8-2018

Offline versus online: A meaningful categorization of ties for retweets

Felicia NATALI
Singapore Management University, felician.2013@phdis.smu.edu.sg

Feida ZHU
Singapore Management University, fdzhu@smu.edu.sg

Offline Versus Online: A Meaningful Categorization of Ties for Retweets

Felicia Natali(✉) and Feida Zhu

Singapore Management University, Singapore, Singapore
{felician.2013,fdzhu}@smu.edu.sg

Abstract. With the recent proliferation of news being shared through online social networks, it is crucial to determine how news is spread and what drives people to share certain stories. In this paper, we focus on the social networking site Twitter and analyse user's retweets. We study retweeting patterns between offline and online friends, particularly, how tweet novelty and tweet topic differ between tweets retweeted by offline friends and those retweeted by online friends.

Keywords: Twitter · Retweet · Offline ties · Online ties

1 Introduction: Retweet and its Drivers

Retweets have long been an important research topic in the social media sphere. With the emergence over the last decade of online social network platforms like Facebook and Twitter, online interactions have produced large volumes of data, offering researchers the opportunity to examine the information users have shared. As a result, information dissemination has become a prominent area of study in the field of social media analysis.

Retweeting is one of the most popular ways of disseminating information on Twitter, a social media and microblogging site that is widely used to circulate news [8]. A retweet is a re-posting of a tweet on your feed, and so the feature allows you and others to share selected tweets with your followers. You can retweet your own tweets or tweets from someone else[1].

Understanding retweets is important since they are used for various practical purposes such as sharing news, promoting political views, marketing products, and tracking real time events. Java et al. attributed the high volume of tweets mostly to daily chatter, although tweets still usually contained a fair amount of news items [6]. Enli and Skogerbø explored Twitter and Facebook as arenas for political communication [3]. Thomases, meanwhile, wrote a guide book about how to create a successful Twitter marketing campaign [15].

Therefore, if the drivers of retweets were understood properly, then harnessing them would bring immense benefits to marketing campaigns and public

[1] Retweet FAQs https://help.twitter.com/en/using-twitter/retweet-faqs.

Published in Social Informatics: 10th International Conference, SocInfo 2018, St. Petersburg, Russia, September 25-28: Proceedings. LNCS 11185, 386-402.
DOI: 10.1007/978-3-030-01129-1_24

policy interventions. Boyd et al. compiled a comprehensive list of the motivations behind retweets. It included making new audiences aware of certain tweets and increasing a listener's visibility [2]. In addition to these internal motivations listed by Boyd et al., a number of external attributes also influence retweets, such as URLs and hashtags, and also Twitter accounts' age and follower count [14]. The study by Kupavskii et al. determined that influential users with high scores on PageRank – a measure of a website page's importance applied to Twitter follow networks – received more retweets [7].

In this study, we aim to find how offline versus online ties can be harnessed to drive retweets. Our study is the first to reveal the retweet patterns of offline friends compared to online friends, and offer another promising way for Twitter users to increase the amount of retweets their tweets receive. In this way, based on the results, marketing and political campaigns can target a specific type of user (offline/online) to increase influence. Our study also highlights the importance of the offline-online categories when discussing retweets, and demonstrate that these categories cannot be replaced by other categories, such as the reciprocated-unreciprocated categories.

2 Literature Study

Types of ties have been known to drive retweets. Past research has looked into how different ties bring about retweets. Most determined that strong ties drove retweets [12,17], although some concluded that weak ties did [13]. Meanwhile, Natali et al. analysed how different ties resulted in different topics getting retweeted [10]. In an extended study of this study[2] that utilized a more extensive data, they discovered that Twitter users did not consider ties when retweeting any topic half of the time – though when they did pay attention to them, the results were largely similar to the previous study. Personal tweets were more likely to be disseminated through strong ties, whereas entertainment and news tweets were more likely to be disseminated through weak ties. These past studies, however, defined strong ties differently. Zhao et al. used the overlap of neighbours as the indicator of strong ties [17], while Peng et al. used mutual mentions, mutual retweets, mutual followers and mutual followees as the indicators of strong ties [12]. Natali et al. and Shi et al., meanwhile, used reciprocity of follow ties to define strong ties [10,13].

In this study, we focus on different categories of ties, namely offline versus online. We aim to find out if offline and online ties can be used in place of other tie categories that were previously utilised in studies that analysed retweets. These categories of ties are reciprocated and unreciprocated. We discover that offline versus online are indeed better tie categories because they can be distinguished more easily by their retweet patterns.

[2] This study is currently unpublished and is a part of a thesis. Please contact authors if you want to know more.

