

Hashtag Popularity on Twitter: Analyzing Co-occurrence of Multiple Hashtags

Nargis Pervin¹(✉), Tuan Quang Phan¹, Anindya Datta¹, Hideaki Takeda²,
and Fujio Toriumi³

¹ National University of Singapore, Singapore, Singapore
{nargisp,phantq,datta}@comp.nus.edu.sg

² National Institute of Informatics, Tokyo, Japan
takeda@nii.ac.jp

³ The Tokyo University, Tokyo, Japan
tori@sys.t.u-tokyo.ac.jp

Abstract. Hashtags increase the reachability of a tweet to manifolds and consequently, has the potential to create a wider market for brands. The frequent use of a hashtag features it in the Twitter trending list. In this study we want to understand what contributes to the popularity of a hashtag. Further, hashtags generally come in groups in a tweet. In fact, an investigation on a real world dataset of Great Eastern Japan Earthquake reveals that 50% of hashtags appear in a tweet with at least another hashtag. How this co-occurrence of hashtags affects its popularity is also not addressed heretofore, which is the focus herein. Results indicate that if a hashtag appears with one or more other similar hashtags, popularity of the hashtag increases. In contrast, if a hashtag appears with dissimilar hashtags, popularity of the focal hashtag decreases. The results reverse when dissimilar hashtags come along with a URL.

Keywords: Twitter · Hashtag · Hashtag co-occurrence · Metacognitive experience

1 Introduction

In August 2007, Chris Messina tweeted on his Twitter account “how do you feel about using # (pound) for groups? As in #barcamp [msg]?” It was claimed as the first ever hashtag [19] on Twitter and since then this became a unique strategy for categorizing messages which can properly lead individuals to conversations and discussions pertaining to a specific topic [7, 15]. Social media is fast paced and no one has the time all day long to sift through his timeline to read everything being posted. That is where hashtags are significant. It can generate immediate, live, and interactive reactions and responses to specific topics. People use hashtags while watching their favorite TV program, listening to a debate on the radio, promoting a product, or running a campaign. It has been shown that when individuals used a hashtag within their tweet, engagement can increase as much as 100% and for brands it could get an increase of 50% [6]. This is because

a hashtag immediately expands the reach of the tweet beyond followers of the tweet author and hence is reachable to anyone interested in that hashtag phrase or keyword.

During the 2012 presidential election, both Obama and Romney used hashtags to campaign through social media. The craze for hashtags is so high that people are willing to pay even \$3,000 to rent a “social media wedding concierge” [8].

From the preceding discussion, it is transparent that the importance of hashtags is enormous, which motivates us to investigate the characteristics of these hashtags. The abundance of information to which we are exposed through online social networks exceeds the amount of information we can consume. Hence, the hashtags compete with each other to attain our limited attention. Users can remember a bounded number of different hashtags at a time, which suggests that one hashtag is remembered by the users at the expense of others [23]. How many users will adopt a hashtag determines its popularity. This adoption solely depends on how people find it meaningful and attractive which is concluded from their metacognitive experience. Metacognitive experiences are those experiences that are related to the current, on-going cognitive endeavor while metacognition refers to a level of thinking that involves active control over the process of thinking that is used in learning situations [12, 13]. The detailed discussion of metacognition can be found in the literature review section.

On inspecting tweets containing hashtags, one can notice that hashtags usually come in groups, i.e., a single tweet contains more than one hashtag. A preliminary analysis on our data set reveals that tweets containing multiple hashtags get diffused more compared to tweets having a single hashtag. Here the decisive question arises whether the characteristics of the hashtags appeared together are random or it carries certain pattern, which is the focus of this study. The popularity of one hashtag might boost the popularity of others when they appear together. For instance, say hashtag h becomes trendy on Twitter. Now, users start using h with h_1 which increases the discoverability of h_1 also. In such circumstances, we believe that there can be three main possibilities: a) popularity of h takes off further, b) hashtag h_1 becomes more popular, and c) hashtag h_1 replaces hashtag h . To understand this phenomenon, it is necessary to investigate the change of popularity of a hashtag h when co-appeared with other hashtags h_1, h_2 , etc. We investigate the popularity of a hashtag measured by the number of distinct users who have adopted / used it and model the popularity using regression technique considering both network variables and content variables of hashtag.

