Proceedings of the Tenth International AAAI Conference on
Web and Social Media (ICWSM 2016)

Understanding Communities via Hashtag Engagement: A Clustering Based Approach

Orianna DeMasi *
University of California, Berkeley
Berkeley, CA 94720
odemasi@eecs.berkeley.edu

Douglas Mason †
Pinterest
San Francisco, CA 94103
douglas@pinterest.com

Jeff Ma
Twitter, Inc.
San Francisco, CA 94103
jma@twitter.com

Abstract

We develop insight into community use of hashtags on social media and find that hashtags with behavior indicative of real world communities are more engaging. To do this, we study the relationship of hashtag usage with user engagement on Twitter. Hashtag engagement is useful as a surrogate measure of how active community members are. We develop a framework for describing hashtag temporal usage, show the existence of 4 broad classes of hashtags, and show that the engagement of a hashtag varies significantly between classes. Periodically used hashtags, such as for TV shows and weekly community chats, are the most engaging, while hashtags relating to events are the least engaging. Looking at how community dynamics vary within this framework reveals that a hashtag being used more frequently is not positively correlated with it being more engaging. We then explore the periodically used hashtags and find negative correlations with diversity of the user base, which implies concentrated communities are the most engaging. We conclude by studying a set of community conversation-oriented hashtags and find these hashtags to be more engaging than other hashtags, regardless of dynamic type. Our findings support the hypothesis that hashtags with stronger community behavior are more engaging.

Introduction

Hashtags have become a cornerstone of online social media. From their birth on the Twitter platform, hashtags have evolved from their basic form of a short string of text preceded by a pound symbol to be a tool for myriad purposes, e.g. ad campaigns and online chats (Yang et al. 2012). In addition to being versatile, they are exceedingly popular and have been adopted on nearly every social media platform.

While hashtags may have been introduced to catalog information, they now can enable users to rally social movements (e.g. #BlackLivesMatter a social movement for racial justice), disseminate public health campaigns (e.g. #ECigTruths from Chicago Department of Public Health), and connect to communities (e.g. #RunChat a community for runners). Finding and connecting users to relevant communities online is of paramount importance for improving user experience, and hashtags can potentially enable such connections. Even though researchers have acknowledged this possibility (Laniado and Mika 2010; Russo and Nov 2010), little research thus far has targeted understanding the community use of hashtags.

In this study we focus on how engaging different types of hashtags are and consider relations with community metrics. We find that hashtags that have a stronger resemblance to real world communities are more engaging.

One of our contributions is a framework for understanding different dynamic types of hashtags. With this framework, we unify previous work on identifying periodic hashtags (Cook, Kenthapadi, and Mishra 2013) and event hashtags (Cha et al. 2010; Lehmann et al. 2012; Crane and Sornette 2008; Lin et al. 2013; Shamma, Kennedy, and Churchill 2011). This unification validates previous observations that there are coherent dynamic types of hashtags (Hsu, Chang, and Chen 2010; Romero, Meeder, and Kleinberg 2011).

Our second contribution is a set of analyses using this framework on a comprehensive cohort dataset. We include a comparison of engagement between hashtag types. Our analyses take steps in the direction of understanding engagement of hashtag types. This understanding is important, not just as a retrospective analysis, but as an actionable way for finding, connecting, and supporting communities.

One of our findings is that periodically recurring hashtags are the most engaging type of hashtag, on average. Previous work that has analyzed peaks in hashtag usage, i.e. events, is minimally actionable, as events are difficult to predict. In contrast, periodic events are predictable, so the ability to identify and understand periodically used hashtags has implications for how to design and implement new features for social media. Such new features could include weekly checkins on relevant content or community features built up around a periodic event. By showing that periodically used hashtags are the most engaging, our work implies that systems implemented around this periodic content would have the most impact. While our work focuses on hashtags, its broader implication is that systems designed to leverage periodic content to connect recurring communities will create a more engaging experience than connecting more ephemeral groups, e.g. groups that connect over events.

The structure of this paper is as follows. We begin by motivating the complexity of hashtags. We then propose a

*Work does as intern as Twitter, Inc.
†Work done as employee at Twitter, Inc.

framework for understanding coarse hashtag usage of all popular hashtags. Using this framework, we show there are 4 broad types of hashtags: events, stable, periodic, and stochastic. These classes have been alluded to in prior work (Hsu, Chang, and Chen 2010; Romero, Meeder, and Kleinberg 2011), but this is the first systematic categorization validating their existence. We then propose a metric of engagement for a hashtag and study this metric within the different hashtag types. We find hashtags with a recurring usage are at least 6.5% more engaging than other types of hashtags and over 100% more engaging than event hashtags, on average.

We build models that consider the relationship of engagement with Tweet volume, author diversity, and Tweet content for each dynamic type. These models show that popular hashtags are not the most engaging and that low diversity groups can be particularly engaging. We then propose a refining framework that partitions the recurring hashtags into 4 more interpretable classes: weekly events, all day weekly events, more frequent and less frequent than weekly events. These subclusters reveal an even stronger negative correlation of diversity with engagement for weekly events, which implies that concentrated groups of users that convene weekly and tweet a lot about a shared interest generate the most engaging hashtags. We conclude by examining engagement for a particular subset of community conversation-oriented "chat" hashtags.

Contributions

Our contributions complement prior work that has looked at temporal usage of hashtags. We extend such work by providing a comprehensive categorization of dynamics and analyzing the implications for engagement and communities.

We provide a novel comprehensive framework for clustering hashtags based on temporal usage. We then analyze the resulting hashtag dynamic types, propose a metric of engagement, and compare this metric of engagement between the hashtag types. This comparison is an important step for designing systems that is missing from prior work. We consider models that relate Tweet volume, author diversity, and Tweet content to hashtag engagement. We also propose and analyze another framework for understanding periodic type hashtags and add a discussion of implications for community engagement. Finally, we consider engagement of community-oriented "chat" hashtags within this clustering.

