A Network Analysis Based Credibility Ranking Model to Combat Misinformation on Twitter

T.H.M De Silva

2020

A Network Analysis Based Credibility Ranking Model to Combat Misinformation on Twitter

A dissertation submitted for the Degree of Master of Science in Computer Science

T.H.M De Silva
University of Colombo School of Computing
2020

UCSC

Declaration

The thesis is my original work and has not been submitted previously for a degree at this or any other university/institute.

To the best of my knowledge, it does not contain any material published or written by another person, except as acknowledged in the text.

Student Name: T.H.M De Silva

Registration Number: 2017/MCS/022

Index Number: 17440224

____

Signature: Date:

This is to certify that this thesis is based on the work of

Mr. T.H.M. De Silva

under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by:

Supervisor Name: Dr. A.R Weerasinghe

____

Signature: Date:

Abstract

Social information has emerged as a key component that drives growth in modern day society. The emergence of Web 2.0 and the global connectivity it brought has caused massive changes in news-reporting and journalism landscapes. Micro-blogging platforms play a key role in global news propagation today. There is a growing need to filter the noise and extract only credible or useful information from the unprecedented volume of data disseminated through these platforms every day. While there is a number of studies carried out on determining credibility of user generated content, there is no one accepted credibility analysis solution. The solutions presented in the past are also difficult to be used in real world applications owing to complexities. In this study, a new methodology for solving the issue of filtering credible information online through analysis of source credibility is proposed. Firstly, a thorough literature review on existing credibility assessment techniques and analysis approaches is conducted. Three user credibility ranking models which follow three slightly different approaches to rank users are proposed and implemented by using data acquired from Twitter social platform. Throughout a detailed study, it is shown that the analysis of perceived credibility of content authors — established through either human input or by assessing the available metadata — is helpful to identify highly credible users within these platforms. The findings from this study show that by preemptively identifying credible users in a network through network analysis methods, it is possible to curb misinformation on micro-blogging platforms.

Acknowledgement

At the outset, I wish to express my sincere gratitude to my supervisor Dr. A.R Weerasinghe, Senior Lecturer at University of Colombo School of Computing – UCSC, for his tremendous encouragement, constant supervision, unwavering support and guidance throughout my research.

Also, I would like to convey my utmost gratitude to all the other academic members of University of Colombo School of Computing – UCSC for the knowledge they passed on to me throughout this degree program.

I also wish to express my candid thanks to my employer — IronOne Technologies — and its CEO, Mr. Marko Garafulic for encouraging me to pursue this program and for partially funding it.

Finally, I would like to thank my colleagues who were supportive and offered valuable insights, my family and friends for their support and encouragement given endlessly to make this project a success.

Table of Contents

List of Figures

List of Tables

Chapter 1 – Introduction

1.1 Research Background

Social Information has become a key component that drives the growth and development of modern-day society. The Internet has presented itself as a platform for distribution of data. It is a collection of networks, where an enormous amount of data is being generated, shared and processed through networks of connected bodies. Along with the emergence of Web 2.0, social networks and micro-blogging platforms have gained popularity as a significant part of the World Wide Web. Social networks are virtual communities where members are able to communicate and share information about other members or events that take place in their everyday lives.

Facebook and Twitter are the market leaders among these existing social platforms, with Facebook hosting 2.45 billion monthly active users while Twitter hosts 330 million monthly active users. Every day, people around the world subscribe to these services for many reasons, such as, getting the latest breaking news updates or to engage in public discourse over a variety of topics ranging from the ever-popular politics, sports to celebrity scandals. Due to the extraordinary popularity these platforms have received, the news consumption by the masses has gradually begun to shift to online and the Internet has emerged as a major source of news, causing a sea of change in the new media landscape [25].

Breaking news situations in particular have thrived the most from these advancements due to people's hunger for speedy updates on situations that rapidly evolve in real time [26]. Nearly 65% percent of more than 2.4 billion internet users receive breaking news from online social media platforms such as Facebook, Twitter, YouTube and Instagram instead of conventional news media sources such as print media and televised news. This unprecedented shift from traditional news media to social media has presented many challenges for traditional journalists and emergency services, as they now have to compete with crowdsourced information. It has led to journalists having a difficult time in striking a balance between the need to be first need the need to be correct causing traditional news services to report unsubstantiated news in a rush to be first [22][23].

Twitter is one of the most popular social networks that saw tremendous growth in the recent years. It remains as one of the most efficient micro-blogging platforms for sharing real time information. As of 2018, it had more than 640 million users and more than 300 million monthly active users [2]. The popularity of Twitter is due to its simple anatomy which helps unidirectional information flow in real time. The users on Twitter can generally post a 280-characters long message at a time, which is called a "tweet". Users on this network do not need to follow each other to get updates from others. This asymmetrical nature of the network has worked in its favor as Twitter has become the leading micro-blogging platform for information diffusion particularly during crisis situations.. Notwithstanding its immense popularity and efficiency as a real time news propagating tool, Twitter suffers from information overload which has allowed malicious parties to misuse the platform for causing harm through numerous unwanted acts such as spam generation, rumors and fake news propagation, to influence the political and social decisions in a society through deliberate or accidental misinformation campaigns carried out. Such incidents have caused damage, created panic and chaos within the society. Due to a lack of fact-checking mechanism in place to filter out these malicious users and their content, Twitter has become a hotbed for rumour propagation and sharing [2].

In many instances, the information generated from online social platforms such as Twitter has influenced, and often shared among wider network in the same platform or among different platforms, without any kind of assessment or evaluation of its credibility. Many people do not verify the trustworthiness of the information being shared. A user is more likely to trust the information generated by a close friend or an acquaintance without questioning the credibility of the information as compared to one shared by strangers. Trust and credibility between users or communities of users can appear as more and more information are generated, shared and acted upon.

The concept of Credibility has been drawing attention since the emergence of Web 20 in late 1990s, which provided the online users with the ability to generate, share and interact content with little or no reference at all to the sources. Credibility as a concept has since been considered in a general sense, by proposing whether an online user can trust an information source or the information itself. Oxford dictionary defines credibility as *"the quality that somebody/something has that makes people believe or trust them",* while the Cambridge dictionary defines it as *"the fact that someone can be believed or trusted"*. The most suited synonyms of credibility are trustworthiness and believability [2].

Research on credibility analysis of online social networks has seen a rapid growth over the last decade. Despite the upward trend observed in such research, there exist challenges in this domain. For example, there is no benchmark for establishing user credibility or the credibility of the content shared on social networks. One of the major challenges is the enormous magnitude of online social network users and the dynamic, clustered structures of these networks [2]. As the user numbers grow, these networks can grow into tremendous sizes, and this may cause obscuring of information related to the users that can be helpful in determining their credibility. Another important challenge that motivates further research in this area is that the trustworthiness of a user and the content they author on social media can often be affected by the relationships they maintain with other users in the community as well as their social standing. The possibility for malicious users to circumvent currently used defences to cause targeted attacks is also another challenge. For instance, it is relatively easy to purchase a large number of fake or bot accounts to increase a user's following on the network to become popular overnight. There exists malicious software which allows creation of bot farms to generate unprecedented amounts of spam. Because of this, the reputation of a Twitter user does not always lead to credibility. It is therefore important to find ways to determine the credibility of the content users might observe on platforms such as Twitter, and to explore methods which could help establish the credibility of the users themselves on the network in order for them to be considered as trustworthy and reliable by other users in the community.

1.2 Problem Definition

Ever since the emergence of social media, micro-blogging platforms have become a part of our daily lives. People around the world subscribe to these platforms for many reasons, such as getting the latest news, engage in public discourse over a variety of topics such as politics, sports, human rights etc. As the number of users of these platforms are at an exponential growth every year, these platforms have become hot targets for fake news, rumours and other misinformation campaigns, both deliberate and accidental.

Twitter is by far the most popular online micro-blogging service recognized worldwide, with 500 million registered users in 2012 and had more than 340 million tweets shared per day [1]. Due to the sheer volume of information it carries, Twitter has become an imperative and timely source carrying mass sentiment and opinions on a wide variety of topics, thus making it a breeding ground for misinformation/disinformation campaigns online. Such campaigns have caused harm to people and organizations in the past due to negative sentiments allowed to form within the communities through misleading information being shared. To address this issue, scholars have come up with promising research on rumours and misinformation detection and verification mechanisms. Prior research on Twitter user credibility assessment is carried out on three levels; tweet-level, user-level and topic-level.

The existing solutions to tackle this issue are two pronged; Manual and Automated [2]. Manual approaches are mostly fact-checking efforts led by journalists, which is time consuming and therefore inefficient. Automated solutions are based on three main disciplines; Machine-learning, Social structure analysis and Weighted algorithm and Information Retrieval (IR) based analysis. While machine learning based automated solutions have shown promise, they suffer from lack of updated, relatively sizable labelled datasets which is pivotal in gaining higher accuracy for classifiers that need training, or to retrieve patterns from data through clustering techniques. Social structure analysis-based solutions have been derived under the perceptions that the users participating in news propagation are inherently credible, which is far from the reality, and therefore cannot guarantee credibility. Weighted algorithms and IR based research have shown promise but is scant.

In this research, the role of perceived user credibility factors on determining a tweet as a trustworthy news from unjustified assumptions or rumours is investigated. It is argued that the credibility of a Twitter user can be determined with enhanced accuracy by attempting to build a novel, hybrid, network analysis-based user credibility ranking model by leveraging existing credibility assessment and analysis techniques.

1.3 Motivation

In April 2019, On Easter Sunday, six deadly suicide bomb attacks were carried out simultaneously in Sri Lanka, killing more than 250 people and injuring more than 600 individuals. The attack carried out by an extremist terrorist outfit threw the peaceful island nation into chaos, and a vast number of rumours, fake news and misinformation campaigns were propagated in social platforms which caused communal violence in the aftermath. The government of Sri Lanka could not contain this situation and opted to enforce a blanket ban on all social platforms for more than two weeks. The lack of an official source to fact-check claims made online within a reasonable time was a hindrance to restore normalcy after the attack. Twitter in particular was a hotbed for sharing unsubstantiated rumours during this period [24].

Motivated by the lack of a mechanism to fact-check rumours during this incident, this research attempts to build a novel credibility ranking model that could help determine the credibility of users who post sensitive breaking news messages on Twitter, and use the said model to identify communities of credible users on Twitter network, which will help curb propagation of rumours and misinformation on Twitter.

1.4 Research Questions

RQ 01: How can we determine a tweet as trustworthy news from an unsubstantiated rumour by assessing the credibility of the users?

SRQ 01: What are the standard approaches used to assess and analyse the credibility of users on Twitter?

SRQ 02: What are the limitations in existing approaches?

SRQ 03: How can we improve the performance (further) of existing approaches?

SRQ 04: How do we rate the credibility of a tweet from a credible source to get a more fine-grained rating?

1.5 Research Aim and Objectives

The aim of this research is to design, implement and evaluate a network analysis-based user trust assessment model for detecting and verifying rumours on Sri Lanka based Twitter. This research will investigate the role of perceived user credibility factors on determining a tweet as a trustworthy news from unjustified assumptions or rumours. This will be achieved by considering the following objectives:

1. Critically review the function of online social platforms, their users, and how they operate.
2. Evaluate the characteristics of the users and the factors that could help determine their credibility.
3. Critically review existing credibility assessment and analysis techniques to pick the most suitable approaches.
4. Collect data to be used in the experiments.
5. Propose a novel trust model for assessing the credibility of Twitter users and develop a prototype of the proposed model.
6. Discuss and evaluate the prototype using proposed evaluation methodology.
7. Submission of a dissertation with research findings.

1.6 Scope

This research attempts to investigate user credibility assessment and analysis factors that could contribute positively in preventing viral spread of rumour and misinformation on Sri Lanka based Twitter. An in-depth literature review will be conducted to find out:

1. Perceived user credibility assessment methods
2. Best performing user credibility analysis approaches

Perceived user credibility assessment methods will be investigated on three fronts;

1. Post level
2. Topic level
3. User level

Based on the findings and the gaps identified in existing literature, a novel, network analysis-based, hybrid credibility ranking model to identify credible users on Twitter is proposed. A prototype will be *designed, implemented* and *evaluated* based on this trust model.

Tweets that were shared on Sri Lanka based Twitter during Easter Attacks incident in 2019 will be collected to be used in building and evaluating this system. Only the tweets in English language will be considered for these experiments.

Twitter's developer tools, necessary data collection and system modelling tools will be reviewed to build the proposed system.

1.7 Summary

This chapter discussed the problem of information overload and the importance of analyzing source credibility on microblogging platforms such as Twitter for tackling the research problem in the absence of an existing fact checking mechanism which has caused unprecedented amounts of rumors, fake news and misinformation propagation on social networks. Initial findings from a research carried out around the research problem and the domain, it is identified that a network analysis-based approach could be utilized to provide a solution. Based on the findings, this research plans to focus on building multiple credibility network models which can successfully rank users on the Twitter platform. The chapter clearly defined the problem and stated the author's motivation pursuing this research. The exact research question is established, and the supporting sub research questions are stated. Thereafter, the aims and objectives of the proposed research is defined. The scope of the research is clearly outlined towards the end. In order to begin the research, a comprehensive literature review is planned with the aim of achieving the objective 1,2, 3 stated in section 1.5. The next chapter discusses in detail the literature review conducted for this research.

Chapter 2 - Literature Review

2.1 Background

Rumour is an old social circumstance used in many social events such as politics and other public spaces [3]. It is more commonly considered as talk which is uncorroborated by any official entity or evidence that points to its validity or truth [4]. Rumours are unverified propositions or hypotheses whereby transmission of messages happens in a way where the receiver is unable to decide on the believability of the information received [3]. It is also referred to as hearsay in many instances. Turenne [3] conducted a study on rumour theory which suggested categorization of rumours under nine different topics which included political attacks, financial disinformation that caused harm to organizations, and to regional and local government stability, defamation and panic alerts that attempted to induce terror.

Misinformation — sometimes known as "Disinformation" — is referred to as a natural language occurrence that has always remained through a mechanism where it is spread from mouth to ear [3]. There is a massive amount of user generated content being created and circulated on the internet today. It is observed that individuals, organizations and bot programs are actively attempting to promote their own propagandas by manipulating this data [5].

The unmoderated nature of social media has paved the way for the spread of rumours online [6]. Due to this phenomenon, news consumers now have a hard task of sifting through a plethora of news items to separate trustworthy news from unjustified assumptions or rumours. There have been situations in the past where the online spread of rumours and misinformation have caused harm to people and organizations [7].

2.2 Credibility Assessment Systems

In recent years, there is a noticeable trend of research being carried out in this domain to investigate user credibility assessment approaches which could be used to build systems that automatically detect and verify rumours and fake news on social media. Castillo et al. proposed that user credibility assessment could be carried out under three levels; Post-level, Topic-level and User-level [8].

Most of the research work has been focused on post level credibility assessment. The studies vary in their approaches for each level, as do the features, methodologies, datasets and the extent to which human participation is expected. The research in this domain also shows instances where hybrid approaches have been looked at, especially on the post level and topic level. Some researchers organized experiments on all three levels [2].

2.2.1 Post-Level Credibility Assessment

At post level, the characteristics of the textual content in a tweet are analysed to determine a credibility score and its trustworthiness [2]. Research on this topic is two-fold; One approach considers the absolute and archival user or topic data available for assessing a tweet [9], while the other approach only takes in real time data and bases its assessment only on the data available within the post. Researchers have used diverse sets of features extracted from various tweet attributes but it all can be classified under three main sets of features; Message characteristics, multimedia features and sentiment features [2].

2.2.2 Topic-Level Credibility Assessment

During a breaking news situation, thousands of user generated content is published on Twitter every minute [6]. Topic level characteristics usually tend to accumulate tweet level cues such as links and hashtag functions of the tweets, the ratio of positive to negative words in a sample, and average sentiment scores obtained for the tweet content [2].

Zhao et al. [10] proposed a trust distribution algorithm which re-estimated trustworthiness repeatedly based on social and contextual properties and passed decisions on trustworthiness of a user for a given topic. They used a similarity-based trust evaluation method by assuming that tweets are trustworthy if they are consistently similar in situational properties against reliable news articles or tweets. While their method achieved better results against a supervised machine learning based keyword match experiment, it suffers from one major drawback. They based their similarity comparison on published news articles. However, during a high impact event, it is highly unlikely that fact-checked news articles will be published in a timely manner. Therefore, their approach will not be effective on systems that attempt to detect and verify rumours in real time.

2.2.3 User-Level Credibility Assessment

User level credibility is assessed by analysing the features extracted from twitter profiles and the content generated from users. Characteristics such as the number of followers, friends, tweets and retweets can determine the propagation of tweets in the network and also affect the reputation of a user [2].

Morris et al. [11] planned a set of experiments to discover the features that could possibly influence a user's beliefs about credibility. Their results concluded that the influence of a user, the user's reputation on the network (for an instance if the user is verified), and their expertise in certain topics as judged by analysing their user bio content were influential in enhancing a user's credibility. It was found out that a user's username had a significant impact on user credibility. They also found out that the use of a Twitter's default display picture contributed to a lower perception of credibility among the participants.

TABLE I:
CREDIBILITY ASSESSMENT ATTRIBUTES IDENTIFIED IN PRIOR LITERATURE UNDER POST, TOPIC AND USER LEVELS [2]

Post-level	Message characteristics	● Tweet length ● # of replies ● # of retweets ● Use of hashtags, mentions or URLs ● # of replicated tweets ● Nouns and verbs used to narrate events
	Multimedia features	Use of media metadata. ● Description ● Title ● Size ● Video duration ● Average # of tags per photo ● Average upload time between consecutive uploads
	Sentiment features	● # of positive phrases ● # of negative phrases
Topic-level	● URLs ● Hashtags ● # of keywords in a sample ● # of nouns and verbs used to narrate an event	

User-level	Explicit attributes	● # of followers ● # of friends ● # of tweets ● # of retweets ● Age of the account
	Latent attributes	● Age bracket ● Gender ● Secondary school degrees ● Beliefs

2.2.4 Hybrid-level Credibility Assessment

Many of the research work in recent past have adopted hybrid approaches to assess credibility using the combined advantages of post level, user level and topic level assessment of credibility. Under such hybrid level assessment, credibility assessment models are able to keep the perception of complete outfit (tweet, topic and user) and relation (structure of the network) to determine precise credibility [2].

Castillo et al. [8] conducted a study to analyse the credibility of news shared on Twitter. They focused on building an automated credibility analysis system using four types of features: Textual (post) features, Author based features, Topic attributes as well as Propagation based features, which considered characteristics such as depth of retweet tree built from retweets of tweets or number of tweets posted in the beginning under a specific topic.

Kim et al. [27] investigated the social network activity on Facebook and Twitter after the floods that occurred in Louisiana in the city of Baton Rouge in 2016. They utilized both post level and user level credibility assessment to analyse the data collected. The results concluded that different roles are played by individuals and organizations in online networks when reactivating to a crisis situation.

2.3 Credibility Analysis Methods

There are three major categories of credibility analysis methods; Automation-based, Manual and Hybrid approaches [2]. This categorization is based on the perspectives of researcher's apprehension of the problem. Some researchers treated the issue as a classification problem which needs standardization through the aid of artificial intelligence. Some other researchers thought of it as a cognitive task which required manual input from humans. There exist methods that aggregate credibility analysis methods from main categories and the sub-categories, which can be considered as hybrid methods [2].

Automation-based approaches can be further divided into Supervised and Unsupervised machine learning techniques, which is gaining traction since of late.

2.3.1 Automation based Credibility Analysis

There are many instances of existing research in academia focusing extensively on automation-based credibility analysis approach to solve the research problem. This section discusses a summary of such work reviewed.

2.3.1.1 Supervised Learning Based

Thakur et al. [12] suggested a two-fold approach to classify rumours depending on personal and non-personal features extracted from tweets. They use Naive Bayes as the algorithm and Support Vector Machines (SVM) to train their classifier. Kwon et al. [13] discussed an approach that makes use of linguistic features to define doubt, negation and guessing to describe the process of "doubt", raised by a person upon receiving a rumour. However, supervised learning approaches depend on labelled data which is hard to acquire.