We postulate that when a hashtag appears with multiple hashtags, it increases the popularity of the focal hashtag. Next, we investigate the nature of these co-appearing hashtags in terms of similarity. Dissimilar hashtags increase the metacognitive difficulty of the users [11], but when used with URLs it adds more information and brings surprisingness to the tweet, which in turn increases the popularity of hashtags. Earlier studies [9] have shown that when hashtags appeared with a URL in a tweet, retweetability of that tweet escalates which is in line with our hypothesis. We examine this phenomenon using the Great

Eastern Japan earthquake data set and also check whether the external event (e.g., earthquake) has any impact on this process.

Our investigation on hashtag popularity supports both the hypotheses. Further, it has been shown that only hashtag content can predict the popularity of hashtags and network variables turned out insignificant in popularity prediction, which is in agreement with [21]. Running the model in three different windows we found that direction of the impact of the independent variables on popularity are same, though the strength of impacts (coefficients) was larger at the time of the event which reduced again back to normal after the event.

Moreover, we have investigated these properties at a granular level. We model hashtag retweetability at the dyad (user-retweeter pair) level. This model will allow us to understand the user level attributes that play role in popularity.

2 Literature Review

What does motivate people to share information? Sharing information with friends is considered to be a communal act in online social network sites. People share YouTube videos, Facebook posts, or tweets on Twitter. While a massive amount of information gets generated online, only a handful of them get noticed and shared. This leads to the straightforward question what makes a piece of content more share-worthy than others. Researches have been carried in the viral-marketing area to unfold the characteristics of the content that goes viral [1–3]. However, the main query lies in why people share information in the first place and what type of content gets shared. Consumers might share some content online for several reasons, e.g., altruistic reasons (e.g., to help others) or for self-enhancement purposes (e.g., to appear knowledgeable, see [25]).

Human reasoning is accompanied by metacognitive experiences. The assumptions about what makes it easy or difficult to think of certain things or to process new information contribute to what exactly people conclude from their metacognitive experiences. Researches showed evidence that people are more likely to advocate a statement as true when the color in which it is printed makes it easy to read (e.g., [12, 13]). [14] describes that accessibility and processing fluency both pertain to the ease of recalling and processing new information. Moreover, repeated exposures lead to the subjective feeling of perceptual fluency, which in turn influences liking [13]. On the other hand, [11] experimentally showed that metacognitive difficulty increases the attractiveness of a product by making it appear unique or uncommon.

In this work we want to investigate why some hashtags go more viral than others? A hashtag is a word or phrase preceded by a hash sign (#), used on Twitter to identify messages on a specific topic. This works as a user-defined index term to link several topics or events together. [26] examined the dual effect of hashtags on Twitter: a) a symbol of a community membership and b) a bookmark. In this paper they investigated which of the two reasons strive people to adopt a hashtag. The prediction using SVM technique incorporates social network variables like indegree, outdegree of nodes (number of people

retweeted the hashtag), relevance, popularity of the hashtags, length (number of characters), age of the hashtag. The dataset used in this study was Twitter data on politics.

Popularity of the hashtag determines how many users will adopt a particular hashtag. Using a 25 week Twitter data, [21] reported hashtag frequency prediction on a weekly basis using regression technique. Features used in the regression model were extracted from the hashtag itself (e.g., number of characters in the hashtag) and their experiment shows that hashtag popularity can be predicted using only the content features of the hashtags instead of using the costly graphical features extracted from tweets. However, [10] claimed that contextual features are more effective than content features which can be explained by the fact that community graph plays an important role in information diffusion. Clarity of hashtag, number of words in a hashtag, user count, tweet count, etc. were used to predict the popularity. However, on Twitter a large number of hashtags are generated every day and people cannot remember all of them. Using an agent-based simulation model, [23] claimed that the users can remember a bounded number of different memes at a time, which suggests that one meme is remembered by the users at the expense of others. The proposed retweet model assumes the finite memory of the users where memes are registered and by the friend and follower links, some other users can read the meme posted. However, a careful investigation of the usage of hashtags needs to be done. On inspecting tweets containing hashtags, one can notice that hashtags usually come in groups, i.e., a single tweet contains more than one hashtag. A preliminary analysis reveals that tweets containing multiple hashtags get diffused more than tweets having a single hashtag. It will be interesting to investigate “Are these characteristics of the hashtags appeared together random or does it carry certain patterns?” Moreover, in the time of emergency, the adoption of hashtags might change. Using a 2011 Japan earthquake data, this chapter investigates how a hashtag becomes popular.

