Measuring Impact of Rumorous Messages in Social Media

by

Tara Matthews

A dissertation presented in fulfillment of the requirements for the Degree of Master in Computer Science

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Under supervision of Dr. Kevin Koidl & Dr. Dave Lewis

Submitted to the University of Dublin, Trinity College, May 2016

Declaration of Authorship

I, Tara Matthews, declare that the following dissertation, except where otherwise stated, is entirely my own work; that it has not previously been submitted as an exercise for a degree, either in Trinity College Dublin, or in any other University; and that the library may lend or copy it or any part thereof on request.

Signed: ________________________________

Date: ________________________________

Summary

This work targets rumorous information spread on social media. It is part of the ongoing efforts within rumour behaviour research, where studies strive to *understand* the unique characteristics and nature of rumours.

This investigation is specifically concerned with the *impact* of a rumour and finds opportunity in an inherent gap in the State of the Art, studying impact on social media itself, rather than impact on the lives / beliefs / surroundings of individuals, in a *'real world'* context. The emergence of the Internet and success of social media has provided new mediums for communication, where these social networks are being used for chat, relevant information, and also news. The credibility of the information we read is now of concern.

This work measures the impact of a selection of rumours published on Twitter, rumours involving related messages of questionable veracity, circulating on the social network, Twitter being the chosen social medium for the research. The formula used to calculate impact is constructed as an accumulative score of social network user engagements, *retweets* and *favourites*, in the case of this study's Twitter application. Hence, resultant impact scores are a reflection of the level of engagement a rumour is receiving, its popularity, and success in attraction and uptake, its impact on social media itself.

The study investigates a suite of features as being potentially influential to social media rumour impact. The features experimented are organised into two categories, *author-based* and *message-based*, involving properties related to the composing author, e.g. the number of *followers* and *friends* linked with the account, and the message composition, e.g. the inclusion of URL links or the length of the message.

This research bridges the State of the Art gap, and measures impact of rumours on social meia itself. It signifies the higher impact of some rumours compared to others, highlighted by statistical significance in the data. Findings are presented related to features influential to impact, following a comprehensive investigation. This work encourages an abundance of further study, leading towards a customisable model for measuring impact on social media. It also lends to *'real world'* application, where commercial tools, highlighting high social media impact of rumours, those with potential to *go viral* or do damage, can be further investigated, by news and verification organisations, journalists, governments, etc.

Acknowledgements

I would like to extend my sincerest thanks and appreciation to Kevin Koidl for his continuous and immense guidance throughout the year. Kevin consistently allowed this study to be much my own work, encouraging my ideas and development throughout the research, while always helping me in spots of difficulty, and steering me in the right direction whenever needed.

To my Dad, who's support is difficult to put into words, I must express my greatest gratitude for providing me with everything I could have ever needed, and his continuous encouragement throughout my years of study. This accomplishment would not have been possible without him, and my Mum also. Thank you both.

Abstract

As social media continues to grow, the connectivity between individuals and organisations becomes tighter, and the availability of data becomes more immediate, constant and abundant. Aside from conversational chat, people are using social media platforms to share relevant information and to report news. We must become aware of unbased rumours and protect ourselves against false information and its potential consequences if believed. As information credibility becomes an increasing concern, there rises an important question of rumour impact. This work is concerned with the study of rumour impact on social media itself.

Information spread on social media has a high potential for impact, due to the real-time nature of these media. News organisations are losing their audiences to lies and unverified stories, costing them both money and reputation. Misinformation can also endanger life if adopted by individuals during times of crisis.

This work measures the impact of a given rumour, impact that will be determined by analysing user engagement measures. Analysis is then conducted in an attempt to understand why some rumours make more impact than others. There are two main categories investigated - *message-based feature analysis* and *author-based feature analysis.*

By measuring and analysing impact, this work strives to understand the dispersal potential of rumours. Questions are posed regarding the higher impact of some rumours compared to others, and message- and author-based features are investigated in a bid to answer such questions. This will indicate just how penetrable a given rumour could be, how many people could potentially be reached, how fast this dispersal can happen, and how composition features can influence success, highlighting the power of social media in the life of a rumour. This lends towards a useful tool for news agencies, social media networks, governments etc, in flagging potentially impactful rumours.

Contents

Declaration of Authorship **i**

Summary **ii**

Acknowledgements **iii**

Abstract **iv**

List of Figures **viii**

List of Tables **x**

1 Introduction **1**

1.1 Motivation . . . 1
1.2 Research Question . . . 2
1.3 Research Objectives . . . 2
1.4 Research Challenges . . . 3
1.5 Overview of Technical Approach . . . 4
1.6 Dissertation Outline . . . 5

2 State of the Art **6**

2.1 Background . . . 6
2.2 Rumour Detection . . . 8
2.3 Rumours spread on Social Media . . . 9
2.4 Rumour Impact . . . 10
2.4.1 Temporal aspects associated with Rumour Impact . . . 12
2.4.2 Contributors to Rumour Impact . . . 12
2.4.2.1 Account-based features . . . 13
2.4.2.2 Message-based features . . . 14
2.5 The Twitter Microblogging Service . . . 15
2.5.1 The Twitter API . . . 17
2.6 Related Work . . . 17
2.6.1 Historical Rumour Behaviour Research . . . 17
2.6.2 Rumour Credibility and Annotation . . . 18

2.6.3 Temporal Patterns in Rumour Frequencies 19
2.7 Conclusion 20

3 Design & Methodology 21
3.1 Requirements Engineering 21
3.1.1 Corpus Generation 21
3.1.2 Impact Measurement 26
3.1.3 Impact Analysis 28
3.1.3.1 Account-Based Features 29
3.1.3.2 Message-Based Features 29
3.1.4 Data Storage and Ethics 30
3.2 Design Structure 31
3.2.1 Rumour Gathering, Feature Retrieval and Storage 32
3.2.1.1 Rumour Gathering 32
3.2.1.2 Feature Retrieval 33
3.2.1.3 Storage 34
3.2.2 Impact Measurement and Impact Analysis 35
3.2.3 Data Visualisation Interface (DVI) 36
3.3 Summary 37

4 Implementation 38
4.1 Rumour Gathering, Feature Retrieval and Storage 39
4.1.1 Node.js 39
4.1.1.1 Node Package Manager 39
4.1.2 Filtering 40
4.1.3 MongoDB 40
4.1.3.1 Mongoose 41
4.2 Impact Measurement and Impact Analysis 41
4.3 Data Visualisation Interface (DVI) 44
4.3.1 DVI Landing Page 44
4.3.2 Real-Time Rumour Search 45
4.3.3 Impact Tool 45
4.4 Summary 46

5 Evaluation 47
5.1 Approach 47
5.2 Rumour Gathering Evaluation 48
5.2.1 Rumour Gathering Results 48
5.2.1.1 Discussion 50
5.3 Feature Retrieval Evaluation 52
5.3.1 Feature Retrieval Results 52
5.3.1.1 Discussion 53
5.4 Impact Measurement & Impact Analysis Evaluation 55
5.4.1 Impact Measurement & Impact Analysis Results 56
5.4.1.1 Account-Based Feature Influence 60
5.4.1.2 Message-Based Feature Influence 63
5.4.1.3 Impactful Rumours compared to Non-Rumours 70

5.4.1.4 Discussion 77
5.5 Summary 78

6 Conclusion **79**
6.1 Objective Assessment 79
6.1.1 Rumour Gathering 79
6.1.2 Impact Measurement 80
6.1.3 Statistical Significance 80
6.1.4 Impact Contribution 81
6.1.5 Final Remarks 82
6.2 Research Contribution 82
6.3 Future Work 83
6.3.1 Impact Benchmarks 83
6.3.2 Longitudinal Impact Studies 84
6.3.3 Collective Impact Studies across multiple Social Media services . . 84
6.3.4 Commercial Applications 84

Bibliography **86**

List of Figures

3.1 *System Components* . . . 31
3.2 *Rumour Gathering, Feature Retrieval and Storage* . . . 32
3.3 *Rumour Gathering Components* . . . 32
3.4 *Feature Retrieval Components* . . . 34
3.5 *Storage Component* . . . 34
3.6 *Impact Measurement and Analysis Components* . . . 35
3.7 *Data Visualisation Interface (DVI) Components* . . . 36

4.1 *System Components* . . . 38
4.2 *Impact Measurement and Impact Analysis Implementation* . . . 43
4.3 *Impact Output* . . . 43
4.4 *DVI - Landing Page* . . . 44
4.5 *DVI - Real-Time Rumour Search Page* . . . 45
4.6 *DVI - Impact Tool Page* . . . 46

5.1 *Evaluation Approach* . . . 47
5.2 *Bucket sizes* . . . 51
5.3 *Follower counts in Higher Impact Rumour vs Lower Impact Rumour* . . . 60
5.4 *Friends counts in Higher Impact Rumour vs Lower Impact Rumour* . . . 61
5.5 *Statuses counts in Higher Impact Rumour vs Lower Impact Rumour* . . . 62
5.6 *Verified counts in Higher Impact Rumour vs Lower Impact Rumour* . . . 63
5.7 *Number of messages containing Hashtags in Higher Impact Rumour vs Lower Impact Rumour* . . . 64
5.8 *Number of messages containing Media in Higher Impact Rumour vs Lower Impact Rumour* . . . 65
5.9 *Number of messages mentioning / tagging another user - in Higher Impact Rumour vs Lower Impact Rumour* . . . 66
5.10 *Number of messages containing a URL in Higher Impact Rumour vs Lower Impact Rumour* . . . 67
5.11 *Average length (chars) of messages in Higher Impact Rumour vs Lower Impact Rumour* . . . 68
5.12 **Followers:** The % of experiments where rumour sets contain more followers vs. the % of experiments where non-rumour sets contain more followers. . . 71
5.13 **Friends:** The % of experiments where rumour sets contain more friends vs. the % of experiments where non-rumour sets contain more friends. . . 72
5.14 **Statuses:** The % of experiments where rumour sets contain more statuses vs. the % of experiments where non-rumour sets contain more statuses. . . 73

5.15 **Media Items (MIs):** The % of experiments where rumour sets include more tweets with MIs vs. the % of experiments where non-rumour sets include more tweets with MIs. 74
5.16 **User Mentions (UMs):** The % of experiments where rumour sets include more tweets with UMs vs. the % of experiments where non-rumour sets include more tweets with UMs. 75

List of Tables

5.1 List of rumour buckets and associated rumour stories. 49
5.2 List of rumour buckets, corresponding regular expression queries and the number of related rumour tweets collected. 50
5.3 *Impact Measurement Variables* 54
5.4 *Message-Based Features* 54
5.5 *Account-Based Features* 54
5.6 *Impact Measurement Scores* 56
5.7 *Cases for which the null hypothesis can be rejected, where* $P \leq 0.05$. . . . 58
5.8 List of non-rumour buckets, associated news stories and the number of related non-rumour tweets collected. 71
5.9 15 experiments - Rumours (R) vs. Non-Rumours (NR). 76

Chapter 1

Introduction

In this chapter, the motivation behind the research is introduced. The research objectives for the thesis are identified, together with potential research challenges that may arise. Finally, a technical approach for the research is proposed and an outline illustrating the structure of the document is provided.

1.1 Motivation

The explosion of social media has characterised Internet growth in recent years. Some original social networks included AOL, chat rooms and Live Journal. While many have come and gone, some more notable than others, its clear social networking is no passing trend, with market leaders such as Facebook, Twitter and WhatsApp, boasting millions of users.

Social networks provide an online voice to just about anyone. Anyone with basic access to the Internet can publish their thoughts, opinions, ideas, through online communities. Today's Internet is flooded with *user-generated content*, and opinionated material in particular [1].

Rumours are prevalent in our society. From the home to the office, they influence our beliefs and behaviours toward others and generally affect the way we see the world [2]. Social psychology defines a *rumour* as a story or a statement in general circulation without confirmation or certainty to facts [3]. There is a myriad of research around rumours in a variety of fields, primarily from a psychological perspective [3–5]. However, the advent of the Internet and social media offers opportunities to transform the way we communicate, giving rise to new ways of communicating rumours to a broad community of users [6].

Much research has been pursued, studying the impact of rumours on peoples lives [7, 8]. There is an increasing need to interpret and act upon rumours spreading quickly through social media, especially in circumstances where their veracity is hard to establish [9]. Analysing the potential *impact* of rumours is often as important as checking their truthfulness.

There is an inherent gap in the State of the Art related to the study of rumour impact on social media itself. There exists an opportunity to formally measure the impact of rumours on social media. An impactful rumour is one that has potential to ferociously penetrate its social network, through high volumes of shares and user uptake or belief. This work considers user engagements as a means for measuring impact and will consider impact as an accumulative score of such engagements, e.g. favourites and retweets in the case of Twitter.

Characteristic to rumours, exists an important temporal aspect. Associated with every rumour, is a lifetime. During a rumour's life, it spreads and is received by individuals that come across it. Many studies have been done focusing on the long term spreading of rumours, and the speed of spread [10–12]. These aspects however, are not considered by this research. Some rumours penetrate their network quickly. They appear in a moment and as quickly, are received and known to many. This immediate potential, characteristic to a rumour, motivates the research question central to this work.

1.2 Research Question

This dissertation poses a question of the impact of rumours generated in short form social media, how to measure the impact of a rumour on social media itself, and an investigation of features influential to such impact.

1.3 Research Objectives

To address this research question, five specific research objectives have been defined:

1. To gather a corpus of Twitter messages, a snapshot body of tweets, considered as rumorous.

2. To conduct an impact measurement of a given rumour, by a formula suggested by this work, calculating an impact score.

3. To look for statistical difference between the means of various pairs of rumours, as a method of highlighting the higher impact of some rumours over others.

4. To assess possible contributors to rumour impact by investigation of *message-based features* and *author-based features.*

5. To conduct an evaluation of the approach, assessment of performance implications and offering of suggestions for future work.

1.4 Research Challenges

Alongside the objectives outlined above, there exists a number of significant research challenges that are specific to this problem domain. These must be addressed for this research to be conducted successfully:

1. Rumour is a vast research area, and rumour theory and study is extensive. It is suggested in [13] that not only do models of explanation change, but rumours themselves also change - not only in content, but perhaps in the way they are believed or disbelieved. Social scientific interest in rumours begins with the birth of modern psychology in the 19th century.

   This work adopts a definition of rumour which is maintained throughout this research. Moreover, the challenge of constructing a corpus containing tweets that accurately conform to this definition, must be overcome.

2. Impact is an ambiguous concept. Therefore, a definition of impact that this work will investigate and use for measurement is formalised.

3. The Twitter API will be used for corpus gathering. This is a service that offers public access to tweets made in the public domain. Therefore, there exists the potential for noisy tweets which could pollute the dataset.

4. Constructing a corpus of messages relating to rumorous topics is challenging. To address this, a formal method for rumour detection is decided upon. This study measures and analyses the impact of a snapshot representation of a rumour. By snapshot, the research involves itself in an investigation on a tweet population collected at a given moment in time.

   Therefore, studies must be done with limited sets of data, limited by size and time of retrieval. The Twitter Search API is part of Twitter's REST API and is the precise service this work will use to gather rumour snapshots. The Twitter Search API searches against a sampling of recent Tweets published only in the past 7 days, highlighting the time constraint.

5. While the tweets are gathered from a public service, there are ethical implications to storing large quantities of user data. In particular, opinions expressed by users must be handled in a respectful manner and any personally identifying content present in the tweets should be used responsibly.

   Information Ethics (IE) is proposed in [14]. IE is more impartial and universal than other non-standard ethics. It considers the entire concept of what may count as a centre of moral claims. A form of moral luck arises as consideration for the patient are included, in this instance - the Twitter user. The principles followed and the actions performed must be independent of position as agent (researcher) or patient, in that choices and behaviour would be the same regardless.

1.5 Overview of Technical Approach

To address the research objectives outlined above, a Twitter API server application is implemented. Data is acquired by specific queries related to topical rumours. The properties required for impact measurement, as well as the features required for impact contribution analysis, are parsed from the Tweet object. Customised Tweet data objects constructed of the original Tweet text and feature data extracted, are stored in a data store.

A comparison of the mean impact scores of various pairs of rumour sets is conducted as the potential for statistical significance is investigated, specifically, statistical difference in the impact of rumours. The investigation then develops to an analysis of rumour populations, and their message-based and author-based features, in a bid to flag those features that are influential to the impact of a rumour.

1.6 Dissertation Outline

A discussion and analysis of the related theory and related work is provided in Chapter Two, together with an explanation for decisions made while performing the research. Chapter Three transforms these decisions into engineering and system requirements, and a formal design for the project. The key details of the implementation are given in Chapter Four. In Chapter Five, an evaluation of the research is provided, along with research findings. Chapter Six is the final chapter, in which conclusions are drawn about the research, a summary of research contributions of the dissertation is detailed, and possible extensions and future work suggested.

Chapter 2

State of the Art

In this chapter, the background for the research is investigated. To further elucidate the motivation behind the research, the potential implications of formally measuring rumour impact are discussed in detail.

The fields of rumour detection and rumour spread are presented, with discussion of key concepts and principles relevant to this research pursuit. Next, *rumour impact* is formally introduced, and a definition provided, the definition that is adopted for this investigation.

The microblogging service, Twitter, is given brief introduction and subsequently justified in its choosing for this research. Finally, research that is closely related from the field of rumour impact within rumour study is reviewed, in the context of the research question.

2.1 Background

Quine questions definition [15], stressing that synonymy required to establish analyticity cannot be obtained from definition. Quine states there are three types of definitions. This study poses a definition of rumour, a pre-existent concept, that shall be explicated to give a more concise, well-defined definition, in the context of this research.

Rumour is a popular research area among a number of varying fields, including psychological studies [16] and computational analyses [17]. Defining and differentiating rumours remains an active topic of discussion [18]. Researchers have attempted to formally define rumour, so as to address the lack of common understanding around the specific categorisation of what is or is not a rumour.

The Oxford English Dictionary (OED) defines a rumour as, *A currently circulating story or report of uncertain or doubtful truth.*[1] A definition of rumour is adopted for the work, as supplied by [18], constructed as an extension of the OED's definition, but with additional descriptions from rumour-related research, providing a more appropriate definition for this research. Thus, the definition embraced for a rumour is *"a circulating story of questionable veracity, which is apparently credible but hard to verify, and produces sufficient skepticism and/or anxiety"*.

Impact analysis is part of a wider community of research, which can be described as rumour behaviour. To study rumour impact means to concern oneself with the potential or consequence of a given rumour. In researching rumour impact, the aim is to understand something about the behaviour of rumours. Moreover, this work strives to understand whether or not a rumour in question will lend to subsequent events of controversy, stress, strife, fatality, etc., i.e. high consequence.

By studying rumour impact, this work targets rumorous information spread on social media, a common goal of Internet users, researchers and news agencies. Governments are also being forced to involve themselves, the Japanese government being one recent example, when they warned about unverified information after the Great East Japan Earthquake in 2011 [19]. Social media allows users to share crisis information with others. Yet, if information spread is false, social media will become a rumour-mill and panic may result [20].

A well-established area of research involves rumour classification and credibility, interested in whether a rumour is true or false, and its believability. Castillo, Mendoza and Poblete focus on automatic methods for assessing the credibility of a given set of tweets [21]. Instead of classification of rumours in terms of truthfulness, what if we could classify impact, identifying those rumours that have potential to do damage or *go viral*, thus being witnessed and possibly believed by many.

This work seeks to address the problem of automated impact determination, classifying rumorous messages in terms of an impact score. The study asks how impactful a particular message is, within social media, with the intention of flagging those messages representing high impact. Research, such as this, leads towards a model that aims to be useful for news agencies, governing bodies etc. in noting messages that have potential for high consequence. Rumorous messages that are low impact - likely to be discarded

[1] http://www.oxforddictionaries.com/definition/english/rumour, accessed: 15/05/2016

and unmemorable are naturally ignored, whether or not they are true or false. Rumours classified as being of higher impact can be subsequently investigated.

2.2 Rumour Detection

Rumour analysis, such as determining rumour impact, is a very challenging task and is not possible without first retrieving a complete set of social conversations (e.g. tweets) that are actually about the rumour. For the research, this first step is taken in the study of social media rumours, and an appropriate dataset that includes a diverse set of stories is identified.

There is a myriad of research involving rumour identification. As part of the *PHEME*[2] project, [18] introduces a method of manually annotating rumours (by people) by reading through the timeline of tweets related to an event and selecting stories that meet the characteristics of a rumour, by the rumour definition they propose.

For the task of annotation, the researchers developed a rumour annotation tool that visualises the timeline of tweets associated with an event. Annotators do not necessarily have prior knowledge of the rumours associated with a given event. They must step through the timeline presented and select whether each tweet is a rumour (a green tick), not a rumour (a red cross), or mark with an orange question mark if they are unsure.

Twitter's Streaming API[3] was used in this research, where a set of keywords were set, and used to filter tweets related to a certain ongoing event. The researchers were therefore listening to live updates from Twitter's public stream and collecting large datasets. Given the size of the collections of tweets, the researchers filtered those tweets that sparked a significant number of tweets, which they assumed to relate to a significant amount of user interest. This made the task of annotation more tractable for the annotators

The *RumorLens* project[4] [22] includes a *Rumour Detector* component. The Detector mines rumours from tweets obtained through the Twitter Garden Hose API (a 10% sample of the overall tweet stream), and produces a continuously-updated, ranked list of tweet clusters that appear to be rumours. The technique the Detector used involves

[2]http://www.pheme.eu, accessed: 15/05/2016
[3]https://dev.twitter.com/streaming/overview, accessed: 15/05/2016
[4]https://www.si.umich.edu/research/research-projects/rumorlens, accessed: 15/05/2016

searching for a set of expressions commonly employed in the flagging of controversial claims but rarely otherwise, e.g. "Rumour has it.., Is this true?", "X is rumoured to..". The authors argue that this method produces higher recall than baselines based on overall trending topics or trending hashtags [23].

The RumorLens community website includes an interface, with an upvote/downvote mechanism that allows users to refine the output of the Rumour Detector component, upvotes indicating that users believe a given candidate rumour is worth investigating further. This allows feedback to be generated and accrued, and aids the Detector in becoming more and more accurate in its identification of rumours.

A novel approach is proposed in [17], in which the authors have a list of rumour examples, and a corresponding list of regular expression queries used to collect rumorous data from Twitter. Handcrafted regexp related to a known rumour are submitted to Twitter and sets of tweets related to that rumour are returned. The goal in this work is to collect and annotate a large dataset that includes all of the tweets that are written about a rumour in a certain period of time.

The aims of the work last discussed, are most akin to this study. However, this study will be within the bounds of a *snapshot* - a particular moment, rather than any extended length of time. For the research last discussed, the authors investigated five rumours. The total populations of tweets received varied quite drastically, with one rumour dataset having a total of 4975 tweets, and another having only 215. This hints at the flimsy nature of the data and associated constraints that will exist within this research investigation.

2.3 Rumours spread on Social Media

This work involves itself with a specific type of rumour, those spread online through social media platforms. Understanding rumours has been the subject of research in psychology for some time [3, 24, 25]. However, research has only recently begun to investigate how rumours are manifested and spread differently online.

Behaviour research and analysis of this type of rumour is of huge importance. While the spread of inaccurate or questionable information has always been a concern, the emergence of the Internet and social media has exacerbated the problem, as it facilitates the spreading of such information to large communities [26].

Highly influential research [11] investigates the dynamics of a rumour process that takes place on top of complex heterogeneous networks. Many systems that are seemingly diverse, such as the Internet, the World Wide Web (WWW), metabolic and protein interaction networks, and food webs, actually share many topological properties [27]. Perhaps the most popular shared property is the Small-World (SW) property, known as "six degrees of separation" [27, 28] - one can go from one node (or element) of the network to another node passing by just a few others.

In Scale-Free (SF) networks, such as social networks, an epidemic disease will pervade regardless of its spreading rate [29]. This might be bad news for epidemiologists and those fighting natural and computer viruses. However, in other applications and networks, it is desirable to spread the "epidemic" as fast and as efficient as possible. Some important examples are marketing campaigns, where marketers deem rumorlike strategies to be most favourable for viral marketing, and social media outlets where rumour composers aim for rapid and immense diffusion.