3 Tie Categories

Strong Ties. Granovetter first introduced the concept of strong ties in his seminal work *The Strength of Weak Ties* [5]. In the study, Granovetter described interpersonal ties as "a (probably) linear combination of the amount of time, the emotional intensity, the intimacy (or mutual confiding), and the reciprocal services which characterize each tie". In addition to this formula, Granovetter emphasized the uniqueness of strong ties. Strong ties had more overlapping friends compared to two individuals selected arbitrarily. Therefore, Granovetter concluded that information that circulated among close friends is usually stale and old.

Measuring Strong Ties on Offline and Online Social Network. There are several ways to measure a tie's strength. The first study to do so is the study by Marsden and Campbell [9]. They discovered that the question of how close a person to another was the best indicator of closeness. Their study applied to the offline setting.

In the online setting, Gilbert and Karahalios authored the most extensive study on the measurement of strong ties on Facebook [4]. They made use of 74 Facebook variables in order to predict strength of ties. Their method achieved a good accuracy. Meanwhile, Backstrom et al. revealed that mutual friends of very intimate friends were rarely unconnected [1]. Their study offered the distance of mutual friends as a potential measure of how intimate two friends are.

Reciprocated versus Unreciprocated. Reciprocated ties have often been used as an easy gauge of strong ties when studying retweets [10,13]. On Twitter, a reciprocated tie appears in a situation where a user follows another user, and he or she is also followed back. On the other hand, an unreciprocated tie appears in a situation where a user follows another user, but he or she is not followed back. When someone follows another person on Twitter, he subscribes to the updates published by that person's account. In this study, the analysis of how reciprocated versus unreciprocated ties retweet will be the baseline for assessing how different the tweet novelty and topic of offline and online ties are.

Offline versus Online. Offline ties are not exactly the same as reciprocated ties, although reciprocated ties can predict offline ties with 73% precision and 65% recall. No one has previously studied how offline versus online friends retweet. In this study, we define *offline friends* as connections on Twitter who have met outside of the internet. The connections include both reciprocated and unreciprocated connections. Meanwhile, *online friends* are connections on Twitter who have never met outside of the internet.

4 Dataset: Two-Hop Retweet Data

Determining whether each tie involved in a full retweet chain is offline or online is impossible; however we can find out if the ties in a retweet chain in an ego network are offline or online.

Therefore, in this study we use a dataset gathered by Xie et al. [16]. It contains the data of 98 Twitter users, including his ego network in 2011 and the list of his Twitter connections (followers or followees) whom he knows in real life. Overall, the dataset has 20030 Twitter users (ego users and their followers/followees) and 23225 edges labeled as an offline or an online friend. Additionally, we conducted another survey in 2015. We included the new survey data in our dataset. It consists of 41 Twitter users who filled out a survey asking them to label who their offline friends are among a random sample of at most one hundred of their connections (followers/followees) on Twitter.

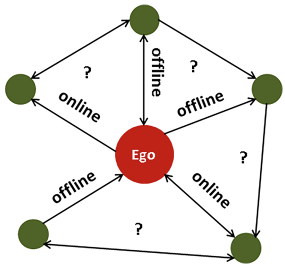


Fig. 1. Ground truth ego networks.

The illustration of the ground truth ego networks can be seen on Fig. 1. From the illustration, the definition of an ego network can be understood clearly. An ego network includes a Twitter user – called *ego user*, depicted by the red circle – and his followers and followees on Twitter. The edges among all of these users are crawled, producing a two-hop follow-networks that are bounded by the ego user and his followers/followees. In the ground truth data, we have the labels of who the offline friends among an ego user's followers or followees are. We procure these labels from the survey answers. However, there is a limitation to our ground truth data. The categories (offline or online) of the edges between the followers or followees of the ego users, are missing. These edges are marked by '?' in Fig. 1. Our experiment and analysis will take into account this limitation.

We crawl the tweets of all the users in our dataset on March 2018. Additionally, we also crawl the latest follow-edges among these users. Temporal changes in offline and online relationships from the year 2011 and 2015 to 2018 can happen. Although those who are offline friends cannot become online by our definition, those who are online friends can become offline. Therefore, the interpretation of the results may downplay the importance of offline friends and exaggerate the importance of online friends.

5 Methodology: Calculating Retweets Depth and Quantifying Retweets Topic

Before proceeding to the methodology, we will recap the issues our research focuses on. In this study, we want to reveal the retweet patterns of offline and online friends on Twitter. Specifically, we want to know the difference in the tweet novelty and retweet topic of offline and online friends. We also want to know whether this difference is greater than the difference between the retweet patterns of reciprocal and unreciprocal friend categories.

