IMPACT OF LOCATION ON SOCIAL MEDIA CREDIBILITY

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

SULIMAN ALADHADH

BSc (Computer Science), Qassim University, Saudi Arabia

MSc (Computer Science), RMIT University

Computer Science and Information Technology

School of Science

College of Science, Engineering, and Health

RMIT University

June, 2019

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; and, any editorial work, paid or unpaid, carried out by a third party is acknowledged.

Suliman Aladhadh
Computer Science and Information Technology
RMIT University
June 19, 2019

Acknowledgement

First and foremost, I would like to thank Allah the Almighty, for all the blessings he has bestowed upon me so that I could complete this thesis. Next, I would like to express my special thanks to my supervisor, Professor Mark Sanderson, for his guidance, support, and motivation throughout my study. It would not have been possible for me to complete this thesis without his kind supervision. I thank my second supervisor, Associate Professor Xiuzhen (Jenny) Zhang, for her useful feedback during my candidature.

My special thanks to my country, the Kingdom of Saudi Arabia, for giving me all that I need to succeed. Next, I would like to acknowledge and thank the support of Qassim University for providing me with the scholarship for my studies. I will never forget the support of the Embassy of the Kingdom of Saudi Arabia and the Saudi Arabian Cultural Mission (SACM) in Canberra while I was in Australia.

My deep gratitude goes to my family. To my parents Lolwah Alfuraih and Saleh Aladhadh, my wife Monirah Almousa, my little daughter and son Maitha and Saleh, and to all of my family members. Thank you for your endless prayers and emotional support that kept me motivated to see this work through to the end. Finally, to my younger brother Abdulaziz, your support for all that I needed during the years of absence from home is really appreciated. I would like to dedicate this thesis to my friends and to everyone has motivated me even with one kind word.

Credits

Portions of the material in this thesis have previously appeared in the following publications:

- **Aladhadh, S.**, Zhang, X. and Sanderson, M., 2014, November. Tweet author location impacts on tweet credibility. In *Proceedings of the 2014 Australasian Document Computing Cymposium, ADCS '14* (pp. 73-76). ACM. This work forms Chapter 3.

- **Aladhadh, S.**, Zhang, X. and Sanderson, M., 2017, September. Beyond the culture effect on credibility perception on microblogs. In *International Conference on Social Informatics, SocInfo '17* (pp. 316-328). Springer, Cham. This work forms Chapter 4.

- **Aladhadh, S.**, Zhang, X. and Sanderson, M., 2018. Location impact on source and Linguistic features for information credibility of social media. *Online Information Review.* This work forms Chapter 5.

This work was supported by Qassim University - Saudi Arabia.

The thesis was written in the TeXworks editor on Mac OS 10.13, and typeset using the LaTeX 2ε document preparation system.

Note

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

Abstract

Social media platforms such as Twitter and Facebook allow users from all over the world to contribute content. However, these users publish content without peer review, and contributions of low quality can create credibility concerns. This reduces the potential social benefits of social media. Social media credibility models rely on popularity, temporal patterns and other collective behaviour of users to study the credibility of user generated content (UGC). However, such approaches do not take into account end user credibility perceptions and factors that may influence a contributor (author), which in turn affects credibility models in social media. Therefore, I studied the factors that influence readers' credibility perception and content credibility. I identified a number of limitations in existing models: most research considers only users' perceptions from one country or culture and then generalises the results to others. I also found these models do not consider author location when assessing credibility. Therefore, I proposed a study on the influence of author, reader and event location on user credibility perception and content credibility in social media.

I propose a model that has been validated using a crowdsourced labelling approach. I ran three controlled experiments mainly varying source-based features (author) and content (text). Further, I applied a linguistic analysis approach to validate the influence of location on content credibility. I also applied a number of statistical analyses to measure the effect of all features. I validated the model using a common social media platform (Twitter) and showed the influence of non-textual features on credibility judgments of readers. Also, I found that reader location represented by culture can determine their credibility perception

in social media. Moreover, I showed how distance between the event and author location can affect sources and credibility distribution in social media.

Location of readers and authors, and the interaction with event locations can be used to improve assessment of credibility in social media. Reader characteristics are found to be important when studying credibility in social media as they can be used to improve user experience in social media. Moreover, an author's location can enhance credibility detection models to assess content accurately as it can differentiate between content with different credibility levels. While I do not claim that only user location can be used to build a standalone credibility system, I conclude that adding geographic location and culture of users can improve the performance of existing credibility models significantly.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Internet users increasingly spend more time on social media.[1] Social media allows users to communicate and share information at any time. In this environment, users are simultaneously information contributors and consumers.

Social media is an important source of information. 66% of Facebook users and 59% of Twitter users obtain news from social media sites [Gottfried and Shearer, 2016] and 69% of news sharing studies in social media used Twitter data because news form most of Twitter's content [KÃijmpel et al., 2015]. The growth of social media is fuelled by an increase in the number of users. Twitter was launched as a social network in early 2006, with approximately 20 million visitors every month in the first year. By the end of 2009, it had around 75 million accounts, with 2.5 million daily posts [Sharifi et al., 2010]. In 2013, the service had 554.7 million active users who posted 58 million tweets daily [Murthy, 2013]. In early 2011, Google began to incorporate social media status updates in its search engine results [Google Social Search, 2011]. In 2011, the Twitter search engine was receiving 1.6 billion daily queries [Siegler, 2011], reflective of the significance of socially relevant content and timely sources of information.

Many entities, such as governments and organisations, analyse social media to obtain

[1] http://www.nielsen.com/us/en/insights/reports/2017/2016-nielsen-social-media-report.html (accessed 2 July 2018)

insights into how people respond and behave. It has been estimated that the United States (US) government will have spent approximately $13 billion on big data analysis by 2021.[2] Other governments are also investing significantly in the analysis of big data to assist with decision making and this is likely to influence future marketing strategies. For example, it is predicted that the market for big data in the global business sector will be valued at $57 billion by 2020.[3]

Social media platforms can distribute information far more rapidly than can traditional sources. For example, the first alarm raised about the impending Japan earthquake was disseminated far more quickly via Twitter posts than by the national news agency [Sakaki et al., 2010]. Social media is an important source of information in times of crisis.

However, there is a potential risk that information communicated might be inaccurate or untrustworthy. For example, many tweets posted in connection with a 2010 earthquake in Chile claimed that a nearby volcano was active and that a tsunami would result, which was subsequently confirmed to be false. Pew Internet Research in 2012 anticipated that false information will result in 'a distribution of harms'.[4]

The absence of barriers of published information in social media results in the uncontrolled dissemination of misleading information and rumours. Thus, the amount of false content available on social media platforms continues to escalate, particularly as people increasingly rely on it to inform them about many significant issues including health, crises and breaking news. Further, social media users and younger generations are affected by inaccurate and harmful content, thus negating many social media benefits.

According to Twitter [Twitter, 2016], 77% of its users are based outside the US and speak more than 35 languages, indicating the diversity of social media users in terms of culture and location. Thus, a full evaluation of location effect on the perception of the credibility of social media content is needed. It has been proposed that the location of an author influences

[2] https://idccommunity.com/government/smart_government/government_spending_will_rise_for_both_big_data_and_mobile_solutions (accessed 5 July 2018)

[3] https://insidebigdata.com/2017/09/09/big-data-important-business/ (accessed 8 August 2018)

[4] http://www.pewinternet.org/2012/07/20/the-future-of-big-data/ (accessed 22 August 2018)

their user generated content (UGC) [Han et al., 2014], but no study has investigated the influence of author location on actual credibility in social media.

As social media has become an important source of information, it is vital to understand how social media readers determine the credibility of information sources. Determining credibility is challenging because credibility is a perception that is influenced by many factors [Fogg et al., 2001]. To date, most research into the credibility of social media content has considered readers and information sources as one group, regardless of the location of the viewers. However, culture and country of the reader have been found to influence reader perceptions of credibility [Yang et al., 2013]. Moreover, the location of the author is considered an important determinant of reader perceiption in relation to gauging the credibility of the information [Morris et al., 2012]; Thomson et al. [2012] demonstrated how the content of authors who included their location in their profile was perceived as more credible than that of those who did not.

Several improvements to existing models in the literature were identified in the current study regarding how to assess the credibility of information. Limitations in the literature were defined following a review thereof and comprised the motivation for this study. In this thesis, I explore the effect of features -particularly location- on the perceptions of readers about the credibility of social media. Location can be viewed from different standpoints:

1. Location of author. This refers to how the author's location affects reader perceptions of credibility; that is, the distance between the topic of the tweet and the location of the author.

2. Location of reader. This pertains to differences in how readers from different locations perceive the credibility of the same information.

3. Location of event. This relates to how distance from the place in which an event is occurring affects the credibility-related behaviour of the author.

Some limitations exist when working in this area. For example, most social media credibility research has used Twitter datasets as representative of social media content, and Twitter-related features like number of hashtags, links, retweets and mentions, as well as text-related features, to measure credibility. However, findings from these studies might be relevant only to the social networks used and cannot be transferred to other social networks. Another limitation is that there is a relationship between the type of data used (e.g. event or domain) and the credibility assessment model performance, and applying the same credibility model in different domains may give different accuracy results. These limitations exist for current social media credibility research. Moreover, social media data are biased in nature: for example, only two percent of Twitter authors generate 50% of Twitter content [Baeza-Yates and Sáez-Trumper, 2015]. However, despite these limitations, social media credibility research has produced models which achieved high accuracy for verifying the content credibility.

1.1 Thesis Aims

This thesis hypothesises that tweets coming from locations close to the place in which event is occurring will be judged more credible and as containing more valuable information than tweets from locations further away. It is also hypothesised that personal differences, such as culture, will affect the interpretation of the presented factors in terms of credibility, and that the ability to determine differences will therefore help with the adaptation of social media retrieval systems in relation to the cultural domain. Thus, the aims of the research were to determine which features help readers to determine the credibility of social media content.

1.2 Research Questions

Three main research questions were identified, each with a number of subquestions:

1. Which features affect readers' perceptions of credibility?

- How does tweet location affect credibility?

2. What is the effect of a readers' location on their perceptions about the credibility of social media information?

    - What is the effect of a reader's culture on their credibility perception in social media?
    - Will a reader's nationality affect their credibility perception of social media among other members of the same culture?
    - What is the effect of the geographical region of countries with the same culture on readers' credibility perception of social media?
    - How do different regions and divisions in large countries influence a reader's credibility perception of social media?

3. What is the effect of location on sources distribution and linguistic features for information credibility of social media?

    - What types of information sources are associated with which events, both inside and outside the country in which the event is taking place?
    - How do linguistic features differ among types of information sources, in terms of the extent of credibility, topic and location?
    - How effective is it to only use the information source to assess credibility?

1.3 Main Contribution

There are three main contributions of this thesis.

1. **Studying readers' perceptions of tweet credibility** I examined the effect of key elements of tweet information on credibility assessment. I identified challenges that readers face when judging credibility based on content alone; they rely heavily on these

features when determining content credibility. I performed an evaluation of different non-textual features and highlighted helpful pieces of information regarding credibility judgments.

2. **A systematic study of credibility perceptions in social media in relation to location with different categories including culture, country and regions** I was able to identify a number of factors that influence social media credibility, and investigated a variety of credibility assessments due to location differences. Previous studies did not include the influence of location on credibility judgments of readers. I discuss how this can maximise the reader experience of credibility assessment of social media content, and the implications on such a system of designing a search engine within and between cultures.

3. **An examination of types of information distributed in different types of events in social media, the linguistic features surrounding credibility, and the influence of location on these two aspects** An attempt was made to determine how the location of the information source affected the credibility of the generated content. Many crowdsourced tasks were run for a number of events to explore the effect of location on the distribution of content; that is, its source, informativeness and credibility. An extensive amount of data annotated by crowdsource workers were also used to study the influence of location on the linguistic features that relate to credibility. The results validate that the content of tweets with the same credibility level (high or low) can carry different linguistic features when generated from sources in different locations (near or far).

This thesis further provides **a review of current approaches to social media credibility research in a novel manner and connects them with factors that influence their performance**. Unlike previous studies, I classified most previous research based on the methodologies used: previous social media credibility surveys were categorised based on the studied domain and included only a few studies related to that domain. This helped in the comparison between the methodologies and identification of the power of each approach. Next, I identified the most important factors that influence the credibility assessment of users – authors and readers – in social media and that can affect the performance of the used methodologies for credibility assessment and create many problems such as the well-known issue of prediction overestimation in the accuracy of the trained model; this can happen when applying them in a domain different from the test domain.

1.4 Thesis Structure

The structure of the remaining thesis chapters is as follows.

Chapter 2 provides a literature survey of research related to this thesis. The chapter has three parts: the first part (Sections 2.2 to 2.4) includes the main concepts and general background about credibility models and moving from offline to online. The second part (Sections 2.5 to 2.6) reviews credibility models in social media, and includes research that explores the effect of culture, language and country on UGC. The third part of the literature review chapter (Section 2.7) presents work related to the distribution of information sources across different locations, and shows how popular sources generate most social media content.

Chapter 3 examines the effect of prominent features in credibility perception of readers, particularly the effect of author location on credibility judgments.

Chapter 4 documents the recruitment of a large number of participants representing different cultures, countries and regions. This chapter outlines the large number of judgments used to analyse the effect of readers' locations on their credibility perceptions.

Chapter 5 connects the source and type of generated content from different perspectives. The chapter analyses in depth the relationship between linguistic features, location and credibility. Further, it attempts to validate a new methodology for classifying social media credibility.

Chapter 6 concludes the thesis findings and provides recommendations for further research.

Finally, at the end of the thesis there are two appendices related to experimental aspects: Appendix A includes ethics approvals and plain language statements for all the experiments, followed by a glossary of acronyms in Appendix B.

Chapter 2

Literature Review

In this chapter, I review previous research related to this thesis. Section 2.1 defines credibility and explains its main components. It presents types of credibility related to computers. Section 2.2 describes early studies of credibility and outlines three types of credibility: source, media and message. Section 2.3 describes the effect of moving from offline to online credibility and the challenges that come with this transition. Section 2.4 presents many approaches and models that have been proposed for online information assessments.

The second part of this chapter (Section 2.5) focuses on credibility in social media. In Section 2.5.1, I present challenges in the assessment of information credibility of social media. Section 2.5.2 presents the main models proposed for assessing social media credibility, which fall into five categories. Section 2.6 describes the main factors that affect content generated by social media authors and readers, including culture, language and country.

Finally, Section 2.7 presents research carried out in relation to location effects on credibility and information sources. Section 2.7.1 reviews research on the effect of a author's location on their behaviour in social media. Section 2.7.2 presents research in the area of defining eyewitnesses in social media. Section 2.7.3 explores the sources included in previous research on social media. Section 2.7.4 reviews research into finding expert sources in such topics in social media.

2.1 Definition

Credibility is a multidimensional concept, commonly defined as believability. It has two main dimensions: trustworthiness and expertise [Hovland et al., 1953, Fogg and Tseng, 1999, Flanagin and Metzger, 2007, Rieh et al., 2014].

Credibility is the perception and judgment an individual holds and makes about information or the source of information. In other words, credibility is a property judged by the receiver of information; it is not a property of the source or information [Flanagin and Metzger, 2008, 2007, Lankes, 2008]. Fogg and Tseng [1999]] defined trustworthiness as perceiving 'goodness or morality of the source', while expertise involves 'perceiving knowledge and skill of the source'. Similarly, Danielson [2006] defined expertise as the perception of the ability of a source to provide information with validity and accuracy, while trustworthiness is the perception of the willingness of a source to provide information with accuracy, taking into consideration the source's expertise.

There are four types of computer credibility according to Fogg and Tseng [1999]: reputed credibility – where the perceiver believes something based on a third party report; presumed credibility – where general assumptions in the perceiver's mind makes the perceiver believe in something; surface credibility – where a perceivers believes something based on simple inspection; and experienced credibility – how much a perceiver believes in something based on first experience.

2.2 Historical Development of Credibility Research

Credibility was originally studied using more formal approaches in the twentieth century; for example by Hovland and colleagues in the communication field in 1951 [Hovland and Weiss, 1951, Hovland et al., 1953]. They began by examining the information source characteristics that influence message acceptance by the receiver. Much research followed that of Hovland and colleagues, with a focus on the credibility of newspaper, television and radio [Meyer, 1988, Newhagen and Nass, 1989]. In communication research, the main focus is on source and media. In this field, credibility is taken to be a characteristic of perception [Rieh and Danielson, 2007]. In the communication field there are three types of credibility [Metzger et al., 2003, Rieh, 2010, Rieh and Danielson, 2007]: message, media and source.

Message credibility refers to how the attributes of a message, such as language and structure, influence a recipient's credibility perception (i.e. will they believe the message). Although the effects of source and message characteristics are linked, when a reader has little information about the source, the message factors begin to have a strong influence on reader credibility perception [Petty and Cacioppo, 1986]. Credibility research has classified message dimensions into three types: structure, content and presentation [Metzger et al., 2003].

Media credibility refers to the credibility of the channel that transfers the message, including TV, radio, newspaper, magazine and Internet. Media credibility mainly refers to which media type a users will use in cases where they receive conflicting stories in different media. People tend to believe stories that come from their frequently used and preferred media [Roper, 1985].

The last type of credibility in the communications field is source credibility. Many researchers have investigated source credibility from the viewpoint of recipients and identified the many factors that influence recipient trustworthiness and expertise. Wilson and Sherrell [1993] found that in more than 100 credibility research of sources, expertise was the most powerful factor in a recipient's persuasion. They attributed the importance of source expertise to its objective (experience, educational level, etc.) and the ease of its evaluation, compared with other source dimensions. Next, i briefly show the challenges came with online credibility assessment.

2.3 Moving from Offline to Online Credibility

There is a long history of assessing credibility of information in different research areas. However, credibility research also refers to different forms of information 'offline'. People have always had problems trusting information generated via different kinds of media, even before the existence of the Internet. In both forms (i.e. online and offline), people need cognitive skills to evaluate the information [Flanagin and Metzger, 2008].

The amount of information distributed online is massive and many people use the web as their primary information source. Many intermediaries have been omitted from this new method of information distribution [Eysenbach, 2008]. Digital media has moved from professional to individual consumer information evaluation [Flanagin and Metzger, 2008]. Thus, credibility assessment in the online environment includes many more challenges than does the 'old' media. The first issue is that there are no standards for the quality of online information, manipulating information is easy and this can interact with reader perceptions of credibility evaluation [Metzger and Flanagin, 2013]. The second issue is source multiplicity. A number of existing layers are embedded in the sources for distribution of online content [Sundar, 2008] and these sources with their related information can cause confusion among information seekers when they perform online searches [Eysenbach and Köhler, 2002]. Other issues, such as ambiguity of source and context are well explained in the literature [Flanagin and Metzger, 2008]. Next, i start showing the main theories in web credibility assessment.

2.4 Online Credibility Assessment Approaches and Theories

This section describes the main models that have been used to study online information credibility. One of the most well-known theories in online credibility, the prominence-interpretation theory, authored by [Fogg, 2003], was described as:

prominence $\times$ interpretation $=$ credibility impact.

This theory explores the overall credibility of a website in two phases: first, when a reader notices the website elements; and second, as the readers' interpretation of the elements they noticed. The theory authors defined prominence as 'an element's likelihood of being noticed when readers assess credibility'. Thus, the elements noticed by readers come into prominence and factors that have not been noticed have no impact on credibility assessment. The prominent phase includes five factors that influence reader notice:

1. reader involvement (ability of reader to inspect website content)
2. topic and content of the web page (politics, health, entertainment, etc.)
3. type of task handled by user (seeking information and other things)
4. reader's experience, such as expert v. beginner
5. individual differences such as education level.

All these factors have an effect on the first component of prominence-interpretation theory. The second component is interpretation, which refers to the reader's assessment process for each element. Three factors influence reader interpretation:

1. readers' assumptions in their mind (e.g. culture and previous experience).
2. knowledge and skills.
3. context (e.g. environment and user expectations).

At the beginning, readers quickly inspect and evaluate elements. This is an iterative process with the readers repeatedly noticing and compiling different elements until they are satisfied with the final conclusion.

A theoretical model for assessing online information in regard to authority was proposed by [Fritch and Cromwell, 2001]. This model provides a set of criteria that readers need to consider when evaluating the authority of Internet information. The model inputs are sources' competence and trustworthiness, document validity and overt-covert affiliation. The inputs are classified into four classes to be evaluated: source, document, institution and affiliation. Each class has its own assessments to evaluate information authority; these assessments are then combined to give an overall cognitive authority for each piece of information. This is an iterative process and the overall cognitive authority can be made more accurate over time.

Dual-process approaches of credibility assessment based on theories of information processing have found that people exhibit large variance in terms of the effort spent evaluating online information credibility, depending on the type of information they are seeking [Metzger, 2007, Taraborelli, 2008]. Thus, this model tells us how people make an effort to assess credibility. When an individual has the motivation and ability to use the available cues and other website features with low effort, this is heuristic processing. In contrast, systematic processing is high effort because it involves analysing the content cues of the message.

The best examples of dual-processing computer credibility assessment are the prototypical models proposed by [Fogg and Tseng, 1999]. They suggested three models: binary evaluation where readers perceive the product as credible or not credible; threshold evaluation where there are upper and lower limits with many levels of perceiving credibility; and spectral evaluation where there are no categories for evaluating credibility but instead there are shade of grey. Choice of model depends on the degree of reader involvement and the ability for online information evaluation.

Rieh [2002] proposed a model that suggests that judgments of information, cognitive authority and quality take place in the information retrieval process; in the web environment this happens between user and information object. In her model, Rieh used Hogarth [1987] judgment and decision-making theory. The model consists of two steps: predictive and evaluation judgment. The predictive judgment refers to a user's expectation of what will happen, whereas evaluative judgment refers to the process by which users express preferences. The study found that knowledge has the greatest influence on user predictive judgment and that source characteristics influence both types of judgment (predictive and evaluative) to assess web cognitive authority.

[Hilligoss and Rieh, 2008] proposed a framework for online credibility assessment for different media and for variety of information seeking tasks. Three different levels of credibility judgment were included in this model: construct, heuristics and interaction. The construct level refers to how a person defines and conceptualises credibility, while heuristics refers to using a 'rule of thumb' to make credibility judgments; it is applicable in many situations. The third level, interaction, refers to use of content and the available cues of source and information for credibility judgments. This study is said to be a framework for assessing online credibility in different contexts.

In this section, i explained the basic theories of online credibility assessment. Next, i start exploring the social media credibility research, which is the main topic of this thesis.

2.5 Credibility in Social Media

The previous section described credibility models for the traditional web, where users begin to interact and exchange information through social media platforms and become the generators of information; that is, user generated content UGC in virtual environments [Kaplan and Haenlein, 2010]. Sharing of personal knowledge with other people has become easier, as people tend to exchange knowledge even when there is no explicit benefit to themselves, as psychology research has shown [Warneken and Tomasello, 2009]. With this new way of interacting, credibility assessment differs from that used for the traditional web. The information credibility evaluation process in social media deals with content, and the characteristics of authors of the content. [Moens et al., 2014]. This means that credibility is connected to UGC and sources, as well as the social relationship between the included entities.

In the following two sections I first present the challenges facing credibility assessment in social media, and then describe existing models for studying credibility in social media.

2.5.1 Challenges in Social Media Credibility Assessment

Evaluating the credibility of online content is challenging, social media credibility assessment has proven more difficult than any previous forms of online content assessment because:

1. Credibility is subjective and varies between people depending on their situation (e.g. demographics, culture, expertise).

2. When readers use social media, authoritativeness and expertise considered to be not important, while other factors include reliability, accuracy, trustworthiness, fairness and completeness were considered to be important [Rieh et al., 2014].

3. The traditional well-known credible sources – such as news agencies, TV and news websites – have changed with the rise of social media. Users of social media themselves become information sources for many events. Differentiation of credibility levels of the millions of daily posts in social media is critical.

4. Social media content is normally disconnected from the original source.

5. In UGC, the challenge is how to define which data really matter for use in retrieval systems to provide the most credible results, and at the same time, to be perceived as credible to the reader.

6. The involvement of multicultural authors and readers from throughout the world [1] leads to different behaviour when using and assessing information of social media [Yang et al., 2013].

7. Context influences social media credibility assessment. For example, assessing news credibility on Twitter [Castillo et al., 2011] will differ from assessing health information [Lederman et al., 2014]: each context has its own criteria and characteristics, which makes it difficult to transfer findings between these contexts.

[1] https://about.twitter.com/company

2.5.2 Credibility Models in Social Media

Establishing author credibility is the most important part of building credibility models in social media, as authors in this case are the main generators of content. Some social media platforms explicitly indicate the credibility of an author. For example, Twitter has a verification sign next to verified accounts, as does Facebook. The verification method makes readers trust verified authors, because the verified authors become authentic sources of information. However, relying on the verification mechanism alone for credibility assessment is not enough for many reasons:

1. Social media platforms cannot authenticate all their users. For example, only $\sim$ 190,000 Twitter's users are verified [Navarra, 2016]. Thus, most social media authors are not verified.

2. These accounts are social media platform dependent.

3. There are many credible authors, but not verified.

Much research in the area of social media credibility has focused on diverse aspects of credibility. Researchers have proposed different models to assess credibility using different methods. In the following section, I classify previous research into five categories: credibility prediction; credibility with similarity; ranking based on credibility; linguistics features associated with credibility; and credibility perception.

2.5.2.1 Credibility Prediction

Many researchers have implemented automated credibility assessment of social media information. One of the early studies in microblog credibility analysed tweet content to predict credibility [Castillo et al., 2011]. The authors hypothesised that social media includes many signals that can help to verify the credibility of information. For the purposes of their experiment, they collected tweets of more than 2,500 topics, each topic has 10000 tweets at

Table 2.1: Examples of Twitter feature categories.

Scope	Feature	Description
Message	LENGTH CHARACTERS	Length of the text of the tweet, in characters
	LENGTH WORDS	... in number of words
	NUMBER OF URLS	Number of URLs contained on a tweet
	CONTAINS USER MENTION	Mentions a user: e.g. @cnnbrk
	IS RETWEET	Is a retweet: contains 'RT'
User	REGISTRATION AGE	The time passed since the author registered his/her account, in days
	COUNT FOLLOWERS	Number of people following this author at posting time
	HAS DESCRIPTION	... a non-empty 'bio' at posting time
	HAS URL	... a non-empty homepage URL at posting time
Topic	Number of tweets	Number of tweets
	AVERAGE LENGTH	Average length of a tweet
	AVERAGE SENTIMENT SCORE	The average sentiment score of tweets
	FRACTION SENTIMENT POSITIVE	The fraction of tweets with a positive score
	FRACTION SENTIMENT NEGATIVE	... with a negative score
	FRACTION POPULAR DOMAIN TOP 100	The fraction of tweets with a URL in one of the top-100 domains
	FRACTION POPULAR DOMAIN TOP 1000	... in one of the top-1,000 domains
	FRACTION POPULAR DOMAIN TOP 10000	... in one of the top-10,000 domains
Propagation	PROPAGATION INITIAL TWEETS	The degree of the root in a propagation tree
	PROPAGATION MAX SUBTREE	The total number of tweets in the largest subtree of the root, plus one
	PROPAGATION MAX | AVG DEGREE	The maximum and average degree of a node that is not the root (2 feat.)
	PROPAGATION MAX | AVG DEPTH	The depth of a propagation tree (0 = empty tree, 1 = only initial tweets,

most. They kept only the news topics, which numbered around 700 news. To build ground truth of credibility, the researchers annotated large number of tweets. They presented each topic with 10 related tweets and asked evaluators to choose the relevant credibility label. At the end, there were two credibility classes: credible and not credible, each class includes around 300 topic.

Next, they placed a set of features into four main categories: message; propagation; topic and user-based. Table.2.1 shows some instances of what each category includes. The features were used to train a supervised classifier to predict the credibility level of each topic. Among many classifiers, J48 decision trees achieved the best results with 86% accuracy. Furthermore, they tested how each individual subset of the four categories' features performed in the credibility task. The features that achieved the best results for the decision tree classifier were the propagation category features, with 75% accuracy. This was while message and user features were the lowest accuracy.

This work is said to have presented the first complete automated credibility model designed in social media and formed the basis for further work in this area. However, this work was performed on the news topic credibility in Twitter, they did not consider the individual tweet's credibility.

In spam-detection research, Benevenuto et al. [2010] tried to detect spammers in Twitter. To do that, they studied features of tweet content and behaviour of spammers to help differentiate between spammers and non-spammers. They labelled 1,065 users (355 spammers and 710 non-spammers) in relation to three different topics. They selected two types of features: content features (number of tweets with URL, number of tweets containing spam words and number of hashtags in a tweet); and user behaviour (including 23 features such as account age). The researchers used a support vector machine (SVM) classifier and were able to classify 96.0% of non-spammers and 70% of spammers. Based on these features, they concluded that spammers exhibit several characteristics: (1) They follow a large number of users but have only a few followers; (2) Their accounts are relatively new because they are blocked by other users and reported as spam; and (3) Some sammers do not follow any other users. However, the work focused on spammer detection, which is known to have special behaviour as in the three above characteristics. Credibility assessment is different from spam detection. In credibility assessments, the same author can generate credible and non-credible content, while spammers are always generating spam.

Castillo et al. [2011]'s work has been extended Kang et al. [2012] by proposing three models for predicting credibility in social media. Each model uses a different strategy, unlike that in Castillo et al. [2011]'s work, which used all features in one model. The first model is a social one, which computes a credibility score based on the social network of a user. The second model is content based and is mainly based on linguistic features in addition to other content features, such as retweet. The last model is a hybrid approach that is a combination of the first two models. To create ground truth of credibility, they collected a Twitter dataset related to seven events. Then, four groups of tweets were presented to crowd source participants who were asked to rate each tweet using a five-point Likert scale. Each group included 10 tweets: Group1 contained only tweet text; Group2 included tweets with low-credibility features for the user, such as very low number of followers; Group3 had tweets with high-credibility indicators such as large numbers of followers; and Group 4 included tweets with true information about the user. In total, 145 participants completed the study. The researchers performed an analysis of variance (ANOVA) and found significant differences among group ratings. The Waikato Environment for Knowledge Analysis (WEKA) machine learning toolkit was then used to evaluate each proposed system's performance in credibility prediction. From among a number of algorithms that were used, they chose J48 as it performed well in comparison to others. They separated the training from the test dataset and 10-fold cross validation was applied. They found the social model achieved the best performance with 88.2% accuracy, followed by the feature hybrid model, with 67.0%. Although this work helped to measure the effect of different features and models for predicting credibility, it did not propose any new features.

[Yang et al., 2012] proposed a model to detect false information in Weibo (a Chinese microblog platform). To complete the classification task they used the same four categories of features introduced by [Castillo et al., 2011] as well as client and location-based features. The client feature refers to the program used to post information while the location feature refers to the place mentioned in the rumours (domestic or abroad). They collected 5,000 posts related to topics identified as rumours and these posts included both true and false information. They trained an SVM classifier for three categories of features and the accuracy results were 72.6% for content features, 72.6% for user features and 72.3% for propagation features. When they added the new features (client and location) for each classifier, the accuracy improved by around 5.0% for each classifier.

[Xia et al., 2012] built a supervised model to classify tweets based on credibility in a crisis. They collected 350 tweets related to an event called 'UK Riots' and five experts annotated the tweets. They found that 52.0% of tweets were credible, 30.0% were not credible and the remainder were ambiguous; only the first two categories of tweets were considered further. To build their classifier, these authors used the features introduced by [Castillo et al., 2011]. WEKA was used to implement the classification. They used a number of algorithms including SVM, Bayesian network, hill climbing and K2. The SVM was the best, with 66.8% accuracy. However, this work was limited to a single event and a small number of tweets.

In the same area of credibility detection in social media, Liu et al. [2015] proposed a real-time rumour detection model. They used the same features as in [Yang et al., 2012], in addition to two new features: 1- belief features, which were used to classify authors into support or deny of an event based on their tweets language; 2- feature aggregation. These researchers collected all tweets related to 421 false events (identified via rumour-tracking websites) and 421 true events. Three classifiers were used – SVM, decision tree and random forest – and among them, SVM was the best performing with 90% accuracy. They found that their system that combined the two types of features (belief and aggregation) achieved a better result than both [Castillo et al., 2011] and [Yang et al., 2012].

2.5.2.2 Credibility with Similarity

A study of credibility in Twitter based on similarity with an authentic source was conducted by [Al-Eidan et al., 2010, Al-Khalifa and Al-Eidan, 2011]. They proposed a model to assess tweets credibility based on the similarities between a tweet and other credible sources outside Twitter. They followed the approach proposed by [Juffinger et al., 2009], which was applied to blog credibility. Al-Khalifa and Al-Eidan [2011] used a set of features to characterise tweets and author profiles with values assigned for five features: similarity of content with a verified source; ratio of inappropriate words (manually compiled from their dataset); URL of the authentic source; verified authors; and author' overall degree (they assign score for each author using service 'TwitterGrader.com'). Then, all tweets were ranked according to their weights into one of three credibility levels: high, average and low. They found that using similarity with the verified source and extra features mentioned above helped the system to obtain better results than similarity alone. However, there are many limitations to this work, the main limitation is using similarity as a predictor of credibility, tweets are UGC and thus will likely not be similar to other verified sources. Also, the research was based on a small amount of data, which may affect the results.

2.5.2.3 Ranking Based on Credibility

Ranking tweets based on credibility is another research area. Gupta and Kumaraguru [2012] used an information retrieval model to determine credibility information in Twitter. Tweets for 14 trend events with their queries were collected from which 500 random English tweets for each event were selected for human annotation into the following classes: definitely credible; seems credible; and definitively incredible. The dataset obtained from the annotation process included around 7,000 tweets. The SVM $_{\text{ranking}}$ algorithm trains a ranking SVM on the training dataset. Based on the trained model, the ranking score of credibility was predicted by the algorithm. They characterised the features which used for computing credibility score into content features (e.g. words, hashtags, retweets) and source features (e.g. number of followers, account age). Next, they used pseudo relevance feedback (PRF), which is an information retrieval technique for enhancing ranking results. In the results of the annotated data, only 17.0% of tweets were credible. For their model evaluation, they used the top 25 tweets ranked by their system, and the baseline was the Twitter search ordered by time recency. They found that their model was able to rank top credible tweets more than the baseline. Moreover, the performance of tweet ranking, using a combination of message and source features, provided a significant improvement than using the features (message or source) individually.

Pal and Counts [2011a] aimed to identify the most authoritative authors on specific topics and proposed a number of features to characterise Twitter authors and compute authority metrics. The tweets were classified into three categories: original tweet, conversational tweet and repeated tweet. They also computed metrics for the tweets' features. To reduce the number of candidates as authorities for a given topic, they used Gaussian mixture model, then they used the Gaussian algorithm to rank authors. They collected around 90 million tweets and selected tweets related to three topics with their queries: world cup, oil spill and iPhone. To evaluate the results of the model, they chose the top 10 results returned by their model and other three baseline models – each model consists of one property type: graph, text, and author –. Then they asked the participants to rate the authoritativeness of the authors using a seven-point Likert scale. Each participant saw 40 tweets with their authors for each topic. In the result, they found that their system achieved better results in finding the authoritative authors than state-of-the-art methods ($p < 0.05$).

[Ravikumar et al., 2012] tried to rank tweets in a high-impact event by considering content trustworthiness and popularity. In their model, they classified Twitter data into three layers: users, tweets and web. They used the links within and between layers for ranking purposes. In the tweet layer, they computed a semantic sense between each pair of tweets. In the user layer, relationships between users were used to compute a score for each user. In the web page layer, they used a page rank algorithm for the links included in tweets. Next, they ranked tweets based on popularity and trustworthiness. They used popularity as an indication of relevance, when a large number of tweets agreed on a given tweet by computing the similarity between them, then it was considered a relevant tweet. For trustworthiness, they assumed that when two authors share the same fact, then the tweet is most likely to be trustworthy, the same approach used in webpage [Balakrishnan and Kambhampati, 2011]. To evaluate the system, Twitter data for six news topics and six queries were collected, and they selected 200 random tweets for each topic. The tweets were then labelled into three classes: -1 to 1, where -1 was for tweets which contain spam, 0 for tweets contain opinions and 1 was for tweets which contain facts. By comparing their system with term frequency-inverse document frequency (TF-IDF), the researchers found that the top k returned tweets by the proposed system were almost always credible, while the TF-IDF top returned results included many untrustworthy tweets.

For the same purpose as the previous research, Gupta et al. [2014] used a semi-supervised learning to rank approach to assess social media credibility in real time. For the purpose of the experiment, they collected Twitter data relating to six events, all of which could be classified as crisis topics. They annotated tweets into different credibility classes: definitely credible, seems credible and definitively incredible. They annotated 500 random tweets using CrowdFlower platform. They extracted a set of features to be used in their ranking model, with all selected features inspired by those used by [Castillo et al., 2011]. To build the system, typical information retrieval approach models were used: Coordinate Ascent, Ada-Rank, Rank-Boost and SVM-Rank. Next, the researchers used NCDG@n (normalised discounted cumulative gain) evaluation metrics to evaluate different ranking models for queries of the six events. They found that Coordinate Ascent and Ada-Rank achieved better results than did the other schemes. The top five features for the credibility model were all content features. Moreover, they made the system real time, so that it can be used as a browser extension, and it has been made available for users to download. Their system gives each tweet a score between 1 (low credibility) and 7 (high credibility). They received feedback from around 1,300 users, with 40.0% agreeing and 60.0% disagreeing with the system score. However, they found that 49.0% of those who disagreed felt that the score needed to be higher than what the system gave.

