We Know What @You #Tag: Does the Dual Role Affect Hashtag Adoption?

Lei Yang[1] Tao Sun[2] * Ming Zhang[2] Qiaozhu Mei[1]
[1] School of Information, the University of Michigan
{yangle, qmei}@umich.edu
[2] School of EECS, Peking University
{suntao, mzhang}@net.pku.edu.cn

ABSTRACT

Researchers and social observers have both believed that hashtags, as a new type of organizational objects of information, play a dual role in online microblogging communities (e.g., Twitter). On one hand, a hashtag serves as a bookmark of content, which links tweets with similar topics; on the other hand, a hashtag serves as the symbol of a community membership, which bridges a virtual community of users. Are the real users aware of this dual role of hashtags? Is the dual role affecting their behavior of adopting a hashtag? Is hashtag adoption predictable? We take the initiative to investigate and quantify the effects of the dual role on hashtag adoption. We propose comprehensive measures to quantify the major factors of how a user selects content tags as well as joins communities. Experiments using large scale Twitter datasets prove the effectiveness of the dual role, where both the content measures and the community measures significantly correlate to hashtag adoption on Twitter. With these measures as features, a machine learning model can effectively predict the future adoption of hashtags that a user has never used before.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

Keywords

Twitter, Hashtag, Dual Role, Prediction

1. INTRODUCTION

Recently, microblogging sites such as Twitter have rapidly grown from yet another social networking service into one of the most prevalent type of social media. According to a recent statistic [36], Twitter has accumulated over 200 million users as well as hundreds of billions of short messages (i.e., tweets). Due to the convenience of posting short messages, thousands of tweets are posted every second and from every corner of the world, assembling the "SMS of the internet" [9] and the collection of "global conversations" [13].

The overwhelming volume of tweets and the tremendously sparse information in each individual tweet have made it extremely difficult for a user to find and track interesting topics. *Hashtag*, as a brand-new organizational object of information, has emerged from this context and soon become widely adopted by Twitter users to facilitate their navigation in this deluge of information. Figure 1 presents two examples of hashtags and the results of searching the two hashtags in Twitter. As simple as a user-composed keyword starting with the # symbol, a hashtag effectively coordinates relevant tweets and people in the microblogging media.

Figure 1: Snapshots of search results of hashtags

The use of hashtags has brought convenience to Twitter users in various ways. As a user-defined index term of content, a hashtag links relevant topics and events together [8], making it much easier to assess the semantics of a tweet. As a result, the exposure of tweets containing certain hashtags are maximized in information retrieval and navigation. For example, tweets related to iPhone are easily retrieved by a single click on the hashtag *#iphone* (see the left panel of Figure 1 (a)). To this end, a hashtag plays the role of a social bookmark: annotating the content, being shared to other users, and assembling the folksonomy.

Is this the only role of a hashtag? Sociologists, media observers, and computer scientists have all noticed another role. Beyond a bookmark of content, a hashtag serves as the symbol of a community [34, 11, 12, 28]. Indeed, a hashtag enables users to identify and participate in online chats designated by the tag [34]. Birds of a feather can be easily found and connected by tracking a particular hashtag, as shown in the right panel of Figure 1. To this end, a hashtag defines a virtual community of users with the same background (e.g., "#Umich"), the same interest (e.g., "#iPhone"), or involved in the same conversation or task (e.g., "#www2012", "#VoteForObama"). A user joins such a community by simply including that hashtag in her own tweets.

*This work was done when Tao Sun was visiting University of Michigan sponsored by China Scholarship Council.

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.
WWW'12,, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1229-5/12/04.

The observations above have clearly indicated a critical hypothesis that when posting a hashtag, the Twitter users are aware of the dual role it plays – as a unique indicator of both the topic of the *content* and the membership of a *community* [25, 15]. In other words, by creating a hashtag, a user either invents and shares a new bookmark (of content), or initializes and spreads a coat of arm (of a community), or both. By adopting an existing hashtag, a user either presents her interest in a topic, or presents her intent to obtain a community membership, or both. Should this hypothesis hold, one would expect the analysis of hashtags contribute significantly to the understanding of user interests and behaviors in microblogging sites.

Interestingly, even before being formally tested, this hypothesis has already been directly or indirectly leveraged in various studies of microblogging. Indeed, in existing literature, hashtags have been effectively utilized as critical features for various tasks of text or social network analysis, including topic detection and tracking [3], text classification [30], community identification [24], and link prediction [35], etc. On one hand, hashtags do perform effectively in most of these tasks as a specific type of features. On the other hand, because of the lack of a formal conclusion on the role of hashtags, it is difficult to explain why the involvement of hashtags works in some tasks but not in others. There's little insight on how to make the best use of hashtags.

