RANKING OF TWEETS BASED ON CREDIBILITY FACTORS

PRADEESH

FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

RANKING OF TWEETS BASED ON CREDIBILITY FACTORS

PRADEESH

DESSERTATION SUBMITTED IN /PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF COMPUTER SCIENCE

**FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR**

2017

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: (I.C/Passport No:)

Registration/Matric No:

Name of Degree:

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Field of Study:

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

ABSTRACT

Twitter is extensively being used to share news, links, images and even have conversations. In Malaysia alone, there are 3.5 million twitter users. As the volume of tweets and users who are increasingly accessing tweets as source of information, they have less information to judge if a tweet is credible or not. The consequences of spreading non-credible tweets can be harmful to the society, nation and to the entire world. To respond to this issue, this research considered ranking tweets by various qualities of a tweet, such as popularity, reliability, timeliness, trustworthiness of web pages and tweets link to provide a more credible Twitter users search results than the current Twitter search which only looks at relevance without looking at the credibility of the tweet. An evaluation of the method on 144,972 tweets from GST which is consists of Malay and English tweets shows that the proposed scoring technique pTRank scores much more better compared to TwitterRank in various ranking evaluations such as in Normalized Discounted Cumulative Gain (nDCG), the system scored a score of 0.393, as opposed to TwitterRank which is at 0.121. The same trend is also noticed with both GST tweets in both the languages and as well as only on English.

ABSTRAK

Kebelakangan ini, Twitter banyak digunakan untuk berkongsi berita, pautan, gambar dan juga mencetuskan perbualan. Dianggarkan terdapat 3.5 juta pengguna Twitter di Malaysia. Jumlah tweet dan pengguna yang merujuk Twitter sebagai sumber informasi semakin meningkat, sedangkan punca sumber itu tidak dapat dipastikan, justeru kesahihan informasi itu diragui. Sebaran tweet yang diragui kesahihannya akan mendatangkan kesan buruk kepada masyarakat, negara dan juga antarabangsa. Oleh itu, isu ini dipilih, dengan penyelidikan terhadap tweet yang diberi ranking berdasarkan beberapa kualiti yang digariskan, seperti populariti, boleh diharapkan, mutakhir, juga kebolehpercayaan laman web dan pautan Tweet. Ini antara langkah ke arah hasil carian pengguna Twitter yang sahih berbanding hasil carian Twitter sekarang yang cuma menyenaraikan Tweet yang relevan tanpa mengambil kira kredibiliti Tweet tersebut. Penilaian kaedah ini terhadap 144,972 Tweet tentang GST, merangkumi Tweet berbahasa Malaysia dan Inggeris menunjukkan teknik pemarkahan pTRank yang disyorkan, mendapat markah yang lebih tinggi berbanding TwitterRank. Penilaian ranking seperti Non Discounted Cumalative Gain (nDCG), mendapat markah 0.393, manakala TwitterRank cuma beroleh markah 0.121. Trend yang sama diperhatikan pada Tweet kedua-dua bahasa berkaitan GST, terutamanya Bahasa Inggeris.

ACKNOWLEDGEMENTS

I owe a great deal of thanks to the number of people who made it possible for me to write this thesis. The process of going through and writing the thesis is not easy when you must juggle your work life and as well as your studies. I truly admit that it would be impossible to complete it without their support.

Firstly and foremost, I would like to thank both my supervisor and my co-supervisor, Dr. Sri Devi Ravana and Dr. Kasturi Dewi Varathan . Without your support and guidance and your mentorship, it would have been impossible for me to complete my thesis. Once again thank you for being agile and understanding

Secondly, I would like to thank my parents. They are the no 1 motivator behind this, encouraging me to pursue my masters in Computer Science. Without their continued encouragement and blessing, it would have been hard for me. After all, they are the loves of my life.

Thirdly, I would like to thank my awesome colleagues and friends who have supported me and being understanding when I need to go on study leave and do my examination and covering for me while I am at work. I would also like to thank Zhang and Sani, who are my Masters buddy for helping me out and giving me encouragement as well

Lastly but not least my gratitude and thanks goes to God almighty. Without His uttermost blessing and spiritual guidance, it would have been a rollercoaster ride. Lastly but not least, to all my lecturers in UM who have thought me. I would like to extend my gratitude to "Matha pitha Guru Deivam" , a Tamil verse for "Mother Father Teacher God"

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

RT	:	Retweet
TREC	:	Text Retrieval Conference
JSON	:	JavaScript Object Notation
AP	:	Average Precision
MAP	:	Mean Average Precision
P	:	Precision
@	:	Mentions (Addressing Another Twitter Account) :
#	:	Hashtags (Topics or Discussion)

CHAPTER 1: INTRODUCTION

1.1 Background Study

Social media such as YouTube, Facebook, Twitter and Instagram have revolutionized the way how humans interact and it is part of Web 2.0 ecosystem which is mainly to empower users by allowing them to generate and consume their own content (Kaplan & Haenlein, 2010). Twitter is particularly unique among other social media platforms as it allows users to express their feelings and thoughts in 140 characters. Although Twitter is primarily used as a communication medium, it has been used as a news medium and as a platform for commercial corporations to reach out to their users (Kwak et al., 2010).

Figure 1.1: A Tweet containing false information that causes anxiety and fear among society

The main challenge with any Web2.0 content is the issue of credibility. Twitter is filled with chatter space and the content which includes news or event updates which is posted by the users themselves (Aditi et al., 2012) as opposed to traditional medium such as newspaper and television whereby it is bound to have undergone vigorous scrutiny (Yang, Counts, Morris,Hoff, 2013).

This has made Twitter infamous for various reasons, death hoaxes are very common in Twitter such as Justin Bieber's alleged death (Hollywoodlife, 2014), which has caused Twitter to be known as a site which is untrustworthy whereby in a study conducted by Semierbach and Oeldorf-Hirsch (2010), the responders viewed the information posted on news sites to be perceived more credible as opposed to the same content being posted in Twitter itself. In another study by Zogby (Gupta, Zhao & Han, 2012), a poll was conducted to find out on trustworthiness in Twitter and only 8% of the people trust content posted in Twitter.

However not everything that are being posted in Twitter are harmful and there are many instances whereby Twitter did shine in bridging and providing information such as during Hurricane Katrina (Hughes,A. L. & Palen, 2009) from people who were on the ground on what was happening and also which was used by journalist to report the news as well. The challenge that is being faced today is that there needs to be a way that to determine rumours and truth that are being posted.

Per the statistics published by *Suruhanjaya Komunikasi dan Multimedia* (2012), 85.7% of all Internet users in Malaysia use Internet to socialize with their friends and there are about 13 million Malaysians are active on social media sites and out of which 65.5% of them log in at least once a day to interact in social media. The risks of spreading false rumors and be risky to the individual, organization or even the country itself. For example, the most recent, MH370 incident have generated over 850,000 tweets during March 11th 2015 as shown in Figure 1.1. Incident through the usage of hashtags such as *#malaysianairlines*, *#prayforMH370,* #MH370. However due to the sheer volume and the fact that it was a major disaster and users wanted information in real-time verifying such tweets as shown in Figure 1.1 would be hard and it leaves room to speculation and false rumors (Aditi et al., 2012). It has become a very serious issue that even the government of Malaysia is looking at on how to curb spread of misinformation in social media (Middleton, 2015).

There were instances that fake news was reported by various media around the world such as in the case of MH370 (NBC, 2014). Thus, makes it even harder for journalist to “fact check” the content posted on social media if it is the truth or not (Hermida, 2010; Farhi, 2009), which may contribute to spreading of misinformation.

However, there are some studies that were done to detect rumours and in credibility in Twitter, however it just focused on the usage of hashtag during high impact events (Castillo et al., 2010; Gupta et al., 2012) such as an earthquake. It only mentions if they are credible or not. In another study, conducted by Morris et al (2013), people also do consider various of other non-text clues when deciding credibility in Twitter, which are clues such as username, the content of the text before deciding if the tweet is credible. Currently the process of screening used by journalist and authorities are manual and there is not a systematic way of screening.

1.2 Problem Statement

The research addresses the following issue, whereby journalists and authorities are unable to verify with precision and credibility on the content and the user who posted in social media without any sort of tool or measure used to judge. Instead they are using their own judgment when publishing news in main paper. In the past literatures, the credibility scoring measurement solely focused on during high-impact events and solely on the features of hashtag, user profile and the number of retweets or the reach (Earle, Bowden, & Guy, 2012; Okazaki & Matsuo, 2012) but little has been done on the features that were found on tweets and the people who tweet behind them. Hence this research addresses the problem by using these features found on tweets and the people to help to determine credibility of a tweet, so that it is not opened interpretation (Boyd, 2016). To

Hence some form of scoring measurement is needed to assist them to able to judge person who posted the tweet by providing a ranking and a scoring without inducing any form of biasness with a ‘yes’ or ‘no’. This research aims to tackle the said problem by extending the work of credibility of tweets in past literatures by extending with 2 new features and as well as using relevance and ranking.

1.3 Research Questions

An in-depth analysis and understanding on underlying Twitter architecture and Web 2.0 would help to address the issue in tackling credibility in Twitter. Thus, the research questions of this research would be defined as following: -

- How additional features from Twitter can be used to further strengthen the credibility of a tweet?
- How to give a score to a user considering credibility of their tweets and popularity?
- How can the scores be represented and ranked for users for them to evaluate the credibility of users without any form of biasness?

1.4 Research Objectives

To answer the research questions, the following statements are the objectives of the research that would be the focus of the thesis.

- To utilize additional features from Twitter that are found in tweets and users profile that can be further to strengthen the credibility of the tweet
- To create an improved credibility score for a Twitter user due to the current poor scoring system by ranking them based on their influence and features found on the user and the quality of the tweets.

- To create an evaluation metric that gives a fair scoring to a person who posted tweets in a *language independent* method whereby the ranking is not influenced or deterred by a language and by utilizing their credibility score.

1.5 Significance of the Study

The significance of contributions from this research are as follows: -

1. This research has proposed two new additional features that are found on Twitter which would help to determine credibility of the tweet
2. This research has also proposed a method of calculating a credibility score by combining the score of features that found on the users and as well using user's influential scores.
3. This research has also helped to create an evaluation metric by ranking them based on their scores which compromises of their credibility score which works across of all tweets regardless of the language it was posted in.

1.6 Thesis outline

This thesis is divided into five chapters. The summary and purpose of each of the chapter are explained as below

Chapter 1 describes the Introduction which includes the motivation, problem statement, research question and the outcome.

Chapter 2 presents the literature information. Related key concepts are explained in the detailed which are pertinent to the proper understanding of the rest of the thesis.

Chapter 3 describes the research method and materials that are used for the proposed solution. Detailed explanation have been provided which includes System Architecture, Algorithms, Twitter API, Features Selection, Building Relevance Judgement and the Final score calculation process

Chapter 4 explains the experiment, the data set and the graphical representation of the scores and the results of the proposed scoring technique and how it fares with other techniques that were discussed in the literature.

Finally, in Chapter 5, conclusions, contributions and future developments of this research work are discussed

CHAPTER 2: LITERATURE REVIEW

In this chapter a comprehensive review of existing literature was performed to support the study in this thesis.

In the following sections, literature on both general and Twitter-specific areas on credibility are reviewed. Starting with various application of social media and their usage pattern. Then we move on to the core focus area of the thesis which is on credibility on various other Web 2.0 systems and ranking done in these systems including some of the works around Twitter that was carried out in the past literature. Later, overview of application of Twitter is discussed on how it helped during various events, which forms the importance of our research, and as well as the features that were used in the past literatures.

2.1 Birth of Social Media in Web 2.0

The foundation of Web 2.0 such as being able to collaborate, tag and storytelling has given birth to social media. Social media is said to be the heart of Web 2.0 that truly empowers the end-user that allows creation and exchange of user generated content (Kaplan & Haenlein, 2010). Social media allows quick and rapid exchange of ideas and discussion to take place. It allows users to quickly share a message, a video or an image with others. However, the concept of social media is not a new, according to Coyle and Vaughn (2008), in their research they have stated that it goes back to the nature of humans to be connected with one another and since social media is about people in its core trying to find unique ways to connect with one another who share common ideas or similarity interests, both of them do blend together. There are various types of social media application such as social networking site, microblogging, image and video content sharing that exists out there and the table below shows the popular social media applications that exists and that are widely being used. Table 2.1, shows the top social media systems which were used in 2015 (Heath, 2015) are highlighted and described further on the type and the purpose it is used for to provide an overall overview of these platforms.

Table 2.1: Popular Social Media Applications

Application Name	**Type**	**Purpose**
Facebook	Social Networking	Social Networking site to connect friends and family members together. It allows users to provide stats update to notify their friends, status update and share pictures
YouTube	Video Content Sharing	Video sharing site that allows anyone to publish their own video
LinkedIn	Social Networking	Social Networking site for professionals that allows them to connect together.
Twitter	Microblogging	Twitter is a microblog that essential allows to communicate in less than 160 characters.

2.1.1 Social Media Usage

Social media usage is very high, there about 3.419 billion Internet Users (Geffen A, 2016) and out of which 2.307 billion users are on various social media platform. According to Kemp (2012), Malaysians are very friendly active on social media as they have the highest number of Facebook friends. Apart from being used by users, businesses and other organizations are seeing the value of using social media and some of the businesses of them have even adapted to use social media for marketing and engagement purposes (Kaplan, 2012). Apart from businesses, healthcare even has tapped social media to listen and respond to patient needs by providing alerts and reminders to take their medication and as well as to interact with them to help to overcome anxiety and depression if the patient is suffering from it (Hawn, 2009).

Thus, by understanding the need and drive for people to connect one another, it lays the groundwork on how social media tools such as Twitter is being used by people to connect one another. In the following section, and introduction on Twitter microblogging platform is explained in depth the usage of Twitter during revolution of government, natural disasters and politics are discussed. These few themes were selected as it the time of the period where Twitter is widely used by people to communicate and disseminate information due to the nature of the event itself and where rumours spread. (Pelvin et al., 2015).

2.2 Application of Twitter

Twitter is a microblogging service that allows to post short messages that are at most 160 characters long. The updates can contain text, images, videos and any other Internet based media. The nature of twitter has its roots from Instant Messaging(IM) services which was started to be used to interact between two parties that was popular among students in high school and colleges. (Quan-Haase, 2008). IM allowed communication between two known people easier over the Internet and made long-distance communication possible (Quan-Haase, 2008). One of the challenges of instant messaging is that two people would need to know each other before a conversation can take place. In additional to that, IM has also paved way a new medium of communication whereby it is different of that is being used in e-mail or phone, which is much more formal of nature.

With the groundwork from IM has been established, Twitter have made it easier to converse with people. Once a user registers for an account, the person can easily post each status update or which is known as a “tweet”. These tweets can be viewable by anyone. It is with this nature of Twitter that allows people to be discovered by other people who could have shared their thoughts. With the openness and ease to communicate 300 billion tweets that have been posted, over which out 32 million were tweeted during World Cup 2014 Finals alone (Team Caffeine, 2014).

This has caused Twitter to be beyond a tool just for conversation it is also being used as a new medium (Java et al., 2007; Pear Analytics, 2009; Naaman et al., 2010), due to the nature of speed and its immediacy. This has opened a new path for the usage of Twitter, whereby it allows anyone to report events and history as it unfolds especially in the times of crisis and emergency which is explained in-depth in the following section.

2.2.1 2007 Wildfires in California

In October of 2007, over 20 wildfires spread in the Santa Barbara County in California up to the US Mexican Border. It has burned about 500,000 hectares of forest. The fire forced approximately 1,000,000 people to evacuate out of their homes and it was one of the large evacuations (MSNBC, 2007). However, a survey that was conducted, one of the main frustration of the locals is that they were not happy with the quality of information provided by the authority, as it was incomplete and they are left in the dark (Sutton et al., 2008). Sutton et al. (2008) also discovered that similar issue occurred when mainstream media covered the event of Hurricane Katrina back in 2005. The main challenge that was faced by a lot of this local news, despite providing an around-the-clock coverage, the news providers were unable to keep up-to-date with rapid changes and residents are left in the dark. To fuel to the frustration of the people, the county emergency website was not able to handle the web traffic that was coming in and it was not be able to catered for everyone. This is the turning point whereby residents and the people around turned to social media to provide updates, several volunteers and residents of the county started posting stories, discussing routes and even providing updates to their friends and family via social media. There are two residents from the San Diego county who have gathered information from friends and other news sources and posted the event as it happens on Twitter (Poulsen, 2007). These people provided very specific details such as where to get help, listing inventory of groceries and supermarkets which were still open (Poulsen, 2007).

This is the event popularized the usage of Twitter, whereby 10% of the affected residents were using Twitter for their source of first time and most of them being first time users of Twitter (Sutton et al., 2008).

2.2.2 2008 China Sinchuan Earthquake

Twitter was the first media that brought the attention of the world to the earthquake that happened in the Sinchuan, China. It was posted by AFP (2008) that a blogger by the name of Robert Scoble posted the event on Twitter. According to Scoble, he was informed of this incident from his friend in China and shared it in Twitter.

Due to the damage the earthquake has done to telecommunications infrastructure, a lot of people turned to the Internet for help to be notified, particular to Tianya Club. It is a very popular forum in China. Qu et al (2009), did a content analysis on the aftermath of the earthquake performed a content analysis. In another work by Li & Rao H.R (2010) have found out that the Twitter helped to propagate and reach out to audience faster that of a main media In the work by Qu et al (2009), the researchers have noted that most of the topics that were being discussed are more of informative in nature whereby people are asking for details and also to updated on the events of the earthquake. Qu et. al (2009) has also noted that when rumours from Twitter that are posted in forum, it is questioned by other members of the board and they have cross-referenced it with multiple sources to dismay that it is a rumour. Finally, once there is sufficient information from members that marked a posting as a rumour, it is removed by the moderator. This way rumors could be contained and moderated, however it was done with manual intervention as opposed to being able to automatically which heavily relies on people's judgement to moderate them.

2.2.3 2011 Arab Spring

Twitter played a very pivotal role in the 2011 Arab Spring or which is also known as the Twitter Revolution crisis (Jurgenson, 2012) whereby what started to be a revolution and anti-government protest and regime changes that occurred in Tunisia in 2011 has spread to many Middle Eastern countries such as Egypt, Libya, Tunisia and even to Syria (Bruns,Highfield,Burgess, 2013).

Twitter was primarily used as the medium to disseminate news by activists and organizers of anti-government protesters (O'Dell, 2011). For instance, in Libya, corruption of Muammar Gadhafi was exposed through various tweets that were done by activists. Since traditional medium was blocked and there was no avenue for people to complain. Although, some of the government in Middle East did initially block access to social media, users quickly used a workaround to spread the information by using various workaround such as using proxies or spoofing their Internet Protocol (IP) Address to be from another country (Rabbat, 2012).

The decentralization of Web 2.0 and Twitter has provided the citizens of Middle East to host their own discussion via hashtag such as #libya and #egypt . This allowed anyone to talk about the topic without worrying about being controlled by one organization. This allowed politicians from both spectrum, people from Middle East, journalist to come together to follow on the event that unfolds in the Middle East. (O'Dell, 2011). However due to the popularity of the hashtag, it was then soon used by spammers to reach out to their audiences as well.

One of the interesting aspect of the 2011 Arab Spring is that, most the tweets were posted in a mixture which consists of English and Arabic as opposed to as Arabic is the common language in these regions in the Middle East. At that upon of time, Twitter was still testing left-to-right support for Arabic and it was only until 2012 that Twitter did officially support it (Twitter,2012). According to Poell and Darmoni (2012), another reason on the usage of mixture of the language is to gain a wider reach to the international audiences on to know what is going on in the particular region and reach out to International media.

The resulting mixed language raises new questions on tackling and classifying tweets from various languages. There is some work that been done to determine the subjects of tweets that were posted in Arabic and English (Papacharissi & de Fatima Oliveria, 2012). The work done by them were mainly preliminary to determine the nature of the topic that was posted. However, to the best extend of this research, there aren't many works that is being done to work with verifying credibility on cross-language tweets and users.

2.2.4 2012 US Election

2008 US Election was heralded to be the first presidential election that made use of Twitter extensively. It was the period of Twitter whereby it was still in its infancy as Twitter was only 2 years only during the 2008 election after it was launched in July 15 2006. However, in 2012 US Elections, social media was used extensive by then US President, Barrack Obama for his re-election and his opponent, Mitt Romney to take helm of the white house. It was this period that the 2012 US election season, broke Twitter's "most heavily tweeted" records. (McKinney et. al., 2014) and this is where campaigners took to social media to gather votes from people in United States by sending mixture of original messages, replies and engaging constantly with their supporters with pictures and videos of their campaign.

As politics and social media combined works as a doubled-edged sword for political misinformation. With this being a conduit, a lot of detractors and due to the structure of politics of being partisan, rumours start emerging and start using Twitter to start propagating false rumors about the candidate that they dislike.

However, spreading rumours during political campaign it is not something new. In an experiment done by DiFonzo et al. (2013), there is a high segregation and clustering within a person's social media circle, due to the nature of how people make friends and become friends in social media (Gilbert & Karahalios, 2009), that is commonly by interests and combined that with that references of an individual, having pre-existing attitude due a certain political party or candidate. The ease of social media makes it easier to spread the hatred or misinformation as well.

There is not much work done on this area of relating political and partisan biasness and in social media as the researchers mostly focused on ascertaining rumor acceptance among people who are into politics, rather than the media systems within which such rumors circulate. (Cacciatore et al., 2014; Nyhan and Reifler, 2010; Weeks and Garrett, 2014).

2.2.5 2014 MH370 Incident

Malaysian Airlines MH370, a Boeing 777 aircraft that was bound for an overnight flight from Kuala Lumpur, Malaysia to Beijing, China went missing on March 7th,2014. The aircraft was carrying 227 passengers and 12 crew members. (NBC, 2014).

Initially when the search and rescue mission began on the next day, rumors in social media started spurring in whereby it started with Chinese social media stating that the aircraft had made an emergency landing in Naning,China (NBC, 2014). This news was picked up by a lot of local news agencies in Malaysia and as well as in overseas.

Another variation of the rumour states that the plane has crashed somewhere between the South China Sea and in Vietnam. Figure 2.2 below shows on the title of the news that was picked up by a local news agency which turned out to be false. (NBC, 2014). This has even caused anger among family members of victims who have vented their anger to news agencies for providing false information (The Week UK, 2016).

Figure 2.1: False Rumour of MH370 landing in China that was reported by ASTRO Awani based on a tweet (Awani, 2014)

The spread of rumours is partially associated with the fact that Malaysian Airlines and the authorities were unable to respond in a timely manner (The Week UK, 2016) which has resulted in social media users engaging in pure speculation to fill in the void and posted unfounded theories. According to Briguglio (2013), the key during a crisis is to keep the public informed by providing timely updates and containing the crisis in order to prevent negative rumor to spread.

Due to the high volume of such rumors, Malaysian Communications and Multimedia Commission (MCMC) had to put up a notice warning the public the harmful effect of spreading rumours and such person that found to spread rumours would be prosecuted by the law as shown in Figure 2.3.

Figure 2.2: Warning Issued in Malay by MCMC informing users not to share any unverified rumours

However, this approach requires the public to report such rumours and incidents to MCMC before action are taken and often these are more reactive measures rather than being proactive and it is proven to be ineffective

In the next section, the issue of credibility in various Web 2.0 system, particularly in Twitter is discussed in-depth

2.3 Credibility

Social Media such as Twitter changed the face of how content are being shared and published is because the communication is in two-ways whereby users can communicate with one another which serves as a complimentary source of information as opposed to search engine.

As Twitter is becoming a medium for people to obtain news and also the platform that is first to report any incidents such as earthquakes, riots and other events often times journalist have struggled with how to incorporate news from social networking platforms such as Twitter into "*established journalism norms and values*" (Hermida, 2010), since it is much more easier to filter news and information without the need of going through section editors . In a traditional medium, such as newspapers, TV News and magazines the content is "*fact checked*" (Hermida, 2010; Farhi, 2009) however without fact checking, it would impact the integrity of the content that is published by the journalist in the print media. One of the major challenge with user generated content is its credibility. Credibility is defined as "*a communicator's positive characteristics that affect that receiver's acceptance of a message*" (Ohanian, 1990). As seen from the past application of Twitter, there are still room for improvement in addressing some of the issues that were brought up during the application of Twitter in the previous section.