2.5.2.4 Linguistics Features Associated with Credibility

Studying the phrases and words around the credibility is an important area of research in social media. O'Donovan et al. [2012] explored linguistic features distribution in social media for different credibility levels. They collected Twitter data related to seven different events. Then, to define the credibility levels of tweets they developed ground truth by hiring participants from Amazon Mechanical Turk to annotate the tweets. A group of tweets were provided to participants with a five-point Likert scale, with 1 being not credible and 5 highly credible. Then, they classified the tweets into two categories: credible (including those scoring 1 and 2) and not credible (including 4 and 5); tweets with a score of 3 were discarded. Next, they studied a number of linguistic features related to each credibility class, such as number of words, pronouns, smile, etc. They found that occurrence of many features was different between the two credibility classes, for example, smile, exclamation, question, etc were higher in non credible class, and they found that positive sentiment not always associated with credible class. Moreover, feature occurrence in the non-credible class was higher than in the credible class, which indicates that the presence of features does not imply credibility.

Mitra et al. [2017b] built a model based on language to predict credibility. They used around 9,000 phrases from several linguistic categories and found some of these categories had predictive power of credibility. They used the CREDBANK corpus, which consists of 66 million tweets for 1,377 events along with credibility annotations [Mitra and Gilbert, 2015]. They used 15 lexicon and non-lexicon measures, including positive and negative emotions. A sentiment analysis tool called Linguistic Inquiry and Word Count (LIWC) was used. This text analysis application counts words in different psychological categories. The statistical predictive model achieved 68.0% accuracy, and the authors found some of their measures were related to low-credibility tweets while others were associated with high credibility. For example, positive and conjunction words were associated with low-credibility tweets.

Kwon et al. [2013] identified the linguistic features related to rumours and non-rumours in Twitter before [Mitra et al., 2017b]. They used 102 topics divided into rumours and non-rumours, with each topic having at least 60 tweets. LIWC was used for language analysis, the classification accuracy for rumour v. non-rumour was higher than that achieved at baseline [Castillo et al., 2011]. However, their model was not solely based on linguistic features; it also included other feature types like structural and temporal. So, we can not compare their results with the findings of other studies that used linguistic features only.

2.5.2.5 Credibility Perception

Another direction in social media credibility research is analysis of the features that affect readers perceptions for judging information credibility. This direction of research is important as it helps to select the features that really matter for end users. A number of experts in social media credibility research presented the needs for new approaches which combine social and computer science methodologies. [Papadopoulos et al., 2016].

Despite the large amount of work on web page credibility perception, most previous studies on information credibility are not compatible with social media content such as Twitter. Social media platforms have their own distinctive features and usage specifications: for example, Twitter has special characteristics such as only 280 characters for each tweet, comparing to the open space in the traditional web pages. Also, users have the ability to choose their identity features; for example, username and profile image. This makes implementing previous results from web pages insufficient.

The information posted on Twitter is considered less credible than when the same information is posted on newspaper websites. Schmierbach and Oeldorf-Hirsch [2012] performed a study to understand how the same information can be perceived differently when it is distributed through two different channels (Twitter and a news website). Their study involved 225 participants and was about a news story taken from the *The New York Times*. The same story was presented in three different forms: long form (the first four paragraphs of the story as they appeared in the newspaper); short form (the first paragraph of the story with 'read more' at the bottom); and Twitter version (Twitter page of *The New York Times* with a list of the same story tweets). Each participant read one form and was asked how he found the source and message credibility, using two seven-point Likert scales (one scale for source and another for message). They found that people consider the first form, as it appears on *The New York Times* website, as more credible than the Twitter form, even though the story posted in Twitter came from *The New York Times* account. This finding shows the lack of credibility in social media (e.g Twitter) from a reader viewpoint.

The first complete study on microblog credibility perception was conducted by [Morris et al., 2012]. These researchers identified 31 features that readers use for credibility assessment, as shown in Table.2.2. They asked the participants what the impact of that feature was on credibility, they used a five-point scale (with 1 being the lowest and 5 being the high-

Table 2.2: Mean ratings of credibility effects of tweet's features, measured by a five-point Likert scale.

Feature	Credibility Impact
non-standard grammar/punctuation	2.71
default user image	2.87
cartoon/avatar as user image	3.22
author is following many users	3.30
logo as user image	3.37
contains shortened URL	3.39
customized Twitter homepage	3.41
author location near you	3.43
contains hashtags	3.48
contains a URL	3.50
author tweets frequently	3.52
contains complete URL	3.57
near top of search result list	3.58
posted recently	3.59
is a reply	3.61
author has many followers	3.65
author bio suggests topic expertise	3.66
is a retweet	3.66
username is related to topic	3.67
author location near topic	3.67
author often mentioned/retweeted	3.69
personal photo as user image	3.70
many tweets w/ similar content	3.71
author often tweets on topic	3.74
account has verification seal	3.92
author is someone you have heard of	3.93
contains URL you clicked thru to	3.93
contains is someone you follow	4.00
verified author topic expertise	4.04
is a RT from someone you trust	4.08
user image, generally	NA

est credibility). Some of these features have credibility influence perceived as higher than others. For example, a tweet that had been retweeted from someone who was followed by a user was given a mean of 4.1, while a tweet with grammatical mistakes received 2.7. In the results of this survey, they found a disparity between the features presented by Twitter and the features considered important in determining a tweet's credibility for the readers. Also, they found that readers were found to be poor at making credibility judgments based on content alone, and user credibility assessment was influenced by other non-content features such as the profile image of the tweet's author.

From the list of defined features the researchers then selected three to run a controlled experiment (message topic, username and user image). They authored 36 tweets on three topics (politics, science and entertainment) with three username styles (traditional, political and internet) and five image types (female photo, male photo, topical photo, generic photo and Twitter default avatar). Half of the tweets were described as real events and information and the other half as plausible events that never took place. A total of 266 participants completed the study and each tweet was scored on two Likert scales from 1 to 5: the first asked about tweet credibility and the second, the tweet's author credibility. To measure the influence of each feature type, statistical analysis was performed using one-way ANOVA. The results revealed significant differences between credibility influence for each feature type. For example, a topical username was perceived as significantly more credible than traditional and Internet styles. Moreover, the message topic influenced reader credibility perceptions in different ways. For example, science tweets were perceived as the most credible (mean 3.9) and politics, the least credible (mean 3.7) ($p < 0.05$). Also, the results revealed readers concern about content credibility when tweets came from stranger authors; that is, someone the user does not follow. In the study, readers could not distinguish the tweets truthfulness by tweets' contents alone, and they needed to use other author features such as username and profile image to assess the tweets credibility. The researchers discussed how incorporating readers' perceptions with interface design increases the credibility of a microblog, knowing that Twitter users spend only 3 seconds on reading a tweet makes the chances of credibility judgment error high [Counts and Fisher, 2011].

A study of the effect of author's username in credibility judgments in Twitter was conducted by [Pal and Counts, 2011b]. In this study, the researchers described how the author's username has an effect on the reader's credibility judgment regarding tweet content, particularly how readers are biased towards an author's name. They selected three topics: world cup, oil spill and iPhone. For each topic, they selected 40 authors with followers numbering between 29 and around 2 million. All selected authors were considered experts on these topics, four tweets from each author were selected on a given topic. These tweets were then shown to participants who were asked to rate two aspects: how interesting they found the tweets; and how authoritative they found the author, each on a seven-point Likert scale. They presented the first 20 authors anonymously without a username and the second 20 authors with a username. The authoritative ratings for anonymous authors were higher than non-anonymous ones. However, authors with a high number of followers were received low rating in the anonymous condition, while they received high ratings when presented as non-anonymous. The results of the study has also classified readers into two classes: heavily and slightly biased. The study only considered one profile feature to show the bias in readers' credibility judgments.

The ratio of followers to follows can affect readers' credibility judgments in Twitter. Westerman et al. [2012] ran a study to understand how the ratio between number of followers and follows affects readers' credibility perceptions towards an author in Twitter. They designed a 3×2 study; examining number of followers (many, moderate and few); and different ratios between followers to follows with two conditions (wide gap, narrow gap). The following definitions were used: many condition = 70,000 followers, moderate = 7000 followers, and few = 70 followers; and wide gap $\approx$ 90.0% and narrow gap $\approx$ 10.0% of follows $<$ followers. To study credibility perceptions, three seven-point Likert scales were provided, each measuring one credibility aspect: trustworthiness, competence and goodwill.

289 participants completed the study, each participant only see one of the six conditions In the results, there was no significant differences between the impact of the number of followers on readers' credibility perception. However, there was a significant effect of the ratio of followers to follows on competence credibility judgments, when the gap is narrow, the authors' competence increase.

Wagner et al. [2012] studies credibility in Twitter by exploring the topical expertise of Twitter authors. To do that, they used Wefollow rank[2] to select 24 Twitter expert authors on the topic of the 'semantic web'. They then presented tweets to 16 participants in three conditions: 1) the last 30 tweets and retweet from the author, author's profile information, and the latest 30 list memberships of the author; 2) only the last 30 tweets and retweet information were presented to participants; and 3) only see the author's profile information and his/her list membership information. For each author, participants were asked how much the author knew about this topic using a five-point Likert scale. By using a two-way ANOVA between author expertise (high/low) and three conditions, they found a significant interaction between the three conditions and the authors expertise as high or low ($p < 0.01$). Also, they found the lowest expertise judgment was given to condition number 2, where participants only see tweets without bio information. The study presents the importance of other features beyond the content on support readers' judgments; however, only small number of topic and participants were includes in this study.

Shariff et al. [2014] studied tweet and author features that Twitter readers use when they assess credibility of news tweets. They collected 400 tweets relating to 20 events and used participants from Crowd Flower platform to determine the credibility level of tweets as follows: definitely credible, seems credible, not credible and cannot decide, participants were asked to leave comments about the reason for their choice. Next, the authors analysed participants' comments and derived the features that readers used to determine credibility. Then, they used the feature categories of Castillo et al. [2011] to classify the derived features. They defined a group of features that influenced readers' perceptions: message based; author

[2]www.wefollow.com

based; and topic-based features. In the results, around 85.0% of tweets were definitely credible, 54.0% of features derived from participants' comments were topic based and only 16.0% were message based. This work used a qualitative methodology to understand the features used by readers when judging credibility of breaking news in Twitter, but the study only considered one topic and did not include other topical areas.

Kang et al. [2015] investigated the influence of author profile features on the readers' credibility perception in social media, and how the findings are transferable between social media platforms Twitter and Reddit. They ran a survey of Twitter readers asking them about the features they considered to indicate information credibility when performing a Twitter search. The authors used Amazon Mechanical Turk to recruit participants, a total of 81 participants completed the study. Content and authors features had the highest influence on reader assessment of credibility. but the visual features such as author image have the highest influence on readers credibility judgments. Further, to clarify the how different microblogging service influence readers credibility, they ran a study on Twitter and Reddit using screen tracking and asked the participants to click on the item they felt has impact on their credibility perceptions, 102 participants completing the study. Each participant saw seven screenshot of pages-three for Reddit and four for Twitter, each sereenshot contains tweets related to a news or a news subreddit. Statistical analysis showed that credibility ratings for Twitter and Reddit were the same.

Next, they defined a set of 12 author and content features of microblog with high impact on readers' credibility perceptions based on participants' feedback, they classified the features into three classes: visual including profile image and the attached photo; network such as number of followers, retweets and mentions; and content which is manly content features such as hashtags, links and sentiment. There were two conditions for each features: present and absent, for example, profile image has two cases as present or not present. Then they authored the needed posts for the experiment, the participants used a five-point Likert scale to rate each tweet's credibility through a controlled experiment. 646 participants from

Amazon Mechanical Turk completed the study. In the result, they found that the features in visual class like profile image have the highest influence of the readers credibility judgments in both platforms. All the studied features in this study were included on previous studies, the new contribution on this study was including two different microblogging platforms in studying credibility.

AlMansour and Iliopoulos [2015] analysed tweet credibility in the Arabic language. They collected around 200 tweets on nine different topics. They presented these tweets – as presented on Twitter – to the participants on the Crowd Flower platform. After they asked the participants to rate each tweet using a five-point Likert scale, with 1 being not credible and 5 being very credible. A total of 52 participants completed the study. The authors pre-defined the number of features in three categories: author, content, and popularity. In total, they used 44 features related to the three categories. Then, the authors classified the tweets into two classes based on credibility, tweets with a rating 1, 2 or 3 were classified as not credible, and tweets with a rating of 4 or 5 were classified as credible. Next, they studied the features that influenced readers' credibility judgments. They found that some features such as spelling errors, and profanity, are related to tweets that are not credible.

This section reviews the research in the area of credibility perceptions on Twitter. The previous research studied the features that readers use when judging credibility of information on Twitter. I can classify the common features that influence readers' credibility perceptions into two types: author and content. Author features are those such as the profile image, profile information, number of followers, and username. Content features include the features in the text such as URLs, hashtags, mentions, replies, sentiments and message topics.

2.6 Factors That Affect Author and Reader Behaviour in Social Media

The factors influencing the credibility judgment of users are variable and include reader and author factors; and the content and media used. With respect to readers and authors, there are many factors that can influence their behaviour, such as culture, country and language.

Readers from different locations can judge the credibility of the same information in social media differently. Location of reader has a big influence on behaviour in social media. Social media readers are distributed in many countries representing different cultures.

Researchers have begun to understand the association between culture and how people interact with computers [Setlock and Fussell, 2010]. For example, Asian users take many considerations into account when choosing communication tools and use of social network [Ji et al., 2010]. Also, previous research has found that people from different countries and languages have different usage patterns. In sections 2.6.1 to 2.6.3, I present an overview of research investigating the effect of culture and related factors (e.g. language and country) on social media users' behaviour.

2.6.1 Culture

Culture influences online behaviour of users in Q&A forums, Yang et al. [2011] found significant differences between users from different cultures in social questioning and answering behaviour, motivation, frequency and content. They conducted a study on users from four countries: India, China, the US and the United Kingdom (UK). These four countries represent two cultures: Asian (India and China) and Western (US and UK), as defined by the authors. They controlled other factors such as demographics and the difference between participants was cultural. A total of 933 participants completed the study from all four countries. The authors found that users from Asian countries more frequently post questions than Western users. Also, they measured the importance of the users receiving an answer when they post a question, Asian users perceive having their questions answered as more important than do Western users ($p < 0.005$). In terms of motivation for asking questions, there were significant differences between cultures: for example, 'it is faster' was one of the most important motivations for Asian users asking questions, while the main motivations for Western users was 'I trust answer from my network'. Culture was also an important factor in encouraging users to answer questions. Asian users were higher in frequency of answering questions than were the UK and US combined. The motivations for answering questions were also influenced by culture. For instance, the belief 'To be connected and I will get help when I help' was significantly higher in the Asian than the Western culture. On the other hand, regarding motivations for not answering questions, 'I don't know the asker' was the main motivation for Western users, while 'I don't know the answer' was the main reason for Asian users. Moreover, the authors found that culture was an important factor in deciding which online tool to use for asking questions. These findings together demonstrate how some cultures need to provide information along with social utility, such as in the Asian culture. Acar and Deguchi [2013] studied the influence of culture on Twitter usage between Japan and the US authors, and found that culture has an impact on the number of questions that authors post on Twitter.

Gupta et al. [2013] wanted to study how cultural differences may affect the way authors create information, and use culture to predict author's behaviour in social media. The authors analysed tweets of authors from different cultures to predict the hashtages and name entities that the author may include in his/her future tweets. For the purpose of the experiment, the researchers collected Twitter data from author of three different countries represent different cultures: US, India and Egypt. Around 1300 authors, along with their friends data were collected for each culture. They only included authors with a valid location entered into their home page. In the results, when they computed the relatedness between the keywords for each culture dataset – relatedness value is the degree of occurrence between the keywords –, they found the India and Egypt authors have a very close results comparing to different result in US users. Moreover, they found the author past tweets played a significant role in predicting the author future keywords for India and Egypt authors comparing to less impact for US authors.

Credibility perceptions of readers in Twitter were influenced by their cultures, as shown by a previous study [Yang et al., 2013]. These researchers ran a study to understand the influence of culture on credibility perception of microblog readers. They selected two countries to represent two cultures: the US represented Western culture and China as an Eastern culture country. To measure differences, a controlled experiment was conducted using microblog data from Twitter and Weibo (Twitter is banned in China). Four features relating to tweet authors were selected: message topic, author name, author image and location. The authors found significant differences in credibility judgments of readers from different cultures, as the same feature may has different influence on the readers of different cultures when judging credibility. Moreover, they found that the influence of the readers' cultures were much greater than the influence of features on credibility judgments.

Gavilanes et al. [2013] studied the relationship between authors behaviour on Twitter and their cultures. The authors focused on three cultural dimensions: Pace of Life, Individualism and Power Distance. Each cultural dimension had a relation to predicting one of three behavioural patterns: temporal pattern, to interact with other authors, and following each others. They collected tweets and data of 2.4 million Twitter authors for experimental purposes. After running the analysis on this large number of data, they found that activities of authors from countries with a high pace of life were temporarily predictable. Furthermore, they found that authors from collectivist culture interact with each other more than the authors in individualistic culture. Also, the results showed a strong relationship between the power distance aspect and the relationships between authors.

The above studies shows the influence of culture on the behaviour of readers and authors in social media, credibility research in social media can use the reader and author cultural background for enhance the credibility models.

Table 2.3: Feature distribution across languages in Twitter (1).

Language	URLs	Hashtags	Mentions	Replied	Retweets
All	21%	11%	49%	49%	13%
English	25%	14%	47%	47%	13%
Japanese	13%	5%	43%	43%	7%
Portuguese	13%	12%	50%	50%	12%
Indonesian	13%	5%	72%	72%	39%
Spanish	15%	11%	58%	58%	14%
Dutch	17%	13%	50%	50%	11%
Korean	17%	11%	73%	73%	11%
French	37%	12%	48%	48%	9%
German	39%	18%	36%	36%	8%
Malay	17%	5%	62%	62%	28%

2.6.2 Language

Language is known as the container of culture [Jiang, 2000]. The language used in social media communication has been found to influence UGC. Previously, it may assumed that the behaviour of English users could be generalised to all social media users; however, research has shown that users of different language behave differently in social media [Honeycutt and Herring, 2009, Hong et al., 2011]. Hong et al. [2011] collected 62 million tweets, they identified 100 languages, with English being the most commonly used language (51.0% of collected tweets). Only the top 10 languages were used for further analysis, as shown in Table.2.3. To measure the effect of language on social media user behaviour, they selected five text features: # hashtag, RT retweet, @ mention, reply and URL. The results reveal large differences between languages in the use of the same content features. For instance, 39.0% and 37.0% of German and French tweets, respectively, included URLs, compared to 13.0% for the Japanese and Indonesian languages. Moreover, hashtag usage across languages also differed greatly. For example, 18.0% of German tweets included hashtags compared with only 5.0% of Japanese tweets. The authors also used hashtags to study how some languages post about the same topics. They considered the top 100 hashtags of each language and found some languages share most hashtags: for example, Indonesian and Malay tweets shared 65 out 100 hashtags. Conversely, among the popular languages (e.g. English), Spanish and English tweets shared 24 out 100 hashtags. Some languages have more social ties than others. In reply features, which indicates one-to-one communication, 59.0% of Korean tweets had a reply compared to 20.0% of Indonesian tweets. The retweet feature describes the broadcast of the tweet to a large audience. Of the Indonesian tweets, 39.0% were retweeted compared with only 7.0% of Japanese tweets.

For the same research aim, Weerkamp et al. [2011] analysed language differences using four features (# hashtag, URL, @ mention and reply) and eight languages (Dutch, English, German, French, Indonesian, Japanese, Portuguese and Spanish). Table.2.4 confirmed the findings of previous research. For example, use of hashtags was highest in German and lowest in Japanese tweets. Moreover, German tweets had the highest usage of URLs and Indonesian tweets the lowest. However, using the @ sign indicates a direct tweet between two persons and is a 'social' feature. Indonesian language had the highest use of mention @ (77.0%) and German language, the lowest (25.0%). The same theory can be applied to reply, as it is more of a social feature than hashtag and URLs. The Dutch and Spanish languages had the highest use of reply (36.0% and 34.0%, respectively), with Indonesian and Portuguese having the lowest (13.0% for both). These findings confirm previous research that has suggested social media usage is mostly influenced by authors cultures which represented by language.

In summary, previous research have shown that social media use differs between languages. Further analysis is needed on the effect of these differences on models that use content-based features across languages.

2.6.3 Country

Country of social media authors has been found to influence their generated content in similar way to culture. Poblete et al. [2011] sought to understand differences in social media usage

Table 2.4: Feature distribution across languages in Twitter (2).

Language	Hashtags	URLs	Mentions	Replied
Dutch	16%	15%	62%	36%
English	14%	30%	50%	25%
German	25%	48%	28%	14%
French	16%	37%	55%	27%
Indonesian	10%	12%	77%	13%
Japanese	4%	11%	48%	26%
Portuguese	11%	10%	45%	13%
Spanish	12%	24%	62%	34%

between countries. The study included ten countries covering most cultures, and around 5 million active authors were selected with more than 4 million tweets collected over a one year period. An active author was defined as one who has a valid location and has at least one tweet in ten days, they used home location of author to determine location. Number of tweets per author differed greatly between countries (e.g. 1,800 tweets in Indonesia compared to 800 in Canada per user). In term of sentiment analysis, happiness level was found to differ between countries; for example, Brazil had continuous happiness for most of the year, unlike other countries. In terms of actual behaviour in the text of a tweet, there are four types of author behaviour: # hashtag, RT retweet, @mention and URL. Table.2.5 presents the use of each of the four features for each country: for example, URLs use in countries such as Netherlands, UK, US, Australia and Canada was higher than in those countries like Indonesia and Japan. Similar results were observed for hashtag use. Together, these results show how country with culture can influence authors' behaviour in social media. This work provides a deep understanding of how people in each country use social media, which can help to improve social media system design (e.g. which features users in each country would like to see). Mocanu et al. [2013] has explored the differences between countries on Twitter

Table 2.5: Feature distribution across countries in Twitte.

Country	Tweets/Users	URLs	Hashtags	Mention	Retweets
Indonesia	1813.53	14.95	7.63	58.24	9.71
Japan	1617.35	16.30	6.81	39.14	5.65
Brazil	1370.27	19.23	13.41	45.57	12.80
Netherlands	1026.44	24.40	18.24	42.33	9.12
UK	930.58	27.11	13.03	45.61	11.65
US	900.79	32.64	14.32	40.03	11.78
Australia	897.41	31.37	14.89	43.27	11.73
Mexico	865.7	17.49	12.83	49.79	12.61
South Korea	853.92	19.67	5.83	58.02	9.02
Canada	806	31.09	14.68	42.50	12.50

from different perspectives as well.

It is important to mention here the increasing amount of research into the relationship between mobility of media and human usage, or what is called 'transmigrant media', as migration has increased in recent years and social networks at the same time are a key feature in many people's lives. How migrants use social media is a current research area [Seto and Martin, 2018, Gomes, 2018, Chang and Gomes, 2017]. Thus, using language or country when analysing social media data can be challenging because some users use languages and applications other than what most people use in their new countries.

From the above, it is evident that social media users – authors and readers – are influenced by factors like culture, language and country. The previous research showed that using of some content features were not the same between authors on social media when they use a different language, or they are from different country with different cultures. Also, the previous study found the culture of readers influenced them when assessing social media information credibility, and when they use Q&A forums. This means these factors need to be taken into account when studying social media credibility. This idea is explored in some parts of this thesis.

2.7 Event and Author Location Influences Credibility and Sources in Social Media

Using the locations of an event and author to predict author credibility is an important factor. During an event, social media platforms sometimes raise the first alarm: authors sharing information about what is happening. Authors differ based on their location with respect to the event location (near or far). Authors posting information about a particular event from the same or a nearby area to an event are known as 'eyewitnesses'. Eyewitnesses in social media are important as they are able to provide first-hand and credible information about the event.

The following sections cover many aspects related to location, sources and credibility. First, I show the influence of author location on the generated content in section 2.7.1, and then examine research that aim to find eyewitness authors in social media in section 2.7.2. I also explore the main information sources that normally contribute to different events in section 2.7.3 and, finally, present some approaches for finding expert sources for different topics 2.7.4.

2.7.1 Location Influence on Author Behaviour in Social Media

There is a strong relationship between what authors write in social media and their location. For instance, tweets from authors in the same event location are found to be different from tweets from distant locations [Morstatter et al., 2014]. Many studies have taken advantage of the influence of location on social media authors behaviour. Thus, some research has begun to predict authors' locations based on their behaviour.

Cheng et al. [2010] proposed a probabilistic model to estimate user location (city level) based on tweet content alone. They crawled more than 1 million Twitter users with around 30 million tweets. They found 74.0% of profile locations submitted by authors were very general and covered a wide geographic space. To link tweet content and location, they considered words appearing at least 50 times. They generated around 500 thousand distinct words and around 25 thousand cities from the tweets content. The used the profile locations entered by users in form of latitude/longitude as the ground truth. Their model was able to estimate the location of 51.0% of authors with 100-mile radius accuracy. One limitation of this work, however, is that they only included authors with 1,000 tweets or more to be able to achieve good results by their model. In the same way Hecht et al. [2011] used only text-based features of tweets and found the authors' country and state could be easily identified only by tweets' content.

For the same purpose Mahmud et al. [2012] presented a model for inferring authors' home locations in Twitter. To predict location, they used three classifiers each with different features based: content, heuristic and time zone models. In the heuristic classier, the authors mentioned their local locations more often than other locations, the highest location occurrences in the author tweets was assumed the authors home location. The second classier was statistical based, which contains words, hashtags and place names in tweets. The third classier was trained on the time zone that authors send their tweets. For the experimental purpose, they collected 1.5 million tweets generated from around 9,500 authors. In the result, after they built a hierarchical classifier based on the three previous classifier. They found a significant improvement with their model in predicting the Twitter authors' locations compared with that of [Cheng et al., 2010].

Schulz et al. [2013] proposed a multi-indicator model for predicting both tweets and author' locations. They used features from the tweet text as well as other information from the user profile. Natural language processing (NLP) techniques were used to estimate locations based on language models, and they used gazetteers to determine locations referred to in tweets. They used 1.3 million geo-tagged tweets for testing and evaluation in their model. They estimated the locations of 54% of tweets to within a 50-km radius, and 79% of user locations to within a 100-mile radius.

Han et al. [2014] extended previous work and proposed a framework relying only on tweets' text-based features to predict Twitter authors' locations. In this research the tested the impact of language on predicting Twitter authors' locations, as previous research used only English language data because identification language tools with high accuracy were available only for English at that time. To test that, they used a large number of multilingual Twitter datasets of fifteen different languages, 23 million tweets: 11 million were English tweets an 12 million were other languages tweets [Han et al., 2012]. In the results, they found that predicting authors' locations for some languages were easier than others, prediction of authors' location from languages was restricted by geographical space. For example, Indonesian and Japanese tweets' authors were much easier to predict than languages spread over large spaces, such as English.

This section presents some research that used the author generated content to predict social media authors location. All of them have shown the strong relationship between an author and his contents as they be able to predict his location by only using what he write. Next, i show the research on the area of identifying eyewitness in social media.

2.7.2 Eyewitnesses in Social Media

Little research has been conducted into how to find eyewitnesses of events in social media. This section differs from Section.2.7.1 as eyewitness research tries to find the authors in affected regions at a time of an event such as in a crisis. Agencies responsible for disaster response have begun to include social media as an important source of information to reach and understand affected people. Social media eyewitness research aims to find authors from the same place as the event, assuming accuracy and credibility of these sources. Eyewitnesses are valuable and in most cases they deliver the first information about the event. The issue with defining eyewitnesses is that of rarity, as described by [Olteanu et al., 2015].

A few researchers in this area have tried to reach eyewitness authors. Truelove et al. [2015] used the 'bushfire' case that occurred in Melbourne in 2013 to determine eyewitnesses, using the keyword 'bushfire' to collect all related tweets from Twitter API, after the data were cleaned and included only on-topic, 461 authors were classified into two categories: witness author (WA) and impact author (IA). Two coders completed the manual classification of the eyewitness authors. After examining the tweets' content they found it was feasible to differentiate WA and IA tweets from other authors' tweets; for example, they found in the 'bushfire' case that WA authors reported on traffic congestion, road closures and emergency response vehicles, while IA authors reported on evacuees and volunteer fire fighters travelling to and from the event. However, this research was a case study that covered only one event with a small dataset, and no model was proposed. The same researchers trained a model to identify WA and IA accounts [Truelove et al., 2016], by using data collected from an Australian Football League match. They collected event-related tweets and used geo-tagged tweets with extra features, including image and text. They found that using the tweet text with linked image increased the number of predicted WA and IA. However, there are many limitations to their model, including the small amount of data analysed, only one event was used to train the model, and accuracy variation of the same model is a common issue when applied to a different event.

Morstatter et al. [2014] aimed to differentiate between tweets from an affected location and tweets from a different location in time of crisis, and to build a model to predict eyewitnesses. They collected tweets related to two crisis events in the US (2013 Boston Marathon Bombing and 2012 Hurricane Sandy), only geo-tagged tweets from the US only were collected. They classified the tweets into two location types (inside and outside region). To build the model NLP techniques were used, they computed the probability distribution of the used words using Jensen-Shannon divergence [Lin, 1991], and found greater divergence within the affected locations compared with those less affected. They then located the tweets from the inside region automatically and included a number of linguistic features like unigrams & bigrams, part of speech (POS) and shallow parsing. They trained a machine learning model for predicting inside and outside tweets; a naive Bayes classifier was employed. The accuracy of predicting tweets as inside or outside was 0.83 for the Boston Marathon Bombing and 0.88 for Hurricane Sandy. This model relied only on linguistic features and was able to achieve good results. By only using tweets content features, the study confirm the ability of linguistic features to help in identifying eyewitness in social media.

Tanev et al. [2017] trained a model to predict eyewitness tweets. Available tweets from 26 crisis were annotated as eyewitness or other [Olteanu et al., 2015]. English and Italian languages were included to measure the influence of language on model accuracy, and they used a set of language features such as lexical, stylistic, word capitalisation and other content features like hashtag, mention and so on. They compared performance of three classifiers: naive Bayes, SVM and random forest. Naive Bayes achieved the best result for the Italian language, with accuracy of 0.69, and random forest was the best for English, at 0.79. One limitation of this work is that only a small number of tweets was used in the training model, as they needed to cut off the number of tweets to balance between eyewitness and non-eyewitness authors.

In this section, I show the main works in the area of eyewitness identification in social media, and we can realise that eyewitness can be identified from their generated content. However, eyewitness are not always available in social media for all events. So, finding other sources who are not from the same place of event but who are close to it and able to provide credible information is needed. It appears that none of the existing credibility research has examined the influence of location on credibility assessment in social media. Moreover, the distribution of information sources in social media across different locations was not explored, and how the behaviour of the same source type differs between locations (i.e. close to or far from the event). Next, I will present main information sources in social media.

2.7.3 Information Sources in Social Media

When readers use social media like Twitter to obtain an update about an event, they are concerned about the source of the information. Many types of sources share information in social media and these sources are a mix of individuals and organisations with high and low credibility status.

Starbird et al. [2010] studied the information sources in social media in an event like a crisis. They studied the 2009 Red River Valley flood that occurred in Canada and the US. They collected Twitter data related to the event from 358 authors and 7,183 tweets. Then they classified the sources into eleven types: national media, alternative media, local media, flood specific service, public service agency, faith organisation, blog, other service organisation, bot, individual and unknown. They found that individuals made up the largest group of sources, at more than one-third of authors, but they did not study the credibility on the information generated by these sources.

Wu et al. [2011] aimed to differentiate between ordinary sources and other sources in social media, and then they classified the other sources into different types. To achieve this, they used the follower graph produced by [Kwak et al., 2010], which included nearly 40 million users and 1 billion edges. The authors classified the sources into: celebrity, media, organisation and blog. They found media was the largest group (41.0%), followed by bloggers (24.0%), organisations (19.0%) and celebrities (16.0%). In the results, they found that the information generated by each source were different in many aspects.

De Choudhury et al. [2012] classified sources in Twitter into four source types: organisations, journalists, ordinary individuals and other. Organisations included all entities related to business, politics, social or other goals, and whether they were commercial or non-profit. Journalists were associated with a news agency or worked for their own interest. Ordinary individuals are those use Twitter for no specific purpose such as posting their daily updates, communicate with friends, etc. The other category included types of sources that did not fit into any one of the three previous categories – it includes a small number of authors and it is not well defined –. They used five types of features: network features (including number of followers and follows), interaction features (the author interaction pattern with their friends including number of retweets, replies, mentions and so on), named entity (the entity names in the content such as place names, derived from OpenCalais toolkit), activity features (containing number of tweets posted by user until time of crawling) and topic distribution (including types of topics that a source is normally interested in). For the purposes of the classification experiment, they collected 1,850 Twitter sources and presented them to participants in Amazon Mechanical Turk for annotation, in addition to 1532 organizations and 1490 journalists were collected from Twellow directory – contains pre-labelled authors –. In total, 4,932 sources were labelled with one-source categories. The data collected for eight events was then applied to test their classifier accuracy. The accuracy of the classifier was 88.7% on the test data and they found the largest source category was ordinary individuals. The results showed how the sources were different to each other, in terms of content such as frequent keywords and number of URLs, mentions and questions. However, the information credibility of these sources has not been tested in this work.

Thomson et al. [2012] studied the credibility of different sources referring to the Fukushima disaster in Japan 2011. They collected tweet data related to the event; the tweets were mostly (60.0%) in Japanese language, with 30.0% in English and 10.0% in other languages. Only data from active sources were collected (those having five tweets or more about the event) and in total 4,950 tweets were annotated. The authors classified tweet sources into 10 categories: traditional media, public institutions, business, non-government organisation (NGO), freelance journalists, researcher, eyewitness, non-local, non-identified source and conspiracy. Next, the authors classified the tweets which were authored by the first seven sources – mentioned above – as highly credible, while the rest of the tweets were of low-credibility. In the results of credibility classification, the found that 67.5% of tweets were highly credible, 8.0% were low and around 25.0% of tweets were classified as neither high nor low. This work said to be the first work studied the sources in social media along with credibility.

Olteanu et al. [2015] explored the sources of information which contribute on Twitter at crisis time. Approximately 28 thousand tweets from 26 crises were used in the study. They ran a series of crowdsourcing annotation tasks by classifying tweets through three phases. The first phase classified the informativeness of the tweet; they included only related or on-topic tweets. The second phase was to classify the information type of the tweets such as donation, warning, etc. The third phase was to categorise the tweet source; they included six categories (government, business, non-profit organisation (NPO), traditional media, eyewitness and ordinary). Over all sources, 42.0% were traditional media, 38% ordinary, 9.0% eyewitnesses, government 5%, 4% NGO and 2% business. From the results we can see that only 9.0% were eyewitness across all crisis events, that show the needing for other information source.

Diverse source categories are used in social media research and types of source differ based on event type. However, there are common sources for most events, such as traditional media, ordinary individuals and journalists. Next, I list some important work in the area of finding the topical experts in social media.

2.7.4 Topical Experts in Social Media

The expert is the author who is more knowledgeable than other authors on a specific topic, and expertise is one of the credibility dimensions. Finding an expert on social media is a challenge in many respects, for example, Twitter includes millions of authors who share information on topics varying from general topics like 'food' to more niche topics like 'barbecue', and finding a Twitter expert on a general topic is not that difficult compared to finding an expert on a more refined topic, especially if you want that expert in a specific location [Cheng et al., 2014a]. Authors post tweets globally, from well-recognised organisations to locally popular community organisations. As a result, the quality of information in social media is highly diverse, and locating sources with highly relevant and credible information on a specific topic has proven to be challenging. Finding experts is the main step in building models to determine authorities of information.

Methodologies to locate experts in social media differ based on both the goal of the study and which social media platform is selected. List is a feature in Twitter that allows users to group experts of different topics. To build a list, a user needs to specify a name with an optional description and then add users to the list as members. For example, a user can create a list named 'Politics' and add 'Barack Obama', 'CNN' and so on as members of the list, as well as additional information about the list.[3]

Bhattacharya et al. [2014] proposed a semantic model to present to topical groups in social media and find experts in these groups. The tweets content and network structures of the experts to build the model. They gathered around 38 million Twitter authors, along with their profile information, following links and the lists to which they were subscribed as members. Only English language tweets were considered. They were able to define more than 500 topical experts of different types of topics. They used a page rank algorithm to rank the top experts in each topic and succeeded in ranking experts for general topics like 'politics', but were unable to do the same for niche topics (i.e. topics with very few experts). Next, they analysed how the experts tweet about their topics of expertise and found that experts in niche topics were tweeting on their specific topic more than experts in general topics. These researchers have conducted many further studies in this area [Ghosh et al., 2012, Sharma et al., 2012, Ghosh et al., 2013, Kulshrestha et al., 2015].

Zafar et al. [2016] proposed a tweet credibility model based on experts. They aimed to identify the most credible tweets for any topic. In their model, they used lists in Twitter to find experts on a topic; authors were counted as experts if they appeared at least 10 times in different lists relevant to topics. The used lists include 50 million authors memberships, and extracted 1.2 million experts based on the previous definition of an expert. They used the TrustRank algorithm [Gyöngyi et al., 2004], which assumes that trusted web pages are linked with other trusted pages. In the same way as applied on Twitter data, they extracted verified authors from the expert list which they created earlier, and collected 83,852 verified authors.[4] The verified authors were the seeds to create more credible authors. They used 25 topics to evaluate their approach using crowdsourcing evaluators. They found their methodology for retrieving the most credible tweets was better than state-of-the-art

[3]https://help.twitter.com/en/using-twitter/twitter-lists
[4]https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts

approaches. Relying on topical experts and ordering the results based on their popularity in topical expert communities is an effective strategy for extracting the most credible and relevant tweets.

