Information Credibility Perception on Twitter

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Shafiza Mohd Shariff

B.Sc. (Hons.) in Computer Science, University Putra Malaysia, Malaysia,

M.Sc in Information Technology, University Technology MARA, Malaysia

School of Science,

College of Science Engineering and Health,

RMIT University

October 2018

Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

Shafiza Mohd Shariff

School of Science

RMIT University

October 2018

Acknowledgments

First and foremost, I wish to praise Allah the Almighty, for whom we ultimately depend on for guidance and protection. I would like to send my utmost gratitude to Him for giving me the opportunity, determination, knowledge, and strength to do my research. Allah's mercy and grace have been with me throughout my life and even more so during my research journey.

I would like to extend my deepest gratitude to my supervisor, Associate Professor Dr Xiuzhen Zhang, for her continuous support, assistance and encouragement. Apart from being an excellent supervisor, she is very knowledgeable of research in her subject area. I really appreciate the ideas, time and support she has given me during the tenure of my research.

I would also like to express my gratitude to my second supervisor, Professor Dr Mark Sanderson. I really appreciate his support, guidance and valuable advice. Indeed, his quick feedback and constructive comments were really inspiring and helpful. He always tries to make time for me, even when he is busy, and for that, I am truly grateful.

I would like to thank my colleagues and friends for the discussions, suggestions and criticism especially my lab mates, Husna Sarirah Husin, Sharin Hazlin Huspi, Evi Yulianti, Fatimah Alqahtani, Lishan Cui and Chen Xi. To the many participants of the user studies that have been conducted for this research, thank you. I am also thankful to my friends and colleagues back in Malaysia, and also my online researchers support group. To my sponsors, Majlis Amanah Rakyat (MARA) and Universiti Kuala Lumpur, thank you for the opportunity given to me to pursue my PhD study at RMIT University, Australia.

Being an overseas student, life would be hard without friends that act as my second family abroad. For that, I would also like to express my gratitude to my friends here in Melbourne especially from Kampung Victoria North Melbourne (KVNM) for supporting and encouraging me throughout my study, the care and concern given for my family's well-being here in Melbourne, and for all the knowledge during our fortnightly knowledge sharing session. I enjoyed spending time with all of them. Thanks for giving me such a joyful time.

Before I was a PhD student, I am foremost a wife, a mother and a daughter. For that, I would like to convey my deepest appreciation and gratitude to the love of my life, my other half, Tengku Mohd Abas for his constant support, sacrifice and love. Without him, I would not be able to focus and have the determination to complete my PhD research. The encouragement and continuous prayer by my parents and family back in Malaysia have also been my strength, and I am truly grateful and blessed with their love. Not forgetting my beautiful children, I am thankful for your sacrifice and support. Your existence is my achievement and encouragement in life. I owe everything to you. Thus this work is dedicated to my family. I love each and every one of you dearly.

Shafiza Mohd Shariff

Credits

Portions of the material described in this thesis have been published in peer-reviewed publications. These publications correspond to the following chapters in this thesis:

- Chapter 3: Shariff, S.M., Zhang, X. and Sanderson, M., 2014. User perception of information credibility of news on Twitter. In European Conference on Information Retrieval (ECIR), pp. 513-518. Springer International Publishing.

- Chapter 4: Shariff, S.M., Sanderson, M. and Zhang, X., 2016. Correlation Analysis of Readers Demographics and Tweet Credibility Perception. In European Conference on Information Retrieval (ECIR), pp. 453-465. Springer International Publishing.

- Chapter 4: Shariff, Shafiza Mohd, Xiuzhen Zhang, and Mark Sanderson, 2017. On the Credibility Perception of News on Twitter: Readers, Topics and Features. Computers in Human Behavior, Volume 75, pp. 785-796. Elsevier Ltd.

Other publication during the candidature that is not directly related to this thesis:

- Shariff, Shafiza Mohd, and Xiuzhen Zhang, 2014. A survey on deceptions in online social networks. International Conference on Computer and Information Sciences (IC-COINS), pp. 1-6. IEEE.

Note

Unless otherwise stated, all fractional results have been rounded to the displayed number of decimal figures.

Contents

List of Figures

List of Tables

Abstract

Information on Twitter is vast and varied. Readers must make their own judgements to determine the credibility of the great wealth of information presented on Twitter. This research aims to identify the factors that influence readers' judgements of the credibility of information on Twitter, especially news-related information. Both internal (within the Twitter platform) and external factors are studied in this research. User studies are conducted to collect readers' perceptions of the credibility of news-related tweets, Twitter features, and the impact of reader characteristics, such as a reader's demographic attributes, their personality and behaviour. Twitter readers are found to depend solely on surface tweet features in making these judgements such as the author's Twitter ID, pictures, or the number of retweets and likes, rather than the tweet's metadata as recommended in previous studies. In this study, surface features are related to cognitive heuristics. Cognitive heuristics are features that the mind uses as shortcuts for making quick evaluations such as deciding the credibility of tweets. There are three main types of cognitive heuristic features found on Twitter that readers use to determine credibility: endorsement, reputation and confirmation. This study finds that readers do not use only one single feature to make credibility judgements but

rather a combination of features. External factors such as a reader's educational background and geolocation also have a significant positive correlation with their perceptions of a tweet's credibility. Readers with tertiary level education, or living in a certain location or environment, such as in a crisis or conflict area, are observed to be more careful in making credibility judgements. Readers who possess *conscientiousness* and *openness to experience* personality traits are also seen to be very cautious in their credibility judgements. Another insight provided by this research is the categorisation of readers' behaviours according to credibility perceptions on Twitter. The behavioural categorisations are defined by readers' behavioural reliance on Twitter's surface features when judging the credibility of tweets. The findings can assist social media authors in designing the surface features of their social media content in order to enhance the content's credibility. Furthermore, findings from this research can help in developing effective credibility evaluation systems by considering readers' personal characteristics.

Chapter 1

Introduction

Online social media (OSM) has been around since 1996, but only after the emergence of Facebook in 2006 did social media become a global phenomenon [Boyd and Ellison, 2007]. Many sites require their users to be honest with the information they provide in their profiles, but users may not necessarily follow these policies. Once an account has been created, social media users will develop connections and build friendship network links with others [Musiał and Kazienko, 2013].

The friendship links allow users to experience online social activities with one another. Common activities include sharing links, files, photos and videos, updating user profiles, writing status updates, commenting on and tagging posts, photos, videos and files, adding and deleting friends and posting on friends' pages/walls [Wilson et al., 2009]. As a result of all these activities, large volumes of information are made publicly accessible. However, not all the information posted online can be trusted and true. There are many negative categories of information found online such as rumours, gossips, deceptive and fake informa-

tion [Alowibdi et al., 2014; Zubiaga et al., 2015]. People have to make the judgement of the online information themselves.

In making a judgement, people rely on some heuristic principles to make the process of assessing an event a simple and quick exercise of judgement [Tversky and Kahneman, 1974]. The use of heuristic principles is explained by Evans [2003] whereby people depend on two systems known as heuristic and analytic systems to make a decision. The two systems are the two major information processing systems described in the dual-processing theory of decision making. The heuristic system is an automatic and implicit process, while the analytic system concerns rational and critical thinking [Evans and Curtis-Holmes, 2005]. When readers evaluate online information for credibility, they automatically use the heuristic system to gain an overview of the credibility of the information. The analytic system may then be engaged depending on factors such as a reader's motivation and willingness to evaluate credibility [Lim, 2009; Rieh and Hilligoss, 2008], a reader's information skills [Lucassen et al., 2013], or language [Eysenbach and Köhler, 2002]. Therefore, it is important for readers to understand the factors that could help them identify and determine the credibility of online content.

1.1 Background

A popular social media platform is Twitter, which allows users to post messages publicly that are currently limited to 280 characters. Other than text, Twitter users can also post up to four pictures, a GIF file and one video. These messages are called tweets. Only Twitter account members can post on Twitter, while non-members can only read the tweets.

Both members and non-members can search for tweets on the platform. Twitter users are encouraged to share freely: anything from self-opinion tweets regarding a topic or sharing articles from other newswires. Almost anything can be shared on Twitter.

With the variety of information now available online, the process of judging whether information is true or false depends entirely on readers. This process is known as credibility perception. Research on credibility mostly resides in the areas of human-computer interaction (HCI), communication and information science. In each research area, the credibility study approach varies. In HCI, credibility studies focus on determining key features that could help developers design credible-looking websites, blogs and social media posts. The interface design could help users in undertaking a credibility assessment process of the different online platforms. Rieh et al. [2014] showed that bloggers adapt this technique when designing their blogs. The bloggers design their blogs depending on how they want their readers to perceive the credibility of their blog.

In the communication field, credibility research focuses on the way information is conveyed depending on the information type, and the media and users' behaviours in response to the information. Kaye and Johnson [2011] described how a blog reader's motivation to read a blog would influence their perception of the blog's credibility level depending on the blog's genre. In this context, Sundar [1999] defined credibility as "global evaluation of the objectivity of the story" (page 380). We can also see a similar pattern in online social media where the users show different credibility perceptions for different genres of tweets [Morris et al., 2012].

While credibility research in the field of communication is similar to that of information science, information science focusses on the criteria users rely on when seeking reliable infor-

mation. Castillo et al. [2011] described the criteria for a credible tweet is based on four main features derived from Twitter: message-based, content-based, user-based and propagation-based. Kang et al. [2012] further added that credibility on social media is best judged by social connections (the number of followers and followees). They found connections achieved a higher predictive accuracy for credibility in their study.

Existing studies have focused on tweet credibility prediction by supervised learning using features from tweet content, the tweet author's social network, and the source of retweets. Prediction models trained from human-annotated credibility ratings are used to predict the credibility of unseen tweets [Castillo et al., 2011], while the tweet credibility prediction model presented in the study by Gupta and Kumaraguru [2012] was used to rank news event tweets by credibility using content-based features such as the number of unique characters in the tweet, and user-based features like the length of the author's username. Continuing on from their work in 2012, Gupta and Kumaraguru, released TweetCred [Gupta et al., 2014], a public tweet credibility prediction tool ranking tweets credibility in real time. Other studies have focused on the utility of individual features for automatically predicting credibility, such as the work by ODonovan et al. [2012] regarding topic-specific tweet collections, and on the credibility verification of tweets for journalists based on the author's influence, the media and information quality and the geolocation of the author as an eyewitness for the news event [Schifferes et al., 2014].

Another class of research has examined the features influencing readers' credibility perceptions of tweets. Examining certain tweet features, Morris et al. [2012] studied just under 300 readers from the US. The authors identified that a tweet written by authors with a

topically-related display name influenced reader credibility perceptions. Meanwhile, Yang et al. [2013] conducted similar research comparing readers from China and the US and found that different cultural backgrounds affect the credibility perceptions of tweets differently, in terms of what and how features were used.

From the two studies above, it is made clear that criteria for credibility perception is mainly about people's personal judgements, and perceptions are subjective. Rieh and Hilligoss [2008] supported this, having identified believability, information quality, and peripheral cues as part of the credibility constructs used to make a judgement. Thus, understanding how readers determine the credibility level of information online is the motivation of this research: to study the factors influencing readers' credibility judgements of tweets, especially when tweets are from authors outside of their trusted friendship network links (the follower-followee relationship), such as tweets retrieved from a search activity. Figure 1.1 shows an example of tweet messages on Twitter regarding the London riot in 2011 that are retrieved through the search request. Although these tweet messages are actually rumours, some may still judge these tweet messages as true (note that the first tweet has been shared three times).

The design of current credibility evaluation system is based on identification and credibility prediction based on the behind-the-scene information that is out of the view of the readers, such as friendship network links and the propagation network of the tweets. This system may not be applicable to readers when deciding credibility levels as the two heuristic-analytic decision-making systems serve as the conceptual basis in credibility judgments. Therefore, it is necessary to understand readers' credibility perception behaviours and how readers use
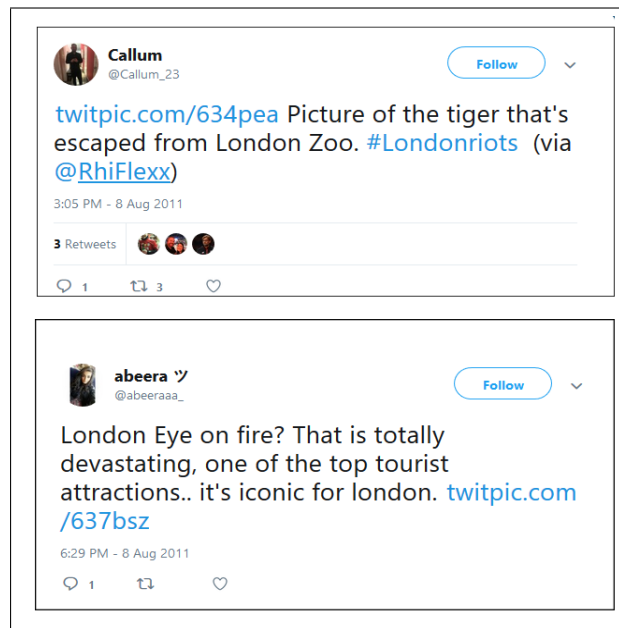
Figure 1.1: Example of tweet messages

heuristics in judging the credibility of a tweet. Furthermore, there is a need to take into account other integrated factors regarding readers such as culture, education, age and gender, and personal characteristics, in understanding the role of personal characteristics in credibility judgments.

1.2 Scope of the Thesis

In this research, we focus on the information seeking behaviour of credible news-related tweets as topical information content is crucial and misinformation could set off panic among the public. As there are many news tweet topics available on Twitter, it is essential to know what features best help people judge credibility or if these features differ depending on the topic. Tweet message surface features are available directly to readers at first glance, and

what is engaged by readers is based on a readers personal characteristics.

In this study, the surface features on the tweet message are: the text including (if any) the user mention, URL and hashtag, author's description such as their picture and Twitter ID, the date the tweet was posted, number of times the tweet has been shared, also known as retweets, and the number of people liking the tweet. Refer to Figure 1.1 for an example of the surface features on tweet messages. First glance refers to when a reader is conducting a heuristic process in using certain features to make a quick decision. Meanwhile, the definition of personal characteristics is based on that established by Donnellan et al. [2009] describing: a) the individual variations in ways of thinking, feeling and acting; and b) adaptation to their environment throughout their life (e.g. work, relationship, culture). Therefore, personal characteristics will cover a reader's demographic data, personality traits and behaviour.

Our research focuses on the scenario where a reader is searching for news topic information on Twitter using the search platform. We chose to focus on the credibility perception of Twitter search results as readers are fed with tweets related to their search queries rather than news stories tweeted by authors within their trust social network [Hu et al., 2012; Petrovic et al., 2013].

1.3 Research Objectives

In this research, there are three research objectives that we would like to achieve. The objectives will provide the description of the specific actions we will take in order to reach the aim of this research in understanding readers' credibility perceptions. The three objectives are as follows:

- The first objective is to identify what features readers use to help them determine the credibility level of information on Twitter. Feature identification aims to compare the features readers use and the automated credibility prediction and evaluation systems on Twitter that have been proposed in previous studies.

- Our second objective is to study the correlation between a reader's credibility perceptions and surface feature preferences with other factors such as news tweets and demographic attributes. Through this goal, we aim to address our hypothesis that a reader's credibility perceptions of different news attributes are related with other factors.

- In the last objective, we want to investigate the connections between a reader's behaviour, reliance on heuristics and a reader's personality in assessing the credibility level of news tweets. The underlying idea is that a reader's personal characteristics may influence their belief regarding the truth of online information. Therefore, the behaviour of readers when perceiving the credibility level of a tweet will indicate the credibility design of tweets.

1.4 Research Questions

A reader must judge the credibility of information on Twitter. While previous studies have focused on the multitude of features found on Twitter, including the visible and metadata features, it is unclear whether those features are also used by readers. In this research, we focus on understanding the factors that impact online content credibility. Specifically, we address the following research questions:

RQ1: How do readers judge the credibility of tweets?

In most existing work, credibility judgements are derived based on the features compiled by crawling the Twitter space and extracting the link relationship between twitterers. The compiled features are used for developing automatic predictions and may not necessarily reflect readers' perceptions of important signals of credibility. We hypothesise that readers use surface features, features that are made apparent to readers, in forming credibility perceptions rather than spending the time to authenticate the author's information or reading comments on a tweet when searching for topic-related tweets as proposed by previous studies. For this research question, identifying a reader's credibility perception of news-related tweets is the main focus.

To achieve RQ1, a user study is conducted to examine Twitter readers' perceptions of the credibility of tweets for news events. News topics and corresponding relevant tweets are selected for the study. Based on readers' comments on how they make credibility judgements of tweets, features are extracted and analysed using association rule analysis to establish the connections between features and credibility levels. The extracted features are then mapped to a detailed list of features described by Castillo et al. [2011]. This is discussed in more detail in Chapter 3.

RQ2: Does the credibility of tweets correlates with factors related to external attributes other than tweet surface features?

A study by Yang et al. [2013] showed that cultural background affects the credibility perceptions of tweets, in terms of what and how features are used. Based on this study, we hypothesise that demographic attributes also influence a reader's credibility perceptions and

the features they use in forming perceptions. So, this research question will examine the relationship between a reader's credibility perceptions and surface feature preferences on Twitter, and other underlying factors such as demographic and news tweets attributes. We further study the correlation between these factors.

To address this research question, a user study is conducted to explore the correlation between readers' demographic attributes, credibility perceptions, and the features used to judge tweet credibility. Only the surface features presented in tweets available directly to readers will be studied based on the findings in RQ1. Correlation analysis is conducted to study the correlation between each demographic attribute with credibility perceptions, news attributes and features. This is discussed in more detail in Chapter 4.

RQ3: Do Twitter readers' personal characteristics play an important role in credibility evaluations of information on Twitter?

Our intuition is that a reader's credibility perceptions of information on Twitter may also be influenced by the attitudes and behaviours of the reader. This intuition is further strengthened by a study by Ahmad et al. [2011] that suggested a reader's personal characteristics influence their belief regarding the truth of information on websites during the information-seeking process. Thus, it is important to study the connections between a reader's behaviour, their reliance on heuristics and their personality in assessing the credibility level of news tweets.

To answer RQ3, a user study was conducted to capture Twitter readers' behaviours in perceiving the credibility of news tweets and readers' personalities. Data collected from the user study was analysed with factor analysis and a multiple regression model. The findings

were then compared to the information-seeking behaviour and decision-making model. This is discussed in more detail in Chapter 5.

1.5 Research Contributions

This study builds on previous research regarding online information credibility perceptions. By studying readers' credibility perceptions of tweet messages and the features they use to make these judgements, this study concludes that:

1. *Readers use Twitter's features found on the tweet messages, known as surface features, rather than a tweet's metadata, behind-the-scenes information about the tweet message, to judge the credibility of tweets.*

2. *Only certain external factors related to demographic attributes correlate with credibility perceptions.*

3. *Personal characteristic influence, to some extent, how readers determine the credibility of online content.*

These contributions are achieved by, first, understanding how a reader makes credibility judgements of information in tweet messages. Although previous research has focused on the behind-the-scene information of a tweet, or the tweet's metadata, such as by investigating the author's social network relationships, this study finds that readers assign more value to the surface features as credibility indicators than a tweet's metadata. Furthermore, the features are used in combination rather than as singular features to help with the heuristic processing inherent to forming credibility perceptions.

Since the features are used as part of the decision-making process to determine credibility levels, this research is able to distinguish three categories of cognitive heuristics: endorsement, reputation and confirmation. Endorsement heuristics describe the way readers perceive information as more credible if other people agree with the information presented. Reputation heuristics refer to the reputation of the authors of the tweet messages and, lastly, confirmation heuristics describe the features that readers use to validate the information in the tweet message. Although previous web credibility studies have also mentioned cognitive heuristics, Twitter has different sets of features not available on the web such as hashtags and user mentions, and the design of tweet messages is dissimilar to messages presented elsewhere on the web.

Aside from the use of tweets' surface features, using the data collected from the study, we are able to establish a relationship between demographic data and credibility perceptions. Education levels and the geolocation of readers were also found to contribute to the way readers view the credibility levels of news tweets. Readers with a higher level of education and in certain living conditions and locations are seen to be more careful in making credibility judgements. Not only that, this research also found that personality traits do have a role to play in readers' credibility perceptions. Readers who have the *conscientiousness* and *openness to experience* personality traits are seen to be very cautious in making credibility judgements.

Moreover, this research provides results regarding readers' credibility perception behaviours. The results show that there are three categories of credibility perception behaviour on Twitter: those who overly depend on Twitter features that relate to confirmation heuris-

tics, those who modestly rely on features relating to cognitive heuristics, and readers who only slightly depend on cognitive heuristics to help them determine the credibility of tweets. This finding is defined by their behavioural reliance on Twitter's surface features when judging the credibility level of tweets.

1.6 Thesis Organisation

The remainder of the thesis is organised as follows:

- **Chapter 2** summarises previous research. The fundamentals of Twitter, the diffusion of information on online social media (OSM) and the relevant literature in assessing the trustworthiness of information on online social media, will be discussed.

- **Chapter 3** describes the methodology used to answer the research questions. Our basis for conducting user studies to investigate readers' credibility perceptions and the factors that impact credibility are discussed.

- **Chapter 4** describes the identification of Twitter features readers use to judge the credibility level of tweets.

- **Chapter 5** describes the correlation between different factors regarding topics, features and demographic attributes that influence readers' credibility perceptions of information posted on Twitter.

- **Chapter 6** discusses the connections between a reader's reliance on heuristics and a reader's personal traits in perceiving the credibility of news tweets.

- **Chapter 7** summarises the findings, concludes the present work and proposes future work based on the outcomes and limitations of this research.

- **Appendix** presents the list of news-related tweet topics and Human Ethics Application approval letters regarding the user studies conducted in this research.

Chapter 2

Research Background

A large body of literature exists on online credibility evaluation, with particular emphasis on online social media (OSM), specifically Twitter. In Section 2.1, we will provide background knowledge on the theoretical part of information credibility, and define the terms related to credibility. Section 2.2 will discuss readers' credibility perception behaviours and factors that have an impact on credibility judgements. This is then followed by a description of credibility evaluation methodologies (Section 2.3) and we summarise the implications of this research and research gaps regarding information credibility perceptions (Section 2.4). Lastly, Section 2.5 provides a summary of this chapter.

2.1 Credibility

In order to discuss previous studies on information credibility assessment, we will first formulate and define some essential terms, and differentiate the different credibility assessments of online information.

2.1.1 Credibility and Trust

The Merriam-Webster dictionary defines credibility as "the quality of being believed or accepted as true, real, or honest". Based on this definition and work by Fogg [2003] regarding factors influencing final judgements, credibility generally regards one's belief in the truth due to an initial impression of a subject. Fogg [2003] distinguishes four types of credibility: presumed, earned, reputed and surface. Presumed credibility reflects general assumptions held by a reader. Earned credibility refers to the experience a reader has with a platform. Reputed credibility is the need for cognition/approval such as links to other reputed websites and other factors. Finally, surface credibility is connected with the design, usability of a platform (e.g. online social media, website) and others. These different types of credibility assessment are taken into account by readers depending on individual differences. Here, credibility is seen as a process where a user decides whether a trust relationship would concur between a trustor and trustee based on the quality of the information and the reputation or quality of a source.

Credibility and trust are closely related. Mayer et al. [1995] defined trust as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (page 712). Wang and Emurian [2005] further listed four characteristics of trust: trustor and trustee (relationship between two parties), vulnerability (involve uncertainty and risk), produced actions (transactions or activities between the two parties) and subjective matter (behaviour or attitude of the two parties regarding the transaction that occurred). Based on the two definitions, trust is viewed as the interper-

sonal relationship between a trustor (the person who trusts) and a trustee (the person being trusted).

Trust in online social media is classified into four types: asymmetric, transitive, context-dependent and personal [Johnson, 2011]. Asymmetry in trust means that two profiles have a different level of trust with one another. If the trust develops from a close mutual friendship, it is called transitive trust. Sometimes users choose to have a different degree of trust depending on circumstances (context-dependent) and, lastly, each user is entitled to have their own opinion and views of the same individuals. In an observation of Facebook users conducted by Dwyer et al. [2007], many users were found to have high confidence in online social media (OSM) to protect their personal information and their social trust levels increase when their friends also use a platform.

Tseng and Fogg [1999] distinguished credibility from trust. In their work, trust is about the dependability and reliability of an object or person, while credibility regards the believability or trustworthiness of information or the quality of the source from where such information comes. The notion of credibility is the quality of being believed or accepted as true, real, or honest, whether it regards the information or the source. Tseng and Fogg [1999] indicate that there are two key elements in the measurement of credibility: trustworthiness and expertise. This concept supports the work by Hovland and Weiss [1951] where information coming from an expert receives higher credibility perception than the information that comes from a questionable source.