If the information is not credible, then it said to be a rumour. Rumours are unverified information that contains valuable statement or public concern for certain groups (DiFonzo and Bordia, 2007). The spreading of rumours, may cause anxiety and unrest to the individual or society (Liu F e. al., 2014). If journalists and mass media report rumours, the repercussions are grave and it could involve in lawsuits by government and mass public as well (Buttery, 2015).

In order to tackle the issue of credibility, there are some work that has been done in different fields of Web 2.0 on what are the features in that medium of Web 2.0 that helps to determine credibility, this is because it is beyond content and the author that needs to be measured for credibility (Rieh & Danielson, 2002). This chapter discusses several ways of tackling credibility in various Web 2.0 systems which are forums, blogs and online shopping sites that were found in the literature and later in-depth on Twitter.

2.3.1 Credibility in Forum

In a study to judge quality and credibility in Internet forums, Saolainen (2011) suggested two main criteria's which are author characteristics which describes the author of the post in terms of *reputation* whereby how reputable the person in the forum based on the number of posts, likes or kudos he has received, *expertise* whereby if the person has any special badges or rewards/token given to him due to his contribution, *identification* and also on the message information content such as *novelty*, *factuality* whereby the posts are factual backed by evidence such as citation or link to external source. However, in this study no formal conclusion could be made due to the limited number of sample size which is only 4739 messages posted in 160 Finnish discussion thread. In additional to that with the decline of users using Internet forum and in favour of another social media (Mou et al., 2013), there is not much research done into this.

2.3.2 Credibility in Blogs

Banks (2008) interviewed 30 leading bloggers and suggested attributes in measuring credibility, which are *focused* as in the topics do not sway and the content of the blog is niche, *authentic* as in content are not reproduction or rewrite of an article, *insightful as* in the blog's author provides his insights or thoughts to the article. Rowse (2006) has added another attribute which is known as *consistency* whereby the author is being consistent and not contradicting themselves. Lastly, another attributed was introduced which is known as *timeliness* of blog content, which is on how often the blog is updated (Banks, 2008; Weil, 2006). However, it is not possible to replicate it in Twitter, were used in blogs are very hard to be applied in tweets due to the nature of tweets being short, whereby it only supports 160 characters whereas a blog entries are long which contains more than 160 characters long.

2.3.3 Credibility of Comments Online Shopping/Review Sites

One of the very successful Web 2.0 e-commerce are online shopping sites such as Amazon and eBay. Both sites have incorporated user-based feedback which allows users to post feedback about the item and as well as about the seller. The review system in e-commerce sites then combines the scores and provides in aggregated score for users to easily access. (Chun Wang, 2008). However, the challenge is that it does not take into consideration of the aggregated score, thus making the results very much biased. In Chun Wang (2008) work, he tried to address this by introducing timeliness factor into credibility of reviews. It considers the timeliness of reviews to be one of the feature to evaluate the credibility of reviews. The results show that method proposed by Chun Wang (2008) method is much more superior of Amazon's default scoring system.

2.3.4 Credibility in Twitter

There is a fair slice of work that was done for Credibility in Twitter. The nature of Twitter is that it is considered as a microblog and as well as a social media tool. (Thelwall, Buckley & Paltoglou, 2011). The advantage of Twitter over other platform is that it contains a rich metadata information despite the message or the content is only having 160-character length. With the growth of Twitter and with various application of Twitter such as in emergencies and politics as discussed earlier in this chapter, the number of rumors and misinformation has also increased. The earliest work on credibility was focused on detecting earthquake by analyzing tweets to complement existing earthquake detection system to improve the accuracy. (Crooks, Croitoru, Stefanidis & Radzikowsk, 2013; Earle, Bowden, & Guy, 2012; Okazaki & Matsuo, 2012). However, their work truly focused on determining if the natural events and their methods cannot be applied into other fields or to be reproduced it was very skewed on the type of event.

One of the pioneer work on this field done by Castillo et. al (2010). The researches there found the need to have a credibility method that can be applied across any field and worked addressed the gap. Castillo et. al (2010) have done work in identifying features that are found in Twitter that helps to determine credibility of an event in Twitter which are *user-based features* such as the number of followers and the number of tweets they have posted and as well *propagation-based features* such as the number of retweets and likes and lastly, *topic-based features* such as if the tweet contains URLs, length of the message. These features were determined by crowdsourcing to Amazon Mechanical Turk and then later applying with a different dataset it with machine learning technique. Liu et. al (2015) extended the work of Castillo by using feature detection in detecting real-time rumour detection in Twitter.

However, the difference between Castillo and Liu is that Castillo's work studied more into credibility of tweet whereas the work of Liu (2014) focused more on rumour propagation, the difference between both is one is more into identifying and classifying whilst the latter is focusing on the propagation itself. Both people's work focused mainly on the usage of hashtag and not on the content of the tweets or any other specific keyword.

There were several other credibility assessments of posts was done however it was focusing on China's equivalent of Twitter, which is known as Sina Weibo. The works done by Wu et al. (2015) and Yang et al. (2012) is to analyze credibility atomic tweet level by grouping similar tweets based on their features and then classify them into similar categories. It is also closer to life scenario especially done by journalists if there were to verify content. The disadvantage of this proposed method is that it does not work with Twitter as the API provided by Twitter does not provide much detail as Sina Weibo. Thus, making the methods that were proposed in their work not being able to be replicated for use in Twitter.

On the other hand, there were similar works done by other researchers. They mainly used with an additional of external list or verification site to validate the credibility of the work. In the research conducted by Quazvinan et al. (2011), the researchers used a database found in About.com's Urban legend site to determine rumour propagation and built a Bayesian classifier based on the linguistic features found on the tweet such as the usage of capital letters, punctuation and exclamation marks. Gupta & Kamaraguru (2012), did a similar approach and used supervised machine learning technique (SVM Ranking) instead of a Bayesian to classify credibility. Finn,Metaxas and Mustafaraj (2014) created a tool which is known as TwitterTrails , that allows users to trace the origins of the rumour by using features such as retweet propagation , following followers ratio and also the content of the tweet and also based on the completeness of their user profile. Their work is only looking at news portal and targeted specifically for journalists who are looking for news to report. Secondly, their work heavily focused on news that were in English language and there was no comparison that were made in their paper on how the system fares with earlier systems as well to determine the strength of the system.

Apart from credibility, another strong area of work that was done is in detecting spam tweets in Twitter. Spam detection in Twitter is heavily focused upon *user-based features* and as well as on *propagation-based features* Spammers are very active in Twitter as well and it said that spam links in tweets are clicked twice as often as compared to e-mail (Sedhai & Sun, 2016). Spams can be reported by twitter users by clicking on the Report functionality in the user profile. However, the manual method of reporting spam is tiresome and often it takes time before Twitter takes any action on the user in Twitter. Twitter does not rely on manual reports and they have certain restrictions in place to combat spam (Song, Lee & Kim, 2011) which are as follows: -

- Following a lot of users in a very short period of time
- Low ratio of number of followers and the number of people the individual is following
- Tweets that are being repeated
- Tweets that contains invalid URLs such as Error 404 or Low Page Rank Score such as that are being posted in a short period

The restrictions that Twitter enforces are no doubt are strict, but the ease of creating accounts in Twitter makes it very easy for spammers to continue to create new account and to spam. Hence, various researchers have proposed methods in combating spammers, these methods do make use of Twitter features and it works by collecting tweets and then using a classifier to classify spams and normal tweets. (Avello,Brenes, 2010; Wang, 2010;

Benevenuto, Magno, Rodrigues, &Almeida, 2010; Markines, Romero, 2010). These works have classified spammers with high accuracy, with an average accuracy of 85%. There were other studies that were done as well but it mainly focused on classification of spam detection but in various usage such as during a major football or concert events that is riding on the popularity of hashtags (Santos et al., 2014).

In 2015, Twitter have improved their spamming algorithm by utilizing Google SafeBrowsing method to automatically flag and remove tweet (Chen et al., 2015). However, much like any other algorithm in the past the algorithm was not in real-time and the dataset had to be collected before it was be able to be classified. In the works by Wang et al. (2015), the researchers there have managed to use features from past literature and by analyzing them in real-time using a classification algorithm to determine if it is spam or not from Twitter. However, their work is still in preliminary stages and even in their initial attempts, it has shown that it is on par with spam detection of Chen et al. (2015).

The major drawback of with all the work that was done the researchers is that their worked is on credibility and spam detection in isolating in credibility and they have archived a great depth in their field of work. However, the researchers work did not focus on ranking of displaying the information from most credible to less credible. In section 2.5, the features that were used in past literature and its purpose are discussed further.

2.4 Ranking

Ranking is a subject that is extensively covered in Information Retrieval (IR) studies. Ranking is important as the general purpose of an IR system is to assist the user in locating the information that they are after (Croft & Laffety, 2015) as it is important as it helps users to determine the ranked of each retrieved document to provide some form of judgement or perception which document is more important over another (Smeaton, 2012). A classic example would be web search engines such as Google, which retrieves web pages, or in another example a student trying to browse through University Malaya library catalog and wants to retrieve a such as the one found in which retrieves journals and article, web search engines such as search queries are ranked in terms of relevance from the most relevance to the least relevance. What these systems do is to avoid a situation which is known as "information overloaded" whereby the user is overwhelmed with the choices of documents that they have (Maes, 1994). In the following subsection, various ranking techniques that were used in past literature for Web 2.0 are discussed.

2.4.1 Ranking Blogs

Blogs is one of the Web 2.0 whereby it is created by the users and not often by journalists or people who govern it Blogs have few things in common across all the blogs which are namely title, date and the content (Juffinger,Grantizer & Lex, 2009) which makes it easier to rank. Existing ranking mechanism for Web search engine such as PageRank algorithm which is used by Google (Page et. al, 1998) are used to rank blogs as it is part of a document. The challenge with PageRank is that often times, spammers with bloggers do boost their ranking by doing which is known as "*linkback*". "*Linkback*" a technique that is used to create to increase the ranking and to trick. Detailed explanation of underlying PageRank is discussed later in this chapter. Google have since made improvements (Search Engine Watch,2013) to their algorithm to penalize blogs that are exploiting this.

There is specific blog based ranking which was worked on such as by Kritikopoulous et al. (2006) that introduced similarities between the bloggers and their blogs based on their content and on the links, that they share. This algorithm gives higher ranking based for the blogger that is well known in the blogosphere by comparing the articles that are linked in his blogs. In the works by Liu et. al (2009), the researchers have extended the work of Kritikopoulous (2006) by introducing PostRank. It is a ranking method that is based structure-based approach like how newsgroup messages are organized. By taking cues from how newsgroup are organized which is based on latest comments, it then takes consideration of comments posted into blogs as part of its ranking methodology.

Kim et. al (2015) decided to take a step further by extending the works of Liu et. al (2009) to include analyze traceback, which are commonly to find out that the blogs or any other external media sites which have linked and took a step further to analyze. His results fared well against the research work that was carried out by Liu and his team.

There is another approach that was done by Bashir (2015). In his work, he proposed to rank blogs based on users' opinion instead of the traditional method which looks at the metadata as done in past literature which was discussed. In his proposal model, he combined the elements of the users such as the keywords there were used, the web client that was used to post and crawled the person's social media if he has a similar name to give an improved ranking. From his analysis, the results perform as well as other query-based expansion system that is out there. There were several drawbacks on this method as it required a huge computational time and as well as heavily reliant on social media profiles that were linked to gauge. However, his work is the one that looked of the person's social media influence.

2.4.2 Ranking Tweets

Ranking tweets is different compared to ranking in traditional Information Retrieval (IR) evaluation method, supposedly there is a tie in the ranking of documents and the tie broken by some random values such as the document's age or the document ID (Wang, Arko & Fang, 2013). This is not the case for tweets as it would not be possible to break such a tie using a document ID or so as it would not provide accurate method. However, researchers have across the field managed to diversify and managed to exploit the reach features of tweets to perform ranking.

The following section explains two methodologies that were used in various literatures that were reviewed in this research. The two methodologies are ranking based on user's characteristics, which looks at the features that are found on the user's account. The other methodology is ranking tweets based on user's network topology which mainly focuses on the person's social network connection and the popularity of the person.

2.4.2.1 Ranking Based on User Characteristics

Yue Wang et. al (2010) have proposed a ranking method for tweets by tie-breaking two similar tweets that have the same score by using term frequency(TF), inverse document frequency (IDF), and document length (DL). In additional, followers count has also added to do ranking. Another ranking method based on temporal retrieval was also done whereby preferring temporal evidence over recency influence, whereby the ranking of the query depends on the time when the query was searched and the time the tweets that were posted. Duan et al. (2010) has also used a learning ranking model using Support Vector Machine (SVM) based on the content of the tweet and exploiting tweet based characteristics such as URL, number of followers. Leavitt et. al (2009) also did a new methodology based on content and responses of 12 popular users to determine to rank Twitter based on their feedback, however the effectiveness of the method was not compared. Another metric has been proposed by Web Ecology project (2011) which measures the influence based on attention such as tweets, retweets and replies that the user receives.

The main disadvantage of using user features purely is that it does consider the whole topological view of the social network whereby it does not consider the interaction between users in a social network which is an important aspect of social media as people.

2.4.2.2 Ranking Based on Network Topology

In ranking systems that use network topology, the relationship between users are considered. Researchers have worked upon the basis of random walk algorithm such as PageRank. The first notable algorithm that has extended the implementation of PageRank is TunkRank. TunkRank works by the assumption that each user is given an influence score that is the number of people would read the tweets based on the people he/she followers. Secondly, the attention that one person reads all the tweets are the same. However, the assumption does not reflect the real-world behavior of the Twitter uses as retweet is not constant as it is based on timeliness. (Tunkelang, 2009).

TwitterRank is another approach by ranking users. It works by taking both topic similarities between users and link structure and combining with an external list, which is the list of top influencers. The researchers of TwitterRank state that the algorithm is much better the current way how Twitter ranks its tweets which is solely based on the number of followers (Jianshu et. al, 2010).

TwitterRank measures the influence taking both topical similarity of users and on how the link structured. The results show that TwitterRank outperformed TunkRank and the default PageRank algorithm and even modified PageRank algorithm that empathizes on topic that was proposed by Haveliwala (2002). In the following section, the algorithm of TwitterRank is discussed in depth as it forms the basis of the proposed algorithm which would be discussed later in Chapter 3.

(a) ***Algorithm***

There are similarities between how a web page is crawled and ranked with PageRank. The same principles can be applied with Twitter as well and this is where the work of Jianshu et al. (2010) when the researchers built TwitterRank.. Figure 2.4 shows the similarities between PageRank and modified PageRank that is tuned for Twitter.

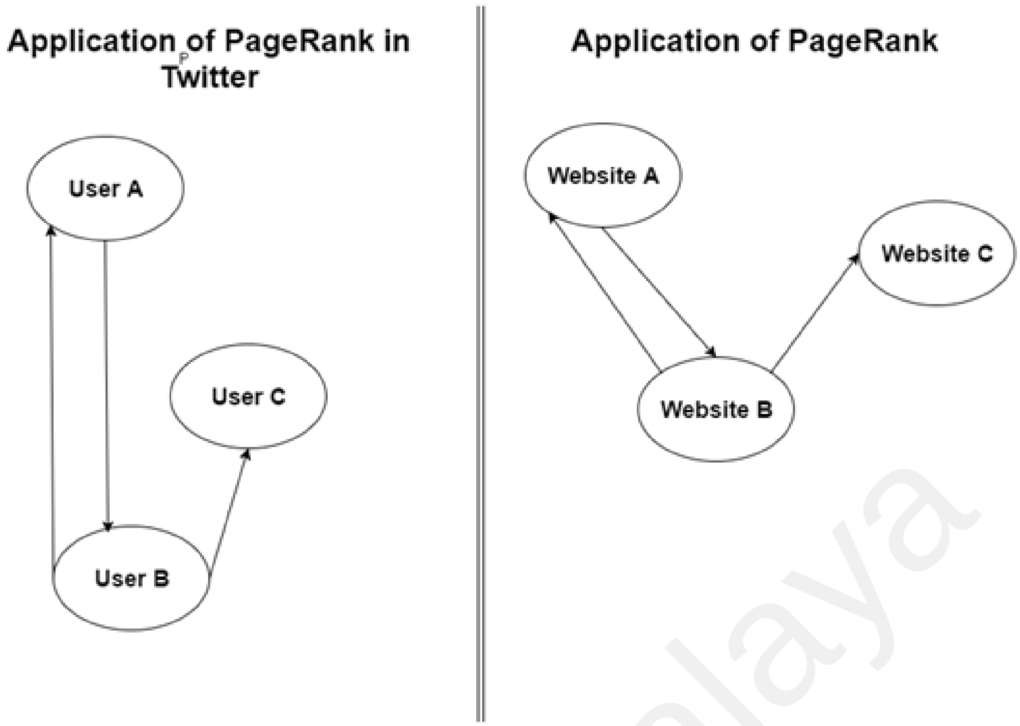

Figure 2.3: Differences between PageRank and Application of PageRank in Twitter

Fundamentally, PageRank works by the concept of voting. In the example given in the right hand side of Figure 2.4, Website A has vouched for Website B and Website B have vouched for A, which means there is a strong relationship between the two-websites. Websites here denotes a node and the line represents if there are link between them or which is known as an edge. Although Website C has been voted by Website B, the score would be much lower as it is not as popular or it does not have as many incoming edges.

This same concept was applied in Twitter by TwitterRank . As shown in left hand side of Figure 2.4 for Twitter, node represents a user and the lines or the edges represent the followers/following relationship depending on the position of the arrow, an inward arrow represents a follower and outward represents a following relationship. From the Figure 2.4, both User A and User B follow each other and User C is only being followed by User B and not by User A. The reciprocity of social relation and the reason behind this due to the way how people behave, to give a real-life example is everyone knows Tiger Woods and it does not necessary mean Tiger Woods needs to know everyone or knows everyone else. This phenomenon of following/follower relationship which is known as “celebrity problem” is discussed in length by Kwak et. al. (2010) in his work.

However as with any graphs, issue arises on cyclic graph. It creates two issues, firstly whereby an infinite loop is created by traversing the whole graph over to calculate the score and there is no way for it stop. Secondly, in the case of PageRank whereby it would create an infinite scoring system as scores are calculated repeatedly as there is no way to stop’ the algorithm, which would cause the score to be inflated which is known as dangling nodes. In order to tackle this, Page et. al. (1998) introduced damping factor to prevent the cycles to be in infinite loop. This is discussed later in Chapter 3.

One of the disadvantage of TwitterRank is that it used a seed list or which is also known as to determine top celebrities or influential people first to build its database and then to start crawling its tweets.

(b) ***Improvisation of TwitterRank***

There are several improvisations were made to TwitterRank. Xiong (2013) improvised the ranking of TwitterRank by creating ranking algorithm which includes topic-sensitive influence ranking which has included mutual information among users, which is known as WTSIRank. In another work by Feng Shi et al. (2013), who implemented a topic-sensitive ranking algorithm by using slightly using a different algorithm. Instead of using the traditional PageRank Algorithm which uses random walk, their work used a weighted value that was mainly determine by the content and the number of published tweets. The transition probability that is to move from one node to another node was calculated using the correlation coefficient between the posts and the richness of content such as including the quality of the link which was posted. They compared their algorithm with WTSIRank, PageRank and TwitterRank and WTSIRank outperformed the rest.

In a comparison that was done by Nuo Li et al. (2015), the researchers there have found that the major disadvantage of all the algorithms is that they have a lot of "stiffness", as in it does not consider the other features that are found on tweets to analyze them but only looking at their relationship between followers and people who follow and just the content of the tweet. In additional to that, the core algorithm for TwitterRank still requires a list of influential Twitter users for it to function, which may introduce biasness in ranking.

In totality, the research done by these researchers mainly focused on influence of a person and the popularity of the topic and ranking them. There is a gap that needs to be addressed which is to combine the ranking methodology, which is not just purely on the influence but also at the aspect of credibility or how truthful is the tweet or the person but rather a means of finding popular twitter users and ranking them and their tweets. This gap is addressed later in Chapter 3 by combining both the scores and as well as ranking them.

In additional to that in all the research work that was done to rank tweets, two popular scoring methods were used, which is Precision (P) and Average Precision (AP) These two scoring methods are popular as they were used in search engine retrieval. (Ishii et al, 2015). The explanation of these scoring method is detailed out later in Chapter 3.

2.5 Features of Twitter Used in Past Literature

In this section, a summary of the features that were used in the literatures that were discussed is explained in detailed to provide an overview of how they work in Twitter. The literature that were reviewed are discussed earlier as part of credibility in Twitter and as well as ranking in Twitter. The features are categorized in two sections, which are account level features, which are found on individual accounts or people in Twitter, and Tweet level features that are found at the atomic level of every tweet. Later in this section,

Table 2.2 provides of the features that were used in past literature that was reviewed are summarized to provide a comprehensive view.

2.5.1 Tweet Features

Tweet features here describes further in details of the features that are found on individual tweets.

2.5.1.1 Coordinates

Coordinates which is part of *user based features,* shows the place or the location of the tweet. It helps in determining the location of the tweet to determine. In previous research, such as the works of Castillo et. al (2010), Yang et al (2012), Stefanidis, & Radzikowski (2013), Liu et al (2015) the researchers have used this feature extensively to narrow down to disaster location to increase the credibility

2.5.1.2 Favorite

Favorite tweets show that the number of likes or how many times it has been favorite by person. Higher number indicates the engagement of how users are taken. It can be used as an effective measure of determining the propagation of tweets and the influence factor of an individual who have posted it. Favorite is also part of *propagation based feature.*

2.5.1.3 Retweet

Retweet is the spread of measure how well if someone agrees or disagrees of the retweet (Boyd et al., 2010), it also could be used as a measure to share or propagate the tweet. Retweet is a powerful tool that allows messages to be reached out to a wider audience. Retweet works by propagation. For instance, if Mary retweets John's tweet, all of Mary's followers can see the tweet. They can they retweet the tweet again. This is the primary medium of how rumors and misinformation are being spread in Twitter. (Finn et al., 2014). Figure 2.5 shows how Twitter shows the people who have retweeted the particular tweet. Retweets are part of *propagation based feature*.

Figure 2.4: Retweet Information from a Tweet

2.5.1.4 Content of Tweet

Content of the tweet mainly compromises message of the tweet. It can contain metadata such as URLs, Images and files. In the works by Castilo et al. (2010) and Liu et al. (2015), the researchers there used lexical analysis to analyze the text to determine if it's a sentence or piece of information or just random words that are being put together. This is because in their works and during emergencies, news media and people who are on the ground would tend to write a longer more coherent sentence. In the works of spam detection by Chen et al. (2015), the content of the tweet is said to be spam if it contains in all capital letters, emoticons and also URLs which have PageRank values less than 2. Content of tweet is part of *topic based features*.

2.5.2 User Profile

User Profile in Twitter is the profile of the person and provides features such as the author's location, author's website and if the user is verified by Twitter. The detailed explanation of each of the feature are further elaborated below. Spammers often times have default information profile with a default picture (Kurt,Chris et al., 2011). Figure 2.4 shows a typical view of a Twitter User Profile. Many spam detection algorithm for Twitter have extensively used the features found on User Profile. (Avello & Brenes, 2010; Wang, 2010).

Figure 2.5: Typical User Profile of a Twitter User . In this example Barack Obama was used

2.5.2.1 Location

The location contains the location of the person. Spam accounts and fake accounts would not specify the location and generally would be blank (Sedhai & Sun, 2016). Location is part of *user based feature.*

2.5.2.2 Account Date Creation

Account creation date is the date when the Twitter account was created and it is part of *propagation based feature* It is an important feature found on Twitter that shows how long the person has been on Twitter According to Kurt,Chris et al. (2011) ,56% of spam accounts which are created are less than 2 days old . Often times new account creation is associated with spam (Song, Lee & Kim, 2011).

2.5.2.3 Verified

Verified indicates the person's twitter account has been verified by Twitter authorities and it is part of *user based feature*. Twitter generally conducts verification of well-known celebrates and other figures around the world (Stever & Lawson, 2013). This is to curb against parody accounts or fake accounts. In the works by Liu et al. (2015), of calculating credibility, the feature had the high weightage when determining if it is credible or not.

