Chapter 1

Introduction

Social networks have been a popular medium for communication, networking, and sharing of opinions and ideas. With traditional media, the reach of social networks was limited to a smaller audience and limited geographical boundaries. However, with the emergence of Web 2.0 in the early 2000s, the Internet allowed user-generated content through online social networks (OSNs), popularly known as social media. It has enhanced the definition of the Internet as the "links between people" in addition to the established Web's definition as the "links between documents" [1]. Facebook, Twitter (now X), and Instagram are some popular examples of social media platforms.

With an increasing shift towards ubiquitous computing, a vast section of the population is generating content at a very rapid rate. According to a report[1], X has 330 million monthly active users worldwide. Similarly, around 1.82 billion people log onto Facebook daily and are considered active daily users[2]. Most of the content on these platforms is informative and helpful to people in daily life as it gives easy access to the news and allows open communication and information gathering [2,3]. However, due to the irresponsible behavior of a community of users, these platforms have also become active propagators of rumors, fake news, misinformation, hate

[1]https://www.statista.com/statistics/274564/monthly-active-twitter-users-in-the-united-states

posts, shame posts, and other negative content. In this thesis, we attempt to deal with a few issues and challenges related to rumors.

There are many definitions of rumor; DiFonzo and Bordia give the most comprehensive definition [4]; rumors are unverified and instrumentally relevant information in circulation. A rumor is an item of circulating information whose veracity status has not yet been verified at the time of posting [5]. These definitions are in accordance with the definitions provided by Oxford Dictionary[3] and Merriam-Webster Dictionary[4]. Apart from rumor, i.e. news without reference, and non-rumor, i.e., news with reference, Sicilia et al. [6] have introduced a new unknown class, which is potentially true news without reference. We use the definition proposed by Zubiaga et al. [7], that is, rumors are unverified pieces of information whose veracity status is unknown at the time of posting throughout this thesis.

Based on various definitions of rumors, we identified some characteristics of rumors, which are as follows.

1. The veracity status of a rumor is unknown at the time of circulation. Veracity may be *true* or *false* or remain *unknown* (in case of long-standing rumors).

2. There is no official responsibility for a rumor, so no one can be held accountable.

3. The topic of a rumor is often related to some emerging news or interesting pieces of stories related to popular events to attract a large number of people and thus spread vigorously.

4. They propagate in the social network either due to an information gap, when the information is released in pieces, or lack of authentic sources of information, or are deliberately fabricated to attract people to believe them.

5. Rumors spread faster than non-rumors.

[3]https://www.lexico.com/en/definition/rumor
[4]https://www.merriam-webster.com/dictionary/rumor

6. Rumors create restlessness and panic in society or are designed purposely to create panic, damage reputation, etc.

The impact of rumors can include, but is not limited to, panic among the masses, threat to reputation, and economic losses [8–11]. For example, on 31 January 2020, a rumor spread on the Chinese microblogging website Sina Weibo that the oral solution of the Chinese patent medicine Shuanghuanglian can inhibit the novel coronavirus COVID-19. Pictures of people queuing outside the medical stores were widely circulated. After the rumor spread on social media, people rushed to buy the medicine offline and online, which left Shuanghuanglian products out of stock. This caused a situation of panic among people and irrational behavior among netizens [12]. So, it is very important to understand rumors and ways to deal with them.

Research into addressing rumors spans a wide range of areas and significant issues, including:

1. Differentiating between a rumor and a non-rumor and developing methods for rumor detection.
2. Determining the veracity and truthfulness of a rumor.
3. Identifying the source or origin of a rumor.
4. Preventing the spread of rumors within social networks.
5. Analyze the impact and damage caused by a rumor.

In this thesis, we specifically focus on two of these issues: identifying the initiator of a rumor and devising strategies to prevent its spread within a social network. This chapter introduces these two issues in Sections 1.1 and 1.2, respectively. In addition, we discuss significant challenges related to these issues in Section 1.3 and motivation to find solutions to these issues in Section 1.4. Section 1.5 lists the objectives of this thesis, and Section 1.6 lists our contributions to achieve those objectives. Lastly, Section 1.7 shows the general structure of the thesis.