3 Dataset Description

In this study, we have used data set from the 2011 Japan earthquake. We used a Twitter dataset collected during the earthquake in 2011 described thoroughly in [20]. Dataset collection procedure has been discussed briefly here:

- First, a set of tweets has been collected from Twitter streaming API for tweets during the event.
- Next, for all these tweets the user details has been crawled using the same API along with the follower IDS.
- For all these users the tweets are collected for 20 days of time period.

The dataset covers a period of 20 days (from 5th March, 2011 to 24th March, 2011), and consists of 362,435,649 tweets posted by 2,711,473 users in Japan. This dataset is remarkable by its completeness: 80% to 90% of all published tweets by these users were present in this dataset. It should be noted that the dataset consists of tweets of Japanese Twitter users. Hence, major proportion of tweets (98%) in the data set is written in Japanese.

4 Solution Details

4.1 Building Research Hypotheses

Synonymous hashtags when appeared together, it will increase the visibility of the tweet contrary to dissimilar hashtags. Dissimilar hashtags will increase the metacognitive difficulty of the users [11], but when used with URLs it would add more information, bring surprisingness to the tweet, and could increase the popularity of hashtags. Thus, we postulate two hypotheses:

Hypothesis 1. *Hashtag popularity increases when it appears with other hashtags.*

Hypothesis 2. *When co-appearing hashtags are dissimilar, presence of URLs increases hashtag popularity.*

We have investigated what happens when the hashtags co-appear. First, we examine whether the co-occurrence of hashtags plays any role in hashtag popularity and then we calculate the distance among the co-appearing hashtags to test whether the distance among the tags has any impact on its popularity. Additionally, we have also considered the interaction effect of the URL and the distance among the co-appearing hashtags. We have modeled hashtag popularity using the content variables of hashtags and the user specific variables. We have presented two models, one with only the hashtag specific variables, and another model with the hashtag specific and dyad specific variables.

4.2 Factors Considered for Hashtag Popularity

To investigate the factors impacting popularity hashtag specific, dyad specific, and control variables are considered as follows:

Hashtag Specific Variables

Length of Hashtag: The hashtag has been extracted from the tweet content by searching words that start with “#”. For all hashtags we counted the number of characters in that hashtag. Very long hashtags are not economical in the Twitter perspective as tweets are limited to only 140 characters. On the other hand, very small hashtags (e.g., abbreviated hashtags containing only 2 or three letters) do not contain sufficient information to understand.

Number of Words: Clarity of the hashtag is important for its adoption. Hashtags, which contain multiple words are easy in order to follow the context from the hashtag itself. However, finding the word segments from a hashtag in the Twitter context is not straightforward as Twitter users use Twitter specific lingual. For the same reason, we counted the number of words in the hashtag by separating the capital letters or other special separator characters (e.g., underscore (-), plus (+) etc.).

Contains Capital Letters: This is a boolean variable computing the presence of capital letters in the hashtag. The value of the variable = 1 if the hashtag contains capital letters and 0 otherwise.

Contains Digits: This is a boolean variable denoting the presence of digits in the hashtag. The value of the variable = 1 if the hashtag contains digits and 0 otherwise.

Contains Other Separators: This is a boolean variable computing the presence of other separators in the hashtag, e.g., underscore (`_`), plus (`+`). The value of the variable = 1 if the hashtag contains other separators and 0 otherwise.

Appeared with Other Hashtag: This determines whether a hashtag appeared with other hashtags or not. If the hashtag appears with other hashtags then the value of the variable is the number of hashtags it appeared with and 0 otherwise. This is a time series variable indicating that the value of the variable determines in a particular time unit, whether the hashtag appeared with others or not.