However, due to the limitation of the dataset explained in Sect. 4, we cannot analyse the whole retweet chain. Therefore the analyses performed will have the following limitations:

1. We can only analyse retweet patterns that happen among Twitter users in an ego network.
2. We can only analyse retweet patterns that go through public accounts, since their edges cannot be crawled otherwise.
3. Only when a retweet passes from or to an ego user, can we know whether a retweet passes through an edge that represents an offline or an online friendship. If the retweet does not come from or go to an ego user, we will only know whether the retweet passes from an ego user's offline or online friend, to another offline or online friend (See Fig. 1).

Given these limitations, there are seven categories of ties that we analyse in this study.

1. *Offline ties* that represent connections on Twitter who know one another offline.
2. *Online ties* that represent connections on Twitter who do not know one another offline.
3. *Offline-to-offline ties* that represent connections on Twitter between an ego user's offline friend and another offline friend.
4. *Online-to-offline ties* that represent connections on Twitter between an ego user's offline friend and an ego user's online friend.
5. *Online-to-online ties* that represent connections on Twitter between an ego user's online friend and another online friend.
6. *Reciprocated ties* that represent connections on Twitter between two users in which the users follow one another.
7. *Unreciprocated ties* that represent connections on Twitter between two users in which only one user follows another.

As Twitter only reveals the original source of a tweet, and not from whom a retweeter retweets, we must make several assumptions to construct a retweet chain. We use these two:

1. *Latest timing.* Twitter generally arranges feed based on chronological order. Although in the past few years, Twitter shows what it considers as the best tweets for you first, more current material will appear afterwards. The tweet of a user who tweets last will be likely to appear on top. Therefore, it makes sense to assume that the followee of a user who retweets something just before the user retweets, is the source of a retweet. If there are no retweeters in the ego network who retweet before the user retweets, the original source of the retweeted tweet is considered. If the original source is a followee, he is

considered as the source of retweet. Otherwise, the source of the retweet is unknown.

2. *Most popular*. Popular people have a lot of followers. They are also most influential. Therefore, it makes sense to assume that a user's followee who tweets or retweets before the user retweets a tweet, and has the most followers is the source of a retweet. When there are no followees who tweet or retweet before the user does, then the source of the retweet is unknown.

Figure 2 is used to illustrate these two assumptions. In the figure, each level represents the time a tweet is retweeted, with t_0 representing the time when the tweet first originates. Therefore, User B is the original source of tweet. The edges are the follow edges that exist among the nodes. Assuming that there are no other follow edges among the nodes outside the system, User C is the most popular. Based on this configuration, the source of retweet for User D is User A based on the latest timing assumption, and User C based on the most popular assumption.

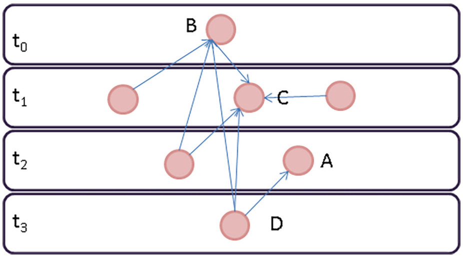


Fig. 2. Illustration of different assumptions for constructing a retweet chain.

In our analysis, we are concerned only with the retweet chain in an ego network. Therefore, all the analyses are based on the assumption that **a retweeter's source of a retweet can only come from the ego network being analysed**. We make such an assumption because we do not know the category of friendship that exists between the source of a retweet outside an ego network and the retweeter, that is, whether it is offline or online. By applying this assumption, we may not get the user who is the true source of a retweet, but we will get the user in an ego network who has the highest likelihood of being the source of a retweet.

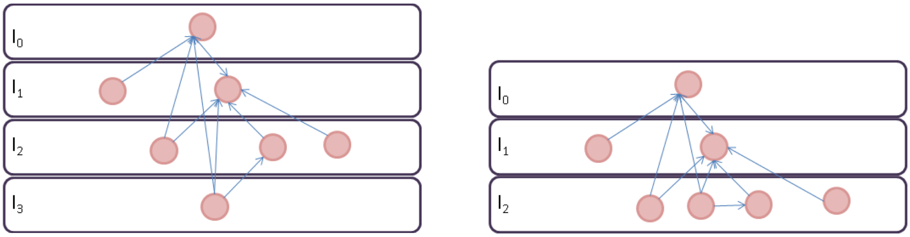
In this study, we need to measure tweet novelty and quantify tweet topics. There are two ways to measure tweet novelty. The first is, how far in time the retweeted tweet is from the original tweet. The second is, the depth of the retweet chain. Now, we will explain these measurements sequentially.

5.1 Tweet Novelty by Duration

In measuring tweet novelty by duration, we measure how far in time the retweeted tweet is from the time when the original tweet is published.