Related Work

There has been a diverse array of academic work using hashtags. Some of this work has leveraged hashtags to find relevant information on Twitter (Conover et al. 2011; Burnap et al. 2014; Giglietto et al. 2015; Letierce et al. 2010; Vieweg et al. 2010). Other work has used hashtags to predict popularity of Tweets (Suh et al. 2010; Naveed et al. 2011; Morchid et al. 2014; Hong, Dan, and Davison 2011; Petrovic, Osborne, and Lavrenko 2011), study information flow (Burnap et al. 2014; Romero, Meeder, and Kleinberg 2011; Starbird and Palen 2012; Weng, Menczer, and Ahn 2013; Naveed et al. 2011), and label data, e.g. with sentiment (Kouloumpis, Wilson, and Moore 2011; Davidov, Tsur, and Rappoport 2010) or political polarity (Conover et al. 2011). Studies have also looked at hashtags themselves, both trying to understand their dynamics (Cha et al. 2010; Lehmann et al. 2012; Crane and Sornette 2008; Lin et al. 2013; Shamma, Kennedy, and Churchill 2011; Yang and Leskovec 2011; Hsu, Chang, and Chen 2010), popularity (Pervin et al. 2015; Ma, Sun, and Cong 2013; Tsur and Rappoport 2012), relevance (Denton et al. 2015) and semantic meaning (Ferragina, Piccinno, and Santoro 2015; Posch et al. 2013; Romero, Meeder, and Kleinberg 2011; Tsur, Littman, and Rappoport 2012). This work highlights the breadth and utility of hashtags as an important feature of social media.

Here, we are most interested in engagement of hashtags by type. We consider types to be classes of temporal usage patterns rather than a priori defined semantic or topic classes (Ferragina, Piccinno, and Santoro 2015; Posch et al. 2013; Romero, Meeder, and Kleinberg 2011; Tsur, Littman, and Rappoport 2012).

Previous studies have looked at hashtag usage dynamics and noted that there are different types. Some studies have acknowledged at least three types of dynamics roughly falling into: continuous activity, periodic activity, or activity concentrated around an isolated time domain. (Hsu, Chang, and Chen 2010; Lehmann et al. 2012). However, these studies did not provide a comprehensive framework for showing these classes exist or showing how to identify and catalog hashtags into each class. They also did not consider differences in engagement between classes.

Other studies have focused on identifying and studying a single class of temporal dynamics. "Peaky" events, such as news have been studied and classified into as many as 6 types of classes (Cha et al. 2010; Lehmann et al. 2012; Crane and Sornette 2008; Lin et al. 2013; Shamma, Kennedy, and Churchill 2011; Yang and Leskovec 2011) . It has been suggested that these classes of events could relate to communities if we consider a group of people interested in an event as a community. Here we are interested in considering communities as groups of users who continue to share relevant and specific information, say on a topic. Furthermore, studying events is more retrospective and less actionable, as events are difficult to predict.

Periodic hashtags, on the other hand, can be predicted and thus could be actionable. Studies have briefly considered subsets from the class of periodic hashtags, for example periodic chats (recurring community conversations that happen during scheduled times on Twitter). Some work has focused on extracting (Cook, Kenthapadi, and Mishra 2013) and analyzing (Budak and Agrawal 2013) these chat hashtags. Another study investigated periodically occurring hashtags for select TV shows (Giglietto et al. 2015) and community engagement with these TV shows. However, these works on chats and TV shows did not compare how engaging community content is with other content.

Engagement with hashtags on Twitter has been broadly framed as a prediction task: predicting the adoption of a hashtag (Tsur and Rappoport 2015), the extent a hashtag will be used (Zhang, Wang, and Li 2014; Weng, Menczer, and Ahn 2013; Tsur and Rappoport 2012; Ma, Sun, and Cong

2013), or how many times a message will be Retweeted (Suh et al. 2010; Naveed et al. 2011; Morchid et al. 2014; Zaman et al. 2010; Hong, Dan, and Davison 2011; Petrovic, Osborne, and Lavrenko 2011). It has been found that using hashtags can be positively correlated with predicting Retweet rates (Suh et al. 2010; Burnap et al. 2014)

and engaging with a real world event (Hu, Farnham, and Talamadupula 2015). Other studies have noted that hashtags can explicitly be used by communities to connect on topics (Cook, Kenthapadi, and Mishra 2013; Budak and Agrawal 2013) and related to a user's community engagement on the Flickr platform (Russo and Nov 2010). However, none of this work has focused on categorizing hashtag types and looking at engagement within different types.

Here we take a distinct approach from previous work by finding a holistic way to globally catalog hashtags. We quantitatively show the existence of 4 clusters (and 4 periodic subclusters) and analyze how engagement is related to dynamic usage of hashtags.

Motivation: Examples of hashtag complexity

As a central feature to social media, hashtags play an important role in online social interactions. Understanding their role is important for understanding engagement. However, achieving this understanding is difficult as the usage of hashtags can be complex.

The semantic meaning of a hashtag can be ambiguous, e.g. #wow (referring to the World of Warcraft game or the exclamation "Wow!") (Romero, Meeder, and Kleinberg 2011). The semantic meaning can also be ambiguous due to temporal dynamics, e.g. the city of a sports team is often used to refer to the team during a match and not the geographic location. Classifications can also be inherently ambiguous as the result of complex definitions, e.g. #Superbowl refers to a specific sports event, so could be correctly classified as an event or as sports. The complexity of separating semantic classes is a well known problem, despite attempts to make exclusive classifications (Ferragina, Piccinno, and Santoro 2015; Posch et al. 2013; Romero, Meeder, and Kleinberg 2011; Tsur, Littman, and Rappoport 2013).

Because of these ambiguities, we adopted an unsupervised approach. Unsupervised learning avoids bias introduced when labeling training examples based upon interpretation of their meaning, e.g. the city name used to refer to a sports team rather than geographic location. Unsupervised learning also exposes a natural classification that we can examine, rather than imposing a hand selected set of classes and labeled training examples that do not cover all cases.

Categorizing Hashtag Usage Types

Examples of different temporal frequency profiles for hashtags are shown in Figure 1. It is clear that there are considerable differences in dynamic patterns. To catalog these different hashtag usage patterns, we describe each pattern by a short list of features and use unsupervised clustering to group similar usage patterns.

