

1

1.1. AN OVERVIEW ON THE PHENOMENON OF RUMOUR SPREADING

In one of the most famous Shakespeare's plays - Henry the Fourth, Part II - he writes "rumour is a pipe, blown by surmises, jealousies, conjectures, and of so easy and so plain a stop, that the blunt monster with uncounted heads, the still-discordant wavering multitude, can play upon it". Shakespeare's words elegantly express how easy, widespread, and vicious the emergence and circulation of rumours could be. Since the play was first written, the phenomenon of rumour spreading is exacerbated and turned into a far-reaching phenomenon to the extent that the World Economic Forum ranked the spread of misinformation as one of the top risks facing the world today [1], and Oxford dictionary picked fake-news as the term of the year in 2016 [2].

Although the rumour spreading is mostly associated with political contexts owing to the excessive use of rumours by political figures to **disparage** their rivals and critics, the scope of this phenomenon is much bigger than politics [3, 4, 5, 6]. It is, in fact, a **domain agnostic** phenomenon that arises in any circumstance in which meanings are **uncertain**; questions are unsettled, information is missing, and lines of communications are absent [6]. From the content perspective, rumours are false or unverified statements about instrumentally important topics. Thus any incident -no matter if it is political or not- could be a subject of **rumour spreading**. People engage in the rumour process since it attributes a ready-made justification to unexplained events. It increases the comprehension and understanding of the situation by offering details and reasons as well as meanings and clarifications. Rumours might also be initiated deliberately as a psychological tool for strategic purposes such as character assassination, influence operations, and financial benefits [7, 8, 3]. The dissemination of rumours, whether **intentional** or inadvertent, may feed on hate, create fear, and raise false hopes [9]. It may tarnish reputation of individuals [4], organisation [3], or even countries [10], provoke riot and unrest [5], shake financial markets [11], influence **decision** making process[12], and disrupt aid operations [13, 7].

The rumour is a collective process **which its** existence is contingent on the circulation [14]. In this vein, the role of media is crucial as it streamlines the communication and increases the rate of reach to the audience [15] (and subsequent exposure to the rumour). Traditionally word-of-mouth and letter were the primary means of communication and rumour spreading [16]. The **introduction** of the technologies such as printing press and radio for the mass communication profoundly affected the rumour spreading [15]. Particularly, the sudden rise of social media in the last decade of the twentieth century has provided a nurturing environment for rumours [17] to thrive and circulate in an unprecedented scale, speed, and scope [18, 19]. The size and diversity of social networks [20] as well as automation mechanisms [21, 22, 23, 24, 25] play a central role in the degree of rumour dissemination. Besides, other factors, such as **lack** of media literacy [26], minimal supervision [17], low barrier to entry [27], and the lack of social media regulation [28] facilitate the creation and circulation of rumours.

The escalation in the rumour diffusion may lead to severe consequences that can influence political, economic, and social well-being [13]. For instance, on April 23 of 2013, the Associated Press Twitter account released a tweet saying "Breaking: Two explosions

in the White House and Barack Obama has been injured.” This tweet went viral by 4000 tweets in less than 5 minutes. The spread of this false news precipitated a big drop (with the value of 140 billion dollars) in the market in a single day. In fact, automated trading algorithms immediately began trading based on the potentials and consequences of the explosion in the white house and the death or injury of U.S. president [11]. This example just shows one case of rumour spreading with dire consequences. There are plenty of rumour dissemination cases in other domains such as elections [29, 30, 31, 32, 33, 34, 35], business issues [12, 6], and healthcare [36, 21] which lead into severe outcomes .