5.2 Calculating the Depth of Retweet Chains

The depth of a retweet chain refers to the deepest level of a retweet chain. Each level represents not the time of a retweet, but the sequence of one. The value can change depending on the assumption that we make. If we stack nodes in Fig. 2 by depth level and, not by the time of a retweet, we will come up with Fig. 3. Figure 3 shows the depth level of different assumptions. The depth of the retweet chain is four if we use the latest timing assumption, and three if use the most popular assumption.



(a) Latest timing assumption. (b) Most popular assumption.

Fig. 3. Levels of depth given different assumptions.

The depth of a retweet chain represents the greatest degree of separation that can be reached by the source of a tweet. When a retweet chain is made by employing the latest timing assumption, the depth of the retweet chain represents tweet novelty, not in terms of duration, but in terms of the length of the chain of direct or indirect friends among whom the tweet has circulated. The deeper the level at which a user retweets, the longer the tweet has circulated among friends who are directly or indirectly connected to the user.

In this study, we calculate the frequency of different tie categories at each level of depth for each assumption. We symbolize this frequency as f_l^c, where l represents the level of depth, a value that can range from one to infinity and c represents the frequency of ties that belong to the category c.

To ensure that the difference in the frequency of ties used for retweets is not due to the difference in the frequency of ties in the networks, we will normalize the frequency by N_c – the frequency of ties that belong to the category c in the networks. We symbolize the normalized f_l^c as $\hat{f}_l^c$ (See Eq. 1). f_l^c represents the proportion of ties in those networks that belong to category c and are used for retweets.

$$\hat{f}_l^c = \frac{f_l^c}{N_c} \tag{1}$$

5.3 Quantifying Retweet Topics

In this study, we also want to find out how well different tie categories can be distinguished by topics. Therefore, we apply Twitter-LDA [18] to extract topics

Table 1. Extracted topics from tweets.

Code	Topic	Sample words
P0	Sexually explicit words	girl, love, baby, hot, fuck
P1	Shows and videos	live, tonight, youtube, video
P2	Global news	new york, trump, people, news
P3	Singapore politics	singapore, lee, pm, pap
P4	Sports	team, great, chicago, race
P5	Singapore news	people, police, singapore, man
P6	Education and Jobs	students, education, school, work
P7	Global politics	trump, president, obama, india
P8	Stocks	latest, price, bitcoin, usd
P9	Traffic and weather	singapore, time, weather, rain
P10	Fun and socialize	song, tonight, happy, guys
P11	Technology	apple, iphone, app, google
P12	Friends and daily life	people, happy, life, day
P13	Social media	tech, social, google, online
P14	Family and finance	money, day, food, children

from the tweets that are retweeted by various tie categories. From implementing Twitter-LDA to process the tweets, we get out 15 topics that are listed in Table 1.

In addition to churning these 15 topics out, Twitter LDA also produces the distribution of these tweet topics that are retweeted by different tie categories.

6 Results: Categorizing Ties for Retweet

In this Section, we will discuss the results of calculating the depth of the retweet chains and quantifying retweet topics of tweets that belong to different tie categories.

6.1 "Offline Versus Online" as the Category of Ties by Tweet Novelty

Table 2 calculates the normalized frequency of ties that belong to category c at depth level l ($\hat{f}_l^c$) expressed in percentage. c can be offline, online, offline-to-offline, online-to-offline, or online-to-online. Therefore, the value 28.33 in the first cell means that 28.33% of offline ties are used to retweet at depth level 1. A user who retweets at depth level one is the start of a retweet chain.

The results show that there are more depth levels produced when the latest timing assumption is used. The results also demonstrates that a greater percentage of offline ties are used to retweet compared to online ties. Meanwhile, the greatest percentage of ties that are used to retweet are the online-to-online

Table 2. Normalized frequency of ties that belong to the offline-online categories at depth level l ($\hat{f}_l^c$) expressed in percentage.

Depth level	Latest timing assumption				
	Off	On	Off-to-off	On-to-off	On-to-on
1	28.33	17.57	22.93	28.19	58.29
2	2.17	0.55	0.94	0.63	1.98
3	0.09	0.02	0.16	0.06	0.31
4	0.02	0.01	0.06	0.01	0.06
5	0.00	0.00	0.03	0.00	0.01
6	0.00	0.00	0.02	0.00	0.00
7	0.00	0.00	0.01	0.00	0.00
8	0.00	0.00	0.01	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00
>= 11	0.00	0.00	0.01	0.00	0.00
	Most popular assumption				
1	28.37	17.82	23.27	28.66	59.41
2	1.53	0.32	0.47	0.35	1.18
3	0.02	0.02	0.03	0.01	0.09
4	0.00	0.00	0.01	0.00	0.01
5	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00

ties. However, when the latest timing assumption is used, offline-to-offline ties have the greatest percentage of retweeting ties compared to other ties at the lower depth levels (depth level $\geq$5). Such results indicate that offline-to-offline ties are more likely to retweet older news that has been retweeted by their direct or indirect Twitter friends at earlier times.