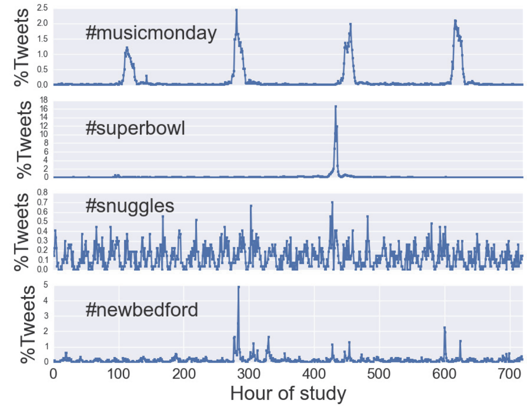

Figure 1: Representative hashtag profiles from each cluster. The percent of total Tweets with the hashtags over time (the normalized distributions of Tweets for each hashtag) show different dynamics between clusters. Each point is the percent of Tweets during the 30 day study period that was made during that hour.

Our approach is in contrast to previous approaches that classified hashtags by semantic content with a supervised method (Ferragina, Piccinno, and Santoro 2015; Posch et al. 2013; Romero, Meeder, and Kleinberg 2011; Tsur, Littman, and Rappoport 2012). Previous approaches that employed unsupervised methods targeted finding one type of dynamic class rather than a global categorization that we create (Cha et al. 2010; Lehmann et al. 2012; Crane and Sornette 2008; Lin et al. 2013; Shamma, Kennedy, and Churchill 2011; Cook, Kenthapadi, and Mishra 2013).

For a hashtag h, the set of Tweets that contained the hashtag during the study period is $T(h)$ and we consider the volume, or overall popularity, of the hashtag to be the logarithm of the number of Tweets with the hashtag during this period

$$V(h) = log(|T(h)|)$$

where $|\cdot|$ denotes the cardinality of a set. We consider $f(h,t)$ to be the normalized time series of a hashtag, i.e. the percent of Tweets that were made with the hashtag during a given time, an hour t, of the study period. Thus, for all hashtags

$$\sum_t f(h,t) = 1.$$

We consider the peak of the time series to be the hour t^h_* during which the hashtag was most popular, i.e. when it received the most Tweets

$$t^h_* = \underset{t}{\operatorname{argmax}} f(h,t).$$

The discrete Fourier transform of this normalized time series is given by

$$\hat{f}(h,\xi) = \sum_t f(h,t) e^{-2\pi i t \xi}$$

where $\hat{f}(h,\xi)$ is the Fourier coefficient of frequency ξ for the normalized time series of hashtag h. The magnitude of the Fourier coefficient gives an indication of how strong of a contribution a signal of period $1/\xi$ makes to the hashtag's usage pattern. Note that we do not consider the Fourier coefficient at frequency $\xi = 0$.

A few frequencies were of particular interest. Let us denote these frequencies as follows: ξ_d a daily signal with 24 hour period, ξ_w a weekly signal with 168 hour period, ξ_m the minimum measurable frequency for a 21 day period, and ξ_n^h the frequency that yielded the maximum magnitude Fourier coefficient that was not one of the previous calendar frequencies for hashtag h, i.e.

$$\xi_n^h = \underset{\xi \notin \{\xi_d, \xi_w, \xi_m\}}{\text{argmax}} |\hat{f}(h, \xi)|$$

where here $|\cdot|$ denotes the magnitude of a complex value. We consider ξ_n^h to be the *non-calendar frequency*

For stability, we computed the Fourier transform on 21 day windows of the time series that overlapped by 3 days and averaged the coefficients in a similar approach to Welch's method for computing a power spectrum. All the features are computed on either the hourly frequency time series $f(h, t)$ or on the discrete Fourier transform of the frequency time series $\hat{f}(h, \xi)$. A much larger set of features was constructed, but the subset of features described below gave the most significant and interpretable clusters.

Coarse features

For an initial coarse categorization of all the hashtags' dynamics, we used the following features:

- The peak volume, the maximum percent of Tweets that occurred in an hour of the study period: $f(h, t_*^h)$.
- The percent of Tweets that happened within 24 hours of the peak volume:
$$\sum_{t=t_*^h-12}^{t_*^h+12} f(h,t).$$
- The percent of volume from the 24 hours around the peak usage that occurred in 4 hours around the peak usage:
$$\left(\sum_{t=t_*^h-2}^{t_*^h+2} f(h,t)\right) \times \left(\sum_{t=t_*^h-12}^{t_*^h+12} f(h,t)\right)^{-1}.$$
- The percent of Tweets that occurred during a weekly window, i.e. the percent of total volume that occurred in 24 hour windows centered around the peak hour t_*^h and separated by 168 hours.
- The ratio of the volume that occurred in 24 hours centered around the peak hour t_*^h and the percent of the volume that occurred weekly.
- The maximum percent of the Tweets that occurred on a single day of the week:
$$\underset{w}{\text{argmax}} \sum_{t \in w} f(h,t), \text{ for } w \in \{\text{monday, tuesday, ...}\}.$$
- An indicator of whether every hour of the study period had a low percentage of the volume $\mathbb{I}[f(h,t) < \delta_0, \ \forall t]$.
- An indicator of whether every day of the week (summed across weeks) had a low percentage of the total volume $\mathbb{I}[\sum_{t \in w} f(h,t) < \delta_1, \ \forall w \in \{\text{monday, tuesday, ...}\}]$.
- The strength of the non-calendar frequency $|\hat{f}(h, \xi_n^h)|$ and the minimum frequency $|\hat{f}(h, \xi_m)|$.
- The difference in the strength of the minimum frequency with the daily frequency $|\hat{f}(h, \xi_m)| - |\hat{f}(h, \xi_d)|$, the weekly frequency $|\hat{f}(h, \xi_m)| - |\hat{f}(h, \xi_w)|$, and the non-calendar frequency $|\hat{f}(h, \xi_m)| - |\hat{f}(h, \xi_n^h)|$.

