UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

Electronics and Computer Science

Dark Retweets: An Investigation of Non-Conventional Retweeting Patterns

by

Norhidayah Azman

Thesis for the degree of Doctor of Philosophy

June 2014

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
Electronics and Computer Science

Doctor of Philosophy

DARK RETWEETS: AN INVESTIGATION OF NON-CONVENTIONAL RETWEETING PATTERNS

by Norhidayah Azman

Retweets are an important mechanism for the propagation of information on the Twitter social media platform. However, many retweets do not use the official retweet mechanism, or even community established conventions, and these "dark retweets" are not accounted for in many existing analyses. In this thesis, a typology of 19 different tweet propagation types is presented, based on seven characteristics: whether it is proprietary, the mechanism used, whether it is created by followers or non-followers, whether it mentions other users, if it is explicitly propagating another tweet, if it links to an original tweet, and the audience that it is pushed to. Based on this typology and two retweetability confidence factors, the degrees of a retweet's "darkness" can be determined. This typology was evaluated over two datasets: a random sample of 27,146 tweets, and a URL drill-down dataset of 262,517 tweets. It was found that dark retweets amounted to 20.8% of the random sample, however the behaviour of dark retweets is not uniform. The existence of supervisible and superdark URLs skew the average proportion of dark retweets in a dataset. Dark retweet behaviour was explored further by examining the average reach of retweet actions and identifying content domains in which dark retweets seem more prevalent. It was found that 1) the average reach of a dark retweet action (3,614 users per retweet) was found to be just over double the average reach of a visible retweet action (1,675 users per retweet), and 2) dark retweets were more frequently used in spreading social media (41% of retweets) and spam (40.6%) URLs, whilst they were least prevalent in basic information domains such as music (8.5%), photos (5%) and videos (3.9%). It was also found that once the supervisible and superdark URLs were discarded from the analysis, the proportion of dark retweets decreased from 20.8% to 12%, whilst visible retweets increased from 79.2% to 88%. This research contributes a 19-type tweet propagation typology and the findings that dark retweets exist, but their behaviour varies depending on the retweeter and URL content domain.

Contents

Declaration of Authorship **xiii**

Acknowledgements **xv**

1 Introduction **1**

1.1 Hypothesis and the Definition of Dark Retweets 5

1.2 Research Publications 6

1.3 Document Structure 7

2 Literature Review **9**

2.1 Blogosphere 9

2.1.1 Power in the Blogosphere 10

2.2 Microblogging 12

2.3 Defining Power and Influence in Blogs and Microblogs 14

2.4 Influence Metrics via Blogs and Tweets 17

2.5 Retweets 21

2.5.1 Retweet Visibility 24

2.5.2 Temporal Spreads 25

2.6 Identifying Retweets Made Using Informal Retweeting Mechanisms 25

2.7 Ethics of Twitter Research 27

2.8 Conclusion 28

3 Pilot Study of Tweet Propagation Types **29**

3.1 Description of Pilot Typology 29

3.1.1 Native Retweets 30

3.1.2 Native Non-follower Retweets 31

3.1.3 RT/Via Retweets 31

3.1.4 RT/Via Non-follower Retweets 32

3.1.5 Replies 32

3.1.6 Non-follower Replies 33

3.1.7 Other Implicit Retweets 34

3.2 Explicit/Implicit and Follower/Non-follower Retweets 34

3.3 Decision Model Based on Typology of Propagation Types 35

3.4 Preliminary Toolkit 37

3.4.1 Architecture 37

3.5 Setup of Pilot Study 44

3.6 Results 46

3.6.1 Proportions of Retweet Types 46

3.6.2 Variability of Proportions Recorded . . . 51
3.6.3 Other Implicit Retweets . . . 52
3.6.4 Replies, Non-follower Retweets and Other Implicit Retweets as a Medium for Propagation . . . 52
3.7 Contribution of Pilot Study . . . 53
3.8 Modifications for Further Evaluation . . . 53

4 Typology of Tweet Propagation Types 55
4.1 Description of Tweet Characteristics . . . 56
4.2 Matrix of Tweet Propagation Types . . . 57
4.2.1 Invalid Groups of Propagating Tweets . . . 58
4.3 Typology of Tweet Propagation Types . . . 60
4.3.1 Original Tweets, Mentions, Replies and PDMF . . . 60
4.3.2 Explicitness and Links to the Originating Tweet . . . 62
4.3.3 Multiple Mechanisms in Tweets . . . 62
4.3.4 Orphan Retweets and Replies . . . 63
4.4 Dark Retweets . . . 64
4.4.1 Retweetability Confidence Factors . . . 64
4.5 Terminology . . . 66

5 Experimental Toolkit 69
5.1 Toolkit Architecture . . . 70
5.2 Data Storage . . . 70
5.3 Scripts and Algorithms . . . 72
5.3.1 Classification Algorithms . . . 73
5.3.2 Catching Dark Retweets . . . 76
5.3.3 Classifying Tweet Typology Types . . . 78
5.3.4 Comparison of Toolkit Against Existing Twitter Analysis Tools . . . 78

6 Proportions of Tweets Based on Typology of Propagation 81
6.1 Sample #1: Random Sampling . . . 82
6.1.1 Data Cleaning . . . 83
6.1.2 Finding . . . 83
6.2 Sample #2: URL Drill-down . . . 86
6.2.1 Distribution of Dark Retweets . . . 87
6.2.2 Supervisible and Superdark URLs . . . 90
6.3 Conclusion . . . 94

7 Exploring Behaviour of Dark Retweets 97
7.1 Reach . . . 97
7.1.1 Results . . . 99
7.1.2 Distribution of Retweeters' Followers . . . 100
7.1.3 Findings . . . 101
7.2 Content Domains . . . 102
7.2.1 Classification of Domains . . . 102
7.2.2 Comparison to Nagarajan et al. (2010) . . . 104
7.2.3 Findings . . . 105
7.3 Conclusion . . . 106

8 Conclusions and Future Work 109
8.1 Research Contributions . . . 109
8.2 Research Hypothesis . . . 111
8.3 Future Work . . . 112
8.3.1 Predicting Future Retweets and Virality . . . 112
8.3.2 Redefining Influence and Power on Twitter . . . 113
8.3.3 Adapting Propagation Typology to Facebook Interactions . . . 114
8.3.4 Detecting Spamming Behaviour on Twitter . . . 116
8.4 Conclusion . . . 117

Appendix A CSV Table Formats for Pilot Toolkit 119
A.1 Table Format A: Data collection . . . 119
A.2 Table Format B: Tweets and followers counts . . . 120
A.3 Table Format C: Followers records . . . 120
A.4 Table Format D: Retweet types . . . 120

Appendix B CSV Table Formats for Main Toolkit 123
B.1 Preparing Streamed Tweets for Classification . . . 123
B.2 Keeping Track of Friends Seen . . . 123
B.3 Recording Each User's List of Followers . . . 124
B.4 Classified Tweets . . . 124
B.5 Statistical Analysis of Classified Tweets . . . 125

Appendix C Description of Binary Values in the Matrix of Tweet Propagation Types 127

Appendix D Matrix of Invalid Tweets 129

References 135

List of Figures

2.1 Diagram of inlinks and outlinks . . . 10
2.2 Example Twitter timelines . . . 14
2.3 Retweet decision model (Zarella, 2009) . . . 24

3.1 Screenshot of user's view of native retweeting . . . 30
3.2 Screenshot of example native retweet . . . 30
3.3 Screenshot of example native retweet made by a third party application . 31
3.4 Screenshot of example RT/Via retweet . . . 32
3.5 Screenshot of user's view of 'Reply' link . . . 33
3.6 Alternative retweet decision model . . . 36
3.7 Preliminary toolkit's activity diagram . . . 38
3.8 CSV files schema . . . 41
3.9 Workflow of CSV table formats . . . 42
3.10 Line chart of retweet types for fundraiser URLs over time (days) . . . 47
3.11 Line chart of retweet types for news URLs over time (days) . . . 47
3.12 Line chart of retweet types for petition URLs over time (days) . . . 48
3.13 Line chart of retweet types for YouTube URLs over time (days) . . . 48
3.14 Pie chart of retweet types for four domains of URLs . . . 50

4.1 Typology of retweet types . . . 61

5.1 Architecture of components in toolkit . . . 70
5.2 Workflow of CSV table formats for main toolkit . . . 71
5.3 Conventional Rt markers detected by the toolkit . . . 74

6.1 Architecture of components in collecting random sampling dataset . . . 83
6.2 Total proportions of random sample . . . 84
6.3 Architecture of components in collecting URL drill-down dataset . . . 86
6.4 Distribution of dark retweet proportions over total retweet volume . . . 87
6.5 Average visible/dark Rt proportions over binned total Rt volume . . . 88
6.6 Cumulative averages for visible and dark retweet proportions . . . 89
6.7 Example repeating tweet texts . . . 93
6.8 Example bots with repeating tweet texts and authors . . . 93
6.9 Example auto-sharing of playlists . . . 93

7.1 Histogram of total followers per retweeter . . . 101

8.1 Facebook's sharing mechanism with *via* acknowledgement . . . 115
8.2 Facebook's sharing mechanism by multiple users . . . 115

List of Tables

1.1 Example repeated tweets . . . 3
1.2 Example replies . . . 3

2.1 First tweets using each retweet variation, as published by Kooti et al. (2012) 22

3.1 Matrix of explicit/implicit and follower/non-follower retweet types . . . 35
3.2 Table Format A - Generated CSV file after the data collection component 42
3.3 Table Format B - Tweets and followers counts . . . 43
3.4 Table Format C - Followers records . . . 43
3.5 Table Format D - Retweet types . . . 43
3.6 Percentage of explicit/implicit and follower/non-follower retweets across all four domains . . . 49
3.7 Proportions of retweet types by domain: Averages and standard deviations 52

4.1 Matrix of tweet propagation . . . 59
4.2 Summary of classifications of visible and dark retweets . . . 66

6.1 Total counts for original and propagating tweet types . . . 85
6.2 Proportion of visible and dark retweets . . . 85
6.3 Dataset comparison between random sampling and URL drill-downs . . . 87
6.4 Proportions of visible/dark Rts over total Rts/URL . . . 88
6.5 Cumulative proportions of visible/dark Rts over total Rts/URL . . . 89
6.6 Quartiles of total retweets and proportions of visible and dark retweets . . 90
6.7 Distribution of URLs in each Rt versus dark Rts % quartiles . . . 91
6.8 Total retweets containing supervisible and superdark URLs . . . 91
6.9 Top 10 supervisible URLs . . . 92
6.10 Manual classification for top 10 supervisible URLs . . . 92
6.11 Top 10 superdark URLs . . . 92
6.12 Manual classification for top 10 superdark URLs . . . 94

7.1 Calculating reach of retweet actions . . . 98
7.2 Average reach per retweet action . . . 99
7.3 Comparison of content domains between Nagarajan et al. (2010) and this thesis . . . 104
7.4 Proportions of visible and dark retweets based on content domains . . . 106

8.1 Mapping taxonomy of propagation types onto Facebook . . . 116

D.1 Matrix of tweet propagation . . . 129

Declaration of Authorship

I, Norhidayah Azman , declare that the thesis entitled *Dark Retweets: An Investigation of Non-Conventional Retweeting Patterns* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:
  - Azman, Norhidayah, Millard, David E. and Weal, Mark J. (2012) Dark Retweets: Investigating Non-conventional Retweeting Patterns. At *4th International Conference on Social Informatics, Lausanne, CH, 05 - 07 Dec 2012.*
  - Azman, Norhidayah, Millard, David E. and Weal, Mark J. (2011) Patterns of Implicit and Non-follower Retweet Propagation: Investigating the Role of Applications and Hashtags. At *Web Science 2011, Koblenz, Germany, 14 - 18 Jun 2011.*
  - Azman, Norhidayah, Millard, David E. and Weal, Mark J. (2010) Issues in Measuring Power and Influence in the Blogosphere. At *Web Science Conference 2010, Raleigh, NC, USA, 26 - 27 Apr 2010.*

Signed:..

Date: 4 July 2014..

Acknowledgements

First and foremost, thank you to my long-suffering supervisor, David Millard, for repeatedly picking me up off the floor throughout the last few years. Your patience is exemplary, and there is no way I could've got here without your unflinching support.

To my second supervisor, Mark Weal, thank you for your articulate insights and meticulous drive. My work has improved leaps and bounds thanks to you.

Thank you to my sponsors, the Malaysian Ministry of Education and Universiti Sains Islam Malaysia. Thanks to the hard-earned money of Malaysian taxpayers, I was given the opportunity to become more than I ever thought I could achieve.

To my mom, Azizah Murad, thank you for your love and support, both emotionally and financially. I hope I'm on my way to become a ground-shaking academic just like you. To my dad, Azman Nazir, enjoy being famous every time my work gets cited.

Thanks to Eric Cooke, my legendary coffee-plying mentor, confidante, guarantor and banker.

To Andrej Kazakov, thank you for being the spark that led me to doing this PhD in the first place.

Thank you to Huw Fryer for being there through thick and thin. No words can express my gratitude for all the help and support you've given me.

To my labmates, thank you for keeping watch over my PC and keeping it turned on when I was away. My work would've died in its tracks if it weren't for you guys.

Finally, to my all family, friends and loved ones, thank you for keeping me sane.

Chapter 1

Introduction

Social media websites such as Facebook and Twitter have been claimed to empower the common citizens in ways which were not ordinarily available using existing mainstream media. The Arab Spring and Obama's presidential elections have been championed as proof of people power, driven by the use of social media (Mansour, 2012; Walker, 2008).

One of the earliest forms of social media came through "web logs", or blogs, which came about in 1999 (Merholz, 1999; The Economist, 2006). The earliest incarnations of blogs consisted of posts which would contain URLs[1] pointing readers to websites deemed interesting by the blog authors (Blood, 2004). This format gradually became more journalistic in style with the rising popularity of Blogger[2], a free blog publishing platform which offered a simple user interface and made it easier for non-technical computer users to publish content online.

Since then, rapid developments have occurred in the world of social media over the following decades. Social networking sites such as MySpace, Friendster and eventually Facebook came into existence, allowing users to publish copious amounts of multimedia content online, with great ease and visibility. The blogosphere also evolved to include microblogging, where users still published content chronologically, but the size of the content itself is restricted to a smaller scale. For instance, Twitter[3] is a microblogging platform which allows you to *tweet*, i.e. publish 140-character text-based posts (called *tweets*). A social network graph is inherently present in Twitter, as you can *follow* another user, thus subscribing to that user's tweets. Once you have logged into Twitter, you will be greeted by a *timeline* which displays all the tweets that have been made by everyone that you are following, in a reverse chronological order.

[1] A URL is a universal resource locator: a formatted string of text that is used by web browsers and other software to identify and access a networked resource on the Internet, such as another website. URLs are more commonly known as web addresses, hyperlinks or links, although URLs are not necessarily exclusive to resources available only the web. Example URL: *http://www.google.com/*

[2] http://www.blogger.com/

[3] http://www.twitter.com/

Since its inception in 2006, Twitter has also evolved rapidly. There are several usage conventions that were introduced within Twitter that have evolved through social norms rather than dictated by the interface. For instance, the term *retweet* can be defined as both a noun and a verb; a noun to signify a "reposted or forwarded message on Twitter", or a verb where to retweet is to "repost or forward (a message posted by another user)" (Oxford University Press, 2013). This term was first used within the user community in 2007 (Kooti et al., 2012) before retweeting became officially implemented as a proprietary mechanism within the Twitter interface in late 2009[4].

Twitter activity has been of particular interest to researchers in power and influence for several reasons. Firstly, the concept of retweets allows patterns of propagation to be observed and used as a measurable proxy for power and influence. Secondly, it is relatively easy to build programs to retrieve data from Twitter using its free and publicly available Application Programming Interface (API)[5]. The Twitter API[6] allows developers and researchers to access data stored in official Twitter servers — such as user profiles, timelines, followers lists and retweet counts — in a standardized way. This allows developers to create new programs built on top of Twitter content, and allows researchers to retrieve data that can be experimented on.

The ability to track how tweets spread using retweets have led to several tweet propagation studies, from investigations of conversational patterns (Boyd et al., 2010) to overall retweet ratios (Cha et al., 2010; Mustafaraj and Metaxas, 2010). Existing studies on tweet propagation have mainly been focusing on retweets made using two conventional mechanisms: tweet texts containing common retweet markers (such as 'RT' or 'via'), or retweets made using the Twitter API's proprietary mechanism. In these cases, retweets could be found in two ways: *a*) simple parsing of the tweet texts to detect those common retweet markers, such as "RT" or "via", or *b*) querying the Twitter API to retrieve the number of proprietary retweets that have been made for any given tweet.

The problem with these two approaches is that these two mechanisms are not the only ways information inside a tweet could be spread. Consider the following two examples of tweets being spread within a particular timeline:

Example 1: Repetitions

The same URL gets repeated by another user who was not the originator, at some point in time **after** the original was made:

[4] "Project Retweet: Phase One" — http://blog.twitter.com/2009/08/project-retweet-phase-one.html

[5] An API is a standard software protocol which facilitates communication between different programs or services.

[6] http://dev.twitter.com/

Table 1.1: Example repeated tweets

User_A	27/06/2012 07:45
Houston New Jobs $ Senior Consultant Network Communications at Discover Financial Services (Houston, TX) http://t.co/vfFFE9v8	

User_B	27/06/2012 07:47
Houston New Jobs $ Senior Consultant Network Communications at Discover Financial Services (Houston, TX) http://t.co/JkJ1Z83C	

In the above example, these tweets do not contain common retweet markers, and do not use Twitter's retweet API. To make retweet identification even more difficult, the URLs aren't identical, but they all point to the same website. In Twitter, as of 2010, all URLs in tweets are automatically shortened to 20 characters max. This is achieved by replacing the original URL with a shorter redirecting one. When these short URLs are typed into web browsers, they redirect the browsers to the original URLs. In the above example, both *http://t.co/vfFFE9v8* and *http://t.co/JkJ1Z83C* point to the same address, which is
http://sqlusa.jobamatic.com/a/jobs/find-jobs/l-Houston,+TX.

Example 2: Replies

The same URL gets tweeted to specific users instead of being broadcast in a retweet to all his/her followers:

Table 1.2: Example replies

User_C	01/02/2013 18:23
@XXL http://t.co/2pOBctmo check out the new MIXTAPE #TheTakeOff a listeners delight.	

User_C	01/02/2013 18:26
@PhunkeyBrewster http://t.co/2pOBctmo dl free mixtape #TheTakeOff #TEAMFNA #BAGITUP #MMP #TEAMNOHOMMO #1STCLASS	

User_C	01/02/2013 18:29
@JayZ_News http://t.co/2pOBctmo CHECK IT OUT FREE DL..	

The '@' symbol here is typically used in Twitter as a prefix before usernames. The combination of @[username] signifies two things, depending on its location. If @[username] is found at the beginning of a tweet text, then this means that the whole tweet is a reply that is being sent specifically to the [username] stated. In the first example of replies shown above, user C has sent a reply to user XXL.

These two examples are a subset of tweets which are propagating URLs, but it is hard to determine if these can be considered to be a reposting of a prior tweet. Boyd et al.

(2010) made a similar observation; they had found tweets which seemed to contain content similar to previously published tweets, but they did not contain any attribution to any prior tweet nor originating author. However, due to the difficulty of determining the provenance of these non-attributed tweets, they were then excluded from the paper's evaluations. It seems that the structure afforded by using common retweet markers and querying the Twitter API for proprietary retweet counts have made these mechanisms the favoured approach in the methodologies of existing retweet research. This could be due to the simplicity and non-ambiguity of getting these retweets, which might not be the case when capturing retweets made using informal mechanisms.

Nonetheless, as shown in the two prior examples, tweets can be spread using informal mechanisms. If we were to exclude them from analyses because of the difficulty to identify them as retweets, then existing insights into how tweets propagate may not form the complete picture. There may be hidden tweet propagation paths that are not currently being investigated, or a tweet's reach may be different than what was initially thought. For example, existing studies do not seem to account for the true reach of replies. The reach of a retweet would extend to all of one's followers, but the reach of a reply only extends to a subset of followers, due to Twitter's visibility rules. When users prefix their Twitter conversations to other users with a dot '.', this usually means that the user wants that tweet to be broadcast to all of their followers. For example, tweeting ".@Bob I think you're right" instead of just "@Bob I think you're right" means that all followers will be able to see that tweet, whereas in the second tweet, only followers that are shared with Bob can see the second tweet. However, the act of prefixing tweets with '.' might not be considered amongst existing research work because retweet markers such as 'RT' and 'via' seem to be more popular and more widely used.

Another potential problem with not accounting for retweets made using informal mechanisms concerns the application of tweet propagation in research areas such as the measuring of power and influence. If a tweet's reach is considered as a quantifiable marker of a tweet's influence, then the existence of dark retweets could mean that a tweet could be more influential than initial estimates. Similarly, if the visibility of replies were to be considered, then the visibility of a particular tweet could be less than that originally envisaged.

Some studies have already begun to explore retweets made using informal mechanisms. For example, (Wu et al., 2011) called them "reintroduction of content", but in their work, all propagated URLs, using both conventional and non-conventional retweeting mechanisms, were considered equally, with no differentiation between the different types of retweets. This seems to be the opposite extreme of completely ignoring retweets made using informal mechanisms.

In this thesis, it is proposed that the act of propagating tweets is more nuanced than just using common retweet markers or using proprietary retweet mechanisms. There are non-conventional ways to send tweets, but there is no consensus as to what would be the best approach in dealing with these dark retweets. In response to this, a typology of tweet propagation types is proposed, incorporating different types of retweets, both visible and dark, to encompass all tweets that are propagated using both conventional and non-conventional retweeting mechanisms. Based on this typology, a study was undertaken to classify two datasets of tweets into visible, dark and orphan[7] retweets, namely using a random sample of 27,146 tweets, and a URL drill-down dataset of 262,517 tweets. Two subsequent studies were then conducted to explore the behaviour of dark retweets across two themes: average reach of retweet actions, and the prevalence of dark retweets based on content domains of URLs.

This research work is situated within the bigger picture of power and influence in social media. Lots of existing research work approximates power and influence by propagation, which in turn is approximated using retweets. This thesis proposes that the study of retweets made using informal mechanisms is important in order to gain a more comprehensive picture of propagation, covering all aspects of tweet propagation and not just those made using conventional retweeting mechanisms. Firstly, ignoring these retweets may mean that existing work on power and influence could be underestimating the actual volume of propagation that is happening. Secondly, a more comprehensive perspective on retweet types — from the different mechanisms available to the different audiences they serve — may lead to a more nuanced picture of power and influence in social media, allowing us to explore whether different forms of power were being exerted by different users in different domains.

1.1 Hypothesis and the Definition of Dark Retweets

This thesis proposes the following hypotheses:

H1: A significant minority of retweets are dark retweets that do not use formal retweeting mechanisms and therefore are difficult to detect

A **minority of retweets** means that in a given dataset, less than 50% will exist as **dark retweets**, which are tweets propagated without using formal retweeting mechanisms. This minority is considered **significant** if it is large enough to impact retweet analyses, such as the average reach of retweet actions.

Formal retweeting mechanisms relate to the methods that are perceived to be the most popular way to make retweets. In this thesis, there are two mechanisms which are considered formal: *a*) Twitter's proprietary retweet mechanism, which can be triggered by clicking the 'Retweet' button on any of their official user interfaces (either on their

[7] Orphan retweets exist when the original tweet has been deleted from Twitter's record.

web pages or mobile apps). This mechanism can also be triggered by third-party apps which use Twitter API's official retweeting method. *b*) Users manually copying and pasting other people's tweets, and then prefixing these tweets with conventional retweet markers such as 'RT' and 'via'.

Tweets made without using the aforementioned mechanisms are considered **difficult to detect**, because several assumptions need to be made in order to determine whether a tweet is a retweet or not, therefore requiring extra detection procedures to deduce a retweet. Furthermore, these tweets would not be included in Twitter's official retweet counts. Manually copied and pasted tweets which include 'RT' or 'via' retweet markers are also excluded from Twitter's official retweet count, but given that this mechanism has been in prolonged use (Kooti et al., 2012), existing retweet studies have been counting these separately in parallel with Twitter's official retweet count.

This hypothesis proposes that a small proportion of dark retweets exist alongside visible ones, but detecting them requires extra effort.

H2: The behaviour of dark retweets is not uniform, and changes depending on the retweeter and the content domains of the URLs spread

The **behaviour of dark retweets** in this thesis relates to several themes, namely 1) the proportions of dark retweet types over a given dataset, 2) the average reach of dark retweet actions, and 3) the prevalence of dark retweets within a given content domain. Changes in these patterns depend on the **retweeter** (retweet author) or the **content domain** (subject matter) of the URL being spread. This hypothesis proposes that dark retweets exhibit non-uniform behaviour across the themes described above. It is important to understand the potential impact of dark retweets in order to inform future tweet propagation studies.

1.2 Research Publications

Parts of this thesis have been published in the following publications:

- Azman, Norhidayah, Millard, David E. and Weal, Mark J. (2012) Dark Retweets: Investigating Non-conventional Retweeting Patterns. At *4th International Conference on Social Informatics, Lausanne, CH, 05 - 07 Dec 2012.*

- Azman, Norhidayah, Millard, David E. and Weal, Mark J. (2011) Patterns of Implicit and Non-follower Retweet Propagation: Investigating the Role of Applications and Hashtags. At *Web Science 2011, Koblenz, Germany, 14 - 18 Jun 2011.*

- Azman, Norhidayah, Millard, David E. and Weal, Mark J. (2010) Issues in Measuring Power and Influence in the Blogosphere. At *Web Science Conference 2010, Raleigh, NC, USA, 26 - 27 Apr 2010.*

1.3 Document Structure

Chapter 2: Literature Review presents a literature review of existing studies on power and influence in social media, the problems faced when quantifying power online, and the role of retweets as an approximation of power. The chapter includes an overview of the history of blogs, its evolution to microblogging, the role of retweets in propagation and the different metrics used to measure power and influence. This is then followed by a review of existing work that have also looked at retweets made using informal mechanisms and the challenges of detecting and working with them.

This is followed **by Chapter 3: Pilot Study** which presents the setup and the results of a preliminary study investigating the prevalence of retweets made using informal mechanisms. The pilot study was run using an experimental toolkit that was developed for classifying tweet propagation types. Based on the findings of this preliminary research, the initial classification was extended into a more comprehensive typology of tweet propagation.

Chapter 4: Typology outlines this enhanced typology of tweet propagation, which is based on seven characteristics: whether it is proprietary, the mechanism used, whether it is created by followers or non-followers, whether it mentions other users, if it is explicitly a retweet, if it links to an original tweet, and the audience which receives the tweet.

An extended toolkit was then built in order to find the different types of tweet propagation, particularly dark retweets. Details of this toolkit is fully explained in **Chapter 5: Experimental Toolkit**, from the architecture used to the scripts created.

Chapter 6: Proportions of Tweets Based on Typology of Propagation then describes the methodology used to observe the proportions of all 19 different tweet types and the result of this experiment. The following **Chapter 7: Exploring Behaviour of Dark Retweets** presents the design and results of the subsequent studies that were run to explore the behaviour of dark retweets in terms of average reach of retweet actions and content domains of URLs.

This thesis then concludes with **Chapter 8: Conclusions and Future Work** with the contributions of this research and possible future work ideas.

Chapter 2

Literature Review

This chapter introduces fundamental concepts from Social Media, including definitions of blogs and microblogs such as Twitter. In this chapter, a number of examples are presented where microblogs have been perceived to wield influence and to have had an important impact on national or global events. This is followed by an outline of some definitions of power, and a discussion on quantifying online power based on various metrics for measuring influence. In particular, the use of retweet propagation as a way of approximating power is discussed.

Given that existing studies acknowledge the existence of tweets which are difficult to be classified as retweets, this chapter provides the academic background of retweets using informal mechanisms. This then leads to the concept of dark retweets, which will be described in Chapter 4: Typology (page 55).

2.1 Blogosphere

The term "blog" was first coined by Peter Merholz as a portmanteau of web logs in 1999 (Merholz, 1999; The Economist, 2006). Since then, blogs had evolved from the 90s' simple filter-style logs of interesting URLs to today's journal-style blog posts (Blood, 2004; Recuero, 2008). Blogs today are commonly defined as websites where posts are published in a usually reverse chronological order. Each *blog post* would be *timestamped* with the date it was published, and would normally allow readers to leave *comments* at the end of the blog post. The side panel of these blog sites would typically contain a list of all previously *archived posts*.

Citations of other blogs involve *permalinks* (URLs pointing to individual blog posts as opposed to the main blog site) and *blogrolls* (a collection of links to other blog sites, usually displayed on the side panel of blog sites if available). The *inlink* for any given blog is defined as any link or citation from external websites that points to that particular blog, whereas an *outlink* refers to a link from that particular blog pointing towards external websites. Figure 2.1 shows an illustration of how inlinks and outlinks are defined.

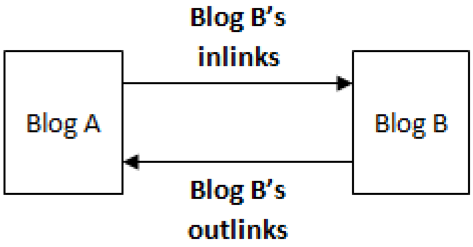


Figure 2.1: Diagram of inlinks and outlinks

2.1.1 Power in the Blogosphere

In the field of commercial marketing, the impact of blogs has been attributed to the "my story" phenomenon, where the stories on blogs gain more perceived credence due to the author being the primary and therefore most authoritative source (Market Sentinel, 2005). Bloggers who are seeking authority perceive blogging as a professional activity; they would seldom be publishing any information that has already been covered by another blogger who is within their same network (Recuero, 2008).

Several studies have been focused on the role of bloggers in exerting some form of power through their blogs. McKenna and Pole (2008) claimed that some of the main activities bloggers engage in includes checking media sources and political advocacy. Bloggers are more likely to be informative rather than to be involved in political activism, and the latter usually manifests in the form of petitions rather than rallies (McKenna and Pole, 2008). A different perspective was also proposed when investigating the impact of blogs: "Being a medium should not be confused with being a cause for change ... people deciding to take action in large numbers, organized by charismatic and capable leaders, will be the cause" (McKenna and Pole, 2008). McKenna and Pole (2008) also found that 45.1% of respondents agreed that the Internet helps people to have more political power. There have been several studies of the perceived ability of blogs to empower people. Sullenger (2006) quoted that the American "Founding Fathers would've loved the Internet" and blogs are empowering people who cannot afford printing presses. The author of Meccawy (2008) herself, as a Saudi woman, felt empowered by the "one-to-many" communication channel available through blogging. Blogs seemed to allow quiet engagement within the confines of censorship to free people's minds, becoming "safety valves" to vent away instead of on the streets, without the need for any proxies (MacKinnon, 2008).

Given the emergence of 'people power' arising from the use of blogs, some parallels have been drawn to the time when the printing press was invented. Blogs had been likened to pamphleting as early as 2001 (Bricklin, 2001). Packer (2004) observed that blogs seemed to "open up ... journalism to a vast marketplace, reminiscent of earlier ages of pamphleting." Analogies were made between blogs and the role of pamphlets as mentioned in the Nobel-winning book 'The Idealogical Origins of the American Revolution' (Klau, 2003, 2009).

The printing press had allowed people access to literature in the form of pamphlets, seemingly eliminating the technical barriers to becoming a popular author (Wiederhold, 1995). There had been cases of pamphlets facilitating successful political activism — Jonathan Swift's 'Drapier's Letters' instigated a nationwide boycott of a coin patent that was later withdrawn. However, there had also been cases where pamphlet-based activism led to imprisonment — Edward Waters got sent to jail after the jury was sent 9 times to get the "right" verdict (Bragg, 2009).

In the US, the freedom of expressions such as the above seems to be strengthened by the First Amendment of the United States Constitution. Sunstein (2007) claimed that it has become a large cultural presence mainly because its provisions for freedom of speech has been used to defend bloggers from several legal defamation cases; some of them ruled how blogs were similar to pamphleting in terms of promoting grassroots activism, hence needed to be protected (Sullenger, 2006). Freedom of the press is based of the notion that everyone should have access to a press, or in this case a blog (Sullenger, 2006).

However, the power of blogs as compared to mass media seems to be debatable. A study on Chinese blogs claimed that there is "no evidence that blogs [cause] political change" (MacKinnon, 2008). Woodly (2008) argued that blogs can only sell reputation; although blogs mobilize opinions and set political agendas, traditional media still reaches more people than most popular websites. Moreover, despite the empowering capabilities of blogs, Zuckerman (2008) argued that bloggers cover stories only if they were primed by mass media. The opinions in Mari (2010) also concurred, claiming that the 2010 UK general elections will be remembered as the year of the TV election — rather than social media — and online discussions would revolve around an established group of politicians, journalists and interest groups using such platforms.

Nonetheless, Farrell and Drezner (2008) found that mainstream media personnel liked to follow blogs for setting news agendas and creating focal points of interest. Castells (2009) observed how newspapers were also beginning to cite stories from the blogosphere, although in an empirical study of keyword propagations, Leskovec et al. (2009) found that only 3.5% of propagations originated from blogs into news media.

Despite the differing views, there have been certain notable cases where blogs have been seen as influential in affecting the outcomes of elections. Packer (2004) claimed that "the blog may be the first innovation from the Internet to make a real difference in election politics." This seemed particularly apparent during the US presidential elections of 2004, 2008 and 2012 (McKenna and Pole, 2008; Ives, 2008; Scherer, 2012; Hall, 2013). In particular, Walker (2008) analyzed the success factors of Obama's online campaign in 2008. This paper claimed that Obama had an "ability to transform the nature of political fundraising and recruitment through the internet", raising USD $55 million — almost USD $2 million daily — in February alone, which was claimed to consist mostly of small Internet donations. In total, Obama had managed to secure USD $500 million online (Scherer, 2012) — more than thirty times the funds that Howard Dean had collected over the Internet in his unsuccessful 2004 campaign to become the Democrat presidential candidate (Walker, 2008). Using some of the technology predated by Dean's campaign, Obama's 2008 campaign had organized more than 8,000 online affinity groups, 750,000 active volunteers, and attracted more than 1.6 million donors (Walker, 2008). Four years later, Obama's campaign upped their game by raising USD $690 million online (Scherer, 2012; Hall, 2013).

Another example of the perceived power of blogs is the resignation of US Senator Trent Lott. One of his speeches was deemed very racist by the blogosphere but this issue was given no mainstream media coverage at the time. However, due to the sustained pressure that seemed to build up amongst bloggers, the press later published the story, which preceded his eventual resignation as Senate Minority Leader (Farrell and Drezner, 2008). Another notable incident is "Rathergate", where Dan Rather from the TV show *60 Minutes* presented a news story based on documents that were then claimed by bloggers to be forged, prompting a retraction of the story later by the CBS network (Munger, 2008).

This perceived empowerment then began to be similarly observed within the sphere of microblogs. The following section describes the subsequent progression from longer-length blogs to microblogging in the form of tweets.

2.2 Microblogging

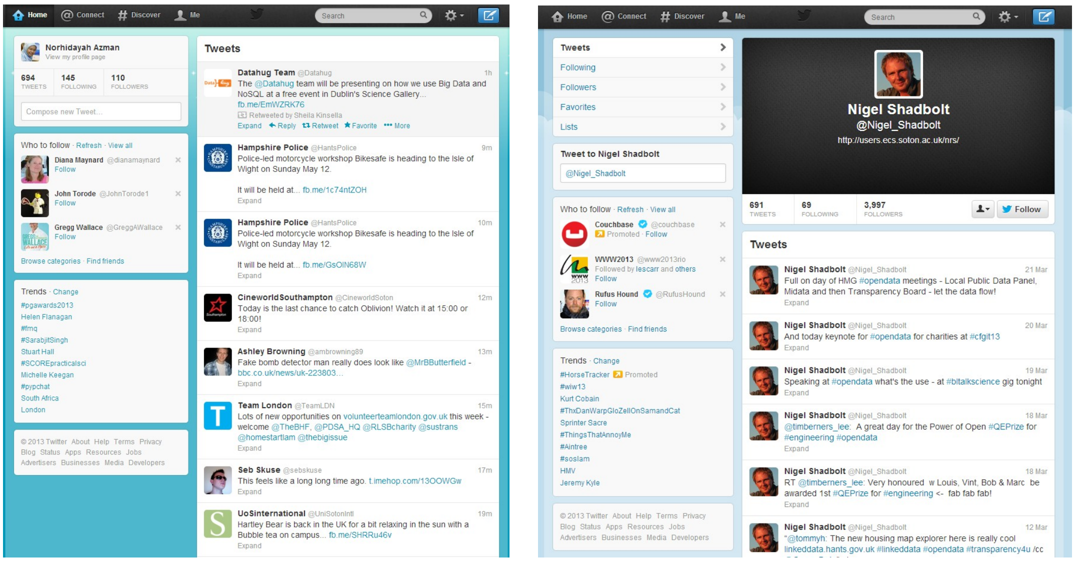
Microblogging is a subset of blogging, where the content and size of the individual posts are considerably smaller than traditional blog posts. In a study on post lengths found in what was deemed to be successful traditional blogs, the average post length ranged between 119 to 2140 words, depending on the blog and its niche (Allsopp, 2010). This is different to microblogs which would contain at most only hundreds of characters. Twitter, a microblogging service that was founded in 2006, only allows 140 characters per post, called a *tweet*. This text-based microblogging service is also similar to the service provided by China's Sina Weibo (weibo is Chinese for microblogging) which was officially launched in 2009. Sina Weibo also uses a 140-character limit for its posts. Several other online services, such as Pinterest and Tumblr, offer variants of microblogging, in the form of photo sharing services where photos with short captions can be logged in chronological order.

Several studies have been done on the uptake of microblogging. As of October 2009, around 19% of surveyed American Internet users claimed to be using Twitter or another service to share updates about themselves or to see updates about others (Lenhart et al., 2010). Fox et al. (2009) found that Internet users between the ages of 18–24 were more likely to use Twitter compared to any other age group. As of March 2013, there were 200 million active users on Twitter, and 400 million tweets were being sent each day (Twitter, 2013).

Similar to blogs, microblogs such as Twitter also display tweets in a reverse chronological order. Each tweet would also be timestamped with the time and/or date of publication. Several usage conventions have emerged through microblogging, particularly Twitter. For example the usage of *hashtags* (#) to tag keywords within posts, aliases or the *at-sign* (@) to *mention* other users, and reposting other people's posts using prefixes such as '//' (Sina Weibo), or 'RT' (Twitter). In Twitter, this reposting concept is called *retweeting*. In parallel with blogs' utility for comments, microblogs allow users to respond directly to other users using several methods: retweets, replies, mentions and direct messages. *Retweets* are tweets which have been reposted onwards to other users, typically using conventional retweeting mechanisms, such as using Twitter's proprietary 'Retweet' button, or to prefix other users' tweets with retweet markers such as 'RT' or 'via'. *Replies* are publicly broadcasted tweets which directly address the participants of a particular conversation. For example, if Users A and B would like to reply to each other, their tweets would be prefixed with the usernames of each other, prefixed with the alias or the *at-sign* (@) e.g. "@UserB hello!." *Mentions* are tweets which contain at least one username in the tweet text, e.g '@username', but it is different to replies due to the placement of the username: replies must always have the username at the start of the text, which is not the case for mentions. Finally, *direct messages* are tweet texts sent privately to another user's inbox, and cannot be publicly viewed by other users. Users can also mark their *favourite* tweets by clicking on the star button on the Twitter user interface.

Twitter allows users to *follow* each other, thus subscribing to each other's tweets. This follower/following relationship between users allow Twitter users to be connected to one another in a network that can be observed and analyzed, making it possible to draw conclusions about power and influence. The *reach* of a Twitter user relates to the number of followers that the user has, as this would be the total size of the immediate audience to the user's tweets.