In response to the detrimental effects of rumour propagation, **those who were affected started to devise confrontation strategies**. Although taking the potential danger of rumour spreading into account and countering this phenomenon was an important action, it was often an intermittent effort with ephemeral impacts. There was no long-term plan behind the confrontation strategies. Whenever a major incident happened or was about to happen, rumours started to thrive and then countering techniques were proposed and practised [5, 37, 4, 7]. **This approach might work previously; however, by the sudden growth of social media in the past decade, the former confrontation approach could not keep up with the rate of rumour supply and circulation**. Because of the changes in the rumour spreading, the countering methods also changed to the extent that variety of stakeholders such as social media platforms, governments, academia, and media organisations began to collaborate and **develop** new solutions. Although a constellation of **counter rumour** strategies **have** been proposed and practised in different levels especially in the past few years, the massive waves of rumours are still sweeping over individuals, organisations, and societal institutions [13]. This is an alarming trend that has to be controlled; otherwise, due to the potential of **the** social media rumours the repercussions might be catastrophic.

1.2. RESEARCH OBJECTIVE AND RESEARCH QUESTIONS

The primary motivation of this dissertation is to tackle the wild spread of rumours in online social media. A clear problem definition is **the** first and foremost prerequisite to this goal. Albert Einstein once said, “If I were given one hour to save the planet, I would spend 59 minutes defining the problem and one minute resolving it”. It is crucial to obtain a good understanding of the problem before taking any action; otherwise, the proposed solution would be inaccurate and irrelevant. The problem definition, in this case, entails determining what exactly has to be curbed and controlled. There are different variations of false and unverified information (e.g., fake-news, disinformation, misinformation, conspiracy theory, etc.) which are recognised by the scholars as similarly harmful phenomena. However, it has to be clarified what is and what is not in the focal point of this **thesis**. Besides, **the** full understanding happens when the targeted phenomenon is studied within the context (i.e., social media). It helps to understand whether and to what extent social media features facilitate the spread of rumours.

By defining and demarcation of the problem, it would be clear what has to be tackled. Because of the relatively long period of rumour confrontation in the societies, it is indispensable to obtain an overview of the past **counter rumour** strategies. It would provide information about the strengths and weaknesses of the rumour responses in the past. Those information could be utilised later in the development of confrontation plan

1

against rumour dissemination. After going through the past strategies and investigating them, it is **a** time to **act** and tackle rumours. There are two broad paradigms of passive and pro-active confrontations. Despite the clear advantages of pro-active approach, the current landscape of rumour confrontation is extensively dominated by passive approach. **Thus to have an effective yet feasible confrontation plan**, both paradigms should be **present** in the rumour response agenda.

The passive paradigm consists of two major strands of short- and long-term strategies. **Short term** strategies often aim to filter rumours using a machine learning technique called binary classification. It is a supervised learning technique in which a model is trained with existing samples of rumours and non-rumours in order to flag unforeseen rumour messages. **In this method, the way of annotating training data-points determines how each class is defined for the classifier. Unlike rumour samples which often are annotated similarly, non-rumours get their labels arbitrarily based on annotators' volition. Such an approach leads to unreliable classifiers which cannot distinguish rumour from non-rumour consistently.** The other set of strategies are the ones with the goal of creating long-term immunity. Those strategies tend to create a resilient society by training people to be more careful and critical about the information they receive. In this vein, the role of academia to assess the effectiveness of training methods or to develop new methods is crucial. Despite a great deal of research in this arena, the amount of progress by academia is not clear yet. This may lead to misjudgements about the performance of the research topic, which can ultimately result in wrong science policies regarding academic efforts for quelling rumours. The other confrontation paradigm is pro-active, which aims to take measure before a rumour begins to spread. This approach has not practised yet; thus, it is essential to take the preliminary steps and develop an early rumour confrontation model with a pro-active perspective.

To address the above-mentioned gaps, the principal objective of this dissertation is defined as follows:

To systematically study the rumour confrontation within online social media by investigation of the rumour construct in social media, scrutinising the past counter strategies, and addressing their shortcomings.

To accomplish this objective, it is required to look into four **nearly independent** topics, which are described in the following:

- First, the main construct of this study, namely rumour needs to be scrutinised. The notion of rumour refers to a complex phenomenon with a controversial conceptualisation which makes its identification rather difficult among the closely related concepts. Without a crystal-clear understanding of rumour, **this dissertation falls apart** as it is like an underlying substrate that glues down different pieces of this manuscript together.
- Second, social media is a major medium for the emergence and the spread of rumours. In the post social media era, the spread of rumours scaled-**up**, accelerated, and diversified. It is essential to understand the properties and mechanisms of this environment that facilitate the emergence and growth of rumours.