Refined recurring features

The coarse clustering separated hashtags into 4 dynamic classes, one of which is periodically recurring hashtags. As will be discussed, these hashtags were found to be the most engaging, so to study them further, we separated the periodically recurring hashtags into even more specific subclusters with the following features:

- The maximum percent of Tweets that occurred in an hour of the study period: $f(h, t_*^h)$.
- The daily support of the hashtag. To get this value we considered the percent of total Tweets that occurred during each hour of the day, summed over days. The support was defined as the number of hours out of 24 that, centered around the peak volume hour, accounted for above a certain fraction δ_2 of the total Tweets with the hashtag.
- The comparisons of strength of the minimum frequency with the daily frequency and the weekly frequency, as used in the coarse features.
- The percent of Tweets that occurred during a weekly window, i.e. the percent of total volume that occurred in 24 hour windows centered around t_*^h and separated by 168 hours.

Selecting the number of clusters

After constructing these features on the Tweets for each hashtag, we use the K-means clustering algorithm and the silhouette score for choosing an appropriate number of clusters. The silhouette score tries to quantify how well clustered the data is by comparing how close each datum is to its cluster versus the nearest neighboring cluster (Rousseeuw 1987). For a set of data points $\{x_i\}$, the silhouette metric is

$$\underset{i}{\text{mean}} \left(\frac{d(x_i, c_i^n) - d(x_i, c_i)}{\max\{d(x_i, c_i), d(x_i, c_i^n)\}} \right)$$

where c_i is the center of the cluster that the ith datum x_i is assigned to, and c_i^n is the center of the closest neighboring cluster. Here $d(\cdot, \cdot)$ is the distance between two points, which we take to be euclidean distance. A larger silhouette score indicates a better clustering, so we choose the number of clusters that maximizes it.

Community metrics

In addition to identifying different hashtag usage patterns, we want to understand the implications of such a clustering, particularly for communities. To look at community dynamics, we consider the two following measures: engagement $E(h)$ and diversity $D(h)$. Engagement quantifies how much users interact with a hashtag. Diversity quantifies how large of a user base a hashtag has relative to its popularity.

Engagement

To quantify how engaging a hashtag h is, we consider a Tweet has "received an engagement" if it has been either Retweeted or Favoured, as these two actions indicate a human interaction. We then consider the engagement score $E(h)$ to be the proportion of Tweets with a hashtag that have received an engagement, or the probability of engagement for that hashtag. Formally this is

$$E(h) = \frac{\sum\limits_{\tau \in T(h)} \mathbb{I}_{[\tau \text{ has Retweet or Favouring}]}}{|T(h)|}$$

where τ is a set in all the Tweets $T(h)$ containing hashtag h, and $\mathbb{I}[\cdot]$ is the indicator function that evaluates to 1 when the argument is true and 0 when false. We use the binary measure of receiving a Retweet or Favouring, instead of looking at the mean number of engagements a Tweet received, as it is robust to the phenomena of a hyper popular Tweet receiving thousands or millions of engagements.

Diversity

To quantify how broadly a hashtag is adopted, we look at the diversity of the user base. A pure measure of adoption, such as the number of users who used a hashtag, can be more representative of how many times a hashtag is Tweeted, rather than a measure of how concentrated the user base is. To capture how broad or diverse the user group is, we normalize by the number of Tweets that used a hashtag. We consider diversity $D(h)$ to be the ratio of the number of users in the set of users who Tweeted the hashtag $U(h)$ over the number of Tweets in the set of Tweets that used the hashtag $T(h)$:

$$D(h) = \frac{|U(h)|}{|T(h)|}.$$

This measure is the reciprocal of the average number of times a user Tweets with the hashtag. High diversity translates to many people using the hashtag very few times. Low diversity corresponds to only a few users Tweeting multiple times with the hashtag. Abnormally low diversity is indicative of a spammer or bot driving the hashtag usage.

Dataset

To evaluate our framework, we consider the corpus of Tweets from all English language Twitter users in the United States that used a hashtag at least once during the 30 day study period starting January 15, 2015. Studying a restricted geographic region diminishes the need to control for time zones and reduces the ambiguity between hashtag uses.

In total, the dataset consists of 19,197,367 users who made 2,529,886,239 Tweets, of which 437,167,710 (roughly 17.3%) contained a hashtag. There were 801,850,909 occurrences of hashtags. Of the hashtags used, 18,149,314 were unique, and they exhibited a long tail of usage frequency. The most popular hashtag (#nowplaying) appeared 3,602,346 times in this period. Because we are interested in usage dynamics, we only look at the top 34,500 most popular hashtags, those that were used in at least 2,000 Tweets during the study period. Tweets were converted to lowercase and hashtags were extracted from Tweets with a regular expression, so capitalization was not considered unique, as it has been previously (Tsur and Rappoport 2015).

Users	19,197,367
Tweets	2,529,886,239
Tweets with #	437,167,710
Hashtag occurrences	801,850,909
Unique hashtags	18,149,314
Popular hashtags	34,500

Table 1: Statistics on all English language users in the US who Tweeted with a pound symbol during the study period.

Removing spammers and bots

An unfortunate problem on social media platforms is the existence of spammers and bots. These are entities that send out an abnormally high number of Tweets with little value and often no receptive audience. Spammers and bots can be difficult to identify because they are designed to mimic real users and avoid efforts to remove them. We are not interested in hashtags predominantly Tweeted by spammers or bots because they do not represent content that should be surfaced. Thus, we remove hashtags that are blatantly generated by spammers or bots.

We removed hashtags that exhibited at least one of the following abnormal behaviors: minimal adoption or zero engagement. We define minimal adoption as hashtags that have extremely low diversity ($D(h) \leq .02$), which was an indicator of a single account, or a small set of accounts, sending out all the Tweets with a given hashtag. Zero engagement was defined as the engagement score for the hashtag being equal to zero ($E(h) = 0$), i.e. out of over 2,000 Tweets not a single one received a Retweet or a Favouring.

Minimal adoption, or unnaturally low diversity, accounted for only 1,581 hashtags being removed and zero engagement accounted for 1,745 hashtags being removed, both small subsets of the total hashtags. Lists of hashtags removed were also manually checked. Low diversity hashtags mostly represented advertisers of pornography, while zero engagement was mostly represented by Islamic propaganda. Ideally, we would have removed spammy users and bots, but identifying such users is ongoing research.

