Chapter 2

Preliminaries and Related Work

This chapter presents an overview of the concepts used throughout this thesis in Section 2.1. We discuss related work for rumor source detection in Section 2.2.1 and various methods for rumor prevention and control in Section 2.2.2. Section 2.3 discusses various datasets we have used for case study and experiments. Section 2.4 discusses various metrics and frameworks we have used to analyze and evaluate our proposed models.

2.1 Preliminaries

2.1.1 Social Network

A social network can be formally defined as a graph $G = (V, E)$ where V represents a set of vertices or nodes, each corresponding to an individual user within the network, and E represents a set of edges or links, each denoting a social relationship or interaction between two nodes. Social networks can be directed, i.e., edges have a direction (e.g., follower-followee relationships in online platforms) or undirected (e.g., friendship). In weighted social networks, the edges carry weights representing the trust between nodes or some similarity score. Social networks are represented using the adjacency matrix and adjacency list.

Adjacency Matrix: An adjacency matrix $A[|V|][|V|]$ is a 2D array with dimensions $|V|\times|V|$, where $|V|$ represents the number of vertices in an undirected graph. If an edge exists between the vertices v_x and v_y, then $A[v_x][v_y] = 1$ and $A[v_y][v_x] = 1$; otherwise, the corresponding values are zero. In a directed graph, if an edge exists from v_x to v_y, then $A[v_x][v_y] = 1$, with zero indicating that there is no edge. For a weighted graph, if an edge exists between v_x and v_y, then $A[v_x][v_y] = w_{xy}$ and $A[v_y][v_x] = w_{yx}$, where w_{xy} and w_{yx} are the weights of the edges from v_x to v_y and from v_y to v_x, respectively; otherwise, the values are zero.

Adjacency List: An adjacency list is a collection of unordered lists that represent a graph. Each list in the adjacency list details the set of neighbors for a specific vertex in the graph.

Social networks have specific properties that differentiate them from random networks. These properties are as follows.

1. **Scale-free networks:** Social networks are scale-free networks, which are a type of complex network distinguished by the presence of a few highly connected nodes, known as hubs, alongside many nodes with relatively fewer connections. These networks exhibit a power law degree distribution, which means that the probability $P(k)$ that a node has k connections (degree) decreases as a power of k, typically described by the equation:

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

   Here, γ is a positive constant, usually within the range $2 < \gamma < 3$. This distribution indicates that, while most nodes have a low degree, a small number of nodes possess a very high degree, forming the network's hubs [48].

2. **Homophily:** Homophily is the tendency of individuals to form connections

with others who are similar to themselves in various attributes such as demographics, beliefs, and behaviors. This phenomenon often results in the formation of closely knit social groups with shared characteristics. The homophily concept is based on the idea of *"Birds of a feather, flock together"* [49].

3. **Small-world effect:** The small-world effect in social networks refers to the phenomenon where most individuals are connected by surprisingly short paths, often just a few degrees of separation [50]. Despite the large size of the network, any two people can typically be connected through a small number of intermediaries. This effect facilitates the rapid spread of information and social influence throughout the network. It also highlights the interconnectedness of social groups, even in large populations.

2.1.2 Centrality Measures

Centrality measures are metrics used in network analysis to quantify the importance or influence of individual nodes (vertices) within a network. These measures help identify the most central nodes, which can be crucial for understanding the structure and dynamics of the network. Different centrality measures capture various aspects of the importance of a node [51]. Some centrality measures that we used in our thesis are defined as follows.

1. **Degree Centrality (DC):** It specifies the number of direct connections of a node to other nodes. A node having a high degree centrality is more popular among its neighbors than one with a low degree centrality. It is calculated as-

$$DC(i) = \sum_{j=1}^{n} X_{ij} \tag{2.1}$$

where i is the i^{th} node whose degree centrality is to be calculated, j are all other nodes, X_{ij} denotes the link between i^{th} and j^{th} node, and n is the total number of nodes. $X_{ij} = 1$ if there is a link between i^{th} and j^{th} node, 0 otherwise.

2. **EigenVector Centrality (EC):** A node becomes more important if the neighbors of this node are also important. Such nodes are highly influential as their neighbors are popular by themselves. Eigenvector centrality assigns relative scores to all nodes in a network based on the principle that connections to high-scoring nodes contribute more to a node's score than connections to low-scoring nodes. It is calculated as-

$$EC(i) = \frac{1}{\lambda} \sum_{i \in M(i)} EC(j) = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} EC(j) \tag{2.2}$$

where A is the adjacency matrix, $M(i)$ is the set of neighbors of i, λ is the eigenvalue of matrix A which can be calculated as below. In cases of multiple eigenvalues, we select the principal eigenvector associated with the largest eigenvalue to calculate the eigenvector centrality.

$$det(A - \lambda I) = 0 \tag{2.3}$$

3. **Betweenness Centrality (BC):** It measures the geodesics, i.e., the shortest path between two nodes, passing through a particular node, of the node. A node with high betweenness centrality controls the information flow in the network and makes it more accessible. It is calculated as

$$BC(i) = \sum_{a \neq i \neq b \in V} \frac{\eta_{ab}(i)}{\eta_{ab}} \tag{2.4}$$

where η_{ab} is the total number of shortest paths from node a to node b and $\eta_{ab}(i)$ is the number of paths that pass through the i_{th} node.