Rieh and Hilligoss [2008] discover a person's experience and similar social connections influence the credibility judgement of information, online or offline. Sikdar et al. [2013b]

describe how online communication in social media shows the type of relationship between the author and the reader, thus explaining why readers tend to believe the online information posted by their group of friends, the people they follow on social media, over those who are not in that friendship group. Even friends and the people readers follow online yield different online information credibility perception levels. Similar findings are mentioned in studies by McKnight and Kacmar [2007]; Rafalak et al. [2014]. These findings show that credibility perception not only focuses on information but also on the trust relationship between reader and author.

In this research, we adopted the notion of credibility by Tseng and Fogg [1999] where credibility is the quality of being believed or accepted as true, real, or honest, whether it regards the information or the source. In this sense, we distinguish that, before trust can be established, credibility must be present. Believing an information or source is credible allows a trust relationship to form. Thus, this thesis studies the factors that would have an effect on the formation of credibility perception so that trust can subsequently be established.

2.2 Credibility Factors

In previous studies, researchers study and propose different factors are involved in the formation of credibility of online information on websites and social media. Most of the factors in the literature are based on peripheral cues and system surface features, as these are the exterior features on the line of sight of the readers, not requiring them to venture deeper in investigating content. These features hold a fairly low-effort assumption regarding how to assess online content. In this section, we will discuss these credibility factors on two platforms,

the web and online social media.

2.2.1 Sources of Information

On websites, the reputation and trustworthiness of a source are important features in determining whether online information on pages such as blogs or article posts are credible, as shown in previous studies [Armstrong and McAdams, 2009; Fogg et al., 2001; Lucassen et al., 2013; McKnight and Kacmar, 2007; Rieh and Hilligoss, 2008; Sundar et al., 2007]. The source credibility features are, however, different on social media. Researchers found that an author's profile and social network on Twitter represent a higher credibility indicator regarding the tweet messages the author posted [AlRubaian et al., 2015; Canini et al., 2011; Castillo et al., 2011; Gupta and Kumaraguru, 2012; Gupta et al., 2012; Kang et al., 2012; Morris et al., 2012; Sikdar et al., 2013b; Yang et al., 2013].

The differences in a source's credibility features between websites and social media are due to the interface design of each platform. On a website, the image and information regarding the author is limited or not available. Readers also consider the site host, such as the name of a news agency or the name of the company, as a source of the information if the details of the author are not made available [Sundar et al., 2007]. In contrast, a social media author's information such as image, gender, location and whom they are connected with on social media is more accessible and thus gives more confidence to the readers of the online information regarding the credibility of the source and the information posted by the source.

2.2.2 Composition of Information

The next credibility factor for online information is the information itself. The main feature of credible online information is the level of relatedness of the information to the topic [Flanagin and Metzger, 2000; Rieh and Hilligoss, 2008; Sundar et al., 2007]. Topic related information is described as being informative and relevant to the keyword search and title of the post [Flanagin and Metzger, 2000] or having similar information to other online information from different websites regarding the topic or keyword [Sundar et al., 2007]. Although topic-related message posts on social media and on websites is still a prominent feature for credibility as shown in previous studies [Aladhadh et al., 2014; Canini et al., 2011; Castillo et al., 2011; Gupta and Kumaraguru, 2012; Kang et al., 2012; Morris et al., 2012; ODonovan et al., 2012; Sikdar et al., 2013b], language is another credibility feature proposed by researchers as a credibility indicator for information on social media.

Since social media is a platform for anyone to post messages about anything, the language used on such posts is informal and contains Internet abbreviations, especially on Twitter due to the limit of characters per post. A post with formal language will be perceived as more credible than one with slang and abbreviations [Gupta and Kumaraguru, 2012; ODonovan et al., 2012; Sikdar et al., 2013b]. A language element that has been proposed is the use of sentiment and semantic words. These words describe the opinion of an author and the positive/negative relation of the post towards the topic. Other than these, social media has special features that can be used as part of the information text that are not found or widely used on the web: hashtags and user mention. The use of features such as hashtags, which start with the symbol '#', and the user mention indicator which starts with the character

'@', also give an indication of the relation between the information with the topic [AlRubaian et al., 2015; Gupta et al., 2012; ODonovan et al., 2012].

2.2.3 External Source Links

The next credibility factor is embedding external sources in the information. Social media users sometimes link external sources on their posts, such as images or links, for example, to other online content. The external sources will give an impression that the information posted is not fabricated [AlRubaian et al., 2015; Castillo et al., 2011; Gupta and Kumaraguru, 2012; Gupta et al., 2012; Kang et al., 2015]. Husin et al. [2013] described that social network users access news agency websites for further information by the hyperlink embedded in the news story posts on online social media such as Twitter, a verification act when the readers are deciding the credibility level of the post's author and evaluating the information [Rieh and Hilligoss, 2008].

Meanwhile, the use of external links on web online content, such as blogs, are mostly linking to articles within the same web platform. However, there have been some websites or blogs that have embedded the URL link to the external source, if the article is based on the source such as Wikipedia [Rieh and Hilligoss, 2008]. Again, this is an option for readers to do their own verification of the credibility of the information.

2.2.4 Interface Design Layout

Other than information structures, the design or features available on a website and social media platform are another credibility feature contributing to the credibility perceptions

of online information. On websites, the layout design of the site gives different credibility perceptions of the online information appearing on those sites [Flanagin and Metzger, 2000]. If the design consists of much commercial content such as advertisements, it gives a lower credibility perception than a layout design that shows a real-world feel and ease of use [Fogg et al., 2001]. The time indicator on the website is also a credibility feature as it will show the recency of the online information posted on the website. The time indicator could also ensure the information is consistent with other information posted on other sites within the same time period [Sundar et al., 2007].

The time indicator is also an important design feature on social media. Consistent with Sundar et al. [2007], who found the importance of knowing when an article is posted on the website, AlRubaian et al. [2015]; Morris et al. [2012] and Kang et al. [2015] also discovered that if a message posted on social media is dated in the middle of the occurrence of a trending issue, the higher the credibility perception of the message post. Other credibility features for online information based on the social media platform design are propagation and count features. Both give an indicator that other social media users find the message posted to be trustworthy. The users can choose to propagate the message using a sharing method, and the number of shares will be shown [AlRubaian et al., 2015; Castillo et al., 2011; Gupta et al., 2012; Kang et al., 2015; Morris et al., 2012; Sikdar et al., 2013b]. Another design feature that has been proposed by Li and Suh [2015] as a credibility indicator on social media is the interaction feature. This feature is the communication between social media users on their own social media page by using the user mention feature in their message post and also in the comment section underneath each message post. Although some researchers argue that

the use of user mention is a credibility indicator on the message post as discussed earlier, Li and Suh [2015] focused on the use of the user mention in the design of social media as it is what makes social media unique.

Generally, the aim of information credibility studies is to find features that make good credibility indicators online. However, there are few studies that consider reader's attributes. The studies that relate to readers mostly look at the use of surface features, including the source features. Demographic attributes of readers are also used to differentiate the population and their credibility judgement [Lucassen and Schraagen, 2011; Morris et al., 2012; Tseng and Fogg, 1999].

2.2.5 Readers' Attributes

Research on characterising users and their behaviour for credibility perceptions in the online community is important and so is discussed separately in this section. The credibility of information on the Internet is rated higher by experienced and savvy users rather than by less experienced users. In addition, the freedom users have in choosing the information from the source they deem credible based on their experience and verification steps also contribute to the high credibility rating of the Internet [Flanagin and Metzger, 2000]. Due to this, Rieh et al. [2014] described in their work how blog authors would design their blog based on the credibility perception of their readers so that the information presented in the blog would be believed and deemed trustworthy. Thus, this would create a trust relationship between the blog authors and their readers when the blog is perceived as credible.

Readers' familiarity with websites and social media sites, and behaviour when accessing

online information also shows a difference in the way readers perceive the credibility level of online information. Readers also use the interface designs of where the online information was published to help them decide the credibility level of the information. The reliance on interface design is a psychological action known as cognitive heuristics. Sundar [2008] and Metzger et al. [2010] both established the use of cognitive heuristics in credibility perception of online information on the web while Yin et al. [2018] discuss the different cognitive heuristics preferences between a reader's gender when perceiving the credibility of information posted or shared on social media. In another study on cognitive heuristics, Lin et al. [2016] examined the ability of three specific heuristic cues based on non-content features: authority, identity and bandwagon cues, in making source credibility judgement. However, the study does not identify and analyse the overall cognitive heuristic cues that readers value, which also include content-based features in credibility assessment.

Demographic attributes also show a relationship with credibility perception. Fogg et al. [2001] describe that demographic profiles are seen to be correlated with the credibility perception of websites regarding specific factors. They discovered that website credibility elements such as interface, expertise and security are influenced by a reader's demographic attributes. In the visual factor, another study found that the user's demographics influenced the perception of online media when shown different manipulation levels of online news photos [Greer and Gosen, 2002]. Although not on the subject of photo manipulation, Kang et al. [2015] found a similar correlation between demographic attributes and visual features on microblogs where young people judged as more credible Twitter posts having these visual features.

There are other credibility studies that report the relationship between user behaviour

and features that contribute to a reader's credibility perception, such as the influence of the tweet author's location [Aladhadh et al., 2014], the reader's demographic attributes [Kang et al., 2015], cognitive heuristics for website credibility assessment [Metzger et al., 2010; Sundar, 2008; Yin et al., 2018], and the effect of heuristic indicators for source credibility assessment on Twitter [Lin et al., 2016]. The research shows that credibility studies that incorporate the reader's element as a feature related to online information credibility are a minority.

2.3 Credibility Evaluation

Assessing the credibility of information is a challenging task since "credibility" is a complex concept, as discussed in Section 2.1.1. Credibility assessments need to be considered relative to both people's credibility judgements and credibility contexts such as environment, situations, expectations, and others [Fogg, 2003; Wathen and Burkell, 2002]. Based on the credibility factors studied by other researchers, three main components affecting perception of the credibility of online information can be identified:

1. Contexts such as language, topic, event or the type of online platform

2. Available features (cues) on the online platform

3. Reader traits and cognitive heuristics such as the selection of cues in making a credibility judgement

In order to gather these credibility factors, researchers have adopted quantitative and qualitative studies. A closed question is often used in quantitative data where a specific

set of answers is provided for the participants to choose, while an open question gives the participants a choice to express themselves freely, producing qualitative data. Section 2.3.1 discusses user studies obtained from previous studies in relation to credibility.

In analysing the data collection, different methods have been used for evaluating credibility. Statistical analysis is one of the methods. The analysis can be examined descriptively and inferentially. A descriptive analysis is mainly on the summarisation process of numbers collected from the user study, such as mean, range and frequency. In an inferential analysis, the data is generalised to show the relationship between variables in the study and to help to make predictions. Some examples of statistical models for doing inferential analysis are chi-square, the t-test and the regression test. In credibility studies, statistical analysis has been used on both descriptive [Morris et al., 2012] and inferential analysis [Aladhadh et al., 2014]. Further discussion on the use of statistical analysis in previous studies is found in Section 2.3.2.

Another evaluation method focussing on quantitative data is the use of machine learning. Machine learning is using algorithms to learn from data without relying on rules-based programming, whereas statistical analysis uses mathematical equations to learn the relationships of variables in the data. Machine learning algorithms are mostly used for predictive modelling problems as the computer tries to find out patterns hidden in the data. In the study of credibility, machine learning has been used to develop a model based on credibility features [Castillo et al., 2011], predict the credibility of online contents [Kang et al., 2012], and rank the credibility of information [Gupta and Kumaraguru, 2012]. Section 2.3.3 discusses how machine learning is used as an evaluation tool in credibility studies.

2.3.1 User Studies

User studies are used to collect credibility judgements as ground truth, and explore features for credibility and study user characteristics and behaviours. The different types of user studies conducted in credibility have been discussed in Section 3. In most studies, respondents are asked to rate the credibility level of information and identify the features that influence them in forming their credibility perceptions. Demographic attributes and familiarity levels with information platforms are also collected in user studies. User studies also seek to obtain credibility ground truth for evaluation purposes [Castillo et al., 2011; Gupta and Kumaraguru, 2012; Sikdar et al., 2013b]. Most of these studies use crowdsourcing platforms to recruit respondents. Table 2.1 summarises the user studies that have been carried out by previous studies in the area of credibility research.

Table 2.1: *User studies conducted in credibility studies*

Literature	Participant	Description
Fogg et al. [2001]	Web users	Collecting demographics data, credibility rating, rating of source's trustworthiness, expertise, interface design.
Yang [2007]	Taiwanese population	Collecting demographics data, credibility rating of topic-related blogs, participants' characteristics, Internet use motivations and behaviour, agreement on credibility belief factors.
Rieh and Hilligoss [2008]	College students	Collecting participants' search habits, search goals, self-report credibility concerns on the web.
Castillo et al. [2011]	Amazon MTurk	Annotating credibility level and newsworthiness of a set of tweet messages for ground truth.
ODonovan et al. [2012]	Amazon MTurk	Annotating credibility level of tweets, collect demographic data, Twitter familiarity and features selection.
Gupta and Kumaraguru [2012]	Not available	Annotating credibility level of tweet messages for ground truth.
Morris et al. [2012]	University alumni and Microsoft employees	Collecting credibility rating of tweets, self-report features, Twitter habits, search habits, demographics data and self-report credibility concerns.
Kang et al. [2012]	Not available	Collecting participant's age, gender, self-report features and credibility rating of tweets (ground truth).
Yang et al. [2013]	Microsoft email list (China, USA)	Collecting participant's demographic data, microblog usage and credibility rating of social media posts.
Sikdar et al. [2013a]	Amazon MTurk	Collecting information credibility rating, newsworthiness level and authors' credibility.
Sikdar et al. [2013b]	Amazon MTurk	Annotating the credibility level of tweet messages for ground truth.
Kang et al. [2015]	Amazon MTurk	Collecting participant's demographic data, self-report Twitter features, Twitter activity rate, usage goal, visual cues, sharing frequency and credibility rating of tweets.

Lin et al. [2016]	University students and from social media	Collecting authors' credibility
Yin et al. [2018]	Social media users	Collecting factual, accurate, and credible annotation, demographic data, authors' credibility

As outlined in the table above and in Table 3.1, there is variation in how user studies are conducted taking into consideration the type of data collected. Most of the user studies show tweet messages along with their cues/features to participants to rate the credibility of messages, and identify feature importance. The messages can be real or manipulated tweets (e.g., user images, user names, etc.). The user study results could then be used to identify the features that have more influence on credibility perception. Results of the user study can be derived from both the statistical and machine learning approaches. This is shown in the studies by Castillo et al. [2011] and Gupta and Kumaraguru [2012], who use machine learning to analyse their user study data, and Kang et al. [2012], Morris et al. [2012], and Gupta and Kumaraguru [2012] (just to name a few), who analyse their user study data with statistical analysis.

2.3.2 Statistical Analysis

Most research on credibility has used various statistical analysis techniques such as descriptive and classical tests to prove their hypotheses and discover credibility factors, as part of their evaluation method. Some research uses both types of tests while others use only one type depending on the focus of the research. Mendoza et al. [2010] used percentage to examine the ability to identify retweets of true tweets and rumours. Thomson et al. [2012] used a

t-test and correlation test to examine the credibility of information on Twitter regarding the Fukushima disaster in regard to source credibility. Table 2.2 shows the list of studies on credibility that employ descriptive analysis in their work.

2.3.3   Machine Learning

Previous studies on credibility use different supervised learning algorithms to automatically classify or rank credibility. Table 2.3 lists previous studies that have used algorithms to engage in credibility classification/ranking accuracy.

Castillo et al. [2011] presented credibility classification with accuracy results of 86% using a J48 decision tree, shown to be better than any other random predictor. The researchers used supervised classifiers for news/chat classification and assessed credibility of 747 news topics. The focus of the prediction model was to detect news tweets that are almost true against other tweets. The features included in the classifier were the tweet message, the tweet authors, the tweet topics (collection of tweets), and the propagation of retweets. Their method was based on topic credibility rather than individual tweets. Crowdsourcing evaluation is used to determine the ground truth.

Kang et al. [2012] focused on individual tweets for their social credibility prediction model and achieved 88% accuracy using the J48 decision tree algorithm. One thousand and twenty-three manually annotated topic-specific (Libya) tweets were used. Three models were built based on source, content and hybrid features (combination of source and content features). Their best accuracy (88.17%) came from a credibility model using source features.

In another research, Gupta and Kumaraguru [2012] adopted supervised algorithms to

Table 2.2: *Statistical analysis used in credibility studies*

Literature	Statistical analysis	
	Descriptive analysis	Inferential analysis
Fogg et al. [2001]	Percentages, mean	Cronbach's alpha (inter-rater agreement), correlation analysis, factor analysis, t-test
Yang [2007]	Percentages, variance, mean, median	exploratory factor analysis, hierarchical regression analysis
Mendoza et al. [2010]	Percentages	Not available
Canini et al. [2011]	Not available	Correlation analysis, ANOVA, Cohen's Kappa (inter-rater agreement)
Thomson et al. [2012]	Percentages, median, standard deviation	Correlation, t-test
Morris et al. [2012]	Percentages	Chi-square test, ANOVA, t-test, correlation analysis
Gupta and Kumaraguru [2012]	Not available	Cronbach alpha, linear regression analysis
Westerman et al. [2012]	Percentages	Correlation analysis, linear contrast analysis
Kang et al. [2012]	Percentages	ANOVA
Sikdar et al. [2013a]	Not available	Correlation analysis, Cohen's Kappa (inter-rater agreement)
Sikdar et al. [2013b]	Not available	Correlation analysis, Cohen's Kappa (inter-rater agreement)
Yang et al. [2013]	Mean, percentages	ANOVA
Aladhadh et al. [2014]	Mean, percentages	Correlation analysis, t-test
Kang et al. [2015]	Mean, percentages	Correlation analysis, ANOVA

Table 2.3: Supervised learning for credibility prediction/analysis

Literature	Supervised Learning Algorithm	Description
Castillo et al. [2011]	J48 decision tree	To predict credibility levels on events shared on Twitter. Their algorithm achieves an accuracy of 86%
Kang et al. [2012]	J48 decision tree	Their social model achieves 88% of accuracy for automatically detecting credible tweets.
Gupta and Kumaraguru [2012]	Rank-SVM and PRF	To rank the tweets based on the credibility of information available in the tweet. Their credibility ranking model achieves an average of 73%.
AlRubaian et al. [2015]	Naive Bayes	Multistage credibility analysis in Twitter. The model achieved 90.3% accuracy, 86.2% precision and 98.8% recall.

rank tweets based on the credibility score for topic-specific (Libya) tweets. An SVM ranking

algorithm was used to rank tweets based on content and source features. Then, using Pseudo

Relevance Feedback (PRF), 34 frequent word unigrams from top tweets were extracted. Next,

the text similarity between the frequent unigrams and the top tweets was used to re-rank the

tweets. Annotators from an online study were shown news links of 14 news events to label 500

tweets per topic. Using Rank-SVM and PRF, Gupta and Kumaraguru [2012] reported a 73%

average NDCG score. Their results showed that an average of 30% tweets from their dataset

contained information, 17% of the informative tweets were credible, and 14% were spam.

The researchers further tested their prediction model on real-time tweets by developing a

Chrome extension tool called TweetCred.

AlRubaian et al. [2015] proposed a multi-stage credibility assessment model that uses

tweet and author features that include an author's relationship network. The dataset focused

on political topics regarding The Islamic State of Iraq and the Levant (ISIS) written in the Arabic language. The multi-stage credibility assessment model successfully classified credible tweets with 98.8% recall and 86.24% precision leading to only 13.76% false positives for the credible label.

2.4 Research Gaps

Most of the proposed methods for credibility evaluation are based on user credibility judgment ratings. Ground truth credibility values are typically gathered using human participants, who act as online content readers and evaluators. We believe that there are differences between readers in how credibility perception is disclosed. Attributes such as demographic profile, personality traits, and attitude toward the online content might impact on readers' perceptions of the information, which could contribute to a deeper understanding about how and why users carry out their credibility judgments. Most research has not paid sufficient attention to this issue.

Further, there was a gap in previous research incorporating user study results in identifying the credibility features importance and readers attributes. Our recommendation is to integrate the user study results with the content analysis results to help identify and correlate the credibility factors that influence credibility perceptions. Another recommendation is to identify how readers use the surface features of tweets discussed in Section 2.2, to make credibility judgements of tweets retrieved from a query action, and distinguish the effect of the reader's cultural background, behaviour and personality in the credibility perception of tweet messages and Twitter feature preferences.

2.5 Summary

In this chapter, we reviewed literature related to our work on credibility perception. We first distinguished the key terms in credibility. Next, the categorisation of credibility factors studied in previous work, approaches to credibility evaluation, and studies related to readers' attributes were described. Discussions in this chapter provided us with a selection of ideas in answering the research questions. The research gaps and research recommendations were also addressed. The coming chapters will include the literature discussed here and the approach in relation to the research questions that will address the research gaps.

Chapter 3

Research Design and Methodology

The present study intends to identify the features Twitter readers use to make credibility judgements, the correlation between a reader's credibility perception with other factors and readers' credibility perception behaviours. In order to answer each of the research questions, we designed three user studies using potential Twitter readers in order to gain representative information regarding the readers' credibility perception of information posted on Twitter. The chapter is organised as follows: approaches for analysing information credibility (Section 3.1), followed by details of the user study (Section 3.2), and finally, a summary of the methodology (Section 3.3).

3.1 Research Methodology

Methodology is a theoretical foundation that determines the appropriate research method in answering research hypotheses [Wahyuni, 2012]. The common research methods are quantitative, qualitative or mixed methods. Quantitative studies are used to determine the rela-

tionships between factors, while qualitative studies focus on developing new theory [Sogunro, 2002]. To determine the type of research method suitable, we need to confirm the aims of this study. In this study, we intend to explore the effect of already established theories and to identify the factors that have an effect on credibility perception. Based on this intention, the appropriate research method to adopt is the quantitative method. Furthermore, most of the previous literature on information credibility uses the quantitative rather than the qualitative approach when collecting credibility judgements [Castillo et al., 2011; Gupta and Kumaraguru, 2012; Kang et al., 2012; Morris et al., 2012; Rieh and Hilligoss, 2008; Sikdar et al., 2013b; Yang et al., 2013].

During a search process of information on Twitter, readers are shown lists of messages that are relevant to the query made. Readers are then required to make their own judgement regarding the credibility of the tweet messages. The reader's credibility perception can be captured as part of a user study. Analysis of interface features influencing a reader's credibility perception [Lucassen and Schraagen, 2011; Morris et al., 2012; Tseng and Fogg, 1999] and the reader's demographic impact on credibility perception was studied [Fogg et al., 2001; Kang et al., 2015; Morris et al., 2012; Song et al., 2016; Yang et al., 2013; Yang, 2007] to better understand readers' credibility perception behaviours [Lin et al., 2016; Yin et al., 2018].

A user study is conducted when a researcher wants to explore, describe or explain a particular phenomenon that mostly involves the participation of humans [Kelly et al., 2009]. In the area of credibility perception, user studies are conducted online. The type of setting used in the user study is determined by the question the research is trying to answer. Most

user studies are conducted in a natural or laboratory setting. In a natural setting, the researchers will observe the participants' behaviour in a normal routine activity environment. However, the researcher has little control over the setting. The laboratory setting is reverse to the natural setting. The user study is conducted in a controlled environment. Later, the user study setting is often used in identifying the effect of one or more variables. Nonetheless, there is a drawback for the laboratory setting. The participants' behaviour could be artificial, and does not represent real life. In our user study, the laboratory setting is used as we aimed to study the effect of variables on readers' credibility perceptions of information on Twitter.

3.2 User Study Design

The first research question examines how readers form a credibility perception of news-related tweet messages, the second question addresses the effect of demographic factors and topic familiarity, while the last question explores the influence of reader-related factors, such as personality and cognitive activities. Therefore, we employed a user study in answering all of these research questions.

3.2.1 Data Collection Methods

Data collection is a vital part of any research evaluation. Researchers often use a mixture of different methods to gather data. These methods include questionnaires, interviews, think-aloud, crowdsourcing and observation. Table 3.1 shows the mixture of data collection methods used by previous researchers in user studies on credibility perception.

