

Dark Retweets: Investigating Non-Conventional Retweeting Patterns

Norhidayah Azman, David E. Millard, and Mark J. Weal

Web and Internet Science, Department of Electronics and Computer Science,
University of Southampton, UK
{nba08r,dem,mjw}@ecs.soton.ac.uk

Abstract. Retweets are an important mechanism for recognising 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 analysis. In this paper, a comprehensive matrix of tweet propagation is presented to show the different nuances of retweeting, 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. Based on this matrix and two assumptions of retweetability, the degrees of a retweet’s “darkness” can be determined. This matrix was evaluated over 2.3 million tweets and it was found that dark retweets amounted to 12.86% (for search results less than 1500 tweets per URL) and 24.7% (for search results including more than 1500 tweets per URL) respectively. By extrapolating these results with those found in existing studies, potentially thousands of retweets may be hidden from existing studies on retweets.

Keywords: retweets, tweet propagation, dark retweets, Twitter, microblogs

1 Introduction

There have been several studies investigating the propagation of tweets, focusing on retweeting as the main mechanism of propagation. Existing work has traditionally looked at both conventional retweeting mechanisms, such as Twitter’s proprietary retweeting mechanism, and manually inserted retweet markers, such as “RT” and “via.”

Retweets form an important part of tweet propagation research, from conversational patterns [1] to overall retweet ratios [2, 3]. However, there are several different nuances to the act of propagating a tweet, mainly due to the different mechanisms and features involved in propagating a tweet. This paper aims to describe these different nuances of retweeting by deconstructing the action into separate characteristics. This paper also introduces “dark retweets”, describes

the different assumptions of retweetability, and also discusses the possible impact they may have on future tweet propagation studies. Our research shows that there are many different ways a user may propagate information across Twitter, and there may be proportions of propagating tweets which may have been missing from existing work.

Section 2 discusses several existing papers on retweets and in particular those which are not explicitly marked as retweets. Section 3 outlines the matrix of tweet propagation, 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 a retweet, if it links to an original tweet, and what is the audience it is pushed to. Section 4 focuses on dark retweets and the different assumptions of retweetability. The following Section 5 presents the evaluation of this matrix, from descriptions of the experimental setup to discussions of the results found.

2 Research Background

Since the introduction of Twitter and microblogs, researchers have been focusing on patterns of propagation across Twitter. *boyd et. al* [1] was one of the earliest studies focusing on retweets. Two datasets were used; one being a random sample of tweets taken in 5-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. This paper also acknowledges 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 this study.

Several researchers have tackled this problem by making assumptions about the existence of propagation paths 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 paper is most similar to the work done by *Adar et. al* [4], which investigated the existence of implicit URL links within the blogosphere. Similarly, the work by *Matsumura et. al* [5] 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* [6] described the *F-cascade* within tweets. It involved users who seemed to have copied a URL that was previously tweeted by someone they follow.

Similarly, in the work done by *Wu et. al* [7], 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* [8], who studied

influence prediction within Twitter. Their metrics indicating influence were not restricted to just retweeting by including retweet markers within 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 conventional and non-conventional retweets.

In the study by Nagarajan et. al [9], 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, 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.

Given the above studies, this paper presents a combination of the theory of implicit links as presented by Adar et. al [4] with the methodology of only using tweets containing URLs in them [7, 8]. The hypothesis of this paper is that there exists a proportion of dark retweets, or tweets which are propagated without using conventional retweeting mechanisms. This may lead to hidden data which would not have been focused upon within existing studies on tweet propagation.

3 Matrix of Tweet Propagation

In this study, tweets are deconstructed using several characteristics. In the following descriptions of these characteristics, the abbreviations stated in parentheses are used in the matrix of tweet propagation shown in Table 1.

Proprietary: The propagation of a tweet is considered proprietary (P) if it was published using methods that were built into the Twitter use case structure. 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 retweet (‘Rt’), a reply (‘@’), or a direct message (DM).

Follower or non-follower: The propagated tweets 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. In this column, a follower relationship is marked as ‘1’ in Table 1, while a non-follower relationship is marked as ‘0’.

Mentions other users: A mention exists in a tweet if its text contains other people’s Twitter usernames in them.