Finding an expert on a specific topic and location is an area of current research. This is more challenging than finding experts in topics only, because it must localise authority and topical expertise at the same time. Cheng et al. [2014a,b] used a massive geo-tagged dataset in Twitter to map source expertise and location for building an expert-finding model in a specific location. For this purpose, they used 15 million geo-tagged lists and 56 topics, and proposed a model to rank the authority experts for a such topic in a given location. For building a ground truth, they selected 56 queries for 56 topics, and asked Amazon Mechanical Turk's participants to evaluate their model's results. The participants saw the top 10 results for each query, and were asked to rate the expertise of each candidate source via a four point Likert scale. In the results, the achieved high accuracy results for identifying the experts in a given topic and location using NDCG@10 and precision@10. However, the limitation of this work is its reliance on geo-tagged information, which is very limited in social media.

This section provides an overview about the work of defining the experts in social media. This is one of the methodologies to find credible sources of information. However, the main limitation in this approach is that finding topical experts is limited to topics and not events, because most studies in this area depend on sources' topical expertise which use Twitter lists. In an event like a crisis, sources will be most likely to be ordinary individuals as shown in the previous section, and the source will not be included in any of these lists, as it is the main methodology for defining experts.

2.8 Research Gaps and Derived Research Questions

RQ1: When I reviewed credibility research in social media in Section 2.5, existing credibility models in 2.5.2 and credibility perception in particular 2.5.2.5, I noticed that many features of social media influence the credibility judgments of readers. Each feature has a different effect on credibility assessment, and sometimes has a complex effect when interacting with other features. Therefore, it is important to study the complexity of the influence of these features on readers' credibility perceptions, in particular for those features that are automatically generated, such as author location. I will address the existing research gap by answering the following research questions:

1. Which factors affect readers' perceptions of credibility?

- How does tweet location affect credibility?

RQ2: After I explored the power of culture, country and language on the behaviour of social media authors and readers in Section 2.6, I found that only a few studies have investigated the effect of these aspects on social media credibility. However, these studies were limited to only a few countries, and conflated the influence of different aspects. Therefore, there is a need to explicitly study the influence of culture, country and region when studying credibility in social media. I try to fill this research gap by developing the following research questions:

2. What is the effect of a reader's location on their perceptions of the credibility of social media information?

- What is the effect of a reader's culture on their credibility perception in social media?
- Will a reader's nationality influence their credibility perception of social media among other members of the same culture?
- What is the effect of the geographical region of countries with the same culture on readers' credibility perception of social media?
- How do different regions and divisions in large countries affect a reader's credibility perception of social media?

RQ3: Section 2.7 presents many aspects related to location in social media: the relationship between author location and generated content 2.7.1; eyewitnesses in social media 2.7.2;

source distribution 2.7.3; and topical experts 2.7.4. From these studies, it was apparent that in social media there is an influence of location on information sources in many respects, such as their locations and way of writing. However, to the best of my knowledge, few studies have explored the influence of distance between the author and event location on credibility assessment in social media. Therefore, there is a need to investigate the effect of location on source and linguistic features for information credibility of social media. The following research questions were designed to fill this research gap:

3. What is the effect of location on sources distribution and linguistic features for information credibility of social media?

- What types of information sources are associated with which events, both inside and outside the country in which the event is taking place?
- How do linguistic features differ among sources of different type, credibility level, topic and location?
- How effective is it to use only information source to assess credibility?

2.9 Summary

This chapter has reviewed the literature related to early studies of online credibility, credibility in social media, the factors that influence authors and readers behaviour in social media, and sources in social media. It provides an overview of credibility components and the main credibility theories related to research in the computing field in general, which includes web credibility, the historical development of credibility research and moving from offline to online environments.

The chapter presented the main approaches in social media credibility research: credibility prediction; credibility with similarity; ranking based on credibility; linguistics features associated with credibility; and credibility perception. Chapter.3 discusses the influence of different features on readers' credibility judgments. Chapter.4 compares the effects of culture, country and regions on readers' credibility judgments. Chapter.5 examines the location effect on source and linguistic features for information credibility in social media. Chapter.6 provides the thesis conclusion and outlines further research.

Chapter 3

Tweet Author Location Impacts on Tweet Credibility

3.1 Introduction

Features influencing reader credibility judgments of web pages have been studied by many researchers [Hargittai et al., 2010, Schwarz and Morris, 2011], the credibility of microblogs content is a new area of research. The first research on microblog credibility perception was conducted by [Morris et al., 2012]. They identified a set of attributes that influenced the perception of readers. They narrowed their study to three attributes: message topic, username, and user image. They found some attributes have a stronger effect on readers' credibility judgment than others. They discussed how incorporating readers' perceptions with interface design increased the perceived credibility of a microblog.

Other research has focused on the effect of an author's username on readers' judgment of credibility of a tweet [Pal and Counts, 2011b]. These authors described how the author's username and number of followers have an effect on the reader's judgment of tweet content. They suggested that the author's name should be considered within the retrieval system. All of these research projects together have shown the importance of non-text features for credibility perception.

In this chapter, I address the first research question of this thesis:

1. Which features affect readers' perceptions of credibility?

    - How does tweet location affect credibility?

The following sections explain the methodology and selected features used in the current study, and then outline the results of the experiment, along with a discussion of the results.

3.2 Methodology

To understand the effect of different tweet features on credibility perception of social media readers, I conducted a crowdsourcing experiment that included several tweet features to study their effect on credibility judgments of readers. I begin by presenting the features selected for the experiment, and the results for each feature; then I measure the interaction between the selected features and location.

3.3 Tweet Features to Examine

Many features influence readers' perception in social media. These features include author-based features such as username and user location; and tweet-based features such as tweet language and tweet length. In this chapter I follow the methodology of Morris et al. [2012], who examined three features: tweet topic, username and user image. To these, I add a fourth feature, tweet location – the location of the tweet author when the tweet was posted. The following sections describe the features.

3.3.1 Tweet Topic

A number of studies have indicated that reader credibility is influenced by the topic of tweets [AlMansour et al., 2014, Kang et al., 2015]. A survey run by Morris et al. [2012] on a number of participants about the features that they use when assessing credibility has also shown that the topic of a tweet is one of the features that readers indicate affects its credibility. I selected topic as one of the features in my experimental design to measure the effect of different topic types. I selected three topics – science, politics and entertainment – that are popular topics in social media and in Twitter in particular.

3.3.2 Username

The username influences social media readers' judgments and causes bias [Pal and Counts, 2011b], and is one of the attributes indicated as important when readers judge credibility in Twitter [Morris et al., 2012]. Therefore, I included three kinds of username in the experiment: traditional (e.g. 'William_Thomas'), topical (e.g. 'The_Politics') and Internet (e.g. 'Bill123') which is neither traditional nor topical, and might combine letters, numbers and symbols. The traditional usernames were distributed equally by gender, as I balanced male and female names. I created a traditional username by selecting popular names in the US for both genders. I used the official social security website in the US to obtain a list of popular names.[1] For all username types I verified that there was no actual Twitter account registered under the same username, to ensure that crowdsource participants would not have any preconceived ideas about any authors.

3.3.3 Profile Image

A user profile image or 'avatar' has been indicated to influence readers' credibility perception [Yang et al., 2012, Sikdar et al., 2013, Gao et al., 2015]. I chose five types of image: male, female, topical, generic, and default.

Similar to [Morris et al., 2012], for male and female images I used the Twitter search engine to obtain real user accounts by searching topics commonly used by men and women. As I did not wish to use photographs from popular accounts, I chose Twitter accounts

[1] https://www.ssa.gov/oact/babynames/ (accessed 08 August 2014)

with 1,000 or fewer followers to ensure that the popularity of account photographs did not influence reader judgments. All photographs were of adults and all were headshots.

For topical photographs, icons were collected from PowerPoint clip art because many topical icons in Twitter are associated with organisational Twitter accounts. These icons needed to be unknown to participants to avoid bias.

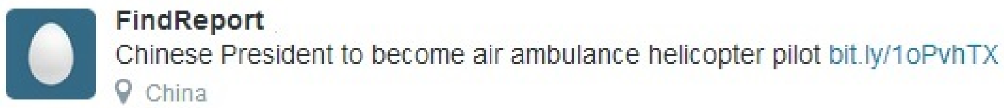
Generic icons were selected from actual Twitter accounts to make it realistic. The default Twitter icon was similar to an egg image and it remained the users' image until they changed it.

3.3.4 Location

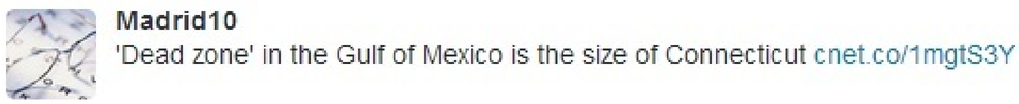
Including the location of the tweets' authors when they tweet is one feature whose effects I aimed to measure on reader credibility judgments. Morris et al. [2012] conducted a survey on 256 Twitter users, and author location was found to be one of the important features that users use when judging tweet credibility. Figure.3.1 (c) and (e) show samples of location tweets. I wish to study the influence of tweet location on readers' perceptions of credibility. The locations included in the tweets were the same country as the event.

3.3.5 Content of Tweet

For the purpose of the experiment, I created a Twitter account and authored original tweets on three topics: science, politics and entertainment. All tweets were written in English with standard spelling and grammar. All tweets were on current events taken from popular news sources, that were all actual events that took place during August 2014. Each tweet was followed by a URL constructed using the Bitly service [bitly.com], a service to make URLs shorter to fit within 140 characters for each tweet; Twitter increased the tweet limit to 280 characters in 2018, after this research was completed.



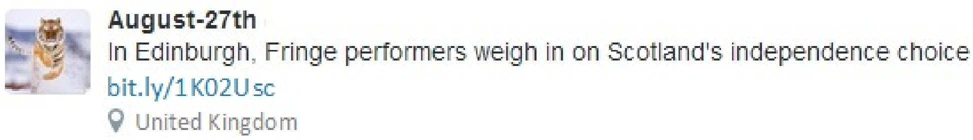
(a)



(b)

(c)

(d)



(e)

Figure 3.1: Samples of the experimental tweets: (a) default image, (b) topical icon, (c) female image, (d) male image and (e) generic icon. Tweets (a) is sample of topical username style, (b) and (e) Internet, and (c) and (d) represent traditional. Note, tweet (a), (c) and (e) have location included.

3.4 Experimental Design

As mentioned in the basic study [Morris et al., 2012], running all combinations of possibilities for all features – message topic, user name, user image, event location and tweet truth – would require ($3\times3\times5\times2\times2$) 180 tweets, which would require a large number of participants. As an alternative, I inserted user images between other features and this reduced the number to ($3 \times 3 \times 2 \times 2$) 36 tweets.

I made half of the tweets describe true events and half describe events that never happened but were possible. This was to ensure that the judgments of participants were influenced by the four features (topic, user image, username and location) rather than the truthfulness of the tweets. I also showed the tweets to some colleagues to determine if people could differentiate between the true and false tweets.

I inserted the experimental data into a Twitter style sheet and saved each tweet as an image. URLs included in the tweets were not clickable and participants were notified of that, to prevent participants checking credibility by clicking on a URL.

I classified these 36 tweets as follows: in each topic area – politics, science and entertainment – there were 12 tweets. In every topic area there were four tweets for each username style (traditional, topical and internet), two of which included the location of the event and two that did not. Within each two was a true and a false tweet; to make it difficult to determine which was which, both pairs of tweets described events from the same country. A participant saw 36 tweets and each tweet was combined with one image from among the five image types.

Note, this is different from the basic study design [Morris et al., 2012]. In that experiment a given participant saw only one of the five user images associated with all tweets; the authors needed to repeat the experiment by adding an extra image type because they found participants did not pay attention to the user image as the same image type appeared with each tweet. Also, I considered that two image types was not enough and a participant needed to see all image types to be able to see the differences.

All participants were anonymous in this study. A statement was provided to the participants before beginning the study, and this included information about the task, what they will do (rating tweet credibility with different levels of credibility), tweet topics, the length of time the task might take and demographic information. Participants had the right to decline to participate and to withdraw at any time.[2]

User images were assigned to each tweet randomly and, for the default image case, the same image appeared each time. Each participant never saw the same tweet, image or username more than once.

3.4.1 The CrowdFlower Platform

CrowdFlower is a platform to run user studies and annotating data for building ground truth. It was used here to recruit participants. Instructions were provided on how to complete a task: participants were notified that any URLs were not clickable and that they should not try to leave the current web page to do additional searches to help them to verify the information in tweets. To ensure the quality of judgments for participants, I inserted five gold questions and informed the participants that if they were able to answer them correctly, their answers would be accepted, I followed the platform instructions for designing gold questions, which were designed in the same way as the dataset. They appear as part of the dataset, but were only used to ensure answer quality.

The tweets were shown to the participants in random order; under each tweet there were two Likert scales with seven points ranging from 'strongly disagree' to 'strongly agree'. The first statement asked participants to give a rating on whether 'this tweet contains credible information', the second statement were about the tweet author 'this author is credible'. Participants were provided with a credibility definition as 'offering reasonable grounds for being believed',[3] which is the definition used in past credibility research [Castillo et al., 2011,

[2]The ethics number for this study is ASEHAPP 47-13.
[3]http://www.merriam-webster.com/dictionary/credible

Shariff et al., 2014]. The tweet was said to be credible when the participants believed the truthfulness of its content.

Within the study and when the participant had finished rating the tweets' and authors' credibility, there were another four questions: demographic information including gender and age, the number of times participants had used Twitter and the technique they used to assess the credibility of tweets.

3.4.2 Participants

The study collected 1,416 judgments from 59 participants, 708 judgments for tweet credibility and the same for tweet author credibility (see Section 3.4.1), each tweet received at least nine judgments for each tweet and its author. The gender distribution was 83.01% male and 16.95% female. The participants' ages were 18-24 years, 27.12%; 25-34, 40.68%; 35-44, 16.95%; and >45 years, 15.25%.

Participants were also asked how they normally read tweets; the results are shown in Figure.3.2. As can be seen from the results, around 95% of participants had experience in using Twitter, which precluded any effect of participants' lack of knowledge on the results. The participants were asked about the techniques they used for clarifying the credibility of tweet content about events; they were provided a text box to explain one or many techniques. I grouped the answers into four categories as follows: used Internet to verify the information (Google, newspaper and review press); examined author features (author profile, author username, location and description); examined tweet language (language of the tweet, grammar, style of writing and included links); and used common sense and general knowledge. Some of the participants used more than one technique at the same time, such as assessing the author name and comparing it with their knowledge. From the participants' responses, there was no dominant technique used to validate the credibility of the tweet contents and answers' distributions were close between the four categories.

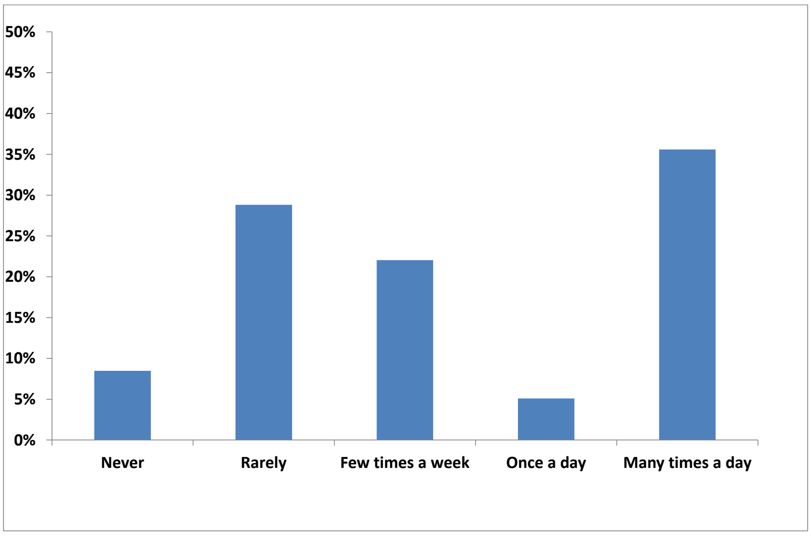


Figure 3.2: Twitter usage among participants.

3.5 Results

Since the participants in the experiment provided one credibility rating for the tweet and another for the author, calculation of correlation coefficient showed the ratings to be closely correlated. I obtained a Pearson correlation value of $R = 0.92$ ($p < 0.001$), which was higher than in the original study [Morris et al., 2012] ($R = 0.85$). The tweet and author credibility means were 4.79 and 4.88, respectively, compared with 3.79 and 3.27 in the original study.

Table 3.1: The level of credibility for each topic.

	Tweet	Author
Politics	4.86	5.00
Science	4.85	4.92
Entertainment	4.79	4.87

3.5.1 Tweet Topic

I calculated the mean ratings for all three topics; the tweet credibility rating for politics was the highest at 4.86, with 4.85 for science and 4.79 for entertainment. For the author credibility ratings there were mean ratings of 5.00, 4.92 and 4.87 for politics, science and entertainment, respectively, as shown in Table.3.1.

3.5.2 User Name

The highest credibility mean was received by the traditional-style, tweet and author credibility means were 5.02 and 5.08, respectively. Then, the topical-style received means of 4.93 for both tweet and author credibility. The lowest credibility means were given to the internet-style, as it received 4.50 for tweet credibility and 4.60 for author credibility, as in Table.3.2.

Table 3.2: The level of credibility for each username style.

	Tweet	Author
Traditional	5.02	5.08
Topical	4.93	4.93
Internet	4.50	4.60

3.5.3 Profile Image

The highest credibility rating was for a male image (mean $_{\text{tweet}}$ = 5.14 and mean $_{\text{author}}$ = 5.26); the lowest image rating was for the default and topical images (mean $_{\text{tweet}}$ = 4.00 and mean $_{\text{author}}$ = 4.11), see Table.3.3. The female image was the second most effective type influencing readers' credibility perception (mean $_{\text{tweet}}$ = 4.92 and mean $_{\text{author}}$ = 4.94) with no difference from the male image ($p > 0.05$). The generic image was the third type (mean $_{\text{tweet}}$ = 4.47 and mean $_{\text{author}}$ = 5.60).

Table 3.3: The level of credibility for each image type.

	Tweet	Author
Male	5.14	5.26
Female	4.92	4.94
Generic	4.47	4.60
Topical	4.00	4.11
Default	4.00	4.11

3.5.4 Location

The mean rating was approximately the same for tweets and author credibility with or without location. Tweets with location were rated 4.85 for credibility and 4.90 for author credibility, compared to tweets without location were received credibility means 4.81 and 4.99 for tweet and author, respectively. No differences were statistically significant (Table.3.4).

Table 3.4: The level of credibility for each location type.

	Tweet	Author
Location	4.85	4.99
No location	4.81	4.90

Next, I investigated the influence of location on tweet topic. I made two tweet groups for each of politics, science and entertainment-one group for tweets with location and one for those without. I then compared the means between each topic group and computed the p-value. The tweet credibility average for the politics topic with location was 5.15 compared to 4.25 for tweets with no location on the same topic. The author credibility average for politics and location was 5.29 compared to 4.42 for the author of tweets on the same topic and no location. After performing a t-test ($p < 0.001$) for both tweet and author credibility, for science there was no significant difference between the two groups. For entertainment tweets, those with no location had higher mean ratings than those with location, for both tweet and author credibility rating. Refer to Table.3.5.

I followed the same procedure for username with location as in Table.3.6. The effect of

Table 3.5: Location v. topic

	Tweet	Author
Location-Politics	5.15	5.29
No location-Politics	4.25	4.42
Location-Science	4.77	4.78
No location-Science	4.94	5.10
Location-Entertainment	4.49	4.46
No location-Entertainment	5.17	5.37

author location on the credibility perception of readers was weak; Internet style was the most influenced type but there were no significant differences between location and no location for either tweet or author.

Table 3.6: Location v. username

	Tweet	Author
Location-Traditional	5.02	5.08
No location-Traditional	4.87	5.00
Location-Topical	4.94	4.93
No location-Topical	5.00	5.13
Location-Internet	4.20	4.23
No location-Internet	4.50	4.79

The same procedure was applied for the profile image with location, as shown in Table.3.7. The image types influenced by location were topical and default. There was a difference between the two location types in topical image mean ratings $_{\text{tweet\_location}} = 5.17$, mean $_{\text{author\_location}} = 5.09$ and mean $_{\text{tweet\_No\_location}} = 4.00$ and mean $_{\text{author\_No\_location}} = 4.56$, $p < 0.05$). Also, there was a significant difference between the location types for the default image (mean $_{\text{tweet\_location}} = 4.14$, mean $_{\text{author\_location}} = 4.10$ and mean $_{\text{tweet\_No\_location}} = 4.07$ and mean $_{\text{author\_No\_location}} = 3.85$, $p < 0.05$).

Table 3.7: Location v. image.

	Tweet	Author
Location-Topical	5.17	5.09
No Location-Topical	4.00	4.56
Location-Female	4.86	4.84
No location-Female	5.44	5.37
Location-Male	5.04	4.99
No location-Male	5.49	5.30
Location-Generic	4.48	4.52
No location-Generic	4.68	4.45
Location-Default	4.14	4.10
No location-Default	4.07	3.85

3.6 Discussion

The results of the experiment indicate the importance of non-textual features in influnce readers' credibility judgments. Not all features have the same effect. For example, user profile images have different effects on readers' credibility judgments. Wei and Stillwell [2017] showed that a user's profile images in social media can be used to predict their intelligence using readers' perceptions. Also, I found that the author's username determines the readers' credibility perception: for example, the traditional style was judged as the most credible type, while Internet was the lowest one. This result confirms the Pal and Counts [2011b] finding regarding the effect of author username on the authoritativeness of the author.

Moreover, the results show that the importance of some features comes when they interact with another feature. For example, I found that location provides power and is more influential than any other feature in political tweets. This indicates that this feature has a greater influence than any other, especially if it is an automatically generated feature such as the tweet geolocation, which authors cannot change. Moreover, the results showed that authors within the country of such an event had a high credibility score. This may be because they are more likely to be witnesses and to give more accurate information. Kumar et al. [2013] studied the importance of author location to identify 'information leaders' at times of events such as earthquakes. Eyewitness accounts in microblogs like Twitter are a popular area of research that is still at an early stage as stated in Chapter.2. Localised tweets may include more accurate information than globalised tweets about an event. Thus, readers perceive tweets that are sent from nearer to an event as highly credible. In Chapter.5 I examine the effect of distance between source and event on credibility.

Now that search engines such as Google and Bing incorporate social media updates in their results, this will add value to search engines when retrieving the most relevant and credible results (tweets) in our case. Social search (search within social networks) itself will need to improve the credibility of retrieved results. In systems like Twitter, popularity is the main feature used to retrieve results for any query topic type. Although popularity is one of the main features providing an indicator of credibility in the social media environment, in many events most tweets come from unpopular authors, which highlights the need to study authors' individual features.

The findings of this study can be used by designers who want to provide user experiences with credible content to final users. For instance, with regard to adapting friend recommendations and social search results for end users, if users want simply to see certain names, images, locations and so on, then presenting the content of those users is one possible solution. The challenge is how to present highly credible content from those authors, and at the same time have end users perceive credibility as well. This is one of the possible areas for future research.

It was clear that readers' credibility judgments were biased because their judgments were based on the author's metadata and not on the content. Thus, one possible way to reduce the existence of bias is to provide the end user with extra information about the author, such as bio, profile location, expertise and registration date.

3.7 Summary

In this chapter, I studied four features that influence user credibility judgments with a particular focus on location and its relationship with other features. I examined two research questions:

1. Which features affect readers' perceptions of credibility?

    - How does tweet location affect credibility?

I investigated how certain features affect readers' perceptions of the credibility of tweets. Using a crowdsourcing experiment, I found that readers perceive the credibility of tweets as influenced more by some features than by others. For example, among five profile image types, a male profile image had a significant effect on readers' credibility perception. Most notably, I discovered that displaying the location of certain types of tweets causes readers viewing these tweets to perceive the tweets as more credible. I found that Twitter users believe politics tweets that include event location more than those that do not include location, and the same is true for the topical style of the username.

Chapter 4

Beyond the Culture Effect on Credibility Perception on Microblogs

4.1 Introduction

The way in which users interact with social media to find information has changed over time and new techniques have now become popular. For instance, microblog users are increasingly using keyword searches and hashtags to find information, rather than relying only on their follows or other users followed by a user-the latter typically being a more trustworthy information source [Morris et al., 2010]. In 2011, 21.28% of Twitter queries contained hashtags [Teevan et al., 2011]. These mechanisms for finding microblog information allow users to see content from strangers that they did not see previously. As a result, end users are becoming wary of content in social media and readers of tweets, for example, are concerned about tweets from authors they do not know [Morris et al., 2012].

With the many credibility challenges that microblog and other social media networks are suffering from, understanding how microblog content is perceived by readers in terms of credibility is essential.

To date, most of research that has examined the effect of microblog features on reader credibility perception has been general, for example, classifying readers as one homogenous group, regardless of their background. Only a few studies have focused on understanding the effect of user differences on microblog credibility perception. In social media usage, users – authors and readers – behaviour has been found to vary based on language, culture and country. Twitter as an example of microblog platforms is used by a massive number of users, who are diverse in term of language, culture, country and location. Twitter supports more than 40 languages and most of its users are from outside the US.[1]

The influence of culture on user behaviour has been well studied [Hofstede, 2011], and found to be an important factor predicting behaviour in social question and answer forums [Yang et al., 2011]. Readers' credibility perceptions are embedded in specific social and cultural contexts, including individual and collective preferences, emotions and other differences. Yang et al. [2013] examined the effect of readers' countries on their credibility perceptions in social media. They compared readers from the US and China with regard to how they judge the credibility of the same social media content, finding significant differences in credibility judgments between microblog readers from the two countries. However, as the research included only two countries from different cultures, it was not able to address whether the measured credibility perception difference was due to the readers' actual locations, countries or culture.

The actual behaviour of authors in social media, such as number of tweets, retweets and inclusion of hashtags differs between authors because of their country and language differences. Poblete et al. [2011] studied the behaviour of millions of authors from different countries using a Twitter dataset considering the use of features such as hashtags, URLs, mentions and retweets. The researchers found differences in the use of these features across countries with similar cultures. Moreover, Weerkamp et al. [2011] analysed tweets in eight languages, six of which were European and two, Asian (Dutch, English, German, French, Portuguese, Spanish, Indonesian and Japanese). Significant behavioural differences were found between authors of different languages, and even between languages of the same

[1] https://about.twitter.com/company (accessed 4 November 2015)

culture. Further, in a very large-scale analysis using more than 60 million tweets, Hong et al. [2011] compared among authors of different languages considering behaviours of using five features: hashtags, URLs, mentions, retweets and replies. They found, for example, that use of 'replies' varied within both the Western culture (usage was 50% for Dutch but only 36% for German) and the Eastern culture (59% for Korean and only 20% for Indonesian). All these studies together indicate that the effects of country, culture and language are clear in social media user behaviour; however, there is a need to differentiate between the effect of each of them, as it is clear that there is ambiguity regarding which is more prominent than the others in user behaviour.

However, there has been no research studying the effect of reader location, including their actual location and culture, in social media credibility. Despite the fact that researchers have studied the effect of country differences on credibility perception in social media, there are a number of limitations that need to be addressed: they examined small numbers of countries and generalised the findings at a cultural level; and the influence of the actual readers' locations has not yet been tested. Most research to date on the credibility perception of social media has treated readers as one group or has considered only the English language. In this chapter, I investigate Arabic and English (American) microblog readers.

This chapter aims to investigate the interrelationships among five components – culture; country; region of groups of countries; regions; and divisions of the same country – and their effects on tweet readers' credibility perception. Readers categorised by different cultures might be expected to have different credibility perceptions, but the credibility perceptions of tweet readers in different countries with the same cultural characteristics have not been investigated. Moreover, in countries that are large in area and population, readers from different regions and divisions may not have the same credibility perceptions.

To the best of my knowledge, this is the first study in social media credibility that differentiates between the effects of culture, country, region and division on how reader perceive the credibility of social media content. Defining the cause of differences is essential to build a credibility model of social media in cultural or national contexts, or both.

Accordingly, this chapter addresses the following research questions:

1. What is the effect of a reader's location on their perceptions about the credibility of social media information?

- What is the effect of a reader's culture on their credibility perception in social media?
- Will a reader's nationality influence their credibility perception of social media among other members of the same culture?
- What is the effect of the geographical region of countries with the same culture on readers' credibility perception of social media?
- How do different regions and divisions in large countries affect readers' credibility perception of social media?

4.2 Methodology

I designed a study using a crowdsourcing platform to examine the credibility perception of readers from the Arabic culture in countries located in the Middle East and North Africa, where most people speak Arabic; and readers from the American culture in different US regions and divisions.

To examine the effect of culture, country, region and division on the credibility perceptions of readers in social media, I used features for a tweet's author including gender, image, username, location and network overlap. These features have previously been found to have a significant effect on credibility perception [Aladhadh et al., 2014, Morris et al., 2012].

An overview of the cultures, countries, regions and divisions included in the study is provided next.

4.2.1 Arabic

Arabic is widely spoken, with around 400 million speakers, and has been ranked as the seventh most – used language on the Internet.[2] Social media networks engage millions of Arabic users who participate and share information. Twitter, is one of the platforms that provides information on different topics – such as breaking news, health and entertainment – includes around 6 million active Arabic users, with a daily production of 17 million tweets. [3]

The current study included eight countries for analysis. I could not include other countries because of a number of limitations: 1. some countries did not have large enough numbers in terms of participants on crowd platforms; and 2. other countries suffered from instability. However, the countries chosen were members of the League of Arab States:[4] Saudi Arabia, United Arab Emirates, Egypt, Jordan, Palestine, Algeria, Morocco and Tunisia. They are representative of the classical Arabic regions,[5] as follows:

- The Arabic Peninsula or Gulf Cooperation Council (GCC) region includes Saudi Arabia and the United Arab Emirates.
- The Levant region includes Jordan and Palestine.
- The Nile valley region includes Egypt.
- The Maghreb or North Africa region includes Morocco, Algeria and Tunisia.

Population numbers vary among these countries. Egypt is the largest Arabian country in terms of population, with a population exceeding 90 million, which is around one-quarter of the total population of the Arabic countries. The combined populations of these eight countries make up 60% of that of all Arabic countries.

[2] https://www.internetworldstats.com/ (accessed 4 March 2017)
[3] http://www.arabsocialmediareport.com/home/index.aspx (accessed 8 November 2016)
[4] http://www.lasportal.org/en/aboutlas/Pages/CountryData.aspx (accessed 10 April 2015)
[5] https://en.wikipedia.org/wiki/Geography_of_the_Arab_League (accessed 4 April 2015)

Many factors lead to the final shape of a culture, including language [Jiang, 2000], economics [Klamer, 1997], religion [Foucault, 2013] and location change [Kitayama et al., 2009]. As the Arab World extends from the Arabian Gulf to Morocco on the Atlantic Ocean, these countries may differ in population, economy and so on. However, all people in these countries historically share the same culture.[6] Wilson [1996] and Dedoussis [2004] considered that these countries in general are homogeneous with regard to culture, Obeidat et al. [2012] treated them as one identity and Hofstede [1991] referred to the Arab culture as the 'Arab group'.

4.2.2 United States

The US has a large population of over 328 million,[7] and they are heavily engaged in social media usage. For example, there are 69 million active Twitter users from the US. [8] The US has four main regions (Northeast, Midwest, South and West) as defined by the US Census Bureau,[9] as shown in Figure.4.1. Each region has a number of divisions as well: there are nine divisions within the US regions as follows:

- The Northeast region includes the 'New England and Middle Atlantic' divisions.
- The Midwest region includes the 'East North Central and West North Central' divisions.
- The South region includes the 'West South Central, East South Central and South Atlantic' divisions.
- The West region includes the 'Mountain and Pacific' divisions.

[6] https://en.wikipedia.org/wiki/Arab_culture#cite_note-1 (accessed 16 May 2016)
[7] https://www.census.gov/popclock/. (accessed 11 May 2015)
[8] https://www.statista.com/statistics/274564/monthly-active-twitter-users-in-the-united-states/. (accessed 28 June 2018)
[9] https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf. (accessed 25 June 2015)

Note: The Pacific region, which includes 'Alaska' and 'Hawaii', was under the 'Pacific' division within the 'West' region.

The 50 states are distributed within the nine divisions as follows:

- New England includes 'the states of Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island and Vermont'.
- Middle Atlantic includes 'the states of New Jersey, New York and Pennsylvania'.
- East North Central includes 'the states of Indiana, Illinois, Michigan, Ohio and Wisconsin'.
- West North Central includes 'the states of Iowa, Nebraska, Kansas, North Dakota, Minnesota, South Dakota, and Missouri'.
- South Atlantic includes 'the states of Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia and West Virginia'.
- East South Central includes 'the states of Alabama, Kentucky, Mississippi and Tennessee'.
- West South Central includes 'the states of Arkansas, Louisiana, Oklahoma and Texas'.
- Mountain includes 'the states of Arizona, Colorado, Idaho, New Mexico, Montana, Utah, Nevada and Wyoming'.
- Pacific includes 'the states of Alaska and California, Hawaii, Oregon and Washington'.

Thus, these states make up the divisions, the divisions make up the regions and the regions make up the US.

This study aimed to understand the influence of readers' locations on their credibility perceptions in social media, and whether regional and divisional differences can influence their credibility judgments. Dividing readers into groups based on their location will help

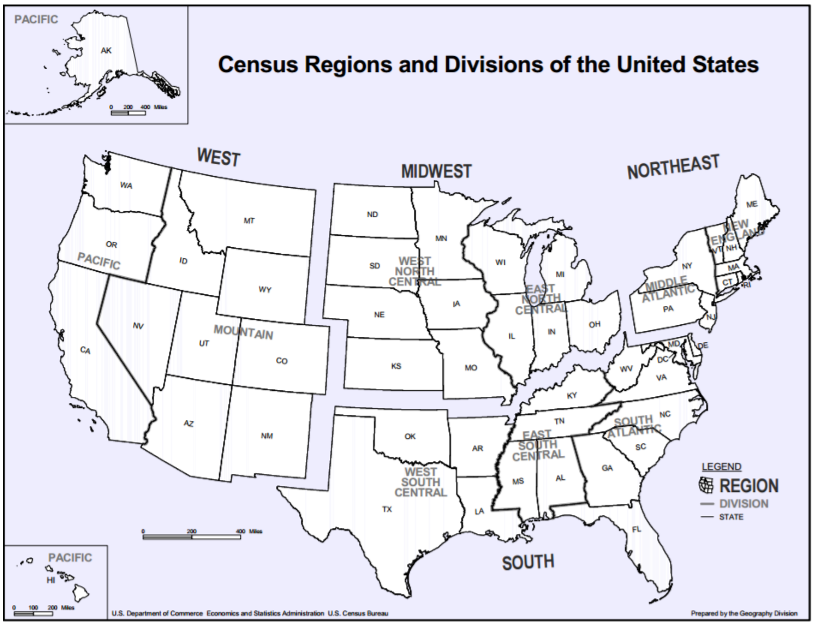


Figure 4.1: US regions and states [Commerce, 2015].

to understand at what level we can say a large country is representative of one credibility perception pattern, which will help in designing the user experience for social media users and supporting their need in evaluating content credibility.

4.2.3 Language

Including different languages in a study of credibility is important because languages have different styles, which might affect readers' perceptions. In Twitter, a user can choose their language and the user interface will automatically become compatible with it. For example, if a user chooses the Arabic language, the text and other features will be oriented from right to the left, as in Figure.4.2: the translation of the tweet is 'New study shows that late and little sleeping might be an early sign of heart disease'. The language of the user determines the interface design in Twitter, including user characteristics such as language, culture, country and region. These factors are principal features in establishing credibility perceptions, the main purpose of this research.

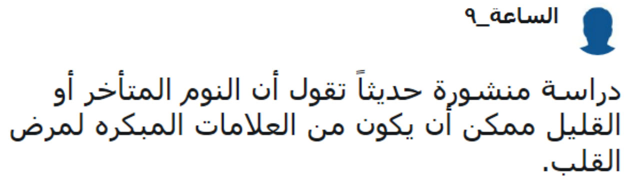


Figure 4.2: Arabic tweet example

4.2.4 Features Examined

I next describe the features used in the experiments. The selected features were related to the authors of tweets, which include their gender, profile image, username, location and network overlap, as explained in the following.

4.2.4.1 Gender

Gender has been studied previously in blogs [Armstrong and McAdams, 2009]: male authors are perceived to be more credible than female authors. Morris et al. [2012] found that gender influences reader credibility perception. They found that readers perceive male-authored posts as significantly more credible than female ones, similar to another study [Yang et al., 2013]. To indicate gender of tweet authors, an image of a male or female was used and usernames were selected that were applicable for each gender.

H1: Male authored tweets will be perceived as more credible than female tweets.
H1a: Culture, country, regions (Arabic and US) and US divisions will change the reader's overall behaviour towards the gender of the author.

4.2.4.2 Profile Image

Profile pictures affect readers' judgments [Kang et al., 2015], as observed in many microblog platforms (Twitter and Reddit). Different types of profile images have different credibility judgment effects. In this study, I followed the methodology of [Yang et al., 2013] in limiting images to two styles: a general (anonymous) image representing male or female; and real photos for both genders, as in Figure.4.3 (b and d). Real photos pertained to the relevant culture, showing the headshot of a young adult. For Arabic authors, I used real images for Arabian males and females, and for authors from the US, I selected Caucasian male and female photos, as in Figure.4.3.

H2: Authors with real profile photos will be perceived as more credible than those with anonymous photos.

H2a: Photos indicating culture, country, region (Arabic and US) and US division will influence reader's credibility perception.