Questionnaires are the most used method to collect data by researchers. Questionnaires

allow quick and direct capture of the participants' responses. Researchers can use closed

or open questions or a mixture of both, in the questionnaire at different stages in the

study [Kelly et al., 2009]. Closed questions are often used in quantitative data where

a specific set of answers is provided for the participant to choose. An open question is

just the opposite of closed questions where participants are given the opportunity to

express themselves freely. Open questions will produce qualitative data.

Interviews are used by researchers to get individualised responses from participants by

asking open questions. Interviews are also used to clarify the meanings of words or other

ambiguities such as the reasons behind a participant's attitude and behaviour [Kelly

et al., 2009]. However, analysing the information is not that easy. The information

gathered from the interview will need to be transcribed first before the analysis process.

Think-aloud data collection is a method involving user study participants thinking and

talking aloud while performing a task. The think-aloud method gives the researcher

an insight into what the participant is thinking, feeling, and finds interesting during

the user study process [Charters, 2003]. However, there are several difficulties of which

researchers need to be aware when conducting this type of data collection. The re-

searcher needs to be objective, take notes and all the while be aware of everything

that the users are saying, as the immediate thinking process changes rapidly as new

thoughts follow the one before.

Crowdsourcing is a request for large scale services, content, or ideas contribution from a

large group of people in an online community. The people are recruited from all over the

world using a web-based crowdsource platform, and they are paid for the work or task they do on the platform [Behrend et al., 2011]. In credibility studies, crowdsourcing has been used to collect credibility annotations and participant's demographics traits. A popular crowdsourcing platform for research purposes is Amazon Mechanical Turk and CrowdFlower.

Observation is a method that has the least interaction with participants among data collection methods. There are two ways to conduct the observation method, real-time or play-back time. In real-time observation, a researcher will sit close to the participants or follows them while all the time watching the participants perform the given activities. In the play-back time observation, a video camera or screen capture software will capture all the activities of the participants and the researchers will watch the video or image for analysis.

3.2.2   Three User Studies

In this study, we have three research questions:

- *RQ1: How do readers judge the credibility of tweets?*

- *RQ2: Does the credibility of tweets correlate with factors related to external attributes other than tweet surface features?*

- *RQ3: Do Twitter readers' personal characteristics play an important role in influencing credibility evaluation of information on Twitter?*

41

Table 3.1: *Data collection methods used in credibility studies*

Authors	Questionnaire	Interview	Think-aloud	Crowdsourcing
Fogg et al. [2001]	✓			
Yang [2007]	✓			
Rieh and Hilligoss [2008]	✓			
Castillo et al. [2011]	✓			✓
Gupta and Kumaraguru [2012]				✓
ODonovan et al. [2012]	✓			✓
Morris et al. [2012]	✓	✓	✓	
Kang et al. [2012]	✓			✓
Yang et al. [2013]	✓			
Sikdar et al. [2013b]				✓
Sikdar et al. [2013a]	✓			✓
Kang et al. [2015]	✓			✓
Lin et al. [2016]	✓			
Yin et al. [2018]	✓			

For each of these research questions, we have designed three user studies to cater for the objective of each question. Although there were three studies conducted, all of them have the same methodology. Readers are shown a set of tweet messages for them to annotate the credibility level, describe the credibility features that influence their credibility perception, and answer some questionnaires (e.g. demographic data, personality test).

In the first research question, it is important to identify how readers determine the credibility level of tweet messages. Based on the study by Castillo et al. [2011], Gupta and Kumaraguru [2012] and Morris et al. [2012], we have designed the user study in two parts: credibility annotation and readers' thoughts regarding the judgement. The credibility annotation part consists of the tweet and information regarding the tweet, such as the topic, topic description, the date the tweet was posted and the author's Twitter ID name. The topic and topic description is given to the readers in order to mimic the search activity, as a reader should have a general idea of what event they are searching for. The author's Twitter ID is given in order to give some identification of who wrote the tweet, as it was indicated by Morris et al. [2012] that readers are concerned with knowing the identity of the author. Meanwhile, the date posted shows the currency of the tweet and whether the tweet is posted within the time frame of the event. This is important as the tweet messages used in this study are related to news. News-related tweets are chosen based on the findings by Morris et al. [2012] that people are more concerned with the credibility of news-related tweets than any other topic.

As for collecting the readers' thoughts on how they determine the credibility level of a tweet, we conducted a pilot test in both an interview and a questionnaire setting. The

participants were divided into two groups. The participants in the interview group were asked what influenced them to make the credibility judgement of each tweet, while in the questionnaire group the same question was asked in writing and the participants needed to write down their answers. Both groups were shown the same tweet messages. We compared the answers and both user study settings showed the participants giving similar direct and short answers. Therefore, for the bigger sampling option, we chose to use the questionnaire setting. We did not give the participants a list of possible options because we wanted the answers to be raw and genuine. Details and analysis of the questionnaires will be discussed in Chapter 4.

For the second research question, the user study was divided into three parts, rather than two parts as in the first user study, as we also added a demographic questionnaire. However, there were changes made to the credibility annotation and readers' credibility features. In the credibility annotation section, instead of the text message, a screenshot of the tweet was shown, so that other features such as the number of likes, number of retweets and a picture of the author could be seen. This change occurred based on the findings in the first user study where some comments given by the readers were related to the said features. Also, due to this change, we made a list of those features from the findings and added more, based on the study by Castillo et al. [2011] to the readers' credibility features section.

We also encouraged the readers to leave comments if their credibility features were not part of the list. To weight the importance of each credibility feature, a four-rating scale was chosen from strongly agree to strongly disagree. The neutral scale was removed as it does not bring meaning to this user study. If a reader has identified a credibility feature as influencing

them to make a credibility judgement, they must be certain of the feature. Further details and analysis will be discussed in Chapter 5.

In the last research question, we wanted to identify readers' personal characteristics, including personality. Therefore, we added another section, personality test, to the user study. To ensure our user study was not exhausting and overwhelming for the readers, the short version of the personality test was chosen. Other than that, the credibility annotation section was changed to a seven-rating scale rather than the credibility level in the first two user studies. The reason for the change is based on the study by Westerman et al. [2012] and Sikdar et al. [2013a]. Having a rating scale of seven also gives wider options and better accuracy in readers' choices [Allen and Seaman, 2007].

We have also changed the way the tweets are shown to the readers. In this user study, we wanted to eliminate readers preconception of knowing the news beforehand. Therefore, thirty simulated news tweets regarding politics, breaking news, and natural disaster news, were shown to the readers. The simulated tweets resembled tweets returned by the Twitter search engine as results for a search with query keywords. Another justification of using simulated tweets was to control the features on the tweets. Details and analysis of the questionnaires will be discussed in Chapter 6.

3.2.3 Sample and Population

All the user studies in this study were conducted in a crowdsourcing environment. We chose crowdsourcing because the majority of work on credibility has used crowdsourcing as shown in Table 3.1. Furthermore, crowdsourcing allows researchers access to a higher number of

participants that are often more heterogeneous than can be accessed by doing an interview or user study in the lab as the participants come from various backgrounds in terms of location, nationality, education, employment, age, gender, language, etc. The cost involved in doing crowdsourcing is relatively low compared to traditional methods [Zuccon et al., 2013].

Moreover, since the participants are from the crowdsourcing platform, it is time-efficient and there is flexibility with time of day, rather than having to find participants, therefore, the user study can be completed much faster than an online study or a traditional user study in the lab [Gadiraju et al., 2017]. Although there are limitations on the use of crowdsourcing, there are ways to minimise the risk of false data collection [Zuccon et al., 2013].

Another rationale for choosing crowdsourcing for our sample population is that users are Internet savvy and have regular interaction with the web and information technologies, and therefore, due to this, they are able to perform different types of tasks on the crowdsourcing platform. Furthermore, having a good sample size and with an appropriate statistical test, a researcher is able to significantly accept or refute a hypothesis. Larger sample sizes also tend to give more reliable findings.

3.3 Summary

In this chapter, we presented methods used by past researchers when looking at the field of credibility. Details of the methodology used in investigating information credibility and credibility perception were provided. In particular, the methodology was designed with the aims to analyse readers' credibility perceptions of online information as well as determine factors affecting readers' credibility perceptions. Multiple sources of data were used in ensuring the

validity of results obtained and the reliability of conclusions made.

Chapter 4

Tweet Features and Credibility

Association Analysis

4.1 Introduction

Twitter readers are concerned with the credibility of tweets relating to breaking news, politics, and disasters [Morris et al., 2012]. In a scenario where an event has occurred and has been reported by news agencies, Twitter is a go-to media platform to get more information [Hu et al., 2012]. Twitter authors not only share news headlines from newswires, but also report real-time events before they reach the press [Kwak et al., 2010]. The readers also use Twitter to get news updates, especially if the event is disaster or crisis-related [Hughes et al., 2008]. News on Twitter comes from a wide variety of sources: some comes from well-known news organisations and government departments, while most comes from members of the public. Consequently, Twitter readers often make their own judgements of the credibility of tweets.

In credibility perception, determining the method readers use to help them make decisions regarding the credibility level of online content is achieved through user studies. We used the same approach in this research. This chapter describes how Twitter readers perceive the credibility of tweet messages and identifies the method readers apply in the credibility perception process. Section 4.2 will discuss the data collection process of using user studies on a crowdsourcing platform, a similar method that is used in the literature discussed in Chapter 3. In the next section (Section 4.3), we will explain how we analyse the data. Section 4.4 describes the analysis process. Findings of the user study are explained in Section 4.5 and the overall discussion is reviewed in Section 4.6. The last section, Section 4.7, will summarise the chapter.

4.2 Data Collection

A user study was conducted on a crowdsourcing platform following other studies on credibility perception [Castillo et al., 2011; Gupta and Kumaraguru, 2012; Kang et al., 2012]. The crowdsourcing platform chosen for this user study is CrowdFlower as this platform allows international users to access the platform, as compared to other platforms like Amazon Mechanical Turk that only allow users with a United States bank account. News event topics, and their relevant tweets are selected for the study, as the study by Morris et al. [2012] shows that users are very much concerned about the credibility of news and crisis events as reported online.

Crowdsourced evaluators are recruited through the CrowdFlower platform to judge the credibility level of tweets, and to describe how they make their credibility judgements.

Through their comments, features are extracted and the predictive association rule analysis is applied to establish associations between features and credibility levels. According to Castillo et al. [2011], the criteria used to determine credible tweets are: the tweet must affirm a fact, be informative for the public, not be self-opinionated, and not be a chat between friends.

Twenty news-related topics reported on major online newswires including BBC, Reuters, CNN, Guardian, and The New York Times, were selected for the experiment. The news events occurred between 1 June 2013 and 15 October 2013. Table 4.1 describes the 20 topics selected. Tweets were collected from the Twitter API tweet search using the topic shown in the left column of Table 4.1. To ensure that redundant tweets are excluded from the tweet set, a manual search was conducted. In total, 400 tweets in English that were manually checked for their credibility for 20 news events were presented to crowdsourced evaluators. The screenshot of the form presented to the evaluators is shown in Figure 4.1.

To ensure the quality of the credibility judgement from the crowdsourced evaluators, other than filtering the crowdsourced valuators by their level (based on number of tasks completed and a good percentage rate for the tasks), the qualification test can also be given to the evaluators. The evaluators need to pass the qualification test before they can continue with the user study. We can conduct the qualification test using gold questions [Behrend et al., 2011; Zuccon et al., 2013]. In this user study, the gold questions consisted of some tweets that did not follow the credibility criteria mentioned by Castillo et al. [2011]. For each of the 20 topics, two gold questions were randomly inserted into the tweet collection. Only evaluators that judged the gold questions correctly and scored a percentage of 80 percent or

Table 4.1: News event-related topics

| Topic	News event description
US Government Shutdown	US Government heads toward a shutdown
Iran-US Relationship	Iranian President takes steps to thaw relationships with the West
Sarin attack in Syria confirmed	United Nations confirms use of chemical weapons in Syria
Shipwreck in Europe	Boat sinks in the Mediterranean, killing dozens
Egypt state of emergency	Egypt declares state of emergency
Train kills dozens in India	Train kills dozens of religious pilgrims in India
Navy Yard shooting	Gunman and 12 victims killed in Washington D.C. Navy Yard shooting
Earthquake in Pakistan	Magnitude 7.7 earthquake kills at least 327 in Pakistan
Terrorist attack mall	Somalian militants terrorise luxury mall
Military ousted president	President Morsi deposed by military after one year in office
NSA whistle blower	Edward Snowden: whistle-blower behind NSA surveillance revelations
UK new prince	The Duchess of Cambridge gives birth to a baby boy
Oil train derails	A train in Quebec derails and explodes
Colorado flood	Colorado flood tragedy
Australia's new prime minister	Australia's new Prime Minister Tony Abbott
Iraq suicide attacks	Suicide bomb attacks on Iraqi school, Sh"ite pilgrims, kill 29
Mexico storm disaster	Mexico storms death toll rises, crop lands damaged
Cyclone hits India	Many evacuated as powerful cyclone hits India
Protest in Egypt	More than 50 people killed at pro-Morsi protest
Riot in Moscow	Rioting erupts in Moscow after killing blamed on migrant
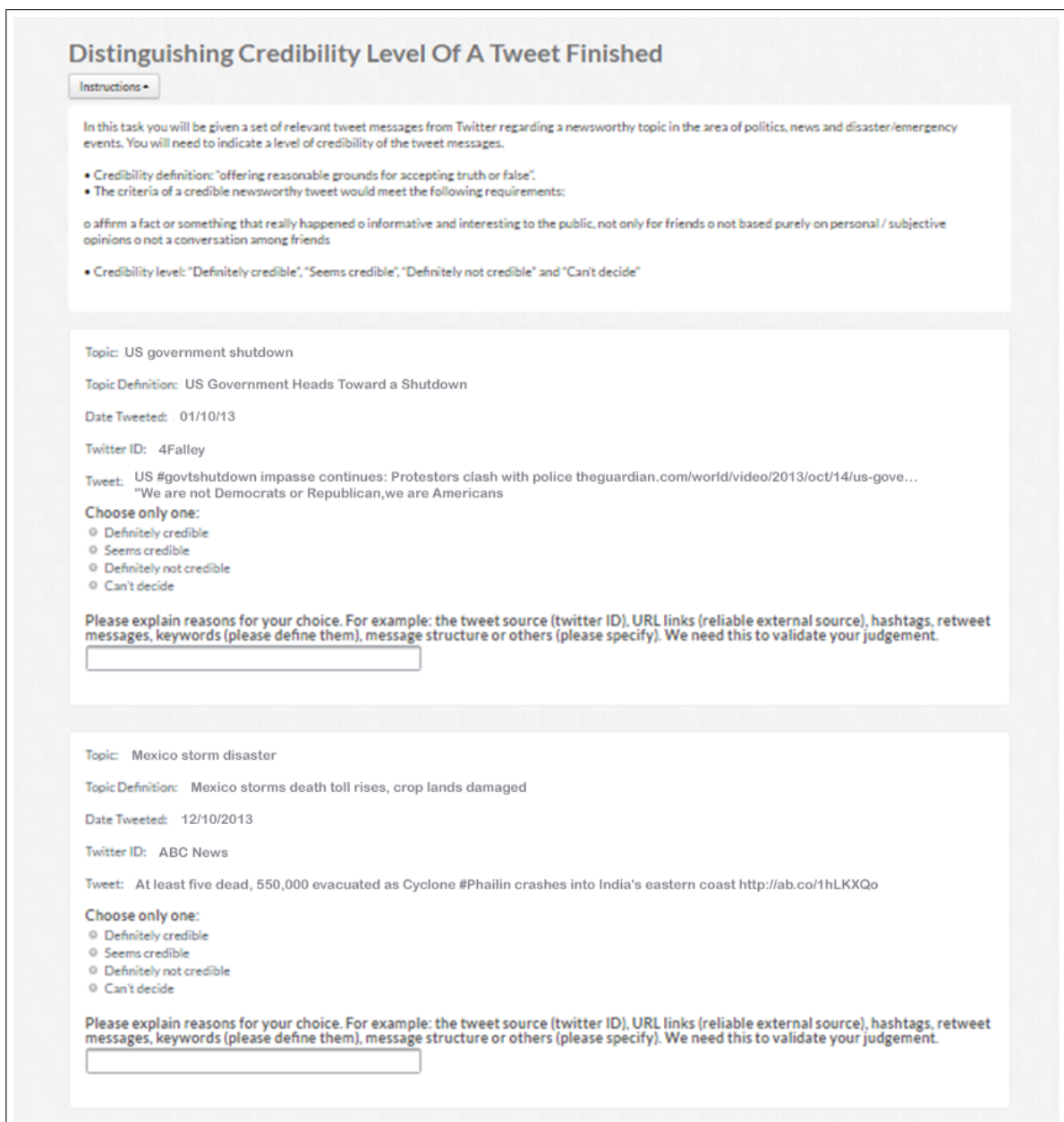
Figure 4.1: User study screenshot to find readers' feature preferences

higher were considered reliable thus their credibility judgements of the tweets are accepted.

While this approach may provide an indication of whether the crowdsourced evaluator did

the task properly and not just provided random answers, the truthfulness of the answers

could not be ascertained.

In this research, apart from the credibility judgement of news related tweets, the method Twitter readers use to judge the credibility level of tweets is studied as well. To retrieve this information, CrowdFlower evaluators were requested to leave textual comments to explain their judgements and describe how they arrived at that decision. The evaluators are known as general Twitter readers for the remainder of this thesis. The comments were manually examined to ensure quality comments were used to analyse the readers' perceptions. Nonsensical comments, such as those containing the word "none" and numbers or words that are out of context for the topic, were removed.

This user study was approved on behalf of RMIT University's College Human Ethics Advisory Network (CHEAN) by a delegated CHEAN committee (Approval number: ASEHAPP 47-13), refer to Appendix B.1 for the approval letter.

4.3 Evaluation Approaches

In this section, we will describe the evaluation analysis applied to the data collected in this user study. Two types of analyses were conducted: the extraction of features from the comments left by the user study participants, and the associations derived from the features.

4.3.1 Content Analysis

Content analysis is a flexible process of analysing text data from open-ended questions using a qualitative research technique. This method helps to interpret the content of text using classified codes and theme identification [Hsieh, 2005]. The normal approach is using la-

tent content analysis, a method for finding hidden information by making inferences in a systematic and objective way from other available materials [Krippendorff, 2004].

There are three distinct methods for content analysis: conventional, directive and summative [Hsieh, 2005]. Summative content analysis analyses content based on keywords. The approach is different from conventional analysis, which starts with observing the data to find the code and the theme. Directive analysis is more focused on theoretical analysis. In this experiment, summative content analysis is the most suitable due to the nature of the open-ended questions, which were allotted one short sentence.

4.3.2   Association Rule Mining

Association rule mining is a rule-based machine learning method that is used in data science for discovering relations between variables in a large database. Association rule mining is perfect for categorical data and aims to extract interesting associations that satisfy predefined minimum support and confidence from sets of items in transaction databases [Agrawal et al., 1993; Tan et al., 2005]. Association rules are used in areas such as retail, marketing, and inventory management. An association rule is written as X$\rightarrow$Y where X, Y $\subset$ I are sets of items called item sets, and X $\cap$ Y $= \emptyset$. X is called antecedent while Y is called consequent. The rule, X$\rightarrow$Y, means X implies Y.

Two measures for association rules are support and confidence. Users predefine thresholds of support and confidence. The threshold is meant to drop rules that are not so interesting or useful. The two thresholds are called minimal support and minimal confidence. Support of an association rule is defined as the fraction of the total number of transactions containing

all items in $X \cup Y$ to the total number of transactions in the database.

$$Support(X \rightarrow Y) = \frac{\# \text{ of transaction containing } (X \cup Y)}{\text{overall transaction (N)}} \tag{4.1}$$

The confidence of an association rule is defined as the measure of the total number of transactions that contain both $X \cup Y$ to the total number of transactions containing X.

$$Confidence(X \rightarrow Y) = \frac{\# \text{ of transaction containing } (X \cup Y)}{\# \text{ of transaction containing } (X)} \tag{4.2}$$

Another metric that has been proposed by Brin et al. [1997] regarding identifying the interesting rules called lift is also used in this study. In general, lift is the ratio between the confidence and the support of the item set in the rule consequent. Specifically, lift is the ratio of the observed support of $X \cup Y$ to that expected if X and Y are independent.

$$Lift = \frac{\text{support } (X \cup Y)}{\text{support } (X) \times \text{support } (Y)} \tag{4.3}$$

A lift value greater than 1 implies that the degree of association between the antecedent and consequent item sets is higher than when the antecedent and consequent item sets are independent.

Association rule mining in this study was administered to find interesting rules that describe the relationship between the readers' demographics and news topics, and the credibility level of news tweets perceived by readers. In this study, the lift metric was applied to determine the rule of interest from the association rule mining.

4.4 Data Analysis

At the conclusion of the user study, a total of 2005 judgements by 98 readers for 400 tweets were collected, where five out of 400 tweets received six judgements and the rest received five judgements each. Due to the nature of the user study being a mixed method survey, combining both quantitative (credibility rating) and qualitative (textual comments) methods, the data analysis was conducted on two levels. First, each comment from the open-ended questions regarding the features readers use to help them decide the credibility level of the tweets was analysed and compiled into categories. The categories were based on the features described in previous studies. The second level saw the categorised features being examined according to the credibility level perceived by the readers. Association patterns between reader's credibility level judgement and the features that they found helpful were identified next.

4.4.1 Deriving Features using Content Analysis

The analysis began with collecting and searching for the occurrences of keywords from the comments left by readers. Overall, 609 noise-free comments were collected. From the 609 comments, 405 comments that states only direct features found on Twitter were calculated using computer-assisted word frequency count. The searches for occurrences of the identified features were based on the features described by Castillo et al. [2011]. Counting was also used by Hsieh [2005] to discover patterns in the data. The features are shown in Table 4.2.

Next, the 204 comments that do not state only direct features were manually checked. In the first step, we looked for comments related to the source, either about the author

Table 4.2: Tweet direct feature occurrences commented on by readers

Features	Number of occurrences
Keywords	341
Link	53
Hashtag	5
Retweet	6

or external sources (i.e. URL links). We then looked for words or descriptions related to the special features on Twitter, that is user mentions and hashtags. Next, other comment descriptions were analysed based on features reported in the previous studies by Castillo et al. [2011] and Gupta and Kumaraguru [2012]. The concept mapping approach [Jackson and Trochim, 2002] for summative content analysis was adopted to group feature keywords. The keyword feature groupings were then shown to three annotators to be evaluated. To ensure a quality feature grouping, only keywords that had groupings agreed to by three annotators were used.

From the analysis conducted, two other features distinct from the ones listed by other researchers were identified: user belief and credibility keyword. User belief refers to the reader's prior belief regarding the relevant topic and is external to Twitter, while in the study by Castillo et al. [2011], all features were derived based on Twitter. Sentences such as "plausible" or "it happened" are examples of comments grouped under user belief. The informative feature was also unanimously agreed upon by the three researchers as part of the user belief feature.

Comments that focus on breaking news and information updates were grouped under the credibility keyword. Due to the finding of this feature, the *keywords* feature was re-

examined by analysing each tweet that reports the *keywords* feature with other studies in the literature [Gupta et al., 2012; Morris et al., 2012]. While this may not be the best method, and it would be nice to be able to pose post-task questions to the readers for further inquiry regarding the comments they gave, the crowdsourcing platform does not allow for this. From the re-examination, it is found that the *keywords* feature can be further categorised into "topic" and "credibility" keywords. Overall, there are eight features that were extracted from the comments reported by the readers. The eight features were then summarised into three categories that best describe the features as shown in Table 4.3.

Table 4.3: Features derived from reader comments for credibility rating

Category	Feature	Number of comments
Topic-based	Topic keyword - *e.g. Prince (UK new prince topic)*	325 (53.4%)
Message-based	Link in tweet - *URLs, URL shortener, image links*	95 (15.6%)
User-based	Display name - *Twitter ID e.g. BBCNews, Anonymous*	88 (14.4%)
User-based	User belief of the topic - *e.g. plausible, professional, it actually happened, facts, informative*	52 (8.5%)
Message-based	Credibility keyword - *e.g. Update, Breaking, Liveupdates*	26 (4.3%)
Message-based	Hashtag - *e.g. #Lampedusa, #Egypt*	11 (1.8%)
Message-based	Retweet - *Contains the letters 'RT in the tweet messages*	8 (1.3%)
User-based	User mention - *e.g. @OMBPress, @cctvnewsafrica*	4 (0.7%)

4.4.2 Analysing Features for Credibility

In the user study, readers were asked to judge the credibility level for each tweet as "Definitely credible", "Seems credible", "Not credible", or "Can't decide". To determine the credibility level for each tweet, a consensus rule was used. If a tweet receives three out of five or four out of six votes for a credibility level, the message is assigned the corresponding credibility rating; otherwise no consensus credibility rating (recall that there are three credibility levels) can be reached for the tweet.