Rumours have a momentary aspect to their character. They are likely and quite often are, the topic of big news or scandal - but only for a moment. They are replaced and forgotten. Rumours can be adopted and believed by many, but their appreciation or uptake is typically incomparably huge, and then it disappears. Social media has facilitated this momentary nature but rapid dispersal of rumours, as it is intrinsically momentary itself - a source of constantly updated, real-time, conversation, news and information.

2.4 Rumour Impact

Rumours are such in that they may be untrue - in part, or in entirety. The impact of rumours in the *real-time web* is extensive, especially if representative of false information. This research studies this notion of impact, and calculates an impact score for sample sets of data related to rumours spreading through social media.

Rumour behaviour including impact study, have been popular research topics for some time. Much research have looked to analysing message characteristics and contextual data. In [7], the impact of identification and disidentification on rumour belief is examined. The research shows that a person's level of identification with the rumour object influences belief in the rumour. By experiment, the authors found that a variation in identification influences the impact of a rumour on an individual's beliefs.

In an organisational or political context, rumours can be especially problematic and potentially dangerous for a company's, party's, or candidate's reputation if they contain negative information about the object of focus [7]. During organisational change, rumours of layoffs, closures, or mergers may create mistrust and lower morale [8].

A novel approach is introduced in [30], which identifies rumours based on temporal, structural and linguistic properties of rumour propagation. Experiments conducted in [20] showed that posting URLs in disaster-related tweets increased rumour-spreading behaviour.

The *RumorLens* project [22] also provides a means for diffusion analysis, allowing an assessment of the potential damage, in terms of how many people tweeted about the rumour or were exposed to it. This project is most like this work, as its aim is to provide analysis on the spread of a rumour through its network. However, to fulfill its goal, *RumorLens* estimates impact only through a state-transition-based visualisation. User states - such as exposure to the rumour - are treated as nodes, and flows between states as links. The system looks at who tweeted about a given rumour.

There is a clear gap in the study of rumour impact. Much investigation has been completed, focusing on impact on a person's beliefs or the personal effects on a person's life, but impact investigation in terms of effects on social media itself is lacking. By *impact* of a rumour, this work refers to its success in attraction and engagement of users in its network. A model by which to measure impact in real-time would be invaluable to many groups of people and organisations. An understanding of the possible contributors to rumour impact, such as message features and composer characteristics, would also be highly useful.

On major social networking sites, Facebook and Twitter being the two most popular [31], users show their appreciation or agreement to a post (Facebook) or tweet (Twitter) with likes (Facebook) / favourites (Twitter) and shares (Facebook) or retweets (Twitter). These can be described as user engagement measures and are central to impact scoring or measurement. The main user engagement measures relevant for this study are:

Retransmission: Retransmission / forwarding / retweeting in the case of Twitter, is powerful in impact calculation. By retransmitting a message, the user directly forwards this post to his or her own followers, and therefore extends message range and dispersal. Retransmission / forwarding of a message, can be thought of as a sign of value. A follower / friend / connection has found your message valuable enough that they wish to share it with their audience.

'Likes', 'Favourites': Notable social media networks such as Facebook and Twitter allow users to *like*, *favour* or show their positive sentiment towards a given message event. A tweet with a high *like* or *favour* count lends to impact as it has been appreciated by a high number of users, and thus its uptake and spread will grow. Likes / favourites can be thought of as a sign of appreciation. Your message resonated with someone else, and they wanted to give a virtual high-five.

Replies: Users may also engage with a tweet or post by replying to the author. This is another field that would represent user engagement. Reply information by way of reply counts are not a data field available through the Twitter API, and has been a frustration to much research across many fields.

2.4.1 Temporal aspects associated with Rumour Impact

It is worth noting that there exists an important temporal aspect to be considered in measuring the impact of rumours. Monitoring a sample selection of tweets over some period of time could reveal important information regarding rumour impact.

However, this will be out of scope for this work, as it would take longer than the time available for this research, and would require a long term study dataset which is not in possession. Moreover, the idea of impact measurement in real-time data, on the real-time web is both exciting and inarguably useful for news organisations, governments etc. for subsequent investigation, verification or debunking. Real-time impact analysis translates to fast veracity investigation and ultimately results in purer news and online information.

2.4.2 Contributors to Rumour Impact

Once impact has been measured, it is essential to further analyse and discuss the potential contributors, if we aim to truly understand this notion of impact. The following properties form a suite of features that are potentially inherent to impact analysis. These features are recognised as being useful in much research involving rumour behaviour analyses [20, 21, 30, 32]. These features can be grouped into two sets - *message-based* *features* and *account-based features*, and may be influential in the level of audience engagement and the impact of a rumour.

2.4.2.1 Account-based features

The following account-based features have been identified as essential for this research:

Author's account: This work is interested in looking at account properties of the composing author, how long a user has had their account, and how popular and active they are on the network. The study looks at when the account was created, whether the account is personalised or default, the number of message posts, and those connected with or the community associated with the account, i.e. friends or followers and followees.

Findings of [21] indicate that more active users tend to spread more credible information, as well as users with newer user accounts but with many followers and followees. Believability has the potential to lend to rumour impact. The number of followers a user has would indicate the popularity of this user, and their immediate network potential, which is of direct interest to our impact measurements.

Hubs make the news available to a big audience, whereas average users quickly convey the information from one neighbour to the next [12].

Verification: Verification is associated with genuineness, verfication that a particular profile in fact belongs to whomever it says it does. The verified badge will appear next to a verified user's name on their profile page, and is used by social media networks such as Twitter and Instagram. Verification is used to *"establish authenticity of identities of key individuals and brands on Twitter... highly sought users in music, acting, fashion, government, politics, religion, journalism, media, sports, business and other key interest areas"* [33].

Katy Perry is the most followed user on Twitter with over 70 million followers [34]. The average verified user has a very high average number of followers, of just over 125,000. In contrast to this, the average number of followers among users across all of Twitter is 208, but 81% of users have fewer than 50 followers [35].

People look to key individuals for a plethora of reasons including inspiration, entertainment, news, anecdotes, information on a topic of which they are known to be knowledgeable, etc. The network potential linked to a verified user could be massive, and could ferociously increase rumour impact on social media, if a given message is associated with a rumorous topic, and spread by the verified user in question.

2.4.2.2 Message-based features

The following message-based features have been identified as essential for this research:

Entities: These are parsed out of the tweet - hashtags, urls, user-mentions, media. In [21], hashtags, urls, and user-mentions are included in their set of *message-based features* which are useful in determination of message credibility. Hence, these entities could also lend to rumour impact, as a rumour appearing more credible is likely to be more believable, increasing user uptake, and subsequent social media impact.

Recent findings by *HubSpot*[5], a marketing and sales platform that helps companies attract visitors, indicate that Tweets with images received 18% more clickthroughs, 89% more favourites and 150% more retweets. Tweets with hashtags are 55% more likely to be retweeted. Link clicks are the biggest way users interact with content, accounting for 92% of all user interaction with Tweets [36].

Question and Exclamation Marks: Question and exclamation marks are also included in the set of *message-based features* in [21]. Asking a question or showing excitement could lend to post engagement and rumour impact.

Quotes: Tweets quoting someone else or information from a news article for example, could be interesting to analyse for impact.

All Caps: A virtual way of shouting. Messages that supply a sense of urgency or drama, could also lend to impact.

Length of a Tweet: It is suggested on many online resources that there is an ideal Tweet length. One such study suggests an ideal length of 100 characters or 120-130 characters if including a link [37].

[5] http://www.hubspot.com/, accessed: 15/05/2016

2.5 The Twitter Microblogging Service

Twitter[6] is to be considered as the microblogging service that will facilitate rumour retrieval and subsequent impact investigation.

Twitter is an online social networking service. Twitter users publish and read short 140-character messages called *tweets*. Registered users can both read and post tweets, but those unregistered can only read them. The Twitter service is accessed through the website interface, SMS or mobile device app [38].

Users may tag their tweet by using a hashtag, and they may target their tweet at a particular user by using an amersat (@user), providing some meta data to the content of the tweet. The Twitter field guide provides a full description of tweet functionality [39]. The simplicity of a tweet lowers time cost and thought burden for the user, a major driving factor behind Twitter's phenomenal growth since its inception in 2006 [40].

Twitter is listed as one of the ten most-visited sites [41], and has been described as *"the SMS of the Internet"* [42]. There are more than 500 million tweets published per day and 200 billion per year, from 320 million active users [43], with only 11% of these having protected accounts [44]. Therefore, not only is the volume of data published by Twitter users massive, the vast majority is also in the public domain.

Based on this state of the art study, a rumour in the context of this thesis has the following characteristics:

1. It is a circulating story.
2. Its veracity is unclear.
3. It has an appearance of clarity, but verifying such is challenging.
4. It instills some sense of anxiety and/or scepticism.

Twitter rumours are akin to how rumours function in the real world, and adhere to the rumour characteristics mentioned above. *YikYak*[7] is a location-based social network that helps people discover their local community. It is another social media service that facilitates the creation and spreading of rumours. It allows users to *"share news, crack*

[6] https://twitter.com/, accessed: 15/05/2016
[7] https://www.yikyak.com/about, accessed: 15/05/2016

jokes, offer support, ask questions, and interact freely". YikYak connects people located in close proximity to each other. It aims to create a community, *"it's home to the casual, relatable, heartfelt, and silly things that connect people with their community".*

Snapchat[8] is yet another popular messaging app that provides means for rumour progation. It is used daily by 100 million users. Snapchat allows its users to send picture, video and messages that self-destruct after a few seconds.

Twitter focuses on network effects that are more prone to the behaviour of rumours in a real-world context. Twitter messages are short, concise in their maximum 140 characters, and unlimited by any lexical ordering restrictions. The rapid growth of Twitter has made it possible for rumours to spread more quickly. The service rapidly gained worldwide popularity after its launch in July 2006, with more than 100 million users posting 340 million tweets a day by 2012 [45]. Twitter allows for small pieces of information to be spread quickly to large audiences [46], allowing rumours to be created and spread in *real-time*, circulating through a wide network, that is largely public.

As mentioned above, Twitter now reports more than 500 million tweets published on its network per day. This incredible volume, characterised by Twitter's popularity and availability to any Internet user, has helped to enable unreliable sources to spread misinformation. Twitter is therefore an excellent case to analyse stories of unclear veracity in social media.

Characteristic to rumours is the sense of anxiety and scepticism they inflict on their witnesses. As stated previously, when misinformation is spread, social media is at risk of becoming a rumour-mill and panic results [20]. Research introduced anxiety as a key element in rumourmongering. For example, anxious students were more likely to report that they heard a rumour [47]. In addition, the likelihood of sharing a rumour was associated with how anxious the rumour made people feel [48, 49].

Twitter rumours may be initiated by any user, and may take a sufficiently long time to verify or quash. Validation may never occur. Inherent to rumours forged on Twitter is just how incredible their potential to disperse and infect is, and how rapidly this can occur. Anxiety and scepticism are therefore central to Twitter rumours, as we are likely to read a version of an originating rumour that has spread out from its source at ferocious speed, and we are not guaranteed any assurance regarding composing authors.

[8]https://www.snapchat.com/, accessed: 15/05/2016

2.5.1 The Twitter API

The merits associated with the *Twitter API*[9] must be mentioned. The Twitter API has been accepted as a useful source for collecting data to be used in exploring a number of problems related to natural language processing and information diffusion [50]. This is fundamentally due to its constant update of new posts and public API. The diversity of Twitter users and tweets, and the public availability to all of this data, make this corpus especially valuable.

This API is the only API that returns results from the entire public Twitter stream and not a randomly selected sample. There are some concerns regarding the rate limit enforced by Twitter.

2.6 Related Work

Popular interest in rumour never seems to wane. Psychological research has been cyclical for many years, while technological research is of very recent interest, where the birth and success of the Internet and social media gives rise to new ways of communicating rumours to large audiences.

With the growing popularity of online social networks and their information propagation potentials, the ability to understand and control the type of information that propagates in the network has become ever more important [30]. There has been a plentiful amount of research conducted related to rumour behaviour as a result.

This work fills a specific gap, studying and measuring the impact of rumours on social media. It is part of the ongoing research efforts performed under the topic of rumour behaviour. Prior discussion provides information regarding other works in *Rumour Impact*. Further exploration into *Rumour Behaviour* analyses is now conducted, and a wider spread of related works discussed.

2.6.1 Historical Rumour Behaviour Research

Rumour research is a topic of historical interest, as it is a problem central to human psychology. World War II saw a burst of interest in the psychology of rumour and

[9]https://dev.twitter.com/overview/api, accessed: 15/05/2016

rumour control [16]. Seminal work was completed in [3], the impetus for which being the authors' concern regarding dangerous rumours that were affecting community morale and national safety - causing needless alarm, and raising extravagant hopes. The next decade witnessed some developmental research [51, 52], which was then followed by a period of inactivity and quiet.

During the 1960s and 1970s, another cycle of interest emerged [16], with famous publications such as a book of sociological background [53] and a report on civil disorders [54], along with other important works [5, 55, 56]. More recently, there has been another round of rumour behaviour research [57–59].

Works mentioned thus far are of a psychological and sociological background, and the intent in mentioning them is to elucidate the importance of this research topic historically. From a technological background, we now concern ourselves with rumour behaviour research as we find ourselves communicating in ways non-traditional, through new mediums, services that we must now strive for understanding, in terms of communicative power, and rumour-mongering potential.

Traditionally, rumour dynamic was studied employing a one-way communication paradigm resembling the telephone game [16]. Influential research conducted of late have studied the dynamic of rumour in discussion groups [60–63], such as a chat group discussion of a rumour in cyberspace over a 6-day period. They have uncovered systematic patterns in both the content and level of individual participation. This offers consistency with the theoretical claim that rumour-mongering is a *"collective, problem-solving interaction that is sustained by a combination of anxiety, uncertainty, and credulity"*. [63].

2.6.2 Rumour Credibility and Annotation

There has been research conducted, focusing on the information credibility of news propagated through social media networks, Twitter for example [21]. It has been shown by many studies that services such as Twitter, are being used to spread misinformation and false rumours, often unintentionally.

Credibility is described in [64] as a *perceived quality* composed of *multiple dimensions.* In [21], credibility is used in the sense of believability: *"offering reasonable grounds for being believed"*. The main objective in this work is to implement automatic assessment of the level of credibility of content posted on Twitter, with the aim of aiding users to assess credibility of information presented to them.

Investigations have been made into constructing schemes for annotating messages with respect to how they contribute to rumorous conversations. In [32], an annotation scheme for capturing rumour-bearing conversational threads is presented, with the main goal being to cover three aspects of the tweet conversation that are key in determining the veracity of a story:

"(i) whether a post supports or denies a story, as used in previous work [32],
(ii) the certainty with which the author of a post presents their view, and
(iii) the evidence that is being given along with the post / statement to back up the author's view".

Other works that have addressed the problem of classifying rumorous tweets include [7, 65–67]. However, these works addressed the task of establishing veracity by looking only at whether individual tweets supported or denied a rumour.

2.6.3 Temporal Patterns in Rumour Frequencies

The problem of modelling the temporal nature of social media explicitly has received lesser attention than other areas of rumour behaviour studies. Interesting work was performed in [68], a study which introduced the problem of modelling frequency profiles of rumours in social media.

The authors demonstrated that through their methods of using joint modelling of collective data over multiple rumours, using multi-task learning, they were able to recognise and predict commonly occurring temporal patterns. They also showed how text data from social media posts added important information, a motivation and aid for much rumour related research, including this thesis.

Another study concerned with the temporal nature of social media involved modelling hashtag frequency time-series in Twitter via a Gaussian-process [69]. Both studies discussed aim to help with identifying those rumours, which, if not debunked early, will likely spread very fast. This is a common concern of much research in rumour dynamics, and is a motivating factor for this thesis.

As mentioned previously, effective early warning systems are of interest to government bodies and news outlets, who struggle with monitoring and verifying social media posts [68], during emergencies and social unrests for example.

2.7 Conclusion

In this chapter, the background and State of the Art were discussed, in order to address the research question. The fields of rumour detection and rumour spread were presented, with focus on concepts and principles central to this research pursuit.

Next, the notion of impact within rumour study was introduced and formally defined. Discussions were made involving what impact entails for this work. The microblogging service Twitter, was then detailed, and justified its popularity for rumour studies.

Finally, research within the area of rumour behaviour analysis was discussed, focusing on the timeline of historical research and the latest motivations, rumour credibility and annotation schemes, and temporal patterns in rumour frequencies. These are studies that closely relate to this work in their concerns and aims.

In concluding the State of the Art, there appears a substantial gap in the area of rumour impact study. While much investigation has been carried out, analysing impact in terms of an individual's belief or on an individual's life, impact in terms of effects on social media itself has yet to be explored.

The goal of this work is to collect and analyse in terms of our impact definition, a suitable dataset of Twitter rumours, written in a certain period of time. An inspection and review of the possible contributors to rumour impact will also prove effective and useful.

Chapter 3

Design & Methodology

Chapter Three provides a detailed description of the design decisions for the research, building on the foundations and stemming from the related work reviewed in Chapter Two. Firstly, a set of requirements are gathered to ensure that the research objectives and challenges outlined in Chapter One are met. Subsequent sections then present the design solution and provide a discussion of the implications of the decisions made during the design process.

3.1 Requirements Engineering

This research involves a complex set of research objectives and challenges. As a result, the requirements are considered in sets relating to specific aims.

3.1.1 Corpus Generation

The first set of requirements relate to the process of corpus generation, detailed by the research objective: *Gather a corpus of Twitter messages, a snapshot body of tweets, considered as rumorous that can be used for impact analysis.*

Snapshot Gathering

The system must be capable of generating a corpus representing snapshot collections of messages.

It is important to carefully define what is meant by the term *snapshot*. This refers to a temporal consideration in the collection process. This research involves itself in an investigation on a message population collected at a given moment in time.

As discussed in the State of the Art, there exists an important temporal aspect to be considered in conducting any study related to rumour behaviour. In general terms, there are two approaches to consider:

1. Monitoring a specific sample of messages over an extended period of time. A gathering of messages is conducted, and then watched and analysed over various intervals during a set amount of time.

2. Collection and analysis in real-time, at any given moment in time. This approach involves a *tap* on the message resource, the Twitter public stream for example, and a collection of the messages readily available at that moment, later to be used for investigation.

There are strengths and weaknesses to both methods. The first approach is inarguably interesting and potentially could reveal important information regarding rumour impact, as the investigator could monitor how a set of rumorous messages progress, in terms of diffusion and/or user engagements. However, it requires dependence on a set group of data, a dataset that would need to be relatively large. It also requires a large period of time, with annotation / analysis at hopefully advantageous intervals. Scheduling and annotation would require much manual involvement.

The second approach requires immediate investigation, with only the data and information available at that time, the *snapshot*. It could be less reliable, and there is no set of progression results to be compared or investigated later. The large dataset and long period of time required by the first approach are out of scope for this research, as neither are available or feasible.

Moreover, the characteristic of immediacy fits well with this work. The idea of impact measurement in real-time data, on the real-time web would be invaluable to various news companies and governing bodies, supplying them with a tool to flag those rumours worthy and needing of further resources and investigation - those deemed impactful at that moment in time. Real-time impact analysis translates to fast veracity investigation and debunking if necessary. Its ultimate vision is purer news and online information. For these reasons, the second approach will be adopted for this research.

Rumorous Corpus

The corpus gathered by the system must consist of messages deemed as rumorous, detected in accordance with the rumour definition decided by the work.

Rumour detection is an extensive field of research on its own. It is important for this work, that a method for detection is intuitively decided upon, to allow for the subsequent impact measurement and analysis that is the core interest and aim of this research. In summary of the methods discussed in the review of related work, the approaches to rumour detection considered fall under three individual methods:

1. Detection by human judges performing manual annotation. This involves a visual timeline / list of messages with options for the human annotator to choose from in marking each message, e.g. *rumour, not a rumour, not sure.*

2. Detection by searching for a set of expressions commonly associated with rumours. These are expressions often employed in the flagging of controversial claims but rarely otherwise, e.g. "Is this true?", "X is rumoured to..".

3. Detection by searching for a handcrafted regular expression relating to a known rumour, a rumour deemed as such by complying to a formal definition.

There are situations in which each method would be suitable. It is important to consider the circumstances present, resources available, and ultimate goals of research when choosing a detection method. The first method requires heavy involvement from human participants. There is resultant consideration of the expertise of the individuals required. This approach suffers from infeasible scalability as every run of the system and endeavor of rumour collection, requires human participants to be available before any real work is done.

The second approach is interesting, and seems to lead to quite a substantial dataset. It is a simple approach to detection, and is attractive if seeking one large body of rumours. However, there is an issue of segmentation or separation associated with this approach. How does one separate this dataset into any sub groups for worthwhile comparison and investigation.

In contrast to this, the last approach supplies natural separation. Once a definition for rumour has been decided upon, known rumours that comply with this definition can be used in rumour search. Choosing keywords, the words associated with the controversial aspects of a rumour, a regular expression can be generated and submitted to the message source (e.g. the Twitter public stream), for the retrieval of messages associated with the rumour. For example:

Rumour: *"The movie 'The Notebook 2' has started filming."*
Keywords: 'Notebook 2', 'Notebook sequel'
Regular Expression: Notebook & (2 | sequel)

In this example, messages related to the rumour, *"The movie 'The Notebook 2' has started filming."*, are expected to be collected. To collect messages related to another rumour, the same process is followed - choosing keywords and crafting a regular expression. This creates unique groups of rumorous tweets relating to specific rumours.

This work defines a unique group of rumorous tweets relating to a specific rumour, collected by and for this study, as a rumour *bucket*. This detection approach will be adopted by this work, as individual buckets of rumours are required, with which impact will be measured and analysed.

Corpus Size

The system approach to corpus generation must be scalable.

One of the issues highlighted by the review of related work was the varying bucket sizes that are a likely reality when collecting data as flimsy as that associated with rumours. There are also rate limits enforced by Twitter in retrieval from their API which also makes for corpus size considerations.

As this work interests itself in those rumours that are in circulation at a given moment in time, many rumour gathering searches must be performed, each search relating to a different known rumour. There could be many depending on how many times searches are performed, and how many rumours are tried. Therefore, the system must be equipt to perform new searches continuously and efficiently, adding rumour buckets to a scalable datastore. This requirement is applied as a principle to all design decisions.

Corpus Quality

The corpus generated by the system must be of good quality.

The quality of the corpus relates to a research challenge detailed in Chapter One, an issue raised as the services of the Twitter API are utilised. The Twitter API operates in the public domain and therefore has the potential to be 'noisy'. The type of 'noise' that concerns this work are those tweets that are duplicated or unrelated to the rumour in question.

Duplicates will be avoided where possible in requests to the API, or removed if still finding their way through. As the Twitter Search API only allows retrieval of data up to 7 days old, and the rumours searched for are those present, in circulation on a given day, the danger of receiving tweets unrelated to the rumour is not high, once a carefully crafted regular expression is chosen. However, unrelated message retrieval is not impossible, messages matching the regular expression, but not related to the rumour in question. For example:

Rumour: *"Khloe Kardashian's father is not who she says it is."*
Keywords: 'Khloe Kardashian', 'father'
Regular Expression: (Khloe Kardashian) & father

Unrelated Rumour Example:
"Khloe Kardashian is having a baby! Who is the father?"

Related Rumour Example:
"Rob Kardashian isn't Khloe Kardashian's father."

This work assumes that a very small percentage of unrelated messages will be received, less than 5%, as searches are performed on popular stories of the day, in retrieval of data available in a seven day window only. This assumption will be asserted by periodic observations of tweet collections, ensuring those deemed as unrelated accumulate to less than 5%. This will have little or no effect on impact measurement.

Corpus Composition

The system must exclude retweets, messages that have been directly forwarded by another user.

This work is interested in the various messages associated with a known rumour. Retweets represent messages directly forwarded from another user. If a large number of people retweet the same rumour message in quick succession, and the API is tapped for tweets related to the rumour in question, lots of the same tweets will be received. These will not be flagged as duplicates as they are not from the same user and have different ids. However, for our research, they are duplicates, and must be ignored to allow us a varying representation of the messages related to a rumour. Therefore, messages that are retweets will be excluded.

Preparation for Impact Measurement

The system must be capable of parsing properties required for subsequent impact measurement, from Tweet objects received.