2.5.2.4 Profile URL

Profile URL displays the URL that has linked to the profile. In the example in Figure 2.4, the Profile URL links to US White House government Web Page. Profile URL provides a way for people who are looking in Twitter to understand more about the user.

2.5.2.5 Following/Followers Ratio

In Twitter, following someone means to subscribed to their tweets and receiving their tweets in the timeline or in the home screen of the person's Twitter. When someone, let's say Mary follows John, she is known as follower of John. Generally, people's perception of influential in Twitter is determined by of followers determines how influential the person is or how respected. (Marwick, 2015). Celebrities, public figures and politicians often have many followers and they follow very little people. The ratio that is calculated helps to determine if the person is passive or an influencer, the higher the following to followers' ratio, it means the person is a passive Jianshu et al. (2010).

2.5.3 Comparison of Twitter Features

In order to provide a summarization view of the chosen literature that were reviewed in this research, Table 2.2 provides the list of features that were used in past literature that was reviewed which encompasses of credibility of tweets and ranking tweets It provides a summarized view of the features that other researches and studies have used these features for their applications. Detailed of the explanation of each of the features are also mentioned after the table to provide the background.

Table 2.2: Features in Twitter Used in Past Literature

	User Profile Features					**Tweet Features**			
Literature	**Following Followers Ratio (User Based Feature)**	**Location (User Based Feature)**	**Verified (User Based Feature)**	**Profile URL (User Based Feature)**	**Account Creation Date (Propagation Based Feature)**	**Retweet Count (Propagation Based Feature)**	**Favorite (Propagation Based Feature)**	**Coordinates (User-based features)**	**Content of Tweet (Topic-based features)**
Avello,Brenes (2010)	✓	×	×	×	×	✓	✓	×	✓ (English)
Chen et al.,2015	✓	✓	✓	✓	×	✓	✓	×	✓ (English)
Gupta et al (2012)	✓	×	✓	✓	×	✓	✓	✓	✓ (English)
Benevenuto,et al (2010)	✓	×	×	×	×	✓	✓	×	✓ (English)
Castilo et al (2010)	✓	×	✓	✓	×	✓	✓	✓	✓ (Spanish)
Feng Shi et al (2013)	✓	×	×	×	×	✓	✓	×	×
Jianshu et al (2010)	✓	✓	✓	✓	×	✓	✓	×	✓ (English)

Liu et al (2015)	√	×	√	×	×	√	√	√	√ (English)
Finn et al (2014)	√	×	×	×	×	√	√	×	√ (English)
Tunkelang, (2009)	√	×	×	×	×	×	√	×	×
Sakaki et al (2012)	√	×	×	×	×	√	√	×	×
Web Ecology Project (2011)	√	×	×	×	×	√	√	×	×
Xiong (2013)	√	×	×	×	×	√	√	×	×
Nuo Li et al (2015)	√	×	×	×	√	√	√	×	√ (Chinese)
Yue Wang et al (2010)	√	×	×	×	×	√	√	√	√ (English)
Dum et al (2010)	√	×	×	×	×	√	√	×	×
Leavitt et al (2009)	√	×	√	√	√	√	√	×	√ (English)
Stefanidis, & Radzikowski (2013)	√	√	√	√	×	√	√	√	√ (English)
Yang et al (2012)	√	√	√	√	×	√		√	√ (English)

2.5.4 New Features & Need of Ranking

There are still more features that can be used. From the feature selection and the research in past literature, the key focus for features for credibility and for ranking are different. The motivation of these features comes from the work of other Web2.0 systems that were discussed such as blogs and forums. In additional to that, to have a combined score and overview, there needs to be new features that encompasses factor of time of tweets as it would impact on credibility and ranking. This is because older tweets would rank low as tweets are very fast. (Sloan et al., 2015). In additional to that past researchers have used (Kurt,Chris et al., 2011) one way of timeliness such as the creation of account to determine if it is a spam or not. In our research, is it further extended to include timeliness of the tweets.

Secondly, there is a need of identifying Twitter Client, which is the client or software that is used to post tweets. It is also considered to be a feature despite it is a software as the motivation comes from the work of Bashir (2015) for ranking blogs based on social media linkage by using blog posting software to determine credibility of blog. Ideally, similar concept can be applied to influence the rank of credibility of a tweet by considering the usage of twitter client as often, spammers use their own programmed Twitter Client (Stieben, 2013). These two new features are further discussed in Chapter 3.

Thirdly, as discussed earlier in Section 2.4.2.2, the research done by these researchers are mixed, as in focusing in respective areas such as in ranking or in credibility. The challenge is that the work done in credibility mainly focused on the truth factor of tweets and the options were presented in binary form (Castilo et. al., 2010), whereas in ranking it was focused on mainly on ranking based on relevance or popularity. Hence, there is a need to combine the various feature based scoring which are user based, propagation based and topic-based features. to provide an overall holistic scoring system. The combination of these two methods are discussed in depth in Chapter 3.

2.6 Summary

From the past literature, there have been various works that has been done in the various fields of Web 2.0 to tackle the issue of credibility and the importance of credibility as well. However, in the case of Twitter, the works done in past literature was mainly focused on the content and event, such as the usage of hashtag and not about the users for credibility. On the other hand, there were works done in terms of ranking tweets based on relevance as well, but it focused on influence of the person and how well the tweets are spread. To the extent of this research, there are not any research out there that combines the factor of credibility and ranking in accessing the scores.

There are various features were used, there are still additional new features can be used. In this research, two new features are introduced which are Twitter Client and as well as Aging of Tweets to address the gap and to improvise the scoring.

CHAPTER 3: RESEARCH METHODOLOGY

The previous chapter gave an idea of overall of the works done in the past by researchers and the gaps that needs to be addressed. In this chapter, the primary focus is on the research methodology that would be used to carry out the experiment. Firstly, the research design of the system that needs to be built is discussed in-depth. Secondly, the algorithm that is used to perform the calculation is explained. Lastly, various ranking methodologies and models of system are evaluated in order to perform the evaluation of the system.

3.1 Research Design Overview

The design of the system is split into three phases based on the research objectives as stated in Chapter 1. Figure 3.1 below illustrates the design of this research.

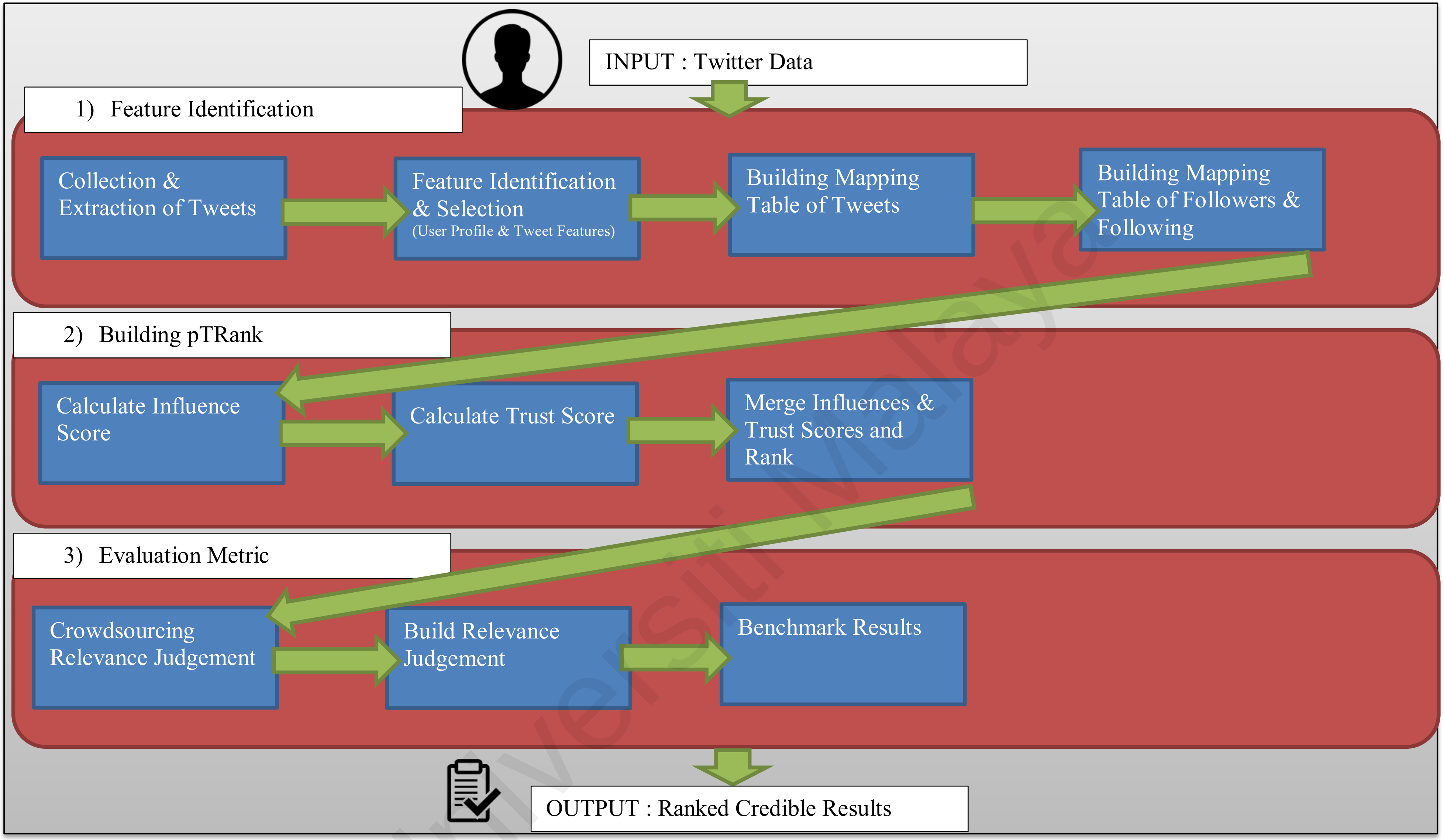

Figure 3.1: Overall Research Design

As seen in Figure 3.1, the design of the system compromises of three section which are identification of features, assignment of the score and finally to create evaluation metric. compromises of identifying and extraction of tweets and their user. Each of the steps are detailed out in the section below. In the first stage of the design, which is the on identification and extraction. it is described in-depth in Section 3.2. As for Assigning and Calculating Score, which contains the core formula and algorithm which was used is described in depth in Section 3.3. Lastly, as for creating an evaluation metric which used to validate the scores, it is described in-depth in Section 3.4.

3.2 Feature Identification in Tweets

For this purpose of research two set of data are collected. One is primarily used as the primary data to conduct the experiment and a secondary dataset is used to validate the result and if the experiment is in line and if the research objective has been met. Once the dataset is collected, it is then extracted to get the features.

3.2.1 Primary Dataset

The Primary Dataset contains of Tweets and its users which were parsed in JavaScript Object Notation (JSON) from the crawler. Over 144,972 tweets with the keyword of GST that were collected from 13th of February 2015 and up to 20th of May 2015. GST was picked as a topic as at that GST was implemented by Malaysian government as part of new tax reform. The period was selected to determine the behavior and the spread of news before implementation of GST and after implementation of GST in Malaysia which was on the 1st of April 2015 (The Star, 2013). The motivation behind choosing GST is because it was a highly debated topic and a lot of people took it to social media to share their views. Majority of tweets were in Malay language, which is 67%, followed by tweets in English which resorts to 30%. The rest of the tweets were in other languages such as in

Chinese, Russian and Spanish. Figure 3.2 shows the breakdown of the language. Majority of the tweets were in Malay language as GST was a highly debated topic in Malaysia.

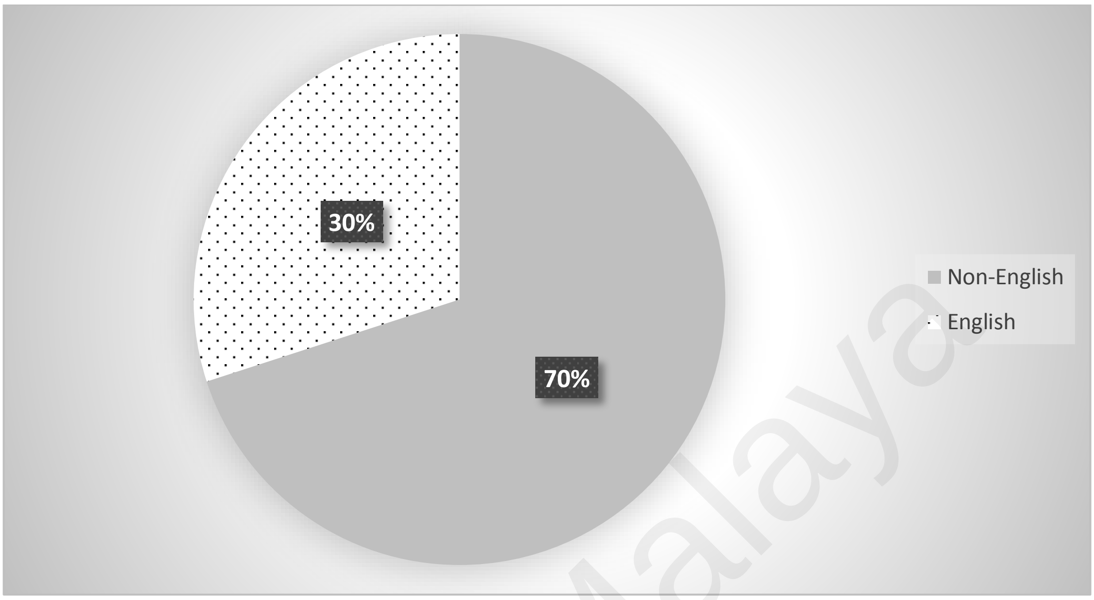

Figure 3.2: Breakdown of Tweets Based on Language for #GST

3.2.2 Verification Dataset

To verify that the scores of the system are the same, a secondary dataset was selected. The dataset that was crawled is #MarsWaters, whereby it was the discovery of water in planet Mars by NASA. The motivation behind picking this topic is because to select a topic which is not related to economy to determine if the scores of the system and the behavior of the system would be the same. Since it was an international topic, most the tweets were in English. Figure 3.3 shows the total breakdown of the tweets. The tweets were crawled from 28th September 2015 to 13th October 2015. 65% of the tweets were in English and the remaining tweets were in languages such as French, Spanish, Arabic and German.

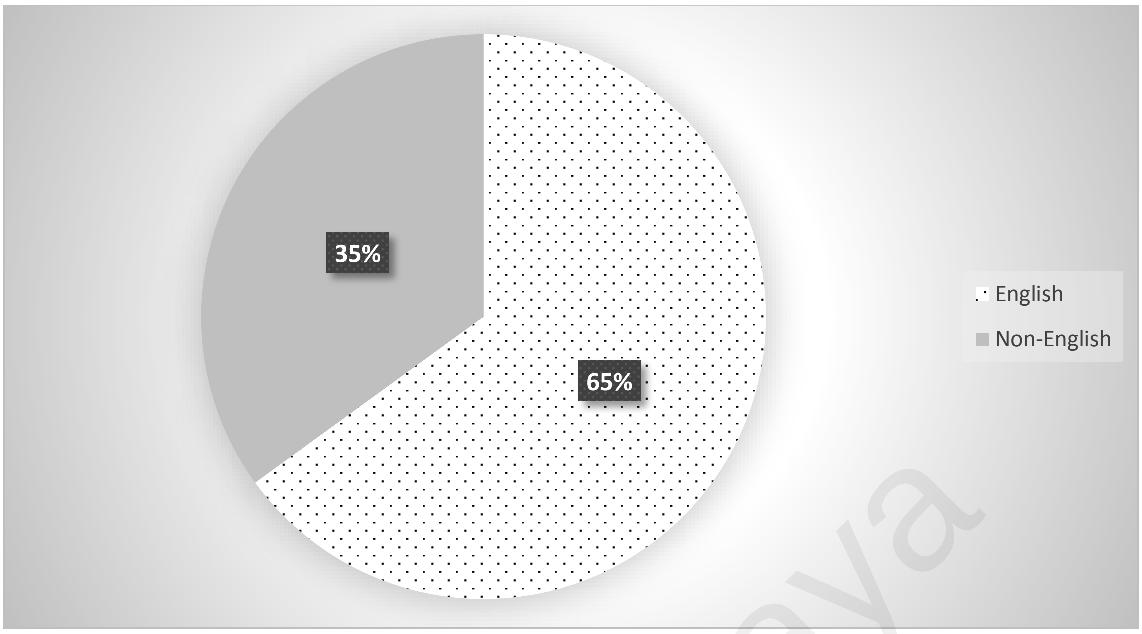

Figure 3.3: Breakdown of Tweets Based on Language for #MarsWaters

3.2.3 Collection & Extraction of Tweets

To gather the data from Twitter, a crawler was used together with Twitter API. Twitter allows application developers to access the data it stores. A crawler is generally used by search engines to index websites. It works by taking a seed value, basically a URL which is known to have links and then the crawler starts to crawl all the links available and the text, once it has crawled it then parses it and then stores it locally (Bahrami, Singhal, Zhuang, 2015). The same concept applies for Twitter crawler instead of crawling the website, it crawls tweets the crawler which was designed for the Twitter API, which is known as REST API. REST API allows developers to access the core Twitter data, allowing developers to retrieve tweets, user information and even timeline information.

Twitter by default throttles the number of tweets which it can be retrieved, which is 800 tweets per request with a maximum tweet of 150 or 350 tweets depending on the developer access level[1] . The restrictions were in place to prevent abuse of Twitter service. To overcome this, tweetf0rm[2] was used as the crawler. It is an open-source software which employs multithreading and multiprocessing with the ability to choose proxy, thus allowing it to be used in distributed computing. Figure 3.4 describes on how this was done using tweetf0rm. To do this two instances of Amazon EC2 servers were employed whereby if one of the instance hits the limit, the other instance continues to crawl. For this research, it was limited to two instances of Amazon EC2 servers due to cost as Amazon charges time-based usage and storage-based usage. This overcomes the restriction placed by Twitter as Twitter sees them two separate crawlers running on two separate machine and allowing to crawl more tweets then the limit was in placed.

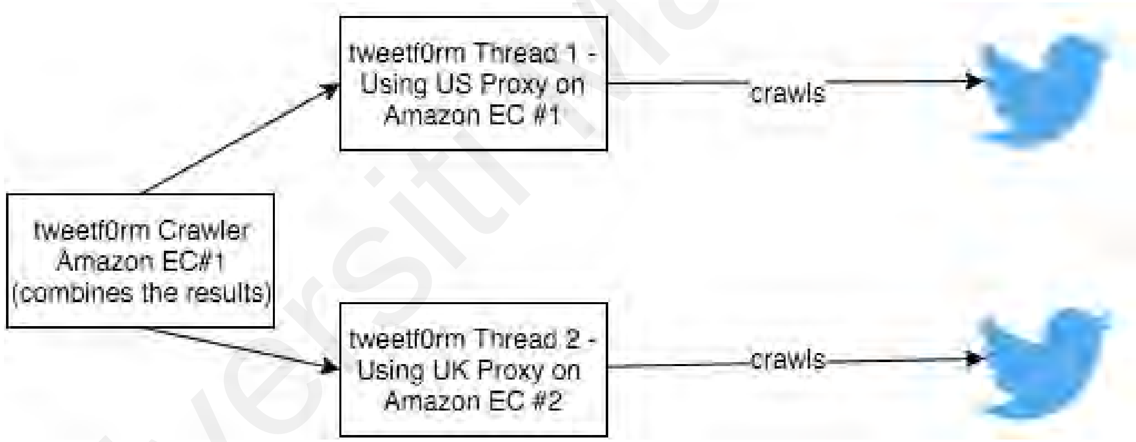

Figure 3.4: Using Multiple Crawlers to Circumvent Twitter's API Limitation

[1] https://dev.twitter.com/rest/public/rate-limiting

[2] https://github.com/bianjiang/tweetf0rm

In addition to crawling Twitter Data, URLs within the tweet and in profile are validated by accessing them via the crawler. Should there be an error such as Error 404 or Error 500, then it would be marked as false, however if the crawling of the website is successful, it would be marked as true. This is used later as part for calculation which is explained in detail in the subsequent section.

3.2.4 Feature Identification

This research uses the all features that were discussed in past literature with one exception, which is the content of the tweet. The reason to exclude this feature is because it was not feasible for the scale of this research, although it will be investigated at a later point. However, URL string from content of the tweet is used to determine the status of URL as in additional to that, to answer our research question of ranking tweets by being neutral to any language and not biased. In this research introduces two additional new features that are Aging of Tweets and Twitter client.

Figure 3.5 shows the features that were selected and used in this literature. The only exception on the feature selection was on the content of tweet. This was not selected because as part of the research objective, this research plans to investigate in identifying and ranking features without being bounded to any specific language. Secondly, it would be investigated at a later stage of time.

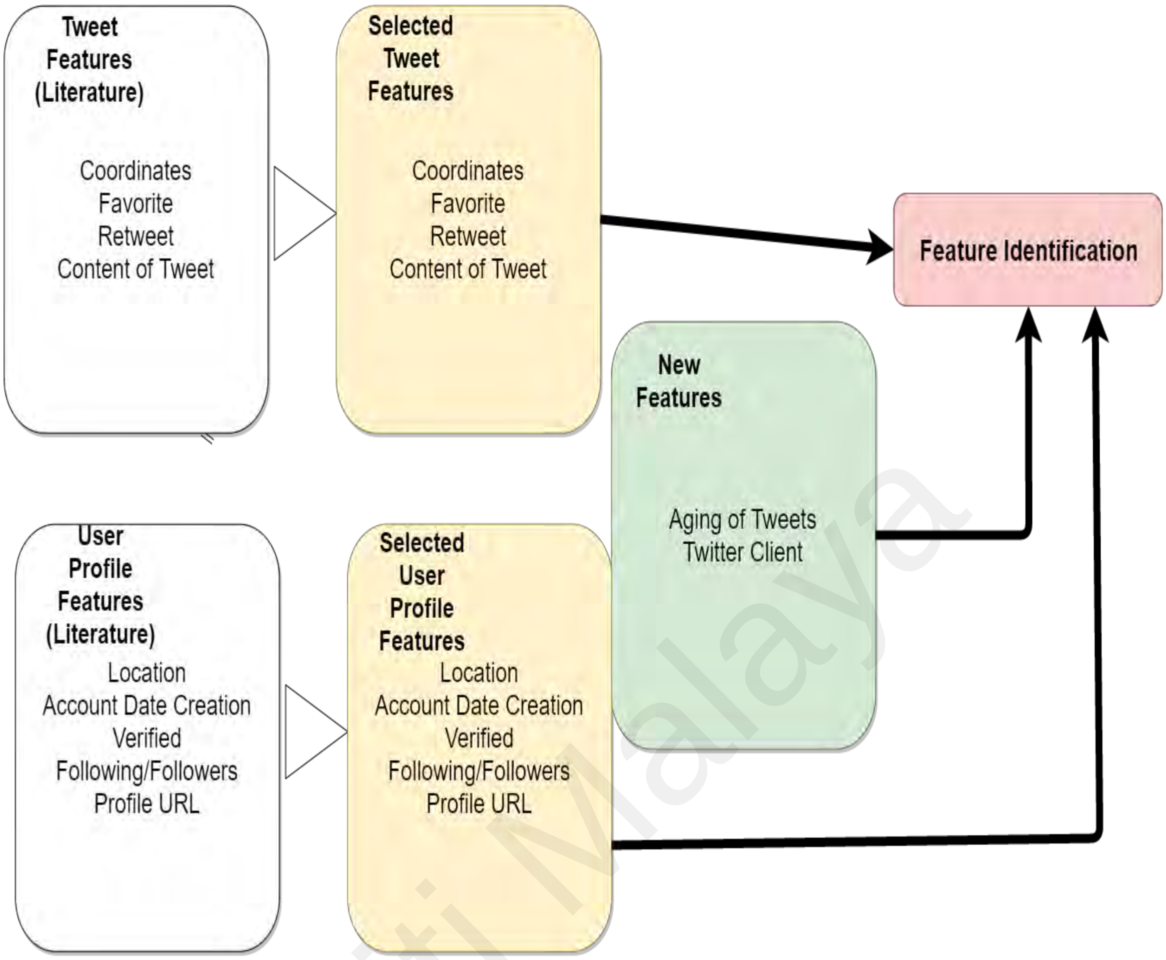

Figure 3.5: Feature Selection To Be Used in Experiment

The features were described earlier in Chapter 2 and the variables used to calculate the feature score are described later in this Chapter to perform the calculations.