Tweets are shown in *timelines* which are displayed in reverse chronological order. There are two different timelines available on Twitter: a user's *home timeline* displays all the tweets made by everyone the user is following, whilst a user *profile timeline* shows all the tweets made by a particular user. Figure 2.2 (page 14) shows examples of both Twitter timelines.



Norhidayah Azman

694 TWEETS 145 FOLLOWING 110 FOLLOWERS

Who to follow · Refresh · View all

Diana Maynard @dianamaynard Follow

John Torode @JohnTorode1 Follow

Gregg Wallace @GreggAWallace Follow

Browse categories · Find friends

Trends · Change

#pgawards2013
Helen Flanagan
#fmq
#SarabjitSingh
Stuart Hall
#SCOREpracticalsci
Michelle Keegan
#pypchat
South Africa
London

© 2013 Twitter About Help Terms Privacy Blog Status Apps Resources Jobs Advertisers Businesses Media Developers

Tweets

Datahug Team @Datahug 1h
The @Datahug team will be presenting on how we use Big Data and NoSQL at a free event in Dublin's Science Gallery... fb.me/EmWZRK76
Retweeted by Sheila Kinsella
Expand Reply Retweet Favorite More

Hampshire Police @HantsPolice 9m
Police-led motorcycle workshop Bikesafe is heading to the Isle of Wight on Sunday May 12.

It will be held at... fb.me/1c74ntZOH
Expand

Hampshire Police @HantsPolice 10m
Police-led motorcycle workshop Bikesafe is heading to the Isle of Wight on Sunday May 12.

It will be held at... fb.me/GsOlN68W
Expand

CineworldSouthampton @CineworldSoton 12m
Today is the last chance to catch Oblivion! Watch it at 15:00 or 18:00!
Expand

Ashley Browning @ambrowning89 13m
Fake bomb detector man really does look like @MrBButterfield - bbc.co.uk/news/uk-223803...
Expand

Team London @TeamLDN 15m
Lots of new opportunities on volunteerteamlondon.gov.uk this week - welcome @TheBHF, @PDSA_HQ @RLSBcharity @sustrans @homestartlam @thebigissue
Expand

Seb Skuse @sebskuse 17m
This feels like a long long time ago. t.imehop.com/13OOWGw
Expand

UoSinternational @UniSotonIntl 19m
Hartley Bear is back in the UK for a bit relaxing in the sun with a Bubble tea on campus... fb.me/SHRRu46v
Expand

(a) Home timeline

Tweet to Nigel Shadbolt

@Nigel_Shadbolt

Who to follow · Refresh · View all

Couchbase @couchbase Promoted · Follow

WWW2013 @www2013rio Followed by lescarr and others Follow

Rufus Hound @RufusHound Follow

Browse categories · Find friends

Trends · Change

#HorseTracker Promoted
#wiw13
Kurt Cobain
#ThxDanWarpGloZellOnSamandCat
Sprinter Sacre
#ThingsThatAnnoyMe
#Aintree
#soslam
HMV
Jeremy Kyle

© 2013 Twitter About Help Terms Privacy Blog Status Apps Resources Jobs Advertisers Businesses Media Developers

Nigel Shadbolt
@Nigel_Shadbolt
http://users.ecs.soton.ac.uk/nrs/

691 TWEETS 69 FOLLOWING 3,997 FOLLOWERS

Tweets

Nigel Shadbolt @Nigel_Shadbolt 21 Mar
Full on day of HMG #opendata meetings - Local Public Data Panel, Midata and then Transparency Board - let the data flow!
Expand

Nigel Shadbolt @Nigel_Shadbolt 20 Mar
And today keynote for #opendata for charities at #cfgit13
Expand

Nigel Shadbolt @Nigel_Shadbolt 19 Mar
Speaking at #opendata what's the use - at #bltalkscience gig tonight
Expand

Nigel Shadbolt @Nigel_Shadbolt 18 Mar
@timberners_lee: A great day for the Power of Open #QEPrize for #engineering #opendata
Expand

Nigel Shadbolt @Nigel_Shadbolt 18 Mar
RT @timberners_lee: Very honoured w Louis, Vint, Bob & Marc be awarded 1st #QEPrize for #engineering <- fab fab fab!
Expand

Nigel Shadbolt @Nigel_Shadbolt 12 Mar
"@tommyh: The new housing map explorer here is really cool linkeddata.hants.gov.uk #linkeddata #opendata #transparency4u /cc

(b) Profile timeline

Figure 2.2: Example Twitter timelines

2.3 Defining Power and Influence in Blogs and Microblogs

In an example of the perceived power of tweets, Guardian columnist Jan Moir felt that she became the victim of a "heavily orchestrated internet campaign" following public disapproval of her column on Stephen Gately's death (Guardian, 2009). More recently, the Arab Spring uprisings in 2010–2011 have been of particular interest to researchers of social networks and power. Mansour (2012) documented the events of the Arab Spring chronologically, and provided breakdowns of the Middle East user demographics and TV, Internet and Twitter penetration rates. The survey participants in this paper claimed that social networking sites were imperative in becoming information sources and mobilizing communities. Yette (2012) coded different tweets according to content, links, reasons for tweeting, languages and hashtags. However, despite these more technical analyses of the Twitter dataset surrounding the Arab Spring, there is still a lack of a definitive way to quantify power and influence online. Blogs and tweets have been perceived to hold power in influencing outcomes in the real world. However, there is a continuing debate in the community about the best ways in which to measure power and influence in microblogging services like Twitter.

Cha et al. (2010) claimed that it is unclear what influence means, particularly due to its different definitions and the lack of empirical evidence. In an attempt to approximate the power of blogs or tweets, existing computer science literature have used propagation as a metric to measure influence, using audience reach (Gill, 2004) and readership traffic (Farrell and Drezner, 2008).

In contrast, social science's discourse on power is not new. In Russell (1975), power is defined as the production of intended effects, thus suggesting an ordinally quantitative characteristic; A has more power than B if A achieves many intended effects and B only achieves a few. More recently, Lukes (2005) critically reviewed various prior work arguing for different definitions of power. The research outlined three different dimensions of power based on these arguments:

First dimension — Overt and Observable: Direct application of power, which is based on observable behaviour in decision-making involving actual, observable conflict. This involves capacity (A can get B to do something that B would not otherwise do), coercion (A secures B's compliance with deprivation threats), authority (B complies in recognition of A's command in terms of B's own values) and force (A strips the choice between compliance and non-compliance).

Second dimension — Covert, Controlling Agendas: Power based on shaping agendas, particularly how decisions are prevented from being taken on potential issues with observable conflict of interest.

Third dimension — Shape Desires and Beliefs: Power through promoting an ideology or a particular philosophy. This could keep potential issues away from consideration, happening in the absence of actual, observable conflict, such as latent conflict. This involves manipulation (a sub-concept of force where B complies without realizing it). Lukes claimed that "we need to attend to those aspects of power that are least accessible to observation" and that "power is at its most effective when least observable."

Philosophical definitions of power, as provided by Russell (1975) and Lukes (2005) seem to contain several nuances; the definitions involve more dimensions such as the visibility of power, and differences in impact. In contrast, communications power pertains to the roles of actors involved in a network, and the reach of an individual becomes an important metric in approximating power (Castells, 2009).

In the context of existing power, several studies have outlined different methods such as follows:

Persuasion: Power is the "relational capacity that enables a social actor to influence asymmetrically the decisions of other social actor(s) in ways that favor the empowered actor's will, interests, and values" (Castells, 2009).

Constraints: "Every tyrant knows that it is important . . . not only to constrain people's actions but also to manipulate their desires, partly by making people fearful, partly by putting certain options in an unfavorable light, partly by limiting information", because unavailable information leads to people ending up not wanting them at all (Sunstein, 2007). Woodly (2008) also observed the power of enforcing constraints; political elites protect their control over popular political epistemologies, shaping and bounding public debate, in a way that serves distribution of power.

Both of these methods of exerting power corroborates Lukes's assertion that there are different types of power, and thus different ways and means of exerting influence (see Section 2.3 on page 14). These methods support the notion that power gets exerted by certain groups of people who are also called the *influentials*. In this context, power propagates in a top-down **hierarchical structure**. Castells (2009) claimed that the person in power decides what is valuable.

However, apart from hierarchical power, there also exists an alternative perspective which acknowledges the role of **peers and networks**. Cha et al. (2010) claimed that there are conflicting views when studying the adoption of online trends; the *traditional view* emphasizes the existence of a persuasive and well-connected select few, while the *modern view* emphasizes decision-making based on opinions of peers rather than influentials.

Gladwell (2001) noted that in the case of determining a child's character, the influence of peers and the community is more important than family. Likewise, Westen (2008) claimed that "political persuasion is about networks and narratives". In parallel to this, Castells (2009) claimed that communication networks are fundamental in making power, where power-holders — not necessarily those in government — are networks of actors using their power within their own areas of influence, through networks built around their interests.

Existing computer science literature is based more on communications power rather than the philosophical definitions of power. Cha et al. (2010) claimed that it is unclear what online influence means, particularly due to its different definitions and the lack of empirical evidence. According to Agarwal et al. (2008), it is difficult to evaluate influence within the blogosphere due to the **"absence of ground truth about influential bloggers"**. To address this situation, existing research studies have been using propagation as a measure for influence, which then becomes a proxy for power. Influence has been equated to various propagation metrics such as audience reach (Gill, 2004) and readership traffic (Farrell and Drezner, 2008).

Given the above considerations, the next subsection discusses the different metrics that have been used in existing studies to approximate power within microblogs, and the challenges present in quantifying online power.

2.4 Influence Metrics via Blogs and Tweets

The following studies have attempted to quantify power and influence within blogs and tweets using the following:

Retweets: Looking at the Arab Spring, Choudhary et al. (2012) studied the Twitter dataset during the Egypt uprising in January 2011, and defined influential twitterers as those whose tweets were retweeted most frequently. This was similar to the approach taken by Stieglitz and Dang-Xuan (2012), where in a dataset of tweets collected during a German 2011 election, users were defined as influential if they were retweeted the most. A more in-depth look at retweets was done by Starbird and Palen (2012); in addition to using retweets as a metric for identifying influential twitterers, they were also used as a metric to pinpoint popular tweets over the course of the revolution, and to quantify the extent of these tweets' reach. Kong et al. (2009) defined an influencer as someone whose tweets "trigger further actions such as RTs and replies from other users". This group was found by calculating the ratio of individual retweets compared to the overall total of retweets. In this ratio model, influence is not approximated solely on the volume of retweets, but by the proportion of users being influenced to make an action such as retweeting. Other ratios have also been used, namely the ratios of retweets and mentions over total tweets, and interactors — total users who have retweeted or mentioned the author — over total followers (Anger and Kittl, 2011).

Total followers and tweets: A strong positive correlation was found between the number of tweets and followers, although no correlation was found between the volume of tweets and influence based on retweets or mentions[1] (Cha et al., 2010). Total followers and tweets as metrics were also used in the study by Zaman et al. (2010), which proposed that retweetability could be predicted based on the relationship between the original author and the retweeter, followed by the number of retweets made by followers and following users. The model used was claimed to be able to predict retweetability up to one-day forward before it loses accuracy.

Inlinks, permalinks, citations and PageRank: One of the earlier studies on measuring authority in the blogosphere was Marlow (2004), which investigated the ranks of blogs using blogrolls and permalinks. It was found that rankings based on blogrolls were prone to favour older blogs and susceptible to selection bias, therefore rankings based on permalinks were proposed as a better proxy for influence (Marlow, 2004). Market Sentinel (2005) looked at influence in the context of issues, producing an "Issue Influence Index" which is similar to

[1] Cha et al. (2010) measured influence via three metrics: indegrees, retweets and mentions. A high volume of any of the three would signify high "influence" in those respective metrics.

PageRank[2]. The study calculated influence based on citations/link counts, PageRank and the percentage within the whole discussion. Agarwal et al. (2008) argued that influential posts are longer, contain more comments and fewer outlinks; the paper focused on recognition (inlinks), activity generation (comments), novelty (outlinks) and eloquence (post length) as measures of influence.

Shared views and topical similarity: Mustafaraj and Metaxas (2010) investigated the senate elections in Massachusetts and the role of real-time search engines which retrieve content from blogs, news and tweets. It was found that people are more likely to retweet items from twitterers whom they agree or share political views (Mustafaraj and Metaxas, 2010). In this study, it was not possible to quantify user responses such as URL clickthroughs, but the possible reach of these URLs suggests the possibility of the medium being exploited to increase awareness for minimum cost. Weng et al. (2010) found that twitterers are more likely to follow those who are interested in similar topics. The paper proposed TwitterRank, an extension of PageRank which also incorporates topical similarity in its rankings. The paper argued that rankings based solely on the number of followers are not accurate because certain following relationships do not necessarily indicate influence. Shared views and topical similarity could be used as factors to explain why a piece of information reaches more users than others, and possibly predict the future volume of propagation and approximated power for any piece of information.

Mutual links and text: Adar et al. (2004) investigated propagation patterns of blog posts according to similarities in links, text and any repeated history of infection. The paper reported that mutual linking blogs are 45% likely to mention one common URL.

Implicit link structures: iRank (Adar et al., 2004) used triplets of URLs, blogs and citations to represent spreading patterns. The study found that via links — links explicitly connecting one blog to another — are quite rare, namely 70% of blog mentions are not attributable to direct links. Therefore, blogs were sorted according to implicit link structures, identifying information sources that later became widely linked to, as the order of blogs within a timeline indicated implicit links. This thesis builds on the findings from the work by Adar et al. (2004), notably the idea of implicit link structures and their parallels in Twitter such as retweets made using informal retweeting mechanisms. Adar et al. (2004) proposed that observations of spreading patterns should include explicit and implicit links. Similarly, this thesis proposes that retweets made using informal retweeting mechanisms should also be considered in tweet propagation studies.

[2] PageRank is an algorithm that ranks the relative popularity of websites by counting the number of their inlinks and outlinks and assigning numerical weights to websites according to those link counts. For example, a website with more inlinks than outlinks will be given a larger weighting than a website with more outlinks than inlinks (Page et al., 1999).

Ratios of tweet mentions and batting averages: Tweetminster (2010) calculated influence as a ratio of total tweet mentions over the total number of a person's tweets, weighted by the time period between account signups. Meanwhile, Aizen et al. (2004) analyzed usage data from the Internet Archive to find the "batting average" — the percentage of hits which led to downloads from the archives. The study found that popularity changes are discrete, sudden, and related to events both online and offline.

Contextual themes — Novelty, conversation and interestingness: Local context is claimed to be important in understanding blogging behaviour (Zuckerman, 2008). Song et al. (2007a) proposed InfluenceRank which combined PageRank with novelty. Opinion leaders were also deemed to be those who formed and reflected mass opinion based on the theory of diffusion of innovations[3] (Song et al., 2007a). Wu and Huberman (2007) found that a dynamic model based on novelty seemed to consistently determine the natural time scale of when attention fades. When interesting stories are passed on to others, they garner even more views, resulting in a positive reinforcement effect, which is claimed to expedite the spread of stories, while fading novelty seemed to cause attention to diminish (Huberman, 2008). An algorithm to quantify novelty amongst news articles was presented in Gabrilovich et al. (2004), using differences in content, structural organization and time as proxies towards quantifying novelty. Meanwhile, Leskovec et al. (2007) proposed a metric for conversation mass — the number of posts which followed after a preceding blogger's post. Choudhury et al. (2009) focused on the 'interestingness' of participants and conversations. The paper found that 'interestingness' mainly affected three variables, namely participation in related themes, participant cohesion and theme diffusion.

Diffusion patterns: Song et al. (2007b) found an asymmetric inter-personal influence, where the spread is in the form of innovation or imitation. Based on the theory of the diffusion of innovations, the paper proposed DiffusionRank, a ranking algorithm that ranks blogs according to how quickly and efficiently information flows through them. This ranking system uses a rate-based information flow model to provide recommendations to users. Adar et al. (2004) found four epidemic profiles, namely chatter, spikes, rapid decay and slower decay. The study found that these profiles correlated to the type of content that was being spread. News editorials and opinion pieces exhibited a slower rate of decay compared to articles coming from

[3] Diffusion of innovations was introduced by Rogers (1962) as a theory to explain the spread of new ideas and technology. The theory proposed four key elements of diffusion: innovation, communication channels, time and social systems. Rogers (1962) also coined the terms early adopter, early majority and late majority to signify the user groups involved in the lifecycle of innovation adoption.

Slashdot[4], which showed the “Slashdot effect”[5] of spiking then decaying rapidly. Gruhl et al. (2004) also investigated chatter and spike topics. In addition, the author adapted theories of infectious diseases like Susceptible-Infected-Recovered (SIR) to model propagation. Leskovec et al. (2007) also used the SIR epidemic model to generate cascade models.

Network characteristics of betweenness centrality and hops: Krauss et al. (2008) looked at various factors such as positivity, intensity, betweenness centrality and time noise, while Matsumura et al. (2010a) proposed the influence diffusion model (IDM) which evaluated the spread of terms in the blogosphere, assigning influence according to the number of hops[6] a term made from one author to another.

Prior friends: So far, all the above metrics describe properties of information propagation; these have been used as proxies to quantify power and influence. Prior friends is another metric that is more related to thresholds for joining networks, from a user network to a propagation network. Kleinberg et al. (2007) found that the threshold for contagion[7] is at most 50%, where the likelihood of someone deciding to join an existing group seemed to increase if existing friends had joined the group before. This finding is very useful to determine the factors of influencing a non-member to join a particular group. In a later paper, Kleinberg (2008) argued that the likelihood of knowing someone is dependent on any shared occupations, cultural backgrounds, or organizational roles. Backstrom et al. (2006) also found that adoption depended on the number of friends who had already adopted. The probability of joining was found to be directly proportional to linkage density within the subsets of friends who had already adopted. Other observed metrics included how friends were connected within the adopters network, the network’s growth itself, and any bursts of changes in membership. The paper found diminishing returns over larger number of friends already in the network. This suggests the existence of a saturation point, where up to a certain point, the number of friends already in the network would no longer influence a non-member to join the network.

[4] Slashdot (http://www.slashdot.org/) is a US-based technology news website that has 4.2 million unique visitors per month as of 2013 (http://slashdotmedia.com/about-slashdot-media/slashdot-org/).

[5] The “Slashdot effect” refers to the event where a popular website posts a link to a smaller one, thus causing a substantial spike in web traffic which overloads or takes down the smaller site’s web servers temporarily.

[6] Hop: the number of times a tweet has been passed on. A single hop means that User A passed a tweet to User B, whilst a double hop means that User A sent a tweet to User B, who then subsequently sent it to User C.

[7] Contagion happens in a network of nodes when a subset of those nodes adopts a new behaviour, and then in turn converts some or all of the remaining nodes into adopting the same behaviour (Kleinberg et al., 2007).

All of these studies have been various attempts in the Computer Science community to approximate and predict power and influence within blogs and tweets through measurable metrics. However, in recent studies, retweets have become pivotal in measuring power and influence; investigating the Arab Spring (Choudhary et al., 2012; Starbird and Palen, 2012), observing sentiments over a German election (Stieglitz and Dang-Xuan, 2012), and identifying influential Twitter users in Korea (Kong et al., 2009) and Austria (Anger and Kittl, 2011). These studies use propagation and reach as an approximation of power, in line with the communications power model described by Castells (2009). This relationship between propagation, retweeting and power means that retweets are key to many power and influence analytics. Therefore, it is important to model retweets properly; having an accurate definition of retweets, and solid methods of retrieving retweets is critical for the study of power and influence in microblogging.

2.5 Retweets

As shown in the previous section, there have been many studies on identifying the flow of information and the metrics which best represent this flow. In particular, the common thread amongst these studies is the importance of passing information across a network, be it posts, user activities or other content.

As briefly described in Section 2.2 (see page 12), the concept of passing on posts to other users have evolved within microblogging communities through mechanisms such as retweets in Twitter and prefixing '//' in front of Sina Weibo posts. For example with Twitter, User A may republish User B's tweet to User A's followers using the retweet mechanism. The conventional mechanism involves clicking on the double arrow icon next to User B's tweet, or by copying and pasting the same tweet and prefixing it with common retweet markers such as 'RT' before it gets republished.

Since the introduction of Twitter and microblogs, researchers have been focusing on patterns of propagation across Twitter, with a particular interest in retweets. In sociology, Murthy (2012) suggested future Twitter research to utilize the interactionist work from Goffman (1981), which presented three themes of conversations: 'ritualization', 'participation framework', and 'embedding'. In particular, Murthy (2012) proposed that the act of retweeting fits well within the 'embedding' theme parameter, where the retweeter embeds the text from an originator then disseminates it.

Boyd et al. (2010) was one of the earliest studies focusing on retweets. This study outlined the evolution of Twitter and some of its usage conventions, including what is a retweet, the construction of retweets, how people retweet and why. The paper found that people have been retweeting as a call for social action — fundraising, demonstrations or collective group identity-making — and crowdsourcing. Two datasets were used; one being a random sample of 720,000 tweets taken in five-minute intervals, and the other being a sample of around 203,000 retweets. From the random sample, the study claimed 3% were retweets, and that the existence of URLs increases the retweetability of that tweet. However, this study explicitly did not consider tweets which used informal retweeting mechanisms due to the difficulty of determining them as retweets. The 3% retweet percentage may be different if these informal retweets were taken into consideration.

Kooti et al. (2012) chronologically described the evolution of retweeting conventions, as shown in Table 2.1, since the launch of Twitter back in 2006. Early adopters were found to constitute a substantial percentage (69.7–86.1%) of the top 1000 highly-connected core users of these retweeting variations. This table includes the term HT which is variously defined as ‘hat tip’ or ‘heard through.’

Table 2.1: First tweets using each retweet variation, as published by Kooti et al. (2012)

Variation	Username	Date	Text
via	@tagami	16/03/2007	@JasonCalacanis (via @kosso) - new Nokia N-Series phones will do Flash, Video and YouTube
HT	@TravisSeitler	22/10/2007	The Age Project: how old do I look? http://tweetl.com/21b (HT @technosailor)
Retweet	@kevinks	01/11/2007	Retweet: @AHealthyLaugh is in the Boston Globe today, for a Stand up show she’s doing tonight. Add the funny lady on Tweeter!
Retweeting	@musicdt	05/01/2008	Retweeting @Bwana: Is anyone streaming live from CES? #ces
RT	@Tdavid	25/01/2008	RT @BreakingNewsOn: “LV Fire Department: No major injuries and the fire on the Monte Carlo west wing contained east wing nearly contained.”
R/T	@samflemming	20/06/2008	r/t: @danwei Live online chat with Chinese President Hu Jintao. http://tinyurl.com/5qqecp. He claims he uses net to know netizen concerns
♺ (recycle icon)	@claynewton	16/09/2008	♺ @ev of @biz re: twitterkeys ☆ http://twurl.nl/fc6trd

The majority of retweets appeared to include URLs in them; more than half of retweets contained URLs (Zarella, 2009; Boyd et al., 2010) while Cha et al. (2010) found that 92% of tweets which had ‘RT’ or ‘via’ in them also had a URL.

Other studies are also formed on the basis of retweets. The study by Webberley et al. (2011) focused on the characteristics of retweets, particularly on retweet chain lengths, follower/following networks, retweet group sizes and retweet time delays. For example on information diffusion van Liere (2010) claimed that the pattern of information brokerage seemed to best describe the diffusion of information seen through retweeting. Whilst in determining rankings, (Kwak et al., 2010) also found that user rankings based on retweets seem to differ as compared to rankings based on total followers and PageRank values. Conover et al. (2011) investigated the political polarization of clusters within retweet and mention networks on Twitter. Retweet clusters were found to be more likely to preferentially spread information within their own communities, whereas this pattern was not detected within mention networks.

In a study comparing Digg[8] and Twitter, Lerman and Ghosh (2010) compared the propagation patterns between these two websites. The act of diffusing could be appropriated by a user's action of voting on Digg or retweeting on Twitter. Both were found to diffuse information via the same network based on friends or followers. Nonetheless, propagation networks in Digg were found to be denser, as links would be voted up/down by connected friends before highly voted stories become featured on Digg's front page to many more unconnected users. In contrast, the paper claimed that tweets spread in Twitter at a slower rate than Digg, but reaches further users within the Twittersphere.

Zarella (2009) proposed a retweet decision model to visualize the decision-making process that users make when deciding whether to retweet something or not. Figure 2.3 (page 24) shows that there are three factors which influence the likelihood of a retweet: the number of followers, the attention of the receivers, and the motivation of the receiver to retweet. The model also shows that if someone does not follow the sender of a tweet, then a retweet will not happen.

However, cases do exist where no visible follower/following paths are apparent, or the retweeter had decided not to attribute tweets according to usual retweet conventions. Galuba et al. (2010a) found 33% of retweets were made by non-followers, whilst Webberley et al. (2011) found that 10% of retweets made within the first hop were published by non-followers. With respect to attributions, in some content domains such as call for action tweets, 95% of tweets were missing attribution data in them (Nagarajan et al., 2010). Therefore, these studies show that Zarella's retweet decision model is inaccurate.

The work by Bastos et al. (2012) observed the proportions of retweets (RT-messages) and replies (AT-messages) across different themes, such as events, technology, politics, altruism, etc. They found that there was no significant correlation between proportions of RT- and AT-messages and follower-followee networks, suggesting that users do look outside their follower-followee networks for content. Similar to Bastos et al. (2012), this thesis proposes looking at the different ways tweets could propagate, involving both retweets and replies. In addition, this thesis also makes a distinction between follower

[8] Digg (http://digg.com/ is a social news aggregator which allows readers to vote any particular web content up (*digging*) or down (*burying*).

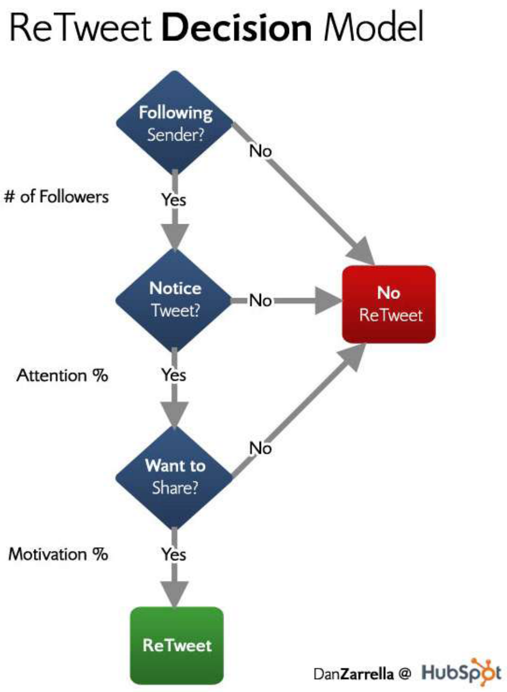


Figure 2.3: Retweet decision model (Zarella, 2009)

and non-follower retweets, resulting in a typology of tweet propagation types, which will be explained in Chapter 4 (page 55).

2.5.1 Retweet Visibility

Boyd et al. (2010) found that not all retweets were formally marked as retweets. Elements such as A following B and B using the same URL A used can suggest a reference, but it was difficult to account for this consistently. Similarly, Nagarajan et al. (2010) found that out of the top 10 tweets in the datasets for the Iran Election and Healthcare, only 20% followed the explicit retweet syntax and credited another author. Particularly, tweets categorized under calls for action, crowdsourcing or collective group identity-making were more likely to be missing author attributions, despite a well-connected follower graph; only 5% of tweets contained attributions data in them. The authors concluded that people might have seen the tweets but were not compelled to attribute the originating author, possibly because calls for action are usually unattributable to individuals, or they did not see the originating tweet. This may mean that retweet patterns are domain- or content-specific, thus further investigation is needed to prove this.

In another study, Galuba et al. (2010a) defined two types of information cascade: an F-cascade, where URL spreads are constrained to followers graphs, and an RT-cascade, where follower graphs are disregarded and only data related to a retweet's originating author is used. The paper found that even though there were large overlaps, 33% of retweets credited people they do not follow. The paper postulated that URLs tweeted by highly connected authors were more likely to be retweeted by their followers, but this causality is likely to be bidirectional. Although more followers could mean more potential retweets, more followers could also be accumulated due to the spread of interesting tweets.

Another paper by Fujiki et al. (2011) also looked at non-follower retweets, proposing a way to improve retweet studies by eliminating bias. In this paper, retweets made by non-followers were given higher weight compared to retweets made by followers.

By combining the findings of Nagarajan et al. (2010) and Galuba et al. (2010a), this research work builds upon the notion that not all retweets contain proper attributes, and when they do, they appear in different ways and in varying proportions.

2.5.2 Temporal Spreads

Temporal analysis of retweets suggested that a message would get retweeted up to 5 hops away from the source within a median of one hour (Kwak et al., 2010). In addition, more than 60% of all retweets were made within the first hour, decreasing rapidly until nothing gets retweeted after 24 hours (van Liere, 2010). This finding is supported by Galuba et al. (2010a) which found that the diffusion delay between URL tweets were distributed log-normally, with a median of 50 minutes.

2.6 Identifying Retweets Made Using Informal Retweeting Mechanisms

Boyd et al. (2010) had acknowledged the existence of tweets which contained texts which were similar to previously published tweets, yet do not contain conventional retweet markers. However, due to the difficulty in determining the provenance of these tweets, they were not focused upon within their study.

In Hoang et al. (2011), it was acknowledged that inferred retweets existed — retweeting relationships could be inferred between pairs of tweets which were not initially classified as retweets. Several researchers have tackled the problem of undetectable tweet propagation paths by making assumptions about the existence of these paths. One such assumption is that propagation paths exist between subsequent tweets based on the timestamps of those published tweets and the content similarity between them. The study that is being presented in this thesis is most similar to the work done by Adar et al. (2004), which investigated the existence of implicit URL links within the blogosphere. Similarly, the work by Matsumura et al. (2010b) also included the assumption that if a collection of blogs which contained the same URL or trackback also contained the same terms, then it was assumed that the first blog in that collection was influencing the subsequent blogs. The paper written by Galuba et al. (2010b) described the F-cascade within tweets. It involved users who seemed to have copied a URL that was previously tweeted by someone they follow.

The work by Myers et al. (2012) attempted to differentiate tweet propagations caused by internal sources (such as a friend seeing and then passing on a URL from another friend within a same network) as opposed to external unseen factors (such as an offline event). Working on a dataset of 3 billion tweets over Jan 2011, this paper found that 71% of URL mentions in tweets happened due to network diffusion, whereas the other 29% happened due to external unseen factors.

In the work done by Wu et al. (2011), the phrase "reintroduction of content" was used to describe intermediary tweets which are similar to previously published tweets but also do not contain conventional retweet markers. Their dataset consisted of tweets which only had URLs in them. In this paper, retweets and reintroductions were treated equivalently, with no separation between the two. The same approach was taken by Bakshy et al. (2011), who studied influence prediction within Twitter. Their metrics indicating influence were not restricted to just retweets containing retweet markers within the tweet texts. Their study also used the approach of using URLs as unique keys which group tweets together, thus all instances of tweets which include the URLs being focused on were considered as a "rebroadcast" of influence. Again, their paper did not differentiate between retweets using formal and informal retweeting mechanisms.

Yang and Counts (2010) looked at the propagation of mentions within tweets, therefore incorporating both retweets and replies, without differentiating between the two. This paper focused on the speed, scale and range of the propagations. The rate a user has been mentioned beforehand seemed to be a good predictor to how far a tweet will propagate in the future.

In the study by Nagarajan et al. (2010), tweets "without indication of retweeting or making references to others" were initially classified as "other" tweets. The paper studied datasets based on three topics: Health Care Reform Debate, the Iran Election, and the ISWC conference. Similarity engines were then used to retrieve tweets similar to the top 10 most frequent tweets in each of these three datasets. This allows tweets without explicit retweeting markers to be grouped together. The retweet patterns of these groups were then subsequently studied. The paper claimed that tweets for calls for action, collective groups and crowdsourcing domains are more likely to have more unmarked, unattributed retweets, as opposed to information sharing tweets.

A similar work on topical domains was done by Bandari et al. (2012), which looked at links — propagated via both original tweets and retweets — across Twitter, and found that technology blogs seemed to be propagated more than traditional news media.

Given the above studies, this thesis presents a combination of the theory of implicit links as presented by Adar et al. (2004) with the methodology of only using tweets containing URLs in them (Wu et al., 2011; Bakshy et al., 2011). The hypothesis of this paper is that there exists a significant minority of dark retweets, or tweets which are propagated without using formal retweeting mechanisms.

Several existing studies, like information brokerage in van Liere (2010), and hashtag adoption in Yang et al. (2012), only looked at retweets containing 'RT' or 'via' keywords in their texts, or metadata labelling tweets as retweets. The approach used by these studies may lead to some retweets being overlooked.

2.7 Ethics of Twitter Research

All tweets published on Twitter can be publicly seen by anyone unless the author opted to protect their user account — *"What you say on Twitter may be viewed all around the world instantly"* as prominently displayed on Twitter's Terms of Service[9]. A protected account makes user profiles and timelines inaccessible to the general public unless granted prior approval. This approval is given when a "follow" request is made and that request is approved by the protected user. An approved follower may then see all the tweets made by the protected user.

In an existing study of 505 million Twitter users, it was estimated that only 5.97% of accounts were protected, leaving a remaining 94% of accounts open to the general public (Gabielkov et al., 2014).

Several papers have discussed the ramifications of Twitter's privacy settings and the affordances allowed towards Twitter research. The main ethical issue underpinning Twitter research lies in the simplicity of retrieving identifying information from Twitter and the responsibilities as a researcher to respect Twitter users' notion of privacy. In particular, there is the inherent assumption that tweets can be viewed unproblematically as a public platform (Thomson, 2012). To this end, several white papers have proposed various recommendations towards a code of conduct for Twitter researchers (Kelley and Cranshaw, 2011; Rivers and Lewis, 2014).

Twitter themselves have enforced their own policy towards the distribution of tweet datasets, which affects researchers handling tweet corpora in their studies. Their policy[10] states that *"[i]f you provide downloadable datasets of Twitter Content or an API that returns Twitter Content, you may only return IDs (including tweet IDs and user IDs)."*

Therefore, tweet data such as tweet texts cannot be distributed as is amongst third parties. The acceptable method of distributing datasets is to publish tweet IDs only,

[9] https://twitter.com/tos

[10] Twitter Developer Rules of the Road: https://dev.twitter.com/terms/api-terms

therefore any researcher can retrieve the corresponding tweet data themselves using the Twitter API.

The experiments carried out for this thesis revolve around aggregated information. All findings in this thesis do not involve presenting identifiable information beyond the boundaries set by Twitter's developers policy.

2.8 Conclusion

As outlined in Section 2.4, retweets are one of the most important metrics used to measure influence in microblogs. If the existence of dark retweets is indeed prevalent amongst all Twitter datasets looking at propagation, then there may be the possibility that existing work on propagation has been missing out on hidden data that had been disregarded due to difficulties in detecting dark retweets.

Based on this motivation, an experimental toolkit was created in order to analyze the different ways a tweet may propagate, in order to potentially provide a more comprehensive overview of tweet propagation. The next chapter presents the setup for the pilot study that was done with over 11,000 tweets, using a preliminary version of the experimental toolkit, and the results of this pilot study.

Chapter 3

Pilot Study of Tweet Propagation Types

This chapter describes the pilot study that was done over 11,000 tweets. A preliminary typology of tweet propagation types was used in this initial study. The findings of this study was then used to develop a more comprehensive typology, which will be explained in Chapter 4: Typology.

In the pilot study, a typology of propagation types was proposed, which included seven different ways of retweeting. This typology was derived from an alternative retweet decision model, which was built upon Zarella's retweet decision model (Zarella, 2009), as outlined in the previous chapter (see Figure 2.3 on page 24).

3.1 Description of Pilot Typology

The pilot study was based on seven different ways a tweet could propagate, namely: 1. native retweets, 2. native non-follower retweets, 3. RT/Via retweets, 4. RT/Via non-follower retweets, 5. replies, 6. non-follower replies, and 7. other implicit retweets.

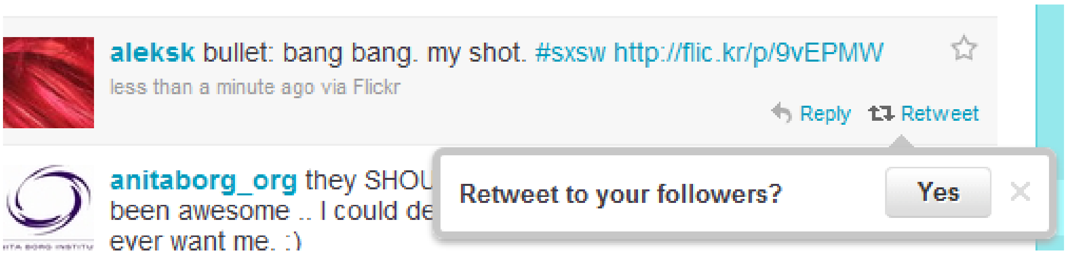
This classification is derived from several prior research papers. The separation of follower and non-follower groups is derived from the works by Galuba et al. (2010b), which described cascade patterns of URLs made by followers and non-followers, and by Fujiki et al. (2011), which looked at eliminating bias by emphasizing the weight of non-follower retweets. The different mechanisms — native, RT/Via, replies and other implicit retweets — were included based on the different mechanisms that currently exist in Twitter and have been studied in various studies looking at 'RT' and 'via' keywords (van Liere, 2010), mentions (Yang and Counts, 2010) and implicit propagation (Adar et al., 2004; Matsumura et al., 2010a; Nagarajan et al., 2010).

The screenshots in this section, as shown in Figures 3.1, 3.2, 3.3, 3.4 and 3.5, were all taken at the time the pilot study was run, which was in March 2011.

3.1.1 Native Retweets

Native retweets are defined as tweets which use the retweet mechanism provided either by Twitter's proprietary user interfaces (webpage, apps) or by the Twitter API. Retweets sent via Twitter API's retweet function using third party applications are also included.

For example, from a user's point of view, the user sees a tweet from someone they follow, and decides to retweet it. The user then clicks on the 'Retweet' link to propagate the same tweet text to his/her followers, as shown in Figure 3.1.



aleksk bullet: bang bang. my shot. #sxsw http://flic.kr/p/9vEPMW

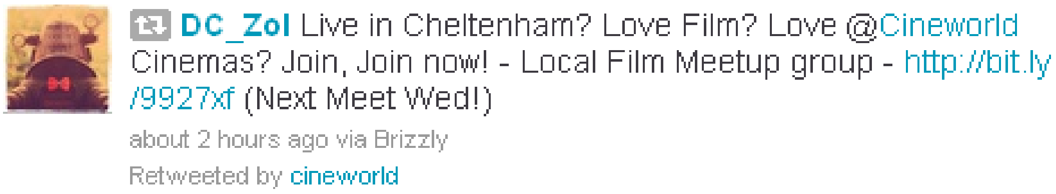
less than a minute ago via Flickr

anitaborg_org they SHOU been awesome .. I could de ever want me. :)

Retweet to your followers?

Figure 3.1: Screenshot of user's view of native retweeting

This propagation method preserves the entirety of the tweet text. When it is retweeted, the user's followers will see the originating author's full tweet. This is followed by a sentence at the bottom stating 'Retweeted by [retweeter's username]' which acknowledges the user who is doing the retweeting. The resulting tweet seen by the user's followers is illustrated by Figure 3.2.