Distance: For the co-appearing hashtags we compute the distance between the hashtag pairs. If more than two hashtags appear with the focal hashtag then the average distance of the hashtag pairs are considered. For each pair of hashtags we have calculated the Levenshtein distance [24] among them.

Inclusion of URLs: Earlier studies [18] have shown that inclusion of URLs in the tweet increases a tweet's retweetability. Our previous study also supports this finding (as described in the previous chapter, Chap. 3). Moreover, we also observed that the presence of both hashtags and URLs in the tweet increases its popularity. Therefore, we compute this variable as a boolean variable denoting the presence of URLs in the tweet. $URL = 1$ if the tweet contains a URL, 0 otherwise. If a hashtag appears in more than one tweet, we compute the average number of times the focal hashtag appeared with URLs. We place $URL = 1$ if the average number of tweets > 0 , 0 otherwise.

DistanceXURL: To examine the moderating effect of the URL on distance of co-appearing hashtags we compute the interaction variable of distance and boolean URL.

$$DistanceXURL = distance \times URL$$

Frequency of Hashtag: For each hashtag h we calculate the frequency of hashtag as the number of times h has been retweeted per minute.

Age of Hashtag: For each hashtag h we compute the age of the hashtag since it has been used by some user. Unit of time used here is an hour.

Dyad Specific Variables

Frequency of Dyad: For each hashtag h we calculate the frequency of hashtag at the dyad level. Hence, dyad frequency (per minute) is computed as the number of times user $u_{retweeter}$ retweets a tweet by u_{author} , that contains hashtag h .

PageRank of Author and Retweeter: Each user on Twitter has a number of followers and followees which can be thought of as incoming and outgoing links from a web page. Similar to web pages, we can also compute the PageRank of a user to enumerate his popularity. However, in our case we formulated the retweet network of the users where direction indicates the reverse direction of information flow from (retweeter \rightarrow author). Instead of using the PageRank computed on the follower-followee network, we computed the PageRank based on the retweet network. In this case, unlike otherwise, computed PageRank determines the activeness and actual influence of the users.

For both author and retweeter of the tweet, we compute the PageRank ($PageRank_{author}$ and $PageRank_{retweeter}$).

Betweenness Centrality of Author and Retweeter: Betweenness centrality is a measure of a node's centrality in a network. It is equal to the number of shortest paths from all vertices to all others that pass through that node. We have measured the betweenness centrality of the users on the retweet network.

For both author and retweeter of the tweet, we compute the betweenness centrality ($betweenness_{author}$ and $betweenness_{retweeter}$).

Relationship between Dyad (Author and Retweeter): On Twitter, a tweet can be retweeted by author's followers or friends. However, if a tweet becomes popular, this can be retweeted by retweeters even if they do not have any relationship with the author of the tweet.

Control Variables

Below are two control variables used in the model:

Day of Week: Day of the week [22] might have an impact on the popularity of the hashtag. [22] finds that day of the week controls traffic on Twitter, while Monday to Thursday the tweet volume increases, Friday it slows down. On Mondays users usually use Monday specific hashtags more frequently (#Monday, #mondayfever). On the other hand, on Saturdays and Sundays people write more fun-filled hashtags like #supersunday, #saturdayale.

Time of the Day: Twitter gets the most traffic during 9am-3pm from Monday to Thursday [22]. We also include this as a control variable in the popularity model (Table 1).

Model Specifications. In this model the dependent variable is the retweet count of tweets containing a specific hashtag for a specific dyad (retweeter \rightarrow author pair).

$$RetweetCount_{i,j,t} = \alpha + \sum_i H(i, t) + \sum_j D(j, t) + \sum_k C(k, t) + \varepsilon$$

$H(i, t)$, $D(j, t)$, $C(k, t)$ refer to the vector of hashtag specific variables, dyad specific variables, control variables respectively.

