

Patterns of Implicit and Non-follower Retweet Propagation: Investigating the Role of Applications and Hashtags

Norhidayah Azman
University of Southampton
Southampton
United Kingdom
nba08r@ecs.soton.ac.uk

David Millard
University of Southampton
Southampton
United Kingdom
dem@ecs.soton.ac.uk

Mark Weal
University of Southampton
Southampton
United Kingdom
mjw@ecs.soton.ac.uk

ABSTRACT

Existing literature on retweets seems to focus mainly on retweets created using explicit, formal retweeting mechanisms, such as Twitter’s own native retweet function, and the prefixing of the terms ‘RT’ or ‘via’ in front of copied tweets. However, retweets can also be made using implicit, informal mechanisms. These include tweet replies and other mechanisms, which use neither the native nor RT/via mechanisms, but their content and timelines suggest the likelihood of being a retweet. Moreover, retweets can also occur with or without a defined follower/following network path between a tweet originator and a retweeter. This paper presents an initial taxonomy of propagation based on seven different ways a tweet may spread: native, native non-follower, RT/Via, RT/Via non-follower, replies, non-follower replies and other implicit ‘retweets’. An experiment has examined this new model, by investigating where tweets containing URLs from the domains of online petitions, charity fundraisers, news portals, and YouTube videos can be classified into the seven different categories. When including other implicit ‘retweets’, more than 50% of all the retweets found across all four domains were classified as implicit retweets, while more than 79% of all retweets were made by non-followers. More work needs to be done on the composition of other implicit ‘retweets’. Initial investigations found hashtags in 99–100% of these tweets, suggesting that retweeting using conventional mechanisms may not be the main method that URLs get propagated across microblogs.

Categories and Subject Descriptors

H.3.1 [Information Systems]: Information Storage and Retrieval—*Content Analysis and Indexing*

General Terms

Algorithms, Measurements, Experimentation

1. INTRODUCTION

Retweets have been acknowledged for their conversational nature [1]. Several studies have been made on the patterns of retweet propagation, particularly in the areas of retweet reach [5], information brokerage [7] and in approximating influence [4, 2].

However, these studies have been focused on retweets made using explicit retweeting mechanisms. These include two conventional retweeting mechanisms; the native retweet function provided by Twitter (native retweet), and also by copy-

ing and pasting prior tweets and then inserting terms such as ‘RT’ or ‘via’ to signify a retweet (RT/via retweet).

There seems to be little emphasis on retweets made using implicit mechanisms. Examples of implicit retweets include those that were spread using the reply function, rather than the retweet function. Implicit retweets could also include those that show no discernable markers signifying a retweet, but the content could have been seen beforehand prior to being subsequently repeated. In these instances, it is difficult to determine the provenance of retweets made implicitly [1].

Another area of interest in this field is the proportion of retweets made by users who do not follow the authors of originating tweets [6]. It was found that 33% of retweets gave credit to users who were not being followed [3].

In this paper, we studied the different ways that a tweet may spread, forming a taxonomy of propagation types. The findings in this paper show that implicit retweets form a large proportion of all types of retweets. This suggests that it may be important to study implicit retweets in more depth when analyzing information propagation in social media. This research also studied the role of hashtags and the Twitter applications involved in implicit retweeting.

Section 2 explains the taxonomy of propagation types based on seven different ways a tweet could spread. Section 3 describes the experiment that was run to validate this taxonomy. The subsequent Section 4 details all the findings, followed by the conclusion in Section 5.

2. TAXONOMY OF PROPAGATION TYPES

The aim of this study is to explore the existence of implicit and non-follower retweets, and comparing them to explicit and follower retweets. These patterns describe the different mechanisms that could be used to propagate an item across an interlinked environment such as microblogs.

From the two aspects described above, a taxonomy of seven different ways a tweet could spread was constructed (see Figure 1): native retweet, native non-follower, RT/via retweet, RT/via non-follower, replies, non-follower replies, and other implicit retweets.

Table 1 shows the classifications of explicit, implicit, follower and non-follower retweet types.

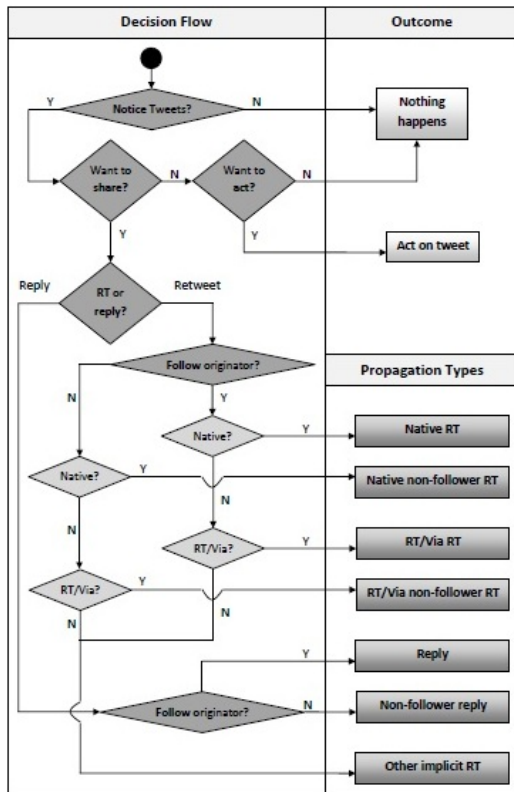


Figure 1: Retweet decision model based on taxonomy of propagation types

	Explicit RTs	Implicit RTs
Follower RTs	N RV	@
Non-follower RTs	NnF RVnF	@nF O

Table 1: Matrix of explicit/implicit and follower/non-follower retweet types

2.1 Native Retweets (N)

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. Third party apps which send retweets via Twitter API’s retweet function are also included in this count.