Table 4.4 lists the distribution of credibility ratings for all tweets. Note that none of the tweets received the judgement of "Can't decide". The results confirm that readers generally trust the information disseminated on Twitter, which mirrors the findings in the study by Castillo et al. [2011].

Table 4.4: Distribution of credibility ratings for 400 tweets

Credibility Level	Number of comments
Definitely credible	342 (85.5%)
Seems credible	2 (0.5%)
Not credible	35 (8.75%)
Can't decide	0 (0%)
No consensus rating	21 (5.25%)

4.5 Result

The objective of this experiment is to identify the tweet features readers reported to use when perceiving the credibility level of news-related tweets. The results will look at the features in regard to tweets that have been perceived as credible, misjudged, or difficult to be judged,

and also the association between features and credibility perception.

4.5.1 Extracting Features for Credible Tweets

The reader comments for 342 that tweets received "Definitely credible" and two tweets that received as "Seems credible" ratings were analysed together with their derived features. Table 4.5 shows that readers perceived these features in general with significantly different weights, where "Topic keyword" was commonly used and "User mention" was rarely used. In contrast, the carefully engineered tens of features in the study by Castillo et al. [2011] were used collectively by machine learning models in predicting topic credibility.

Table 4.5: Features derived from reader comments for credible tweets

Feature	Number of comments
Topic keyword	315 (54%)
Link in tweet	95 (16.3%)
Display name	88 (15%)
User belief of the topic	44 (7.5%)
Credibility keyword	26 (4.5%)
Hashtag	8 (1.4%)
Retweet	6 (1%)
User mention	2 (0.3%)

4.5.2 Analysing Misjudged and Difficult-to-Judge Tweets

The 35 tweets with the "Not credible" rating in Table 4.4 were analysed next. These tweets were misjudged by readers, as all tweets in the study were manually verified as credible. The politics news topics 'Iran and US relationship' and 'US Government shutdown' had the largest number of misjudged tweets. The tweets from both topics were often statements that consisted of a question rather than statements, or titles for news articles from reliable

news agencies with short URL links, which may be why readers had a misperception of their credibility. Although having a link in a tweet is an important feature for credibility perception (see Table 4.6), the language features of tweets also play an important role for a reader's perception of credibility.

Twenty-one difficult-to-judge tweets, tweets that readers could not reach consensus on, were also analysed. Ninety-six percent of these difficult tweets were made up of breaking news (42.8%) and politics news (42.8%). In further observation of these tweets, it was found that the tweets were mostly lacking links to external sources. The analysis is consistent with the high association between having a link in a tweet and the tweets credibility level shown in Table 4.6.

Table 4.6: Top association rules

Credibility Level	Number of comments
Link in Tweet = available 74 => Credible 72	97.7%
Hashtag = yes 8 => Credible 8	97.6%
Retweets = yes 6 => Credible 6	97.6%
Twitter display name = yes, User belief = yes 3 => Credible 3	96.2%
Twitter display name = yes 88 => Credible 81	91.0%
User belief = yes, Topic keyword = yes 36 => Credible 27	77.4%
User belief = yes 44 => Credible 33	76.7%

4.5.3 Feature and Credibility Association Analysis

To uncover the relationships between features and tweet credibility, association rule mining was applied to the 379 tweets in Table 4.4 with consensus ratings of "Definitely credible", "Seems credible" and "Not credible" based on the features in Table 4.3. Using the WEKA Predictive Apriori package [Scheffer, 2001] the best 100 association rules were mined of the

form "feature set => credibility" with an accuracy threshold of 70%. Table 4.6 lists the top association rules, where numbers of comments supporting the left and right-hand sides are shown. According to the table, for all top rules the right-hand side is always the "Credible" rating readers tend to believe in the information conveyed in tweets yet do not consistently give the "Not credible" rating. Moreover, "Link in Tweet", "Twitter display name" and "User belief" are important features that often lead readers to give a "Credible" rating for tweets. From Tables 4.3 and 4.6, it can be seen that the "Topic keyword" feature, the most important feature commented on by readers, does not form a strong association rule; only when combined with "User belief" it give high accuracy in predicting credible tweets. The "Link in Tweet" feature is used as an indicator of credibility.

4.6 Discussion

Readers describe a direct approach based on the first impression in the choice of features they use to decide the credibility level of tweets. In previous work, when readers were asked about their general approach and features they use to help them decide the credibility level of tweets, the readers listed features they plainly saw on the tweet. None of the readers described looking deeper into the credibility of the author or the news topic. This behaviour or these preferences were also reported by Morris et al. [2012], where they had to prompt their participants to click on the URL links provided on the tweet or to click on the author's name to get into the author's profile page.

We studied the user perceptions of credibility for news tweets on Twitter via a user study on the CrowdFlower platform. By analysing user credibility judgements and comments, eight

features were identified, where "Twitter display name", "Link in tweet" and "User belief" in the tweet topic are most important. "User belief" is also described by Rieh and Hilligoss [2008] as part of the credibility features used to judge the credibility of Wikipedia. In the credibility association analysis, we found strong associations between features and tweet credibility using one of the machine learning approaches, association rule mining, varying from other studies in the literature. We further found that politics and breaking news are more difficult for users to consistently provide credibility ratings.

4.7 Summary

In this chapter, the readers' perception of credibility of news tweets on Twitter were analysed via a user study on the CrowdFlower platform. By analysing reader's credibility judgements and comments, eight features were identified: display name, link in tweet and user belief in the tweet topic were found to be the most important. By feature and credibility association analysis, strong associations were found between features and tweet credibility.

To summarise:

- Readers reported the use of eight features in the process of perceiving the credibility level of news-related tweets.

- User belief, an external feature was found to be one of the most important features in judging the credibility level of tweets.

- Features reference was highly associated with the perception of the tweet message credibility level.

Chapter 5

Factors in Credibility Perception

5.1 Introduction

This chapter examines the relationship between readers' demographics, news topics and tweet features with readers' credibility perceptions, and further examines the correlation among these factors. The correlation between readers' demographic attributes, news topics and features, and readers' credibility judgements is examined. We also compare the credibility prediction tool with readers' credibility perceptions of tweets, to identify the similarities or differences between the two.

In the following sections, we describe our data collection and data analysis methodologies (Section 5.2). Similar to the previous study, we designed a user study on the crowdsourcing platform, CrowdFlower, to collect reader's demographic data and credibility judgements of tweets. We will focus only on tweet surface features. Section 5.3 discusses the evaluation approach, while the findings of this study will be discussed in Section 5.4. A discussion regarding the findings will be given in Section 5.5, and Section 5.6, summarises this chapter.

5.2 Data Collection

This section discusses the process of data collection. Continuing on from the previous study, there are two elements of credibility collected in this study, humans' credibility perceptions and machine credibility prediction ratings. These are collected in order to measure whether there is a difference between the two since the features valued in the two methodologies differ. In the previous study, we found that readers mostly focus on the tweet surface features while the machine prediction model looked at more than just surface features. The credibility prediction model also focussed on metadata, propagation networks and other behind-the-scene information [Castillo et al., 2011; Gupta and Kumaraguru, 2012]. A simpler way to make the comparison is comparing the credibility level of the same tweet messages judged by readers and predicted by the TweetCred, real-time credibility prediction Chrome extension tool developed by Gupta et al. [2014] based on their earlier research [Gupta and Kumaraguru, 2012; Gupta et al., 2012].

Other than that, we also collected the demographic data of our readers in order to answer our second research question. The first subsection will explain the method we used to choose the tweets to be used in the user study. The second and third subsections explain the method we used to collecting credibility ratings from both the credibility prediction tool and readers, as well as readers' demographic data.

5.2.1 Tweet Message Collection

We compiled tweets from three news categories: breaking news, political news, and natural disaster news, the same categories used in past studies [Morris et al., 2012; Yang et al.,

2013] and in our previous user study. Each news category consisted of five world news topics reported by news agencies including BBC, Reuters and CNN from 2011 until May 2014. A list of the 15 news topics used in this study can be found in Appendix A. Each news topic must have had more than 30 unique tweets retrieved from the search result to ensure that the topic is fairly well-known by people and not an isolated news topic. The tweets could be retweeted as long as they were not a retweet of the same message within the collection. This is to ensure that the user study participants did not see a repetitive tweet message. Overall, 1510 tweets were added to the collection.

The news topics were evenly divided between trending and not trending topics. Trends were determined from the trending list on Twitter and "What the Trend", a Twitter page that is part of the HootSuite Media, which lists Twitter's trending topics. The tweets were manually examined to ensure they were topic related tweets and randomly picked from the topic search results by Twitter. We also included some known rumour tweets from the same news topic as reported in snopes.com. In the news tweet collection, two writing styles of tweets were included - a style expressing the author's opinion or emotion towards the topic and another reporting factual information. These writing styles were used after results from a pilot user study indicated that readers also find tweets expressing an author's feelings on a topic as credible.

5.2.2 TweetCred Credibility Rating

TweetCred is a Chrome extension tool that gives a real-time credibility score about a tweet based on more than 45 features including meta-data, content-based simple lexical features,

content-based linguistic features, author, external link URL's reputation, author network, etc. The credibility rating predicted by TweetCred is displayed next to the author's display name or beside the date. Figure 5.1 shows the credibility rating of five out of seven on the credibility scale predicted by TweetCred.

Figure 5.1: Credibility rating of a tweet predicted by TweetCred

5.2.3 Readers' Credibility Rating

This study aimed for a global participation therefore, a crowdsourcing platform was used to recruit participants. Conducting online surveys on the crowdsourcing platform allowed us to get a large number of international participants within a short time at a lower cost than a traditional survey. Furthermore, the nature of Twitter that allows anyone (both account and non-account holder) to search for tweet messages and view them made it possible for us to conduct the survey on the crowdsourcing platform as anyone can be a Twitter reader

through a search request. We conducted a user study of 1510 tweets on 60 news topics on breaking news, natural disaster and political news that were judged by 754 participants.

The online survey was divided into two parts. The first part regarded the basic demographic questions: gender, age and education level. The screenshot for the user study's demographic questions is shown in Figure 5.3. Country information was supplied by the crowdsourcing platform as part of the crowdsourcing worker's information upon registration. The workers from here on were regarded as tweet readers searching for information as shown in Figure 5.2.
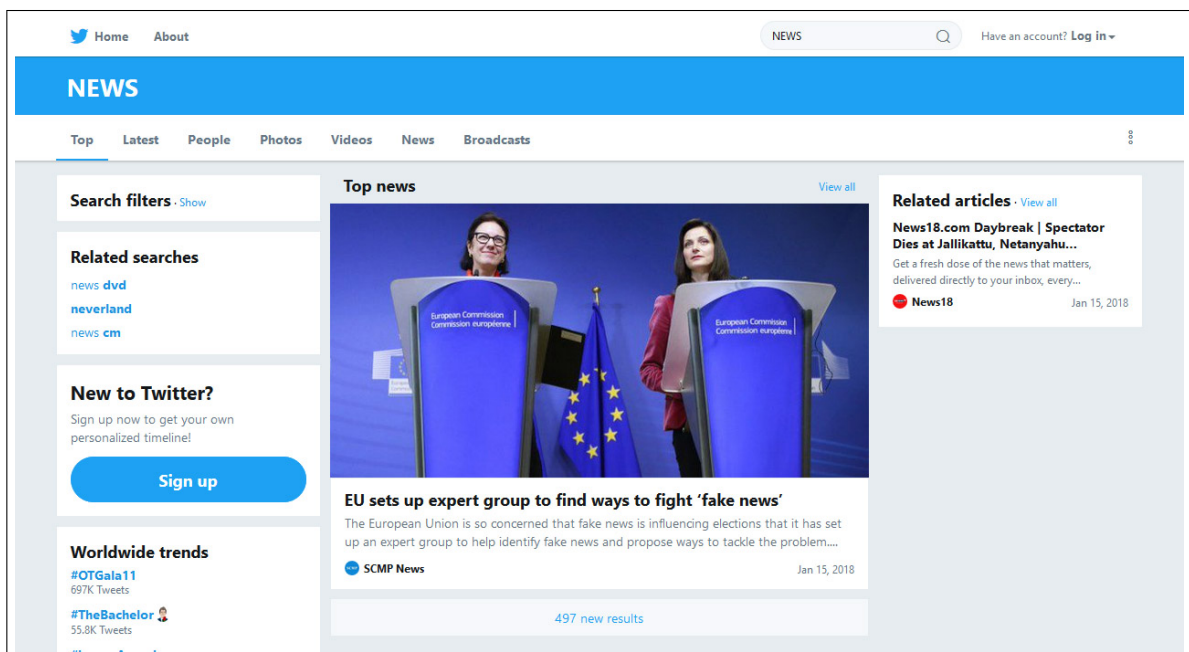

Figure 5.2: A screenshot of using Twitter's search platform

The second part of the questionnaire was on credibility perception. A number of pilot studies have been conducted to determine the optimal number of tweet judgements the readers were willing to make. Twelve judgements per reader were chosen, and we set a total

of seven judgements per tweet by different readers. The tweets were shown to the readers as they would be shown in a Twitter search result page, retrieved in response to a search topic. Readers were also shown the topic and topic description so that they would have some knowledge regarding the topic, the same as they would have if they were the ones doing the search. The readers were asked to rate the perceived credibility level of the tweet. Four levels were listed: very credible, somewhat credible, not credible and cannot decide, which is based on the studies by Castillo et al. [2011] and Gupta and Kumaraguru [2012].

If the readers judged a tweet as having a positive credibility level, the 'very credible' or 'somewhat credible' levels, we prompted the readers with a list of features reported in previous research by Castillo et al. [2011], and in our previous study, as shown in Table 4.3. Figure 5.4 is a screenshot of the design layout. The readers were also encouraged to describe other features in a free text interface. If readers judged a tweet as having a negative credibility level, 'not credible' or 'cannot decide', the readers were asked to justify their credibility judgement. The two different methods were chosen based on the outcome of our pilot study where free text was found to provide more insight regarding the way readers make a negative credibility judgement. This user study was approved by RMIT University's College Human Ethics Advisory Network (CHEAN) by a delegated CHEAN committee (Approval number: ASEHAPP 47-13), refer to Appendix B.1 for the approval letter.

As the user interfaces for positive and negative judgements were different, we must ensure that judgements are not biased towards the positive credibility level, as it is an option design, while the negative credibility level requires the readers to type their answer. A total of 227 readers chose at least once the negative credibility level within the set of 12 random tweets

Figure 5.3: Tweet message and demographic questions

Figure 5.4: Credibility perception and feature selection

shown to them. The majority of readers chose at least one negative credibility judgement. The highest number of negative credibility judgements was three. From our manual analysis, readers who perceived tweets as 'not credible' or 'cannot decide' were seen to be reliable with their judgements, since the negative credibility judgements were chosen no matter the tweet display order. The study also found that readers still rate a tweet as either 'not credible' or 'cannot decide' a second time even after they know the nature of the negative credibility judgement design. Therefore, we concluded that the user study's distinct design would not incur any bias on the credibility judgements.

To ensure the quality of answers by readers, a set of gold questions were shown to the readers, a common technique on crowdsourcing [Zuccon et al., 2013]. The readers were required to answer the gold questions at a minimum 80% qualifying level before they were allowed to progress with the user study. The gold questions were standard awareness questions, e.g. determining whether a topic and a tweet message were about the same news topic. The gold questions were not counted as part of the user study.

Lastly, the credibility level given by readers and by an automated credibility prediction tool were compared. Since seven readers judged each tweet, the credibility judgements were aggregated, and a consensual vote determined the credibility level. Tweets that did not have a consensus judgement were discarded from the list. However, at the time the TweetCred real-time credibility scores were collected, there were some tweets from the final readers' credibility perception dataset that were no longer available on Twitter. The tweets were inaccessible due to them being deleted by the author or by Twitter for certain reasons such as privacy, sensitivity and legality. Tweets that were no longer available on Twitter had to

be discarded from the comparison list.

5.3 Data Analysis Approaches

This section explains the analysis approaches that we performed on our data collection in order to find the answer to our second research question. The first is a statistical analysis named Cohen's Kappa, an inter-rater agreement for categorical data. This analysis is used to find agreement between the human's credibility perception and the machine credibility ranking prediction tool. As our main aim in this user study was to find the correlation between demographic data and credibility perception, news category and credibility features, the second statistical analysis applied was chi-square test of independence. Continuing on from the previous user study, we analysed the association between the various variables to find interesting patterns in the data using association rule mining (explained in Section 4.3.2).

5.3.1 Cohen's Kappa

Cohen's Kappa statistical analysis is used to find agreement between two independent observers rating the same set of things. This is different from Fleiss Kappa which is used to find agreement between multiple raters and Krippendorff's alpha which is useful when there were multiple raters and multiple possible ratings [McHugh, 2012]. In this study, we only had two raters, the readers and the credibility prediction tool. Therefore, Cohen's Kappa was the best inter-rater agreement analysis for our study.

In Cohen's Kappa, if the two raters randomly assign their ratings, there is a chance that their ratings would sometimes agree with one another. The Kappa's calculation is based on

the difference between the observed agreement ratings compared to the expected agreement ratings by chance alone. Kappa's outcome result is standardised to be between the -1 and 1 scale. One is perfect agreement, 0 is exactly as expected by chance, and negative values indicate agreement less than chance. Cohen's Kappa was used to determine the difference of agreement between the credibility rating predicted by TweetCred and readers' credibility level majority vote of the same tweets. The tool scored the credibility level of a tweet using a 7-scale rating. The 7-scale rating was also categorised into three categories of credibility level: scores 1 and 2 are low, scores 3 - 5 are medium, and scores 6 and 7 are high [Gupta et al., 2014]. Then, the tool's score was categorised into the same credibility level used in the study where low is set as not credible, medium as somewhat credible and high as very credible. Afterwards, the agreement matrix table was built to perform the Cohen's Kappa analysis. The result from the statistical analysis was then compared to the Cohen's Kappa agreement interpretation range described by Viera and Garrett [2005] to identify the agreement level between the two.

5.3.2 Chi-Square Test of Independence

The chi-square test of independence calculated the difference between observed data counts and expected data counts. The cut-off acceptance for the relationship was based on the accepted probability value (p-value) of 0.05. The chi-square statistic test can be calculated as follows, where O_i and E_i are the observed value and expected value for cell i of the contingency table:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \tag{5.1}$$

In this study, in addition to the correlation analysis regarding a single demographic attribute and credibility judgements, how combinations of demographic attributes correlate with credibility judgements was also analysed. Therefore, multi-way chi-square tests were also performed. Let V_1, ..., and V_k be k binary variables, the contingency table to calculate the χ^2 for these k binary variables is $(V_1, \bar{V}_1) \times (V_2, \bar{V}_2) \times ... \times (V_k, \bar{V}_k)$. For example, when there are three binary variables A, B and C, to find out if variables A and B were correlated with variable C, the χ^2-statistic would be $\chi^2(\text{ABC}) + \chi^2(\text{AB}\bar{\text{C}})$ [Brin et al., 1997]. Note that the chi-square statistic is upward-closed this means that the χ^2 value of ABC would always be greater than the χ^2 value of AB. Therefore, if AB is correlated, adding in variable C, ABC must also be correlated. Refer to Brin et al. [1997] for proof of the theorem.

In the study's problem setting, the theorem to prevent false discoveries for multi-way chi-square analysis was applied. Assuming that A and B were independent variables for demographic attributes and C is the dependent variable for credibility levels, if A and B were correlated, even if A, B, and C were correlated, the study would not be able to tell if the association between the credibility level and the demographic attributes is due to an actual effect or to the non-independence of observations. To solve this issue, chi-square analysis was first applied between individual demographic attributes and the credibility judgements. If the result is insignificant, multi-way correlation analysis for the combination of demographic attributes will be applied. To this end, the correlation for pairwise demographic attributes was first analysed. If the attributes were significantly correlated, the analysis will not continue the

χ^2 test between the pair and credibility judgements. The correlation between demographic attributes and features readers use for credibility judgements was similarly analysed. The cell in the contingency table that influences the χ^2 value was also measured. The interest or dependence of a cell (c) is defined as I(c)= Oc/Ec. The further away the value is from 1, the higher influence it has on the χ^2 value. Positive dependence is when the interest value is greater than 1, and a negative dependence is those lower than 1 [Brin et al., 1997].

In this study the demographic data collected from the readers were used for chi-square analysis; refer to Table 5.1. The readers' demographic data, except for gender, were also categorised into binary and categorical settings based on other research [Fogg et al., 2001; Greer and Gosen, 2002] to examine any correlation between demographic attributes or combinations of demographic attributes with tweet credibility perception. We partitioned the demographic data to ensure that the credibility perception was tested with the different methods of demographic data categorisation found in previous research, so that our findings would become a credibility perception baseline for a multiple set of demographic data categories. The different ways of partitioning demographic data are as follows:

- Age: Binary {Young adult ($\leqslant$ 39 years old), Older adult ($\geqslant$ 40 years old)} and Categorical {Boomers (51-69 years old), Gen X (36-50 years old), Gen Y(21-35 years old), Gen Z (6-20 years old)} [McCrindle et al., 2010]

- Education: Binary {Below university level, University level} and Categorical {School level, Some college, Undergraduate, Postgraduate}

- Location: Binary {Eastern hemisphere, Western hemisphere (divided by the prime

meridian)} and Categorical {Asia-Pacific, Americas, Europe, Africa}

Correlation analysis for each single demographic attribute for all the different slicing with credibility judgements or features was conducted.

5.4 Data Analysis Results

A total of 10,571 credibility judgements for 1510 news tweets were collected from the user study. Only 9828 judgements from 819 readers were accepted for this study because only those readers answered the demographic questions and completed all 12 judgements.

Any credibility judgements that were found to not describe the features used to make the credibility judgements or contained nonsensical comments were discarded. The study also discarded judgements of two readers from the Oceania continent, and three readers that did not have any educational background, due to their low values undermining the required minimal expected frequency to apply the χ^2 analysis. The final dataset for analysis constituted 754 readers with 9048 judgements.

5.4.1 Overall Demographics

The final collection of data included readers from 76 countries with the highest number of participants coming from India (15%). Countries were grouped into continents due to the countries' sparsity. Out of the 754 readers, the majority (69.0%, n=521) of readers were male, similar to prior work that used crowdsourced readers [Kang et al., 2015]. Most of the readers were in the age group of 20-29 years old (43.4%, n=327). In regard to the readers' educational background, the majority had a university degree (38.1%, n=287). Table 5.1

shows the readers' demographic profiles.

Table 5.1: Demographic attribute distribution

Demographic	Value	Frequency	%
Gender	Male	521	69.2
	Female	233	30.8
Age	16-19 years old	58	7.7
	20-29 years old	327	43.4
	30-39 years old	243	32.2
	40-49 years old	89	11.8
	50 years and older	37	4.9
Education	High school	127	16.8
	Technical training	58	7.7
	Diploma	81	10.7
	Professional certification	50	6.6
	Bachelor's degree	287	38.1
	Master's degree	137	18.2
	Doctorate degree	14	1.9
Location	Asia	275	36.5
	Europe	247	32.8
	South America	130	17.2
	North America	65	8.6
	Africa	37	4.9

5.4.2 News Topics

Other than the demographic data one the Twitter readers, this study also aimed to determine whether the news topics of the tweets would affect the readers' credibility perceptions. The news topics were indirectly presented to the readers as to how they regarded the news type: breaking news, natural disaster news and political news, the year the news occurred, and the trending attribute. After pre-processing the raw data to get the final dataset for further analysis, the ratio between the tweets for each news topics was found, as is shown in Table 5.2.

This result ensured us that there would be no biased analysis regarding the news topics because the proportions were fairly equivalent.

Table 5.2: Tweets news topic distribution

News Characteristic	Value	Frequency	%
News Type	Breaking News	509	33.8
	Natural Disaster	500	33.2
	Politics	499	33.0
Year	2011	374	24.8
	2012	375	24.9
	2013	377	25.0
	2014	382	25.3
Trending	Trending	781	51.8
	Not Trending	727	48.2

5.4.3 Features

The features reported by readers were tweet surface features as listed on the user study. For features reported in free text, a summative content analysis was applied based on the list of features identified beforehand [Hsieh, 2005], the same process as in the previous user study. Table 5.3 (Column 2) lists the features reported by readers when making their credibility judgements. Since the features are sparse, it is difficult to analyse their influence on the readers' credibility judgements. Therefore, the features were grouped into five categories:

- **Author**: features pertaining the person who posted a tweet, including the Twitter ID, display name, and the avatar image;

- **Transmission**: features in a tweet message for broadcasting the messages on Twitter;

- **Auxiliary**: auxiliary information external to the textual message, including URL links, pictures, or videos;

- **Topic**: words and phrases indicating the search topic or news type, including search keywords and alert phrases such as "breaking news";

- **Style**: writing style of a tweet, including language style as well as message style as expressing opinion or stating facts.