Table 1. Variables affecting hashtag popularity

	Variable	Meaning
Dependent variable	numDistinctUsers	number of distinct users who adopted/used the hashtag
Network variables	PageRank	average PageRank of the users using the hashtag in retweet network
	betweenness	average betweenness centrality of the users using the hashtag in retweet network
Content variables	hasCaps	value = 1 if the hashtag contains capital letters, = 0, otherwise
	hasDigits	value = 1 if the hashtag contains digits, = 0, otherwise
	hasOther	value = 1 if the hashtag contains other separators, = 0, otherwise
	numWords	number of words in the hashtag
	length	length of the hashtag
	appearedWithOthers	number of hashtags #h appeared with
	distance(h,H)	average distance of #h with all hashtags in H
	isURL	boolean variable indicating if the tweet contains URLs
Control variables	timeOfDay	24 hours have been divided into 5 time-windows, morning(7am-10am), noon (11am-3pm), afternoon (4pm-7pm), evening (8pm-11pm), night (12am -6am)
	dayOfWeek	day of the week is coded as dummy variable
	tagAge	time since the tweet is composed (in hour)
	tagFreq	frequency of the hashtag per unit time (minute)
Dyad specific variables	dyadFreq	dyad (retweeter → author) frequency per unit time (measured in minute)
	relationship	boolean variable denoting follower-followee relationship, value = 1 if relationship exists and 0 otherwise

$$H(i, t) = [hasDigits, hasCaps, numWords]'_i \\ + [distance, appearedWithOthers, isURL, distanceXisURL]'_{i,t}$$

$$D(j, t) = [Pagerank_{author}, betweenness_{author}, Pagerank_{retweeter}, \\ betweenness_{retweeter}]'_{j,t} + [relationship]_j$$

$$C(k, t) = [dayOfWeek, timeOfDay, age]'_{k,t}$$

Table 2. Hashtag popularity model at the dyad level

Variable	Pre-event	During-event	Post-event
<i>dyadFreq</i>	0.001***	0.002***	0.005***
<i>length</i>	-0.004***	-0.004***	-0.004***
<i>numWords</i>	-0.006***	0.002	0.017***
<i>length</i> ²	0.001***	0.001***	0.001***
<i>numWords</i> ²	0.001**	0.000	-0.005***
<i>tweetCount</i>	0.009***	0.046***	0.016***
<i>hasCaps</i>	0.013***	0.010***	0.029***
<i>hasDigits</i>	-0.028***	-0.027***	-0.039***
<i>hasOther</i>	0.005***	0.024***	0.029***
<i>appearedWithOthers</i>	0.008***	0.003***	0.001***
<i>tagAge</i>	-7.7E-05***	-7E-05***	-5.3E-05***
<i>tagFreq</i>	3.63E-05***	1.07E-05***	9.83E-06***
<i>distance</i>	-0.001***	-0.004***	-0.002***
<i>isURL</i>	0.085***	0.093***	0.104***
<i>distancesXisURL</i>	0.004***	0.012***	0.006***
<i>PageRank_{author}</i>	0.037	0.214	-0.672
<i>betweenness_{author}</i>	0.001E-10***	0.001E-10***	-3.01E-10***
<i>PageRank_{retweeter}</i>	0.317	-2.173***	-1736.72***
<i>betweenness_{retweeter}</i>	0.001E-09***	0.001E-09***	-1.21E-09***
<i>relationship</i>	0.772***	0.868***	0.911***

***p < 0.01, **p < 0.05, *p < 0.1

5 Data Analysis and Findings

The Great Eastern Japan earthquake dataset has been used to examine this phenomena. The dataset consists of 1.3 million observations with 521028 hashtags from 0.1 million users. The model investigates the effect of URLs in hashtag popularity at two levels - first at the hashtag level and second at the dyad (user-retweeter pair) level.

5.1 Data Preparation

Using the Twitter dataset, we found the hashtags from each tweet by simply searching words that start with “#”. From the primary tweet dataset we prepared a dataset where each row contains the timestamp of the tweet, the list of hashtags in the tweet, author of the tweet, boolean variable indicating whether the tweet contains URLs. Our tweet dataset (5th-24th March) has been divided into three time windows, pre-earthquake (5th – 10th March), during-earthquake (11th – 16th March), and post-earthquake (17th – 24th March).