In this work, we take the initiative to conduct a systematic empirical analysis of how the dual role of a hashtag affects hashtag adoption. Our goal is to formally test whether the content role and the community role affect users' behavior of adopting hashtags that they have never used before. To achieve this, we focus on how hashtags are adopted and used by Twitter users, and investigate whether this behavior can be explained as the behavior of bookmarking the content and/or the behavior of joining a community. Should the hypothesis hold, the following observation would be expected: 1) the major factors that affect the behavior of content tagging and joining communities will also significantly affect the adoption of hashtags; and 2) by carefully measuring the magnitude of these factors, it is feasible to make predictions of future adoption of hashtags.

The contribution of our work is not to investigate how to utilize hashtags to enhance particular data mining tasks. Instead, our study provides a foundation of the rationality of using hashtags in these tasks: why it works; and in what context it might work better. Accordingly, our goal is not to differentiate the role of an individual hashtag, but to provide a macroscopical analysis of the two roles. The predictive analysis in our work also serves as a feasibility study of hashtag recommendation, which provides insights to the future construction of recommender systems of hashtags.

The rest of the paper is organized as follows. We start with a discussion of related work in Section 2. In Section 3 we introduce comprehensive measures to quantify the major factors of content tagging and joining communities, followed by a demonstrative analysis of how these measures correlate with the hashtag usage (Section 4). A formal regression analysis is presented in Section 5 to test the predictive power of these measures to the adoption of hashtags. In Section 6, we develop machine learning methods to predict the future adoption of hashtags of individual users, which serves as a feasibility study of hashtag recommendation. We conclude our work in Section 7.

2. RELATED WORK

To the best of our knowledge, this is the first formal analysis to test the predictive power of the dual role of hashtags on hashtag adoption. Between the two roles, the role of content tagging has connected hashtags to existing social tagging systems such as del.icio.us [1] and Flickr [2]. Those tags are also generated by the users to annotate content such as Web pages or photos and are being shared to other users.

The second role as the symbol of a community membership, on the other hand, has differentiated hashtags from traditional social tags. Playing this role, hashtags are employed as a mechanism to participate in online conversations [19] and join virtual communities [34, 12]. This role has been carefully examined in individual events such as scientific conferences [22, 11]. Letierce et al. [22] concluded that annotating messages by hashtags reveals a strong desire from the users to join in the discussion related to the conference. Ebner et al. [11] argued that conference communities in Twitter are built based on the adoption of one specific hashtag. Therefore, a hashtag is not only a tag of content but also the tag of a community. The *community* role of a hashtag presents in its functionalities to *identify* a community, *form* a community, and allow users to *join* a community - a role never played by traditional social tags.

In general, social tags (such as those in del.icio.us, CiteULike, Flickr, Last.fm and YouTube) serve for two purposes [25, 15]: *i) organizational*, to facilitate description or navigation, and *ii) social*, to facilitate resource exposure and sharing. This is quite different from the user's intent of joining a community, which presents the community role of a hashtag. Because of the conversation nature of Twitter, the social interactions in Twitter communities are far more frequent than those in other social media. In other words, a Twitter community started by hashtags has much richer inner-community communications, compared to a flat set of people who happen to use the same bookmark.

Our measurement of the factors of content tagging or joining communities has largely benefited from existing literature. The factors affecting content tagging are inspired from the studies of traditional content tagging systems [25, 15], especially tag recommender systems [20, 14]. The factors of joining communities are inspired from the literature of social network analysis, in which this is modeled as a cascading behavior [1, 10], or the adoption of innovation [32, 5].

The predication analysis in our work is in general related to the prediction of user behavior in online communities [33, 18, 17, 23]. Sen et al. [33] mentioned that tag selection in MovieLens is affected by community influence and evolving personal tendency such as preferences and beliefs about the tags they apply. Heymann et al. [18] predicted tag usage in Del.icio.us based on page text, anchor text, surrounding hosts, and other tags for URLs. Hecht et al. [17] predicted location information through users' tweets.

Finally, our work is also related to the recommender systems in microblogging communities, including the recommendation of followee [16], content [7] and conversations [6] in Twitter. Our work presents a feasibility study of hashtag recommendation, instead of aiming at constructing a particular recommender system.

[1] http://www.delicious.com
[2] http://www.flickr.com

3. THE PROBLEM AND MEASURES

Our fundamental hypothesis is that a hashtag plays the dual role of both the signature of the content and the symbol of a community membership. Should it hold, the user's behavior of adopting a hashtag would be interpreted as either the behavior of bookmarking a content, or the behavior of adopting a community membership, or both. In this section, we present comprehensive measures to quantify the major factors of tag selection and joining communities.