DC_Zol Live in Cheltenham? Love Film? Love @Cineworld Cinemas? Join, Join now! - Local Film Meetup group - http://bit.ly/9927xf (Next Meet Wed!)

about 2 hours ago via Brizzly

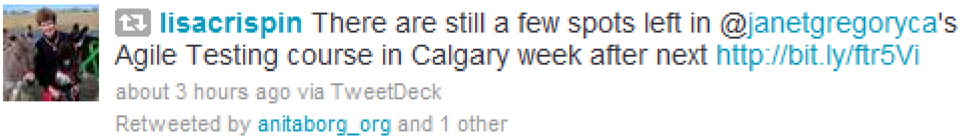
Retweeted by cineworld

Figure 3.2: Screenshot of example native retweet

Another type of native retweeting is done by third party applications which use Twitter API's own proprietary retweet function. In this case, the user sees a tweet via a third party application such as TweetDeck[1], and then retweets it. TweetDeck then sends this request via Twitter API's retweet function. The end result is similar to Figure 3.2, but the timestamp will display 'about # hours ago via [third party application name]'.

In the pilot study's toolkit, native retweet counts were found by querying the Twitter API using the 'retweeted_by' function, sending a unique tweet ID number as a parameter.

[1] As of May 2011, TweetDeck is now owned by Twitter.



lisacrispin There are still a few spots left in @janetgregoryca's Agile Testing course in Calgary week after next http://bit.ly/ftr5Vi

about 3 hours ago via TweetDeck

Retweeted by anitaborg_org and 1 other

Figure 3.3: Screenshot of example native retweet made by a third party application

The API returns a list of users who have used the native retweet mechanism to propagate that unique tweet. The 'retweeted_by' function includes the retweets that were made using this native mechanism only.

3.1.2 Native Non-follower Retweets

Native non-follower retweets are native retweets which were made by non-followers. These retweets were made using Twitter's native retweeting mechanism, by users who do not follow the originating author of the original tweet.

For example, User B makes a native retweet of User A's tweet. User C follows User B, so sees this retweet and decides to make a *native retweet* of User A as well. In this instance, User C follows User B, but *does not follow* User A. Therefore, the connection between User C and User A is a **non-follower** relationship, and User C's retweet is classified as a **native non-follower retweet**. These retweets happen due to Twitter's native retweet architecture, which links a retweet straight back to the originating author, User A, instead of the intermediary users such as User B.

3.1.3 RT/Via Retweets

RT/Via retweets (see Figure 3.4 on page 32) are defined as tweets which repeat prior tweets and include any of the following RT/Via markers within the tweets:

> "rt @", "rt@", "rt:@", "rt: @" , "retweet @", "via @", "retweet :@", "r/t", "rt:", "RT @", "RT@", "RT:@", "RT: @", "RETWEET @", "VIA @", "RETWEET :@", "R/T", "RT:"

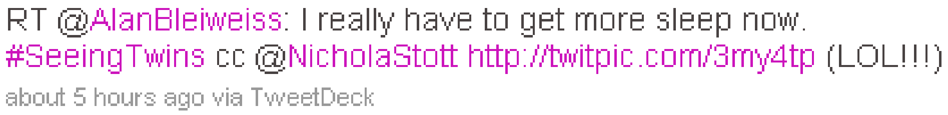
These markers were chosen to replicate the approach used by Boyd et al. (2010). This study focused only on variants of 'RT' and 'via' because "these two variants . . . provide a diverse dataset of retweets" (Boyd et al., 2010).

For example, from a user's point of view, he/she sees a tweet from someone they follow, and decides to retweet it by copying the tweet text and then prefixing it with any of the aforementioned markers.

These tweets include those made by third-party apps which cut and paste prior tweets and prefixes any of the above RT/Via markers onto the tweets, before posting them via the Twitter API. This is different to using Twitter API's own proprietary retweet function.

This retweeting mechanism allows the user to modify the tweet text in various ways. For example, the user can append the text with his/her own commentary, which may be positive, negative or neutral. This retweeting mechanism has also been outlined in Boyd et al. (2010), where the tweet text might be truncated or even have its meaning altered. This is due to retweeters modifying the text in order to fit in comments and the actual tweet within the limits of 140 characters.

In order to record RT/Via retweets, the pilot study's toolkit parsed each tweet text and looked for any occurrences of the above RT/Via markers. If any of these markers existed in the text, the toolkit then checked whether this tweet had been recorded as a native retweet. If the tweet text contained an RT/Via marker, and it hadn't been classified as a native retweet, then the toolkit recorded it as an RT/Via retweet.


RT @AlanBleiweiss: I really have to get more sleep now. #SeeingTwins cc @NicholaStott http://twitpic.com/3my4tp (LOL!!!)

about 5 hours ago via TweetDeck

Figure 3.4: Screenshot of example RT/Via retweet

3.1.4 RT/Via Non-follower Retweets

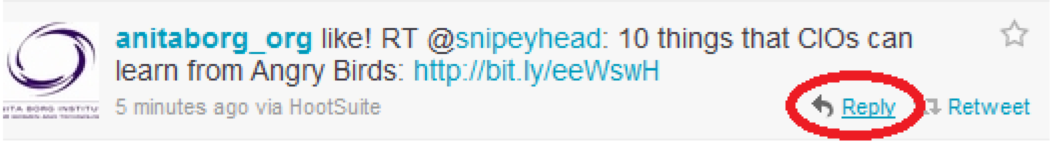
RT/Via non-follower retweets are defined as RT/via retweets made by non-followers.

For example, User B looks up search results for a hashtag on Twitter, and sees a tweet by User A, then decides to make an *RT/via retweet* of User A. In this instance, User B *does not follow* User A. Therefore, the connection between User B and User A is a **non-follower** relationship, and User B's retweet is classified as an **RT/via non-follower retweet**.

3.1.5 Replies

Replies are defined as tweets which begin with a mention to another user. This pilot study focused specifically on tweet replies where one user sends a URL to another user directly.

From a user's point of view, the user sends a tweet addressed to another specific user by clicking on the 'Reply' link, as shown in Figure 3.5.


anitaborg_org like! RT @snipeyhead: 10 things that CIOs can learn from Angry Birds: http://bit.ly/eeWswH

5 minutes ago via HootSuite

Figure 3.5: Screenshot of user's view of 'Reply' link

The difference between a *tweet reply* and a *direct message* is that replies can be viewed publicly by followers of both users involved and also the general public browsing through either user's Twitter profile page. This is unlike direct Twitter messages that can only be seen by the two users involved in the correspondence.

The main reason why replies are considered interesting, and therefore included in this typology, is that this preliminary toolkit found several instances where URLs were being propagated via replies. A subset of these replies included URLs which had been seen before. For example, User A saw a URL propagated by another user, and then User A sends that same URL to his/her followers via replies.

Myers et al. (2012) claimed that when "one user follows another, he/she can see all of their tweets, include URLs that they post, and it is through this relationship that contagions spread on Twitter." This claim that a follower can see all of somebody else's tweet is true only if that follower is looking up the timeline displayed on that user's profile. In the case of replies, they can only be seen by mutual followers — if User A replies to User B, this reply can be seen by User C in his home timeline only if he/she is following both Users A and B. If User C only follows one user and not the other, then this reply would not be visible in User C's home timeline, but if User C looks up User A's profile timeline, which is presumed to be public[2], then the reply would be visible to User C. Due to this unique visibility factor, replies are classified separately from native and RT/Via retweets, which are broadcast to all followers rather than just a subset.

This research work is aimed at investigating all the possible ways a tweet could spread, thus replies have been included in this study's observations.

3.1.6 Non-follower Replies

Non-follower replies are defined as replies, i.e. tweets beginning with a user mention, but the user mentioned in the tweet text does not follow the tweet's author.

For example, User B makes a reply to User A, but User A *does not follow* User B. Therefore, the connection between User A and User B is a **non-follower** relationship, and User B's reply is classified as an **non-follower reply**.

[2] if User A's profile is set to protected, and User C does not follow User A, then User C cannot see User A's tweets at all.

3.1.7 Other Implicit Retweets

Other implicit retweets are defined as retweets that do not conform to any of the classifications described above — specifically, they are not original tweets, native retweets, RT/via retweets nor replies.

For example, User B sees A's tweet, and decides to copy the entire tweet text but does not include any retweet markers such as "RT" or "via" and does not acknowledge that the originating tweet came from User A. However, User B *follows* User A, so it is possible to deduce that User B made the tweet after seeing it originate from User A. This tweet is classified as an **other implicit retweet** but only if a *follower* relationship can be detected by User B towards User A.

In another example, User D follows User C. User D saw a tweet by User C who had just signed a petition online and shared the link on his timeline. User D then signs the petition and shares the same link on his own timeline, but without mentioning User C and omitting any retweet markers in the text. Since User D has a follower relationship to User C, this tweet is also classified as an other implicit retweet.

3.2 Explicit/Implicit and Follower/Non-follower Retweets

Out of all seven retweet types defined here, the first four categories were defined altogether as **explicit retweets**, where the retweets were made via known retweet mechanisms such as Twitter's native retweets, or by manually inserting retweet terms such as 'RT' and 'via' into tweet texts before being retweeted. Retweets which do not conform to these mechanisms were classified as **implicit retweets**. Implicit retweets include replies, and other implicit retweets: tweets that can be reasonably assumed to be retweets, because they copy key information that a given user received earlier in their feed.

Both the native and RT/via retweet types were then divided into **follower** and **non-follower** categories. These were found by identifying whether the person retweeting is a follower of the originating author being that is retweeted.

The matrix in Table 3.1 illustrates the groupings of explicit and implicit retweets, and which retweet type corresponds to which follower/non-follower classification.

	Explicit retweets	*Implicit retweets*
Follower retweets	Native RT/Via	Replies
Non-follower retweets	Native non-follower RT/Via non-follower	Non-follower replies Other implicits

Table 3.1: Matrix of explicit/implicit and follower/non-follower retweet types

3.3 Decision Model Based on Typology of Propagation Types

Zarella (2009) had proposed a retweet decision model which did not take into account tweet propagation paths involving non-followers (see Figure 2.3 on page 24). An alternative retweet decision model is proposed in this pilot study, building upon Zarella's model and incorporating the typology of propagation types as described in Section 3.1 (page 29).

This alternative decision model differs mainly in terms of a user's **attention** upon reading a tweet, the subsequent **action** taken by that user — whether to retweet or to make a different action — and the **mechanism** chosen by that user to execute the action. This is in addition to considering the originating tweet's provenance, namely whether a **follower/following path** exists between the user retweeting and the user being retweeted.

Figure 3.6 (page 36) illustrates the alternative retweet decision model based on the pilot study's preliminary typology of propagation types. This alternative retweet decision model is based on four characteristics, as described below:

Attention: Does the user notice the tweet?

Action: Does the user want to share/act on the tweet?

Mechanism: Does the user want to retweet or reply tweets to other users?

Follower/following path: Is the user doing the retweeting also a follower of the originating tweet's author?

The seven different retweet types are primarily made of **tweets that were intended to be shared from one user to another**, after it has already been seen by the sharer beforehand. These can be made either **via retweets or replies**, and they can be classified according to whether any **follower/following paths exist or not** between the users doing the retweeting or replying.

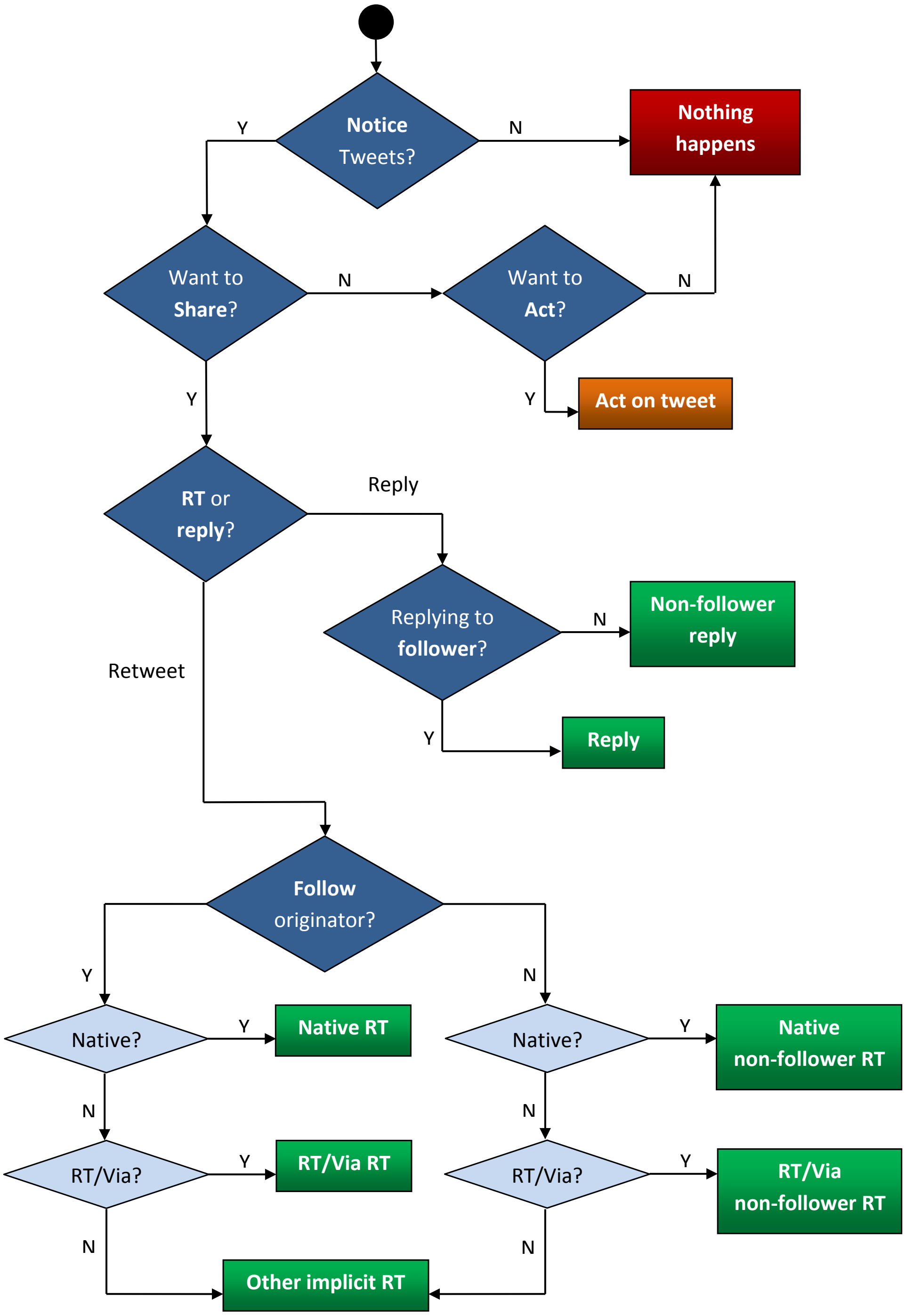


Figure 3.6: Alternative retweet decision model

3.4 Preliminary Toolkit

This section details the outline of the preliminary toolkit that was built to facilitate the pilot study which will be explained in Section 3.5.

This preliminary toolkit focused on searching for all tweets related to a specific URL. From a dataset of tweets containing some manually-selected URLs, this preliminary toolkit then outputs the different values for different quantitative metrics, such as total followers, visible follower/following networks and explicit versus implicit retweets. These values form the intermediate dataset which then gets used in further statistical analyses and data visualizations. This approach is similar to the one used by Wu et al. (2011), as their dataset was also restricted to tweets containing URLs only.

The methodology of Myers et al. (2012) is also similar to this thesis, where the toolkit collects all users mentioning the URLs gathered. Then, each user's follower/following network is iteratively requested from the Twitter API and then stored.

3.4.1 Architecture

A suite of scripts was created to facilitate the collection of tweets as described above. Scripts are used to allow easier reuse in terms of mixing and matching several scripts performing different analyses together. The scripts were modularized according to the different functions needed in this toolkit.

For this research work, several scripts were run sequentially to create the different outputs, depending on the metrics being focused on. The flow of this sequence was divided into three main functional components, namely data collection, intermediate data processing, and output generation.

Data collection: Search for tweets which contain the URL being searched by the user.

Intermediate data processing: Process the collected tweets to find retweets, then count their frequencies, record timestamps and identify existing follower/following networks.

Output generation: Use the processed intermediate data to make statistical calculations and draw pie charts or timeline charts.

Activity Diagram

The logic flow of this preliminary toolkit is best described by the activity diagram in Figure 3.7 (page 38). The diagram illustrates how the links are linked together to form the whole toolkit.

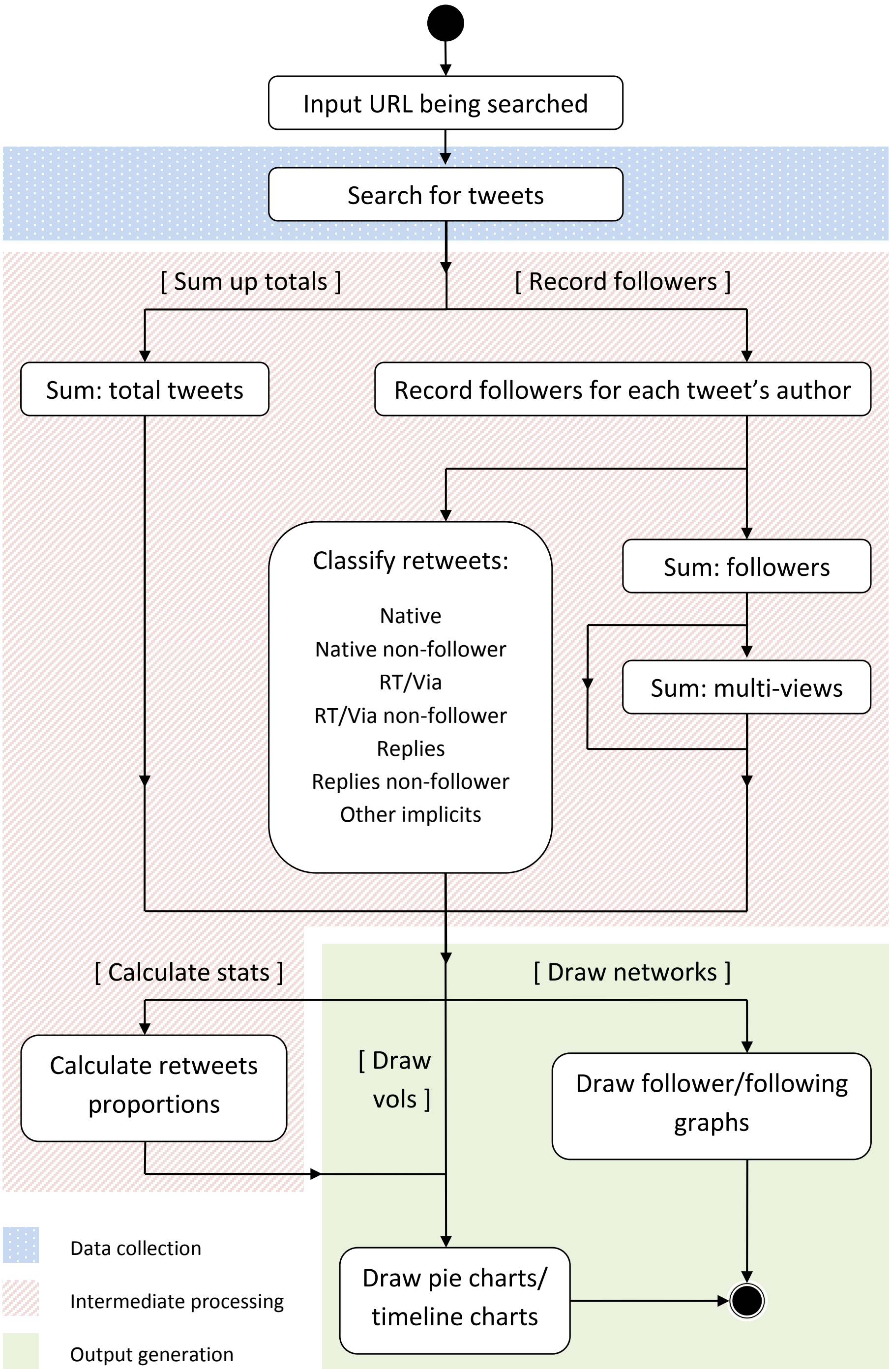


Figure 3.7: Preliminary toolkit's activity diagram

User View

From the point of view of the user — the person **running the suite of scripts** in this toolkit — the user firstly inputs a search query in the form of a URL. Given the input URL, the scripts will then do several types of counts:

URLs: Count the total occurrences of tweets containing the input URL.

Followers: Count how many followers could potentially have seen tweets containing the input URL.

Multiviews: Count how many followers could have potentially seen tweets containing the input URL more than once[3].

From the dataset of collected tweets, the user could also lookup each tweet's author and identify the followers for that author. This information could then be used for analyzing the follower/following network of the twitterers involved in the collected tweets dataset. For example, using the toolkit, the user could classify retweets according to different types:

Native retweets: Retweets using the Twitter framework's proprietary retweeting mechanism. Retweets using this mechanism are called native retweets in this pilot study.

Native non-follower retweets: Retweets which use the native retweet mechanism, but the person retweeting does not follow the author of or usernames mentioned in the originating tweet.

RT/Via retweets: Retweets which contain the usual proforma of retweets, such as the terms "RT" or "via".

RT/Via non-follower retweets: Retweets containing terms such as "RT" or "via" but the person retweeting does not follow the author of or usernames mentioned in the originating tweet.

Replies: Tweets beginning with a mention to another user, for example: '@user_name *check this URL out: http://www.xyz.com/*'.

Non-follower replies: Tweets beginning with a user mention but the users mentioned in the tweet text does not follow the tweet's author.

Other implicit retweets: Retweets which fall outside of the categories described above, i.e. they are not original tweets, native retweets, RT/via retweets nor replies. These

[3] If the retweet was made natively using Twitter's proprietary mechanism, then the client would hide multiple retweets and only display the retweet once.

include repeated tweets or tweets with no acknowledgements to any originating authors. Repeated tweets are identified by using an algorithm that checks whether one tweet has happened after somebody else's tweet, and whether a follower path exists between those two users (see Algorithm 4 on page 44).

After the collected tweets have been classified into groups, the preliminary toolkit would then be used to calculate the proportions of each retweet type. The preliminary toolkit can also produce different graphs using the CairoPlot[4] drawing package, such as pie charts, showing the proportions of all retweets, and a timeline chart to show the volume of retweets over time.

APIs and Python

This preliminary toolkit used the Twitter API as the main data source, via a collection of Python scripts.

In addition, several other APIs were used to collect the data needed and to process them into intermediary data, such as BackType[5], which was similar in function to Twitter Search, allowing tweets to be searched based on keywords/phrases/URLs, but had the added advantage of resolving shortened URLs. This means that if a user searches for a long URL, BackType will be able to determine which tweets are using shortened links that point to that exact long URL. This API was used primarily in the data collection component of the toolkit.

Since the pilot study was conducted, Twitter has acquired BackType and has now incorporated its URL resolving properties into its own Twitter Search API.

Data Backend

CSV files were used to store the data that was collected and processed throughout this preliminary toolkit. After each functional component (see Section 3.4.1 on page 37), different formats of CSV files were generated by the scripts. These CSV files could then be fed back into subsequent scripts for further processing.

This file format was chosen due to the flexibility of CSV files in naming and accessing column headings. In addition, it provides a lower programming overhead for processing the data backend as compared to a dedicated database such as SQL.

The main characteristic of this toolkit's data backend was that given the multiple scripts that this toolkit consists of, there were only **four** main CSV table formats that were required for these scripts to operate correctly:

[4] http://cairoplot.sourceforge.net/

[5] http://www.backtype.com/

Table Format A: Data collection (Twitter data)

Table Format B: Tweets and followers counts (Counts)

Table Format C: Followers records (Followers)

Table Format D: Retweet types (Retweets by author or by day)

This preliminary toolkit is formed of three functional components: data collection, intermediate data processing and output generation. All the generated CSV table formats corresponds to this preliminary toolkit's *data collection* and *intermediate data processing* components. These CSV table formats are defined by the column headings that are required in order for those CSV files to be processed by subsequent scripts.

For example, during the runtime of this preliminary toolkit, immediately after the first component of data collection, the output CSV file gets generated in a table format that is suitable for processing during the second component of intermediate data processing. The same applies to the output CSV files generated after the data processing component; the table formats generated after this second component complies to the operational requirements of scripts within the third component of output generation. This workflow is illustrated by Figure 3.9 (page 42).

Figure 3.8 outlines the relationships between all the CSV files generated by this toolkit's suite of scripts.

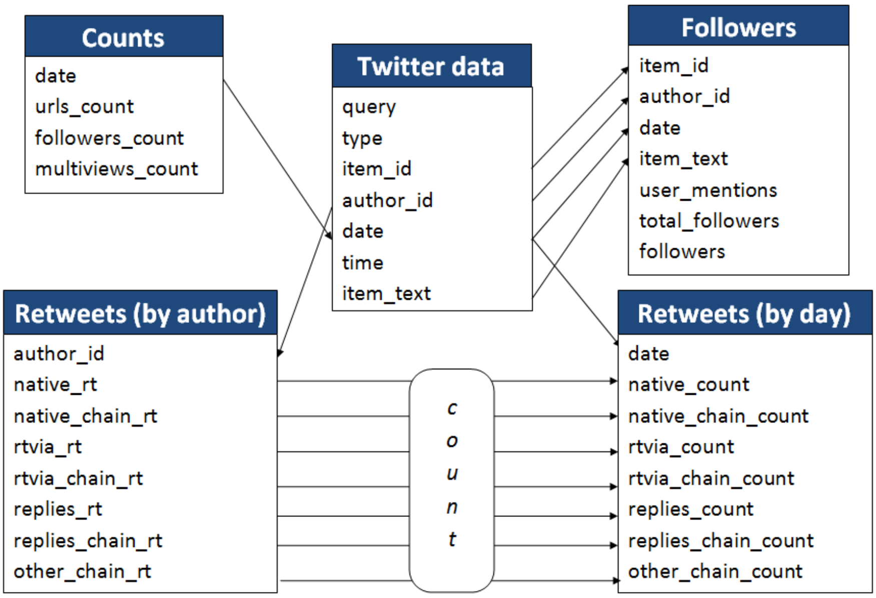


Figure 3.8: CSV files schema

The CSV table formats themselves are relatively flexible. The suite of scripts in this preliminary toolkit will operate correctly as long as the CSV files being used conform to these table formats, for example the CSV files must at least contain columns with the exact headings as illustrated below. Each row in table formats A and C stores data

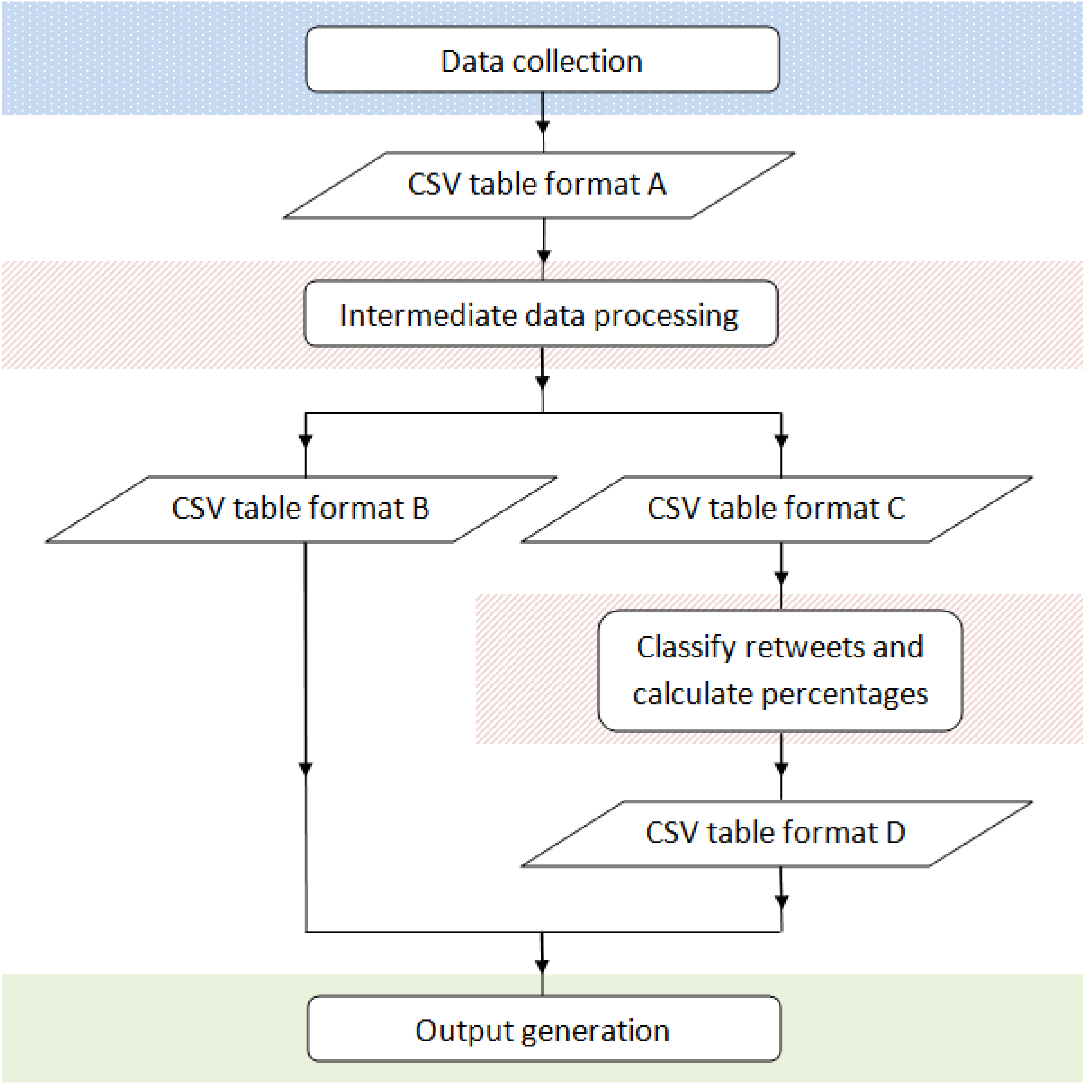


Figure 3.9: Workflow of CSV table formats

for individual tweets, while each row in table format B stores data per day, and finally data format D stores retweet types per author per row. Full explanations of each data column heading is provided in Appendix A (page 119). Particularly for table format D, Appendix A.4 (page 120) shows why *chain* was used instead of *non-follower* for the table headings, and the meanings of *n_rt*, *nc_rt* and the rest of Table 3.5 (page 43).

query	*type*	*item_id*	*author_id*	*date*	*time*	*item_text*
...	...	...	...	...	...	...

Table 3.2: Table Format A - Generated CSV file after the data collection component

date	*urls_count*	*followers_count*	*multiviews_count*
...	...	...	...

Table 3.3: Table Format B - Tweets and followers counts

item_id	*author_id*	*date*	*item_text*	*user_mentions*	*total_followers*	*followers*
...	...	...	...	...	...	...

Table 3.4: Table Format C - Followers records

author_id	*n_rt*	*nc_rt*	*r_rt*	*rc_rt*	*rp_rt*	*rpc_rt*	*oc_rt*
...	...	...	...	...	...	...	...

Table 3.5: Table Format D - Retweet types

Algorithms for Identifying Retweets

Most of the scripts involved in this preliminary toolkit performed simple RESTful[6] GET requests to either the BackType or Twitter APIs to collect data. The scripts either read in existing CSV files to perform analyses on them, or wrote the analyzed data into new CSV files.

Algorithms 1 to 4 (pages 43–44) were used in this preliminary toolkit to identify and classify retweets into the seven different retweet types.

```
$all\_natives \longleftarrow$ all native retweets recorded by Twitter API;
foreach $this\_native\_rt$ in $all\_natives$ do
    if $this\_native\_rt's$ author follows $current\_author$ then
        $all\_native\_rts \longleftarrow this\_native\_rt's\ author$;
    else
        $all\_native\_chain\_rts \longleftarrow this\_native\_rt's\ author$;
    end
    $parsed\_authors \longleftarrow$ parse all other authors mentioned in this tweet;
    foreach $this\_parsed\_author$ in $parsed\_authors$ do
        if $this\_parsed\_author$ follows $current\_author$ then
            $all\_native\_rts[current\_author] \longleftarrow this\_parsed\_author$
        else
            $all\_native\_chain\_rts[current\_author] \longleftarrow this\_parsed\_author$
        end
    end
end
```

Algorithm 1: Storing native retweets

[6] REpresentational State Transfer software architecture which is widely used for distributing information on the Web.

```
$all\_rtvias$ ⟵ all common retweet conventions;
foreach $this\_rt\_syntax$ in $all\_rtvias$ do
    if $this\_rt\_syntax$ in $tweet\_text$ then
        $parsed\_authors$ ⟵ parse other authors also mentioned in this tweet;
        foreach $this\_parsed\_author$ in $parsed\_authors$ do
            if $current\_author$ follows $this\_parsed\_author$ then
                $all\_rtvia\_rts[this\_parsed\_author]$ ⟵ $this\_parsed\_author$
            else
                $all\_rtvia\_chain\_rts[this\_parsed\_author]$ ⟵ $this\_parsed\_author$
            end
        end
    end
end
```

Algorithm 2: Storing RT/via retweets

```
if first char of $item\_text$ == '@' then
    $parsed\_authors$ ⟵ parse all authors mentioned in this tweet;
    foreach $this\_parsed\_author$ in $parsed\_authors$ do
        if $this\_parsed\_author$ follows $current\_author$ then
            $all\_replies\_rts[current\_author]$ ⟵ $this\_parsed\_author$
        else
            $all\_replies\_chain\_rts[current\_author]$ ⟵ $this\_parsed\_author$
        end
    end
end
```

Algorithm 3: Storing tweet replies

```
if $big\_followers\_set$ is empty then
    $seen\_set$ ⟵ add $current\_author$'s followers;
else
    $parsed\_authors$ ⟵ parse other authors also mentioned in this tweet;
    for $this\_parsed\_author$ in $parsed\_authors$ do
        $check1$ ⟵ $current\_author$ doesn't follow $this\_parsed\_author$;
        $check2$ ⟵ this retweet isn't a native or RT/via retweet;
        if $check1 = True$ and $check2 = True$ then
            $all\_other\_chain\_rts[this\_parsed\_author]$ ⟵ $this\_parsed\_author$
        end
    end
end
```

Algorithm 4: Storing 'other implicit retweets'

3.5 Setup of Pilot Study

This section details the setup of the pilot study that was run using the preliminary toolkit described in Section 3.4. This is followed by a description of the preliminary findings, particularly in the typology of propagation types, the proportions of retweet types found, the variability of those findings and the role of implicit retweets.

These findings are then discussed further, particularly on the role of replies, non-follower retweets and 'other implicit retweets' as mediums for propagation.

Domains of URLs

In the work by Nagarajan et al. (2010), they investigated whether different content domains of URLs would exhibit different retweet networks, specifically dense retweet networks — which contain retweet and author attributions — and sparse retweet networks — which are missing retweet and author attributions. The following experiment uses the same approach of grouping URLs according to content domains, namely the content types of the webpages pointed to by these URLs. Based on these content domains, the proportions of all seven retweet types are recorded in order to observe any variations in patterns. In this pilot study, tweets which were collected contained the URLs from **four** domains, namely *online petitions*, *charity fundraisers*, *news portals*, and *YouTube videos*. These domains were arbitrarily chosen to represent some of the types of URLs being propagated across Twitter.

Online petitions: Tweets containing petition URLs usually contain calls of action, persuading readers to visit the URL and sign a petition to show support for a cause.

Charity fundraisers: Tweets containing these URLs generally persuade readers to support a charity by visiting a fundraising webpage and donating money online.

News portals: Tweets pointing to news webpages highlight stories which tweet authors want to comment on, or just to attract more attention to.

YouTube videos: Tweets containing YouTube URLs show videos which are of interest to the tweet authors, and may be of interest to his/her followers.

Data Collection and Processing

Five different URLs were chosen for each domain, giving a total of **20 URLs overall**. These URLs were also arbitrarily chosen to represent the different websites available within each domain.

The preliminary toolkit was used to **collect tweets** containing those URLs, and also **record the follower/following networks** of all the Twitter users involved within the collected dataset of tweets. This resulted in a dataset of over 11,000 tweets.

Using the above data, the toolkit then **identified** all retweets and replies, and **classified** them according to the seven types as explained in Section 3.1 (page 29).

In order to observe retweet proportions' cumulative rates of increase, the toolkit **generated line graphs** to show their growth patterns over a given time window. **Pie charts were generated** to show the proportions of all the replies and retweet types found in the overall dataset.

3.6 Results

As outlined in Section 3.5 (page 44), this pilot study involved running the toolkit on a set of URLs grouped according to different domains. The following subsections describe the findings that were found from this experiment.

3.6.1 Proportions of Retweet Types

The line charts in Figures 3.10–3.13 (pages 47–48) show the growth of each retweet type, according to the four domains, while the pie charts in Figures 3.14(a)–3.14(d) (page 50) show the proportions of each retweet type, according to the four domains.

Timelines of Cumulative Growth

Across all four domains, the category 'other implicit retweets' seems to form the largest cumulative growth as compared to all the other retweet types.

Fundraisers: Illustrated by Figure 3.10 (page 47). Cumulative totals for 'other implicit retweets' overtake all other types from Day 6 onwards, forming an exponential growth pattern and reaching around 4000 tweets over 21 days. In contrast, all the other retweet types only accumulate up to 400 tweets or less in the same time period.

News: Illustrated by Figure 3.11 (page 47). Cumulative totals for 'other implicit retweets' overtake all other types from Day 1 onwards, forming an exponential growth pattern.

Petitions: Illustrated by Figure 3.12 (page 48). Cumulative totals for 'other implicit retweets' overtake all other types from Day 2 onwards. Growth looks linear between Days 1–7 before stagnating. Interestingly, the native and native non-follower retweets form a consistent pattern, whereby they both form similar linear growth patterns, accumulating between 50-100 tweets in total.

YouTube: Illustrated by Figure 3.13 (page 48). Cumulative totals for 'other implicit retweets' overtake all other types from Day 1 onwards, forming a decelerating growth pattern. All other retweet types accumulate up to 130 tweets or less.