1.1 Rumor Source Identification

The rumor source identification problem involves identifying the origin of a rumor or piece of information on a social network. This problem lies within the broader field of information diffusion and network analysis. The goal is to identify the origin or initial source of a rumor and to analyze the dynamics of an information cascade in a social network. This is crucial for controlling the spread of rumors, understanding how rumors spread, and mitigating potential negative impacts.

Social networks are often modeled as graphs $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_n\}$ is the set of nodes (vertices) in the network and E is the set of edges, where each edge $(v_i, v_j) \in E$ represents a connection or relationship between nodes v_i and v_j. Rumors are assumed to propagate through the network from the source to its neighbors, who then pass it on to their neighbors, and so on. A rumor propagation model $\mathcal{P}$ defines how the rumor spreads through the network from one node to another. It can be epidemic model, independent cascade model (IC), or more complex variants of these. A subset of nodes $I \subseteq V$, called the *infected nodes* are the nodes influenced by rumors at a certain point in time. The objective of the rumor source identification problem is to identify one or more nodes $s \in V$ (referred to as the *source node* or *initiator*) that maximizes the likelihood of being the origin of the observed infection pattern I, in the propagation model $\mathcal{P}$ [13,14]. Formally, let $\mathcal{L}(s \mid I, \mathcal{P}, G)$ be a likelihood function that gives the probability that node $s \in V$ is the source of the rumor, the goal is to find:

$$s^* = \arg\max_{s \in V} \mathcal{L}(s \mid I, \mathcal{P}, G) \tag{1.1}$$

where s^* is the node (or nodes) in V that maximizes the likelihood of being the source of the rumor.

Identifying the source of the rumor is challenging for several reasons. The large size and complex structure of real-world social networks including loops, clusters,

etc. make the problem computationally hard. In some cases, there might be more than one source of the rumor, further complicating the detection process.

1.2 Rumor prevention and control

OSNs have become a breeding ground for the massive spread of rumors. Therefore, preventing and controlling rumors on social networks during crises, such as natural disasters, political unrest, or pandemics, is crucial, as they can severely affect social media users, as discussed in various scenarios [15]. The problem of rumor prevention and control in social networks focuses on strategies and methods to prevent the spread of harmful rumors or mitigate their impact once they spread. Similarly to the rumor source detection problem, rumor propagation in social networks is also modeled using graphs where nodes represent individuals and edges represent their connections. Rumors are modeled to spread from one node to another following specific probabilistic or deterministic rules, such as those defined by epidemic models or their variants. Researchers' communities rely on one of the two following methods to prevent and control rumor propagation and percolation into a social network.

Containment: Once a rumor starts spreading, containment strategies aim to limit its spread by isolating or blocking connections between infected nodes and the rest of the network. This blocking occurs at both the node level [16–18] and the link level [19]. In addition, detecting potential sources of rumors early and preventing them from spreading by blocking or neutralizing these sources [20].

Counter-Rumor Propagation: Another approach involves spreading counter information to negate the effects of the rumor. This involves circulation of a counter-rumor message denying the rumor or the correct news. This is often done by some authority, a responsible person, fact checkers, or news agencies. They often provided the correct information about the rumor with evidence or links to evidence. This

kind of rumor control uses both the self-media-based perspective [21, 22] and the diffusion model-based approach [23–27].

This problem involves optimizing objectives, such as minimizing the number of nodes infected by the rumor or maximizing the effectiveness of counter-rumor methods.