3.2.4.1 Aging of Tweets

Aging of tweets is used to determine the time lapse between a person's tweet (Boyd, Golder, Lotan, 2010). As generally an active person in Twitter would be tweeting at least every day. The motivation behind this feature to be used comes from the realm of credibility detection in e-commerce whereby Chun Wang (2008) used Time Decay Based Ranking (TDBR) to penalize old comments, which were posted a long time, which may influence the credibility. Likewise, the same temporal information can be used to use to determine if a person is dormant or active in Twitter, which would make an important metric. In our research, a person is said to be dormant if he or she has not posted a tweet within a period of 2 days. This was because a 30-day sampling period was done during World Cup Final in 2014 and it was noticed that on average out of 661 million users, only 303 million tweets were posted, which roughly equates to 2.1 tweets per user. (Edwards, 2016).

3.2.4.2 Twitter Client

Twitter Client is a software that is used to post tweets and in our research considered as part of a feature as it helps in determining if a tweet was posted by a human or a robot. Often, spammers use their own Twitter Client through the API provided by Twitter to quickly post and replicate the same message (Stieben, 2013). In this research, a list of twitter clients was selected to be as a trusted twitter client as this was based on a study that conducted to find out popular Twitter Clients that were used by people (Mayo, 2012). The list was then validated with list of popular Twitter clients used by Twitter Influencers (Maximise Social Business, 2013), as these sets help to provide a list of popular clients that were used. The clients that did match on both list were selected to be the trusted Twitter Clients.

3.2.5 Storing & Building Relationship Table

Once the tweets are crawled, it is then stored as JSON object which would be used for processing and storage of tweets. Figure 3.6 shows a typical response of JSON object.

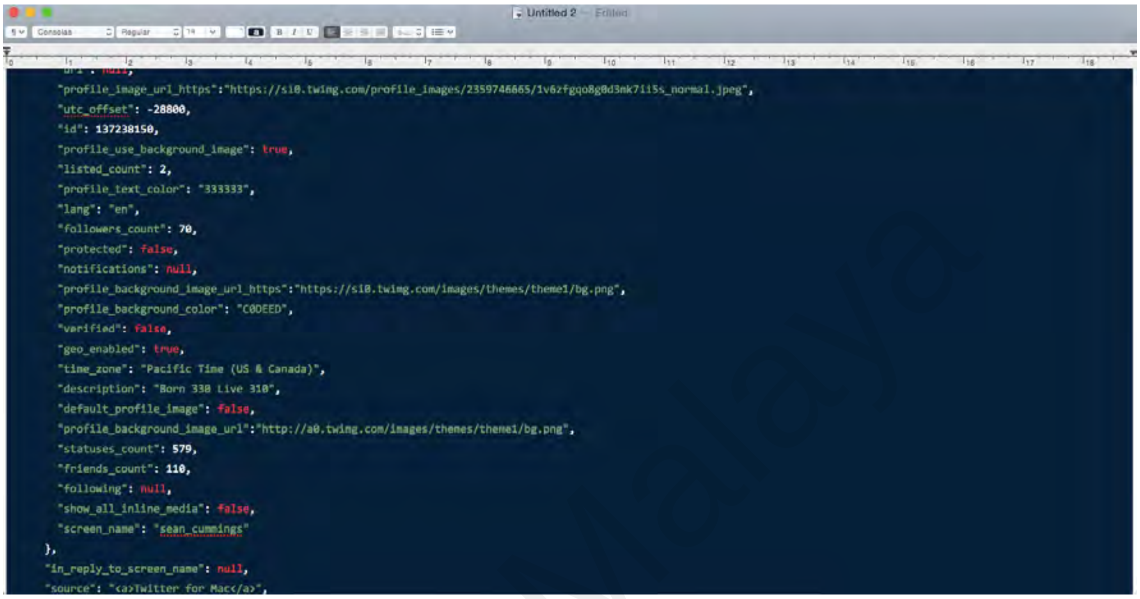

Figure 3.6: A Typical JSON Format Response from Twitter which includes all the data relevant about the user and the tweet

A JSON Object contains a key-value pair, whereby for every key there would be a value. This in result makes it much easier to be processed and stored, as all the JSON Object contains the same set of keys (Novak, 2004).

To store the JSON response, MongoDB was used. MongoDB is essentially a noSQL database and essentially the data structured is stored as the same way as how JSON object are being stored. This makes it easier to perform calculation without the needs of performing a lot of conversation (Bae, Han, Song, 2014). Figure 3.6 below shows the ease of using objects from MongoDB into programming. For example, to retrieve the ID or Twitter User Profile from the JSON, it would be as simple as to use getDB(Tweets).obj(ID). The ID here corresponds to the Twitter ID found in the JSON response in Figure 3.7 below which is 37539828.

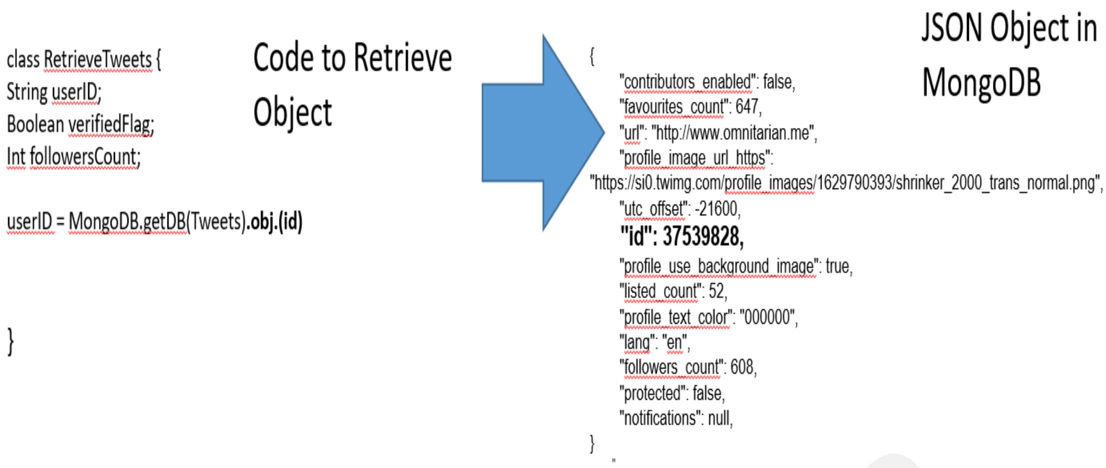

Figure 3.7: Object Retrieval via JSON using MongoDB

With the ability to easily retrieve objects, the data is then indexed to optimize performance. The general principle of how the system is indexed is shown below in Figure 3.8.

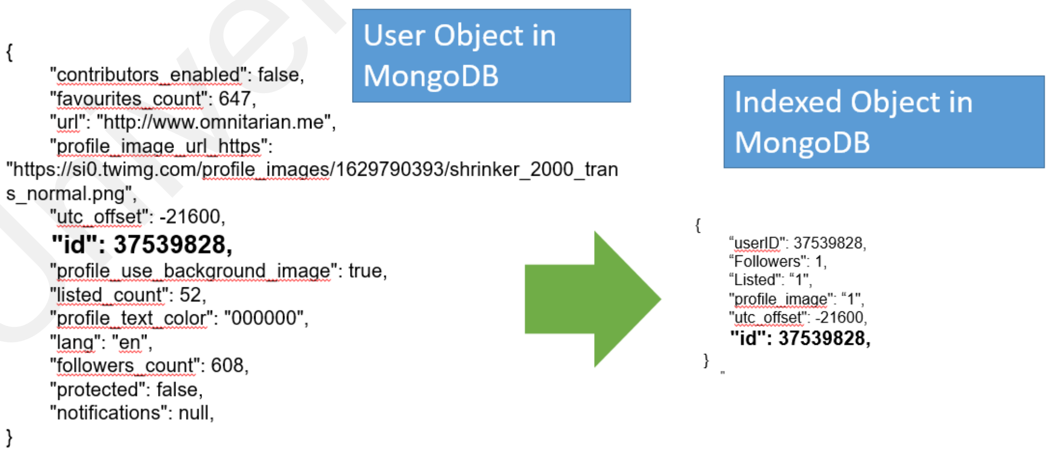

Figure 3.8: Indexing of Objects in MongoDB

Suppose if a feature is presented in a Twitter User, it is then given a score of 1, otherwise it is given a score of 0. When an object of a person is retrieved from the database. Indexing JSON data format helps to improve the performance of retrieval systems as opposed to the traditional way of searching and looking an object one by one and retrieving it (Preoutic-Pietro et al., 2012). This would further reduce the computational time that is needed to compute and would help up in speeding the calculation.

Thus, the two indexing tables, which are the following and followers table, and the feature matrix table are built to perform comparison which the information from the tables would be used later in scoring of the users in our system. The detailed table structure of the two tables are explained in the next subsection.

3.2.5.1 Building Follower/Following Table

The follower/following table contains relationship information between two people on twitter. This can be illustrated in Figure 3.9.

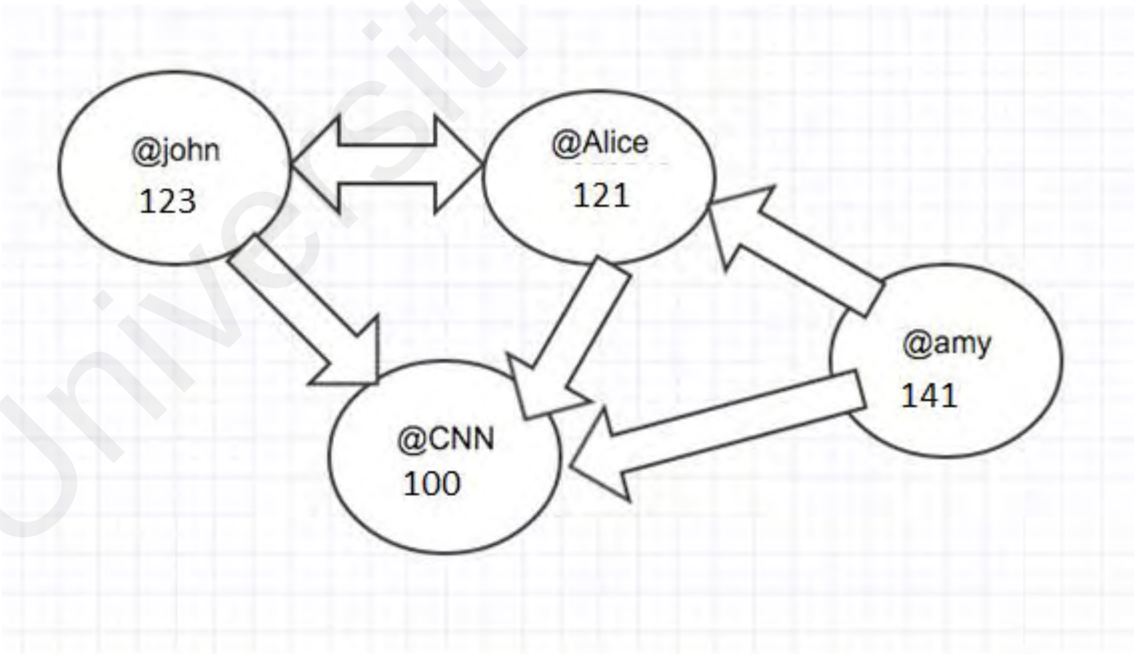

Figure 3.9: Relationship Between Twitter Users showing a One Way Relationship and Two Way Relationship

If John, denoted by his twitter ID of 123 is being followed by Alice who is denoted by her Twitter ID of 121 and she follows back, then it said to be there is a two-way relationship and this is a given of score of 1. However, if there is only one-way relationship, it is given a score of 0, for instance in the case whereby Alice, Amy and John are following CNN, but CNN is not following any of them back. The scores of 1 and 0 indicate the strength of the relationship if it is a one-way or a two-way relationship. The information from Figure 3.9 is then is then mapped back as shown in Table 3.2.

Table 3.1: Follower/Following Mapping Table

Twitter User ID	Following ID	Following Status
123	100	0
	121	1

3.2.5.2 Building Feature Matrix Table

Feature Matrix Table consists of the features that were discussed in Section 3.2.4 The Feature Matrix consists of two sub-tables, one is for the author (user) himself and the other one is for the tweets that the user has posted. The tables are linked together by using User ID.

(a) ***Feature Matrix Table Author***

The Feature Matrix Table extracts the features which were discussed in Section 3.2 in User Profile. A score of 1 is given if the feature is found, otherwise it is given a score of 0. Table 3.3 denotes the feature matrix table.

Table 3.2: Feature Matrix Table

Twitter User ID	Schema_Name	Score
15056260	Created_At	0
	Location	1
	Verified	1
	Profile_URL	1

The list of the features that are used in Table 3.3 is from the works of past literature, all the features that were used are continued to use here as they were used extensively. The list of schema name are the abbreviations of the features that were described earlier in Section 2.5. The explanation of each of the schema ID and to its corresponding feature and its value are described as follows

i Created_At (Account Date Creation)

Created At consists of the date when the account was originally created. An account is said to be a spam account if it is less than 2 days from the date of the tweet (Kurt,Chris et al., 2011). It is considered as a spam and then it is marked as 0.

ii Location (Location)

Location indicates the location of the person. Ideally a lot of spam or new accounts would not specify a location and would put in a default location. Value of 0 is given to indicate there is no coordinate and likelihood of it being a spam account and value of 1 is given if it contains a coordinate.

iii Verified (Verified)

Verified indicates if the account has been verified by Twitter authorities. Twitter generally conducts verification of celebrities and well known figures around the world (Stever & Lawson, 2013). This is to combat against parody account or people impersonating as them. If the account is said to be verified, it is given a value of 1, otherwise it is given a value of 0.

iv Profile_URL (Profile URL)

Profile URL indicates if the current person has a URL linked to the profile. This is done by scanning the said URL if it has a link back to the person's Twitter account. This is to verify if the website the person has put is his Twitter profile. If there is not, it would be given a score of 0.

(b) ***Feature Matrix Table for Tweets***

Feature Matrix Table for Tweets consists of atomic level of tweets posted by author. These two tables are linked by using Twitter User ID as their primary key.

Table 3.3: Tweet Feature Table

Tweet User ID	Tweet ID	Schema_Name	Score
15056260	638219521388470272	Source	0
		Coordinates	1
		Favorited	1
		Statues_Count	1
		Retweeted	0
		URL	1

Table 3.4 shows the Feature Matrix Table and the list of features. The list of features is from Tweet Feature from the past literatures, with additional of the feature that was introduced the list of schema name are the abbreviations of the features that were described earlier in Chapter 2.5. The explanation of each of the schema ID and to its corresponding feature and its value are described below

i Source (Twitter Client)

Source indicates the Twitter Client that was used by the person when the tweet was posted it is given a score of, if it matches list of twitter clients that was selected. The list of Twitter Client was based on a study that conducted to find out popular Twitter Clients that were used by people out there (Mayo, 2012). The list was then validated with list of popular Twitter clients used by Twitter Influencers (Maximise Social Business, 2013) to provide a comprehensive list. If the client is not in the list it is given a score of 0

ii Coordinates (Coordinates)

Source If the tweet contains coordinates or the location of the person, it is given a score of 1. This is to indicate that the tweet has coordinates and it can be used as a measurement to measure credibility of a tweet

iii Favorited (Favorite)

If the tweet has been favorite by people at least more than 5 times, it would be given a score of 1 as it shows that people agree or acknowledged the tweet (Can et al., 2015), if the number is not met, a score of 0 is given.

iv Retweeted (Retweet)

If the tweet has been retweeted at least more than 5 times, it would be given a score of 1, as it would have gain momentum and there is a reach out (Can et. al., 2015) as the tweet that retweeted could have been agreed or disagreed by other party. If it is not, otherwise a score of 0 would be given which indicates that it could be just a chatter or conversation.

v URL

In our research, URL is used as an additional measurement to check for spams and for credibility. In previous research this is done together as part of content of the tweet to verify it (Santos et al., 2014). Suppose if the page is available and is active, it is given a score of 1 otherwise if it cannot be found and error is returned such as Error 404 for Page Not Found, then it is given a score of 0.

vi Statues Count (Aging of Tweets)

Statuses Count indicate on the activity stream of the person or the aging of the tweets. If the person does not tweet in a 2-day period, it is given a score of 0, this is because of the retweet pattern that was observed in behavior of the user. The value of 2 day was selected due to the phenomenon of user behavior and is described further in Section 3.3.1.

3.3 Building pTRank

In our approach, which combines both ranking based approach and on credibility to calculate the overall score of credibility for a user which is known as pTRank. The idea behind this is to harness the strength of user based features and as well as the diversity of ranking algorithm to form the basis of our scoring system. To calculate the rank of the user, the algorithm would be defined to be as Influence(X), where X is the user. This is because in the past literature such as the works of TwitterRank (Jianshu et al,2010), Influence was the key method that was used in ranking, hence the name. In additional to that, to measure credibility of the person, the trust matrix scoring system which is known as Trust(X) is calculated. The scores are later combined to become the evaluation score of the user which would be known as Credibility(X), where X is the user. The system that calculates the credibility(X) score is known as pTRank. The motivation behind this is that in the past literature credibility scores did not factor in ranking. In this subsection, the scoring algorithm, merging and smoothing algorithm, which is used to merge both the scores of Trust(X) and Influence(X), is explained in-detailed.

3.3.1 Calculating Influence(X) Score

The person's influence score would be used as the base scoring system. Generally, the term influence means the ability to "*change or affect someone or something and the power to cause changes without directly forcing them to happen*" (Meltwater Inc, 2014). However, there is a challenge in the world of Twitter, a person or a celebrity such as

Justin Bieber[3] or Donald Trump[4] do have millions of followers, although it shows that these people have a huge following group, it would be effective to measure the people who are engaging (i.e communicating with the person) such as retweeting and replying.

In order to provide a better scoring system of ranking people in Twitter which is currently based on the number of followers that the person have rather than using influence to measure how good people engage (Tunkelang, 2009), which is known as the TunkRank proposed a model as shown in Equation 3.1:

$$Influence(X) = \sum Y\epsilon Followers(X) \frac{1 + \rho \cdot Influence(Y)}{|Following(Y)|} \tag{3.1}$$

Where {X,Y} are two individual twitter accounts and where p is the constant probability that X will retweet the tweet from Y once he has read it. $Followers(X)$ denotes the number of person who X follows and their user ID, whereas $Following(Y)$ denotes the number of person who is being followed by person Y and their user ID.

The model's assumption is that that there is less value in a person who follows a lot of people, as he/she would not have the time to read the tweets and thus have a lesser chance.

To give an illustration and to provide a better understanding of the model. Let us assume that there are two people, Ali and El, El follows Ali but Ali does not follow El. Ali has 2 followers, one is El and one is Miya. . Both have no followers. This view is represented in Figure 3.10.

---

[3] https://twitter.com/justinbieber

[4] https://twitter.com/realDonaldTrump

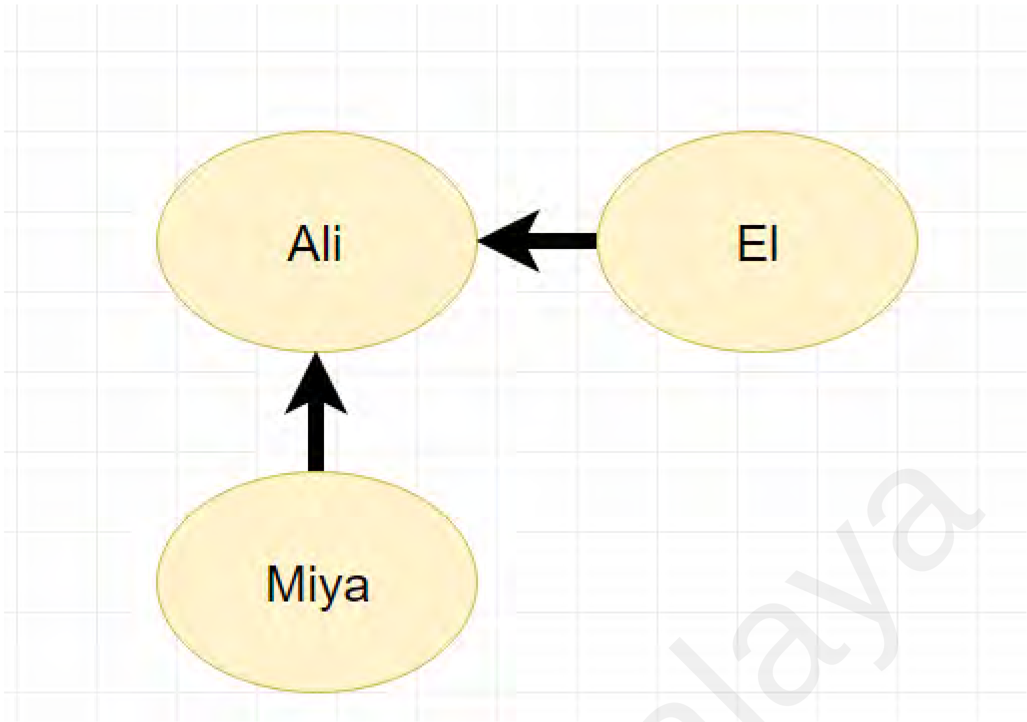

Figure 3.10: Calculating Ali's Influence Score

To calculate Influence score for Ali, with the assumption that probability of someone retweeting the tweets is set to 0.05.

$$Influence(Ali) = \sum \frac{1 + 0.05 \cdot Influence(Miya)}{|2|} + \frac{1 + 0.05 \cdot Influence(El)}{|2|} \quad (3.2a)$$

$$Influence(Ali) = \sum \frac{1 + 0.05 \cdot 0}{|2|} + \frac{1 + 0.05 \cdot 0}{|2|} = 2 \quad (3.2b)$$

From Equation 3.2a, to calculate the Influence score for Ali, the influence score for Ali and his followers needs to be calculated. When the scores for Ali's followers, Miya and El are calculated it is 0, is because both do not have any followers. Hence, in Equation 3.2b, the Influence(X) score for Ali is 2 Higher scores indicate that the person has higher influential capability as his retweets can reach a much wider audience. Examples of people having very high scores include US President, Barack Obama (Greenberg, 2010).

As discussed in the literature review, one of the shortfall of the TunkRank model is that, the assumption of probability of retweeting is constant. However, this is not the case in real-world scenario (Donlinar, 2014). One of the key factor of retweets is timeliness, this is because tweets have short life time and the probability of retweets are higher if it is within the first few minutes of the tweet was posted. This is shown in Figure 3.11 on how the rate of retweet is impacted by passage of time whereby the original formula does not take into consideration.