This work is interested in measuring the impact of a rumour on social media itself. The study suggests a formula that will be built using properties that are adopted as defining rumour impact, properties that are available from Tweet objects. This work tends towards a customisable model for impact measurement that can be applied over other social media platforms, using other properties depending on the specific case or application. The properties necessary for impact measurement will be parsed from Tweet objects received. The system will store these properties with each data item (each rumour message).

3.1.2 Impact Measurement

The next set of requirements is related to the measuring of impact, detailed by the research objective: *Conduct an impact measurement of a given rumour, by a formula suggested by this work, calculating an impact score.*

Impact Calculation

The system must be capable of computing an impact score, by the formula suggested for the work.

There is an inherent gap in the state of the art related to the study of rumour impact on social media itself. Following the review of related works which focused on impact on the lives of individuals, their beliefs, their social environments, there boasted an opportunity to formally measure rumour impact, and stand as a stepping stone towards a generally applicable and customisable model for such measurement involving solely impact on social media.

This research introduces a formula for calculating impact on social media, and by this formula, will measure the impact of a variety of rumours.

To measure impact, this work suggests and implements the following formula:

$$impact = retweets + favourites \tag{3.1}$$

By impact of a rumour, we refer to its success in attraction and engagement of users in its social network. This formula comprises jargon specific to Twitter. However, this formula could be applied generally across other social media platforms. For example, 'shares' and 'likes' on Facebook.

Users show their appreciation or agreement to a tweet by *favouriting* or *retweeting*. These can be described as user engagement measures, as by these methods, users engage with a tweet that resonates with them. Twitter users may also reply to a tweet. This too is a form of user engagement. However, reply information by way of reply counts are not a data field available through the Twitter API, as outlined by the State of the Art review. For this reason, replies are not included in the impact formula. Retweets and favourites are inherent to impact measurement as these offer a direct reflection of user appreciation, and the popularity of a given tweet.

Impact is calculated by retrieving all messages stored in a specific rumour bucket, and performing the calculations necessary, resulting in an impact score reflecting the above formula, (3.1).

3.1.3 Impact Analysis

The next set of requirements is related to the analysis of impact, subsequent to impact scoring, detailed by the research objectives:

1. *Look for statistical difference between the means of various pairs of rumours, as a method of highlighting the higher impact of some rumours over others.*

2. *Investigate possible contributors to rumour impact by investigation of message-based features and author-based features.*

Statistical Difference

The system should provide t-test functionality, allowing for a two-sample location test of the null hypothesis such that the means of two rumour populations are equal.

A t-test is a statistical hypothesis test that can be used to determine if the variances of two sets of data are significantly different from each other. Rumour sets that are statistically different from each other can be taken as statistically significant, rejecting the null hypothesis[1]. Therefore, the means of the rumour populations are not equal, with one representing high impact, in contrast to the other set.

In order to make sense of the rumour data collected, in terms of separating and scoring rumours according to their social media impact, and to perform subsequent impact influence analysis, the approach that will be adopted by the study is to perform a t-test on two sample sets of rumours, of a specified size, randomly chosen from their respective rumour populations. This will supply us with a T-value. Given a T-value and the degrees of freedom, *(sample size * 2) - 2*, a P-value is obtained. A significance level, α, the probability of making a Type I error is set to be small - 0.05, the most widely used significance level. The P-value is compared to α.

1. A small p-value (≤ 0.05) indicates strong evidence against the null hypothesis, so it is rejected.

2. A large p-value (> 0.05) indicates weak evidence against the null hypothesis (fail to reject).

[1] http://www.socialresearchmethods.net/kb/stat_t.php, accessed: 15/05/2016

Influences to Rumour Impact

The system should investigate a set of features (message-based) and (account-based) in comparison of rumours flagged as being statistically different, attempting to find those features that influence impact.

By the same method used to parse the properties required for impact measurement, the system will collect a wide range of features that will be stored with each rumour data item, and later retrieved for impact contribution analysis. The features that are collected for each rumour item are as follows:

3.1.3.1 Account-Based Features

This work is interested in looking at whether account properties of the composing author influence the impact of a rumour.

1. **user.created_at:** The year the account was created identifying how long a user has had their account.
2. **user.default_profile, user.default_profile_image:** This identifies whether a user has personalised their account or is using default settings.
3. **user.followers_count:** The number of people following this user.
4. **user.friends_count:** The number of people this user follows.
5. **user.statuses_count:** The number of messages this user has posted.
6. **user.verified:** Identifies key individuals such as celebrities, journalists, politicians etc. Boolean.

3.1.3.2 Message-Based Features

This work is interested in looking at whether message properties, features of the rumour tweet itself, are influential to rumour impact.

1. **entities.hashtags:** Boolean. True if message contains one or more hashtags.
2. **entities.media:** Boolean. True if message contains one or more media items, such as an image.

3. **entities.urls:** Boolean. True if message contains one or more url links.
4. **entities.user_mentions:** Boolean. True if message contains one or more user mentions, e.g. @user.
5. **'retweet':** Boolean. True if message contains the word 'retweet', e.g. "please retweet...".
6. **All Caps:** Boolean. True if the message is all uppercase.
7. **'?':** Boolean. True if message contains a question mark.
8. **'!':** Boolean. True if message contains an exclamation mark.
9. **' "" ':** Boolean. True if message contains opening and closing quotation marks.
10. **message.length:** The length of the message.

3.1.4 Data Storage and Ethics

The final requirement does not apply to any specific research objective, but is highlighted as one of the challenges to the research.

Data Storage

The system must conform to the legal and ethical requirements for data storage.

The nature of the research undertaken involves the collection of large amounts of user data from the Twitter API, a public service. There are ethical implications to storing large quantities of user data. Opinions expressed by users must be handled in a respectful manner and any personally identifying content present in the tweets should be used responsibly.

From a legal perspective, the research must comply with both the *Communications (Retention of Data) Act, 2011* [70] and the Twitter Developer Agreement [71]. In accordance with Twitter's policies, confidential information will not be published. Personally identifying message properties such as the user id and name are not collected as part of the rumour gathering process as these are not fields required by the research. Therefore, ethical compliance is maintained.

3.2 Design Structure

Guided by the requirements of the engineering process, software that will enable the research question to be addressed can now be designed. The requirements discussed in the previous section form the building blocks for the system design, each subcomponent representing a specific functionality that must be provided by the system. The proposed design reflects the structure required to meet the engineering requirements. The system structure is as follows:

- **Rumour Gathering, Feature Retrieval and Storage:**
  - **Gathering:** Gathering of tweets related to rumours of interest from the Twitter API, and performing of necessary filtering.
  - **Feature Retrieval:** Parsing of required metadata from tweets received, and translation to feature logic.
  - **Storage:** Storage of rumour items with all necessary features in the database.
- **Impact Measurement and Impact Analysis:** Impact scoring of random sample size sets from rumour buckets, followed by t-test calculations, and potential impact contributor investigation.
- **Data Visualisation Interface:** A web application supplying visual representation of rumour impact measurement and feature analysis.

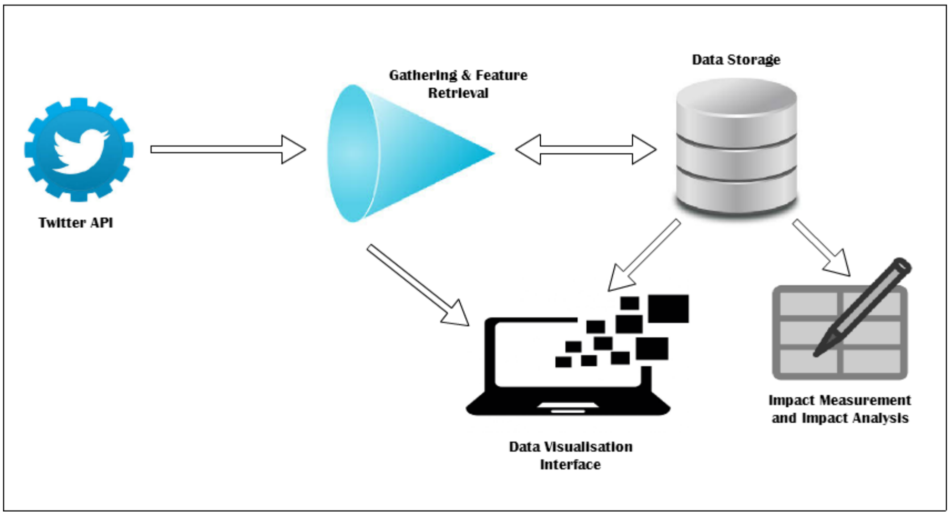

FIGURE 3.1: *System Components*

3.2.1 Rumour Gathering, Feature Retrieval and Storage

The rumour gathering, feature retrieval, and storage components are pictured in Figure 3.2. The *Gathering* and *Feature Retrieval* processes form two separate system components, both interacting with a *Data Store*.

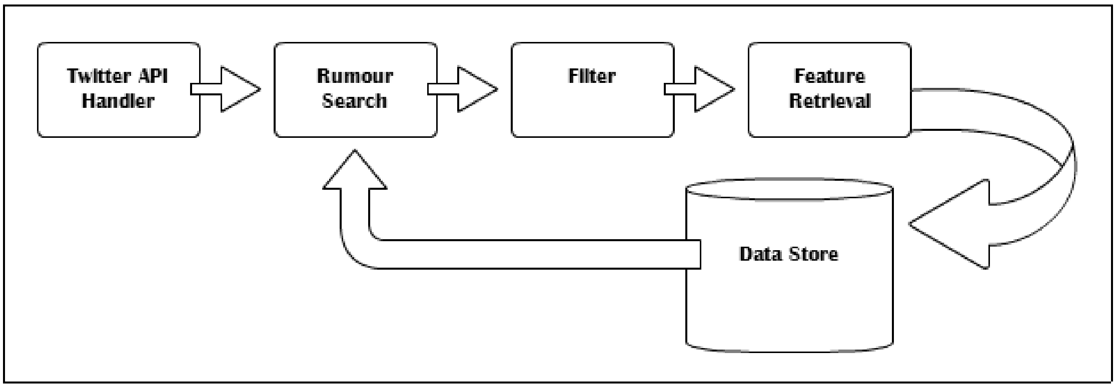

FIGURE 3.2: *Rumour Gathering, Feature Retrieval and Storage*

3.2.1.1 Rumour Gathering

The design structure for the processes of rumour gathering and data filtering consists of the subset of components, pictured in Figure 3.3.

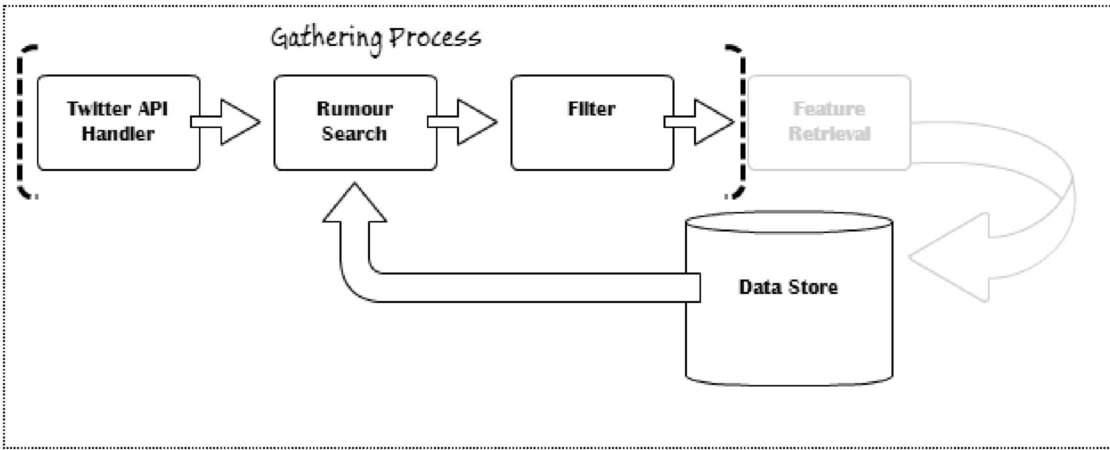

FIGURE 3.3: *Rumour Gathering Components*

- **Twitter API Handler:** Forms the interface between the tool and the Twitter API. The handler is responsible for maintaining necessary details for this interaction: authentication, persistent connection, etc.

- **Rumour Search:** Initiates the rumour search process, requesting rumours by regular expression queries to Twitter's Search API.

- **Filter:** Ensures the *corpus quality* requirement is met by removing duplicate tweets that squeeze through the Search process.

The gathering process involves key design decisions with regard to rumour collection. This component must be capable of submitting query strings to Twitter's Search API, queries that will reflect regular expressions, constructed of keywords specific to the desired rumour.

There are pre-processing requirements to design for. Firstly, in accordance with the *corpus size* requirement, the tool must be capable of handling many different search requests, and keeping rumour buckets organised. To comply with the *corpus composition* requirement, retweets should be excluded. To ensure the *corpus quality* requirement is met, pre-proceesing is required to avoid searching in the same time range more than once, i.e. searching only for tweets older than the last set received, to avoid receiving the same set time and time again, and to gather the most advantageous *snapshot*.

The requirements related to high numbers of rumour searches, as well as pre-processing and filtering needs, illustrate the necessity of having access to the Data Store for the *Gathering* component of the system.

3.2.1.2 Feature Retrieval

This part of the system provides the data for answering the research question, and fulfills the impact measurement and impact analysis engineering requirements.

The feature retrieval process is responsible for taking a Tweet object received from the Search API, funnelled through the *Gathering* process, and parsing the metadata required for impact scoring, namely *retweet count* and *favourite count*, as decided by this work. It also obtains those features required for subsequent impact investigation, the list of other features for analysis as potential influencers to our impact measurement.

After preparation of impact measurement properties and impact analysis features, the data can be organised together, along with the associated tweet id and text, and stored in the system's shared data store.

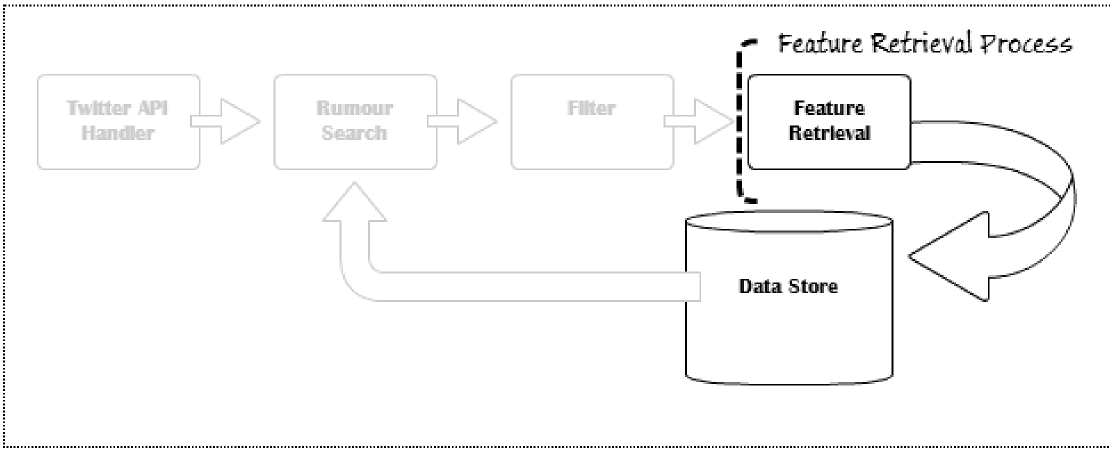

FIGURE 3.4: *Feature Retrieval Components*

3.2.1.3 Storage

The system requires an adaptable data store which can be efficiently accessed during the *Gathering* process, and added to following the *Feature Retrieval* process.

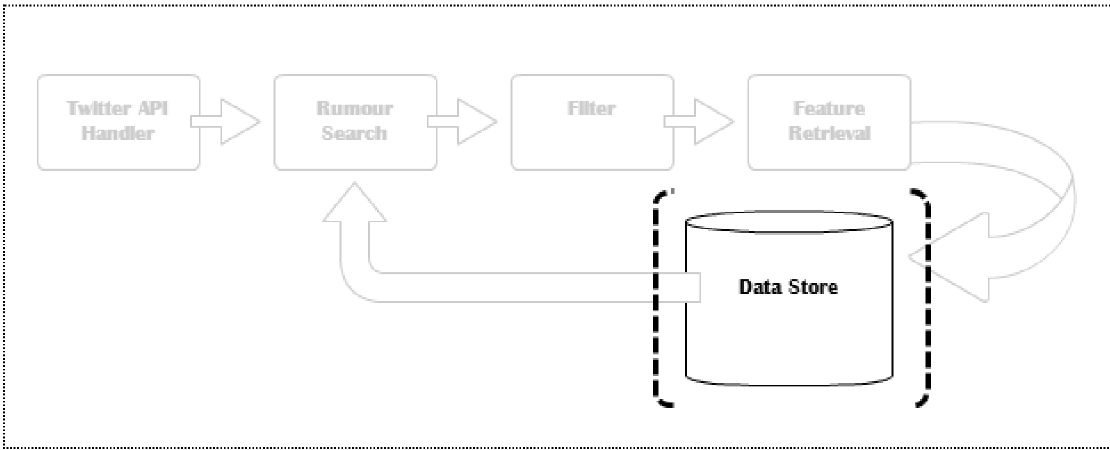

FIGURE 3.5: *Storage Component*

3.2.2 Impact Measurement and Impact Analysis

This part of the system allows the *impact measurement* and *impact analysis* requirements to be fulfilled, and is pictured in figure 3.6.

This component involves the following functionality:

1. Select a random sample of specified sample size.
2. Compute impact score by formula (3.1), section 3.1.2.
3. Perform t-test calculations on various pairs of rumour random sample sets, of equal sample size, in looking for statistical difference.
4. Perform investigations on properties of rumour sets, properties gathered during feature retrieval, investigating the possibility of these features having influence on impact.

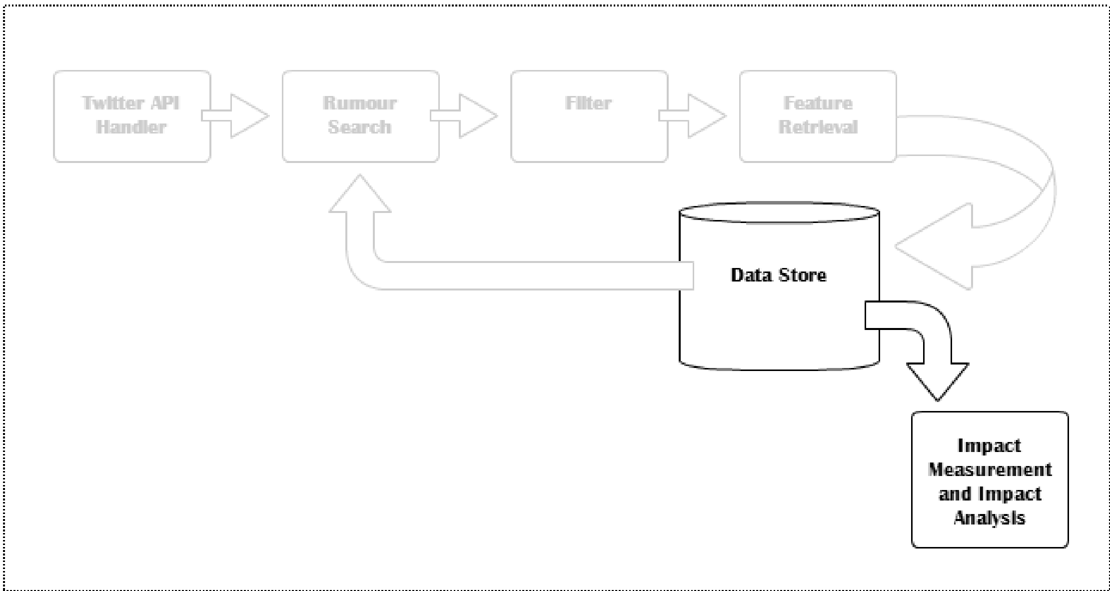

FIGURE 3.6: *Impact Measurement and Analysis Components*

Core decisions that must be made here involve the methods by which random sampling will be performed. This component requires access to the data store, as it obtains random samples from rumour buckets stored within. Statistical methods are also required here, and for these functions, the component shall use a well-established statistics library.

3.2.3 Data Visualisation Interface (DVI)

The DVI is intended as a graphic visualisation of the various buckets of rumours collected and constructed from the *Gathering* and *Feature Retrieval* processes, allowing for visual display of impact scoring and the properties used in this calculation, and the list of features used for contribution analysis. The DVI also includes a real-time rumour search interface. See figure 3.7.

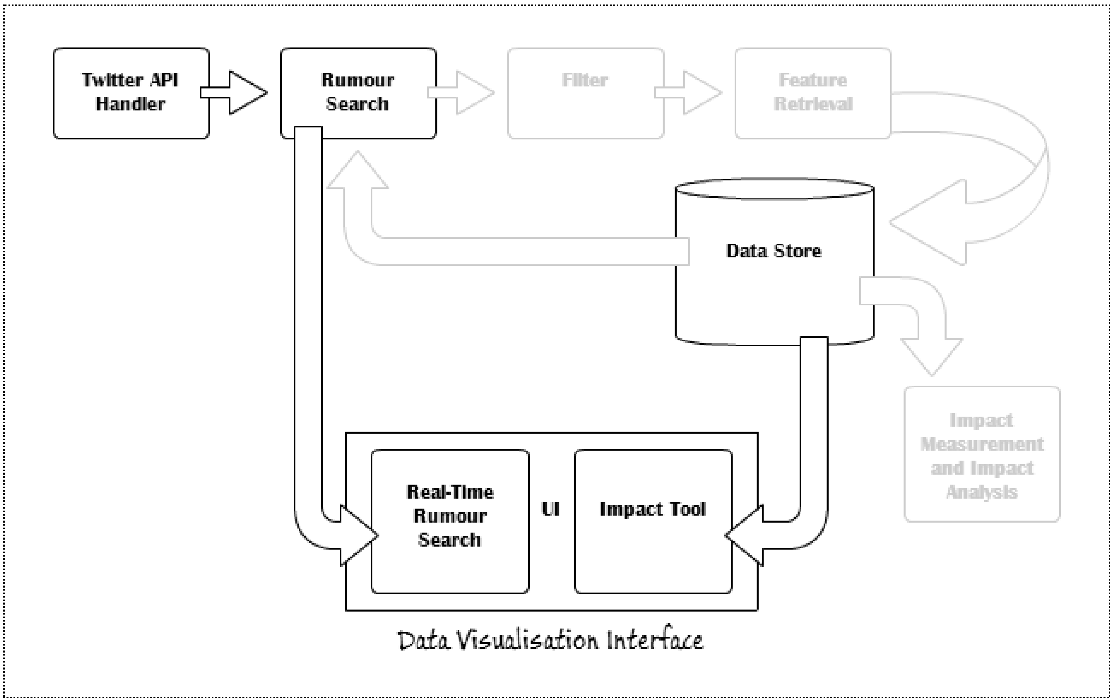

FIGURE 3.7: *Data Visualisation Interface (DVI) Components*

- **Impact Tool:** Displays a list of rumour buckets collected, with clickable links to illustrate impact scoring and properties, and features gathered for influence analysis.

- **Database:** Data store of rumour buckets.

- **Real-Time Rumour Search:** Demonstrates the method used for rumour collection by taking user inputted keyword strings and performing a live impact scoring. This subcomponent uses a simplified *Gathering* process, where pre-processing and filtering steps related to duplicate tweets, are not included, as the intended functionality involves a quick Search API rumour retrieval demonstrating a collection process, and impact scoring on the fly.

The DVI forms a useful stepping stone in demonstrating a resource that could be beneficial to individuals and organisations interested in performing live rumour impact analysis, and also visualising impact data associated with rumours collected. The tool will be built as a dynamic web application.

The impact tool must be efficient and graphically pleasing and organised, as it will illustrate the data used for impact investigation. The live rumour search tool must be capable of passing user inputted queries to server side logic, and translating these to search queries for the Twitter API.

3.3 Summary

In this chapter, a design for a system allowing the research question to be addressed was proposed. By beginning with a process of requirements engineering, a set of requirements were crafted as a framework to achieving the research objectives. The following sections gave detail firstly, on a high level design of the system, followed by detailed descriptions and discussions of subcomponents.