4. **Closeness Centrality (CC):** It measures the average shortest path between two nodes. A node is more important if it can reach other nodes in the shortest way possible. It measures the closeness of a node to all other nodes. This

feature helps to disseminate influence and information in the network faster. The more a node is central to other nodes, the lower its total distance from them. It is calculated as

$$CC(i) = \frac{1}{\sum_{v_j \in V \wedge i \neq j} d(i,j)} \tag{2.5}$$

where $d(i,j)$ is the distance between nodes i and j.

2.1.3 Multi-Criteria Decision Making

Multi-Criteria Decision Making (MCDM) is a branch of decision-making that involves evaluating and selecting options based on multiple criteria or attributes. It is used in situations where decisions are complex and involve several conflicting or complementary factors that must be considered simultaneously. Unlike single-criterion decision-making, MCDM involves multiple criteria, which could be quantitative or qualitative. MCDM techniques often involve ranking the available options according to their performance on the different criteria, allowing decision makers to prioritize the best choices [52, 53]. MCDM methods incorporate the decision maker's preferences, which may be subjective. Different MCDM approaches can weight criteria differently depending on the importance ascribed to each criterion.

They have four entities-alternatives, criteria, criteria-weight, and performance measures. These methods are applied to problems of various domains where a decision is based on multiple criteria. For example, building information modeling for architecture, engineering and construction [54], COVID-19 pandemic study [55], material selection for optimal design [56], ranking of sustainable sources of renewable energy [57, 58], and many others. The two popular MCDM methods that we used in our thesis are defined below.

1. **Technique for Order Preference by Similarity to Ideal Solution (TOPSIS):** Hwang and Yoon proposed this approach in 1981 [59]. It is a ranking-based MCDM method that selects alternatives that are closest to the ideal solution on the positive side and the farthest from the ideal solution on the negative side. While the negative ideal increases costs, the positive ideal increases profit. This method inputs a $n * m$ decision matrix consisting of n alternatives i.e. nodes of the network and m criteria i.e. centrality measures as selected in [60–65]. Criteria are assigned some positive weights according to their importance in reaching the goal, i.e. influential node detection. A TOPSIS score is generated for each alternative based on which we can determine its rank. In simple TOPSIS, no criteria-weight is assigned to any criteria, so the same weight is assigned to all [60, 61, 64, 65]. In weighted TOPSIS, the weights are first calculated [62, 63, 66] and then assigned to the criteria.

2. **Analytic Hierarchy Process (AHP)-** This MCDM method is proposed by Saaty in 1977 [67]. It models a decision problem as a hierarchical structure of goal, criteria, sub-criteria (optional), and alternatives. Each lower level is prioritized according to its immediate upper level. Figure 2.1 shows the hierarchical modeling of influential node identification problems in complex networks for the AHP method. Here, the goal is to identify the most influential k nodes in the network, the criteria are centrality measures selected in [68], and the alternatives are the individual nodes of the network. After structuring the problem, a pairwise comparison matrix is constructed showing the priority of a criterion over other criteria and then ranking the criteria in the order of importance according to the AHP method. The construction of a pairwise comparison matrix requires human expertise and intensive human efforts. Hence, this method is useful for ranking the importance of criteria, as they are limited in number, but for ranking alternatives, it is not a practical approach, as alternatives are larger in number.