2.3.1.2 Unsupervised Learning Based

Unsupervised learning approaches that make use of clustering algorithms to detect similarity in user's tweets to analyse user behaviour have been attempted [14]. However, the effectiveness of this method has not been proven with real cases. Most of the dataset sizes available are relatively small to be able to draw general conclusions as they are not indicative of the patterns of social behaviour.

2.3.1.3 Graph Based

Gupta et al. [28] used a PageRank based algorithm which they called EventOptCA to model relationships between users, tweets and events. Theirs was an approach to analyze credibility of tweets during an event using graph-based optimization. They calculated these relations iteratively until final credibility scores were calculated. Gupta et al. first built a basic trust analysis model and they enhanced it by updating event credibility scores using graph regularization. Graph regularization is a technique that comes under the paradigm of Neural Graph Learning [29]. This enhanced approach out-performed basic credibility analysis model and the methods used were significantly more accurate.

Lu et al. [20] suggested that key nodes in a microblogging network have the power to influence. Key nodes perform important roles in terms of network's structure and function. They proposed an interesting rumour spread-mechanism model, ISR (Ignorants, Spreaders and Rejectors), to derive an algorithm based on a user's direct capability of influence and region of influence, which outperformed PageRank algorithm in ranking users based on calculated trust scores.

This research is heavily motivated by the above-mentioned work as their approaches looks promising for a network analysis-based solution to overcome the research problem.

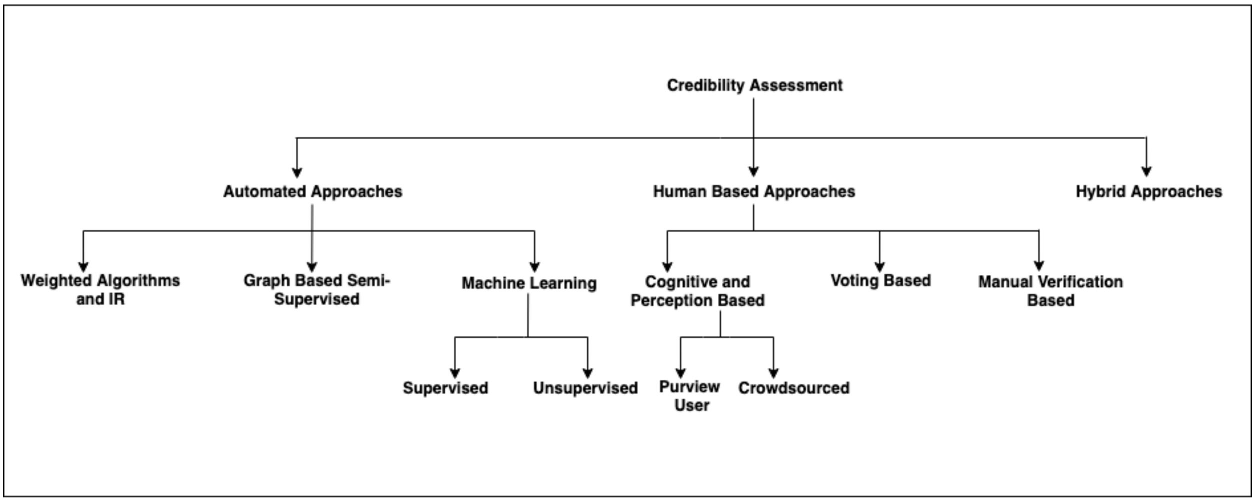

Figure 2.1: Classification of credibility assessments in microblogs [2]

2.3.2 Human based Credibility Analysis

This section discusses credibility assessment approaches that assess user credibility on all three levels manually. Most of the work done in this area rely on human subjects for a final judgement during evaluation.

2.3.2.1 Fact Check Based

Human-based credibility assessments mainly comprise of manual fact-checking performed by social media services themselves or journalists. These approaches suffer from limitations such as general inefficiency due to the time and labour needed for manual fact-checking, as well as their lack of ability to catch sight of hostile activities promptly [2].

FactCheck.org [15] and PolitiFact [16] are established fact checking websites in the United States that attempt to verify the accuracy of rumours posted online, but it is not done in a timely fashion.

WatchDog Sri Lanka [17] is a fact-checking organization in Sri Lanka which aims to verify accuracy of rumours and fake news posted on Sri Lanka based Twitter. This relatively new service was established in the aftermath of a deadly bombing incident that took place in Sri Lanka during Easter weekend in April 2019. Their verification approach is manual fact-checking based on five verified Sri Lanka based Twitter news correspondents.

2.3.2.2 Voting Based

Voting based credibility assessment is an efficient method to assess user level credibility, where the users/accounts can be ranked based on a trust value generated for a specific tweet. Canini et al. [18] proposed an interesting approach to build a refined credibility model which took follower relationships as votes of confidence by users within a network, which were further distinguished between social structure and reputation. Their algorithm attempted to determine the trustworthiness of user-generated messages on Twitter for any given topic. Their algorithm had the potential to suggest intriguing users on Twitter. The idea of aggregating topic-based content analysis and network structure related knowledge could prove useful in building highly credible communities of Twitter users for different topics of interest. One of the drawbacks of their approach was that their algorithm required the knowledge of the follower graph.

2.3.2.3 Cognitive and Perception Based

The ability to perform cognitive tasks has been considered as an important factor that affects people’s day to day lives. The concept of cognitive ability differs across different social science disciplines. For an instance, in psychology, cognitive ability is referred to as the competency to perform mental operations such as problem solving, habituation, ability to understand, reasoning, knowledge gathering and making constructs [30].

This ability helps individuals to identify the cognitive task paired with the data propagation [2]. Due to the reasoning abilities that comes with cognitive ability of individuals, they are naturally good at perceiving credibility of things and beings they surround with. Hence, this is an important factor which can help determine the credibility in micro-blogging networks. Through observation of factors and the importance given by the humans to such factors in determining a source is credible or not, it is possible to develop models that could emulate human reasoning and ability to make connections when determining credibility.

An internet survey organized by Yang et al. [19] attempted to make comparisons between user views about US based Twitter and China's equivalent Sina Weibo. They proposed a general framework to analyse whether a given tweet tends to be a rumour based on three-pronged sets of features; Content, Account and Network analysis. Under Account analysis, they compared relevance of multiple perceived credibility factors such as, gender, account name, display picture, country of origin and user interests. Based on the findings, they determined that the age of the user account, the follower ratio and the repost ratio performed the best with evaluated data in determining user credibility.

Mendoza et al. [31] studied an event about an earthquake happened in Chile to conduct a survey and establish the reliance of Twitter for information propagation during a crisis situation. They looked at how verified news and false rumours were spread through Twitter during the incident.

A statistical evaluation of attributes was performed by O'Donovan et al. [32] which compared attributes such as diverse topics, credibility, length of tweet chains and dyadic pairs. They found out that URLS, mentions, retweets and the length of tweet were good indicators of credibility.

Most of the research carried out in this domain seems to suggest that the said characteristics are important to assess credibility of online social networks. Thus, this research too will utilize the frequently used features in prior literature that are considered most relevant to determine credibility of a user.

2.3.3 Hybrid Credibility Analysis

It is observed that a number of researchers have chosen hybrid approaches to analyse credibility with the hope of aggregating the advantages of automation and human based approaches. The methods, approaches, assessment factors selection varies from one approach to another [2]. Some researchers have used machine learning approaches with human perception-based approaches while others used graph analysis, perception-based approaches and clustering techniques all in one.

TABLE II:
SUMMARY OF HYBRID CREDIBILITY ANALYSIS METHODS ATTEMPTED IN PRIOR RESEARCH

Authors	Event level	Post level	User level	Models/Algorithms/Approaches
Kang et al [33]	+	+	-	Social graph, perception, Bayesian J48 tree
Qazvinian et al [34]	-	+	+	Crowdsourcing, Bayes classifier
AlRubian et al [35]	+	+	+	Human expertise, naive Bayes, Pairwise comparison
Kumar and Geetha Kumari [36]	+	-	+	PageRank algorithm, statistical method

AlRubian et al. [35] proposed a reputation-based source credibility assessment method using existing models and by introducing a set of new features. Their approach combined analysis of Twitter's social graph structure and user's sentiment to identify and evaluate credible sources on Twitter which were topically relevant. They measured the credibility of a Twitter user based on how popular she/he is and how sentimental the user is regarding a relevant topic. The results of their approach showed significant accuracy at locating users who were considered credible on a relevant, pre-defined topic. One drawback of their approach was that they did not take the dynamic nature of the Twitter eco-system into consideration, especially the follower relationships and the possibility of highly credible users endorsing less credible users through retweets and the impact it could have on their credibility analysis model.

Table III lists the relationships identified between tweets and users in Twitter network by AlRubian et al. These relationships can be important in selecting features to be used in the model proposed by this research.

TABLE III:
RELATIONSHIPS BETWEEN TWEETS AND USERS IN TWITTER NETWORK. [35]

	User	Tweet
User	Follows / is followed by Mention Replies to Retweets to	Posts Retweets Marks as favourite Replies
Tweet	Posted by Retweeted by Marked as favourited by Replied by	Replies / is replied from Retweets / is retweeted from

2.4 Summary

This chapter discussed the prior research work carried out in the problem domain. It is discovered that there are three main approaches to solve the problem of credibility in online social networks. The research in this domain is dominated by post level credibility assessment systems. However, topic level and user level credibility assessment approaches have shown encouraging results and promise. In analysing user credibility, many different approaches have been investigated in existing literature. Each approach has their own advantages and drawbacks. After careful consideration, it is decided that a graph based, network analysis approach coupled with a novel credibility assessment which takes into consideration user level credibility features discussed in this chapter is most appropriate for building the proposed solution. The following chapter will outline the exact path this research plan on taking to achieve its research goals.

Chapter 3 – Methodology

3.1 Introduction

The methodology of a planned research is the path chosen to explore the research question and to discover a solution by using the knowledge gathered from previous literature review chapter. This chapter discusses various research methodologies that are commonly used by the researchers in computer science domain. The specific research methodology, research design and the design approach used in this research is acknowledged and discussed. The chapter starts off by providing a complete introduction to the research undertaken. It then describes how the research is planned out based on the knowledge gained in literature review. In addition to the details of the bona fide research methodology adopted in solving the research problem, this chapter also discusses the datasets that are collected, their appropriateness and the exact methods in which they are used to build and evaluate the proposed model. Furthermore, it attempts to interpret the results obtained through experiments designed and implemented in this research. It then justifies the selection of an appropriate evaluation plan to evaluate the performance of the proposed model and attempts to critically evaluate the performance of the model based on that plan.

3.1.1 Representation of the Problem

The aim of this research study is to find a solution to the problem of assessing the credibility of users on Twitter using a hybrid credibility analysis approach. This approach includes network structure analysis coupled with credibility analysis of the nodes in the social graph based on factors humans believe as important to determine user trustworthiness. Studies have shown that Twitter is one of the most active crowd-sourcing news platforms during breaking news situations. The social structure on Twitter is such that no two persons have to follow each other to receive updates from one another. This is an advantage as far as the reach of a news item is considered because a seemingly believable news can be propagated to a large audience. It is, however, problematic when it comes to fact checking this information for its truthfulness and to differentiate between news as confirmed legitimate, fake news or simply rumors.

As discussed in previous chapters, the existing solutions to verify news during developing situations are scant, difficult and time consuming. The solution proposed is to analyze the network of users by taking their social relationships into consideration and to judge their credibility based on the publicly available user metadata. It will help to build up a credible network model of Twitter users in a given community whom can be trusted to share credible news on the network, thus minimizing the chance of fake news or rumor propagation. The solution was developed in three incremental phases with each phase ending by producing a credibility network model.

The results of the three credibility models will be assessed based on the evaluation plan to determine the success of the proposed solution.

3.2 Selection of Research Methodology

A research methodology can be thought of as a framework within which a research study is conducted [37]. It is the path through which a researcher formulates the research problem, objectives and present results from the data collected during the study [38]. The method to be used in a research study must be clearly outlined and the selection justified in order for the results of the research study to be given credibility. The choice of research method will most likely be determined by the purpose of the research and the research question [37].

There are two major types of research methodologies used in academia today. They are quantitative and qualitative research methods. These methods operate through different approaches, based on underlying goal the researcher aspires to achieve. According to Newman et al. [39], the qualitative research approach is used when attempting to observe and interpret reality *"with the aim of developing a theory that will explain what was experienced."*

The quantitative research approach is applicable for situations where "*one begins with a theory (or hypothesis) and tests for confirmation or disconfirmation of that hypothesis."*

Quantitative and qualitative research are each appropriate for different types of research problems, according to Cassell and Symon [40]. Hence, the research question determines the path to take for the researcher to achieve expected results.

3.2.1 Qualitative Research Method

Qualitative research provides a thorough description and an apprehension into human experiences. It is a comprehensive approach which involves discovery [41]. Creswell describes qualitative research as an evolving representation which occurs within a natural setting [42]. It allows the researcher to acquire a high level of detail through greater involvement in the actual experiences. The investigation of a social experience through the viewpoint of a participant is one unique attribute of qualitative research [41].

Those employing qualitative research tend to place importance and value on the human explanative aspects of knowledge on social phenomenon and the significance of researcher's own elucidations and understanding of the phenomenon being studied [43].

Qualitative research is based on inductive, rather than deductive reasoning [41]. In following a qualitative research approach, the researcher attempts to explain the questions that arise from observed constituents. There is no single explanation that is better than the other [44].

A qualitative researcher's underlying objective is to observe, describe and understand human behavior in a natural environment beginning with an assumption or prior knowledge. The success of qualitative research depends on the skills and expertise of the researcher. According to Patton [45], while the reliability of quantitative research is determined by the apparatus constructed, in qualitative research *"the researcher is the instrument"* [46].

Reliability and validity are processed independently in quantitative studies. However, these terms are not looked at separately in qualitative research. Instead, terminology that enclose both these terms, such as credibility, transferability and trustworthiness is used [46].

3.2.2 Quantitative Research Method

Quantitative research methodology is defined as "the general approach the researcher takes in carrying out the research project.", according to Leedy & Omrod [47]. It involves the collection of data, out of which information is extracted in a quantifiable manner and can be subjected to statistical treatment in order to support or oppose *"alternative knowledge claims"* [48].

Quantitative research is associated with three historical trends. The research design, test and measurement strategies, and statistical analysis. It also involves data collection which is numerical in nature quintessentially, and mathematical models are often preferred as a methodology of data analysis by the researchers [41].

Quantitative methodology identifies new variables or attempts to discover new relationships amongst and between different variables. According to Lichtman [44], quantitative research deals with hypothesis and testing of the said hypothesis with the measurable, numeric data. It is an experimental approach that focuses on the outcome over the process used to obtain results. Hence, the research itself is independent of the researcher. Thus, the data can be used to measure reality objectively. Quantitative research produces meaning through interpretation of collected data in an unbiased manner.

Table IV lists some basic comparisons between Qualitative and Quantitative research methodologies.

TABLE IV:
COMPARISONS BETWEEN QUALITATIVE RESEARCH METHODOLOGY AND QUANTITATIVE RESEARCH METHODOLOGY [37][44][40]

Qualitative Research	Quantitative Research
Is concerned with behavior and situation	Concerned with cause and effect
Is focused on Interpretation	Focused on quantification
Gives room for flexibility and less rigidity	Less flexibility and rigidly defined
Longitudinal research design	Cross sectional research design
Emphasis on the richness of qualitative data	Emphasis on statistical data
Numbers may not be of importance but might be involved in different instances	Involve numbers most of the time
Treat those studied as participants and informants	Treats those being studied as anonymous objects to be measured
Takes place in naturalistic settings	Takes place in experimental settings
Uses inductive approach	Uses deductive approach
Progresses from specific to general	Progresses from general
Works with observations	Involves sampling and surveys
Understands and interprets the meaning of human interaction	Statistical test to decide whether or not to reject the null hypothesis
Relies on interpreting and understanding behaviour	Relies on hypothesis testing and analysis
Relies on interpreting and understanding behaviour	Relies on hypothesis testing and analysis
Is geared towards process rather than outcome	Is geared towards outcome rather than process
Is concerned with emergent themes	Is driven by specific hypothesis

3.2.3 Mixed Method

The mixed method approach to research began to emerge in the mid to late 1900s according to Tashakkori and Teddlie [49]. In a mixed method approach, researchers incorporate methods of collecting or analyzing data from qualitative and quantitative approach to research [50]. The goal for researchers using mixed methods approach is to benefit from the strengths and to keep the weaknesses found in quantitative and qualitative approaches to a minimum [50].

This research mostly uses quantitative methods as the research question deals with the subject area of assessing the perceived credibility of users within a micro-blogging network and to achieve better accuracy. For this purpose, the collection of data in this research will be mainly of numerical nature with more importance given to collecting data which depicts a user's relationships amongst and between other users in the network, as well as user's activity based statistical data which will be used to assess credibility.

However, since this research is based on perceived user credibility factors, it is imperative that the results of the experiments be evaluated from a qualitative standpoint as well. Humans are inherently good at perceiving and understanding credibility and trust, as well as ranking more than assigning a score due to their superior cognitive capabilities. Therefore, evaluating the results of experiments conducted in this research with the involvement of humans makes sense.

It is not uncommon to use mixed methods in a research study of this nature, and it only shows the diversity of this research problem. As such, appropriate methods are used as needed at appropriate stages.

3.3 Selection of Research Design and Approach

According to Cormack [51], the research design represents the major methodological propulsion of the research by being the distinctive and specific approach, which is best equipped to answer the research questions. The research questions, the aim and objectives of the research therefore influence the selection of an appropriate research design [52].

Burns and Grove [53] stated that the purpose of the research design is to achieve a substantial control of the research and to improve the validity of the research by examining the research problem. According to van Wyk [54], research design talks about what data is required, what methods will be employed for data collection and the analysis of this data, and how all this will answer the research questions.

This study attempts to investigate the factors which could help determine the credibility of nodes (users) in micro-blogging networks through analysis of primary data. Experiments are designed to retrieve the required quantitative data for this purpose. The credibility of users on micro-blogging platforms can depend on many factors as has been pointed out in the previous chapter. Thus, the data collected will need to be analyzed closely to find any causal links between factors or variables that relate to the research problem. This study will thus take an inductive research approach to interpret the data and to build up theories surrounding the credibility assessment of users on micro-blogging platforms such as Twitter.

3.3 Approach for Building Credibility Network Models

As mentioned in the previous chapter, the solution prototype consists of three phases. The first phase is building a model that could identify the key actors on the network who can be considered as influential within a community. This is done by running a graph analysis algorithm on the follower graph and obtaining their rankings. A variation of the PageRank algorithm is used for this purpose, the details of which will be discussed in the subsequent sections.

The second phase is building the novel, credibility assessment-based network model. It is done by firstly building a credibility scoring system which analyzes a Twitter user's profile metadata to calculate a credibility score for that user, and then by introducing the derived credibility score of individual users as their node weight to PageRank algorithm to influence the existing user ranking in the graph.

The third and the final phase is to build a model by seeding users who are deemed as "most credible" within the Sri Lankan Twitter community for receiving breaking news. This is done by firstly taking feedback from the community as to who they consider as highly credible within the network. Based on the results, the idea is to seed these credible people into the network and analyze it to identify other credible users within the network. The following section describes the general overview of the selected approach.

3.3.1 Overview of the System Architecture

This section aims to present the architectural and functional design of the proposed credibility network models. The first phase of this process is the Data collection. From the preliminary dataset that is collected, a subset of data is used to collect a secondary set of Twitter user data which is pivotal to build the user network. The populated user network is then analyzed, initially based on the follower relationships that exist among the users in the network.

It is observed in prior literature that the key nodes in a network have the potential to be influencers in that network [20]. From the existing research on this domain, it is found out that the identification of such nodes using network analysis techniques has proved useful to observe key players in a network. Therefore, as a baseline measure, this system attempts to build an influencer network by running a network analysis algorithm on the follower graph built using the datasets mentioned above. The PageRank algorithm developed by Google to index webpages based on their popularity is widely used for this purpose in the existing academia [2, 36, 55]. As such, this system also utilizes the original PageRank algorithm to build its first model.

It should be noted that this model cannot be called a "credibility network model" as it only takes into account the follower relationships among users in order to rank them. Therefore, this model will only attempt to identify influential users within the network, which is referred to as **Model 0**.