1.3 Major challenges faced in dealing with rumors

1.3.1 Challenges in identifying rumor initiators

Finding the people involved in the spreading of rumors is a difficult task. Often these people are left unidentified and face no legal consequences, which encourages rumor-initiating behavior in society even more. So, there is a growing need to identify these people in the network. Several attempts have been made to detect the rumor initiators on social media, focusing on the network structure and considering rumor initiator detection as a maximum likelihood problem. Many proposed solutions are limited to degree regular trees [13], star networks [14], regular degree graphs with a single cycle [28], and randomly increasing trees [29], including binary search trees, recursive trees, plane-oriented recursive trees, and undirected trees. However, in real life, social network structures are not that simple. Millions of users are linked to each other with billions of links and link cycles. So, these solutions are less effective on real-world social networks. In addition, these networks only consider social networks as nodes and links between nodes and do not consider the node properties, link properties, and other semantics of the networks. The other problem is that social networks are structurally the same but semantically different [30]. Facebook has its definition of the user as a node connected to another node via a symmetric link, while X has its definition of the user as a node connected to another node via a directed link. There is no consensus on guidelines for the development of OSNs [31]. So, if the same person spreads a rumor on different OSNs, then we have to find the

rumor initiator on all of the OSNs separately. Thus, there is a duplication of effort in finding the same rumor initiator.

1.3.2 Challenges in preventing rumors

Rumor prevention and control on social networks is challenging due to the complex structure of networks, uncertainty about spread, and non-linear propagation of rumors. Existing research has provided solutions for the prevention and control of rumors in social networks. These solutions are based on self-media perspectives [21,22] as well as diffusion control models. In self-media-based approaches, establishing the credibility of self-media is difficult. In addition, subjective judgments can lead to inaccurate rumor debunking and are prone to echo chambers and confirmation bias. These factors make it challenging to efficiently debunk rumors and assess the effectiveness of prevention and control efforts. Therefore, formal rumor prevention and control models are needed to accurately measure the extent of rumors and evaluate rumor prevention strategies.

The rumor diffusion control models are based on two strategies- First, blocking the nodes and edges that are responsible for rumor spread [16, 17, 19, 20, 32]. These methods were able to control rumors to an extent; however, they lacked the perspective of users' interest in rumors. It is often the case that a person is interested or not interested or only interested upto some extent in a particular rumor. These interests are guided by a person's age, occupation, location, topics, and many more. Some researchers have considered the interest of users to block the rumor [33,34] but they have considered interest as a factor of age, location, gender, etc. These factors are important to consider when trying to block rumors. However, one important factor that is not considered is the user's interest in the topic of the rumor and its role in the network diffusion.

The second strategy to control rumor diffusion is by providing a counter-rumor

diffusion mechanism [18, 23–26, 35]. These have provided significant research directions to address the problem; however, we have identified a few research gaps. First, most existing research focuses on modeling and analyzing how rumors spread and how they can be controlled after significant dissemination. However, relatively few studies emphasize proactive strategies to prevent the spread of rumors before they become widespread. Second, a key challenge that arises in rumor management is the issue of time lag. Specifically, determining how long the system remains in an 'infected' state (that is, where the rumor is actively spreading) before initiating a counter-rumor diffusion process. This timing is critical because it directly influences the ultimate state of the system, which is either a preventive state or a controlled state. A preventive state is achieved when the spreading of counter-rumor begins early enough to stop the rumor before it spreads to a large portion of the network. In this state, the rumor does not gain momentum, user exposure is minimal, and public trust is less likely to be compromised. Preventive states are ideal because they contain the problem early using fewer resources and cause less damage to the public. They are particularly valuable in high-stakes scenarios, such as public health emergencies or political elections, where false rumors can rapidly lead to harmful real-world consequences. On the other hand, a controlled state occurs when the counter-rumor process is initiated only after the rumor has already spread significantly within the network. In this scenario, efforts are focused on damage control by limiting further spread, correcting rumor, and restoring credibility. While the spread of the rumor is eventually curbed, the controlled state implies that harm (e.g., confusion, panic, social division) has already occurred to some degree. Furthermore, interventions at this stage tend to be more resource intensive, requiring widespread dissemination of corrective information, collaboration with authorities, and sometimes even platform-level interventions (e.g., content moderation, account suspension). Third, most researchers have advocated that the high trustability or credibility of the government or authorities helps to counterfeit the rumor. However, when the public lacks trust in the authorities, the process to counterfeit the rumor is not much effective. So, there is a problem of distrust in authorities by the public.