4.2.4.3 Username

The name presented next to an author's profile image has been found to be affect readers' credibility perception. Pal and Counts [2011b] found the perception of readers regarding the quality of content are often based solely on the author's name, tweets with a popular authors' usernames are rated more highly credible than when the same tweets are anonymous (no author username). In this study, two name styles were used: a topical username (e.g. Politics_News) and an Internet style name (e.g. Morning_7am), which is neither traditional nor topical.

H3: Topical usernames will be perceived as more credible than Internet style usernames.

H3a: Usernames indicating culture, country, region (Arabic and US) and US division will influence credibility perceptions of readers.

4.2.4.4 Location

The author's location has been found to be an important feature in microblog credibility perception, especially in political topic [Aladhadh et al., 2014]. Authors from the same country in which an event occurred are perceived as more credible by readers than those with no location indicated in their tweets. Another study used liberal and conservative locations to study the interaction between location of authors and readers' microblog credibility perceptions' [Yang et al., 2013]. This classification was based on a Gallup Poll, which lists the most liberal and conservative states in the US.

In this study, I considered the size of a location: authors from large locations tend to share information more than those from small locations [Kang et al., 2015]. Each tweet identified the location of authors, including the name of their country/state and city. For Arabic tweets, I chose two large cities and two smaller cities for each country: for example, the large cities in Saudi Arabia were Riyadh and Jeddah; and the small cities were Al Qunfudhah and Yanbu. For Arabic countries I used Wikipedia to define the large and small cities based on population: the highest population cities were counted as large cities, and the lowest population cities became the small cities. For US, I adopted a methodology from previous research [Yang et al., 2013]. I chose 16 states; two large cities were chosen from each of eight states and two small cities from each of the other eight. For example, Seattle, Washington was classified as large and Manti, Utah as small. Wikipedia defines large and small cities based on population.

H4: Tweets authored in large locations will be perceived as more credible than those authored in small locations.

H4a: H4 will be affected by culture, country, region (Arabic and US) and US division.

H4b: Location is a topic-dependent feature, as found by [Aladhadh et al., 2014]. Location styles will vary based on topic, with significant interactions.

4.2.4.5 Network Overlap

Twitter provides a social network among users defined by users following other users. Poblete et al. [2011] studied social connections among Twitter users from 10 countries, and found that the connectivity among users in some countries such as South Korea, Japan and Canada was significant compared to that in some other countries including the US.

I aimed to measure the effect of connectivity on the credibility perceptions of readers. In other words, to examine how social proximity between tweet authors and readers may influence reader credibility judgments. I used two conditions-overlap and no overlap-inspired by [Yang et al., 2013]. I stated to readers ‘Imagine this is the number of your friends who are following this author’. I generated a random number of friends, as shown in Figure.4.3.

H5: Tweets authored by authors with a network overlap with the reader will be perceived as more credible than tweets with no overlap.

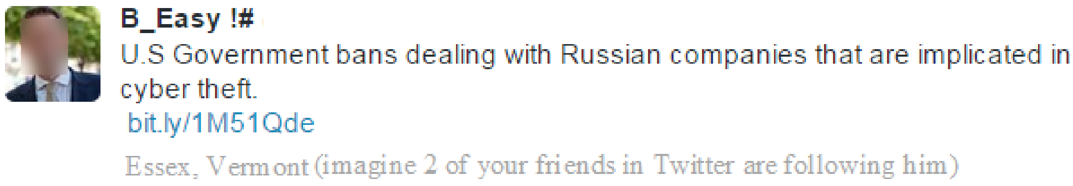
H5a: Culture, country, region (Arabic and US) and US division will have a significant interaction with **H5**.

H5b: Overlapping styles will vary based on topic, with significant interactions, as has been found in past work [Yang et al., 2013].

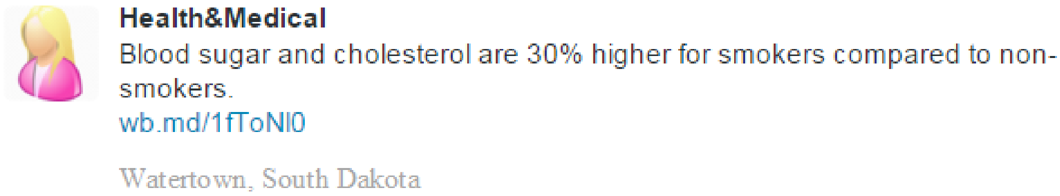
4.3 Experimental Design

I followed the methodology developed by [Yang et al., 2013]. I examined five author features: gender, location (large, small), username (topical, Internet), profile image (general, photo), and network overlap (overlap, no overlap). For each feature there were two conditions. Using a Latin square design, the number of tweets was 32 ($2 \times 2 \times 2 \times 2 \times 2$). For each language – English and Arabic –, I authored 32 political tweets and 32 health tweets. Accordingly, I authored 128 tweets for the purpose of this experiment. Each participant from each culture read the tweets of one topic, either politics or health. All political tweets were about local events in the respective countries to make the tweet more relevant to the participants. The health tweets were written in English and the same tweets were translated to Arabic [Morris et al., 2012, Yang et al., 2013].

The experiment consisted of two parts: judging tweet credibility; and a survey on reader demographic information. When judging credibility, participants were asked whether they thought ‘This tweet contains credible information’. They answered using a seven-point Likert scale from strongly disagree to strongly agree. Each participant needed to rate 32 tweets on one topic, tweets were presented randomly. For the survey part, I asked participants about their demographic information (gender, age and educational level). I also asked them about



(a)



(b)

(c)

Figure 4.3: Sample tweets: a) US, politics, male, real photo, Internet style username, large location, overlap; b) US, health, female, anonymous photo, topical style username, small location, no overlap; c) Arabic, health, female, real photo, topical style username, small location, overlap.

their usage and opinion of the credibility of Twitter as a news source for the seven topics.[10]

4.3.1 Tweet Contents

All simulated tweets were false but plausible, thus precluding participants' previous knowledge from affecting their judgments. Groups of native Arabic and English speakers reviewed all tweets to ensure there were no grammatical mistakes that would affect readers' credibility judgments. I repeated the checking and re-authoring process until all tweets were deemed readable and plausible. This approach has been taken in prior studies [Aladhadh et al., 2014, Morris et al., 2012].

All author features (gender, image, username, location and network overlap) were randomly combined for each tweet. Location and network overlap needed to be added to the tweet. I presented network overlap information next to location. Each username, photo and location was presented only once, so I prepared a sufficient set for each feature.

4.3.2 Experiment

I recruited participants via the CrowdFlower platform[11]. The platform gives an option to choose participants countries. I restricted participants to the nine studied countries. Note that Palestine is defined as a country by CrowdFlower. I specified language capability for the Arabic participants to be Arabic. Moreover, each participant was asked to specify their language. Only those who responded 'Arabic' were further considered for this study. For participants from the US, only those who nominated their first language as 'English' were used. A number of gold questions were inserted into the study, and only those participants who answered these questions with >80% accuracy were included further.

[10] The ethics number for this study is ASEHAPP 35-15.
[11] https://www.crowdflower.com

Table 4.1: Participant distribution across cultures (Arabic and English), countries (Arabic countries), Arabic regions, US regions and US divisions.

Cultures	Freq.	Arabic Countries	Freq.	Arabic Regions	Freq.	US Regions	Freq.	US Divisions	Freq.
Arabic	543	Saudi Arabia (SAU)	56	Gulf Cooperation Countries, GCC (SAU & UAE)	91	West	115	East North Central	51
		United Arab Emirates (UAE)	35					East South Central	22
		Jordan (JOR)	34	Levant (JOR & PSE)	56	Midwest	66	Middle Atlantic	56
		Palestine (PSE)	22					Mountain	30
American	405	Egypt (EGY)	243	Nile valley (EGY)	243	Northeast	101	New England	25
		Morocco (MAR)	24					Pacific	80
		Algeria (DZA)	78	North Africa (MAR & DZA & TUN)	153	South	123	South Atlantic	79
		Tunisia (TUN)	51					West North Central	21
								West South Central	41

Table 4.2: Participant demographics.

Demographic	Arabic	American
Male	72%	63%
Female	28%	37%
18-24	35%	35%
25-34	38%	37%
35-44	15%	17%
45-above	12%	11%
Less than high school	2%	1%
High school	14%	11%
Diploma	22%	32%
Bachelor degree	49%	44%
Master degree	11%	11%
PhD	2%	1%

4.4 Results

I received 30,336 judgments from 948 participants. Table.4.1 shows the spread of participants among the two cultures, the Arabic countries, and US regions and divisions. Table.4.2 shows that the distribution of participants across the two cultures was almost equal and the genders, age distributions and educational levels for the two cultures had similar proportions.

I asked participants about the nature of their usage of Twitter as a news source on seven topics. Each topic was rated on a five-point Likert scale from never to >90%. No significant differences were found in overall usage between Arabic and English. Table.4.3 presents the mean usage for each culture on different topics, along with results of a t-test across all topics. Figure.4.4 shows that usage behaviour of microblogs as a news source differed more

Table 4.3: Twitter usage means across seven topics.

Topic	Arabic$_{use}$	English$_{use}$	P
Breaking news	2.81	3.03	0.012*
Celebrity	3.18	2.81	0.000**
Emergency	2.74	2.93	0.040*
Health	2.58	2.38	0.022*
Politics	2.92	2.81	0.240
Product	2.62	2.73	0.198
Science & Technology	3.12	2.65	0.000**

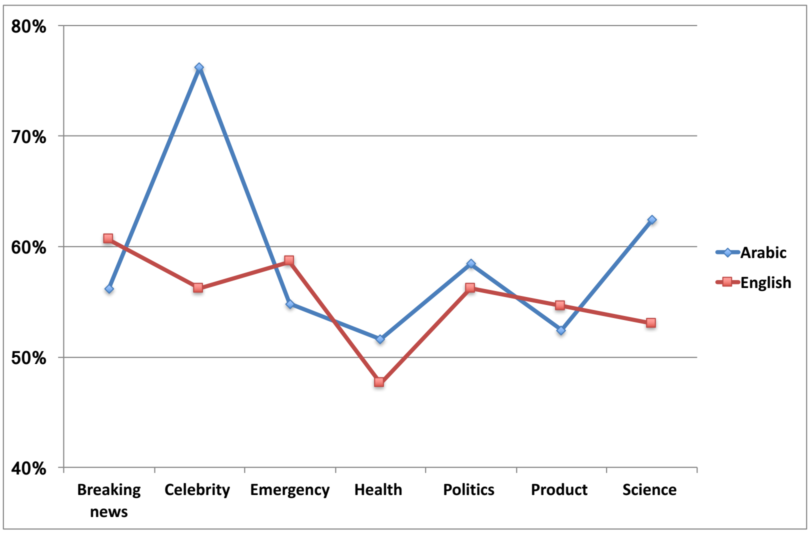


Figure 4.4: Means of microblog usage importance as a news source for different topics.

between the two cultures for the topics celebrity and science & technology ($p < 0.001$) than for breaking news, emergency and health ($p < 0.05$); there were no differences for politics and products.

Participants were asked about the credibility of tweets on each topic. They used a five-point Likert scale with options graded from 'not credible' to 'very highly credible'. The significant differences are shown in Table.4.4.

The usage and credibility ratings for all topics across cultures showed that the credibility

Table 4.4: Twitter credibility means across seven topics.

Topic	$Arabic_{credibility}$	$English_{credibility}$	P
Breaking news	3.04	3.23	0.001*
Celebrity	3.27	3.06	0.003**
Emergency	3.01	3.32	0.000**
Health	3.11	3.02	0.187
Politics	2.98	2.85	0.053
Product	3.07	3.19	0.056
Science & Technology	3.45	3.31	0.056

rating was correlated with usage rating, with a Pearson correlation coefficient of (R Arabic = 0.44 and R English = 0.53, $p < 0.001$). The credibility rating was higher than the usage rating for all topics in both cultures. Readers from both cultures found tweet content to have credible information even when they did not often use Twitter to find news on that topic. For example, in the US culture,'Science & Technology' was one of the least used topics, but at the same time was one of the most credible. Similarly, Arabic participants did not use Twitter as a news source for 'Health' as much as other topics, but they found health information on Twitter to be credible. Usage of Twitter as a news source for the seven topics was similarly high in both cultures. These results improve confidence that participants' assessments were not affected by a lack of knowledge of the platform.

Figures.4.5 and 4.6 show the amount of time participants spent reading and posting tweets. The two groups had a similar distribution among usage categories. The biggest reading category size was 'few times a week', while 'rarely' was the biggest group in posting behaviour. This also makes the sample homogenous in terms of Twitter usage among participants.

4.4.1 Results of the Experiment

I used a mixed design ANOVA (within and between predictors) to analyse and test the effect of all features and their interactions on the credibility rating of tweets. This was done five times: between cultures (Arabic and US); among Arabic countries; among Arabic regions;

among US regions; and among US divisions.

Sections.4.4.2 to 4.4.6 provide ANOVA results for author features and their interaction with cultures, Arabic countries, US regions, Arabic regions and US divisions are . ANOVA analyses were performed to test all hypotheses and the effect of the experimental manipulations on readers' credibility judgments, where the demographic variables of gender, age and educational level were controlled. This was followed up with pairwise t-tests when appropriate and Bonferroni corrections were used to mitigate the effect of multiple comparisons. Please note that for all results in the tables * p<0.05 and ** p<0.001.

For cultures, Arabic countries, US regions, Arabic regions and US divisions, the results for each hypothesis from the ANOVA model are presented as means with p-values. Where an interaction was significant, the direction of differences was checked and means with p-values are reported in tables for readability. Arabic country codes are used instead of complete names.

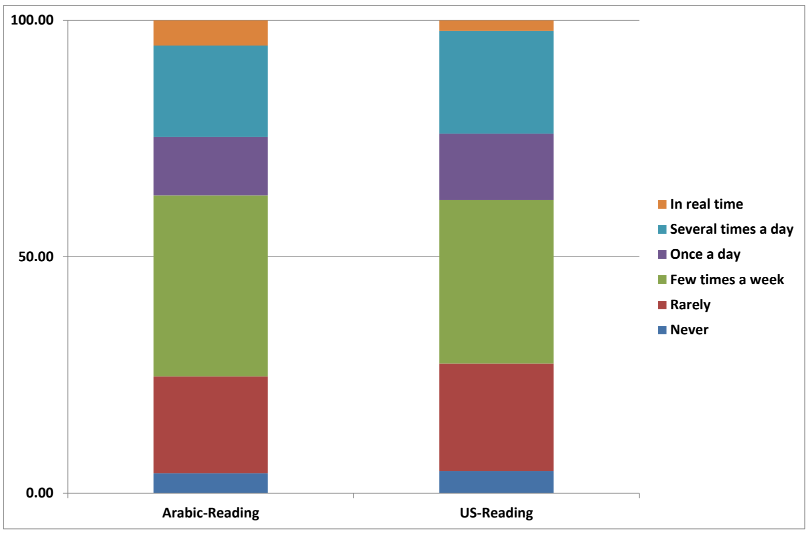


Figure 4.5: Distribution of participants' reading in each culture.

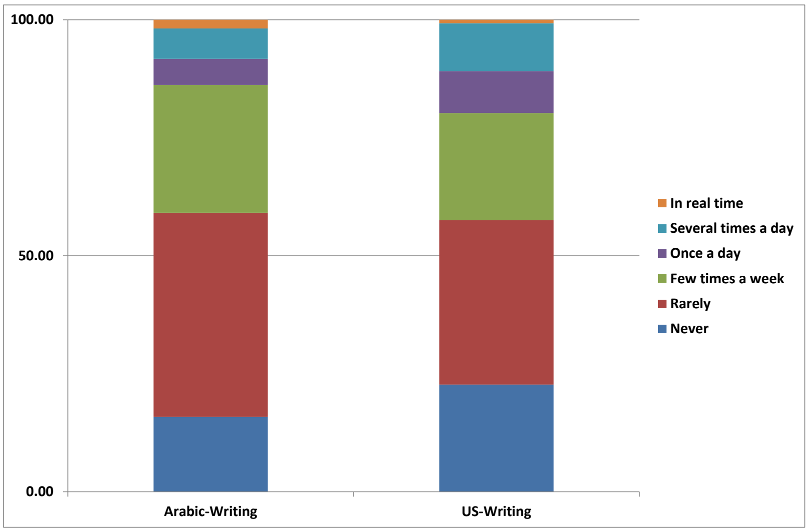


Figure 4.6: Distribution of participants' writing in each culture.

4.4.2 Interaction of culture with author's profile features

As Table.4.5 shows, the difference between cultures was significant in credibility judgments. These differences are now explored for each feature.

4.4.2.1 Gender

(H1: supported) Readers regarded tweets from males (mean $_{\text{male}}$ = 4.63) as more credible than female tweets (mean $_{\text{female}}$ = 4.34) with $p<0.001$.

(H1a: not supported) The interaction between culture and gender was not significant ($p = 0.284$) and culture did not affect readers' credibility perceptions according to an author's gender. The effect of topic interaction with gender and culture was explored, revealing a significant interaction ($p<0.001$). The two cultures had similar credibility perceptions regarding male authors in politics, whereas for the health topic there was a significant difference between the two cultures: Arabic readers perceived the credibility of male authors as higher than did US readers. The opposite was true for female authors: for political tweets

Table 4.5: ANOVA results for cultures.

Factor	DF	*F*	P
Culture	1	9.450	0.002*
Topic	1	88.170	0.000**
Gender	1	157.020	0.000**
Image	1	15.080	0.000**
Username	1	70.580	0.000**
Location	1	0.250	0.617
Overlap	1	39.850	0.000**
Culture*Gender	1	1.150	0.284
Culture*Image	1	84.180	0.000**
Culture*Username	1	0.002	0.962
Culture*Location	1	0.759	0.384
Culture*Overlap	1	5.980	0.015*
Topic*Culture	1	0.913	0.340
Topic*Gender	1	8.470	0.004*
Topic*Image	1	0.886	0.347
Topic*Username	1	0.886	0.690
Topic*Location	1	8.070	0.005*
Topic*Overlap	1	10.110	0.002*
Culture*Topic*Gender	1	17.020	0.000**
Culture*Topic*Image	1	3.610	0.580
Culture*Topic*Username	1	12.350	0.000**
Culture*Topic*Location	1	31.130	0.000**
Culture*Topic* Overlap	1	116.220	0.000**

the two cultures showed a significant difference as Arabic readers saw female authors as more credible than did US readers whereas the two cultures had the same perception of female authors for the health topic, as shown in Table.4.6

Table 4.6: Interaction of culture with gender and topic.

	Politics			Health		
	Arabic	American	$P_{\text{two cultures}}$	Arabic	American	$P_{\text{two cultures}}$
Male	4.41	4.41	0.957	5.00	4.71	0.000**
Famale	4.17	3.93	0.006*	4.71	4.56	0.107
$P_{\text{two styles}}$	0.000*	0.000*	-	0.000**	0.002*	-

4.4.2.2 Profile Image

(H2: supported) The difference in readers' perceptions of credibility because of real or generic photos was significant ($p<0.001$). People perceived tweets with real photos (mean $_{\text{Photo}} = 4.52$) as more credible than those with generic ones (mean $_{\text{Generic}} = 4.45$).

(H2a: supported) The interaction between culture and profile image was significant ($p<0.001$). A real image was perceived as significantly more credible by Arabic readers than by US ones (Table.4.7), whereas US readers were more accepting of the use of an anonymous image than Arabic readers.

Table 4.7: Interaction of culture with image.

	Arabic	American	$P_{\text{two cultures}}$
Real	4.69	4.45	0.000**
Generic	4.45	4.35	0.000**
$P_{\text{two styles}}$	0.000**	0.000**	-

The interaction between profile image, culture and topic was not significant (Table.4.5). The results for Arabic readers were consistent with the findings in Table.4.6, for both topics ($p<0.01$). US readers for political tweets showed similar results in that the real image was more credible than an anonymous image ($p < 0.001$). For health tweets there was no significant difference between the two types.

4.4.2.3 Username

(H3: supported) Tweets with a topical author name were judged more credible (mean $_{\text{Topical}} = 4.57$) than tweets with an Internet style (mean $_{\text{Internet}} = 4.40$, $p<0.001$).

(H3a: not supported) The interaction between culture and username was not significant ($p<0.962$). However, adding topic to the interaction between culture and username made the interaction significant ($p<0.001$). In politics, Arabic readers were more tolerant of the

Table 4.8: Interaction of culture with username and topic.

	Politics			Health		
	Arabic	American	$P_{\text{two cultures}}$	Arabic	American	$P_{\text{two cultures}}$
Topical	4.33	4.28	0.082	4.98	4.69	0.001*
Internet	4.25	4.06	0.015*	4.73	4.58	0.064
$P_{\text{two styles}}$	0.024*	0.000*	-	0.000**	0.007*	-

use of an Internet style than were US readers, while the opposite was true for health, see Table.4.8.

4.4.2.4 Location

(H4: not supported) Tweets authored in large and small locations had no effect on readers' credibility perceptions (mean $_{\text{large}}$ = 4.49, mean $_{\text{small}}$ = 4.48, p = 0.617).

(H4a: not supported) The interaction between location and culture was not significant (p = 0.384). However, the interaction between culture, location and topic was significant ($p<0.001$). For the Arabic culture, large locations were perceived to provide more credible tweets than small locations regarding the political topic, while small locations were seen as the most credible for health topics. For both topics there was no difference between the two location types in the US culture, as Table.4.9 shows.

Table 4.9: Interaction of culture with location and topic.

	Politics			Health		
	Arabic	American	$P_{\text{two cultures}}$	Arabic	American	$P_{\text{two cultures}}$
Large	4.37	4.14	0.003*	4.79	4.66	0.098
Small	4.20	4.20	0.963	4.92	4.61	0.000**
$P_{\text{two styles}}$	0.000**	0.138	-	0.000**	0.274	-

(H4b: supported): The interaction between location and topics was significant ($p<0.05$) for politics: a large location was perceived to have a higher credibility than a small location, but the two location types did not differ in credibility for health topics (see Table.4.10).

Table 4.10: Interaction of location and topic.

	Large	Small	$P_{\text{two styles}}$
Politics	4.26	4.20	0.014*
Health	4.73	4.77	0.112
$P_{\text{two topics}}$	0.000**	0.000**	-

4.4.2.5 Network Overlap

(H5: supported) The difference between tweets authored by someone having or not having a friend connection was significant (mean $_{\text{overlap}}$= 4.54, mean $_{\text{no overlap}}$= 4.43, $p<0.001$).

(H5a: Supported) The interaction between cultural context and network overlap was significant ($p<0.05$). Both cultures perceived authors with an overlap to be significantly more credible than those without (see Table.4.11). However, US readers were more affected by network overlap than Arabic readers, as they found that tweets with overlap were significantly more credible than tweets with no network overlap. There was a less significant effect of network overlap in the Arabic culture.

Table 4.11: Interaction of culture with network overlap.

	Arabic	American	$P_{\text{two cultures}}$
Overlap	4.61	4.48	0.030*
No_overlap	4.54	4.33	0.000**
$P_{\text{two styles}}$	0.030*	0.000**	-

This interaction was still significant ($p<0.001$) when topic was added. **Table.4.12 shows that the behaviour of Arabic readers became complex when topic was added to the interaction.** In politics, tweets with no overlap were significantly the most credible type ($p<0.001$), while in health the opposite was found and tweets with overlap were the most credible type ($p<0.001$). However, US readers had consistent credibility perceptions in both topics, with a significant difference for politics ($p < 0.001$).

(H5b: supported) The interaction between network overlap and topic was significant

Table 4.12: Interaction of culture with network overlap and topic.

	Politics			Health		
	Arabic	American	$P_{\text{two cultures}}$	Arabic	American	$P_{\text{two cultures}}$
Overlap	4.20	4.32	0.139	5.01	4.65	0.000**
No_overlap	4.38	4.03	0.000**	4.70	4.62	0.376
$P_{\text{two styles}}$	0.000**	0.000**	-	0.000**	0.593	-

($p<0.001$). For the health topic, there was a large difference between overlap and no overlap, as tweets with network overlap were judged more credible than those with no overlap. However, for politics the difference between the two styles was small in comparison to the health topic, see Table.4.13.

Table 4.13: Interaction of network overlap and topic.

	Large	Small	$P_{\text{two styles}}$
Politics	4.26	4.20	0.021*
Health	4.83	4.66	0.000**
$P_{\text{two topics}}$	0.000**	0.000**	-

4.4.3 Interaction of Arabic Countries with Authors' Profile Features

Table.4.14 presents the results of the mixed design ANOVA that was applied between the Arabic countries.

4.4.3.1 Gender

(H1: supported) Male-authored tweets were perceived more credible than female tweets ($\text{mean}_{\text{male}} = 4.68$, $\text{mean}_{\text{female}} = 4.42$, $p<0.001$).

(H1a: supported) The interaction between country and gender was significant ($p=0.030$). Credibility ratings differed significantly between male and female authors ($p<0.001$) in some countries (Algeria, Egypt, Saudi Arabia and Tunisia); other countries had the same gender credibility perceptions.

Table 4.14: ANOVA results for Arabic countries.

Factor	DF	*F*	P
Arabic_Countries	7	1.340	0.225
Topic	1	17.800	0.000**
Gender	1	31.770	0.000**
Image	1	43.440	0.000**
Username	1	20.590	0.000**
Location	1	1.300	0.253
Overlap	1	2.130	0.145
Countries*Gender	7	2.240	0.030*
Countries*Image	7	1.320	0.236
Countries*Username	7	0.735	0.642
Countries*Location	7	1.730	0.098
Countries*Overlap	7	1.110	0.354
Countries*Topic	7	2.110	0.040*
Gender*Topic	1	0.021	0.884
Image*Topic	1	0.110	0.741
Username*Topic	1	3.080	0.079
Topic*Location	1	9.940	0.002*
Topic*Overlap	1	59.050	0.000**
Countries*Topic*Gender	7	0.371	0.919
Countries*Topic*Image	7	1.270	0.259
Countries*Topic*Username	7	1.890	0.068
Countries*Topic*Location	7	0.838	0.556
Countries*Topic* Overlap	7	1.710	0.103

The interaction of country and gender with topic was not significant (p=0.919). The means for gender across both topics for all countries showed that all countries had the same credibility perceptions on both topics – that is, males were judged more credible than female.

4.4.3.2 Profile Image

(H2: supported) Tweets from authors with a real image were more credible than tweets from authors with an anonymous image ($\text{mean}_{\text{photo}} = 4.66$, $\text{mean}_{\text{generic}} = 4.44$, $p<0.001$).
(H2a: not supported) The interaction between country and profile image was not significant (p=0.236), nor was the effect of tweet topics on interaction between country and image (p=0.698). For all countries and both topics, a real image was perceived as more credible

than an anonymous one.

4.4.3.3 Username

(H3: supported) A topical username style was perceived more credible than an Internet username style (mean $_{\text{topical}}$ = 4.64, mean $_{\text{internet}}$ = 4.46, $p<0.001$).

(H3a: not supported) The interaction with country did not show any significant difference ($p=0.642$) and this was true for all countries. A follow-up pairwise comparison showed that some countries did not show a significant difference between the two styles (Algeria, North Africa and Palestine). Interactions between country, username and topic were not significant. For both topics, all countries were consistent with these general findings.

4.4.3.4 Location

(H4: not supported) No significant differences were found between the two location types. Arabic readers viewed tweets from large and small locations as having the same credibility level (mean $_{\text{large}}$ = 4.57, mean $_{\text{small}}$ = 4.53, $p = 0.253$).

(H4a: not supported) The interaction of country with location was not significant ($p=0.098$): topic did not affect the interaction of country and location ($p=0.556$); rather, all countries had the same credibility perception towards location styles for both topics.

(H4b: supported) Interaction between location and topic was significant ($p<0.001$). Large locations as sources were perceived as significantly more credible than small locations with regard to political tweets, while both location types prompted similar credibility ratings regarding the health topic, as shown in Table.4.15.

Table 4.15: Interaction of location and topic for Arabic countries.

	Large	Small	$P_{\text{two styles}}$
Politics	4.42	4.70	0.000**
Health	4.72	4.79	0.214
$P_{\text{two topics}}$	0.004*	0.000**	-

4.4.3.5 Network Overlap

(H5: supported) The two network overlap styles were rated at the same credibility level (mean $_{\text{overlap}}$= 4.58, mean $_{\text{no overlap}}$= 4.52, p>0.05).

(H5a: not supported) There was no effect of the difference in countries for the overlapping feature (p>0.05): even the three-way interaction of country, network overlap and topic showed no significant difference (p>0.05).

(H5b: supported) The interaction between network overlap and topic was significant (p<0.001): the tweets with network overlap were perceived to be highly credible regarding the health topic, while no-network-overlap tweets were perceived as highly credible on the topic of politics, as shown in Table.4.16.

Table 4.16: Interaction of network overlap and topic for Arabic countries.

	Overlapping	No overlapping	$P_{\text{two styles}}$
Politics	4.24	4.46	0.000**
Health	4.92	4.59	0.000**
$P_{\text{two topics}}$	0.000**	0.178	-

4.4.4 Interaction of US Regions with Author's Profile Features

Table.4.17 shows the ANOVA results for the four US regions. Next, I present the results for each of the five author profile features within the US regions based on the hypotheses presented in Section.4.2.

4.4.4.1 Gender

(H1: supported) US readers perceived tweets from males as more credible than those from females (mean $_{\text{male}}$=4.57, mean $_{\text{female}}$= 4.25, p<0.001).

(H1a: not supported) The interaction of gender with region was not significant. All regions saw male authors as significantly (p<0.001) more credible than females. Adding topic as an interaction to the gender and region interaction did not produce a significant

Table 4.17: ANOVA results for US regions.

Factor	DF	*F*	P
US_regions	3	1.540	0.202
Topic	1	26.150	0.000**
Gender	1	83.040	0.000**
Image	1	14.140	0.000**
Username	1	35.170	0.000**
Location	1	0.202	0.653
Overlap	1	37.910	0.000**
Regions*Gender	3	0.159	0.924
Regions*Image	3	1.840	0.138
Regions*Username	3	1.240	0.294
Regions*Location	3	1.580	0.192
Regions*Overlap	3	2.030	0.108
Regions*Topic	3	0.212	0.888
Gender*Topic	1	21.170	0.000**
Image*Topic	1	10.450	0.001*
Username*Topic	1	2.230	0.136
Topic*Location	1	8.640	0.003*
Topic*Overlap	1	25.920	0.000**
Regions*Topic*Gender	3	0.383	0.766
Regions*Topic*Image	3	1.360	0.253
Regions*Topic*Username	3	0.422	0.737
Regions*Topic*Location	3	4.280	0.005*
Regions*Topic* Overlap	3	4.600	0.004*

interaction. For both topics, male tweets were more credible than female tweets ($p<0.001$) in politics for all regions.

4.4.4.2 Profile Image

(H2: supported) Readers in US regions perceived tweets with a real image to be more credible than tweets with a generic image (mean $_{\text{Photo}}$ =4.47, mean $_{\text{Generic}}$= 4.36, $p<0.001$). **(H2a: not supported)** Regional differences did not affect reader credibility perception towards image type ($p>0.05$). Also, with the three-way interaction between image, region and topic, there were no differences in readers behaviours.

4.4.4.3 Username

(H3: supported) A topical username style in the US was perceived as significantly more credible than the Internet style (mean topical = 4.50, mean Internet = 4.33, $p<0.001$).
(H3a: not supported) The interaction between username and region was not significant ($p = 0.294$). This result was observed in all regions other than the West.

By adding topic to the interaction, there was no significant difference. In politics, all regions showed a significant difference except West . In health only the Northeast was significant. Thus, US credibility perceptions regard the two username styles were similar for both topics.

4.4.4.4 Location

(H4: not supported) US readers had the same credibility perceptions as Arabic readers; that is, they viewed large and small locations in the same way (mean large = 4.41, mean small = 4.42, $p>0.05$).
(H4a: not supported) The interaction between location and regions was not significant ($p=0.192$) and there were no significant differences between the two location types for any region.
(H4b: supported) There was a significant interaction between location and topic ($p<0.05$). As Table.4.18 presents, in politics small locations were perceived as more credible than were large locations, while there was no difference between the two locations for the health topic. Moreover, the three-factor interaction (region, location and topic) revealed a significant effect of location type on the credibility of tweets ($p<0.05$). Across all regions in both topics, there were no significant differences, except for politics in the Midwest, where a small location was seen as more credible than a large location ($p<0.05$).

Table 4.18: Interaction of location and topic for US regions.

	Large	Small	$P_{\text{two styles}}$
Politics	4.14	4.22	0.014*
Health	4.68	4.62	0.086
$P_{\text{two topics}}$	0.000**	0.000**	-

4.4.4.5 Network Overlap

(H5: supported) US readers perceived tweets from authors with friends overlapping as more credible than those with no overlap (mean $_{\text{overlap}}$ = 4.49, mean $_{\text{no\_overlap}}$= 4.34, p<0.001).

(H5a: sot supported) The interaction between region and network overlap showed no significant difference (p=0.108). All regions saw tweets with network overlap as more credible than tweets with no network overlap, significant results so for the Midwest and Northeast regions.

(H5b: supported) The interaction between network overlap and topic was significant (p<0.001). Also, the three-way interaction (region, network overlap and topic) was significant (p = 0.004). The post t-test showed that some regions (Midwest, Northeast, South) saw political tweets with network overlap as more credible than such tweets with no network overlap (p<0.05). Table.4.19 shows that readers in all regions perceived tweets with network overlap as more credible than tweets with no network overlap for both topics, except readers from the South region, who exhibited the opposite behaviour.

Table 4.19: US region v. overlap v. topic

	Overlap $_{\text{Politics}}$	No_overlap $_{\text{Politics}}$	P	Overlap $_{\text{Health}}$	No_overlap $_{\text{Health}}$	P
Midwest	4.29	4.00	0.000**	4.8	4.7	0.283
Northeast	4.46	4.15	0.000**	4.8	4.67	0.072
South	4.25	4.89	0.000**	4.39	4.55	0.017
West	4.25	4.13	0.086	4.66	4.61	0.399

Table 4.20: ANOVA results for Arabic regions.

Factor	DF	*F*	P
Arabic_regions	3	2.734	0.043*
Topic	1	35.755	0.000**
Gender	1	55.310	0.000**
Image	1	78.321	0.000**
Username	1	30.157	0.000**
Location	1	2.272	0.132
Overlap	1	3.701	0.055
Regions*Gender	3	1.936	0.123
Regions*Image	3	1.223	0.300
Regions*Username	3	0.447	0.720
Regions*Location	3	0.835	0.475
Regions*Overlap	3	0.116	0.950
Regions*Topic	3	2.434	0.064
Gender*Topic	1	0.634	0.426
Image*Topic	1	0.414	0.520
Username*Topic	1	4.071	0.044*
Topic*Location	1	19.866	0.000**
Topic*Overlap	1	91.623	0.000**
Regions*Topic*Gender	3	0.205	0.893
Regions*Topic*Image	3	2.013	0.111
Regions*Topic*Username	3	0.964	0.409
Regions*Topic*Location	3	1.242	0.294
Regions*Topic* Overlap	3	2.967	0.000**

4.4.5 Interaction of Arabic Regions with Author's Profile Features

After exploring the effect of the culture, country and region within a country on the credibility perception of readers, I examined this in more depth by grouping the countries into regions. **In this section I test the regions from the Arabic area; each region might include more than one country.** Table.4.20 shows the ANOVA results for the interaction between Arabic regions and profile features.

4.4.5.1 Gender

(H1: supported) Male tweets were perceived as more credible than female tweets (mean $_{\text{male}}$=4.67, mean $_{\text{female}}$= 4.39, $p<0.001$).

(H1a: not supported) The interaction of gender with region was not significant. All Arabic regions saw male authors as significantly ($p<0.001$) more credible than females, except that for Levant there was no significant difference. Adding topic as an interaction to the gender and region interaction was not significant. For both topics, male tweets were perceived more credible than female tweets, except in Levant. However, the differences between the genders for politics was high in the Gulf Cooperation Council and North African regions ($p<0.001$), and less significant in the Nile region ($p<0.05$). The results for health were the opposite, as Gulf Cooperation Council and North African readers showed less significant differences between the genders than did Egypt, ($p<0.05$ and $p<0.001$, respectively).

4.4.5.2 Profile Image

(H2: supported) Readers in Arabic regions perceived tweets with the real image as more credible than tweets with the generic image (mean $_{\text{Photo}}$ = 4.65, mean $_{\text{Generic}}$ = 4.41, $p<0.001$).

(H2a: not supported) Regional differences did not affect readers' credibility perceptions towards image type ($p>0.05$). Also, with the three-way interaction between image, region and topic, there were no differences among readers' perceptions, as shown in Table.4.20.

4.4.5.3 Username

(H3: supported) A topical username style in Arabic regions was perceived as significantly more credible than the Internet style (mean topical = 4.62, mean Internet = 4.45, $p<0.001$).

(H3a: not supported) The interaction between username and region was not significant ($p=0.720$). All readers from all regions had the same credibility perception: the topical style was more credible than the Internet style.

The three-way interaction between username, region and topic was not significant ($p=0.409$). For politics, there was no difference between the two styles in any region except Gulf Cooperation Council ($p<0.05$) . For health, all regions saw the topical username as significantly more credible than the Internet style ($p<0.05$), except Levant reported no differences between the two styles.

4.4.5.4 Location

(H4: not supported) Arabic readers at the regional level had the same credibility perceptions as Arabic readers at the country level; that is, they viewed large and small locations in the same way (mean large = 4.56, mean small = 4.51, $p=0.132$).
(H4a: not supported) The interaction between location and Arabic regions was not significant ($p=0.475$) and there were no significant differences between the two location types for any Arabic region.
(H4b: supported) There was a significant interaction between location and topic as readers' perceptions differed across topics ($p<0.001$). There were significant differences for politics, as the large locations were more credible than small locations ($p<0.001$), while there was no difference between the two locations for the health topic, as in Table.4.21. The three-factor interaction (Arabic region, location and topic) did not show any significant effect of location type on the readers credibility perceptions ($p>0.05$).