The literature review from previous chapter also discussed prior work in academia on user credibility analysis approaches that have been explored based on metadata obtained from Twitter users. Metadata is publicly available user profile data provided by the Twitter platform to identify a specific user on the network. The existing research work on this domain has yielded interesting results [19]. Motivated by the results of their work, this research attempts to design a credibility scoring algorithm to assess a user's credibility.

The proposed algorithm will take into account a series of credibility factors which are considered as "important" in analyzing a user's reputation or credibility in prior research work. The system will utilize this proposed algorithm to calculate a credibility score for each user in the follower graph. It will then run the chosen network analysis algorithm by providing initial weights to each node in the network. The initial weight introduced will be the credibility score generated for a user by the algorithm mentioned above. Based on this, the system will build its second user ranking model — which can be called the first credibility ranking model. It will be referred to as **Model 1** in subsequent sections.

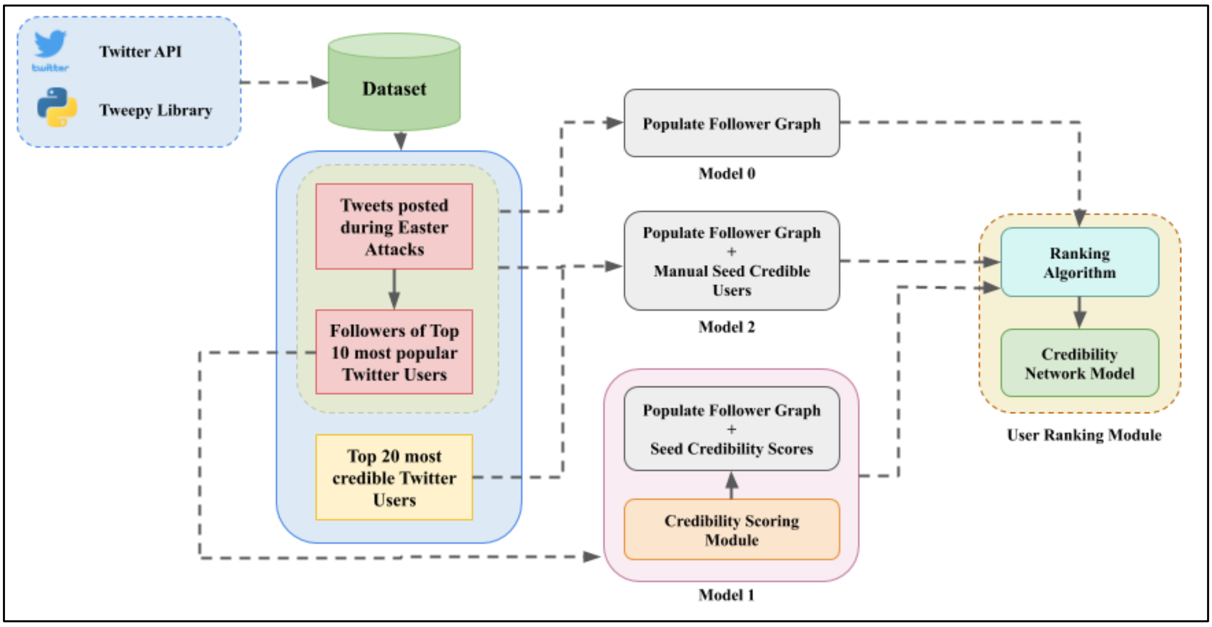

Figure 3.1: Proposed design of the credibility ranking models.

As discussed in literature, it is observed that Twitter users often follow other users because they believe the updates from the user whom they follow are of interest to them [2]. It can be argued that in doing so, the users place a certain level of trust in the profile they opt to follow.

It Is also observed that humans are good at perceiving credibility because of their superior cognition and reasoning skills. Therefore, the system takes these facts and observations into consideration to build its second model. It does so by running Personalized PageRank algorithm, a variation of Google's PageRank, on the user graph and by seeding a select set of users as "credible" within the network. A survey is conducted among Sri Lankan Twitter community asking to participants to provide the top ten "most credible" Twitter users they follow to receive breaking news. The top 20 most credible users are selected from the survey results and are seeded as "credible" users in the network by providing them with an initial weight. The system will then run Personalized PageRank algorithm on the graph and provide a credibility network model, which is referred to as **Model 2**. The proposed design of this system is depicted in figure 3.1.

In summary, the proposed system will build three separate network models based on factors such as,

- Follower relationships,
- Credibility scores derived from an algorithm proposed,
- Community seeded credibility based on a survey result.

All three modules will generate influencer/credibility ratings for each user in the network using PageRank algorithm as base network analysis algorithm with some models running a variation of it with node weights introduced to it. Finally, the most accurate model is selected based on the evaluation results. Evaluation is to be conducted based on baselines from past work as well as from a survey conducted among active Twitter users within Sri Lankan Twitter community.

3.4 Twitter User Ranking - Model 0 (Popularity Based)

The user ranking model consists of three major modules. They are **Data Collection** module, **Pre-processing and Graph population** module and **User Ranking** module.

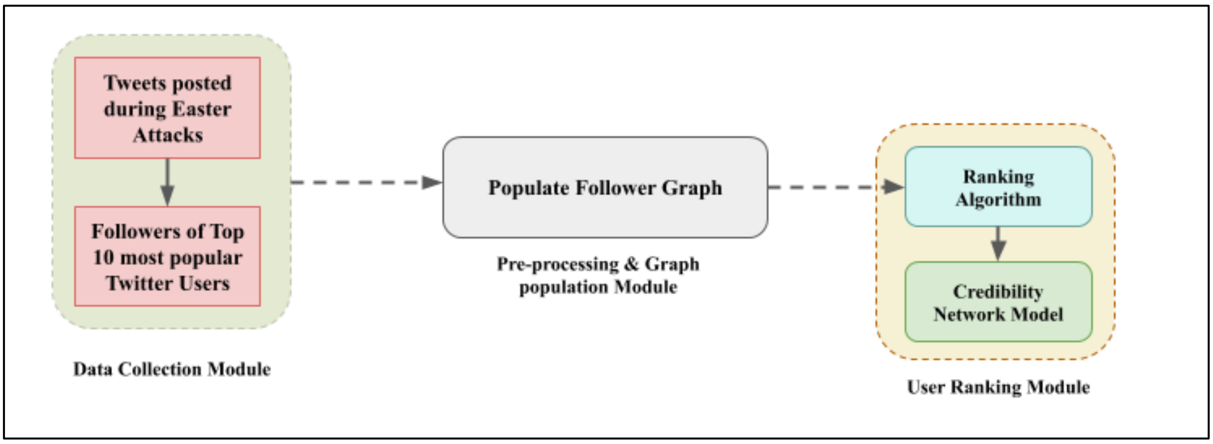

Figure 3.2: Proposed design of Model 0

3.4.1 Data Collection

One of the main objectives of this research is to provide a solution to assess the credibility of Twitter users within the Sri Lankan Twitter community during a breaking news situation with the Easter Attacks that happened in Sri Lanka during April 2019 as a sample study. For this purpose, a fresh dataset is collected as opposed to using a publicly available dataset. This is because there is no publicly available existing dataset collected covering the

Easter Attacks incident. The network analysis also requires a full universe for accuracy of its findings. Therefore, the adopted network analysis approach in this research requires users and their friends in the Twitter network up to two levels as shown in figure 3.3.

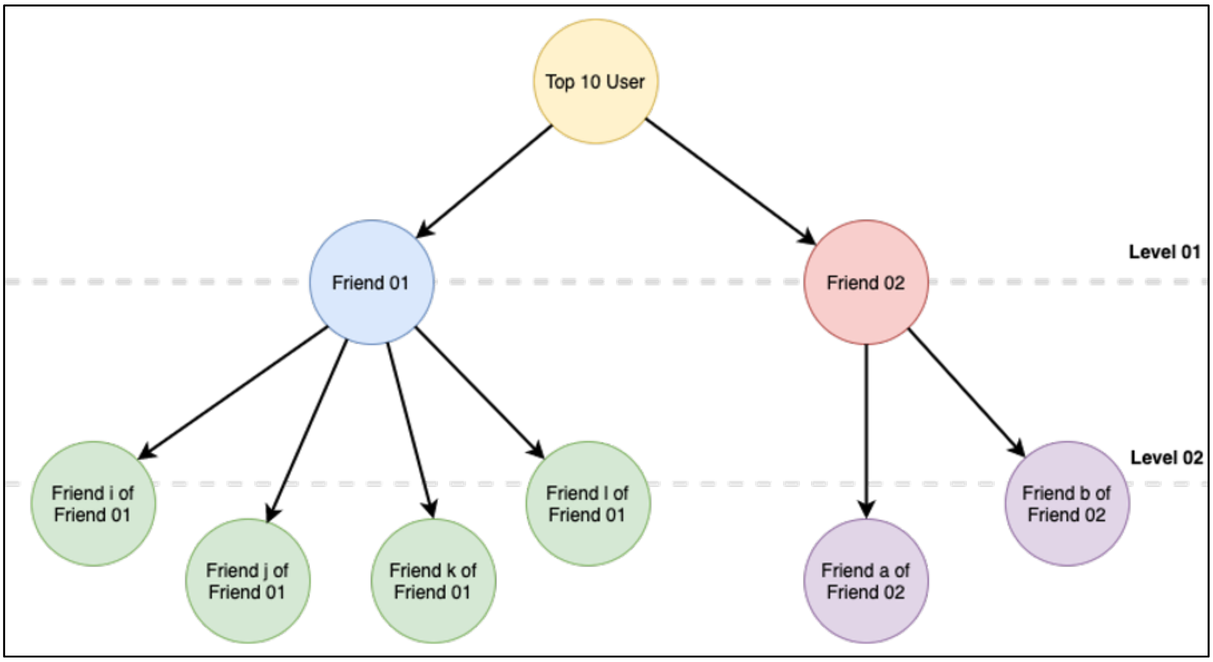

Figure 3.3: Follower relationships between a Top 10 user and her friends

Data collection phase for this research consists of two major phases. They are described in detail in sections 3.4.1.1 and 3.4.1.2.

3.4.1.1 Phase 01

In the first phase, a dataset (Dataset A) containing the tweets that were posted during 2019 April Easter Attacks in Sri Lanka are collected. This is done by extracting tweets that were posted from the 21st of April 2019, when the bomb attacks were carried out, to 28th of April 2019, with an expectation of recording a weeklong state of verified news, rumors, and all the chaos that ensued in the immediate aftermath.

The tweets posted under the hashtags, #EasterSundayAttacks, #EasterSundayAttacksLK, #EasterAttacks, #EasterAttacksLK, #EasterAttacksSL, #srilanka and #lka were collected.

Next, the above data set is refined to extract the unique number of users who engaged in posting tweets under the aforementioned hashtags. To generate a sample Twitter user network needed for analysis, this research considers a subset of 10 users from this unique list, who are considered as top 10 most influential users based in Sri Lanka who were active during the incident. For this purpose, the following criterion is used.

Location: Either Sri Lanka or Colombo, Sri Lanka.
Sorted: by the number of followers they have. (In descending order)

3.4.1.2 Phase 02

In the second phase, the top 10 users identified above are used to build the Twitter user graph. Accordingly, using Twitter's Public API, a dataset (Dataset B) containing the friends of each top 10 user were collected. For each friend of this top user, friends of them were also collected. This way, it is possible to generate a graph with a significant number of users which can be useful to conduct experiments. For each of these users, publicly available metadata are also collected. The list of metadata attributes collected can be found in Table V.

TABLE V:
LIST OF METADATA ATTRIBUTES COLLECTED FOR EACH USER

User ID	Age of the account
User's Screen name	User's bio
Number of followers	User's location
Number of friends	Verified status
Number of tweets	Number of lists

Twitter's public API has limitations when it comes to offering data retrieval services. Due to this, the number of friends collected for each friend of a Top 10 user is limited to 100 friends.

Data Collection Methods

Twitter Data: Twitter Search API, Twitter Streaming API with Tweepy, a library written in Python language [56]. The size of the datasets collected and their characteristics can be found in section 4.2.1.

3.4.2 Pre-processing and Graph Population Module

The dataset B collected above includes millions of users who are connected through their follower relationships all the way to the top 10 most influential users who were selected as the parent nodes of the follower graph. This dataset needs to be cleaned in order to filter the required user information to build the graph. Once the dataset is cleaned and stored in an appropriate data structure, the network graph can be populated using a python-based network analysis tool known as NetworkX.

3.4.3 User Ranking Module

As described previously, this particular model attempts to identify and rank popular users in a network through analyzing the social structure. For this purpose, this research employs the famous web page ranking algorithm developed by Google called **PageRank** algorithm.

In this approach, the PageRank algorithm takes into account the number of followers a person has on the network as well as the number of friends a particular user follows on the network. The algorithm will analyze the graph and provide ratings that obey power law as measurements of popularity of the nodes in the network. These values can be sorted in the descending order to obtain a ranking list of users from highest ranked to the lowest.

The next section discusses the concept behind PageRank algorithm and how it is used in the context of this research.

3.4.3.1 PageRank Algorithm

The PageRank algorithm was introduced by Larry Page, one of the co-founders of Google. The algorithm measures the relative importance of a web page compared to all other web pages found online. It was used entirely to improve Google's search engine indexing functionality at the time. In measuring the PageRank value, the number of links pointing to a web page is considered as an important factor [20].

The PageRank algorithm is devised around the concept that:
"more important websites are likely to receive more links from other websites" [20].

There are two underlying assumptions that make the PageRank algorithm. They are;

1. A number assumption.
2. A quality assumption.

The core concept behind this algorithm is the introduction of a random walk model. It assumes that a walker occupies a certain node at a specified time and traverses the graph through its edges. Hence, PageRank of a node in the network is thought of as the probability of a walker occupying each node when it jumps to a random node with a uniform probability. This allows PageRank algorithm to measure the significance of nodes based on the link structure of the graph [59].

This research uses PageRank algorithm to rank users on Twitter network. Twitter users follow other or are followed by other users on the network. The PageRank value of a given user in the network is calculated based on two assumptions [20][55].

Assumption 1:

A user is more important if he/she has more followers on the network. (number assumption)

Assumption 2:

A user receives more weight on the network if important users follow him/her. (quality assumption)

Based on the two assumptions made above, there are two steps to calculate a user's rank using PageRank algorithm:

1. **Initial step** – Create a network among the users by using "follower-friend" relationships that exist among them. Initially the PageRank value of every user is the same.

2. **Recursive calculation** – After a certain rounds each user will receive their final PageRank value. At the end of each iteration, all users receive a new PageRank score.

During each round of PageRank calculations, the following takes place:

1. Each user on the network distributes their PageRank value among the friends he/she follows. Therefore, each outgoing link receives an equal weight.

2. Each user gets an updated PageRank value which is an aggregation of all the weights from inbound links it receives from its followers [20].

Equation A defines the PageRank algorithm:

$$PR(u_i) = \frac{1-d}{N} + d \sum_{u_j \in M(u_i)} \frac{PR(u_i)}{L(u_{j)}}$$

(A)

where u_1, $u_2, \dots u_n$ are the users being considered, $M(u_j)$ is the set of users that share links with u_j, $L(u_j)$ is the number of friends of user u_j, and N is the total number of users. d is known as the damping factor that takes a value of 0.85 under most circumstances.

The results obtained from this model will help establish a baseline for identifying popular users in the Sri Lankan Twitter community who are enjoying a comfortable user following and therefore have the potential to exert more influence to propagate news within the network.

3.5 Twitter User Ranking - Model 1 (User Metadata Based)

This model comprises of four modules. They are **Data Collection** module, **Credibility Scoring** module, **Graph Population and Credibility Seeding** module and **User Ranking** module. It represents the novel approach taken in this research of attempting to seed credibility into a network based on a credibility score derived for each user in the network. This credibility score is calculated by assessing the user metadata that provides information about user's history and implicit clues as to their credibility within the Twitter network. Figure 3.4 shows the proposed design of Model 1.

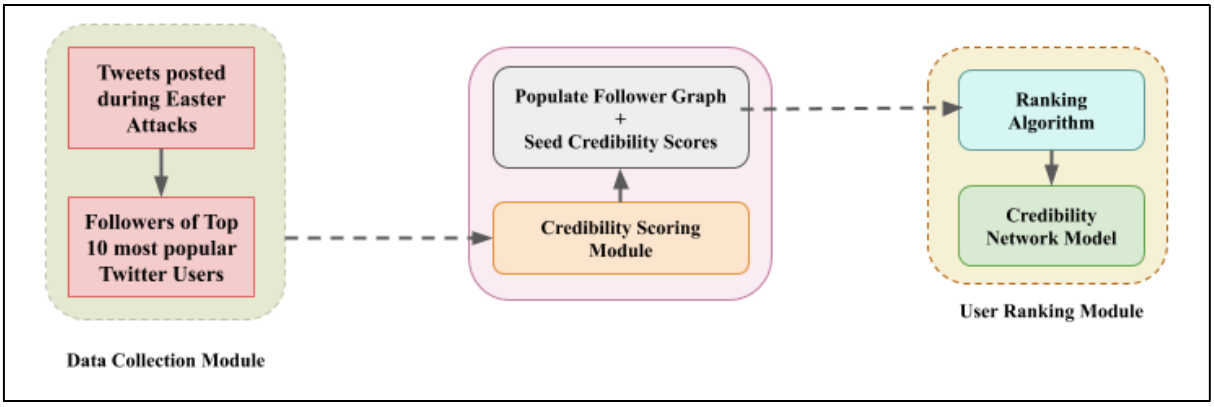

Figure 3.4: Proposed design of Model 1

3.5.1 Data Collection

This module makes use of Dataset B which was extracted for use in building Model 0. However, this model will extract user metadata from the initial dataset and store them separately to aid Credibility Scoring module which will be implemented to calculate user credibility rating.

3.5.2 Credibility Scoring Module

The credibility scoring module is tasked with calculating a credibility rating for each and every user in the Twitter user network. When calculating this credibility score, it takes into account a set of user credibility features identified in prior literature as important. The existing work carried out in this domain has suggested some best performing user credibility features as well as some other features with ordinary performance. In this research, a mix of such credibility features is used to derive a user credibility rating with an enhanced accuracy. This is in line with the research objective of obtaining a more fine-grained credibility rating for all users on the network, including users who are already deemed credible according to existing credibility assessment mechanisms in place.

TABLE VI:
LIST OF USER CREDIBILITY FEATURES USED IN CREDIBILITY SCORING

User credibility feature	Weightage
Number of followers	1.0
Number of friends	1.0
Number of tweets	1.0
Age of the account	1.0
Verified status	1.0
User’s Location	0.5
Number of lists	0.5
User’s bio	0.5

Table VI lists the complete set of user level credibility features used in this model. It also features the weights allocated for each feature to reflect their relative importance when calculating the final credibility score. A weight of 1.0 is assigned for features identified as more important while the rest of the features are allocated a weightage of 0.5. Prior research work found out that features such as number of followers, friends, tweets as well as the age of the account and the verified status offered by the Twitter platform are good indicators of a Twitter user’s credibility [19]. Thus, these features are given a weight 1.0. A user’s location, bio and the number of lists he/she is subscribed to were also considered as features that could be of importance albeit not as much effective as the features described earlier. Thus, in order to obtain a finely-grained rating, these features were also included in the experiment by giving a weightage of 0.5 to reflect their relative lesser importance as indicated in prior literature.

3.5.2.1 Credibility Scoring Matrix

In order to produce a credibility rating for a user, each and every feature listed in Table VI is evaluated separately and a feature-level credibility score will be assigned. For this purpose, a credibility scoring matrix is developed. This matrix will determine how the score is distributed for each feature a particular user will be evaluated against. The credibility scoring matrix is described in Table VII.

