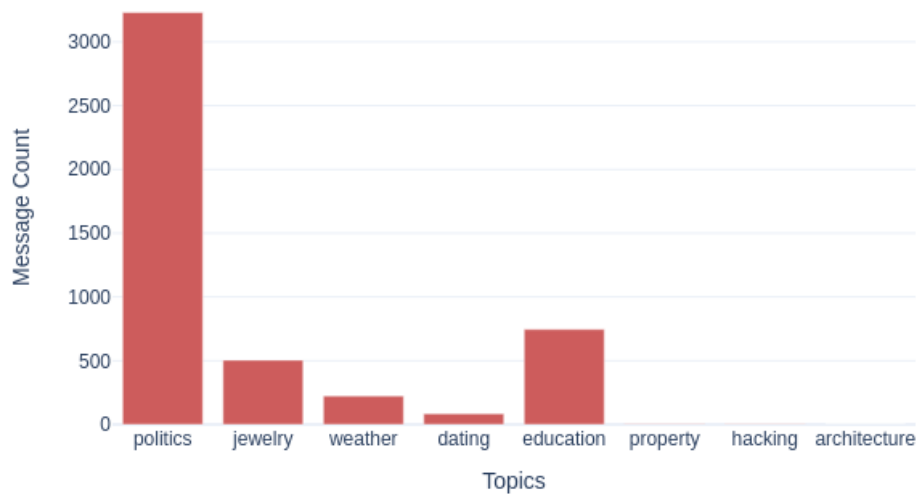


Figure 15: Figure 9: Topic distribution across all messages in the network - example 2

The final graph below shows the engagement increase for the most popular topics in accordance to the targeted topics. Inauthentic nodes are drumming up engagement through the flooding of high engagement content

Average Engagement by Topic

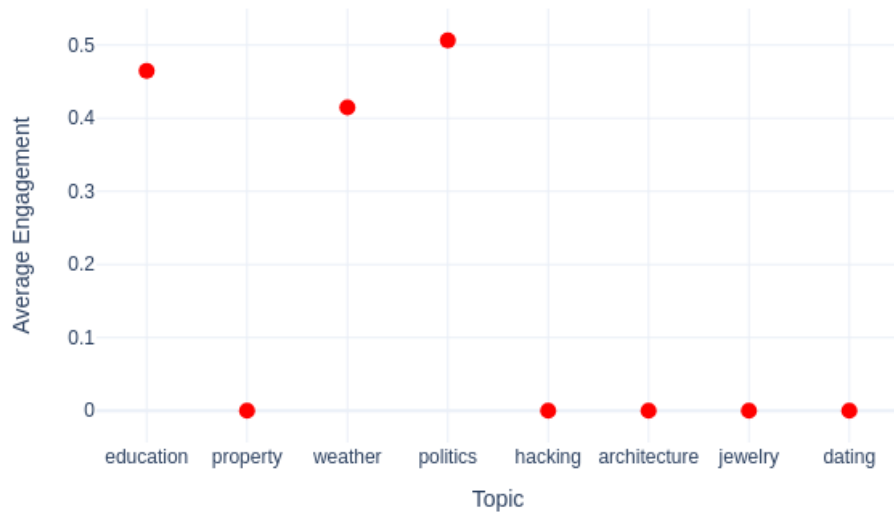


Figure 16: Figure 9: Topic distribution across all messages in the network - example 2

## 6 Future work

More work could be done in regards to different aspects of this project. This thesis focuses on the addition of a layer of complexity to an existing base model, in regards to potential areas from improvement the current base model can still be improved, improvements can be made to the new layer of complexity or engineer another layer of complexity to capture different metrics.

### 6.1 Improve existing model's metrics

The base model features a good baseline for simulating inauthentic agent infiltration through pooling. Specific behaviours can be simulated to explore different attack vectors and techniques such as discussed by Gan. C et. al [10].

### 6.2 Improve topic bias layer

The topic bias layer added in this thesis can be expanded upon through a more complex implementation of the user agents but creating different tiers of related personal interests as presented in [28]. This way each agent may capture the behaviour of human interest more accurately.