Explicit: A tweet is considered to be explicitly propagated by a user if a retweet marker such as ‘RT’ or the ‘@’ reply marker was written explicitly in the tweet text. Proprietary retweets/replies and manually marked retweets/replies are considered explicit, while those without any retweet/reply markers are considered as implicit. For example, “*Done! RT @User_X Sign 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.

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.

Tweet pushed to: all or some people: This denotes the difference between the visibility of a retweet and a reply. Retweets are pushed onto the timelines of all the followers of the retweeter. In Table 1, a ‘11’ value means the tweet is pushed to all and some of the authors’ followers. This visibility changes for replies addressed to a specific Twitter user. They are 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. This is marked as ‘01’ in Table 1.

Using these seven characteristics, a binary matrix was constructed to illustrate all possible combinations of these characteristics. This process resulted in a 2^{10} matrix, containing 1024 rows. 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. Table 1 shows the valid rows after this evaluation was completed. The abbreviations used under the Categories column in Table 1 come from the characteristics described in Section 3. The categories in Table 1 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). In Table 1, there are rows which are coloured in three shades of grey. The lightest shade of grey corresponds to Orphan Retweets and Replies, which will be described in Section 3.5. The rows which are coloured in the two darkest shades of grey correspond to Dark Retweets, which will be described in Section 4.

3.1 Original Tweets and Mentions

Table 1 shows 18 different groups. The base group is Original Tweets, which a) do not seem to have been made using any proprietary retweeting or replying

Table 1. Matrix of Tweet Propagation

Categories	Proprietary	Mechanism			Explicit	F/nF	Link to original	Mentions other users	Push	
		Rt	@	DM					All	Some
Original Tweet	0	0	0	0	0	0	0	0	0	0
PRtF	1	1	0	0	1	1	1	1	1	1
PRtnF	1	1	0	0	1	0	1	1	1	1
Rt@F	0	1	0	0	1	1	1	1	1	1
Rt@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	0	1	0	1	1	1
	0	1	0	0	1	1	0	0	1	1
	0	1	0	0	0	1	0	0	1	1
RtnF	0	1	0	0	1	0	0	1	1	1
	0	1	0	0	0	0	0	1	1	1
	0	1	0	0	1	0	0	0	1	1
	0	1	0	0	0	0	0	0	1	1
P@F	1	0	1	0	1	1	1	1	1	1
	1	0	1	0	1	1	1	1	0	1
P@nF	1	0	1	0	1	0	1	1	1	1
	1	0	1	0	1	0	1	1	0	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
@F	0	0	1	0	1	1	0	1	1	1
	0	0	1	0	1	1	0	1	0	1
@nF	0	0	1	0	1	0	0	1	1	1
	0	0	1	0	1	0	0	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
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

mechanisms, *b*) do not seem to be explicitly propagating another tweet, and *c*) therefore do not have any links to originating tweets nor users.

There also exists the Original Mentions group, which is similar to Original Tweets, but contains mentions in its texts. In this paper, the rows for Original Mentions were omitted from Table 1 for brevity, and Original Mentions were grouped together with Original Tweets, under the assumption that Original Mentions behave similarly to Original Tweets. This is because they also do not seem to have used any retweeting or replying mechanisms.

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 (contain metadata that links to the tweet that is being retweeted or replied to). Therefore these proprietary retweets and replies are marked in Table 1 with the value of 1 under the Explicit and Link to Original columns.

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

3.3 Multiple Mechanisms in Tweets

Several categories include two mechanisms, such as Rt@ and @Rt. 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 with 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, which is opposite to all proprietary Twitter retweets or replies.

However, there exists certain retweets which are not marked by the Twitter REST API as being made using Twitter’s proprietary retweet mechanism. Even so, they still contain metadata linking to originating tweets. On further inspection, these tweets were found to be retweets that were manually created after the proprietary replying 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 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 are not similar to a normal retweet. Section 3 has already discussed the visibility of retweets and replies. This difference in reach may have an implication to future retweet propagation studies.

3.4 Limited Visibility of Direct Messages

The PDMF category in Table 1 concerns direct messages (DM) which can only be accessed by the parties involved in private interactions. Due to this private nature of DMs, we could not study DM propagations in more detail.