TABLE VII:
CREDIBILITY SCORING MATRIX

Feature	Credibility score					
	0	1	2	3	4	5
Follower ratio (No. of followers/ No. of friends)	< 0.5	0.5 – 1.0	1.0 – 2.0	2.0 – 5.0	5.0 – 10.0	> 10.0
No. of tweets	< 10	10 – 100	100 – 1000	1000 - 5000	5000 – 10000	> 10000
Account age	< 1 month	1 – 6 months	6 months – 2 years	2 – 5 years	5 – 10 years	> 10 years
Is verified	No	-	-	-	-	Yes
Has a Bio	No	-	-	-	-	Yes
Has a location	No	-	-	-	Yes	Sri Lanka/ Colombo
No. of lists	< 2	2 – 10	10 – 30	30 - 60	60 – 90	> 90

The credibility scoring matrix allocates a score of 5 to users who have their location set as either Sri Lanka, Colombo or a combination of the two keywords. This is done with the intention of giving more importance to users who are from Sri Lanka, since one of the objectives of the research is to identify credible users in Sri Lankan Twitter community. However, the existence of a bio and a location for any user provides a level of trust according existing literature. The user credibility scoring matrix is motivated from prior work in this domain [10, 19].

A few examples of credibility evaluated using the scoring matrix shown in Table VII is annexed in Appendix B.

3.5.2.2 Credibility Scoring Algorithm

The algorithm used for calculating a user's credibility score based on their credibility features is as follows:

Algorithm 1

U <- All Users
For each user in U
    Calculate individual credibility score for each user level credibility feature
    Calculate product of individual credibility feature where
    $$credibilityProduct_{feature} = credibilityScore_{feature} * credibilityWeight_{feature}$$
    Obtain average credibility score where
    $$credibilityScore_{average} = \frac{\sum credibilityProduct_{feature}}{\sum credibilityFeature}$$
    Generate user credibility rating R where
    R $= credibilityScore_{average}$ normalized to a value between a 0-1

The implementation details of the credibility scoring algorithm defined above is annexed in Appendix A.

3.5.3 Graph Population & Credibility Seeding Module

The user credibility scoring module will generate a set of users with their corresponding credibility rating. These values are injected into this credibility seeding module. This module will first populate the graph using network structure knowledge and then seed credibility for each and every user in the network through the credibility ratings it receives. This is done by generating a dictionary of nodes and their corresponding credibility ratings. This dictionary is provided to user ranking module for ranking the users based on their network structure as well as credibility.

3.5.4 User Ranking Module

In this model, the network analysis is performed by personalizing the same PageRank algorithm that is used in Model 0. PageRank algorithm can be influenced to obtain different results based on a personalization vector which can be applied to the default PageRank algorithm. This is known as **Personalized PageRank (PPR)** algorithm and the details about this version of the algorithm is discussed in the next section. The personalization vector generated by the credibility scoring module — which is detailed in section 3.5.2 — consists of all users in the network with their credibility ratings derived by credibility scoring module. The results generated from PPR is filtered and sorted to obtain a list of credible Twitter users in the populated network.

3.5.4.1 Personalized PageRank Algorithm

Personalized PageRank (PPR) is similar to original PageRank algorithm, in which it is also defined using a random walker model. However, it assumes that a traversing walker keeps jumping to a set of pre-defined nodes, instead of all nodes. Therefore, in PPR, the result is skewed towards these restarting nodes [59]. Thus, it can be said that PPR is a measure of the closeness of each node in the graph to the restarting nodes. A node that receives a higher PPR score can be considered as a node that is close to the restarting nodes which are established as important in the network. The Personalized PageRank can be defined as an update of original PageRank algorithm as shown in equation B:

$$PPR(u_i) = \frac{1-d}{N} + d \sum_{u_j \in M(u_i)} \frac{PPR(u_j)}{L(u_j)} W(u_j)$$

(B)

where, $W(u_j)$ is the normalized weight factor computed for user u_j by applying link analysis.

3.6 Twitter User Ranking - Model 2 (Community Seeded)

This model consists of three modules. The **Data Collection** module, **Credibility Seeding** module and the **User Ranking** module. It is the first attempt at introducing perceived credibility as a feature into the follower network to analyze and extract a list of users who are ranked based on their credibility. In this approach, an initial credibility rating is seeded into a subset of Twitter users chosen as credible based on results from an online survey conducted in which participants are from Sri Lankan Twitter community.

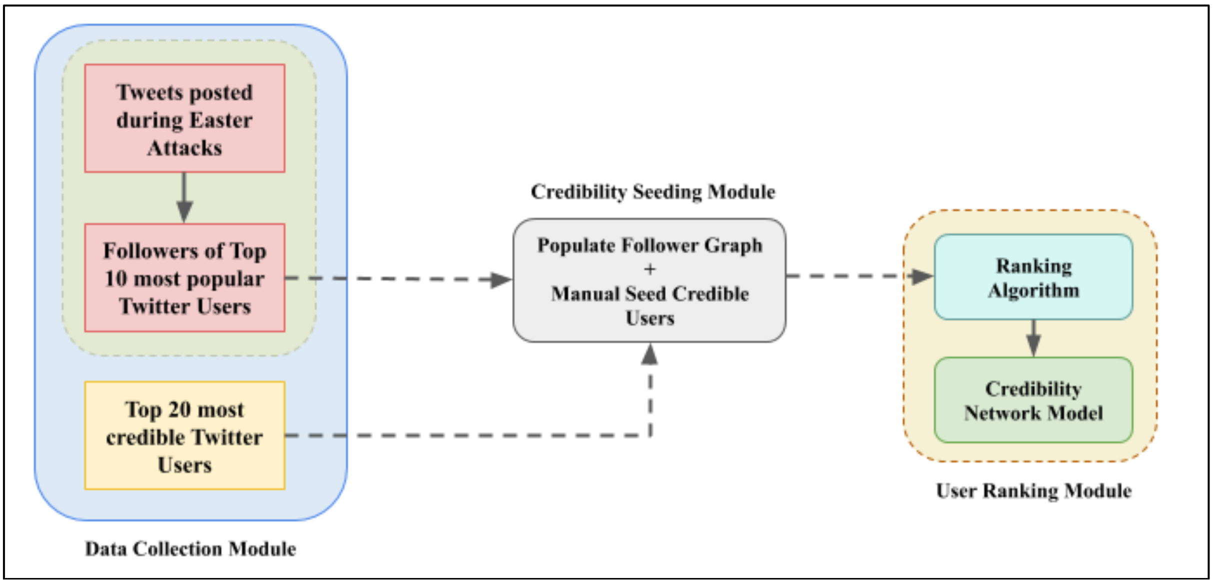

Figure 3.5: Proposed Design of Model 2

The idea behind this approach is to take advantage of the humans' ability to give credibility judgements using their superior reasoning skills.

3.6.1 Data Collection

This module makes use of the datasets collected in Model 0 to populate the user graph. However, this research will conduct a survey among Sri Lankan Twitter users to gather a list of credible users.

3.6.1.1 Survey to gather Credible Users

To obtain a list of credible Twitter users on the network, a survey is conducted among Sri Lankan Twitter community. All participants are asked to provide a list of top 10 Twitter personalities in Sri Lanka they follow for receiving reliable news updates during a breaking news situation.

A select number of user profiles which are considered as highly reliable by the survey respondents is used as credible users in the experiments.

3.6.2 Community Seeded Credibility Ranking Module

As described in previous section, this approach banks on seeding a set of highly credible Twitter users to the network in order to analyze and generate credibility ratings for every user in it. An initial node weight of order of dimension is assigned to each of these users in the network based on a weight allocation matrix which takes the number of confidence votes received by these identified credible users in the survey into consideration. The weight matrix used for weight allocation is found in Table XII of section 4.5.

The proposed model is built on the assumption that people tend to place a certain level of trust in users they opt to follow on Twitter. Similar to Model 1, this model also uses the Personalized PageRank algorithm to analyze the network. As such, it can be argued that the nodes with higher PPR scores are credible users on the network whom these seeded credible users share links with. It is possible to build up a network of credible users in this manner.

The community seeded credibility ranking module will take a populated network graph, a list of credible users and the corresponding weights allocated for these users. It will then perform PPR on the graph and provide a list of PPR scores obtained by each node in the graph. This output can be analyzed to extract credibility ranking among users in the network.

3.7 Implementation of the Proposed System

This research proposed three user ranking models based on popularity and credibility of a user within the network to address the research problem. The proposed solution can help identify the credibility of users on Twitter easily based on the ranking or a rating value they receive after credibility analysis. The content they author can be considered as credible or not depending on their credibility ranking within the community. Our assumption is that higher the credibility ranking of a user, more credible the content they post during breaking news situations. This section discusses the introduction of a proper implementation method to build the proposed ranking models in the architecture defined in the previous section. The architecture of the system comprises multiple important components. Beginning from data collection, data refining and analysis to building the three user ranking models, this section lays out a detailed overview of selecting the tools used for information retrieval, data cleansing techniques, selecting a network analysis tool which can run the chosen PageRank algorithm on Twitter user network.

3.7.1 Selection of Programming Languages, APIs and Other Tools

3.7.1.1 Python

Python is a general purpose interpreted, object-oriented and high-level programming language created by Guido van Rossum from National Research Institute for Mathematics and Computer Science in the Netherlands [60]. It is chosen as our preferred programming language to implement the proposed system due to its simplicity, support for functional programming on top of object oriented programming and also for the large collection of in-built libraries that makes data retrieval and pre-processing easier compared to other similar high-level programming languages which run on heavy boilerplate code.

3.7.1.2 Twitter's Standard Search API

Since it was decided that this research would be conducted on data obtained by Twitter platform, a research into selecting a data source for retrieving the required data was conducted. It was found out that Twitter provides free access to its REST API to collect its own data through a developer platform. The free tier of this API access program comes with limitations, which was considered as sufficient in this case to collect the data required.

Twitter's Standard Search API allows queries against the indices of recent or popular Tweets in their database. It behaves similarly to the Search UI features offered in Twitter's mobile or web clients, but with certain limitations. The Twitter Search API searches against a subset of recent Tweets published in the past 7 days. In this research context, this API is used to collect the users needed to populate the user network. A sample of 10 users with the highest following who tweeted during the Easter Attacks incident was selected. The friends of these users, and the friends of friends were retrieved using this API. The date range limitations imposed by Twitter's Standard API meant collecting data about the Easter Attacks incident which had happened months ago was not possible through this service. In order to circumvent this problem, a workaround solution provided by Jefferson Henrique is used as described in the next section.

3.7.1.3 Old Tweets Extractor Tool

Twitter's official API suffers from a time constraint limitation which prevents users from retrieving tweets older than seven days. In order to bypass this issue, an open sourced solution is presented by Jefferson Henrique, which makes use of Twitter's Search functionality implemented in its Web UI. Since this research banks on tweets that were posted in April 2019, it was decided to use this tool for the data collection purposes [61]. In addition to the tweets themselves, this tool provides a list of statistics related to the tweets, such as the number of retweets, likes it received, the timestamp of the tweet, the number of followers of the author etc. The designed experiments make use of these additional data to filter the dataset as described in section 3.7.2.

3.7.1.4 Tweepy

Tweepy is an easy-to-use Python based library for accessing Twitter's Standard API. [62] It acts as a high level wrapper interface which provides a full suite of Twitter data retrieval methods. Since it provides out-of-the-box support for collecting Twitter user data as required in this research for generating the user graph, it was decided to utilize this tool for data collection purposes.

3.7.1.5 NetworkX

NetworkX is a Python library for creation, manipulation and study of the structure dynamics and functions of complex networks [63]. It provides a number of standard graph algorithms built-in, including the PageRank algorithm proposed in the section 3.4.3.1 for performing network analysis. NetworkX has been used in prior research in network analysis for identifying influential nodes [64].

3.7.2 Data Collection and Preprocessing

This research aims to identify the credibility of Sri Lankan Twitter users who published content in the aftermath of the deadly Easter Attacks incident. Therefore, the author focused on filtering the tweets that are relevant to the incident. The tweets related to Sri Lanka are usually posted under the hashtags, #lk, #lka, #SriLanka, #SL, #srilanka among many other variations. In addition, it was observed that a considerable amount of tweets were also posted under the hashtags such as, #EasterSundayAttacks, #EasterAttacks, #EasterSundayAttackLK and other variations with the two keywords 'Easter Sunday'. It was also observed that the initial tweets circulated about this incident mostly used the standard hashtags which were variations of 'Sri Lanka' as stated above. Therefore, considering both these facts, the author carefully selected a set of hashtags for collecting and filtering the tweets as stated in section 3.4.1.

3.7.2.1 Tweets Dataset

The dataset was filtered to remove duplicates and extracted a unique list of users who posted tweets during the incident. From this list, the Top 10 most influential users based in Colombo, Sri Lanka were chosen, who were selected based on their number of followers. The data filtering method described is illustrated in Figure 3.6.

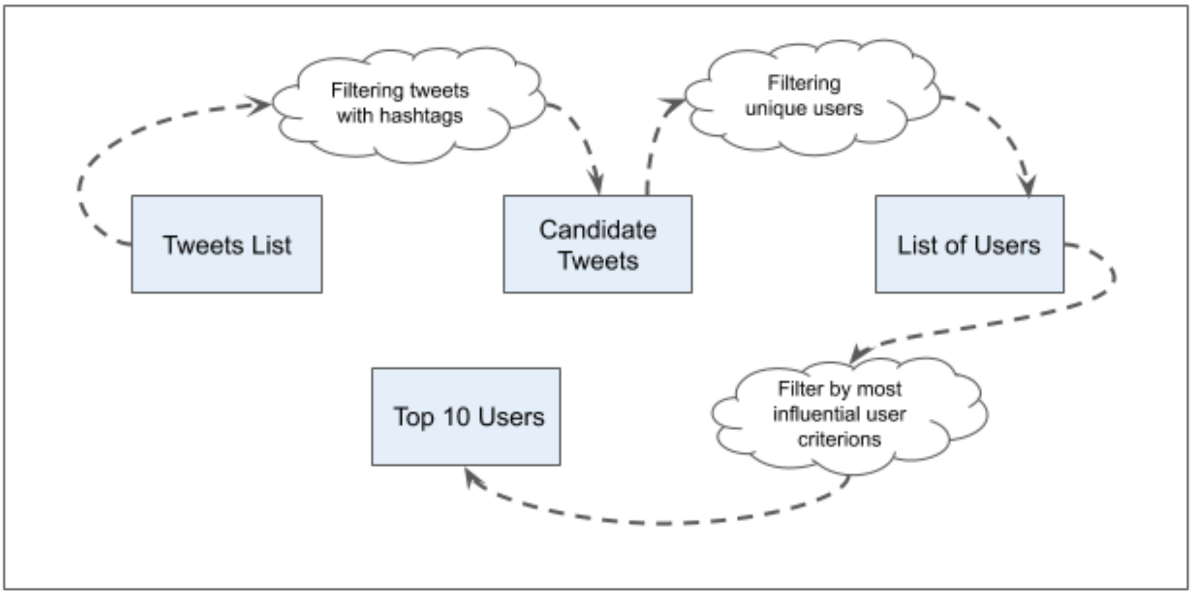

Figure 3.6: Tweets filtering process

3.7.2.2 Users Dataset

The Top 10 list of users is used to generate the user graph required in this research for network analysis. The author used the Tweepy library to get a list of friends each Top 10 user has on the network. Tweepy API provides the following functionality to achieve this objective:

API.friends([*id/user_id/screen_name*][, *cursor*][, *skip_status*][, *include_user_entities*])

This function returns a list of Twitter User objects who are being followed by the relevant Top 10 user.

Since this research proposed acquiring two depth levels of friends for each Top 10 user as stated in section 3.4.1, the above mentioned Tweepy API functionality was used to obtain friends of friends as well. Due to the limitations imposed on the number of service calls that can be made using Twitter's Standard API, it was decided to limit the number of users obtained as friends of friends to a maximum of 100 friends per every primary friend of a Top 10 user.

A Twitter User object returned from the above Tweepy API contains a list of metadata regarding the user. This research only considers a select set of metadata for the analysis purposes. Therefore, filters are applied to remove all the unnecessary metadata from the user object.

3.7.2.3 Credible Seeds Survey Data

In order to seed credibility into the user graph in the proposed Model 2 as stated in section 3.6, the author conducted a survey among the Sri Lankan Twitter community. The participants were asked to nominate their Top 10 most preferred Twitter profiles for obtaining credible news during breaking news situations. The survey responses were cleaned and filtered for establishing clarity and validity of the data . Further filtering was performed to extract a unique list of Twitter user profiles and counted the votes of confidence received by each profile. After careful observation of the data, it was decided that the profiles which received more than three votes of confidence are considered as a candidate for being a credible user on the network upon whom the community can rely on for receiving credible information. A list of 20 such user profiles were extracted. The survey result filtering process is illustrated in Figure 3.7.

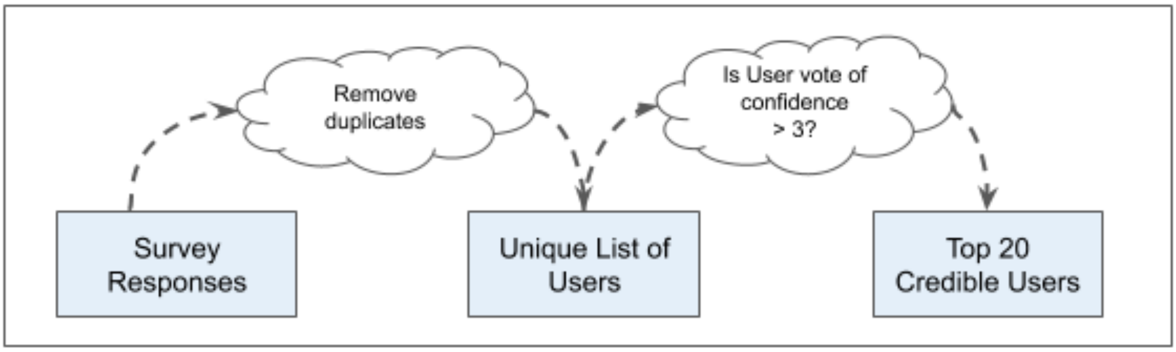

Figure 3.7: Credible users survey results filtering process

3.7.3 Implementation of Ranking Models

3.7.3.1 PageRank algorithm using NetworkX

In section 3.4.3.1, the author discussed in detail the PageRank algorithm and its proposed usage in identifying influential and credible people within the Twitter community. The NetworkX graph analysis tool is implemented as a Python library and houses a variety of graph analysis algorithms which can be used out-of-the-box. Therefore, NetworkX's implementation of this algorithm is used for conducting the analysis. The function definition of the PageRank algorithm implementation in NetworkX [65] is as follows:

pagerank(*G*, *alpha=0.85*, *personalization=None*, *max_iter=100*, *tol=1e-08*, *nstart=None*, *weight='weight'*)

The parameters used in the above function definition is described below:

Parameters:

G : graph ← A NetworkX graph. Undirected graphs are converted to a directed graph by adding two directed edges for each undirected edge.

alpha : float, optional ← Damping factor for PageRank. Default value is 0.85.

personalization : dict, optional ← This is the "personalization vector" consisting of a dictionary with a key for every node and a non-zero personalization value for each node. The default value is a uniform distribution. This personalization vector is manipulated in this research to obtain Personalized PageRank (PPR) values in Model 2 and Model 1.

max_iter : integer, optional ← Maximum number of iterations that will be performed.

tol : float, optional ← Maximum tolerated error value. The PageRank algorithm will stop once the sum of the error values of all nodes is below this value.

nstart : dictionary, optional ← Starting value of PageRank iteration for each node.

weight : key, optional ← Edge data key to use as a weight. If no weight specified, weights are set to 1.

3.7.3.2 Format of Input Data

In this implementation of the proposed system, the network user relationship data are stored using a nested JSON object which depicts a tree structure. JavaScript Object Notation (JSON) is a standard format used for representing structured data based on JavaScript object syntax [66]. This notation is commonly used in web applications for data transmission. Due to its organized key/value based structure and the author's familiarity with its notation, it was decided to store the user data; both individual metadata and the relationship data between users in the network, in JSON formatted files. A sample JSON data structure of relationships between users is shown in Figure 3.8:


```
{
   "friends":[
      {
         "friends":[
            {
               "screen_name":"WebGossiplk",
               "user_id":"3283337208"
            },
            {
               "screen_name":"Cinema_srilanka",
               "user_id":"949185210762973184"
            },
            {
               "screen_name":"adaderanabizsin",
               "user_id":"1031859864044093440"
            },
            {
               "screen_name":"adaderana_biz",
               "user_id":"1032112163559624704"
            },
            {
               "screen_name":"FMDeranaRadio",
               "user_id":"399287824"
            },
            {
               "screen_name":"tvderana",
               "user_id":"221391483"
            },
            {
               "screen_name":"adaderana",
               "user_id":"176337215"
            },
            {
               "screen_name":"AdaDerana_24",
               "user_id":"1081474421779456001"
            }
         ],
         "user_id":"183904448",
         "screen_name":"adaderanasin"
      }
   ],
   "user":"adaderana"
}
```


Figure 3.8: Sample JSON structure of User relationships

3.7.3.3 Output of the ranking model

The PageRank algorithm run on NetworkX initially returns a dictionary of nodes with PageRank as value. This list is sorted in descending order of the PageRank value to obtain a list of rankings from most popular/credible to least popular/credible depending on the model. The list is then filtered against the unique list of users who participated in discourse during the Easter Attacks incident. A list of rankings is obtained for each ranking model in a similar manner. The results are carefully analyzed and interpreted as discussed in the next chapter.