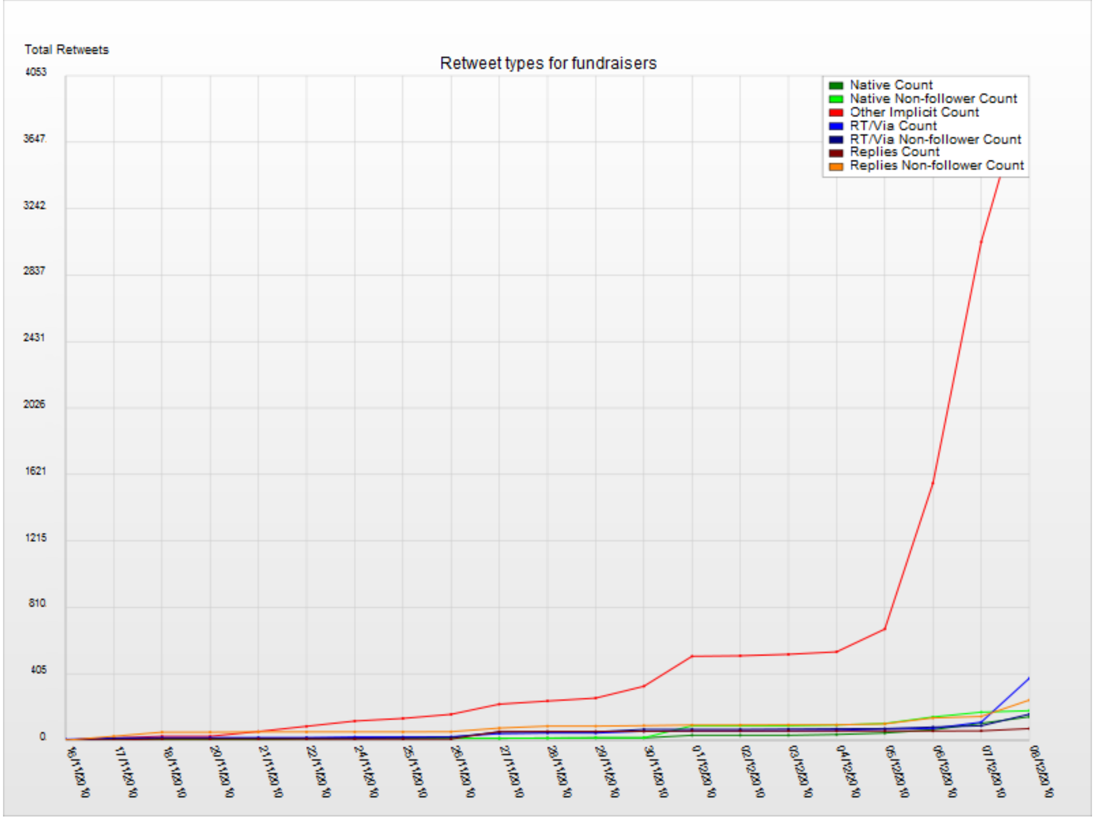


Figure 3.10: Line chart of retweet types for fundraiser URLs over time (days)

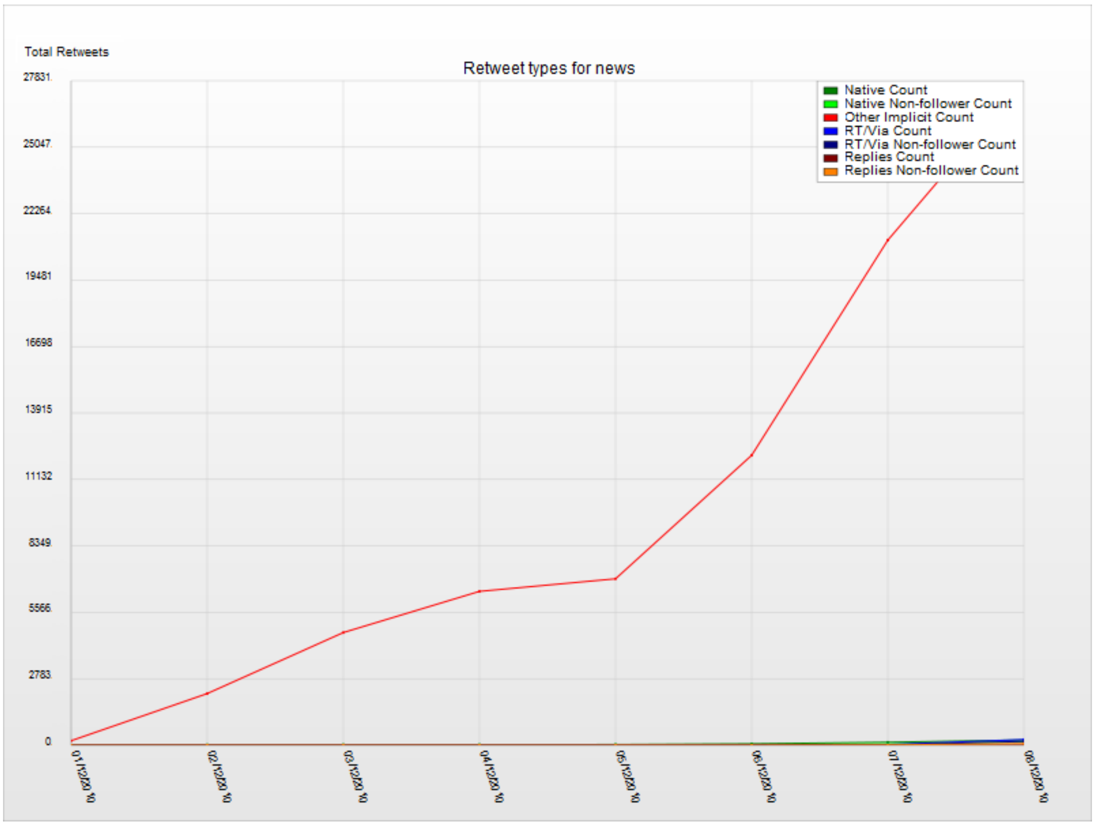


Figure 3.11: Line chart of retweet types for news URLs over time (days)

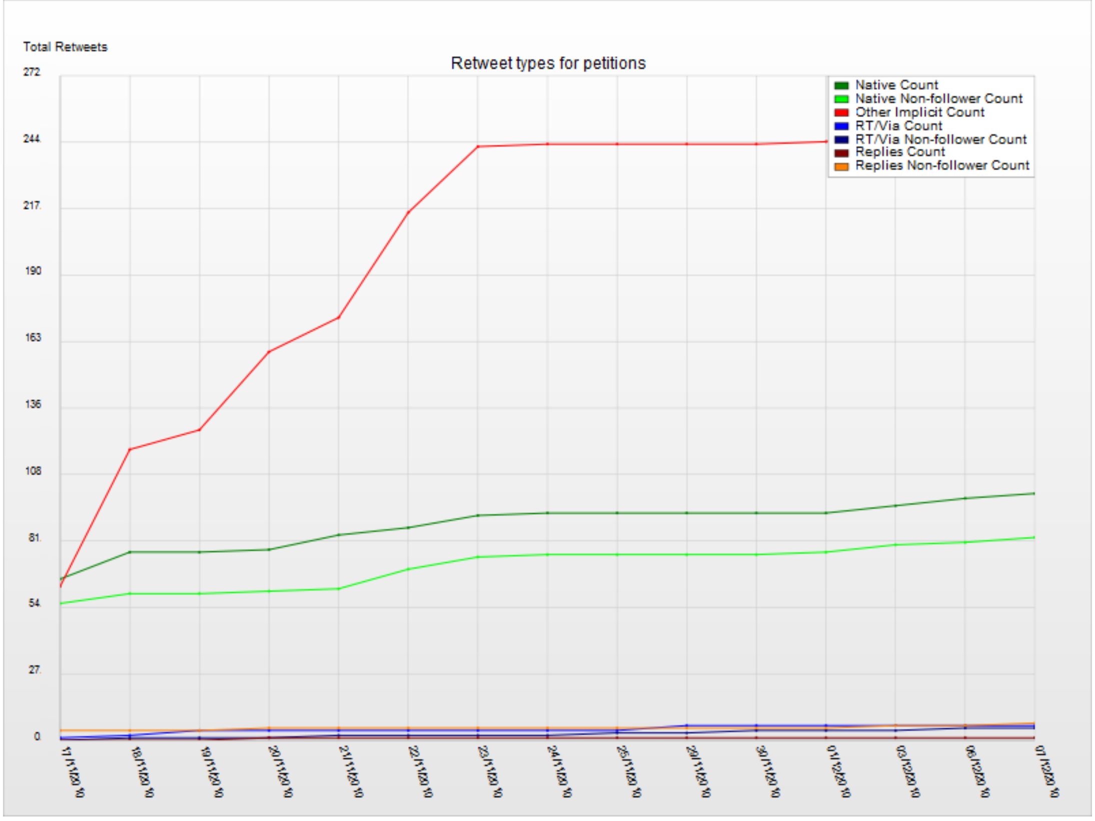


Figure 3.12: Line chart of retweet types for petition URLs over time (days)

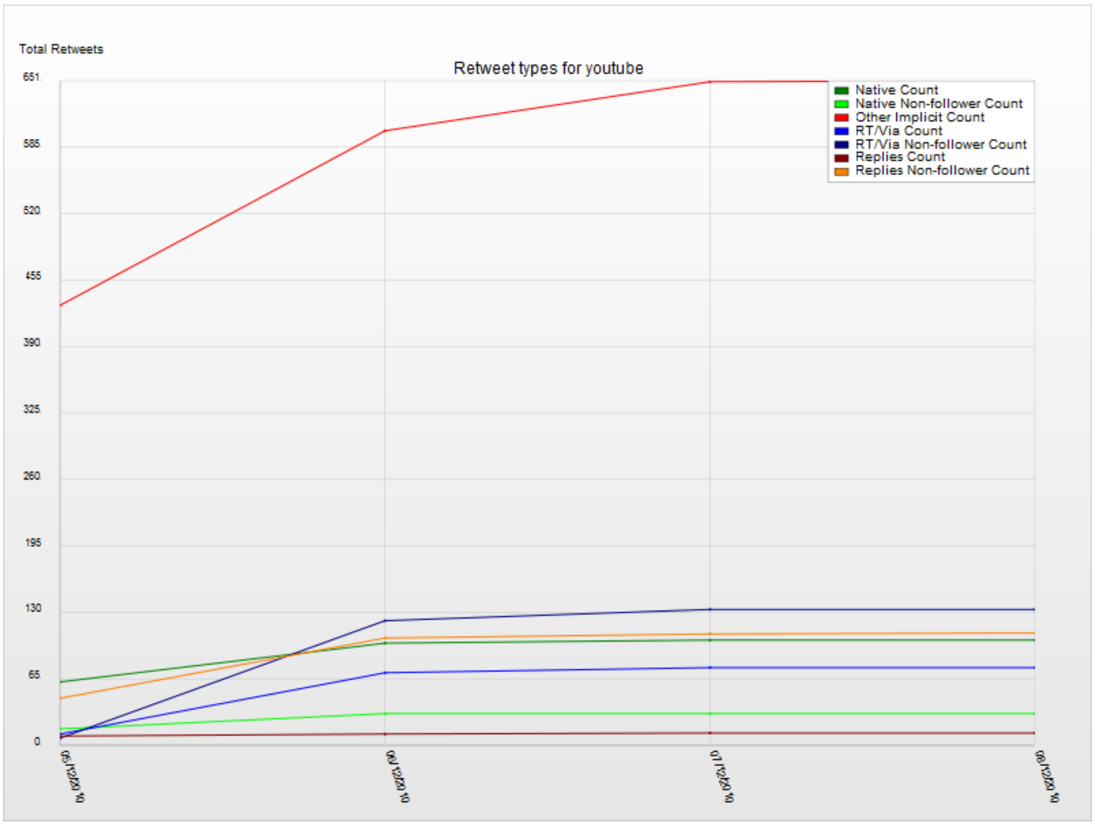


Figure 3.13: Line chart of retweet types for YouTube URLs over time (days)

Pie Charts of Retweet Type Proportions

Across all four domains, the proportions of 'other implicit retweets' seem to be consistently bigger than explicit retweets. Moreover, non-follower retweets also form bigger proportions as compared to follower retweets.

Table 3.6 shows the breakdown of explicit/implicit retweets and follower/non-follower retweets for each domain. The classification of all seven retweet types into the explicit/implicit and follower/non-follower groups has been discussed in Section 3.2 (page 34).

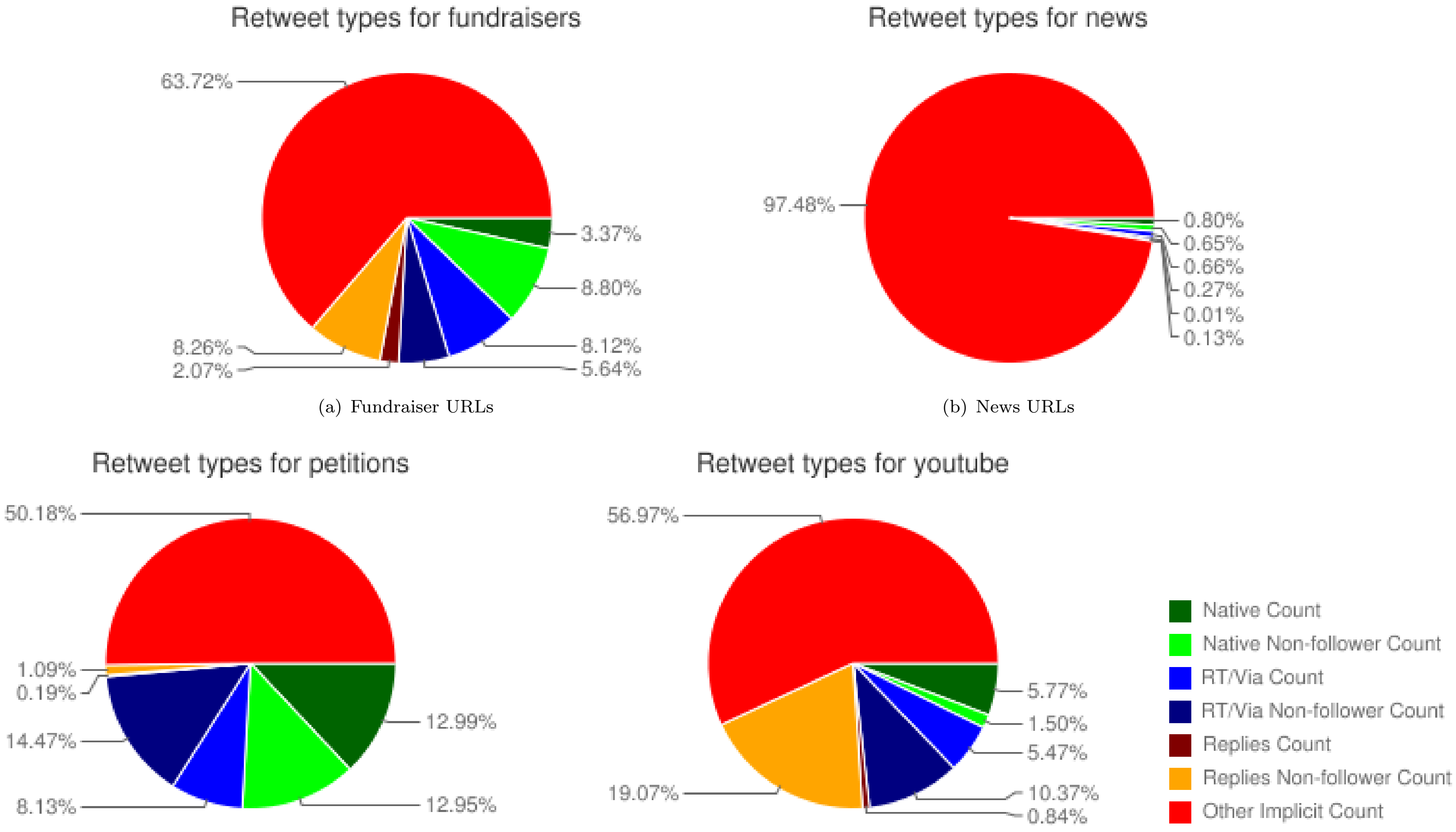
Domain	Retweets (%)			
	Explicit	*Implicit*	*Follower*	*Non-follower*
Fundraisers Fig 3.14(a) (pg 50)	25.9	74.1	13.6	86.4
News Fig 3.14(b) (pg 50)	2.4	97.6	1.5	98.5
Petitions Fig 3.14(c) (pg 50)	48.5	51.5	21.3	78.7
YouTube Fig 3.14(d) (pg 50)	23.1	76.9	12.1	87.9
Overall	25.0	75.0	12.1	87.9

Table 3.6: Percentage of explicit/implicit and follower/non-follower retweets across all four domains

In the domain of news, an extremely high proportion of implicit retweets (97.6%) could be seen. This seems to suggest that explicit retweet mechanisms such as native and RT/via retweets do not seem to be a popular way of propagating news URLs across the sample of tweets that were found.

In contrast, tweets containing petition URLs have an even split of explicit and implicit retweet types. This may suggest that the type of URL being propagated could be a determining factor in how subsequent retweets are made, particularly which mechanism would be used.

Also particularly interesting is the spread of retweet type proportions in the domain of petitions. Figure 3.14(c) (page 50) illustrates how the breakdown of retweet types seem to be slightly more uniform (four retweet types spread between 8–15% each) as compared to the other three domains which seem to have more variable proportions.



(a) Fundraiser URLs

(b) News URLs

(c) Petition URLs

(d) YouTube URLs

Figure 3.14: Pie chart of retweet types for four domains of URLs

Summary of Findings on Proportions of Retweet Types

Across all four domains, two observations consistently emerge:

- There are more implicit retweets as opposed to explicit retweets (proportions of implicit retweets range from 51.5% to 97.6%)
- There are more non-follower retweets as opposed to follower retweets (proportions of non-follower retweets range from 78.7% to 98.5%)

The growth of 'other implicit retweets' across all four domains appear in exponential or linear patterns of growth.

Two extreme cases can be seen from the findings of this experiment. In the case of news URLs, nearly 90% of all retweets do not use explicit retweet mechanisms, nor via follower paths. This is opposite to petition URLs. Although 'other implicit retweets' and non-follower retweets still account for a majority of retweets found, the difference margins between them are smaller. Therefore, these cases seem to suggest that different retweeting mechanisms are chosen depending on the domain of the URLs being propagated.

All these findings were found using the toolkit that classifies retweets according to the typology of propagation types, as described in Section 3.1 (page 29).

3.6.2 Variability of Proportions Recorded

This pilot study found that cumulatively, the 'other implicit retweets' tend to consistently outnumber the other retweet types by a fairly large amount. However, a closer look into the averages and standard deviations of these proportions reveal a high level of variability in the values recorded, as detailed in Table 3.7 (page 52). This table uses several abbreviations, and their meanings are explained below:

N = **Native**
RV = **RT/Via**
R = **Replies**
F = **follower**
NF = **non-follower**

The results in Table 3.7 suggest a high volatility in the statistics observed.

RT types		Fundraisers (%)		News (%)		Petitions (%)		YouTube (%)	
		Avg	*S Dev*	*Avg*	*S Dev*	*Avg*	*S Dev*	*Avg*	*S Dev*
N	*F*	3.37	4.69	0.80	0.58	12.99	13.04	5.77	4.06
	NF	8.80	21.38	0.65	0.80	12.95	14.16	1.50	1.51
RV	*F*	8.12	13.34	0.66	1.05	8.13	25.05	5.47	5.05
	NF	5.64	6.23	0.27	0.68	14.47	33.67	10.37	10.37
R	*F*	2.07	5.68	0.01	0.02	0.19	0.69	0.84	0.65
	NF	8.26	14.47	0.13	0.28	1.09	2.02	19.07	18.05
Implicit		63.72	27.75	97.48	2.92	50.18	29.90	56.97	14.37

Table 3.7: Proportions of retweet types by domain: Averages and standard deviations

3.6.3 Other Implicit Retweets

As mentioned above, throughout all four domains, the proportion of ‘other implicit retweets’ was consistently bigger than any other retweet type observed by this toolkit.

This suggests that when looking at the propagation of retweets, looking only at explicit retweets such as native and RT/Via retweets may not offer a complete picture of a full propagation pattern. Based on the above findings, more than half of the URLs were found to be tweeted without using the above retweet mechanisms. Therefore, more work needs to be done to identify how these ‘dark retweets’ propagate.

When the dataset of retweets are analyzed manually, these ‘other implicit retweets’ seem to consist of either one of the following characteristics:

- Verbatim copies of other tweets
- Unknown retweet markers
- Non-Latin characters

These manual observations of the make-up of ‘other implicit retweets’ are useful for identifying the components of these retweets in order to refine the typology into smaller, more detailed tweet propagation types.

3.6.4 Replies, Non-follower Retweets and Other Implicit Retweets as a Medium for Propagation

Prior work on retweets concentrate on studying retweets which conform to a pre-defined retweet mechanism, such as native and RT/via retweets. There seems to be little work done on the role of replies, and implicit or non-follower retweets, particularly where

Twitter users get acknowledged in retweets made by other people who are not their followers.

From the findings in this pilot study, the prevalence of ‘other implicit retweets’ seem to suggest that a large proportion of people do not follow normal retweet mechanisms. This observation raises questions as to how tweets are normally perceived to be propagated; tweets do not seem to spread only via retweets, but they could also spread via implicit means such as verbatim copying or using non-conventional retweet markers in their tweets. All this suggests the possibility of a different way for tweets to propagate across the Twitter network.

3.7 Contribution of Pilot Study

The contribution of this pilot study is two-fold:

Typology of propagation types: The pilot study proposed an initial typology of propagation patterns, consisting of seven different categories. This typology provided a detailed breakdown of the different ways that a thread, idea, or URL spreads across a set of microblogs.

Initial analysis of retweet types: The preliminary toolkit was used to classify the retweet types found and record their overall proportions. Of all seven propagation types, implicit retweets, particularly ‘other implicit retweets’, seem to have the biggest proportion of retweet types. From the observations of retweet type proportions, on average implicit retweets accounted for more than 50% of all retweets across all the four domains studied.

From these contributions, the analysis of ‘other implicit retweets’ patterns is an area which could be further improved. There is no uniform pattern underpinning the spread of ‘other implicit retweets.’ More work is needed to break down the ‘other implicit retweets’ that were found into more identifiable patterns.

3.8 Modifications for Further Evaluation

More investigation was needed into why other ‘other implicit retweets’ seem to propagate more than other retweet types. In this pilot study, implicit retweets accounted for more than 50% of all retweets across the URL domains of fundraisers, news, petitions and YouTube videos. The next experiment looks at **investigating the composition of ‘other implicit retweets’** by breaking this category down into more granular categories.

The variations in the results from the experimental dataset must also be considered. From Section 3.6.2 (page 51), we've seen that for the statistics on the proportions of retweet types found, there was a high degree of volatility; some of the standard deviations were bigger than the averages. In the next phase of evaluation, **a larger dataset of URLs** would be used to see if the size increase leads to more consistent results. In the subsequent main experiment, the difference between the proportions of implicit retweets in the pilot study and dark retweets in the main experiment will be discussed further in the following chapter.

The next chapter presents a more complete typology that describes implicit tweets in much greater detail, allowing subsequent experiments to explore dark retweeting behaviour in more detail.

Chapter 4

Typology of Tweet Propagation Types

The previous chapter outlined the pilot study which was based on classifying tweets using a typology of seven types of propagation. Across a majority of the tweet classified, the largest proportions were recorded as ‘other implicit retweets’ which is not particularly descriptive. Therefore, the seven types were then expanded to include more granular classifications of tweets. This was done in order to break down the ‘other implicit retweets’ propagation type into smaller more descriptive groups.

The expansion of the typology was done by breaking down the tweet entity into several different characteristics: 1. whether it is proprietary, 2. the mechanism used, 3. whether it is created by followers or non-followers, 4. whether it mentions other users, 5. if it is explicitly propagating another tweet, 6. if it links to an original tweet, and 7. the audience that it is pushed to.

The combinations of these seven characteristics became the basis for more granular classifications. As a result, the initial typology of seven propagation types was expanded further — this is to eliminate cases where large proportions of tweets were classified into a non-descriptive propagation type such as ‘other implicit retweets.’

By using binary values of 0s (false) and 1s (true) for each of the seven characteristics, 1024 different combinations were produced, of which 49 were valid tweet combinations. These 49 combinations were then grouped into 22 tweet categories, each containing one or more valid tweet combinations. Out of these categories, 19 were considered as retweets, or tweet propagation types, including visible, dark and orphan retweets. Out of these 19, only 18 were detectable using the experimental toolkit described in this thesis.

This chapter describes all these categories and tweet propagation types, plus the processes involved in identifying them.

4.1 Description of Tweet Characteristics

In this study, tweets were deconstructed using seven tweet characteristics as briefly outlined above. These characteristics were then labelled as follows:

- Proprietary
- Propagation mechanism
- Explicit
- Follower or non-follower
- Links to original tweet
- Mentions other users
- Tweet pushed to: all or some followers

In the pilot study, the initial seven types of propagation were based on two characteristics only, namely propagation mechanism and follower/non-follower. The additional five characteristics assist in breaking down the typology into more granular types of propagation.

In the following descriptions of these seven characteristics, the abbreviations stated in parentheses are used in the matrix of tweet propagation to be shown in Table 4.1 (page 59).

Proprietary: The propagation of a tweet is considered proprietary (P) if it was published using methods that were built into the Twitter infrastructure. For example, a retweet is considered proprietary if it was made using Twitter's proprietary methods, either by *a*) clicking the retweet button on its official user interfaces (e. g. web page, mobile apps), or *b*) third party apps utilizing the Twitter API's proprietary retweeting method.

Propagation mechanism: Tweets can either be propagated as a push-to-all retweet ('Rt'), a push-to-some reply ('@'), or a push-to-one direct message (DM).

Explicit: This characteristic concerns whether a user explicitly intends to *propagate* a tweet. A tweet is considered to be explicitly propagated if 1) a proprietary retweet was made, or 2) a retweet marker such as 'RT' was used, or 3) the '@' reply marker was written explicitly in the tweet text[1]. For example, *"Done! RT @User_X Sign*

[1] Explicit retweets include proprietary retweets and manually marked retweets. Implicit retweets include those without any retweet markers. A reply would be considered as an explicit reply, and cannot exist as an implicit reply

this petition! http://bit.ly/SmgF" would be considered as an explicit retweet, while *"@User_Y Please sign this petition: http://bit.ly/SmgF"* would be considered as an explicit reply.

Follower or non-follower: A retweet/reply/DM can be made by either a follower (F) or a non-follower (nF). This relates to the relationship between the author of the originating tweet and the person propagating that tweet.

Links to original tweet: If a propagating tweet contains metadata that links to the originating tweet, then the originating tweet's unique ID is stored. The Twitter API automatically stores this metadata when its proprietary retweet or reply mechanism is used. However, there exists non-proprietary tweets which also return this metadata. This denotes that the original proprietary retweet/reply have been modified into a non-proprietary state.

Mentions other users: A mention exists in a tweet if its text contains other people's Twitter usernames in them, preceded with an at-sign '@'.

Tweet pushed to: all or some followers: This encapsulates the difference between the visibility of a retweet and a reply. Retweets are pushed onto the timelines of all the followers of the retweeter. This visibility changes for replies; replies are addressed to a specific Twitter user that is mentioned at the beginning of a tweet text. This reply is only pushed to the timelines of mutual followers of the reply creator and the person being addressed to. For example, if User A makes a reply to User B, then the reply will only appear on the timelines of those who follow both Users A and B. In theory, it is possible for anyone to see this reply by looking up User A's personal page on Twitter, which lists all the tweets made by User A. However, this requires extra effort from those who don't follow Users A nor B, hence it is assumed that there exists a state where a tweet is visible only to some people but not all.

4.2 Matrix of Tweet Propagation Types

A binary matrix were constructed to illustrate all valid combinations of the seven tweet characteristics. This process resulted in a 2^{10} matrix, containing 1024 rows – one binary digit for each characteristic, apart from Mechanism which uses three digits (100, 010 or 001). Each row was then manually evaluated to identify if it is possible for any single tweet to possess the combination of characteristics as recorded in that row. This left 49 valid rows after this evaluation was completed.

These 49 valid rows were then grouped into 22 tweet categories, with each containing one or more valid tweet combinations. The tweet categories were made mainly by grouping the rows according to the characteristics of Proprietary, Mechanism and Follower/Non-follower. For example, PRtF denotes proprietary (P) retweets (Rt) made by followers (F), while @nF denotes a non-proprietary reply (@) made by a non-follower (nF).

The breakdown of these categories are as follows:

- 3 original categories: tweets, mentions and replies
- 10 visible retweets: PRtF, PRtnF, RtP@F, RtP@nF, RtF, RtnF, P@RtF, P@RtnF, @RtF and @RtnF
- 6 dark retweets: Rtf (dark), Rtnf (dark), P@F (dark), P@nF (dark), @F (dark) and @nF (dark)
- 1 direct message: PDMF
- 2 orphan categories: Orphan Rt and Orphan @

Out of the 19 non-original categories, only 18 could be observed using the experimental toolkit due to PDMFs, or proprietary (P) direct messages (DM) made by followers (F). These involve direct messages which are unobservable empirically. They are included in the main typology but discounted from subsequent experiments in this thesis. All 22 tweet categories are shown in Table 4.1 on page 59.

This table shows the 49 valid combinations of those seven characteristics, all in binary form. The meanings for all the binary values of 1s and 0s in this table are described in more detail in Appendix C (page 127). The abbreviations used under the Categories column in Table 4.1 come from the seven tweet characteristics described in the previous sub-section.

In Table 4.1, there are rows which are shaded in grey, signifying these propagation types as dark retweets, which will be discussed in Section 4.4 (page 64). The rows which are displayed in italics — PDMF, Orphan Retweets and Orphan Replies — are tweet propagation types which are included in the typology, but for various reasons are not included as dark retweets. The limited visibility problem of PDMF tweets will be discussed in Section 4.3.1 (page 60), whilst the incomplete data related to Orphan Retweets and Orphan Replies will be discussed in Section 4.3.4 (page 63).

4.2.1 Invalid Groups of Propagating Tweets

As mentioned before, a 2^{10} matrix containing 1024 rows was derived from seven characteristics. The valid rows have been shown in Table 4.1, but there were 975 rows which were considered invalid, therefore excluded from consideration. Appendix D (page 129) shows the invalid rows in full detail. These invalid groups contain binary values for

Table 4.1: Matrix of tweet propagation

Categories	Proprietary	Mechanism			Explicit	F/nF	Link to original	Mentions other users	Push	
		Rt	@	DM					All	Some
Original Tweets	1	0	0	0	0	0	0	0	1	1
	0	0	0	0	0	0	0	0	1	1
Original Mentions	1	0	0	0	0	1	0	1	1	1
	1	0	0	0	0	0	0	1	1	1
	0	0	0	0	0	1	0	1	1	1
	0	0	0	0	0	0	0	1	1	1
Original Replies	1	0	1	0	1	1	1	1	1	1
	1	0	1	0	1	1	1	1	0	1
	1	0	1	0	1	0	1	1	1	1
	1	0	1	0	1	0	1	1	0	1
	0	0	1	0	1	1	0	1	1	1
	0	0	1	0	1	1	0	1	0	1
	0	0	1	0	1	0	0	1	1	1
	0	0	1	0	1	0	0	1	0	1
PRtF	1	1	0	0	1	1	1	1	1	1
PRtnF	1	1	0	0	1	0	1	1	1	1
RtP@F	0	1	0	0	1	1	1	1	1	1
RtP@nF	0	1	0	0	1	0	1	1	1	1
RtF	0	1	0	0	1	1	0	1	1	1
	0	1	0	0	1	1	0	0	1	1
RtnF	0	1	0	0	1	0	0	1	1	1
	0	1	0	0	1	0	0	0	1	1
P@RtF	1	1	1	0	1	1	1	1	1	1
	1	1	1	0	1	1	1	1	0	1
P@RtnF	1	1	1	0	1	0	1	1	1	1
	1	1	1	0	1	0	1	1	0	1
@RtF	0	1	1	0	1	1	0	1	1	1
	0	1	1	0	1	1	0	1	0	1
@RtnF	0	1	1	0	1	0	0	1	1	1
	0	1	1	0	1	0	0	1	0	1
Rtf (dark)	0	1	0	0	0	1	0	1	1	1
	0	1	0	0	0	1	0	0	1	1
Rtnf (dark)	0	1	0	0	0	0	0	1	1	1
	0	1	0	0	0	0	0	0	1	1
P@F (dark)	1	0	1	0	1	1	1	1	1	1
	1	0	1	0	1	1	1	1	0	1
P@nF (dark)	1	0	1	0	1	0	1	1	1	1
	1	0	1	0	1	0	1	1	0	1
@F (dark)	0	0	1	0	1	1	0	1	1	1
	0	0	1	0	1	1	0	1	0	1
@nF (dark)	0	0	1	0	1	0	0	1	1	1
	0	0	1	0	1	0	0	1	0	1
PDMF	*1*	*0*	*0*	*1*	*1*	*1*	*0*	*1*	*0*	*1*
	1	*0*	*0*	*1*	*1*	*1*	*0*	*0*	*0*	*1*
Orphan Rt (Ori Not Found)	*1*	*1*	*0*	*0*	*1*	*0*	*1*	*0*	*1*	*1*
Orphan @ (User Not Found)	*0*	*1*	*1*	*0*	*1*	*0*	*0*	*0*	*1*	*1*
	0	*1*	*1*	*0*	*1*	*0*	*0*	*0*	*0*	*1*
	0	*0*	*1*	*0*	*1*	*0*	*0*	*0*	*1*	*1*
	0	*0*	*1*	*0*	*1*	*0*	*0*	*0*	*0*	*1*

permutations which cannot exist within a dataset of valid tweets. The invalid groups that were found are:

- Replies without mentions
- Overlapping mechanisms
- No mechanism but explicitly propagating
- No mechanism but links to original tweet
- No mentions but is a follower (of unknown source)
- Original tweets pushed to all but not some
- Original tweets not pushed to anyone
- Non-proprietary direct messages
- Proprietary retweet without attribution

- Implicit proprietary retweet
- Implicit proprietary replies
- Implicit non-proprietary replies
- Implicit proprietary direct messages
- Non-follower direct messages
- No audience
- Proprietary retweet without a link to original tweet
- Non-proprietary retweet with a link to original tweet
- Proprietary retweet not seen by all
- Tweet seen by all but not some users
- Non-proprietary retweet seen by some users but not all
- Direct messages seen by all
- Direct messages with a link to original tweet

4.3 Typology of Tweet Propagation Types

A directed tree graph was created to map out an overview of all these seven characteristics, as shown in Figure 4.1 (page 61). The leaves of the tree, circled with thicker lines, denote the different tweet propagation types in this typology. The grey-shaded leaves are the retweet propagation types which do not use formal retweeting mechanisms, thus considered as "dark retweets", which will be explained further in Section 4.4 on page 64.

4.3.1 Original Tweets, Mentions, Replies and PDMF

Out of the 22 tweet categories shown in Table 4.1, the original tweets, mentions and replies group consists of tweets which: *a*) do not seem to have been made using any proprietary retweeting or replying mechanisms, *b*) do not seem to be explicitly propagating another tweet, and *c*) therefore do not have any links to originating tweets nor users.

This classification includes tweet texts with mentions in them (original mentions), and replies between two or more users which do not seem to be explicitly propagating another tweet (original replies). Considering that original tweets, mentions and replies do not

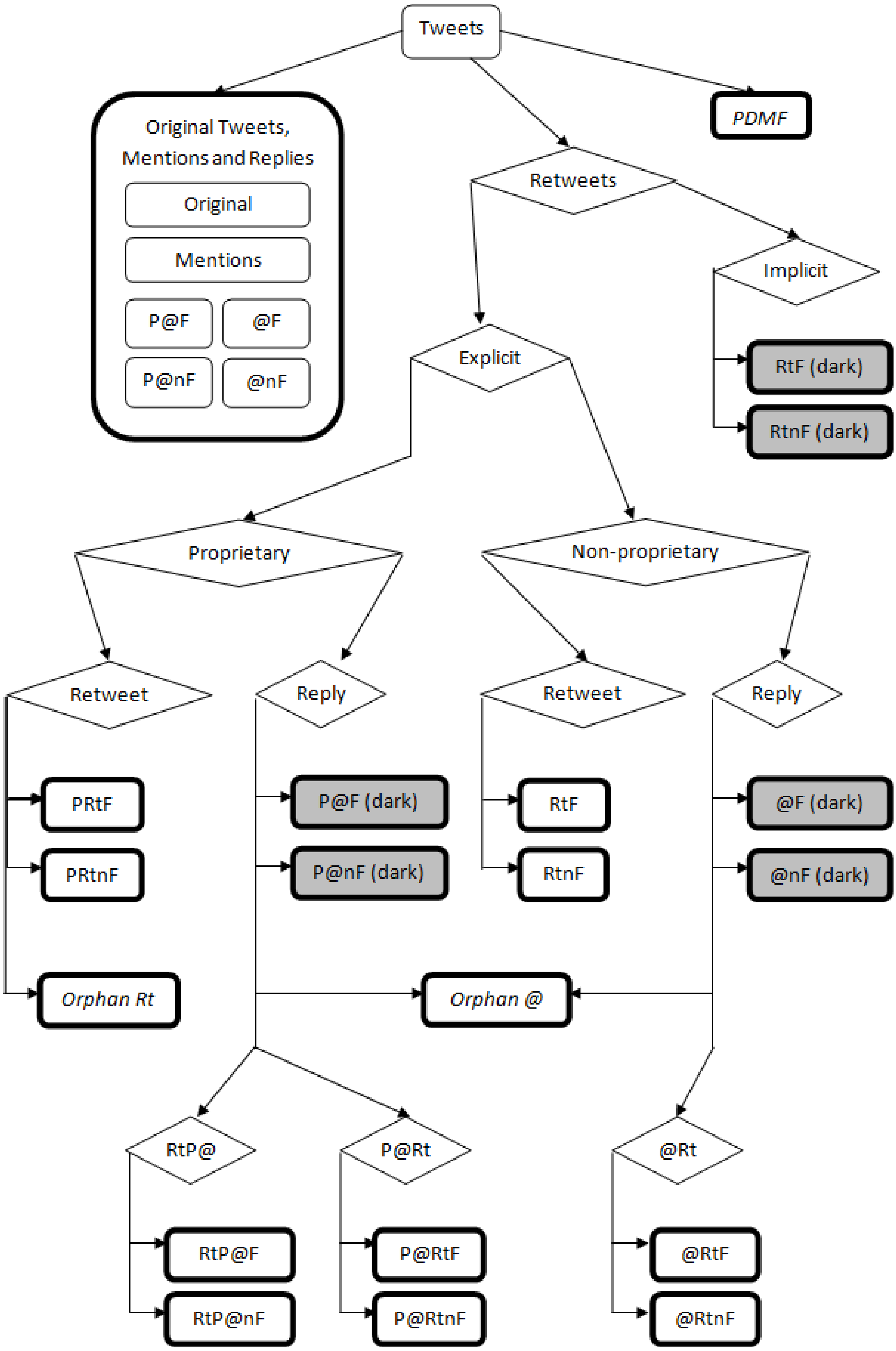


Figure 4.1: Typology of retweet types

serve to propagate any messages onwards, this leaves this typology with **19 different tweet propagation types** in total.

The PDMF category in Table 4.1 concerns direct messages (DM) which can only be accessed by the parties involved in private interactions. Due to the limited accessibility of this private nature of DMs, DM propagations could not be studied in more detail.

4.3.2 Explicitness and Links to the Originating Tweet

Tweets made using proprietary retweeting or replying methods are considered to cause two other characteristics to be true, namely Explicit (explicitly propagating another tweet), and Links to Original (containing metadata that links to the tweet that is being retweeted or replied to). Therefore these proprietary retweets and replies are marked in Table 4.1 with the value of 1 under the Explicit and Link to Original columns.

Non-proprietary tweets are also considered to be explicit but only if they include retweet and/or reply markers within their texts.

4.3.3 Multiple Mechanisms in Tweets

Several categories include two mechanisms, such as RtP@[2] and P@Rt[3]. Although the main common factor between these categories is the existence of multiple mechanisms when creating these tweets, there are distinct differences between these groups according to the order of the mechanisms used.

The Rt@ category was created specifically for retweets that were made using Twitter's proprietary replying mechanism. Manually typing in retweet markers in front of copied and pasted tweets has been the traditional way of creating retweets before Twitter's proprietary retweeting mechanism was created. Manual retweets allow users to modify the text of the tweet in order to add responses or other new content into the retweet. This modification ability does not exist within Twitter's proprietary retweeting mechanism, which propagates tweets in its original form. A completely manual retweet – where the user manually types in 'RT @User_B' and then copies User B's tweet – would not contain any metadata that links to another tweet, unlike all proprietary Twitter retweets or replies which do.