Chapter 4

Implementation

In this chapter, the implementation details of the system are discussed. The technologies used are described, along with a justification for their choosing. While the system has logically separated components, namely a subsystem of *Rumour Gathering and Feature Retrieval*, *Storage*, and a *Data Visualisation Interface*, each with separated design considerations, the implementation process and technologies chosen reflects the requirements of the system as a whole.

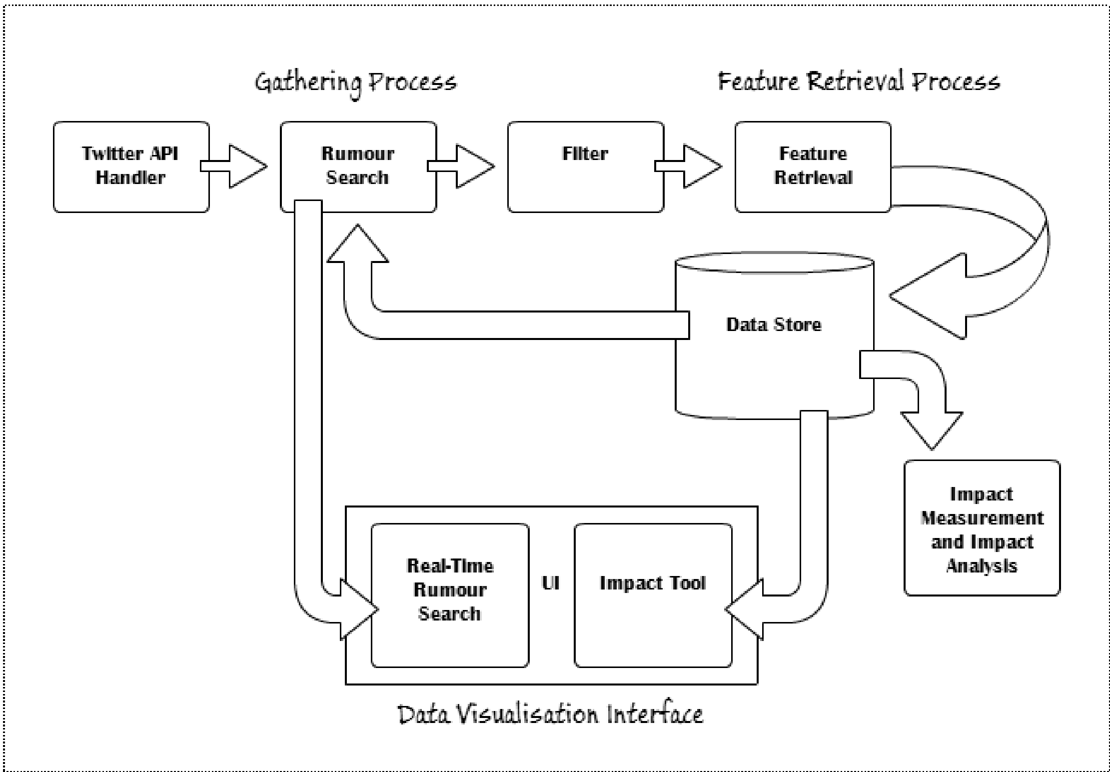

FIGURE 4.1: *System Components*

4.1 Rumour Gathering, Feature Retrieval and Storage

4.1.1 Node.js

Node.js is an open source environment for developing server-side applications, and was selected as the server technology for our system. Node.js applications are written in JavaScript, and can be run within the Node.js runtime on OS X, Microsoft Windows, and Linux.

The decision to use Node.js was driven by a number of factors. Firstly, Node.js provides a rich library of various JavaScript modules which greatly simplifies the development of web applications. As the design had to facilitate the development of a data visualisation interface, a web application component, adequate server choice was required.

This application is data driven, with constant communication between itself and the data store. Speed of execution is important in server choice, as there are some delays to be expected during database access and associated functionality. Being built on Google Chrome's V8 JavaScript Engine, Node.js is very fast in code execution.

Node.js is proving itself as a perfect technology partner, in applications similar to ours, with comparable requirements:

- Data Analysis Applications
- JSON APIs based Applications

4.1.1.1 Node Package Manager

Node Package Manager (npm) is the default package manager for Node.js. It allows Javascript developers to share the code that they've created to solve particular problems, and for other developers to reuse that code in their own applications. Npm runs through the command line and manages dependencies for an application. Npm is an invaluable resource to the design of our system, and allows us to make use of various packages.

Specific to the *Gathering* process, npm includes a package called *twitter*, an asynchronous client library for the Twitter REST and Streaming API's. Using this library, allows us the functions required for the subcomponents - *Twitter API Handler* and *Rumour Search.*

To begin, a new Twitter client is initiated, with valid Twitter developer credentials in the form of a set of consumer and access tokens/keys. Twitter requires all applications to register with them, before generating the required keys. There exists a security responsibility in maintaining the privacy of these keys.

The use of Twitter's Search API, part of the REST API, is required. Npm makes it easy to use the REST API as all the developer needs to do is pass the endpoint and parameters necessary for the REST API method. The 'query' is an inherent parameter to this work and is where the keyword regular expression string to be queried in the Twitter API is included.

4.1.2 Filtering

Firstly, to exclude retweets from the resultant rumour tweet set, 'retweet' is included as a variable in the 'exclude' parameter. To make sure the receiving of most duplicated tweets is avoided, the lowest id of the last tweet set received is included as a variable in the 'max_id' parameter.

The Search API only delivers a maximum of 100 tweets per request. By keeping track of the IDs of tweets received, the lowest ID (related to the oldest tweet - in terms of when composed) received in the current search can be used, as the highest ID parameter for the subsequent search.

In the next search, only tweets older than this ID are received. This allows the tool to avoid duplicate tweet retrieval. In addition, filtering is also performed subsequent to each search request, where the ID of each tweet received is checked for within its bucket. If the ID exists, this tweet is removed as it is a duplicate. This fully ensures the high quality of the corpus that will be presented to the *Feature Retrieval* and *Impact Measurement and Impact Analysis* processes.

4.1.3 MongoDB

MongoDB is a document-oriented database. Classified as a NoSQL database, MongoDB drops the traditional table-based relational database structure, in favour of JSON-like documents with dynamic schemas. Each tweet retrieved from the Twitter API is associated with a large volume of metadata, metadata that is fundamental to addressing the research question of impact measurement and impact analysis.

MongoDB is a natural choice as it provides scalable storage without the need for the complex schema definition that would be required were a SQL database used. Its merits for our system include:

- **JSON:** MongoDB uses JSON documents in order to store records. JSON is the default format of data retrieved from the Twitter API. Translating tweet metadata to database documents is therefore simplified as both make use of the same format. A JSON database, such as MongoDB, returns query results that can be easily parsed, with little or no transformation, directly by JavaScript in our case - reducing the amount of logic needed in the application layer, and specifically within the *Impact Measurement and Impact Analysis* component.

- **Ad Hoc Queries:** MongoDB supports a wide range of queries. Queries can return specific fields of documents, which is of interest to use during *Impact Measurement and Impact Analysis.*

- **mLab:** Database-as-a-Service for MongoDB. This service allows the provisioning of MongoDB on demand on AWS, Amazon, or Google. Database-as-a-Service provides a database solution, allowing us to focus on our application and research instead of operations. mLab's interface is invaluable, as it provides a visual viewing resource for data collections (tables).

4.1.3.1 Mongoose

Mongoose is another npm package the system makes use of. Mongoose allows for elegant mongoDB object modelling in node.js applications. Mongoose allows for straight-forward connection to your database. Mongoose uses models, special constructors compiled from Schema definitions. Instances of these models represent documents which can be saved and retrieved from the database. All document creation and retrieval from the database is handled by these models.

4.2 Impact Measurement and Impact Analysis

This component is the core of our system. Here, impact measurements are performed by applying the definition and associated formula suggested for impact. Impact scores are compared, and the data collected is used in investigating potential influencers to rumour impact on social media. The process is illustrated in figure 4.2, and is detailed as follows:

- **Sample size and rumour bucket selection:** Firstly, the sample size is decided - the number of rumour tweets that will be taken from each bucket. The buckets that will be measured must also be selected.

- **Retrieve random sample:** Given the selected sample size n, a process of generating n random numbers within the range, *0 - ((total bucket size) - 1)*, is conduced. These are then used as a set of random indices. The rumours at these random indices in the rumour bucket are collected. By following this process, a randomly chosen sample is assured.

- **Compute Impact Scores:** For each rumour selected, its impact score is calculated by the formula suggested by this work, formula (3.1), section 3.1.2. This individual score is then added to an *impact set* that is used in subsequent t-test analysis.

- **T-test for Statistical Difference:** Given the two impact sets collected, constructed of the impact scores computed for *sample size* rumours, in each rumour set, giving two impact sets, t-test analysis is now conducted, which provides a T-value.

  Given this T-value and the degrees of freedom, *(sample size * 2) - 2*, a resultant P-value is obtained. The P-value is compared to α, the significance level (0.05), which determines whether statistical difference has been found. If statistical difference is found, the null hypothesis that both rumour sets are essentially the same, can be rejected, and efficient evidence that one rumour bucket is more impactful than the other is supplied. See figure 4.3.

  Another npm package was used for the statistical requirements of the system, namely *simple-statistics*, an implementation of descriptive, regression, and inference statistics.

- **List of Features & Counts:** As impact sets for rumour samples are built, feature sets are also built, constructed of counts or averages, depending on the feature. For example, figure 4.3 shows statistical t-test data, as well as example feature lists with a handful of the features we've gathered. 'Followers' is the total number of followers, the accumulative total, from the *sample size* rumours in the set. The 'Av. tweet length' is the average tweet length of the *sample size* rumours.

- **Impact Output:** This is the output of the *Impact Measurement and Impact Analysis* component, and supplies us with statistical significance data and feature data for contribution analysis. See figure 4.3.

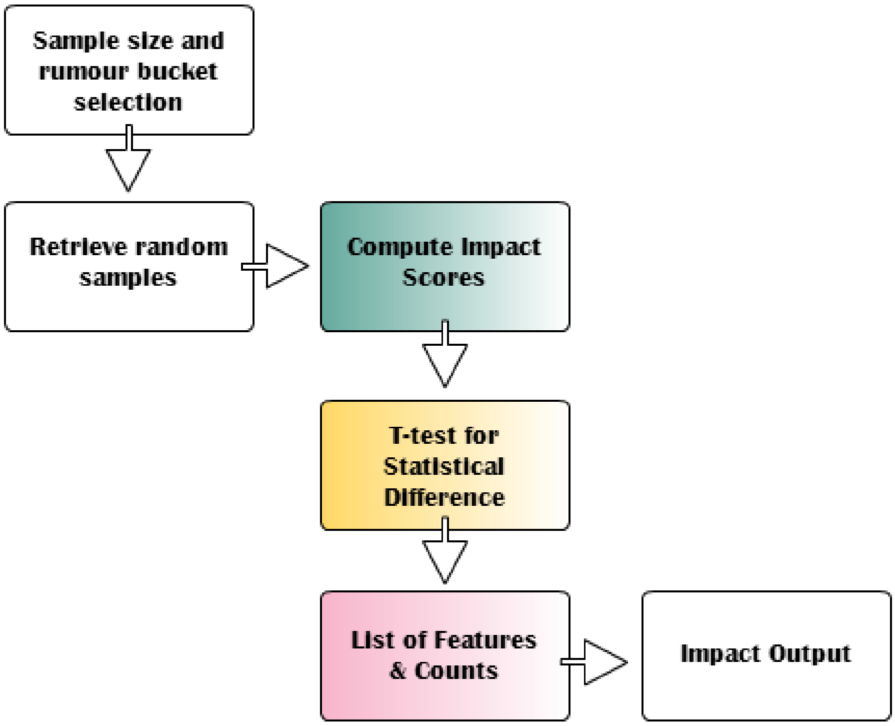

FIGURE 4.2: *Impact Measurement and Impact Analysis Implementation*

```
T-Value: 2.3603873774083293
Sample Size: 40
DF: 78
Two-tailed P value: 0.0207
Statistically different

Rumour A: kim_divorce          Rumour B: rob_blac
Followers: 365                 Followers: 222
Friends: 926                   Friends: 19
Statuses: 16310                Statuses: 12068
Verified: 0                    Verified: 0
Includes Hashtags: 8           Includes Hashtags: 9
Includes Media: 19             Includes Media: 11
Includes User Mentions: 1      Includes User Mentions: 6
Includes URLs: 39              Includes URLs: 39
Includes word 'retweet': 1     Includes word 'retweet': 0
Av. tweet length: 125.5        Av. tweet length: 129.9
```

FIGURE 4.3: *Impact Output*

4.3 Data Visualisation Interface (DVI)

The DVI constitutes a simple front end interface component, and is constructed using commonplace web app technologies - HTML, CSS, and AJAX, allowing for asynchronous communication with the server.

Ajax is a group of technologies. HTML and CSS are used in combination for mark up and style information. The DOM is accessed with JavaScript to dynamically display and allow interaction with the information presented. JavaScript and the XMLHttpRequest object provide a method for exchanging data asynchronously between the browser and server to avoid full page reloads.

4.3.1 DVI Landing Page

The landing page, pictured in figure 4.4, gives brief detail of the motivation of the application, *"The impact of rumours on social media itself. Measurement in terms of an impact score. Analysis investigating message-based and account-based features"*. The user is presented with two paths of action, with clickable buttons:

1. *R-T Rumour Search*
2. *Impact Tool*

FIGURE 4.4: *DVI - Landing Page*

4.3.2 Real-Time Rumour Search

This page, pictured in figure 4.5, displays a search bar for the user to input the keywords related to the rumour they are interested in searching for. The web app consumes this input and passes it to the server, which translates the keywords to a regular expression query be inputted to the Twitter Search API through the *Search* component.

The main use of this tool is that it returns the number of rumour tweets that were returned from the Twitter Search API. Recall, the maximum number that can be returned in a single request is 100. This tool has been useful for the research as it gave indication of the amount of data that would be available were this rumour investigated for the processes of impact measurement and impact analysis.

If 100 tweets were returned, or close to this number, sufficient interest in this rumour can be assumed, as many Twitter users are tweeting about it. If this number is 0, or not close to 100, this rumour is not attracting enough attention to really be considered as a circulating story, or be sufficient for conducting impact methods.

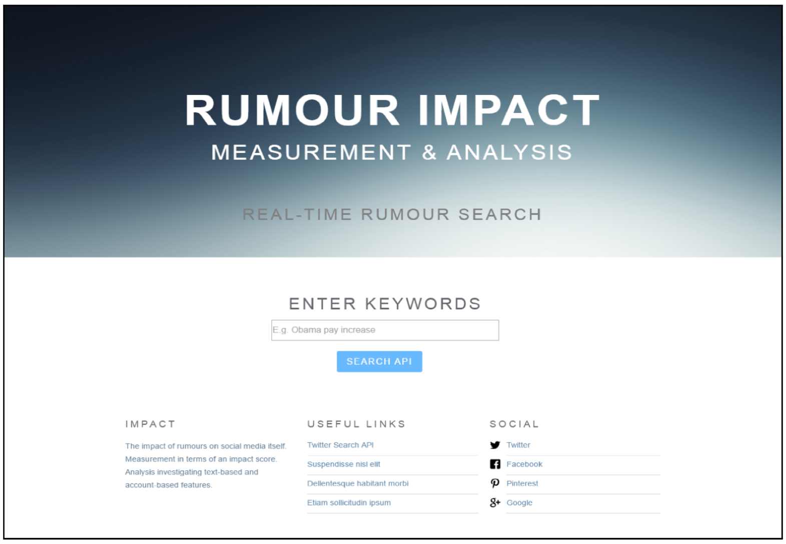

FIGURE 4.5: *DVI - Real-Time Rumour Search Page*

4.3.3 Impact Tool

This page, pictured in figure 4.6, displays a list of the rumours this study has been working with for impact methods of measurement, and contribution analysis. This tool is extremely useful to the research as it displays all the bucket data, impact data and feature data associated with a rumour that has been gathered and processed. The user simply clicks the rumour they are interested in, and a table of results is returned, with impact scoring, and total feature counts and averages, depending on applicability.

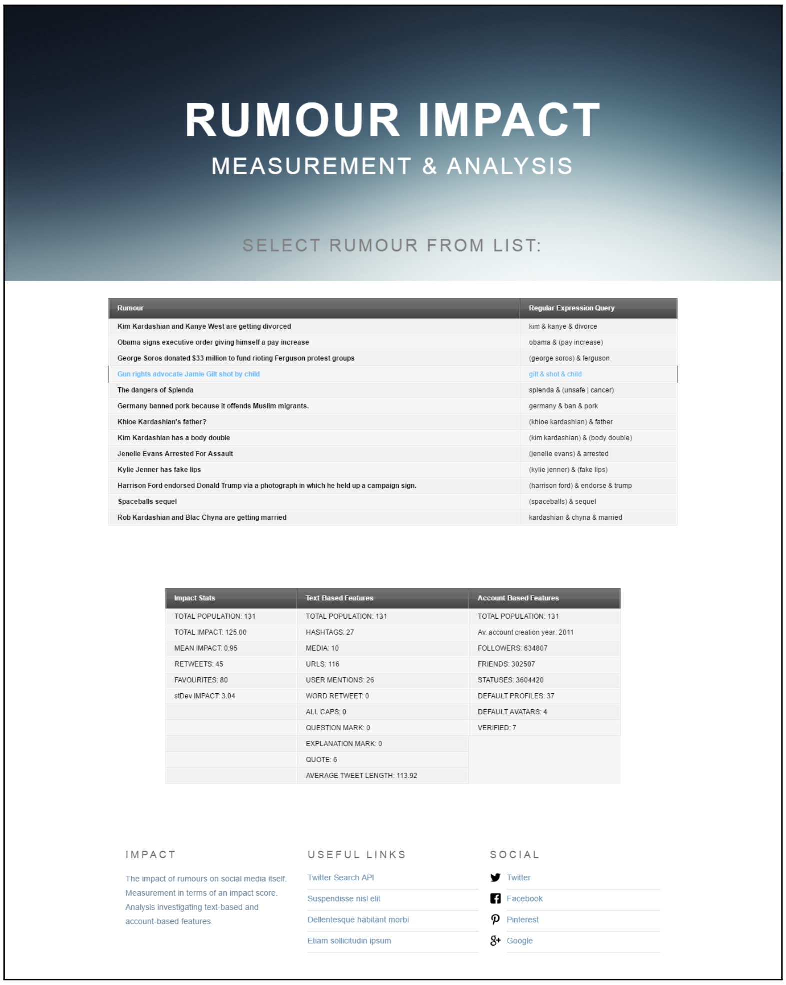

FIGURE 4.6: *DVI - Impact Tool Page*

4.4 Summary

Chapter Four detailed the technology choices made in creating the system. The chapter also contained details of the implementation of the filtering, impact measurement, impact analysis, real-time rumour search, and impact data visualisation processes.

Chapter 5

Evaluation

In this chapter, the results of the research are evaluated against the research objectives set out in Chapter One. Firstly, a plan is proposed, detailing the approach to evaluation the chapter shall follow. Applicable research functions and system components are then evaluated according to this plan.

5.1 Approach

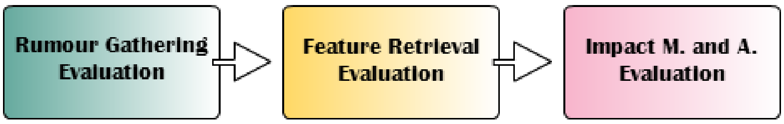

FIGURE 5.1: *Evaluation Approach*

A modular approach is taken for evaluation where the research objectives are organised into specific aims, reflective of their modular nature. Thus, the approach is as follows:

- **Rumour Gathering:** To assess the rumour corpus generation procedure, and the effectiveness of the system's *Rumour Gathering* component.
- **Feature Retrieval:** To assess the *Feature Retrieval* process, presenting feature details obtained.
- **Impact Measurement and Impact Analysis:** To present impact scores calculated for rumour buckets collected, and signify those of higher impact, with subsequent investigation, identifying properties that are influential to impact, thence addressing the primary research question.

5.2 Rumour Gathering Evaluation

The rumour corpus, the dataset built of rumour buckets, is gathered with the aim of addressing the research question. Therefore, the criteria by which to assess the *Rumour Gathering* process are reflective of the requirements of the primary research question, which strives to measure and analyse the impact of rumorous messages on social media. Consequently, desirable corpus properties are:

- **Corpus size:** The number of rumour buckets collected, each filled with as many tweets related to the rumour as are available at the time of search, in compliance with the notion of a *snapshot*.

- **Corpus composition and corpus quality:** The dataset should be free of retweets (forwarded messages), duplicates, and tweets should be truly related to the rumour in question.

Ultimately, the *Rumour Gathering* process must be assessed on whether it has been successful in rumour detection and associated gathering functionality, providing a rumorous dataset that enables the research question of the dissertation to be addressed.

5.2.1 Rumour Gathering Results

Rumours were chosen and collected at one point in time between February and April 2016, by the detection approach discussed in section 3.1.1. Each query represents a known rumour, fitting to the rumour definition adopted by the research, as suggested by previous rumour behaviour research [18], a *"circulating story of questionable veracity, which is apparently credible but hard to verify, and produces sufficient skepticism and/or anxiety"*.

This work is interested in rumours of the day, rumours that were circulating stories at the time of search. To find rumours and build a *snapshot*, circulating rumour stories fitting our rumour definition were found using online resources, similar to the approach adopted in previous work [17], sources such as: Snopes.com[1], starcasm[2] and CELEBUZZ[3].

Table 5.1 presents rumour buckets collected - bucket names and associated rumours.

[1]http://www.snopes.com/, accessed: 15/05/2016
[2]http://starcasm.net/, accessed: 15/05/2016
[3]http://www.celebuzz.com/, accessed: 15/05/2016

Bucket Name	Rumour
calvert_leaving	Is Jamie Calvert leaving 'Teen Mom 2'?
evans_arrested	'Teen Mom 2' Janelle Evans arrested
ford_trump	Harrison Ford endorses Trump holding a sign
germany_pork	Germany bans pork under Sharia law
giant_rat_ldn	Giant rat found in London
gilt_shot	Gun rights advocate Jamie Gilt shot by child?
gunz_vasectomy	Did Peter Gunz (rapper) have a vasectomy?
khloe_parentage	Who is Khloe Kardashian's father?
kim_divorce	Kim Kardashian & Kanye West are getting divorced
kim_doppelganger	Does Kim Kardashian have a body double?
kim_fake_butt	Kim Kardashian's butt is fake
kylie_jenner_lips	Kylie Jenner's lips are fake
manson_trump	Charles Manson endorsed Trump from prison
nazi_submarine	Nazi submarine found in Great Lakes
notebook_sequel	'The Notebook 2' is filming
obama_pay_incr	Obama gives himself a pay increase
oprah_pregnant	Oprah pregnant with first child
pawnstars_arrest	'Pawn Stars' Austin Russell arrested
destroy_earth	Newly discovered planet could destroy earth
poppins_sequel	Disney announced Mary Poppins sequel
rob_blac	Rob Kardashian & Blac Chyna married?
soros_ferguson	George Soros funds Ferguson protests
spaceballs_sequel	A Spaceballs sequel?
splenda_unsafe	Splenda is unsafe for humans and is carcinogenic
tilapia_dangerous	Eating farm-raised tilapia is dangerous
trump_tshirts	Trump supporters t-shirts "Make America White Again"

Table 5.1: List of rumour buckets and associated rumour stories.

Table 5.2 presents regular expression queries used for detection and the number of tweets collected in each rumour bucket. The size of the rumour buckets varies greatly, as highlighted as a likelihood in the related work review in Chapter Two. This has added complexity to the research, where the temporal characteristics of rumours, and the decision to collect in *snapshot* process disallowing a large dataset built over time, has affected bucket sizes as expected. See figure 5.2. The mean size of rumour buckets is noted below, along with the standard deviation. The large standard deviation reflects just how spread out the sizes are.