### 6.3 Implement new complexity layer

Once again the base model (SimSoM) provides a good basis for message sharing behaviour within the network, but a layer of complexity can be included for a more dynamic approach as currently basic model assumes all nodes will communicate with a distribution of any other connected node. The following paper by Gatti. M et al in 2014 [12] discusses in great detail complex methods of simulating user behaviour.

## 7 Project management

### 7.1 Implementation schedule report

Task	Difficulty Expected	Actual Difficulty	Reason
Network generation	Medium	Medium	Implementing the initial network generation was easier than expected
Agent generation	Large	Medium	Due to available libraries and frameworks, agent generation complexity was reduced
Agent behaviour	Large	Large	The complexity of modeling realistic agent behavior met expectations
Message generation	Medium	Large	Unanticipated challenges arose in generating diverse and realistic messages
Topic bias	Large	Medium	Initial concerns over modeling topic bias were mitigated with refined algorithms
UI	Small	Small	UI development proceeded as planned without significant issues

## 7.2 Software structure

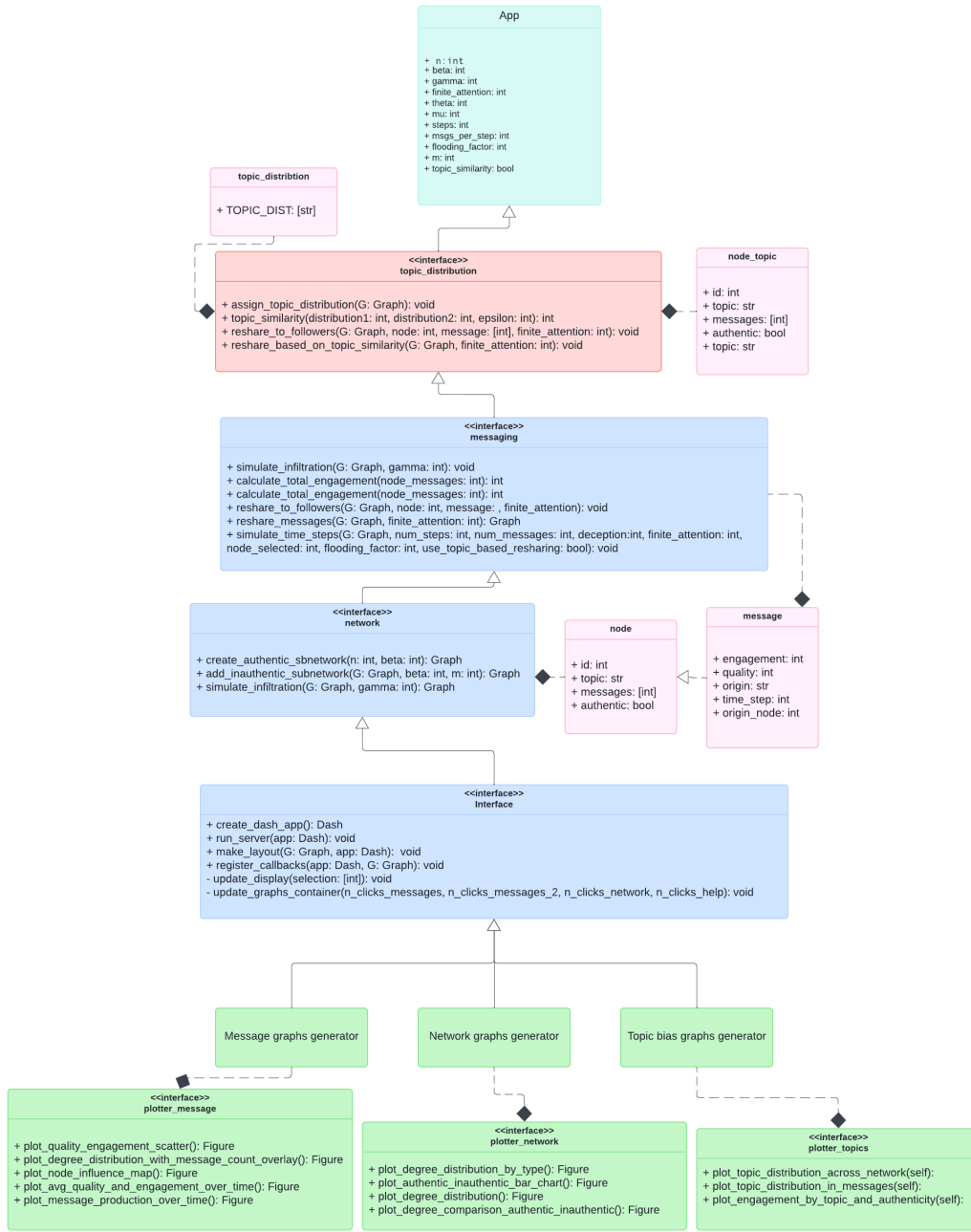


Figure 17: Gantt chart

## 7.3 Gantt chart

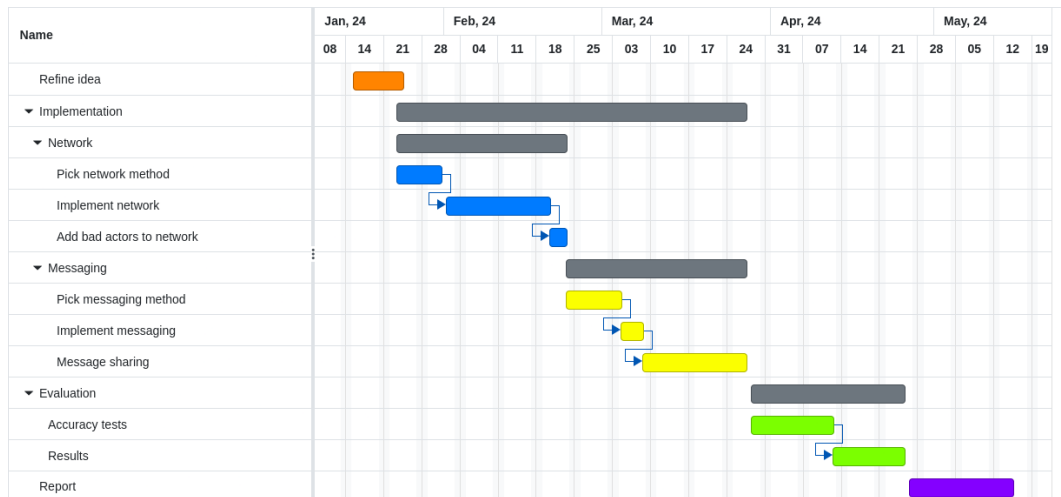


Figure 18: Gantt chart

## 7.4 risk assessment

Risk	Probability	Severity	Risk exposure	Mitigation
Underestimating task	3	4	12	Set a clear and detailed outline.
Implementation complexity	4	4	16	Take time to understand the design and focus on critical infrastructure.
Poor health	2	5	10	Implement critical sections first to mitigate impact of lost time.
Scope creep	3	3	9	Finish implementations before moving onto other tasks.
Resource availability	2	2	4	Contact ECS for extra machines.
Requirements change	1	5	5	Ensure initial requirements meet the problem criteria.
Data loss	1	6	6	Back up data onto cloud services or physical drives.