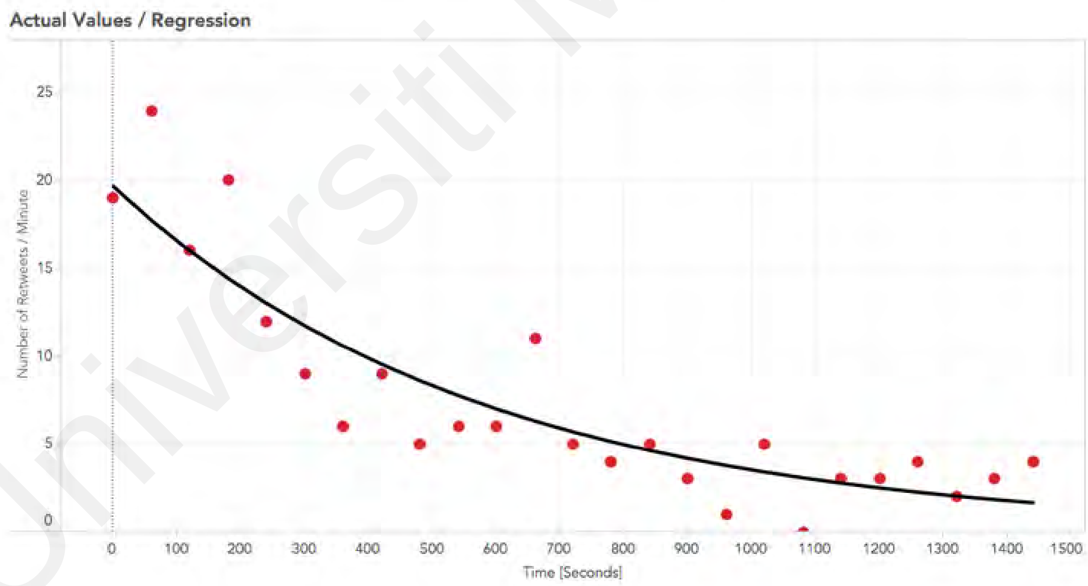

Figure 3.11: Retweet count as passage of time. Retweets are high during initial posting of the tweet (Donlinar,2014)

3.3.1.1 Calculating Passage of Time of a Tweet

To create an accurate representation of the Influence score, the probability of aging of tweets to be considered. Aging of tweets can be represented by the having a projection of probability of retweet by considering of passage of time and as well as behavior of outside the Twitter User's network. This is because as time passes by, the probability of someone retweeting goes lower. This same measure can be used to determine an active user or active topic in Twitter, which would strength the Influence score. Miller (2015) proposed the following equation (Equation 3.3)

$$\rho = \frac{1}{2\bar{n}}\left(\sqrt{\left(\frac{P(1,\lambda t)}{P(2,\lambda t)}\right)^2 + 4\bar{n}\frac{R(t)/n_1}{P(2,\lambda t)}} - \frac{P(1,\lambda t)}{P(2,\lambda t)}\right) \tag{3.3}$$

Where

$$P(1,\lambda t) = 1 - e^{\lambda t}$$

$$P(2,\lambda t) = 1 - e^{\lambda t} - \lambda t e^{\lambda t}$$

λ is the reload rate whereby each user read through all the tweets and decides to retweet them. The reload rate is modelled after a Poisson distribution. $P(1,\lambda t)$ and $P(2,\lambda t)$ denotes the probability of first-order followers, which means the followers that are followed by the person and second order followers, which are followers outside of the person's network retweeting them.

In this research, the mean value of followers, $\bar{\bar{n}}$ is set as 90.9 (Myers et. al., 2014) and the value of λ is set to be as 1 hour[5]. The idea behind setting the value of λ is to be aligned with the Twitter API refresh limit whereby Twitter allows a maximum refresh of an hour.

---

[5] https://dev.twitter.com/rest/public/rate-limiting

An example of using time-based ranking is shown in the equation 3.4 and equation 3.5 respectively. Suppose, there are two tweets, which are Tweet A and Tweet B and it was posted by Ali and El. Respectively Tweet A was published by Ali. Ali had 1000 followers and within the 30 minutes' period, he did manage to get 25 retweets from his followers, whereas Tweet B which was posted by El 30 minutes ago, did manage to get 25 retweets. El has about 2000 followers. With the same assumption of, $\bar{\bar{n}}$ is set as 90.9 and reload rate, λ is set to be 1 hour and using the formula above

$$\rho(Ali)^{\alpha} = \frac{1}{181.8}\left(\sqrt{\left(\frac{P(1,1)}{P(2,1)}\right)^2 + 4(90.9)\frac{25/1000}{P(2,0.5)}} - \frac{P(1,1)}{P(2,1)}\right) = 0.02167 \tag{3.4}$$

$$\rho(El)^{\alpha} = \frac{1}{181.8}\left(\sqrt{\left(\frac{P(1,0.5)}{P(2,0.5)}\right)^2 + 4(90.9)\frac{25/2000}{P(2,0.5)}} - \frac{P(1,0.5)}{P(2,0.5)}\right) \tag{3.5}$$

$$= 0.02183$$

From the calculations, tweet that was posted by El as stated in Equation 3.4 is the winner, although the tweet posted by Ali comes very close. The gap would be much wider supposing if Ali tweeted it about 3-4 hours ago. This shows how close the probability of the tweets is being retweeted by Ali or El.

In this research, the time, is calculated as the average of all the time of the tweet posted by the user X which was compared with the timestamp of the tweet was crawled. equation 3.6 shows the value of how the value of t is calculated.

$$t(X) = \frac{\sum(\Delta t_c - \Delta t_p)}{Tweets(X)} \tag{3.6}$$

Δt_c denotes the timestamp when the tweet was crawled from the crawler and Δt_p is the timestamp of the actual tweet when it was posted by the tweets. Tweets(X) shows the total number of Tweets posted by the user.

3.3.1.2 Calculating Influence(X) Formula with Passage of Time and Retweet

By taking consideration Equation 3.1 and Equation 3.3, with value of the $\bar{\bar{n}}$ is set as 90.9 an improved version of TunkRank formula is derived. The new, equation 3.7 are as follows

$$Influence(X) = \sum Y\epsilon Followers(X) \frac{1 + \rho^{\alpha} \cdot Influence(Y)}{|Following(Y)|} \tag{3.7}$$

whereby the value of ρ^{α} is described in Equation 3.8a and 3.8b respectively. The formula stated in 3.8a is the formula that was propose by Miller (2015) described in Equation 3.3

$$\rho^{\alpha} = \frac{1}{2\bar{n}}\left(\sqrt{\left(\frac{P(1,\lambda t)}{P(2,\lambda t)}\right)^2 + 4\bar{n}\frac{R(t)/n_1}{P(2,\lambda t)}} - \frac{P(1,\lambda t)}{P(2,\lambda t)}\right) \tag{3.8a}$$

$$\rho^{\alpha} = \frac{1}{181.8}\left(\sqrt{\left(\frac{P(1,\lambda t)}{P(2,\lambda t)}\right)^2 + 4\bar{n}\frac{R(t)/n_1}{P(2,\lambda t)}} - \frac{P(1,\lambda t)}{P(2,\lambda t)}\right) \tag{3.8b}$$

The value of t is described in Equation 3.5 and the value of λ is 1 hour and the value of $\bar{n}$ is 90.9. The possible values that Influence(X) is between 0 to 1, whereby 1 being the most influential and 0 being not influential. The new equation forms the basis of Influence(X) score that would be used to calculate the Influence score of a person.

3.3.2 Calculating Trust(X) Score

Calculating the Trust score would tell us on how credible the person is and their followers based on the features that were identified and discussed in Section 3.2.4

The scoring of Trust(X) was based on the modified PageRank algorithm by Jianshu et al, (2010), which was used in TwitterRank. PageRank link-based analysis algorithm used by Google to rank webpages (Page, 1998). An overview of TwitterRank and PageRank is discussed in-depth in Chapter 2. Although the concept of the algorithm from TwitterRank was used and it was then further enhanced to address one of the limitation of TwitterRank which is using seed list or a pre-populated list of influencers. In this research, this is addressed by utilizing User Profile features. In the section below, the issue of tackling damping factor and modified TwitterRank algorithm is discussed.

3.3.2.1 Damping Factor

$$p = \frac{1-d}{N} \tag{3.9}$$

The problem with PageRank and TwitterRank is that any graphs solution, issue arises on cyclic graph. It creates two issues, firstly whereby an infinite loop is created by traversing the whole graph over to calculate the score and there is no way for it stop. Secondly, in the case of PageRank whereby it would create an infinite scoring system as scores are calculated repeatedly as there is no way to stop' the algorithm, which would cause the score to be inflated which is known as dangling nodes. To tackle this, Page et. al. (1998) introduced damping factor to prevent the cycles to be in infinite loop as shown in Equation 3.9, whereby the constant d which is a probability that would help to skip the node which is not in path and it is also known as the teleportation factor. In other words, teleportation factor means 'jumping' exploring to nodes whereby there are no any path associated to it. Damping factor is calculated by subtracting its value from 1 and then dividing with the number of nodes, N to normalize. The ideal value for d, damping value is set to be 0.85 (Abdullah, 2004). This is the exact same value that was used in TwitterRank and modified TwitterRank algorithm, thus it would be used in this research as well. (Jianshu et al., 2010; Xiong, 2013).

3.3.2.2 Calculating Trust(X) Score Using Modified TwitterRank algorithm

The original formula of TwitterRank is written as below in equation 3.10

$$Tscore_{t+1}(X) = p + e \sum_{Y \in R_X} \frac{TScore_t(Y)RN_{XY}}{N_y} \tag{3.10}$$

Whereby p is the damping factor and e is the Damping value, R_x denotes the people that a user, say X follows, $RN_{XY\ are}$ the popularity score of both X and Y which are added and combined. The popularity list in the case of TwitterRank is obtained from TwitterCounter. (Jianshu et al., 2010). The popularity score can be anywhere from 0 to 100 whereby 100 being the most popular. Lastly N_y denotes if Y has that feature, for example if Y is a verified account. The scores are calculated until the individual score is stabilized. Then the algorithm proceeds on calculating of all the pairs of individual scores which are based from the list.

However, one of the downside of TwitterRank as stated is the 2 is the use of external influence list instead of calculating the score each user individually. The usage of external list influences the score and requires the use of an external source provider which may change from time to time and is not reliable (Zubiaga et. al., 2015). In this research's approach, the equation is changed to use the richness of features that were discussed in Chapter 3.2. Therefore, to support this, Equation 3.11 was modified from Equation 3.10 as below to cater this this, the modification was to replace RN_{XY} with TN_{XY} which factors in all the features that were discussed.

$$Pscore_{t+1}(X) = p + e \sum_{Y \in R_X} \frac{PScore_t(Y)TN_{XY}}{N_y} \tag{3.11}$$

where e is the Damping factor, R_x denotes the people that X follows, TN_{XY} are the number of features that both X and Y have in common which are features described in described in Section 3.2.5.2 (b). Lastly N_y denotes if Y has that feature, for example if Y is a verified account. The above equation is calculated and repeated for every individual in the Twitter user list until the scores have been stabilized for that feature, in this example it is on verified account. Then it is repeated for the 6 features that are mentioned in Section 3.2.4 which are account creation date, frequency of tweets, verified profile, the location until all the features are calculated and accounted for.

The original algorithm of PScore is further improved by including a factor of weightage. The motivation and the reasoning behind of including weightage is to determine how strong or a weak the features which were described earlier to further improve the scoring and ranking.

With this understanding, a new equation is derived from Equation 3.11. This is done by adding a new weightage factor,ω. The new equation formulates the Trust(X) score. Equation 3.12 shows the equation

$$Trust(X) = \sum_{x=k}^{X} Pscore_{t+1}(X).\,\omega \tag{3.12}$$

where X is total number feature, k denotes the current feature that was selected, and PScore is the score that was calculated earlier and finally weightage factor,ω. The sum of all individual weightages needs to equal to 1, in other words Trust(X) scores have a value between 0 to 1.

In order to find the ideal weightage factor ω for each of the feature which are mentioned in Section 3.2.5.2 (b), as part of the Feature Matrix Table. A simple assumption was made that all of the features have an equally important weightage, since there are only 6 features , thus giving an equal score of 0.1667 for each . To illusrate this, Figure 3.12 shows the weightages of each of the features.

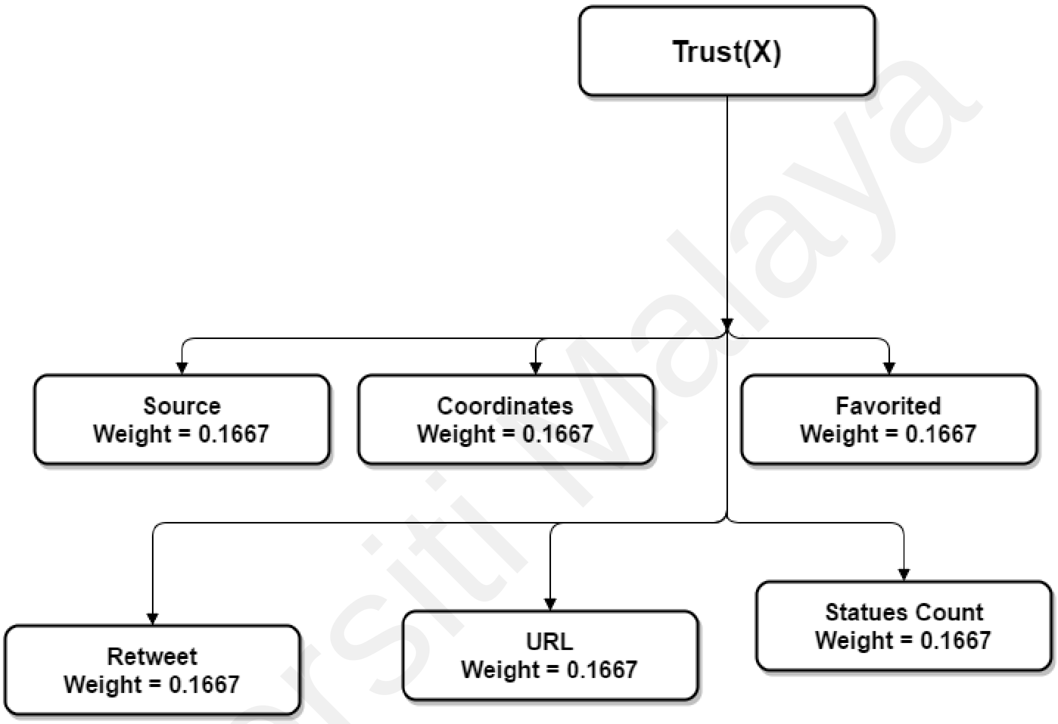

Figure 3.12: Various Weightage Weight, ω against the Feature Matrix Table

However, upon looking at how difference feature that was weighted differently in past literature calculation (Castilo et al., 2010; Gupta et. al., 2012; Nuo Li et. al., 2015), the same principles were applied in this case. Figure 3.12 shows the list of features that were given a higher scoring or weightage as part of algorithm in past literature, it has one exception which is for Statues_Count (Aging of Tweets) , Source (Twitter Client) features and URL (feature) .In this case, it is given as the highest scores weightages as these were two new features that were introduced in this research. .

To calculate the various weightages, found in past literature, first the feature is given a point of 1 to all the features indicate it was present. Suppose if the feature is weighted higher over another feature, for example if retweeted is higher compared to coordinates, then a point of 0.1 is added to retweet and a point of 0.1 is subtracted from coordinates. This is then repeated for all the features in all the literatures that were discussed in Section 2.5.3. This is represented in Figure 3.13

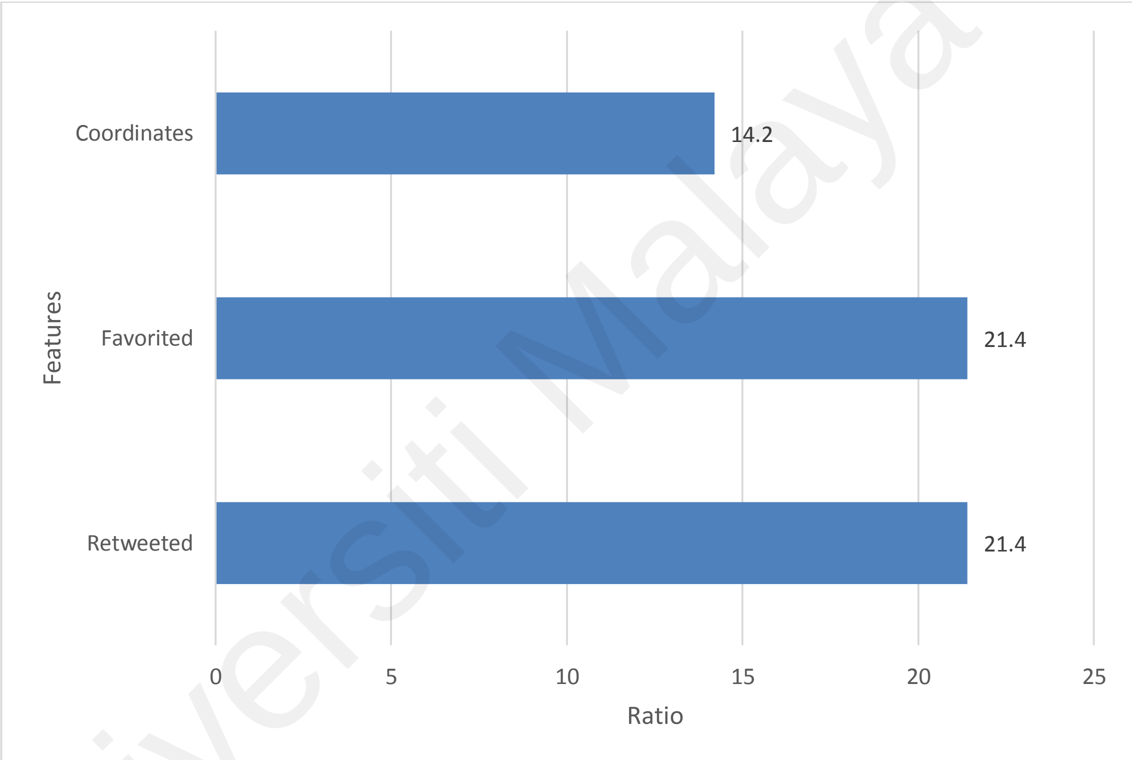

Figure 3.13: Selection of Top 3 Features That Were Used In Past Literature

From Figure 3.14 that both Favorited and Retweet feature has the highest feature score. For the case of the new features which are Source, URL and Statues Count that were introduced, a small set of sample of 100 tweets from the GST and 100 tweets from #MarsWaters dataset were used to determine the weightage of this said feature. If the feature is presented and it matches the criteria of Feature Matrix Table in Section 3.2.5.2 (b), then it is given 1 point otherwise it is given a 0-point Figure 3.14 shows the total scores that were found in the new feature.

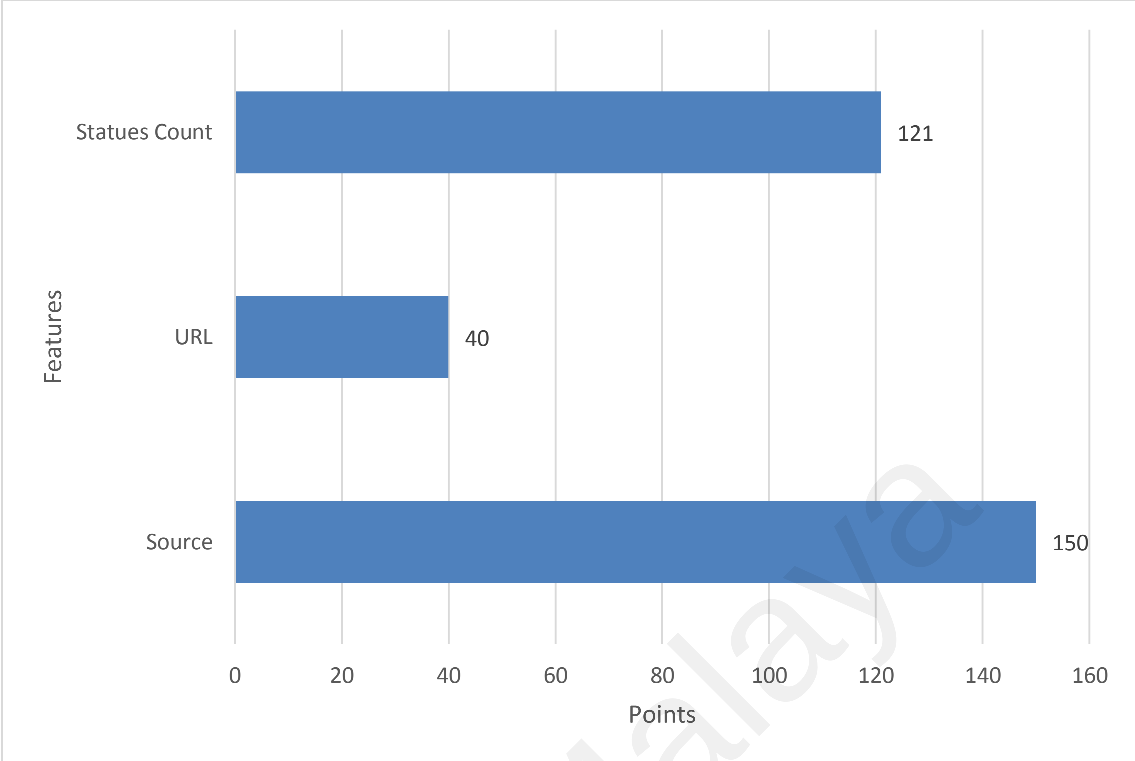

Figure 3.14: Selection of New Features Based On Points

It can be clearly seen in Figure 3.13 that the two-prominent feature that were found are Source and Statues Count as they have yielded the highest points which are 150 and 121 respectively. These two features were found in 95% of the tweets dataset that was sampled. Hence it can be deduced that these are two favorable features that would able to carry the highest.

To assign scores, Turueswell's empirical 80/20 rule arises (Burrell, 1985). In this context, it means that 80% of the scores should be assigned to the highest scoring features that are found. With this principle, the new weightages ω , which would be used to calculate Trust(X) scores are shown in Figure 3.15. The 4 features that are given high equal scores are Favorited, Retweet, Statues Count and Source. The two lowest scoring features are Coordinates and URL as from Figure 3.13 and Figure 3.14, they have yielded the lowest scores. The reason for that are these features were not presented in a lot of the dataset.

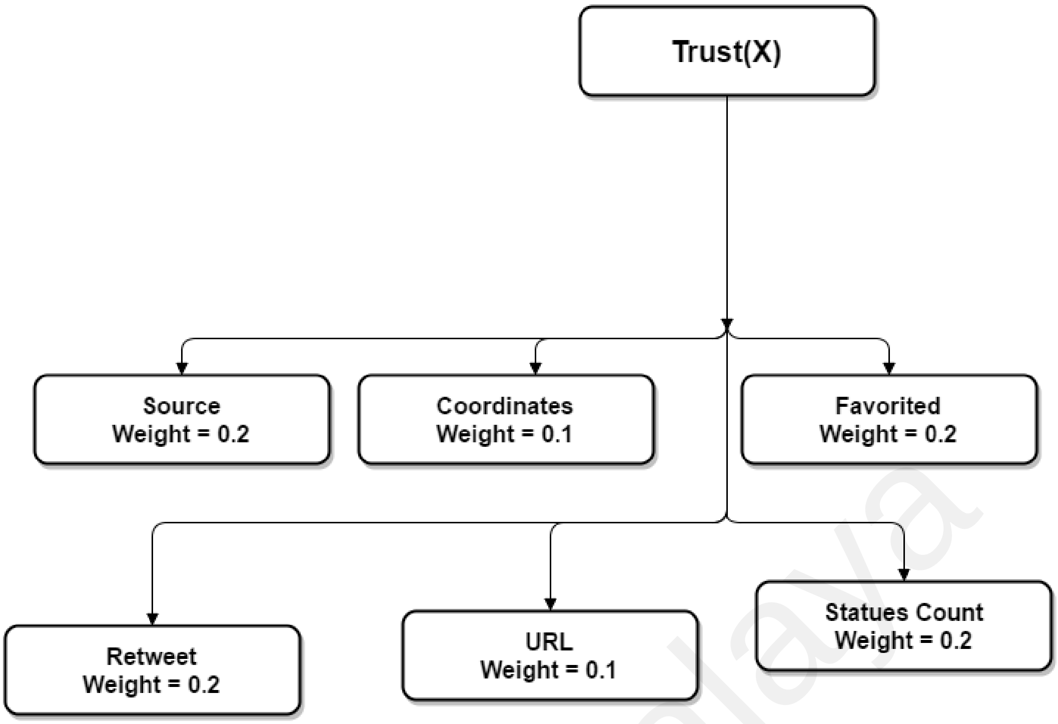

Figure 3.15: Summing up of Trust(X) scores with different weightage

3.3.3 Merging and Smoothing Scores

Trust score and Influence score each produces its own unique score and to provide a holistic score and secondly to answer the research question in calculating a score for the user, smoothing technique (Ravana, 2011) is introduced and used to merge two different scores individual scores for a Twitter User. These two scores are added together with various value of α , alpha to measure the degree of Influence and Trust scores effect the overall scoring as stipulated in the equation 3.13 to calculate the overall Credibility score, which is also known as pTRank, the scoring algorithm proposed in this research.

$$Credibility(X) = \alpha\big(Trust(X)\big) + (1 - \alpha)(Influence(X)) \tag{3.13}$$

where alpha value, α is between 0 to 1

For the purposes of the test and to determine the ideal weightage of systems, several scores of α would be tested. The reason behind this to determine if scoring is influence by Influence(X) or Trust(X) or both are equally. The higher the alpha score which is closing to 1, it is much more favorable to Trust(X), if the value of alpha is closing to zero, then the score for credibility is favorable to Influence(X). The different weightage is used to conduct experiment to prove if the assumption is true. This is discussed later in Chapter 4.