3.1 Basic Notations

We focus on the behavior of a Twitter user u to post a hashtag h at time t in one of his tweets d, which is a short document with a limit of 140 characters. Adopting the common practice in information retrieval, we represent a tweet as a bag of words, i.e., $d = \{w_1, ..., w_N\}$. We denote D as a large collection of tweets, U as a set of twitter users, and H as a set of hashtags.

Multiple types of relationships exist between Twitter users: a user u can *follow* another user v, become a *friend* (i.e., followee) of v, or *mention* v in one of his own tweets, or quote and redistribute one of v's tweets by *retweeting* it. In our task, we particularly focus on the *retweeting* relationship. Rather than the following-followee relationships, a retweeting relationship is a much stronger indicator of social influence, which is a major driving force of cascading behavior in online communities [10]. Indeed, it has been proved in literature that while the number of followers sheerly represents a user's popularity, the *following* relationship reveals little about the social influence of a user [4]. In contrast, the *retweeting* relationship, akin to the practice of forwarding interesting blog posts and links via email, can be understood as a form of adoption of innovation and diffusion of information, and reveals salient aspects determining the influence [4].

Networks of users can be constructed according to a specific type of relationship. For example, a *retweet network* of users can be defined as a directed graph $G_{rt} = (V, E)$, where every vertex $u \in V$ denotes a Twitter user and every directed edge $(u, v) \in E$ denotes that the user v has retweeted at least one tweet of the user u. The direction of the edge (i.e., $u \rightarrow v$) corresponds to the direction of information diffusion. An edge can be further weighted by $rt_{u,v}$, the number of times that u was retweeted by v. The larger $rt_{u,v}$ is, the more frequently v has been influenced by u.

For a user u, we then denote the set of users he has retweeted as I_u, corresponding to the in-links of u in G_{rt}. The set of users who have retweeted u is denoted as O_u, corresponding to the out-links of u. Clearly, the behavior of users in I_u can directly influence the behavior of u, while the behavior of u can directly influence users in O_u.

Definition 1: Hashtag Adoption.

We define a hashtag h as *used* by a user u if it appears in one of her tweets. We call a hashtag usage an "*adoption*" if an **existing** hashtag h is for the **first** time used by a user u in her tweet d. Thus we exclude the scenario that h is invented by u. The adoption is *spontaneous* if the tweet d is composed by the user herself instead of by retweeting others.

The adoption of hashtags is the user behavior we focus on in this study. Given a time-stamped collection of tweets D, a set of users U, and a retweeting network G_{rt}, we now present formal measures to quantify the major factors of selecting a content tag and adopting a community membership.

3.2 Measures Related to Content Tagging

We start with the role of a hashtag as an indicator of content. A user uses a hashtag to characterize his interest, to annotate the topic of a tweet, and to share the content to his peers. To this end the adoption of a hashtag is similar to the adoption of a bookmark in Del.icio.us, or a tag in Flickr, or a user-selected keyword in weblogs. What factors affect the user's selection of such a content tag? How can they be measured? Clues can be found from literature, especially from those about recommender systems of tags [14].

To summarize, a tag recommender system usually filters tags by optimizing a number of objectives, closely corresponding to the major factors of how a user selects a tag. These factors include: 1) the *popularity* of the tag; 2) the *relevance* to the content; and 3) the closeness to the user's personal *preference*. While the popularity of a tag is easy to measure, the relevance objective is usually modeled by information retrieval methods such as adaptive filtering [2]. The preference objective, on the other hand, is usually optimized using collaborative filtering techniques (e.g., [31]).

Largely inspired by this body of existing literature, we discuss how to measure the two corresponding characteristics of a hashtag: the magnitude of *relevance* to the content and the closeness to the user *preference*.

Relevance

We first measure the *relevance* between a hashtag and the content it annotates. It is worth giving attention that there are two different scenarios in tag recommendation. In some cases when the document to be annotated is known (e.g., suggesting bookmarks for a particular web page), the relevance is measured between the tag and the document. In more other cases, the recommendation is made independent to a particular document (e.g., suggesting tags on the front-page of Delicious). In such cases, the relevance is judged between the tag and a dynamic profile of the user. The user profile is usually constructed by accumulating all historical content generated or consumed by the user. Such a method is known as adaptive filtering in information retrieval [2]. Note that the extremely sparse information in the tag string is usually ineffective in the assessment of relevance. It is common practice to construct and use a rich representation of a tag based the context - all documents it annotated.