5.2 Data Analysis

We formed the retweet network from the tweets in our database, where the nodes represent the users and directed links represent the reverse of direction (retweeter \rightarrow user) of information flow. Network variables such as PageRank, betweenness centrality are measured using the retweet network.

Besides network variables, hashtag contents have been analyzed. Since tweets are limited to 140 characters, each character in the tweet is very costly. Therefore, very long hashtags are not preferable. Moreover, long and complex hashtag increases the cognitive load and are not easy to understand [16,17]. For the same reason, the hashtag is analyzed and the number of words in the hashtag is counted. The intuition behind this is that number of words in a hashtag increases its clarity and the hashtag itself carries more contextual information about the tweet itself. If the words are separated by special characters or by capital letters it is easy to determine the words in the hashtag.

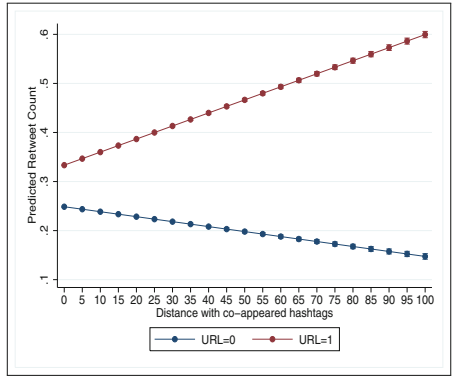
There tends to be more than one hashtag in a tweet. A preliminary analysis has shown that if a hashtag appears with others, popularity of the focal hashtag increases. Moreover, we included the distance among the co-appearing hashtags to examine the effect of distance on its popularity. Previous studies have experimented that inclusion of URLs and hashtags increase the chance of retweetability [4]. In this study, we have included the boolean variable for URL to examine its effect on similarity/dissimilarity of hashtags.

Next, we examine whether there is an effect of an event on popularity and Japan earthquake data is used for that reason. The effects of the variables are investigated in all the three time periods (pre-, during- and post-event time-windows) independently.

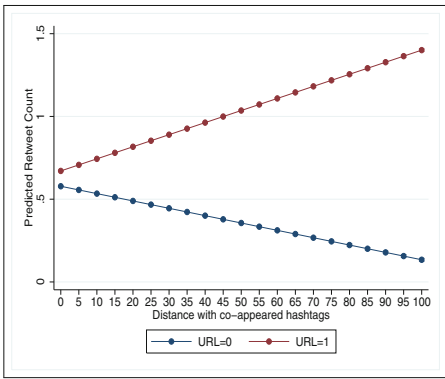
5.3 Findings and Discussion

To understand the effect of hashtag popularity at user level, we modeled retweetability of a hashtag for dyads, where each dyad consists of the user who tweeted and the one who is retweeted. The model is run with the hashtag and dyad specific variables along with the control variables as listed earlier. Retweet count (per hashtag per dyad) is considered as the dependent variable. We have divided the dataset into three different time-windows and regression technique has been used in all cases. Results for the three time-windows have been shown in Table 2. The findings show that dyad specific and hashtag specific variables considered in the model have significant impacts on retweet count. The dyad frequency have significant positive impact on retweet count per unit time, which suggests that the users retweet hashtags from Twitter users they usually retweet from. Moreover, in the Twitter world if the user and retweeter has follower-followee relationship, then retweet count of a hashtag increases opposed to retweet practices from non-follower/friend relationship.

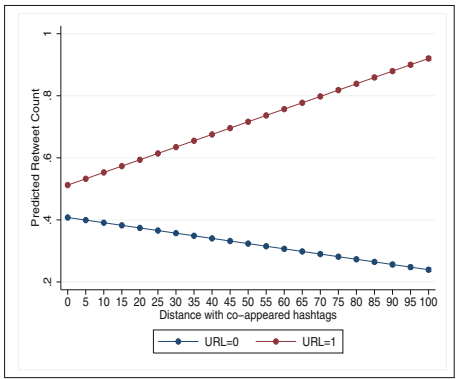
It is also clear from our dataset that length does not have a significant impact on popularity, however, the number of words in a hashtag has inverse u-shaped impact. While the number of words has a positive impact on popularity, too many



(a) Pre-event Time Window



(b) During-event Time Window



(c) Post-event Time Window

Fig. 1. Interaction plot on distance and URLs in pre-, during-, and post-event window (dyad Level)

words in a hashtag have a negative impact. Intuitively, this is comprehensible as the number of words increases the clarity of hashtag at first, but as the number of words grows in abundant the hashtag becomes complex. Further, the presence of digits or capital letters in the hashtag has negative impact on popularity, but popular hashtags mostly contain other separators to segregate the words in hashtag phrases.