1.4 Motivation for the thesis

1.4.1 Motivation for rumor initiator detection

To address the challenges in section 1.3.1, we propose an ontology-based solution to find the rumor initiator in the OSNs. Ontology is widely used in the social network domain for various purposes such as modeling P2P sharing of annotations [36], modeling social tagging mechanisms and affiliation networks [37], modeling security and privacy concerns in OSNs [38], rumor modeling [39], modeling rumor detection mechanism [40] etc. The reason to choose ontology for modeling OSNs is manifold. Firstly, ontology lays the foundation for the knowledge graph of OSNs which can accommodate real-life social networks with millions of nodes and billions of links and link cycles. Secondly, ontology provides semantics to the data by annotating meta-data. The proposed ontology model is a generic domain ontology model that semantically equates all OSNs and builds a consensus on concepts among them. This is very helpful in finding *one solution for all* and provides possible interoperability. Third, ontology is empowered with reasoning abilities that can be exploited to answer domain-related queries for intended purposes, i.e. finding rumor initiators in OSNs *a posteriori*, finding all OSNs where rumor is initiated, etc.

1.4.2 Motivation for rumor prevention models

Motivation for rumor blocking model- Rumors are pieces of stories. These are related to some event or person. Events or persons are usually associated with some topic or an overlap of many topics. In social networks, people are intended towards some topics depending upon their likes or dislikes. For example, a person can be a fan of sports and avoid politics. In addition, people can be interested in more than one topic, more or less. This assumption is very obvious and depicts the real world. In [41], the authors have briefly described various methods and techniques to extract users' interest from microblogging websites. These interests are used to find whether a particular post is recommended to a user depending on his/her interests.

The mining of user interests is highly popular in recommender systems where some news items or items are recommended to users based on their profile, past history, or reviews and ratings [42,43]. Information diffusion is a possible application of interest mining, where users are recommended posts or rumors based on their interest in the topic. To address the challenges related to rumor blocking methods in Section 1.3.2, we propose a "user's interest in topic" - based rumor control model for social networks. It is quite likely that a user is not interested in a rumor circulating in a social network based upon its topic and thus decides not to spread it. It is an obvious fact that a user is more likely to be connected to other users with similar likes and dislikes. So, if a user blocks a rumor by not passing it further, he/she can effectively control the spread of the rumor. Different people take rumors differently according to their interest which is further guided by their age, location, job, etc. Users can have multiple overlapping interests. We consider three factors for rumor control - users' interest in the topic of rumor, influence of a user on its neighbors, and trust between the users. These are quite crucial factors, as a user tends to believe a rumor coming from an influential person and a trusted neighbor more than a non-influential and non-trustworthy neighbor. Based on these factors, we propose two strategies for rumor control, viz. node level blocking and edge level blocking. Node level blocking considers the topic interest and influence of a person while edge level blocking considers the trust among nodes and mutual interest in the topic of rumor. These strategies are incorporated into our proposed rumor blocking model, Susceptible-Infected-Recovered-Blocked (SIRB), which is an extended variant of the popular epidemic model Susceptible-Infected-Recovered (SIR) [44].

Motivation for rumor prevention model- To address the challenges related to counter-rumor based methods in Section 3.2.1, we propose an integrated multi-criterion decision-making (MCDM) based approach for rumor prevention on social networks. This approach is two-fold. First, we select influential or key nodes from the social network. This is done to ensure faster dissemination of counter-rumor messages in the network. Second, on detecting a potential rumor, a counter-rumor

diffusion process is initiated, which ultimately drives the whole system into preventive mode. A counter-rumor is a message or information that is used to debunk the rumor. We propose a Susceptible-Infected-Recovered-Prevented-Agent (SIRPA) model as a counter-rumor diffusion model which is a modified variant of the classical epidemic model susceptible-infected-recovered (SIR) model. The reason for identifying key nodes is to perform targeted immunization rather than random immunization. The former technique has been shown to be much more effective in accelerating the diffusion of information for scale-free social networks than later when it comes to immunization [45]. MCDM method is used to select keynodes that are popular, easily accessible, and capable of faster diffusion. The keynodes solve the time-lag problem as they initiate the counter-rumor diffusion process upon encountering any rumor at the early stage. In cases where they do not have information to counter the rumor, they still have a say on whether some post is a rumor. So, either the system will end up with some counter-rumor message or a warning message that this post is a rumor.