Table 4.21: Interaction of location and topic for Arabic regions.

	Large	Small	$P_{\text{two styles}}$
Politics	4.38	4.20	0.000**
Health	4.73	4.82	0.052
$P_{\text{two topics}}$	0.000**	0.000**	-

4.4.5.5 Network Overlap

(H5: not supported) Arabic readers perceived tweets from authors with overlapping friends as having the same credibility as those with no overlap (mean $_{\text{overlap}} = 4.56$, mean $_{\text{no\_overlap}} = 4.50$, $p>0.05$).

(H5a: not supported) The interaction between region and network overlap was not significant ($p=0.950$). There was no significant difference between the two overlapping styles in any region.

(H5b: supported) The interaction between network overlap and topic was significant ($p<0.001$). In politics, tweets with network overlap were perceived as less credible than tweets with no network overlap; the opposite was true for health, where tweets with network overlap were perceived as more credible than tweets with no network overlap, as in Table.4.22. Moreover, the three-way interaction (Arabic region, topic and network overlap) was significant ($p<0.001$). The follow-up comparison showed that some regions were consistent with the general finding (H5b), except that for the North Africa region there was no difference between the two styles for political tweets.

Table 4.22: Interaction between network overlap and topic for Arabic regions.

	Overlapping	No overlapping	$P_{\text{two styles}}$
Politics	4.18	4.40	0.000**
Health	4.94	4.61	0.000**
$P_{\text{two topics}}$	0.000**	0.018*	-

4.4.6 Interaction of US Divisions with Author's Profile Features

This section examines the effect of smaller geographical regions of readers on their credibility perceptions in Twitter. Table.4.23 presents the ANOVA results for the interaction between author's profile features and US divisions.

Table 4.23: ANOVA results for US divisions.

Factor	DF	*F*	P
US division	8	1.611	0.120
Topic	1	22.690	0.000**
Gender	1	42.293	0.000**
Image	1	6.421	0.012*
Username	1	34.116	0.000**
Location	1	0.715	0.398
Overlap	1	22.913	0.000**
Divisions*Gender	8	0.852	0.557
Divisions*Image	8	1.204	0.295
Divisions*Username	8	2.472	0.013*
Divisions*Location	8	2.609	0.009*
Divisions*Overlap	8	1.283	0.251
Divisions*Topic	8	0.212	0.888
Gender*Topic	1	0.737	0.659
Image*Topic	1	4.281	0.039*
Username*Topic	1	0.930	0.335
Topic*Location	1	12.222	0.001*
Topic*Overlap	1	16.110	0.000**
Divisions*Topic*Gender	8	0.535	0.830
Divisions*Topic*Image	8	0.612	0.768
Divisions*Topic*Username	8	0.659	0.728
Divisions*Topic*Location	8	3.567	0.001*
Divisions*Topic* Overlap	8	1.762	0.083

4.4.6.1 Gender

(H1: supported) US readers perceived tweets from males as more credible than those from females (mean $_{\text{male}}$ = 4.60, mean $_{\text{female}}$ = 4.30, $p<0.001$).

(H1a: not supported) The interaction of gender with US divisions was not significant. All readers from different divisions saw male authors as significantly more credible than females ($p<0.001$), with the exception of readers from the Mountain, New England and West North Central divisions who reported no differences in their credibility perception with regard to gender. Adding topic as an interaction to the gender and region interaction did not produce a significant result ($p=0.083$). For both topics, male tweets were more credible than female tweets , with p>0.001 for politics for all regions.

4.4.6.2 Profile Image

(H2: supported) Readers in US divisions perceived tweets with a real image as more credible than tweets with the generic image (mean $_{\text{Photo}}$ = 4.50, mean $_{\text{Generic}}$= 4.40, p<0.001). **(H2a: not supported)** Division differences did not affect readers' credibility perceptions towards image type (p=0.295). Moreover, in the three-way interaction between image, US division and topic, there were no differences among readers' credibility perceptions (p=0.768).

4.4.6.3 Username

(H3: supported) A topical username style was perceived by US readers as significantly more credible than the Internet style (mean topical = 4.56, mean Internet = 4.34, p<0.001). **(H3a: supported)** The interaction between username and US division was significant (p<0.05). However, only four divisions resulted in significant differences between the two username styles (East North Central, Mountain, New England and West South Central), all of which regarded the topical username as more credible than the Internet style. Also, the three-way interaction (divisions, topic and username) showed no significant difference (Table.4.24).

Table 4.24: Interaction of US divisions with username.

	Topical	Internet	$P_{\text{two styles}}$
East North Central	4.60	4.29	0.000**
East South Central	4.57	4.26	0.090
Middle Atlantic	4.58	4.49	0.183
Mountain	4.52	4.29	0.034*
New England	4.95	4.38	0.000**
Pacific	4.48	4.40	0.156
South Atlantic	4.14	4.08	0.305
West North Central	4.52	4.49	0.803
West South Central	4.68	4.37	0.001*

4.4.6.4 Location

(H4: not supported) US readers had the same credibility perceptions in that they viewed large and small locations in the same way (mean large = 4.44, mean small = 4.46, p=0.398). **(H4a: supported)** The interaction between location and US divisions was significant ($p<0.05$): divisions showed no significant differences in credibility perceptions between the two location types, with the exception of Mountain (see Table.4.25). Moreover, the interaction between location and division with topic was significant ($p<0.05$).

Table 4.25: Interaction of US divisions with location.

	Topical	Internet	$P_{\text{two styles}}$
East North Central	4.43	4.47	0.594
East South Central	4.43	4.40	0.810
Middle Atlantic	4.58	4.49	0.100
Mountain	4.25	4.56	0.000**
New England	4.68	4.65	0.804
Pacific	4.45	4.43	0.617
South Atlantic	4.09	4.14	0.322
West North Central	4.45	4.56	0.342
West South Central	4.57	4.48	0.185

(H4b: supported) The interaction between location and topic was significant ($p<0.05$): only the politics topic was influenced by location style, as shown in Table.4.26.

Table 4.26: Interaction of location and topic for US divisions.

	Large	Small	$P_{\text{two styles}}$
Politics	4.10	4.23	0.002*
Health	4.78	4.70	0.065
$P_{\text{two topics}}$	0.000**	0.000**	-

4.4.6.5 Network Overlap

(H5: supported) US readers perceived tweets from authors with friends overlapping as more credible than those with no overlap (mean $_{\text{overlap}}$ = 4.53, mean $_{\text{no\_overlap}}$= 4.37, p<0.001).

(H5a: not supported) The interaction between divisions and network overlap was not significant (p=0.251). Readers from all regions saw tweets with network overlap as more credible than tweets with no network overlap, with significant results in East North Central, Middle Atlantic, Pacific and West South Central (p<0.05).

(H5b: supported) The interaction between network overlap and topic was significant (p<0.001). Network overlap was an important feature for US readers to judge credibility of political tweets, as they saw tweets with overlapping friends as more credible then tweets with no overlap. In contrast, in health there was no effect of overlap style on readers' credibility judgments (see Table.4.27).

Table 4.27: Interaction between network overlap and topic for US divisions.

	Overlapping	No overlapping	$P_{\text{two styles}}$
Politics	4.30	4.02	0.000**
Health	4.75	4.72	0.590
$P_{\text{two topics}}$	0.000**	0.000**	-

4.4.7 Effect Size

I was interested in comparing the effects of each factor across the five ANOVA models. Effect size was independent of sample size. Thus, to understand the factors' importance on effects, I calculated the Effect Size of each factor in ANOVA models using the Pearson's correlation coefficient r [Field, 2013]:

$$r = \sqrt{\frac{F(1+df_R)}{F(1+df_R)+df_R}}$$

where $F(1 + df_R)$ is the F-ratio for each factor, and df_R is the degrees of freedom for the model. A threshold can be applied when classifying r effect size [Cohen, 1988, 1992]: $r = 0.10$ is a small effect size, $r = 0.30$ is a medium effect size and $r = 0.50$ is a large effect size. Table.4.28 shows r for each individual, two and three-way interaction for each of the five categories (cultures, Arabic countries, US regions, Arabic regions and US divisions). It is evident that all author profile features had effects between small and medium (with the exception of location in cultures and US groups). In cultural three-way interactions, all had effects (medium for overlap). Moreover, in both Arabic and English, as the classification became broad, the effect size became bigger. For example, there were four medium effects for Arabic regions compared to one for Arabic countries, and four medium effects for US regions but only two for US divisions.

Table 4.28: For effect size (Pearson r), medium effects are in dark blue, small effects are in light blue and grey is for insignificant effects. Note: 'group' in the table refers to the corresponding type (Cultures, Arabic Countries, US Regions, Arabic Regions or US Divisions).

	Cultures	**Arabic Countries**	**US Regions**	**Arabic Regions**	**US Divisions**
Group	0.0995	0.0503	0.0621	0.0713	0.0643
Topic	0.2922	0.1807	0.2485	0.2502	0.2353
Gender	0.3776	0.2384	0.4159	0.3060	0.3138
Image	0.1253	0.2759	0.1854	0.3573	0.1277
Username	0.2637	0.1939	0.2852	0.2309	0.2846
Location	0.0162	0.0496	0.0225	0.0650	0.0429
Network Overlap	0.2012	0.0634	0.2952	0.0828	0.2364
Group*Gender	0.0348	0.0650	0.0200	0.0600	0.0468
Group*Image	0.2861	0.0499	0.0679	0.0477	0.0556
Group*Username	0.0014	0.0373	0.0558	0.0288	0.0796
Group*Location	0.0283	0.0572	0.0629	0.0394	0.0818
Group*Network Overlap	0.0793	0.0458	0.0713	0.0147	0.0574
Group*Topic	0.0310	0.0631	0.0231	0.0672	0.0233
Topic*Gender	0.0943	0.0063	0.2250	0.0344	0.0435
Topic*Image	0.0306	0.0144	0.1601	0.0278	0.1045
Topic*Username	0.0306	0.0762	0.0747	0.0869	0.0489
Topic*Location	0.0920	0.1360	0.1459	0.1892	0.1749
Topic*Network Overlap	0.1029	0.3174	0.2475	0.3823	0.1999
Group*Topic*Gender	0.1330	0.0265	0.0310	0.0195	0.0371
Group*Topic*Image	0.0617	0.0490	0.0584	0.0612	0.0397
Group*Topic*Username	0.1136	0.0597	0.0325	0.0424	0.0412
Group*Topic*Location	0.1786	0.0398	0.1032	0.0481	0.0955
Group*Topic*Network Overlap	0.3310	0.0568	0.1070	0.0742	0.0673

4.4.8 General Findings

In this section, I compare the current results with those from a previous study [Yang et al., 2013], as both studies have used the same methodology to study the interaction between

Table 4.29: Comparision of the present results with [Yang et al., 2013] for the five factors.

Author's Factor	**Factor type**	**Mean $_{2013}$**	**P**	**Mean $_{2017}$**	**P**
	Male	3.87		4.63	
Gender	Female	3.75	<0.001**	4.34	<0.001**
	Photo	3.88		4.52	
Image	Generic	3.74	<0.001**	4.45	<0.001**
	Topical	3.86		4.57	
Username	Internet	3.76	<0.001**	4.40	<0.001**
	Liberal $_{2013}$/Large $_{2017}$	3.92		4.49	
Location	Conservative2013/Small $_{2017}$	3.71	<0.001**	4.48	>0.05
	Yes	3.92		4.54	
Overlap	No	3.71	<0.001**	4.43	<0.001**

culture and credibility in social media; in addition to the other factors that were included in our study include country and regions. The main points can be summarised from Table.4.29:

- All the author factors studied in both studies had a large effect on the credibility perception of readers, except that location in our results had no effect on readers' perceptions. This is likely to be because of the use of different location classification in the current study.

- Culture had a powerful effect on users' credibility perceptions in social media.

- In 2013 [Yang et al., 2013], the means ranged between 3.71 and 3.92, while in our study 2017 they were between 4.34 and 4.63. The credibility perceptions of readers towards microblogs grew dramatically in the intervening four years. Understanding the reasons for this increase in credibility towards microblogs is a suitable topic for further research.

- The overall mean view of US readers increased from 3.48 in 2013 to 4.40 in 2017. This suggests there was a significant increase in the credibility behaviour of US readers over those four years. The Twitter platform was used in both studies for US readers.

The overall view of Arabic readers was 4.57 and Chinese readers was 3.75, both groups being representative of Eastern culture. It is uncertain whether the increased microblog credibility perception was caused by the use of different platforms (Twitter v. Weibo).

- The mean difference in the overall credibility view between the US and China in [Yang et al., 2013] was 0.27, while the mean difference between Arabic and US in the current study was 0.17. This indicates that some cultures are closer in credibility perception than others, but the differences are still significant.

4.4.9 Top and Bottom Features

In Table.4.30, features are ordered from high to low by their means. The changed order between the 2013 and 2017 studies and between Arabic and US shows how credibility perceptions towards profile features were not static between readers. However, if one views the top and bottom five features in 2013 and 2017, it is clear that those factors are shared between the groups. For Arabic and English, four out of five features are shared between the top and bottom. This means that these features were highly influential in determining readers' credibility judgments.

Table 4.30: Features order based on credibility means for the current (2017) study and that of (2013); Arabic and English cultures in the current study.

2013	**2017**	**Arabic culture**	**US culture**
1 - Liberal location	1 - Male gender	1 - Male gender	1 - Male gender
1 - Overlapping	2 - Topical username	2 - Photo image	2 - Topical username
3 - Photo image	3 - Overlapped	3 - Topical username	3 - Overlapping
4 - Male gender	4 - Photo image	4 - Overlapping	4 - Photo image
5 - Topical username	5 - Large location	5 - Large location	5 - Small location
6 - Internet username	6 - Small location	6 - Small location	6 - Large location
7 - Female gender	7 - Generic image	7 - No_Overlapping	7 - Generic image
8 - Generic image	8 - No_overlaapping	8 - Internet username	8 - No_Overlapping
9 - Conservative location	9 - Internet username	9 - Generic image	9 - Internet username
9 - No_overlapping	10 - Female gender	10 - Female gender	10 - Female gender

4.5 Discussion

Research has documented the effectiveness of cultural difference in determining microblog credibility perception. Yang et al. [2013] for example, reported that participants from the US and China showed significant differences in their perceptions. However, previous studies were limited by small numbers of either countries or participants, or did not use the same microblog data between countries in their methodology. In this study, I used the same data for all participants and included a large number of participants compared to previous studies.

All author features included in this study were shown to have a distinct effect on readers' credibility judgments. The results indicate that, regardless of country, culture is the most significant effect on users' credibility assessments. The two- or three-way interactions of culture with other features were significant. This finding confirms the hypothesis that culture has a strong influence on readers' behaviour in credibility perceptions of readers of social media [Yang et al., 2013]. However, the current results indicate that 'culture' is not necessarily restricted to one country, as in previous research findings –many countries may be included as one culture.

4.5.1 Arabic Countries and Regions

The results show that readers from different countries, but belonging to the same culture, mostly share general credibility patterns, with no effect of country difference on their credibility behaviour (i.e. all eight countries had the same credibility preference). However, the results for the Arabic regions showed a significant interaction among these regions indicating that 'local culture' has more effect on readers' credibility judgments than country; albeit a much smaller effect than between totally different cultures. A retrieval system used in social media can lead to optimised search results by focusing more on cultural preferences.

4.5.2 US Regions and Divisions

Readers from different regions in the US have the same credibility perceptions, and regional differences do not affect that behaviour. There was no significant interaction between region and any of the five features in this study, as shown in Section.4.4.4. However, regional differences affected two features: location and network overlap when interacting with a tweet topic.

The difference between the nine US divisions had no significant result as shown in Section.4.4.6. However, username and location had a significant interaction with US divisions ($p<0.05$). Moreover, the three-way interaction of division, location and topic was significant as well. It was evident that location is a two-level division and regions influenced readers' perceptions significantly. In the US, large and small location categories strongly influenced the credibility perception of readers.

From the results presented above in Sections 4.5, 4.5.1 and 4.5.2, the following points can be summarised:

- Cultural differences influence the credibility judgments of readers more than any other location classifications.
- In countries with similar identifiable culture, the effect of the country level on readers' perceptions is miniscule. However, when the classification becomes broader, such as the regional differences affecting local culture, the regional cultural influence is greater than that of the country on readers' credibility perceptions.
- In large countries like the US, the smaller location representation of readers (division) has a stronger effect than the large one (region) on readers' credibility perceptions.
- The second and third points confirm the first, as culture is more important than actual location when it comes to understanding social media credibility perceptions.

4.5.3 Topical Context

The credibility effect of the features included in this study are context dependent rather than caused by a single effect. For example, location and network overlap were found to be the most important topical context-dependent features. At all five levels (cultures, Arabic countries, US regions, Arabic regions and US divisions), location and network overlap had a significant interaction with topic. Large locations were important for political tweets and small locations important for health tweets. The type of tweet location presented to readers depended on the tweet topic. This finding reflects that of [Aladhadh et al., 2014] who found a tweet's location increases its credibility for political tweets. However, as only a few authors include their location on their posts ($\sim$ 1% of tweets in Twitter are geo-tagged [Morstatter et al., 2013]), predicting tweet location is a current area of research and is a predictable feature [Han et al., 2014].

Network overlap is an important feature influencing credibility of health tweets: people see authors followed by their friends in health tweets as a credible source. However, this is not the case for political tweets. Therefore, for health tweets it is important to focus on social overlapping to retrieve results for readers to support them in making credibility judgments.

Only two out of five features were found to have a significant interaction with culture. However, four features had a significant interaction with culture when topical context was added in a three-way interaction. This indicates that cultural differences are hidden between topical contexts. For some factors (such as username), there was no significant interaction with culture or with topic but the three-way interaction username, topic and culture was significant. This indicates that for some features, the way they affect readers' credibility judgments are complex. There is a need to optimise current retrieval algorithms to make results more topic based for different culture 'localisation'.

The effect of topical context is powerful, even on readers from the same country. Regional differences among US readers did not make any significant difference to readers' judgment for all features. However, credibility perceptions of US readers from different regions are significantly different for location and network overlap based on the topic of a tweet. This means that credibility perception varies based on topic. In other words, topic determines the level of importance of a feature as a predictor of credibility perception. Topic might be a more influential feature on readers' credibility judgment than other features. It is necessary to build categories for the proportion of features in different microblog topics and use that to enhance credibility of retrieved results in social media. This is similar to the categories of information distribution in different crisis types [Imran and Castillo, 2015]

Employing readers' credibility perceptions to design the user experience in microblogs in regard to retrieving credible information for final users has been found to be effective [Yang et al., 2013]. In this study, I validated that culture is an essential factor in studying credibility in social media, and I differentiate the effects of five levels (cultures, Arabic countries, US regions, Arabic regions and US divisions).

4.6 Summary

In this chapter, I examined the extent to which differences between cultures, Arabic countries, US regions, Arabic regions and US divisions interact with Twitter profile features (author gender, profile image, username, location and network overlap) influence readers' credibility judgments. I examined the following research questions:

1. What is the effect of a reader's location on their perceptions about the credibility of social media information? :

- What is the effect of a reader's culture on their credibility perception in social media?

  Culture was the most prominent of the five categories.

- Will a reader's nationality influence their credibility perception of social media among other members of the same culture?

The country category had the smallest effect on reader credibility judgments.

- What is the effect of the geographical region of countries with the same culture on readers' credibility perception of social media?

  The region category had a stronger influence on reader credibility perceptions than country alone, but a much weaker effect than culture.

- How do different regions and divisions in large countries influence a reader's credibility perception of social media?

  The division category had a stronger effect than regions on reader judgment, but culture of a whole country like the US had an even stronger effect.

I sought to examine how readers' culture, country, region and division influences their credibility perceptions. I found that culture can be used to customise social search engines to help assess content credibility, but including a country's attributes (regions and division) along with its culture can be even more effective. Moreover, profile features have a significant effect on credibility judgments by readers. These findings will inform designers of interfaces and algorithms about readers' needs. Despite the large effect of culture on readers' credibility perceptions as found in this study, a number of categories – such as corporate accounts – were not included in this study.

Chapter 5

Location Impact on Source and Linguistic Features for Information Credibility of Social Media

5.1 Introduction

Social media content is 'noisy'-mixed with high and low-credible information. Assessing social media information credibility is challenging, because content is influenced by many factors including source type and author location. Social media content with the same credibility status might not share characteristics such as linguistic features.

Credibility can be measured through knowledge about an information source. For example, identifiable news sources are likely to be more credible than anonymous sources. However, sources with the same credibility level can also be influenced by their location – for example, near to or far from the event location – which can influence their generated content.

Next, I introduce the thee aspects related to this chapter: information source; location; and credibility with linguistic features.

5.1.1 Information Source

Sources in social media research are generally classified based on their location as remote or local. Remote sources share information about the event from a distance. Local sources are able to obtain first-hand information from an event site and are called eyewitnesses; however, many limitations exist for such sources. Increasing the number of credible sources that can be used to find credible information is essential for assessing social media information credibility.

Considering the source of information is a critical part of assessing its credibility. In social media, many researchers have attempted to categorise sources into high and low-credibility sources to reach credible information. Tweets are posted by sources varying from globally well-recognised organisations to locally popular community organisers [De Choudhury et al., 2012], and from specialists in such a domain to fake accounts that steal the identity of other people. Consequently, the quality of information is hugely diverse; finding sources who have highly relevant and credible information about such an event as the source of information is challenging.

Methodologies to find authentic sources in social media vary; for example, topical content and network structure can be used to rank sources based on credibility for a given topic [Canini et al., 2011]. Other research has included author-related data such as tweet content and author profile to build their models, in addition to using social media expert groups. These groups are in the form of topical expertise directories, such as lists and skills membership on Twitter and LinkedIn, respectively [Wagner et al., 2012, Bhattacharya et al., 2014, Ghosh et al., 2012, Bastian et al., 2014]. For example, lists in Twitter are an organisational feature created by authors to group experts on such topics. Previous research has classified sources as high and low based on their topical expertise, local authority and expertise [Cheng et al., 2014a]. However, previous research has focused only on the information source regarding a general topic, rather than a particular event that may be limited by time and location, such as a crisis event.

5.1.2 Author and Event Location

Knowledge about whether a source is local or remote can help enhance credibility assessment. Currently, it is common practice for traditional stakeholders (e.g. national press) to contact social media sources close to the place of an event to receive updates [Dailey and Starbird, 2014]. Thus, information coming from the same region as an event is likely to be richer than remote content.

Using the location of microblog post to predict content credibility is important. During events, social media platforms often provide the first alarm: people begin sharing information about what is happening. Sources from the same event location share specific and accurate information whereas those further from the event location share general content [Kumar et al., 2013].

Sources posting information about a particular event from the same or nearby area are known as 'eyewitnesses' and are presumed to have accurate information. Much research has attempted to reach eyewitness authors (e.g. [Diakopoulos et al., 2012, Olteanu et al., 2015]) but there are many limitations to those studies:

- Very few witnesses to an event share information on social media [Truelove et al., 2015].
- Witness authors are identified via GPS coordinates attached to their tweets; however, $\sim 1\%$ of tweets in Twitter are geo-tagged [Morstatter et al., 2013].

To date, no research has examined the influence of the distance between source and event locations on the credibility of information.

5.1.3 Credibility and Linguistic Features

The language of tweets generated from the same event's region differs from the language of tweets from outside that region [Morstatter et al., 2014, Kumar et al., 2014, Cheng et al., 2010]. Research on social media has shown that content of the same credibility level (whether credible or not) shares common features [Castillo et al., 2013]. Research has also revealed an overestimation of prediction in current credibility models when applied in different domains [Boididou et al., 2014, Aker et al., 2017]. However, no study has investigated the effect of location on the behaviour of information sources, and the influence of the author location on assessing content credibility .

The linguistic features related to different credibility levels in web-based articles have been studied: Popat et al. [2016, 2017] presented an approach to identify true and false textual claims. They studied linguistic styles by using a set of lexicons. They assumed the language of a high-credibility article is unbiased and objective, while subjective language relates to low-credibility articles. Also, they included the reliability of the source of the article and examined the effectiveness of using language features with other factors to identify the credibility of a given claim. Horne and Adali [2017] compared real and fake news, finding a significant difference between them in structure of the title and other language features. A large-scale study of true and false news distributed between 2006 and 2017 on Twitter revealed significant differences in language features of author responses to false and true news [Vosoughi et al., 2018]. Moreover, linguistic features were found to be an important factor in identifying experts in Twitter [Horne et al., 2016]. The language of a tweet is influenced by the source's location [Cheng et al., 2010, Han et al., 2014]; and including linguistic features with other factors such as location in social media credibility models can enhance limitations in existing models [Boididou et al., 2014].

Social media content is influenced by the location of the source, which affects the number of contributed sources and the writing style. No research has investigated the distribution of sources of such an event like crisis or political events locally and remotely, and the effect of the source's location on linguistic features for content of different credibility, as these are readily available features that can be used for determining credibility in social media.
The research questions addressed in this chapter are:

1. What is the effect of location on sources distribution and linguistic features for infor-

mation credibility of social media?

- What types of information sources are associated with which events, both inside and outside the country in which the event is taking place?
- How do linguistic features differ among sources of different type, credibility level, topic and location?
- How effective is it to only use the information source to assess credibility?

5.2 Methodology

I followed the following steps to complete this study:

1. I defined a number of dimensions to classify social media messages at the time of events: location, informativeness, source and credibility.
2. I collected Twitter data for sets of events occurring between 2016 and 2017 using Twitter API. These events were taken from three topics: entertainment, politics and crisis. Each event had 1,200 tweets in the dataset.
3. I ran a set of crowdsourcing tasks to evaluate and characterise these messages.
4. I analysed the interaction between the type of an source's location and their message characteristics based on a message's informativeness, the used source, credibility and semantics.

5.2.1 Location Dimension

I build on the hypothesis that being a local source might increase the credibility of information. My definition of the source's proximity to the event location is that it was within the same country in which the event occurred, similar to [Kumar et al., 2013], while remote sources were those outside that country. Then, to evaluate the effect of distance from the event, I compared the two categories (local and remote) according to different content types.

There are two ways to determine sources' location in Twitter. The first is to use GPS coordinates associated with tweets at the time of post. However, <%1 of tweets are geotagged [Morstatter et al., 2014] and in any case, location changes as sources move. The other way to determine tweet location is via the profile location entered by sources. In this research I employed the second option, as have many other researchers [Sakaki et al., 2010, Mislove et al., 2011, Thomson et al., 2012, Poblete et al., 2011, Kumar et al., 2013]. In this work I am interested in the home country of the source, not their current location. Thus, I chose to use the author profile location as it more accurately reflects the home country of the source.

Hecht et al. [2011] found 66% of Twitter authors had a valid geographical location, 16% had empty locations and only 18% had invalid locations. Thus, the profile location field is free text that can include not the author geographic location entered by authors. To determine a author's home country and then classify them to one of the two location classes (local and remote), I took the following steps:

Local: To define a author's location inside the country of an event, I used country names in multiple formats. For example, for an event in the UK I used all formats used in Geonames.[1]. Also, I used the large cities of each country of an event, using the relevant Wikipedia entry to define the large cities of each country.

Remote: For authors outside the country of an event, I used the list of all countries' names in different formats (without the names of the country of the event). Also, I used the list of the top 100 largest cities (excluding cities in the country of the event): the author had to be located in one of them to be classified as a remote author.

[1]http://www.geonames.org/

5.2.2 Content Dimension

To assess the content broadcast during different events I reviewed previous research that analysed social media content and credibility and created a set of categories that had been included in other research [Olteanu et al., 2015, Starbird et al., 2010, Vieweg et al., 2010, Gupta and Kumaraguru, 2012, Imran et al., 2013]. These categories are broad to fit with different events, the large number of messages in this kind of study and the limitation of using crowdsource workers to complete annotations.

5.2.2.1 Informativeness

In this step I assessed the status of tweets' informativeness regarding an event. This process is a subjective task as it depends on the person seeking the information, and the context of the event to understand the implications. For this dimension I followed the precedent set by [Vieweg et al., 2010, Gupta et al., 2014], to measure how a tweet provides understanding about a situation. The following criteria were used.

- Related to the event and informative – when a tweet includes information about the situation and helps a reader to understand what is happening.
- Related to the event, but not informative – when a tweet includes information about the situation but does not help one to understand what is happening.
- Not related to the event.
- Not applicable.

5.2.2.2 Source

When people receive an update about an event in social media they look for the source of information, which is the main concern regarding the quality of shared information.

The authors contributing on social media are diverse and distributed as individuals or organisations, and each category has subcategories. Thus, I selected sources that are always included in different events; these sources were also included in previous research [Olteanu et al., 2015, Thomson et al., 2012, Wu et al., 2011].

For this dimension, based on previous research I developed a categorisation schema employing the following types of sources:

- government: information published by officials.
- businesses: information published by profit-making businesses.
- Non-profit organisation organisations: information published by non-profit organisations.
- traditional media: information published by news agencies, TV and radio.
- journalist: information published by journalist associated with an organisation or freelance.
- eyewitness: information reported by an eyewitness to the event, or their family, friends.
- politician: an individual politician working in government.
- academic, specialist, researcher an individual working in a university or think-tank.
- digerati: a word derived from mixing 'digital' and 'literati', which refers to an individual who is famous in social media and has a strong influence.
- celebrities: individual who is famous for any reason, for example, a singer, actor or media presenter (not in technology).
- outsiders: ordinary or non-identified sources.

5.2.2.3 Credibility Assessment One (Main Method)

Assessing the credibility of information is a subjective process; in this research I used methodology commonly used in social media credibility research [Castillo et al., 2011, 2013, Gupta et al., 2014]. I asked annotators to label tweets based on their credibility, according to one of the following categories:

- definitely credible
- seems incredible
- definitely incredible
- I can't decide.

The credibility definition used in this research and provided to the annotators was 'offering reasonable grounds for being believed',[2] accompanied by an explanation for each category. This definition was used in previous research [Castillo et al., 2011]. Applying this definition to this research, a given tweet was said to include credible information when the annotator believed the truthfulness of the tweet's information. The criterion for a tweet to be a credible was that is presented a fact, was informative and was not a personal point of view [Castillo et al., 2011, Shariff et al., 2014]. I noted that 'seems incredible' somewhat replicated 'seems credible' as both indicate whether the tweet is credible or not. Thus, I only retained one of these to force evaluators to select from the other options as in [Castillo et al., 2011, Gupta and Kumaraguru, 2012].

5.2.2.4 Credibility Assessment Two (Source-based Credibility)

Another way of assessing the credibility of information (not commonly used for assessing social media credibility) is mainly based on the source of information. This can be an authentic source with high credibility or a less credible source; this can influence the quality of posted information. Thus, source credibility levels are distributed between high and low based on their authenticity status. To validate this methodology and compare it with the first methodology (the main approach), I followed the categories established in previous work

[2]http://www.merriam-webster.com/dictionary/credible

for source credibility assessment [Thomson et al., 2012] where sources were classified as high or low credibility as follows:

- governments (high)
- businesses (high)
- non-profit (high)
- traditional media (high)
- journalists (high)
- eyewitnesses (high)
- politicians (high)
- academics, specialists, researchers: (high)
- digerati (high)
- celebrities (high)
- outsiders (low)

This method is not very accurate as it blindly classifies source credibility (e.g. some types of sources are not always credible), but the main concept here is that I compare between authentic (or known) sources and unknown sources. Also, these sources are small size in the used dataset which have no impact on the results. Please note that I used this methodology only for the purpose of comparison with the main method of assigning credibility; therefore, the discussion in this chapter is based on the main method results, although I discuss the effect of source-based credibility in Section 5.5.3.

Table 5.1: The LIWC categories used to analyse tweet content.

Feature	**Example**
Function words, includes sub-categories such as personal and impersonal pronouns, and auxiliary verbs.	it, to, no, very
Affect words, includes sub-categories such as positive and negative emotion.	happy, cried
Social processes, includes sub-categories such as family, friend, and male&female references.	mate, talk, they
Cognitive processes (cogproc), includes sub-categories such as insight, causation and certainty.	cause, know, ought
Perceptual processes (percept), includes sub-categories such as see, hear, and feel.	look, heard, feeling
Biological processes (bio), includes sub-categories such as body, health, and ingestion.	eat, blood, pain
Core drives and needs (drives), includes sub-categories such as affiliation, achievement and power.	ally, win, superior
Relativity (relativ), includes sub-categories such as motion, space, and time.	area, bend, exit
Informal language (informal), includes sub-categories such as swear words, nonfluencies, and assent.	damn, btw, umm
Authentic, pronoun, word count (WC), Qmark, exclam, hashtags and URL.	count each category in a tweet

5.2.2.5 Linguistic Feature Analysis

There are many types of features in Twitter, which are generally classified as content-based, social-based or network-based features [Castillo et al., 2013, O'Donovan et al., 2012, Kang et al., 2012]. Although credibility research in social media has grown rapidly, there have been few investigations into linguistic features associated with credibility.

I used a tweet's language as an indicator of credibility for many reasons.

- The linguistic features of tweets have been found to be important predictors of credibility in social media events [Mitra et al., 2017b, Kwon et al., 2013].
- The language of tweets is influenced by the source's location [Cheng et al., 2010, Han et al., 2014] and this allows us to study the effect of language and location on credibility.
- Text is available in all tweets, whereas other features might be absent.

I used the sentiment analysis tool LIWC, which analyses text by counting words in different psychological categories [Tausczik and Pennebaker, 2010]. Its dictionary of categories includes almost 6,400 words and word stems. Use of LIWC is common in social media data analyses [Nguyen et al., 2013, Golbeck, 2016] and credibility research in specific areas such as [Gupta and Kumaraguru, 2012, Kwon et al., 2013, Zeng et al., 2016, Rosso and Cagnina, 2017, Mitra et al., 2017b]. The categories used here are listed in Table.5.1 and are those that have been used in previous research. All but the last of these categories have subcategories.

For example, the affect feature is the main category, and includes two subcategories: positive and negative emotions. The negative emotion includes three subsub categories anxiety, anger and sadness. Thus, the current analyses includes general categories. LIWC does not provide the definitions of the categories; however, the full list of both categories, and subcategories with example are available at.[3] The results were normalised [0-1]. I investigated whether any of the categories dominantly appeared in tweets of each source, credibility level, location and topic.

5.3 Data Collection

I collected data from different events on different topics. For each event I identified the most used keywords and hashtags employed during the event window. I then submitted these to the Twitter streaming API,[4] to crawl event-related tweets. These tweets were most likely to be representative of the discussion of the event in Twitter. Table.5.2 provides information about the numbers of authors and tweets for each event. I collected events in the same way as previous research [Vieweg et al., 2010, De Choudhury et al., 2012, Thomson et al., 2012]. In this study I excluded retweeted messages, and only included active authors who had three or more tweets about the event (most authors only had one tweet), similar to [Vieweg et al., 2010, Starbird and Palen, 2010, Thomson et al., 2012]. This threshold for sampling was applied to reduce noise by capturing active authors.

I used crowdsourcing to complete the annotation process for informativeness and source. The crowdsource workers were given instructions for how to complete each task, including the event name with a short description and the link to an outside source to read about the event in more detail. This is further detailed in the following subsections. Also, I provided examples about each event to help them understand the task. I used CrowdFlower to complete the tasks (recently they change the name to 'Figure eight').[5]

[3]http://liwc.wpengine.com
[4]https://dev.twitter.com/streaming/overview
[5]https://www.figure-eight.com/

Table 5.2: A description of the events used in this study.

EVENT NAME	COUNTRY	YEAR	POSTS	USERS	ACTIVE AUTHORS
Apple Event	US	2016	1 407 577	771 081	106 067
Search keywords:#AppleEvent					
Summer Olympics	Brazil	2016	3 094 539	1 404 981	236 962
Search keywords: #Rio2016					
Oscar Academy Award	US	2017	3 133 296	1 463 300	274 144
Search keywords: #Oscar #Oscar2017					
Italy Earthquake	Italy	2016	512 798	320 306	30 177
Search keywords: #ItalyEarthquake #PrayForItaly					
London Attacks	UK	2017	303 884	212 765	16 564
Search keywords:#LondonAttacks #Prayforlondon #Londonstrong #Westminster					
Cyclone Debbie	Australia	2017	89 954	33 129	6292
Search keywords:#debbie #cyclone #CycloneDebbie #tcdebbie Qld Queensland cyclone #BigWet					
Presidential Debate	US	2016	5 443 507	1 819 068	355 895
Search keywords: #debatenight					
Presidential Election	US	2016	1 295 766	892 094	77 903
Search keywords: #ElectionDay #ElectionNight #USElection2016					

5.3.1 Description of the Task

Here I outline the task descriptions as provided to the workers who were asked to categorise tweets for informativeness, source and credibility.[6]

5.3.1.1 Informativeness

Please read the following tweet, check the link inside if needed and select the most appreciate category:

- The tweet is related to the event and is informative if it includes information about the event that is useful and helps you understand what happened.
  - @CNN: The Oscars weren't afraid to get political https://t.co/pE5zvwe6me https://t.co/XK2FdGvTF8

[6]The ethics number for this study is ASEHAPP 35-15.

- The tweet is related to the event but is not informative if it mentions the event but doesn't help you understand what happened.
  - Did Antoine Fuqua just direct a Walmart short film or am I crazy? Oscars2017
- The tweet is not related to the event.
  - Thanks so much I love you Oscars2017?
- The tweet is not applicable or has some problems, such as not being readable, or other issues.