3.5 Orphan Retweets and Replies

As seen in Table 1, Orphan Rt and Orphan @ categories exist due to certain missing 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 on a separate trial dataset, 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 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 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 1.

4 Dark Retweets

A dark retweet is defined as a retweet which is propagated using non-conventional retweeting methods. The term “dark” is used to denote how undetectable or invisible a retweet is. In Table 1, several rows were shaded in different shades of grey. The darkness of the shades signify the degree of difficulty in detecting the retweet. There are two main assumptions of retweetability, consisting of varying degrees:

Assumption #1. Is Tweet A a retweet of Tweet B? This assumption becomes more concrete if Twitter API returns a looked up tweet as a proprietary retweet, and metadata of the originating tweet exists. This assumption also becomes moderately concrete for non-proprietary retweets if retweet markers such as ‘RT’ or ‘via’ exists within the tweet text. However, this latter assumption is still debatable – is this manually marked, non-proprietary retweet referring to an original tweet that does indeed exist?

Assumption #2. Who was the originating author? This assumption becomes more concrete if a tweet looked up via Twitter API results in metadata identifying the author of the tweet that is being retweeted. This assumption also becomes moderately concrete for non-proprietary retweets if they contain the username of the perceived author of the originating tweet. However, this latter assumption is still debatable – is the manually mentioned user the correct author of the originating tweet?

The more we have to assume about these two questions, then the “darker” a retweet becomes. For example, orphan retweets, as described in Section 3.5, are assumed to be retweets due to Twitter API’s metadata that claims this to be true. However, the originating tweets no longer exist, therefore violating Assumption #1.

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 that is being referred to (which relates to Assumption #1), nor any identifying information of an originating author (which relates to Assumption #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 [1, 7, 9]. In Table 1, retweets which are not explicitly marked as retweets are considered as dark retweets. The darkest retweets are shown by the darkest-shaded rows, where the Explicit, Link to Original and Mentions Other Users columns are all marked with 0s.

Replies are also considered within the context of dark retweets, because a tweet reply is still an explicit way of propagating tweets, albeit without using conventional retweeting mechanisms. Manual inspections of URLs which return lots of replies seem to show that there are authors who send out multiple replies of similar tweets, rather than making a general retweet. The advantage of this is that a user is able to send a tweet directly to a non-follower. Therefore in the context of tweet propagation, replies are also included in this study. In this paper, tweets are considered as dark retweets depending on its degree of “darkness” compared to other conventional retweets.

Referring to Table 1, the darkest shade of grey denotes that both Assumptions 1 and 2 are difficult to presume within these types of tweets, whilst the second darkest shade denotes that only one of the two assumptions are difficult to presume. However, for brevity, this paper combines the rows coloured by the two darkest shades of grey into one category which represents Dark Retweets.

The following section outlines the proportion of visible, dark and orphan retweets, based on the shades of grey illustrated in Table 1.

5 Matrix Evaluation

Based on the matrix that was described in Section 3, a study was done to evaluate the proportions of tweets which fell within the different categories. The objective of this evaluation was mainly to deduce the extent of dark retweets, involving tweets which propagate without conforming to conventional retweeting methods. These dark retweets may be missed out by existing research of tweet propagation which focus only on retweets made by proprietary Twitter methods or manual insertions of retweet markers in tweet texts.

The hypothesis of this experiment is that dark retweets do exist, and therefore suggests that existing studies on tweet propagation may be missing some further hidden data.

5.1 Experimental Setup

An experiment was run over 2,348,936 (2.3 million) tweets, spanning over 49 days (12/05/12–29/06/12). These tweets were collected at random from Twitter’s Streaming API on its Spritzer setting, which provides a random sample containing 1% of current tweets being published globally in real time.

The dataset used in this study only focused on tweets which had a URL in them. This approach is similar to the one used by Wu et. al [7], as their dataset had a similar restriction as well. To get these tweets, 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 are then classified by querying the Twitter REST API for more details of each tweet. This includes the follower/following information for each particular Twitter user that gets seen within these collected tweets.

A suite of Python scripts were created to perform two main tasks; data collection and data processing. During data collection, the scripts access Twitter’s real-time Streaming API and collect unique URLs from random tweets. They then use Twitter’s Search API to collect all the tweets containing each of the unique URLs collected. Due to Twitter’s rate-limiting policy, the search results were limited to the most recent 1500 tweets, or tweets published in the last 7 days if the 1500 limit did not get reached.