Results

Using the above described framework for describing hashtag usage, we used K-means clustering to find clusters of

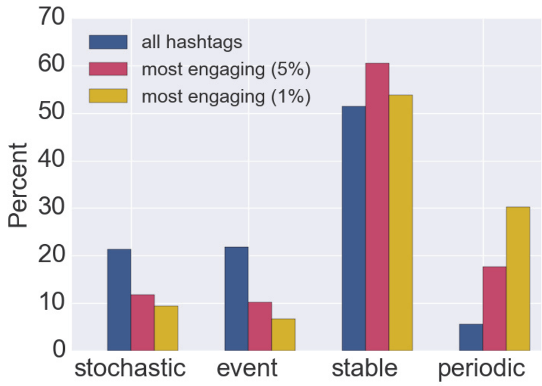

Figure 2: Distribution of hashtags by type. Periodically recurring hashtags account for the fewest number of hashtags. However, they account for a large portion of the top 5% and even larger portion of the top 1% most engaging hashtags.

dynamic patterns in the dataset. The number of clusters was chosen to maximize the silhouette distance.

Dynamic types

A coarse clustering of the entire dataset was made by using the coarse features listed above. This stage resulted in 4 global clusters. The clusters were validated by extensive manual inspection of a randomly selected subset of hashtags. This inspection revealed that the clusters could be interpreted as: stable (daily chatter), periodically recurring, single event, and irregular stochastic patterns. Example time series from each cluster are shown in Figure 1.

The percent of hashtags of each type can be seen in Figure 2. The most represented type of hashtag is stable, or hashtags that were used in a relatively constant number of Tweets each day in the study period. These stable hashtags are representatively of "chatter" on twitter and make up over 50% of the hashtags. The periodically occurring hashtags account for less than 10% of the hashtags. However, it is periodically recurring hashtags that will be shown to be most engaging.

Refined periodically recurring subtypes

To further investigate the most engaging class of hashtags, periodically recurring hashtags, we used the set of refining features discussed above and K-means to refine the cluster of recurring hashtags into subclusters. The silhouette distance indicated that 4 subtypes of periodically recurring clusters exist. Upon extensive manual inspection of randomly selected subsets of hashtags, we validated the clustering and found that the subclusters can be described as all-day events (e.g. #monday), concentrated weekly events (e.g. TV shows and weekly chats), periodic events with either strong imbalance between the events, or less than weekly (e.g. fortnightly), and events that are either more frequent than weekly or have significant support on other days of the week (e.g. daily chats and popular TV shows with lots of mid-week usage).

Discussion

In the above sections, we described a framework for categorizing popular hashtags in a meaningful way. Now we examine the importance of these categories and what they imply for communities on social media.

Engagement varies between dynamic classes

We consider the engagement of a hashtag $E(h)$ to be the probability that a Tweet with that hashtag received a Retweet or a Favouring, a human engagement. Higher engagement means that a larger portion of Tweets with that hashtag received some human engagement and implies that the hashtag more successfully engages users on a broad scale.

Figure 2 shows the distribution of hashtag types in the top 5% and 1% of most engaging hashtags. We see that, while periodically recurring hashtags are the least populous, they comprise a significant portion of the top 5% and top 1% most engaging hashtags. Periodic hashtags are less than 10% of the total popular hashtags in the dataset, but they account for roughly 30% of the top 1% most engaging hashtags.

In addition to looking at the representation of each dynamic class in the most engaging hashtags, we statistically compare the distributions of engagement for hashtags in each type. A table comparing the distribution of engagement scores for hashtags in each cluster is in Table 2.

Table 2 reveals that the clusters have statistically significantly different distributions of engagement. This result is important; it means that the dynamics of how a hashtag is being used is related to how engaging the hashtag is. The cluster of periodically recurring hashtags is most engaging, and the cluster of event hashtags is least engaging. Periodic hashtags are on average at least 6.5% more engaging than other types of hashtags and over 100% more engaging on average than event hashtags. This result is influential for designing systems. Periodic content is more easily predicted than events, so it implies that periodic content could be leveraged to connect users with more engaging content and perhaps other users who share interest in a recurring hashtag.

Volume does not increase engagement and lower diversity can be more engaging

In addition to noting that the dynamic classes of hashtags have different average levels of engagement, we consider the correlation of volume $V(h)$ and diversity $D(h)$ with engagement for each type of hashtag. We quantify the relationship of volume and diversity with a linear model within each cluster that predicts the engagement of a hashtag. In addition to diversity and volume, we also consider the content of Tweets with each hashtag. The content we consider is

- the percent of Tweets with a URL
- the percent of Tweets that have a mention, i.e. are directed at another user by using an @ symbol
- the average number of hashtags that a Tweet with that hashtag has.

Model parameters are given in Table 3. These models show that, while little of the variance is explained by these models, there is not a positive linear correlation between

	event	stochastic	stable	periodic
event	- (-) -%	- (-) -%	- (-) -%	- (-) -%
stochastic	0.039 (0.000) 140.733%	- (-) -%	- (-) -%	- (-) -%
stable	0.089 (0.000) 193.668%	0.050 (0.000) 137.614%	- (-) -%	- (-) -%
periodic	0.101 **(0.000) 206.265%**	0.063 **(0.000) 146.565%**	0.012 **(0.001) 106.504%**	- (-) -%

Table 2: Results of statistical tests comparing distributions of hashtag engagement within clusters. The table format is "distance between cluster means (p-value) percent increase %", where the distance is the mean of the row cluster minus the mean of the column cluster. Tests show that hashtags in periodic cluster are statistically more engaging and hashtags in events cluster are less engaging than all other clusters.

how popular a hashtag is, or the number of Tweets with the hashtag, and how engaging the hashtag is. There is a lack of positive correlation of engagement with popularity for all dynamic hashtag types. No positive correlation means that popularity is not the same as engagement. This result is not intuitive and implies that, when designing systems to leverage hashtags to find and surface engaging content, sophistication is need. Simply looking at the most frequently Tweeted hashtag is insufficient.

These models also give insight into the relationship of engagement with diversity of the hashtag's user base. Most striking are a strong negative relation of diversity with engagement for periodically recurring hashtags and a strong positive relation of diversity with engagement for the stable, or consistently used, hashtags. These relations imply that concentrated groups of user who repeatedly Tweet a hashtag are more engaging for periodic content, while a broad user base is more engaging for consistently Tweeted hashtags.