These five categories were based on the feature categorisation by Castillo et al. [2011].

Table 5.3: Features reported by readers to judge credibility of news tweets

Category	Feature	Description
Author	Tweet author	Twitter ID or display name e.g. `Sydneynewsnow`
Transmission	User mention	Other Twitter users' Twitter IDs mentioned in the tweet starting with the @ symbol e.g. `@thestormreports`
	Hashtag	The # symbol used to categorise keywords in a tweet e.g. `#Pray4Boston`
	Retweet	Contain the letters RT (retweet) in the tweet and the retweet count
Auxiliary	Link	Link to outside source - URLs, URL shortener
	Media	Picture or video from other sources embedded within the tweet
Topic	Alert phrase	Phrase that indicates new or information update regarding a news topic - e.g. `Update`
	Topic keyword	The search keyword regarding a news topic e.g. `Hurricane Sandy`
Style	Language	The language construction of the tweet (formal or informal English)
	Author's opinion	Tweet that conveys the author's emotion or feeling towards the news topic
	Fact	Factual information on the tweet regarding the news topic

5.4.4 Findings

The Difference between Human Perception and Machine Predicted Credibility Levels

Overall, 1317 tweets were used for this analysis rather than 1510 tweets. One hundred and ninety-three tweets had to be deleted from the dataset as 113 tweets were no longer available on Twitter, and 80 tweets did not have a consensus judgement by a majority vote among the readers. The agreement between the two lists of credibility levels of news tweets was calculated using Cohen's Kappa. The Cohen's Kappa agreement test shows that both humans and the tool had a slight agreement regarding the credibility level of news tweets where, Cohen's kappa = 0.04. The agreement matrix between the two is shown in Table 5.4.

Table 5.4: The agreement matrix between readers' credibility perception and automated credibility prediction

		TweetCred			
		Very credible	Somewhat credible	Not credible	Total
Readers	Very credible	256	654	67	977
	Somewhat credible	51	230	50	331
	Not credible	1	4	3	8
	Total	308	888	120	1316

Although the credibility level from both the tool and readers was more on the credible side, it is clear that readers are more trusting of news tweets. Meanwhile, the automated tool gave mixed credibility prediction with 'somewhat credible' being more prominent than the other two credibility levels.

Correlation Analysis

The correlation analysis for individual demographic attributes for each data setting (as described in 5.3.2): Original (O), Binary (B), Categorical (C), and the credibility perceptions are shown in Table 5.5. At the original data setting, Education and Location were significantly correlated with credibility judgement, $\chi^2 = 49.43, p < 0.05$ and $\chi^2 = 80.79, p < 0.05$. Only Location is significantly correlated at all levels of partitioning.

A post hoc analysis of the interest value of cells in the contingency table $Education \times Credibility$ for the original data found that the cell that contributes most to the χ^2 value is readers with a 'Professional certification', who commonly gave 'not credible' judgements. In regard to the contingency table $Location \times Credibility$, a correlation between the readers from the African continent and the 'cannot decide' credibility perception was found in the original and the categorical data setting with a positive dependence. Both cells' interest values were far from 1, indicating strong dependence. In the contingency table for $Location \times Credibility$ in the binary data setting, the interest value in each cell was close to 1 therefore, there was no strong dependence.

Next, multi-way correlation analysis between combinations of demographic attributes and credibility judgements were conducted. Since Location is significantly correlated at all data levels, due to the upward closeness of χ^2 statistics (see Section 5.3.2), combinations including Location will not be analysed. The correlation result for the other demographic attribute pairs is shown in Table 5.6. In analysing the combination of demographic attributes, Bonferroni corrections of the *p-values* $p < 0.003$ were applied [Wright, 1992].

Table 5.6b shows that only for the binary setting the (Age, Education) pair was not sig-

Table 5.5: Demographic profiles and credibility perception chi-square results

Demographic	Data setting	Credibility	
		χ^2	p-value
Gender	Original	1.51	0.680
	Binary	1.51	0.680
	Categorical	1.51	0.680
Age	Original	4.87	0.249
	Binary	4.68	0.197
	Categorical	9.84	0.132
Education	Original	***49.43	9.2E-5
	Binary	4.78	0.189
	Categorical	12.29	0.197
Location	Original	***80.79	2.918E-12
	Binary	***39.62	1.286E-8
	Categorical	***80.33	1.388E-13

*$p - value < 0.05$, **$p - value < 0.01$, ***$p - value < 0.001$

nificantly correlated. Therefore, the correlation of the (Age, Education) pair with credibility judgements was further analysed. The correlation analysis outcome for *Age × Education × Credibility* is $\chi^2 = 3.70$,p>0.003, accepting the null hypothesis. The result indicates that the joint independent demographic attributes of Age and Education in the binary setting do not correlate with the credibility judgements.

To determine the correlation between the news topics and readers' credibility perceptions, the chi-square test of independence was continuously applied. Table 5.7 shows the correlation result between tweets news topics and readers credibility perception.

As had been hypothesised, all the news types were significantly correlated with credibility judgements. A post hoc analysis was executed to determine the interest value for each contingency table that contributes the most to the significant χ^2 value. In the contingency

Table 5.6: *Chi-square result for demographic attribute pairwise correlation*

(a) (Age, Gender) & (Education, Gender)

		Gender	
		χ^2	p-value
Age	O	***107.71	2.242E-22
	B	***77.40	1.065E-13
	C	***82.18	5.227E-16
Education	O	***105.89	1.324E-9
	B	***48.67	2.572E-12
	C	***61.80	2.421E-13

*$p - value < 0.05$, **$p - value < 0.01$, ***$p - value < 0.001$

(b) Age, Education

	Education					
Age	O		B		C	
	χ^2	p-value	χ^2	p-value	χ^2	p-value
O	***1791.23	7.752E-305	***763.96	4.911E-164	***1579.96	0.0E-0
B	***105.89	1.474E-20	2.18	0.133	**47.96	2.171E-10
C	***1732.96	0.0E-0	***749.53	1.351E-154	***1549.49	7.751E-305

*$p - value < 0.05$, **$p - value < 0.01$, ***$p - value < 0.001$

Table 5.7: *News characteristic correlation with readers' credibility perception*

News characteristics	Credibility	
	χ^2	p-value
News Type	**93.75	5.039E-18
Year	**61.89	5.775E-10
Trending	*8.09	0.044

*$p - value < 0.05$, **$p - value < 0.001$

table of *News type* X *Credibility*, tweets that report 'breaking news' being perceived as 'very credible' by readers were found to show a strong positive dependence on the chi-square result. As for the contingency table *Trending* X *Credibility*, the strong dependence comes

from 'trending' tweets with 'very credible' judgement by the readers. In the last contingency table, *Year* X *Credibility*, tweets that received a 'somewhat credible' judgement from the readers and reporting news occurring in 2014 have a positive dependence in the correlation between the two variables.

Afterwards, a multi-way correlation analysis between the combination of readers' demographics and news types with readers' credibility perceptions was conducted. In analysing the combination of demographic attributes and news characteristics, as well as conducting the multi-way correlation test, Bonferroni corrections of the p-values where $p < 0.001$ were applied due to multiple hypotheses being tested. The study discovered that all attributes from both variables did not correlate with each other in all demographic data settings. Table 5.8 shows the correlation result between readers' demographics in the original data setting and news types, since all other data settings achieved a similar result. Thus, we only focussed on the demographics original data setting combination with news types for the multi-way correlation test.

Table 5.8: Chi-square result between readers' demographics and news topics

		News Characteristic					
		News Type		Year		Trending	
		χ^2	p-value	χ^2	p-value	χ^2	p-value
Demographic	Gender	1.24	0.538	4.84	0.184	0.34	0.560
	Age	4.36	0.823	9.48	0.662	6.63	0.157
	Education	14.45	0.273	23.5	0.172	6.24	0.397
	Location	3.99	0.858	13.6	0.327	2.01	0.734

The results of the correlation analysis are shown in Table 5.9. Not all multi-way corre-

lation test results have a significant correlation based on the corrected p-value (p<0.001).

The result indicates that readers' demographics paired with the news topics do not signifi-

cantly correlate with credibility perceptions. However, the variables' individual attributes -

readers' age and their geolocation - paired separately with the year the news occurred, were

found to correlate significantly with the readers' credibility perceptions. The combination

of the tweets news type and the readers' locations also showed a significant correlation with

credibility perception.

Table 5.9: Correlation between readers' demographics and news topics with credibility perception

Demographic	News characteristic	Credibility	
		χ^2	p-value
Gender	News type	6.94	0.326
	Year	8.43	0.491
	Trending	7.38	0.061
Age	News type	35.53	0.061
	Year	*53.06	0.033
	Trending	18.59	0.099
Education	News type	47.81	0.090
	Year	64.56	0.154
	Trending	16.92	0.529
Location	News type	*38.35	0.032
	Year	*55.16	0.021
	Trending	17.17	0.143

$^*p-value < 0.05$

The association between the readers' demographics and news topics at the item set level

with regard to the readers' credibility perceptions using association rule mining was further

investigated. The minimum support to 1% and the consequent towards the credibility per-

ception as per the objective of the association rule mining were set. The extracted rules were

then pruned for redundant association using the algorithm proposed by Ashrafi et al. [2004] regarding the items in the antecedent that have the same item in the consequent.

Table 5.10 shows the top 10 association rules ordered by lift, the ratio between the confidence and the support of the item set in the rule consequent. From the table, the most interesting rule is of female readers with a higher education level (Bachelor's degree) rating trending political news as 'somewhat credible', as shown in the first row. Meanwhile, trending news that occurred in 2012 and 2013 (this user study was conducted in 2014) was associated with the 'very credible' credibility level. Although female readers find trending political news as 'somewhat credible', they perceived trending natural disaster news topics as 'very credible' (Row 4), while male readers tended to perceive trending breaking news topic that mainly occurred in 2013 as 'very credible' (Row 8).

To easily view the interesting antecedent and consequent rules ordered by lift,the rules were visualised using grouped matrix plot, as shown in Figure 5.5. This visualisation grouped the rules based on similar antecedents that were statistically dependent on the same consequent. The antecedents consisted of the most important item in the group, the number of other items in the group and the number of rules displayed as the column labels. The row labels on the right-hand side (RHS) are the consequent item shared by the groups. For example, in the first column, the first three association rules shown in Table 5.10 were found to have been grouped together since the rules have the same consequent of 'somewhat credible' credibility level. Politics is the most important item in the group among eight other items.

In Figure 5.5, 346 non-redundant association rules were placed into 10 groups. The lift value, represented by the colour of each balloon, is the aggregated interest measures of

Table 5.10: Associations between demographics and news characteristics for credibility perception

Association Rules	Support (%)	Lift
{news type=Politic, trending=Trended, gender=Female, education=Bachelor's degree} ⇒ {credibility=Somewhat credible}	1.0	1.6
{news type=Politic, trending=Trended, age=20-29 years old, location=Europe} ⇒ {credibility=Somewhat credible}	1.0	1.5
{gender = Male, education = High school, location = Asia} ⇒ {credibility = Somewhat credible}	1.7	1.5
{news type = Natural disaster, gender = Female, location = North America} ⇒{credibility = Very credible}	1.1	1.4
{news type = Breaking news, trending = Trended, year = 2013,age = 20-29 years old} ⇒ {credibility = Very credible}	1.1	1.3
{news type = Breaking news, year = 2012,location = South America} ⇒{credibility = Very credible}	1.1	1.3
{year = 2012,gender = Male, education = Technical training} ⇒ {credibility = Very credible}	1.0	1.3
{news type = Breaking news, trending = Trended, year = 2013,gender = Male} ⇒ {credibility = Very credible}	1.7	1.3
{year = 2013, gender = Male, education = Technical training} ⇒ {credibility = Very credible}	1.0	1.3
{gender = Male, education = Technical training, location = Europe} ⇒ {credibility = Very credible}	1.6	1.3

each group. The darkest colour containing the most interesting rules, directs to the top left corner on the left-hand side (LHS). The size of the balloon shows the aggregated support value. The group where political news is the most important item combined with eight more items, which would most likely be perceived as 'somewhat credible' by readers were the most interesting rules. Other than the known rules, as mentioned above, regarding 'breaking news being perceived as 'very credible' and male readers perceiving news tweet as 'very credible', the study also found that readers from 'South America' were also likely to perceive news

tweets as 'very credible'.
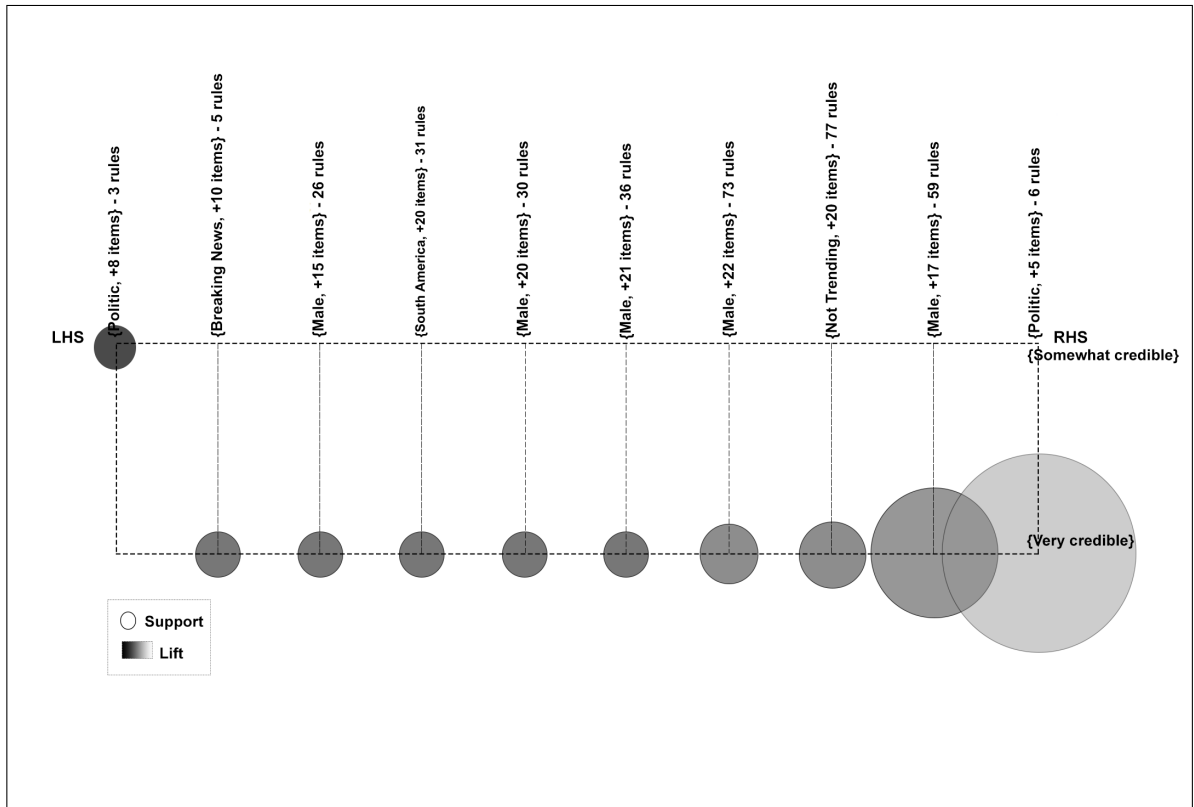
*Figure 5.5: Grouped matrix-based for 346 association rules with **k=10 groups***

Table 5.11 shows that all demographic attributes were significantly correlated with credibility perception features reported by readers. In the last column of Table 5.11, demographic attributes and the Transmission feature, more than 20% have expected values less than 5. Therefore, Fisher's Exact Test was used [McDonald, 2009]. The table is based on demographic data at the original setting, and similar results were obtained for data in binary and categorical settings. As all demographic attributes were correlated with credibility perception features, due to the upward closeness of chi-square statistics, any combination of demographic attributes was also correlated with credibility perception features.

Table 5.11: *The chi-square correlation between demographics and features used in credibility perception*

Demographic	Feature Categories				
	Author	Topic	Style	Auxiliary	Transmission
Gender	0.01	**18.15**	**23.27**	1.59	0.59*
Age	**16.63**	**26.65**	**41.99**	8.65	1.00*
Education	11.12	**31.87**	**50.12**	**16.53**	**0.03***
Location	**46.87**	**83.81**	**67.35**	**13.60**	1.00*

*Calculated using Fisher's Exact Test

Topic and Style features had the most significant correlation with the demographic profiles while the Transmission feature has the least significant correlation with demographic attributes. *Age* and *Location* were significantly correlated with Author, and *Education* and *Location* were correlated with Auxiliary features. Meanwhile, only *Education* had significant correlation with Transmission.

The study also investigated the combination of features readers reported to use when perceiving the credibility level of tweets. Using association mining to find the frequent combinations of features [Hahsler et al., 2007], the study found that Transmission, Author and Auxiliary were frequently used with other features. Table 5.12 shows the frequent features that meet the support threshold of 1%, or 90 times. The support threshold refers to a feature's frequency of occurrence in the dataset. A low support threshold would help to eliminate uninteresting patterns [Tan et al., 2005].

5.5   Discussion

This study provides insight into reader perceptions of information credibility of news on Twitter, in terms of the interaction among reader demographics, news attributes and tweet

Table 5.12: Frequent pattern mining of feature categories

Frequent patterns	Support (%)
Topic	14.1
Style	12.7
Topic, Style	6.1
Auxiliary, Style	5.2
Auxiliary, Topic	4.7
Auxiliary, Topic, Style	4.6
Auxiliary,Topic,Style,Transmission	3.7
Auxiliary,Topic,Transmission	2.7
Author	2.7
Author,Auxiliary,Topic,Style,Transmission	2.6
Topic,Style,Transmission	2.5
Style,Transmission	2.0
Auxiliary,Style,Transmission	1.9
Author,Topic,Style	1.8
Author, Style	1.8
Topic,Transmission	1.6

features. Our user study was conducted on a crowdsourcing platform, inviting participants from different continents and of various demographics. The richness of the data allowed us to evaluate the correlation betweeb readers' demographics, news attributes and tweet features with their perceptions of the credibility of news tweets.

However, first, we sought to understand whether there is a difference between human perception and machine-predicted credibility levels. In a previous study, we saw that the readers were more focussed on tweet surface features than investigating other features of a tweet, such as the tweet's metadata; for example, opening up the page of the tweet's author and looking through the author's connections with other Twitter users (follower-followee relationships) in order to determine the credibility of the author in relation to a particular news genre. Therefore, we investigated whether there was a difference in the credibility levels

of tweets as based on readers' credibility perceptions and the credibility scores computed using a supervised automated ranking algorithm that determines the credibility of a tweet based on more than 45 features, including the tweet's metadata. The credibility prediction tool used in this comparison was TweetCred, developed by Gupta et al. [2014].

The differences in credibility levels between readers' perceptions and by automatic predictions are obvious. Readers were found to give more 'very credible' judgements while the automatic credibility prediction tool produced more 'somewhat credible' ratings. The investigation revealed that automated credibility prediction indeed uses metadata or features that are not readily available to readers in forming their credibility perceptions of tweets. However, at the binary level (credible and not credible), both the tool and readers agreed that news tweets are believable (credible), even if they are rumours. For the few rumour tweets that we have placed in the user study, we found that both the reader and the automated tool only labelled about 15% of these tweets as 'not credible' and 85% of the tweets as 'credible'. This result helps explain why so much misinformation and rumour tweets are propagated on Twitter [Bruno, 2011; Jin et al., 2013; Sakamoto et al., 2014; Starbird et al., 2014].

Next, we studied how readers' demographic data contributed to the way readers view the credibility levels of news tweets. Our study discovered that a readers' educational background and their geolocation have a significant correlation with credibility judgements. This finding is different from other studies [Greer and Gosen, 2002; Kang et al., 2015; Yang et al., 2013], as these studies did not reveal a significant correlation between tweet credibility perception and educational background. From the analysis, readers with a 'Professional certificate' and who perceived tweets as 'not credible' are the ones that contributed the most to the significant χ^2

result. It is likely that educational background is connected with experience and, thus, such readers are more careful in making credibility judgements. Another possible reason may be the absence or a low number of higher education level participants in this study.

Although other researchers found location correlated with credibility judgements in general, the dataset of international readership further shows that readers from Africa, especially, have positive dependence on the 'cannot decide' credibility judgement. Conflicts in a country may play a role in the sceptical attitudes held by readers from these countries towards the media [Cozzens and Contractor, 1987]. Therefore, tweets that readers find ambiguous resulted in indecisive judgements on the tweet's credibility [Rassin and Muris, 2005]. Other demographic attributes, namely age and gender, were not correlated with tweet credibility perception, which is a result similar to the work by Cassidy [2007]. Moreover, the combination of age and gender did not have any significant correlation with tweet credibility perception either.

News topics, including the news type, the year the news is taking place and the trending level of the news, also had a significant association with readers' credibility perceptions. The study further found that trending news topics and breaking news were news areas more likely to be found 'very credible'. The study by Morris et al. [2012] showed that their participants developed more confidence in the credibility of a tweet that posted similar content to other tweets such as trending topic tweets. Furthermore, Twitter is one of the fastest social platforms in reporting breaking news and spreading news, thus it is likely that Twitter readers find breaking news tweets highly credible [Broersma and Graham, 2013; Hu et al., 2012]. However, tweets reporting news events occurring in 2014 (the year when our user study was

conducted) had a positive correlation with the 'somewhat credible' credibility level. This is likely due to news awareness by readers being progressively updated. The result gives a new view regarding the way readers perceive the credibility level of current news events as old news events, since other research uses only news tweets from a certain time frame within the same year [Gupta and Kumaraguru, 2012; Hu et al., 2012; Kang et al., 2015; Kwak et al., 2010].

Our study also found that selected paired attributes correlated with readers' credibility perceptions of news tweets. At the item level, readers have different perceptions of news. While natural disaster and breaking news were perceived as 'very credible' by both genders, female readers found it difficult to rate political news as highly credible. This is not due to the fact that female readers are not devoted readers of political news as they are often portrayed in fictions. To the contrary, female readers are more critical in their judgements regarding politics which has made them more cautious in believing political news, as reported by Zboray and Zboray [1996]. Furthermore, in the statistics retrieved from the Organisation for Economic Co-operation and Development (OECD) regarding women in politics in 2017, it was found that the average percentage of women as members of parliament is 28.8% and in certain countries, the percentage was more than 40% (e.g. Mexico, Sweden and Iceland) [OECD, 2017]. These statistics show that women are interested in politics. Although young readers lack life experience, they are capable of assessing the credibility of news tweets of different news types, in pretty much the same way as experienced readers [Rieh and Hilligoss, 2008].

The composition of tweets also gives a different impression to readers from different

demographic backgrounds when perceiving the credibility level of tweets. Our study found that all demographic attributes are significantly correlated with topic features: topic keyword and news alert phrase, and the tweet writing style. More than 26% of credibility judgements relied on Topic and Style features. Features that were used in broadcasting tweets – the auxiliary and author feature – were mostly combined with other features. The discovery of the use of combination of features by readers in this study is an important contribution of this research, as these features were previously studied separately.

5.6 Summary

This chapter provides insight regarding reader perception of information credibility of news on Twitter, in terms of the interactions among reader demographics, news topics and tweet features. The user study was conducted on a crowdsourcing platform, and inviting participants from different continents and of various demographics gives richness to the data that allows us to evaluate the correlations between reader demographics, news topics and tweet features with reader credibility perceptions of news tweets. Although the focus of this study is on understanding Twitter readers and whether news topics and features affect readers' credibility judgements, we have also proven that there is quite a difference between machine credibility prediction and readers' credibility perceptions. The difference is based on readers' tendency to focus on tweet surface features.

To summarise:

- readers' educational backgrounds and geolocations have significant correlation with their credibility perceptions, and furthermore, the news characteristics are also signifi-

cantly correlated with readers' credibility perceptions.

- Readers use a combination of features to make decisions regarding tweet credibility where the search topic keyword and the writing style of tweets were most helpful in perceiving tweet credibility.

- Readers tend to be more trusting of information shared on Twitter, possibly due to the limited explicit author information available on Twitter.