In our task, we adopt the intuition of adaptive filtering by measuring the "user-level" relevance instead of a "tweet-level" relevance. This is not only because a tweet is extremely short, but also because we care about whether a user will adopt a hashtag sometime in the future, instead of whether she will adopt a hashtag in a particular tweet. This leads to the following family of measures:

Definition 2: Relevance Measures.

The relevance between a hashtag h and a user u is measured by a similarity function of two profiles, $sim(D_h, D_u)$, where the hashtag profile D_h integrates all tweets containing h and the user profile D_u consists of all tweets posted by u up to a given time point.

Preference

We then measure how close a hashtag is tied to the per-

sonal *preference* of a user. Such a personalized preference plays a critical role in tag recommender systems [14]. Among all the tags relevant to the content, a user usually prefers the ones she has used before, or the ones that are similar to those she has used. In the problem of hashtag adoption, we are interested in hashtags that have not been used by the user.

Definition 3: Preference Measures.

The degree to which a hashtag h is close to the personal preference of u is measured by an aggregate function $f(\cdot)$ of the similarity between h and all hashtags ever used by u. Let H_u be the set of hashtags u has ever used up to a given time point, a preference measure takes the form of $f(\{sim(h, h')|h' \in H_u\})$.

Note that the similarity between two hashtags is computed based on a richer profile of two hashtags instead of the tag string. The profile is constructed using either all the tweets h appeared in, D_h, or all users who have used h, U_h. Any reasonable aggregate function $f(\cdot)$ introduces an instantiation of the preference measure, such as the maximum, minimum, sum, or a weighed average.

It is worth mentioning that while the relevance measure reflects the intuition behind many content filtering services [2], the preference measures reflect the intuition behind a common approach to collaborative filtering - to first establish the similarity among items, and then recommend items that are similar to the items already rated by the user [31]. Please also note that both the relevance and the preference are time-variant, as both the semantics of a hashtag and the personal interest/preference of a user change fast over time.

3.3 Measures Related to Joining Community

Hashtags also play a relatively implicit role as the symbol of a community membership. While a user can also use traditional types of social tags such as Flickr tags to find people with similar interests, a *community* clearly differs from a flat set of people by the rich interconnections among the members. Indeed, we calculated the average level of interactions within a "hashtag community," based on the number of retweets, replies, and mentions between community members. The level of interaction is significantly higher than among random users. Based on our computation, the network density of hashtag communities is around a thousand times larger than the network density among users using the same tag in Flickr, LiveJournal, or YouTube.

To this end, the user behavior of adopting a hashtag can be interpreted as the adoption of a community membership (i.e., to join a community). In social network analysis, the behavior of joining a community is usually treated as an instantiation of information diffusion or cascading behavior [10, 1], similar to the *adoption of innovation* [32, 5]. By adopting the membership of a community, the newcomer adopts the norm of the community.

Our second task is to quantify the major factors that affect the behavior of joining communities. We extract a summary of several factors from the literature investigating the adoption of community membership: 1) how large the community is; 2) how prestigious the community members are; 3) how much the user is influenced by his friends already in the community; and 4) how actively the community members interact. Apparently, the size of a "hashtag community" is highly correlated to the popularity of the hashtag. The activity of a community is either correlated to the popularity of the hashtag, or can be interpreted as a form of indirect social influence, where the community members propagate their influence through the inner-community links. We focus on two essential factors of the adoption of community membership: *prestige* and *influence*.

Prestige

We start with the measurement of the *prestige* of community members designated by a hashtag. To do this, we first propose a way to measure the prestige of individual users. The measurement of prestige is a classical problem in social network analysis, leading to many possible instantiations [10, 27]. Intuitively, if a user is followed/retweeted by many users, she appears to be more prestigious; she appears to be even more prestigious if her followers are prestigious. In this work, we adopt the PageRank score [29], which has been proven to be a robust measure of prestige. Similar to the context of ranking Web pages or scientific papers, PageRank of Twitter users can be computed based on either the following/followee graph or the retweet graph. We believe the retweet network is a better choice because retweeting implies an endorsement of another user's content, similar to linking to a Web page or citing a scientific article.

Definition 4: Prestige Measures.

Let U_h be a hashtag community, essentially a set of users who have used the hashtag h. The prestige of a hashtag community is measured by the result of an aggregate function of the prestige score of each existing member of the hashtag community, $f(\{p_u|u \in U_h\})$. p_u is realized as the PageRank score of the vertex u in the retweet network G_{rt}. Once again, any reasonable aggregate function $f(\cdot)$ introduces an instantiation of the prestige measure.