Very complicated paragraph! My points:
1. It is unnecessary to talk about binary classification! I would leave that part out

2. A figure would help understand the strategies:

1. Passive: 1.1 Short-term.. 1.2 Long-term
2. Pro-active

3. Try to break the paragraphs into some!

- Third, in order to tackle rumour spreading in social media, it is of the utmost importance to have an overarching view regarding the as-is situation. This means we need to have a comprehensive and critically analysed list of the past counter-rumour strategies juxtaposed in a common framework.
- Fourth, it is essential to address the shortcomings of the past counter-rumour strategies to develop new confrontation strategies. Due to the long-standing vulnerability of human-being to rumours and the high rate of diffusion, strategies based on short- and long-term approaches should be taken into account.

The topics mentioned earlier, and their associated research questions are followed during this dissertation. For each of the above topics, one or more research questions are raised, whose answers can help address the objective of the dissertation. The first question is about the main building block of this research and take the rumour conceptualisation into account. The second one is primarily about the role of social media and its exclusive mechanisms in the facilitation of rumour spreading. The third question tends to address the as-is situation of rumour confrontation by the analysis of the past counter rumour strategies. The last three questions are about tackling rumours. In the following, the research questions are discussed in more detail.

RQ1. WHAT IS THE RUMOUR AND HOW IS IT DIFFERENTIATED FROM ITS CONCEPTUAL SIBLINGS?

This question is posed regarding the epistemic crisis of rumour and its conceptual siblings as they are often used interchangeably. There are many concepts in the English language implying false or unverified information. Despite the years of academic research on those concepts, there is a considerable disagreement between the proposed definitions and they are often conflated into one idea. Besides, the lack of consensus drains the community's efforts in countering the surge of false information. To address this question, we delve into the literature of rumour spreading to understand the notion of this phenomenon as a process (not a product) along with its conceptual relatives.

RQ2. WHETHER AND TO WHAT EXTENT SOCIAL MEDIA STREAMLINE THE SPREAD OF RUMOURS?

Rumour spreading is a long-standing phenomenon between human-beings. The development of communication technologies facilitated the spread of rumours by introducing features such as distant mass communication and synchronicity. However, by the emergence and the growth of social media, it suddenly turned to a large-scale phenomenon with catastrophic consequences. This research question tends to investigate the anatomy of social media to understand the mechanisms and properties that could promote the spread of rumours. To address the question, two sets of experiments are designed to ascertain the role of social bots and recommendation systems as two exclusive mechanisms in social media platforms.

Questions should include an explanation of challenges and gaps, along with the importance of solving them. Almost all research questions lack such a structure, and include a part of the solution you are going to provide in the thesis. The solutions should be presented in the contribution section.

1

RQ3. WHAT IS THE AS-IS SITUATION REGARDING THE RUMOUR RESPONSE IN THE PAST AND HOW DID EACH STRATEGY CONFRONT RUMOURS?

The **purpose** of questioning the as-is situation is to shed light on the past efforts in countering rumours to ascertain the flaws and shortcomings of current control approaches. It works as a bird-eye view which allows understanding what aspects of rumour confrontation has been targeted extensively and what aspects are highly neglected. **To address this question, we first review the literature of rumour control since the second world war, then we set all the collected counter strategies into a common framework and analyse their effectiveness.**

RQ4. **How IS THE ACADEMIC RESPONSE REGARDING THE SPREAD OF RUMOURS?**

One of the important aspects of confrontation with rumour spreading that deserves special attention is to create immunity against the rumours by training the mind. Academia is on the front-line of developing immunity-based response; however, it is not known whether the past academic efforts could do justice to the significance of this confrontation approach. What this research question is bringing up is to measure the readiness of academia regarding rumour spreading. Depending on the outcome of the assessment, the next steps could be devised. **To address this question, a computational approach based on the combination of emergence framework and bibliometrics analysis is adopted. In this approach, the performance of the academia is evaluated based on research outputs.**

RQ5. HOW COULD WE IDENTIFY RUMOURS IN SOCIAL NETWORKS AUTOMATICALLY, CONSISTENTLY AND IN A TIMELY MANNER?