During the experiment, several URLs returned exactly 1500 search results via Twitter’s Search API. These URLs were classified as URLs which may have more than 1500 search results, but further results could not be retrieved due to Twitter’s rate-limiting policy. Section 5.3 will discuss in more detail the ramifications of these limits with respect to the outcomes of this research.

5.2 Evaluation Results

Based on the matrix of tweet propagation as described in Section 3, an experiment was run to observe the proportions of visible, dark and orphan retweets in the above dataset of 2.3 million tweets. Out of this dataset, 820,318 (34.92%) were classified as either visible, dark or orphan retweets. For search results less than 1500 tweets per URL, 15,840 URLs were analyzed. This total increased to 16,976 URLs when the other rate-limited URLs were included.

The results are shown in this paper in the form of pie charts (Figure 1) and line graphs (Figure 2). As shown in Figure 1, Visible Rts form the biggest proportion of retweets overall, but interestingly the proportion changes when search results from rate-limited URLs were included. When the dataset is restricted to search results less than 1500 tweets per URL, dark retweets only accounted for 12.86% of all retweets, but this value increased to 24.7% when the restriction did not apply.

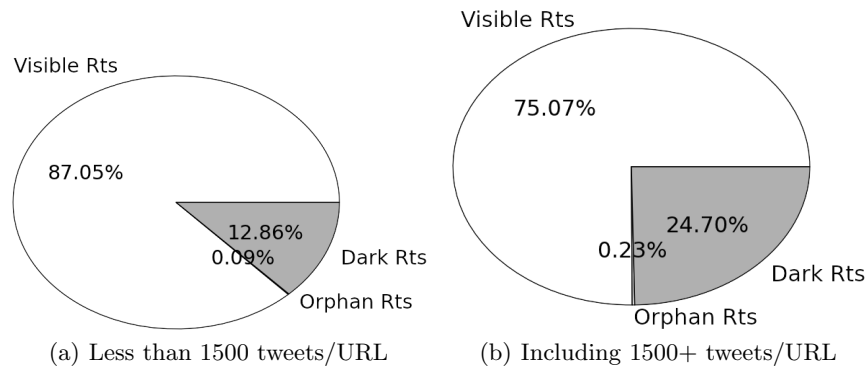


Fig. 1. Pie charts of visible, orphan and dark Rts according to size of search results (total tweets) per URL

Table 2 shows the proportions of visible, dark and orphan tweets as found in Figure 1. This table contains more than the 18 categories outlined in Table 1 because the categories of RtF and RtnF are split into visible and dark groupings. Table 2 shows that these dark RtF and RtnF categories accounted for 0% tweets. Further tweet content analysis would be helpful in capturing these two categories.

For the purpose of this paper, only the URLs with up to 1500 search results were considered. Section 5.3 will later present the reasoning for this approach, in addition to further discussions of the impact of Twitter Search API's rate limit on findings such as the above. Figure 2 shows line graphs which plot the proportions of visible, dark and orphan retweets over the total tweets found per URL. These graphs appear to show a positive correlation between total tweets and the proportions of both visible and dark retweets. However, there seems to be no strong correlation between total tweets and orphan retweets. To verify

Table 2. Proportions of Tweets Found Based on Table 1

Tweets		All %	Rts only %
Ori		65.08	-
Visible	PRtF	12.97	37.14
	PRtnF	8.15	23.33
	Rt@F	0.00	0.00
	Rt@nF	0.13	0.37
	RtF	3.11	8.90
	RtnF	1.56	4.46
	P@RtF	0.00	0.00
	P@RtnF	0.08	0.23
	@RtF	0.15	0.42
@RtnF	0.08	0.23	
Dark	RtF_d	0.00	0.00
	RtnF_d	0.00	0.00
	P@nF	3.82	10.93
	P@F	0.00	0.00
	@F	1.64	4.71
	@nF	3.16	9.06
Orphan	OrphanRt	0.00	0.00
	Orphan@	0.08	0.23

these observations, the Kendall tau-b correlation co-efficients, represented by τ_b , were calculated for the relationships between visible, dark and orphan retweets against the total tweets found per URL. A value of 1 for τ_b would signify a perfect positive correlation between two variables. A value of -1 would signify a similar perfect association but for a negative correlation. A value of 0 would denote that no correlation was found. Calculations made showed that at $\tau_b = 0.739$, visible retweets seem to have a largely positive correlation against total tweets per URL. Meanwhile, dark retweets have a moderately positive correlation against total tweets per URL ($\tau_b = 0.453$), while orphan retweets have a very small positive correlation against total tweets per URL ($\tau_b = 0.120$).