Influence

We then propose measures for the social *influence*, also known as the "neighborhood effect." If many of my friends have adopted an innovation, or have joined a community, so will I [1]. To measure this effect, a critical issue is how to identify the "neighborhood," or the people who have an influence on the behavior of the user. Once again, we prefer to utilize the retweeting relationship instead of the following/followee relations. This is because a retweeting behavior provides a much stronger indication of cascading behavior and social influence.

We quantify the social influence of a hashtag community on a user by aggregating the influence of its individual members on that user. Different realizations of individual influence define different instantiations of the influence measure.

Definition 5: Influence Measures.

Let U_h be a hashtag community, essentially the set of users who have used hashtag h. The magnitude of influence U_h has on a user u is defined as a function of the influence of individual users $v \in U_h$ on user u given the retweet network structure G_{rt}, $g(\{s(v, u)|v \in U_h\}, G_{rt})$. The strength of influence of a single user v on u, $s(v, u)$, is either realized as a constant if $v \in I_u$ and zero otherwise, or proportional to the prestige of v, or proportional to the frequency that u has retweeted v's tweets, $rt_{v,u}$ (as a surrogate of the frequency that v has influenced u before), etc.

The function $g(\cdot)$ can be realized as any reasonable aggregate function of all the individual influences $s(v, u)$. In this case, the inference measure excludes the effect of the community members who are not u's direct in-neighbors. A more general realization of $g(\cdot)$ propagates the influence of each community member through her out-neighbors to the whole retweet network. In other words, the influence of a particular user and a community of users is now "smoothed" on the retweet network G_{rt}. Such a measure also reflects the inner-connectivity of a hashtag community.

Like relevance and preference, prestige and influence measures are also time-variant because the retweeting behaviors change frequently over time.

3.4 Other Measures

Our discussion has been focused on factors of tagging content and joining communities, namely *relevance*, *preference*, *prestige*, and *influence*. Each of the four factors can be instantiated as a family of various measures, according to the particular selection of functions. Beside these four, we are aware of many other potential factors of hashtag adoption. They characterize either the property of the hashtag (e.g., the popularity of the hashtag), or the user (e.g., whether she has been actively posting), or the community defined by the hashtag (e.g., the size of the community). In fact, some of these factors, such as the popularity of hashtag and the size of the hashtag community, are highly correlated (with a correlation higher than 0.9). Not surprisingly, the number of people adopting a hashtag is highly correlated to the number of times the hashtag is used.

Factors like these may be highly correlated to the behavior of hashtag adoption. The pure analysis of only these factors, however, will not directly help us identify the two roles of the hashtag. They are either correlated with both roles, or neither of them. Nevertheless, these factors should still be included in our analysis, in order to comprehensively understand the effect of the content-specific and community-specific factors in the context of hashtag adoption. Below we present a broader range of measures that are likely to predict the adoption of hashtags but not relevant to testing the dual role:

POPULARITY: the number of times a hashtag has been posted; or the number of users who adopted the hashtag.

LENGTH: the number of characters in the hashtag string.

FRESHNESS: the age of a hashtag, or the length of period since its first burst.

DEGREE: the number of people a user has retweeted or been retweeted by.

ACTIVENESS: the frequency or the ratio of tweets, retweets, and hashtags a user has posted in the past.

We refer to these measures as role-unspecific measures, which would provide a baseline prediction model not specific to either of the roles. Together with the four families of content and community specific measures (role-specific features), these general measures will be instantiated in Section 4 and used in the regression and prediction analysis.

4. DATA AND FEATURE INSTANTIATION

We have introduced formal measures to quantify role-specific factors of hashtag adoption, including the relevance, preference, prestige, and influence. Do the measures make sense? Are these factors truly correlated with the use of hashtags? As a proof of concept, we first conduct a straightforward exploratory analysis to demonstrate the correlation between each of the factors and the user behavior of hashtag usage. We start with an introduction of the datasets and the specific instantiations of the measures used in our analysis.

4.1 Datasets

Two real world datasets are collected from Twitter.com, corresponding to two different sampling strategies.

The first dataset collects all the tweets posted by a sample of political *users* between March 25th 2007 to December 13th 2010 [24]. The sample frame of the political users consists of House, Senate and gubernatorial candidates during the midterm elections in the United States (i.e., year 2010). The twitter accounts of these candidates are matched by querying Google with the names of the candidates and keyword "Twitter." Then a manual process is applied to inspect and filter the top 3 results so that only the Twitter accounts operated by the candidates or their staff are kept in the collection. Different from [24], we do not further filter the set of users according to their activity level. Hence, the dataset, notated as the POLITICS dataset, contains $373,439$ tweets and $9,343$ hashtags posted by $1,029$ users.