## 8 Bibliography

### References

- [1] Awrad Mohammed Ali et al. “Synthetic generators for cloning social network data”. In: *Proceedings of SocInfo* (2014).
- [2] Sumit Kumar Banshal et al. “Power Laws in altmetrics: An empirical analysis”. In: *Journal of Informetrics* 16.3 (2022), p. 101309. ISSN: 1751-1577. DOI: <https://doi.org/10.1016/j.joi.2022.101309>. URL: <https://www.sciencedirect.com/science/article/pii/S175115772200061X>.
- [3] Ginestra Bianconi et al. “Triadic closure as a basic generating mechanism of communities in complex networks”. In: *Phys. Rev. E* 90 (4 Oct. 2014), p. 042806. DOI: 10.1103/PhysRevE.90.042806. URL: <https://link.aps.org/doi/10.1103/PhysRevE.90.042806>.
- [4] Axel Bruns. “After the ‘APIcalypse’: social media platforms and their fight against critical scholarly research”. In: *Information, Communication & Society* 22.11 (2019), pp. 1544–1566. DOI: 10.1080/1369118X.2019.1637447. eprint: <https://doi.org/10.1080/1369118X.2019.1637447>. URL: <https://doi.org/10.1080/1369118X.2019.1637447>.
- [5] Maira A de C Gatti et al. “A simulation-based approach to analyze the information diffusion in Microblogging Online Social Network”. In: *2013 Winter Simulations Conference (WSC)*. 2013, pp. 1685–1696. DOI: 10.1109/WSC.2013.6721550.
- [6] Gábor Csányi and Balázs Szendrői. “Structure of a large social network”. In: *Phys. Rev. E* 69 (3 Mar. 2004), p. 036131. DOI: 10.1103/PhysRevE.69.036131. URL: <https://link.aps.org/doi/10.1103/PhysRevE.69.036131>.
- [7] Brian Dopson, Cardavian Lowery, and Deepti Joshi. “Collection and analysis of social media datasets”. In: *J. Comput. Sci. Coll.* 30.2 (Dec. 2014), pp. 254–261. ISSN: 1937-4771.
- [8] Shangbin Feng et al. “TwiBot-20: A Comprehensive Twitter Bot Detection Benchmark”. In: *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. ACM, Oct. 2021. DOI: 10.1145/3459637.3482019. URL: <https://doi.org/10.1145/3459637.3482019>.
- [9] Alceu Ferraz Costa et al. “RSC: Mining and Modeling Temporal Activity in Social Media”. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’15. Sydney, NSW, Australia: Association for Computing Machinery, 2015, pp. 269–278. ISBN: 9781450336642. DOI: 10.1145/2783258.2783294. URL: <https://doi.org/10.1145/2783258.2783294>.
- [10] Chenquan Gan et al. “Advanced Persistent Threats and Their Defense Methods in Industrial Internet of Things: A Survey”. In: *Mathematics* 11.14 (2023). ISSN: 2227-7390. DOI: 10.3390/math11143115. URL: <https://www.mdpi.com/2227-7390/11/14/3115>.
- [11] Diego Garlaschelli and Maria I. Loffredo. “Patterns of Link Reciprocity in Directed Networks”. In: *Phys. Rev. Lett.* 93 (26 Dec. 2004), p. 268701. DOI: 10.1103/PhysRevLett.93.268701. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.93.268701>.
- [12] Maíra Gatti et al. “Large-scale multi-agent-based modeling and simulation of microblogging-based online social network”. In: *Multi-Agent-Based Simulation XIV: International Workshop, MABS 2013, Saint Paul, MN, USA, May 6-7, 2013, Revised Selected Papers*. Springer. 2014, pp. 17–33.
- [13] Bonan Hou et al. “Modeling and simulation of large-scale social networks using parallel discrete event simulation”. In: *SIMULATION* 89.10 (2013), pp. 1173–1183. DOI: 10.1177/0037549713495752. eprint: <https://doi.org/10.1177/0037549713495752>. URL: <https://doi.org/10.1177/0037549713495752>.
- [14] Mohsen Jamali, Gholamreza Haffari, and Martin Ester. “Modeling the temporal dynamics of social rating networks using bidirectional effects of social relations and rating patterns”. In: *Proceedings of the 20th International Conference on World Wide Web*. WWW ’11. Hyderabad, India: Association for Computing Machinery, 2011, pp. 527–536. ISBN: 9781450306324. DOI: 10.1145/1963405.1963480. URL: <https://doi.org/10.1145/1963405.1963480>.