Chapter 6

Personality and Behavioural Factor Analysis

6.1 Introduction

In previous chapters, it is shown that Twitter readers use surface features when critically analysing credibility and paying attention to the quality of the references. Previous studies have discovered that readers depend on heuristics approaches for web credibility assessment [Metzger et al., 2010; Rieh and Hilligoss, 2008; Sundar, 2008] and heuristic indicators for source credibility assessment on Twitter [Lin et al., 2016]. The heuristic approach is a reasoning process when making quick decisions. The reasoning can be based on experience, knowledge, intuition and common sense. It can also be based on the personality of a person [Moore et al., 1997]. Heuristics can also reflect how people behave in the decision-making process. Examining readers' reliance on tweet surface features to perform credibility judge-

ments can help us better understand heuristic cues on Twitter. As there is a relationship between personality and heuristic cues, it is important to ascertain the role of reader's personality and their behaviour when perceiving the credibility levels of online information on Twitter.

The chapter is organised as follows: A description of the main two themes in this study – personality and readers' behaviour (Section 6.2) – is given, followed by details on the data collection methodology (Section 6.3) and data analysis (Section 6.4). Next, is an outline of the results of our analysis examining the impact of personality and readers' behaviour on credibility perceptions (Section 6.5). A discussion of the findings is in Section 6.6 and, lastly, Section 6.7 summarises the chapter.

6.2   Personal Characteristics

We use the term 'personal characteristics' to refer to the individual differences between people in the way they think and interpret things, their feelings, and behaviour that show the stability adaptation in life from relatively early childhood through to the end of their life. This definition was adapted from the study by Donnellan et al. [2009]. In this user study, the personal characteristics that we examine in our third research question were the readers' personality traits, dependency on heuristic cues and their credibility perception behaviour. This section will first define the personality traits and readers' behaviour. Heuristic cues were refer to the tweet surface features and the features that were described in Chapters 4 and 5.

6.2.1 Personality

There are several theories on personality, including the Eysenck's Three Factor Model [Eysenck, 1967], the Big Five Personality Traits [McCrae and Costa, 2003; Widiger and Costa, 2013] and the Alternative Big Five [Goldberg, 1990]. In this study, the Big Five Personality Traits was chosen. The Big Five Personality Traits are as follows:

- *Agreeableness* is defined as being trusting, straightforward and selfless. Having a high degree of trust in others and a strong desire to aid others. A willingness to concede to others is also observed with this personality.

- *Emotional stability* relates to a person's level of calm and self-confidence. A person with high levels of emotional stability would be more comfortable, relaxed and unemotional, whereas a person with low levels would have mood swings, angst and get irritated easily.

- *Extraversion* is the tendency to seek stimulation in the company of others, and to express positive emotions. There were two types of personality in this trait. Extroverts tend to be more outgoing and socially active while introverts were reserved.

- *Conscientiousness* is defined as being organised as opposed to being spontaneous. It is also described as being reliable, and planning ahead in the pursuit of long-term goals. The opposite of this (low conscientiousness) is a person who is more easy-going, tolerant, and less bound by rules and plans.

- *Openness to experience* looks at a person's imagination, curiosity, and interest in seeking new experiences of culture or new ideas. A person with high openness is more

adventurous and creative, as opposed to an individual with low openness that is more conservative and traditional.

Previous research has shown that personality is correlated with information-seeking behaviour [Halder et al., 2010], trust [Quijano-Sanchez et al., 2010], and online social behaviour [Adali and Golbeck, 2014; Kosinski et al., 2014]. Although there have been various personality studies, there has been little research on the impact of personality traits with regard to credibility perceptions of information on Twitter.

6.2.2 Readers' Behaviour

Readers' behaviour in making credibility assessments of information on Twitter is based on their selection of features as indicators of the credibility level of a tweet message. The feature indicators Twitter readers' use include:

- the existing features provided on Twitter, e.g., hashtag, author's display name;

- counts of favouritism, e.g., votes, retweets;

- social network relationships, e.g., follower-followee relationship;

- geolocation;

- linguistics, e.g. abbreviation, punctuation.

Several studies have attempted to look for patterns in readers' behaviour regarding the use of credibility indicators [Kang et al., 2015; Morris et al., 2012]. However, previous studies

lacked interpretive information of the credibility assessment process and focussed on the descriptive outcome.

Metzger et al. [2010] and Sundar [2008] found that readers depend on cognitive heuristics when assessing the credibility level of online information. Cognitive heuristics were one or more biased mental shortcuts or cues people use to make a decision when in an uncertain situation, such as deciding the credibility level of information online [Lockton, 2012]. These shortcuts or cues can be features available on online platforms such as blogs, websites and online social media. Lang [2000] described that people would select certain features to encode, store and retrieve information rather than processing all of the information. Zhuang et al. [2016] studied readers' behaviour in relation to the perception of an information retrieval system and found that each person is unique, and they were likely to have different strategies to form perceptions of information. The concept is similar to judging the credibility level of tweets.

To understand the association and effect of personality and behaviour on the credibility perceptions of Twitter readers on news-related tweets, a user study was conducted to capture Twitter readers' behaviour in perceiving the credibility of news tweets and readers' personality. The following sections will explain how the user study was designed, conducted and analysed with factor analysis and the multiple regression model. We then show the findings of our user study.

6.3 Data Collection

In answering the third research question, we continued on in this research with the methodology used in previous user studies, the use of questionnaires and crowdsourcing. Both credibility and psychology studies have applied this methodology. We have discussed the use of questionnaires and crowdsourcing in credibility studies in previous chapters. For psychology studies, Bachrach et al. [2014] proposed a personalised recommendation system for tourist destination attractions. Personal characteristics are one of the features in their recommendation system, and they collected this information by asking crowdsourced workers to provide personal characteristic data, including answering personality test questions. Another study that applied personality tests in crowdsourcing platform as part of their user study was the study conducted by Halko and Kientz [2010]. Their study explores the relationship between personality and persuasive technologies. Thus, we decided to implement the same methodology in our credibility perception user study in order to explore the relationship and role of personality traits in readers' credibility perceptions.

6.3.1 Study Design

In this study, the readers were given a scenario: "Imagine you have read or heard about a news event. You wanted to found out more about the event or the current situation of the news by searching Twitter with query keywords. You are shown tweet messages returned by the Twitter search engine.". This scenario was similar to our previous user studies. The scenario was shown to the readers to help them establish the knowledge to perform a credibility judgement of tweets when a search request is made on Twitter regarding an event.

Thirty simulated news tweets regarding politics, breaking news, and natural disaster news, the query keywords and a simple description about the news were shown to the readers. The intent of the simulated tweet was to control the features on the tweets and to eliminate preconceptions among participants from knowing the news beforehand, a concept based on the work by Yang et al. [2013]. Each of the simulated tweet messages was previously indicated as plausible by seven annotators. The agreement percentage between the annotators was 85.7%.

This user study was divided into three sections. The readers were required to answer some demographic questions in the first section (refer to Figure 6.1). The categories for each demographic data were based on previous studies and also the reports published by the Organisation for Economic Cooperation and Development (OECD). In the second section, the readers were shown a simulated tweet message, topic search keyword and simple description regarding the topic. In this section, the readers were also instructed to state their credibility perception of each tweet and report the features they prefer to use to determine the credibility level as shown in Figure 6.2.

The readers were asked to judge credibility using a scale of 1 to 7 with 1 being "not credible at all" and 7 being "highly credible". Afterwards, the readers disclosed the features that they used to help them make the credibility judgement of the tweets. A feature selection list based on a list of features shown to the participants that have been reported in previous work [Castillo et al., 2011] and the list of features collected in the previous chapter (Section 5.3) were shown to the participants. The feature selection list was programmed to show the features to the readers randomly for each tweet to avoid selection bias. The participants

Figure 6.1: Demographic section

were also encouraged to give reasons for their credibility judgement in a free text interface.

The last section of the user study was a personality test. Readers had to answer self-descriptive five personality dimension questions using a recognised and validated Ten-Item Personality Inventory (TIPI), a short version of the Big 5 Personality Test [Gosling et al., 2003]. TIPI comprised of the positive and negative adjectives representing each personality traits (see Figure 6.3 for the personality test questions adopted from the study by Gosling

Figure 6.2: Tweet message and credibility perception

et al. [2003]).

Existing studies have used this test to study the relationship between personality and various research areas, such as creativity [Batey et al., 2010], recommendation systems [Bachrach et al., 2014], and video gaming preferences [Johnson and Gardner, 2010]. The participants indicated their agreement with the personality statement comprised of positive and negative personality adjectives using a scale of 1 (strongly disagree) to 7 (strongly agree). The personality trait value is the average of the positive and negative personality statement agreement ratings.

Figure 6.3: Ten-Item Personality Inventory (TIPI) personality test

This user study was approved by RMIT University's College Human Ethics Advisory Network (CHEAN) by a delegated CHEAN committee (Approval number: ASEHAPP 36-16); refer to Appendix B.2 for the approval letter.

6.4 Data Analysis

The initial dataset consisted of the credibility judgements of news tweets from 1000 readers. Readers who did not judge all 30 tweets or did not give meaningful feature descriptions were excluded from the dataset. Readers whose judgements were the same for all questions were also deleted to remove response bias. These judgements were deleted because they would provide an imbalanced indication of the credibility perception of the tweets. Furthermore, the quality of the judgements would be questionable when the scores were the same for all tweets. After the data clean-up process was completed, the final dataset comprised of responses from 900 readers and was exported to SPSS for further analysis. To analyse the dataset, several statistical analysis techniques were used. The next subsections will describe these techniques.

6.4.1 Reliability Test

Reliability is concerned with how well a user study has been conducted and the extent of measurement error. A reliable study allows for a good reproduction of the survey data. To ensure that our user study was designed and conducted properly, we applied a reliability test to the data collected from our user study. There were two types of reliability tests conducted in this paper: Cronbach alpha and composite reliability. The Cronbach alpha is a measure to assess the internal consistency of a set of items, while the second reliability test is a test to measure the internal consistency of the construct indicators indicated by each factor in factor analysis. The descriptions for each reliability tests were as follows:

Cronbach alpha

Internal consistency describes the extent to which all the items in a questionnaire test measure the same construct and therefore are connected to the correlation of the items within the test. By using Cronbach's alpha, the reliability test will determine whether the designed questionnaire is accurately measuring the correct latent variable [Moss et al., 1998]. Cronbach's alpha is commonly used to see if questionnaires with multiple Likert scale questions were reliable.

Composite reliability

Composite reliability is generally a function to determine the reliability for multidimensional measures [Widhiarso and Ravand, 2014]. The test will measure the variances of the individual components, the weights assigned to individual components and the correlations between the components. Composite score reliability can be used to test the reliability for both weighted and unweighted dimensions of the multidimensional measure, such as factor analysis without bias.

6.4.2 Multiple Regression

Multiple regression models were probabilistic models that include two or more independent variables (IV) in an equation to predict the relationship between the independent variables and a dependent variable for each subject. The equation for regression is:

$$Y = B_0 + B_1 X_1 + B_2 X_2 + + B_k X_k \tag{6.1}$$

where Y is the predicted regression value, X represents the multiple independent variables and B represents the contribution of each independent variable during regression [Tabachnick and Fidell, 2014].

This technique has been used in many disciplines, such as psychology [Härtl et al., 2010], sales and marketing [Rahbar and Abdul Wahid, 2011] and information technology [Erdur-Baker, 2010]. In relation to applying multiple regression analysis for personality-related studies, Tan and Yang [2014] used this method to look at how personality traits affect users' Internet application usage, while Wang et al. [2015] was able to predict the use of social networking sites from personality traits.

There are different ways of doing multiple regression analysis (e.g. hierarchical, stepwise) and each of these different techniques will answer different types of questions. For example, in hierarchical multiple regression analysis, a researcher would be able to ascertain the significance of certain independent variables (IV) in relation to a dependent variable (DV) based on some rules like a theoretical model. Meanwhile, stepwise techniques are applied to studies that have independent variables that do not have a theory to support them. Therefore, to determine which of the IV have a significant effect in the model, the IV are put into the model one at a time to see if they meet the statistical criteria. If the variable no longer contributes to the regression model significantly, IV will be deleted from the equation [Tabachnick and Fidell, 2014]. In this study, we used the stepwise multiple regression analysis technique as we do not have a theoretical model to refer to regarding the relationship between personality traits and credibility perceptions of tweet messages.

For the multiple regression model, the Big Five Personality traits would be the indepen-

dent variables and the dependent variable would be the credibility perception of news-related tweets by each reader. The statistical criteria used in this analysis for an IV to be included in the regression equation was the standard p-value<0.05. The objective of this analysis was to predict the personality that has a relationship with tweets' credibility perception.

6.4.3 Factor Analysis

Factor analysis is a technique used to reduce the dimensionality of observed variables and seek the underlying unobservable variables that were reflected in the observed variables [Bartholomew et al., 2011]. Factor analysis uses the basic statistics correlation coefficient to interrelate and discover patterns in a set of variables. Many fields have applied this technique.

There were two commonly used factor analysis techniques, Confirmation Factor Analysis (CFA) and Exploratory Factor Analysis (EFA). Confirmatory factor analysis (CFA) is used to statistically verify the factor that is being tested from a hypothesis based from an empirical research, or a theory. The Exploratory Factor Analysis (EFA) will analyse which variables were statistically grouped together based on the factor loadings once the correlation matrix between variables has been computed [Kline, 2014]. Therefore, in this study, EFA was the suitable technique as we wanted to discover the underlying factors influencing the variables in the dataset. The factors in this study were not established prior from previous research or theory: thus CFA was not an appropriate analysis tecnhiques.

The EFA is used to assess the correlation between the features and then to identify the distinctive groups, known as factors, according to the features. Grouping tweet features will help to identify the cognitive heuristics. A readers' cognitive heuristics characterises

his/her behaviour in evaluating the credibility of information and sources online. To discover the cognitive heuristics, the reported features by readers were first sorted into 900 groups containing readers' id, feature type and the average count of each self-reported feature for 30 tweets. The number of factors to be applied was determined by the data points that were shown by the elbow break of the scree plot – the point where the slope of the curve is clearly levelling off [Cattell, 1978].

6.5 Result

Before we discuss the findings, we will first verify the reliability of the dataset we collected and analysed in the user study. To assess the internal consistency of the judgements, Cronbach's α test was performed. The overall credibility judgements resulted in Cronbach's $\alpha = 0.93$, while the internal consistency score for personality test was $\alpha = 0.69$. Thus, all Cronbach Alpha tests were above the acceptable cut-off of 0.60 [Moss et al., 1998] and therefore acceptable for further analysis.

The findings were divided into two subsections. First, the correlation between readers' personality traits and the credibility assessment of different news types, and the regression model between readers' credibility perceptions and readers' personality are presented. In the next subsection, the cognitive heuristics on Twitter and readers' credibility perceptions are presented.

6.5.1   Overall Demographics

The final collection of data for this study included readers from 64 countries. The countries were then grouped into continents due to the countries' sparsity. Out of the 900 readers, the majority (71.9%, n=647) of readers were male. Most of the readers were in the age group of 26-35 years old (43.1%, n=388). In regard to the readers' education backgrounds, the majority had or were pursuing a tertiary education at the bachelor's degree level (35.2%, n=317). Table 6.1 shows the readers' demographic profiles. The skewness of gender and age was somewhat expected as other studies has found similar results on crowdsourced populations [Kang et al., 2015; Tan and Yang, 2014]. Regarding the education level of readers, it is not surprising that a tertiary education level was held by the majority of readers. Based on the report on education by OECD [2018], many countries, especially in the North and South America (the majority of our readers) have higher than average education levels.

6.5.2   Personality

A correlation analysis using Pearson's correlation coefficient was conducted to determine the relationship between readers' credibility perceptions of different news types (politics, breaking news, and natural disaster) and readers' Big 5 personality traits (*extraversion, openness to experience, conscientiousness, emotional stability,* and *agreeableness*). Table 6.2 shows the correlation between readers' stated levels of personality traits and their stated levels of credibility across news types, and the overall credibility perceptions.

From Table 6.2, *agreeableness, conscientiousness, openness to experience* and *emotional stability* have weak positive significant correlation ($r = 0.1$ to 0.3) with credibility perceptions

Table 6.1: Demographic profile distribution

Demographic	Value	Frequency	%
Gender	Male	647	71.9
	Female	253	28.1
Age	18-25 years old	286	31.8
	26-35 years old	388	43.1
	36-45 years old	157	17.4
	46-55 years old	48	5.3
	56-65 years old	14	1.6
	65 years and older	7	0.8
Education	No schooling	16	1.8
	High school	179	19.9
	Diploma	131	14.5
	Professional certification	105	11.7
	Bachelor's degree	317	35.2
	Master's degree	137	15.2
	Doctorate degree	15	1.7
Location	Europe	274	30.4
	South America	272	30.2
	North America	132	14.8
	Asia	130	14.4
	Africa	88	9.8
	Oceania	4	0.4

of political news, breaking news and natural disaster news, and extraversion does not correlate with the credibility of other news types. The weak correlation for this correlation coefficient analysis regarding personality traits and credibility perceptions is common in psychological and personality studies based on the review study by Meyer et al. [2001]. These results indicate that people with high *agreeableness, emotional stability, conscientiousness* and *openness to experience* personality dimension were more perceptive and willing to trust news information on Twitter. Another possibility for the weak correlation is the population participating in this user study. Based on the study by Feitosa et al. [2015] that compares crowdsourc-

Table 6.2: *Pearson's correlation (r) between readers' personality traits and credibility perceptions*

News type / Personality	Agree- ableness	Emotional stability	Conscien- tiousness	Openness to experi- ence	Extraver- sion
Politics	*0.120	*0.139	*0.144	*0.149	0.053
Breaking news	*0.109	*0.139	*0.127	*0.151	0.055
Natural dis- aster	*0.143	*0.138	*0.176	*0.187	0.050
Overall cred- ibility	*0.114	*0.123	*0.134	*0.145	0.071

*$p - value < 0.01$

ing with traditional data collection methods in regard to personality studies, crowdsourcing under-performed when the participants came from non-native English-speaking countries, which is what we have in our data.

We then performed a stepwise multiple regressions test to predict the relationship between the five personality traits and Twitter readers' credibility perception of news tweets. Table 6.3 shows the regression model having a statistically significant relationship between the personality traits and the credibility of news-related tweets. The significant personality traits included in the regression model were *openness to experience* and *conscientiousness*. The regression model indicates that people with high openness to experience and conscientiousness personality traits are more likely to perceive news-related tweets as credible.

The estimated proportion of variation that fit the regression model in this study's dataset was $R^2 = 0.032$. Although the proportion of variation (R^2) in this dataset was only 3.2%, the regression model found in this study was still statistically significant and acceptable, as

Table 6.3: Multiple regression

Dependent variable = Credibility perception				
	Personality traits (IV)	Beta	t	VIF
$R^2 = 0.032$	Openness to experience	0.120	**3.291	1.236
F=**14.598	Conscientiousness	0.088	*2.491	1.236

*$p - value < 0.05$, **$p - value < 0.01$

described by Colton and Bower [2002]. The low degree of variation in personality trait values (low R^2 score) may be due to the readers' delusion in answering the self-report personality test as discussed by McFarland and Ryan [2000]. In the study by McFarland and Ryan [2000], they found that people are delusional when it comes to creating a good image about themselves, thus giving false answers regarding their personality judgements.

6.5.3 Readers' Behaviour

Reliability analysis was conducted and the internal consistency between the reported features was $\alpha = 0.68$. The first step in EFA is deciding the number of factors. Four components could be identified from an initial examination of a scree plot (see Figure 6.4) based on the eigenvalues > 1.0 [Kaiser, 1960]. The four components were determined based on the distinctive break shown by the dotted red line. The accounted variance of the four components was 54.12%. The four-factor model demonstrated a moderate Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO = 0.67), indicating that the factors have a fair amount of variance. The test checks if the variables can be grouped efficiently.

Another test, which checks if there is redundancy between the variables that can be
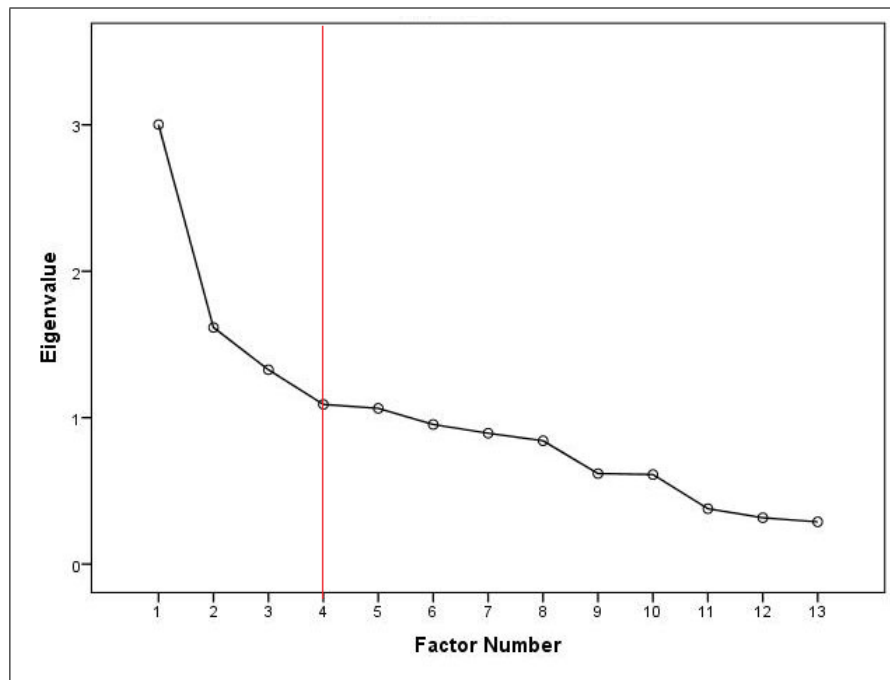
Figure 6.4: Factor analysis scree plot list

summarised with a smaller numbers of factors, is the Bartlett's Test of Sphericity. If the factors were correlated, it means the EFA is useful in summarising the information available in the data. For this dataset, the result of Bartlett's Test of Sphericity was $\chi^2 = 2487.79$, df $=$ 91, $p < 0.001$, a significant result that suggested relation-ships existed among the dimensions. The EFA is conducted using maximum likelihood extraction method for normally distributed variables and the oblique rotation method Promax with the power of four to allow the factors to correlate.

Table 6.4 displays the factor weights for the four readers' behavioural factors and the variables/features for each factor were highlighted. The variables must have a factor loading score of > 0.32 to form the factors [Tabachnick and Fidell, 2014]. Four variables were dropped from the factor groupings as they did not fit the factor loading score criteria, which

116

were media, writing style, sentiment, and language structure.

Table 6.4: Factors in readers' cognitive heuristics in credibility perception

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**
	Endorsement	Confirmation	Reputation	Relevance
Retweet	1.036	0.043	-0.080	-0.096
Number of votes	0.624	-0.079	0.067	0.191
Media	0.169	0.137	0.094	0.032
Link to external source	-0.017	0.688	-0.026	-0.298
User mention	0.044	0.642	0.048	0.195
Hashtag	-0.020	0.559	-0.041	0.324
Author's image	-0.007	-0.165	0.848	0.146
Author's username	0.018	0.234	0.708	-0.226
Writing style	-0.070	-0.011	0.079	-0.016
Alert phrase	0.036	-0.190	0.037	0.625
Topic keyword	-0.018	0.038	-0.117	0.454
Sentiment	0.064	-0.043	0.120	0.256
Language structure	-0.037	0.127	0.067	-0.171

Extraction method: Maximum Likelihood
Rotation Method: Promax with Kaiser Normalisation

Factor 1 seemed to represent the action of believing the tweet is trustworthy due to the fact other people liked and shared the tweet. This factor is similar to the effect of the *endorsement heuristic* where people tend to follow the dominant perception of the tweet, a behaviour found in an online auction or online shopping portal [Melnik and Alm, 2002]. For Factor 2, the features were likely related to the action of examining the news information with the existence of an external source, mentions of other Twitter users or the use of a topic indexing feature (#hashtag). This heuristic was found to be a *confirmation heuristic* rather than the other heuristics found in the literature related to credibility perceptions [Feist and Gorman, 2012]. Further investigation of the comments reported by readers supported the confirmative

behaviour of this heuristic factor. Readers claimed the need to use the features to confirm the credibility level of the news information in the tweet message. Further examination of readers' comments suggest that this is due to the feature/s in this factor being unavailable in the tweet. There were comments like "There is no legitimate account mentioned and no news link" and "No source link to cross-check."