- **MEAN** Bucket Size: 139
- **STDEV** Bucket Size: 156

Bucket Name	Regular Expression Query	#tweets
calvert_leaving	calvert & leaving & (teen mom 2)	8
evans_arrested	(teen mom 2) & evans & arrested	287
ford_trump	ford & trump & sign	124
germany_pork	germany & bans & pork	596
giant_rat_ldn	giant & rat & london	11
gilt_shot	gilt & shot & (by child)	131
gunz_vasectomy	gunz & vasectomy	85
khloe_parentage	(khloe kardashian) & father	100
kim_divorce	(kim kardashian) & (kanye west) & divorce	98
kim_doppelganger	(kim kardashian) & (body double)	101
kim_fake_butt	(kim kardashian) & (fake butt)	11
kylie_jenner_lips	(kylie jenner) & (fake lips)	334
manson_trump	(charles manson) & endorse & trump	62
nazi_submarine	(nazi submarine) & (great lakes)	40
notebook_sequel	notebook & (2 | sequel)	28
obama_pay_incr	obama & (pay (rise | increase))	217
oprah_pregnant	oprah & pregnant	77
pawnstars_arrest	(pawn stars) & russell & arrested	55
destroy_earth	(new planet) & (destroy earth)	109
poppins_sequel	poppins & (2 | sequel)	11
rob_blac	(rob kardashian) & (blac chyna) & married	184
soros_ferguson	(george soros) & ferguson	183
spaceballs_sequel	spaceballs & (2 | sequel)	142
splenda_unsafe	splenda & (unsafe | cancer)	594
tilapia_dangerous	tilapia & dangerous	3
trump_tshirts	trump & t-shirts & (make america white again)	34

Table 5.2: List of rumour buckets, corresponding regular expression queries and the number of related rumour tweets collected.

The *Gathering Process* is also responsible for implementing pre-processing and filtering requirements, see section 3.2.1.1. The application of these methods excludes retweets (forwarded messages) from entering the dataset, as well as duplicated tweets. Tweets entering the dataset that are not truly related to the rumour in question, are also of concern. Periodic checks were performed on the dataset to ensure this was not occurring to any noteworthy level.

5.2.1.1 Discussion

With the data presented, it is now possible to assess and evaluate the rumour gathering and corpus generation process.

A dataset of rumour buckets has been gathered. In comparison with previous work that followed similar rumour detection and tweet retrieval methods [17], a much larger variety of buckets has been collected, 26 compared to 5. However, this application is suggestive in requiring a much greater variety to allow us to address the research question, and associated objectives. The sizes of this study's rumour buckets vary substantially. Some compare to previous work, and some are much smaller. This suggests the complexity of the *snapshot* approach, and the consequential data limits as indicated by the research challenge detailed in section 1.4, where collection was performed *in the moment* rather than once per hour, or over any extended period of time. *Noisy* tweets - retweets and duplicates, have been excluded from the dataset.

Thus, the *Rumour Gathering* implementation has achieved its objective of creating a rumorous corpus, the foundation that enables the research question to be addressed.

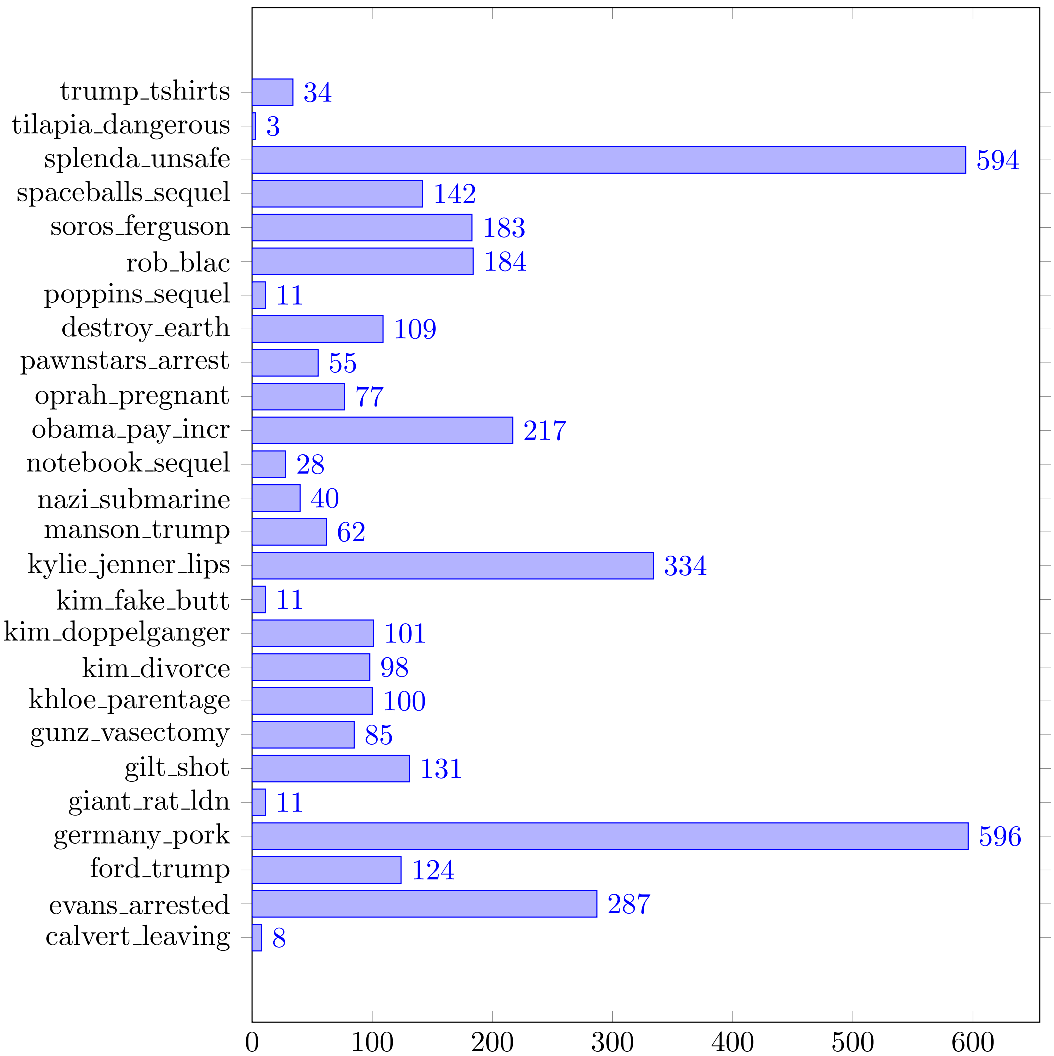

FIGURE 5.2: *Bucket sizes*

5.3 Feature Retrieval Evaluation

The *Feature Retrieval* process, following *Rumour Gathering*, aims to allow for the research question to be addressed. This part of the system is responsible for parsing the tweet metadata obtained from the Twitter API. The metadata obtained allows for impact measurement by the formula suggested by the work. The *Feature Retrieval* process is also responsible for translating applicable metadata to feature logic. This part of the system is fundamental to the research question, allowing for impact measurement and impact analysis.

The effective fulfillment of this task is the successful collection and storing of tweet properties that will be used for impact measurement and those that will be investigated as potential influencers to rumour scores obtained.

5.3.1 Feature Retrieval Results

Subsequent to rumour gathering, tweet features of interest to the research were also collected and organised, in a format that would allow efficient impact measurement and impact analysis in later stages of the research.

Table 5.3 details *favourites* and *retweets*, the user engagements by which *impact* is formalised and defined. Table 5.4 details *Message-Based Features* and table 5.5 details *Account-Based Features* - counts or averages where applicable. While features were gathered for each rumour bucket listed in tables 5.1 and 5.2, the data presented here only represents half of the buckets gathered, the 13 that are larger in size.

The features retrieved are as follows:

Impact Measurement Variables

To measure impact, this work implemented formula (3.1), as detailed in section 3.1.2, which requires the following variables to be collected:

- **Retweets:** Accumulative #**Retweets**.
- **Favourites:** Accumulative #**Favourites**.

Message-Based Features

- **HT:** #tweets in the bucket containing **Hashtags**
- **M:** #tweets in the bucket containing **Media**
- **URL:** #tweets in the bucket containing **URLs**
- **UM:** #tweets in the bucket containing **User Mentions**
- **R:** #tweets in the bucket containing the word **'retweet'**
- **AC:** #tweets in the bucket entirely **All Caps**
- **?:** #tweets in the bucket containing a **Question Mark**
- **!:** #tweets in the bucket containing an **Exclamation Mark**
- **Q:** #tweets in the bucket containing a **Quote**
- **Av.L:** Average **Length** of the tweets in the bucket

Account-Based Features

- **Y:** Average **Creation Year** of the accounts associated with the bucket
- **FO:** Accumulative #**Followers**
- **FR:** Accumulative #**Friends**
- **S:** Accumulative #**Statuses**
- **DP:** Accumulative #**Default Profiles**
- **DA:** Accumulative #**Default Avatars**
- **V:** #**Verified** accounts associated with the bucket

5.3.1.1 Discussion

With the data at hand, the evaluation of the feature retrieval process can now commence. The variables required for impact measurement have been collected for every rumour tweet received through the gathering process. A wide selection of features have also been gathered, features noted as being useful in other studies of rumour behaviour, as highlighted by the State of the Art.

Bucket Name	Size	Retweets	Favourites
evans_arrested	*287*	272	513
ford_trump	*124*	104	105
germany_pork	*596*	605	493
gilt_shot	*131*	45	80
khloe_parentage	*100*	165	319
kim_divorce	*98*	64	70
kim_doppelganger	*101*	7	30
kylie_jenner_lips	*334*	64	248
obama_pay_incr	*217*	501	416
rob_blac	*184*	54	103
soros_ferguson	*183*	389	281
spaceballs_sequel	*142*	18	99
splenda_unsafe	*594*	370	420

Table 5.3: *Impact Measurement Variables*

Bucket Name	Size	HT	M	URL	UM	R	AC	?	!	Q	Av.L
evans_arrested	*287*	83	102	280	12	0	0	87	0	1	104.97
ford_trump	*124*	18	12	110	71	0	0	24	0	8	97.44
germany_pork	*596*	156	79	569	115	0	0	100	0	17	107.45
gilt_shot	*131*	27	10	116	26	0	0	0	0	6	113.92
khloe_parentage	*100*	51	51	90	2	0	0	0	0	0	107.31
kim_divorce	*98*	27	38	96	8	5	0	0	0	2	121.67
kim_doppelganger	*101*	12	30	92	5	0	0	0	0	6	123.28
kylie_jenner_lips	*334*	114	50	289	17	0	0	199	0	5	109.1
obama_pay_incr	*217*	24	13	128	80	0	0	100	0	10	117.2
rob_blac	*184*	43	50	179	26	0	0	0	0	0	126.75
soros_ferguson	*183*	31	13	169	32	0	0	0	0	2	106.58
spaceballs_sequel	*142*	3	0	114	6	0	0	0	0	0	99.94
splenda_unsafe	*594*	173	73	550	114	0	0	197	0	11	98.07

Table 5.4: *Message-Based Features*

Bucket Name	Size	Y	FO	FR	S	DP	DA	V
evans_arrested	*287*	2013	3011983	715931	16447580	159	3	9
ford_trump	*124*	2011	767542	361021	3485078	41	6	2
germany_pork	*596*	2012	2226140	1066305	31710003	248	23	2
gilt_shot	*131*	2011	634807	302507	3604420	37	4	7
khloe_parentage	*100*	2013	140413	60857	6490747	74	7	0
kim_divorce	*98*	2014	315434	250111	4427747	57	0	0
kim_doppelganger	*101*	2013	969098	205493	6480350	46	0	5
kylie_jenner_lips	*334*	2013	14089288	411349	28255599	186	32	13
obama_pay_incr	*217*	2012	501664	376337	3963232	113	10	2
rob_blac	*184*	2013	7778250	756902	15792014	74	2	7
soros_ferguson	*183*	2011	501124	394641	11184692	61	11	1
spaceballs_sequel	*142*	2014	322135	83652	2528130	15	5	0
splenda_unsafe	*594*	2012	11174880	1033550	40865809	250	17	10

Table 5.5: *Account-Based Features*

Thus, the *Feature Retrieval* implementation has achieved its objectives. User engagement measures, the variables used for impact measurement have been collected, as has an array of properties that will be investigated as potential contributors to impact. The success of this task allows for the data supply that is inherent to the research question.

5.4 Impact Measurement & Impact Analysis Evaluation

The rumour buckets gathered through the *Rumour Gathering* process have allowed the detection and collection of rumorous tweets, supplying a rumorous corpus. The *Feature Retrieval* process carried out the necessary task of building the data required for impact measurement, and obtained a selection of tweet property data which would allow for impact analysis and impact contribution investigation.

While this data is fundamental to our research, it does not address the research question on its own. At present, success only ranges as far as efficient data collection. It is now time to address the core question this research poses, and assess our approach to impact measurement and impact analysis, impact on social media itself. This section of the evaluation procedure is the most important, and will directly constitute the ultimate success of the research.

In order to conduct this significant evaluation, the following course of action will be taken, and results will be presented within the following structure:

1. A presentation of the impact scores for rumour buckets collected.

2. T-test analyses, and ultimate presentation of statistical significance found, where statistical difference between rumours has highlighted the higher impact of some rumours, compared to others.

3. An investigation of potential influencers to rumour impact - By presentation of the feature properties in rumours deemed more impactful, in comparison to those deemed less so.

4. Further investigation of influencers to rumour impact - By comparing impactful rumours with non-rumours.

Bucket Name	Size	Total Impact	Mean Impact	SD Impact
germany_pork	*596*	1098	1.84	11.83
splenda_unsafe	*594*	790	1.33	12.84
kylie_jenner_lips	*334*	312	0.93	4.55
evans_arrested	*287*	785	2.74	11.66
obama_pay_incr	*217*	917	4.23	24.44
rob_blac	*184*	157	0.85	6.66
soros_ferguson	*183*	670	3.66	21.96
spaceballs_sequel	*142*	117	0.82	6.57
gilt_shot	*131*	125	0.95	3.04
ford_trump	*124*	209	1.69	9.23
kim_doppelganger	*101*	37	0.37	1.04
khloe_parentage	*100*	484	4.84	46.15
kim_divorce	*98*	134	1.37	4.83

TABLE 5.6: *Impact Measurement Scores*

5.4.1 Impact Measurement & Impact Analysis Results

The first set of results relates to the evaluation stage:

"A presentation of the impact scores for rumour buckets collected".

It was decided to exclude the smaller buckets for impact evaluation. Therefore, the study now considers only the larger 13 buckets gathered by the system. Table 5.6 presents the impact scores calculated by the formula suggested by this work, formula (3.1), section 3.1.2, *impact = retweets + favourites* .

It is difficult to directly compare impact scores as bucket sizes vary so substantially, preventing a clear benchmark to be inferred. However, for convenience, table 5.6 presents the data with buckets placed in decreasing order, in terms of size. This will naturally place buckets of comparable size in close proximity. In addition, a mean impact score has been noted for each bucket, which conveys the average impact score of the rumour tweets within the bucket. The standard deviation of impact scores for rumour tweets within each bucket is also noted.

It is worth noting the observations that are apparent from this raw data. The largest rumour bucket collected is *germany_pork*, the bucket related to the rumour, *"Germany bans pork under Sharia law"*. This bucket has obtained the highest impact score of 1098. However, the mean impact of the rumour tweets within the bucket is only 1.84, which is not one of the highest mean impact scores obtained. Therefore, it van be assumed that this high impact score is largely reflective of the bucket size, with more tweets lending to a higher accumulative score.

The mean impact scores are important, as they supply an indication of the respective impact of each rumour tweet within each bucket. The mean scores allow us data that is comparable. As the *snapshot* method was followed in rumour gathering, only tweets at one point in time could be collected. A direct result of this is that the timing of a rumour has had a great effect on the amount of tweets that were available for retrieval. To put this in simple terms, if rumour A became topical only today, and rumour B became topical 3 days ago, it is very likely that there will be a lot more rumour tweets to collect for rumour B. Interestingly, *khloe_parentage* is one of the smallest buckets of rumours collected, but represents the highest mean impact score, reflective of the high impact of the rumour tweets within the bucket.

The standard deviation of impact scores illustrates how spread out the values are within each rumour bucket. This is a reflection of how much the data can vary, even within the same rumour. A higher standard deviation signifies the widespread nature of the impact scores, and elucidates once again, how shaky rumour data truly is. Taking the *khloe_parentage* example again, standard deviation is very high, depicting how substantially the rumour tweets within the bucket vary, in terms of impact.

The next set of results relates to the evaluation stage:

"T-test analyses, and ultimate presentation of statistical significance found, where statistical difference between rumours has highlighted the higher impact of some rumours, compared to others".

T-tests were conducted on impact scores for every pair of rumours presented in table 5.6. A random sample, of size 40, was chosen for all tests. The decision regarding 40 as the size of the random sample was made in an attempt to detect the most meaningful difference as possible, with consideration of the varying bucket (total population) sizes.

By performing T-test analyses, this work endeavored to find statistical significance in the data, allowing us to flag rumours that were higher in impact compared to others. Therefore, the null and alternative hypotheses were as follows:

- H_0: There is no significant difference between specified populations, or no difference among rumour sets, regarding their impact.
- H_1: There is significant difference between specified populations. The rumour sets are different in terms of impact.

Experiment	Rumour A	Rumour B	T	P
A	*kim_divorce*	*rob_blac*	2.36	0.02
B	*kim_divorce*	*spaceballs_sequel*	2.24	0.03
C	*obama_pay_incr*	*rob_blac*	2.34	0.02
D	*obama_pay_incr*	*spaceballs_sequel*	3.6	0
E	*soros_ferguson*	*rob_blac*	2.02	0.047
F	*soros_ferguson*	*spaceballs_sequel*	2.27	0.047
G	*evans_arrested*	*kim_doppelganger*	2.39	0.02

TABLE 5.7: *Cases for which the null hypothesis can be rejected, where* $P \leq 0.05$.

As is to be expected, with data as flimsy as that associated with rumours, and under the limits of our *snapshot* process, statistical significance was not found in a large proportion of cases. However, the study was successful in obtaining statistical significance in some cases, highlighting those rumours that were statistically different in terms of impact. Only by studying rumours that are significantly different, can the higher impact of some rumours be signified, and investigations of influential features for such impact performed.

Once again related to the changeable nature of rumours, the study was cautious in immediately accepting results obtained through t-test analyses. The t-test calculation is reflective of the sample populations it is presented with. Being aware of how flimsy rumour data is, and how varying individual impact scores can be (see SD Impact, Table 5.6), many t-tests on pairs of samples were performed, which at first appeared to be statistically different, but for which assurance was needed.

As the variances differed, H_0 was rejected for certain rumours, those detailed in table 5.7, after receiving similar t-value results with many t-test calculations, within the same rumour populations, selecting different random sample sets.

Table 5.7 presents examples of those rumours where statistical significance was found, with a t-value and P-value, representative of one of the t-test calculations associated with the pair. Rumour A is of higher impact than Rumour B.

- **Degrees of Freedom:** *(sample size * 2) - 2 = (40 * 2) - 2 =* 78
- **Significance Level** α**:** 0.05, the most widely used significance level.

Given a T-value and the degrees of freedom, a P-value is obtained. The P-value is compared to α. A small P-value (≤ 0.05) indicates strong evidence against the null hypothesis, so it is rejected. The study concludes that there exists statistical difference between the impact of respective rumours.

The next set of results relates to the evaluation stage:

"An investigation of potential influencers to rumour impact - By presentation of the feature properties in rumours deemed more impactful, in comparison to those deemed less so".

Recall the *Message-Based Features* and *Account-Based Features* gathered through the feature retrieval process. After observing the raw data collected, it was decided to reduce the number of features that would be taken any further through the process of impact analysis. The key objective to the following analyses is to flag those features that are influential to rumour impact, i.e. contributing to a higher impact score, obtained through the formula suggested by the work.

The features that will not be considered are as follows:

Message-Based Features

- *Word 'retweet':* Eliminated due to minimal existence in the dataset. See table 5.4 - a total of 5 rumour tweets in the total dataset.
- *All caps:* Eliminated due to non-existence in the dataset.
- *Question mark:* Eliminated due to insufficient presence. Absent from many of the rumour buckets where statistical difference was found.
- *Exclamation mark:* Eliminated due to non-existence in the dataset.
- *Quote:* Eliminated due to minimal existence in the dataset. See table 5.4.

Account-Based Features

- *Year:* Eliminated as raw data appears minimal for comparison to be made.
- *Default Profile:* Eliminated as it appears a very large number of Twitter users maintain the default wallpaper for their accounts. Therefore, not a substantial indicator for any difference or comparison.
- *Default Avatar:* Eliminated due to minimal existence in the dataset. See table 5.5.

The approach taken in feature analysis is to take each feature that will be investigated individually. The study then presents those rumours once again, those incurring statistical difference related to impact, and compares each feature's existence in the higher impact rumour compared to that of the lower impact. Those features that are more substantial in higher impact rumours can then be concluded as contributing / influential to impact.

5.4.1.1 Account-Based Feature Influence

This work evaluates whether account properties of the composing author influence rumour impact. The features that are analysed are *followers*, *friends*, and *statuses*, representing how popular and active the authors are. The influence of *verified* accounts is also investigated, those accounts belonging to key individuals that Twitter have signified by placement of the verified badge. These accounts, belonging to politicians, celebrities, journalists tend to have many followers, and attract significant user attention. In the figures to follow, figure 5.3, 5.4, 5.5, and 5.6, the letters along the y-axis represent the experiments listed in table 5.7, those for which statistical difference was found.

Followers

The study investigates whether the number of followers associated with the composing authors involved in a rumour, has an effect on the impact of the rumour. Followers are those people who have connected with a Twitter account. Someone who thinks you're interesting can follow you. Following is not mutual, you don't have to follow back.

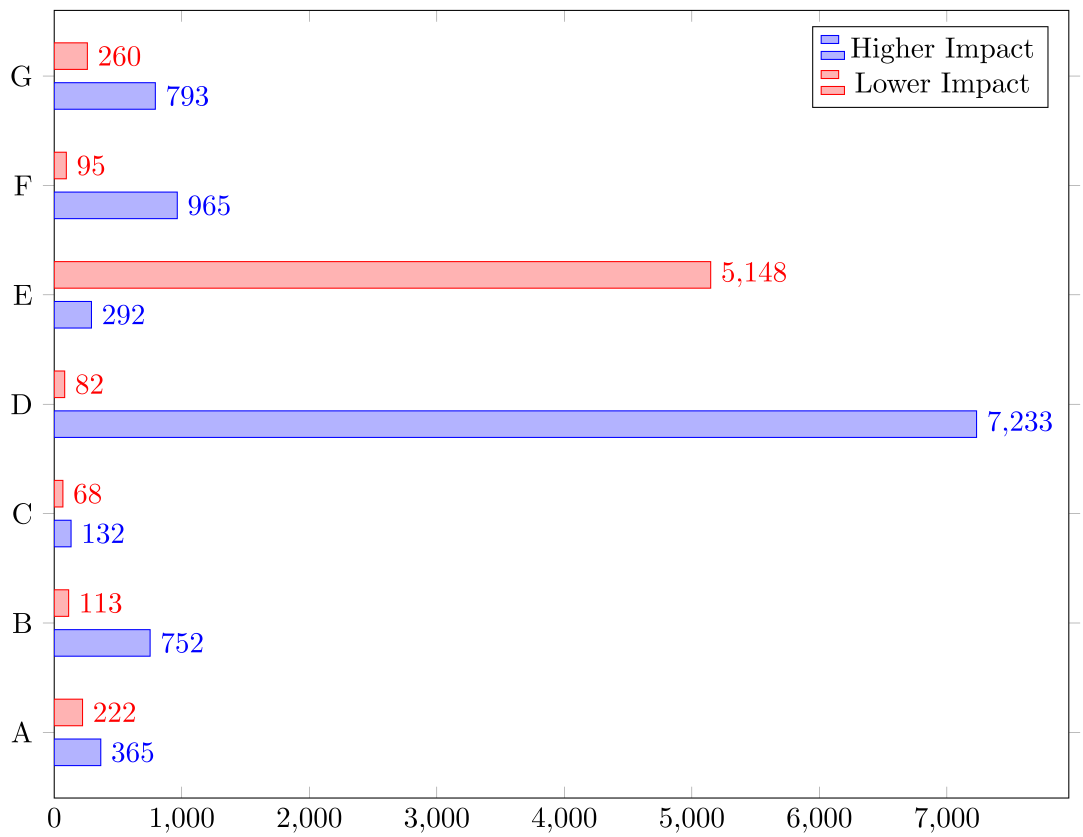

Figure 5.3: *Follower counts in Higher Impact Rumour vs Lower Impact Rumour*

Out of the seven experiments where statistical difference was found, highlighting the higher impact of one rumour compared to another, six cases have higher follower counts in the higher impact rumour than for the lower impact rumour. *Experiment E* is the one exception where the lower impact rumour has more followers in the set than the higher impact rumour. Six cases out of seven represents 86%. Under the specific conditions of our experiments, *it can be concluded that followers influence the impact of a rumour.*

Friends

The study investigates whether the number of friends (followees) associated with the composing authors involved in a rumour, has an effect on the impact of the rumour. Friends are those people you have connected with, the people you follow, who do not necessarily follow you back.