A previous study by Natali et al. discovered that a user's offline friends were more highly connected on Twitter than a user's online friends [11]. Therefore, we can conclude that friends who are likely to be offline (offline-to-offline ties) are more likely to retweet older news. Meanwhile, although a Twitter user's online friends are not as connected as their offline friends [11], they are the best circulator of information on Twitter networks at higher depth levels (depth level $\leq$4). These results support Granovetter's theory that strong ties confine information circulation within local clusters [5]. As such, novel news typically comes from weak ties.

However, when we measure the tweet novelty by duration (in weeks), we discover that online ties and online-to-online ties dominate the distribution of tweets across different number of weeks except for the first week when offline ties dominate. The results (See Table 4) show that tweet novelty matters to offline and online friends not so much in terms of duration but in terms of the number of direct and indirect friends among whom the tweet has circulated.

6.2 "Offline Versus Online" as the Category of Ties by Topic

We plot the topic distribution of tweets retweeted by ties that belong to the offline-online categories on Fig. 4. Twitter-LDA gives us f_t^c, the frequency of tweets of topic t retweeted by ties belonging to category c. We normalize the frequency by f_t, the total frequency of tweets of topic t.

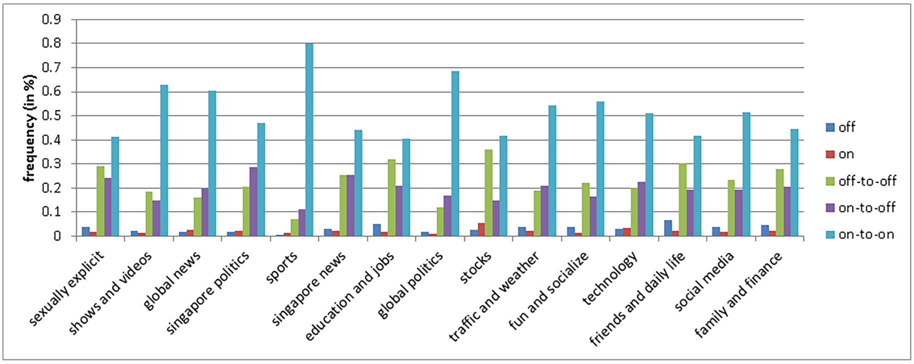


Fig. 4. Frequency of tweets by offline-online categories.

Across all topics, online-to-online ties dominate retweets, confirming the results in Sect. 6.1 that show these types of ties prompt the most retweets. The results also demonstrate that a high frequency of offline ties usually indicates a high frequency of offline-to-offline ties. This phenomenon appears in many topics, including "sexually explicit","shows and videos", "education and jobs","fun and socialize", "friends and daily life", "social media", and "family and finance". We conclude that these topics are more likely retweeted by offline ties, or the friends a user engages with outside of the internet.

Additionally, "global news", "Singapore politics", "sports", and "technology" are topics that are likely to be retweeted by online-to-offline ties or online ties. Meanwhile, other topics point to mixed results. Although the topics of "Singapore news", "global politics", and "traffic and weather" are more likely to be retweeted by offline ties than online ties, they are more likely to be retweeted by online-to-offline ties than offline-to-offline ties. Meanwhile, although the topic "stocks" is more likely to be retweeted by online ties than offline ties, it is more likely to be retweeted by offline-to-offline ties than online-to-offline ties.

When we compare these results to the research work conducted by Natali et al. [10], we can see some similarities as well as discrepancies. Natali et al. discovered that personal tweets were more likely to be disseminated through the stronger ties (reciprocated ties). In our study, personal topics such as "fun and socialize", "friends and daily life", and "family and finance", are also more likely to be disseminated through stronger ties (offline ties). However, while Natali et al. showed that entertainment tweets were more likely to be circulated through weaker ties (unreciprocated ties), our study demonstrates that entertainment-focused topics ("shows and videos") are more likely to be circulated by stronger ties (offline ties). Yet, a different entertainment topic, "sports" is more likely to be disseminated by weaker ties (online ties).

6.3 "Reciprocated Versus Unreciprocated" as the Category of Ties by Tweet Novelty

In order to discover how the different retweet patterns of "offline versus online" ties compare to those observed in "reciprocated versus unreciprocated" ties, we must analyse the retweet patterns of reciprocated and unreciprocated ties using the same dataset. Table 3 calculates the normalized frequency of ties that belong to category c at depth level l ($\hat{f}_l^c$) expressed in percentage. c can be reciprocated or unreciprocated.