- [15] Andreas Birkbak Jessamy Perriam and Andy Freeman. “Digital methods in a post-API environment”. In: *International Journal of Social Research Methodology* 23.3 (2020), pp. 277–290. DOI: 10.1080/13645579.2019.1682840. eprint: <https://doi.org/10.1080/13645579.2019.1682840>. URL: <https://doi.org/10.1080/13645579.2019.1682840>.
- [16] N Gizem Kocak, Seçil Kaya, and Evrim Erol. “Social media from the perspective of diffusion of innovation approach”. In: *The Macrotheme Review* 2.3 (2013), pp. 22–29.
- [17] Tobias Lemke and Michael W Habegger. “Foreign Interference and Social Media Networks: A Relational Approach to Studying Contemporary Russian Disinformation”. In: *Journal of Global Security Studies* 7.2 (Apr. 2022), ogac004. ISSN: 2057-3170. DOI: 10.1093/jogss/ogac004. eprint: <https://academic.oup.com/jogss/article-pdf/7/2/ogac004/43510245/ogac004.pdf>. URL: <https://doi.org/10.1093/jogss/ogac004>.
- [18] Yury A. Malkov and Alexander Ponomarenko. “Growing Homophilic Networks Are Natural Navigable Small Worlds”. In: *PLOS ONE* 11.6 (June 2016), pp. 1–14. DOI: 10.1371/journal.pone.0158162. URL: <https://doi.org/10.1371/journal.pone.0158162>.
- [19] Nikolaos Mavridis. “Artificial agents entering social networks”. In: *A Networked Self*. Routledge, 2010, pp. 299–311.
- [20] Monica Murero. “Coordinated inauthentic behavior: An innovative manipulation tactic to amplify COVID-19 anti-vaccine communication outreach via social media”. In: *Frontiers in Sociology* 8 (2023). ISSN: 2297-7775. DOI: 10.3389/fsoc.2023.1141416. URL: <https://www.frontiersin.org/articles/10.3389/fsoc.2023.1141416>.
- [21] David F. Nettleton. “Data mining of social networks represented as graphs”. In: *Computer Science Review* 7 (2013), pp. 1–34. ISSN: 1574-0137. DOI: <https://doi.org/10.1016/j.cosrev.2012.12.001>. URL: <https://www.sciencedirect.com/science/article/pii/S1574013712000445>.
- [22] Andrew T. Stephen and Olivier Toubia. “Explaining the power-law degree distribution in a social commerce network”. In: *Social Networks* 31.4 (2009), pp. 262–270. ISSN: 0378-8733. DOI: <https://doi.org/10.1016/j.socnet.2009.07.002>. URL: <https://www.sciencedirect.com/science/article/pii/S0378873309000367>.
- [23] Stefan Stieglitz et al. *Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts*. 2017. arXiv: 1710.04044 [cs.HC].
- [24] Bao Tran Truong et al. *Quantifying the Vulnerabilities of the Online Public Square to Adversarial Manipulation Tactics*. 2023. arXiv: 1907.06130 [cs.CY].
- [25] Alex Hai Wang. “Detecting Spam Bots in Online Social Networking Sites: A Machine Learning Approach”. In: *Data and Applications Security and Privacy XXIV*. Ed. by Sara Foresti and Sushil Jajodia. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 335–342. ISBN: 978-3-642-13739-6.
- [26] Mudasir Ahmad Wani, Muzafar Ahmad Sofi, and Suheel Yousuf Wani. “Why Fake Profiles: A study of Anomalous users in different categories of Online Social Networks”. In: *International Journal of Engineering, Technology, Science and Research* 4 (2017), pp. 320–329.
- [27] Kurt Wirth, Ericka Menchen-Trevino, and Ryan T. Moore. “Bots By Topic: Exploring Differences in Bot Activity by Conversation Topic”. In: *Proceedings of the 10th International Conference on Social Media and Society*. SMSociety ’19. Toronto, ON, Canada: Association for Computing Machinery, 2019, pp. 77–82. ISBN: 9781450366519. DOI: 10.1145/3328529.3328547. URL: <https://doi.org/10.1145/3328529.3328547>.
- [28] Junkai Zhou et al. *Knowledge Boundary and Persona Dynamic Shape A Better Social Media Agent*. 2024. arXiv: 2403.19275 [cs.CL].