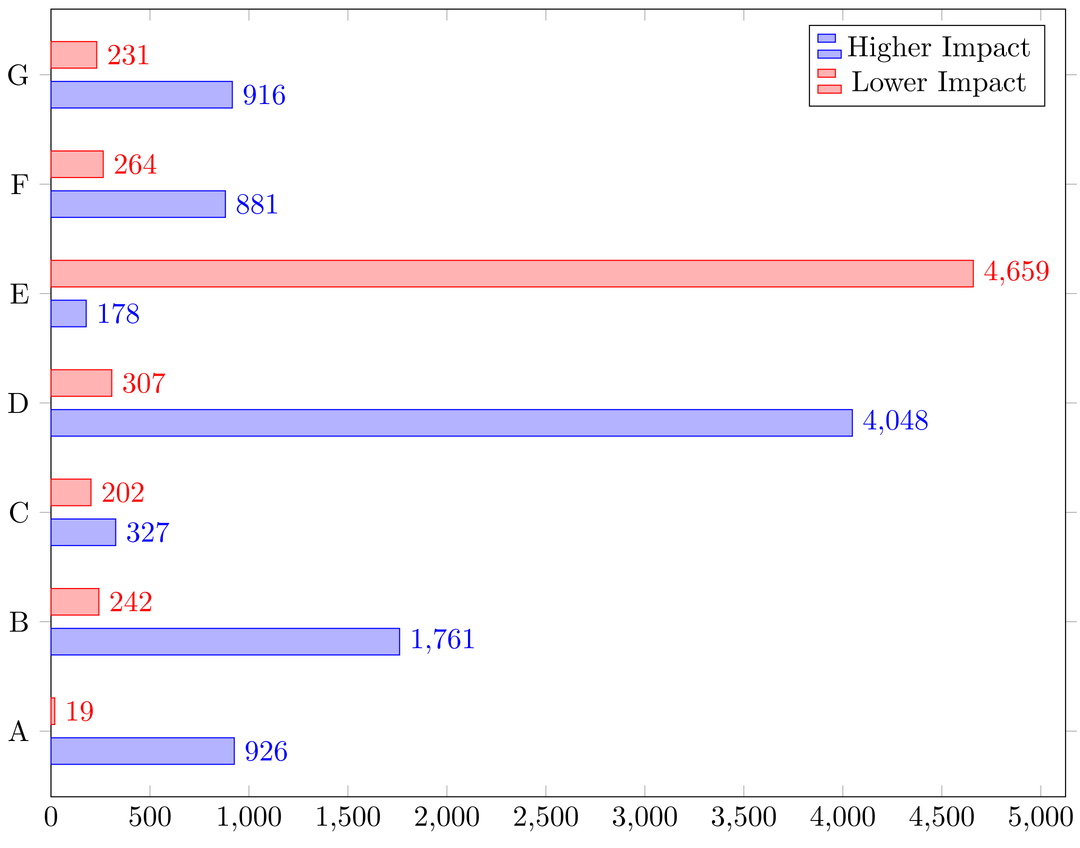

Figure 5.4: *Friends counts in Higher Impact Rumour vs Lower Impact Rumour*

A similar result to *Followers*. Out of the seven experiments where statistical difference was found, highlighting the higher impact of one rumour compared to another, six cases have higher friends counts in the higher impact rumour than for the lower impact rumour. *Experiment E* is the one exception where the lower impact rumour has more friends in the set than the higher impact rumour. Six cases out of seven represents 86%. Under the specific conditions of our experiments, *it can be concluded that friends influence the impact of a rumour.*

Statuses

The study investigates whether the total number of statuses (messages / tweets) that the authors have composed in the lifetime of their accounts, has an effect on the impact of the rumour.

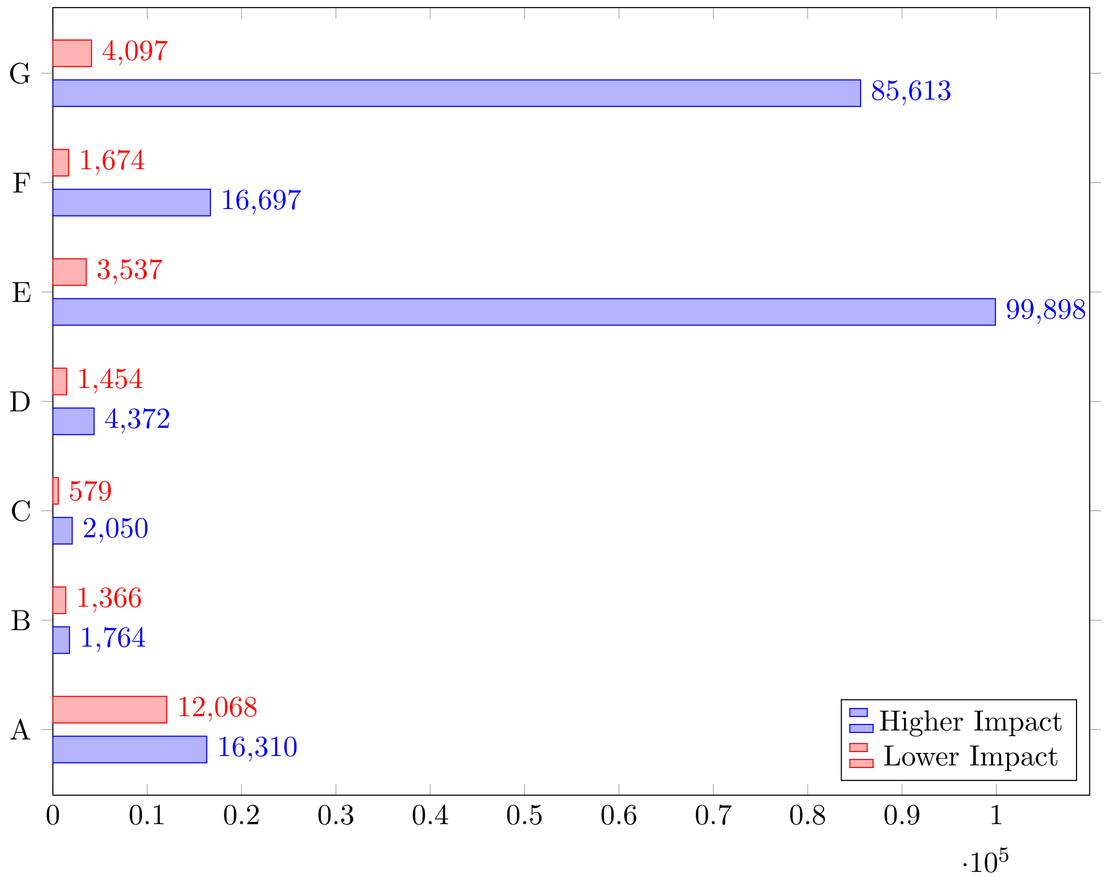

FIGURE 5.5: *Statuses counts in Higher Impact Rumour vs Lower Impact Rumour*

In all seven experiments where statistical difference was found, highlighting the higher impact of one rumour compared to another, total statuses counts associated with the authors are higher in the higher impact rumour compared to the lower impact rumour. This result represents 100%. Under the specific conditions of our experiments, *it can be concluded that the total number of statuses associated with author accounts influences the impact of a rumour.*

Verified Accounts

The study investigates whether authors who are *verified*, influence the impact of a rumour. These accounts involve notable individuals and normally have many followers associated with them.

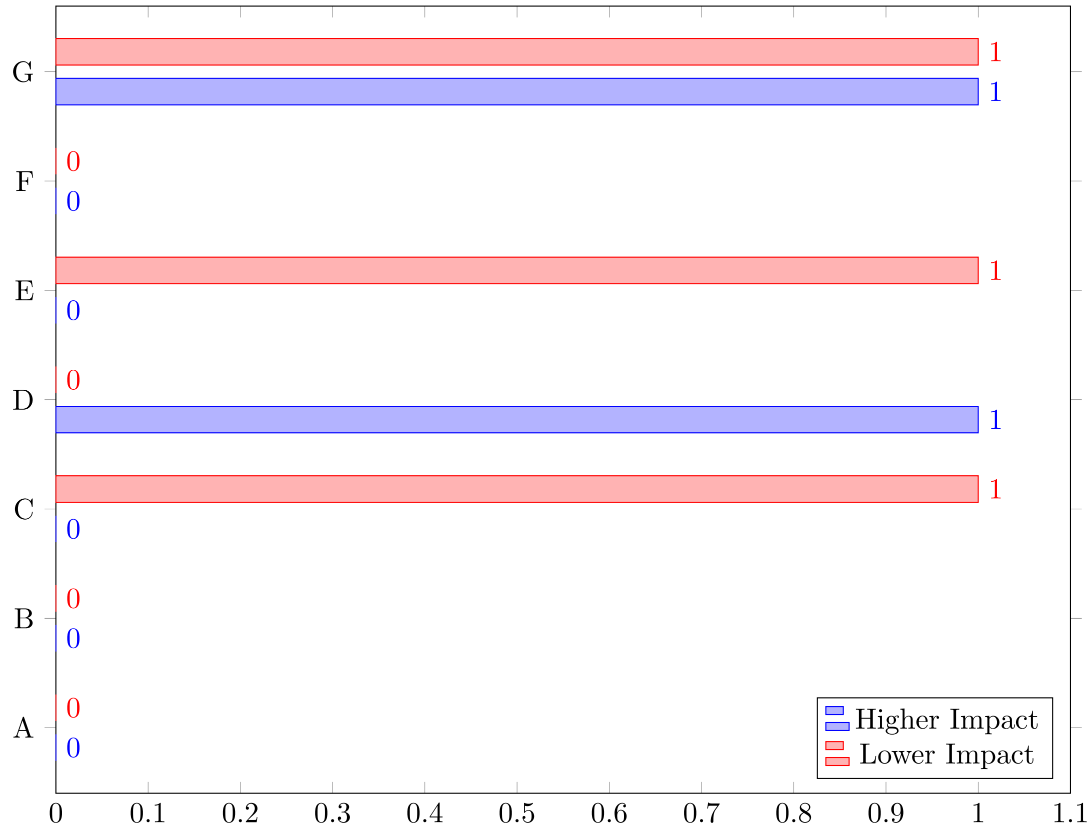

Figure 5.6: *Verified counts in Higher Impact Rumour vs Lower Impact Rumour*

These analyses are inconclusive, but are included as they seem to lend some interesting insight. In each rumour bucket, there was either one verified account or none. The rumour that included a verified author was not necessarily the more impactful one, see *experiment C* and *experiment E*. This is an interesting finding as it disproves what may be likely to believe, that accounts associated with celebrities and other key individuals would automatically influence higher impact. Under the specific conditions of our experiments, *it can be concluded that verified accounts do not influence the impact of a rumour.*

5.4.1.2 Message-Based Feature Influence

This work evaluates whether message properties, features of the rumour tweet itself, are influential to rumour impact. The message properties for which substantial data was found are related specifically to Twitter *entities*, namely *Hashtags, Media, URLs, User Mentions*. The study also investigates the idea of there being an *ideal* tweet length, and if this affects rumour impact. In the figures to follow, figure 5.7, 5.8, 5.9, 5.10 and 5.11, the letters along the y-axis represent the experiments listed in table 5.7, those for which statistical difference was found.

Hashtags

The study investigates whether the inclusion of hashtags influences the impact of a rumour. For each rumour tweet, the system simply noted *true* if a hashtag was present, and did not count the number of hashtags in each message. Therefore, the counts in the figure below are simply the number of rumour tweets within each rumour bucket (sample size 40) that contained a hashtag, be that one or more.

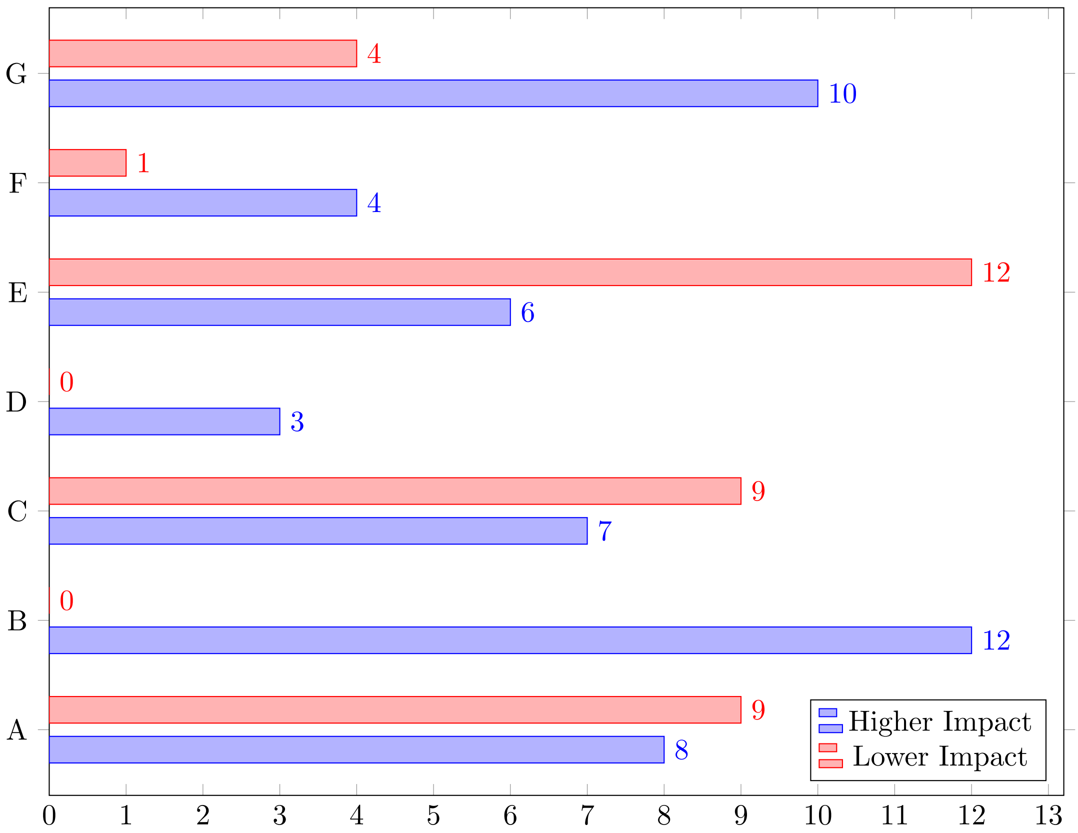

FIGURE 5.7: *Number of messages containing Hashtags in Higher Impact Rumour vs Lower Impact Rumour*

Out of the seven experiments where statistical difference was found, highlighting the higher impact of one rumour compared to another, four cases have more messages containing hashtags in the higher impact rumour than for the lower impact rumour.

This result represents 57%. This result is not sufficient to verify the significance of hashtags for rumour impact. Under the specific conditions of the experiments of this study, *it can be concluded that the inclusion of hashtags in rumour messages does not influence the impact of a rumour.*

Media

The study investigates whether the inclusion of media influences the impact of a rumour, media items being limited by Twitter, to photographs / images. For each rumour tweet, the system simply noted *true* if a media item was present, and did not count the number of media items in each message. Therefore, the counts in the figure below are simply the number of rumour tweets within each rumour bucket (sample size 40) that contained a media item (image), be that one or more.

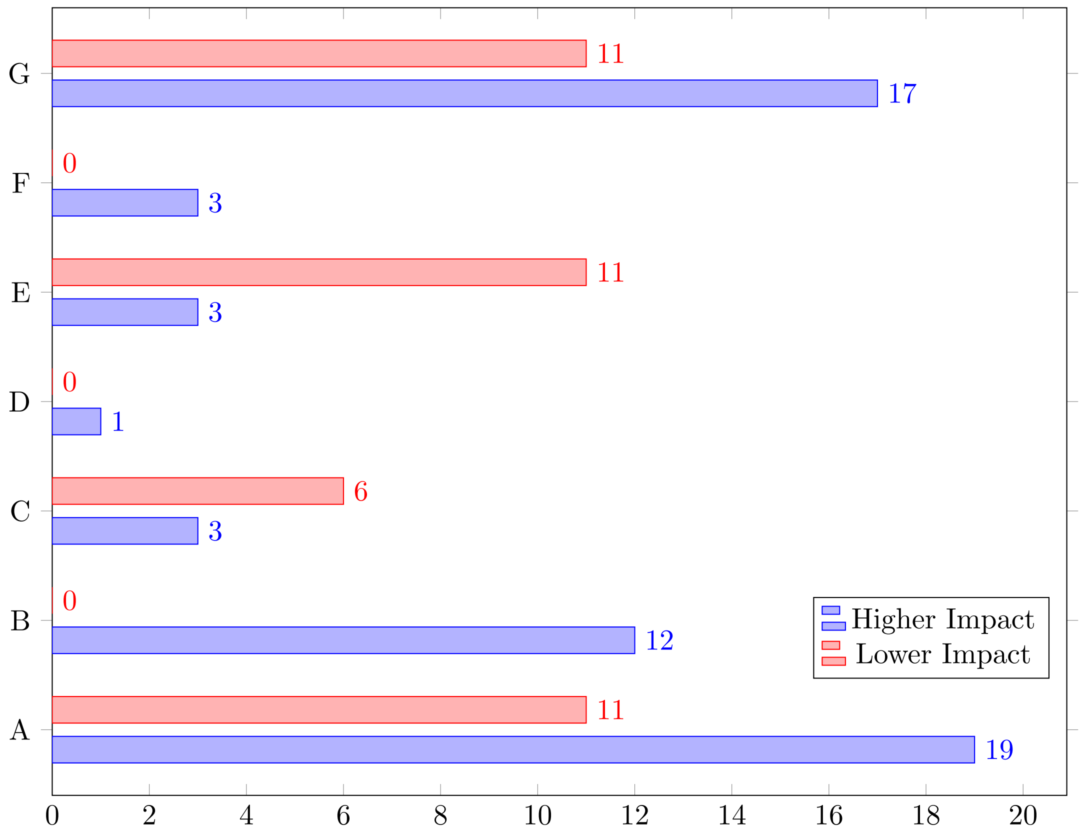

FIGURE 5.8: *Number of messages containing Media in Higher Impact Rumour vs Lower Impact Rumour*

Out of the seven experiments where statistical difference was found, highlighting the higher impact of one rumour compared to another, five cases have more messages containing media items in the higher impact rumour than for the lower impact rumour.

This result represents 71%. This result is not quite conclusive but it is believed that with more data and more analyses, this percentage is likely to increase, and become more conclusive. Under the specific conditions of the experiments of this study, *it can be concluded that the inclusion of media in rumour messages is likely to influence the impact of a rumour.*

User Mentions

The study investigates whether the inclusion of a user mention influences the impact of a rumour, user mentions tagging another Twitter user in the message (@user). For each rumour tweet, the system simply noted *true* if a user mention was present, and did not count the number of user mentions in each message. Therefore, the counts in the figure below are simply the number of rumour tweets within each rumour bucket (sample size 40) that contained a user mention, be that one or more.

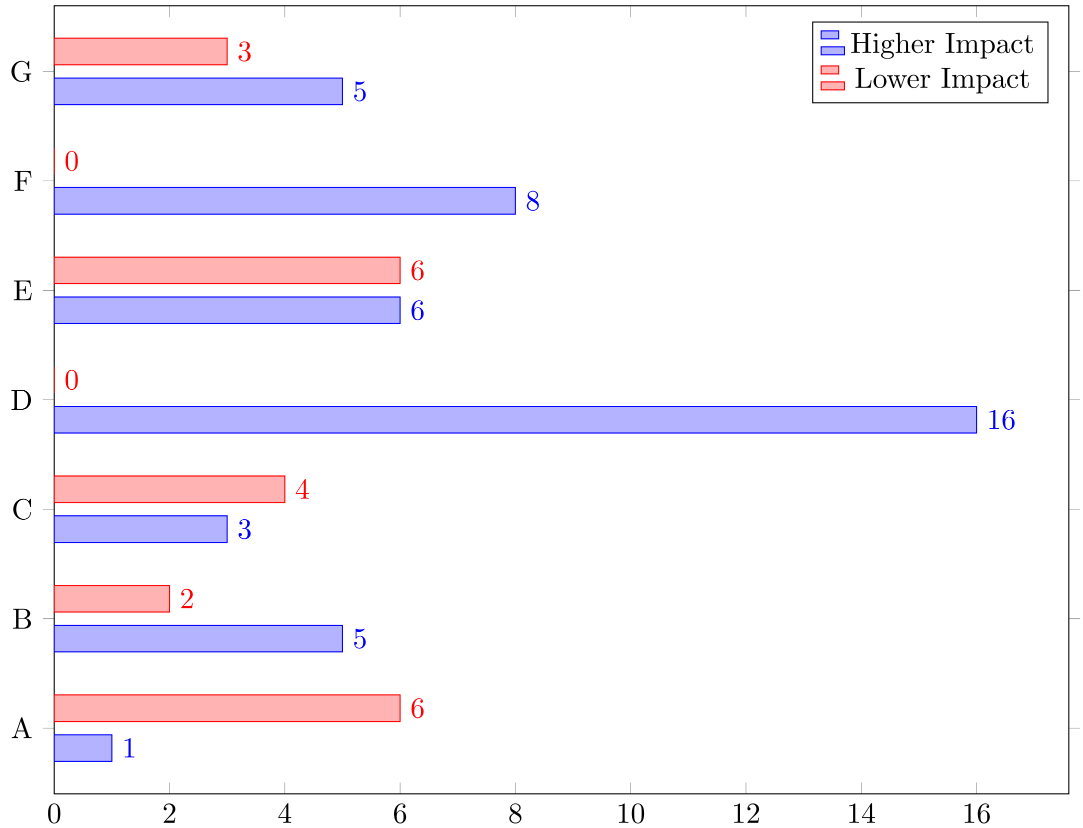

FIGURE 5.9: *Number of messages mentioning / tagging another user - in Higher Impact Rumour vs Lower Impact Rumour*

A similar result to *Media*. Out of the seven experiments where statistical difference was found, highlighting the higher impact of one rumour compared to another, five cases have more messages containing user mentions in the higher impact rumour than for the lower impact rumour. This result represents 71%. Again, this result is not quite conclusive but it is believed that with more data and more analyses, this percentage is likely to increase, and become more conclusive. Under the specific conditions of the experiments of this study, *it can be concluded that the inclusion of user mentions in rumour messages is likely to influence the impact of a rumour.*

URLs

The study investigates whether the inclusion of URL links influences the impact of a rumour, URLs linking to another web source.