The strongest relationship of content with engagement is a significant negative correlation of mentions with engagement for the periodic hashtags. This negative relationship implies that hashtags where the users are Tweeting at each other and not directing their Tweets towards a broader audience are less engaging.

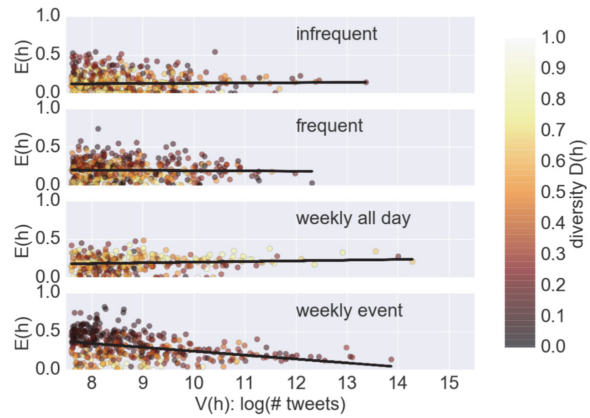

Figure 3: Distribution of periodic type hashtags in subclusters. Color is the diversity $D(h)$ of each hashtag. Low diversity implies a concentrated user group. Hashtags for weekly events with low diversity are the most engaging.

Subclustering of periodically recurring hashtags shows engaged concentrated user groups

To better understand the most engaging periodically recurring subcluster, we use the above mentioned framework to refine the periodically recurring cluster into 4 subclusters. The distribution of subclusters is shown in Figure 3. The subclusters can roughly be described as weekly events (e.g. #dadchat), weekly all day events (e.g. #throwbackthursday), infrequent events (i.e. event happens less than weekly or skipped a week), and frequent events (i.e. more than weekly, such as daily news hours).

Figure 3 shows that for weekly concentrated events (e.g. TV shows and community chats), lower diversity hashtags have a higher engagement. We quantify the correlation of engagement with volume, diversity, and content measures with a linear model in each subcluster. Model parameters are given in Table 4. We consider the same measures of Tweet content for the subcluster models as we did for the cluster models: percent Tweets with a link, percent Tweets with a mention, average number of hashtags in Tweets.

These models show that, while little of the variation is explained by Tweet content, there is a particularly strong negative correlation with diversity for the weekly events subcluster. Lower diversity with higher engagement hints at focused community structure. Concentrated communities that convene weekly around a shared interest and Tweet repeatedly about it are the most engaging. This result shows that to find engaging weekly event hashtags, looking at the size of the user base or the number of Tweets is insufficient.

Community-oriented "chats" are more engaging

We have seen that temporal usage patterns relate to engagement, with recurring hashtags being the most engaging. We have also seen that volume is not positively correlated with engagement. This implies that recurring hashtags with moderate user bases, community tags, are the most engaging. We can further explore this implication by looking at explicitly community conversation-oriented "chat" hashtags.

Prior work has targeted hashtags for community chats and recognized the importance of such hashtags (Cook, Kenthapadi, and Mishra 2013; Budak and Agrawal 2013). This prior work defined chats as periodically scheduled events that happen in coordination on Twitter with the intent of communication. This work also highlighted that exhaustively finding such hashtags can be difficult. However, we can easily access a subset of the chat hashtags by searching for the word "chat" in the hashtag.

	coeff.	p-value
Cluster: Stochastic (Adj. R^2 = 0.197)		
Intercept	0.4650	*
$V(h)$	-0.0223	*
$D(h)$	0.0155	*
% links	-0.0726	*
% mentions	-0.0787	*
# hashtags	-0.0120	*
Cluster: Event (Adj. R^2 = 0.145)		
Intercept	0.3214	*
$V(h)$	-0.0113	*
$D(h)$	-0.0541	*
% links	-0.0704	*
% mentions	-0.0408	*
# hashtags	-0.0138	*
Cluster: Stable (Adj. R^2 = 0.233)		
Intercept	0.231	*
$V(h)$	-0.0070	*
$D(h)$	0.1430	*
% links	-0.0585	*
% mentions	-0.0080	.007
# hashtags	-0.0017	.001
Cluster: Periodic (Adj. R^2 = 0.416)		
Intercept	0.7356	*
$V(h)$	-0.0216	*
$D(h)$	-0.2149	*
% links	-0.0726	*
% mentions	-0.2324	*
# hashtags	-0.0246	*

Table 3: Linear model parameters for relation of engagement $E(h)$ with the volume, diversity, and Tweet content features within clusters. (Note * means that the value was $< .001$.)

	coeff.	p-value
Subcluster: Infrequent (Adj. R^2 = 0.446)		
Intercept	0.6580	*
$V(h)$	-0.0187	*
$D(h)$	-0.1637	*
% links	-0.0731	*
% mentions	-0.2725	*
# hashtags	-0.0417	*
Subcluster: Frequent (Adj. R^2 = 0.281)		
Intercept	0.6059	*
$V(h)$	-0.0177	0.002
$D(h)$	-0.1461	*
% links	-0.0986	*
% mentions	-0.1726	*
# hashtags	-0.0462	0.001
Subcluster: Weekly all day (Adj. R^2 = 0.182)		
Intercept	0.3218	*
$V(h)$	0.0057	0.243
$D(h)$	0.0025	0.921
% links	-0.0101	0.672
% mentions	-0.1740	*
# hashtags	-0.0639	*
Subcluster: Weekly event (Adj. R^2 = 0.532)		
Intercept	0.9997	*
$V(h)$	-0.0477	*
$D(h)$	-0.4286	*
% links	-0.0965	0.001
% mentions	-0.1553	*
# hashtags	-0.0596	0.006

Table 4: Linear model parameters for relation of engagement $E(h)$ with the volume, diversity, and Tweet content features within subclusters of periodically recurring cluster. (Note * means that the value was $< .001$.)