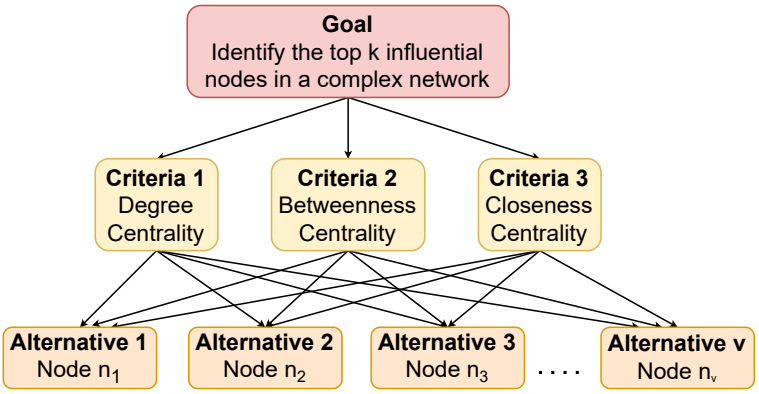

FIGURE 2.1: AHP Method

2.1.4 Information Diffusion Model

Information diffusion models are widely used to assess the influence of selected nodes on the diffusion of information and communication among nodes. Different models have their own rules regarding propagation in the social network. The most popular models used are compartmental models, independent cascade model, and linear threshold model. Compartmental models are popularly known as epidemic models because they simulate the behavior of epidemic spread in a people's network. In these models, the nodes are divided into various compartments or states, which can be in one state at a time. The two most popular compartment models are the Susceptible-Infected (SI) model and the Susceptible-Infected-Recovered (SIR) model [69]. There are also many variants of these models. In the SI model, the nodes are initially in a susceptible state, i.e., they have not received any infection yet but are prone to infection. Upon contacting an infected node, the susceptible nodes become infected with a rate of spread of infection α. SI model can be represented using mean-field equations as follows-

$$\frac{dS}{dt} = -\frac{\alpha SI}{n} \tag{2.6}$$

$$\frac{dI}{dt} = \frac{\alpha SI}{n} \tag{2.7}$$

Here, S is the expected number of susceptible nodes, I is expected number of infected nodes, and α is the rate of diffusion.

The SIR model extends the SI model considering that the nodes recover after some time from infection with a recovery rate β. At any time, all the nodes are in either of the three states- susceptible, infected, and recovered. The SIR model can be represented as-

$$\frac{dS}{dt} = -\frac{\alpha SI}{n} \tag{2.8}$$

$$\frac{dI}{dt} = \frac{\alpha SI}{n} - \beta I \tag{2.9}$$

$$\frac{dR}{dt} = \beta I \tag{2.10}$$

These two models are widely used by researchers to perform the evaluation of their proposed MCDM method to find the influential nodes. Figure 4.2 shows a state transition diagram of the SI and SIR model.

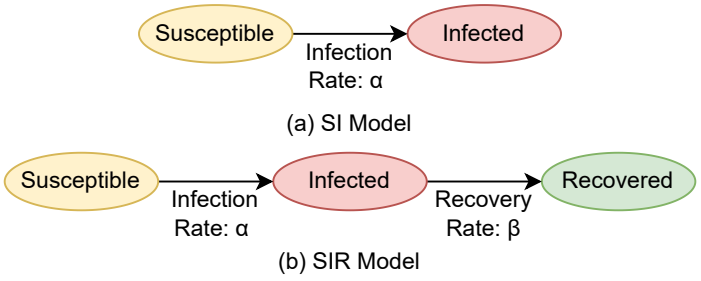

FIGURE 2.2: State transition diagram for SI and SIR models

2.1.5 Ontology and Associated Concepts

Ontology is defined as "a formal, explicit specification of a shared conceptualization" [70]. It is a knowledge representation technique widely used to add semantics to existing domain concepts to represent data in machine-accessible format. It bridges the gap between the semantics of concepts of different vendor-specific applications or services under the same domain to ensure consensus and interoperability [71]. Various concepts of an ontology are as follows.

- **Classes:** The categories or types of things within the domain.
- **Individuals (or Instances):** The specific entities that belong to some classes.
- **Attributes (or Properties):** The characteristics or properties of classes and individuals.
- **Relationships:** The connections between classes and / or individuals.
- **Rules/Constraints:** Logical statements that provide more complex relationships and conditions within the ontology.

By modeling the concepts of a domain into ontology classes, establishing the relationship among these classes, and incorporating both implicit and explicit rules, we get an ontology model of that domain. Modeling a domain knowledge as ontology provides the following advantages.

1. The ontology provides a formal way to represent the data, that is, it represents the data in a machine-accessible format. So, it can be used to provide intelligence to machines that are highly useful for automation.
2. The ontology provides semantics to the data by explicitly stating the concepts, relationships, axioms, and rules.
3. The ontology provides interoperability that ensures the generality of the domain. People generally have multiple OSN accounts. If someone wants to initiate a rumor, they are likely to initiate it on one or multiple platforms. So, opting for a single solution is more convenient than developing an algorithm for initiator detection on multiple OSNs. Thus, interoperability feature of ontology is very useful for our purpose.
4. Ontology has inherent reasoning powers to perform explicit queries for information extraction and knowledge extraction. Using these reasoning abilities, our proposed model can extract information that is used to find the rumor

initiator. This feature makes it different from traditional information models that cannot be reasoned to extract information [72].

2.2 Related Work

2.2.1 Related work for Rumor Initiator Detection

In this subsection, we observe how ontologies have been used in social networks for different purposes. We also provide a study of existing techniques that have been used to detect rumor initiators in social networks.