2.2 Native Non-follower Retweets (NnF)

Native non-follower retweets are native retweets which include non-follower paths. For example, user B made a native retweet to his or her followers, and the tweet text was referencing user A. However, user B is *not following* user A, thus this connection between users B and A is categorized as a **native non-follower retweet**.

2.3 RT/Via Retweets (RV)

RT/Via retweets are defined as tweets which repeat prior tweets and include any of the following tokens within the tweets: “rt @”, “rt@”, “rt:@”, “rt: @” , “retweet @”, “via @”, “retweet :@”, “r/t”, “rt.”, “RT @”, “RT@”, “RT:@”, “RT:

@”, “RETWEET @”, “VIA @”, “RETWEET :@”, “R/T”, or “RT.”.

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

2.4 RT/Via Non-follower Retweets (RVnF)

RT/Via non-follower retweets are defined as the non-follower path that is observed amongst RT/Via retweets. For example, user E made an RT/Via retweet of a text which referenced user D. However, user E is *not following* user D. Therefore, the connection between the two users is categorized as an **RT/Via non-follower retweet**.

2.5 Replies (R)

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

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 taxonomy, is that this research’s experimental 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 sees a URL propagated from another user, and then User A sends that same URL to followers via replies.

This propagation type may not traditionally be considered as a retweet type, but considering that this toolkit only records replies which contain *URLs that had already been seen before* by the user making the reply, thus this reply becomes a valid point of reference in determining the provenance and propagation of URLs across tweets.

This research work acknowledges that the visibility of replies could be difficult to determine accurately. If user A tweets ‘hello!’, then all of A’s followers would see that tweet appear in their subscribed timelines of tweets.

However, when a reply is made, then that tweet would only appear in the timelines of mutual followers of **all** the parties involved in that particular correspondence. For example, if users A and B were replying tweets amongst themselves, then only mutual followers of both A and B would see those replies in their subscribed timelines.

Nonetheless, if someone were to browse the Twitter profile pages of either A and B, then anyone could see those replies publicly.

Given the complexities of tweet visibility pertaining to replies, it can be slightly harder to correctly gauge how many people

could potentially have seen a reply displayed on Twitter. Although a reply is publicly visible to non-followers, it is more likely to be seen by mutual followers of the parties involved in the reply, as opposed to the general public.

2.6 Non-follower Replies (@nF)

Non-follower replies similarly begin with a user mention. As an example, user J sent a tweet containing a URL to user K. Then, user K made a reply to user L. In this case, user K is *not following* user J, thus this connection is categorized as a **non-follower reply**.

2.7 Other Implicit ‘Retweets’ (OnF)

Other implicit ‘retweets’ are defined as tweets which do not conform to any of the above six classifications described. For example, user Y is *not following* user X, but makes a retweet without using an explicit mechanism; the retweet was not made using native, RT/via or reply mechanisms. These also include tweets which seem to have no apparent attribution or acknowledgement of prior tweets or originators from any time before the retweet was made. For example, tweets being copied verbatim would be categorized as an other implicit ‘retweet’.

Some assumptions had to be made when looking at other implicit ‘retweets’. Some of these could be original tweets which were created without referencing any prior tweets. Moreover, a user could click on a ‘share this’ button on a third-party website, which then shares a particular URL link to that user’s Twitter followers without using any explicit retweet mechanisms. Therefore, this research acknowledges that some of the tweets considered as an other implicit ‘retweet’ may include original tweets, or come from third-party sources such as non-Twitter web pages, user e-mails and so on.

This research holds the assumption that Twitter users reference these URLs after seeing prior tweets which prompt them to visit these websites. This assumption seems intuitive in the case of online petitions, because user A could sign a petition and then tweet the petition’s URL link to follower user B. Then, it is possible for user B to also sign the petition and subsequently post the same URL link to user B’s own followers, without acknowledging user A’s prior tweet.

3. EXPERIMENT SETUP

An experiment was run to validate this taxonomy against a collection of tweets containing URLs from four chosen domains, namely *online petitions*, *charity fundraisers*, *news portals*, and *YouTube videos*. These domains were chosen after manually observing the types of URLs that seem to be propagating across Twitter.

The experimental toolkit built for this research work consists of a suite of Python scripts. The toolkit is exploratory in nature and used mainly to investigate the types of retweets found off of the Twitter API.