The second dataset is constructed from a random sample of *tweets* at a much larger scale. This sample is drawn from a roughly 5% sampled stream of tweets between June 1 2009 and December 31 2009. This dataset, notated as the STREAM dataset, includes roughly 19 million users, 49 million unique hashtags, and 476 million tweets.

4.2 Instantiation of Measures

Given a time period Δ, a user u, and a hashtag h which was created before Δ, we construct a user-hashtag pair $\langle u, h\rangle$. We then use the hashtag usage, i.e., the number of times h is used by u during Δ, to surrogate the level of interest of the user on the hashtag. If h hasn't been used by u during and before Δ, the level of interest will be set to zero.

We now present all the particular instantiations of the general measures proposed in Section 3. These instances are used in the explanatory analysis to demonstrate their correlation to hashtag usage. They will also be used as independent variables in the regression and features in the prediction model to predict hashtag adoption.

Note that most of the measures, including the the four role-specific measures, are sensitive to time. In our analysis, we carefully select time windows of various lengths to instantiate these measures. We introduce additional notations related to the selection of time windows: Δ as a time window, $n^{\Delta}_{u,h}$ as the frequency u uses h during Δ, $nr^{\Delta}_{u,h}$ as the frequency of hashtag usage in retweets, and $D^{\Delta}_{u,h}$ as set of tweets containing h that are posted by u during Δ. Table 1 summarizes all features instantiated from both the role-unspecific measures (i.e., baseline features) and the four role-specific measures. Please note that the similarity of two hashtags is computed based on either the tweets or the users using the hashtags, instead of based on the hashtag strings.

4.3 A Demonstrative Analysis

Now that we have the particular instantiations of the adoption behavior and the features according to the role-specific factors, we present a simple correlation analysis to demonstrate the relationship between these role-specific factors and a user's degree of interests in a hashtag (surrogated by the frequency of using a hashtag).

Table 1: The complete list of instantiations (features) of the general measures. [i]