2.2.1.1 Ontologies for Social Networks and Rumor Modeling

Earlier works have emphasized providing semantics to social networks. Towards this goal, many social network ontologies have been proposed. In Jung and Euzenat [36], a three-layered architecture for a semantic social network is proposed in which three networks are superposed, i.e., social network, ontology network, and concept network in the context of P2P sharing of annotations. This model exploits the knowledge structure used by individuals to infer social relationships and share knowledge. However, this model fails for unequal networks, i.e., networks with nearly the same concept but an entirely different meaning. For example, Facebook and Twitter are two OSNs having the same concepts of users, posts, etc. However, these two networks are different in their operations i.e., friendship and microblogging, respectively. So, both are unequal networks at the semantics level. Also, the classical social network measures are computed for obtaining social features like betweenness, closeness, etc. which can be computationally expensive if calculated online. Mika [73] has presented a three-layered view of the semantic social networks in the form of the tripartite model of actors, concepts, and instances with social dimensions. This work aims at facilitating social tagging mechanisms and affiliation networks, which comprise overlapping objects and concepts based on overlapping sets of communities. However, other problems, such as ambiguous associations and

knowing too much, have not been addressed. Flink system by Mika [37], is a web-based presentation of the social networks of the researchers' community. It exploits the Friend of a Friend (FOAF[1]) for social intelligence. FOAF is an RDF-based vocabulary for representing social network data in a standardized format. It describes people, relationships, online identities, and social groups using subject-predicate-object triples. FOAF serves as an early initiative in the Semantic Web movement, enabling decentralized and distributed sharing of social connections across the web. However, Flink system does not address the multiplexity of social relationships, i.e., users can have different significance to relationships in various aspects. For example, a user can be a friend to another user or a follower to another user. In Hamasaki et al. [74], actor-actor ties on the same tripartite model as of Mika's model [73] are explained. Mika's model [73] was aggregated with adjacent networks, in which each tie represents a relation between actors such as knows, collaborates, and being a friend to, solves word sense disambiguity and data sparsity problems. In Carminati et al. [38], authors have identified five categories of social network data and proposed an ontology based on them. This includes personal information, relationships, social network resources, the relationship between users and resources, and actions that can be performed in a social network. This social network ontology focused on exploiting the reasoning power of ontologies and the semantic web for security and privacy concerns in social networks. They proposed a social network access control mechanism using the semantic web tools OWL and SWRL. In Shen et al. [75], a visual analysis tool OntoVis is proposed in which information in an ontology for social networks is used to understand the social processes and behaviors in large, heterogeneous social networks. This work is limited to extracting information from social network ontologies with undirected relationships. Acharya and Mohbey [76] have proposed a differential privacy-based social network detection system over spatial-temporal proximity for POI Recommendation in location based social networks to solve the issues in cyber security using the ontological data model for Points of Interest.

[1]https://www.w3.org/wiki/Good_Ontologies#The_Friend_Of_A_Friend_.28FOAF.29_ontology

In Declerck et al. [39], ontological modeling of rumors is proposed for the Pheme project[2], built on top of PROTON ontology[3]. This Pheme ontology models the annotations of the Pheme project and is used to find the veracity status of rumors. Wongthongtham and Salih [77] have proposed an ontology model to extract semantics of texual data and define the domain of data for identifying the credibility domain. Sarfraz et al. [40] have proposed an ontology for rumor detection in OSN. Their approach defines the core ontology constructs and considered roles, ratings, confidence levels, location, and post categorization for detecting rumors.

All of these proposed ontologies are intended for different problems of social networks, and none addressed the problem of finding the rumor initiator in OSNs.

2.2.1.2 Detection of Rumor Initiator in Social Networks

In Shah and Zaman [13], Spencer and Srikant [14] and Pei-Duo et al. [28], the problem of rumor initiator detection is considered a maximum likelihood estimation problem. In Shah and Zaman [13], authors have proposed rumor centrality measures to solve this problem; however, their approach is limited to a degree-regular graph that has all nodes of the same degree with no cycles or regular trees and Susceptible Infected Removed (SIR) epidemic spread model. Pei-Duo et al. [28] have also used the same method as that of Shah and Zaman [13] but on a degree-regular graph with a single cycle. In Spencer and Srikant [14], the approach is limited to star networks and irregular trees having almost all nodes of degree two and deals only with the Susceptible Infected (SI) model. In Fuchs and Pei-Duo [29], an asymptotic formula for detection probability of rumor center on random increasing trees is derived. It works for binary search trees, recursive trees, plane-oriented recursive trees, and undirected trees. In Zhang and Aggarwal [78], Rumor Initiator Detector (RID) is introduced that finds the initiator of rumor in a signed network using the Asymmetric Flipping Cascade (AFC) diffusion model. The AFC model has represented signed networks as trust and distrust networks having nodes of opinions- trust, distrust,