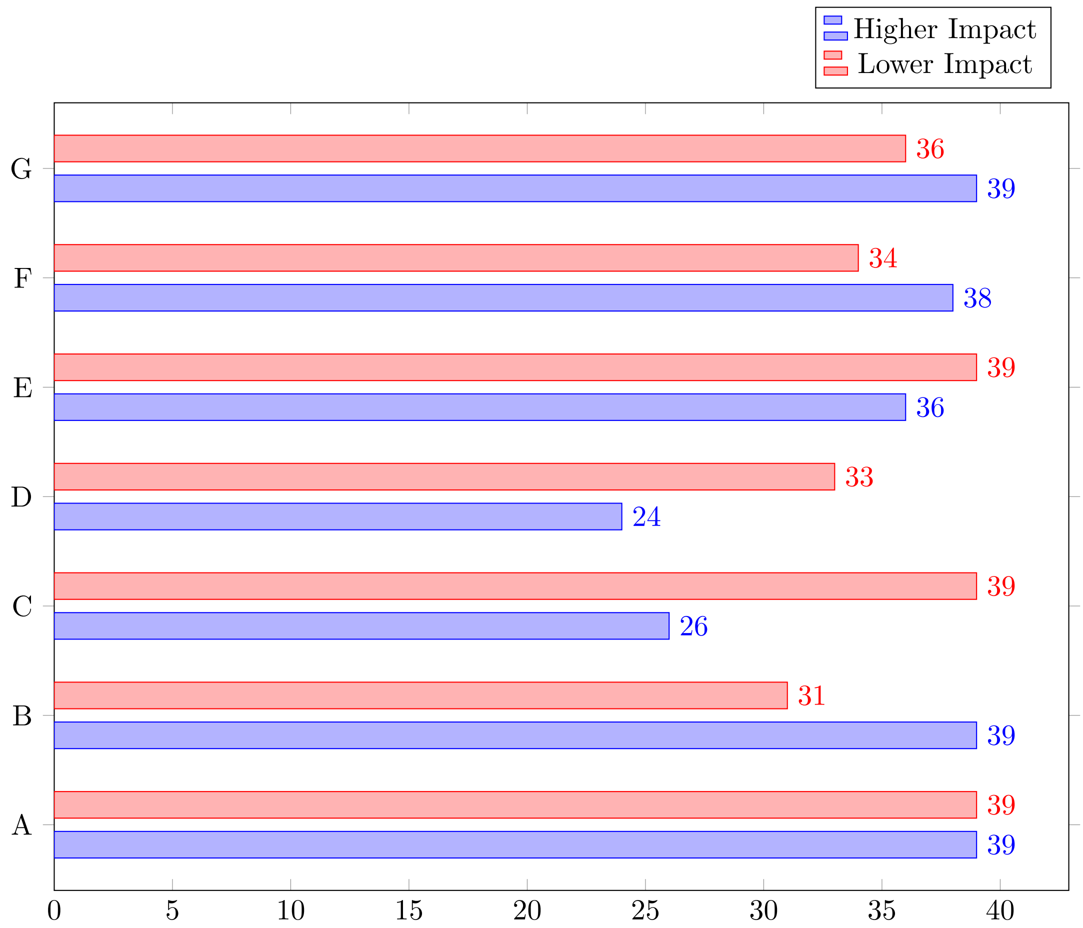

FIGURE 5.10: *Number of messages containing a URL in Higher Impact Rumour vs Lower Impact Rumour*

For each rumour tweet, For each rumour tweet, the system simply noted *true* if a URL was present, and did not count the number of URLs in each message. Therefore, the counts in figure 5.10 are simply the number of rumour tweets within each rumour bucket (sample size 40) that contained a URL, be that one or more.

Out of the seven experiments where statistical difference was found, highlighting the higher impact of one rumour compared to another, only three cases have more messages containing user mentions in the higher impact rumour than for the lower impact rumour. This result represents 43%. Under the specific conditions of the experiments of this study, *it can be concluded that the inclusion of URLs in rumour messages does not influence the impact of a rumour.*

Tweet Length

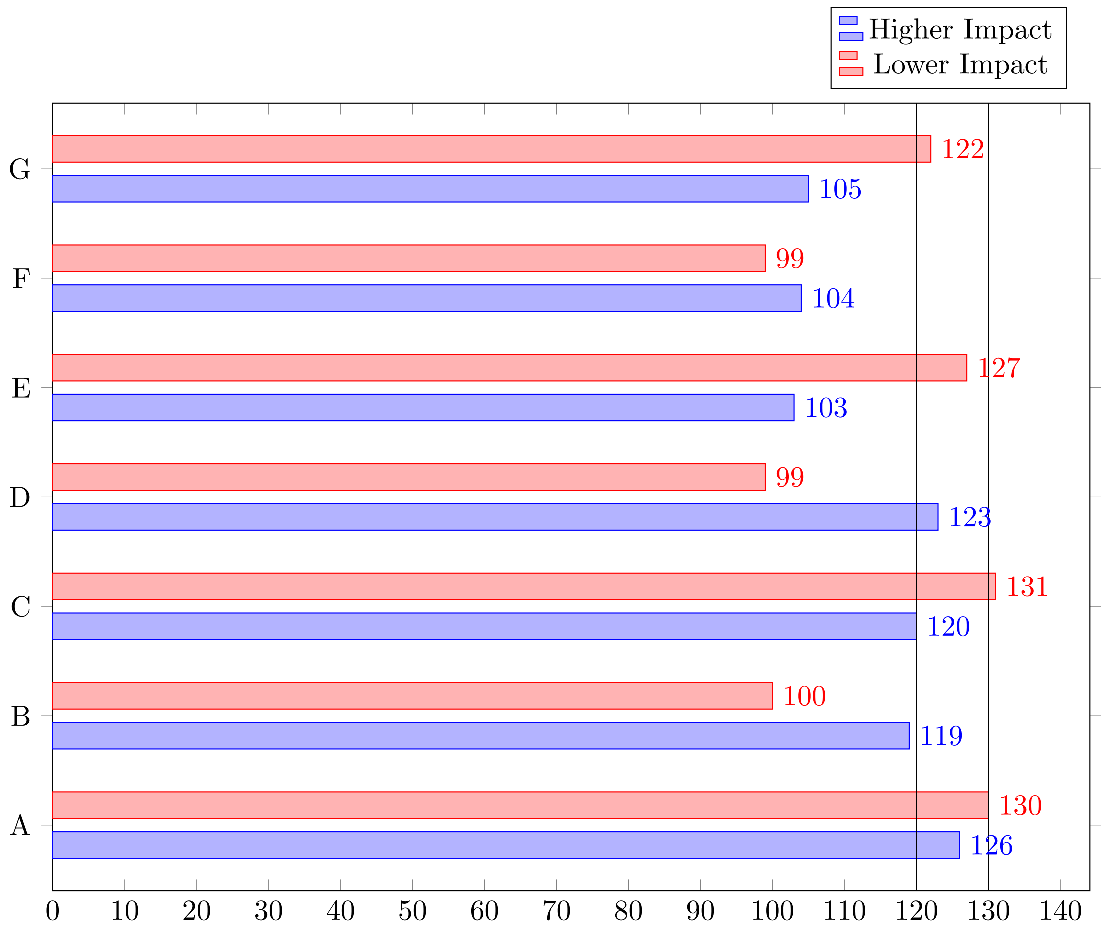

FIGURE 5.11: *Average length (chars) of messages in Higher Impact Rumour vs Lower Impact Rumour*

Online sources are highlighting the idea of there being a *sweet spot* related to the length of tweets and engagements by users [37].

This is suggested to be in the range 120-130 characters when URL links are included. As the vast majority of rumour messages collected contain links, see table 5.4, the average length of higher impact and lower impact rumours are compared against this 120-130 character *sweet spot*, see figure 5.11.

This study's results do not conform to this 120-130 character range. Out of the seven higher impact rumours, only three are within these bounds - *A: 126, C: 120, D: 123*. The same result, three out of seven is also found for the lower impact impact rumours within experiments - *A: 130, E: 127, G:122*. Therefore, these results signify that the higher impact rumour is not necessarily of length suggested to be optimal, 120-130 characters. This is an interesting finding as it suggests rumours behave out of the norm compared to *regular* tweets.

Under the specific conditions of the experiments of this study, *it can be concluded that crafting messages within the range 120-130 characters does not influence the impact of a rumour.*

Summary

Subsequent to analysing the feature properties of higher impact rumours compared to lower impact rumours, under the conditions of the data and experiments of this study, findings are now summarised.

Findings of this comprehensive study suggests,

- ***followers*** are influential to rumour impact.
- ***friends*** are influential to rumour impact.
- ***total statuses*** is influential to rumour impact.
- ***verified accounts*** are not influential to rumour impact.
- ***hashtags*** are not influential to rumour impact.
- ***media*** is likely to be influential to rumour impact.
- ***user mentions*** are likely to be influential to rumour impact.
- ***URLs*** are not influential to rumour impact.
- ***length 120-130 chars*** is not influential to rumour impact.

The final set of results relates to the evaluation stage:

"Further investigation of influencers to rumour impact - By comparing impactful rumours with non-rumours".

5.4.1.3 Impactful Rumours compared to Non-Rumours

The final stage of evaluation is inspired by a question of the specific nature of rumours, compared to all other messages. Following the study's investigation and findings regarding influential features lending to the impact of rumours, analyses is now extended as the research asks if there is a measurable difference between rumours and non-rumours - not related to the text but related to impact and features.

For choosing non-rumours, news stories from the credible source - *BBC News*[4] were selected. These stories were chosen on the 26th April 2016, and involve factual events, i.e. no question regarding veracity.

Table 5.8 presents non-rumour buckets collected - bucket names, associated news stories, and the number of tweets collected in each non-rumour bucket.

The steps taken for these analyses is as follows:

- T-test analysis between sample size (40) messages of impactful rumour and sample size (40) messages of non-rumour.
- Where statistical difference is not found, i.e. the rumour and non-rumour represent the same impact, the presence of features investigated previously is compared.

By performing this investigation, the research is ultimately asking, *"are rumours and non-rumours essentially the same or is there something that can be measured that makes them different"*.

A summary of findings related to rumours vs non-rumours is presented in the coming discussions. The rumours compared to the five non-rumours listed in table 5.8, are *kim_divorce*, *obama_pay_increase*, and *evans_arrested*, the three higher impact rumours of the data, noted in table 5.7. This gives a total of fifteen experiments:
*(3 rumours) * (5 non-rumours).* For a full table of results, see table 5.9.

The five features are presented individually, the five deemed to be *influential* and *likely influential.* In each case, the study assesses whether the feature is more prominent in the rumour sets compared to the non-rumour sets. A feature more prominent in the rumour sets allows something that can be measured specific to rumours and their impact.

[4] http://www.bbc.com/news, accessed: 15/05/2016

Bucket Name	News Story	#tweets
china_apple	China shuts Apple's film and book services	241
marathon_space	Tim Peake 'runs' London Marathon from space	298
prince_chart	All-Prince top five in midweek chart	266
scottish_power	Scottish Power to pay 18m GBP in fines	288
tyrone_dublin	County Tyrone man shot dead in Dublin	102

TABLE 5.8: List of non-rumour buckets, associated news stories and the number of related non-rumour tweets collected.

Followers

Following feature analysis, it was found that followers are influential to rumour impact. By comparing the numbers of followers associated with impactful rumours against non-rumours, the study now asks if there are more followers associated with rumours compared with non-rumours.

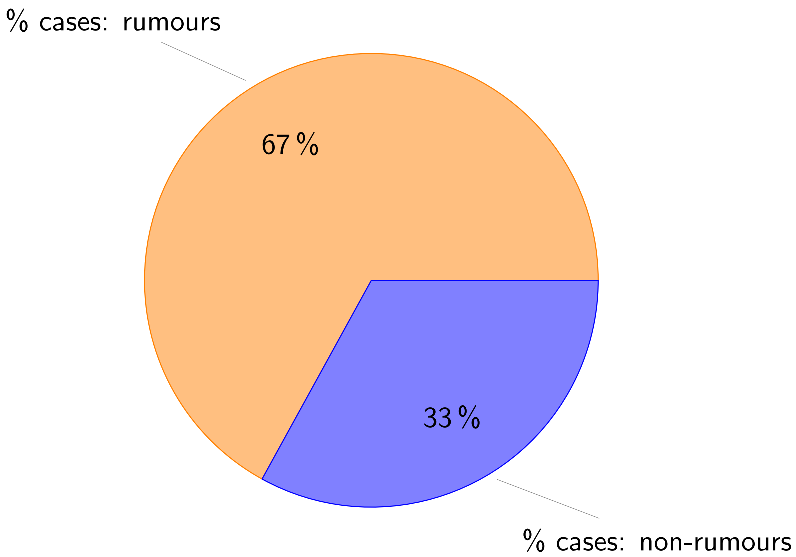

FIGURE 5.12: **Followers:** The % of experiments where rumour sets contain more followers vs. the % of experiments where non-rumour sets contain more followers.

Out of fifteen experiments, ten cases had more followers in the rumour set compared to the non-rumour set. This result represents 67%. This result is not quite conclusive but it is believed that with more data and more analyses, this percentage is likely to increase, and become more conclusive.

Under the specific conditions of the experiments of this study, *it can be concluded that the number of followers is likely to be an acceptable measure for rumour impact, a measure unique to messages deemed rumorous, highlighting the different nature of rumours compared to other messages.*

Friends

Following feature analysis, it was found that friends are influential to rumour impact. By comparing the numbers of friends associated with impactful rumours against non-rumours, the study now asks if there are more friends associated with rumours compared with non-rumours.

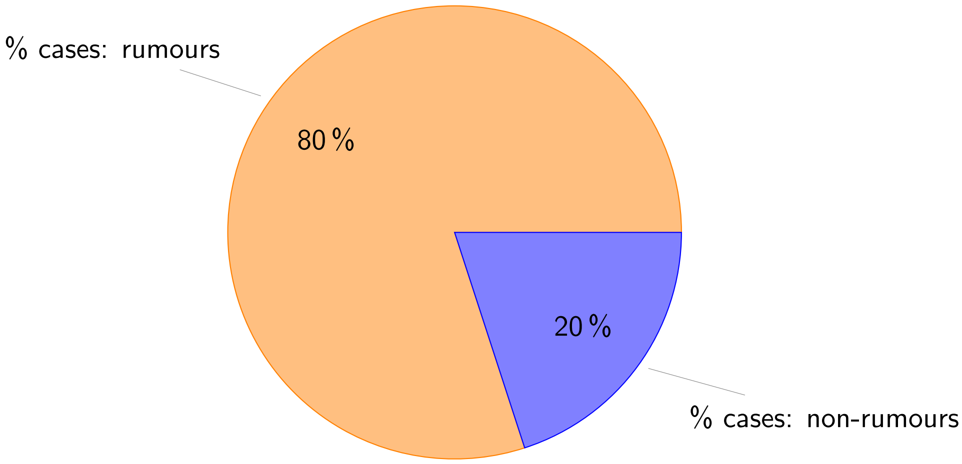

Figure 5.13: **Friends:** The % of experiments where rumour sets contain more friends vs. the % of experiments where non-rumour sets contain more friends.

Out of fifteen experiments, 12 cases had more followers in the rumour set compared to the non-rumour set. This result represents 80%. Under the specific conditions of the experiments of this study, *it can be concluded that the number of friends is an acceptable measure for rumour impact, a measure unique to messages deemed rumorous, highlighting the different nature of rumours compared to other messages.*

Statuses

Following feature analysis, it was found that the total number of statuses associated with the accounts of the rumour authors, i.e. total number of messages composed in the lifetime of the accounts, is influential to rumour impact. By comparing the numbers of statuses associated with the authors of impactful rumours against non-rumours, the study now asks if there are more statuses associated with the authors of rumours compared with non-rumours.

Out of fifteen experiments, only three cases had more total statuses associated with the authors in the rumour set compared to the non-rumour set. This result represents 20%. In other words, the authors posting non-rumours, related to credible news, tend to post more often in general, compared to those who post rumorous messages.

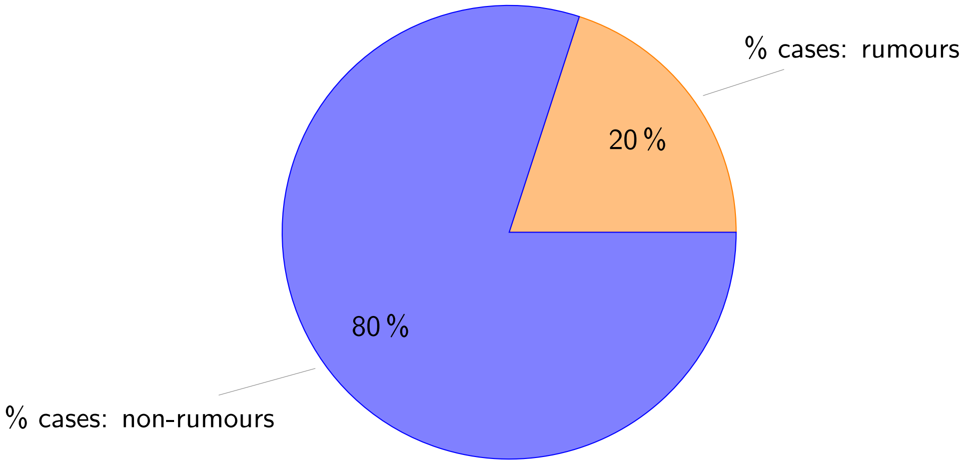

FIGURE 5.14: **Statuses:** The % of experiments where rumour sets contain more statuses vs. the % of experiments where non-rumour sets contain more statuses.

In the previous evaluation section, the higher impact rumours had authors with higher total statuses counts than that for the lower impact rumours, in 100% of experimental cases.

Under the specific conditions of the experiments of this study, *it can be concluded that the total number of statuses associated with the author accounts is a likely measure for the impact of all types of messages and is not unique to rumours.*

Media

Following feature analysis, it was found that the inclusion of media items (images) is likely to be influential to rumour impact. By comparing the numbers of tweets containing media items, associated with impactful rumours against non-rumours, the study now asks if there are more tweets containing media items associated with rumours compared with non-rumours.

Out of fifteen experiments, ten cases had more tweets containing media items in the rumour set compared to the non-rumour set. This result represents 67%, see figure 5.15. This result is not quite conclusive but it is believed that with more data and more analyses, this percentage is likely to increase, and become more conclusive.

Under the specific conditions of the experiments of this study, *it can be concluded that the inclusion of media items is likely to be an acceptable measure for rumour impact, a measure unique to messages deemed rumorous, highlighting the different nature of rumours compared to other messages.*

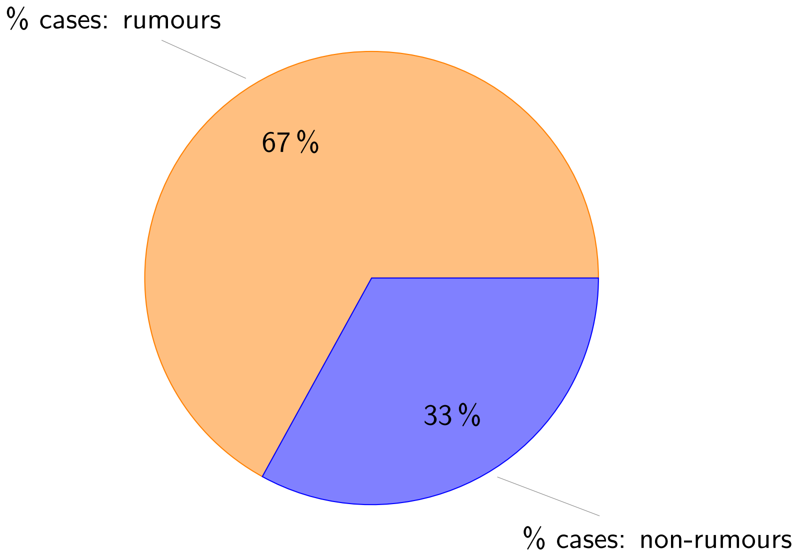

Figure 5.15: **Media Items (MIs):** The % of experiments where rumour sets include more tweets with MIs vs. the % of experiments where non-rumour sets include more tweets with MIs.

User Mentions

Following feature analysis, it was found that the inclusion of user mentions (tagging another user, @user) is likely to be influential to rumour impact. By comparing the numbers of tweets containing user mentions, associated with impactful rumours against non-rumours, the study now asks if there are more tweets containing user mentions associated with rumours compared with non-rumours.

Out of fifteen experiments, eight cases had more tweets containing user mentions in the rumour set compared to the non-rumour set. This result represents 53%, see figure 5.16. In other words, the authors posting non-rumours, related to credible news, tend to include user mentions as commonly (slightly more) as authors posting rumorous messages.

Under the specific conditions of the experiments of this study, *it can be concluded that the inclusion of user mentions is a likely measure for the impact of all types of messages and is not unique to rumours.*

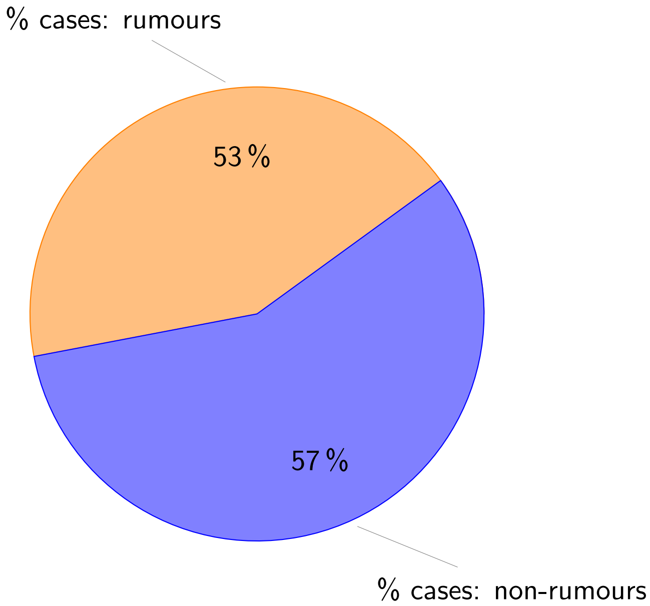

Figure 5.16: **User Mentions (UMs):** The % of experiments where rumour sets include more tweets with UMs vs. the % of experiments where non-rumour sets include more tweets with UMs.

Summary

Subsequent to analysing the feature properties of higher impact rumours compared to non-rumours, under the conditions of the data and experiments of this study, findings are now summarised.

Findings of this comprehensive study suggests,

- **followers** are likely to be an acceptable measure for rumour impact, unique to other messages, highlighting the different nature of rumours.

- **friends** are an acceptable measure for rumour impact, unique to other messages, highlighting the different nature of rumours.

- **total statuses** is a likely measure for the impact of all types of messages and is not unique to rumours.

- **media** is likely to be an acceptable measure for rumour impact, unique to other messages, highlighting the different nature of rumours.

- **user mentions** are a likely measure for the impact of all types of messages and is not unique to rumours.

Table 5.9: 15 experiments - Rumours (R) vs. Non-Rumours (NR).

R & NR No Statistical Diff in Impact (Higher count between R & NR in bold)								
	P		Bucket	FO	FR	S	M	UM
A	0.57	Rumour	*evans_arrested*	2030	**4998**	15677	**15**	1
		Non-Rum.	*china_apple*	**6795**	3058	**109785**	10	**4**
B	0.27	Rumour	*evans_arrested*	**258453**	**339**	14531	9	0
		Non-Rum.	*marathon_space*	111	39	**33829**	9	**7**
C	0.15	Rumour	*evans_arrested*	**423**	**296**	11777	**12**	1
		Non-Rum.	*prince_chart*	150	0	**125922**	0	**3**
D	0.09	Rumour	*evans_arrested*	**13782**	278	30093	**13**	1
		Non-Rum.	*scottish_power*	5266	**2685**	**738684**	2	**2**
E	0.17	Rumour	*evans_arrested*	169	**1082**	171	**21**	**2**
		Non-Rum.	*tyrone_dublin*	**2604**	485	**305417**	2	1
F	0.13	Rumour	*kim_divorce*	**9423**	**132**	14136	14	5
		Non-Rum.	*china_apple*	45	2	**35287**	**15**	**6**
G	0.87	Rumour	*kim_divorce*	**1203**	10	**30561**	**14**	0
		Non-Rum.	*marathon_space*	276	**220**	919	5	**3**
H	0.31	Rumour	*kim_divorce*	**752**	**1761**	1764	**18**	**1**
		Non-Rum.	*prince_chart*	116	53	**130655**	0	0
I	0.23	Rumour	*kim_divorce*	**728**	**393**	**6491**	**15**	3
		Non-Rum.	*scottish_power*	6	66	88	1	3
J	0.14	Rumour	*kim_divorce*	**752**	**1761**	1764	**18**	**3**
		Non-Rum.	*tyrone_dublin*	334	256	**239432**	3	0
K	0.08	Rumour	*obama_pay_incr*	9	**26**	856	2	**15**
		Non-Rum.	*china_apple*	**262**	10	**37071**	**4**	6
L	0.6	Rumour	*obama_pay_incr*	**5377**	**5747**	16368	3	**12**
		Non-Rum.	*marathon_space*	282	242	**48292**	**8**	6
M	0.13	Rumour	*obama_pay_incr*	**290**	**2005**	**12656**	**6**	**15**
		Non-Rum.	*prince_chart*	14	13	4917	1	0
N	0.24	Rumour	*obama_pay_incr*	65	**66**	484	2	**18**
		Non-Rum.	*scottish_power*	**110**	15	**206737**	2	1
O	0.19	Rumour	*obama_pay_incr*	90	271	417	2	**15**
		Non-Rum.	*tyrone_dublin*	**52380**	**1763**	**11233**	**3**	0

- **P:** Resultant P-value in experiment, P is compared to α. A P-value greater than 0.05 indicates strong evidence for the null hypothesis, signifying that the rumour and non-rumour are not different in terms of impact.