3.8 Summary

This chapter laid out the complete methodology followed by this research to achieve its aims and goals. As such, it started off by presenting the exact problem to which a solution is devised through the planned methodology. It then justified the selection of proposed research methodology and research design. The exact approach to the solution is then discussed in detail through the subsequent sections. There are three planned user ranking models proposed which will attempt to rank Twitter user credibility using three different approaches. This research will then evaluate the results from these models to obtain the best performing credibility network model. The algorithms that will be used to achieve this are discussed in detail. It then discusses the implementation of the proposed solution in detail and justifies the selection of tools for building it. The next chapter will present the results from the experiments and attempt to interpret them. It will also critically evaluate the study carried out.

Chapter 4 - Results and Evaluation

4.1 Introduction

The purpose of this chapter is to summarize the results obtained by the experiments designed and implemented according to the research methodology outlined in Chapter 3. The proposed research discussed three user ranking models — two of which are credibility based user ranking models — to identify credible users among the Sri Lanka based Twitter community. This chapter attempts to interpret and critically analyse the results obtained in each model and to evaluate the said ranking models for their performance. The chapter begins by laying out the results obtained from the data collection phase as described in section 3.4.1. It then presents the results obtained by each user ranking model along with the author's observations about the results. It then lays out the proposed evaluation procedure in detail. Based on the evaluation methods proposed, the models are evaluated and the results are critically analysed to identify the best performing models. A discussion on evaluation results will follow.

4.2 Data Collection

Based on the system architecture defined in section 3.3.1, the data collection required for credibility analysis is done in two phases. In the first phase, tweets related to Easter Attacks incidents are collected. In the second phase, friends and friends of friends for a list of users identified as top most influential, who tweeted during the incident are collected. Table VIII lists a few instances of tweets collected in the first phase.

4.2.1 Data Collection Results

A total of 117,704 tweets were collected in phase one. A total of 2,184,563 unique users and their metadata were collected in phase two. Figure 4.1 shows the distribution of collected tweets by the hashtags.

TABLE VIII:
FEW INSTANCES OF TWEETS COLLECTED UNDER EASTER ATTACKS HASHTAGS

User	Timestamp	Tweet
@Jayashantha	9:20 AM April 21, 2019	An Explosion reported at the #Kochchikadechurch #lka #srilanka
@NewsNow360	9:31 AM April 21, 2019	#SriLanka: A second explosion was reported at the St. Sebastian's Church in Katuwapitiya, Negombo a while ago. Few casualties reported from the incident. #lka
@munza14	10:15 AM, April 21, 2019	Explosions at St. Anthony's Church, Kochchikade, several other churches in Negombo & Batticaloa and at Shangri-La Hotel, Colombo as well as Cinnamon Grand.#LKA #SriLanka OH GOD! WHAT JUST HAPPENED!
@sheronrip1	2:55 PM, April 21, 2019	#8th Explosion in Dematagoda, Sri Lanka. #Facebook and #whatsapp are down #LKA #SriLanka #SriLankaBlast #SriLankaExplosions #ColomboBlast
@newsradiolk	2:10 PM, April 24, 2019	A controlled explosion has been carried out by the Bomb Disposal Squad on an unattended motorbike in Gaspaha Junction, Pettah. Police say no explosives were found. #SriLanka #lka #EasterSundayAttackLK
@colombogazette	6:37 PM, April 26, 2019	House to house search operations launched http://colombogazette.com/2019/04/26/house-to-house-search-operations-launched/ via @colombogazette #Srilanka #lka #EasterSundayattacksSL #EasterSundayattacksLK #EasterSundayattacks

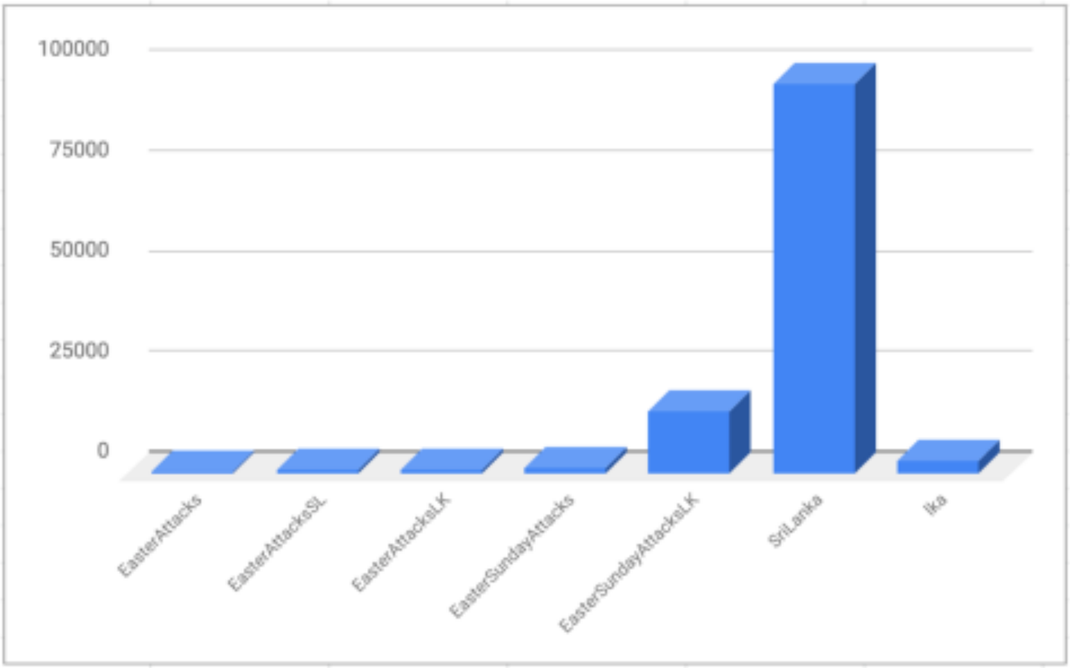

Figure 4.1: Distribution of collected tweets by hashtags

A majority of tweets were posted under the hashtag #srilanka, followed by #EasterSundayAttacksLK. Table IX lists the top ten most influential users identified based on their number of followers.

TABLE IX:
TOP 10 INFLUENTIAL USERS BASED ON FOLLOWER COUNT

User ID	Screen Name	No. of Followers
637463766	Almashoora	61,662
3032260688	MangalaLK	119,045
91046596	USEmbSL	122,782
2448362660	IamDimuth	176,449
1356593833	RajapaksaNamal	326,409
1968865952	HarshadeSilvaMP	263,445
66329707	Dailymirror_SL	436,998
41786801	SriLankaTweet	97,574
176337215	adaderana	401,511
554917917	Jino_Offl	142,214

The user list mentioned in Table IX is used to extract their friends' networks and use it to populate the user graph needed for analysis. The top10 list comprises mainly of reputed journalists and news outlets in Sri Lanka, local politicians and other celebrities.

4.2.2 Survey Results

A survey was conducted among Sri Lankan Twitter users to identify highly credible users in the network as stated in section 3.7.2.3. This survey yielded 95 responses, which were refined to extract the top 20 most credible users for receiving breaking news on Sri Lankan Twitter. Figure 4.2 depicts the top 20 most credible Twitter users and the votes of confidence they received according to the survey result:

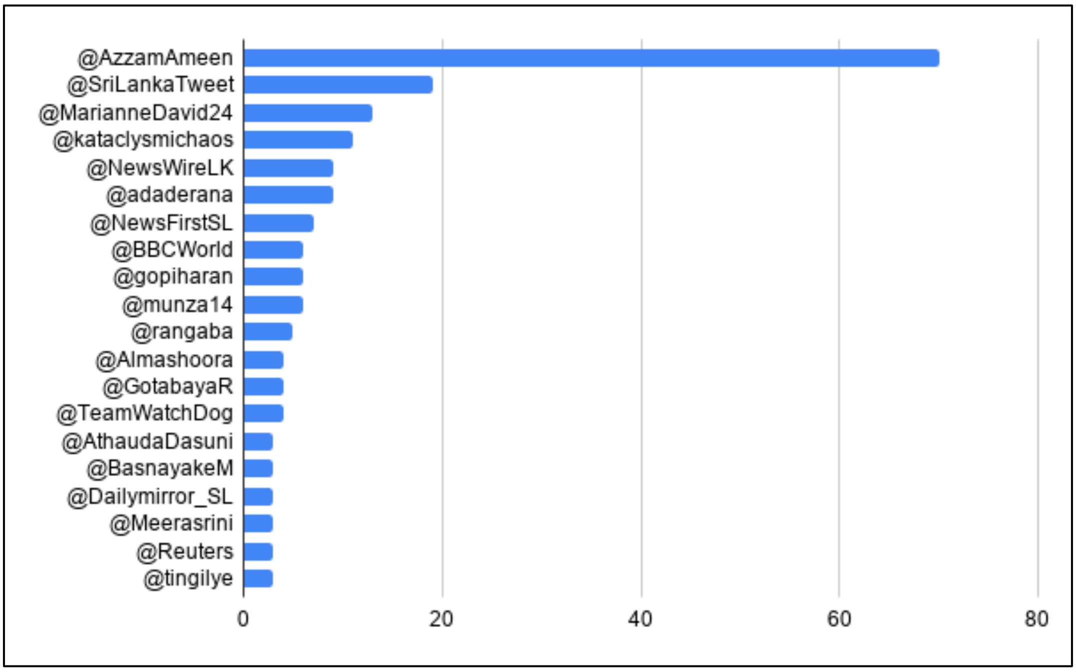

Figure 4.2: Top 20 credible users from survey

The list of highly credible Twitter profiles consists mostly of well-known journalists in Sri Lanka and news media outlets, both local and international. Figure 4.2 show that Azzam Ameen, A journalist from Newswire.lk— who has also worked as Sri Lanka's news correspondent for BBC News — is considered as the most credible Twitter profile for obtaining credible news in the local arena, by an overwhelming number of votes of confidence. The results also suggest that a majority of users subscribe to Sri Lanka based Twitter news correspondents to receive credible news updates. This indicates that if a Twitter profile is of a journalist or a news media person — who is ideally from the reader's country of origin — it will help the reader to trust the content being shared by these profiles more.

The users identified as credible from the survey are given a credibility weightage of order of dimension as stated in section 3.6.2, before Personalized PageRank algorithm — detailed in section 3.5.4.1 — is run on the user graph for building Model 2.

4.3 Model 0 (Popularity based) – Results and Analysis

The first of the three user ranking models was built based on the popularity of the users in the network. As detailed in section 3.4, PageRank algorithm is used to calculate a user's PageRank value — which can be thought of as a popularity rating — based on primarily the number of followers a user has on Twitter.

As stated in section 4.2.1, a full user graph consisting of more than two million unique Twitter users — all who can be traced back to having direct and indirect relationships with the top 10 influential users identified — is used to perform network analysis. After filtering the resulting dataset by the users who participated in Easter Attacks related discourse, a total of 12,394 users with their PageRank values were identified. They were sorted from highest to lowest ranked to identify the most popular users in the network. Table X lists the Top 20 most popular users in Sri Lankan Twitter along with their normalized PageRank values as calculated by Model 0.

TABLE X:
TOP 20 RANKED USERS FROM MODEL 0

Rank	Screen Name	PageRank	Rank	Screen Name	PageRank
1	SriLankaTweet	0.0000623881	11	HarshadeSilvaMP	0.0000102701
2	adaderana	0.0000226206	12	CNN	0.0000095824
3	GotabayaR	0.0000221499	13	adaderanasin	0.0000091819
4	PresRajapaksa	0.0000220117	14	cnnbrk	0.0000080613
5	AzzamAmeen	0.0000212155	15	chaturaalwis	0.0000080379
6	Dailymirror_SL	0.0000203038	16	MangalaLK	0.0000078222
7	RajapaksaNamal	0.0000164489	17	bbcsinhala	0.0000070561
8	BBCBreaking	0.0000107404	18	IamDimuth	0.0000066622
9	BBCWorld	0.0000104604	19	IslandCricket	0.0000065396
10	SriLanka	0.0000104575	20	irajonline	0.0000060004

It is observed that topmost popular users in Sri Lanka based Twitter are mostly reputed journalists and news media outlets. A number of government officials and politicians also rank higher in the list. Some government institutes, Non-Governmental Organizations also can be found in the top bracket for most popular Twitter profiles. This pattern can be observed in the rest of the ranked list as well.

The top ranked user who goes by 'SriLankaTweet' screen name on Twitter is a freelance journalist based in Sri Lanka and seems to be having a significantly higher PageRank value than the second ranked user profile on the list. This user had a friend count of 22,600 at the time of data collection. When collecting friends of friends as candidate users for populating the user network, it is possible that a large number of secondary friends came from this user's friends — which is around 16,000 friends higher than the next user with highest number of friends from the list of initial top ten influential users. Therefore, the PageRank algorithm has placed more relative importance to this node in the network. This is an indication of the algorithm's reliance on its user relationships as discussed in section 3.4.3.1. Thus, the unusually high PageRank score can be justified.

Figure 4.3 depicts an instance of the user network visualized with only 20% of the nodes present. A holistic visualization of the complete graph could not be rendered due to computational limitations.

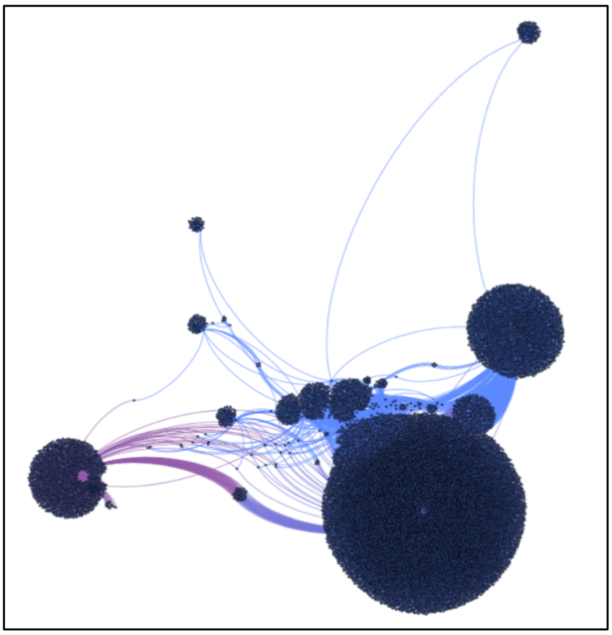

Figure 4.3: An instance of a user network analysis visualization

4.4 Model 1 (User Metadata based) – Results and Analysis

As discussed in section 3.5, this research proposed an algorithm to calculate a user's credibility based on user's metadata. The credibility score obtained for each user in the network was used as node weights for users when populating the network for analysis. This model used the calculated credibility score to manipulate the personalization vector of PageRank algorithm to extract credibility ranking of users on the network as discussed in section 3.5.2.2. The top 20 ranked list of credible Twitter profiles generated from Model 1 is found in Table XI.

It was observed that the results generated from Model 1 did not deviate significantly from the results generated by Model 0 and Model 2 as far as the top 20 lists generated by all three models are concerned. However, the results from Model 1 list showed deviations from the rankings generated by other models as the list rankings broke away from top and moved towards the mid-field. It was also observed that some of the journalists and media outlets touted as popular and credible by Models 0 and Model 2 received much lower rankings in Model 1.

TABLE XI:
TOP 20 CREDIBLE USERS FROM MODEL 1

Rank	Screen Name	PageRank	Rank	Screen Name	PageRank
1	SriLankaTweet	0.0000646125	11	HarshadeSilvaMP	0.0000129816
2	GotabayaR	0.0000248614	12	SriLanka	0.0000126819
3	adaderana	0.0000248449	13	CNN	0.000012375
4	PresRajapaksa	0.0000247233	14	adaderanasin	0.0000113252
5	AzzamAmeen	0.000023927	15	cnnbrk	0.0000111786
6	Dailymirror_SL	0.0000230153	16	tingilye	0.0000091903
7	RajapaksaNamal	0.0000191604	17	MangalaLK	0.0000105337
8	PresRajapaksa	0.0000220132	18	chaturaalwis	0.000010019
9	BBCBreaking	0.0000138576	19	IamDimuth	0.0000093737
10	BBCWorld	0.0000135777	20	IslandCricket	0.0000092511

It was also observed that the PageRank values for the user profiles had increased from the values they received by popularity-based Model 0, which showed the effect of credibility score-based node weights had on the positioning of a user within the network in this model.

4.5 Model 2 (Community seeded) – Results and Analysis

The community seeded credibility ranking module was an attempt at imparting credibility aspect into the Twitter user network as discussed in detail in section 3.6. A survey was conducted to identify a list of highly credible users on Sri Lankan Twitter. Based on the survey results, the selected set of credible users were given initial node weights. Table XII lists the weight matrix used to assign weights for the top 20 credible users.

TABLE XII:
WEIGHT MATRIX FOR SURVEY BASED CREDIBLE USERS

Vote Range	Weight Allocation
< 3	0.5
3 – 10	5
10 – 50	50
50+	500

The decision to allocate weights for credible users in this experiment was primarily influenced by the difference of votes obtained by top 20 highly credible users obtained from the survey result. The first ranked user in this list had obtained more than three times the votes garnered by the second ranked user. Therefore, it was decided to award weights to these users in an order of dimension for each votes range defined to better reflect the significance of the overwhelming confidence received by the top half of the list of highly credible list of Twitter users. The weights were adjusted a few times to reflect the survey result better and for the ranking model results to align closer with the reality.

Through the use of personalization vector of PageRank algorithm as described in section 3.5.4.1, these credible users were fed into the network and network analysis was performed. The top 20 credible users identified by Model 2 is listed in Table XIII.

It was observed that the user ‘AzzamAmeen’, who obtained the most votes from survey is ranked first in this credible list of users. It is true that the author fine-tuned the weightage allocation to put this user on the top considering the overwhelming confidence he received by the survey participants. Therefore, this result isn't significant by itself. However, it was also noted that a number of other credible users have obtained higher rankings among the top 20 most credible list of users identified by Model 2. A quick comparison between the top 20 lists generated by Model 0 and Model 2 shows that there are new entries to the list of credible users as identified by Model 2. This was an important observation since the manually added weights did not affect the rankings overwhelmingly and showed that such adjustments could provide ranking results that could reflect better credibility among existing popularity based ranking models. It was also noticed the rankings of users who were present in top 20 lists of both Model 0 and Model 2, were different.

TABLE XIII:
TOP 20 CREDIBLE USERS FROM MODEL 2

Rank	Screen Name	PageRank	Rank	Screen Name	PageRank
1	AzzamAmeen	0.0004758831	11	BBCBreaking	0.0000107418
2	SriLankaTweet	0.0001074466	12	SriLanka	0.000010459
3	kataclysmichaos	0.000050629	13	HarshadeSilvaMP	0.0000102715
4	MarianneDavid24	0.0000491636	14	Reuters	0.0000097211
5	adaderana	0.0000267181	15	CNN	0.0000095838
6	GotabayaR	0.0000262474	16	tingilye	0.0000091903
7	Dailymirror_SL	0.0000244013	17	adaderanasin	0.0000091833
8	PresRajapaksa	0.0000220132	18	cnnbrk	0.0000080628
9	RajapaksaNamal	0.0000164504	19	chaturaalwis	0.0000080393
10	BBCWorld	0.0000145579	20	MangalaLK	0.0000078236

4.6 Evaluation

Evaluation process of a research involves critically examining the developed research prototype. For this purpose, collection and analysis of information about the prototype, including the activities, features, and outcomes of the research is important. The purpose of evaluation is to make discernments about the solution, to improve its effectiveness, and/or to derive conclusions [67].