Information regarding the tweet's author that can be seen on the tweet message at first glance (without going to the author's page) is the author's name and the author's image representation, also known as the avatar. Both indicators were grouped together under Factor 3. The author's credibility indicators were used to determine the reputation of the tweet author, and thus this heuristic cue is known as the *reputation heuristic*. The last factor, Factor 4, describes readers' behaviour in connecting the tweet message to news relevance features. Although the items in this factor could also be seen as readers' confirmation bias in regard to the search action [White, 2013], further investigation of readers' comments showed that the items are concerned with (query) relevance. Some examples of comments given by the readers were "The updated information is relevant to the topic", "The keyword is the same" and "Tweet is irrelevant to the query". Therefore, this type of behaviour can be described as *relevance heuristic*.

To measure the internal consistency of the construct indicators, composite reliability was used. Table 6.5 shows the composite reliability for each factor, and all factors were found to be acceptable except for *relevance heuristic* (CR = 0.45). For the *relevance heuristic*, due to its low composite reliability score and that the average variance extracted (AVE) value was less than the recommended value of 0.5, the factor was dismissed from further analysis.

Table 6.5: Reliability scores for each factor

Factor	Composite Reliability (CR)	Average Variance Extracted (AVE)
Endorsement	0.836	0.731
Confirmation	0.701	0.442
Reputation	0.756	0.610
Relevance	0.453	0.298

For the *confirmation heuristic*, although its AVE value was also < 0.5, based on the work by Fornell and Larcker [1981], if $CR > 0.6$, $AVE > 0.4$, the factor is acceptable. Therefore, the *confirmation heuristic* was accepted as part of the cognitive heuristics for credibility perceptions on Twitter.

The main feature for each factor (bolded in Table 6.4) was used to assess how the readers behaved when evaluating the credibility level of news tweets. The main feature was submitted to a Cluster Analysis using Ward's hierarchical clustering method. The Ward's hierarchical clustering method was used because the clustering method was based on proximity of two closest identities, until only a single cluster remains. The number of clusters was determined after the results were manually inspected and was based on the dendrogram result. The dendrogram or tree diagram is a visualised representation of the similarity among entities in a dataset. By looking at the distance-based decision rule scree plot, the number of clusters can be identified (based on the distinctive break, "elbow" similar method in Figure 6.4). After determining the number of the clusters from the dendrogram, the main features were input to the k-means clustering algorithm to be clustered based on the nearest means. For this data, 900 readers are best distributed into 3 clusters (see Table 6.6).

119

Table 6.6: Means of readers' behaviour factor by cluster

Cluster	n	Retweet	Link to external source	Author's image
1	322	4.649	22.916	3.761
2	412	2.816	3.973	3.058
3	166	11.434	8.765	8.813

In Table 6.6, the first cluster shows readers who rely on external links to help them with their credibility perceptions. A link to an external source is a feature that relates to the confirmation heuristic. A low average reliance on cognitive heuristics can be found in the second group. Analysing their comments, some readers in this group were found to report the use of their beliefs when deciding the credibility level of the tweets. However, this study cannot ascertain the low usage of surface features is due to readers' belief as only $\sim 15\%$ of the readers in the study left an optional comment regarding the way they make credibility judgements. The third group, with the lowest number of readers, relied more on the retweet feature. The retweet feature is categorised under the *endorsement heuristic*.

From further investigation, we found that readers in the second group perceives the least number of tweet messages as credible (16 out of 30). With the tweets used on the user study being plausible but non-factual fake news, it can be assumed that the readers in the second cluster did not overly-rely on the cognitive heuristics when perceiving credibility, which enabled them to judge the tweet messages based on their content rather than the surface features. Comparing the credibility perceptions of each cluster, readers relying heavily on *confirmation heuristics* were found to be the ones that made the highest number of credible

judgements for news-related tweet messages (23 out of 30).

6.6 Discussion

Based on the Big Five Personality Inventory, four out of the five personality traits, except for the *extraversion* trait, were found to have significant correlation with the perceived credibility of news-related tweets, politics, breaking news, and natural disaster news types. Further investigation revealed that the personality traits known as *openness to experience* and *conscientiousness* were the main personality traits associated with the credibility perceptions of tweet messages.

The results in this study were consistent with research on the association between personality and decision making [Lauriola and Levin, 2001; LePine et al., 2000]. Readers with high *openness to experience* and *conscientiousness* personalities consider assessing the credibility of tweets as a risk-taking activity. The consistency in the association between personality traits and credibility perceptions and decision-making probably stems from the consequences of disseminating false information online if a tweet message is perceived wrongly. However, it is beyond the scope of this study to discuss in further detail the different facets of the different personality traits in relation to credibility perceptions.

Readers' behaviour also affected their judgements of tweet credibility. Using factor analysis, the study discovered three cognitive heuristics associated with credibility perceptions: *endorsement*, *confirmation*, and *reputation*. The *endorsement heuristic* is based on the number of retweets and votes of the tweet message. Morris et al. [2012] identified that retweet features available on Twitter, whether an author is retweeting a tweet from another author or

the number of retweets made, is one of the top credibility indicators for readers. The number of votes has also been discussed in E-commerce and online branding marketing as a feature that improves the appeal of products for sale, such as an online rating mechanism. Metzger et al. [2010] described the same heuristic as the *endorsement heuristic*, describing the way people tend to perceive information as more credible if other people show their agreement with the information.

The second cognitive heuristic is known as the *confirmation heuristic*, consisting of three features: link to an external source, user mention and hashtags. In previous Twitter credibility studies, the three features were described as individual feature for credibility perceptions. A hashtag can be used to categorise the tweet into groups, linking relevant topics and events together [Davidov et al., 2010]. Links on a tweet relate to the original source of information, albeit shown as a full or shortened URL by Castillo et al. [2013] and Morris et al. [2012]. For the last feature, user mentions were a tag-like feature on the user level. The name value of the user mentions builds up influence on Twitter that helps the mentioned user get responses from others [Cha et al., 2010; Jiang et al., 2015]. Combined, these three features give an opportunity for readers to validate the information on the tweet message more confidently, as what is found in this research. Furthermore, *confirmation heuristic* on Twitter, as discovered in this study, is different from other credibility heuristic on the web. The difference can be seen on the usage or lack of it on the web. On the web, such as blog or websites, the user mention features were non-existence as there were no social networking relationship available and the use of hashtag does not have the same impact that it has on Twitter (i.e indicating trending topics or acting as a keyword search that can be used to verify information).

The features grouped together under *reputation heuristic* describe the authors of the tweet messages. Studies have shown that an author's representation such as their username and image were among the top credibility indicators on online media including Twitter [Johnson, 2011; Kang et al., 2015; Morris et al., 2012]. This factor has also been discussed in heuristic research known as the *authority heuristic* on both the web and social media [Lin et al., 2016; Sundar, 2008] while Metzger et al. [2010] refer to the same heuristic cues as the *reputation heuristic*.

In terms of how *cognitive heuristics* are used by readers, three types of readers were found in the dataset. The first group of readers perceive the highest number of tweets as credible based on *confirmation heuristics* especially based on the external link embedded in the tweets. West et al. [2012] describe how people who use this type of cognitive heuristic tend to have a bias blind spot, which explains the high number of credible judgements given to fake news used in this study. Another group of readers did not appear to rely on cognitive heuristics as heavily. Lastly, the third group of readers used all three cognitive heuristics moderately to perceive the credibility level of tweet messages with *endorsement heuristics* especially the retweets feature.

People who rely on social media to receive news updates locally and globally often have to determine the trustworthiness of the news content they are presented with. This study shows that readers have difficulty in determining the credibility level of news information on social media, particularly considering the rise of fake news on social media. This study suggests that readers mostly rely on cognitive heuristics to determine whether news-related tweet messages are credible or not while also being influenced by particular personality traits.

123

6.7 Summary

In this study, readers' personalities and behaviours are examined to see whether these aspects have a role in the credibility perceptions of news tweets. The credibility perception user study is designed based on a search scenario where a tweet message is disseminated and could be from any author. The results suggest that the self-reported levels of some personality traits correlate with readers' perceptions of the credibility of three types of news-related tweets: politics, breaking news and natural disaster news. It was found that readers' reliance on cognitive heuristics in perceiving the credibility of tweets is different from that of the web [Metzger et al., 2010]. On the web, readers look for any feature that does not fit with their expectations, whether it be the design or functionality of a webpage [Metzger et al., 2010]. Since Twitter's design is uniform, this feature is not available for readers to refer to. In addition, three different types of readers' behaviour regarding the use of cognitive heuristics when perceiving the credibility level of tweet messages were found.

To summarise:

- The *openness to experience*, and *conscientiousness* personality traits have a correlation with readers' credibility perceptions of news-related tweets.

- Readers also use three types of cognitive heuristics based on Twitter features to assess the credibility of tweet messages: *endorsement*, *reputation*, and *confirmation heuristic*.

- There were three categories of behaviour for readers' credibility perceptions on Twitter: readers who mainly depend on confirmation heuristics, modestly rely on cognitive heuristics and readers who only slightly depend on cognitive heuristics, constituting a

new model to describe readers.

Chapter 7

Conclusion

In this thesis, we conducted three user studies to answer research questions on readers' credibility perceptions of information on Twitter. Specifically, the thesis looked at how readers judge the credibility of tweets; the correlation of factors related to external attributes with credibility perceptions; and the role of readers' personal characteristics in influencing credibility evaluations of information on Twitter. In our literature review, we found that most previous studies on information credibility focussed on credibility evaluation models regarding Twitter features and sources. Credibility perception is subjective and external factors may influence reader perception. These factors may affect readers' attitudes and preferences, and how they interpret the truthfulness of information shared on Twitter.

The chapter is organised as follows: summary of findings (Section 7.1 - Section 7.3), limitations of the study (Section 7.4), followed by directions for future research (Section 7.5).

7.1 Credibility Features

The first research question investigated how readers determine the credibility level of information on Twitter. We designed a within-subjects user study based on the credibility perception methods from Morris et al. [2012], Castillo et al. [2011] and Gupta and Kumaraguru [2012]. Twenty news event topics that occurred between 1 June 2013 and 15 October 2013, and 400 relevant tweets collected from the Twitter API using query terms related to news topics were selected for the study. We asked the readers to annotate the credibility levels of the tweet messages and describe the credibility features that influenced their credibility judgements.

Through the readers' comments, features were extracted using summative content analysis and analysed using predictive association rule analysis to establish associations between features and credibility levels. Eight features were identified, where topic query keyword, display name, link in tweet and user belief in the tweet topic were found to be the most important. By feature and credibility association analysis, we found strong associations between features and tweet credibility. The associations also showed that readers commonly combine features, especially topic keyword and the reader's belief, in perceiving a tweet message as credible. We further found that politics and breaking news are more difficult for users to consistently rate as credible. The lack of a link to external sources in a tweet was found to negatively affect credibility perceptions, giving inconsistent judgments.

7.2 Factors Correlated with Credibility Perceptions

The second research question addressed the relationship between subjective factors and credibility perceptions. Continuing on from the first user study, a larger user study was conducted.

A total of 1510 tweets returned from 15 news-related topics covering breaking news, political news, and natural disaster news, were judged by 754 participants on a crowdsourcing platform. The readers were asked to fill in a questionnaire divided into two sections. The first section looked at basic demographic questions and the second section focussed on the credibility annotation of tweet messages and describing the credibility features used in credibility judgements.

This study explored the correlation between readers' demographic attributes, credibility judgements, news topics and features used to judge tweet credibility. Correlation analysis was conducted to study the correlation between each demographic attribute with credibility judgements or features. Association rule mining was administered to find interesting rules that describe the relation between the readers' demographics and news topics as well as the credibility level of news tweets as perceived by readers. This study also compared the credibility ratings between readers and an automated tweet credibility prediction tool called TweetCred in order to identify differences in credibility judgement between the two.

The findings of this study showed that the difference between readers' credibility perceptions and the automated credibility prediction tool stemmed from readers' behaviours in making credibility judgements; they tended to use surface features shown in tweets rather than conduct a deeper investigation of the information shown, such as the tweet's metadata, which was embedded in the credibility prediction tool. Readers' geolocations and educational backgrounds were also found to have a significant correlation with readers' credibility perceptions of news-related tweets. Trending news and breaking news were found to be most favourably perceived by readers. The study also found that selected paired attributes corre-

lated with readers' credibility perceptions of news tweets. The findings provide new insight on the relationship between different factors and credibility perceptions.

7.3 How Personal Characteristics Influence Credibility Perceptions

The third research questions examined the role of a reader's personal characteristics in perceiving the credibility level of information on Twitter. A user study was conducted to capture Twitter readers' behaviours in perceiving the credibility of news tweets and the readers' personalities from a personality test. Thirty simulated news tweets regarding politics, breaking news, and natural disaster news were shown to the readers to avoid biasing credibility with prior knowledge of the news. The features in the tweets were also controlled and resembled tweets returned by the Twitter search engine as results for searches with query keywords.

Data collected from 900 readers answering the user study was analysed with factor analysis and the multiple regression model. The results show that *openness to experience*, and *conscientiousness* personality traits had the greatest effect on readers' credibility perceptions of news-related tweets. We also found three types of cognitive heuristics were used in determining the credibility of a tweet message: *endorsement*, *reputation*, and *confirmation heuristics*.

The *endorsement heuristic* is based on the number of retweets and votes on the tweet message. The *confirmation heuristic* consists of three features: link to the external source, user mention and hashtags. Lastly, the *reputation heuristic* describes the authors of the tweet messages. The *confirmation heuristic* is reported as a novel heuristic for credibility perceptions on Twitter. This study further suggests three categories of behaviour for readers'

credibility perceptions on Twitter: readers who overly depend on the confirmation heuristic, modestly rely on the cognitive heuristic, and readers who only slightly depend on the cognitive heuristics.

7.4 Limitations of the Study

There are several limitations within this study. The first limitation is the skewed gender distribution in our dataset. While conventional qualitative user studies allow researchers to control the gender balance in user studies, user studies on crowdsourcing platforms often leads to gender imbalance (Kang et al., 2015; Peer et al., 2016). The implications of this imbalance in credibility analysis are not clear and need to be further examined.

A second limitation is the fact that the study focussed on three news types only; breaking news, natural disaster and politics. Perhaps with more variety of news, as found in a traditional news layout, such as sports and entertainment news, we could identify if the credibility features would differ according to the seriousness of the news presented. Furthermore, only news information shared on Twitter was focussed on in this study and not news presented on other social media platforms.

Lastly, individual credibility features were not extensively explored in this study. For example, we controlled for the effect of information search by asking participants to perceive the credibility of tweets from a single tweet message on a particular topic. Also, only some general behaviours were analysed, which focus on the surface features of the tweet messages such as the writing style, number of retweets and likes, and author features. The method of data collection chosen may have limited the credibility judgement behaviour that we were

able to observe.

Despite these limitations, the results of this research contribute to our understanding of readers' credibility perceptions in terms of use of credibility features, the correlation between external factors and credibility perceptions, and the role of readers' personal characteristics in forming credibility perceptions on Twitter.

7.5 Future Work

This study has addressed a number of research questions, however there are still many opportunities for further research. Currently we analysed readers' credibility perceptions by analysing the feedback of the readers in our user study where the readers were given screenshots of tweet messages. It may also be useful to analyse credibility perceptions during a real-world event by combining the experiment with user information behaviour; for example, using eye tracking devices to collect user information behaviour data during the process of making credibility judgements. A reader could be asked to perform the information search and determine which tweet messages listed seem most credible to them and what is the feature that has the most influence in helping them determine this. Not only that, other information retrieval factors such as time spent on the result pages, search experience, and task completion time can also be studied in relation to credibility perceptions.

In a crisis situation such as natural disasters, social media like Twitter has been used to find information from local news agency and even experiences shared by witnesses. [Mendoza et al., 2010] showed in their study that Twitter activity is related to the significance of an event. With the right use of features related to the crisis, tweets will propagate faster and

longer on social media, getting a larger and wider audience. Therefore, suggestion for future research could be comparing the reader's satisfaction with information searches in finding credible information, specifically on crisis-related material. Based on readers' preferences, and establishing a dynamic credibility ranking system that incorporates readers' personal characteristics and demographics could improve the search return results of online information. Emergency responders and organisations can also establish a method to innovatively monitor and display crisis-related tweets for general readers, or personalised tweets for specific groups of readers. The system could then help to increase the utilisation of social media data for crisis response and management.

In this thesis, we focussed only on Twitter. Twitter was chosen as it is a prominent news sharing platform and due to the limited number of characters allowed on the platform (screenshots are easily read), and data are accessible for research through the Twitter API. It would be interesting to apply similar methodologies to other social media platforms; for example, on Facebook the credibility features would be quite different from those on Twitter. There are similarities and differences between these social media services. The readers of social media platforms may also be of different groupings and backgrounds. Between-subject user study designs can help to identify and compare the different credibility perceptions among readers of different social media platforms. Understanding reader perceptions of different social media platforms can also help to develop better credibility assessment tools.

References

Adali, S. and Golbeck, J. (2014). Predicting personality with social behavior: a comparative study. *Social Network Analysis and Mining*, 4(1):1–20.

Agrawal, R., Imieliński, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, 22(2):207–216.

Ahmad, R., Wang, J., Hercegfi, K., and Komlodi, A. (2011). Different people different styles: Impact of personality style in web sites credibility judgement. In *Symposium on Human Interface*, pages 521–527. Springer Berlin Heidelberg.

Aladhadh, S., Zhang, X., and Sanderson, M. (2014). Tweet author location impacts on tweet credibility. In *Proceedings of the 2014 Australasian Document Computing Symposium*, page 73. ACM.

Allen, I. E. and Seaman, C. A. (2007). Likert scales and data analyses. *Quality progress*, 40(7):64–65.

Alowibdi, J. S., Buy, U. A., Yu, P. S., and Stenneth, L. (2014). Detecting deception in online social networks. pages 383–390, Beijing, China. IEEE.

AlRubaian, M., Al-Qurishi, M., Al-Rakhami, M., Rahman, S. M. M., and Alamri, A. (2015).

A multistage credibility analysis model for microblogs. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1434–1440.

Armstrong, C. L. and McAdams, M. J. (2009). Blogs of information: How gender cues and individual motivations influence perceptions of credibility. *Journal of Computer-Mediated Communication*, 14(3):435–456.

Ashrafi, M. Z., Taniar, D., and Smith, K. (2004). A new approach of eliminating redundant association rules. In *International Conference on Database and Expert Systems Applications*, pages 465–474. Springer Berlin Heidelberg.

Bachrach, Y., Ceppi, S., Kash, I. A., Key, P., Radlinski, F., Porat, E., Armstrong, M., and Sharma, V. (2014). Building a personalized tourist attraction recommender system using crowdsourcing. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1631–1632. International Foundation for Autonomous Agents and Multiagent Systems.

Bartholomew, D., Knott, M., and Moustaki, I. (2011). Latent variable models and factor analysis. John Wiley & Sons, Ltd.

Batey, M., Furnham, A., and Safiullina, X. (2010). Intelligence, general knowledge and personality as predictors of creativity. *Learning and individual differences*, 20(5):532–535.

Behrend, T. S., Sharek, D. J., Meade, A. W., and Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior research methods*, 43(3):800.

Boyd, D. M. and Ellison, N. B. (2007). Social network sites: Definition, history, and schol-

arship. *Journal of Computer-Mediated Communication*, 13(1):210–230.

Brin, S., Motwani, R., and Silverstein, C. (1997). Beyond market baskets: Generalizing association rules to correlations. In *ACM SIGMOD Record*, volume 26:2, pages 265–276. ACM.

Broersma, M. and Graham, T. (2013). Twitter as a news source: How dutch and british newspapers used tweets in their news coverage, 2007–2011. *Journalism Practice*, 7(4):446–464.

Bruno, N. (2011). Tweet first, verify later: How real-time information is changing the coverage. In *Reuters Institute for the Study of Journalism*. University of Oxford.

Canini, K. R., Suh, B., and Pirolli, P. L. (2011). Finding credible information sources in social networks based on content and social structure. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 1–8. IEEE.

Cassidy, W. P. (2007). Online news credibility: An examination of the perceptions of newspaper journalists. *Journal of Computer–Mediated Communication*, 12(2):478–498.

Castillo, C., Mendoza, M., and Poblete, B. (2011). Information credibility on twitter. In *Proceedings of the 20th International Conference on World Wide Web*, pages 675–684.

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

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences.* Springer Science & Business Media, first edition.

Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, P. K. (2010). Measuring user influence

in twitter: The million follower fallacy. *ICWSM*, 10(10-17):30.

Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal OLD*, 12(2).

Colton, J. A. and Bower, K. M. (2002). Some misconceptions about r2. *International Society of Six Sigma Professionals*, EXTRAOrdinary Sense(3(2)):20–22.

Cozzens, M. D. and Contractor, N. S. (1987). The effect of conflicting information on media skepticism. *Communication Research*, 14(4):437–451.

Davidov, D., Tsur, O., and Rappoport, A. (2010). Semi-supervised recognition of sarcastic sentences in twitter and amazon. In *Proceedings of the fourteenth conference on computational natural language learning*, pages 107–116. Association for Computational Linguistics.

Donnellan, M. B., Conger, K. J., McAdams, K. K., and Neppl, T. K. (2009). Personal characteristics and resilience to economic hardship and its consequences: Conceptual issues and empirical illustrations. *Journal of Personality*, 77(6):1645–1676.

Dwyer, C., Hiltz, S. R., and Passerini, K. (2007). Trust and privacy concern within social networking sites: A comparison of facebook and MySpace. In *Proceedings of AMCIS*, volume 2007.

Erdur-Baker, Ö. (2010). Cyberbullying and its correlation to traditional bullying, gender and frequent and risky usage of internet-mediated communication tools. *New media & society*, 12(1):109–125.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in cognitive sciences*, 7(10):454–459.

REFERENCES

Evans, J. S. B. and Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, 11(4):382–389.

Eysenbach, G. and Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? qualitative study using focus groups, usability tests, and in-depth interviews. *Bmj*, 324(7337):573–577.

Eysenck, H. (1967). *The biological basis of personality*, volume 689. Transaction publishers.

Feist, G. J. and Gorman, M. E. (2012). *Handbook of the psychology of science.* Springer Publishing Company.

Feitosa, J., Joseph, D. L., and Newman, D. A. (2015). Crowdsourcing and personality measurement equivalence: A warning about countries whose primary language is not english. *Personality and Individual Differences*, 75:47–52.

Flanagin, A. J. and Metzger, M. J. (2000). Perceptions of internet information credibility. *Journalism & Mass Communication Quarterly*, vol. 77 no. 3(September 2000):515–540.

Fogg, B. J. (2003). Prominence-interpretation theory: Explaining how people assess credibility online. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '03, pages 722–723, New York, NY, USA. ACM.

Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., and Treinen, M. (2001). What makes web sites credible?: a report on a large quantitative study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 61–68. ACM.

Fornell, C. and Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1):39–50.

Gadiraju, U., Möller, S., Nöllenburg, M., Saupe, D., Egger-Lampl, S., Archambault, D., and Fisher, B. (2017). Crowdsourcing versus the laboratory: Towards human-centered experiments using the crowd. In Archambault D., Purchase H., H. T., editor, *Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments*, volume 10264 of *Lecture Notes in Computer Science*, pages 6–26. Springer, Cham.

Goldberg, L. R. (1990). An alternative "description of personality": the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216.

Gosling, S. D., Rentfrow, P. J., and Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528.

Greer, J. D. and Gosen, J. D. (2002). How much is too much?: Assessing levels of digital alteration of factors in public perception of news media credibility. *Visual Communication Quarterly*, 9(3):4–13.

Gupta, A. and Kumaraguru, P. (2012). Credibility ranking of tweets during high impact events. In *Proceedings of the 1st Workshop on Privacy and Security in Online Social Media*, PSOSM '12, pages 2:2–2:8, New York, NY, USA. ACM.

Gupta, A., Kumaraguru, P., Castillo, C., and Meier, P. (2014). TweetCred: Real-Time Credibility Assessment of Content on Twitter. In Aiello, L. M. and McFarland, D., editors, *Social Informatics: 6th International Conference, SocInfo 2014, Barcelona, Spain, November 11-13, 2014. Proceedings*, pages 228–243. Springer International Publishing, Cham.

Gupta, M., Zhao, P., and Han, J. (2012). Evaluating event credibility on twitter. In *SDM*, pages 153–164.

REFERENCES

Hahsler, M., Grün, B., and Hornik, K. (2007). Introduction to arules – mining association rules and frequent item sets. In *SIGKDD Explor*, volume 2, pages 1–28.

Halder, S., Roy, A., and Chakraborty, P. (2010). The influence of personality traits on information seeking behaviour of students. *Malaysian Journal of Library & Information Science*, 15(1):41–53.

Halko, S. and Kientz, J. A. (2010). Personality and persuasive technology: An exploratory study on health-promoting mobile applications. In *Persuasive Technology*, pages 150–161. Springer Berlin Heidelberg.