However, there exists certain retweets which are not marked by the Twitter REST API as being made using Twitter's proprietary retweet mechanism, but they still contain metadata linking to originating tweets. On further manual inspection, these tweets were found to be retweets that were manually created *after* the proprietary replying

[2] RtP@: Non-proprietary retweet made using a proprietary reply

[3] P@Rt: Proprietary reply made using a proprietary/non-proprietary retweet

mechanism was used. For example, User A would like to retweet some text written by User B, but instead of clicking on the 'Retweet' button, User A clicks the 'Reply' button next to User B's tweet. This action causes User A's input textbox for new tweets to be automatically filled with '@User_B', and this allows User A to copy and paste User B's tweet, prefix 'RT' or other retweet markers in front of the whole text, or modify the text slightly and prefix it with 'MT' (modified tweets). This retweeting style would not be classified by the Twitter API as a proprietary retweet, therefore in Table 4.1, the Rt@ categories contain 0 under the Proprietary column.

For @Rt categories, these tweets were intended to become replies, where the tweet texts begin with a mention to another Twitter user. However, the tweet texts also contain retweet markers such as 'RT' or 'via'. These @Rt categories are particularly interesting because the reach of these replies is not similar to a normal retweet, as discussed above. This difference in reach may have an implication to future retweet propagation studies.

4.3.4 Orphan Retweets and Replies

As seen in Table 4.1, Orphan Rt and Orphan @ categories exist due to certain missing tweet elements.

A retweet is considered as an Orphan Rt if the Twitter API labels it as a proprietary retweet, but the metadata related to the author of the originating tweet is missing. On further manual checks, it was found that this is because the tweet that is being retweeted no longer exists. Interestingly, the Twitter API response does not delete the metadata linking to the unique ID of the deleted tweet, but returns an empty response for the originating author's metadata instead. In Table 4.1, the Link to Original column is marked with 1 but Mentions Other Users is marked with 0.

Similarly, an orphan reply (Orphan @) exists when the person being replied to (the username prefixed at the start of the tweet text) no longer exists. When orphan replies are looked up via the Twitter API, the metadata for linking to originating tweets and also originating authors become unavailable. In Table 4.1, the Link to Original and Mentions Other Users are both marked with 0s.

Due to the unique characteristics of these orphan categories, they are grouped separately to all the other categories in Table 4.1.

An existing study on deleted tweets was done by Hazim Almuhimedi and Acquisti (2013) which presented various analyses ranging from total proportions (2.4% of all tweets in their dataset) to how fast tweets were being deleted (8.45 hours on average). In this thesis, the existence of deleted tweets is acknowledged by including orphan retweets and replies in the proposed typology of tweet propagation types.

4.4 Dark Retweets

A dark retweet is defined as *a tweet that is propagating another tweet, but does not use conventional retweeting mechanisms such as a) using Twitter's proprietary retweeting mechanism, or b) using common retweet markers such as "Rt" and "via" within the tweet text.*

4.4.1 Retweetability Confidence Factors

Differentiating between an original tweet and a dark retweet is non-trivial, as dark retweets do not have any of the conventional markers that would identify them as a retweet. In this thesis, there are two factors which influence the degree of confidence that a tweet is a retweet:

Factor #1: Confidence that Tweet A is propagating Tweet B. We assume with *strong confidence* that Tweet A is a retweet of Tweet B if Tweet A is queried via the Twitter API, and then the API returns some metadata related to the originating tweet: Tweet B. We assume with *moderate confidence* that Tweet A is a retweet of Tweet B if Tweet A is not a proprietary retweet, but retweet markers such as "RT" or "via" exist within the text of Tweet A. However, this latter assumption is still debatable – is this manually marked, non-proprietary retweet referring to an original tweet that does indeed exist, i.e. does Tweet B really exist?

Factor #2: Confidence that the originating author can be identified. We assume with *strong confidence* that Tweet A is a retweet of Tweet B if Tweet A is looked up via Twitter API and this results in metadata identifying the author of Tweet B. We assume with *moderate confidence* that Tweet A is a retweet of Tweet B if Tweet A is not a proprietary retweet, but Tweet A contains the username of the perceived author of Tweet B. However, this latter assumption is still debatable – is the manually mentioned username the correct author of the originating tweet?

For any given retweet, the less we have to assume about the validity of these two factors, then the more confidence we have that this tweet is a retweet.

In the following chapters of this thesis, the detection of retweets relies on evidence provided by:

- the metadata relating to a particular tweet, and
- the content of the tweet text.

There are several levels of difficulty in validating both factors. For Factor #1, if the originating tweet can be automatically detected by the Twitter API, then it is easy to validate this assumption.

However, this becomes harder when no metadata is available, such is the case with non-proprietary retweets. If the non-proprietary retweet contains a mention of another user — presumably the originating author — then this user's timeline can be processed to see if the originating tweet could be found.

The difficulty becomes hardest when there are no user mentions in the tweet text. Determining the provenance of the originating tweet would then require even more additional assumptions.

All of the above difficulty levels are also applicable when validating Factor #2 i.e determining the originating author.

In the case of copied tweets with no attributions or retweet markers, they may not be considered as retweets because there is no evidence that suggests the existence of an originating tweet (which relates to Factor #1), nor any identifying information of an originating author (which relates to Factor #2). However, there have been several studies which have documented the existence of tweets which propagate across Twitter without using retweet markers nor giving proper attribution to originating authors (Boyd et al., 2010; Wu et al., 2011; Nagarajan et al., 2010). In Table 4.1, propagating tweets which are not explicitly marked as retweets are considered as dark retweets.

Fake retweets may also occur, and in this research work, they can only be detected if an originating tweet or author could not be found. The Orphan Rts group in Table 4.1 already encompasses some of these fake retweets, but they can only detect proprietary retweets which have had their originating tweets deleted from the system. This research work does not look at fake retweets which were intentionally created. For example, User A may send a retweet of a particular message that was supposedly made by User B, but this not true because User B had not sent such a message. The detection of these fake retweets would require extra computational algorithms (natural language processing, pattern matching) and API requests (users timelines). Therefore, the detection of intentional fake retweets are beyond the scope of this research.

Replies are also considered within the context of dark retweets. A reply by itself denotes a one-way communication between two or more specific users. If the reply is not explicitly propagating another tweet, then the reply is classified as an original reply. However, if the reverse is true, where a reply is seen to be passing on content from a prior tweet, then this reply is considered as an implicit retweet. The advantage of using replies is that a user is able to send a tweet directly to a non-follower. However, replies would not be classified by the official Twitter API as a retweet, therefore making Factor #1 less concrete. In this study, in the context of tweet propagation, a reply would be categorized as a dark retweet if the tweet text itself does not contain any conventional retweet markers such as "RT" or "via" and so on.

In this thesis, tweets are considered as dark retweets when there is less confidence that a tweet may be a retweet, in comparison to other conventionally-made retweets. This confidence is derived from evidence such as a tweet's metadata and textual content. Referring to the matrix of propagation types in Table 4.1, the dark retweets include six tweet propagation types, namely RtF (dark), RtnF (dark), P@F (dark), P@nF (dark), @F (dark) and @nF (dark). This is because these groups require more assumptions to be made with respect to Factors #1 and/or #2 before they can be identified as retweets.

4.5 Terminology

For the remainder of this thesis, a retweet is defined as a propagating tweet which passes on a message from a prior tweet. However, there are different types of retweets which are referenced in this thesis. The following terms and phrases denote specific meanings:

A visible retweet relates to tweets which are passing on a message from a prior tweet using the retweet mechanism only. This may be done using Twitter's proprietary retweeting mechanism, or by inserting retweet markers such as "Rt" or "via" in the tweet text.

A dark retweet relates to tweets which are passing a message from a prior tweet but does not use conventional retweeting mechanisms such as Twitter's proprietary retweeting mechanism, or by inserting retweet markers such as "Rt" or "via" in the tweet text. These dark retweets may be done using other mechanisms such as replies, or none at all.

Table 4.2: Summary of classifications of visible and dark retweets

Visible Rts	Dark Rts
PRtF	RtF (dark)
PRtnF	RtnF (dark)
Rt@F	P@F (dark)
Rt@nF	P@nF (dark)
RtF	@F (dark)
RtnF	@nF (dark)
P@RtF	
P@RtnF	
@RtF	
@RtnF	

Table 4.2 summarizes the tweet propagation types which are considered as visible or dark retweets in this thesis.

In the next chapter, the process of evaluating this typology is presented, discussing the experimental toolkit that was developed, the technical components involved and the algorithms that were required.

Chapter 5

Experimental Toolkit

In order to investigate the 18 different tweet propagation types[1], an experimental toolkit was developed based on extensions and modifications of the previous toolkit used in the pilot study. Several functionalities were required from this toolkit, as follows:

- Data collection: retrieve and store volumes of tweets larger than the 11,000 tweets that was collected during the pilot study (see Chapter 3: Pilot Study on page 44).
- Classification: read and process all tweets, and classify them into original tweets and different tweet propagation types, according to several rules and characteristics. The classification is constrained to tweets which have unique static keywords in them, and in the instance of this thesis, the focus is on URLs. This classification may also work on tweet datasets containing other unique static keywords such as hashtags, although this specific analysis is currently outside the scope of this thesis.
- Data analysis: perform experiments targeted towards specific research areas – such as temporal analysis and reach – and generate output such as statistical or graphical representations of the data. This toolkit would also need to be flexible and extensible in order to accommodate any future analyses which have not been implemented for now.

Based on the findings of the pilot study, further modifications were done to the toolkit in order to accommodate a more robust typology, which was described in Chapter 4: Typology (page 55). This chapter describes the components involved in the revised main toolkit, along with all the algorithms created to facilitate the experiments in the post-pilot phase of this research work. Issues which arose during the development work

[1] There are 19 different tweet propagation types including PDMF: proprietary direct messages (DMs) made by followers. As these cannot be viewed publicly, this propagation type is omitted from all subsequent analyses.

are also discussed, together with the solutions that were chosen for implementation and the justifications behind them.

The Python scripts used in this toolkit are available for download at the following URL: https://github.com/coolster1/dark-rt-toolkit

5.1 Toolkit Architecture

Building on the pilot study, the main experiment also focused on tweets containing URLs in them, mainly to facilitate the detection of dark retweets. However, the difference between the preliminary toolkit and the main toolkit is the source for retrieving these URLs. Instead of manually-created seed URLs, in the main toolkit the Streaming API was used to collect random URLs. Then, all the tweets that contained each of these URLs were collected using the Twitter Search API. Using the search results, the tweets were then classified by querying the Twitter REST API for more details of each tweet. This includes the follower/friends[2] information for each particular Twitter user that gets seen within these collected tweets.

Figure 5.1 illustrates the components involved in the architecture of this toolkit.

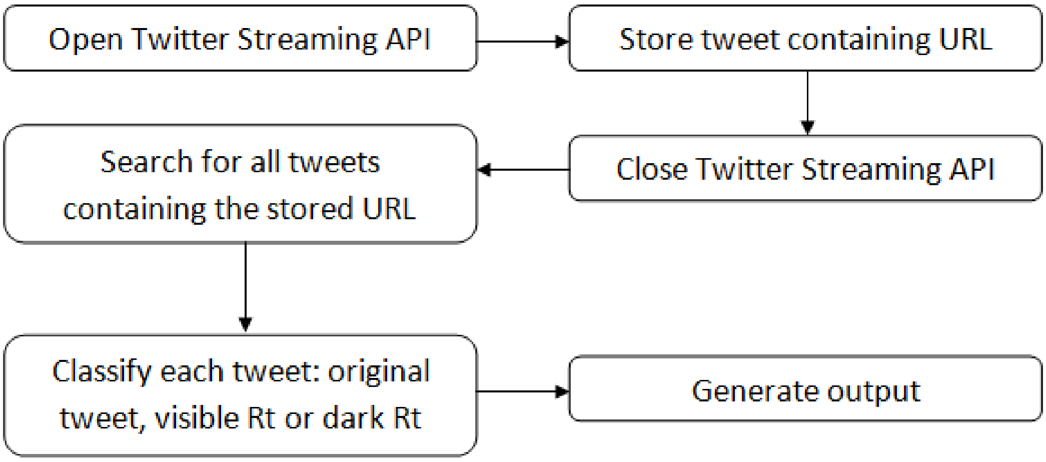

Open Twitter Streaming API
Store tweet containing URL
Close Twitter Streaming API
Search for all tweets containing the stored URL
Classify each tweet: original tweet, visible Rt or dark Rt
Generate output

Figure 5.1: Architecture of components in toolkit

5.2 Data Storage

Similar to the preliminary toolkit, all the retrieved data is stored in comma-separated files (CSVs). Manipulating CSV files were considered to be easier than using databases, considering that at the time the pilot was run (see previous Chapter 3: Pilot Study on page 29), the dataset contained only thousands of tweets. Therefore, the resulting

[2] The term "friend" is used by Twitter as the opposite of a follower — User A is a friend of User B if User B follows User A.

CSV files were not particularly large, thus manageable in terms of storage space and processing speed.

Different formats of CSV files were generated, corresponding to the type of function being performed by the toolkit. These CSV table formats were defined by their column headings, which are required to allow processing by subsequent components.The CSV table formats themselves are relatively flexible. The suite of scripts in this toolkit will operate correctly as long as the CSV files being used conforms to these table formats, for example the CSV files must at least contain columns with the exact headings as explained below.

This workflow illustrated by Figure 5.2 illustrates the flow of CSV table formats used in this main toolkit. Full explanations of each data column heading is provided in Appendix B (page 123).

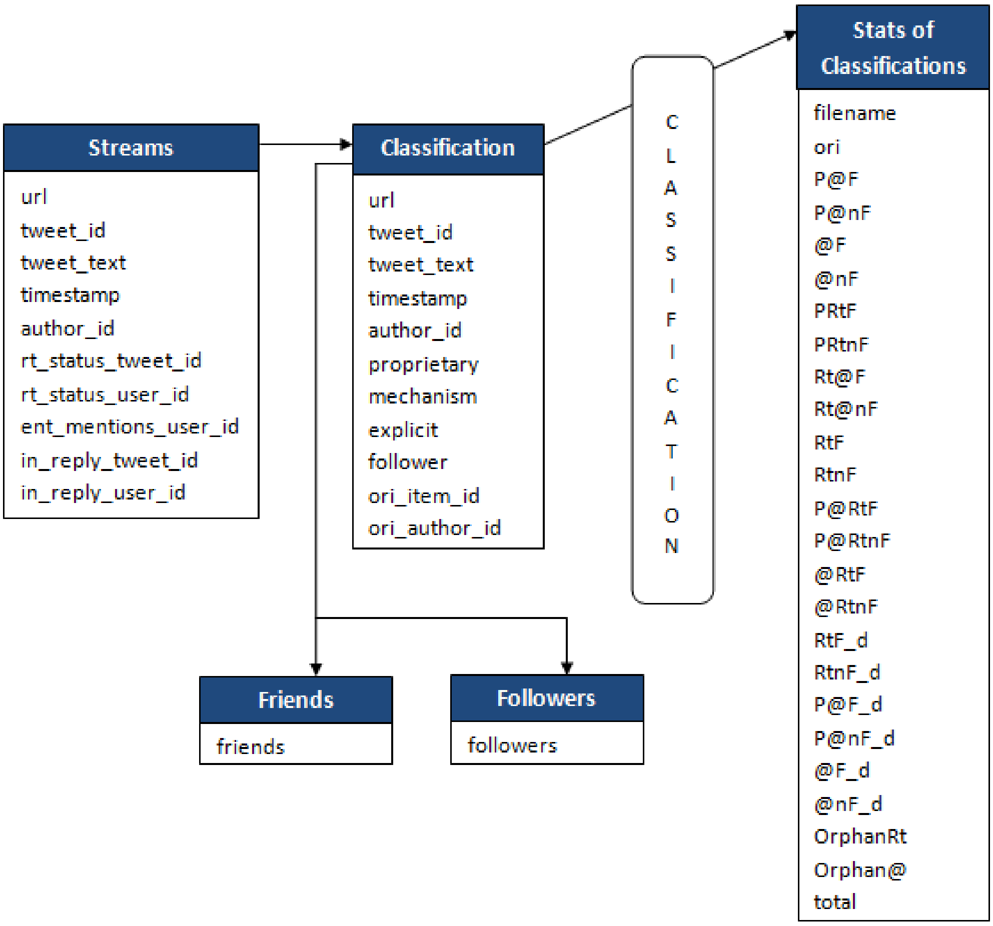


Figure 5.2: Workflow of CSV table formats for main toolkit

In the beginning of the research work, this method of data storage was adequate for the 11,000+ tweets that were being processed during the pilot project. However, as the research progressed to include substantially bigger datasets of hundreds of thousands of tweets, the increase in CSV filesizes resulted in increases in lookup and processing times as well. Each row in the CSV files had to be read and processed sequentially, thus resulting in longer processing times.

A database such as SQL could be used as a potential alternative to CSV files. The use of databases would allow direct access to relevant rows using primary keys. Moreover, the algorithms for accessing the relevant rows would be cleaner and more elegant when a query language such as SQL is used. The database could also be normalized in order to eliminate redundant information being copied across different files, as is the case in the CSV storage method being deployed currently.

Another suitable improvement to this toolkit setup is to utilize Hadoop[3], an open-source software framework for distributed data storage and processing. Hadoop is used by several large-scale Twitter-based projects, such as those run by Yahoo! Research and Tweetminster. This framework allows faster processing across millions of lines of text, as opposed to the limited processing power of a standard desktop computer.

However, due to legacy reasons, both databases and Hadoop were not used for this research work. This was because the existing toolkit had been optimized to read and write CSV files. It was thought that it would take more overhead time to familiarize and then migrate the existing data storage strategy from using CSV files to using a database and Hadoop. Adapting this toolkit to utilize databases and Hadoop would be a future work plan in this research.

5.3 Scripts and Algorithms

This main toolkit was created using the Python scripting language, similar to the preliminary toolkit. This is due to its easier learning curve and the good availability of client scripts online for the APIs that were used.

As described in Section 5.1, there are several components to this toolkit, from collecting tweets from the Streaming API, to classifying tweets into different types, to generating statistical and graphical representations of the results.

All the scripts used in this toolkit are available on Github via the following URL: `https://github.com/coolster1/dark-rt-toolkit`

The next subsections focus on the algorithms behind the scripts which perform the classification of tweets into different types according to the typology as described in Chapter 4: Typology (page 55).

[3] Hadoop is not primarily classified as a database like NoSQL, but contains a filesystem which is capable of handling large-scale parallel computation.

5.3.1 Classification Algorithms

The scripts work by looking for each of the seven characteristics identified in the previous Chapter 4. A tweet can then be classified by analyzing how it fits with all the characteristics taken together.

A tweet is considered to be explicitly propagating another tweet if it is a proprietary retweet or reply, or contains conventional retweet markers. Algorithm 5 describes how the explicit variable is set.

```
Data: Tweet ID & tweet text
if (Tweet ID is returned as a Proprietary Rt) or mechanism == @ then
    explicit = True;
else
    if Tweet text contains RT/via/HT/MT then
        explicit = True;
    else
        explicit = False
    end
end
```

Algorithm 5: Catching the Explicit characteristic

Algorithm 6 illustrates how proprietary retweets and replies are detected by the toolkit. All tweet IDs are queried via the REST API to see whether it is a proprietary retweet or reply. If this is true, then the tweet is considered proprietary.

```
Data: Tweet ID & tweet text
if Twitter API returns Tweet ID as a Proprietary Rt then
    proprietary = True;
else
    if (Text begins with @) & (Twitter API returns Tweet ID as a Proprietary Reply)
    then
        proprietary = True;
    else
        proprietary = False;
    end
end
```

Algorithm 6: Catching the Proprietary characteristic

Algorithm 7 detects if a follower/friends network exists between the author of the current tweet, and the originating author (in the form of Twitter API metadata), or all users mentioned in the tweet text (if originating author is unknown).

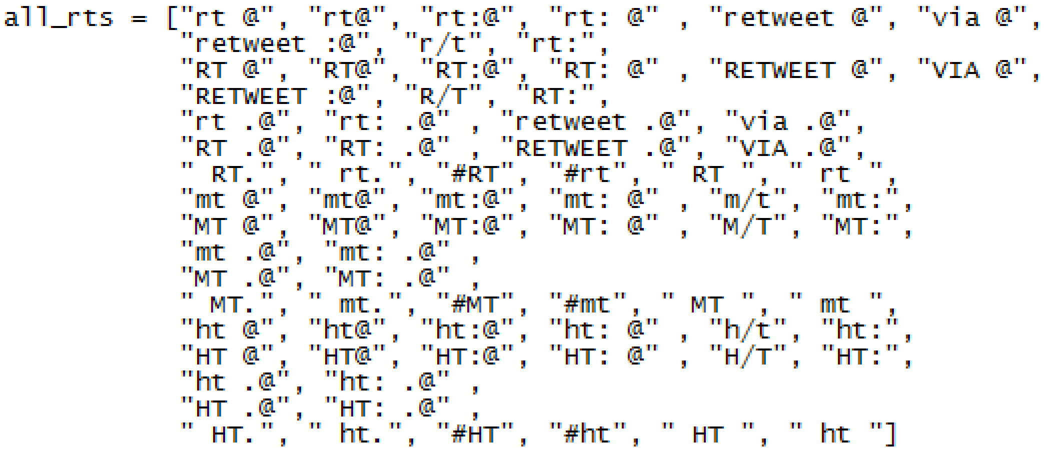
Algorithm 8 illustrates how to catch the tweet's propagating mechanism; retweet, reply, Direct Message or no mechanism (signifying an original tweet).

At first pass, the tweet ID is queried via the Twitter REST API to see if the tweet's a proprietary retweet, and thus setting the mechanism variable to 'Rt'.

```
Data: User IDs of author or all mentions in tweet text
if Author follows author or user mentioned then
    follower = True;
else
    follower = False;
end
```

Algorithm 7: Catching Follower/Non-follower characteristic

At the second pass, the tweet text is parsed to detect if any conventional Rt markers such as “RT”, “via”, “MT”, or “HT” can be detected within the text. Figure 5.3 shows the conventional Rt markers which were queried for. If any of these markers exist within the tweet text, then this also sets the mechanism variable to ‘Rt’.



```
all_rts = ["rt @", "rt@", "rt:@", "rt: @" , "retweet @", "via @",
           "retweet :@", "r/t", "rt:",
           "RT @", "RT@", "RT:@", "RT: @" , "RETWEET @", "VIA @",
           "RETWEET :@", "R/T", "RT:",
           "rt .@", "rt: .@" , "retweet .@", "via .@",
           "RT .@", "RT: .@" , "RETWEET .@", "VIA .@",
           " RT.", " rt.", "#RT", "#rt", " RT ", " rt ",
           "mt @", "mt@", "mt:@", "mt: @" , "m/t", "mt:",
           "MT @", "MT@", "MT:@", "MT: @" , "M/T", "MT:",
           "mt .@", "mt: .@" ,
           "MT .@", "MT: .@" ,
           " MT.", " mt.", "#MT", "#mt", " MT ", " mt ",
           "ht @", "ht@", "ht:@", "ht: @" , "h/t", "ht:",
           "HT @", "HT@", "HT:@", "HT: @" , "H/T", "HT:",
           "ht .@", "ht: .@" ,
           "HT .@", "HT: .@" ,
           " HT.", " ht.", "#HT", "#ht", " HT ", " ht "]
```

Figure 5.3: Conventional Rt markers detected by the toolkit

The third pass checks if the tweet text begins with a “@”, thus setting the mechanism variable to ‘@’ if this is true. Finally, if the mechanism variable has still not been set after these passes, then the mechanism variable remains in its default value which is ‘0’.

```
Data: Tweet ID & tweet text
if Twitter API returns Tweet ID as a Proprietary Rt then
    mechanism = Rt;
else
    if Tweet text contains RT/via/HT/MT then
        mechanism = Rt;
    else
        if Tweet text begins with @ then
            mechanism = @;
        else
            mechanism = 0
        end
    end
end
```

Algorithm 8: Catching the Mechanism characteristic

This toolkit also records the use of multiple mechanisms, such as replies and conventional retweets, as described in Section 4.3.3 on page 62. The algorithm recording these cases is illustrated by Algorithms 9 and 10.

For example, in Algorithm 9, if a tweet was detected as a reply but the mechanism had already been set as an 'Rt' prior to being detected as a reply, then the mechanism variable would be prefixed to '@Rt'. This category exists because certain replies contain conventional Rt markers in the tweet text. However the existence of a user mention at the beginning of the text ("@userA hello") means that it functionally serves as a reply and can only be seen by a subset of users, unlike a normal Rt.

```
Data: Tweet text
if Tweet text begins with @ then
    if mechanism is empty then
        mechanism = @;
    else
        if mechanism == Rt then
            mechanism = @Rt;
        end
    end
end
```

Algorithm 9: Catching Multiple Mechanisms I

Similarly, in Algorithm 10, if a tweet is marked as a non-proprietary retweet, but metadata returned by Twitter's REST API contains a reference to an originating tweet, then the mechanism variable is set to 'Rt@'. This tweet type was recreated using manual investigations via Twitter's proprietary web page. It was found that this particular instance of multiple mechanisms can happen when a user begins to create a non-proprietary Rt by clicking the proprietary 'Reply' button on Twitter, and then prefixing the username generated in the input textbox with an 'Rt' or other conventional retweet markers.

```
Data: proprietary & ori_tweet_id variables
if proprietary == False & ori_tweet_id exists then
    mechanism = Rt@;
end
```

Algorithm 10: Catching Multiple Mechanisms II

The ori_author_id variable relates to originating authors or mentioned users. This toolkit deals with user mentions in different ways, as illustrated by Algorithm 11. If the ori_tweet_id and ori_author_id variables have both been present in Twitter API metadata, then this means that the tweet is a proprietary RT or @. In this case, the list of mentioned users in the text is ignored.

However, if both of these variables are empty, then the list of users mentioned are all stored under the ori_author_id variable.

```
Data: ori_tweet_id & ori_author_id & tweet text
if ori_tweet_id & ori_author_id are both empty then
    if Tweet text contains any mentions then
        auth_id = List of all mentions;
    end
end
```

Algorithm 11: Catching the Mentions Other Users characteristic

5.3.2 Catching Dark Retweets

Repeated tweets are considered as an implicit and therefore dark retweet. In these instances, tweets which are presumed to have been seen previously by successive users are then propagated by copying them without attributing original authors, thus bypassing conventional retweeting mechanisms.

The subsequent algorithms are simplified versions of the actual code used to catch repeated tweets. The classification of tweets follows a two-step process, where the first step has been described above in Algorithms 5–11, and the second step involves overlaying the processed data with further information about repeated tweets.

The main function of this second step is to traverse the follower/friends network of the authors and friends of each author sequentially, and then snowballing the list of friends as the toolkit reads and passes each tweet in chronological order. This process is illustrated by Algorithm 12.

```
Data: auth_id
if auth_id has not been seen before then
    Request friends list from Twitter API;
    if auth_id account is not protected then
        Store auth_id's friends list;
    else
        No friends list to store;
    end
end
Write friends list to disk;
Add auth_id to list of seen authors;
```

Algorithm 12: Storing Current Author's Friends List

As an example, assume that we have a dataset of two chronologically ordered tweets — Tweet A, followed by Tweet B — written respectively by User X, who has 10 friends, and User Y, with 5 friends. After the first tweet has been read, User X and his/her list of friends will be stored into the list of authors seen and a big friends list, respectively. Therefore, the authors seen list will contain 1 entry, and the big friends list will have 10 entries.

Then, after the second tweet has been read, the same happens with User Y and his/her friends, so the authors seen list will now contain 2 entries, and the big friends list will have 15 entries, assuming that the two users' do not have any mutual friends.

This process is then run iteratively throughout the whole dataset, increasing the number of entries in the authors seen list and the big friends list after each tweet has been read.

In order to determine whether a tweet is an implicitly propagating tweet or not, then the toolkit checks the ori_auth_id variable for that tweet, which stores the user ID(s) of originating author(s), as shown in Algorithm 13.

```
Data: mechanism, auth_id
if mechanism does not start with Rt then
    if ori_auth_id contains more than 1 ID then
        forall the ori_auth_id do
            if looped_auth follows a previously seen author then
                if mechanism is empty then
                    mechanism = ImpRt
                else
                    Concatenate 'Imp' to existing mechanism;
                end
            else
                Leave mechanism unchanged;
            end
        end
    else
        if auth_id follows a previously seen author then
            if mechanism is empty then
                mechanism = ImpRt
            else
                Concatenate 'Imp' to existing mechanism;
            end
        end
    end
else
    Leave mechanism unchanged;
end
Write updates to disk;
```

Algorithm 13: Catching Repeated Tweets

First of all, the toolkit checks for **tweets which have not been marked as an 'Rt'**. Once established, the algorithm then parses ori_auth_id variables which contain more than one user ID in them. After that, the algorithm checks whether each user ID had been following a previous author that had been seen in previous passes. If this is true, and the mechanism variable is empty, then it is set to 'ImpRt', or the keyword 'Imp' is prefixed at the front of any existing mechanism variable. Therefore, an '@' mechanism will become 'Imp@' instead. The same checks are made for non-Rts which have only one user ID stored under the ori_auth_id variable. The keyword 'Imp' — signifying *implicit* — was chosen to denote retweets and replies which were not explicitly marked as retweets — hence making them dark retweets.

Once the mechanism variables have been checked and may or may not be modified, all these changes are then stored back onto disk.

The reason behind this two-step approach to classifying the tweets is because the detection of repeated tweets requires extra time and requests to Twitter's REST API, therefore it has been separated out of the preliminary classification step outlined in Algorithms 5–11.

5.3.3 Classifying Tweet Typology Types

After Algorithms 5–13 have been run, then there is another script which then groups the tweets according to the 18 different categories[4] in the typology of tweet propagation types (see Table 4.1 on page 59). This grouping is done based on the seven characteristics as described in Section 4.1 (page 56).

The algorithm for this classification is simply an iterative check for the binary values for all seven characteristics: Proprietary, Mechanism, Follower/Non-Follower (F/nF), Mentions Other Users, Explicit, Links to Original Tweet, and Tweet Pushed to: All or Some Users. Each tweet typology type would have a unique permutation of the binary values for all seven characteristics, as shown in the binary table in Section 4.2 (page 57). All tweets would be classified according to their binary value permutations.

5.3.4 Comparison of Toolkit Against Existing Twitter Analysis Tools

There are several Twitter analysis tools that are currently available both for free and for purchase online. Some tools would focus on tracking information such as the number of followers and retweets for a given Twitter username, such as Simply Measured[5] and Twitonomy[6], whilst tools such as Klout[7] claim to measure a user's influence by measuring various metrics such as retweets, mentions and followers. There are many other Twitter analysis tools available, with features ranging from analyzing the reach of tweets[8], to mapping followers geographically on a map[9], to deducing tweet sentiments[10].

[4] The main tweet propagation typology contains 19 different types, but due to the restricted visibility of PDMFs (see Section 4.3.1 on page 60), only 18 types could be observed empirically using this experimental toolkit.

[5] http://simplymeasured.com/

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

[7] http://klout.com/home

[8] TweetReach: http://tweetreach.com/

[9] TweepsMap: http://tweepsmap.com/

[10] Sentiment140: http://www.sentiment140.com/

An existing Twitter analysis tool which is most similar to the toolkit in this thesis is TweetCharts[11], which can also classify tweets into retweets, replies and mentions when given specific queries such as URLs, rather than just Twitter usernames.

Given the specific requirements of the research work in this thesis, the toolkit presented in this chapter was developed mainly to classify all the tweets containing specific URLs into different tweet propagation types, which include visible, dark and orphan retweets. The reach of tweets is also calculated according to the visibility of replies which is limited to mutual followers — this may defer from existing Twitter analysis tools which may overlook this. Therefore, the main difference between this thesis's toolkit and existing tools online is that the toolkit was designed specifically to evaluate the typology of tweet propagation types (see Figure 4.1 on page 61).

Future iterations of this toolkit may be developed towards measuring influence by incorporating both visible and dark retweets. However, this is currently beyond the scope of this thesis.

The next chapter presents the design and results of the main evaluation study. Based on the typology of tweet propagation types proposed in this thesis, the following chapter describes the first experiment of observing the tweet propagation types found in two datasets of tweets, namely a random sampling dataset and a URL drill-down dataset.

[11] `http://tweetcharts.com/`

Chapter 6

Proportions of Tweets Based on Typology of Propagation

Chapter 4 (page 55) presented a typology of tweet propagation types, consisting of 22 different groups. These groups were merged into two main detectable propagation types: visible and dark retweets.

Based on the matrix described in that chapter, a larger study was carried out to evaluate the proportions of tweets which fell within these different groups. Using the experimental toolkit described in Chapter 5 (page 69), the objective of this evaluation was to reveal the extent of visible and dark retweets. This follows up on the findings from the pilot study (see Chapter 3 on page 29), which established the existence of dark retweets.

Twitter's Streaming API was used as the main data source for this research work. Set on its Spritzer setting, this API provides a random sample containing 1% of current tweets being published globally in real time.

In this evaluation, two different sampling methods were used: **random sampling** and **URL drill-downs**. A random sample was used to observe the proportion of visible and dark retweets that appear publicly on Twitter. In contrast, URL drill-downs were used to observe the probability of a given URL being propagated via visible or dark retweets.

The main difference between the two sampling methods is in the *scope of tweets* which get classified into the typology. In the random sample, tweets are selected at random from the Twitter stream, meaning that the number of sampled tweets is close to the number of sampled URLs. For the URL drill-downs, all the tweets which contain a given URL gets classified by the same toolkit, meaning that the number of sampled tweets is much higher than the number of sampled URLs, as each URL is represented by many tweets.

The random sampling provides a random snapshot of the twittersphere but the execution is more expensive in terms of resources — many API calls were made to Twitter even though a small number of tweets were being classified. The final random sample collected for evaluation consists of 27,146 tweets from 25,273 unique URLs.

The URL drill-downs illustrate the retweet behaviours of each individual URL. In this sampling method, classification is done on all the tweets containing a particular URL. Depending on the search results returned by the Twitter API, these results range from 1 tweet to 18,000 tweets[1] per URL. Therefore, the URL drill-downs provide a larger corpus of 262,517 tweets, but they come from a considerably smaller pool of only 747 URLs.

The following sections describe in more detail the objectives, methodologies and findings of the main typology classification over random sampling and URL drill-downs.

6.1 Sample #1: Random Sampling

As mentioned above, random sampling was used to observe the proportions of visible and dark retweets occurring in a random sample of the whole Twitter stream. More specifically, given a random tweet, what is the probability that the tweet will be a visible or dark retweet?

Using the experimental toolkit, a random tweet containing a URL was retrieved from Twitter's Streaming API. This one tweet was then classified according to the tweet propagation typology, and then this process was repeated until the sample period had ended.

Figure 6.1 (page 83) illustrates the components involved in acquiring the random sample.

Although only one tweet gets classified, there are various subprocedures within the toolkit which requires multiple Twitter API calls. Therefore, although a relatively small amount of tweets were classified, the whole data collection still took a considerable long time due to the expensive resource-hungry execution of the toolkit. The random sample was collected between 28 Oct 2013 – 12 Jan 2014 (72 days), producing 27,146 tweets from 25,273 unique URLs.

[1] Twitter API's search function is rate-limited to return either 1) a maximum of 18,000 tweets per query, or 2) tweets from the last 7 days only, whichever comes first.

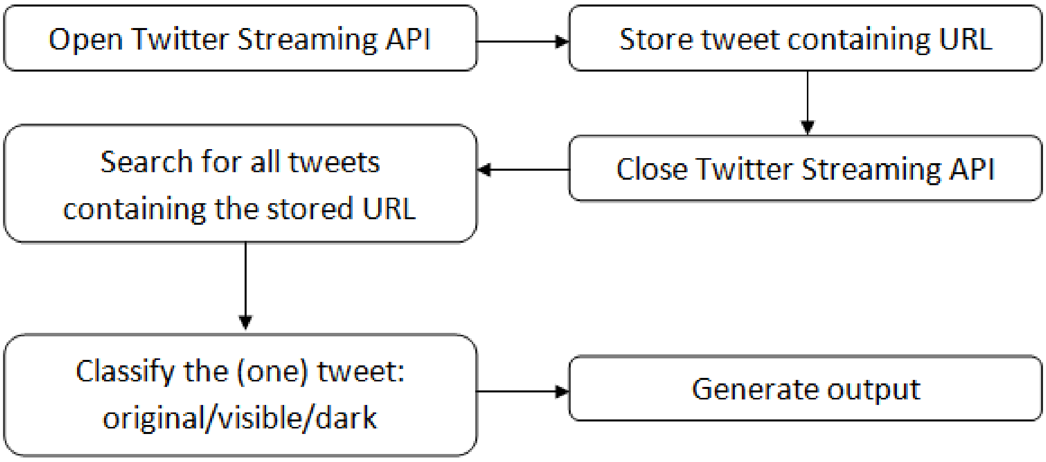

Open Twitter Streaming API
Store tweet containing URL
Search for all tweets containing the stored URL
Close Twitter Streaming API
Classify the (one) tweet: original/visible/dark
Generate output

Figure 6.1: Architecture of components in collecting random sampling dataset

6.1.1 Data Cleaning

The experimental toolkit used for this research work was optimized for retweet markers written solely in ASCII characters[2]. Therefore, if a tweet is written in Arabic, Japanese and so on, then it reduces the accuracy of this toolkit's classification. This would result in increased detections of false positives. In order to mitigate this effect, the above dataset of 27,146 tweets was produced after removing non-ASCII tweets from the dataset.

The ASCII data cleaning was done by using a Python module called *chardet*[3] which detects the character encoding for any given string. For each string, the chardet module returns the detected character set, such as ASCII and UTF-8 and the confidence level of its detection in the form of a percentage. For example, the string "Hello" would result in an ASCII encoding at a confidence level of 100%.

For this data cleaning procedure, a tweet is considered acceptable for this dataset when an ASCII encoding is detected at more than 80% level of confidence. Before the ASCII data cleaning was carried out, the dataset collected 57,962 tweets in total. Once the non-ASCII tweets were discarded, the dataset was left with 27,146 tweets.

For the rest of this research work, all datasets subsequently collected were also cleaned to remove all non-ASCII tweets.

6.1.2 Finding

Using the experimental toolkit, the cleaned dataset of ASCII-only random 27,146 tweets were classified according to the typology of tweet propagation types, consisting of 18

[2] American Standard Code for Information Interchange (ASCII) is a character-encoding scheme consisting of 128 alphanumerical Latin characters. These include a-z, A-Z, 0-9 and some basic punctuation symbols.

[3] *chardet* can be downloaded from https://pypi.python.org/pypi/chardet

detectable different ways a tweet could spread.

Figure 6.2 illustrates the proportions of original tweets versus retweets, and the proportion of visible versus dark retweets in this random sample.

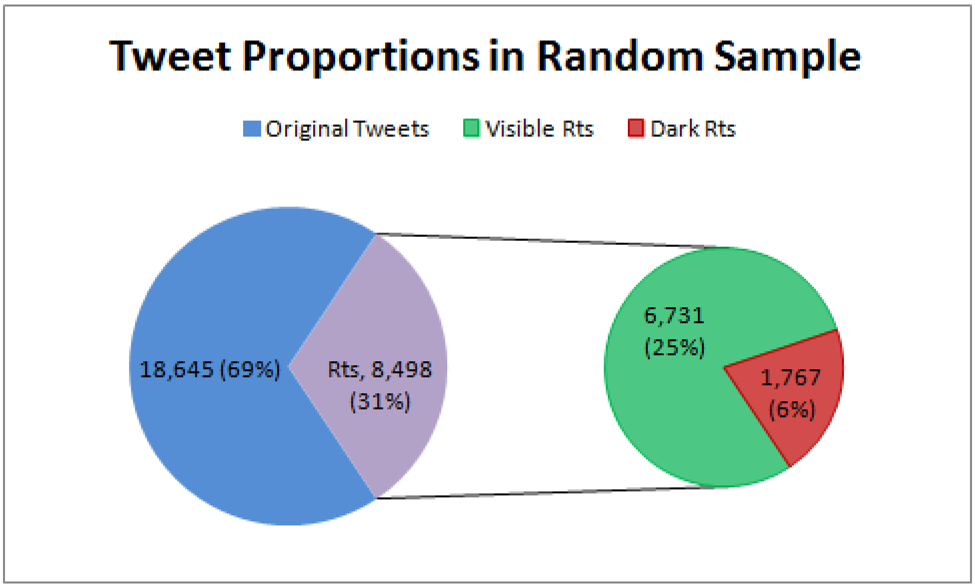


Figure 6.2: Total proportions of random sample

The proportions of tweets for each classification is displayed in Table 6.1 (page 85).