A properly planned evaluation enables the researcher to demonstrate the success or progress of the research solution. The information that will be collected will allow the researcher to better communicate the solution prototype's effectiveness and impact to the end users, and also to the research community for any current and potential future enhancements. In the subsequent sections the author prepares a concise evaluation plan that will test the functionality and the effectiveness of the implemented research prototype empirically.

4.6.1 Evaluation Method

In devising an evaluation method, the author took into consideration the research problem and the exact approach taken to answer the research question. This research attempted to determine the credibility of Twitter users based on a hybrid method incorporating Twitter user graph analysis with multiple credibility evaluation mechanisms.

As discussed in the literature review, prior work in this research area is dominated by automated solutions that use labelled data and machine learning. However, in this research, the author opted to use a network analysis-based method to rank users in a network for their credibility. The credibility aspect of this research was introduced in two different ways.

A Twitter user's metadata were used to calculate a credibility score using a novel algorithm in section 3.5.2.2. This credibility score was fed into the Twitter network for graph analysis as described in section 4.4 to identify credible users on the network. In section 4.5, the author conducted experiments to impart credibility into a Twitter network by seeding a set of users into the network, who were selected as credible by the community.

Due to the novelty in this approach, and the use of a fresh dataset for analysis, the evaluation of this research project was not as straight-forward as it seemed at first. The lack of similar research in academia which follow a network analysis-based approach for credibility detection, and the complexities that arose in porting algorithms and systems developed in prior research to match with the work carried out in this research meant there was no fairly easy approach for evaluating the models against any recent work in this domain. Therefore, the author looked into literature about conducting manual human evaluations to test the effectiveness of the models built for research carried out in this domain in the past.

Castillo et al. [8] in their 2011 research about information credibility on Twitter adopted a manual method for evaluating the credibility assessment solution they built. They recruited a group of evaluators on Amazon Mechanical Turk and asked them to rank a set of tweets under four credibility ranking levels. This method proved to be successful in determining the effectiveness of their solution. Motivated by their work, the author planned a similar manual evaluation approach for this research.

Accordingly, the author devised an experiment to conduct a survey among the Sri Lanka based Twitter community. The participants were given a set of Twitter user profiles and asked to rank these users based on the credibility perceived by the evaluator on each user for receiving credible news updates during breaking news situations. The results of this survey were evaluated against the results obtained from credibility ranking models described in sections 4.3, 4.4 and 4.5. The Spearman's rank correlation analysis was performed for results obtained from the models against the results from the survey. Based on results, a model which demonstrated most accurate credibility rankings against the survey result was identified.

4.6.2 Selection of Evaluation Method

In the previous section, it was pointed out that there is a dearth in existing research which follow similar approach to solve the research problem. Thus, the use of existing research work to evaluate the research outcomes is difficult. Prior research work in this research domain has utilized ranking based manual user evaluations to good effect. It is also observed that humans are good at ranking than rating goods or people. For an instance, humans find ranking a given set of products based on a criterion than giving them a rating, such as on a Likert scale. This is because rating requires respondents to consider how they feel about individual items, and to measure positive or negative response to a statement, whereas ranking involves comparing individual elements to each other [68]. Although both these methods show trade-offs between them, the author opted for the ranking option since it aligned better with the evaluation goals of this research.

4.6.3 Manual Evaluation Survey

The author decided to choose the result list from popularity-based Model 0 to select users to be used in manual evaluation. PageRank based analysis for important node identification has been researched in the past and has been proven as a good indicator of popularity [20]. As such, the author considered the top 75 most popular Twitter users from Model 0 to select candidate users to be listed in the survey.

In choosing Twitter user profiles for evaluation, the author carefully picked a mixture of profiles which enjoyed higher ranks in the results consistently across ranking models, as well as profiles which did not appear next to each other on credibility ranking result lists. This was to avoid confusion and to aid evaluators to make their credibility judgements with relative ease. Thus, user profiles with a fair amount of credibility value difference between them were picked.

It was decided to use 12 user profiles for the evaluation experiment. The reason for selecting only 12 user profiles was to make the evaluation process as less cumbersome as possible for the participants. The survey was conducted through an online survey platform. It received 32 responses. Figure 4.4 shows the average rankings received by the user profiles from human evaluators.

The credibility rankings were obtained from Model 0, Model 1 and Model 2 for the 12 user profiles used in manual evaluation.

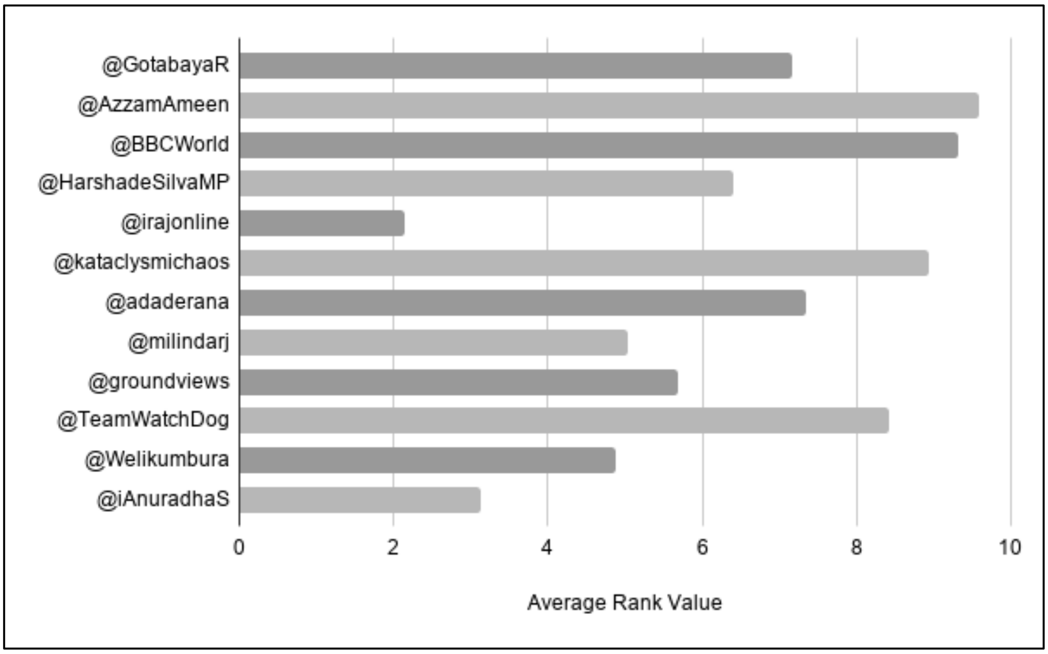

Figure 4.4: Evaluation survey results

4.6.4 Correlation Analysis

Since this research opted for a rank comparison method to evaluate the accuracy of its credibility ranking models, the author had to choose a suitable rank correlation coefficient for this purpose. The use of measures of correlation is usually observed in studies that observe relationships between two variables. These studies assume that neither variable is functionally dependent upon the other [69]. A correlation coefficient is a quantitative measure of the strength of the correspondence.

4.6.5 Spearman's Rank Correlation Coefficient

The Spearman correlation is used to evaluate monotonic relationships between two continuous or ordinal variables. A monotonic relationship is where the variables may change together, but not necessarily in a persistent rate. It is often used to evaluate relationships between ordinal variables and is based on the ranked values for each variable than the raw values of data. Figure 4.5 represents examples from monotonic and non-monotonic relationships.

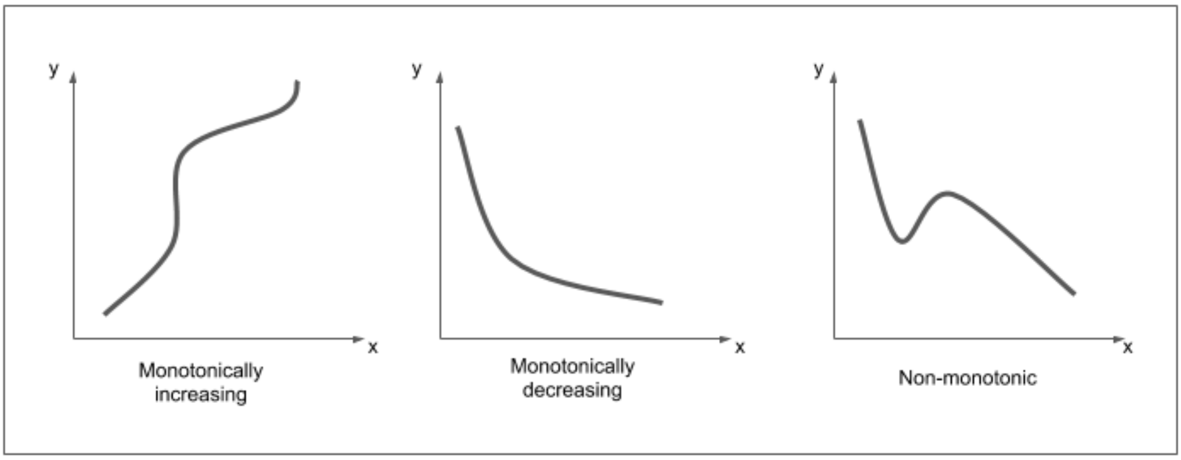

Figure 4.5: Monotonic and non-monotonic relationships

4.6.5.1 Definition of Spearman's Rank Correlation Coefficient

The Spearman's rank correlation coefficient (ρ) is the non-parametric version of the Pearson correlation coefficient.

The formula for ρ when no tied rank is present:

$$\rho = 1 - \frac{6 \sum {D_i}^2}{n(n^2 - 1)}$$

(C)

ρ : Coefficient of rank correlation

D_i : Difference in ranks between paired values of X and Y

$$D_i = rg(X_i) - rg(Y_i)$$

n : Sample size

The Spearman's correlation coefficient ρ is constrained as follows:

$$-1 \leq \rho \leq 1$$

The closer ρ is to ± 1, the stronger the monotonic relationship between variables. Conversely, the closer ρ is to 0, the weaker the associations between the variables.

4.6.6 Evaluation of the Models

As discussed in Chapter 3, this study developed three user ranking models based on three factors: popularity, metadata-based credibility of users in the network, and a community seeded credibility. As observed in sections 4.3-4.5, each ranking model generated list of user rankings that were different from each other. Therefore, there was a need to evaluate these results from each model against the rankings obtained from a survey conducted among active users within Sri Lankan Twitter community. Thus, each credibility ranking model was evaluated with Spearman's rank correlation coefficient to detect and identify associations between the ranks from models and the survey results.

4.6.6.1 Evaluation of Rankings generated by Model 0 (Popularity based)

In order to evaluate Model 2 with the survey results, the results obtained from survey were first converted to ranks. The conversion of responses to average ranking values was done by using equation D.

$$\text{Average Rank of a User} = \frac{1}{n}\sum_{0}^{n} x_n w_n$$

(D)

Where n is the total response count, x is the response count for answer choice and w is the weight of ranked position.

Table XIV shows the ranking values obtained for the survey results using equation D. The average ranking values were converted to new ranks based on Spearman's correlation coefficient.

TABLE XIV:
CONVERSION OF SURVEY RESULTS TO RANKS

User	**Average Survey Ranking**	**New Rank (Y_i)**
@AzzamAmeen	9.59	1
@BBCWorld	9.31	2
@kataclysmichaos	8.94	3
@TeamWatchDog	8.41	4
@adaderana	7.34	5
@GotabayaR	7.16	6
@HarshadeSilvaMP	6.41	7
@groundviews	5.69	8
@milindarj	5.03	9
@Welikumbura	4.88	10
@iAnuradhaS	3.13	11
@irajonline	2.13	12

Secondly, the PageRank values generated for the 12 users in Model 0 were converted to ranks. The results of the conversion are reflected in Table XV.

TABLE XV:
CONVERSION OF MODEL 0 RESULTS TO RANKS

User	PPR value from Model 0	New Rank (X_i)
@adaderana	0.0000226206	1
@GotabayaR	0.0000221499	2
@AzzamAmeen	0.0000212155	3
@BBCWorld	0.0000104604	4
@HarshadeSilvaMP	0.0000102701	5
@irajonline	0.0000060004	6
@kataclysmichaos	0.0000055706	7
@milindarj	0.0000039677	8
@groundviews	0.0000039326	9
@TeamWatchDog	0.000003039	10
@iAnuradhaS	0.000002636	11
@Welikumbura	0.0000022733	12

By applying equation C,

$$\rho = 1 - \frac{6\sum D_i{}^2}{n(n^2-1)}$$

$$\rho = 1 - \frac{6\times 138}{12(12^2-1)}$$ where n = 12 (evaluated number of users is 12)

$$\boldsymbol{\rho = 0.517}$$

The derived Spearman’s rank correlation coefficient for Model 0 when evaluated against the survey result is 0.517. This is a positive correlation and indicates a significant association between the rankings.

4.6.6.2 Evaluation of Rankings generated by Model 1 (Metadata based credibility)

In order to evaluate the rankings generated by Model 1, the personalized PageRank values obtained from the model are converted to ranks. Table XVI depicts the conversion of these values to ranks.

TABLE XVI:
CONVERSION OF MODEL 1 RESULTS TO RANKS

User	PPR value from Model 1	New Rank (X_i)
@GotabayaR	0.0000221499	1
@adaderana	0.0000248449	2
@AzzamAmeen	0.000023927	3
@BBCWorld	0.0000135777	4
@HarshadeSilvaMP	0.0000129816	5
@irajonline	0.0000086304	6
@kataclysmichaos	0.0000077949	7
@milindarj	0.0000066792	8
@groundviews	0.000006157	9
@TeamWatchDog	0.0000050201	10
@Welikumbura	0.0000043355	11
@iAnuradhaS	0.0000041299	12

The Spearman's rank correlation coefficient ρ is calculated by applying the equation C:

$$\rho = 1 - \frac{6\sum D_i{}^2}{n(n^2-1)}$$

$$\rho = 1 - \frac{6\times138}{12(12^2-1)}$$ where n = 12 (evaluated number of users is 12)

$$\boldsymbol{\rho = 0.517}$$

The derived Spearman's rank correlation coefficient for Model 1 when evaluated against the survey results was 0.517. This was a positive correlation. It was also interesting to note that the Spearman's rank correlation coefficient of Model 0 and Model 1 were exactly the same.

4.6.6.3 Evaluation of Rankings generated by Model 2 (Community seeded credibility)

To compare the ranking correlations between the survey result and the values generated by Model 2, the personalized PageRank values obtained by the 12 users in Model 2 are converted to ranks. The results of the conversion are reflected in Table XVII.

TABLE XVII:
CONVERSION OF MODEL 2 RESULTS TO RANKS

User	PPR value from Model 2	New Rank (X_i)
@AzzamAmeen	0.0004758831	1
@kataclysmichaos	0.000050629	2
@adaderana	0.0000267181	3
@GotabayaR	0.0000262474	4
@BBCWorld	0.0000145579	5
@HarshadeSilvaMP	0.0000102715	6
@TeamWatchDog	0.0000071366	7
@irajonline	0.0000060018	8
@milindarj	0.0000039691	9
@groundviews	0.0000039341	10
@iAnuradhaS	0.0000026374	11
@Welikumbura	0.0000022748	12

By applying equation C,

$\rho = 1 - \frac{6 \sum D_i^2}{n(n^2-1)}$

$\rho = 1 - \frac{6\times 52}{12(12^2-1)}$ where n = 12 (evaluated number of users is 12)

$\boldsymbol{\rho = 0.818}$

The derived Spearman's rank correlation coefficient for Model 2 when evaluated against the survey result was 0.818. This was a positive correlation which is closer to +1, which indicates a strong association between the rankings.

4.7 Comparison of the Ranking Models

As observed in the section 4.6, the association of credibility rankings with the results from manual survey conducted to evaluate the models showed positive correlations in all three models. This bode well for the accuracy for the models built. Table XVIII represents a summary of the Spearman's rank correlation coefficient (ρ) values obtained for the three user ranking models.

TABLE XVIII:
COMPARISON OF SPEARMAN'S CORRELATION COEFFICIENT BETWEEN MODELS

	Model 0 (Popularity based)	Model 1 (Metadata based)	Model 2 (Community seeded)
Spearman's rank correlation coefficient (ρ)	**0.517**	**0.517**	**0.818**

In section 4.6.5.1 it was learnt that the closer a Spearman's rank correlation coefficient (ρ) is to zero, the weaker the relationship between associated variables. However, the ranking models built in this study all showed positive correlations, with ρ value more than 0.5 for all models. This indicated acceptable levels of accuracy for the all three models concerned. Model 2 in particular, which was developed using a novel community seeded credibility approach as discussed in section 3.6, showed a high positive correlation of 0.818 which was highly encouraging. The popularity-based Model 0 and the user metadata-based Model 1 showed identical correlation coefficients of 0.517. These results all lead to the conclusion that the models can be used to detect credibility of Twitter users with high accuracy. Thus, Model 2 was selected as the best performing credibility ranking model out of the three ranking models.

4.8 Summary

This chapter presented and discussed in detail the results obtained through the experiments conducted in this study. This study attempted to build three different credibility ranking models to identify credible users on Sri Lanka based Twitter. Firstly, the results from the data collection phase were presented and discussed. Then, the credibility ranking results generated by each ranking model proposed in section 3.3 were presented and the results were critically reviewed. It then defined a comprehensive approach to evaluate the models using establish statistical analysis measurements. A survey was conducted among regularly active users in the Sri Lanka based Twitter to gather data required for a manual evaluation proposed. Spearman's rank coefficient of correlation was applied to compare the models with manual evaluation survey data. It was found out that all three ranking models show positive correlations with the test data. The community seeded credibility ranking model detailed in section 3.6 was identified as the best performing model. The next chapter presents the concluding remarks on this study.

Chapter 5 - Conclusion and Future Work

5.1 Introduction

The aim of this study was to provide a mechanism to rank the users on Sri Lankan Twitter based on their credibility in order for general users to obtain credible updates during high impact breaking news situations. Therefore, this study focused on determining the credibility of a Twitter user and to impart that credibility into the content they generated online. In this research, the concepts of credibility, credibility assessment techniques and credibility analysis approaches have been discussed in length. A survey of the existing credibility assessment and analysis techniques and approaches in prior academia has been conducted. Based on the findings, a conceptual framework has been developed and presented to solve the research problem as outlined throughout the dissertation. The approach taken was to perform network analysis to build three user credibility ranking models for identifying credible users on the Twitter network thereby preventing the propagation of misinformation during breaking news situations. The proposed models have been designed and developed successfully. A proper evaluation based on a survey conducted among active users on Sri Lanka based Twitter and a sound evaluating method was carried out to verify the effectiveness of the Credibility Ranking models built. The evaluation yielded interesting results which proved the suitability of the solution developed to address the research problem.

5.2 Conclusions about Research Objectives

In this study, three credibility ranking models to rank the users of Sri Lanka based Twitter were developed using three factors: popularity of a user, credibility assumed by analyzing the user metadata and the community seeded credibility. These three factors were analyzed separately in each of the ranking models proposed and built as detailed in sections 3.4-3.6. A data collection mechanism was implemented to collect the necessary tweets and user data for building the user graphs as discussed in section 3.7.2. An algorithm was introduced in section 3.5.2.2 to calculate credibility by analyzing the user metadata. Sections 4.3-4.5 elaborates the process followed to develop each ranking model to generate Credibility Ranking values for the users on the network.

As discussed in section 4.6.1, a survey was conducted among the Sri Lankan Twitter community to gather user input for a manual evaluation of the Credibility Ranking models. The participants were asked to rank 12 users based on the criterion that asked them who among the users would a participant trust more for obtaining the most credible updates during a breaking news situation. The participants ranked users from 1 to 12 based on who they believed most credible to least credible with rank 1 depicting highest credibility and 12 depicting the least credible user among the list of users. An evaluation approach was devised to evaluate the results from Ranking Models (popularity based, metadata based and community seeded) with the survey results. For this purpose, Spearman's rank correlation coefficient (ρ) was used. Spearman's rank correlation coefficient measures the strength and direction of the association between two variables.