Abbreviation	Description	Mathematical Presentation
Role-unspecific features (baseline)		
N.usrTweet.Δ	Number of u's tweets (including retweets)	$\lvert D_u^\Delta \rvert$
N.rt.Δ	Number of u's retweets	$\lvert \{d \lvert d \in D_u^\Delta \text{ and } d \text{ is an RT}\} \rvert$
N.uniTag.Δ	Number of unique hashtags used in u's tweets	$\lvert \{h \lvert n_{u,h}^\Delta > 0\} \rvert$
N.uniTag.rt.Δ	Number of unique hashtags used in u's retweets	$\lvert \{h \lvert nr_{u,h}^\Delta > 0\} \rvert$
N.uniTag.nonRT.Δ	Number of unique hashtags used in u's non-retweets	$\lvert \{h \lvert n_{u,h}^\Delta - nr_{u,h}^\Delta > 0\} \rvert$
N.tag.Δ	Number of times that u uses any hashtag in tweets	$\sum_h n_{u,h}^\Delta$
N.tag.rt.Δ	Num. of times that u uses any hashtag in retweets	$\sum_h nr_{u,h}^\Delta$
N.tag.nonRT.Δ	Num. of times that u uses any hashtag in non-retweets	$\sum_h n_{u,h}^\Delta - \sum_h nr_{u,h}^\Delta$
Prop.tag.Δ	The proportion of u's tagged tweets	$\frac{\lvert D_{u,h}^\Delta \rvert}{\lvert D_u^\Delta \rvert}$
Prop.tag.rt.Δ	The proportion of u's tagged retweets	$\frac{\lvert \{d \lvert d \in D_{u,h}^\Delta \text{ and } d \text{ is RT}\} \rvert}{\lvert D_u^\Delta \rvert}$
Prop.tag.nonRT.Δ	The proportion of u's tagged non-retweets	$\frac{\lvert \{d \lvert d \in D_{u,h}^\Delta \text{ and } d \text{ is not RT}\} \rvert}{\lvert D_u^\Delta \rvert}$
Indegree.Δ	Number of users u retweeted	$\lvert I_u^\Delta \rvert$
Outdegree.Δ	Number of users who retweeted u	$\lvert O_u^\Delta \rvert$
Prestige.usr.Δ	Prestige (PageRank score) of user u	p_u^Δ
Popularity.Δ	Number of tweets containing h	$\sum_u n_{u,h}^\Delta$
N.tagUsr.Δ	Number of users who have used h	$\lvert \{u \lvert n_{u,h}^\Delta > 0\} \rvert$
Length	Number of characters in the string of h	length(h)
Freshness.Δ	Time since h's first burst	EndDate(Δ) - Date(h.1st.burst [ii])
Role-specific features		
Relevance		
Relevance.Δ	Cosine similarity between the profile of u and h	$\cos(D_u^\Delta, D_h^\Delta)$
Preference		
Pref.sum.Δ	Sum of similarity between h and hashtags in u's tweets	$\sum_{h' \in H_u^\Delta} sim(h, h')$ [iii]
Pref.avg.Δ	Avg. similarity between h and hashtags in u's tweets	Pref.sum.Δ $/\lvert H_u^\Delta \rvert$
Pref.max.Δ	Max. similarity between h and hashtags in u's tweets	$\max_{h' \in H_u^\Delta} sim(h, h')$
Pref.sum.rt.Δ	Sum of sim. between h and hashtags in u's retweets	$\sum_{h' \in H_u^\Delta(rt)} sim(h, h')$
Pref.avg.rt.Δ	Avg. sim. between h and hashtags in u's retweets	Pref.sum.rt.Δ $/\lvert H_u^\Delta(rt) \rvert$
Pref.max.rt.Δ	Max. sim. between h and hashtags from u's retweets	$\max_{h' \in H_u^\Delta(rt)} sim(h, h')$
Pref.sum.nonRT.Δ	Sum of sim. between h and hashtags in u's non-retweets	$\sum_{h' \in H_u^\Delta \setminus H_u^\Delta(rt)} sim(h, h')$
Pref.avg.nonRT.Δ	Avg. sim. between h and hashtags in u's non-retweets	Pref.sum.rt.Δ $/\lvert H_u^\Delta \setminus H_u^\Delta(rt) \rvert$
Pref.max.nonRT.Δ	Max. sim. between h and hashtags in u's non-retweets	$\max_{h' \in H_u^\Delta \setminus H_u^\Delta(rt)} sim(h, h')$
Prestige		
Pres.sum.Δ	Sum prestige of users who have used h	$\sum_{u \in \{u \lvert n_{u,h}^\Delta > 0\}} p_u^\Delta$
Pres.avg.Δ	Average prestige of users who have used h	Pres.sum.Δ $/\lvert \{u \lvert n_{u,h}^\Delta > 0\} \rvert$
Pres.max.Δ	Maximum prestige of users who have used h	$\max_{u \in \{u \lvert n_{u,h}^\Delta > 0\}} p_u^\Delta$
Pres.wt.Δ	Sum of prestige, weighted by the freq. of hashtag usage	$\sum_{u \in \{u \lvert n_{u,h}^\Delta > 0\}} p_u^\Delta n_{u,h}^\Delta$
Influence		
Inf.abs.num.Δ	Number of u's in-neighbors who have used h	$\lvert \mathcal{I}_u^\Delta \rvert$, where $\mathcal{I}_u^\Delta = \{v \lvert v \in I_u^\Delta, n_{v,h}^\Delta > 0\}$
Inf.abs.prop.Δ	Proportion of u's in-neighbors who have used h	$\lvert \mathcal{I}_u^\Delta \rvert / \lvert I_u^\Delta \rvert$
Inf.rtf.*.Δ [iv]	Influence calculated by retweet frequency	$s(x,u) = rt_{x,u}^\Delta$
Inf.ratio.*.Δ	Influence calculated by retweet ratio	$s(x,u) = Pr_{x,u}^\Delta, Pr_{x,u}^\Delta = \frac{rt_{x,u}^\Delta}{\sum_{\{v \in I_u^\Delta\}} rt_{v,u}^\Delta}$
Inf.wtRatio.*.Δ	Influence calculated by by hashtag usage and retweet ratio	$s(x,u) = Pr_{x,u}^\Delta n_{x,h}^\Delta$
Inf.pres.*.Δ	Influence calculated by prestige	$s(x,u) = p_x^\Delta$
Inf.wtPres.*.Δ	Influence calculated by hashtag usage and prestige	$s(x,u) = p_x^\Delta n_{x,h}^\Delta$
Inf.smooth.sum.Δ	Sum of smoothed influence with retweet graph	See ref. [26] [v]

i. A different time window Δ leads to different versions of these features except for the "Length" feature.
ii. The identification of a burst follows [21].
iii. $sim(h, h')$ is computed based on either $\cos(D_h^\Delta, D_{h'}^\Delta)$ or $\cos(U_h^\Delta, U_{h'}^\Delta)$.
iv. The wildcard (*) here stands for *sum*, *avg*, and *max*, e.g., *sum* indicates $\sum_{x \in \mathcal{I}_u^\Delta} s(x,u)$.
v. The way to compute the smoothed influence follows the algorithm proposed in [26].