The results of this classification exercise show that 79.2% of total retweets were visible, whereas the remaining 20.8% of those retweets were dark, as shown in Table 6.2 (page 85). Due to the very small number of orphan tweets (only 3 tweets over a dataset of 27,146), this category was not included as part of retweet calculations. Out of this 20.8%, the category of dark RT/via retweet made by followers — RtF (dark) — contribute the biggest proportion of the dark retweets volume.

This rough 79/21 visible/dark retweets split is radically different to the findings in the pilot study in Chapter 3. In the pilot study, the overall visible/dark split was 25/75, where the majority of the retweets were classified as dark. This is most likely due to the choice of URLs during the pilot study; having a small sample of 20 URLs overall may have made it more susceptible to being skewed by URLs which return a disproportionate volume of retweets which are dark. Further on in this chapter, the distribution of dark retweets over URLs will be discussed in Section 6.2.1, whilst the existence of supervisible and superdark URLs will be discussed in Section 6.2.2.

In conclusion, for a given random retweet, there was roughly an 79% probability that it would be classified as a visible retweet, and a 21% chance that it would be a dark retweet.

Table 6.1: Total counts for original and propagating tweet types

Tweet types		Total tweets
ORIGINAL	Original	18,247
	P@F	0
	P@nF	178
	@F	156
	@nF	64
TOTAL ORIGINAL		**18,645**
VISIBLE	PRtF	3,560
	PRtnF	2,189
	Rt@F	0
	Rt@nF	22
	RtF	595
	RtnF	345
	P@RtF	0
	P@RtnF	4
	@RtF	12
	@RtnF	4
TOTAL VISIBLE		**6,731**
DARK	RtF_d	1,675
	RtnF_d	0
	P@F_d	0
	P@nF_d	29
	@F_d	39
	@nF_d	24
TOTAL DARK		**1,767**
ORPHAN	OrphanRt	0
	Orphan@	3
OVERALL TOTAL		**27,146**

Table 6.2: Proportion of visible and dark retweets

TOTAL VISIBLE	6,731	79.2%
TOTAL DARK	1,767	20.8%
TOTAL RTS	8,498	100.0%

6.2 Sample #2: URL Drill-down

URL drill-downs were carried out to observe the patterns of tweet propagation for each URL found. Specifically, given a random URL, what is the probability that the URL will be propagated using a visible or dark retweet?

The URL drill-downs provide a more focused view which allows us to see whether different URLs exhibit different propagation patterns. Similar to random sampling, Twitter's Streaming API was used to retrieve tweets containing URLs. However, in contrast to random sampling, instead of classifying only one tweet per URL, the URL drill-downs involve classifying all tweets which contain the specified URL. Therefore, the main difference between the random sampling and the URL drill-downs is the ratio of tweets being classified for each URL: approximately 1:1 for random sampling, but 1:many for URL drill-downs.

Figure 6.3 illustrates the components involved in acquiring the URL drill-down dataset.

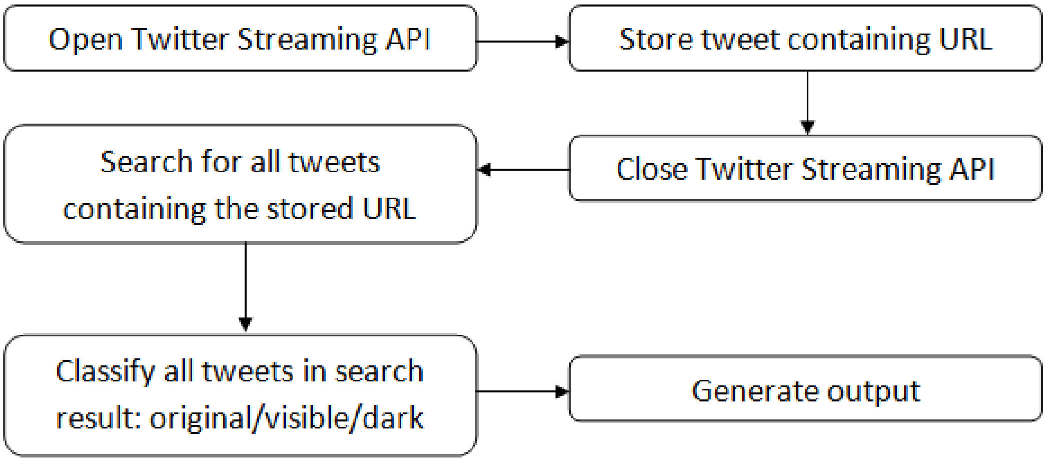

Open Twitter Streaming API
Store tweet containing URL
Close Twitter Streaming API
Search for all tweets containing the stored URL
Classify all tweets in search result: original/visible/dark
Generate output

Figure 6.3: Architecture of components in collecting URL drill-down dataset

In URL drill-downs, fewer requests were made to Twitter's Streaming API, because for each URL retrieved, more tweets were gathered and classified. These tweets were retrieved by requesting all the tweets containing the URL via Twitter's Search API. This request can return up to 18,000 tweets per URL, as per the rate limit set by the Twitter API. This method results in a dataset containing considerably more tweets compared to the random sample, namely 262,517 tweets[4]. However, URL drill-downs involve a smaller number of only 747 URLs. This dataset was collected between 28 Oct 2013 – 9 Jan 2014 (69 days).

The differences between the two datasets are outlined in Table 6.3.

[4] The final dataset size of 262,517 tweets was produced after removing all non-ASCII tweets from the raw dataset of 443,919 tweets.

Table 6.3: Dataset comparison between random sampling and URL drill-downs

Dataset	Tweets	Unique URLs
Random sampling	27,146	25,273
URL drill-downs	262,517	747

6.2.1 Distribution of Dark Retweets

Using the dataset for URL drill-downs, Figure 6.4 illustrates the distribution of dark retweet proportions over total retweet volume. Each datapoint denotes a unique URL, the y-axis denotes the total retweet volume per URL, whilst the x-axis denotes the proportion of dark retweets over total retweets. Therefore, the datapoints on the top-right quadrant of the graph show URLs which are disproportionately numerous and are very dark compared to the rest of the dataset. Since these few URLs are associated with so many tweets, they could skew the results of any simple analysis. Therefore more complex analyses are required.

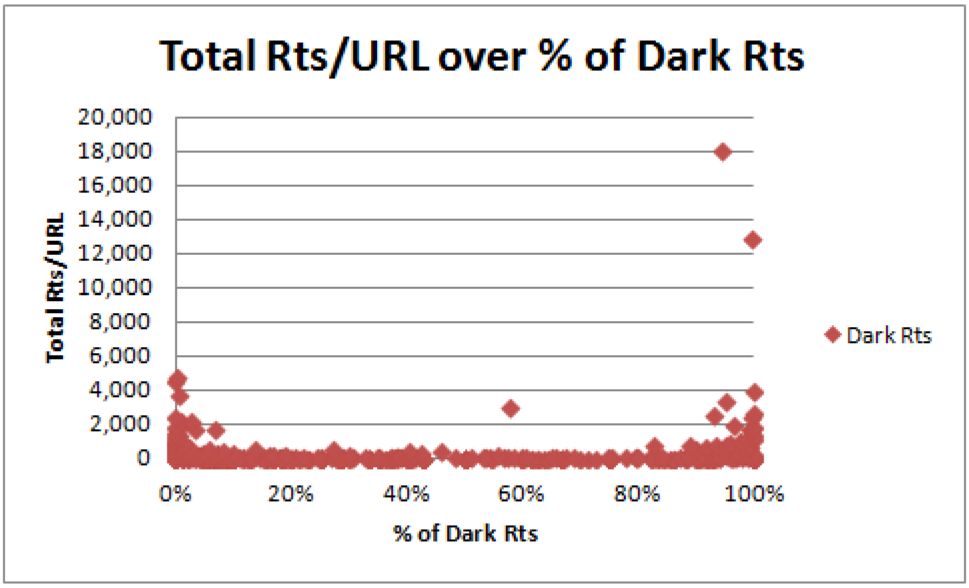


Figure 6.4: Distribution of dark retweet proportions over total retweet volume

Effect of Retweet Volume on Averages

In order to investigate the effect of URLs with different volumes of total retweets, Figure 6.5 (page 88) shows the differences in average proportions of dark retweets when the total retweet volume was binned into 10, 100, 1,000, 10,000 and 100,000 tweets. For each bin, the average dark retweet proportion for all URLs with the corresponding bin's retweet volume is calculated. For example, for the first bin of value 10, all URLs with between 1-10 total retweets are grouped together and the average dark retweet proportions for those URLs are calculated. This is repeated for each of the bins. This graph shows that

URLs with large numbers of total retweets show a disproportionately large average of dark retweets, in comparison to all the other bins.

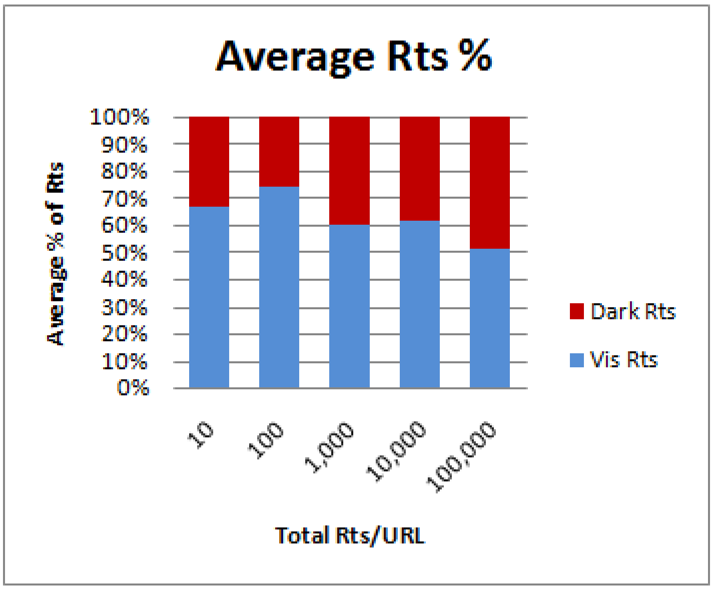


Figure 6.5: Average visible/dark Rt proportions over binned total Rt volume

Figure 6.5 was generated based on the data shown in Table 6.4. This table shows that the largest bin of retweets/URL, the 100,000 bin, involves 30,898 retweets but only 2 URLs. This means that those 2 URLs account for roughly 20% of overall retweets in the whole URL drill-downs dataset, and these are both unusually dark.

Table 6.4: Proportions of visible/dark Rts over total Rts/URL

Rt Bins	Vis Rts	Dark Rts	Total URLs	Total Rts	Avg Rts/URL
25	66.4%	33.6%	376	2,477	7
50	66.0%	34.0%	78	2,862	37
100	80.7%	19.3%	121	9,345	77
250	76.5%	23.5%	76	11,046	145
500	81.5%	18.5%	38	13,509	356
1,000	57.3%	42.7%	26	18,198	700
2,000	58.9%	41.1%	17	24,861	1,462
3,000	42.5%	57.5%	7	16,557	2,365
5,000	61.6%	38.4%	6	23,358	3,893
10,000	0.0%	0.0%	0	0	0
100,000	3.2%	96.8%	2	30,898	15,449
	TOTAL		747	153,111	205

The figure and table shows preliminary evidence that the number of overall retweets can affect the calculations of retweet proportions. Using these values, a cumulative table was derived to see the extent of how total retweets affect the averages of visible and dark retweet proportions. In Table 6.5 (page 89), the individual bins are accumulated to show the snowball effect of these totals — how do the different proportions for each individual bin affect the averages for the overall dataset?

Table 6.5: Cumulative proportions of visible/dark Rts over total Rts/URL

Rt Bins	Vis Rts	Dark Rts	Total URLs	Total Rts	Avg Rts/URL
25	66.4%	33.6%	376	2,477	7
50	66.2%	33.8%	454	5,339	12
100	75.4%	24.6%	575	14,684	26
250	75.9%	24.1%	651	25,730	40
500	77.8%	22.2%	689	39,239	57
1,000	71.3%	28.7%	715	57,437	80
2,000	67.6%	32.4%	732	82,298	112
3,000	63.4%	36.6%	739	98,855	134
5,000	63.0%	37.0%	745	122,213	164
10,000	63.0%	37.0%	745	122,213	164
100,000	51.0%	49.0%	747	153,111	205

Based on the values in Table 6.5, the changes in the cumulative averages are illustrated in Figure 6.6. Through manual inspection, this figure shows a natural division along the 500 Rts/URL bin. For URLs containing up to 500 retweets each, the averages for visible and dark retweets tend towards an 79/21 split, almost similar to the 79/21 split seen in the random sampling dataset from the previous analysis.

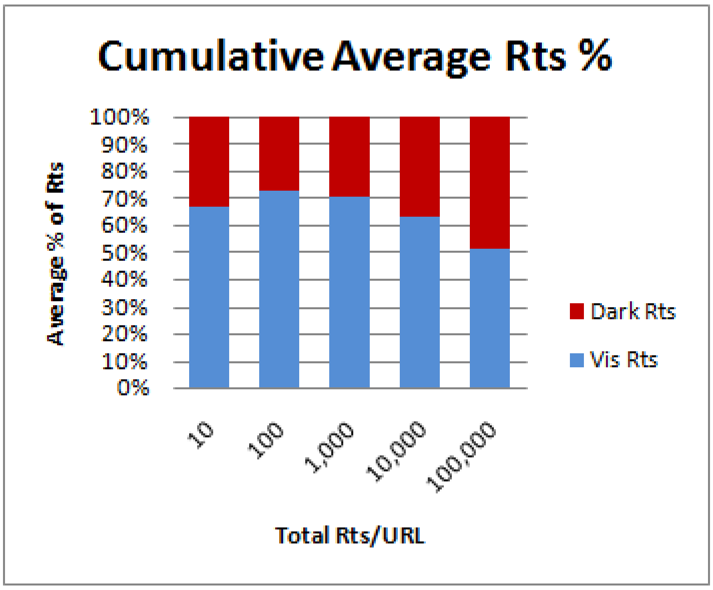


Figure 6.6: Cumulative averages for visible and dark retweet proportions

However, after the 500 Rts/URL bin, the overall average for visible retweets decreases up until the largest bin. This shows that for URLs containing large numbers of retweets, they are more likely to contain dark retweets, and thus skew the overall averages of visible and dark retweets.

6.2.2 Supervisible and Superdark URLs

The analysis above shows that there exists a small sample of URLs which are very dark or very visible, skewing the average proportions of visible and dark retweets. The terms **supervisible** and **superdark** URLs are introduced to identify them. A URL is considered supervisible/superdark if: 1) the URL results in a disproportionately large number of retweets, and 2) a disproportionately large percentage of those retweets are predominantly visible or dark.

Statistical Definition of Superdark and Supervisible URLs

In order to create a quantitative definition of what supervision and superdark URLs mean, a statistical analysis was done to see the distribution of quartiles within the URL drill-downs dataset. Using these quartiles over an ordered dataset, the upper quartile splits the highest 25% from the lowest 75% of the dataset. Therefore, the upper quartiles for two variables, namely total retweets and proportions of visible/dark retweets, are taken into consideration in order to quantitatively identify supervisible or superdark URLs.

Table 6.6: Quartiles of total retweets and proportions of visible and dark retweets

Quartile	Total Rts	Vis Rts %	Dark Rts %
Q1	4	33.3%	0.0%
Q2	23	95.2%	4.8%
Q3	96	100.0%	66.7%

Based on Table 6.6, the quantitative definition of these super URLs are as follows:

- Supervisible: URLs with more than 96 retweets where 100% of those retweets are visible.
- Superdark: URLs with more than 96 retweets where more than 66.7% of those retweets are dark.

Effect of Superdark and Supervisible URLs

Table 6.7 shows the distribution of URLs according to the quartiles as outlined in Table 6.6. This distribution table shows that 67 URLs were categorized as supervisible and 48 were superdark according to the quantitative definition as described above.

Table 6.8 shows that these upper quartile supervisible and superdark URLs actually constitutes 62.9% of the overall retweet count in the URL drill-downs dataset.

Table 6.7: Distribution of URLs in each Rt versus dark Rts % quartiles

			Unique URLs			
% of dark rts	>Q3	100.0%	56	40	36	48
	Q2–Q3	66.7%	30	78	58	27
	Q1–Q2	4.8%	0	3	32	45
	<Q1	0.0%	105	62	60	67
			4	23	96	17,999
			<Q1	Q1–Q2	Q2–Q3	>Q3
			Total retweets			

Table 6.8: Total retweets containing supervisible and superdark URLs

	URLs	Retweets
Supervisible	67	67,999
Superdark	48	28,373
TOTAL RTS		96,372
% of overall Rts		62.9%

Therefore, this research work has shown that in a random sample, roughly 79% of the retweets found were visible, with the remainder 21% being dark. However, once the dataset focuses on the individual URLs themselves, this study has shown that different URLs show different proportions of visible and dark retweets. A small subset of those URLs exhibit characteristics which identify them as supervisible and superdark URLs, which could disproportionately skew the average proportions for an overall dataset.

Investigating Superdark and Supervisible URLs

The top 10 supervisible and superdark URLs respectively were manually inspected to identify the content domains of these URLs. The URLs chosen for manual inspection were the ones which contained the most amount of retweets. Tables 6.10 (page 92) and 6.12 (page 94) show the result of this manual classification.

Tables 6.10 and 6.12 show that 17 of the 20 URLs contain predominantly more retweets instead of original tweets. In addition, manual inspections of these retweets seem to show that they are caused due to automation in various forms, such as bots, repetition, or auto-sharing of playlists.

Figure 6.7 shows an example of the same tweet text being repeated.

Figure 6.8 shows an example of a bot generating the same tweet text repeatedly, using the same authors repeatedly.

Finally, Figure 6.9 shows an example of a playlist being shared automatically as a different song is being broadcast.

Table 6.9: Top 10 supervisible URLs

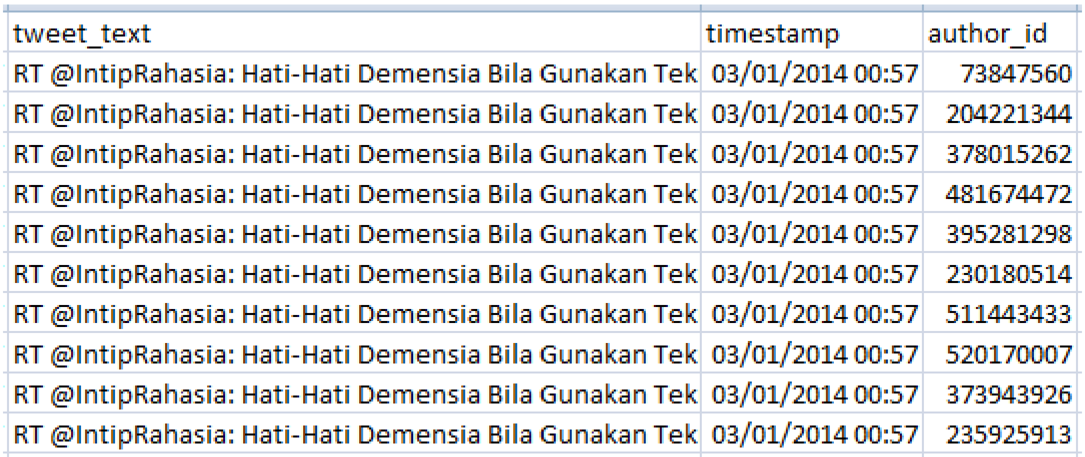
	URL
1	http://www.flickr.com/photos/abroaderview-volunteers
2	http://pictwitter.tv/15tHxlz
3	http://cakrawanita.blogspot.com/2013/06/hati-hati-demensia-bila-gunakan.html
4	http://beritaterhangat2013.blogspot.com/2013/12/5-tanaman-yang-membawa-keberuntungan-di.html
5	http://pic-twitr.com/1dNu3Dp
6	http://pictwitter.tv/19eM4Wb
7	http://smarturl.it/StoryOfMyLife
8	http://www.infowars.com/brand-obama-totalitarianism-2-0/
9	http://pic-twltter.cc/1a8J1gU
10	http://pic-twltter.cc/17eCwIs

Table 6.10: Manual classification for top 10 supervisible URLs

URL	Vis Rts	Ori %	Rts %	Domain	Mechanism
1	4,500	0.0%	100.0%	Photos	Bots retweeting - repeated authors
2	2,472	0.2%	99.8%	URL taken down	Bots retweeting - repeated authors
3	1,899	0.0%	100.0%	Blog	Repeated texts
4	1,497	0.1%	99.9%	Blog	Repeated texts
5	1,262	0.2%	99.8%	URL taken down	Repeated texts
6	1,060	0.1%	99.9%	URL taken down	Repeated texts
7	1,018	0.5%	99.5%	Music - iTunes	Repeated texts
8	900	0.0%	100.0%	Opinion piece	Repeated texts
9	851	0.5%	99.5%	URL taken down	Repeated texts
10	700	0.0%	100.0%	URL taken down	Repeated texts

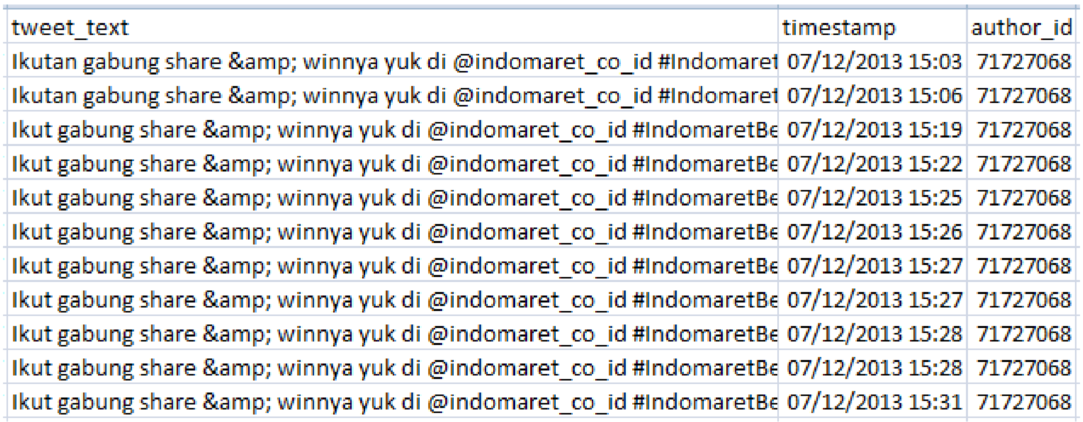
Table 6.11: Top 10 superdark URLs

	URL
1	http://saboom.okm.info/en/index.php
2	http://beautifulyou.indomaret.co.id
3	http://www.nurido.de/334-schufa-freier-kredit-ohne-bonitatsprufung-fur-deutschland/
4	http://pakistancyberforce.blogspot.com/2013/12/blog-post.html
5	http://VoteJordanHibbs.com
6	http://www.performertrack.com/event-reg-lc120413seminar.html
7	http://rdo.to/vampires
8	http://tunein.com/radio/Party-1025FM-(El-Party)-s166407/
9	http://ift.tt/1d7MxwM
10	http://www.player-webservic.com/9890



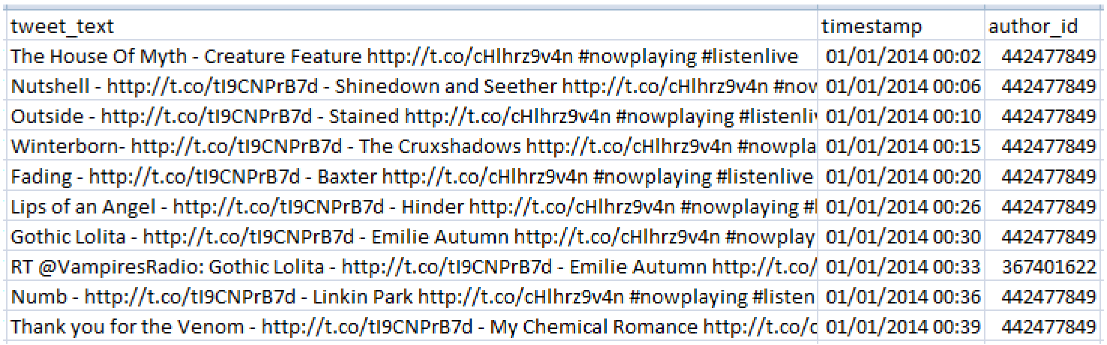
tweet_text	timestamp	author_id
RT @IntipRahasia: Hati-Hati Demensia Bila Gunakan Tek	03/01/2014 00:57	73847560
RT @IntipRahasia: Hati-Hati Demensia Bila Gunakan Tek	03/01/2014 00:57	204221344
RT @IntipRahasia: Hati-Hati Demensia Bila Gunakan Tek	03/01/2014 00:57	378015262
RT @IntipRahasia: Hati-Hati Demensia Bila Gunakan Tek	03/01/2014 00:57	481674472
RT @IntipRahasia: Hati-Hati Demensia Bila Gunakan Tek	03/01/2014 00:57	395281298
RT @IntipRahasia: Hati-Hati Demensia Bila Gunakan Tek	03/01/2014 00:57	230180514
RT @IntipRahasia: Hati-Hati Demensia Bila Gunakan Tek	03/01/2014 00:57	511443433
RT @IntipRahasia: Hati-Hati Demensia Bila Gunakan Tek	03/01/2014 00:57	520170007
RT @IntipRahasia: Hati-Hati Demensia Bila Gunakan Tek	03/01/2014 00:57	373943926
RT @IntipRahasia: Hati-Hati Demensia Bila Gunakan Tek	03/01/2014 00:57	235925913

Figure 6.7: Example repeating tweet texts



tweet_text	timestamp	author_id
Ikutan gabung share &amp; winnya yuk di @indomaret_co_id #Indomaret	07/12/2013 15:03	71727068
Ikutan gabung share &amp; winnya yuk di @indomaret_co_id #Indomaret	07/12/2013 15:06	71727068
Ikut gabung share &amp; winnya yuk di @indomaret_co_id #IndomaretBe	07/12/2013 15:19	71727068
Ikut gabung share &amp; winnya yuk di @indomaret_co_id #IndomaretBe	07/12/2013 15:22	71727068
Ikut gabung share &amp; winnya yuk di @indomaret_co_id #IndomaretBe	07/12/2013 15:25	71727068
Ikut gabung share &amp; winnya yuk di @indomaret_co_id #IndomaretBe	07/12/2013 15:26	71727068
Ikut gabung share &amp; winnya yuk di @indomaret_co_id #IndomaretBe	07/12/2013 15:27	71727068
Ikut gabung share &amp; winnya yuk di @indomaret_co_id #IndomaretBe	07/12/2013 15:27	71727068
Ikut gabung share &amp; winnya yuk di @indomaret_co_id #IndomaretBe	07/12/2013 15:28	71727068
Ikut gabung share &amp; winnya yuk di @indomaret_co_id #IndomaretBe	07/12/2013 15:28	71727068
Ikut gabung share &amp; winnya yuk di @indomaret_co_id #IndomaretBe	07/12/2013 15:31	71727068

Figure 6.8: Example bots with repeating tweet texts and authors



tweet_text	timestamp	author_id
The House Of Myth - Creature Feature http://t.co/cHlhrz9v4n #nowplaying #listenlive	01/01/2014 00:02	442477849
Nutshell - http://t.co/tI9CNPrB7d - Shinedown and Seether http://t.co/cHlhrz9v4n #nov	01/01/2014 00:06	442477849
Outside - http://t.co/tI9CNPrB7d - Stained http://t.co/cHlhrz9v4n #nowplaying #listenli	01/01/2014 00:10	442477849
Winterborn- http://t.co/tI9CNPrB7d - The Cruxshadows http://t.co/cHlhrz9v4n #nowpla	01/01/2014 00:15	442477849
Fading - http://t.co/tI9CNPrB7d - Baxter http://t.co/cHlhrz9v4n #nowplaying #listenlive	01/01/2014 00:20	442477849
Lips of an Angel - http://t.co/tI9CNPrB7d - Hinder http://t.co/cHlhrz9v4n #nowplaying #l	01/01/2014 00:26	442477849
Gothic Lolita - http://t.co/tI9CNPrB7d - Emilie Autumn http://t.co/cHlhrz9v4n #nowplay	01/01/2014 00:30	442477849
RT @VampiresRadio: Gothic Lolita - http://t.co/tI9CNPrB7d - Emilie Autumn http://t.co/	01/01/2014 00:33	367401622
Numb - http://t.co/tI9CNPrB7d - Linkin Park http://t.co/cHlhrz9v4n #nowplaying #listen	01/01/2014 00:36	442477849
Thank you for the Venom - http://t.co/tI9CNPrB7d - My Chemical Romance http://t.co/c	01/01/2014 00:39	442477849

Figure 6.9: Example auto-sharing of playlists

Table 6.12: Manual classification for top 10 superdark URLs

URL	Dark Rts	Ori %	Rts %	Domain	Mechanism
1	17,021	0.0%	100.0%	Porn	Bots retweeting - repeated authors
2	12,890	3.3%	96.7%	Competition	Bots retweeting - repeated authors
3	3,970	55.0%	45.0%	Spam - loans	Bots retweeting - repeated authors
4	3,235	65.1%	34.9%	Blog	Repeated texts
5	2,673	0.7%	99.3%	URL taken down	Bots retweeting - repeated authors
6	2,443	0.1%	99.9%	Advert - seminar	Bots retweeting - repeated authors
7	2,399	0.9%	99.1%	Music - radio	Auto playlist sharing
8	1,933	0.4%	99.6%	Music - radio	Auto playlist sharing
9	1,859	53.1%	46.9%	URL taken down	Bots retweeting - repeated authors
10	1,769	17.3%	82.7%	Music - radio	Auto playlist sharing

The content domains of these URLs range from music providers to porn to blogs. Several URLs were also inaccessible at the time of manual inspection, therefore could not be classified.

In this subsection, this research work shows that superdark and supervisible retweets seem to originate from a variety of automated tweeting mechanisms.

6.3 Conclusion

In this chapter, two different sampling methods were applied to classify two datasets into the tweet propagation typology. A random sampling dataset was retrieved to identify the proportions of visible and dark retweets in randomly selected tweets. In the random sampling dataset, it was found that roughly 79% of retweets were visible, whilst the remaining 21% were dark.

Using the URL drill-downs dataset, this study has shown that there are variations to the proportions of visible and dark retweets depending on each individual URL. In addition, the total volume of retweets per URL could also disproportionately skew the average percentages of dark/visible retweets in the overall dataset. The dataset only approaches the 79/21 visible and dark retweets split — similar to random sampling — when URLs with up to only 500 retweets were considered.

Using the upper quartile of total retweets and proportions of dark/visible retweets, the terms supervisible and superdark URLs were introduced. These URLs contain large numbers of retweets which are predominantly visible/dark. These very visible/dark URLs were found to contribute 62.9% of the total overall retweets in this dataset. Manual inspections seem to show that automatically generated retweets are the main cause for these URLs which highly skew the overall averages of retweet proportions.

The analysis has shown that there is a relatively high proportion of dark retweets, and that for some URLs, dark retweets are in the majority. It has also revealed the existence of supervisible and superdark URLs which are significant in terms of number of tweets, and yet behave at the extremes. The next chapter looks at the importance of these findings for our understanding of behaviour on Twitter, by investigating how dark retweet behaviour varies according to the retweeter and the URL content domains.

Chapter 7

Exploring Behaviour of Dark Retweets

In Chapter 6, it was shown that in a random sampling dataset, roughly 21% of the retweets found were dark. The existence of potentially overlooked dark retweets may impact the conclusions of existing retweet propagation studies. For instance, the study by Suh et al. (2010) involved 74 million tweets, where 11.15% were found to be retweets[1]. Out of this subset, 28.4% of those retweets were found to contain URLs in their texts. If we apply the 21% proportion to the retweets in their study, it implies that there were around 2.1 million dark retweets that were overlooked. Even if we only include the retweets containing URLs, that still implies over 600,000 dark retweets missed.

However, this 79/21 pattern of visible/dark retweets is not uniform — in Chapter 6, different URLs were shown to exhibit different proportions of visible versus dark retweets. In particular, URLs which were very visible or very dark — the supervisible and superdark URLs — were skewing the average proportions across the dataset. In this Chapter 7, the behaviour of dark retweets is explored further, particularly to investigate the variation of dark retweet behaviour across different domains, and the impact of dark retweet behaviour on average reach. The aim is to answer two questions: whether different retweet actions result in varying numbers of average users reached, and whether certain content domains are more likely to be propagated using dark retweet instead of visible ones.

7.1 Reach

In the work by Kwak et al. (2010), the reach of a retweet was defined as the total recipients of the retweet that were not following the original author. This means that in

[1] Retweets were defined as tweets which "used a set of textual markers" including "RT @", "via @" and "HT @" among others (Suh et al., 2010).

a retweet chain A (original tweet) → B (retweet) → C (retweet), the total reach of this chain will be the followers of B and C who do not also follow A — effectively the sum followers of B and C which do not overlap with A's.

Gathering full retweet chains is difficult given Twitter API's rate limits, and was beyond the scope of this thesis. Therefore, instead of using Kwak's definition of reach, a related measure was calculated in this thesis, namely the impact of each retweet action. For every given retweet action, the total reach is calculated by summing up the total followers for each retweeter, whilst the average is calculated by dividing the total reach by the number of retweeters. Table 7.1 shows an example of how the total and average reach for retweet actions were calculated.

Table 7.1: Calculating reach of retweet actions

Rt action	Retweeters	Followers	Total reach	Average reach
PRtF	A	85	$A + B + C + D = 2643$	$\frac{12643}{4} = 661$
	B	2423		
	C	32		
	D	103		
RtnF	E	45	$E + F + G = 8811$	$\frac{8811}{3} = 2937$
	F	8523		
	G	243		

In this experiment, the reach of each retweet action is the sum total of all the followers who could potentially see a tweet. As described in the typology of tweet propagations (see Section 4.2 on page 57), the visibility of a tweet varies depending on whether it is a retweet, in which case all followers of a retweet author could potentially see the tweet, or a reply, where only mutual followers of the users mentioned in the tweet could potentially see that tweet. The total reach for all tweet propagation types were calculated to see if there were any underlying patterns between visible and dark retweets.

In addition to conventional retweets, the reach of replies were also taken into account. As replies can only be seen in the home timelines of mutual followers, then this experiment factored in the number of mutual followers for the two parties involved in any given reply. This was done by counting the mutual followers between the reply author and the recipient, as opposed to only counting the total followers of the author. For example, if User A sends a reply to User B, and they both share 5 mutual followers, then the reach for that reply will be 5 instead of the total followers that User A has. Therefore, replies are expected to reach less users than broadcasts via conventional retweets.

This analysis calculates the reach of the retweet actions found in the random sampling dataset of 27,146 tweets, as used in the previous chapter. The calculations for the mutual followers were done by requesting the list of followers for each user from the Twitter API.

Each followers list was limited to 25,000 followers per user in order to reduce the time and API requests taken by this experiment[2].

7.1.1 Results

The dataset for this experiment contained all of the non-original tweets within the whole 27,146 tweets corpus. As illustrated in Table 6.1 (page 85), there were 18,645 original tweets within the 27,146 tweets corpus, leaving 8,498 visible and dark retweets, and 3 orphan tweets. In this analysis, orphan tweets were ignored due to its negligible number (only 3 out of the total 27,146 tweets). Therefore, this study focuses only on the 8,498 visible and dark retweets.

Based on this tweet corpus, Table 7.2 shows the total tweets that were analyzed across the different propagation types — visible and dark — plus their respective total users reached, the median reach, and the average reach.

Table 7.2: Average reach per retweet action

Tweet type		Rts	Reach	Median Reach	Avg reach/Rt
VISIBLE	PRtF	3,560	5,954,397	486	1,673
	PRtnF	2,189	3,041,149	270	1,389
	Rt@F	0	0	0	0
	Rt@nF	22	89,573	1,160	4,072
	RtF	595	1,200,414	469	2,018
	RtnF	345	985,855	593	2,858
	P@RtF	0	0	0	0
	P@RtnF	4	1,678	182	420
	@RtF	12	795	27	66
	@RtnF	4	96	4	24
TOTAL VISIBLE		**6,731**	**11,273,957**	**412**	**1,675**
% VISIBLE		**79.2%**	**63.8%**	-	-
DARK	RtF_d	1,675	6,368,752	1,009	3,802
	RtnF_d	0	0	0	0
	P@F_d	0	0	0	0
	P@nF_d	29	7,295	30	252
	@F_d	39	9,125	27	234
	@nF_d	24	912	9	38
TOTAL DARK		**1,767**	**6,386,084**	**920**	**3,614**
% DARK		**20.8%**	**36.2%**	-	-
OVERALL TOTAL		**8,498**	**17,660,041**	**477**	**2,078**

[2] Each Twitter API request only returns up to 5,000 followers' IDs per request. Therefore, up to five requests were made to the API per user in order to reduce the time taken and the number of requests used up before the API's rate limits were enforced. Out of the 5,504 users stored in this dataset, 208 (3.8%) were found to have 25,000 followers.

In Kwak et al. (2010), it was found that every retweet would reach 1,000 users on average, irrespective of how many followers the originating author had.

Since the study by Kwak et al. (2010) was calculating total reach, including all retweets in a chain, it is reasonable to expect that the average reach of a retweet action measured in this thesis, namely individual retweets in those chains, would be less. However, the data shows that the average reach for visible retweet actions were marginally higher (1,675 users per retweet) than the finding in Kwak et al. (2010). For dark retweets, the average reach (3,614 users per retweet) was just over 2 times the average reach size of visible retweets. This seems to suggest that dark retweets reach more users on average compared to visible retweets.

This seems to be caused mainly by RtF_d tweets[3] — a dark RtF would reach 3,802 followers on average. It is not immediately clear why dark RtF tweets seem to reach more followers in comparison to conventional visible retweets. For the whole dataset, the total median reach for a retweet action was only 477. The difference between this low median and the high average indicates that certain retweet actions, such as RtF_d tweets, were being made by a minority of retweeters who reached a disproportionately large number of followers, thus skewing overall average reach calculations.

In addition, this analysis does not distinguish between unique users, thus all potential followers are added up equally. Therefore, it is possible that this analysis may overestimate the total number of users who might have been shown the same tweet more than once. In practice, current Twitter clients hide multiple instances of the same tweet, but the reach calculation only does a simple addition of all the followers for all authors instead of keeping track of unique followers. However, the finding in Kwak et al. (2010) also did not make any distinctions between unique users or otherwise, therefore it is assumed that their findings was based on reaching non-unique users as well.

7.1.2 Distribution of Retweeters' Followers

As mentioned previously, one potential reason why dark retweet actions had twice the average reach as compared to visible ones would be if dark retweets were being favoured more by popular authors, with higher volumes of followers.

Figure 7.1 (page 101) shows a histogram of the total followers for all of the retweeters in this dataset. This histogram was done by observing the total followers for each retweeter, and then binning the frequencies of those retweeters. Each bin represents the total followers per retweeter, whilst the y-axis represents the percentage of retweeters corresponding to each bin. These percentages have been normalized according to overall total retweeters, so that adding up the total percentages across all bins will result in

[3] RtF_d = RtF (dark) = Dark RT/via retweets made by followers = Tweets which propagates URLs previously received by followers but with no conventional retweet markers such as "RT" or "via."