[2]https://www.weblyzard.com/pheme/
[3]https://ontotext.com/documents/proton/Proton-Ver3.0B.pdf

inactive and unknown. This model requires a specific network state and gives an NP-hard solution for the exact identification of rumor initiators in a general graph. In Xu and Chen [79], a monitor-based approach to find the rumor source in Independent Cascade(IC) model is proposed. This monitor reports the data it receives, which is observed for a polynomial-time algorithm to compute the rumor quantifier and calculate a reachability-based score. This score is also used to rank the importance of nodes as the source of rumors. In this approach, the focus is on the network structure rather than text-based features and user characteristics. Seo et al. [80] have proposed an approach in which monitor nodes are injected into the network whose job is to report the data they receive. They show with a sufficient number of monitor nodes, it is possible to recognize most rumors and their sources with high accuracy. To detect the rumor source, the source must be close to positive monitors and far from negative monitors. A more comprehensive survey on rumor initiator detection is provided by Shelke and Attar [81].

All of these proposed methods worked on the structure of the social network, that is, all users are considered nodes, and the edges define their associations with each other. However, user, temporal, and post-based features are not considered in any of the existing research. In addition, the networks analyzed are limited in structure, such as cycles, directions, and degrees, which is not the case with real-world social networks.

2.2.2 Related work for Rumor Prevention and Control

2.2.2.1 Self- Media Perspective based Research Work

Zhou and Feng [21] have emphasized self-media and have proposed four measures to manage and control rumors. These are enhancing the supervision of government and authorities, strengthening law enforcement, improving literacy and strengthening the public's ability to comprehend the authenticity of the information. Kostka et al. [22] have advocated crowd-sourcing of users' reactions to the rumorous post.

These reactions are categorized into four types, namely, support, denial, appeal for more information and comment. This work does not talk about rumor prevention and there was also extensive labor involved to get the reactions of so many users. Even if the reactions are obtained, there may be a polarity of ideas because of which the reactions may not be considered neutral.

The self-media perspective-based works have certain limitations. Firstly, the credibility and trustworthiness of self-media sources is hard to establish. Secondly, the perspectives are often limited by the subjective judgement of the self-media and lack of information or ignorance may lead to inaccurate debunking of the rumors. Third, these approaches may suffer from echo chambers and confirmation bias problems. Using these approaches, it is hard to debunk rumors efficiently and assess the amount of rumor prevention and control in the system. So, we need some formal rumor prevention and control models that are capable to assess the amount of rumor prevailing in the system and the effect of counter-rumor strategies in preventing the rumors.

2.2.2.2 Diffusion Model-based Research Work

In the literature, there are two ways to control the diffusion and percolation of rumor in a social network. The first way is to block the rumors in the network either at the edge level or at the node level. The second way is the counter-rumor diffusion method in which some counter-rumor information is introduced in the network that competes against the spread rate of rumor. In this section, we discuss various rumor control models proposed by researchers that are used to represent and analyze the diffusion of rumors in a social network. These models are motivated by traditional epidemic models. Some of these models focus on the analysis of the rumor diffusion process in the social network, while other models not only analyze, but also propose a scheme to control rumors in the social network.

Rumors spread in the same way as viral diseases. So, the study of rumor diffusion models is highly motivated by epidemic models. One of the most popular epidemic

models that researchers have widely used as the base model is the *Susceptible-Infected-Recovered (SIR)* model [69]. In this model, nodes in a social network are classified into three categories- *Susceptible, Infected and Recovered*. Susceptible nodes are prone to receive infection and have not yet encountered one, infected nodes have received infection and become potential spreaders, and recovered nodes are those who were infected earlier, but now have recovered. Epidemic models allow for the mathematical representation of epidemic spread and control theory. Later, Daley Kendall [82] (DK model) and Maki Thompson [83] (MT model) modified the SIR model for the rumor domain. Based on DK and MT models, various variants of these models have been proposed.

Various methods have been proposed by the researchers' community to block rumors in a social network. Zheng and Pan [16] considered rumor blocking as the Least Cost Rumor Community Blocking Optimization (LCRCBO) problem and proposed a Minimum Vertex Cover Based Greedy (MVCBG) algorithm. This algorithm finds a minimal set of nodes (Minimal Vertex Cover) that can be removed from the network to remove complete connectivity in the network. These nodes are removed along with their incoming and outgoing edges. This ensures not only the containment of rumors within community, but also ensures that the anticipated count of influenced nodes does not surpass a specified positive integer, K. Yan et al. [19] considered rumor blocking as the Rumor Spread Minimization (RSM) problem where they propose removal of an edge set from the network to minimize the rumor spread. They proposed a heuristic-based approximation algorithm to identify this set of edges. Lin et al. [20] proposed a crowd-blocking approach where users can implement rumor control strategies in collaboration with other users. To motivate users for this, they proposed two incentive mechanisms based on the Stackelberg game and real-time reverse auction-based mechanisms. This approach requires awareness in the crowd about the rumor. Yan et al. [17] proposed the Minimizing Influence of Rumors (MIR) problem for rumor control where they selected an initial set of blockers and used the independent cascade model to study their effect on rumor blocking.