The massive flow of rumours in social media has made the manual inspection of the transmitted messages impossible. One of the alternative approaches that could be used is computational rumour detection which is scalable and fast. **For this question, we tend to develop a machine learning-based model that can identify rumours accurately and in a timely manner. Despite the common practice (and as we later argue the common mistake) of the field, namely, using a binary classifier, we utilise a novel classification approach called one-class classification as rumour detection is not a binary problem.**

RQ6. HOW COULD WE TAKE PREEMPTIVE MEASURES REGARDING RUMOURS IN SOCIAL MEDIA AND CONFRONT THEM PROACTIVELY?

Despite the discrepancies between the **counter rumour** strategies, they share a similar confrontation **style**. They develop resilience against rumours in a retrospective manner. They tacitly assume the inflow of rumours always recycles the past rumourmongering techniques. Thus if the new rumours use novel techniques, it would be pretty hard to rein them. In order to address these issues, we could switch to the pro-active confrontation **style**, which simply means looking at the rumour process from **the rumourmonger eye**. This would allow to think like adversaries, discover their plans before execution, and develop preemptive measures. **To address this question, we tend to develop a data model that can capture the functional mechanisms of a rumour campaign from a ru-**

mourmonger perspective.

1

1.3. CONTRIBUTIONS AND GUIDE TO READERS

In particular, this dissertation makes six different contributions to the field of rumour studies by a systematic study on the rumour confrontation within the social media. Chapter 2 contributes a comprehensive conceptualisation regarding the notion of rumour and its conceptual siblings. Chapter 3 investigates the role of social media in the rumour promotion by measuring the extent that recommendation systems and social bots (as two major exclusive mechanisms of social media) streamline **spread** of rumours. Chapter 4 evaluates the as-is situation of rumour confrontation by presenting past counter rumour strategies, and then setting them in the epidemic control framework. Chapter 5 contributes to the rumour immunisation approach by measuring the readiness of the academia regarding rumour spreading through a bibliometric approach. Chapter 6 contributes to the mitigation of rumour transmission- and rumour exposure-rate by proposing a novel approach to computational rumour detection based on **one-class classification**. Finally, Chapter 7 proposes a pro-active approach to rumour confrontation by developing an operational level model that can capture the underlying mechanisms of rumour campaigns. The following outline puts forward the list of contributions as well as their corresponding chapters and research questions.

- **Chapter 2** In response to RQ1, we examine the epistemic crisis between different variations of false and unverified information. We delve into the literature and infer that rumour, misinformation, disinformation, propaganda, conspiracy theory, pseudoscience, and fake-news belong to the same conceptual family as they follow a similar development process. The genesis of each concept might find its origins in different uses, but after the first generation of transmission, different variations start to look alike. This would help the scientific community to pool their knowledge and resources on confrontation with rumour spreading instead of endless discussions on the categorisation of false and unverified information. It is also discussed that gossip and legend do not belong to the rumour family no matter how similar **is their development process**. This would also brief the community to include different variation of rumours and leave out gossips and legends when they study rumours. Chapter 2 provides a more detailed explanation regarding this process-based view to **the** rumour and its conceptual siblings.
- **Chapter 3** In response to the RQ2, namely the role of social media in the promotion of rumours, we investigate two exclusive social media automation mechanisms of social bots and recommendation systems as they are alleged to play a central role in the spread of rumours. To this end, we analysed 700,000 tweets in 2018 Dutch municipal election, and 1,000 YouTube videos about conspiratorial topics. Our analysis along with a handful of studies in this domain show that the automation mechanisms in online social media platforms have a clear impact on the spread of rumours; however, this effect mediates by a variety of factors such as location, time, and rumour topic. In Chapter 3 **the** data collection, **the** experiments and **the** results are thoroughly discussed.