100%. The bins are not uniform in size — this is to allow a more granular observation towards the lower end of the scale.

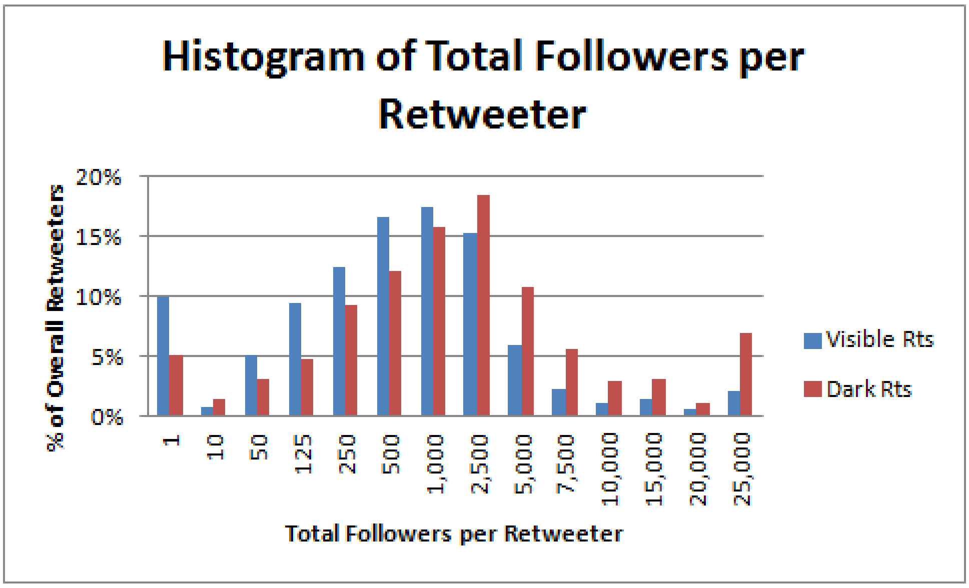


Figure 7.1: Histogram of total followers per retweeter

This histogram shows a non-uniform distribution of followers for both visible and dark retweets. For visible retweets, the distribution peaks at the 501–1,000 bin, whilst the dark retweets distribution peaks at the following 1,001–2,500 bin. This shows a slight skew towards the right for dark retweets, signifying that dark retweets were made by slightly more retweeters with larger followers, in comparison to visible retweets.

The distribution of total followers per author for all 18 different tweet propagation types[4] could be a potential area for future research. In addition, future work could also look at the tweet propagation patterns of users categorized based on their total followers. For instance, a comparison could be made to see whether users with high volumes of followers would use different tweet propagating patterns in comparison to less popular users.

7.1.3 Findings

In this study on average reach of retweet actions, it was found that dark retweets' average reach was just over twice as large as visible retweets. An analysis done on the total followers for all retweeters showed that in comparison to visible retweets, dark retweets were made by people who tended to have larger follower numbers.

However, these findings were derived under the definition of reach for a given retweet action as the total followers for all retweeters. This differs from the definition by Kwak et al. (2010), which involved followers who weren't following the original tweet's author.

[4] Excluding proprietary direct messages made by followers (PDMF).

Despite this difference in definitions, this work on average reach indicates that actual retweet chains may be underestimated in existing research work such as in Kwak et al. (2010), and that retweets chains may be longer than previously found.

7.2 Content Domains

This second study investigates whether certain content domains are more likely to be propagated using dark retweets rather than visible ones. This study is based on the work by Nagarajan et al. (2010), which claimed that out of four content domains in its study, three categories — call for action, collective groups and crowdsourcing tweets — were more likely to contain sparse author attribution, as opposed to information sharing tweets. Sparse retweet networks contain tweets which could be implied as a retweet/repost/copy, but did not contain an author attribution. In this thesis, tweets with these same characteristics have been classified as dark retweets.

In another study looking at the propagation patterns based on different user types, Cha et al. (2010) found differences between Twitter accounts for news sources and celebrities; mainstream news organizations[5] were consistently retweeted over a diverse set of topics, while celebrities were better at attracting mentions within other people's tweets. The research work in this thesis aims to observe if different content domains propagate in different patterns, not unlike the findings found in Nagarajan et al. (2010) and Cha et al. (2010).

Following a similar methodology to the studies above, this analysis attempts to classify URLs in the drill-down dataset, consisting of 262,515 tweets, into several content domains. This classification is done by manually accessing the URLs and grouping them according to content types. Once the classification is done, then the visible/dark retweet proportions for each content domain is observed.

7.2.1 Classification of Domains

Using the 262,515 tweets URL drill-down dataset, all the URLs in this dataset were extracted. From this list of 747 URLs, the supervisible URLs (67) and superdark URLs (48) were removed in order to minimize any skewing of the results. The remaining 632 URLs were then grouped into domains based on the content of the URLs. The classification was done manually by *a*) accessing the URLs with a web browser, then *b*) deducing the content type of each URL manually, and then *c*) grouping them into a domain class appropriate to the content. Once this manual classification was completed, the proportions of retweet types for all these domains were then recorded.

[5] Cha et al. (2010) did not elaborate on what constitutes as a mainstream news organization.

After the manual classification was done, eleven content domains were finally chosen to be representative of the 632 URLs. The composition of these content domains is as follows:

Social media Web pages offering social networking topological services, for example finding new followers and tracking unfollowers[6].

News Web pages displaying time-stamped reports on current events.

Blog Web pages displaying time-stamped articles in reverse-chronological order. These articles tend to include more opinion pieces and are open to comments from readers.

Music Web pages displaying content related to music, such as audio files, playlists, iTunes, online radio stations, etc.

Photos Web pages displaying content related to photos, such as graphic files, memes, photo-sharing web sites like Flickr, etc.

Videos Web pages displaying content related to videos, such as video files, video-sharing web sites like YouTube, etc.

Retail Web pages which promote retail services, ranging from one specific product to online auction sites. These could potentially fall under the spam category if they were distributed in an unsolicited manner, but considering that these pages were still online and not taken down, then the assumption used is that these retail pages are legit and thus not considered to be spam.

Competition Web pages promoting competitions which encourage the public to generate retweets and mentions as a form of voting.

App Web pages offering mobile application downloads and information.

Collective action Web pages which encourage an action from the mass public. These URLs range from online voting pages to petition sites.

Government Web pages from official governmental agencies providing information for the public.

Events Web pages promoting and allowing sign-ups for events such as seminars, etc.

Literature Web pages displaying works of literature such as poems, stories, etc.

Spam Web pages which were taken down using spam notices[7]

URL taken down Web pages which could not be resolved or led to 404 ‘page not found’ errors.

[6] unfollowers: Twitter users who used to follow another user but unfollowed him/her later on.

[7] When ISPs or official web services take down a web page for suspected spam content, the missing content would be replaced by notices saying that these pages were suspected to contain spam.

7.2.2 Comparison to Nagarajan et al. (2010)

The dataset in this thesis does not use the same keywords nor tweets that were used by Nagarajan et al. (2010), but using their methodology, this thesis has also grouped URLs to emulate the domains that were presented in that study.

There were four content domains classified in the study by Nagarajan et al. (2010). However, in this thesis, a more granular classification was conducted on the dataset. Table 7.3 shows the comparison between the content domains in Nagarajan et al. (2010) and in this thesis.

Table 7.3: Comparison of content domains between Nagarajan et al. (2010) and this thesis

Nagarajan et al. (2010)	This thesis
Information sharing	Social media Spam App Government Events Literature URL taken down Blog Retail Competition News Music Photos Videos
Call for social action Collective group identity-making Collective action	Collective action

There were a few instances where the content was deemed to be unrelated and unsuitable for classification under the domains as stated in Nagarajan et al. (2010). In these cases, extra domains which weren't found in that study were added and used instead. This was done to better reflect the different types of content for the 632 URLs in this dataset.

Differentiating between call for social action, collective group identity-making and crowd-sourcing was also difficult. In Nagarajan et al. (2010), the examples for the three domains were given as follows:

- *call for some sort of social action: "show support for democracy in Iran add green overlay to your Twitter avatar with 1 click".*
- *collective group identity-making: "Join @MarkUdall @RitterForCO and @BennetForCO to support an up or down vote on the public option"*

- *crowdsourcing: "Tell John Boehner that you are one of millions of Americans who supports a public option"*

Given that these three content domains seemed quite similar in nature, and that the three content domains in Nagarajan et al. (2010) exhibited the same characteristic of producing sparse retweet networks, therefore it was decided to combine these three domains into the collective action domain.

One possible attempt to ensure that URLs relating to call for social action, collective group identity-making and crowdsourcing domains remain separated is to emulate the data collection methodology used in Nagarajan et al. (2010), which was to collect tweets related to specific events, such as the Iran election or healthcare reform debate. Seeded sampling methods were deployed, where the seeds for subsequent tweet searches were already pre-determined, namely four event-based hashtagged tweets. Due to the events' discursive nature, it might have been potentially easier to identify separate URLs for the three content domains. Given the randomized nature of this thesis's dataset, it might have been possible that this dataset was not populated with many URLs related to these three content domains. Searching tweets using seeded events could be a possible methodology for future work in this area. In addition, future work could focus on the difference between random and seeded sampling methods to determine whether they could affect the outcomes of Twitter experimentation.

7.2.3 Findings

The results of this manual classification are presented in Table 7.4 (page 106), particularly the content domains that were chosen and their respective URLs and tweets. The tweet propagation patterns for each group were generated using the same mechanism as the main study on retweet proportions (see Section 6.1 on page 82).

As shown in Table 7.4, the miscellaneous domain was derived by merging all the content domains which contained less than 10 URLs each. These low-frequency content domains are: app, collective action, government, events, and literature.

Based on this study, basic information URLs such as news, music, photos and videos seem to be propagated using more visible retweets compared to the rest of the content domains. These findings for basic information URLs such as music, photos and videos seem to corroborate some of the findings in Nagarajan et al. (2010) — in this study, information sharing tweets were more likely to contain dense retweet networks containing author attributions, which are synonymous to visible retweets.

In comparison, social media and spam URLs seem to be distributed using more dark retweets. Manual inspections of the tweets related to these domains show that some form of automatic tweet generation is in place, resulting in tweet texts which are almost

Table 7.4: Proportions of visible and dark retweets based on content domains

Content Domain	URLs	Vis Rts	Dark Rts	Vis Rts %	Dark Rts %
Social media	19	549	381	59.0%	41.0%
Spam	50	4,431	3,026	59.4%	40.6%
Miscellaneous*	21	302	140	68.3%	31.7%
URL taken down	41	770	301	71.9%	28.1%
Blog	96	1,692	499	77.2%	22.8%
Retail	22	447	99	81.9%	18.1%
Competition	19	2,636	328	88.9%	11.1%
News	88	4,797	541	89.9%	10.1%
Music	45	8,557	790	91.5%	8.5%
Photos	57	4,578	242	95.0%	5.0%
Videos	174	18,694	749	96.1%	3.9%
TOTAL	632	47,453	7,096	87.0%	13.0%
*App	7	20	20	50.0%	50.0%
*Collective action	7	134	54	71.3%	28.7%
*Government	4	53	38	58.2%	41.8%
*Events	2	3	28	9.7%	90.3%
*Literature	1	92	0	100.0%	0.0%

identical between each other. These tweets do not contain any conventional retweet markers, therefore classified as dark retweets. The relationship between auto-senders and dark retweets may be useful towards formulating possible filtering mechanisms in the future.

Another notable finding is the overall total split of visible versus dark retweets: 87% to 13%. This differs from the 79/21 split as shown in the previous Chapter 6. This could be due to the exclusion of supervisible and superdark URLs which was shown to have skewed the average proportions in the previous chapter.

7.3 Conclusion

In the previous Chapter 6, it was shown that there was a 79/21 visible/dark retweets split in a random dataset of tweets. However, the behaviour of dark retweets is not uniform — supervisible and superdark URLs could change the proportions of visible/dark retweets in a dataset. This chapter focuses on the behaviour of dark retweets by exploring their variations in average reach and content domains.

This resulted in the finding that the average reach of dark retweets (3,614 users per retweet) was just over 2 times the size of visible ones (1,675 users per retweet). This reach is dependent on the number of followers for each retweeter — in this thesis's dataset, dark retweets seemed to be made more by retweeters with higher numbers of followers.

In terms of content domains, social media and spam URLs seemed to make use of dark retweets more (41% and 40.6% respectively) compared to basic information URLs such as music (8.5%), photos (5%) and videos (3.9%).

The same study also found that once the supervisible and superdark URLs were discarded from analysis, the proportion of visible retweets increased from 79.2% to 87%, whilst dark retweets decreased from 20.8% to 13%.

These two experiments show that there exists a complexity to the behaviour of dark retweets — the proportions of visible/dark retweets in a given dataset varies according to the retweeter and the content domain of the URL.

Chapter 8

Conclusions and Future Work

The motivation behind this thesis was the observation that dark retweets were possible to create, but there is no body of work that clearly defines what they are, how to detect them and how to classify them. In this thesis, a typology of tweet propagation types was presented in order to study the phenomenon. A pilot study was run based on a preliminary typology which had only seven retweet propagation types. This study found that more than half of all the tweets in the pilot dataset was categorized as 'other implicit retweets'. This early work inspired an expanded typology of 19 different tweet propagation types.

Using two datasets — a random sample of 27,146 tweets, and a URL drill-down dataset of 262,517 tweets — the proportions of each propagation type were observed to explore the extent of dark retweets as compared to visible ones. This was then followed by two studies looking at the behaviour of dark retweets across two themes, namely average reach of retweet actions and content domains.

This final chapter discusses the contributions of the research work in this thesis, followed by possible future work that could extend this research.

8.1 Research Contributions

This research work has several contributions:

- A comprehensive typology of tweet propagation types, as presented in Chapter 4 (page 55), consisting of 19 different ways a tweet could spread. This typology was developed based on seven characteristics: whether it is proprietary, the mechanism used, whether it is directed to followers or non-followers, whether it mentions other users, if it is explicitly propagating another tweet, if it links to an original tweet, and what is the audience it is pushed to.

- The introduction of “dark retweets”. In Section 4.4 (page 64), two retweetability confidence factors were presented, and the term dark retweets was introduced to encompass tweets which would require higher degrees of assumptions compared to other visible retweets. Essentially, dark retweets are considered to be tweets which were not propagated using proprietary retweet mechanisms or prefixing copied and pasted tweets with conventional retweet markers (e.g. ‘RT’/‘via’).
- An experimental toolkit developed using Python, as shown in Chapter 5 (page 69), that can retrieve random URLs from the Twitter Streaming API, and do various analyses on tweets containing those URLs. These analyses include getting the proportions of 18 different tweet propagation types — excluding the inaccessible private proprietary direct messages made by followers (PDMF) — within a retrieved dataset, illustrating the lifespan of propagated tweets, and calculating their average reach. This toolkit is downloadable on Github via the following URL: `https://github.com/coolster1/dark-rt-toolkit`
- A novel analysis of the extent of dark retweeting. Specifically, a 20.8% proportion of dark retweets (see Section 6.1 on page 82) was found within the random sample of 27,146 tweets. This gives a rough 79/21 split of visible/dark retweets in the dataset. However, this average was susceptible to a subset of URLs, namely supervisible and superdark URLs, which were skewing the average proportions of visible and dark retweets found in the dataset. Once the supervisible and superdark URLs were discarded from analysis, the proportion of dark retweets decreased from 20.8% to 13%, whilst visible retweets increased from 79.2% to 87%.
- An evaluation of the behaviour of dark retweets on various aspects:
  - In Section 7.1.3 (page 101), the average reach of a dark retweet action (3,614 users per retweet) was found to be just over double the average reach of a visible retweet action (1,675 users per retweet).
  - In Section 7.2.3 (page 105), dark retweets were found to be more frequently used in spreading social media and spam URLs. Visible retweets were more prevalent amongst basic information domains such as music, photos and videos.

This research work was published in three conference papers:

- Azman, Norhidayah, Millard, David E. and Weal, Mark J. (2010) Issues in Measuring Power and Influence in the Blogosphere. At *Web Science Conference 2010, Raleigh, NC, USA, 26 - 27 Apr 2010.*

- Azman, Norhidayah, Millard, David E. and Weal, Mark J. (2011) Patterns of Implicit and Non-follower Retweet Propagation: Investigating the Role of Applications and Hashtags. At *Web Science 2011, Koblenz, Germany, 14 - 18 Jun 2011.*

- Azman, Norhidayah, Millard, David E. and Weal, Mark J. (2012) Dark Retweets: Investigating Non-conventional Retweeting Patterns. At *4th International Conference on Social Informatics, Lausanne, CH, 05 - 07 Dec 2012.*

8.2 Research Hypothesis

The above contributions have been made based on the following hypotheses:

H1: A significant minority of retweets are dark retweets that do not use conventional mechanisms and therefore are difficult to detect

A typology consisting of 19 different tweet propagation types was developed based on seven tweet characteristics, namely: 1) whether it is proprietary, 2) the mechanism used, 3) whether it is created by followers or non-followers, 4) whether it mentions other users, 5) if it is explicitly propagating another tweet, 6) if it links to an original tweet, and 7) the audience that it is pushed to. The term “dark retweets” was introduced based on low confidence levels of two retweetability confidence factors, namely that 1) Tweet A is propagating Tweet B, and 2) the originating author can be identified. This typology was then validated across a dataset of 27,146 tweets.

A proportion of dark retweets were found over this dataset, namely 20.8%. This minority is considered as a significant minority because 1) it was found to increase the average reach of retweet actions, and 2) it was more prevalent amongst tweets propagating social media and spam URLs.

Detecting these tweet propagations requires extra detection procedures, which are different to simply requesting the Twitter API’s official retweet counts and counting manually created RT/via tweets.

H2: The behaviour of dark retweets is not uniform, and changes depending on the retweeter and the content domains of the URLs spread.

Despite the rough 79/21 visible/dark retweet split found in the random sample, this average is susceptible to distortion by supervisible and superdark URLs. These URLs contain disproportionately large numbers of retweets which are highly visible or dark. In the URL drill-down dataset, 67 supervisible and 48 superdark URLs were accounting for

96,372 retweets or 62.9% of overall retweets. Once these supervisible/superdark URLs were omitted from analysis, the visible/dark retweet split changed to 87/13.

Two additional studies were undertaken to examine the behaviour of dark retweets in terms of average reach and content domains.

Firstly, the experiment in Section 7.1.3 (page 101) was based on the work by Kwak et al. (2010) which looked at the average reach of a retweet. In this experiment, the average reach of a dark retweet action (3,614 users per retweet) was found to be just over double the average reach of a visible retweet action (1,675 users per retweet).

Secondly, as presented in Section 7.2.3 (page 105), all the URLs found in the dataset were manually grouped into different content domains, and the proportions of dark retweets found were recorded. It was found that the content domains with the most dark retweets were social media (41%) and spam (40.6%), whilst those with the least were basic information domains such as music (8.5%), photos (5%) and videos (3.9%).

8.3 Future Work

Based on the findings of this research work, there are several other threads of research that could subsequently be carried out:

8.3.1 Predicting Future Retweets and Virality

The existence of dark retweets could affect the findings of several studies on predicting future retweet rates and virality. Studies such as Hoang et al. (2011) define virality based on retweet count and likelihood, emphasizing the importance of retweets. In another example, Petrovic et al. (2011) proposed a time-sensitive model that predicts whether a tweet will be retweeted or not, and compared the results of this proposed model to the baseline of retweets as detected by two human subjects. Their work indicated that the time-sensitive model's performance was equivalent to the humans' performance. However, a retweet was defined as a retweet that was made using the proprietary Twitter mechanism only.

Similarly, the work by Macskassy and Michelson (2011) attempted to predict retweetability based on four models: general time-based power law, recent communication, on-topic and homophily. The last two models relate to the Twitter profiles of the users in the dataset, which were matched to Wikipedia articles in order to form a topical profile of each user. Their study showed that the on-topic model constituted the biggest probability of retweetability (51.6%), and a combination of all four models provided a better result in detecting a tweet's retweetability.

Several studies have also combined different models such as tweet content, social relationships and temporal factors; predictions of retweet patterns were shown to be improved in comparison to other baselines which did not combine different factors such as social relationships (Yang et al., 2010; Peng et al., 2011; Hong et al., 2011).

Pfitzner et al. (2012) investigated the role of tweet sentiment in predicting retweets. The study found no significant polarity of sentiments between original tweets and retweets, and that emotional divergence, i.e. the difference between positive and negative sentiments, can be a noticeable factor in deciding retweetability.

Ruan et al. (2012) proposed a prediction model that incorporated multiple factors such as network structure, user interaction, content characteristics and past activity. Retweets were considered as subsets of network and content features. Using these factors, particularly network features and past activity data, the model was able to predict the volume of future tweets.

The common thread across all of the above papers is that retweets form a substantial part of their analyses. They were either metrics of measurements which then became the basis of future predictions, or they became the units of those future predictions themselves. The existence of dark retweets could potentially mean that there were missing measurement data, or that future dark retweets were not included in the predictions. The difference in dark retweets' reach may help explain why certain prediction models performed worse than others.

8.3.2 Redefining Influence and Power on Twitter

Visible and hidden power were concepts described in the literature review in Chapter 2. Dark retweets not only impact existing calculations of power and influence, but they suggest that it may be possible to differentiate between different types of influence by taking into consideration the visibility of attribution.

Several studies into the roles of influencers have seemed to be based primarily on retweeting behaviours. Lumezanu et al. (2012) investigated the role of retweeting in spreading propaganda, particularly in identifying hyperadvocates via retweeting activity, using datasets from the Nevada Senate race (#nvsen) and the US debt ceiling debate (#debtceiling). Different patterns occurred in different community types. In communities with higher averages of tweets sent and higher retweet proportions, hyperadvocates were found to send more messages within short time windows, and more proportions of retweets. In communities with smaller averages of tweets sent and smaller retweet proportions, then hyperadvocates seemed to do more quick retweeting.

In a similar work, Tinati et al. (2012) defined user profiles based on total retweets and the propagation patterns they produced.

It is possible that due to dark retweets, different user profiles may exhibit retweeting behaviours that are different to the ones these papers have already detected. For instance, a certain user profile may prefer using dark retweets as opposed to visible ones, or that they prefer to use dark retweets only in particular instances, similar to the findings of the content domains experiment as done in Section 7.2.3 (page 105).

8.3.3 Adapting Propagation Typology to Facebook Interactions

If we were to use the above taxonomy of propagation onto a different medium such as social networking sites, then we could possibly map the different retweet types into different propagation types with similar characteristics but applicable to sites such as Facebook.

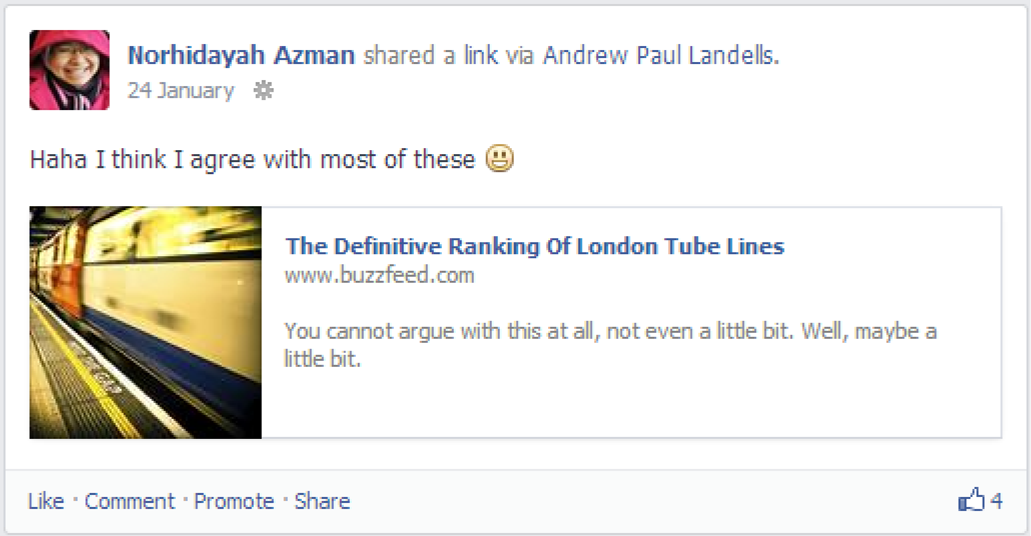
Facebook's *'Share'* function allows users to share items that have been posted by their friends. These items include URLs, uploaded pictures or videos. For example, User A posts a link to URL X on Facebook, which then gets seen by User B. User B can then share the same URL to his/her own friends by clicking the *'Share'* link at the bottom of User A's post. This sharing mechanism seems to be the likeliest approximation to Twitter's retweet function, due to its ability to share resources and propagate them onwards to friends, while still retaining the ability to comment on the shared resource.

Initial observations of resource sharing on Facebook seem to show that it is difficult to differentiate between proprietary sharing mechanisms (where users use Facebook's own *'Share'* link to propagate resources), or a Facebook equivalent of the RT/Via mechanism (where users simply cut and paste the same resource whilst bypassing the *'Share'* function). Facebook's sharing mechanism is more likely to be similar to the proprietary mechanism, rather than the RT/Via mechanism.

Meanwhile, replies could possibly be approximated by the action of writing on a friend's wall, because such wall-to-wall posts are displayed as a conversation between the parties involved on Facebook's news feeds. Similar to Twitter, these wall-to-wall posts could also be seen by mutual friends, and also users viewing the walls of the parties involved in the conversation.

Nonetheless, approximating non-follower and other "dark" propagations is different than with retweets. Firstly, Facebook's sharing mechanism automatically inserts a *via [user]* in the post of every resource shared, providing an automated acknowledgement or track-back to the person being referenced. This feature can easily facilitate the identification of proprietary follower propagation paths, as illustrated by Figure 8.1 (page 115).

Secondly, if a group of users are found to be sharing the same resource, such as a link to the same URL, Facebook automatically groups them together into a single resource post, as illustrated by Figure 8.2 (page 115).



Norhidayah Azman shared a link via Andrew Paul Landells.
24 January

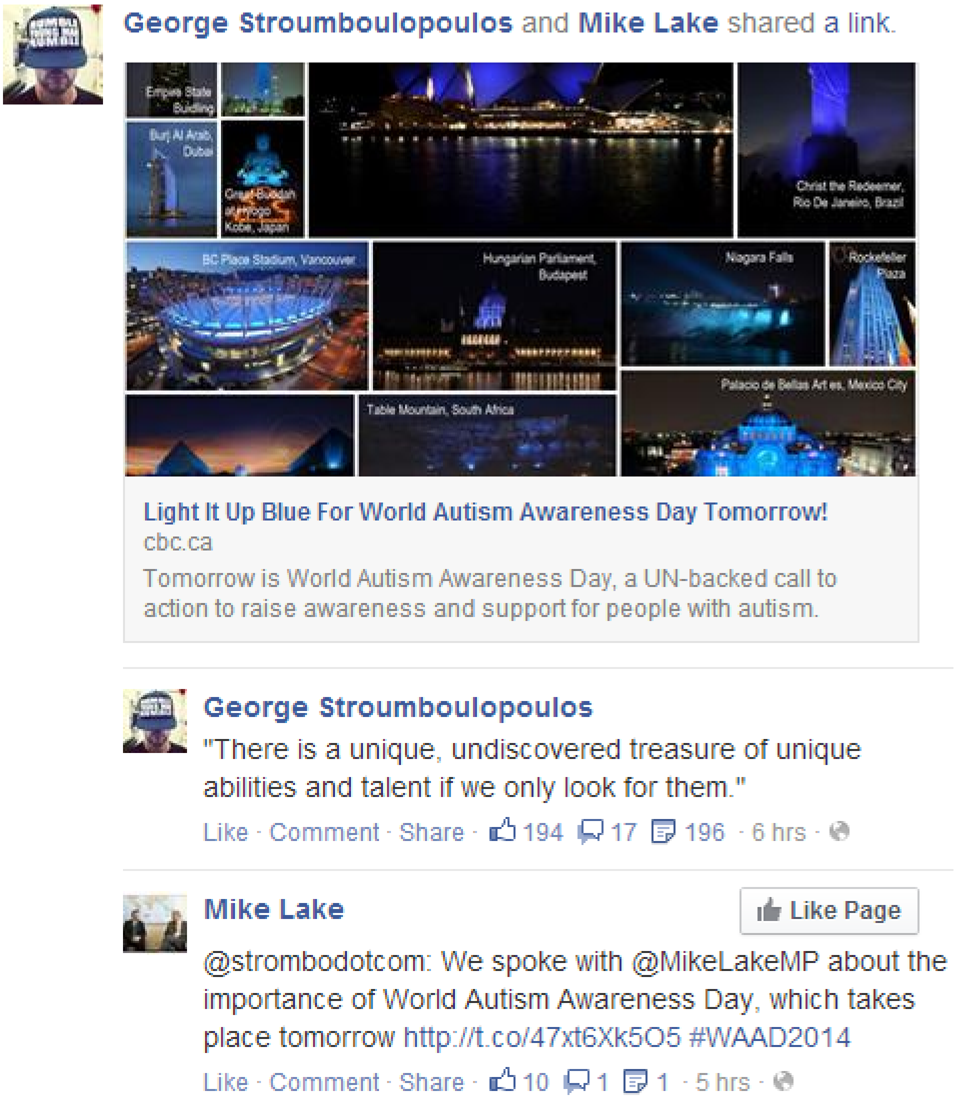
Haha I think I agree with most of these 😃

The Definitive Ranking Of London Tube Lines
www.buzzfeed.com

You cannot argue with this at all, not even a little bit. Well, maybe a little bit.

Like · Comment · Promote · Share 👍 4

Figure 8.1: Facebook's sharing mechanism with *via* acknowledgement



Empire State Building
Burj Al Arab, Dubai
Great Buddah at Hyogo Kobe, Japan
Christ the Redeemer, Rio De Janeiro, Brazil
BC Place Stadium, Vancouver
Hungarian Parliament, Budapest
Niagara Falls
Rockefeller Plaza
Palacio de Bellas Art es, Mexico City
Table Mountain, South Africa

George Stroumboulopoulos and **Mike Lake** shared a link.

Light It Up Blue For World Autism Awareness Day Tomorrow!
cbc.ca
Tomorrow is World Autism Awareness Day, a UN-backed call to action to raise awareness and support for people with autism.

George Stroumboulopoulos
"There is a unique, undiscovered treasure of unique abilities and talent if we only look for them."
Like · Comment · Share · 👍 194 17 196 · 6 hrs ·

Mike Lake Like Page
@strombodotcom: We spoke with @MikeLakeMP about the importance of World Autism Awareness Day, which takes place tomorrow http://t.co/47xt6Xk5O5 #WAAD2014
Like · Comment · Share · 👍 10 1 1 · 5 hrs ·

Figure 8.2: Facebook's sharing mechanism by multiple users

However, the closed architecture of Facebook means that information such as friend networks, their statuses and their shared resources could only be obtained if a particular user's privacy setting allows that information to be shared publicly. Due to the various privacy settings that can be customized in Facebook, it seems more difficult to obtain information on who could potentially see which resource. There seems to be no way of determining such custom privacy settings. Therefore, approximating non-follower propagation becomes difficult because it seems harder to determine if any user had seen any particular URL being shared by somebody else beforehand.

To gain access to this information, a Facebook application would have to be built and users would have to allow the application to collect such data from their profiles. This could lead to a possible skewness of the data collected.

Table 8.1 shows the possible mappings that could be made between different tweet propagation types and how they could apply to Facebook interactions.

Retweets	*Facebook*
Proprietary RT/Via	'Share' link
Replies	Wall-to-wall posts
Proprietary non-follower RT/Via non-follower Non-follower replies Dark retweets	Possible to identify, but difficult to collect big datasets due to privacy settings

Table 8.1: Mapping taxonomy of propagation types onto Facebook

8.3.4 Detecting Spamming Behaviour on Twitter

Chapter 6 described superdark and supervisible URLs which have a disproportionate influence on dark retweeting measures. Many of the retweets containing superdark URLs were classified as spam coming from automated processes and bots. Chapter 7 also showed that tweets containing spam URLs had a higher proportion of dark retweets than URLs from other content domains. This indicates that dark retweeting patterns could be very important in the area of spam detection.

A study by Tao et al. (2013) has focused on detecting near-duplicates within a tweet corpus in order to improve filtering performance. In this thesis, the relationship between dark retweets and spambots was briefly discussed in Section 7.2.3 (page 105). Determining whether dark retweets were more likely to contain spam tweets could provide an alternative perspective to this research area and inform further studies similar to the above.

8.4 Conclusion

Social media has become an integral part of our daily lives. The advent of Facebook, Twitter and other social networking sites has allowed millions of users across the globe to connect with each other, share their views and discuss a multitude of topics.

This shared global discourse has led to interesting studies in identifying influence. Since the conversations are public and machine-readable, researchers have developed quantitative techniques for answering questions such as what are the hottest topics being currently discussed? Who talks and listens to whom?

What counts as influence online is a subject of ongoing debate, but typically quantitative researchers have used proxy measures such as the volume and reach of retweets in order to generate their metrics. However, as shown in this thesis, relying solely on visible retweets may present a less whole picture of tweet propagation. As presented in Chapter 6, in a random dataset, 20.8% of retweets were dark, giving a rough 79/21 visible/dark retweet split. This shows that studies based solely on visible retweets may be under-representing the total tweet propagation paths that may exist. By omitting dark retweets from their studies, longer propagation chains may have been broken, therefore limiting the potential length of an observed propagation chain.

However, despite this significant overall proportion of dark retweets, we must be cautious as their behaviour is non-uniform. In Chapter 6, it was shown that supervisible and superdark URLs were skewing average proportions of the overall dataset. These URLs were exhibiting disproportionately large numbers of retweets, where disproportionate percentages of them were either visible or dark.

Therefore, in Chapter 7, the variability of dark retweet behaviour was explored further. Using two experiments, this chapter investigated whether different retweet actions displayed different averages of reach, and whether certain content domains contained more dark retweets than others. It was found that dark retweet actions reached on average twice the amount of users as compared to visible retweet actions. In addition, dark retweets were found to be more prevalent amongst tweets spreading social media and spam URLs in comparison to URLs of music, videos and photos. This study also observed that once the supervisible and superdark URLs were discarded from analysis, the proportion of dark retweets decreased from 20.8% to 13%, whilst visible retweets increased from 79.2% to 87%.

The work undertaken in this thesis has shown that dark retweets are a significant phenomenon, but that their behaviour is non-uniform. Due to the difficulty in detecting them, it seems that seems that they have been ignored in many existing studies. In order to get a bigger, more complete picture of how and where information spreads, we need more data scientists to embrace the dark side of retweets.

Appendix A

CSV Table Formats for Pilot Toolkit

The following tables describe the column headings for the CSV files that would get generated by the pilot toolkit. The CSV schema illustrating the relationships between table formats A, B, C and D can be seen on page 41.

A.1 Table Format A: Data collection

query: The URL being searched for. This URL would be repeated for every row.

type: Type of data being collected, in this case ’tweet’. This column was added to allow future developments such as collecting information from blogs (therefore the column would be labelled as ’blog’) or news sources (’news’).

item_id: Unique identifier for the data collected, in this case the tweet ID.

author_id: Unique identifier for the author, in this case the twitterer’s user ID.

date: The date the post was made, in this case the tweet’s published date.

time: The time the post was made, in this case the tweet’s published time.

item_text: The full text of the post, in this case the tweet’s full text.

No primary key was designated, as all the CSV files would be read sequentially row-by-row throughout the whole file.

A.2 Table Format B: Tweets and followers counts

date: The date the post was made, in this case the tweet's published date.

urls_count: Total tweets containing the query URL.

followers_count: Total followers who have potentially seen the query URL.

multiviews_count: If any particular follower could potentially see the same query URL more than once, then multiviews_count gets incremented.

A.3 Table Format C: Followers records

item_id: Unique identifier for the data collected, in this case the tweet ID.

author_id: Unique identifier for the author, in this case the twitterer's user ID.

date: The date the post was made, in this case the tweet's published date.

item_text: The full text of the post, in this case the tweet's full text.

user_mentions: List of user IDs mentioned in each tweet.

total_followers: Total followers of the tweet's author.

followers: List of the author's followers.

A.4 Table Format D: Retweet types

For the column headings in this CSV table format, the word **'chain'** was used instead of the term **'non-follower'**, in order to make the coding process faster and more efficient[1]. Throughout the pilot study, the term *non-follower* was used when describing retweet types concerning users disconnected from follower/following networks, but the term *chain* was used when naming the table formats' column headings and also scripting variables.

date or author_id: Tweet date, or unique user ID of author, depending if the table counts total retweets by date or by author.

native_rt (n_rt): List of authors making native retweets.

native_chain_rt (nc_rt): List of authors making native non-follower retweets.

rtvia_rt (r_rt): List of authors making RT/Via retweets.

[1] *non-follower* contains 12 characters, as opposed to *chain* which has only 5.

rtvia_chain_rt (rc_rt): List of authors making RT/Via non-follower retweets.

replies_rt (rp_rt): List of authors making replies.

replies_chain_rt (rpc_rt): List of authors making non-follower replies.

other_chain_rt (oc_rt): List of authors making implicit retweets.

Appendix B

CSV Table Formats for Main Toolkit

B.1 Preparing Streamed Tweets for Classification

url: The URL being searched for.

tweet_id: Unique tweet ID.

tweet_text: The full text of the tweet.

timestamp: The date and time the tweet was published.

author_id: The author's unique Twitter user ID.

rt_status_tweet_id: Tweet ID of the originating tweet, as provided by the Twitter API. This data is provided when the current tweet is a native retweet.

rt_status_user_id: User ID of the originating tweet's author, as provided by the Twitter API. This data is provided when the current tweet is a native retweet.

ent_mentions_user_id: User IDs for all users mentioned in the tweet text.

in_reply_tweet_id: Tweet ID of the originating tweet, as provided by the Twitter API. This data is provided when the current tweet is a native reply.

in_reply_tweet_id: User ID of the originating tweet's author, as provided by the Twitter API. This data is provided when the current tweet is a native reply.

B.2 Keeping Track of Friends Seen

friends: The friends' unique Twitter user IDs.

This CSV format is mainly used as a utility file to group together lists of friends that the toolkit has already seen. Only one utility CSV file is created for each specific URL, but it gets updated constantly for every author for each tweet that gets seen by the toolkit.

B.3 Recording Each User's List of Followers

followers: The followers' unique Twitter user IDs.

This CSV format is used to create individual CSV files for each author seen by the toolkit, therefore each unique user ID will have a CSV file created containing the list of all that user's followers.

B.4 Classified Tweets

url: The URL being searched for.

tweet_id: Unique tweet ID.

tweet_text: The full text of the tweet.

timestamp: The date and time the tweet was published.

author_id: The author's unique Twitter user ID.

proprietary: One-digit binary value signifying whether the tweet was made using Twitter's proprietary methods or not.

mechanism: Three-digit binary value signifying whether the tweet made was an original tweet, a retweet or a reply.

explicit: One-digit binary value signifying whether the tweet made was explicitly propagating another tweet or not.

follower: One-digit binary value signifying whether the tweet's author is a follower of the user being retweeted/referenced/replied to.

ori_item_id: Unique tweet ID of the tweet being retweeted/referenced/replied to.

ori_author_id: Unique Twitter user ID of the author being retweeted/referenced/replied to.

This CSV format is used to classify each tweet classified by the toolkit. This file format then becomes the basis for further grouping of the tweets according to the tweet propagation typology as described in Chapter 4.