They proposed a two-stage strategy, viz. Generating Canadidate Set and Selecting Blockers (GCSSB) for this purpose. They also provided a study of their proposed work on tree networks and obtained an optimal solution for them. Guo et al. [32] emphasized that a user's perception of rumor is not based on just one feature, rather on the overall evaluation of all its features. Therefore, the rumor's influence spread can be divided into the spread of its multiple features. Based on this, they proposed a multi-feature diffusion model (MF-model) by formulating a multi-feature rumor blocking (MFRB) problem on multi-layer networks.

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 much more. Proposing a model based on users' interest works well as interested users generally behave in a similar way and one does not have to take account of uninterested users for rumor analysis and control. However, in the literature, there are some rumor control models that consider users' interest. Jeong et al. [33] proposed a rumor control model called ISS (Ignorant, Spreader, and Stifler). This model aims to reduce the number of infectors in the system through three control strategies: telling the truth, punishing spreaders, and deleting messages. Two scenarios are considered: one where user interest factors such as age, sex, location, and occupation are taken into account, and the other where no interest factors are considered. The results indicate that when interest levels are high, more emphasis is needed on truth telling before ignorant individuals receive rumors and on punishing spreaders. In contrast, deleting rumor-related information from the media is crucial at points of low interest. A general interest factor is applied to all nodes, regardless of individual interests. Tang et al. [34] proposed an ISS model for rumor diffusion that emphasizes the interest of users in a dynamic friend network. This model calculates the interest between two users as the interest distance, which is an XOR of two binary codes. This study suggests that ordinary news can turn into

rumors through complex diffusion. It finds that random social networks facilitate the spread of rumors better than dynamic friend networks, and smaller global clustering coefficients in social networks lead to wider propagation of rumors.

These studies do not consider user interest in the topic of the rumor and its role in the network diffusion. In chapter 4, we propose a model that considers this factor and develop a rumor blocking approach based on this model.

Among counter-rumor diffusion based methods, Bao et al. [23] proposed a variant of the SIR model, called the Susceptible-Positively Infected- Negatively Infected-Recovered (SPNR) model which divides the infected nodes into positive and negative categories based on users' opinions. They suggest a rumor control method guided by opinions, where positively infected nodes (nodes refuting the rumor) are increased in the network so as to control the rumor. This model is based on the fact that the maximum users follow the belief of influential users. Thus, they recommend inserting positively infected nodes and connecting them with others for rumor control. Zhang et al. [24] proposed the Official Refutation Information (ORI) model for rumor dissemination, where each rumor competes with its corresponding official refutation information. They found that high government credibility ensures perceived credibility; an informed public can independently dispel rumors; longer delay between rumor and truth causes greater damage; using multiple media for ORI dissemination is more effective. Okada et al. [25] proposed an expanded SIR model for rumor prevention by integrating correction tweets. The model includes five node types: susceptible, infected but unaware of corrections, infected and spreading without corrections, aware of corrections but not spreading, and aware of corrections and spreading. Corrections by self-media groups may introduce chaos and new ideas. Ojha et al. [26] introduced the Susceptible-Verified-Exposed-Infected-Recovered (SVEIR) model, incorporating user verification to authenticate and distinguish between verified and unverified users (Exposed). The model combats fake news with mean-field equations to monitor spreaders and fake news. Kumar et al. [35] introduced the Susceptible-Infected-Recovered-Anti-spreader (SIRA) model,

which accounts for simultaneous operation of both supporters and deniers of information. Their model addresses rumor propagation and control strategies involving anti-spreaders. Guo et al. [18] introduced the SEIMR (Susceptible-Expose-Infective-Media-Remover) rumor controlling model, incorporating media reports and time-lag effects. It explores how media reports influence the status of the nodes switching between susceptible, infected, and exposed.

In addition to these, there are various agent-based rumor diffusion models proposed by researchers that give insight into how agents can be used for rumor diffusion and counterfeiting. Mazzoli et al. [84], a multi-agent-based rumor diffusion network for scale-free networks is proposed. The model represented three types of diffusion – the spontaneous spreading of information, collective influence, and communication persuasion. Instead of detecting news as rumor or not, the agent calculates the news reliability. The authors observed that the correction does not take over the misinformation cascade and users keep on sharing false news. Another observation is a polarization of the criticism over the news spread, and this depends on the reliability of the news given at the beginning of the simulation. Serrano et al. [85] proposed an ABSS (Agent-Based Social Simulation) model for rumor spread. They proposed the hypothesis that when a user becomes recovered, it will not influence their neighbors to recover. A variant of SIR as neutral-infected-vaccinated-cured is proposed. Agents initialize infected users and after some time a random infected node starts counter-rumor spreading.