3.4 Creating Evaluation Metric

An important aspect of Information Retrieval system is the evaluation. Evaluation allows different retrieval systems to be compared with one and another using a standardized measure of scoring. Generally, in any information retrieval system, the documents are retrieved in ranked order, whereby the top documents which the system thinks it is relevant based on the user input is rank is on top.

In many Information Retrieval systems, the output is in an orderly binary vector whereby 1 indicates the document that was retrieved did match the criteria and 0 indicates that the document did not match the criteria (Liu, 2009). In this research, three evaluations which are Precision (P), Average Precision (AP) and as well as Normalized Discounted Cumulative Gain metric are used and details of each evaluation metric and the selection for them are discussed next. Normalized Discounted Cumulative Gain (nDCG) is a new evaluation metric that was introduced. This was done to provide an additional scoring method to validate the results of the system. Secondly in this research, documents that are retrieved are considered as users that are relevant to the topic that was retrieved. The concept of obtaining and building relevance judgement are discussed first and later in this section, the evaluation metric which would be used as part of benchmarking the system are discussed in detailed.

3.5 Using CrowdSourcing to Build Relevance Judgment

To provide a neutral un-biased ranking and results and to adhere to the assumptions that are made for most IR test collection crowdsourcing was considered as a platform to obtain an unbiased relevance judgment (Soboroff, 2007). Crowdsourcing is the use of people by enlisting services using Internet (Davtyan et al., 2015). Two major crowdsourcing services which were considered are CrowdFlower[6] and Amazon Turk[7]. For this research, CrowdFlower was selected because of that it is possible to submit tasks from Malaysia. Crowdsourcing works by giving a certain task to a group of people which are known as workers. Each of these workers would have perform a task to complete. Then these workers are paid once the tasks are completed. On the other hand, Amazon Turk has a similar concept to CrowdFlower, however it was not considered and used in this research as it was only opened to people who are in United States and Canada.

3.5.1.1 Dividing of Tweets to Tasks for Crowdsourcing

To submit to CrowdFlower system to be judged by people who are known as workers. The workers judge the top 1,200 Twitter Users for the topic #GST and top 200 twitter users for #MarsWaters. The number of users for the system that needs to be ranked were obtained from top results from baseline, TwitterRank and from the newly proposed system, which shall be known as pTRank. Explanation of benchmarking with various system are explained in Section 3.7. The technique that was used here is known as the *pooling technique* whereby relevant documents using several retrieval systems are obtained and the top-ranked documents for each of them are combined. This technique is used in Text Retrieval Conference (TREC) to build relevance judgement. In this research,

---

[6] http://www.crowdflower.com/

[7] https://www.mturk.com/mturk/welcome

documents in this context refers to collection of users and the tweets. The advantage that it provides is that relevant tweets and users can be measured within the same task which makes it easier.

Therefore, for this experiment, and to give context to the workers, the top 5 tweets for each of the twitter users were obtained from the systems. Then for each of the twitter user and corresponding 5 tweets are segmented into unit. Each of the units, 2 workers are required to judge the set of documents. Each worker is then asked the question on which of the tweet is credible and asked to provide justification. The questionnaire can be found in the appendix. The questions were validated by using the principles of questionnaire design (Sudman & Bradburn, 1982). As shown in Figure 3.16, for the workers to be paid, 5 units must be completed and these workers would have to answer the questions that are presented.

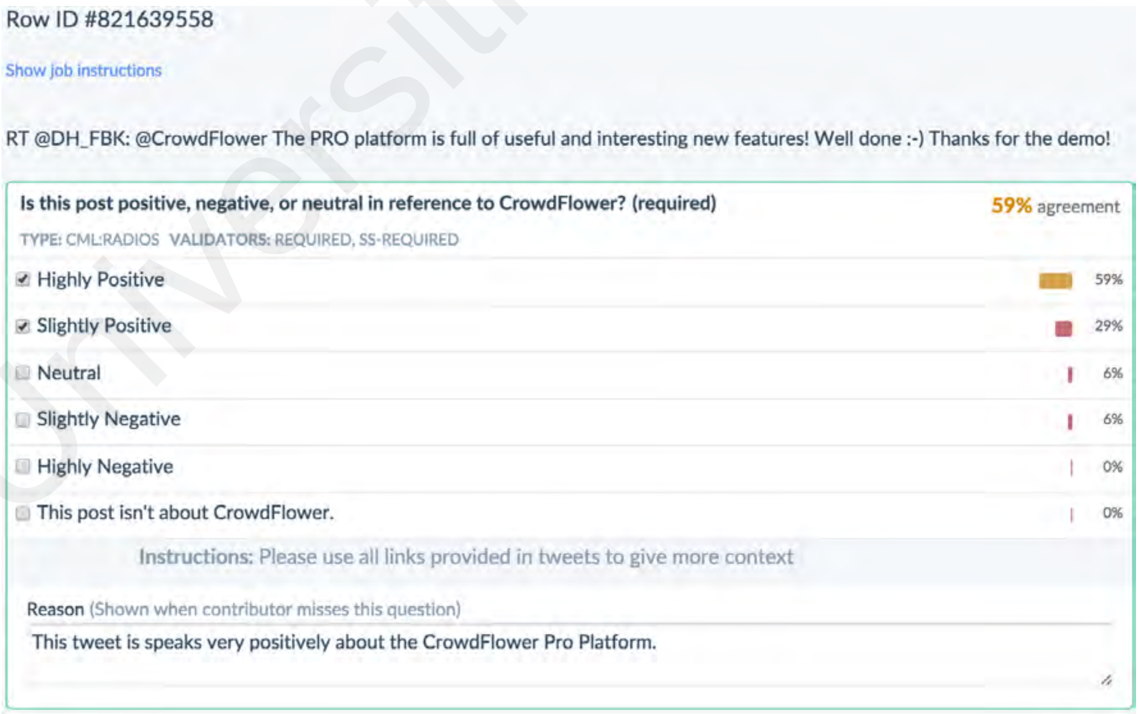

Figure 3.16: Task that needs to be completed by a worker in order to be paid by CrowdFlower

3.5.1.2 Cost and Selection of Workers for CrowdFlower

The primary dataset of this experiment consists of tweets which were in multiple language, and most the tweets were in Malay language. The composition of the languages that were collected are part of Section 3.2. Hence with the various languages part of data collection, the workers needed to be someone who understand the command of the language very well. Therefore, for GST dataset the workers are selected from Malaysia, Singapore and Brunei and those who do understand Malay and English language. As for secondary dataset, since most the dataset was in English, there were no specific criteria, if the worker understands English and the same selection of people were used as well.

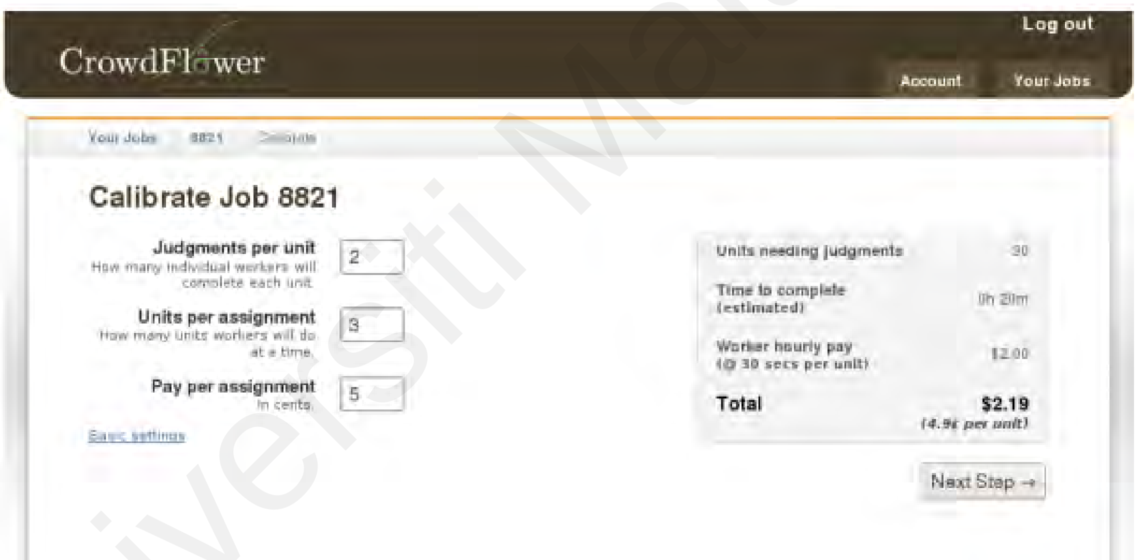

Figure 3.17: CrowdFlower Job Calibration on calculating the number of jobs and the total cost

As for the costs, it costs USD$0.12 per judgment for GST, which accounts to USD$144 and for MarsWaters, it costs USD$0.08 per judgment which accounts to USD$16. Overall this coasted USD$160 for the entire task. This is much cheaper as opposed to having a full-time person or dedicated team doing this, with the scalability, it allowed the tasks to be completed in just 2 weeks. Figure 3.17 shows on the ease of use in CrowdFlower in selecting a job and calculating the task based on a specific need.

3.6 Building Relevance Judgement

Relevance Judgment or which are known as "right answers" are important as it provides how well the set of documents meets the needs of the user of the documents that they are seeking for. There are 5 assumptions which are made for most IR-test collections which are relevance needs to be topical, the judgment must be in binary list whereby a document is said to be relevant or non-relevant. Thirdly, the relevance of a document is not impacted by relevance of another document and it is of independent. Fourthly, relevant judgment is consistent across the judges or people who are judging it. Lastly, the judgment must be stable over time. (Saracevic, 2007).

3.6.1 Compiling Crowd Flower Results to Relevance Judgement

To perform evaluation a relevance judgment file is built. It consists of the score the person is tweeting related to the topic, whereby a score of 0 is given if the person is not relevant and a score of 1 if the person is relevant. Figure 3.18 shows the sample of how these sample look like. The format confines to standard QREL relevance assessment that is used in services such as TREC.

TOPIC	REL	TWEET ID	JUD
0001	C	0000010176	0
0001	A	0000010187	1
0001	C	0000010218	0
0001	C	0000010219	0
0001	C	0000010220	0
0001	A	0000010187	1
0001	A	0000010221	1
0001	C	0000010222	0
0001	C	0000010231	0

Figure 3.18: Relevance Judgment of GST Tweets

Each field is TAB separated. The list has 4 columns which are Topic ID (TOPIC), Relevance Assessment (REL), Tweet User ID (Tweet ID) and Relevance Assessment (JUD). The file is sorted by topic ID in ascending order. The first column represents Topic ID which shows the subject. In this research, there are only two Topic IDs – one is for GST and MarsWaters which is denoted by 0001 and 0002 respectively. Then it is followed by the relevance assessment that was done by humans in three grades which are "relevant (A)", "partially relevant (B)" and non-relevant (C)" (TREC, 2000). After which is the Twitter User ID which is stated and lastly the last column is the relevance assessment which is either 0 for non-relevant and 1 represents relevant user.

With the relevance judgment is in place, it would be used to judge the scoring of systems in Chapter 4.

3.7 Benchmarking scores

To access the effectiveness of the system, following Information Retrieval evaluation method would be used. TREC provided an official evaluation metric to measure the performance of the systems for their social media track which are Mean Average Precision (MAP) and Percision@30 (P30) (Tomlinson, 2013).

3.7.1 Precision

Precision is the popular use of scoring of the performance of information retrieval system (Singhal,2001). It was also used in Jianshu et al. (2010) for TwitterRank. Precision is defined as the total number of relevant documents retrieved over the total number of retrieved items, which can be calculated as follows in equation 3.14

$$Precision = \frac{Total\ Number\ of\ Relevant\ Documents\ Retrieved}{Total\ Number\ of\ Retrieved\ Document} \tag{3.14}$$

Precision makes it very easy for it to be computed as the formula is very simple and straight forward. Generally, in order provide faster results to the user, it is commonly truncated at a certain cut-off point which is known as k. Hence the term P@k., equation 3.15 shows the formula that is used to calculate precision at k Common cut-off points are 10, 20, 30 and 50 respectively.

$$Precision(k) = \frac{Total\ Number\ of\ Relevant\ Documents\ Retrieved\ (k)}{Total\ Number\ of\ Retrieved\ Document\ (k)} \quad (3.15)$$

Precision at 30 (P@30 was selected as well, since it was used in evaluation of TREC microblog retrieval systems. (Tomlinson,2013).

3.7.2 Average Precision

Average Precision (AP) combines the strength of precision and recall. Recall which is also known as sensitivity is the ratio of total relevant documents that are successfully retrieved. The equation for AP is defined in equation 3.16

$$Recall = \frac{Total\ Number\ of\ Relevant\ Documents\ Retrieved}{Total\ Number\ of\ Releveant\ Documents} \quad (3.16)$$

However, one of the challenges with recall is that it is not popular as the formula requires the comparison to be done with the all the relevant documents that were retrieved by the system.

Average Precision addresses this by combining the value of precision and recall into one single value which compromises both aspect. To calculate AP at evaluation depth, k it is as follows in equation 3.17

$$AP@k = \frac{1}{R}\sum_{i=1}^{k} r_i \frac{\sum_{i=1}^{j} r_j}{i} \quad (3.17)$$

where R denotes the total number for relevant documents for the topic, the value of r_i is 1 if the document is relevant, otherwise it would be 0. The inner equation of $\frac{\sum_{i=1}^{j} r_j}{i}$ calculates the precision of top k documents where the documents are relevant.

3.7.3 Normalized Discounted Cumulative Gain (nDCG)

Average Precision takes into consideration of the rank of retrieved documents, however the weightage of each document is not fixed, however it is calculated based on the number of the relevant documents that were retrieved and at which position it was returned in. To have precise scoring on weight of the ranks, discounted cumulative gain (DCG) was introduced to address this. The formula of DCG is defined as in equation 3.18

$$DCG@k = \sum_{i=1}^{k} \frac{r_i}{log_b(i+1)} \qquad (3.18)$$

$$if\ i > b - 1$$

where k denotes the depth of the evaluation and r_i is the relevance of documents at rank i. In the works by Järvelin and Kekäläinen (2002), the value of b was suggested to be 2. It basically measures the importance of the relevant documents to the users. The same value is used for this research as well.

Search results generally vary on the query that was possible and plus comparing two search results system using DCG alone cannot be consistently producing the right result (Croft. Metzler,Strohman, 2010) , hence Järvelin and Kekäläinen (2002) introduced an effective version of DCG that is normalized so that the scores that are generated are in within the range of [0-1]. It provides an effective method of comparing multiple performance of different retrieval system. This is known as Normalized Discounted Cumulative Gain (nDCG) The equation for Normalized Discounted Cumulative Gain (nDCG) is as follows in equation 3.19 where it is represented as a ratio of Discounted Cumulative Gain (DCG) at position k and Inverse Discounted Cumulative Gain (IDCG) at position k

$$nDCG@k = \sum_{i=1}^{k} \frac{DCG@k}{IDCG@k} \tag{3.19}$$

This can be represented in its equational form as shown in equation 3.20, whereby the value of b is set to 2.

$$nDCG@k = \sum_{i=1}^{k} \frac{\sum_{i=1}^{k} r_i w_i}{\sum_{i=1}^{\min(k,R)} w_i} \tag{3.20}$$

Where

$$\sum_{i=1}^{\min(k,R)} w_i = \frac{1}{log_2(i+1)}$$

$\sum_{i=1}^{k} r_i w_i$ is the Discounted Cumulative Gain, $DCG@k$ and $\sum_{i=1}^{\min(k,R)} w_i$ is the inverse Discounted Cumulative Gain (IDCG).

3.8 Summary

In this chapter, the methodology for conducting the experiment was laid out. It started off by collecting tweets for topic by using a crawler that crawled Twitter and collected results for #GST and as well as for #MarsWaters. From the past literature, two new features were identified to improve the reliability of ranking and relevance based on credibility, which are the use of Twitter Client and as well as Aging of Tweets. The algorithms are based on the combination of using a hybrid approach that is by combining the method of user based characteristics and network. Influence(X), which is based on user-based characteristics is discussed in whereby it is measured based on the method of how the features found on the user's social media. Secondly, Trust(X) score is introduced in calculating the trustworthiness of the user by looking at his network and features that are found in the people that he or she follows. The scores are then combined and normalized to form Credibility score, which also will be computed as pTRank which is the proposed algorithm. Finally, to validate the results later, the tweets were which collected were judged by people using crowd-sourcing and if they are relevant to the topic, the results would be known as relevance judgment and to benchmark them with other systems, several Information Retrieval techniques were discussed and used.

CHAPTER 4: RESULTS AND DISCUSSION

This chapter explains the experiments that were conducted as part of this research and the results of the experiment that were conducted from this research are discussed. This chapter discusses of the various baseline that is used and as well as the types of evaluation that were performed and how the system fares.

4.1 Building Baseline System and TwitterRank

In order to carry out the prove and to evaluate the effectiveness of the new scoring technique, pTRank , the baseline needs to be established which both Twitter Default Search (TDS) and TwitterRank are used .In the following section , the brief overview of implementation and setup of TwitterRank and Twitter default search are explained.

4.1.1 Building Twitter default search

The purpose of using Twitter default search (TDS) is to act as a baseline. This is to provide a relative comparison on how the original Twitter ranks. Twitter does not publish its algorithm nor shares the method of ranking. To address this, a simulated Twitter search is built, which is Twitter default search (TDS). TDS works by mimicking the default behavior of how Twitter retrieves results (Srijith, 2010) by ranking tweets with the highest number of retweets in reverse chronological order.

4.1.2 TwitterRank

Default TwitterRank algorithm was compiled and used from source code. There was a slight modification that was made to read from the crawled tweets, instead of crawling from Twitter and building the tweet database which is the default behavior of TwitterRank itself. By default, TwitterRank filters out non-English tweets during ranking based on the language identifier found in the tweets.

4.2 Internal Evaluation

pTRank would be used to evaluate on the performance against other baselines. For it to be compared with other baselines, several runs with variation needed to calculate Credibility(X) scores needs to done to determine the value ideal value of a α that provides the high credibility score of Credibility(X). Hence, the tests that would be performed is known as internal evaluation to evaluate the best performing pTRank scoring technique.

As discussed earlier in Section 3.2, two different weightage scores are introduced. In other to test them out, two sets of pTRank are introduced, which is pTRank1 that favors for weightages, ω of features that are equally important and pTRank2 for features that have a higher weightage value, ω.

Various values of α values were tested which are ranging from 0.0 to 1.0. This is done to determine which combination of features yields the best score. Figure 4.1 shows the scores with various values of α. The y-axis in the graph in Figure 4.1 represents the scores which are ranging from 0 to 1 to and x-axis represents the various values of α.

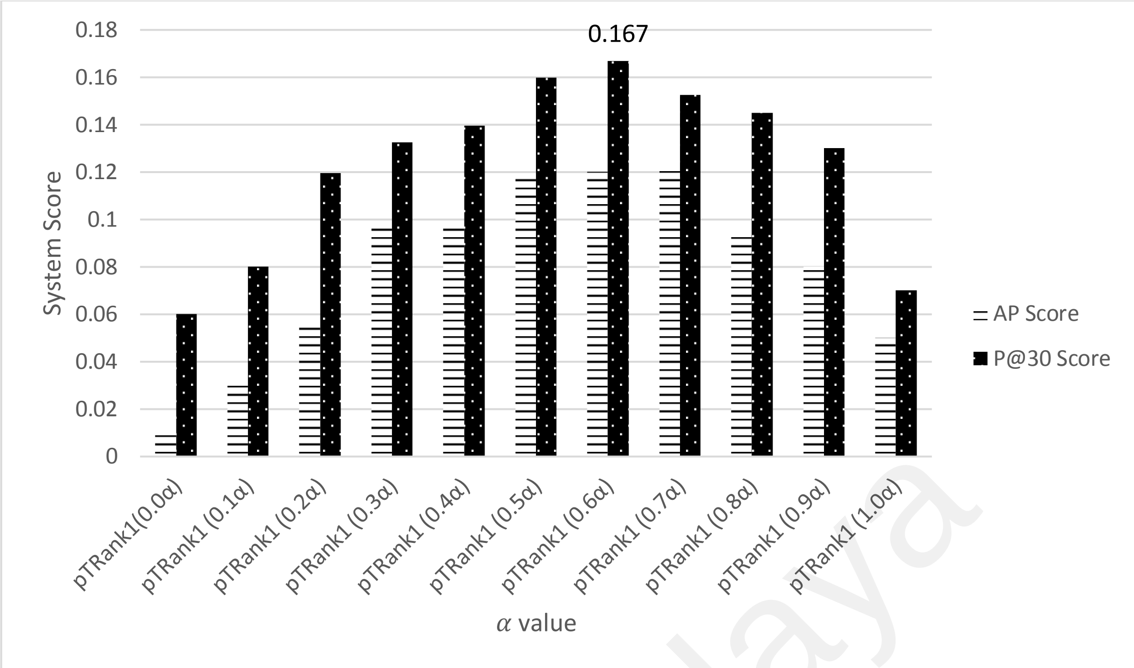

Figure 4.1: pTRank1 scores using different combinations of is α, alpha values with GST Dataset

From the experiment, the best performing score for pTRank1 is when the α value is set at 0.6. This shows that equal weightage of Trust and Influence element are important. The same experiment is repeated for pTRank2 as shown in Figure 4.2. The x-axis and y-axis are the same as Figure 4.1. However, this time the different weightage, ω with various values of α as well.

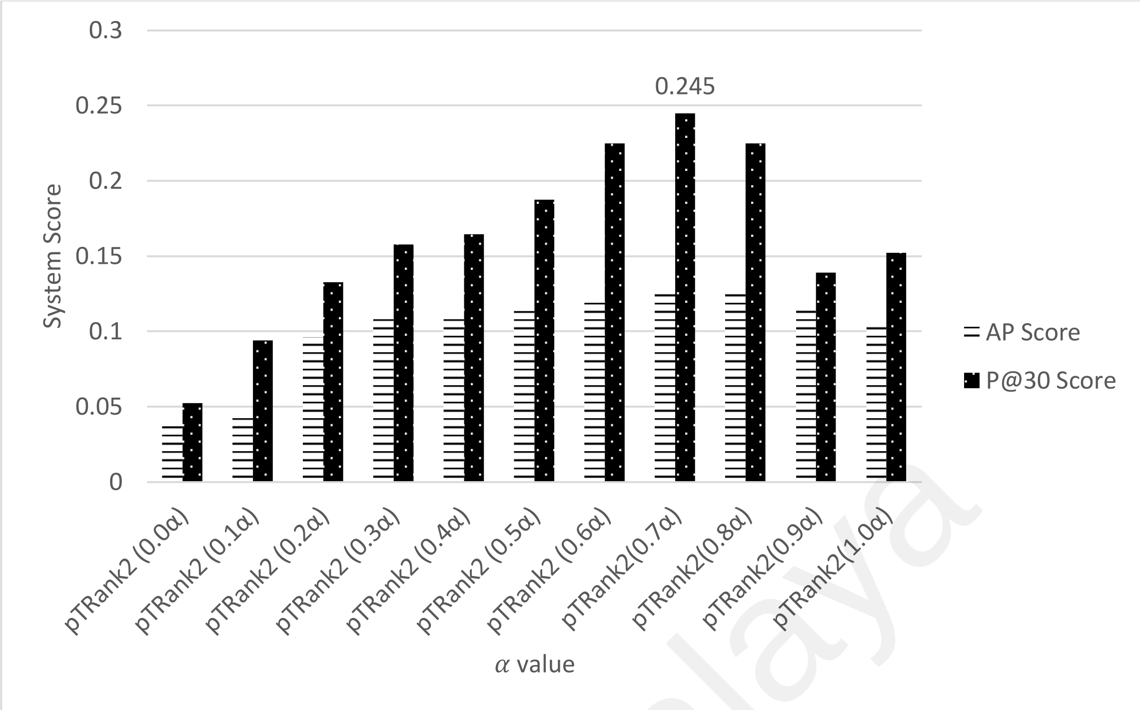

Figure 4.2: pTRank2 scores using different combinations of is α, alpha values with GST Dataset

From the graphs above, it can be deducted that the ideal value of α is between 0.6 for pTRank1 which yields 0.245 for P@30and for pTRank2 value of alpha is α for 0.7, which yields 0.245 for P@30. This is because from the results, when the score is 0.6 and 0.7 respectively, it shows that Trust(X) is an important element and it is more important compared to Influence(X). This can be said the same for AP scores as well. The summary of the best performing combination of pTRank1 and pTRank2 are represented with the table below. However, when the alpha scores are very low or very high, the score of Credibility(X) tends to perform poorly. This is because when biasness is induced (i.e extreme Trust or Influence) scores have started to drop and become unfavorable.