5.3.1.2 Source

Please read the tweet posted at the time of Cyclone Debbie 2017 in Australia, visit the tweet in Twitter, check the link inside the tweet if needed and select the most appreciate source of information as:

- government: information published by an official such as police or hospital
  - @BOM_Qld: Radar loop from the #Mackay radar shows the eyewall and eye of #CycloneDebbie as it tracks towards the coast.
- non-profit organisation: information published by administration of NGOs and not-for-profit organisations such as Red Cross and UNICEF
  - @RACQOfficial: Don't risk your safety, stay off the roads. #FloodedForgetIt #BigWet #bnetraffic
    https://t.co/jccsBrBrpb
- business: information published by a profit-related business or enterprise
  - @AEMO_Media: We are working with @PowerlinkQLD to prepare for TC Debbie and keeping a close eye on the situation. #CycloneDebbie

- traditional and/or Internet media: information published by source news organisations or web blogs, such as TV and radio
  - @ABCemergency: All #Brisbane schools to close today as former #CycloneDebbie heads south #bigwet
- eyewitness: information reported by an eyewitness to the event, or their family, friends and so on
  - So our fence has come down. Tried to save it but it's too windy too strong. https://t.co/fp69EZB2QK #Mackay #bowen #CycloneDebbie
- journalist: associated with an organisation or freelance
  - #CycloneDebbie blew the feathers off a cockatoo. https://t.co/hmELUDw65S
- academic, researcher or specialist: an individual working in a university or think-tank
  - @climatrisk: Stay safe FNQ a real frightnener. Hope everything is battened down #CycloneDebbie
    https://t.co/Ea9z5ASZMl
- politician: an individual politician working in government
  - @AnnastaciaMP: #CycloneDebbie is now crossing the coast between #Bowen and #AirlieBeach. Stay safe everyone. https://t.co/9u2mY2zguY
- digerati: an individual who is popular in the area of social media and technology
  - Absolutely APPALLED at @Avis car hire charging us $158 to extend our car rental as #CycloneDebbie is stopping us from reach
- celebrity: individual who is famous for any reason, for example, a singer, actor or media presenter (not in technology)

  - The team at SDBHQ, and the entire #BlakeArmy is thinking of everyone in North Qld. Stay safe Queenslanders, SDBHQ xo

- ordinary individual: authors on Twitter posting updates on their daily lives, or non-identified sources

  - Good luck North Queensland. Batten down the hatches. Don't drive in flood waters. Look after each other! STAY SAFE! #cyclonedebbie

All the sources used in this studied included their profile locations. The same source type can be local or remote at the same time with regard to a particular event. For example, 'the Red Cross' was engaged from two locations in 'Cyclone Debbie Australia':

- Australian Red Cross (local) (@RedCrossAU: Affected by #CycloneDebbie? Let your family know you are ok. Register at https://t.co/NruW5WjXtO https://t.co/EGChoXeh9X).

- Papua New Guinea Red Cross (remote) (@PNGRedCross: High tide in 5 hrs. #RedCross running evac centres. The latest from #TVNZ KimberleeDowns #CycloneDebbie #TCDebbie https://t.co/BvLXnOL7kN).

The same is true for other organisations including government: a government account can sometimes engage in an international event, for example:

- This tweet about 'Rio2016' from 'Road authority' of Uganda government located at Kampala Uganda: (@UNRA_UG: Wishing #TeamUganda at the #OlympicGames all the entire best! We're your fans and you have our support! Cheers #UGA #RIO2016).

- This tweet about 'Cyclone Debbie Australia' from 'Met Office Storms' of UK government located at Exeter, UK: @metofficestorms: Rainfall radar image of #CycloneDebbie which is slow moving off the coast of #Queensland. Peak wind gust 117 mph at Hamilton Island.

- This tweet about ‘Italy earthquacke’ from ‘Italy UN’ of the Italian government located in New York: @ItalyUN_NY: At least 241 dead & 2.5k displaced so far after #ItalyEarthquake. We are very grateful for the solidarity in #NYC
- This tweet about ‘Rio2016’ from ‘Dept of Sport’ of the Indian Ministry of Youth located in India:@IndiaSports: Indian players / teams event schedule, fixtures for #RioOlympics on Day 4. #Rio2016 #Olympics https://t.co/AvwhX9haVe

The same applies to traditional media sources, all of which include their profile location, for example:

- This tweet from ‘BBC’ at ‘Rio2016’ had the profile location London, UK: @BBCWorld: If Michael Phelps was a country https://t.co/wTxvBP8CL5 #Rio2016.
- This tweet authored by ‘The New York Times’ about ‘oscar2017’ had the profile location New York City: @nytimes: The #Oscars audience wonders: Who is Gary from Chicago? https://t.co/HDbLOGzn6z.

5.3.1.3 Credibility (Main Method)

Please read the tweet posted at the time of ‘the event name’, check the link inside the tweet if needed and determine the credibility level of the tweet as:

- definitely credible
- seems incredible
- definitely incredible
- I can’t decide.

5.3.2 Characteristics of the Tasks

All tasks in this study involved tweets written in English. The crowdsource workers were from the same country as the relevant event unless they were too few, in which case workers from neighbouring countries were included (this happened rarely). This procedure was followed because local workers are more likely to have situational awareness of the event, and to understand the dialects, locations, entities and culture of the overall place of the event. Moreover, as per CrowdFlower platform guidelines, 40âĂŞ50 tweets for each event and task were annotated by the research team. Any worker with less than 80% agreement with the annotated tweets was classified as untrusted. A trusted judgment by a trusted worker took a mean of 8.0 seconds to be made for the informativeness task; for the source task mean completion time was 22.8 seconds; and for the credibility task, 18.0 seconds. The overall agreement between workers' judgments for 100 randomly selected tweets by the platform was 72.5% for the informativeness task, 81.2% for the source task and 81.2% for the credibility assessment. For each tweet in all tasks at least three judgments were collected, and the final label was calculated by the majority.

Each step of the annotation, which included 1,200 tweets for a single event in each task (informativeness, source or credibility), was completed by around 15âĂŞ25 workers. Each worker was limited to 300 judgments for each task and could not exceed this limit, as recommended by the crowdsourcing platform.

The first classification task was to define event-related tweets. Some tweets may contain event keywords but be unrelated to the event. Thus, for each event 600 related tweets were annotated from each location type selected randomly. For all events, tweets were labelled until the limit was passed; only related tweets (related & informative and related but not informative) were retained, and then classified based on their source and credibility.

5.3.3 Evaluation of the Tasks

Subjectivity is involved in the classification process; for example, tweet content might affect results, especially in large-scale studies such as the present one. To evaluate the effect of this subjectivity in the results, I followed the methodology used by [Olteanu et al., 2015]. Independently, two coders from the research team labelled 100 tweets selected randomly from all events. They classified the tweets based on informativeness and source type. The coders had full background information about all events, read the tweets as displayed on Twitter and visited any links within a tweet and the author profile of the tweet.

I applied Cohens' Kappa (k) to measure the inter-coder agreement, with k formulated as:

$$k = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

where Pr(a) is the number of times that the coders agree and Pr(e) is the probability that the coders agree by chance [Carletta, 1996]. The results were ($k = 0.80$) for the informativeness task and ($k = 0.89$) for the source task. Both values indicate substantial and excellent agreement between coders' labels.

This was followed by a comparison of labels for the tweets on which both coders agree with labels provided by crowdsource workers. The result for informativeness was $k = 0.77$; for source, $k = 0.79$; and for credibility, $k = 0.81$. The results also showed substantial agreement for all tasks. Next, I checked the agreement for each author individually with workers. This included the labels with no agreement between the two authors. The results also indicated high agreement: $k = 0.78$ and 0.64 for informativeness; $k = 0.79$ and 0.72 for source; and $k = 0.80$ and 0.74 for the credibility task.

From the previous experiment it was noted that the crowdsource workers provided a reliable set of collective labels in social media labelling tasks. This conclusion is similar to those of previous studies that used crowdsourcing for labelling [De Choudhury et al., 2012, Diakopoulos et al., 2012]. This study received 28,800 labels (8 ÃŮ 1,200 ÃŮ 3) for each task-informativeness, source and credibility.

5.4 Results

In this section I present the analysis performed on the data received from the crowdsource workers. I first present the distribution of the content across locations. Then I examine the proportion of the linguistic features among different sources and credibility levels, and the influence of other factors (location and topic) on the tweets' linguistic distribution. Please note that for all results in the tables * $p<0.05$ and ** $p<0.001$.

5.4.1 Content *v.* Location Dimensions

I begin by presenting the content distribution for each content type and then the content categories across locations.

5.4.1.1 Informativeness

The distribution of messages that were found to be related to the event (including the first two categories 'related and informative' and 'related but not informative') was on average 91%. The proportion of related messages based on location was similar: the average for local related messages was 88%, while the average for remote messages was 91%. The effect of distance on the related messages was therefore weak.

The informative messages (only the first category of the informativeness task was considered 'related and informative') gave an average 46%, similar to [Gupta et al., 2014]. The effect of distance to event location on informativeness of tweets was strong. The percentage of informative remote tweets was higher than that of local tweets: the average was 43% for the local location, but 49% for remote locations. This indicates that tweets from outside the country of an event relay more informative information than those within that country, ($p < 0.001$).

5.4.1.2 Sources

Table.5.3 shows the numbers of sources and their distribution in local and remote locations from all events. For each source I present the proportion in both location categories. I applied a sign test [Gibbons and Chakraborti, 2011] over sources and location to determine whether the distribution of sources is different between the two locations. Also, I classified these events into three topics: entertainment (Apple Event, Summer Olympics and Oscars), politics (Debate Night and US Election) and crisis (Italy Earthquake, London Attacks and Cyclone Debbie). I show the proportion of each source category across topics. Table.5.4 presents the sources for each event in different locations.

Table 5.3: The overall sources, and their distribution across locations.

	All	SD all	Local	SD local	Remote	SD remote	P two locations
Government	330.000	57.078	201.000	34.668	129.000	22.931	*
Non_Profit	258.000	36.394	158.000	22.050	100.000	15.005	
Business	141.000	13.917	62.000	5.946	79.000	8.543	**
Traditional_Media	2090.000	130.001	777.000	60.523	1313.000	78.991	**
Eyewitness	8.000	1.309	4.000	1.069	4.000	1.069	
Journalist	625.000	69.221	374.000	46.775	251.000	22.557	**
Academic	78.000	6.319	38.000	4.496	40.000	3.586	
Politician	256.000	49.702	154.000	24.33	102.000	26.612	*
Digerati	259.000	26.104	143.000	14.446	116.000	12.259	
Celebrity	328.000	35.21	160.000	18.189	168.000	17.542	
Ordinary	5227.000	222.014	2729.000	122.092	2498.000	108.390	*

Next, I present the distribution of each source's type in general, locally, remotely, and across topics:

- *Government*: 3.4% of sources in all events were government, 4.2% local and 2.7% from remote sources (at crisis events to support a foreign country or in 'Rio2016 Olympic' supporting their national team) ($p < 0.05$). As expected, the government sources in the country of the event were higher than those outside. Government accounts in entertainment, politics and crisis events were 3.3%, 0.3% and 5.7%, respectively.

- *Non_profit organization*: 2.7% of all sources in this study were NGOs, 3.3% of which were local and 2.1%, remote ($p > 0.05$). The distribution of these sources differed between topics: 4.1% in entertainment; 1.2% in politics; and 2.3% in crises.

- *Business*: 1.5% were business sources. Local business sources were 1.3%, while remote

Table 5.4: Source distribution of tweets for each event locally and remotely.

	Government	**Non_Government**	**Business**	**Traditional_Media**	**Eyewitness**	**Journalist**	**Academic**	**Politician**	**Digerati**	**Celebrity**	**Ordinary**
Apple_Local	0	2	13	71	0	20	2	2	43	14	433
Apple_Remote	0	1	25	128	0	14	0	0	41	9	382
Rio_Local	77	72	18	104	3	80	2	1	11	14	218
Rio_Remote	39	47	20	195	0	47	4	2	12	15	219
Oscar_Local	1	16	12	114	0	39	2	13	20	59	324
Oscar_Remote	3	10	12	174	0	31	1	0	14	56	299
Debate_Local	2	11	5	57	0	25	3	72	22	17	386
Debate_Remote	1	1	1	78	0	25	6	78	20	32	358
Election_Local	2	13	6	33	1	31	15	38	33	30	398
Election_Remote	1	3	4	85	0	22	4	8	16	28	429
Italy_Local	8	6	2	34	0	4	6	4	4	21	511
Italy_Remote	8	13	4	177	1	11	11	5	4	20	346
London_Local	30	15	1	166	0	26	6	13	7	3	333
London_Remote	14	16	5	147	0	21	8	9	4	4	372
Cyclone_Local	81	23	5	198	0	149	2	11	3	2	126
Cyclone_Remote	63	9	8	329	3	80	6	0	5	4	93

business sources were 1.6% ($p < 0.001$). Entertainment events were included in the highest number of business sources (3.0%) compared with politics (0.7%) and crisis (1.0%) events.

- *Traditional and Internet media*: these formed the second largest source, at 22.0%. The traditional sources used locally were 16.0% compared with 27.0% remotely ($p < 0.001$). The traditional sources in crisis events were 29.0%, entertainment events 22.0% and political events 11.0%. The number of traditional sources used by remote authors was higher than that used by local people in almost all events: 'Cyclone Debbie' in Australia 2017 was the event with the greatest proportion of traditional media sources (44.0%).

- *Journalist*: 6.5% of sources in all events were journalists, with 8.0% local and 5.0% remote ($p < 0.001$). The journalists in crisis events were the largest group (8.0%), followed by entertainment (6.0%) and then politics (4.0%). The proportion of journalist sources was the highest in two events: the Summer Olympics 2016 recorded 11.0% and Cyclone Debbie 2017 recorded 19.0%.

- *Academic and researcher*: 1.0% of sources were researchers and academics, and their distribution was the same (1.0%) both locally and remotely ($p > 0.05$). The distribution of this source between the two locations was almost identical for all events from different topics.

- *Eyewitness*: these were the least estimated sources in the current dataset (<1%). This is because this type of source is most likely related to crisis events with a diffused and progressive nature [Olteanu et al., 2015].

- *Politician*: 3.0% of sources were politicians, with local sources making up 3.0% and remote sources, 2.0% ($p < 0.05$). As expected, the largest proportion (8.0%) of this source type was in political events. The US Presidential Debate was the event with the largest proportion of political sources (13.0%) among the event sources.

- *Digerati*: the proportion of digerati sources was 3.0% , and their distributions locally and remotely were 3.0% and 2.0%, respectively ($p > 0.05$). Digerati were most closely associated with technology and blogging (4.0%) in both entertainment and politics, but only 1.0% for crises.

- *Celebrity*: these accounted for 3.4% of sources from all included events. 3.0 % of local and 4.0% of remote sources were celebrities ($p > 0.05$). As with digerati, celebrities as a source were a larger proportion of authors: 5.0% for both entertainment and political events compared with only 2.0% for crisis events.

- *Ordinary*: among all source categories, the ordinary source provided the largest source (54.0%). Ordinary sources locally were 57.0% and remotely, 52.0% ($p < 0.05$). The results indicate that ordinary individuals are the majority of sources for most events. This finding concurs with those of [Olteanu et al., 2015, De Choudhury et al., 2012]. In times of crisis, 50.0% of sources are ordinary and there is a large difference in the distribution of these sources locally (54.0%) and remotely (45.0%), compared with events in entertainment (52.0%) and politics (67.0%). However, Cyclone Debbie was an event with few ordinary sources: 18.0

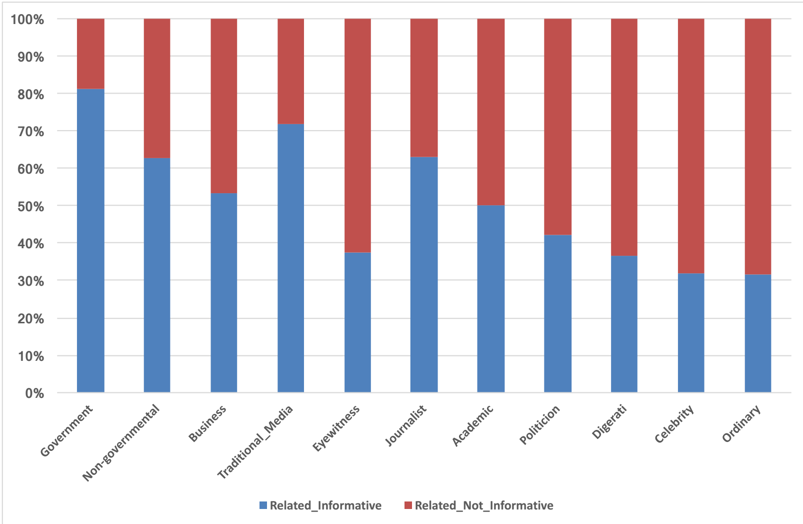


Figure 5.1: The interaction of sources and informativeness (informative and not informative).

5.4.1.2.1 Source v. Informativeness

I studied the interaction between sources and informativeness: for example, Figure.5.1 shows the source distribution within informativeness categories (related and informative & related but not informative). It is evident that some sources, such as government, non-government, business, traditional media and journalist, have more tweets in the 'related and informative' category, while other sources were common in the 'not informative' category (e.g. politician, digerati, celebrity and ordinary).

5.4.1.3 Credibility (Main Method)

The results for the credibility annotations of tweets were 44.1% (4,236 tweets) in the 'definitely credible' category; 54.5% (5,229 tweets) 'seems incredible'; only 1.3% (129 tweets) 'definitely incredible'; and <1% (6 tweets) 'I can't decide'. Next, since 'seems incredible' and 'definitely incredible' belong to an incredible category, I combined them into one class called 'incredible', like [Castillo et al., 2011]. The 'I can't decide' tweets were discarded, so

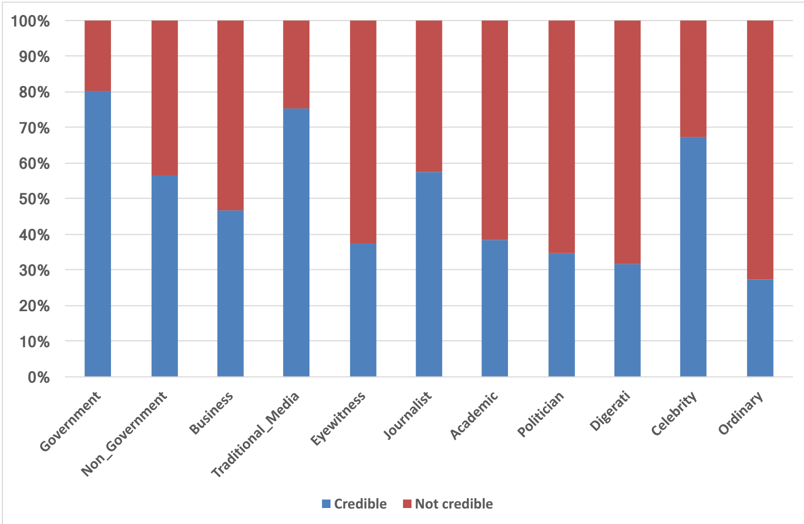


Figure 5.2: The interaction of source and credibility.

I ended up with two credibility classes: ‘credible’ and ‘incredible’.

With regard to the effects of location on content credibility, only 41.0% of local tweets were credible compared with 48.0% of remote ones ($p < 0.001$). It is clear that the content of remote sources was judged more credible than that of local ones ($p<0.001$).

5.4.1.3.1 Source v. Credibility

After annotation of the credibility of tweets, I studied the distribution of sources in the credibility classes. Figure.5.2 shows the distribution of sources in the credibility classes, including that government, non-government, traditional media, journalist and celebrity tweets mostly belonged to the credible class. In contrast, most (>50%) tweets from individual, academic, politician and digerati sources belonged to the incredible class. The other sources were evenly distributed between the two credibility classes in general.

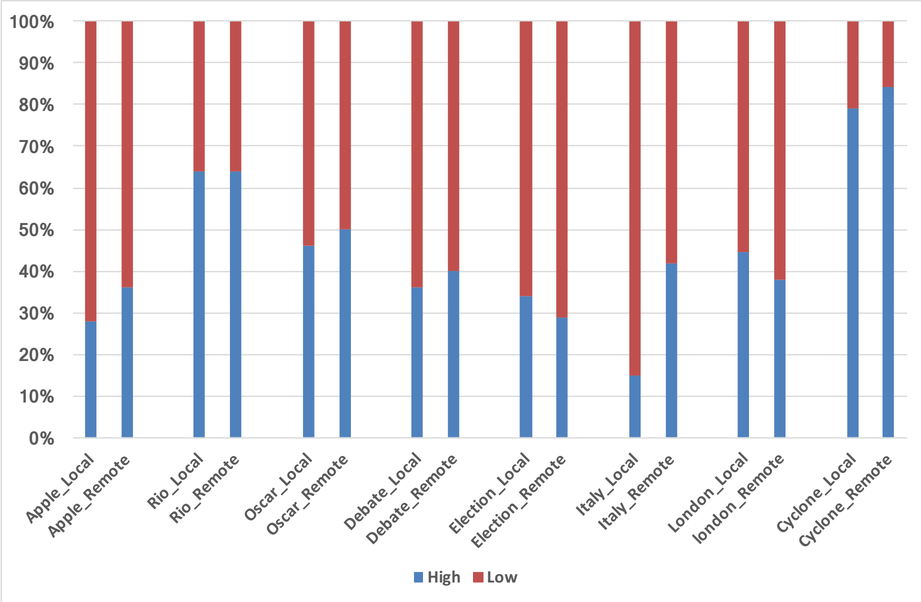


Figure 5.3: High and low-credibility source proportions.

5.4.1.4 Credibility (Source-based Credibility)

In this section I analyse the distribution of the sources of different credibility classes, including the effect of location on the credibility distribution in social media. I used the same categories for credibility classification as those used by [Thomson et al., 2012]. That is, tweets were categorised as high or low credibility based on the source of the tweet: highly credible if the source was highly regarded, and of low credibility if it came from a low-credibility source.

One key component of credibility judgments considered here is source credibility. The classical treatment of credibility considers the source of information as a key determinant of its reliability, which previous studies did not explicitly consider. This enables comparison to be made with other methodologies involving users' judgments.

Figure.5.3 shows the credibility class (high and low) for different events. For each event I present the high and low credibility classes from local and remote locations. This shows how location can determine distribution of credibility classification for most events. The proportion differences for both credibility classes between locations are significant ($p<0.001$).

For example, for Cyclone Debbie and the Italy Earthquake the number of credible sources from remote locations was greater than those from the local location.

5.4.1.5 Linguistic Feature Analysis

Having studied the distribution of informativeness, source and credibility with respect to event, location and topic, in this section I address the second research question: How do linguistic features differ among sources of different type, credibility level, topic and location? I next analyse the features of tweet content by applying multivariate analysis of variance (MANOVA) tests. MANOVA was used to measure the difference between entities such as sources, credibility classes and location when there are many outcome variables-many linguistic features, in the present case. MANOVA creates a linear composite of the outcome variables and determines whether this composite is different between entities as reflected in a significant Pillai's trace result. The MANOVA test also provides results for separate ANOVAs for each individual outcome; even if the MANOVA result is not significant the individual ANOVA for some outcomes may still be significant. I used MANOVA because all of the outcomes were related-all being linguistic features that can provide overall results-so the separate ANOVA can be examined for each outcome.

5.4.1.6 Source and Linguistic Feature Distribution

In this section I present the features across different sources, and source interactions with location and topic. For each factor and factorial interaction, I provide the Pillai's trace result, which indicates how all of the outcome variables (linguistic features) together differ between groups such as source, credibility, location and topic. I follow this by investigating the individual ANOVA results for each linguistics outcome.

5.4.1.6.1 Source

I investigated how the sources differed from each other. The analysis included 9 of the 11 sources (eyewitness and researcher were excluded as there were too few samples: 8 and 78, respectively). The Pillai's trace was significant ($p < 0.001$). The separate ANOVA results showed significant effects for all features ($p < 0.001$) except 'percept' and 'drives' ($p < 0.05$). This means that different sources have different styles of writing, and feature distributions differed in their content. For example, interaction of 'traditional media' linguistic features and location shows three features differ between the two locations 'social', 'pronoun' and 'URL' ($p < 0.05$), while the three-way interaction ('traditional media', location and topic) presents eight features with significant ($p < 0.05$).

Table 5.5: ANOVA results for source interactions with other factors.

	Source v. Location		**Source v. Topic**		**Source v. Location v. Topic**	
	F	**P_{value}**	**F**	**P_{value}**	**F**	**P_{value}**
Function	0.927	0.493	4.864	0.000**	1.903	0.016*
Affect	0.164	0.995	2.749	0.000**	1.956	0.012*
Social	1.359	0.209	7.105	0.000**	2.633	0.000**
Cogproc	1.033	0.409	3.093	0.000**	1.016	0.436
Percept	0.369	0.937	5.676	0.000**	0.928	0.535
Bio	0.224	0.987	0.610	0.879	0.860	0.616
Drives	2.386	0.014*	2.796	0.000**	2.905	0.000**
Relativ	0.822	0.583	8.779	0.000**	2.193	0.004*
Informal	0.377	0.934	9.015	0.000**	2.952	0.000**
Authentic	0.462	0.883	6.083	0.000**	1.385	0.138
WC	0.672	0.717	4.866	0.000**	3.041	0.000**
Pronoun	1.026	0.413	4.135	0.000**	0.864	0.612
Qmark	0.974	0.454	4.930	0.000**	1.363	0.150
Exclam	1.088	0.368	0.640	0.854	0.996	0.457
Hashtags	0.823	0.582	6.298	0.000**	1.494	0.092
URL	0.874	0.538	12.803	0.000**	2.837	0.000**

5.4.1.6.2 Source *v.* Location

After finding significant differences between sources in all features in Section 5.4.1.6, I examined the effect of location on a source's feature distribution. The Pillai's trace result showed no significant interaction between source and location. However, the separate ANOVA showed that 'drives' had a significant interaction (see Table.5.5). This means that the same source type had a similar feature distribution regardless of the source's location.

5.4.1.6.3 Source *v.* Topic

The topic factor influenced feature distribution across different topics. I studied how feature distribution differed for the same source in different topics. Pillai's trace gave a significant result ($p < 0.001$), showing there were significant interactions with all features ($p < 0.001$) except the 'bio' and 'exclam' categories, which had no significant interactions (see Table.5.5).

5.4.1.6.4 Source *v.* Location *v.* Topic

The influence of location on feature distribution with regard to source was weak as shown in the interaction between source and location. Here I include the third factor 'topic' in considering the interaction between source and location. The feature distribution differed considerably between topics: the three-way interaction gave a significant Pillai's trace result ($V = 0.09$, $F(256,9460) = 1.691$, $p < 0.001$). Many features have significant interactions, as shown in Table.5.5. This result indicates that the feature distribution of source tweets across locations changed according to topic.

5.4.1.7 Credibility and Linguistic Feature Distribution

In this section, I examine how the language used differed between credibility levels, and the influence of location and topic on the distribution of these features across credibility classes, as in Table.5.6.

5.4.1.7.1 Credibility

I analysed the distribution of features between the two credibility classes: credible and incredible.

Table 5.6: ANOVA results for credibility interactions with other factors.

	Credibility v. Location		**Credibility v. Topic**		**Credibility v. Location v. Topic**	
	F	**P_{value}**	**F**	**P_{value}**	**F**	**P_{value}**
Function	0.207	0.649	7.070	0.001*	13.602	0.000**
Affect	1.221	0.269	4.578	0.010*	11.765	0.000**
Social	4.216	0.040*	135.564	0.000**	0.842	0.431
Cogproc	0.737	0.391	12.615	0.000**	5.331	0.005*
Percept	2.470	0.116	20.508	0.000**	0.514	0.598
Bio	0.573	0.449	4.879	0.008*	3.822	0.022*
Drives	7.927	0.005*	23.746	0.000**	9.494	0.000**
Relativ	2.763	0.097	105.807	0.000**	7.862	0.000**
Informal	3.429	0.064	34.155	0.000**	8.056	0.000**
Authentic	4.166	0.041*	52.025	0.000**	3.669	0.026*
WC	7.169	0.007*	23.504	0.000**	9.492	0.000**
Pronoun	5.834	0.016*	66.657	0.000**	6.798	0.001*
Qmark	0.288	0.591	46.399	0.000**	0.683	0.505
Exclam	6.871	0.009*	2.521	0.008*	1.414	0.243
Hashtags	0.647	0.421	8.981	0.000**	2.375	0.093
URL	7.090	0.008*	28.862	0.000**	13.034	0.000**

The Pillai's trace's result showed a significant effect of credibility level on the linguistic feature distribution ($V = 0.184$, $F(16,9573) = 134.773$, $p < 0.001$). There were many significant differences between the two credibility classes in all features ($p < 0.001$) except 'drives' and 'authentic'. Moreover, the occurrence of some features in the 'incredible' class was higher than in the credible class (e.g. function, affect, social, cogproc, bio, informal, pronoun, Qmark, exclam and hashtag; $p < 0.001$). This indicates that credible content is not necessary associated with the presence of more features, confirming findings reported by [O'Donovan et al., 2012]. In contrast, only a few features (percept, relative, word count [WC], URL, $p < 0.001$) occurred in the credible class more often than in the incredible class (see Table.5.7).

5.4.1.7.2 Credibility *v.* Location

In this section, I examine the interaction between credibility and location.

The interaction between credibility and location was significant as shown by Pillai's trace ($p < 0.001$). Individual univariate ANOVA results for the outcome variables revealed significant interactions for 'social', 'drives', 'authentic', 'WC', 'pronoun', 'exclam' and 'URL'

Table 5.7: Interaction between the two credibility levels: Credible v. Incredible with the linguistic features.

	Credible mean	**Incredible** mean	**P** Two credibility classes
Function	0.364	0.321	0.000**
Affect	0.098	0.080	0.000**
Social	0.157	0.134	0.000**
Cogproc	0.129	0.095	0.000**
Percept	0.043	0.050	0.000**
Bio	0.029	0.020	0.000**
Drives	0.092	0.094	0.381
Relativ	0.136	0.171	0.000**
Informal	0.158	0.149	0.000**
Authentic	0.185	0.180	0.379
WC	0.477	0.555	0.000**
pronoun	0.149	0.081	0.000**
Qmark	0.047	0.011	0.000**
Exclam	0.015	0.008	0.000**
Hashtags	0.122	0.115	0.001*
URL	0.130	0.180	0.000**

($p < 0.05$). Table.5.8 presents significant interactions between credibility and location for the features found to be affected by the interaction. A Bonferroni correction was applied to mitigate the multiple comparison effect.

Table 5.8: Interaction between the two factors: credibility v. location with the linguistic features.

	Credible			**Incredible**		
	Local	**Remote**	**P**two locations	**Local**	**Remote**	**P** two locations
Social	0.136	0.132	0.212	0.164	0.150	0.000**
Drives	0.100	0.089	0.000**	0.092	0.092	0.909
Authentic	0.186	0.174	0.193	0.175	0.192	0.041*
WC	0.548	0.561	0.004*	0.479	0.476	0.394
Pronoun	0.082	0.079	0.437	0.157	0.141	0.000**
Exclam	0.008	0.008	0.893	0.018	0.013	0.000**
URL	0.166	0.194	0.000**	0.125	0.136	0.016*

The credible class had three features that were affected by location: ‘drives’ was significantly more frequent in tweets from the local than remote locations, while ‘WC’ and ‘URL’ were more frequent in remote than in local tweets. Location influenced five features in the incredible class: two (social and URL) were more common in local tweets, and the others (authentic, pronoun and exclam) were the opposite. In both classes, 'URL' was more common in remote location than local ones. To verify the cause of these differences between location types for the same credibility class and determine whether source distribution difference had any effect, I ran three-way interactions that included credibility, source and location. The differences between the two credibility classes in the six features remained, which indicates a negligible effect of source distribution on the results.

Table 5.9: Interaction between the two factors: credibility v. topic with the linguistic features.

	Entertainment			**Politics**			**Crisis**		
	Credible	**Incredible**	**P** two levels	**Credible**	**Incredible**	**P** two credibility	**Credible**	**Incredible**	**P** two levels
Function	0.275	0.339	0.000**	0.365	0.398	0.000**	0.323	0.354	0.000**
Affect	0.074	0.090	0.000**	0.090	0.102	0.007*	0.076	0.103	0.000**
Social	0.100	0.137	0.000**	0.187	0.156	0.000**	0.116	0.179	0.000**
Cogproc	0.063	0.109	0.000**	0.139	0.151	0.021*	0.083	0.128	0.000**
Percept	0.053	0.054	0.795	0.044	0.046	0.664	0.052	0.029	0.000**
Bio	0.017	0.029	0.000**	0.020	0.022	0.450	0.024	0.035	0.000**
Drives	0.098	0.078	0.000**	0.097	0.101	0.337	0.087	0.099	0.001*
Relativ	0.194	0.131	0.000**	0.136	0.139	0.582	0.230	0.140	0.000**
Informal	0.153	0.166	0.000**	0.151	0.137	0.001*	0.142	0.172	0.000**
Authentic	0.126	0.181	0.000**	0.145	0.181	0.002*	0.269	0.193	0.000**
WC	0.551	0.446	0.000**	0.542	0.494	0.000**	0.570	0.493	0.000**
Pronoun	0.050	0.130	0.000**	0.138	0.161	0.000**	0.053	0.155	0.000**
Qmark	0.009	0.048	0.000**	0.018	0.028	0.022*	0.006	0.065	0.000**
Exclam	0.010	0.021	0.000**	0.010	0.017	0.000**	0.002	0.008	0.000**
Hashtags	0.123	0.119	0.149	0.118	0.133	0.000**	0.104	0.114	0.001*
URL	0.224	0.140	0.000**	0.129	0.096	0.000**	0.186	0.154	0.000**

5.4.1.7.3 Credibility *v.* Topic

The Pillai's trace result revealed a significant interaction between credibility and topic with features ($p < 0.001$). The separate univariate ANOVAs showed that all features had a significant interaction with credibility and topic. Table.5.9 shows that the features found to be important via the ANOVA test differed between the credibility classes for different topics. Crisis was the most affected topic, followed by entertainment and politics. Some features, including 'WC' and 'URL' were highly significant in the credible class for all topics. Some features, such as ‘function’, ‘affect’, ‘cogproc’, ‘pronoun’, ‘Qmark’ and ‘exclam’ were frequent in the incredible class for all topics. Moreover, some features, such as ‘informal’ and ‘social’ were common in the incredible class in entertainment and crisis tweets, while the opposite applied to the politics topic as it was common in the credible class. Conversely, some features differed between the two credibility classes for some topics: for example, ‘bio’ was common in the incredible class for entertainment and crisis, and ‘percept’ was common in the credible class in crisis but not in political tweets.

5.4.1.7.4 Credibility *v.* Location *v.* Topic

In this section I examine the three-way interaction between credibility, location and topic. The Pillai’s trace results revealed a significant interaction among the three factors (V = 0.014, F(32,9575) = 4.115, p < 0.001). This result shows that the features of different topics differ significantly when interacting with credibility and location. Separate univariate ANOVAs on the outcome variables showed significant interactions involving 11 features (function, affect, drive, relative, informal, WC and URL, p < 0.001.) and (cogproc, bio and authentic, p < 0.05).

Table 5.10: Linguistic features distribution between the two locations for each topic in credible tweets.

	Credible								
	Entertainment			**Politics**			**Crisis**		
	Local	**Remote**	**P** two locations	**Local**	**Remote**	**P** two Locations	**Local**	**Remote**	**P** two Locations
Function	0.278	0.272	0.633	0.366	0.363	0.848	0.349	0.297	0.000**
Affect	0.070	0.078	0.163	0.087	0.093	0.266	0.085	0.066	0.000**
Cogproc	0.060	0.067	0.378	0.137	0.140	0.671	0.095	0.71	0.000**
Bio	0.014	0.021	0.047*	0.022	0.019	0.414	0.023	0.024	0.848
Drives	0.105	0.090	0.007*	0.097	0.098	0.868	0.097	0.077	0.000**
Relativ	0.139	0.158	0.008*	0.136	0.135	0.967	0.239	0.220	0.004*
Informal	0.152	0.155	0.528	0.150	0.152	0.830	0.136	0.148	0.018*
Authentic	0.121	0.131	0.535	0.141	0.148	0.657	0.295	0.243	0.000**
WC	0.531	0.571	0.000*	0.546	0.539	0.452	0.567	0.574	0.320
Pronoun	0.045	0.055	0.165	0.136	0.141	0.507	0.066	0.041	0.000**
URL	0.220	0.227	0.439	0.118	0.141	0.008*	0.159	0.213	0.000**

Table.5.10 presents the mean differences between the features in the credible class for

different locations and topics. Of the three topics, crisis was the most affected by location classes: it had 9 out 11 features with significant differences between the two locations. Seven of these features were common in local sources, and only two features were common remotely (informal and URL).

In the credible class, the features in entertainment and politics were less sensitive than the crisis topic to the location change. 'Bio', 'relative' and 'WC' in entertainment, 'URL' in politics and 'cogproc', 'informal' and 'URL' in the crisis topic were significantly more common for remote than local location tweets. Interestingly, for politics and crisis, 'URL' was significantly common at a remote location. This finding shows that remote tweets most likely include third party information.

Table.5.11 presents the mean differences in features in the incredible class for different topics and locations. The table shows how tweets from different locations classified as incredible compared in their linguistic feature distributions.

Five features had significant differences between local and remote classes in political tweets compared with four for entertainment and crisis. All features with a significant difference in political tweets were more common locally than remotely; the opposite was true for significant features in entertainment except 'function was higher locally', which shows that the influence of location differs between topics. For the crisis topic, 'informal' and 'pronoun' features were more common in local that remote tweets; the opposite was true for the other two features (relative and WC).