B.5 Statistical Analysis of Classified Tweets

filename: The CSV file containing all the tweets for a particular URL.

ori: Total tweets which are classified as original tweets.

P@F: Total tweets which are classified under P@F.

P@nF: Total tweets which are classified under P@nF.

@F: Total tweets which are classified under @F.

@nF: Total tweets which are classified under @nF.

PRtF: Total tweets which are classified under PRtF.

PRtnF: Total tweets which are classified under PRtnF.

RtP@F: Total tweets which are classified under RtP@F.

RtP@nF: Total tweets which are classified under RtP@nF.

RtF: Total tweets which are classified under RtF.

RtnF: Total tweets which are classified under RtnF.

P@RtF: Total tweets which are classified under P@RtF.

P@RtnF: Total tweets which are classified under P@RtnF.

@RtF: Total tweets which are classified under @RtF.

@RtnF: Total tweets which are classified under @RtnF.

RtF_d: Total tweets which are classified under RtF (dark).

RtnF_d: Total tweets which are classified under RtnF (dark).

P@F_d: Total tweets which are classified under P@F (dark).

P@nF_d: Total tweets which are classified under P@nF (dark).

@F_d: Total tweets which are classified under @F (dark).

@nF_d: Total tweets which are classified under @nF (dark).

PDMF: Total tweets which are classified under PDMF.

OrphanRt: Total tweets which are classified under OrphanRt.

Orphan@: Total tweets which are classified under Orphan@.

total: Total sum of all tweets.

Based on the classified tweets, they get grouped according to the different tweet propagation types as labelled by the columns above.

Appendix C

Description of Binary Values in the Matrix of Tweet Propagation Types

The binary values below correspond to Table 4.1 (page 59) which illustrates the matrix of tweet propagation types. All the binary values of 1s and 0s denote different meanings, according to the seven characteristics in which these values fall under.

- Proprietary
  - A '1' denotes that tweets were made using proprietary functions, i.e. by using the official Twitter UI's functions, or 3rd party apps which use the Twitter API's proprietary function calls.
  - A '0' denotes that none of Twitter's proprietary functions were used.
- Mechanism
  - Rt (100): retweet
  - @ (010): reply
  - DM (001): direct message
- Follower/Non-Follower (F/nF)
  - 1: Tweet was made by a follower
  - 0: Tweet was made by a non-follower
- Mentions Other Users
  - 1: Tweet mentions the username (i.e. @username) of the originator (i.e. author of the tweet being retweeted/replied/referred to).

  - 0: No mention of the originator.

- Explicit
  - 1: Tweet was explicitly flagged as a retweet/reply/DM.
  - 0: Tweet was not explicitly flagged as a retweet/reply/DM.

- Links to Original Tweet
  - 1: Tweet contains metadata which identifies the originating tweet.
  - 0: Tweet does not contain metadata which identifies the originating tweet.

- Tweet Pushed to: All or Some Users
  - 11: Tweet was pushed to all followers.
  - 01: Tweet was pushed to some followers but not all.

Appendix D

Matrix of Invalid Tweets

The following rows in Table D.1 show the binary values for permutations which cannot exist within a dataset of valid tweets. These invalid tweet groups were excluded from the typology of tweet propagation types, as described in Chapter 4: Typology.

Table D.1: Matrix of tweet propagation

Categories	Proprietary	Mechanism			Explicit	F/nF	Link to	Mentions	Push	
		Rt	@	DM			original	other users	All	Some
Replies without mentions	1	0	1	0	1	1	1	0	1	1
	1	0	1	0	1	1	1	0	0	1
	1	0	1	0	1	0	1	0	1	1
	1	0	1	0	1	0	1	0	0	1
	0	0	1	0	1	1	0	0	1	1
	0	0	1	0	1	1	0	0	0	1
Overlapping mechanisms	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	0	1	1	1
	1	1	1	1	0	1	1	1	1	1
	1	1	1	1	0	1	0	1	1	1
	1	1	1	1	1	1	1	0	1	1
	1	1	1	1	1	1	0	0	1	1
	1	1	1	1	0	1	1	0	1	1
	1	1	1	1	0	1	0	0	1	1
	1	1	1	1	1	0	1	1	1	1
	1	1	1	1	1	0	0	1	1	1
	1	1	1	1	0	0	1	1	1	1
	1	1	1	1	0	0	0	1	1	1
	1	1	1	1	1	0	1	0	1	1
	1	1	1	1	1	0	0	0	1	1
	1	1	1	1	0	0	1	0	1	1
	1	1	1	1	0	0	0	0	1	1
	1	1	1	0	1	1	1	1	1	0
	1	1	1	0	1	1	0	1	0	1
	1	1	1	0	1	1	0	1	1	1
	1	1	1	0	0	1	1	1	1	1
	1	1	1	0	0	1	0	1	1	1
	1	1	1	0	1	1	1	0	1	1
	1	1	1	0	1	1	0	0	1	1
	1	1	1	0	0	1	1	0	1	1
	1	1	1	0	0	1	0	0	1	1
	1	1	1	0	1	0	1	1	1	0
	1	1	1	0	1	0	0	1	0	1
	1	1	1	0	1	0	0	1	1	1
	1	1	1	0	0	0	1	1	1	1
	1	1	1	0	0	0	0	1	1	1
	1	1	1	0	1	0	1	0	1	1
	1	1	1	0	1	0	0	0	1	1
	1	1	1	0	0	0	1	0	1	1
	1	1	1	0	0	0	0	0	1	1
	1	1	0	1	1	1	1	1	1	1
	1	1	0	1	1	1	0	1	1	1

Continued on next page

Table D.1 – continued from previous page

Categories	Proprietary	Mechanism			Explicit	F/nF	Link to original	Mentions other users	Push	
		Rt	@	DM					All	Some
Overlapping mechanisms (continued)	1	1	0	1	0	1	1	1	1	1
	1	1	0	1	0	1	0	1	1	1
	1	1	0	1	1	1	1	0	1	1
	1	1	0	1	1	1	0	0	1	1
	1	1	0	1	0	1	1	0	1	1
	1	1	0	1	0	1	0	0	1	1
	1	1	0	1	1	0	1	1	1	1
	1	1	0	1	1	0	0	1	1	1
	1	1	0	1	0	0	1	1	1	1
	1	1	0	1	0	0	0	1	1	1
	1	1	0	1	1	0	1	0	1	1
	1	1	0	1	1	0	0	0	1	1
	1	1	0	1	0	0	1	0	1	1
	1	1	0	1	0	0	0	0	1	1
	1	0	1	1	1	1	1	1	1	1
	1	0	1	1	1	1	0	1	1	1
	1	0	1	1	0	1	1	1	1	1
	1	0	1	1	0	1	0	1	1	1
	1	0	1	1	1	1	1	0	1	1
	1	0	1	1	1	1	0	0	1	1
	1	0	1	1	0	1	1	0	1	1
	1	0	1	1	0	1	0	0	1	1
	1	0	1	1	1	0	1	1	1	1
	1	0	1	1	1	0	0	1	1	1
	1	0	1	1	0	0	1	1	1	1
	1	0	1	1	0	0	0	1	1	1
	1	0	1	1	1	0	1	0	1	1
	1	0	1	1	1	0	0	0	1	1
	1	0	1	1	0	0	1	0	1	1
	1	0	1	1	0	0	0	0	1	1
	0	1	1	1	1	1	1	1	1	1
	0	1	1	1	1	1	0	1	1	1
	0	1	1	1	0	1	1	1	1	1
	0	1	1	1	0	1	0	1	1	1
	0	1	1	1	1	1	1	0	1	1
	0	1	1	1	1	1	0	0	1	1
	0	1	1	1	0	1	1	0	1	1
	0	1	1	1	0	1	0	0	1	1
	0	1	1	1	1	0	1	1	1	1
	0	1	1	1	1	0	0	1	1	1
	0	1	1	1	0	0	1	1	1	1
	0	1	1	1	0	0	0	1	1	1
	0	1	1	1	1	0	1	0	1	1
	0	1	1	1	1	0	0	0	1	1
	0	1	1	1	0	0	1	0	1	1
	0	1	1	1	0	0	0	0	1	1
	0	1	1	0	1	1	1	1	1	1
	0	1	1	0	1	1	0	1	1	0
	0	1	1	0	1	1	1	1	0	1
	0	1	1	0	0	1	1	1	1	1
	0	1	1	0	0	1	0	1	1	1
	0	1	1	0	1	1	1	0	1	1
	0	1	1	0	1	1	0	0	1	1
	0	1	1	0	0	1	1	0	1	1
	0	1	1	0	0	1	0	0	1	1
	0	1	1	0	1	0	1	1	1	1
	0	1	1	0	1	0	0	1	1	0
	0	1	1	0	1	0	1	1	0	1
	0	1	1	0	0	0	1	1	1	1
	0	1	1	0	0	0	0	1	1	1
	0	1	1	0	1	0	1	0	1	1
	0	1	1	0	1	0	0	0	1	0
	0	1	1	0	1	0	1	0	0	1
	0	1	1	0	0	0	1	0	1	1
	0	1	1	0	0	0	0	0	1	1
	0	1	0	1	1	1	1	1	1	1
	0	1	0	1	1	1	0	1	1	1
	0	1	0	1	0	1	1	1	1	1
	0	1	0	1	0	1	0	1	1	1
	0	1	0	1	1	1	1	0	1	1
	0	1	0	1	1	1	0	0	1	1
	0	1	0	1	0	1	1	0	1	1
	0	1	0	1	0	1	0	0	1	1
	0	1	0	1	1	0	1	1	1	1
	0	1	0	1	1	0	0	1	1	1
	0	1	0	1	0	0	1	1	1	1

Continued on next page

Table D.1 – continued from previous page

Categories	Proprietary	Mechanism			Explicit	F/nF	Link to original	Mentions other users	Push	
		Rt	@	DM					All	Some
Overlapping mechanisms (continued)	0	1	0	1	0	0	0	1	1	1
	0	1	0	1	1	0	1	0	1	1
	0	1	0	1	1	0	0	0	1	1
	0	1	0	1	0	0	1	0	1	1
	0	1	0	1	0	0	0	0	1	1
	0	0	1	1	1	1	1	1	1	1
	0	0	1	1	1	1	0	1	1	1
	0	0	1	1	0	1	1	1	1	1
	0	0	1	1	0	1	0	1	1	1
	0	0	1	1	1	1	1	0	1	1
	0	0	1	1	1	1	0	0	1	1
	0	0	1	1	0	1	1	0	1	1
	0	0	1	1	0	1	0	0	1	1
	0	0	1	1	1	0	1	1	1	1
	0	0	1	1	1	0	0	1	1	1
	0	0	1	1	0	0	1	1	1	1
	0	0	1	1	0	0	0	1	1	1
	0	0	1	1	1	0	1	0	1	1
	0	0	1	1	1	0	0	0	1	1
	0	0	1	1	0	0	1	0	1	1
	0	0	1	1	0	0	0	0	1	1
No mechanism but explicitly propagating	1	0	0	0	1	1	1	1	1	1
	1	0	0	0	1	1	0	1	1	1
	1	0	0	0	1	1	1	0	1	1
	1	0	0	0	1	1	0	0	1	1
	1	0	0	0	1	0	1	1	1	1
	1	0	0	0	1	0	0	1	1	1
	1	0	0	0	1	0	1	0	1	1
	1	0	0	0	1	0	0	0	1	1
	0	0	0	0	1	1	1	1	1	1
	0	0	0	0	1	1	0	1	1	1
	0	0	0	0	1	1	1	0	1	1
	0	0	0	0	1	1	0	0	1	1
	0	0	0	0	1	0	1	1	1	1
	0	0	0	0	1	0	0	1	1	1
	0	0	0	0	1	0	1	0	1	1
	0	0	0	0	1	0	0	0	1	1
No mechanism but links to original tweet	1	0	0	0	0	1	1	1	1	1
	1	0	0	0	0	1	1	0	1	1
	1	0	0	0	0	0	1	1	1	1
	1	0	0	0	0	0	1	0	1	1
	0	0	0	0	0	1	1	1	1	1
	0	0	0	0	0	1	1	0	1	1
	0	0	0	0	0	0	1	1	1	1
	0	0	0	0	0	0	1	0	1	1
No mentions but is a follower of unknown source	1	0	0	0	0	1	0	0	1	1
	0	0	0	0	0	1	0	0	1	1
Original tweets pushed to all not some	1	0	0	0	0	0	0	0	1	0
	0	0	0	0	0	0	0	0	1	0
Original tweets pushed to some but not all	1	0	0	0	0	0	0	0	0	1
	0	0	0	0	0	0	0	0	0	1
	1	0	0	0	0	1	0	1	0	1
	1	0	0	0	0	0	0	1	0	1
	0	0	0	0	0	1	0	1	0	1
	0	0	0	0	0	0	0	1	0	1
Original tweets not pushed to anyone	1	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	1	0	1	0	0
	1	0	0	0	0	0	0	1	0	0
	0	0	0	0	0	1	0	1	0	0
	0	0	0	0	0	0	0	1	0	0
Non-proprietary DM	0	0	0	1	1	1	1	1	1	1
	0	0	0	1	1	1	0	1	1	1
	0	0	0	1	0	1	1	1	1	1
	0	0	0	1	0	1	0	1	1	1
	0	0	0	1	1	1	1	0	1	1
	0	0	0	1	1	1	0	0	1	1
	0	0	0	1	0	1	1	0	1	1
	0	0	0	1	0	1	0	0	1	1
	0	0	0	1	1	0	1	1	1	1
	0	0	0	1	1	0	0	1	1	1
	0	0	0	1	0	0	1	1	1	1
	0	0	0	1	0	0	0	1	1	1
	0	0	0	1	1	0	1	0	1	1
	0	0	0	1	1	0	0	0	1	1
	0	0	0	1	0	0	1	0	1	1

Continued on next page

Table D.1 – continued from previous page

Categories	Proprietary	Mechanism			Explicit	F/nF	Link to	Mentions	Push	
		Rt	@	DM			original	other users	All	Some
	0	0	0	1	0	0	0	0	1	1
Proprietary Rt without attribution	1	1	0	0	1	1	1	0	1	1
	1	1	0	0	1	1	0	0	1	1
Implicit proprietary Rt	1	1	0	0	0	1	1	0	1	1
	1	1	0	0	0	1	0	0	1	1
Implicit proprietary replies	1	0	1	0	0	1	1	1	1	0
	1	0	1	0	0	1	0	1	1	0
	1	0	1	0	0	1	1	0	1	0
	1	0	1	0	0	1	0	0	1	0
	1	0	1	0	0	0	1	1	1	0
	1	0	1	0	0	0	0	1	1	0
	1	0	1	0	0	1	1	1	1	0
	1	0	1	0	0	1	0	1	1	0
Implicit non-proprietary replies	0	0	1	0	0	1	1	1	1	0
	0	0	1	0	0	1	0	1	1	0
	0	0	1	0	0	1	1	0	1	0
	0	0	1	0	0	1	0	0	1	0
	0	0	1	0	0	0	1	1	1	0
	0	0	1	0	0	0	0	1	1	0
	0	0	1	0	0	0	1	0	1	0
	0	0	1	0	0	0	0	0	1	0
Implicit proprietary DMs	1	0	0	1	0	1	1	1	1	0
	1	0	0	1	0	1	0	1	1	0
	1	0	0	1	0	1	1	0	1	0
	1	0	0	1	0	1	0	0	1	0
	1	0	0	1	0	0	1	1	1	0
	1	0	0	1	0	0	0	1	1	0
	1	0	0	1	0	1	1	1	1	0
	1	0	0	1	0	1	0	1	1	0
Non-follower DMs	1	0	0	1	1	0	1	1	1	0
	1	0	0	1	1	0	0	1	1	0
	1	0	0	1	1	0	1	0	1	0
	1	0	0	1	1	0	0	0	1	0
No audience	1	1	0	0	1	0	1	1	0	0
	0	1	0	0	1	1	0	1	0	0
	0	1	0	0	0	1	0	1	0	0
	0	1	0	0	1	1	0	0	0	0
	0	1	0	0	0	1	0	0	0	0
	0	1	0	0	1	0	0	1	0	0
	0	1	0	0	0	0	0	1	0	0
	0	1	0	0	1	0	0	0	0	0
	0	1	0	0	0	0	0	0	0	0
	1	0	1	0	1	1	0	1	0	0
	1	0	1	0	1	1	0	0	0	0
	1	0	1	0	1	0	0	1	0	0
	1	0	1	0	1	0	0	0	0	0
	0	0	1	0	1	1	0	1	0	0
	0	0	1	0	1	1	0	0	0	0
	0	0	1	0	1	0	0	1	0	0
	0	0	1	0	1	0	0	0	0	0
	1	0	0	1	1	1	0	1	0	0
	1	0	0	1	1	1	0	0	0	0
Proprietary Rt without link to original tweet	1	1	0	0	1	1	0	1	1	1
	1	1	0	0	1	1	0	1	1	0
	1	1	0	0	1	1	0	1	0	1
	1	1	0	0	1	1	0	1	0	0
	1	1	0	0	1	0	0	1	1	1
	1	1	0	0	1	0	0	1	1	0
	1	1	0	0	1	0	0	1	0	1
	1	1	0	0	1	0	0	1	0	0
Proprietary Rt not seen by all	1	1	0	0	1	1	1	1	0	1
	1	1	0	0	1	0	1	1	0	1
Seen by all but not some	1	1	0	0	1	1	1	1	1	0
	1	1	0	0	1	0	1	1	1	0
	0	1	0	0	1	1	0	1	1	0
	0	1	0	0	0	1	0	1	1	0
	0	1	0	0	1	1	0	0	1	0
	0	1	0	0	0	1	0	0	1	0
	0	1	0	0	0	1	0	0	1	0
	0	1	0	0	1	0	0	1	1	0
	0	1	0	0	0	0	0	1	1	0
	0	1	0	0	1	0	0	0	1	0
	0	1	0	0	0	0	0	0	1	0
	1	0	1	0	1	1	0	1	1	0
	1	0	1	0	1	1	0	0	1	0
	1	0	1	0	1	0	0	1	1	0

Continued on next page

Table D.1 – continued from previous page

Categories	Proprietary	Mechanism			Explicit	F/nF	Link to original	Mentions other users	Push	
		Rt	@	DM					All	Some
Seen by all but not some (continued)	1	0	1	0	1	0	0	0	1	0
	0	0	1	0	1	1	0	1	1	0
	0	0	1	0	1	1	0	0	1	0
	0	0	1	0	1	0	0	1	1	0
	0	0	1	0	1	0	0	0	1	0
	1	0	0	1	1	1	0	1	1	0
	1	0	0	1	1	1	0	0	1	0
	1	0	0	0	0	1	0	1	1	0
	1	0	0	0	0	0	0	1	1	0
	0	0	0	0	0	1	0	1	1	0
	0	0	0	0	0	0	0	1	1	0
Non-proprietary Rt seen by some but not all	0	1	0	0	1	1	1	1	0	1
	0	1	0	0	1	1	0	1	0	1
	0	1	0	0	0	1	0	1	0	1
	0	1	0	0	1	1	0	0	0	1
	0	1	0	0	0	1	0	0	0	1
	0	1	0	0	1	0	0	1	0	1
	0	1	0	0	0	0	0	1	0	1
	0	1	0	0	1	0	0	0	0	1
	0	1	0	0	0	0	0	0	0	1
DMs seen by all	1	0	0	1	1	1	1	1	1	1
	1	0	0	1	1	1	0	1	1	1
	1	0	0	1	1	1	1	0	1	1
	1	0	0	1	1	1	0	0	1	1
Non-proprietary Rt with link to original tweet	0	1	0	0	0	1	1	1	1	1
	0	1	0	0	1	1	1	0	1	1
	0	1	0	0	0	1	1	0	1	1
	0	1	0	0	0	0	1	1	1	1
	0	1	0	0	1	0	1	0	1	1
	0	1	0	0	0	0	1	0	1	1
Proprietary reply without link to original tweet	1	0	1	0	1	1	0	1	1	1
	1	0	1	0	1	1	0	1	0	1
	1	0	1	0	1	1	0	0	1	1
	1	0	1	0	1	1	0	0	0	1
	1	0	1	0	1	0	0	1	1	1
	1	0	1	0	1	0	0	1	0	1
	1	0	1	0	1	0	0	0	1	1
	1	0	1	0	1	0	0	0	0	1
Non-proprietary reply with link to original tweet	0	0	1	0	1	1	1	1	1	1
	0	0	1	0	1	1	1	1	0	1
	0	0	1	0	1	1	1	0	1	1
	0	0	1	0	1	1	1	0	0	1
	0	0	1	0	1	0	1	1	1	1
	0	0	1	0	1	0	1	1	0	1
	0	0	1	0	1	0	1	0	1	1
	0	0	1	0	1	0	1	0	0	1
DMs with link to originating tweet	1	0	0	1	1	1	1	1	0	1
	1	0	0	1	1	1	1	0	0	1

References

Adar, E., Zhang, L., Adamic, L. A., and Lukose, R. M. (2004). Implicit structure and the dynamics of blogspace. In *Workshop on the Weblogging Ecosystem, WWW2004*, New York, NY.

Agarwal, N., Liu, H., Tang, L., and Yu, P. S. (2008). Identifying the influential bloggers in a community. In *Proceedings of the international conference on Web search and web data mining*, pages 207–218, Palo Alto, California, USA. ACM.

Aizen, J., Huttonlocher, D., Kleinberg, J., and Novak, A. (2004). Traffic-based feedback on the web. *Proceedings of the National Academy of Sciences*, 101(suppl_1):5254–5260.

Allsopp, G. (2010). Bloggers: This is how long your posts should be. [internet]. Available at: http://www.viperchill.com/blog-post-length/ [Accessed June 2014].

Anger, I. and Kittl, C. (2011). Measuring influence on twitter. In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, i-KNOW '11, pages 31:1–31:4, New York, NY, USA. ACM.

Backstrom, L., Huttenlocher, D., Kleinberg, J., and Lan, X. (2006). Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 44–54, Philadelphia, PA, USA. ACM.

Bakshy, E., Hofman, J. M., Mason, W. A., and Watts, D. J. (2011). Identifying 'influencers' on Twitter. In *Fourth ACM International Conference on Web Search and Data Mining*, Hong Kong, CN. ACM.

Bandari, R., Asur, S., and Huberman, B. (2012). The pulse of news in social media: Forecasting popularity. AAAI.

Bastos, M., Travitzki, R., and Puschmann, C. (2012). What sticks with whom? twitter follower-followee networks and news classification. Dublin, IE. AAAI.

Blood, R. (2004). How blogging software reshapes the online community. *Comms. ACM*, 47(12):53–55.

Boyd, D., Golder, S., and Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *Hawaii International Conference on System Sciences*, volume 0, pages 1–10, Los Alamitos, CA, USA. IEEE Computer Society.

Bragg, M. (2009). *In Our Time – A Modest Proposal.* Radio podcast, *BBC*, [internet] 29 January. Available at: http://www.bbc.co.uk/programmes/b00h3650 [Accessed June 2014].

Bricklin, D. (2001). Pamphleteers and web sites. [internet]. Available at: http://www.bricklin.com/pamphleteers.htm [Accessed June 2014].

Castells, M. (2009). *Communication Power.* OUP Oxford.

Cha, M., Haddadi, H., Benevenuto, F., and Gunmadi, K. P. (2010). Measuring user influence in twitter: The million follower fallacy. In *International AAAI Conference on Weblogs and Social Media Fourth International AAAI Conference on Weblogs and Social Media*, Washington, DC, USA. AAAI.

Choudhary, A., Hendrix, W., Lee, K., Palsetia, D., and Liao, W.-K. (2012). Social media evolution of the egyptian revolution. *Commun. ACM*, 55(5):74–80.

Choudhury, M. D., Sundaram, H., John, A., and Seligmann, D. D. (2009). What makes conversations interesting?: themes, participants and consequences of conversations in online social media. In *Proceedings of the 18th international conference on World wide web*, pages 331–340, Madrid, ES. ACM.

Conover, M., Ratkiewicz, J., Francisco, M., Goncalves, B., Menczer, F., and Flammini, A. (2011). Political polarization on twitter. Barcelona, ES. AAAI.

Farrell, H. and Drezner, D. (2008). The power and politics of blogs. *Public Choice*, 134(1):15–30.

Fox, S., Zickuhr, K., and Smith, A. (2009). Twitter and status updating, fall 2009 | pew research center's internet & american life project. [internet]. Available at: http://www.pewinternet.org/Reports/2009/17-Twitter-and-Status-Updating-Fall-2009.aspx [Accessed June 2014].

Fujiki, S., Yano, H., Fukuda, T., and Yamana, H. (2011). Retweet reputation: A bias-free evaluation method for tweeted contents. Barcelona, ES. AAAI.

Gabielkov, M., Rao, A., and Legout, A. (2014). Studying social networks at scale: Macroscopic anatomy of the twitter social graph. In *The 2014 ACM International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '14, pages 277–288, New York, NY, USA. ACM.

Gabrilovich, E., Dumais, S., and Horvitz, E. (2004). Newsjunkie: Providing personalized newsfeeds via analysis of information novelty. In *In WWW2004*, pages 482—490, New York, NY, USA.

Galuba, W., Aberer, K., Chakraborty, D., Despotovic, Z., and Kellerer, W. (2010a). Outtweeting the twitterers- predicting information cascades in microblogs. In *3rd Workshop on Online Social Networks*, Boston, MA, USA. USENIX.

Galuba, W., Aberer, K., Chakraborty, D., Despotovic, Z., and Kellerer, W. (2010b). Outtweeting the twitterers- predicting information cascades in microblogs. In *3rd Workshop on Online Social Networks*, Boston, MA, USA. USENIX.

Gill, K. (2004). How can we measure the influence of the blogosphere? In *Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics*, New York, NY, USA.

Gladwell, M. (2001). *The Tipping Point: How Little Things Can Make a Big Difference*. Abacus, new ed edition.

Goffman, E. (1981). *Forms of Talk*. University of Pennsylvania Press.

Gruhl, D., Guha, R., Liben-Nowell, D., and Tomkins, A. (2004). Information diffusion through blogspace. In *Proceedings of the 13th international conference on World Wide Web*, pages 491–501, New York, NY, USA. ACM.

Guardian, T. (2009). Jan Moir responds to criticism of her Daily Mail article on Stephen Gately. *The Guardian*, [internet] 16 October. Available at: http://www.guardian.co.uk/media/2009/oct/16/jan-moir-stephen-gately-response [Accessed June 2014].

Hall, S. (2013). The power of marketing personalization in President Obama's re-election. [internet]. Available at: http://blog.hubspot.com/blog/tabid/6307/bid/34268/The-Power-of-Marketing-Personalization-in-President-Obama-s-Re-Election.aspx [Accessed June 2014].

Hazim Almuhimedi, Shomir Wilson, B. L. N. S. and Acquisti, A. (2013). Tweets are forever: a large-scale quantitative analysis of deleted tweets. In *Proceedings of the ACM 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, pages 897–908, New York, NY, USA. ACM.

Hoang, T.-A., Lim, E.-P., Achananuparp, P., Jiang, J., and Zhu, F. (2011). On modeling virality of twitter content. In Xing, C., Crestani, F., and Rauber, A., editors, *Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation*, volume 7008 of *Lecture Notes in Computer Science*, pages 212–221. Springer Berlin Heidelberg.

Hong, L., Dan, O., and Davison, B. D. (2011). Predicting popular messages in twitter. In *Proceedings of the 20th international conference companion on World wide web*, WWW '11, pages 57–58, New York, NY, USA. ACM.

Huberman, B. (2008). Crowdsourcing and attention. *Computer*, 41(11):105, 103.

Ives, B. (2008). How Barack Obama is using the Web to further engage voters. *EContent*, 31(5):12–13.

Kelley, P. G. and Cranshaw, J. (2011). Conducting research on Twitter: A call for guidelines and metrics. [internet]. Available at: http://patrickgagekelley.com/papers/twitter-pmj.pdf [Accessed June 2014].

Klau, R. (2003). Rick Klau's weblog: Topical, polemical, and short. [internet]. Available at: http://tins.rklau.com/2003/04/topical-polemical-and-short.html [Accessed June 2014].

Klau, R. (2009). Rick Klau's weblog: Twitter: Topical, polemical and short. [internet]. Available at: http://tins.rklau.com/2009/03/twitter-topical-polemical-and-short.html [Accessed June 2014].

Kleinberg, J. (2008). The convergence of social and technological networks. *Comms. ACM*, 51(11):66–72.

Kleinberg, J., Nisan, N., Roughgarden, T., Tardos, E., and Vazirani, V. (2007). Cascading behavior in networks: Algorithmic and economic issues. In *Algorithmic Game Theory*, pages 613—632. Cambridge University Press.

Kong, H., Park, H., and Han, S. (2009). The talkative, the popular, and the influential korean twitter users. In *Social Networks Interoperability 1st International Workshop 2009*, Shanghai, CN.

Kooti, F., Yang, H., Cha, M., Gummadi, K., and Mason, W. (2012). The emergence of conventions in online social networks. Dublin, IE. AAAI.

Krauss, J., Nann, S., Simon, D., Fischbach, K., and Gloor, P. (2008). Predicting movie success and academy awards through sentiment and social network analysis. In *Proc. European Conference on Information Systems (ECIS)*, Galway, IE.

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

Lenhart, A., Purcell, K., Smith, A., and Zickuhr, K. (2010). Social media and young adults | pew research center's internet & american life project. [internet]. Available at: http://pewinternet.org/Reports/2010/Social-Media-and-Young-Adults/Part-3/4-Twitter-among-teens-and-adults.aspx [Accessed June 2014].

Lerman, K. and Ghosh, R. (2010). Information contagion: An empirical study of the spread of news on digg and twitter social networks. Washington, DC, USA. AAAI.

Leskovec, J., Backstrom, L., and Kleinberg, J. (2009). Meme-tracking and the dynamics of the news cycle. In *Proc. 15th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, 2009*, Paris, France. ACM.

Leskovec, J., Mcglohon, M., Faloutsos, C., Glance, N., and Hurst, M. (2007). Cascading behavior in large blog graphs. In *SIAM International Conference on Data Mining*, Minneapolis, MN, USA.

Lukes, S. (2005). *Power: A Radical View*. Palgrave Macmillan, 2nd edition.

Lumezanu, C., Feamster, N., and Klein, H. (2012). #bias: Measuring the tweeting behavior of propagandists. Dublin, IE. AAAI.

MacKinnon, R. (2008). Flatter world and thicker walls? blogs, censorship and civic discourse in china. *Public Choice*, 134(1):31–46.

Macskassy, S. and Michelson, M. (2011). Why do people retweet? anti-homophily wins the day! Barcelona, ES. AAAI.

Mansour, E. (2012). The role of social networking sites (SNSs) in the january 25th revolution in egypt. *Library Review*, 61(2):128–159.

Mari, A. (2010). Social media will not be integral to 2010 elections. [internet]. Available at: http://www.computing.co.uk/computing/news/2260462/social-media-integral-2010 [Accessed June 2014].

Market Sentinel (2005). Measuring the influence of bloggers on corporate reputation. White paper.

Marlow, C. (2004). Audience, structure and authority in the weblog community. In *In International Communication Association Conference*, New Orleans, LA, USA.

Matsumura, N., Yamamoto, H., and Tomozawa, D. (2010a). Finding influencers and consumer insights in the blogosphere. In *International AAAI Conference on Weblogs and Social Media Fourth International AAAI Conference on Weblogs and Social Media*, Washington, DC, USA. AAAI.

Matsumura, N., Yamamoto, H., and Tomozawa, D. (2010b). Finding influencers and consumer insights in the blogosphere. In *International AAAI Conference on Weblogs and Social Media Fourth International AAAI Conference on Weblogs and Social Media*, Washington, DC, USA. AAAI.

McKenna, L. and Pole, A. (2008). What do bloggers do: an average day on an average political blog. *Public Choice*, 134(1):97–108.

Meccawy, M. (2008). Bloggers and the emergence of a new tribalism in Saudi Arabia: An insider's experience. [internet]. Available at: http://meccawy.com/site/wp-content/uploads/2008/06/media_conf_paper_v3.pdf [Accessed June 2014].

Merholz, P. (1999). Peterme.com. [internet]. Available at: http://web.archive.org/web/19991013021124/http://peterme.com/index.html [Accessed June 2014].

Munger, M. (2008). Blogging and political information: truth or truthiness? *Public Choice*, 134(1):125–138.

Murthy, D. (2012). Towards a sociological understanding of social media: Theorizing twitter. *Sociology*.

Mustafaraj, E. and Metaxas, P. (2010). From obscurity to prominence in minutes: Political speech and Real-Time search. In *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, Raleigh, NC, US.

Myers, S. A., Zhu, C., and Leskovec, J. (2012). Information diffusion and external influence in networks. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '12, pages 33–41, New York, NY, USA. ACM.

Nagarajan, M., Purohit, H., and Sheth, A. (2010). A qualitative examination of topical tweet and retweet practices. In *International AAAI Conference on Weblogs and Social Media Fourth International AAAI Conference on Weblogs and Social Media*, Washington, DC, USA. AAAI.

Oxford University Press (2013). Oxford dictionaries. [internet]. Available at: http://oxforddictionaries.com/definition/english/retweet [Accessed June 2014].

Packer, G. (2004). The revolution will not be blogged. *Mother Jones*, (May/June).

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab. Previous number = SIDL-WP-1999-0120.

Peng, H.-K., Zhu, J., Piao, D., Yan, R., and Zhang, Y. (2011). Retweet modeling using conditional random fields. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 336–343.

Petrovic, S., Osborne, M., and Lavrenko, V. (2011). Rt to win! predicting message propagation in twitter. Barcelona, ES. AAAI.

Pfitzner, R., Garas, A., and Schweitzer, F. (2012). Emotional divergence influences information spreading in twitter. Dublin, IE. AAAI.

Recuero, R. (2008). Information flows and social capital in weblogs: a case study in the brazilian blogosphere. In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, pages 97–106, Pittsburgh, PA, USA. ACM.

Rivers, C. M. and Lewis, B. L. (2014). Ethical research standards in a world of big data. [internet]. Available at: http://f1000research.com/articles/3-38/v1/pdf [Accessed June 2014].

Rogers, E. (1962). *Diffusion of innovations*. Free Press of Glencoe.

Ruan, Y., Purohit, H., Fuhry, D., Parthasarthy, S., and Sheth, A. (2012). Prediction of topic volume on twitter. In *WebSci'12*, Evanston, Illinois, USA.

Russell, B. (1975). *Power: A New Social Analysis*. Routledge, new ed edition.

Scherer, M. (2012). Exclusive: Obama's 2012 digital fundraising outperformed 2008. [internet]. Available at: http://swampland.time.com/2012/11/15/exclusive-obamas-2012-digital-fundraising-outperformed-2008/ [Accessed June 2014].

Song, X., Chi, Y., Hino, K., and Tseng, B. (2007a). Identifying opinion leaders in the blogosphere. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 971–974, Lisbon, Portugal. ACM.

Song, X., Chi, Y., Hino, K., and Tseng, B. L. (2007b). Information flow modeling based on diffusion rate for prediction and ranking. In *Proceedings of the 16th international conference on World Wide Web*, pages 191–200, Banff, Alberta, Canada. ACM.

Starbird, K. and Palen, L. (2012). (how) will the revolution be retweeted?: information diffusion and the 2011 egyptian uprising. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, pages 7–16, New York, NY, USA. ACM.

Stieglitz, S. and Dang-Xuan, L. (2012). Political communication and influence through microblogging–an empirical analysis of sentiment in twitter messages and retweet behavior. In *System Science (HICSS), 2012 45th Hawaii International Conference on*, pages 3500–3509.

Suh, B., Hong, L., Pirolli, P., and Chi, E. H. (2010). Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In *IEEE Second International Conference on Social Computing*, pages 177–184, Minneapolis, MN, USA. IEEE.

Sullenger, D. W. (2006). Silencing the blogosphere: A first amendment caution to legislators considering using blogs to communicate directly with constituents. *Richmond Journal of Law & Technology*, 13:1.

Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton University Press.

Tao, K., Abel, F., Hauff, C., Houben, G.-J., and Gadiraju, U. (2013). Groundhog day: Near-duplicate detection on twitter. In *Proceedings of the 22nd international conference on World Wide Web*, WWW '13, New York, NY, USA. ACM.

The Economist (2006). It's the links, stupid. *The Economist*.

Thomson, K. (2012). Ethics of Twitter research. In *Proceedings of the Interdisciplinary Research Ethics Workshop*, Birmingham, UK.

Tinati, R., Carr, L., Hall, W., and Bentwood, J. (2012). Identifying communicator roles in twitter. In *Proceedings of the 21st international conference companion on World Wide Web*, WWW '12 Companion, pages 1161–1168, New York, NY, USA. ACM.

Tweetminster (2010). Can word-of-mouth predict the general election result? A Tweetminster experiment in predictive modelling. [internet]. Available at: http://www.scribd.com/doc/29154537/Tweetminster-Predicts [Accessed June 2014].

Twitter (2013). Celebrating #twitter7. weblog post, [internet]. Available at: http://blog.twitter.com/2013/03/celebrating-twitter7.html [Accessed June 2014].

van Liere, D. (2010). How far does a tweet travel?: Information brokers in the twitterverse. In *Proceedings of the International Workshop on Modeling Social Media*, pages 1–4, Toronto, Ontario, CA. ACM.

Walker, M. (2008). The year of the insurgents: the 2008 US presidential campaign. *International Affairs*, 84(6):1095–1107.

Webberley, W., Allen, S., and Whitaker, R. (2011). Retweeting: A study of message-forwarding in twitter. In *Mobile and Online Social Networks (MOSN), 2011 Workshop on*, pages 13–18.

Weng, J., Lim, E., Jiang, J., and He, Q. (2010). TwitterRank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 261–270, New York, NY, USA. ACM.

Westen, D. (2008). *The Political Brain The Role Of Emotion In Deciding The Fate Of The Nation*. PublicAffairs, reprint edition.

Wiederhold, G. (1995). Digital libraries, value, and productivity. *Comms. ACM*, 38(4):85–96.

Woodly, D. (2008). New competencies in democratic communication? blogs, agenda setting and political participation. *Public Choice*, 134(1):109–123.

Wu, F. and Huberman, B. A. (2007). Novelty and collective attention. *Proceedings of the National Academy of Sciences*, 104(45):17599–17601.

Wu, S., Hofman, J. M., Mason, W. A., and Watts, D. J. (2011). Who says what to whom on Twitter. In *Proceedings of the 20th international conference on World wide web (WWW '11)*, Hyderabad, IN. ACM.

Yang, J. and Counts, S. (2010). Predicting the speed, scale, and range of information diffusion in twitter. Washington, DC, USA. AAAI.

Yang, L., Sun, T., Zhang, M., and Mei, Q. (2012). We know what @you #tag: does the dual role affect hashtag adoption? In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, pages 261–270, New York, NY, USA. ACM.

Yang, Z., Guo, J., Cai, K., Tang, J., Li, J., Zhang, L., and Su, Z. (2010). Understanding retweeting behaviors in social networks. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, CIKM '10, pages 1633–1636, New York, NY, USA. ACM.

Yette, L. (2012). A call to action: Twitter's power to mobilize during the arab spring. [Masters thesis]. Available at: http://www.american.edu/soc/communication/upload/Laila-Yette.pdf [Accessed June 2014].

Zaman, T. R., Herbrich, R., Gael, J. V., and Stern, D. (2010). Predicting information spreading in twitter. Whistler, CA.

Zarella, D. (2009). The science of ReTweets: viral content sharing on twitter. [presentation]. Available at: http://www.hubspot.com/Portals/53/docs/science-of-retweets-201003.pdf [Accessed June 2014].

Zuckerman, E. (2008). Meet the bridgebloggers. *Public Choice*, 134(1):47–65.