Table 4.1: AP and P@30 Scores of pTRank1 and pTRank2

Technique	AP Score	P@30 Score
pTRank1	0.122	0.153
pTRank2	0.126	0.245

Clearly from the graph and from the Table 4.1 pTRank2 with the alpha score of 0.7 is a much better performer compared to pTRank1 with the score of 0.6 score. In other to ensure that the same pattern is observed, another round of test was conducted with the best performing two ranking method with P@10, P@20 and as well as NDCG@20 with pTrank1 and pTRank2 with alpha score of 0.7 as these are the two best performing scoring technique.

In addition, to determine if the two new features that were introduced which is Twitter Client and Aging of Tweets did indeed improve the score. The same set of test was conducted but without factoring of the two features with the top performing pTRank1 and pTRank2 values based on Table 4.1. The new scores are described below in Table 4.2

Table 4.2: AP and P@30 Scores of pTRank1n and pTRank2n Without Two New Features

Technique	AP Score	P@30 Score
pTRank1n	0.102	0.113
pTRank2n	0.110	0.203

It can be clearly seen that the scores that the scores were obtained which was much lower, which is 0.102 for pTRank1 as opposed to 0.122 for AP score. In the case of pTRank2n scores, it did score 0.110 as opposed to pTrank2 which scored 0.127 for AP score. It can be concluded that removing the two new features from scoring method, the scoring was 15-19% lower as opposed to including them.

Figure 4.3 shows the score of the two scoring techniques inclusive of the two new features with other benchmarking scores. In Figure 4.3 below, the x-axis represents the score of the scoring technique whereas the y-axis represents the various benchmarking methodologies that were used.

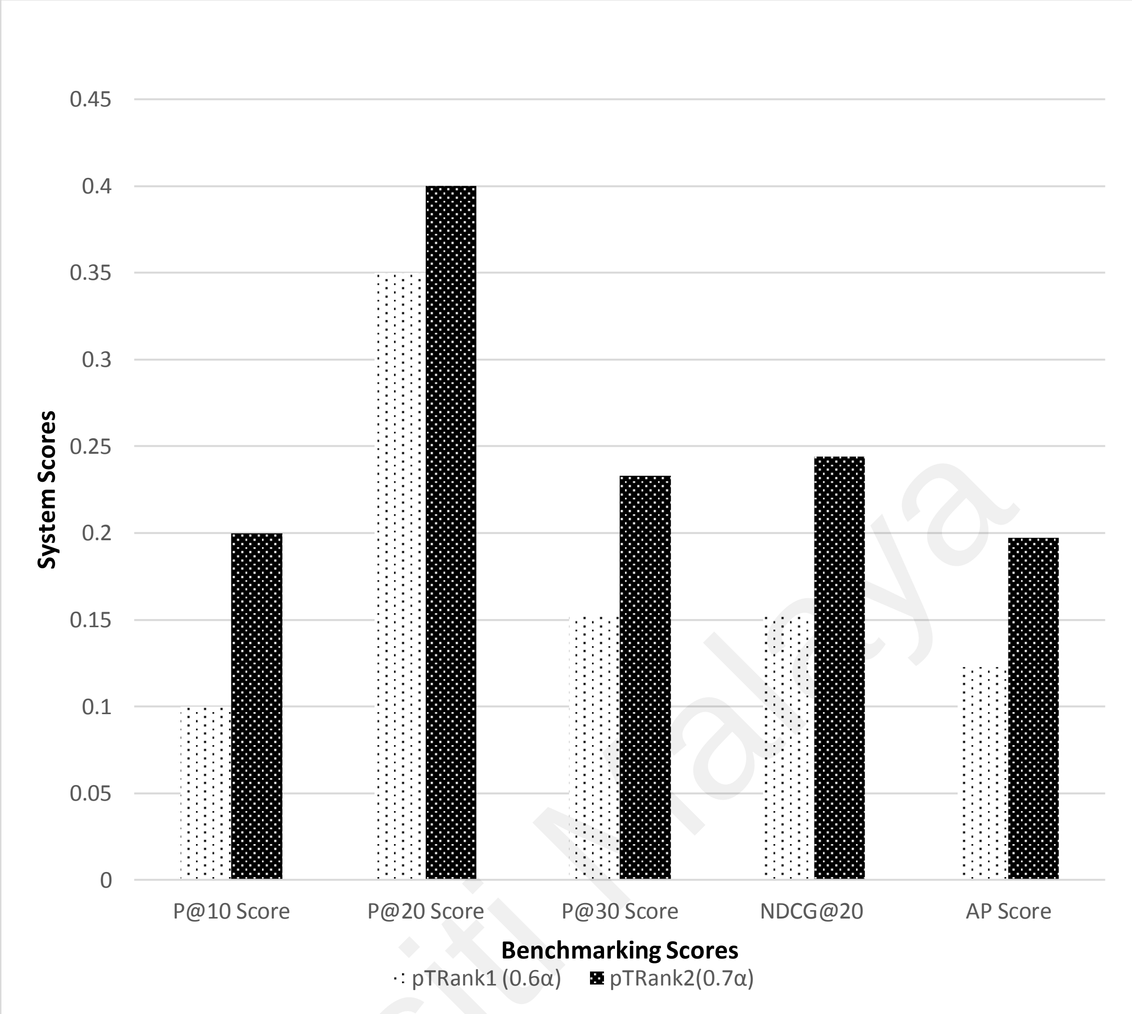

Figure 4.3: Comparison of Two Scoring Techniques, namely pTRank1 & pTRank2 for GST Dataset

From Figure 4.3, the same pattern can be observed as well that pTRank2 outperforms the pTrank1 in various other ranking mechanism which also includes NDCG as well. From the results, above, pTRank2 with alpha score of 0.7 and this would be used as the scoring technique that would be evaluated with baseline and TwitterRank.

4.3 External Evaluation

With internal testing is in place and the best performing scoring technique and the scoring are tuned. pTRank2 with alpha score of 0.7. External evaluation is then conducted, the purpose of the external evaluation is to benchmark ptTRank with other 2 baseline techniques which are mainly TDS and as well as TwitterRank , to see the best performing technique. The same GST dataset that was used in evaluating pTRank is used to benchmark with these scoring techniques.

Figure 4.4 shows the scores of each system. From the results in diagram in Figure 4.4, it can be clearly seen that pTrank2 is outperforming the TwitterRank and Default Twitter Search (Baseline).

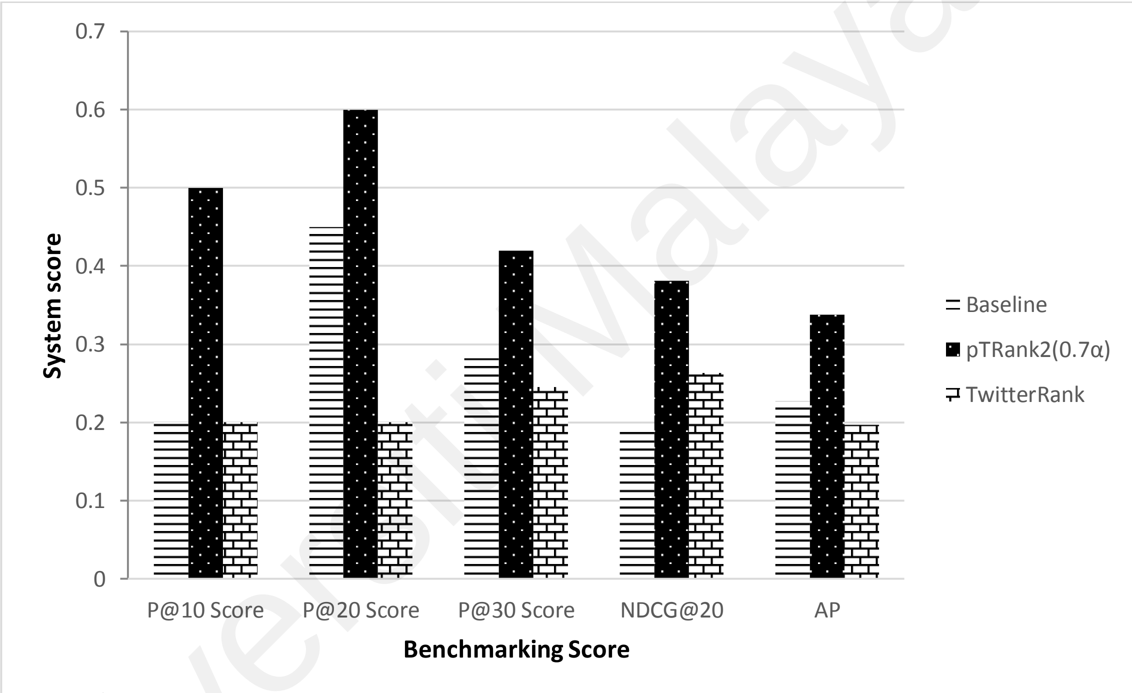

Figure 4.4: Comparison of pTRank with Baseline Scoring Technique using #GST Dataset

Our proposed scoring method, pTRank2 outperforms other techniques in all the benchmarking scores. The baseline which is TDS yields very bad in all the search due to the way on how Twitter favors popular tweets over content, which also means that popular tweets may not be credible. To ensure that the results are consistent across the technique and it is not influenced by data, a second test with the secondary dataset was used to determine if all the techniques do respect the pattern. In our secondary dataset, MarsWater data was used results are shown in Figure 4.5

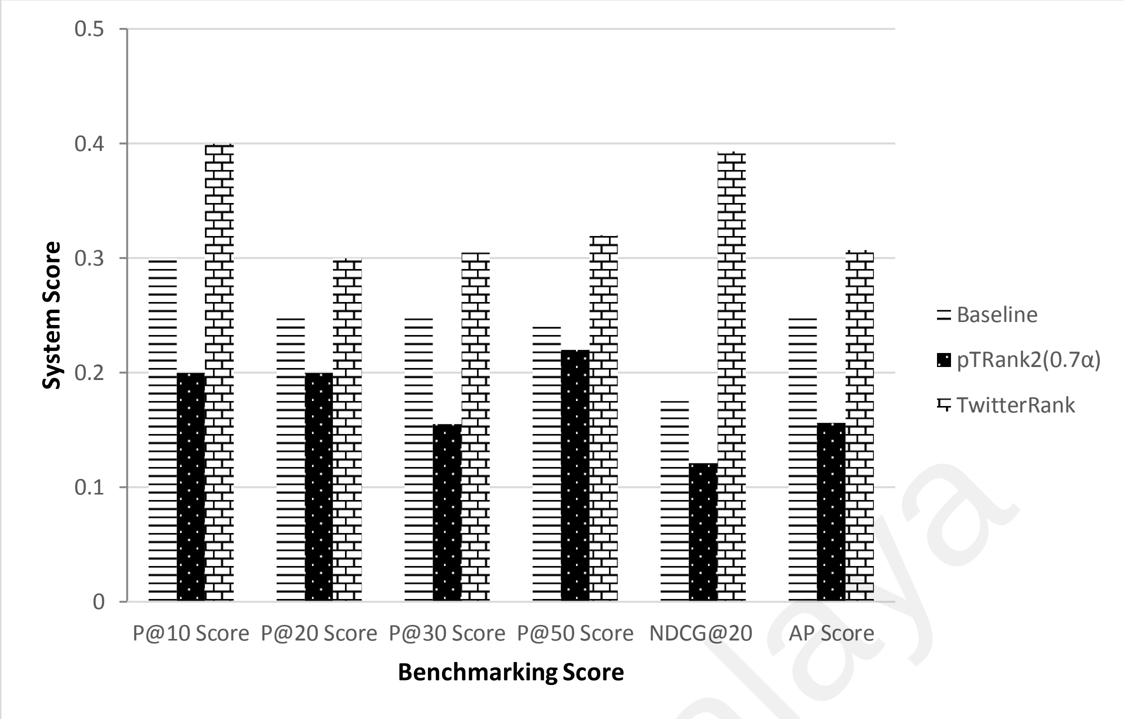

Figure 4.5: Comparison of pTRank with Baseline with Filtered GST Dataset containing English tweets

In the secondary dataset, a variation was noticed with pTrank2 techniques compared to the first dataset. There are several reasons to this, firstly it is because relevance judgment files were judged by people from South East Asia and there is a huge majority of tweets in the collection which are mostly in European languages such as French, Italian and German. These were not considered to be relevant by the crowd sourcing workers which did not understand these languages and hence marked them as irrelevant even though it could be relevant, hence causing the score to be lower for pTrank2. Secondly, the default behavior of TwitterRank filters out non-English tweets and pTRank considers non-English tweets as well, hence creating a biasness in the score.

From our observation, the results shown in Figure 4.5 did not provide a conclusive result. Hence, a second round of test that was conducted. In second round of test, filtering was done to the dataset to remove non-English tweets with the language identifier as by default Twitter contains a language identifier that indicates the language. Tweets which are tagged as en_US, en_UK, en_AU are considered for this.

The test is conducted again with both the datasets, both GST and MarsWaters to eliminate the factor of the judgment that was done by crowd-sourcing and to ensure that all techniques have the same level of playing field. It can be clearly seen that pTRank2 manages to outperform TwitterRank and TDS for GST dataset in Figure 4.6.

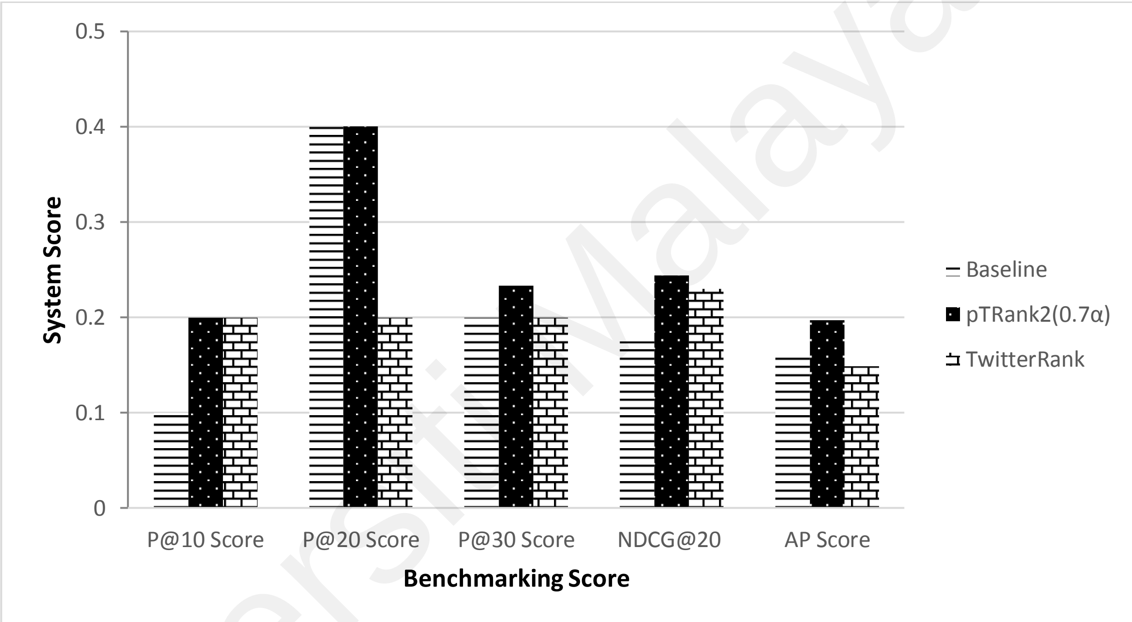

Figure 4.6: Comparison of pTRank with Baseline with Filtered GST Dataset containing English tweets

The same pattern can be observed with MarsWater dataset as well, the results are in Figure 4.7. It can be clearly seen that pTRank2 is on par with TwitterRank in terms of the scoring, and the same pattern can be observed. Although there in certain tests such as NDCG@20, there is a slight variation in the techniques whereby TwitterRank ranks slightly higher which is at 0.393 and our proposed system is at 0.378. The observation that was noticed was that TwitterRank did pick an additional 3 tweets which influenced the scoring. Apart from that our proposed technique works very well across all the benchmarking scoring technique

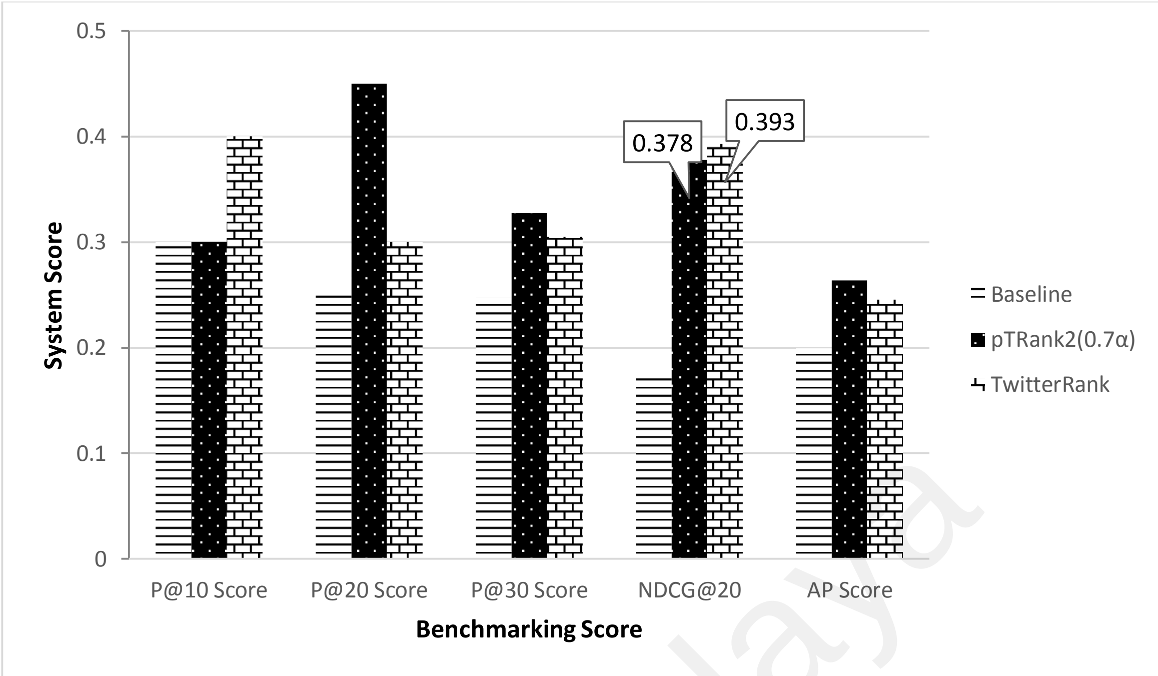

Figure 4.7: Comparison of pTRank with Baseline with Filtered MarsWaters containing English tweets

4.4 Summary

In this chapter, the results of our proposed technique, pTRank have been discussed and examined for two different test data collection. Firstly, an internal evaluation of the technique performed, whereby picking the right alpha value score which is used to balance the score between Trust and Influence and the weight of influence. These two techniques are known as pTRank1 and pTRank2. From the results, it was observed that pTRank2 with the alpha score of 0.7 yielded the best results. To prove that the new techniques did improve the scores, the best performing technique was put to the test again without considering Twitter Client and Aging of Tweets. It can be seen that by taking into consideration of the two new features the scores have improved by 15-19% respectively in AP scoring. With the best performing technique that was selected, it was then put to the test of other techniques which are Twitter Default Search (TDS) and TwitterRank. When it is put to the test, pTRank performs very well for GST dataset and outperformed TwitterRank, however it faired very poorly for MarsWaters dataset. Upon closer inspection, it was due to two reasons, one is because the relevance judgement for #MarsWaters dataset was judged by English speaking people and the dataset contained a lot of non-English tweets. This did not occur for GST which contained a mixture of language, because it was judged by people South East Asia countries who speak Malay and English. Secondly, TwitterRank filters out non-English tweets by default, thus giving it a higher favorable score.

Hence a second test was conducted, this time around it was done with both datasets by filtering out non-English tweets. However, upon standardizing it and removing non-English tweets, the technique did perform better when it comes to AP Scoring and P30 whereby scoring 0.263 for AP, and 0.328 for P30 respectively, compared to TwitterRank which only yielded 0.245 for AP and 0.305. Clearly these results confirm the effectiveness of our proposed algorithm.

CHAPTER 5: CONCLUSION AND FURTHER WORK

In this concluding chapter, the work of this thesis and its contribution are discussed and as well as the limitation and the challenges that were faced and as well as to discuss on the important directions of future work

5.1 Contribution

This research has contributed in the following areas:

1. Two features were identified, which are found in tweets that was used to determine credibility of the tweet, which were aging of tweets and Twitter Client. These two new features were not presented in past literature and this features helped to improve the score of the system by 15-19% respectively.
2. This research has contributed a new way calculating Credibility score which is by using the user's influence score by utilizing the features that were presented. The scores were combined by using the Trust and Influence scores together and then was merged using a smoothing technique.
3. An evaluation metric for benchmarking various techniques was created for the comparison of Default Twitter Search (TDS) and TwitterRank. The proposed technique, pTRank fared better by having the edge when it comes to Percision@30 scores, Mean Average Precision compared to TDS and TwitterRank in the tests that were conducted, both in Multilanguage Tweets and as well as English only tweets.

5.2 Limitation of Study

This study is limited to the dataset which was obtained from Twitter using the Twitter Rest API. The Rest API does not gather the whole dataset with the given keyword and there is a limit to the number of tweet that can be pulled.

Secondly, this study does not look at the content of the tweet, which may help to provide much better insight. This is because this study wishes to answer the research objective of how to rank tweets considering relevance and judgment by extending the feature set found in Twitter.

5.3 Problems Faced

The main problem that was encountered was on structuring the huge data and processing them. As JSON objects are large, it required a huge amount of processing power to store and parse them. This has taken a considerable amount of time to crawl and then store them in a format which could be easily extracted and manipulated. Secondly, the algorithm was implemented in a very crude way in Python without any optimization which meant a lot of processing tasks have taken a longer time to process, this results in longer time for the results to be obtained.

5.4 Benefits Of This Study in Real World

The work contributed by this research would greatly help authorities and even journalist alike. As stated back in Chapter 1, part of the motivation of this study managed to contribute in broader aspects of other fields, for example help journalist to determine credible source when writing a news article, political analyst in understanding political. It can even be used in government organizations to track people who spread slander and those who try to de-stabilize or bring treat to the national security of a said country. It can even be used in health environment in detecting spread of rumours for a disease or so. It truly has a wider range of application, which can be used.

5.5 Future Work

Here are some of the future work that can be done to enhance the existing algorithm to improve the ranking and the algorithm.

- Usage of Lexical Analysis & Tokenization - The results earlier showed that when using lexical analysis such as in the case of TwitterRank , as it was on par with some of the scenarios with pTRank. Lexical analysis or analyzing the content of the tweet would have enhanced and would provide a better-insight of the nature of the tweet. Findings from sentiment analysis can be applied in here to determine the nature and the tone of the message such as the works of Bollen, Mao and Pepe (2011) to determine the sentiment of the user.
- Real Time Ranking and Retrieval – The challenge with the current implementation is the it was done on the data that was already crawled and manually downloaded but not on real-time feed. A way to overcome this would be to work with Twitter to increase the API capacity which some studies have done in the past (Small et al, 2011). Another alternative is to look at the information that can be extracted from web based on search, although the information would be limited but it would be able to provide some real-time information.
- Considering Other Social Media Network Presence & External Ranking – Since people are well connected these days with the emergence of Web 2.0, a person has multiple social media profile (Kemp, 2015). For example, the same person who is on Twitter is also on Facebook and Instagram. The rich features in other social media can also be used as a measurement to measure the credibility of the person as well.
- Scoring System Improvement – Currently, the scoring system can be improved. The binary value of 0 and 1 only tells if the feature is presented but

it does not provide the granularity to the scoring. By adding this granularity, it would provide a much more accurate ranking, however the challenge is striking the right balance between performance and accuracy.