Comparing Table.5.10 and .5.11, indicates the effect of location on the linguistic feature distributions of the two credibility levels for the same topic. For example, there was almost no effect of location for political tweets in the credible class, whereas in the incredible class there was a noticeable influence of location. For crisis tweets this effect was in the opposite direction; it was strongly affected by location in the credible class with less effect in the incredible class. The influence of location on entertainment was the same for both credibility levels.

All these observations reveal the effect of author location on the linguistic features in credibility classes. Moreover, the observations indicate that the influence of location on credibility is not the same for different topics.

Table 5.11: Linguistic features distribution between the two locations for each topic in incredible tweets.

	Incredible								
	Entertainment			**Politics**			**Crisis**		
	Local	**Remote**	**P** two locations	**Local**	**Remote**	**P** two locations	**Local**	**Remote**	**P** two locations
Function	0.354	0.324	0.000**	0.412	0.384	0.015*	0.350	0.359	0.320
Affect	0.094	0.086	0.101	0.110	0.093	0.006*	0.100	0.106	0.233
Cogproc	0.107	0.112	0.430	0.156	0.146	0.209	0.124	0.131	0.322
Bio	0.031	0.026	0.054	0.021	0.024	0.496	0.035	0.036	0.696
Drives	0.073	0.082	0.028*	0.108	0.094	0.011*	0.096	0.102	0.170
Relativ	0.125	0.136	0.050	0.140	0.137	0.732	0.130	0.150	0.001*
Informal	0.158	0.174	0.000**	0.142	0.133	0.125	0.179	0.165	0.002*
Authentic	0.173	0.188	0.222	0.178	0.185	0.654	0.185	0.202	0.200
WC	0.441	0.451	0.114	0.519	0.469	0.000**	0.479	0.507	0.000**
Pronoun	0.130	0.130	0.918	0.179	0.144	0.000**	0.162	0.148	0.017*
URL	0.123	0.158	0.000**	0.103	0.090	0.128	0.149	0.160	0.130

5.4.1.8 Source-based Credibility and Linguistic Feature Distribution

In this section, I examine how source credibility can affect the language used, since not all sources have the same credibility level. I grouped the sources into high and low. Ordinary sources were classified as low class as they are not authentic and cannot be classified into one of the authentic sources. The high credibility class contains all sources except ordinary, which are called trusted sources as in [Dailey and Starbird, 2014].

I include this section to help address the third research question and compare the two credibility assessment methodologies. In the discussion I compare between the two methods and show how effective is the second one.

5.4.1.8.1 Credibility

. I analysed the distribution of features between the two credibility classes credible and incredible, in addition to the interaction between credibility and other factors – location and topic –, as in Table.5.12.

Table 5.12: ANOVA results for credibility (source-based) interactions with other factors.

	Credibility v. Location		Credibility v. Topic		Credibility v. Location v. Topic	
	F	**Sig**	**F**	**Sig**	**F**	**Sig**
Function	4.151	0.042*	10.958	0.000**	7.993	0.000**
Affect	0.202	0.653	0.237	0.789	3.865	0.021*
Social	0.534	0.465	47.713	0.000**	0.112	0.894
Cogproc	4.422	0.036*	1.466	0.231	0.115	0.891
Percept	0.354	0.552	27.651	0.000**	0.545	0.580
Bio	0.894	0.345	0.876	0.416	0.418	0.658
Drives	4.871	0.027*	1.807	0.164	6.211	0.002**
Relativ	1.742	0.187	55.443	0.000**	8.320	0.000**
Informal	1.027	0.311	52.400	0.000**	11.510	0.000**
Authentic	2.388	0.122	32.958	0.000**	5.413	0.004**
WC	0.491	0.484	17.844	0.000**	1.385	0.250
Pronoun	0.972	0.324	20.677	0.000**	0.617	0.539
Qmark	0.873	0.350	37.326	0.000**	1.504	0.222
Exclam	1.886	0.170	3.746	0.024*	2.040	0.130
Hashtags	0.165	0.685	8.500	0.000**	2.401	0.091
URL	4.593	0.032*	37.340	0.000**	8.880	0.000**

Using Pillai's trace, there was a significant effect of credibility class on linguistic features distribution: $V = 0.09$, $F(15,9574) = 70.024$, $p<0.001$. The two credibility classes differed in many features (i.e. function, affect, social, cogproc, bio, drives, relative, WC, pronoun, exclam, Qmark, and URL, $p<0.001$; drives and hashtag, $p<0.05$). Some features in the incredible class was higher than in the credible class (e.g. as function, social, cogproc, affect, hashtag, Qmark and bio). In contrast, in the credible class, only a small number of features (URL, WC and relative) occurred more often than in the incredible class ($p<0.001$).

5.4.1.8.2 Credibility *v.* Location

This section explore the interaction between credibility and location with distribution of features. For example, tweets in the same credibility class but from different location classes may have similar feature distributions, or the location factor may influence the feature distributions.

Table 5.13: Interaction between credibility classes and locations with different features.

	High			**Low**		
	Local	Remote	$P_{\text{two locations}}$	Local	Remote	$P_{\text{two locations}}$
Function	0.343	0.313	0.000**	0.365	0.352	0.017*
Cogproc	0.105	0.095	0.021*	0.125	0.127	0.590
Drive	0.099	0.91	0.010*	0.090	0.091	0.689
Url	0.165	0.195	0.000**	0.120	0.136	0.000**

Pillai's trace results showed that the interaction of credibility and location with all dependent variables (features) was not significant ($p > 0.05$). However, separate univariate ANOVAs on the outcome variables revealed significant interactions with 'function', 'cogproc', 'drives' and 'URL' ($p < 0.05$). Table.5.13 shows the features found to be affected by the interaction between credibility and location. A Bonferroni correction was applied to mitigate the multiple comparison effect.

In the credible class, all features were significantly more common in local location than remote tweets, except for 'URL'. In contrast, in the incredibility class only two features had significant differences between the two location types (function and URL). 'Function' was more common locally than remotely, but the opposite was true for 'URL'. In both classes, 'URL' was higher at a remote location than at a local one. This indicates that remote authors shared more URLs, regardless of their credibility, suggesting that the influence of location is greater than the influence of credibility in this case.

The differences in Table.5.13 indicate the influence of a source's location on their behaviour even when they were classified into the same credibility class. The credible class included multiple sources that might have caused the observed differences. However, this is not likely to be the case because (a) the balance of the high credible sources was almost the same between the two locations; and (b) although the low credibility class included one source type, two features differed significantly across locations.

Table 5.14: Interaction between credibility classes and topics with different features.

	Entertainment			Politics			Crisis		
	High	**Low**	$\mathbf{P}_{\text{two credibility}}$	**High**	**Low**	$\mathbf{P}_{\text{two credibility}}$	**High**	**Low**	$\mathbf{P}_{\text{two credibility}}$
Function	0.291	0.339	0.000**	0.358	0.394	0.000**	0.337	0.342	0.392
Social	0.124	0.123	0.919	0.176	0.169	0.163	0.127	0.171	0.000**
Percept	0.051	0.055	0.176	0.040	0.047	0.059	0.051	0.029	0.000**
Relativ	0.140	0.135	0.193	0.144	0.133	0.063	0.216	0.148	0.000**
Informal	0.172	0.151	0.000**	0.152	0.140	0.003*	0.145	0.170	0.000**
Authentic	0.138	0.180	0.000**	0.148	0.171	0.055	0.260	0.197	0.000**
WC	0.534	0.441	0.000**	0.546	0.503	0.000**	0.563	0.496	0.000**
Pronoun	0.078	0.122	0.000**	0.546	0.158	0.000**	0.073	0.142	0.000**
Qmark	0.025	0.041	0.000**	0.018	0.026	0.070	0.012	0.061	0.000**
Hashtag	0.123	0.118	0.105	0.119	0.129	0.015*	0.103	0.115	0.000**
Url	0.216	0.129	0.000**	0.142	0.097	0.000**	0.182	0.157	0.000**
Exclam	0.013	0.021	0.000**	0.013	0.014	0.617	0.003	0.007	0.002*

5.4.1.8.3 Credibility *v.* Topic

The MANOVA results indicated a significant interaction of credibility and topic with features ($p<0.001$). The individual ANOVAs presented that eleven features had a significant interaction with credibility and topic (function, social, percept, relative, informal, authentic, WC, Qmark, hashtag and URL, $p<0.001$; exclam, $p<0.05$). The features distribution of the credibility classes were not the same across topics, as in Table.5.14. The features differences between the two credibility classes in criss topic more than any other topics. Some features, such as 'WC' and 'URL' were highly significant in the credible class for all topics. Some features, such as 'function' were common in the incredible class in entertainment and political tweets, but there was no difference between the two credibility classes for crisis. Moreover, some features, such as 'informal' were high in the credible class for entertainment and politics, while the opposite was true for the crisis topic, as 'informal' was common in the incredible class.

5.4.1.8.4 Credibility *v.* Location *v.* Topic

As Pillai's trace showed no significant interaction between credibility and location, in this section I examine the three-way interaction between credibility, location and topic.

The MANOVA results identified a significant three-factor interaction ($V = 0.008$, F(32,9575) = 2.31, p<0.001). These results show that the features differ significantly among credibility classes when interacting with topics and location. Individual ANOVAs on the linguistic features showed significant interactions for seven features (function, relative, informal and RUL, p<0.001; affect, drives and authentic, p<0.005).

Table 5.15: Interaction between locations and topics with different features for credible tweets.

	Credible								
	Entertainment			**Politics**			**Crisis**		
	Local	**Remote**	**P** two locations	**Local**	**Remote**	**P** two locations	**Local**	**Remote**	**P** two locations
Function	0.298	0.289	0.106	0.371	0.346	0.063	0.361	0.312	0.000**
Affect	0.074	0.074	0.937	0.064	0.084	0.986	0.087	0.071	0.002*
Drives	0.089	0.087	0.735	0.104	0.103	0.886	0.103	0.082	0.000**
Relativ	0.127	0.154	0.000**	0.149	0.139	0.314	0.222	0.210	0.054
Informal	0.172	0.173	0.904	0.154	0.151	0.682	0.140	0.151	0.014*
Authentic	0.124	0.151	0.051	0.158	0.138	0.315	0.279	0.240	0.004*
Url	0.208	0.224	0.023*	0.129	0.156	0.014*	0.158	0.206	0.000**

Table.5.15 provides the means of features in the credible class for different locations and topics. Crisis was the most affected topic by location: it had six out seven features with significant differences between the two location types. Four of these features were highly significant locally, and only two features were common in remote texts (informal and URL).

The entertainment and political topics were less sensitive to change of location in credible class. The features 'relative' and 'URL' for entertainment and 'URL' for politics were significantly more common in remote than local location tweets. For all three topics, 'URL' was significantly more common in remote location texts.

Table.5.16 shows the means for features in the incredible class for the three topics and two locations. This table demonstrates how tweets from sources of different topics in different locations classified as having low credibility can have the same or different linguistic features.

While no differences were found for politics, three features showed significant differences between local and remote classes for each of the entertainment and crisis topics; two were common to both topics. The 'informal' feature was significantly more common in remote location tweets on the topic of entertainment, but the opposite was true for crises. The second feature was 'URL', which was significantly more common in remote location tweets on both topics. 'Function' was the third entertainment feature and it was more common in local than remote tweets ($p<0.001$). The 'relative' feature was the third feature in crisis topic tweets, and it was high for remote sources ($p<0.005$).

Table 5.16: Interaction between locations and topics with different features for incredible tweets.

	Incredible								
	Entertainment			**Politics**			**Crisis**		
	Local	**Remote**	**P** two locations	**Local**	**Remote**	**P** two locations	**Local**	**Remote**	**P** two locations
Function	0.356	0.321	0.000**	0.398	0.389	0.340	0.340	0.345	0.586
Affect	0.096	0.092	0.326	0.106	0.098	0.147	0.099	0.103	0.503
Drives	0.079	0.084	0.191	0.101	0.091	0.051	0.090	0.097	0.134
Relativ	0.132	0.138	0.347	0.132	0.135	0.687	0.139	0.157	0.003*
Informal	0.142	0.160	0.000**	0.142	0.137	0.398	0.179	0.162	0.000**
Authentic	0.184	0.177	0.596	0.160	0.182	0.134	0.193	0.202	0.474
Url	0.109	0.149	0.000**	0.100	0.094	0.433	0.149	0.165	0.039*

5.5 Discussion

Around the world, many events occur daily, and social media has become a significant forum that people use to read and share information about such events [Kwak et al., 2010]. Events are combined from many topics, regardless of the type of event or the authors who contribute posts about these events, who can be close to or far from the location of the event. Information posted about an event varies in credibility and can include inaccurate and false information, such as rumours.

In reviewing the literature, no data were found regarding the association between credibility and factors such as location, topic and linguistic features. Olteanu et al. [2015] found that source and information types differ among events, and location affects author behavior in terms of language use when a tweet author broadcasts from the affected region at the time of the event [Morstatter et al., 2014, Poblete et al., 2011]. The content features of tweets are found to differ among credibility levels [Mitra et al., 2017b, O'Donovan et al., 2012]. However, no previous research has investigated how the distribution of information sources differs within and outside the country of an event, and how author location can influence linguistic feature distribution among credibility levels.

5.5.1 Source Distribution

The current study found that the location of an author influences the distribution of sources that contribute at the time of an event, and this influence also the credibility assessment. As we saw in Table.5.3, the distribution of many source types differs across locations. This finding can influence authors' classification models [De Choudhury et al., 2012], where they do not consider the authors' classification in relation to sources' location. For example, during a crisis, it is necessary to define the type of information source to easily find the needed information, as in [Imran and Castillo, 2015].

5.5.1.1 Source Distribution and Informativeness across Locations

Table.5.3 shows source distribution locally and remotely, and Figure.5.1 presents the informativeness status of different sources. These types of result can provide an expectation of what different types of source and informative status will contribute at the time of an event from different locations, which can be beneficial to many people. For example, decision makers tend to turn to Twitter when they need to take an action, such as during crisis [Olteanu et al., 2015]. Also, it is a common practice for journalists to deal with a huge amount of data manually to develop a news story [Dailey and Starbird, 2014, Heravi and McGinnis, 2015], so it is important for them to have knowledge of the types of source that Twitter might provide during different events. For example, classifying sources along with information types that might be provided for each type can help journalists to anticipate what kind of information they will get from social media.

5.5.2 The Relationship between Credibility and Linguistic Features

I contend that this research presents the importance of textual features for identifying credibility in social media, and demonstrates the influence of an author's location on their style of writing. These findings can inform a wide range of systems: for example, news reporting systems that try to reach a credible source like a journalist or eyewitness at the time of and event [Diakopoulos et al., 2012]; or a system for fact checking that differentiates between high and low-credibility content [Popat et al., 2017]. While I do not claim that a stand-alone system with textual features only will enable rumour detection, it can be used as an extra credibility signal.

Most research that has attempted to classify credibility in social media like Twitter used features beyond the content of the tweets, such as social and network-based approaches [Castillo et al., 2013], temporal approaches [Mitra et al., 2017a] or the popularity of tweets [Mendoza et al., 2010]. As all of these are useful, the previous features need sometimes after posting the content to be collected [Zhao et al., 2015]. Using linguistic features as a marker is a key factor in identifying incredible content, preventing possible damage that may occur [Bovet and Makse, 2019].

5.5.2.1 The Influence of Location on Credibility Linguistic Features

Existing credibility prediction models face difficulty when applied to different events [Boididou et al., 2014, Aker et al., 2017], as performance accuracy is overestimated. The results reported here show that there are significant differences in content for the same credibility level and topic when generated from different locations. For example, Table.5.10 shows many differences in linguistic features for the same topic from a different location. This finding demonstrates the influence of location on feature distribution, which might indicate some ambiguity underlying credibility classifier variance across topics.

When we look at features proportions of tweets, we can see some commonalities regardless of credibility. For example, URLs feature strongly in tweets from remote locations at both the credible and incredible level. This suggests that remote authors always share third party information as an external source in their tweets: using URLs as an indicator of credibility at the time of an event might not be very accurate because it might relate more to the place of the author than to credibility. However, Popat et al. [2016] shows the effect of the reliability of the web source hosting such articles on increasing the performance of credibility model prediction. The same approach can be used for social media by studying the reliability of URLs in messages and their effect on the performance of credibility models.

Location influences the language used for tweets at the same credibility level. For example, Table.5.11 shows that for all topics incredible tweets do not always have the same features when generated from different locations. This differs slightly from the findings of [Kwon et al., 2013], who found that rumours on different topics share linguistic features.

On the other side, there are many significant differences between the two location types in credible tweets for topics like crisis, while there is less effect of location on incredible tweets for the same topic. This is an interesting finding, which means that credible tweets have different characteristics when generated close to the affected region. This might have implications for eyewitness identification research [Morstatter et al., 2014].

As this study found that incredible tweets share the same linguistic features in social media; the same behaviour in different contexts, including financial fraud and web pages. Humpherys et al. [2011] investigated the text of hundreds of financial disclosures and found some linguistic differences between their content, such as that fraudulent disclosures use more words than non-fraudulent ones. The researchers were able to achieve high classification accuracy using linguistic features. Wawer et al. [2014] used linguistic features to judge the credibility of web pages, finding that trusted words are associated with government web pages, for example. The researchers' classification models achieved accurate results.

Research has shown how textual features have more power to assess credibility, compared to other features like visual ones, which can sometimes lead to incorrect judgments by authors [Zubiaga and Ji, 2014]. Thus, implementing a credibility model that can handle the text of different contexts will be beneficial.

In this research, I mainly focused on the source and credibility of content, regardless of information type. The information may be credible but not highly informative, so future research could study the content credibility in relation to informativeness status as in [Kumar et al., 2013]. Also, the future research can study credibility with time, as in some cases such as crisis, time is a very important factor. We know that a very few authors generate most of the social media content: for example, 2% of Twitter authors generate 50% of the tweets [Baeza-Yates and Sáez-Trumper, 2015]. This creates further challenges for finding informative and credible sources at the same time because of the scarcity of information sources.

5.5.3 The Effectiveness of the Source-based Credibility Method

In this section I discuss the extent to which we can trust the results from the second method of assessing credibility (source-based), as it blindly classifies tweets regardless of the information they contain. To do this I compare the results with the results obtained using the first method (via crowdsourcing), which is the most common methodology for studying the credibility of social media content. The purpose of this is to find an alternative way to annotate content credibility with less effort (quicker and less costly than the traditional method of crowd annotation).

Examining the results for the two credibility assessment methods in Sections.5.4.1.7 and 5.4.1.8, it is evident that with both methods, the two credibility classes differ significantly in most of their linguistic features. Also, the effects of location and topic are evident in the distributions of the studied features. However, there are some differences between the methods in terms of the number of features that show significant differences. For example, in the interaction between credibility and location, there were seven features with significant differences in method one, but only four differences for method two; with two features in common between them. The same was found for the interactions credibility v. topic and credibility v. location v. topic.

In general, the first method found more differences between the two credibility classes than did the second one. However, it is clear that the second method is able to differentiate the two credibility classes, even when interacting with location or topic. Gadiraju et al. [2015] reported many challenges when using crowdsourcing for data labelling, including increasing task rewards, time-consuming, the malicious actions of some crowdsourcing workers and the existence of many ethical constraints. Thus, the second method might become an alternative way of labelling the credibility of social media data because it based on information source, while the models for classifying information sources in social media already exist such as [De Choudhury et al., 2012].

5.6 Summary

In this chapter, for a diverse set of events, I considered the effect of location on the information source and credibility level in social media. After developing the hypothesis-driven by previous research-that the location of authors affects their behavior, I included the location of both event and author to clarify their effect on credibility status. The research questions investigated in this chapter were:

1. What is the effect of sources' location and linguistic features on information credibility of social media? :

- What types of sources are expected for different events from both in- and outside the country?

- How do linguistic features differ among sources of different type, credibility level, topic and location?
- How effective is it to only use the information source to assess credibility?

I found that the distribution of some sources differs significantly between locations. For the second research question, I found that tweets within the same credibility level have different linguistic features based on their distance from an event and the topic of an event. Future work should include other features to derive a complete list of common and different behaviors between sources across locations. Moreover, information type with authors' location can influence content credibility, and this is part of proposed future research. For the third research question, I found that assessing the credibility of social media content based on the source might be used as an alternative approach to annotation, which may save much effort. However, further research is needed to accurately classify sources into corresponding credibility classes.

Chapter 6

Conclusion

In this thesis, I addressed the challenge of credibility in social media by answering the following three main research questions:

- Which features affect readers' perceptions of credibility?
- What is the effect of readers' location on their perceptions about the credibility of social media information?
- What is the effect of location on sources distribution and linguistic features for information credibility of social media?

6.1 Solutions and Contributions

In this section, I explain the solutions that have been contributed by this thesis.

6.1.1 RQ1: Features affecting readers' perceptions of credibility

- **I studied a number of non-textual features to examine readers' credibility perceptions.**

The non-textual features of tweets have a powerful influence on readers' credibility judgments. The influence of these features can vary based on the contextual topic of a tweet. This research studied six features: topic of a tweet, and author's gender, username, profile image, tweet location and network overlap. The study included different types of features to understand reader perceptions. It found that readers are affected by feature differences: different types of the same feature have a different effect on readers' credibility judgments. For example, a male profile image was the image type that most influenced readers' credibility perceptions. The effect of author location was influenced by a tweet's topic: political tweets that included an author location were perceived as highly credible compared with other tweets. In contrast, some features were associated with low credibility, such as the default profile image and the Internet style username. I presented this contribution in Chapter 3 and Section 4.2.4.

6.1.2 RQ2: Effect of readers' location on credibility:

- **I ran a number of controlled experiments, including the features mentioned in the first contribution.**

These experiments were run in two languages, Arabic and English. Participants represented two cultures from multiple countries. Crowdsourcing platforms were used to evaluate the effect of tweet features on readers' credibility perceptions. The results showed the importance of the included features in determining the credibility judgments of readers from different cultures and countries. They also showed the effect of a controlled experimental methodology on detecting readers' credibility perceptions accurately, as I used the same data belonging to each culture. This allowed me to avoid the confusion seen in previous studies between the influence of culture and country. This contribution was presented in Section 4.3.

- **I used effect size as an additional method for analysing results.**

For deep analysis and further investigation, I performed a effect size analysis. This method can provide the effect size for each feature in different readers' location classification levels and provides the actual effect of features on credibility judgments of readers. No other credibility research has applied this method to their results. The results present different understandings relating to the effect of features on credibility perception. Significance tests indicate differences between the two groups, while effect size inform about the magnitude of a difference. This contribution was presented in Section 4.4.7.

- **I performed a comparative analysis between the influence of culture and country on social media credibility perceptions.**

A social media user belongs to a geographical location, which is part of a country, and the country is representative of a culture. This classification is vital as it has been found to influence social media users in general, as readers or authors. In this research, I classified consumers (readers) into five levels: cultures, regions of countries, countries, regions and divisions within the same country. Unlike in previous research, the results showed the effect of each level separately from mixing with other levels. Although culture has the greatest influence on readers' credibility judgments, the lower classification levels including country and regions within countries have some effects as well. The results indicate that together, culture and country level are important when designing a credible social search system. This contribution was presented in Sections 4.4.2 to 4.4.6.

6.1.3 RQ3: Effect of distance between source and event location on social media credibility:

- **I evaluated the relationship between location and number of content types.**

The influence of location on the behaviour of social media authors is clear, especially in relation to generated content. In this research, I studied the relationship between location and three types of content. The results showed that the influence of location is prominent in all three content types. The proportions of the same information type, 'informative tweets' and 'credible tweets', differ between the two location types. This contribution was presented in Section 5.2.2.

- **I performed an estimation of the types of source that contribute to different events.**

I examined the distribution of the most popular information sources, both within and outside the country of an event. The results showed that the ratio of sources locally and remotely differed for most source types. These differences influenced the quality of the generated content including informativeness and credibility. This contribution was presented in Section 5.4.1.2.

- **I studied the effect of source location on linguistic feature distribution when studying credibility.**

The results of this thesis included a large effect of author (source) location on the semantic features of tweet content. This contribution was presented in Sections 5.4.1.6 and 5.4.1.7.

- **I employed MANOVA to analyse LIWC semantic feature distribution in tweet content.**

LIWC is a powerful tool in the area of data science. It has the ability to analyse large amounts of data and includes more than 90 linguistic features. MANOVA helped me to compare these features for different credibility classes, locations and topics. The results showed the effectiveness of using statistical analysis models to understand which factors influence semantic feature distribution, so that this can be included in credibility models. This contribution was presented in Section 5.4.1.5.

- **I tested an alternative way of annotating tweet credibility.**

Labelling data for research is a costly process in terms of effort, money and time. In this thesis, I tested an alternative method for annotating tweet credibility by using their source reputation. I used a basic procedure for classifying sources into different credibility levels. I compared this method with the human annotator method and found they produced similar results. The results indicate the effectiveness of this procedure and the findings in many cases were close to the human annotation results. This approach may provide a basic way of labelling information credibility in social media and is worthy of further research. This contribution was presented in Section 5.4.1.8.

6.2 Limitations

This thesis, like all research, has limitations. The main limitations of this work are described below.

First, most credibility research in social media uses Twitter datasets. The main reason for this is the access that Twitter provides to researchers for data collection. One might question whether these findings are transferable among social media environments, or are only applicable to a specific dataset or event, for example. The MediaEval workshop and taskforce attempted to explore such a question for image credibility [Boididou et al., 2016, Ionescu et al., 2016].

Second, I applied the results to two cultures and countries from the same culture, and used them to study the effect of culture on social media credibility. Previous research explored the Chinese culture and credibility [Yang et al., 2013]. Thus, studying other cultures and countries can provide a more broad understanding of similarities and differences between them, so that the findings can be applicable at the international level. Moreover, in this thesis for the last research question I used only English language tweets because it is the dominant language in social media. However, including other languages is important to include in further research as language differences have been found to enhance prediction in social media [Han et al., 2014], and generated content differs between languages in social media [Weerkamp et al., 2011].

Third, the included authors in chapters 3 and 4 when studying the credibility perception were individuals; I did not include other types of sources such as organisations. Even though the majority of Twitter authors are individuals, including organisation authors might provide new understanding.

Fourth, the dataset used in Chapter 3 was quite unbalanced with regard to participant gender; however, the results were consistent with previous research findings, which also included differences between participant demographics [Morris et al., 2012].

6.3 Future Work

A possible direction for future research is implementing a credibility model using some non-textual features as part of credibility prediction. All credibility models are based on textual or metadata features. Sometimes textual features can be short or not helpful for credibility assessment. For example, the profile image can be used to estimate author credibility in the same way that [Wei and Stillwell, 2017] did to estimate intelligence based on the profile image. The same can be applied to credibility assessment.

Another possible direction is to use author location along with semantic features to predict the most credible authors, for example, train a model on credible local sources in different contexts, then use this model to find these local sources in the absence of eyewitnesses. Such models may be helpful at critical times to reach a source quickly. The same type of model can also be applied for other types of source for different purposes.

Moreover, it would be valuable to include languages other than English for further analysis of the effect of distance between source and event. Although, no research has investigated the influence of language differences on credibility in social media, I believe that this could increase the accuracy of credibility models. This is an interesting area for further investigation. For example, applying existing credibility models (which use linguistic features for assessing credibility) in different languages may lead to a new understanding of the limitations of credibility models. Moreover, including language and country together as representative of culture would be valuable in future research; minimising the scope of trained data can help reduce the overestimation problems with current credibility models.

When performing future crowdsourcing tasks, it is vital to follow platform instructions to achieve the best design for your task. Also, since credibility is subjective, it is necessary to make the task very clear for workers, in terms of how they should judge credibility, by providing examples and training them. Otherwise, it is very hard to find agreement among workers. Moreover, it is always recommended to run around 10% of data at the begin to ensure that the task is clear and the gold questions work adequately.

A further potential direction for future work is to combine credibility with other types of content. To date, credibility research has been general (e.g. obtaining credible information for a given topic), but assessing credibility based on type of information is needed as well (e.g. in times of crisis, information types can be classified into donation, caution& advice and so on [Olteanu et al., 2015]). Therefore, assessing credibility in addition to the type of information has the potential to shorten the time to reach affected people. Moreover, it is important to investigate the relationship between credibility and time, thus to explore when credible information begins to be shared; this has not been studied yet.

Appendix A

Ethics Approvals and Plain Language Statements

Here are the ethics approvals for the experiments run in this thesis, and we include the plain language statements as well. First we obtained ethics approvals from the College Human Ethics Advisory Network at RMIT University, then we run the experiments.

RMIT University

19 August 2014

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 7096
Fax +61 3 9925 6506
www.rmit.edu.au

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

Dear Jenny

ASEHAPP 47 – 13 ZHANG-SHARIFF Query-biased Credibility Ranking of Tweets

Thank you for requesting an amendment and extension to your Human Research Ethics project titled: ***Query-biased Credibility Ranking of Tweets*** which was originally approved by Science Engineering and Health CHEAN in September 2013 and extended to 13th June 2014. Thank you also for providing an annual report describing progress on the project.

I am pleased to inform you that the CHEAN has **approved** your extension and amendment as outlined in your request and your Human Research Ethics project is now approved until **13th December 2014**

The CHEAN notes and thanks you for providing all documentation that incorporates these amendments. This documentation will be appended to your file for future reference and your research may now continue.

The committee 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 Human Research Ethics Sub-Committee.

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 Supervisor/s:

Figure A.1: Ethics Approval for experiments in chapter 3.

RMIT UNIVERSITY

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 7096
Fax +61 3 9925 6506
www.rmit.edu.au

3rd July 2015

Mark Sanderson
Building 14 Level 9, Room 17
School of Computer Science &
Information Technology
RMIT University

Dear Mark

ASEHAPP 35-15 SANDERSON-ALADHADH Tweet Author Location Impacts on Tweet Credibility

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 **12 Months** from the date of this letter to **3rd July 2016** 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,

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

Cc CHEAN Member: Amanda Kimpton School of Health Sciences RMIT University
Student Investigator/s: Suliman Aladhadh School of Computer Science & IT RMIT University Other
Investigator/s: Jenny Zhang School of Computer Science & IT RMIT University

Figure A.2: Ethics Approval for experiments in chapter.4.

RMIT UNIVERSITY

معلومات المشارك

عزيزي المشارك,
انت مدعو للمشاركه في مشروع بحثي بواسطة جامعة ار ام اي تي. ارجو انت تقرأ هذه الورقه بتمعن وكن متأكد من فهما قبل ان تقرر المشاركة. اذا لديك اي استفسار عن المشروع تكرما تواصل مع أحد الباحثين.

من المشمول في هذا المشروع ؟

باحثين من جامعة ار ام اي تي يقومون بالبحث. وهذا البحث يقوم به الطالب سليمان العضاض كجزء بحثه لدكتوراه في علوم الحاسب.
هو تحت اشراف البرفسور مارك ساندرسون و د اي .جيني زهانج من ار ام اي تي.
لجنة أخلاقيات البحوث البشرية في ار ام اي تي وافقت على هذا المشروع البحثي.
اذا لديك اي اسأله, تفضل بالتواص مع سليمان العضاض على
Suliman.aladhadh@rmit.edu.au

ماهو هذا المشروع ؟

هذا البحث هو للمساعده في استخدام خصائص التغريدات للحصول على التغريدات والمستخدمين الموثوقين عند عمل بحث عن موضوع معين. بما انه تويتر أحد اصبح مصادر معلومات وخاصه ، في مايتعلق بالأخبار، من المهم التأكد من صحة المعلومات داخل التغريده. من هذه التجربه نحن نهدف إلى معرفة انواع المصداقيه في التغريدات وخصائصها؛ خصوصا كيف يفهمالناس المصداقية في الحكم على المعلومات المشاركه في وسائل التواصل الإجتماعي،خصوصا ان التغريده تحتوي على ١٤٠ حرف فقط.

اذا وافقت على المشاركة , ماذا سوف يتطلب علي عمله ؟

انت تحتاج لتحديد مستوى المصداقية للتغريدات الموجودة في استمارة البحث. التغريدات هي عشوائية من مجموعة تغريدات في موضوع معين، لكل تغريدة تحتاج ان تحدد مستوى مصداقيتها باستخدام مقياس 7 نقاط ليكرت (من غير موافق بشدة إلى موافق بشدة). بعد تقييم جميع التغريدات سوف تذهب لاستبيان قصير عن استخدامك لتويتر. اخيرا سوف يتم اخذ المعلومات الديموغرافية الاساسية وهي : مجموعة العمر , المستوى التعليمي , الجنس و لغتك الأولى.

ماهي المخاطر المحتملة او السلبيات ؟

تقييم مصداقية التغريدات ربما ليس ممتع لدرجه كبيره ولكن هذه هي المخاطر الموجودة. التغريدات مدروسه ولايوجد فيها اي محاذير او مخاطر ثقافية. اذا لم ترد ان تقيم اي تغريده لاي سبب فقط قم بتجاوزها لما بعدها. يمكنك التوقف في اي وقت.

ماهي الفوائد المتعلقة بالمشاركة ؟

ربما لايوجد فائده مباشره لك للمشاركة في هذه الدراسه. لكن البيانات المحصله من هذه الدراسة ربما تساعد للمشاركه بالمعرفة العامة لإنتاج نموذج يساعد مستخدمين وسائل التواصل الإجتماعي الوصول للمعلومات الصحيحة.

ماذا سوف يحدث للمعلومات التي سوف اقدمها ؟

عندما تكمل الاستبانه والتقرير يأخذ من منصة الاستبانة, فقط الباحثين سوف يصلون للبيانات. سوف تحفظ في خادم ار ام اي تي بحماية رقم سري. المعلومات المتعلقة بالبيانات ونتائج البيانات سوف تستخدم في نشر الأوراق العلمية و المأتمرات ورسالة الدكتوراه .
الرجع للنشر في موقع المعنيين رساله في مخزن ار ام اي تي وهي مفتوحه بشكل عام بالكتبه على الانترنت للأوراق البحثية. البيانات البحثية سوف تبقى محفوظة في ار ام اي تي ل 6 سنوات بعد النشر قبل ان تدمر. بينما الورقه النهائية والرسالة سوف تبقى على الانترنت.

بسبب طبيعة جمع البيانات, لن يكون هناك نموذج تفويض مكتوب منك. عةضا عن ذلك, نحن نعتبر انك اطيتنا تفويض ضمني بإكمالك مهمة تقييم مصداقية التغريدات.

أمان الموقع

المستخدمين يجب ان يكونوا حذرين من ان شبكة الانترنت العالمية هي شبكة عامة غير آمنه وهذا يزيد من اختمالية المخاطر من كشف عملايات المستخدمين , اعتراضها او تعديلها من طرف ثالث او البيانات التي يحملها المستخدم تحتوي على فيروسات او عيوب.

Figure A.3: Statement of plain language (Arabic) for experiments in chapter 4.

أمان البيانات

هذا المشروع سيستخدم موقع خارجي لانشاء و جمع وتحليل البيانات المجمعة في شكل استبيان. الموقع المستخدم هو
www.crowdflower.com
اذا وافقت على المشاركة في الاستبيان, الاستجابات التي تعطيها سوف يتم تخزينها في خادم المضيف المستخدمه من
CrowdFlower.
لن يجمع اي بيانات شخصية في الاستبيان لذلك سوف تجمع كبيانات. عند تكتمال جمع وتحليل البيانات, سوف ننقل البيانات الى خوادم ار ام اي تي حيث سوف تحفظ بأمان ل خمس (5) سنوات. البيانات في خادم كرودفلور سوف تمسح وتدمر.

ماهي حقوقي كمشارك ؟

- الحق في الانسحاب من المشاركة في اي وقت
- لك الحق في ان البيانات غير المعالجة تسحب وتتلف
- لك الحق في الأسئلة (بالإيميل) في اي وقت

من يجب ان اتواصل معه اذا لدي اي استفسار ؟

تفضل بالتواصل مع سليمان العضاض على
Suliman.aladhadh@rmit.edu.au

ماهي المسائل التي يجب ان انتبه لها قبل ان اقرر المشاركة من عدمها ؟

بسبب طبيعة جمع البيانات, لن يكون هناك نموذج تفويض مكتوب منك. عۊضا عن ذلك, نحن نعتبر انك اطيتنا تفويض ضمني بإكمالك مهمة تقييم مصداقية التغريدات.

تفضلوا بقبول فائق الاحترام,

Mark Sanderson (mark.sanderson@rmit.edu.au)
Xiuzhen Jenny Zhang (xiuzhen.zhang@rmit.edu.au)
Suliman Aladhadh (Suliman.aladhadh@rmit.edu.au)

If you have any concerns about your participation in this project, which you do not wish to discuss with the researchers, then you can contact the Ethics Officer, Research Integrity, Governance and Systems, RMIT University, GPO Box 2476V VIC 3001. Tel: (03) 9925 2251 or email human.ethics@rmit.edu.au

Figure A.3: Statement of plain language (Arabic) for experiments in chapter 4. (Continued)

INVITATION TO PARTICIPATE IN A RESEARCH PROJECT

PARTICIPANT INFORMATION

Project Title: *Location and culture Impact on Tweet Credibility*

Dear participant,

You are invited to participate in a research project being conducted by RMIT University. Please read this sheet carefully and be confident that you understand its contents before deciding whether to participate. *If you have any questions about the project, please ask one of the investigators.*

Who is involved in this research project?

Researchers at RMIT are conducting the project. This research is performed by Suliman Aladhadh as part of his Phd in Computer Science. He is under supervision of Prof. Mark Sanderson and Dr. Xiuzhen Jenny Zhang of RMIT. RMIT Human Research Ethics Committee has approved this research project.
If you have any questions, please contact Suliman Aladhadh on suliman.aladhadh@rmit.edu.au

What is the project about?