Most of the proposed models are mainly for rumor control [23, 24, 33, 35, 86–88], and only a few models are for rumor prevention [25, 26, 89]. However, these rumor prevention models do not make a distinction between rumor control and prevention. So, there is a need to specify the rumor prevention and control process and develop the rumor prevention model accordingly. In addition, some researchers have advocated the role of government and authorities in debunking the rumor. However, in situations where the credibility and trust in the government and authorities is less, we need to find a solution where the rumor debunking comes from an unbiased

source. Agents are one solution to this problem [84, 85]. Lastly, the selection of the rumor counter nodes should be such that maximum spread of the rumor can be prevented. So, there is a need for targeted immunization. In Chapter 5, we first find the keynodes for targeted immunization which are later used to run the proposed counter-rumor propagation model for rumor prevention.

2.3 Datasets

In this section, we provide a brief introduction to the datasets utilized for case study, simulations, and analysis throughout this thesis. By offering a comprehensive overview of the datasets, we aim to provide insight on the empirical foundations of the thesis by demonstrating how they contribute to the validation and testing of the proposed models. Various datasets used throughout this thesis are listed as follows.

Pheme dataset- In chapter 3, we have used the Pheme dataset [90] to demonstrate the proposed ontology model in a case study of the Twitter social network. Pheme is a public dataset that is a collection of social media data, primarily from Twitter, curated for the purpose of studying rumors and misinformation. It was developed as part of the PHEME project[4], which focused on the automatic detection of rumors on social media during events of crisis. This dataset comprises a set of Twitter rumors and non-rumors shared amid unfolding news situations. It encompasses rumors and non-rumors associated with nine distinct events where each rumor is labeled with its corresponding veracity status (categorized as true, false, or unverified). Each of the rumors and non-rumors contain the source tweet and reaction tweets. In the source tweets, there is information about the user who initiated the tweet. Thus, we can extract information about the rumor initiator from the source tweet.

Network generating models- We have used the Barabási Albert (BA) model [91] and Erdős Rényi (ER) model [92] for the generation of synthetic networks in

[4]https://www.pheme.eu/

chapters 4 and 5. BA model is a generative model that produces scale-free networks with few nodes with many connections and many nodes with only a few connections. It follows a power-law degree distribution which leads to the emergence of highly connected nodes i.e. key nodes. The BA networks also have small-world property which means that any two nodes are not far apart in terms of the number of connections along the shortest path between them. These properties make a resemblance between BA networks and real-world social networks, so it is highly used to study phenomena like information diffusion in social networks. ER model is a randomly generating network model for generating random networks, which are networks with arbitrary connections between nodes. In the ER model, each possible edge between nodes is included in the graph with a fixed probability p which is the same for all possible edges. It is used to study the structural properties of networks and model random network phenomena like information diffusion in random networks. Both of these models are used to generate networks of 1000 nodes in size. The degree distribution of both networks is shown in figure 2.3 for the BA network and in figure 2.4 for ER network. While the BA network resembles more to social networks as it follows the power-law degree distribution, the ER network provides a random simulation network and follows the poisson degree distribution.

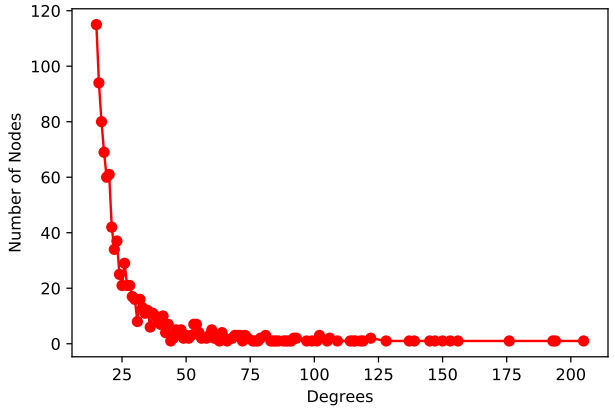

FIGURE 2.3: Degree Distribution Plot for BA Network

Twitch dataset- We have also simulated the proposed rumor prevention and control models in chapters 4 and 5 on four real datasets from the Twitch social

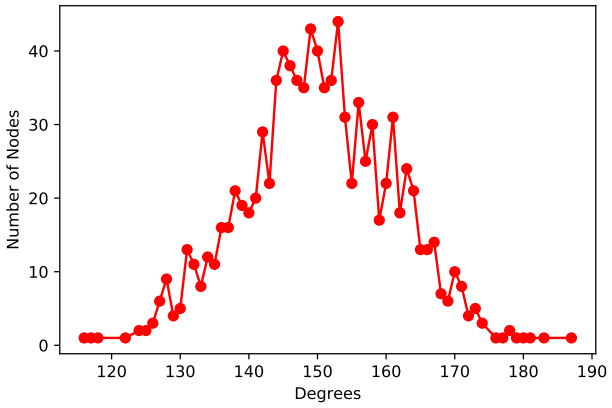

FIGURE 2.4: Degree Distribution Plot for ER Network

network [93]. Twitch networks are publicly available datasets of Twitch user-user networks of gamers who stream in a certain language. Nodes are the users themselves and the links are mutual friendships between them. Each dataset contains gamers from a country. We have used datasets for England (EN), Espanol (ES), Portugal (PT) and Russia (RU). The details of the data set are provided in Table 2.1.