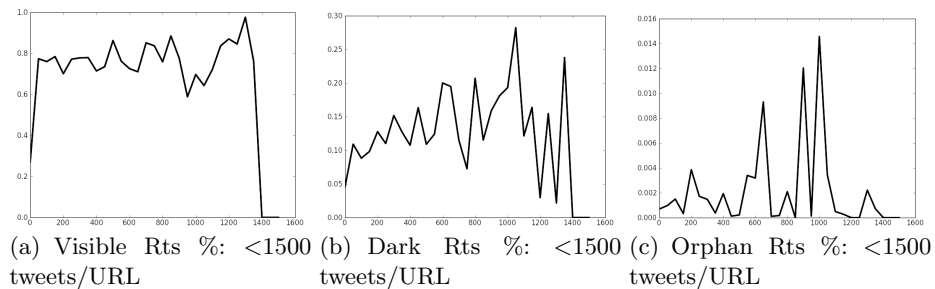


Fig. 2. Proportions of visible, orphan and dark Rts according to size of search results (total tweets) per URL

5.3 Twitter Search API’s Limit of 1500 Tweets

As explained in Section 5.2, two distinct findings emerged when the dataset used was separated into two, namely search results which were not rate-limited, and those which included them. The proportions of dark retweets for search results less than 1500 tweets per URL amounted to 12.86% of all retweets. However, this value increases to 24.7% when rate-limited URLs were also taken into account.

The findings for URLs that were rate-limited were not focused upon in this paper. This is because we could only see an incomplete subset of the tweets which would have included these URLs. This limitation means that it was more likely for us to underestimate the proportion of retweets that were found. This is true particularly for dark retweets, which could only be detected if the complete timeline of all related tweets were found. Only then could the tweets be ordered based on each tweet’s timestamps. Since the ability to see complete timelines is compromised by the 1500 search limit, therefore it is possible that we would underestimate the amount of dark retweets found, and overestimate the amount of original tweets found. Therefore, any estimation for dark retweets would be a conservative estimate.

Since the proportion of dark retweets seems to rise substantially when the rate-limited URLs are included (from 12.86% of all retweets to 24.7%), then this seems to suggest that dark retweets are more prevalent amongst rate-limited URLs. On manual inspection, it seems that the search results for rate-limited URLs tend to contain tweets published by spambots, or autosenders tied to web services. The importance of recording dark retweets may be slightly diminished when they are considered to be more popular amongst spambots. However, it would be difficult to automatically discount all tweets containing rate-limited URLs as spam.

5.4 Impact on Existing Research

The existence of dark retweets may impact the findings of other existing studies on tweet propagation. In this paper, 12.86% of dark retweets were found for URLs which returned up to 1500 search results each. Table 3 shows the extrapolation of our findings with two other papers, namely those made by Suh et. al [10] and Cha et. al [2]. These figures were derived by matching the proportions found in these papers with the findings as presented in this paper.

The main difference between the work by Suh et. al [10] and this paper is that the dataset used in our research is restricted to tweets containing URLs only. Therefore if we were to extrapolate the proportions of retweets with URLs in them, as found by Suh et. al [10], then 0.47%, or 347,800 tweets, would potentially be hidden from their dataset.

This extrapolation becomes more difficult to make when the retweets dataset is not restricted to tweets containing URLs only, such as the case with Cha et. al’s study [2]. However, in their paper, it was claimed that 92% of the retweets they found contained URLs in them. The proportion of retweets that they found was not given in their paper, therefore we could only extrapolate the proportion of

retweets with URLs as stated by them, amounting to 92% of all retweets. From this percentage, 13.6% of those retweets could be extrapolated as dark retweets, therefore potentially hidden from Cha et. al’s evaluation.