The results (See Table 3) show that at all depth levels a higher percentage of reciprocated ties are used to retweet when compared to unreciprocated ties. At level one, the percentage is even greater than one hundred, meaning that on average, each tie is used more than one time to retweet. It is also important to remember that the information that flows through reciprocated ties can go two ways, naturally increasing the likelihood of any information passing through. However, even if we increase the frequency of unreciprocated ties in Table 3 by a factor of two, the frequency of reciprocated ties that is used to retweet is still higher at all depth levels.

Similarly, when duration of tweet (in weeks) is used to measure novelty, reciprocated tie dominates the distribution of tweets (See Table 5 in the Appendix).

Therefore, we cannot distinguish reciprocated-unreciprocated ties by tweet novelty, unlike how we can distinguish offline-online ties.

6.4 "Reciprocated Versus Unreciprocated" as the Category of Ties by Topic

We plot the topic distribution of tweets retweeted by ties that belong to the reciprocated-unreciprocated categories on Fig. 5. Twitter-LDA gives us f_t^c, that is the frequency of tweets of topic t retweeted by ties that belong to category c. We normalize the frequency by f_t, the total frequency of tweets of topic t.

Across all topics, reciprocated ties are used more than unreciprocated ties to retweet. Although these results contradict the results of the research by Natali et al. [10], they are not necessarily invalidated because the dataset used in this study is different than the one used by Natali et al. The contexts of the two studies are also different. In this study we examine the retweets in ego networks, whereas Natali et al. analysed the retweets that span beyond an ego network within a time period.

Table 3. Normalized frequency of ties that belong to the reciprocated-unreciprocated categories at depth level l ($\hat{f}_l^c$) expressed in percentage.

Depth level	Latest timing assumption	
	Reciprocated	Unreciprocated
1	137.25	24.88
2	6.89	0.57
3	1.12	0.11
4	0.28	0.03
5	0.12	0.01
6	0.06	0.00
7	0.03	0.00
8	0.02	0.00
9	0.01	0.00
10	0.01	0.00
>= 11	0.02	0.00
	Most popular assumption	
1	141.41	25.23
2	4.13	0.31
3	0.25	0.04
4	0.03	0.00
5	0.01	0.00
6	0.00	0.00

In conclusion, reciprocated-unreciprocated ties also cannot be distinguished by topics just as how they cannot be distinguished by tweet novelty. Meanwhile, offline-online ties can be distinguished by both criteria.

6.5 Putting It All Together: "Offline or Not and Reciprocated or Not" as Categories of Ties

We also want to know whether combinations of the above tie categories will improve the categorization of ties by making each category more distinguishable from one another.

Table 6 in the Appendix calculates the normalized frequency of ties that belong to category c at depth level l ($\hat{f}_l^c$) expressed in percentage. c can be any of the 10 categories made by combining the offline-online categories and the reciprocated-unreciprocated categories.

The results show that reciprocated-unreciprocated categories can help to explain the behaviours of ties in retweeting. At depth level one, a higher percentage of unreciprocated ties is used to retweet compared to reciprocated ties, regardless of which offline-online categories the ties belong to. Meanwhile, at the

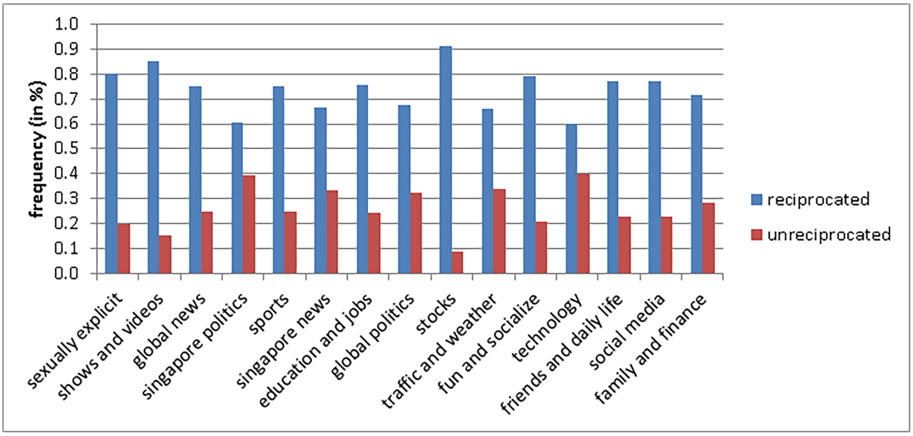


Fig. 5. Frequency of tweets by reciprocated-unreciprocated categories.

depth level two, a higher percentage of reciprocated ties is used to retweet. The results for level three and four are mixed. For some categories a higher percentage of reciprocated ties retweets more, while for other categories a higher percentage of unreciprocated ties retweet more. At the level beyond five, reciprocated offline-to-offline ties are the ones mostly used for retweet.