- **FO:** Accumulative #**Followers**
- **FR:** Accumulative #**Friends**
- **S:** Accumulative #**Statuses**
- **M:** #tweets in the bucket containing **Media**
- **UM:** #tweets in the bucket containing **User Mentions**

5.4.1.4 Discussion

The experiments conducted within the processes of *Impact Measurement & Impact Analysis* have addressed the research question posed in the dissertation. Firstly, the 13 larger buckets of data collected through the *Gathering* process were presented, and the impact scores calculated for each given detail. The formula for impact based on user engagements, specifically *retweets* and *favourites* for this study's Twitter analyses, was used to measure impact, impact scores for rumours, for which comparisons were then attempted.

Direct comparisons in the raw data for impact scores was difficult, due to substantially varying bucket sizes, consequential of the *snapshot* method of gathering, and the shaky nature of rumour data. T-test analyses was performed, allowing more conclusive comparison between rumours and their associated impact scores. The favourable outcome of t-test analyses was that statistical difference between rumours, in terms of their impact, would be found. Statistical difference was not found between the majority of rumour sets, but some were conclusive and deemed significant / different.

Obtaining statistical difference between rumours flagged those that were of higher impact compared to those of lower impact. The study was then in a position to apply feature analysis and compare the presence of *message-based* and *account-based* features in the higher impact rumour, and the lower impact rumour.

Many of the features were omitted from further analyses as their presence in the total dataset was unsubstantial. The study analysed a total of nine features, and compared their presence in the higher impact rumour and lower impact rumour. Features with larger presence in the higher impact rumour, such as *followers* were resultantly deemed *influential* to rumour impact.

Analyses was taken one step further as the study asked whether influential features are specifically influential to the impact of rumours, or whether they are of influence to the impact of all tweets in general. Features deemed influential to general tweet impact are still important measures for rumour impact, but not uniquely so. Other features deemed uniquely influential to rumour impact, allow us something to measure, specific to rumours, which the study believes behave differently to non-rumours.

5.5 Summary

In this chapter, an evaluation plan was first presented, which facilitated the assessment of each individual element of the system and research, from the perspective of the primary research question of the dissertation.

The first process to be evaluated was that of *Rumour Gathering*, which was demonstrated to have satisfied the research objective of effective rumour corpus generation, meeting the corpus requirements of the research. The contribution of the *Feature Retrieval* process was then evaluated, and deemed successful, following efficient retrieval of variables required for impact measurement, and a wide range of rumour tweet properties for impact contribution analysis.

The final part of the evaluation considered the contribution made by the *Impact Measurement & Impact Analysis* processes, and its role in answering the research question posed by the dissertation. The results demonstrated that the impact of many rumours had been measured, and features influential to such impact were deduced.

Chapter 6

Conclusion

Chapter Six concludes this dissertation. Its purpose is to demonstrate that the research objectives outlined in Chapter One have been met, to express the research contribution this work represents, and to discuss future work opportunities.

6.1 Objective Assessment

6.1.1 Rumour Gathering

Research Objective: *To gather a corpus of Twitter messages, a snapshot body of tweets, considered as rumorous.*

Before performing any study in rumour analysis, such as rumour impact on social media, in the case of this work, a sufficient collection of rumorous conversations was first required. This initial step was taken, and methods for rumour detection deliberated. Following a review of detection techniques, a keyword regular expression method was implemented.

This method allowed for the retrieval of messages associated with known rumours fitting the rumour definition of the work. Twitter was chosen as the rumour source, with the Twitter API facilitating collection. A *Gathering* tool was built, adhering to *snapshot* requirements involving instant rumour search, immediate consumption and processing of rumorous messages, at a single moment in time. The result was an adequate rumorous corpus, the foundation that enabled the research question to be addressed.

6.1.2 Impact Measurement

Research Objective: *To conduct an impact measurement of a given rumour, by a formula suggested by this work, calculating an impact score.*

This work's aim in analysing impact was to make progress within an inherent gap in the related State of the Art. Social media has provided new mediums for communication, and thus, new facilities for rumours to thrive. The study involved itself in measuring impact in terms of that on social media itself, referring to impact as a measure of message success in attraction and engagement of users in its network.

To this end, user engagements were selected as the variables which would be used in constructing the impact formula. Twitter users engage with a message, showing their appreciation, by *retweeting* and *favouriting*. Thus, the impact formula suggested by this work was, $impact = retweets + favourites$, as introduced in section 3.1.2, formula (3.1).

The system parsed the properties necessary for impact measurement from Tweet objects received during the *Feature Retrieval* process. These impact variables were stored with each data item (each rumour message). Impact was calculated for a given rumour by retrieving all messages within its rumour bucket, and performing necessary calculations, resulting in an impact score reflecting the above formula.

6.1.3 Statistical Significance

Research Objective: *To look for statistical difference between the means of various pairs of rumours, as a method of highlighting the higher impact of some rumours over others.*

To compare rumours in terms of their impact, this work strove to find statistical significance in its data. This would allow us to signify those rumours statistically different to each other in terms of impact. T-test hypothesis testing was included in the *Impact Measurement & Impact Analysis* system tool. T-test calculations were performed on various pairs of rumour random sample sets, in search of statistical difference.

These analyses were inherent to the research as they provided the means for rumour impact comparison, and subsequent analysis into the features that potentially influenced these differences in impact. The result was that statistical difference in much of the data was not found, rumour data being so unreliable and shaky by nature. However, and to this study's success, statistical difference between a number of cases was highlighted, for which the null hypotheses could be rejected, signifying that the rumours involved were in fact, statistically different in terms of impact.

6.1.4 Impact Contribution

Research Objective: *To assess possible contributors to rumour impact by investigation of message-based features and author-based features.*

In pursuit of this work's goal, involving the measurement of impact on social media, making effort to understand this impact was a related aim of the work. Rumour data, as flimsy as it is, was analysed in a bid to make sense of such data, as the work attempted to find properties that influenced social media impact, in the data provided by the study.

A suite of features, constructed of those deemed useful in related studies within rumour behaviour analysis, was constructed and organised into the categories, *account-based features* and *message-based features*, see section 2.4.2. The *Feature Retrieval* process was responsible for parsing metadata from rumorous messages received, and translating applicable metadata to feature logic.

T-test hypothesis testing highlighted the higher impact of some rumours compared to others, cases for which statistical difference was found. Investigations could now get under way into understanding such impact, and determining whether any of the features in the feature suite had an influence on the higher impact of some rumours. Interesting findings were the result of the investigations, with features deemed as being *influential*, *likely to be influential* and *not influential* to the impact of a rumour on social media itself.

These analyses were extended as a question of the special nature of rumours compared to all other messages was brought to the fore. The work enquired after some measurable difference between rumours and non-rumours - a measurable difference related to impact and features, i.e. some features that could be measured specific to rumours highlighting their special nature.

6.1.5 Final Remarks

Research Objective: *To conduct an evaluation of the approach, assessment of performance implications and offering of suggestions for future work.*

The meaningful output of this research has been a demonstration of the potential that exists within the study of rumour impact, for rumour behaviour analyses, which ultimately strives to *understand* something about the nature of a rumour. The work argued that classification of rumours in terms of their impact could be both fruitful and useful, as organisations such as news companies, continue to battle against the mounds of unverified information that exists on online social mediums.

Classifying in terms of impact allows those rumours which have potential to do damage or *go viral*, to be highlighted and put forward for further investigation, be that verification, debunking etc. Rumours that are frivolous, of low impact, receiving little user engagement, are naturally ignored as they are not worth any extra time or resources being spent on them, regardless of their truthfulness or falsehood.

Real-time classification methods for rumours in social media is of ongoing, and increasing importance, as social media continues to grow, and the abundance of information available is ever increasing and ever more immediate. As we understand more and more about the special nature of rumours, the data associated becomes rich enough to introduce tools for measuring the impact of rumours spreading through social media, allowing interested parties to expend time and effort on those of interest, those likely of high impact potential.

6.2 Research Contribution

The following discussion details the contributions of the research. Firstly, a dataset of rumours, associated with known rumours, sourced from online sites, has been collected. The dataset represents a rumorous corpus sufficient for experimental analyses within rumour behaviour, such as those associated with rumour impact. Feature data associated with every rumour message in the dataset has been collected and stored with the rumorous text, enriching the resultant dataset of rumours.

A formula has been suggested for calculating the impact of rumours on social media itself. The formula is built of user engagements, a direct action by users to show their appreciation, and thus, boost the success of a rumour in terms of attraction and impact. The formula presented by the work appears Twitter specific in its variable names, *retweets* and *favourites*. However, this may be customised / altered / extended, to reflect user engagements or other application specific properties on other social media, e.g. *shares* and *likes* on Facebook, *likes*, *comments* and *sends* on Instagram, etc.

Sense has been made of the rumour data collected by the study, and statistical significance has been highlighted. The work has proved that there exists statistical difference between some rumours in terms of social media impact, granting comparisons and subsequent analyses to be made.

The research expresses important findings related to rumour author features and rumour composition features that the study suggests are influential (and not influential) to the impact of rumours on social media. The study also suggests that rumours act differently to other messages, and that there exists measurable differences in terms of influential features unique to rumour impact.

In the context of the greater research area, the work bridges the void that exists in the State of the Art, and provides study, measurement, and analyses of rumour impact on social media itself, a gap created as rumours find new environments to thrive, resulting from the success and growth of social media. This study acts as an encouraging step towards building a customisable model for measuring impact, that can be applied over the various social media platforms that exist. This work introduces an exciting opportunity for extensive research in rumour impact on social media itself, where research should be continued.

6.3 Future Work

There are exciting opportunities for future work, potential extensions to the research undertaken in this dissertation.

6.3.1 Impact Benchmarks

The data constraints of this research resulted in a rumour dataset comprising rumour buckets of significantly varying population sizes. This made it difficult to directly compare the impact scores of rumours within the dataset, where bucket sizes varied from just 3 messages to 596. It would be useful to collect much more data, associated with lots more rumours, to enrich the dataset, filling it with many more buckets, allowing for impact score comparison between sets of similar size.

This would lead towards studies that could investigate and highlight benchmarks for rumour impact, i.e. a benchmark score for high impact, medium impact, low impact, for example. Benchmark scores would provide meaningful comparison, where a rumour searched in real-time with associated impact calculated, could be compared against such a benchmark to provide meaning to its impact score.

6.3.2 Longitudinal Impact Studies

Time constraints were also a restrictive factor of this study. A *snapshot* method for rumour collection and analysis was adopted, whereby rumours were collected at a given moment in time, and were immediately analysed as they stood in that moment. This was an exciting opportunity for this study as it interested itself in the real-time nature of rumours, and the performance of impact measurements in real-time.

However, a longitudinal research study that involves repeated observations of the same rumour over long periods of time, could prove beneficial. This would encourage studies involving the changing impact of social media rumours over their lifetime, and could prove influential as user engagements and feature counts increase, or possibly cease at some point during the longitudinal observation.

6.3.3 Collective Impact Studies across multiple Social Media services

This work focused on Twitter, as Twitter data is so rich for such a study, and reflective of rumours in the real world. Future work could consider cross-social-media studies, where rumours are collected from more than one social medium and investigated together, as the one dataset. This could prove highly interesting and pose questions of the influence of the specific microblogging service on social media rumour impact. There are opportunities to study whether rumours behave the same in all social media platforms, or whether there are medium specific differences.

6.3.4 Commercial Applications

As well as research opportunities, this work has a commercial and industrial relevance, where news companies and verification specific organisations, are paying more and more attention to rumours related to news and relevant information, spiralling through online social media. This work included a stepping stone towards an *Impact Tool*, by building the *Data Visualisation Interface* system component, which involved a graphical representation of impact scores and features, and live rumour search and impact scoring functionality.

An interface such as this demonstrates the potential for commercial products which provide such services for those industries interested. Following the November 2015 Paris attacks, FIRSTDRAFT NEWS, a "newsgathering and verification source for journalists", stated that the media needed to start debunking rumours in real time, proven by the horrors of Paris[1].

As always happens in chaotic, breaking news situations, rumours emerged quickly and were shared widely. A tool following this study could be used to highlight those rumours attracting high levels of user engagements, those of high impact on their social media. Time and money can then be spent on quickly debunking or verifying these rumours that have potential to do damage. Following the Paris attacks, news organisations published reflective round-up pieces, along the lines of, "The online rumours about Paris you shouldnt believe". However, these follow up articles published after an event, are too late. Live services where debunking or verification happens in real-time is now vital.

Governments too could benefit from some implementation of such a service, as they try to protect their citizens from false warnings and information during times of crises, e.g. the Great East Japan Earthquake in 2011 [19]. The continuation of this study has *real world* relevance, and could save both money and lives.

[1]https://medium.com/1st-draft/too-little-too-late-the-horror-of-paris-proves-the-media-need-to-debunk-rumours-in-real-time-2d52da2a6eb0.kljv03nyf, accessed: 15/05/2016

Bibliography

[1] Yuping Liu-Thompkins and Michelle Rogerson. Rising to stardom: An empirical investigation of the diffusion of user-generated content. *Journal of Interactive Marketing*, 26(2):71–82, 2012.

[2] Jean Noel Kapferer. Rumors. new brunswick, 1990.

[3] Gordon W Allport and Leo Postman. The psychology of rumor. 1947.

[4] Leon Festinger, Dorwin Cartwright, Kathleen Barber, Juliet Fleischl, Josephine Gottsdanker, Annette Keysen, and Gloria Leavitt. A study of a rumor: its origin and spread. *Human Relations*, 1948.

[5] Ralph L Rosnow and Gary A Fine. *Rumor and gossip: The social psychology of hearsay.* Elsevier, 1976.

[6] Stephan Lewandowsky, Ullrich KH Ecker, Colleen M Seifert, Norbert Schwarz, and John Cook. Misinformation and its correction continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3):106–131, 2012.

[7] Sabine A Einwiller and Michael A Kamins. Rumor has it: The moderating effect of identification on rumor impact and the effectiveness of rumor refutation1. *Journal of applied social psychology*, 38(9):2248–2272, 2008.

[8] Nicholas DiFonzo and Prashant Bordia. A tale of two corporations: Managing uncertainty during organizational change. *Human Resource Management*, 37(3-4): 295–303, 1998.

[9] Michal Lukasik, Trevor Cohn, and Kalina Bontcheva. Classifying tweet level judgements of rumours in social media. *hospital*, 796(487):132, 2015.

[10] Anatol Rapoport and Lionel I Rebhun. On the mathematical theory of rumor spread. *The bulletin of mathematical biophysics*, 14(4):375–383, 1952.

[11] Yamir Moreno, Maziar Nekovee, and Amalio F Pacheco. Dynamics of rumor spreading in complex networks. *Physical Review E*, 69(6):066130, 2004.

[12] Benjamin Doerr, Mahmoud Fouz, and Tobias Friedrich. Why rumors spread so quickly in social networks. *Communications of the ACM*, 55(6):70–75, 2012.

[13] Pamela Donovan. How idle is idle talk? one hundred years of rumor research. *Diogenes*, 54(1):59–82, 2007.

[14] Luciano Floridi. Information ethics: On the philosophical foundation of computer ethics. *Ethics and information technology*, 1(1):33–52, 1999.

[15] Willard V Quine. Main trends in recent philosophy: Two dogmas of empiricism. *The philosophical review*, pages 20–43, 1951.

[16] Ralph L Rosnow and Eric K Foster. Rumor and gossip research. *Psychological Science Agenda*, 19(4), 2005.

[17] Vahed Qazvinian, Emily Rosengren, Dragomir R Radev, and Qiaozhu Mei. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1589–1599. Association for Computational Linguistics, 2011.

[18] Arkaitz Zubiaga, Maria Liakata, Rob Procter, Kalina Bontcheva, and Peter Tolmie. Towards detecting rumours in social media. *arXiv preprint arXiv:1504.04712*, 2015.

[19] C Ogiue. An examination of rumors and false rumors during the great east japan earthquake (!”), 2011.

[20] Yuichi Tanaka, Yuta Sakamoto, and Hiroki Honda. The impact of posting urls in disaster-related tweets on rumor spreading behavior. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, pages 520–529. IEEE, 2014.

[21] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684. ACM, 2011.

[22] Paul Resnick, Samuel Carton, Souneil Park, Yuncheng Shen, and Nicole Zeffer. Rumorlens: A system for analyzing the impact of rumors and corrections in social media. In *Proc. Computational Journalism Conference*, 2014.

[23] Zhe Zhao, Paul Resnick, and Qiaozhu Mei. Early detection of rumors in social media from enquiry posts. In *International World Wide Web Conference Committee (IW3C2)*, 2015.

[24] Floyd H Allport and Milton Lepkin. Wartime rumors of waste and special privilege: why some people believe them. *The Journal of Abnormal and Social Psychology*, 40(1):3, 1945.

[25] Nicholas DiFonzo and Prashant Bordia. Rumor, gossip and urban legends. *Diogenes*, 54(1):19–35, 2007.

[26] Alex Koohang and Edward Weiss. Misinformation: toward creating a prevention framework. *Information Science*, 2003.

[27] Steven H Strogatz. Exploring complex networks. *Nature*, 410(6825):268–276, 2001.

[28] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-worldnetworks. *nature*, 393(6684):440–442, 1998.

[29] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200, 2001.

[30] Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. Prominent features of rumor propagation in online social media. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, pages 1103–1108. IEEE, 2013.

[31] eBizMBA. Top 15 most popular social networking sites — april 2016. `http://www.ebizmba.com/articles/social-networking-websites`.

[32] Arkaitz Zubiaga, Maria Liakata, Rob Procter, Kalina Bontcheva, and Peter Tolmie. Crowdsourcing the annotation of rumourous conversations in social media. In *Proceedings of the 24th International Conference on World Wide Web Companion*, pages 347–353. International World Wide Web Conferences Steering Committee, 2015.

[33] Faqs about verified accounts. `https://support.twitter.com/articles/119135`.

[34] Twitter top 100 most followers. `http://twittercounter.com/pages/100`, .

[35] Digital Information World Irfan Ahmed. 30+ twitter marketing stats: 2015 - infographic. `http://www.digitalinformationworld.com/2015/01/twitter-marketing-stats-and-facts-you-should-know.html`.

[36] HubSpot Lindsay Kolowich. 12 data-backed tips to increase your conversion rate on twitter. `http://blog.hubspot.com/marketing/twitter-stats-for-more-conversions`.

[37] BufferSocial Belle Beth Cooper. 10 surprising new twitter stats to help you reach more followers. `https://blog.bufferapp.com/10-new-twitter-stats-twitter-statistics-to-help-you-reach-your-followers`.

[38] Twitter via sms faqs. `https://support.twitter.com/groups/34-apps-sms-and-mobile/topics/153-twitter-via-sms/articles/14014-twitter-via-sms-faq#`, .

[39] Tweets. `https://dev.twitter.com/overview/api/tweets`, .

[40] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 56–65. ACM, 2007.

[41] The top 500 sites on the web. `http://www.alexa.com/topsites`, .

[42] Business Standard Leslie D'Monte. Swine flu's tweet tweet causes online flutter. `http://www.business-standard.com/article/technology/swine-flu-s-tweet-tweet-causes-online-flutter-109042900097_1.html`.

[43] Twitter usage / company facts. `https://about.twitter.com/company`, .

[44] Goldee Udani. An exhaustive study of twitter users across the world.[online] available http://www. beevolve. com/twitter-statistics. *Oc tober*, 2012.

[45] Twitter turns six. `https://blog.twitter.com/2012/twitter-turns-six`, .

[46] Jacob Ratkiewicz, Michael Conover, Mark Meiss, Bruno Gonçalves, Snehal Patil, Alessandro Flammini, and Filippo Menczer. Detecting and tracking the spread of astroturf memes in microblog streams. arxiv preprint. *arXiv preprint arXiv:1011.3768*, 2010.

[47] Susan Anthony. Anxiety and rumor. *The Journal of social psychology*, 89(1):91–98, 1973.

[48] Marianne E Jaeger, Susan Anthony, and Ralph L Rosnow. Who hears what from whom and with what effect a study of rumor. *Personality and Social Psychology Bulletin*, 6(3):473–478, 1980.

[49] Ralph L Rosnow, James L Esposito, and Leo Gibney. Factors influencing rumor spreading: Replication and extension. *Language & Communication*, 8(1):29–42, 1988.

[50] Albert Bifet and Eibe Frank. Sentiment knowledge discovery in twitter streaming data. In *Discovery Science*, pages 1–15. Springer, 2010.

[51] Kurt Back, Leon Festinger, Bernard Hymovitch, Harold Kelley, Stanley Schacter, and John Thibaut. The methodology of studying rumor transmission. *Human Relations*, 1950.

[52] Stanley Schachter and Harvey Burdick. A field experiment on rumor transmission and distortion. *The Journal of Abnormal and Social Psychology*, 50(3):363, 1955.

[53] Tamotsu Shibutani. *Improvised news: A sociological study of rumor*. Ardent Media, 1966.

[54] United States. National Advisory Commission on Civil Disorders and Otto Kerner. *Report of the National Advisory Commission on Civil Disorders, March 1, 1968*. US Government Printing Office, 1968.

[55] Edgar Morin. *Rumour in Orleans*. Pantheon, 1971.

[56] Terry Ann Knopf. *Rumors, race, and riots*. Transaction Publishers, 1975.

[57] Gary Alan Fine and Patricia A Turner. *Whispers on the color line: Rumor and race in America*. Univ of California Press, 2001.

[58] Allan J Kimmel. *Rumors and rumor control: A manager's guide to understanding and combatting rumors*. Routledge, 2004.

[59] Fredrick Koenig. Rumor in the marketplace. *The social psychology of commercial hearsay*, 1985.

[60] Prashant Bordia. Studying verbal interaction on the internet: The case of rumor transmission research. *Behavior Research Methods, Instruments, & Computers*, 28 (2):149–151, 1996.

[61] Prashant Bordia and Nicholas DiFonzo. Problem solving in social interactions on the internet: Rumor as social cognition. *Social Psychology Quarterly*, 67(1):33–49, 2004.

[62] Prashant Bordia, Nicholas Difonzo, and Artemis Chang. Rumor as group problem solving development patterns in informal computer-mediated groups. *Small Group Research*, 30(1):8–28, 1999.

[63] Prashant Bordia and Ralph L Rosnow. Rumor rest stops on the information highway transmission patterns in a computer-mediated rumor chain. *Human Communication Research*, 25(2):163–179, 1998.

[64] BJ Fogg and Hsiang Tseng. The elements of computer credibility. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 80–87. ACM, 1999.

[65] Eni Mustafaraj Markus Strohmaier Harald Schoen Gayo-Avello, Panagiotis Takis Metaxas, Daniel Peter Gloor, Carlos Castillo, Marcelo Mendoza, and Barbara

Poblete. Predicting information credibility in time-sensitive social media. *Internet Research*, 23(5):560–588, 2013.

[66] Rob Procter, Farida Vis, and Alex Voss. Reading the riots on twitter: methodological innovation for the analysis of big data. *International journal of social research methodology*, 16(3):197–214, 2013.

[67] Kate Starbird, Jim Maddock, Mania Orand, Peg Achterman, and Robert M Mason. Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing. *iConference 2014 Proceedings*, 2014.

[68] Michal Lukasik, Trevor Cohn, and Kalina Bontcheva. Point process modelling of rumour dynamics in social media. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL*, pages 518–523, 2015.

[69] Lisa M Giocomo, Eric A Zilli, Erik Fransén, and Michael E Hasselmo. Temporal frequency of subthreshold oscillations scales with entorhinal grid cell field spacing. *Science*, 315(5819):1719–1722, 2007.

[70] Office of the Attorney General. Communications (retention of data) act 2011. `http://www.irishstatutebook.ie/eli/2011/act/3/enacted/en/print`.

[71] Twitter developer agreement policy. `https://dev.twitter.com/overview/terms/agreement-and-policy`, .