On the other hand, while hashtag frequency (for a specific dyad) has a positive impact, which suggests that user retweets tweet containing specific hashtags many times. Above all, we have examined the interaction effect of hashtag dissimilarity with presence of URLs and we receive similar impact as seen in model 1 at hashtag level.

Figure 1a plots the interaction effect of URLs on distance in the pre-event time-window at the dyad level.

It shows that when there is a URL in a tweet (along with a hashtag), the retweet count of that hashtag by a specific dyad is more compared to when

there is no URLs in the tweet. In this case, when the distance among the co-appearing hashtags is higher, introduction of a URL results in higher retweet count, compared to when a URL appears with similar hashtags.

To determine if the patterns characterizing the significant interactions conform to the directions as proposed in the research hypotheses, we have plotted the interaction effects (Fig. 1a, b, and c) for all three time-windows. This procedure was introduced by [5] for all interaction cases. Figure 1a, b, and c show the disordinal (or crossover) interaction of URLs on the relationships of hashtag similarity with hashtag popularity. Figure 1b and Fig. 1c plot the interaction effect of URLs on similarity with retweetability of a hashtag at dyad level in the during-event and post-event windows respectively. Similar to hashtag level analysis, one can note that the retweet count at dyad level has similar result as in Fig. 1a. The appearance of URLs when the hashtags are similar has significantly less impact compared to when the hashtags are dissimilar.

Overall, we can see that the presence of URLs with similarity (or dissimilarity) of hashtags has significant impact at dyad level, which suggests that choice of hashtag is driven by individual metacognitive experiences.

6 Conclusion

Hashtag in a tweet starts with a # symbol and is used before a relevant keyword or phrase in a tweet, which facilitates to categorize the tweets into different topics. Consequently, it becomes convenient to search them in a Twitter search. However, in practice hashtags mostly come in groups, i.e., one can find more than one hashtag in a tweet. Are these co-appearing hashtags random or do they carry certain patterns? In this study, we have investigated the characteristics of the co-appearing hashtags. Findings show that the popularity of a hashtag increases when a hashtag appears with other hashtags. Moreover, the similarity / dissimilarity of the hashtags plays crucial role in hashtag popularity. Results indicate that when similar hashtags appear together the hashtag popularity increases as opposed to dissimilar hashtags. To our surprise when the dissimilar hashtags appear with a URL, then the effect is reversed. This phenomenon can be explained by the fact that when dissimilar items co-appear it increases the meta-cognitive load and introduces confusion, but with the provision of extra information (e.g., URL), this becomes surprising and interesting to users resulting adoption of those hashtags together. These findings can help to diffuse new hashtags by coupling with similar popular hashtags or adding the pinch of surprise with dissimilar hashtags and a URL. It also can help the practitioners implement efficient policy making for product advertisement with brand hashtags.

In this study we have analyzed how these effects change due to an event and observed that due to event the trend of the effect was the same. However, to generalize the effect of the event on this phenomenon, it will be interesting to investigate on a separate data set. We plan to investigate this in our future research.