The results can be explained by the theory of weak ties [5]. At depth level one most tweets are novel, and therefore, the weaker (unreciprocated) ties of each offline-online category are used to retweet. Meanwhile, at depth level five and above, the tweets are old, and therefore, reciprocal offline-to-offline, the strongest category of ties is used to retweet. Although offline-online categories alone cannot distinguish retweet behaviour at depth level one, combinations of offline-online and reciprocal-unreciprocal categories can do so.

Meanwhile, when we look at the tweet novelty in terms of duration (in weeks), a different pattern emerges. Although tweets at week one are also circulated mainly by unreciprocated ties, offline-to-offline ties do not dominate the circulation of old tweets. Therefore, duration of tweets in weeks is again shown not to be as good as the depth of retweet chains to influence the types of ties used.

We also plot the topic distribution of tweets retweeted by ties that belong to the combined categories. We do not show the figure on this paper due to the page limit. Moreover, the results are also inconsequential. In the combined categories, offline-reciprocated and online-unreciprocated ties are more likely to be retweeted across all topics. We can conclude that combined categories cannot be distinguished by topics as well as the offline-online categories can be.

6.6 The Effect of Temporal Changes in Offline/Online Relationships

As we have explained on Section Dataset, temporal changes in offline and online relationships from the year 2011 and 2015 to 2018 can happen. Therefore, the interpretation of the results may downplay the importance of offline friends and exaggerate the importance of online friends. In summary, offline friends are proven to be more important than what the results show in circulating older news and personal tweets. Meanwhile, the huge importance of online-to-online ties in circulating news may be exaggerated.

7 Conclusion

Overall, we have analysed the retweet patterns, specifically tweet novelty and tweet topics, of offline and online ties on Twitter ego networks. We compare our results with the analysis of retweet patterns of reciprocal and unreciprocal ties. We have shown that offline ties and friends who are likely to be offline (offline-to-offline ties) are the ones who tweet old news. The age of tweets should be measured by how many times they have been circulated among direct and indirect friends on Twitter, not by duration. However, in general, online-to-online ties play the most important role in circulating information on Twitter.

Offline ties are also more likely to retweet about family and friends, while online ties are more likely to retweet news. On the other hand, reciprocated and unreciprocated ties show similar retweet patterns. Hence, offline versus online is a more reliable tie category with regard to retweets than reciprocated versus unreciprocated. In terms of practical application, someone who wants to increase a tweet's shelf life and popularise personal tweets should focus more effort on targeting offline friends. On the other hand, online ties should be harnessed for any new marketing campaigns. Our study highlights the importance of the offline-online network paradigm for retweets that cannot be replaced easily, such as by the reciprocated-unreciprocated network paradigm.

A Appendix: Supplementary Results

Table 4. Normalized frequency of ties that belong to the offline-online categories at week w expressed in percentage.

Week	Latest timing assumption				
	Off	On	Off-to-off	On-to-off	On-to-on
1	32.21	28.00	27.33	41.96	83.39
2	0.28	0.37	0.20	0.39	0.68
3	0.08	0.14	0.08	0.13	0.30
4	0.06	0.12	0.05	0.08	0.16
5	0.04	0.08	0.04	0.05	0.14
6	0.02	0.03	0.02	0.05	0.10
>=7	0.33	0.78	0.30	0.79	2.40
	Most popular assumption				
1	31.50	28.03	26.89	42.14	83.45
2	0.26	0.37	0.19	0.39	0.69
3	0.07	0.14	0.07	0.12	0.31
4	0.05	0.12	0.05	0.08	0.16
5	0.04	0.08	0.04	0.05	0.14
6	0.02	0.03	0.02	0.04	0.10
>=7	0.32	0.77	0.30	0.79	2.40

Table 5. Normalized frequency of ties that belong to the reciprocated-unreciprocated categories at week w expressed in percentage.

Week	Latest timing assumption	
	Reciprocated	Unreciprocated
1	140.02	24.99
2	1.25	0.14
3	0.47	0.06
4	0.24	0.04
5	0.20	0.02
6	0.14	0.01
>= 7	2.79	0.33
	Most popular assumption	
1	140.74	24.96
2	1.27	0.14
3	0.48	0.05
4	0.24	0.04
5	0.20	0.02
6	0.14	0.01
>= 7	2.72	0.37

Table 6. Normalized frequency of ties that belong to the combined categories at depth level l ($\hat{f}_l^c$) expressed in percentage.