Searching for "chat" hashtags returns a set of 197 hashtags, some of which are not periodic. They are distributed between clusters, with most in the periodically recurring cluster. This distribution is expected because some chats target a broader definition of community where chat happens on a topic anytime (e.g. #phdchat and #edchat, hashtags for ongoing conversations about things related to doctoral students and education, respectively) and other chats happen at pre-scheduled weekly times (e.g. #dadchat, a weekly online chat for fathers).

Figure 4 shows the chat hashtags in red overlaid on their dynamic class. Nearly all of the community-oriented chat hashtags have a higher engagement than expected for their volume, even though they do not have a relatively large volume. Statistically comparing the average engagement for the chat hashtags with the average engagement for other hashtags in each dynamic type, some of which could be unidentified chats, confirms that engagement is higher for chats. Table 5 shows the statistical comparison for each dynamic class of chat hashtags with other hashtags in the dynamic class. Periodically recurring chats are on average 130% more engaging than other periodically occurring hashtags (e.g. TV shows). Chats that happen as events, e.g. an esteemed person hosting a conversation on Twitter, are also significantly more engaging than other event type hashtags. These observations support the broader observation that community-oriented hashtags are more engaging. They also indicate that different types of periodically occurring hashtags exist.

Conclusions and Future Work

We have presented an exploration of how hashtag usage relates to engagement. To do this, we built a framework for categorizing hashtags by usage type. We studied this categorization on a cohort of Twitter users and found that there are 4 broad hashtag types: stochastic, stable, event, and periodically recurring usage. Recurring hashtags that people come back time and again to follow are the most engaging on average. Events, while popular by volume, are not as broadly engaging as weekly interest-oriented content on average.

We used linear models to quantify the relationship of engagement with volume, diversity, and Tweet content mea-

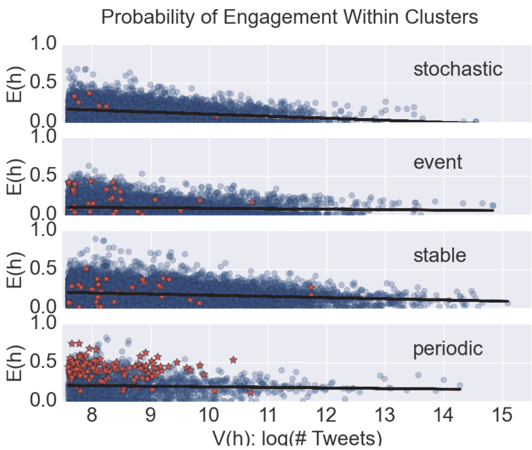

Figure 4: Models predicting engagement of a hashtag from its popularity show no positive correlation between engagement and popularity. Hashtags with "chat" in them are displayed as red stars.

Cluster	chat μ	non-chat μ	percent	p-value
event	0.213	0.095	224.1%	$<$ .001
stochastic	0.240	0.134	178.6%	0.062
stable	0.223	0.185	120.8%	0.100
periodic	0.414	0.179	231.6%	$<$.001

Table 5: Comparison of mean (denoted by μ) engagement of "chat" hashtags with other hashtags in each dynamic type. Community-oriented chat hashtags are statistically more engaging than other periodic and event hashtags.

sures. These models revealed that volume is not positively correlated with broad engagement - popular hashtags are not necessarily the most engaging. These models also revealed a negative relationship of diversity with engagement for periodically recurring hashtags. Upon further investigation of the most popular periodic hashtags, we found there to be 4 subclusters and strong negative correlations of engagement with diversity, particularly for weekly recurring events. This implies concentrated engaged user groups, communities, forming around hashtags. We further explored this by looking at community conversation-oriented "chat" hashtags and found these hashtags to be more engaging than other hashtags, regardless of dynamic type.

We found community effects to be alive and well, and more importantly, engaging on Twitter. Our results emphasize complex relations between engagement and hashtags. These results also imply that content relating to concentrated parties, communities, is the most engaging, but more work is needed to fully identify such content than just looking at what is popular by volume.

Future work will look at the connectedness of the social graph that forms around users of each hashtag. Because Tweets are public and can be viewed without users being connected, community discussions could easily be participated in without following or being followed by any other participants. We thus suspect that hashtag community follow graphs could be sparse.

We would also like to explore the dynamic nature of hashtag type, e.g. during some periods a hashtag could have a constant temporal pattern, while at other times it could have an event pattern. We would like to study if a hashtag's engagement changes with its temporal usage type.

Finally, we would like to consider additional measures of engagement and diversity. Our measure of engagement controlled for bias from individual Tweets being extremely engaging or high volume hashtags receiving more total engagements as a function of exposure. However, by not accounting for the absolute number of engaged Tweets with a hashtag, our models celebrate hashtags that have more consistent or broad appeal, which could be easier to acheive by low voume hashtags.

Acknowledgments

We would like to thank Twitter, Inc. for financial support and Gillian Chin, Albert Magyar, Dimitris Papailiopoulos, Benjamin Recht, and Aneesh Sharma for valuable feedback.

References

Budak, C., and Agrawal, R. 2013. On participation in group chats on twitter. In *Proc. WWW*, 165–176. ACM.

Burnap, P.; Williams, M. L.; Sloan, L.; Rana, O.; Housley, W.; Edwards, A.; Knight, V.; Procter, R.; and Voss, A. 2014. Tweeting the terror: modelling the social media reaction to the woolwich terrorist attack. *Social Network Analysis and Mining* 4(1):1–14.

Cha, M.; Haddadi, H.; Benevenuto, F.; and Gummadi, P. K. 2010. Measuring user influence in twitter: The million follower fallacy. In *Proc. ICWSM*. AAAI.

Conover, M.; Ratkiewicz, J.; Francisco, M.; Gonçalves, B.; Menczer, F.; and Flammini, A. 2011. Political polarization on twitter. In *Proc. ICWSM*. AAAI.

Cook, J.; Kenthapadi, K.; and Mishra, N. 2013. Group chats on twitter. In *Proc. WWW*, 225–236. ACM.

Crane, R., and Sornette, D. 2008. Robust dynamic classes revealed by measuring the response function of a social system. *Proc. the National Academy of Sciences* 105(41):15649–15653.

Davidov, D.; Tsur, O.; and Rappoport, A. 2010. Enhanced sentiment learning using twitter hashtags and smileys. In *Proc. of the 23rd International Conference on Computational Linguistics: Posters*, 241–249. ACL.