Härtl, K., Engel, J., Herschbach, P., Reinecker, H., Sommer, H., and Friese, K. (2010). Personality traits and psychosocial stress: quality of life over 2 years following breast cancer diagnosis and psychological impact factors. *Psycho-oncology*, 19(2):160.

Hovland, C. I. and Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public opinion quarterly*, 15(4):635–650.

Hsieh, H.-F. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9):1277–1288.

Hu, M., Liu, S., Wei, F., Wu, Y., Stasko, J., and Ma, K.-L. (2012). Breaking news on twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2751–2754. ACM.

Hughes, A. L., Palen, L., Sutton, J., Liu, S. B., and Vieweg, S. (2008). "sight-seeing" in disaster: an examination of on-line social convergence.

Husin, H. S., Thom, J. A., and Zhang, X. (2013). News recommendation based on web usage and web content mining. In *ICDE Workshops 2013*, pages 326–329, Brisbane,

REFERENCES

Australia. IEEE.

Jackson, K. M. and Trochim, W. M. (2002). Concept mapping as an alternative approach for the analysis of open-ended survey responses. *Organizational Research Methods*, 5(4):307–336.

Jiang, B., Sha, Y., and Wang, L. (2015). Predicting user mention behavior in social networks. In *National CCF Conference on Natural Language Processing and Chinese Computing*, pages 146–158. Springer.

Jin, F., Dougherty, E., Saraf, P., Cao, Y., and Ramakrishnan, N. (2013). Epidemiological modeling of news and rumors on twitter. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis*, page 8. ACM.

Johnson, D. and Gardner, J. (2010). Personality, motivation and video games. In *Proceedings of the 22Nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, OZCHI '10, pages 276–279, New York, NY, USA. ACM.

Johnson, K. A. (2011). The effect of twitter posts on student's perceptions of instructor credibility. *Learning, Media and Technology*, 36(1):21–38.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1):141–151.

Kang, B., Höllerer, T., and O'Donovan, J. (2015). Believe it or not? analyzing information credibility in microblogs. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 611–616. ACM Press.

REFERENCES

Kang, B., O'Donovan, J., and Höllerer, T. (2012). Modeling topic specific credibility on twitter. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces*, pages 179–188. ACM.

Kaye, B. K. and Johnson, T. J. (2011). Hot diggity blog: A cluster analysis examining motivations and other factors for why people judge different types of blogs as credible. *Mass Communication and Society*, 14(2):236–263.

Kelly, D. et al. (2009). Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends® in Information Retrieval*, 3(1–2):1–224.

Kline, P. (2014). *An easy guide to factor analysis*. Routledge.

Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., and Graepel, T. (2014). Manifestations of user personality in website choice and behaviour on online social networks. *Machine learning*, 95(3):357–380.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Sage, 2nd edition.

Kwak, H., Lee, C., Park, H., and Moon, S. (2010). What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600.

Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of communication*, 50(1):46–70.

Lauriola, M. and Levin, I. P. (2001). Personality traits and risky decision-making in a controlled experimental task: an exploratory study. *Personality and Individual Differences*, 31(2):215 – 226.

REFERENCES

LePine, J. A., Colquitt, J. A., and Erez, A. (2000). Adaptability to changing task contexts: Effects of general cognitive ability, conscientiousness, and openness to experience. *Personnel psychology*, 53(3):563–593.

Li, R. and Suh, A. (2015). Factors influencing information credibility on social media platforms: Evidence from facebook pages. *Procedia Computer Science*, 72:314–328.

Lim, S. (2009). How and why do college students use wikipedia? *Journal of the American Society for Information Science and Technology*, 60(11):2189–2202.

Lin, X., Spence, P. R., and Lachlan, K. A. (2016). Social media and credibility indicators: The effect of influence cues. *Computers in Human Behavior*, 63:264–271.

Lockton, D. (2012). Cognitive biases, heuristics and decision-making in design for behaviour change. Available at SSRN: https://ssrn.com/abstract=2124557.

Lucassen, T., Muilwijk, R., Noordzij, M. L., and Schraagen, J. M. (2013). Topic familiarity and information skills in online credibility evaluation. *Journal of the American Society for Information Science and Technology*, 64(2):254–264.

Lucassen, T. and Schraagen, J. M. (2011). Factual accuracy and trust in information: the role of expertise. *Journal of the American Society for Information Science and Technology*, 62(7):1232–1242.

Mayer, R. C., Davis, J. H., and Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review*, 20(3):709–734.

McCrae, R. R. and Costa, P. T. (2003). *Personality in adulthood: A five-factor theory perspective*. Guilford Press.

McCrindle, M., Wolfinger, E., and others (2010). Generations defined. *Ethos*, 18(1):8.

REFERENCES

McDonald, J. H. (2009). *Handbook of biological statistics*, volume 2. Sparky House Publishing Baltimore, MD, third ed. edition.

McFarland, L. A. and Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, 85(5):812.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. In *Biochemia Medica*, volume 22, pages 276–282.

McKnight, D. H. and Kacmar, C. J. (2007). Factors and effects of information credibility. In *Proceedings of the Ninth International Conference on Electronic Commerce*, ICEC '07, pages 423–432, New York, NY, USA. ACM.

Melnik, M. I. and Alm, J. (2002). Does a seller's ecommerce reputation matter? Evidence from eBay auctions. *The journal of industrial economics*, 50(3):337–349.

Mendoza, M., Poblete, B., and Castillo, C. (2010). Twitter under crisis: Can we trust what we RT? In *Proceedings of the first workshop on social media analytics*, pages 71–79. ACM.

Metzger, M. J., Flanagin, A. J., and Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, 60(3):413–439.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., and Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American psychologist*, 56(2):128.

Moore, S. R., Smith, R. E., and Gonzalez, R. (1997). Personality and judgment heuristics: Contextual and individual difference interactions in social judgment. *Personality and Social Psychology Bulletin*, 23(1):76–83.

143

REFERENCES

Morris, M. R., Counts, S., Roseway, A., Hoff, A., and Schwarz, J. (2012). Tweeting is believing?: understanding microblog credibility perceptions. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 441–450.

Moss, S., Prosser, H., Costello, H., Simpson, N., Patel, P., Rowe, S., Turner, S., and Hatton, C. (1998). Reliability and validity of the PAS-ADD checklist for detecting psychiatric disorders in adults with intellectual disability. *Journal of Intellectual Disability Research*, 42(2):173–183.

Musiał, K. and Kazienko, P. (2013). Social networks on the internet. *World Wide Web*, 16(1):31–72.

ODonovan, J., Kang, B., Meyer, G., Hollerer, T., and Adalii, S. (2012). Credibility in context: An analysis of feature distributions in twitter. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 293–301. IEEE.

OECD (2017). *Women in politics*. OECD Publishing, Paris.

OECD (2018). *Education at a Glance 2018*. OECD Publishing, Paris.

Petrovic, S., Osborne, M., McCreadie, R., Macdonald, C., Ounis, I., and Shrimpton, L. (2013). Can twitter replace newswire for breaking news? In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*.

Quijano-Sanchez, L., Recio-Garcia, J. A., and Diaz-Agudo, B. (2010). Personality and social trust in group recommendations. In *2010 22nd IEEE International Conference on Tools with Artificial Intelligence*, volume 2, pages 121–126. IEEE.

Rafalak, M., Abramczuk, K., and Wierzbicki, A. (2014). Incredible: Is (almost) all web

content trustworthy? analysis of psychological factors related to website credibility evaluation. In *Proceedings of the 23rd international conference on world wide web*, pages 1117–1122. ACM.

Rahbar, E. and Abdul Wahid, N. (2011). Investigation of green marketing tools' effect on consumers' purchase behavior. *Business strategy series*, 12(2):73–83.

Rassin, E. and Muris, P. (2005). Indecisiveness and the interpretation of ambiguous situations. *Personality and Individual Differences*, 39(7):1285–1291.

Rieh, S. Y. and Hilligoss, B. (2008). College students' credibility judgments in the information-seeking process. *Digital media, youth, and credibility*, pages 49–72.

Rieh, S. Y., Jeon, G. Y., Yang, J. Y., and Lampe, C. (2014). Audience-aware credibility: From understanding audience to establishing credible blogs. *Proc. of the International Conference on Web and Social Media (ICWSM)*.

Sakamoto, Y., Li, H., and Tanaka, Y. (2014). Rumors on social media during emergencies. `https://ssrn.com/abstract=2473601`. Accessed on 2015-09-13.

Scheffer, T. (2001). Finding association rules that trade support optimally against confidence. In De Raedt, L. and Siebes, A., editors, *Principles of Data Mining and Knowledge Discovery*, pages 424–435. Springer Berlin Heidelberg.

Schifferes, S., Newman, N., Thurman, N., Corney, D., Göker, A., and Martin, C. (2014). Identifying and verifying news through social media: Developing a user-centred tool for professional journalists. *Digital Journalism*, 2(3):406–418.

Sikdar, S., Kang, B., ODonovan, J., Höllerer, T., and Adah, S. (2013a). Understanding information credibility on twitter. In *2013 International Conference on Social Computing*,

REFERENCES

pages 19–24. IEEE.

Sikdar, S. K., Kang, B., ODonovan, J., Hollerer, T., and Adal, S. (2013b). Cutting through the noise: Defining ground truth in information credibility on twitter. *Human*, 2(3):151–167.

Sogunro, O. A. (2002). Selecting a quantitative or qualitative research methodology: An experience. *Educational Research Quarterly*, 26(1):3.

Song, L., Lai, J., and Li, J. (2016). Identifying factors affecting individual perceived credibility on sns. In *Proceedings of the The 3rd Multidisciplinary International Social Networks Conference on SocialInformatics 2016, Data Science 2016*, page 2. ACM.

Starbird, K., Maddock, J., Orand, M., Achterman, P., and Mason, R. M. (2014). Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing. In *iConference 2014 Proceedings*, pages 654–662. iSchools.

Sundar, S. S. (1999). Exploring receivers' criteria for perception of print and online news. *Journalism & Mass Communication Quarterly*, 76(2):373–386.

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. *Digital media, youth, and credibility*, 73100.

Sundar, S. S., Knobloch-Westerwick, S., and Hastall, M. R. (2007). News cues: Information scent and cognitive heuristics. *Journal of the Association for Information Science and Technology*, 58(3):366–378.

Tabachnick, B. G. and Fidell, L. S. (2014). *Using Multivariate Statistics*. Pearson Education Limited/Allyn & Bacon, Inc., 6 edition.

Tan, P.-N., Steinbach, M., and Kumar, V. (2005). Association analysis: basic concepts and

algorithms. *Introduction to data mining*, pages 327–414.

Tan, W.-K. and Yang, C.-Y. (2014). Internet applications use and personality. *Telematics and Informatics*, 31(1):27–38.

Thomson, R., Ito, N., Suda, H., Lin, F., Liu, Y., Hayasaka, R., Isochi, R., and Wang, Z. (2012). Trusting tweets: The fukushima disaster and information source credibility on twitter. In *Proceedings of the 9th International Conference on Information Systems for Crisis Response and Management*.

Tseng, S. and Fogg, B. (1999). Credibility and computing technology. *Communications of the ACM*, 42(5):39–44.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131.

Viera, A. J. and Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam Med*, 37(5):360–363.

Wahyuni, D. (2012). The research design maze: Understanding paradigms, cases, methods and methodologies. *Journal of Applied Management Accounting Research*, 10(1):69–80.

Wang, J.-L., Jackson, L. A., Wang, H.-Z., and Gaskin, J. (2015). Predicting social networking site (SNS) use: Personality, attitudes, motivation and internet self-efficacy. *Personality and Individual Differences*, 80:119–124.

Wang, Y. D. and Emurian, H. H. (2005). An overview of online trust: Concepts, elements, and implications. *Computers in human behavior*, 21(1):105–125.

Wathen, C. N. and Burkell, J. (2002). Believe it or not: Factors influencing credibility on the web. *Journal of the Association for Information Science and Technology*, 53(2):134–

144.

West, R. F., Meserve, R. J., and Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of personality and social psychology*, 103(3):506.

Westerman, D., Spence, P. R., and Van Der Heide, B. (2012). A social network as information: The effect of system generated reports of connectedness on credibility on twitter. *Computers in Human Behavior*, 28(1):199–206.

White, R. (2013). Beliefs and biases in web search. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 3–12. ACM.

Widhiarso, W. and Ravand, H. (2014). Estimating reliability coefficient for multidimensional measures: A pedagogical illustration. *Review of psychology*, 21(2):111–121.

Widiger, T. A. and Costa, P. T., editors (2013). *Personality disorders and the five-factor model of personality (3rd ed.).* American Psychological Association.

Wilson, C., Boe, B., Sala, A., Puttaswamy, K. P., and Zhao, B. Y. (2009). User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems*, pages 205–218.

Wright, S. P. (1992). Adjusted p-values for simultaneous inference. *Biometrics*, pages 1005–1013.

Yang, J., Counts, S., Morris, M. R., and Hoff, A. (2013). Microblog credibility perceptions: Comparing the usa and china. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW)*, pages 575–586. ACM.

Yang, K. C. (2007). Factors influencing internet users' perceived credibility of news-related

blogs in taiwan. *Telematics and Informatics*, 24(2):69–85.

Yin, C., Sun, Y., Fang, Y., and Lim, K. (2018). Exploring the dual-role of cognitive heuristics and the moderating effect of gender in microblog information credibility evaluation. *Information Technology & People*, 31(3):741–769.

Zboray, R. J. and Zboray, M. S. (1996). Political news and female readership in antebellum boston and its region. *Journalism History*, 22(1):2.

Zhuang, M., Toms, E. G., and Demartini, G. (2016). The relationship between user perception and user behaviour in interactive information retrieval evaluation. In *European Conference on Information Retrieval*, pages 293–305. Springer.

Zubiaga, A., Liakata, M., Procter, R. N., Bontcheva, K., and Tolmie, P. (2015). Towards detecting rumours in social media. Austin, Texas, USA.

Zuccon, G., Leelanupab, T., Whiting, S., Yilmaz, E., Jose, J. M., and Azzopardi, L. (2013). Crowdsourcing interactions: using crowdsourcing for evaluating interactive information retrieval systems. *Information retrieval*, 16(2):267–305.

Appendix A

List of News-Related Topics for Second User Study

Table A.1: News event-related topics

Topic	Description	Year	Trendiness
Al-Qaeda leader shot	Osama bin Laden was killed in Pakistan by US Navy Seal in Operation Neptune Spear	2011	Trending
Asylum seeker trade	Prime Minister Julia Gillard says an imminent deal with Malaysia to trade asylum seekers	2011	Not trending
Credit card scandal	Labor MP Craig Thomson alleged of misused a union credit card	2011	Not trending
Egyptian protests	Syria government attack on protesters at Deraa	2011	Trending
England riots	Series of riots in cities and towns across England that leads to looting, violence and destruction	2011	Trending
Hurricane Irene	Hurricane Irene hit US East Coast	2011	Trending
Japan disaster	Earthquake, tsunami hit Japan and nuclear emergency	2011	Trending
New South Wales election	NSW election result with a complete defeat of Kristina Keneally's ALP at the hands of Barry O'Farrell	2011	Note trending
New Zealand earthquake	65 people died in the earthquake that devastated Christchurch	2011	Not trending
Norway terror attack	Eight people died in a bombing in Oslo and 69 young people died on nearby Utoya island	2011	Not trending
Perth bushfires	Twin bushfires rage out of control in Perth	2011	Note trending
Queensland flood	Tropical cyclone Yasi hit the coast of north Queensland	2011	Trending
Royal wedding	Prince William and Catherine Middleton wedding	2011	Trending

Steve Jobs death	Apple co-founder died after long battle with cancer	2011	Not trending
US congresswoman shot	U.S. Representative Gabrielle Giffords and eighteen others were shot during a constituent meeting held in a supermarket parking lot	2011	Trending
Attack in Benghazi	US embassy at Benghazi attacked	2012	Not trending
Aurora theatre shooting	Aurora theatre shooting during Dark Knight preview	2012	Trending
Costa Concordia shipwreck	Italian cruise disaster resulting 32 deaths	2012	Not trending
Derecho storm	Derecho thunderstorm traveled from Indiana, across the Midwest, and into the Mid-Atlantic states. The storm caused 22 deaths and widespread damage across its 800-mile track	2012	Not trending
Egyptian protests	Egyptian protesters against the Muslim Brotherhood and Egyptian President Mohammed Morsi	2012	Trending
Hurricane Isaac	Hurricane Isaac, a tropical storm that slowly marched across the Atlantic toward the U.S. causing severe damage in the Caribbean and the U.S. Gulf Coast	2012	Trending
Hurricane Sandy	Hurricane Sandy, kills at least 117 people in the United States and 69 more in Canada and the Caribbean	2012	Trending
Myanmar election	Myanmar elections 2012: Aung San Suu Kyi claims victory	2012	Not trending
Pakistan avalanche	Avalanche at a Pakistani military base near the Siachen Glacier, killing 129 soldiers and 11 civilians	2012	Not trending

Queensland election	Labor government lost in the Queensland election after 14 years in power, and the Liberal National government took over	2012	Not trending
Sandy Hook shooting	Sandy Hook Elementary School shooting killing 26 people	2012	Trending
Social media campaign on Kony	A social media campaign to shine a light on Ugandan warlord Joseph Kony	2012	Trending
SOPA protest	Protest against two proposed laws in the United States Congress, the Stop Online Piracy Act (SOPA) and the PROTECT IP Act (PIPA)	2012	Trending
Typhoon Bopha	Typhoon Bopha typhoon hit the southern Philippines, setting off floods and landslides and killing at least 77 people	2012	Not trending
US election	The Democratic nominee, President Barack Obama, and his Vice President, Joe Biden, were elected to a second term	2012	Trending
Australia caught spying Indonesia	Indonesia summoned Australia's ambassador to give an explanation of reports about Australias spying activities	2013	Not trending
Australia's new Prime Minister	Australia's new Prime Minister, Tony Abbott, won the vote by a wide margin	2013	Not trending
Boston bombing	Bombsexplode near the finish line at the world's oldest and most prestigious marathon inBoston	2013	Trending
Colorado flood	Colorado unprecedented flash flood	2013	Not trending
Iran-US relationship	Iranian president takes steps to thaw relations with the west	2013	Not trending

Mexico double hurricane	Mexico was hit with hurricane Ingrid and Manuel where death toll rises to 123, crop lands damaged	2013	Not trending
Nairobi Mall shooting	Somalian militants terrorize luxury mall	2013	Not trending
Navy Yard shooting	Gunman and 12 victims killed in D.C. Navy Yard Shooting	2013	Not trending
NSA whistle-blower	Edward Snowden is the whistle-blower behind the NSA surveillance revelations	2013	Trending
Pakistan earthquake	Magnitude 7.7 Earthquake kills at least 327 in Pakistan	2013	Trending
Royal baby	The Duchess of Cambridge gives birth to a baby boy	2013	Trending
Snow at Middle East	Rare snowstorm in the Middle East	2013	Not trending
Train derailed	A Train in Quebec derails and explodes killing 47	2013	Not trending
Typhoon Haiyan	At least 10,000 people are thought to have died in the central Philippine province of Leyte after Typhoon Haiyan	2013	Trending
US government shutdown	US Government shutdown after congress failed to agree by late September 2013 on the budget for the fiscal year beginning October 1	2013	Trending
Afghanistan election	Afghanistan election crisis after one candidate demanded a halt to vote counting, suspended cooperation with election authorities and called for a UN commission to mediate the case	2014	Not trending
Al Jazeera journalists arrested	3 Al Jazeera journalists jailed In Egypt	2014	Trending

Flood at Afghanistan	Flood in Afghanistan kills 100 people	2014	Not trending
Gaza attacked	Israel military launched more than 6,000 air strikes on Gaza killing 2,251 people	2014	Trending
Hailstorm in Russia	Surprise hailstorm causes panic on Russian beach	2014	Not trending
Hillary Clinton said she's poor	Hillary Clinton, wife of former democratic U.S. President Bill Clinton shared that they went broke after leaving the White House	2014	Trending
MH17 shot down	MH 17 shot down in Ukraine, killing all on board	2014	Trending
MH370 missing	MH370 flight to China went missing	2014	Not trending
Mount Everest avalanche	Mount Everest deadliest avalanche kills 12	2014	Trending
Niagara falls during polar votex	Niagara falls was partially frozen during the polar votex	2014	Not trending
Nigerian girls abducted	63 abducted women and girls escape Boko Haram	2014	Trending
NSA double agent	Double agent' arrested in Germany for passing information on NSA inquiry to the US	2014	Not trending
Sewol ferry disaster	A passenger ferry sank off the southern coast of South Korea	2014	Trending
Thailand military coup	The Thai government has been overthrown in a military coup	2014	Not trending
Wildfire in US	San Diego County wildfires were a swarm of 20 wildfires that erupted during May 2014	2014	Trending

Appendix B

Approval Letters for Human Ethics Application

B.1 Approval Letter for ASEHAPP 47-13

This is an approval letter for a human ethics application to conduct a crowdsourcing-based experiment described in Chapter 3 Section 3.2 and Chapter 4 Section 4.2.3.

RMIT
UNIVERSITY

13th September 2013

Xiuzhen (Jenny) Zhang
Building 14 Level 9, Room 5
School of Computer Science & IT
RMIT University

Dear Jenny

ASEHAPP 47 – 13 <u>ZHANG-SHARIFF</u> Query-biased Credibility Ranking of Tweets

Thank you for submitting your amended application for review.

I am pleased to inform you that the CHEAN has approved your application for a period of **3 Months** from the date of this letter to **13th December 2013** and your research may now proceed.

The CHEAN would like to remind you that:

All data should be stored on University Network systems. These systems provide high levels of manageable security and data integrity, can provide secure remote access, are backed up on a regular basis and can provide Disaster Recover processes should a large scale incident occur. The use of portable devices such as CDs and memory sticks is valid for archiving; data transport where necessary and for some works in progress. The authoritative copy of all current data should reside on appropriate network systems; and the Principal Investigator is responsible for the retention and storage of the original data pertaining to the project for a minimum period of five years.

Annual reports are due during December for all research projects that have been approved by the College Human Ethics Advisory Network (CHEAN).

The necessary form can be found at: www.rmit.edu.au/staff/research/human-research-ethics

Yours faithfully,

Linda Jones
Chair, Science Engineering & Health
College Human Ethics Advisory Network

Cc   CHEAN Member: Susana Gavidia-Payne School of Health Sciences RMIT University
       Student Investigator/s: Shafiza Mohd Shariff School of Computer Science & IT RMIT University

B.2 Approval Letter for ASEHAPP 36-16

This is an approval letter for a human ethics application to conduct a crowdsourcing-based

experiment described in Chapter 5 Section 5.2.

RMIT University

Science Engineering and Health

College Human Ethics Advisory Network (CHEAN)

Plenty Road
Bundoora VIC 3083

PO Box 71
Bundoora VIC 3083
Australia

Tel. +61 3 9925 4620
Fax +61 3 9925 6506
● www.rmit.edu.au

3 June 2016

Associate Professor Xiuzhen Jenny Zhang
Building 14 Level 9 Room 5
School of Science
RMIT University

Dear Associate Professor Zhang,

ASEHAPP 36-16 Information credibility of Tweets

I am pleased to inform you that the CHEAN has approved your application for a period of **12 Months** from the date of this letter to **3 June2017** and your research may now proceed.

The CHEAN would like to remind you that:

All data should be stored on University Network systems. These systems provide high levels of manageable security and data integrity, can provide secure remote access, are backed up on a regular basis and can provide Disaster Recover processes should a large scale incident occur. The use of portable devices such as CDs and memory sticks is valid for archiving; data transport where necessary and for some works in progress. The authoritative copy of all current data should reside on appropriate network systems; and the Principal Investigator is responsible for the retention and storage of the original data pertaining to the project for a minimum period of five years.

Please Note: Annual reports are due on the anniversary of the commencement date for all research projects that have been approved by the CHEAN. Ongoing approval is conditional upon the submission of annual reports failure to provide an annual report may result in Ethics approval being withdrawn.

Final reports are due within six months of the project expiring or as soon as possible after your research project has concluded.

The annual/final reports forms can be found at:
www.rmit.edu.au/staff/research/human-research-ethics

Yours faithfully,

Associate Professor Barbara Polus
Chair, Science Engineering & Health
College Human Ethics Advisory Network

Cc   CHEAN Member: Dr Toh Yen Pang
      Student Investigator/s: Shafiza Mohd Shariff, School of Computer Science and Information Technology Other
      Investigator/s: Professor Mark Sanderson, School of Computer Science and Information Technology