The Ranking values generated from each model and the survey result were used as the two variables to calculate the ρ value for each model comparison. According to the evaluation results presented in section 4.6.6, a Ranking Model with highest Spearman's correlation coefficient (closer to +1) was found. The community seeded credibility Ranking Model scored 0.818 in the evaluations and thus was determined as the best performing ranking model. This ranking approach used a set of credible users determined by the community as inherently credible to analyze the rest of the user network and rank other users who may have shared links with these credible users. It is observed that this method involves a degree of manual input. However, this is justified by the high-performance scores of the model compared to the other two ranking models. In the perspective of credibility perception of users on Twitter network, the results from this study show that taking initial cues from the community with regards to whom they consider as highly credible on the network can help identify other credible users who may not be as popular or influential within the network. This is one of the new contributions made by this study towards the research domain.

There were two other credibility Ranking Models developed which utilized other potential credibility factors such as popularity of a user in the network and the user metadata-based credibility scores. The evaluation of these models presented identical Spearman's rank correlation coefficient values, which were both positive and were leaning towards the high association value of +1. (0.517). These results indicated that the use of such techniques to verify the credibility of users would still be useful in the absence of any other such mechanism. However, the results also confirmed the observations from previous research work that popularity alone cannot guarantee credibility on social networks. The results from the survey had some interesting facts, where the users who were highly ranked both from popularity-based Model and User metadata-based Model performing weakly in the survey result which proves the claim that popularity doesn't guarantee credibility. It also makes a strong case that perhaps the metadata-based credibility factors such as the number of followers, friends, the age of the account, number of posts may not be sufficient to make credibility judgements correctly. This is another contribution made towards this research area from this study.

However, the author prefers to point out the observation that the data collection approach for this study banked on obtaining only ten primary accounts and the resulting user graph was based on the friends and their friends' network. Prior research in this domain has mentioned a weakness of network analysis approaches as the inability to use the full user network to perform network analysis due to practical limitations. In the absence of credibility knowledge fed into a user graph as has been done in Model 2, the absence of a full universe seems to hurt the performance of metadata-based Credibility Ranking Model 1.

This brings us to the research question defined in Chapter 1 this study had attempted to answer through the implementation of three user ranking models. The research question read:

How can we determine a tweet as trustworthy news from an unsubstantiated rumor by assessing the credibility of the users?

In this study, the author focused his efforts on analyzing the user network on Twitter to make credibility judgements based on three available credibility assessment factors: popularity, metadata-based credibility and community seeds. It was found out that the community seeds based Ranking Model showed encouraging results. Thus, by utilizing this ranking model, it is possible to provide a ranking for the Twitter user's social standing in the network based on their credibility. The readers can refer to a tweet author's ranking to obtain an idea of how credible the tweet author has been on the network before making credibility judgements on the tweet content. If the tweet author's rank is comparatively low, it can be assumed that the credibility of their content can be low as well, which will force the readers to exercise more caution when spreading the piece of news among their own Twitter circles. This approach can potentially help to minimize the propagation of misinformation or fake news on Twitter network, thereby answering the research question established in the beginning.

To summarize the **contributions of this research**:

- It conducted a **thorough literature review on existing research work** carried out analyzing credibility of users on Twitter and other social networks. Based on the findings, a survey was conducted to evaluate the existing credibility assessment and analysis techniques and the summary of results are reflected in Chapter 2.
- **A fresh dataset of tweets that were circulated during the unfortunate Easter Sunday attacks in 2019** were collected. More than one hundred thousand tweets were collected. Based on the popular users who tweeted during the incident, a user network amounting to more than two million unique users who shared links between them either as primary level friends or secondary level friends were collected along with user metadata for each user. This dataset can be used for performing other experiment of this nature in future in this research area.
- **Three user ranking models, with two of the models having the capability to rank users based on their credibility** were built based on three factors: user's popularity, community perception based seeded credibility and user metadata based credibility.
- The empirical results from experiments revealed that **a user's credibility on the network can be determined with higher accuracy using a community seeded credibility ranking model** as detailed in section 3.6 and asserted in section 4.6.6. The use of a subset of highly credible users on the network through community recommendation can help to identify credible users and communities surrounding these highly credible people.
- The empirical results from the study asserted some of the facts observed in prior research work in this area. The moderate performances shown by popularity based Model 0 and the user metadata based Model 1 as observed in sections 4.6.6.1 and 4.6.6.2 showed that popularity measures and user metadata alone cannot guarantee online credibility.

5.3 Conclusions about Research Problem

In section 1.1, this study discussed the importance of following credible users on Twitter to obtain credible news during breaking news situations because of the unprecedented amounts of misinformation being shared owing to a lack of a fact-checking mechanism offered by the microblogging platform. One way of evaluating the credibility of a tweet is to evaluate the credibility of the user who posted it. The credibility ranking models proposed in this study introduced a new method to rank Twitter users based on their credibility as perceived by the community. As such, it can be concluded that the community seeded credibility Ranking Model introduced in this study addresses the research problem to a greater extent. However, there is room for improvement for better results from this solution through future work described in section 5.5.

5.4 Limitations

This study primarily considered tweets related to Sri Lanka during a high impact event occurred in 2019. It only considered the tweets posted under a select set of hashtags and therefore the tweets are from a Sri Lankan context. However, Twitter users may not always use hashtags during high impact events. Therefore, the data collected may have missed out on retrieving some of the important tweets posted during the early hours since the incident.

The number of secondary friends of the influential users were capped at 100 secondary users for every user due to the API restrictions imposed by Twitter. The time constraints and computational complexities too contributed to this decision. Therefore, the results obtained are subject to this limitation.

The number of credible users seeded into the network in best performing community seeded model is only 20, which may not be sufficient for a network with more than 2 million nodes. The inclusion of at least a 1-2% of credible users into this model could help to provide more accurate results.

5.5 Future work

The performances of Ranking Models built in this research can be further improved through inclusion of global data. The network analysis-based techniques depend on the size of the network. Therefore, availability of more data can increase the accuracy of the credibility ranking values generated by the models.

The existing solutions can be extended further to analyze how the user credibility rankings change between a high impact event and normal times. This distinction can help the users to make credibility judgements on a Twitter user and the content posted by that user based on dynamic situational context rather than on a static credibility ranking value.

It is observed that a large majority of research work carried out in this domain are focusing on the tweet content to make credibility judgements. Machine learning based labelled data are used vastly for this purpose. One of the drawbacks of labelled data is that the annotator's bias affecting the performance of the trained model. Therefore, the possibility of a hybrid solution where the content-based credibility models can be superimposed with network analysis-based user credibility ranking models can help to negate the bias to a larger extent while improving the overall accuracy of the predicted credibility values.

List of References

[1] S. Huang, P. Li and H. Kao, "Identification of item features in microblogging data", *2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, 2015. Available: 10.1109/taai.2015.7407082

[2] M. Alrubaian, M. Al-Qurishi, A. Alamri, M. Al-Rakhami, M. Hassan and G. Fortino, "Credibility in Online Social Networks: A Survey", *IEEE Access*, vol. 7, pp. 2828-2855, 2019. Available: 10.1109/access.2018.2886314

[3] N. Turenne, "The rumour spectrum", *PLOS ONE*, vol. 13, no. 1, p. e0189080, 2018. Available: 10.1371/journal.pone.0189080

[4] G. Michelson and S. Mouly, "Rumour and gossip in organisations: a conceptual study", 2000. .

[5] P. Metaxas, "Web Spam, Social Propaganda and the Evolution of Search Engine Rankings", *Lecture Notes in Business Information Processing*, pp. 170-182, 2010. Available: 10.1007/978-3-642-12436-5_13

[6] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata and R. Procter, "Detection and Resolution of Rumours in Social Media", *ACM Computing Surveys*, vol. 51, no. 2, pp. 1-36, 2018. Available: 10.1145/3161603

[7] S. Vosoughi, M. Mohsenvand and D. Roy, "Rumor Gauge", *ACM Transactions on Knowledge Discovery from Data*, vol. 11, no. 4, pp. 1-36, 2017. Available: 10.1145/3070644

[8] C. Castillo, M. Mendoza and B. Poblete, "Information credibility on twitter", *Proceedings of the 20th international conference on World wide web - WWW '11*, 2011. Available: 10.1145/1963405.1963500

[9] H. Al-Khalifa and R. Al-Eidan, "An experimental system for measuring the credibility of news content in Twitter", *International Journal of Web Information Systems*, vol. 7, no. 2, pp. 130-151, 2011. Available: 10.1108/17440081111141772

[10] L. Zhao, T. Hua, C. Lu and I. Chen, "A topic-focused trust model for Twitter", *Computer Communications*, vol. 76, pp. 1-11, 2016. Available: 10.1016/j.comcom.2015.08.001

[11] M. Morris, S. Counts, A. Roseway, A. Hoff and J. Schwarz, "Tweeting is believing?", *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12*, 2012. Available: 10.1145/2145204.2145274

[12] H. Thakur, A. Gupta, A. Bhardwaj and D. Verma, "Rumor Detection on Twitter Using a Supervised Machine Learning Framework", *International Journal of Information Retrieval Research*, vol. 8, no. 3, pp. 1-13, 2018. Available: 10.4018/IJIRR.2018070101

[13] S. Kwon, M. Cha and K. Jung, "Rumor Detection over Varying Time Windows", *PLOS ONE*, vol. 12, no. 1, p. e0168344, 2017. Available: 10.1371/journal.pone.0168344

[14] M. Abbasi and H. Liu, "Measuring User Credibility in Social Media", *Social Computing, Behavioral-Cultural Modeling and Prediction*, pp. 441-448, 2013. Available: 10.1007/978-3-642-37210-0_48

[15] "FactCheck.org", *En.wikipedia.org*, 2019. [Online]. Available: https://en.wikipedia.org/wiki/FactCheck.org.

[16] "PolitiFact", *En.wikipedia.org*, 2019. [Online]. Available: https://en.wikipedia.org/wiki/PolitiFact.com.

[17] "Watchdog Sri Lanka : Sniffing out fake news", *Watchdog.paladinanalytics.com*, 2019. [Online]. Available: https://watchdog.paladinanalytics.com.

[18] K. Canini, B. Suh and P. Pirolli, "Finding Credible Information Sources in Social Networks Based on Content and Social Structure", *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, 2011. Available: 10.1109/passat/socialcom.2011.91

[19] Y. Yang, K. Niu and Z. He, "Exploiting the topology property of social network for rumor detection", *2015 12th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2015. Available: 10.1109/jcsse.2015.7219767

[20] J. Lu and W. Wan, "Identification of Key Nodes in Microblog Networks", *ETRI Journal*, vol. 38, no. 1, pp. 52-61, 2016. Available: 10.4218/etrij.16.0115.0732

[21] Winnie, O. (2017). Assessing the credibility of online social network messages.

[22] Pew Research Center for the People and the Press. (2009). *Press Accuracy Rating Hits Two Decade Low*. [online] Available at: https://www.people-press.org/2009/09/13/press-accuracy-rating-hits-two-decade-low/

[23] Pew Research Center for the People and the Press. (2012). *Further Decline in Credibility Ratings for Most News Organizations*. [online] Available at: https://www.people-press.org/2012/08/16/further-decline-in-credibility-ratings-for-most-news-organizations/

[24] The Straits Times. (2019). *Fake news rampant after Sri Lanka attacks despite social media ban*. [online] Available at: https://www.straitstimes.com/asia/south-asia/fake-news-rampant-after-sri-lanka-attacks-despite-social-media-ban

[25] Chakraborty, A., Kulshrestha, J., Babaei, M. and Gummadi, K. (2018). Media bias monitor: Quantifying biases of social media news outlets at large-scale. *Twelfth International AAAI Conference on Web and Social Media*, pp.290–299.

[26] Kwak, H., Lee, C., Park, H. and Moon, S. (2010). What is Twitter, a social network or a news media?. *Proceedings of the 19th international conference on World wide web - WWW '10*.

[27] J. Kim and M. Hastak, "Social network analysis: Characteristics of online social networks after a disaster", *International Journal of Information Management*, vol. 38, no. 1, pp. 86-96, 2018. Available: 10.1016/j.ijinfomgt.2017.08.003.

[28] A. Gupta and P. Kumaraguru, "Credibility ranking of tweets during high impact events", *Proceedings of the 1st Workshop on Privacy and Security in Online Social Media - PSOSM '12*, 2012. Available: 10.1145/2185354.2185356 [Accessed 19 January 2020].

[29] T. Bui, S. Ravi and V. Ramavajjala, "Neural Graph Learning", *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining - WSDM '18*, 2018. Available: 10.1145/3159652.3159731 [Accessed 19 January 2020].

[30] G. Huang, Y. Xie and H. Xu, "Cognitive Ability: Social Correlates and Consequences in Contemporary China", *Chinese Sociological Review*, vol. 47, no. 4, pp. 287-313, 2015. Available: 10.1080/21620555.2015.1032161

[31] M. Mendoza, B. Poblete and C. Castillo, "Twitter under crisis", *Proceedings of the First Workshop on Social Media Analytics - SOMA '10*, 2010. Available: 10.1145/1964858.1964869

[32] J. ODonovan, B. Kang, G. Meyer, T. Hollerer and S. Adalii, "Credibility in Context: An Analysis of Feature Distributions in Twitter", *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, 2012. Available: 10.1109/socialcom-passat.2012.128 [Accessed 19 January 2020].

[33] B. Kang, J. O'Donovan and T. Höllerer, "Modeling topic specific credibility on twitter", *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces - IUI '12*, 2012. Available: 10.1145/2166966.2166998

[34] V. Qazvinian, E. Rosengren, D. R. Radev and Q. Mei, "Rumor has it: Identifying Misinformation in Microblogs", 2011.

[35] M. Alrubaian, M. Al-Qurishi, M. Hassan and A. Alamri, "A Credibility Analysis System for Assessing Information on Twitter", *IEEE Transactions on Dependable and Secure Computing*, pp. 1-1, 2017. Available: 10.1109/tdsc.2016.2602338

[36] K. Kumar and G. Geethakumari, "Detecting misinformation in online social networks using cognitive psychology", *Human-centric Computing and Information Sciences*, vol. 4, no. 1, 2014. Available: 10.1186/s13673-014-0014-x

[37] M. Oghenefejiro Winnie, "Assessing the credibility of online social network messages", 2018.

[38] K. Jilcha Sileyew, "Research Design and Methodology", *Text Mining - Analysis, Programming and Application [Working Title]*, 2019. Available: 10.5772/intechopen.85731.

[39] I. Newman and C. Benz, *Qualitative-quantitative research methodology*. Carbondale, Ill: Southern Illinois Univ. Press, 2006, pp. 1-12.

[40] C. Cassell and G. Symon, *Qualitative methods on organizational research: a practical guide*. London: Sage, 1994.

[41] C. Williams, "Research Methods", *Journal of Business & Economics Research (JBER)*, vol. 5, no. 3, 2011. Available: 10.19030/jber.v5i3.2532.

[42] J. Creswell, *Research design: Qualitative & quantitative approaches*. 1994.

[43] J. Ritchie and J. Lewis, *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. 2013.

[44] M. Lichtman, *Qualitative research in education*. Los Angeles: SAGE, 2010.

[45] M. Patton, *Qualitative evaluation and research methods*. Newbury Park: Sage, 1999.

[46] N. Golafshani, "Understanding reliability and validity in qualitative research", *Qualitative Report 8(4), 597-607*, 2003.

[47] P. Leedy and J. Ormrod, *Practical Research*. Upper Saddle River: Prentice-Hall, 2001.

[48] J. Creswell, *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Pearson Education, 2002.

[49] A. Tashakkori and C. Teddlie, *SAGE handbook of mixed methods in social & behavioral research*. Thousand Oaks, Calif.: SAGE Publications, 2010.

[50] R. Johnson and A. Onwuegbuzie, "Mixed Methods Research: A Research Paradigm Whose Time Has Come", *Educational Researcher*, vol. 33, no. 7, pp. 14-26, 2004. Available: 10.3102/0013189x033007014

[51] D. Cormack, *The research process in nursing*. Oxford: Blackwell Science, 2005.

[52] P. Brink and M. Wood, *Advanced design in nursing research*. Thousand Oaks, Calif.: Sage Publications, 1998.

[53] N. Burns and S. Grove, *The practice of nursing research*. St. Louis, Mo.: Elsevier, Saunders, 2009.

[54] B. van Wyk, "Research design and methods Part II."

[55] V. Kandiah and D. Shepelyansky, "PageRank model of opinion formation on social networks", *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 22, pp. 5779-5793, 2012. Available: 10.1016/j.physa.2012.06.047

[56] "Tweepy", *Tweepy.org*. [Online]. Available: https://www.tweepy.org/.

[57] J. Henrique, "Jefferson-Henrique/GetOldTweets-python", *GitHub*, 2017. [Online]. Available: https://github.com/Jefferson-Henrique/GetOldTweets-python.

[58] G. Suresh, K. Suresh and S. Thomas, "Design, data analysis and sampling techniques for clinical research", *Annals of Indian Academy of Neurology*, vol. 14, no. 4, p. 287, 2011. Available: 10.4103/0972-2327.91951

[59] S. Park, W. Lee, B. Choe and S. Lee, "A Survey on Personalized PageRank Computation Algorithms", *IEEE Access*, vol. 7, pp. 163049-163062, 2019. Available: 10.1109/access.2019.2952653.

[60] "Python Tutorial - Tutorialspoint", *Tutorialspoint.com*, 2020. [Online]. Available: https://www.tutorialspoint.com/python/index.htm.

[61] J. Henrique, "Jefferson-Henrique/GetOldTweets-python", *GitHub*, 2018. [Online]. Available: https://github.com/Jefferson-Henrique/GetOldTweets-python.

[62] "Tweepy", *Tweepy.org*, 2020. [Online]. Available: https://www.tweepy.org/.

[63] "NetworkX — NetworkX documentation", *Networkx.github.io*. [Online]. Available: https://networkx.github.io/.

[64] T. Araujo, P. Neijens and R. Vliegenthart, "Getting the word out on Twitter: the role of influentials, information brokers and strong ties in building word-of-mouth for brands", *International Journal of Advertising*, vol. 36, no. 3, pp. 496-513, 2016. Available: 10.1080/02650487.2016.1173765.

[65] "PageRank Algorithm — NetworkX 2.4 documentation", *Networkx.github.io*. [Online]. Available:

https://networkx.github.io/documentation/stable/reference/algorithms/generated/networkx.algorithms.link_analysis.pagerank_alg.pagerank.html#networkx.algorithms.link_analysis.pagerank_alg.pagerank.

[66] "Working with JSON", *MDN Web Docs*. [Online]. Available: https://developer.mozilla.org/en-US/docs/Learn/JavaScript/Objects/JSON#:~:text=JavaScript%20Object%20Notation%20(JSON)%20is,page%2C%20or%20vice%20versa).

[67] M. Patton, QUALITATIVE EVALUATION METHODS. 1987.

[68] P. Cleave, "Recognising the Difference Between Rating and Ranking Questions", *Blog.smartsurvey.co.uk*, 2016. [Online]. Available: https://blog.smartsurvey.co.uk/recognising-the-difference-between-rating-and-ranking-questions.

[69] J. Zar, "Spearman Rank Correlation", *Encyclopedia of Biostatistics*, 2005. Available: 10.1002/0470011815.b2a15150.