1

- **Chapter 4** In response to RQ3, we collect, review, and analyse ~~the~~ major **counter rumour** strategies that were dispersed in the literature. Our focus is on the organisational and governmental response to tackle the rumours since the world war II. To understand why those strategies could not steadily rein in rumour spreading, we analyse them using epidemic control framework due to the strong similarity between the propagation of disease and information. We conclude that the ephemeral reactions, the absence of a comprehensive plan, and neglecting the **immunisation based** solutions are amongst the reasons for the failure of response to rumour dissemination. Chapter 4 provides detailed explanations regarding the counter strategies and their analysis.
- **Chapter 5** In response to RQ4, we use the theory of emergence to assess the readiness of academia regarding rumour spreading. Based on this theory, five dimensions of novelty, growth, coherence, impact, uncertainty and ambiguity determine the status of an emerging phenomenon. In this research, we first need to quantify the academic efforts regarding rumour spreading to be able to measure it. To this end, we collect more than 21,000 scientific papers about rumours. The next step is the operationalisation of the emergence dimensions. After this phase, and measuring the degree of emergence in the topic of rumour spreading, we could observe an increasing trend for the growth, the coherence and the impact and a decreasing trend for the novelty. To propel this research domain and encourage academia to contribute more to this arena, we propose an external push strategy meaning arranging dedicated publication venues such as journals and conferences for this field of research. In Chapter 5, ~~the~~ data collection, analysis, and results are explained and discussed in detail.
- **Chapter 6** In response to RQ5, we first pose a major issue regarding the binary classification as the predominant approach in computational rumour detection. we argue that unlike rumour samples which are often annotated similarly, non-rumours get their labels arbitrarily based on annotators' volition. Because of that, binary classification may lead to unreliable outcomes. To tackle this issue, we propose to use ~~the~~ novel classification approach called one-class classification (OCC). Unlike the binary classification, the training in OCC is only based on one class. We apply seven once-class classifiers from three different learning paradigms and compare their performance. Our results show that this approach can recognise rumours with a high level of F1-score. Chapter 6 provides detailed explanations regarding data, features, and experiments.
- **Chapter 7** In response to RQ6, we propose a proactive rumour confrontation approach which provides us with an opportunity of looking at the rumour campaigns from an adversarial perspective and developing preemptive measures. We develop this model in a **step by step** manner. We start from a coarse-grained model (by combining the misinformation machine model and Maathuis Cyber Operation Model), then we operationalise it based on the literature, real cases, and expert interviews in an iterative manner. Finally, we give a formal presentation of the model using OWL. In Chapter 7 the model development and verification is explained in detail.

Do you want to add the disaster data you collected as a contribution?

1.4. ENGINEERING SOCIAL TECHNOLOGIES FOR A RESPONSIBLE DIGITAL FUTURE

1

This section explains the relevance of this thesis with the TU Delft's research program of "Engineering Social Technologies for a Responsible Digital Future". The technological developments are accelerating across a large number of domains, from health to finance and communication [38]. This rapid development is like a double-edged sword which comes with perils and promises. Although on the surface, technologies often offer a lot to fix problems and improve humans life, but underneath they may lead to more troubles. Thus, there is an urgent need for the investigate those technologies (and the changes inflicted by them), in order to take appropriate measures before it gets too late.

One of those technologies with far-reaching implications on our lives is social media. It has removed the physical barriers and allows multilateral synchronous communication with long-distance locations. It also provides us with the opportunity of multi-media message transmission. However, all those features could also serve the mischievous function of rumour spreading which may lead to catastrophic repercussions. Hence it is of the utmost importance to protect and secure this technology by countering irresponsible usages. Due to the multidisciplinary nature and the large scale of the problem, it should be addressed by social technologies which incorporate both social and computational aspects of the problem. This problem is addressed based on the principles mentioned above. It harvests a socio-technical approach to benefit social media without any concern regarding rumours.

The introduction ends quite abruptly! You basically cover everything, but need to end it a bit more smoothly