Denton, E.; Weston, J.; Paluri, M.; Bourdev, L.; and Fergus, R. 2015. User conditional hashtag prediction for images. In *Proc. SIGKDD*, 1731–1740. ACM.

Ferragina, P.; Piccinno, F.; and Santoro, R. 2015. On analyzing hashtags in twitter. In *Proc. ICWSM*. AAAI.

Giglietto, F.; Boccia Artieri, G.; Gemini, L.; and Orefice, M. 2015. Understanding engagement and willingness to speak up in social-television: A full-season, cross-genre analysis of tv audience participation on twitter. *Cross-Genre Analysis of TV Audience Participation on Twitter (May 27, 2015)*.

Hong, L.; Dan, O.; and Davison, B. D. 2011. Predicting popular messages in twitter. In *Proc. WWW*, 57–58. ACM.

Hsu, M.-H.; Chang, Y.-H.; and Chen, H.-H. 2010. Temporal correlation between social tags and emerging long-term trend detection. In *Proc. ICWSM*. AAAI.

Hu, Y.; Farnham, S.; and Talamadupula, K. 2015. Predicting user engagement on twitter with real-world events. In *Proc. ICWSM*. AAAI.

Kouloumpis, E.; Wilson, T.; and Moore, J. 2011. Twitter sentiment analysis: The good the bad and the omg! *Proc. ICWSM* 11:538–541.

Laniado, D., and Mika, P. 2010. Making sense of twitter. In *Proc. ISWC*. Springer. 470–485.

Lehmann, J.; Gonçalves, B.; Ramasco, J. J.; and Cattuto, C. 2012. Dynamical classes of collective attention in twitter. In *Proc. WWW*, 251–260. ACM.

Letierce, J.; Passant, A.; Breslin, J. G.; and Decker, S. 2010. Using twitter during an academic conference: The #iswc2009 use-case. In *Proc. ICWSM*. AAAI.

Lin, Y.-R.; Margolin, D.; Keegan, B.; Baronchelli, A.; and Lazer, D. 2013. # bigbirds never die: Understanding social dynamics of emergent hashtags. In *Proc. ICWSM*. AAAI.

Ma, Z.; Sun, A.; and Cong, G. 2013. On predicting the popularity of newly emerging hashtags in twitter. *Journal of the American Society for Information Science and Technology* 64(7):1399–1410.

Morchid, M.; Dufour, R.; Bousquet, P.-M.; Linarès, G.; and Torres-Moreno, J.-M. 2014. Feature selection using principal component analysis for massive retweet detection. *Pattern Recognition Letters* 49:33–39.

Naveed, N.; Gottron, T.; Kunegis, J.; and Alhadi, A. C. 2011. Bad news travel fast: A content-based analysis of interestingness on twitter. In *Proc. WebSci*, 8. ACM.

Pervin, N.; Phan, T. Q.; Datta, A.; Takeda, H.; and Toriumi, F. 2015. Hashtag popularity on twitter: Analyzing co-occurrence of multiple hashtags. In *Social Computing and Social Media*. Springer. 169–182.

Petrovic, S.; Osborne, M.; and Lavrenko, V. 2011. Rt to win! predicting message propagation in twitter. In *Proc. ICWSM*. AAAI.

Posch, L.; Wagner, C.; Singer, P.; and Strohmaier, M. 2013. Meaning as collective use: predicting semantic hashtag categories on twitter. In *Proc. WWW*, 621–628. ACM.

Romero, D. M.; Meeder, B.; and Kleinberg, J. 2011. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proc. WWW*, 695–704. ACM.

Rousseeuw, P. J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20:53–65.

Russo, P., and Nov, O. 2010. Photo tagging over time: A longitudinal study of the role of attention, network density, and motivations. In *Proc. ICWSM*. AAAI.

Shamma, D. A.; Kennedy, L.; and Churchill, E. F. 2011. Peaks and persistence: modeling the shape of microblog conversations. In *Proc. ACM conference on Computer supported cooperative work*, 355–358. ACM.

Starbird, K., and Palen, L. 2012. (how) will the revolution be retweeted?: information diffusion and the 2011 egyptian uprising. In *Proc. ACM conference on computer supported cooperative work*, 7–16. ACM.

Suh, B.; Hong, L.; Pirolli, P.; and Chi, E. H. 2010. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *Social computing (socialcom), 2010 IEEE second international conference on*, 177–184. IEEE.

Tsur, O., and Rappoport, A. 2012. What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities. In *Proc. WSDM*, 643–652. ACM.

Tsur, O., and Rappoport, A. 2015. Dont let me be #misunderstood: Linguistically motivated algorithm for predicting the popularity of textual memes. In *Proc. ICWSM*. AAAI.

Tsur, O.; Littman, A.; and Rappoport, A. 2012. Scalable multi stage clustering of tagged micro-messages. In *Proc. WWW*, 621–622. ACM.

Tsur, O.; Littman, A.; and Rappoport, A. 2013. Efficient clustering of short messages into general domains. In *ICWSM*. AAAI.

Vieweg, S.; Hughes, A. L.; Starbird, K.; and Palen, L. 2010. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Proc. of the SIGCHI conference on human factors in computing systems*, 1079–1088. ACM.

Weng, L.; Menczer, F.; and Ahn, Y.-Y. 2013. Virality prediction and community structure in social networks. *Scientific reports* 3.

Yang, J., and Leskovec, J. 2011. Patterns of temporal variation in online media. In *Proc. WSDM*, 177–186. ACM.

Yang, L.; Sun, T.; Zhang, M.; and Mei, Q. 2012. We know what@ you# tag: does the dual role affect hashtag adoption? In *Proc. WWW*, 261–270. ACM.

Zaman, T. R.; Herbrich, R.; Van Gael, J.; and Stern, D. 2010. Predicting information spreading in twitter. In *Workshop on computational social science and the wisdom of crowds, NIPS*, volume 104, 17599–17601. Citeseer.

Zhang, P.; Wang, X.; and Li, B. 2014. Evaluating important factors and effective models for twitter trend prediction. In *Online Social Media Analysis and Visualization*. Springer. 81–98.