This research is to help use tweets features to predict credible tweets and authors for a query. As Twitter has become one of the sources of information especially regarding news, it is important to ensure the credibility of the information in the tweet messages. From this experiment, we aim to learn the credibility classification in tweet messages and their characteristics; particularly how people perceive credibility in judging information shared in social media, specifically tweets that only have 140 characters.

If I agree to participate, what will I be required to do?

You will need to determine the credibility level of relevant tweets posted in the questionnaire. The tweets are random relevant tweets on events related topic. For each tweet you will need to determine its credibility level by 7 points Likert scales (from strongly disagree to strongly agree). After completing rating of tweets you will go to a quick questionnaire about your normal Twitter using. Finally, basic demographic information will be collected: gender, age group, education level and first language. The estimated time to complete the survey is about 10 to 12 minutes.

What are the possible risks or disadvantages?
Judging credibility of tweets might be not that so exiting but that is the extent of the risks. The tweets have been screened and do not tackle culturally sensitive issues. If however, you prefer not to judge a particular tweet for any reason, just skip the tweet and move over the next one. You can stop the participation any time.

What are the benefits associated with participation?
There may be no direct benefits to you for participating in this study. However, the data collected in the study may help to contribute to public knowledge of proposing model to help social media users find credible information.

What will happen to the information I provide?

Once the questionnaires have been completed and the report generated from the questionnaire platform, only the researchers will have access to the data. It will be kept on the computer server in RMIT,

Figure A.4: Statement of plain language (English) for experiments in chapter 4.

password protected. Information regarding the data collection and the result of the data will be used for paper publication, conference and the PhD student's thesis. A reference to publication on their respective website or thesis in the RMIT Repository is publicly accessible online library of research papers. The research data will be kept securely at RMIT for 5 years after publication, before being destroyed. Whereas the final research paper / thesis will remain online.

Because of the nature of data collection, we are not obtaining written informed consent from you. Instead, we assume that you have given implied consent by your completion of the tweets credibility judging tasks given to you.

Security of the website

Users should be aware that the World Wide Web is an insecure public network that gives rise to the potential risk that a user's transactions are being viewed, intercepted or modified by third parties or that data which the user downloads may contain computer viruses or other defects.

Security of the data

This project will use an external site to create, collect and analyse data collected in a survey format. The site we are using is www.crowdflower.com. If you agree to participate in this survey, the responses you provide to the survey will be stored on a host server that is used by CrowdFlower. No personal information will be collected in the survey so none will be stored as data. Once we have completed our data collection and analysis, we will import the data we collect to the RMIT server where it will be stored securely for five (5) years. The data on the CrowdFlower host server will then be deleted and expunged.

What are my rights as a participant?

- The right to withdraw from participation at any time.
- You have the right to have any unprocessed data withdrawn and destroyed.
- You have the right to ask questions (via email) at any time.

Whom should I contact if I have any questions?
Please contact Mark Sanderson (mark.sanderson@rmit.edu.au)

What other issues should I be aware of before deciding whether to participate?
Because of the nature of data collection, we are not obtaining written informed consent from you. Instead, we assume that you have given consent by your completion of the tweets credibility judging tasks given to you.

Yours sincerely,

Prof. Mark Sanderson (mark.sanderson@rmit.edu.au)
Dr. Xiuzhen Jenny Zhang (xiuzhen.zhang@rmit.edu.au)
Suliman Aladhadh (suliman.aladhadh@rmit.edu.au)

If you have any concerns about your participation in this project, which you do not wish to discuss with the researchers, then you can contact the Ethics Officer, Research Integrity, Governance and Systems, RMIT University, GPO Box 2476V VIC 3001. Tel: (03) 9925 2251 or email human.ethics@rmit.edu.au

Figure A.4: Statement of plain language (English) for experiments in chapter 4. (Continued)

College Human Ethics Advisory Network (CHEAN)
College of Science, Engineering and Health

Email: seh-human-ethics@rmit.edu.au
Phone: [61 3] 9925 4620
Building 91, Level 2, City Campus/Building 215, Level 2, Bundoora West Campus

27 March 2017

Professor Mark Sanderson
School of Science
RMIT University

Dear Prof Sanderson

ASEHAPP 35-15 Tweet Author Impacts on Tweet Credibility

Thank you for requesting an amendment to your Human Research Ethics project titled: **Tweet Author Impacts on Tweet Credibility**, which was originally approved by Science Engineering and Health CHEAN in 2015 for a period of **12 months**.

I am pleased to inform you that the CHEAN has **approved** your amendment as outlined in your request. Your extension has also been approved until 2 January 2018.

The CHEAN notes and thanks you for providing all documentation that incorporates these amendments. This documentation will be appended to your file for future reference and your research may now continue.

The committee 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.

Figure A.5: Ethics Approval for experiments in chapter 5

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 Other Investigator/s: Suliman Aladhadh, School of Science Jenny Zhang School of Science

RMIT University Page 2 of 2

Figure A.5: Ethics Approval for experiments in chapter 5. (Continued)

RMIT UNIVERSITY

INVITATION TO PARTICIPATE IN A RESEARCH PROJECT

PARTICIPANT INFORMATION

Project Title: *Location and culture Impact on Tweet Credibility*

Dear participant,

You are invited to participate in a research project being conducted by RMIT University. Please read this sheet carefully and be confident that you understand its contents before deciding whether to participate. *If you have any questions about the project, please ask one of the investigators.*

Who is involved in this research project?

Researchers at RMIT are conducting the project. This research is performed by Suliman Aladhadh as part of his Phd in Computer Science. He is under supervision of Prof. Mark Sanderson and Dr. Xiuzhen Jenny Zhang of RMIT.
RMIT Human Research Ethics Committee has approved this research project.
If you have any questions, please contact Suliman Aladhadh on suliman.aladhadh@rmit.edu.au

What is the project about?

This research is to help use tweets features to predict credible tweets and authors for a query. As Twitter has become one of the sources of information especially regarding news, it is important to ensure the credibility of the information in the tweet messages. From this experiment, we aim to learn the credibility classification in tweet messages and their characteristics; particularly how people perceive credibility in judging information shared in social media, specifically tweets that only have 140 characters.

If I agree to participate, what will I be required to do?

You will need to determine the credibility, informativeness and sources of tweets. The tweets will we show you are random relevant tweets on events related topics. There is no time frame for this process because it depends on how many judgments you want to complete, up to a limit of 300 judgments.

What are the possible risks or disadvantages?

Judging credibility of tweets might be not that so exiting but that is the extent of the risks. The tweets have been screened and do not tackle culturally sensitive issues. If however, you prefer not to judge a particular tweet for any reason, just skip the tweet and move over the next one. You can stop the participation any time.

What are the benefits associated with participation?

There may be no direct benefits to you for participating in this study. However, the data collected in the study may help to contribute to public knowledge of proposing model to help social media users find credible information.

What will happen to the information I provide?

Once the questionnaires have been completed and the report generated from the questionnaire platform, only the researchers will have access to the data. It will be kept on the computer server in RMIT, password protected. Information regarding the data collection and the result of the data will be used for

Figure A.6: Statement of plain language for experiments in chapter 5.

paper publication, conference and the PhD student's thesis. A reference to publication on their respective website or thesis in the RMIT Repository is publicly accessible online library of research papers. The research data will be kept securely at RMIT for 5 years after publication, before being destroyed. Whereas the final research paper / thesis will remain online.

Because of the nature of data collection, we are not obtaining written informed consent from you. Instead, we assume that you have given implied consent by your completion of the tweets credibility judging tasks given to you.

Security of the website

Users should be aware that the World Wide Web is an insecure public network that gives rise to the potential risk that a user's transactions are being viewed, intercepted or modified by third parties or that data which the user downloads may contain computer viruses or other defects.

Security of the data

This project will use an external site to create, collect and analyse data collected in a survey format. The site we are using is www.crowdflower.com. If you agree to participate in this survey, the responses you provide to the survey will be stored on a host server that is used by CrowdFlower. No personal information will be collected in the survey so none will be stored as data. Once we have completed our data collection and analysis, we will import the data we collect to the RMIT server where it will be stored securely for five (5) years. The data on the CrowdFlower host server will then be deleted and expunged.

What are my rights as a participant?

- The right to withdraw from participation at any time.
- You have the right to have any unprocessed data withdrawn and destroyed.
- You have the right to ask questions (via email) at any time.

Whom should I contact if I have any questions?
Please contact Mark Sanderson (mark.sanderson@rmit.edu.au)

What other issues should I be aware of before deciding whether to participate?
Because of the nature of data collection, we are not obtaining written informed consent from you. Instead, we assume that you have given consent by your completion of the tweets credibility judging tasks given to you.

Yours sincerely,

Prof. Mark Sanderson (mark.sanderson@rmit.edu.au)
Dr. Xiuzhen Jenny Zhang (xiuzhen.zhang@rmit.edu.au)
Suliman Aladhadh (suliman.aladhadh@rmit.edu.au)

If you have any concerns about your participation in this project, which you do not wish to discuss with the researchers, then you can contact the Ethics Officer, Research Integrity, Governance and Systems, RMIT University, GPO Box 2476V VIC 3001. Tel: (03) 9925 2251 or email human.ethics@rmit.edu.au

Figure A.6: Statement of plain language for experiments in chapter 5. (Continued)

Appendix B

Glossary

UGC: User Generated Content

WEKA: Waikato Environment for Knowledge Analysis

NLP: Natural Language Processing

TF-IDF: Term Frequency-Inverse Document Frequency

POS: Part of Speech

SVM: Support Vector Machine classifier

PRF: Pseudo Relevance Feedback

NCDG: Normalized Discounted Cumulative Gain

LIWC: Linguistic Inquiry and Word Count

ANOVA: Analysis of Variance

MANOVA: Multivariate analysis of variance

Bibliography

A. Acar and A. Deguchi. Culture and social media usage: Analysis of japanese twitter users. *International Journal of Electronic Commerce Studies*, 4(1):21, 2013.

A. Aker, A. Zubiaga, K. Bontcheva, A. Kolliakou, R. Procter, and M. Liakata. Stance classification in out-of-domain rumours: A case study around mental health disorders. In *Proceedings of the 9th International Conference on Social Informatics*, pages 53–64. Springer, 2017.

R. M. B. Al-Eidan, H. S. Al-Khalifa, and A. S. Al-Salman. Measuring the credibility of arabic text content in twitter. In *Fifth IEEE International Conference on Digital Information Management*, pages 285–291. IEEE, 2010.

H. S. Al-Khalifa and R. M. B. Al-Eidan. An experimental system for measuring the credibility of news content in twitter. *International Journal of Web Information Systems*, 7 (2):130–151, 2011.

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

A. A. AlMansour and C. S. Iliopoulos. Using arabic microblogs features in determining credibility. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 1212–1219. ACM, 2015.

A. A. AlMansour, L. Brankovic, and C. S. Iliopoulos. A model for recalibrating credibility in different contexts and languages-a twitter case study. *International Journal of Digital Information and Wireless Communications (IJDIWC)*, 4(1):53–62, 2014.

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

R. A. Baeza-Yates and D. Sáez-Trumper. Wisdom of the crowd or wisdom of a few?: An analysis of users' content generation. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, pages 69–74. ACM, 2015.

R. Balakrishnan and S. Kambhampati. Sourcerank: relevance and trust assessment for deep web sources based on inter-source agreement. In *Proceedings of the 20th International Conference on World Wide Web*, pages 227–236. ACM, 2011.

M. Bastian, M. Hayes, W. Vaughan, S. Shah, P. Skomoroch, H. Kim, S. Uryasev, and C. Lloyd. Linkedin skills: large-scale topic extraction and inference. In *Eighth Conference on Recommender Systems*, pages 1–8. ACM, 2014.

F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida. Detecting spammers on twitter. In *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)*, number 2010, page 12. CEAS Conference, 2010.

P. Bhattacharya, S. Ghosh, J. Kulshrestha, M. Mondal, M. B. Zafar, N. Ganguly, and K. P. Gummadi. Deep twitter diving: Exploring topical groups in microblogs at scale. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 197–210. ACM, 2014.

C. Boididou, S. Papadopoulos, Y. Kompatsiaris, S. Schifferes, and N. Newman. Challenges of computational verification in social multimedia. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 743–748. ACM, 2014.

C. Boididou, S. Papadopoulos, D. Dang-Nguyen, G. Boato, M. Riegler, S. E. Middleton, A. Petlund, and Y. Kompatsiaris. Verifying multimedia use at mediaeval 2016. In *Working Notes Proceedings of the MediaEval Workshop*. CEUR-WS.org, 2016.

A. Bovet and H. A. Makse. Influence of fake news in twitter during the 2016 us presidential election. *Nature communications*, 10(1):7, 2019.

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

J. Carletta. Assessing agreement on classification tasks: the kappa statistic. *Computational linguistics*, 22(2):249–254, 1996.

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

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

S. Chang and C. Gomes. Digital journeys: A perspective on understanding the digital experiences of international students. *Journal of International Students*, 7(2):347–366, 2017.

Z. Cheng, J. Caverlee, and K. Lee. You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 759–768. ACM, 2010.

Z. Cheng, J. Caverlee, H. Barthwal, and V. Bachani. Who is the barbecue king of texas?: a geo-spatial approach to finding local experts on twitter. In *The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 335–344. ACM, 2014a.

Z. Cheng, J. Caverlee, H. Barthwal, and V. Bachani. Finding local experts on twitter. In *23rd International World Wide Web Conference, Companion Volume*, pages 241–242. ACM, 2014b.

J. Cohen. *Statistical power analysis for the behavioral sciences 2nd edn.* Erlbaum Associates, Hillsdale, 1988.

J. Cohen. A power primer. *Psychological bulletin*, 112(1):155, 1992.

U. D. o. Commerce. Census regions and divisions of the United States, 2015. URL https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us.

S. Counts and K. Fisher. Taking it all in? visual attention in microblog consumption. In *Proceedings of the Fifth International Conference on Weblogs and Social Media*, pages 97–104. AAAI Press, 2011.

D. Dailey and K. Starbird. Journalists as crowdsourcerers: Responding to crisis by reporting with a crowd. *Computer Supported Cooperative Work*, 23(4-6):445–481, 2014.

D. R. Danielson. Web credibility. In *Encyclopedia of human computer interaction*, pages 713–721. IGI Global, 2006.

M. De Choudhury, N. Diakopoulos, and M. Naaman. Unfolding the event landscape on twitter: classification and exploration of user categories. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 241–244. ACM, 2012.

E. Dedoussis. A cross-cultural comparison of organizational culture: evidence from universities in the arab world and japan. *Cross Cultural Management: An International Journal*, 11(1):15–34, 2004.

N. Diakopoulos, M. De Choudhury, and M. Naaman. Finding and assessing social media information sources in the context of journalism. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2451–2460. ACM, 2012.

G. Eysenbach. Credibility of health information and digital media: New perspectives and implications for youth. *Digital media, youth, and credibility*, pages 123–154, 2008.

G. Eysenbach and C. Köhler. 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. *British Medical Journal*, 324(7337):573–577, 2002.

A. Field. *Discovering statistics using IBM SPSS statistics.* sage, 2013.

A. J. Flanagin and M. J. Metzger. The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society*, 9(2):319–342, 2007.

A. J. Flanagin and M. J. Metzger. Digital media and youth: Unparalleled opportunity and unprecedented responsibility. *Digital media, youth, and credibility*, pages 5–27, 2008.

B. Fogg, J. Marshall, O. Laraki, A. Osipovich, C. Varma, N. Fang, J. Paul, A. Rangnekar, J. Shon, P. Swani, et al. 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, 2001.

B. J. Fogg. Prominence-interpretation theory: explaining how people assess credibility online. In *Extended abstracts of the 2003 Conference on Human Factors in Computing Systems*, pages 722–723. ACM, 2003.

B. J. Fogg and H. Tseng. The elements of computer credibility. In *Proceeding of the CHI '99 Conference on Human Factors in Computing Systems*, pages 80–87. ACM, 1999.

M. Foucault. *Religion and culture.* Routledge, 2013.

J. W. Fritch and R. L. Cromwell. Evaluating internet resources: Identity, affiliation, and cognitive authority in a networked world. *JASIST*, 52(6):499–507, 2001.

U. Gadiraju, G. Demartini, R. Kawase, and S. Dietze. Human beyond the machine: Challenges and opportunities of microtask crowdsourcing. *IEEE Intelligent Systems*, 30(4): 81–85, 2015.

Q. Gao, Y. Tian, and M. Tu. Exploring factors influencing chinese user's perceived credibility of health and safety information on weibo. *Computers in Human Behavior*, 45:21–31, 2015.

R. O. G. Gavilanes, D. Quercia, and A. Jaimes. Cultural dimensions in twitter: Time, individualism and power. In *Proceedings of the Seventh International Conference on Weblogs and Social Media*. The AAAI Press, 2013.

S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi. Cognos: crowdsourcing search for topic experts in microblogs. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 575–590. ACM, 2012.

S. Ghosh, M. B. Zafar, P. Bhattacharya, N. K. Sharma, N. Ganguly, and P. K. Gummadi. On sampling the wisdom of crowds: random vs. expert sampling of the twitter stream. In *22nd ACM International Conference on Information and Knowledge Management*, pages 1739–1744. ACM, 2013.

J. D. Gibbons and S. Chakraborti. Nonparametric statistical inference. In *International encyclopedia of statistical science*, pages 977–979. Springer, 2011.

J. Golbeck. Detecting coping style from twitter. In *International Conference on Social Informatics*, pages 454–467. Springer, 2016.

C. Gomes. Digital media, diversity and the physical world. In *Siloed Diversity*, pages 41–63. Springer, 2018.

Google Social Search. Official Blog, 2011. URL http://bit.ly/2tm4LXJ.

B. Y. J. Gottfried and E. Shearer. News Use Across Social Media Platforms 2016. Technical report, Pew Research Center, 2016. URL http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/.

A. Gupta and P. Kumaraguru. Credibility ranking of tweets during high impact events. In *Proceedings of the 1st workshop on privacy and security in online social media*, page 2. ACM, 2012.

A. Gupta, K. P. Sycara, G. J. Gordon, and A. Hefny. Exploring friend's influence in cultures in twitter. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 584–591. ACm, 2013.

A. Gupta, P. Kumaraguru, C. Castillo, and P. Meier. Tweetcred: Real-time credibility assessment of content on twitter. In *International Conference on Social Informatics*, pages 228–243. Springer, 2014.

Z. Gyöngyi, H. Garcia-Molina, and J. O. Pedersen. Combating web spam with trustrank. In *(e)Proceedings of the Thirtieth International Conference on Very Large Data Bases*, pages 576–587. VLDB Endowment, 2004.

B. Han, P. Cook, and T. Baldwin. Geolocation prediction in social media data by finding location indicative words. In *COLING 2012, 24th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers*, pages 1045–1062. Indian Institute of Technology Bombay, 2012.

B. Han, P. Cook, and T. Baldwin. Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research*, 49:451–500, 2014.

E. Hargittai, L. Fullerton, E. Menchen-Trevino, and K. Y. Thomas. Trust online: Young adults' evaluation of web content. *International journal of communication*, 4:27, 2010.

B. Hecht, L. Hong, B. Suh, and E. H. Chi. Tweets from justin bieber's heart: the dynamics of the location field in user profiles. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 237–246. ACM, 2011.

B. R. Heravi and J. McGinnis. Introducing social semantic journalism. *The Journal of Media Innovations*, 2(1):131–140, 2015.

B. Hilligoss and S. Y. Rieh. Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Inf. Process. Manage.*, 44(4):1467–1484, 2008.

G. Hofstede. *Cultures and Organizations: Software of the mind.* McGraw-Hill, 1991.

G. Hofstede. Dimensionalizing cultures: The hofstede model in context. *Online readings in psychology and culture*, 2(1):8, 2011.

R. M. Hogarth. *Judgement and choice: The psychology of decision.* Wiley, 1987.

C. Honeycutt and S. C. Herring. Beyond microblogging: Conversation and collaboration via twitter. In *42st Hawaii International International Conference on Systems Science*, pages 1–10. IEEE Computer Society, 2009.

L. Hong, G. Convertino, and E. H. Chi. Language matters in twitter: A large scale study. In *Proceedings of the Fifth International Conference on Weblogs and Social Media.* The AAAI Press, 2011.

B. D. Horne and S. Adali. This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *The 2nd International Workshop on News and Public Opinion at ICWSM*, pages 591–600. AAAI Press, 2017.

B. D. Horne, D. Nevo, J. Freitas, H. Ji, and S. Adali. Expertise in social networks: How do experts differ from other users? In *Proceedings of the Tenth International Conference on Web and Social Media*, pages 583–586. AAAI Press, 2016.

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

C. I. Hovland, I. L. Janis, and H. H. Kelley. *Communication and persuasion; psychological studies of opinion change.* Yale University Press, 1953.

S. L. Humpherys, K. C. Moffitt, M. B. Burns, J. K. Burgoon, and W. F. Felix. Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, 50(3):585–594, 2011.

M. Imran and C. Castillo. Towards a data-driven approach to identify crisis-related topics in social media streams. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1205–1210. ACM, 2015.

M. Imran, S. Elbassuoni, C. Castillo, F. Diaz, and P. Meier. Extracting information nuggets from disaster-related messages in social media. In *10th Proceedings of the International Conference on Information Systems for Crisis Response and Management.* ISCRAM Association, 2013.

B. Ionescu, A. Gînsca, M. Zaharieva, B. Boteanu, M. Lupu, and H. Müller. Retrieving diverse social images at mediaeval 2016: Challenge, dataset and evaluation. In *Working Notes Proceedings of the MediaEval 2016 Workshop.* CEUR-WS.org, 2016.

Y. G. Ji, H. Hwangbo, J. S. Yi, P. P. Rau, X. Fang, and C. Ling. The influence of cultural differences on the use of social network services and the formation of social capital. *International Journal of Human–Computer Interaction*, 26(11&12):1100–1121, 2010.

W. Jiang. The relationship between culture and language. *English Language Teaching (ELT) journal*, 54(4):328–334, 2000.

A. Juffinger, M. Granitzer, and E. Lex. Blog credibility ranking by exploiting verified content. In *Proceedings of the 3rd ACM Workshop on Information Credibility on the Web*, pages 51–58. ACM, 2009.

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

B. Kang, T. Höllerer, and J. O'Donovan. 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, 2015.

A. M. Kaplan and M. Haenlein. Users of the world, unite! the challenges and opportunities of social media. *Business horizons*, 53(1):59–68, 2010.

S. Kitayama, H. Park, A. T. Sevincer, M. Karasawa, and A. K. Uskul. A cultural task analysis of implicit independence: comparing north america, western europe, and east asia. *Journal of personality and social psychology*, 97(2):236, 2009.

A. Klamer. *The value of culture: On the relationship between economics and arts.* Amsterdam University Press, 1997.

J. Kulshrestha, M. B. Zafar, L. E. Noboa, K. P. Gummadi, and S. Ghosh. Characterizing information diets of social media users. In *Proceedings of the Ninth International Conference on Web and Social Media*, pages 218–227. AAAI Press, 2015.

S. Kumar, F. Morstatter, R. Zafarani, and H. Liu. Whom should i follow?: identifying relevant users during crises. In *Proceedings of the 24th ACM conference on hypertext and social media*, pages 139–147. ACM, 2013.

S. Kumar, X. Hu, and H. Liu. A behavior analytics approach to identifying tweets from crisis regions. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pages 255–260. ACM, 2014.

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

S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang. Prominent features of rumor propagation in online social media. In *13th International Conference on Data Mining*, pages 1103–1108. IEEE, 2013.

A. S. KÃijmpel, V. Karnowski, and T. Keyling. News sharing in social media: A review of current research on news sharing users, content, and networks. *Social Media + Society*, 1 (2):1–14, 2015.

R. D. Lankes. Credibility on the internet: shifting from authority to reliability. *Journal of Documentation*, 64(5):667–686, 2008.

R. Lederman, H. Fan, S. Smith, and S. Chang. Who can you trust? credibility assessment in online health forums. *Health Policy and Technology*, 3(1):13–25, 2014.

J. Lin. Divergence measures based on the shannon entropy. *IEEE Trans. Information Theory*, 37(1):145–151, 1991.

X. Liu, A. Nourbakhsh, Q. Li, R. Fang, and S. Shah. Real-time rumor debunking on twitter. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, pages 1867–1870. ACM, 2015.

J. Mahmud, J. Nichols, and C. Drews. Where is this tweet from? inferring home locations of twitter users. In *Proceedings of the Sixth International Conference on Weblogs and Social Media*. AAAI Press, 2012.

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

M. J. Metzger. Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. *JASIST*, 58(13):2078–2091, 2007.

M. J. Metzger and A. J. Flanagin. Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, 59:210–220, 2013.

M. J. Metzger, A. J. Flanagin, K. Eyal, D. R. Lemus, and R. M. McCann. Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Annals of the International Communication Association*, 27(1):293–335, 2003.

P. Meyer. Defining and measuring credibility of newspapers: Developing an index. *Journalism quarterly*, 65(3):567–574, 1988.

A. Mislove, S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, and J. N. Rosenquist. Understanding the demographics of twitter users. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, page 5th. AAAI Press, 2011.

T. Mitra and E. Gilbert. CREDBANK: A large-scale social media corpus with associated credibility annotations. In *Proceedings of the Ninth International Conference on Web and Social Media*, pages 258–267. AAAI Press, 2015.

T. Mitra, G. P. Wright, and E. Gilbert. Credibility and the dynamics of collective attention. *PACM on Human-Computer Interaction*, 1(CSCW):80:1–80:17, 2017a.

T. Mitra, G. P. Wright, and E. Gilbert. A parsimonious language model of social media credibility across disparate events. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 126–145. ACM, 2017b.

D. Mocanu, A. Baronchelli, N. Perra, B. Gonçalves, Q. Zhang, and A. Vespignani. The twitter of babel: Mapping world languages through microblogging platforms. *PloS one*, 8 (4):e61981, 2013.

M.-F. Moens, J. Li, and T.-S. Chua. *Mining User Generated Content*. Chapman & Hall/CRC, 2014.

M. R. Morris, J. Teevan, and K. Panovich. A comparison of information seeking using search engines and socialănetworks. In *Proceedings of the Fourth International Conference on Weblogs and Social Media*, 2010.

M. R. Morris, S. Counts, A. Roseway, A. Hoff, and J. Schwarz. Tweeting is believing?: understanding microblog credibility perceptions. In *CSCW '12 Computer Supported Cooperative Work*, pages 441–450. ACM, 2012.

F. Morstatter, J. Pfeffer, H. Liu, and K. M. Carley. Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. In *Proceedings of the Seventh International ICWSM Conference on Weblogs and Social Media*. AAAI Press, 2013.

F. Morstatter, N. Lubold, H. Pon-Barry, J. Pfeffer, and H. Liu. Finding eyewitness tweets during crises. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 23–27. Association for Computational Linguistics, 2014.

D. Murthy. *Twitter: Social communication in the Twitter age.* John Wiley & Sons, 2013.

M. Navarra. Twitter is verifying way more accounts, and here?s the data to prove it, 2016. URL https://thenextweb.com/twitter/2016/07/22/twitter-verification-rises/.

J. Newhagen and C. Nass. Differential criteria for evaluating credibility of newspapers and tv news. *Journalism Quarterly*, 66(2):277–284, 1989.

D.-P. Nguyen, R. Gravel, R. B. Trieschnigg, and T. Meder. "how old do you think i am?" a study of language and age in twitter. In *Proceedings of the Seventh International Conference on Weblogs and Social Media*. AAAI Press, 2013.

B. Obeidat, R. Shannak, R. Masa'deh, and I. Al-Jarrah. Toward Better Understanding for Arabian Culture : Implications based on Hofstede ' s Cultural Model Toward Better Understanding for Arabian Culture : Implications Based on Hofstede ' s Cultural Model. *European Journal of Social Sciences*, 28(4):512–522, 2012.

J. O'Donovan, B. Kang, G. Meyer, T. Höllerer, and S. Adali. Credibility in context: An analysis of feature distributions in twitter. In *2012 International Conference on Privacy,* *Security, Risk and Trust, PASSAT 2012, and 2012 International Conference on Social Computing, SocialCom*, pages 293–301. IEEE, 2012.

A. Olteanu, S. Vieweg, and C. Castillo. What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 994–1009. ACM, 2015.

A. Pal and S. Counts. Identifying topical authorities in microblogs. In *Proceedings of the Forth International Conference on Web Search and Web Data Mining*, pages 45–54. ACM, 2011a.

A. Pal and S. Counts. What's in a@ name? how name value biases judgment of microblog authors. In *Proceedings of the Fifth International Conference on Weblogs and Social Media*. AAAI Press, 2011b.

S. Papadopoulos, K. Bontcheva, E. Jaho, M. Lupu, and C. Castillo. Overview of the special issue on trust and veracity of information in social media. *Transactions on Information Systems*, 34(3):14:1–14:5, 2016.

R. E. Petty and J. T. Cacioppo. *The elaboration likelihood model of persuasion*, pages 1–24. Springer, 1986.

B. Poblete, R. Garcia, M. Mendoza, and A. Jaimes. Do all birds tweet the same?: characterizing twitter around the world. In *Proceedings of the 20th ACM CIKM international conference on Information and knowledge management*, pages 1025–1030. ACM, 2011.

K. Popat, S. Mukherjee, J. Strötgen, and G. Weikum. Credibility assessment of textual claims on the web. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*, pages 2173–2178, 2016.

K. Popat, S. Mukherjee, J. Strötgen, and G. Weikum. Where the truth lies: Explaining the credibility of emerging claimsăon the web and social media. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 1003–1012, 2017.

S. Ravikumar, R. Balakrishnan, and S. Kambhampati. Ranking tweets considering trust and relevance. In *Proceedings of the Ninth International Workshop on Information Integration on the Web*, page 4. ACM, 2012.

S. Y. Rieh. Judgment of information quality and cognitive authority in the web. *JASIST*, 53(2):145–161, 2002.

S. Y. Rieh. Credibility and cognitive authority of information. *Encyclopedia of Library and Information Sciences, 3rd Ed*, pages 1337–1344, 2010.

S. Y. Rieh and D. R. Danielson. Credibility: A multidisciplinary framework. *ARIST*, 41(1): 307–364, 2007.

S. Y. Rieh, M. R. Morris, M. J. Metzger, H. Francke, and G. Y. Jeon. Credibility perceptions of content contributors and consumers in social media. *Proceedings of the Association for Information Science and Technology*, 51(1):1–4, 2014.

B. W. Roper. *Public attitudes toward television and other media in a time of change: The fourteenth report in a series.* Television Information Office, 1985.

P. Rosso and L. C. Cagnina. Deception detection and opinion spam. In *A Practical Guide to Sentiment Analysis*, pages 155–171. Springer, 2017.

T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World Wide Web*, pages 851–860. ACM, 2010.

M. Schmierbach and A. Oeldorf-Hirsch. A little bird told me, so i didn't believe it: Twitter, credibility, and issue perceptions. *Communication Quarterly*, 60(3):317–337, 2012.

A. Schulz, A. Hadjakos, H. Paulheim, J. Nachtwey, and M. Mühlhäuser. A multi-indicator approach for geolocalization of tweets. In *Proceedings of the Seventh International Conference on Weblogs and Social Media*. AAAI Press, 2013.

J. Schwarz and M. R. Morris. Augmenting web pages and search results to support credibility assessment. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 1245–1254, 2011.

L. D. Setlock and S. R. Fussell. What's it worth to you?: the costs and affordances of CMC tools to asian and american users. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, pages 341–350. ACM, 2010.

W. Seto and F. Martin. Transmigrant media: Mediating place, mobility, and subjectivity. *International Journal of Cultural Studies*, page 1367877918812470, 2018.

S. M. Shariff, X. Zhang, and M. Sanderson. User perception of information credibility of news on twitter. In *Advances in Information Retrieval - 36th European Conference on IR Research*, pages 513–518. Springer, 2014.

B. Sharifi, M. Hutton, and J. K. Kalita. Experiments in microblog summarization. In *Proceedings of the 2010 IEEE Second International Conference on Social Computing, SocialCom / IEEE International Conference on Privacy, Security, Risk and Trust, PASSAT 2010*, pages 49–56, 2010.

N. K. Sharma, S. Ghosh, F. Benevenuto, N. Ganguly, and P. K. Gummadi. Inferring who-is-who in the twitter social network. *Computer Communication Review*, 42(4):533–538, 2012.

M. Siegler. At 1.6 Billion Queries Per Day, Twitter Finally Aims To Make Search Personally Relevant, 2011. URL https://techcrunch.com/2011/06/01/new-twitter-search-relevant/.

S. Sikdar, B. Kang, J. O'Donovan, T. Höllerer, and S. Adali. Understanding information credibility on twitter. In *International Conference on Social Computing, SocialCom 2013*, pages 19–24, 2013.

K. Starbird and L. Palen. *Pass it on?: Retweeting in mass emergency*. International Community on Information Systems for Crisis Response and Management, 2010.

K. Starbird, L. Palen, A. L. Hughes, and S. Vieweg. Chatter on the red: what hazards threat reveals about the social life of microblogged information. In *Proceedings of the ACM conference on Computer supported cooperative work*, pages 241–250. ACM, 2010.

S. S. Sundar. The main model: A heuristic approach to understanding technology effects on credibility. *Digital media, youth, and credibility*, 73-100, 2008.

H. Tanev, V. Zavarella, and J. Steinberger. Monitoring disaster impact: detecting micro-events and eyewitness reports in mainstream and social media. In *Proceedings of the 14th ISCRAM Conference*, pages 592–602. ISCRAM Association, 2017.

D. Taraborelli. How the web is changing the way we trust. *Current issues in computing and philosophy*, pages 194–204, 2008.

Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1): 24–54, 2010.

J. Teevan, D. Ramage, and M. R. Morris. #twittersearch: a comparison of microblog search and web search. In *Proceedings of the Forth International Conference on Web Search and Web Data Mining*, pages 35–44. AAAI Press, 2011.

R. Thomson, N. Ito, H. Suda, F. Lin, Y. Liu, R. Hayasaka, R. Isochi, and Z. Wang. Trusting tweets: The fukushima disaster and information source credibility on twitter. In *Proceedings of the 9th International ISCRAM Conference*, pages 1–10. Vancouver: Simon Fraser University, 2012.

M. Truelove, M. Vasardani, and S. Winter. Towards credibility of micro-blogs: characterising witness accounts. *GeoJournal*, 80(3):339–359, 2015.

M. Truelove, M. Vasardani, and S. Winter. Introducing a framework for automatically differentiating witness accounts of events from social media. *R@ Loc 2016*, page 13, 2016.

I. Twitter. TWITTER USAGE / COMPANY FACTS, 2016. URL https://about.twitter.com/company.

S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1079–1088. ACM, 2010.

S. Vosoughi, D. Roy, and S. Aral. The spread of true and false news online. *Science*, 359 (6380):1146–1151, 2018.

C. Wagner, V. Liao, P. Pirolli, L. Nelson, and M. Strohmaier. It's not in their tweets: Modeling topical expertise of twitter users. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*, pages 91–100. IEEE, 2012.

F. Warneken and M. Tomasello. The roots of human altruism. *British Journal of Psychology*, 100(3):455–471, 2009.

A. Wawer, R. Nielek, and A. Wierzbicki. Predicting webpage credibility using linguistic features. In *23rd International World Wide Web Conference, Companion Volume*, pages 1135–1140. ACM, 2014.

W. Weerkamp, S. Carter, and M. Tsagkias. How people use twitter in different languages. In *Proceedings of the ACM Web Science*, number 1, page 1. ACM, 2011.

X. Wei and D. Stillwell. How smart does your profile image look?: Estimating intelligence from social network profile images. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, WSDM '17, pages 33–40. ACM, 2017.

D. Westerman, P. R. Spence, and B. V. D. Heide. 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, 2012.

E. J. Wilson and D. L. Sherrell. Source effects in communication and persuasion research: A meta-analysis of effect size. *Journal of the Academy of Marketing Science*, 21(2):101, 1993.

M. E. Wilson. Arabic speakers: Language and culture, here and abroad. *Topics in Language Disorders*, 16(4):65–80, 1996.

S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts. Who says what to whom on twitter. In *Proceedings of the 20th International Conference on World Wide Web*, pages 705–714, 2011.

X. Xia, X. Yang, C. Wu, S. Li, and L. Bao. Information credibility on twitter in emergency situation. In *Intelligence and Security Informatics - Pacific Asia Workshop*, pages 45–59, 2012.

F. Yang, Y. Liu, X. Yu, and M. Yang. Automatic detection of rumor on sina weibo. In *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, page 13. ACM, 2012.

J. Yang, M. R. Morris, J. Teevan, L. A. Adamic, and M. S. Ackerman. Culture matters: A survey study of social q&a behavior. In *Proceedings of the Fifth International Conference on Weblogs and Social Media*. The AAAI Press, 2011.

J. Yang, S. Counts, M. R. Morris, and A. Hoff. Microblog credibility perceptions: comparing the usa and china. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 575–586. ACM, 2013.

M. B. Zafar, P. Bhattacharya, N. Ganguly, S. Ghosh, and K. P. Gummadi. On the wisdom of experts vs. crowds: Discovering trustworthy topical news in microblogs. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 437–450. ACM, 2016.

L. Zeng, K. Starbird, and E. S. Spiro. # unconfirmed: Classifying rumor stance in crisis-related social media messages. In *Tenth International Conference on Web and Social Media*. AAAI Press, 2016.

Z. Zhao, P. Resnick, and Q. Mei. Enquiring minds: Early detection of rumors in social media from enquiry posts. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1395–1405. ACM, 2015.

A. Zubiaga and H. Ji. Tweet, but verify: epistemic study of information verification on twitter. *Social Network Analysis and Mining*, 4(1):163, 2014.