References

1. Aral, S., Muchnik, L., Sundararajan, A.: Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Nat. Acad. Sci.* **106**(51), 21544–21549 (2009)
2. Berger, J., Milkman, K.L.: What makes online content viral? *J. Mark. Res.* **49**(2), 192–205 (2012)
3. Berger, J., Milkman, K.: Social transmission, emotion, and the virality of online content. Wharton research paper (2010)
4. Boyd, D., Golder, S., Lotan, G.: Tweet, tweet, retweet: conversational aspects of retweeting on twitter. In: *Proceedings of the 2010 43rd Hawaii International Conference on System Sciences, HICSS 2010*, pp. 1–10. IEEE Computer Society, Washington, DC, USA (2010)
5. Cohen, J., Cohen, P., West, S., Aiken, L.: *Applied Multiple Regression/Correlation Analysis for the Social Sciences*. L. Erlbaum Associates, Hillsdale (1983)
6. Cooper, S.: Big mistake: making fun of hashtags instead of using them. October 2013. <http://www.forbes.com/sites/stevecooper/2013/10/17/big-mistake-making-fun-of-hashtags-instead-of-using-them/>. Accessed 2nd August 2014
7. Doctor, V.: What’s the point of all these hashtags? December 2012. <http://www.hashtags.org/platforms/twitter/whats-the-point-of-all-these-hashtags/>. Accessed 2nd August 2014
8. Hathaway, J.: Now you can get a \$3,000 “social media concierge” for your wedding, March 2014
9. Hughes, A., Palen, L.: Twitter adoption and use in mass convergence and emergency events. *Int. J. Emerg. Manage.* **6**(3), 248–260 (2009)
10. Ma, Z., Sun, A., Cong, G.: On predicting the popularity of newly emerging hashtags in twitter. *J. Am. Soc. Inf. Sci. Technol.* **64**(7), 1399–1410 (2013)
11. Pocheptsova, A., Labroo, A.A., Dhar, R.: Making products feel special: when metacognitive difficulty enhances evaluation. *J. Mark. Res.* **47**(6), 1059–1069 (2010)
12. Reber, R., Schwarz, N.: Effects of perceptual fluency on judgments of truth. *Conscious. Cogn.* **8**(3), 338–342 (1999)
13. Reber, R., Winkielman, P., Schwarz, N.: Effects of perceptual fluency on affective judgments. *Psychol. Sci.* **9**(1), 45–48 (1998)
14. Schwarz, N.: Metacognitive experiences in consumer judgment and decision making. *J. Consum. Psychol.* **14**(4), 332–348 (2004)
15. Shirley, T.: Why hashtags are so important. August 2014. <http://www.thelasthurdle.co.uk/hashtags-important/>. Accessed 2nd August 2014
16. Song, H., Schwarz, N.: If it’s hard to read, it’s hard to do processing fluency affects effort prediction and motivation. *Psychol. Sci.* **19**(10), 986–988 (2008)
17. Song, H., Schwarz, N.: If it’s difficult to pronounce, it must be risky fluency, familiarity, and risk perception. *Psychol. Sci.* **20**(2), 135–138 (2009)
18. 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, SOCIALCOM 2010*, pp. 177–184. IEEE Computer Society, Washington, DC, USA (2010)
19. Sweeney, D.: Can you legally own a twitter hashtag? March 2012. <http://www.socialmediatoday.com/content/can-you-legally-own-twitter-hashtag>. Accessed 2nd August 2014

20. Toriumi, F., Sakaki, T., Shinoda, K., Kazama, K., Kurihara, S., Noda, I.: Information sharing on twitter during the 2011 catastrophic earthquake. In: Proceedings of the 22nd International Conference on World Wide Web companion, WWW 2013 Companion, pp. 1025–1028. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2013)
21. Tsur, O., Rappoport, A.: What’s in a hashtag?: content based prediction of the spread of ideas in microblogging communities. In: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, WSDM 2012, pp. 643–652. ACM, New York, NY, USA (2012). <http://doi.acm.org/10.1145/2124295.2124320>
22. TweetSmarter: The ultimate guide to finding the best time to tweet (2011). <http://blog.tweet smarter.com/retweeting/when-is-the-best-time-to-tweet/>
23. Weng, L., Flammini, A., Vespignani, A., Menczer, F.: Competition among memes in a world with limited attention. *Scientific reports* 2 (2012)
24. Wikipedia: Levenshtein distance, July 2014
25. Wojnicki, A.C., Godes, D.: Word-of-mouth as self-enhancement. HBS marketing research paper (2008)
26. Yang, L., Sun, T., Zhang, M., Mei, Q.: We know what@ you# tag: does the dual role affect hashtag adoption? In: Proceedings of the 21st International Conference on World Wide Web. pp. 261–270. ACM (2012)