There are various studies which use retweets as the basis of their studies, but considering that our studies have restricted the dataset that we used to tweets containing URLs only, then it is not possible to be certain that any extrapolations will hold.

Table 3. Extrapolation of our Findings with Existing Research

Studies		Suh	Cha
Existing work	Dataset size	74 million	1.76 billion
	Original tweets	88.85%	-
	All Rts	11.15%	-
	Rts with URLs	28.4% of retweets	92% of retweets
Extrapolation from our work	Original tweets	96.36%	-
	Visible retweets	3.17%	92% of retweets
	Dark Retweets	0.47%	13.6% of retweets
	Hidden Retweets	347,800	-

The impact of this hidden data is the current focus of ongoing research work. One such focus is on the perceived reach of a retweet. In the work done by Kwak et. al [11], it was claimed that an average retweet could reach 1000 users, irrespective of who the originating author is. Considering the different visibility properties of dark retweets, their existence may or may not affect this number. Further research could reveal the actual impact of dark retweets on tweet propagation studies such as the above.

6 Conclusion

This paper has presented a comprehensive matrix of tweet propagation, which was deconstructed into seven different characteristics, namely proprietary, mechanism, follower/non-follower, mentions other users, explicit, links to original tweet, and push audience. The paper has discussed the different nuances of tweet propagation, such as the availability of metadata linking to originating tweets. The different visibilities of retweets and replies also meant that the reach of these tweets were different. In addition, several retweets seemed to be using multiple retweeting mechanisms, which impacts the total reach of their tweets. Orphan retweets and replies also exist due to the subsequent deletion of originating tweets and users at a later time. The concept of dark retweets was introduced, based on two assumptions of retweetability pertaining to a retweet’s provenance; an originating tweet exists and it was made by an originating author.

The experiment that was run to evaluate the matrix of tweet propagation showed that over 2.3 million tweets, dark retweets amounted to 12.86% (for

search results less than 1500 tweets per URL) and 24.7% (for search results including more than 1500 tweets per URL) respectively. By extrapolating the results found in several existing papers [10, 2], the potential hidden data as a result of dark retweets amounted to between 0.47–13.6%.

Several threads of future work is planned for this research, such as conducting a deeper statistical analysis on the frequencies found in Table 2, by looking at the stability of percentages found over time. This is in addition to identifying other external factors impacting these percentages, such as location, working hours, external offline events, etc. Another future work thread involves quantifying the actual impact of dark retweet detection on existing studies on tweet propagation.

References

1. boyd, d., Golder, S., Lotan, G.: Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In: Hawaii International Conference on System Sciences, Los Alamitos, CA, USA, IEEE Computer Society (2010) 1–10
2. Cha, M., Haddadi, H., Benevenuto, F., Gunmadi, K.P.: 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, AAAI (2010)
3. Mustafaraj, E., Metaxas, P.: 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 (2010)
4. Adar, E., Zhang, L., Adamic, L.A., Lukose, R.M.: Implicit structure and the dynamics of blogspace. In: Workshop on the Weblogging Ecosystem, WWW2004, New York, NY (2004)
5. Matsumura, N., Yamamoto, H., Tomozawa, D.: 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, AAAI (2010)
6. Galuba, W., Aberer, K., Chakraborty, D., Despotovic, Z., Kellerer, W.: Outtweeting the twitterers- predicting information cascades in microblogs. In: 3rd Workshop on Online Social Networks, Boston, MA, USA, USENIX (2010)
7. Wu, S., Hofman, J.M., Mason, W.A., Watts, D.J.: Who says what to whom on Twitter. In: Proceedings of the 20th international conference on World wide web (WWW '11), ACM (2011)
8. Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J.: Identifying ‘influencers’ on Twitter. In: Fourth ACM International Conference on Web Search and Data Mining, ACM (2011)
9. Nagarajan, M., Purohit, H., Sheth, A.: 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, AAAI (2010)
10. Suh, B., Hong, L., Piroli, P., Chi, E.H.: Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In: IEEE Second International Conference on Social Computing, IEEE (2010) 177–184
11. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? In: Proceedings of the 19th international conference on World wide web, Raleigh, North Carolina, USA, ACM (2010) 591–600