5.6 Concluding Remarks

To conclude this thesis managed to discover two new features which were aging of tweets and as well as Twitter Client. that were proved useful in research. Secondly, this research also helped to improve credibility score by considering both relevance and influence score. The importance of having a credibility score helps, users to evaluate if the tweets are genuine or fake, which can help them in determine to ignore or respond to the tweet. If this information is not presented, a person may just spread the rumour which can bring devastating consequences to the society.

In additional to that, this research has also paved away for a lot of people such as law enforcements, journalists, health professionals to validate tweets that were posted online on relating on an issue or a pandemic. For example, government officials can use to dismiss rumors that are being spread in Twitter to its citizen or given organizations such as World Health Organizations (WHO) can use to dismiss about rumors of a diseases or a pandemic.

Lastly but not least, the work here has also paved way for others to explore in this area of exploiting twitter's rich features. In additional to that also have paved for cross-language Twitter based retrieval, as Twitter is after all used by various people from around the world.

REFERENCES

Abdullah, I. B. (2010). Incremental pagerank for twitter data using hadoop (Doctoral dissertation, Master's Thesis. University of Ediburgh. 2010. Accessed September 10. 2013. Reference Source)

Alexander, B. (2006). Web 2.0. A New Wave of Innovation for Teachning and learning, 32-44.

Awani. (2014). Fasa baharu pencarian pesawat MH370 bermula. Retrieved March 26, 2017, from http://www.astroawani.com/berita-mh370/fasa-baharu-pencarian-pesawat-mh370-bermula-45404

Bae, J. H., Han, N. G., & Song, M. (2014). Twitter Issue Tracking System by Topic Modeling Techniques. Journal of Intelligence and Information Systems,20(2), 109-122.

Bahrami, M., Singhal, M., & Zhuang, Z. (2015, February). A cloud-based web crawler architecture. In Intelligence in Next Generation Networks (ICIN), 2015 18th International Conference on (pp. 216-223). IEEE.

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Everyone's an influencer: quantifying influence on twitter. WSDM?11.

Banks, M. A. (2008). Blogging heores: Interviews with 30 of the World's Top Bloggers. Indianapolis, IN: Wiley Publishing, Inc.

Bashir, S. (2015). Ranking entities on the basis of users' opinions.Multimedia Tools and Applications, 1-23.

Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010, July). Detecting spammers on twitter. In Collaboration, electronic messaging, anti-abuse and spam conference (CEAS) (Vol. 6, p. 12).

Bollen, J., Mao, H., & Pepe, A. (2011). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. ICWSM, 11, 450-453.

Boulos, M. N., & Wheeler, S. (2007). The emerging Web 2.0 social software: An enabling suite of sociable technologies in health and health care education 1. Health Information & Libraries Journal, 24(1), 2-23.

Boyd, C. (2016, March 30). Fake Flight MH370 Videos Being Shared on Twitter and Facebook. Retrieved March 26, 2017, from https://blog.malwarebytes.org/fraud-scam/2014/03/fake-flight-mh370-videos-being-shared-on-twitter-and-facebook/

Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In System Sciences (HICSS), 2010 43rd Hawaii International Conference on (pp. 1-10). IEEE.

Briguglio, P. (2013). Crisis communications – The golden hour. Retrieved November 21, 2015, from http://www.commpro.biz/crisis-communications-central/crisis-communications-the-golden-hour/

Bruns, A., & Highfield, T. (2016). May the best tweeter win: the Twitter strategies of key campaign accounts in the 2012 US election. In Die US-Präsidentschaftswahl 2012 (pp. 425-442). Springer Fachmedien Wiesbaden.

Bruns, A., Highfield, T., & Burgess, J. (2013). The arab spring and social media audiences english and arabic twitter users and their networks. American Behavioral Scientist, 57(7), 871-898.

Buchwalow, I. B., and Böcker, W. (2010). Immunohistochemistry: basics and methods. Berlin: Springer Verlag.

Burrell, Q. L. (1985). The 80/20 rule: Library lore or statistical law?. Journal of Documentation, 41(1), 24-39.

Buttry, S. (2015). Tips on verifying, debunking and carefully handling rumors. Retrieved June 18, 2016, from https://stevebuttry.wordpress.com/2015/03/01/tips-on-verifying-debunking-and-carefully-handling-rumors/

Can, E. F., Oktay, H., & Manmatha, R. (2013). Predicting retweet count using visual cues. In Proceedings of the 22nd ACM international conference on Conference on information & knowledge management (pp. 1481-1484). ACM.

Chen, C., Zhang, J., Chen, X., Xiang, Y., & Zhou, W. (2015, June). 6 million spam tweets: A large ground truth for timely Twitter spam detection. In 2015 IEEE International Conference on Communications (ICC) (pp. 7065-7070). IEEE.

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2010). Who is tweeting on Twitter: human, bot, or cyborg?. In Proceedings of the 26th annual computer security applications conference (pp. 21-30). ACM.

Churchill, D. (2009). Educational applications of Web 2.0: Using blogs to support teaching and learning. British journal of educational technology, 40(1), 179-183.

Comscore. (2015). 2015 U.S. Digital Future in Focus. Retrieved June 26, 2016, from https://www.comscore.com/Insights/Presentations-and-Whitepapers/2015/2015-US-Digital-Future-in-Focus

Croft, B., & Lafferty, J. (2013). Language modeling for information retrieval (Vol. 13). Springer Science & Business Media.

Croft, W. B., Metzler, D., & Strohmann, T. (2010). Search engines. Pearson Education.

Crooks, A., Croitoru, A., Stefanidis, A., & Radzikowski, J. (2013). # Earthquake: Twitter as a distributed sensor system. Transactions in GIS,17(1), 124-147.

Davtyan, M., Eickhoff, C., & Hofmann, T. (2015). Exploiting Document Content for Efficient Aggregation of Crowdsourcing Votes. InProceedings of the 24th ACM International on Conference on Information and Knowledge Management (pp. 783-790). ACM.

DiFonzo N, Bourgeois M, Suls J, et al. (2013) Rumor clustering, consensus, and polarization: dynamic social impact and self-organization of hearsay. Journal of Experimental Social Psychology 49(3): 378–399.

Doerr, B., Fouz, M., & Friedrich, T. (2012). Why rumors spread so quickly in social networks. Communications of The ACM.

Earle, P. S., Bowden, D. C., & Guy, M. (2012). Twitter earthquake detection: earthquake monitoring in a social world. Annals of Geophysics, 54(6).

Edwards, J. (2016). Leaked Twitter API data show the number of tweets is in serious decline. Retrieved May 1, 2016, from http://www.businessinsider.my/tweets-on-twitter-is-in-serious-decline-2016-2/#YdHsRSH1rVzDbSyd.99

Emplo Inc. (2014). DNA of the social networks | emplo.com. Retrieved June 26, 2016, from http://emplo.com/en/blog/dna-of-the-social-networks/

Fake Flight MH370 Videos Being Shared on Twitter and Facebook | Malware Bytes. (2014, March 14). Retrieved from https://blog.malwarebytes.org/fraud-scam/2014/03/fake-flight-mh370-videos-being-shared-on-twitter-and-facebook/

Farhi, Paul (2009) The Twitter explosion, American Journalism Review, 31(3), 26-31.

Gilbert, E., & Karahalios, K. (2009). Predicting tie strength with social media. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 211-220). ACM.

Göbel, S., & Munzert, S. (2016). Political Advertising on the Wikipedia Market Place of Information. Available at SSRN.

Greenberg, A. (2013). A Better Way To Filter Twitter's Spambots? Ask Google. Retrieved March 26, 2017, from http://www.forbes.com/sites/firewall/2010/07/09/a-better-way-to-filter-twitters-spambots-ask-google/#135895837ec1.

Gupta, A., & Kaushal, R. (2015). Improving Spam Detection in Online Social Networks. Paper presented at Cognitive Computing and Information Processing (CCIP), 2015 International Conference, Noida, India.

Gupta, A., & Kumaraguru, P. (2012). Credibility ranking of tweets during high impact events

Gupta, M., Zhao, P., & Han, J. (2012, January). Evaluating Event Credibility on Twitter. In SDM (pp. 153-164).

Hafizah Abdul Mansor, N., & Illias, A. (2013). Goods and Services Tax (GST): A New Tax Reform in Malaysia. International Journal of Economics Business and Management Studies, 2(1).

Han, B., Cook, P., & Baldwin, T. (2014). Text-Based Twitter User Geolocation Prediction. Journal of Artificial Intelligence Research, 49, 451-500

Heath, N. M. (2015). The 16 social media apps everyone should have. Retrieved June 25, 2016, from http://www.businessinsider.com/best-social-media-apps-2015-11?IR=T

Hermida, Alfred (2010) Twittering the news: The emergence of ambient journalism. Journalism Practice, 4(3), 297-308

Hollywoodlife. (2014). Justin Bieber Dead? New Death Hoax Hits Twitter. Retrieved March 26, 2016, from http://hollywoodlife.com/2014/01/09/justin-bieber-dead-died-death-hoax-twitter/

Huang, C. (2011, June). Facebook and Twitter key to Arab Spring uprisings: report. In The National (Vol. 6).

Hughes, A. L., & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. International Journal of Emergency Management, 6(3-4), 248-260.

Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. ACM Transactions on Information Systems (TOIS), 20(4), 422-446

Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., & Stumme, G. (2007). Tag recommendations in folksonomies. In Knowledge Discovery in Databases: PKDD 2007 (pp. 506-514). Springer Berlin Heidelberg

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities. Journal of The American Society for Information Science and Technology.

Juffinger, A., Granitzer, M., & Lex, E. (2009). Blog credibility ranking by exploiting verified content. In Proceedings of the 3rd workshop on Information credibility on the web (pp. 51-58). ACM.

Jurgenson, N. (2012). When atoms meet bits: Social media, the mobile web and augmented revolution. Future Internet, 4(1), 83-91.

Kamel Boulos, M. N., & Wheeler, S. (2007). The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education1. Health Information & Libraries Journal, 24(1), 2-23.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. Business horizons, 53(1), 59-68.

Kemp, S. (2012). Social, Digital and Mobile in Malaysia - We Are Social UK. Retrieved June 18, 2016, from http://wearesocial.com/uk/blog/2012/01/social-digital-mobile-malaysia

Kıcıman, E. (2010). Language differences and metadata features on Twitter. In Web N-gram Workshop (p. 47)..

Kumar, S., Morstatter, F., & Liu, H. (2014). Crawling Twitter Data. In Twitter Data Analytics (pp. 5-22). Springer New York.

Kwak, H., Lee, C., Park, H., & Moon, S. B. (2010). What is Twitter, a social network or a news media?

Lenhart, A., & Madden, M. (2007). Teens, privacy and online social networks: How teens manage their online identities and personal information in the age of MySpace.

Li, J., & Rao, H. R. (2010). Twitter as a Rapid Response News Service: An Exploration in the Context of the 2008 China Earthquake. The Electronic Journal of Information Systems in Developing Countries, 42.

Liu, C., Li, M., & Wang, Y. M. (2009). Post-rank reordering: resolving preference misalignments between search engines and end users. In Proceedings of the 18th ACM conference on Information and knowledge management (pp. 641-650). ACM.

Liu, F., Burton-Jones, A., & Xu, D. (2014). Rumors on Social Media in disasters: Extending Transmission to Retransmission. In PACIS (p. 49).

Liu, X., Nourbakhsh, A., Li, Q., Fang, R., & Shah, S. (2015). Real-time rumor debunking on twitter. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (pp. 1867-1870). ACM.

Longueville, B. D., Smith, R. S., & Luraschi, G. (2009). "OMG, from here, I can see the flames!": a use case of mining location based social networks to acquire spatio-temporal data on forest fires. doi:10.1145/1629890.1629907

Longueville, F. D., Hountondji, Y., Henry, S., & Ozer, P. (2010). What do we know about effects of desert dust on air quality and human health in West Africa compared to other regions? Science of The Total Environment. doi:10.1016/j.scitotenv.2010.09.025

Maes, P. (1994). Agents that reduce work and information overload.Communications of the ACM, 37(7), 30-40.

Malaysian Communication and Multimedia Commission (2013). Internet Users Survey 2012

Mangold, W. G., & Faulds, D. J. (2009). Social media: The new hybrid element of the promotion mix. Business horizons, 52(4), 357-365.

Markines, B., Cattuto, C., & Menczer, F. (2009). Social spam detection. In Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web (pp. 41-48). ACM.

Martinez-Romo, J., & Araujo, L. (2013). Detecting malicious tweets in trending topics using a statistical analysis of language. Expert Systems with Applications, 40(8), 2992-3000.

Marwick, A. E. (2015). Instafame: Luxury selfies in the attention economy. Public Culture, 27(1 75), 137-160.

Mayo, B. (2012). How Many People Use Twitter's Own Apps Retrieved March 26, 2016, from http://benjaminmayo.co.uk/how-many-people-use-twitter-s-own-apps.

Maximise Social Business. (2013). The Top 25 Twitter Clients Used by the Top 50 Social Media Influencers [INFOGRAPHIC]. Retrieved May 13, 2016, from http://maximizesocialbusiness.com/top-25-twitter-clients-top-50-social-media-influencers-9349/

McCreadie, R., Macdonald, C., & Ounis, I. (2011). Crowdsourcing blog track top news judgments at TREC. In Proceedings of the workshop on crowdsourcing for search and data mining (CSDM) at the fourth ACM international conference on web search and data mining (WSDM) (pp. 23-26).

McKinney MS, Houston B and Hawthorne J (2014) Social watching a 2012 republican presidential primary debate. American Behavioral Scientist 58(4): 556–573.

Mendoza, M., Poblete, B., & Castillo, C. (2010). Twitter under crisis: can we trust what we RT? 1st Workshop on Social Media Analytics (SOMA 10).

Middleton, R. (2015). Malaysia: Government to meet social media giants to stem 'rise of false information' Retrieved December 15, 2015, from http://www.ibtimes.co.uk/malaysia-government-meet-social-media-platform-giants-stem-rise-false-information-1515736

Miller, E. (2015). Inferring Tweet Quality From Retweets. Retrieved October 14, 2015, from http://www.evanmiller.org/inferring-tweet-quality-from-retweets.html

MSNBC. (2007). ‘We can’t stop it’. Retrieved June 18, 2016, from http://www.nbcnews.com/id/21431682/ns/us_news-2007_california_wildfires/t/we-cant-stop-it/

Muessig, K. E., Bien, C. H., Wei, C., Lo, E. J., Yang, M., Tucker, J. D., Hightow-Weidman, L. B. (2015). A Mixed-Methods Study on the Acceptability of Using eHealth for HIV Prevention and Sexual Health Care Among Men Who Have Sex With Men in China. J Med Internet Res Journal of Medical Internet Research, 17(4).

Muessig, K. E., Nekkanti, M., Bauermeister, J., Bull, S., & Hightow-Weidman, L. B. (2015). A systematic review of recent smartphone, Internet and Web 2.0 interventions to address the HIV continuum of care. Current HIV/AIDS Reports, 12(1), 173-190.

Myers, S. A., Sharma, A., Gupta, P., & Lin, J. (2014). Information network or social network?: the structure of the twitter follow graph. InProceedings of the 23rd International Conference on World Wide Web (pp. 493-498). ACM..

Naaman, M., Boase, J., & Lai, C. (2010). Is it really about me?: message content in social awareness streams.

NBC. (2014). Social Media Spread False Reports that Malaysia Flight Landed Safely. Retrieved February 15, 2015, from http://www.nbcnewyork.com/news/national-international/NATL-Social-Media-Spread-False-Reports-of-Safe-Landing-249158601.html

O’Connor, P. (2008). User-generated content and travel: A case study on Tripadvisor. com. Information and communication technologies in tourism 2008, 47-58.

O'Dell, J. (2011). How Egyptians Used Twitter During the January Crisis [INFOGRAPHIC]. Retrieved May 15, 2016, from http://mashable.com/2011/01/31/egypt-twitter-infographic/#d_4k3xhZOsqU

Oh, O., Agrawal, M., & Rao, H. R. (2011). Information control and terrorism: Tracking the Mumbai terrorist attack through twitter. Information Systems Frontiers. doi:10.1007/s10796-010-9275-8

O'Keefe, D. J. (1990). Persuasion: Theory and research. Newbury Park, CA: Sage Publications.

o'Reilly, T. (2009). What is web 2.0. " O'Reilly Media, Inc.".

Papacharissi, Z., & de Fatima Oliveira, M. (2012). Affective news and networked publics: The rhythms of news storytelling on# Egypt. Journal of Communication, 62(2), 266-282.

Pear Analytics. (2009). Twitter Study - August 2009. Retrieved from http://pearanalytics.com/wp-content/uploads/2012/12/Twitter-Study-August-2009.pdf

Pervin, N., Phan, T. Q., Datta, A., Takeda, H., & Toriumi, F. (2015). Hashtag popularity on twitter: Analyzing co-occurrence of multiple hashtags. In International Conference on Social Computing and Social Media (pp. 169-182). Springer International Publishing.

Rabbat, N. (2012). The Arab revolution takes back the public space. Critical Inquiry, 39(1), 198-208.

Rieh, S. Y., & Danielson, D. R. (2007). Credibility: A multidisciplinary framework. Annual review of information science and technology, 41(1), 307-364.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. In Proceedings of the 19th international conference on World wide web (pp. 851-860). ACM.

Santos, I., Minambres-Marcos, I., Laorden, C., Galán-García, P., Santamaría-Ibirika, A., & Bringas, P. G. (2014). Twitter content-based spam filtering. In International Joint Conference SOCO'13-CISIS'13-ICEUTE'13 (pp. 449-458). Springer International Publishing.

Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. Journal of the American Society for Information Science and Technology, 58(13), 2126-2144.

Savolainen, R. (2011). Judging the quality and credibility of information in Internet discussion forums. Journal of the American Society for Information Science and Technology, 62(7), 1243-1256.

Schmierbach, M., & Oeldorf-Hirsch, A. (2010). A little bird told me, so I didnt believe it: Twitter, credibility, and issue perceptions. Paper presented at the 93rd annual conference of the Association for Education in Journalism & Mass Communication.

Search Engine Watch. (2013). 6 Major Google Changes Reveal the Future of SEO. Retrieved May 15, 2016, from https://searchenginewatch.com/sew/opinion/2301719/6-major-google-changes-reveal-the-future-of-seo https://moz.com/google-algorithm-change

Sedhai, S., & Sun, A. (2016). Effect of Spam on Hashtag Recommendation for Tweets. In Proceedings of the 25th International Conference Companion on World Wide Web (pp. 97-98). International World Wide Web Conferences Steering Committee.

Sheedy, C. S. (2011). A Case Study of Social Media Use in the 2011 Egyptian Revolution. Use in the 2011 Egyptian Revolution. Masters of Arts in Public Communication Thesis.

Singhal, A. (2001). Modern information retrieval: A brief overview. IEEE Data Eng. Bull., 24(4), 35-43.

Sloan, L., Morgan, J., Burnap, P., & Williams, M. (2015). Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data.

Small, S. G., Lim, D., Appel, K., Kalic, D., Kemmer, M., Purcell, D., & Tran, C. (2011). 10 Weeks to TREC: STIRS Siena's Twitter Information Retrieval System.

Smeaton, A. (Ed.). (2012). Information retrieval and hypertext. Springer Science & Business Media.

Soboroff, I. (2007). A Comparison of Pooled and Sampled Relevance Judgments in the TREC 2006 Terabyte Track.

Song, J., Lee, S., & Kim, J. (2011). Spam filtering in twitter using sender-receiver relationship. In Recent Advances in Intrusion Detection (pp. 301-317). Springer Berlin Heidelberg.

Sreenivasan, N. D., Lee, C. S., & Goh, D. H. (2010). Tweet Me Home: Exploring Information Use on Twitter in Crisis Situations.

Starbird, K., Palen, L., Hughes, A. L., & Vieweg, S. (2010). Chatter on the red: what hazards threat reveals about the social life of microblogged information.

Stever, G. S., & Lawson, K. (2013). Twitter as a way for celebrities to communicate with fans: Implications for the study of parasocial interaction. North American journal of psychology, 15(2), 339.

Stieben, D. (2013). Is That Girl Really Following You On Twitter? Spambots Exposed. Retrieved May 13, 2015, from http://www.makeuseof.com/tag/is-that-girl-really-following-you-on-twitter-spambots-exposed/

Sudman, S., & Bradburn, N. M. (1982). Asking questions: a practical guide to questionnaire design.

Team Caffeine. (2014). 10 Remarkable Twitter Statistics for 2015. Retrieved June 26, 2016, from http://lorirtaylor.com/twitter-statistics-2015/

The Star (2013). Budget 2014: GST at 6% on April 1, 2015. Retrieved May 13, 2015, from http://www.thestar.com.my/news/nation/2013/10/25/budget-2014-gst/

The Week UK. (2016). MH370 conspiracy theories: What happened to the missing flight? Retrieved June 26, 2016, from http://www.theweek.co.uk/mh370/58037/mh370-conspiracy-theories-what-happened-to-the-missing-flight

Thomas, K., Grier, C., Ma, J., Paxson, V., & Song, D. (2011). Design and Evaluation of a Real-Time URL Spam Filtering Service.

Thomas, K., Grier, C., Song, D., & Paxson, V. (2011). Suspended accounts in retrospect: an analysis of twitter spam. In Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference (pp. 243-258). ACM.

Thomas, K., McCoy, D., Grier, C., Kolcz, A., & Paxson, V. (2013). Trafficking Fraudulent Accounts: The Role of the Underground Market in Twitter Spam and Abuse. In USENIX Security (Vol. 13, pp. 195-210).

TREC. (2000). Data - English Relevance Judgements File List. Retrieved May 13, 2016, from http://trec.nist.gov/data/qrels_eng/

Twitter (Ed.). (2013). This app. sending spam tweets! - Rules and Policies ... Retrieved February 20, 2016, from https://twittercommunity.com/t/this-app-sending-spam-tweets/13750

Twitter. (2012). Right-to-left languages on Twitter | Twitter Blogs. Retrieved June 18, 2016, from https://blog.twitter.com/2012/right-to-left-languages-on-twitter

Twitter. (2014). The Streaming APIs. Retrieved May 13, 2016, from https://dev.twitter.com/streaming/overview

Wang, A. H. (2010). Don't follow me: Spam detection in twitter. In Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference on (pp. 1-10). IEEE

Wang, B., Zubiaga, A., Liakata, M., & Procter, R. (2015). Making the most of tweet-inherent features for social spam detection on Twitter.

Wang, Y., Darko, J., & Fang, H. (2013). Tie-breaker: A New Perspective of Ranking and Evaluation for Microblog Retrieval. In TREC.

Web Ecology Project. (2011). Highlights and perspectives of soil biology and ecology research in China. Soil Biology & Biochemistry.

Wu, K., Yang, S., & Zhu, K. Q. (2015). False rumors detection on sina weibo by propagation structures. In 2015 IEEE 31st International Conference on Data Engineering (pp. 651-662). IEEE.

Xiong, X. (2013). Research on Key Issues of Spreading Behavior in Microblogging Network. PLA Information Engineering University.

Yang, F., Liu, Y., Yu, X., & Yang, M. (2012). Automatic detection of rumor on Sina Weibo. In Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics (p. 13). ACM.

Yang, J., Counts, S., Morris, M. R., & Hoff, A. (2013). Microblog credibility perceptions: Comparing the usa and china. In Proceedings of the 2013 conference on Computer supported cooperative work (pp. 575-586). ACM.

Yardi, S., Romero, D., & Schoenebeck, G. (2009). Detecting spam in a twitter network. First Monday, 15(1).

Zhai, C. X., Cohen, W. W., & Lafferty, J. (2003). Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. InProceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval (pp. 10-17). ACM.

Zhang, W., & Gelernter, J. (2014). Geocoding location expressions in Twitter messages: A preference learning method. JOSIS, 9.

Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E., Yan, H., & Li, X. (2011). Comparing Twitter and Traditional Media Using Topic Models. doi:10.1007/978-3-642-20161-5_34

Zubiaga, A., Spina, D., Martinez, R., & Fresno, V. (2015). Real ‐ time classification of Twitter trends. Journal of the Association for Information Science and Technology, 66(3), 462-473.