Depth level	Latest timing assumption									
	Reciprocated					Unreciprocated				
	Off	On	Off-to-off	Off-to-on	On-to-on	Off	On	Off-to-off	Off-to-on	On-to-on
1	26.96	8.03	23.05	19.97	53.06	45.85	32.80	22.17	44.54	70.24
2	2.29	0.49	1.00	0.67	2.27	0.62	0.64	0.59	0.55	1.31
3	0.09	0.01	0.17	0.07	0.34	0.21	0.05	0.06	0.05	0.23
4	0.02	0.01	0.06	0.01	0.07	0.07	0.01	0.03	0.01	0.05
5	0.00	0.00	0.04	0.00	0.01	0.00	0.00	0.00	0.00	0.01
6	0.01	0.00	0.02	0.00	0.00	0.00	0.01	0.01	0.00	0.00
7	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
>= 11	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Most popular assumption									
1	27.07	8.01	23.41	20.39	53.99	45.02	33.50	22.34	45.11	71.81
2	1.63	0.37	0.50	0.37	1.31	0.34	0.25	0.32	0.29	0.89
3	0.03	0.02	0.04	0.01	0.08	0.00	0.03	0.02	0.00	0.13
4	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.02
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

References

1. Backstrom, L., Kleinberg, J.: Romantic partnerships and the dispersion of social ties: a network analysis of relationship status on Facebook. In: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW 2014), pp. 831–841. ACM (2014)
2. Boyd, D., Golder, S., Lotan, G.: Tweet, tweet, retweet: conversational aspects of retweeting on Twitter. In: The 43th Hawaii International Conference on System Sciences (HICSS 2010), pp. 1–10. IEEE Publishing (2010)
3. Enli, G.S., Skogerbø, E.: Personalized campaigns in party-centred politics. Inf. Commun. Soc. **16**(5), 757–774 (2013)
4. Gilbert, E., Karah, K.: Predicting tie strength with social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2009), pp. 211–220. ACM (2009)
5. Granovetter, M.S.: The strength of weak ties. Am. J. Sociol. **78**(6), 1360–1380 (1973)
6. Java, A., Song, X., Finin, T., Tseng, B.: Why we Twitter: understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis (KDD 2007), pp. 59–65. ACM (2007)
7. Kupavskii, A., et al.: Prediction of retweet cascade size over time. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pp. 2335–2338. ACM Press (2012)
8. 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 (WWW 2010), pp. 591–600. ACM Press (2010)

9. Marsden, P.V., Campbell, K.E.: Measuring tie strength. Soc. Forces **63**(2), 482–501 (1984)
10. Natali, F., Carley, K.M., Zhu, F., Huang, B.: The role of different tie strength in disseminating different topics on a microblog. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2017), pp. 203–207. ACM (2017)
11. Natali, F., Zhu, F.: A comparison of fundamental network formation principles between offline and online friends on Twitter. In: Wierzbicki, A., Brandes, U., Schweitzer, F., Pedreschi, D. (eds.) NetSci-X 2016. LNCS, vol. 9564, pp. 169–177. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-28361-6_14
12. Peng, H.k., Zhu, J., Piao, D., Yan, R., Zhang, Y.: Retweet modeling using conditional random fields. In: IEEE 11th International Conference conference on Data Mining Workshops (ICDMW 2011), pp. 336–343. IEEE Computer Society (2011)
13. Shi, G., Shi, Y., Chan, A.K., Wang, Y.: Relationship strength in service industries. Int. J. Market Res. **51**(5), 659–685 (2009)
14. Suh, B., Hong, L., Pirolli, P., Chi, E.H.: Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In: Proceedings of the 2010 IEEE Second International Conference on Social Computing, pp. 177–184. IEEE Publishing (2010)
15. Thomases, H.: Twitter Marketing: An Hour a Day. Wiley, New Jersey (2009)
16. Xie, W., Li, C., Zhu, F., Lim, E.P., Gong, X.: When a friend in Twitter is a friend in life. In: The 4th ACM Web Science Conference (WebSci 2012), pp. 344–347. ACM (2012)
17. Zhao, J., Wu, J., Feng, X., Xiong, H., Xu, K.: Information propagation in online social networks: a tie-strength perspective. Knowl. Inf. Syst. **32**(3), 589–608 (2012)
18. Zhao, W.X., et al.: Comparing Twitter and traditional media using topic models. In: Clough, P., et al. (eds.) ECIR 2011. LNCS, vol. 6611, pp. 338–349. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-20161-5_34