TABLE 2.1: Dataset Description for Twitch Dataset

Dataset	#Nodes	#Edges	Density	Transitivity
EN	7,126	35,324	0.002	0.042
ES	4,648	59,382	0.006	0.084
PT	1,912	31,299	0.017	0.131
RU	4,385	37,304	0.004	0.049

2.4 Ontology Model Evaluation using OQuaRE Framework

The proposed ontology model for finding the rumor initiator in Chapter 3 is evaluated for quality and acceptability using the OQuaRE (Ontology Quality Requirements and Evaluation) framework [46, 47]. The OQuaRE ontology evaluation methodology is based on SQuaRE (Software product Quality Requirements and Evaluation), a standard model for evaluating software quality and requirement specification. The OQuaRE model reuses SQuaRE characteristics and sub-characteristics

for the evaluation of the ontology. The characteristics and sub-characteristics of OQuaRE are listed in Table 2.2.

Table 2.2: Characteristics in OQuaRE

Characteristics	Sub-characteristics	Associated Metrics
Structural	Formal Relations Support, Cohesion, Tangledness, Redundancy	ANOnto, RROnto, TMOnto, LCOMOnto
Functional Adequacy	Schema and Value Reconciliation, Consistent Search and Query, Knowledge Acquisition, Clustering and Similarity, Indexing and Linking, Inference, Controlled Vocabulary, Guidance and Decision Trees	ANOnto, RROnto, AROnto, INROnto, NOMOnto, LCOMOnto
Maintainability	Modularity, Reusability, Analysability, Changeability, Modification Stability, Testability	WMCOnto, DITOnto, NOCOnto, RFCOnto, NOMOnto, LCOMOnto, CBOOnto
Reliability	Recoverability, Availability	WMCOnto, DITOnto, NOMOnto, LCOMOnto
Operability	Learnability	WMCOnto, LCOMOnto, RFCOnto, NOMOnto, CBOOnto, NOCOnto
Compatibility	Replaceability	WMCOnto, DITOnto, NOCOnto, NOMOnto
Transferability	Adaptability	WMCOnto, DITOnto, RFCOnto, CBOOnto

In the table, associated metrics for each characteristic are also shown. Each characteristic is evaluated using a subset of sub-characteristics measured using associated metrics. These metrics are calculated as follows.

$$LCOMOnto = \Sigma path(C_{leaf_i}, Thing)/m \tag{2.11}$$

$$WMCOnto = (\Sigma(P_{C_i}) + \Sigma(R_{C_i})/\Sigma C_i \tag{2.12}$$

$$DITOnto = max(path(C_{leaf_i}, Thing) \tag{2.13}$$

$$NACOnto = \Sigma(SupC_{leaf_i})/\Sigma C_{leaf_i} \tag{2.14}$$

$$NOCOnto = \Sigma(R_{C_i})/(\Sigma C_i - R_{Thing)} \tag{2.15}$$

$$CBOOnto = \Sigma Sup_{C_i}/(\Sigma C_i - R_{Thing}) \tag{2.16}$$

$$RFCOnto = (\Sigma(P_{C_i}) + Sup_{C_i})/(\Sigma C_i - R_{Thing}) \tag{2.17}$$

$$NOMOnto = \Sigma(P_{C_i})/\Sigma C_i \tag{2.18}$$

$$RROnto = \Sigma(P_{C_i})/(\Sigma R_{C_i} + \Sigma C_i) \tag{2.19}$$

$$AROnto = \Sigma Att_{C_i}/\Sigma C_i \tag{2.20}$$

$$INROnto = \Sigma R_{C_i}/\Sigma C_i \tag{2.21}$$

$$CROnto = \Sigma I_{C_i}/\Sigma C_i \tag{2.22}$$

$$ANOnto = \Sigma A_{C_i}/\Sigma C_i \tag{2.23}$$

$$TMOnto = \Sigma R_{C_i}/\Sigma C_i - \Sigma C(DP)_i \tag{2.24}$$

Based on the metrics, the OQuaRE framework calculates the values of sub-characteristics, averages them to calculate the value of characteristics. It again averages the values of the characteristics to provide a global average ontology score. This ontology score is used to determine whether the ontology is acceptable or not. The OQuaRE scale system classifies the global ontology score as 1-not acceptable, 3-minimally acceptable, and 5-exceeds the requirements.