1.5 Thesis Objective

The main goal of this thesis is to propose novel methods for identifying the source of the rumor and its prevention in the OSN.We attempted to find answers to the following research questions (RQs) as part of our thesis.

- **RQ1:** Is it possible to create a model for social networks that can handle all kinds of network structure and semantics? Can this model be flexible enough to represent the different types of connection and interaction within various social networks while still keeping the unique features of each one?
- **RQ2:** Is it possible to create a unified approach that can identify the source of a rumor on different OSN platforms?

- **RQ3:** Is it possible to block a rumor before it reaches its intended audience by leveraging individual user's interest in the topic of the rumor?
- **RQ4:** Is it possible to identify and select key nodes in a social network from multiple perspectives to implement more effective targeted immunization? In addition, can targeted immunization accelerate the spread of counter-rumor messages?
- **RQ5:** How can we design a counter-rumor diffusion model that keeps the network in a rumor preventive state, rather than merely shifting into a controlled state?

We have also added some RQs in the chapters 4 and 5 to more specify RQ3, RQ4 and RQ5 and provided explanations for them in the chapters. The answers to the above research questions are given in Chapter 6.

1.6 Contributions to the Thesis

In this thesis, we contribute three chapters to address rumor source identification and rumor prevention and control problems in OSNs. The contributions made in this thesis are listed below.

1.6.1 Ontology based model for rumor source identification

The main contributions to this chapter are as follows:

1. We propose a three-layered ontology-based architecture for modeling the concepts of online social networks (OSNs) and used it to develop an upper-level domain-specific ontology as well as a lower-level vendor-specific ontology.
2. The proposed ontology model is populated and queried to identify the rumor initiator in OSNs *a posteriori*, considering two scenarios: one where the veracity of the rumor is known and another where it remains unknown.

3. The proposed model is evaluated for its quality and acceptability by utilizing the OQuaRE framework [46, 47] which is a SQuaRE based ontology quality evaluation framework.

1.6.2 SIRB: Rumor Blocking Model

Our research contributions to this chapter are summarized as follows:

1. We introduce two rumor blocking strategies: node-level and link-level blocking on the basis of factors such as user interest in the topic, their influence within the network, and the mutual trust between neighboring users.

2. We propose a compartmental model based on propagation structure, SIRB, for blocking rumors on social networks. This model integrates proposed blocking strategies and is evaluated for its effectiveness in blocking the spread of rumors.

3. We present an empirical study to demonstrate the significance of the proposed SIRB model for synthetic and real social network datasets.

1.6.3 SIRPA: Rumor Prevention Model

Our research contributions to this chapter are summarized as follows.

1. We present a Multi-Criteria Decision Making (MCDM) approach for identifying key nodes in a social network, utilizing the Analytic Hierarchy Process combined with the Technique for Order Preference by Similarity to an Ideal Solution (AHP-TOPSIS) algorithm to ensure targeted immunization.

2. We propose a propagation structure-based counter-rumor diffusion model, SIRPA, to prevent rumors on social networks. This model integrates the proposed AHP-TOPSIS method for providing targeted immunization and is evaluated for its effectiveness in preventing the spread of rumors.

3. We present an empirical study to demonstrate the significance of the proposed SIRPA model for synthetic and real social network datasets.