Five different URLs were chosen for each domain, giving a total of **20 URLs overall**. The experimental 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. A

total of 11,846 tweets were collected, involving 7118 unique users.

4. FINDINGS

Across all four domains, the proportions of implicit retweets seem to be consistently bigger than explicit retweets. Moreover, non-follower retweets also form bigger proportions as compared to follower retweets (see Table 2).

Domain	Retweets (%)			
	Exp	Imp	F	nF
Fundraisers	25.9	74.1	13.6	86.4
News	2.4	97.6	1.5	98.5
Petitions	48.5	51.5	21.3	78.7
YouTube	23.1	76.9	12.1	87.9

Table 2: Proportions of explicit/implicit and follower/non-follower retweets across all four domains

Across all four domains, two observations seem to consistently emerge: 1. There are more implicit retweets as opposed to explicit retweets (proportions of implicit retweets range from 51.5% to 97.6%), and 2. There are more non-follower retweets as opposed to follower retweets (proportions of non-follower retweets range from 78.7% to 98.5%)

In the news domain, we can see an extremely high proportion of implicit retweets (97.6%), while tweets containing petition URLs seem to have a somewhat 50/50 spread 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.

The proportions found by this toolkit also support the idea that there exists alternative propagation pathways that could be useful in quantifying explicit and implicit power.

Given this preliminary finding, further analysis was carried out to investigate the composition of other implicit ‘retweets’, particularly at hashtags and Twitter applications. For each domain, the proportions of retweets containing hashtags were recorded, and the types of Twitter applications that were used were identified.

In this analysis, hashtags were found between 98–100% of other implicit ‘retweets’. This suggests that users could have seen particular URLs from people they do not follow via hashtags before subsequently retweeting them (see Table 3).

The top three Twitter applications found for each domain is listed in Table 4. There seems to be no pattern that could be identified from these rankings. Interestingly, in the domain of YouTube URLs, the top three applications are all proprietary interfaces created by Twitter.

Domain	With # (%)	Without # (%)
Fundraisers	98.6	1.4
News	98.8	3.2
Petitions	100.0	0.0
YouTube	100.0	0.0

Table 3: Proportions of implicit retweets containing hashtags

D	1st	2nd	3rd
F	TweetDeck	Tweet Button	Foursquare
N	dlvr.it	DestroyTwitter	TweetMeme
P	HootSuite	CoTweet	Twitter for iPhone
Y	twitter.com	Twitter for iPhone	Twitter for Android

Table 4: Rankings of Twitter applications making implicit retweets

4.1 Implicit and Non-follower Retweets as a Medium for Propagation

As mentioned above, throughout all four domains, the proportion of 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 tweeted without using the above retweet mechanisms. Therefore, more work needs to be done to identify how these ‘dark’ retweets propagate.

When the ‘dark’ retweets are analyzed manually, the tweets seem to consist of either one of the following characteristics: *verbatim copies of other tweets, unknown retweet markers, or non-Latin characters.*

Prior work on retweets seems to be concentrated on studying retweets which conform to explicit, pre-defined retweet mechanisms. 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 experiment, the prevalence of implicit retweets seems 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 seem to spread not just via retweets, but they could also spread via implicit means such as verbatim copying or using non-conventional retweet tokens in their tweets. All this suggests the possibility of a different way for tweets to propagate across the Twitter network.

5. CONCLUSION

This paper has contributed a taxonomy of propagation types based on seven different methods of spreading a tweet. This

taxonomy provides a novel way of viewing retweeting behaviour and it may lead to a re-evaluation of some existing research.

In the experiment to validate this taxonomy, more than 50% of all retweets were classified as implicit retweets across all four domains of URLs experimented on. In addition, more than 79% of all retweets were made by non-followers. Hashtags usage seems to be a plausible explanation to this high percentage. On all four domains, more than 99–100% of other implicit ‘retweets’ contained hashtags in them.

This initial taxonomy is currently being refined, particularly to separate original tweets from implied retweets within the other implicit ‘retweet’ category.

At present, this research has shown that implicit retweeting plays a significant role in the propagation of information through the social media space and must be considered when looking at issues such as influence within social microblogging.

6. REFERENCES

- [1] d. boyd, S. Golder, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In *Hawaii International Conference on System Sciences*, pages 1–10, Los Alamitos, CA, USA, 2010. IEEE Computer Society.
- [2] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gunmadi. 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] W. Galuba, K. Aberer, D. Chakraborty, Z. Despotovic, and W. Kellerer. Outtweeting the twitterers- predicting information cascades in microblogs. In *3rd Workshop on Online Social Networks*, Boston, MA, USA, 2010. USENIX.
- [4] H. Kong, H. Park, and S. Han. The talkative, the popular, and the influential Korean Twitter users. In *Social Networks Interoperability 1st International Workshop 2009*, 2009.
- [5] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600, Raleigh, North Carolina, USA, 2010. ACM.
- [6] M. Nagarajan, H. Purohit, and A. Sheth. 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.
- [7] D. van Lieere. 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, Canada, 2010. ACM.