Appendices

Appendix A: Code Listings

A.1 Data Collection: Retrieve Users

```
import tweepy
import csv
import pandas as pd
import sys

reload(sys)
sys.setdefaultencoding('utf8')

#### input your credentials here
consumer_key = '####'
consumer_secret = '####'
access_token = '####'
access_token_secret = '####'

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth, wait_on_rate_limit = True)

csvFile = open('users.csv', 'a')
csvWriter = csv.writer(csvFile)
csvWriter.writerow(["screen_name", "created_at", "statuses_count", "followers_count", "friends_count", "verified",
"location", "listed_count"])

with open("usernames.txt", "r") as filestream:
  for usernames in enumerate(filestream):
  print(usernames)
users = api.lookup_users(screen_names = [usernames])
for user in users:
  print(user.screen_name, user.created_at, user.statuses_count, user.followers_count, user.friends_count, user.verified,
user.location, user.listed_count)
csvWriter.writerow([user.screen_name, user.created_at, user.statuses_count, user.followers_count, user.friends_count,
user.verified, user.location, user.listed_count])
```

A.2 Data Collection: Store as Graphs

```
import json
import tweepy
import csv
import pandas as pd
import sys
import time

reload(sys)
sys.setdefaultencoding('utf8')

consumer_key = '####'
consumer_secret = '####'
access_token = '####'
access_token_secret = '####'

class Friend:
  def __init__(self, screen_name, user_id):
  self.screen_name = screen_name
self.user_id = user_id

def to_dict(self):
  return {
```

```
    "screen_name": self.screen_name,
    "user_id": self.user_id
  }

class User:
  def __init__(self, username, friends):
  self.username = username
self.friends = friends

def to_dict(self):
  return {
    "user": self.username,
    "friends": [friend.to_dict() for friend in self.friends]
  }

def extract_users(file):
  users_list = []
root_json = json.load(file)# users = root_json['data']
for user in root_json:
  user_with_friends = get_user_with_friends(user)
users_list.append(user_with_friends)
return users_list

def get_user_with_friends(user):
  friends_list = []
friends_object = user['data']
friends = friends_object['data']
username = user['name']
for friend in friends:
  friend_id = friend['id']
friend_screen_name = friend['screen_name']
new_friend_object = Friend(friend_screen_name, friend_id)
friends_list.append(new_friend_object)
new_user_object = User(username, friends)
return new_user_object

def divide_chunks(l, n):
  for i in range(0, len(l), n):
  yield l[i: i + n]

def get_info_for_user(user):
  friends = user.friends
new_friends_list = []
chunk_size = 98
friends_in_chunks = list(divide_chunks(friends, chunk_size))
print('{} friend list chunks available..'.format(len(friends_in_chunks)))
for index, chunk in enumerate(friends_in_chunks):
  print('chunk {} is being processed..'.format(index))
new_friends = retrieve_user_data(chunk)
new_friends_list.extend(new_friends)
new_user = User(user.username, new_friends_list)
print('new user {} with {} friends'.format(new_user.username, len(user.friends)))
return new_user

def retrieve_user_data(chunk):
  auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth, wait_on_rate_limit = True)
ids = get_ids_from_users(chunk)
print('the user ids are being retrieved..')
print('{}'.format(ids))
friends = []
try:
users = api.lookup_users(user_ids = ids)
for user in users:
  friend = Friend(user.id_str, user.screen_name)
print('friend name: {}, id: {}'.format(user.screen_name, user.id_str))
friends.append(friend)
except tweepy.TweepError as e:
```

```
  print('Something went wrong, quitting...', e)
return []
finally:
return friends

def get_ids_from_users(chunk):
  ids = []
for friend in chunk:
  ids.append(friend['id'])
return ids

def extract_json_array(file):
  root_json = json.load(file)
users = root_json['data']
with open('edited.json', 'w') as outfile:
  jsondata = json.dump(users, outfile, indent = 4)

def main():
  updated_users_list = []
files = [1, 2, 3, 4, 5]
for index, file in enumerate(files):
  with open('file_{}.json'.format(index)) as file:
  users = extract_users(file)
print('{} users were extracted..'.format(len(users)))
for user in users:
  new_user = get_info_for_user(user)
updated_users_list.append(new_user)
for new_user in updated_users_list:
  serializable_users = [user.to_dict() for user in updated_users_list]
with open('easterattack_refined_{}.json'.format(index), 'w') as outfile:
  jsondata = json.dump({
    "data": serializable_users
  }, outfile, indent = 4)

sys.exit()

if __name__ == '__main__':
  main()
```

A.3 Data Collection: Save Friends of Friends Data

```
import json
import jsonpickle
from json
import JSONEncoder
import tweepy
import csv
import pandas as pd
import sys
import time

reload(sys)
sys.setdefaultencoding('utf8')

consumer_keys = ['#####']
consumer_secrets = ['#####']
access_tokens = ['#####']
access_token_secrets = ['#####']

class Friend(object):
  def __init__(self, screen_name, user_id):
  self.screen_name = screen_name
self.user_id = user_id

def to_dict(self):
```

```
  return {
    "screen_name": self.screen_name,
    "user_id": self.user_id
  }

class User(object):
  def __init__(self, username, friends):
  self.username = username
self.friends = friends

def to_dict(self):
  return {
    "user": self.username,
    "friends": [friend.to_dict() for friend in self.friends]
  }

# class UserEncoder(JSONEncoder): #def
default (self, o): #return 0. __dict__

class FriendOfFriend(object):
  def __init__(self, screen_name, user_id, friends):
  self.screen_name = screen_name
self.user_id = user_id
self.friends = friends

def to_dict(self):
  return {
    "screen_name": self.screen_name,
    "user_id": self.user_id,
    "friends": [friend.to_dict() for friend in self.friends]
  }

def extract_user(file):
  json_data = json.load(file)
root_json = json_data['data']
root_user_object = root_json[0]
root_user_object_friends = getUserFriends(root_user_object['friends'])
new_user_object = User(root_user_object['user'], root_user_object_friends)
return new_user_object

def getUserFriends(friends):
  user_friends = []
for friend in friends:
  user_friend_id = friend['screen_name']
user_friend_screen_name = friend['user_id']
user_friend_friends = get_friends_details(friend['friends'])
print('length of user friend of friends array is {}'.format(len(user_friend_friends)))
new_friend_object = FriendOfFriend(user_friend_screen_name, user_friend_id, user_friend_friends)
user_friends.append(new_friend_object)
return user_friends

# def getUserFriendsOfFriends(friends): #user_friends_of_friends = []#
for friend in friends: #friend_of_friend_id = friend['user_id']# friend_of_friend_screen_name = friend['screen_name']#
new_friend_of_friend_object = Friend(friend_of_friend_screen_name, friend_of_friend_id)#
user_friends_of_friends.append(new_friend_of_friend_object)# print('size {}'.format(len(user_friends_of_friends)))# return
user_friends_of_friends

def divide_chunks(l, n):
  for i in range(0, len(l), n):
  yield l[i: i + n]

def get_friends_details(friends):
  new_friends_list = []
chunk_size = 98
friends_in_chunks = list(divide_chunks(friends, chunk_size))
print('{} friend list chunks available..'.format(len(friends_in_chunks)))
for index, chunk in enumerate(friends_in_chunks):
  print('chunk {} is being processed..'.format(index))
api_index = (index + 1) % 10
```

```
new_friends = retrieve_friends_data(chunk, api_index)
new_friends_list.extend(new_friends)
return new_friends_list

def retrieve_friends_data(chunk, api_index):
  auth = tweepy.OAuthHandler(consumer_keys[api_index], consumer_secrets[api_index])
auth.set_access_token(access_tokens[api_index], access_token_secrets[api_index])
api = tweepy.API(auth, wait_on_rate_limit = True)
ids = get_ids_from_users(chunk)
print('the user ids are being retrieved..')
print('{}'.format(ids))
friends = []
try:
users = api.lookup_users(user_ids = ids)
for user in users:
  friend = Friend(user.id_str, user.screen_name)
print('friend name: {}, id: {}'.format(user.screen_name, user.id_str))
friends.append(friend)
except tweepy.TweepError as e:
  print('Something went wrong, quitting...', e)
return []
finally:
return friends

def get_ids_from_users(chunk):
  ids = []
for friend in chunk:
  ids.append(friend['user_id'])
return ids

def main():
  files = [1, 2, 3, 4, 5, 6, 7, 8, 9, 0]
for index, file in enumerate(files):
  with open('top_{}.json'.format(index + 1)) as file:
  user = extract_user(file)
serializable_user = user.to_dict()
with open('top_{}_final.json'.format(index + 1), 'w') as outfile:
  jsondata = json.dump(serializable_user, outfile, indent = 2)

if __name__ == '__main__':
  main()
```

A.4 User Analysis: Calculate Credibility Score using Metadata (Model 1)

```
import csv
import pandas as pd
from datetime
import datetime

def getAgeFromCreatedDate(date_str):
  start_date = datetime.strptime(date_str, '%Y-%m-%d %H:%M:%S')
end_date = datetime(2019, 04, 21)
num_months = (end_date.year - start_date.year) * 12 + (end_date.month - start_date.month)
return num_months

class TwitterUser(object):
  def __init__(self, id, screen_name, bio, followers, friends, statuses, lists, age, verified, location):
  print('creating user {}'.format(id))
self.id_str = '{}'.format(id)
self.screen_name = screen_name
self.bio = bio
self.followers_count = followers
self.friends_count = friends
self.statuses_count = statuses
self.listed_count = lists
print('user created at {}'.format(age))
self.age = getAgeFromCreatedDate(age)
```

```
self.verified = verified
self.location = location

def load_user_data(filename):
  df = pd.read_csv(filename, lineterminator = '\n', error_bad_lines = False)
del df['index']# df['created_at'] = pd.to_datetime(df['created_at']).dt.strftime('%Y-%m-%d')# print(df.info)
users = []

for index, row in df.iterrows(): #print('id: {}, name: {}, location: {}'.format(row[0], row[1], row[9]))
user = TwitterUser(row[0], row[1], row[2], row[3], row[4], row[5], row[6], row[7], row[8], row[9])
users.append(user)

return users

def getUserFromId(id, users):
  user = next(iter(filter(lambda x: x.id_str == id, users)), None)
return user

# Mandatory attributes# Age of the account# Number of followers# Number of friends# Verified Status# Number of tweets

# Optional Attributes# Location# Number of lists# Has a bio

def calculateAgeScore(age):
  if age <= 0:
  return 0
elif age > 0 and age <= 6:
  return 1
elif age > 6 and age <= 24:
  return 2
elif age > 24 and age <= 60:
  return 3
elif age > 60 and age <= 120:
  return 4
elif age > 120:
  return 5
else :
  return 0

def calculateFollowerRatioScore(followers, friends):

  if friends == 0:
  return 0

ratio = followers / friends

if ratio < 0.5:
  return 0
elif ratio >= 0.5 and ratio <= 1:
  return 1
elif ratio > 1 and ratio <= 2:
  return 2
elif ratio > 2 and ratio <= 5:
  return 3
elif ratio > 5 and ratio <= 10:
  return 4
elif ratio > 10:
  return 5
else :
  return 0

def calculateTweetCountScore(tweets):
  if tweets <= 10:
  return 0
elif tweets > 10 and tweets <= 100:
  return 1
elif tweets > 100 and tweets <= 1000:
  return 2
elif tweets > 1000 and tweets <= 5000:
  return 3
```

```
elif tweets > 5000 and tweets <= 10000:
  return 4
elif tweets > 10000:
  return 5
else :
  return 0

def calculateVerifiedStatusScore(isVerified):
  if isVerified:
  return 5
else :
  return 0

def calculateLocationScore(location):
  if pd.isnull(location):
  return 0
elif "sri lanka" in location.lower() or "srilanka" in location.lower() or "colombo" in location.lower():
  return 5
elif pd.isnull(location) is False:
  return 4

def calculateUserListsScore(lists):
  if lists < 2:
  return 0
elif lists >= 2 and lists <= 10:
  return 1
elif lists > 10 and lists <= 30:
  return 2
elif lists > 30 and lists <= 60:
  return 3
elif lists > 60 and lists <= 90:
  return 4
elif lists > 90:
  return 5

def calculateUserBioScore(bio):
  if pd.isnull(bio):
  return 0
else :
  return 5

def calculateUserTrustScore(user): #mandatory credibility factors
ageScore = calculateAgeScore(user.age)
followerRatioScore = calculateFollowerRatioScore(user.followers_count, user.friends_count)
verifiedUserScore = calculateVerifiedStatusScore(user.verified)
tweetCountScore = calculateTweetCountScore(user.statuses_count)

mandatory_attributes = [ageScore, followerRatioScore, verifiedUserScore, tweetCountScore]

# optional credibility factors
locationScore = calculateLocationScore(user.location)
listCountScore = calculateUserListsScore(user.listed_count)
bioScore = calculateUserBioScore(user.bio)

optional_attributes = [locationScore, listCountScore, bioScore]

weighted_mandatory_trust_score = 0
mandatory_attribute_weight = 1
for attribute in mandatory_attributes:
  weighted_mandatory_trust_score += mandatory_attribute_weight * attribute
weighted_mandatory_trust_score = weighted_mandatory_trust_score / len(mandatory_attributes)

weighted_optional_trust_score = 0
optional_attribute_weight = 0.5
for attribute in optional_attributes:
  weighted_optional_trust_score += optional_attribute_weight * attribute
weighted_optional_trust_score = weighted_optional_trust_score / len(optional_attributes)

final_weighted_score = weighted_mandatory_trust_score + weighted_optional_trust_score
```

```
normalized_final_weighted_score = getNormalizedTrustScore(final_weighted_score)
return normalized_final_weighted_score

def getNormalizedTrustScore(x):
  minX = 0
maxX = 7.5
return (x - minX) / maxX - minX

def main():
  users = load_user_data('input1.csv')

csvFile = open('output1.csv', 'a')
csvWriter = csv.writer(csvFile)
csvWriter.writerow(["id_str", "screen_name", "credibility_score"])

unidentified_users = []

for user in users:
  userObject = getUserFromId(user.id_str, users)

if userObject is None:
  unidentified_users.append(user.id_str)
continue

print(user.screen_name)

csvWriter.writerow([user.id_str, user.screen_name, format(calculateUserTrustScore(userObject), '.4f')])

if len(unidentified_users) > 0:
  unidentified_users_csvFile = open('unidentified_users.csv', 'a')
unidentified_users_csvWriter = csv.writer(unidentified_users_csvFile)
unidentified_users_csvWriter.writerow(["id_str"])
for obj in unidentified_users:
  unidentified_users_csvWriter.writerow([obj])

if __name__ == '__main__':
  main()
```

A.6 Perform PageRank Analysis

```
import json
import csv
import sys
import time
import networkx as nx
import matplotlib.pyplot as plt
import pprint

class User(object):
  def __init__(self, username, pagerank):
  self.username = username
self.pagerank = pagerank

class CredibleUser(object):
  def __init__(self, id, screen_name, score):
  self.id = id
self.screen_name = screen_name
self.score = score

def format_decimal(value, decimals = 2):
  from decimal
import Decimal# divide value by 10 ** decimals;
this is just scaling
value = Decimal(value).scaleb(-decimals)
return "{:.{d}f}".format(value, d = decimals)
```

```
def populateGraph(filename, G):
  with open(filename) as file:
  js = json.load(file)# object = js[0]
user_object = js['user']
friends_object = js['friends']
G.add_node(user_object)

for eachFriend in friends_object: #first level
friend_name = eachFriend['screen_name']
G.add_edge(user_object, friend_name)# friend_user_id = eachFriend['user_id']
friend_of_friends = eachFriend['friends']

for eachFriendOfFriend in friend_of_friends: #second level
friend_of_friend_name = eachFriendOfFriend['user_id']
G.add_edge(friend_name, friend_of_friend_name)

def get_credible_tweeps_dict(filenames):
  credible_tweeps = dict()
for filename in filenames:
  with open(filename) as csv_file:
  csv_reader = csv.reader(csv_file, delimiter = ',')
next(csv_reader)
for row in csv_reader: #user = CredibleUser(row[0], row[1], row[2])
credible_tweeps[row[1]] = float(row[2])# credible_tweeps[row[1]] = 0.2# print(len(credible_tweeps))

return credible_tweeps

def main():

  #create a di - graph.
G = nx.DiGraph()

# load files.

filenames = ['top_1_final.json', 'top_2_final.json', 'top_3_final.json', 'top_4_final.json', 'top_5_final.json', 'top_6_final.json',
'top_7_final.json', 'top_8_final.json', 'top_9_final.json', 'top_10_final.json']
for filename in filenames:
  populateGraph(filename, G)

nodes = list(G.nodes)
personalized_dict = {}#
credible_tweeps_list = ['AzzamAmeen', 'SriLankaTweet', 'MarianneDavid24', 'kataclysmichaos', 'NewsWireLK', 'adaderana',
'NewsfirstSL', 'BBCWorld', 'gopiharan', 'munza14', 'rangaba', 'Almashoora', 'GotabayaR', 'TeamWatchDog',
#'AthaudaDasuni', 'BasnayakeM', 'Meerasrini', 'Reuters', 'tingilye', 'Dailymirror_SL']# val = 0.000000000000002

#
for node in nodes: #if node in credible_tweeps_list: #personalized_dict['{}'.format(node)] = val#
else :#personalized_dict['{}'.format(node)] = 0

credible_tweeps_dict = get_credible_tweeps_dict(["user_credibility_sheet_1.csv", "user_credibility_sheet_2.csv"])
print('credible tweeps length: {}'.format(len(credible_tweeps_dict)))

# sys.exit()

# credible_list = {
  'AzzamAmeen': val,
  'SriLankaTweet': val,
  'MarianneDavid24': val,
  'kataclysmichaos': val,
  'NewsWireLK': val,
  'adaderana': val,
  'NewsfirstSL': val,
  'BBCWorld': val,
  'gopiharan': val,
  'munza14': val,
  #'rangaba': val,
  'Almashoora': val,
  'GotabayaR': val,
  'TeamWatchDog': val,
```

```
  'AthaudaDasuni': val,
  'BasnayakeM': val,
  'Meerasrini': val,
  'Reuters': val,
  'tingilye': val,
  'Dailymirror_SL': val
}

print(nx.info(G))# pagerank = nx.pagerank(G, max_iter = 500)
pagerank = nx.pagerank(G, max_iter = 500, personalization = credible_tweeps_dict)

new_dict_pr = {
  k: v
  for k,
  v in sorted(pagerank.items(), key = lambda item: item[1], reverse = False)
}

# print('new_dict_pr list length: {}'.format(len(new_dict_pr)))# sys.exit()

dict2 = {}
for key, value in new_dict_pr.items():
  dict2['{}'.format(key)] = format(value, '.15f')

sorted_dict2 = {
  k: v
  for k,
  v in sorted(dict2.items(), key = lambda item: item[1], reverse = True)
}

csvFile1 = open('user_pagerank_model3_1.csv', 'a')
csvFile2 = open('user_pagerank_model3_2.csv', 'a')
csvFile3 = open('user_pagerank_model3_3.csv', 'a')

csvWriter1 = csv.writer(csvFile1)
csvWriter2 = csv.writer(csvFile2)
csvWriter3 = csv.writer(csvFile3)

csvWriter1.writerow(["screen_name", "page_rank"])
csvWriter2.writerow(["screen_name", "page_rank"])
csvWriter3.writerow(["screen_name", "page_rank"])

index = 0
for key, value in sorted_dict2.items():
  print('{} user: {}  PR: {}'.format(index + 1, key, value))
if index + 1 <= 1040000:
  csvWriter1.writerow([key, value])
elif index + 1 > 1040000 and index + 1 <= 2080000:
  csvWriter2.writerow([key, value])
else :
  csvWriter3.writerow([key, value])
index += 1

sys.exit()

if __name__ == '__main__':
  main()
```

Appendix B: Examples of User Credibility Evaluation using Credibility Scoring Matrix

Username	@AzzamAmeen	
Feature	**User Statistics**	**Credibility Score**
Follower ratio	343.9	5
No. of Tweets	24744	5
Age of the Account	> 10 years	5
Verified Account	Yes	5
Has a Bio	Yes	5
Has a Location	Yes, Colombo	5
No. of Lists	409	5

Username	@GotabayaR	
Feature	**User Statistics**	**Credibility Score**
Follower ratio	4794.6	5
No. of Tweets	2067	3
Age of the Account	> 5 years	4
Verified Account	Yes	5
Has a Bio	Yes	5
Has a Location	Yes, Sri Lanka	5
No. of Lists	178	5

Username	@munza14	
Feature	**User Statistics**	**Credibility Score**
Follower ratio	50.78	5
No. of Tweets	21377	5
Age of the Account	> 9 years	4
Verified Account	No	0
Has a Bio	Yes	5
Has a Location	Yes, Sri Lanka	5
No. of Lists	73	4

Username	@himalkk	
Feature	**User Statistics**	**Credibility Score**
Follower ratio	4.9	3
No. of Tweets	76813	5
Age of the Account	> 10 years	5
Verified Account	No	0
Has a Bio	Yes	5
Has a Location	Yes, Colombo	5
No. of Lists	94	5