1.7 Thesis organization

This thesis is organized into six chapters. Figure 1.1 illustrates the overview and interconnection between each chapter. Chapter 1 briefly introduces rumors and various issues and challenges with rumors. This chapter also presents limitations, motivation, thesis objective, and a list of contributions. In Chapter 2, we present a review of the literature and preliminary information related to the thesis. Chapter 3 addresses the first issue we are dealing with, that is, identifying the source of a rumor in OSNs. Here, we demonstrate our proposed ontology-based approach for the detection of rumor sources in OSNs. In addition, in Chapter 4 and Chapter 5, we address the second issue and present two models for the prevention and control of rumors. In Chapter 4, we present a rumor-blocking method, and in Chapter 5, we present a rumor prevention model using a counter-rumor diffusion mechanism. Finally, in Chapter 6, we conclude the thesis with future research work.

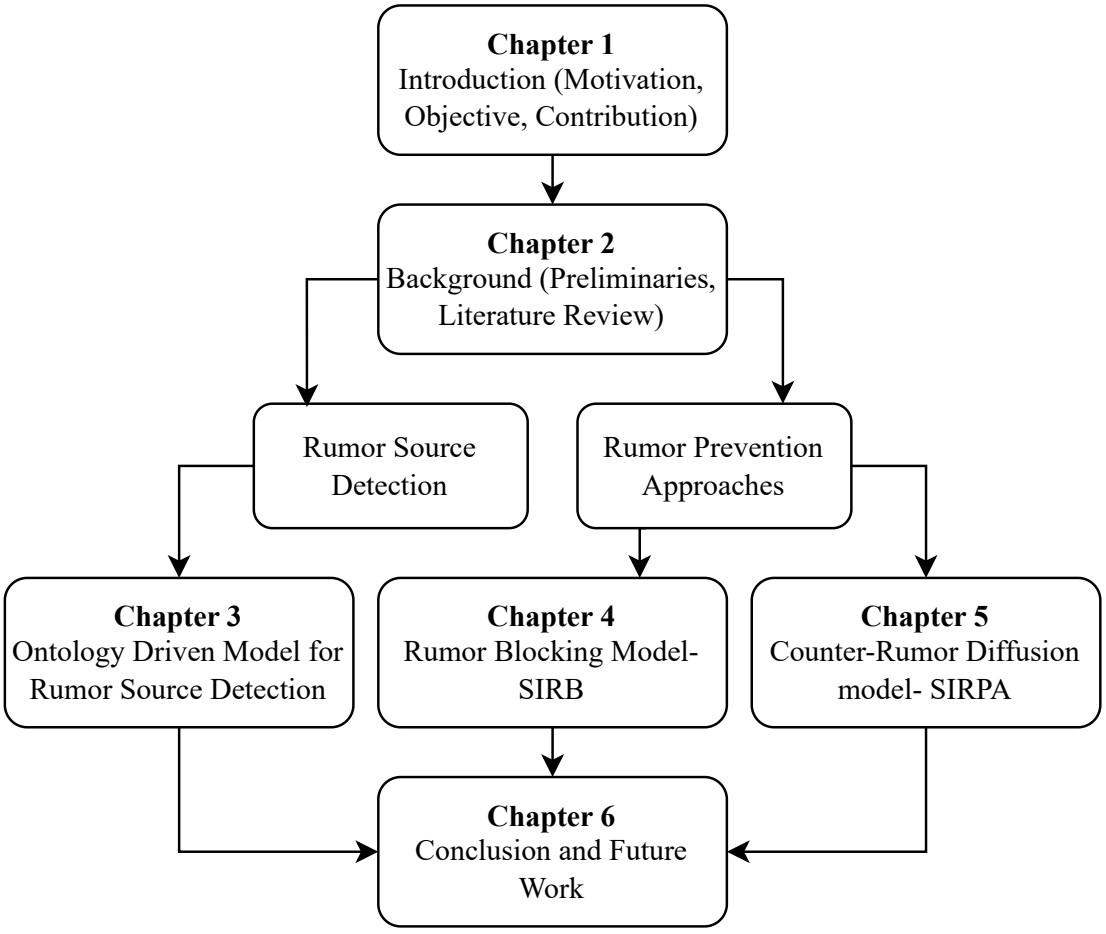

FIGURE 1.1: Thesis structure