As a proof of concept, we collected users who have posted more than 5 hashtags during a 10-day time window between November 1st and November 10th 2009 from the STREAM dataset, and randomly sampled 30,106 users from them. We then collected all hashtags tweeted by these 30,106 users during this period, from which we obtained 269,712 user-hashtag pairs. We then obtained 40,427 user-hashtag pairs from the POLITICS dataset between August 1st and October 30th 2010. Note that identical time window is used to calculate the hashtag usage and the features.

To make it easier to visualize, we bin the values of each feature into buckets, so that each bucket contains the same number of user-hashtag pairs. The plots in Figure 2 present the correlation between the degree of interest in a hashtag and the role-specific factors. Not surprisingly, the degree of interest of a user in a hashtag highly correlates with all four factors: relevance to the content, personal preference of the user, community prestige, as well as social influence. Intuitively, this suggests that the factors of selecting a content tag and the factors of joining a community are both affecting the behavior of using a hashtag.

The result of this demonstrative analysis looks promising. However, the simple correlation analysis does not provide evidence on whether these factors have the predictive power of the adoption of a hashtag, or whether they would make a joint effect in reality. To draw convincing conclusions, we proceed with a formal regression analysis.

5. REGRESSION ANALYSIS

Beyond the intuitively promising results from the correlation analysis, we are looking for formal evidences on 1) whether each of the proposed measures has a predictive power of hashtag adoption; 2) if yes, how significant the effect is; and 3) whether the effect remains significant when the factors interplay with each other.

We thus conduct a regression analysis with the hashtag adoption behavior as the response (dependent) variable. Three groups of users are constructed from our two datasets:

1) a group of users selected from the POLITICS dataset, namely the POLITICS group;

2) an interest-based group of users selected from the STREAM dataset, namely the MOVIE group;

3) a group of users randomly sampled from the STREAM dataset, namely the RANDOM group.

5.1 Experiment Setup

We formulate the regression analysis such that the dependent variable corresponds to whether or not a user will adopt a hashtag in a time period *in the future*. The independent variables correspond to particular instantiations of the measures in Section 3, either role-specific or role-unspecific. Since it is impossible to observe future events, we hold out the observations from a time window of 10 days, Δ_1, to surrogate the hashtag adoption behaviors in the future. Values of all independent variables are computed only based on the observations in a time period Δ_0 *prior to* Δ_1.

We prepare the data of our regression analysis as follows:

1) Given a dataset and a time window Δ_1, we extract either the complete set of users or a sample of users, all of which have been using hashtags recently (i.e., within 10 days prior to the held-out period Δ_1). The sampling process will be discussed in the next subsection.

2) For each user u selected in 1), we collect all hashtags posted by u in the held-out period Δ_1 (i.e., in the "future") but have not appeared in any tweet of u before Δ_1. These hashtags are "adopted" by the user in the held-out period. Every such pair $\langle u, h\rangle$ is formatted as a data point in the regression analysis, with the outcome variable set to 1.

3) For each user u selected, we randomly sample a set of hashtags from up to 10 days prior to Δ_1, which have **not** been used by u in either Δ_1 or ever before. These hashtags are representatives of those that exist but are **not** adopted by the user. Every such pair $\langle u, h\rangle$ is also formatted as a data point, with the outcome variable set to 0.

In the regression analysis, we use 10 days between November 11th 2009 and November 20th 2009 as the held-out period Δ_1 for the MOVIE and the RANDOM group. For the POLITICS group, we partition the period between August 1st and October 30th 2010 into nine equivalent time windows and employed each of them as a held-out period. The length of each held-out period $Delta_1$ is thus also 10 days. For all data points sampled through the above procedure, we compute the values of all the independent variables (i.e., one particular instantiation for each role-specific or role-unspecific measure) based on all the tweets posted within 30 days before Δ_1 and the snapshot of the retweet network right before Δ_1. In other words, Δ_0 is designed as the 30-day time window prior to Δ_1, with respect to each particular realization of Δ_1. Below we provide more details on how the users and user-hashtag pairs are sampled.

5.1.1 Sample Users

The first task of data preparation is to select the set of users from a given dataset. As for the POLITICS dataset, we simply select all users who have used at least one hashtag within 10 days before the held-out period Δ_1, representing active hashtag users. For the STREAM dataset, however, it is computationally costly to run the analysis on all the 17 million users. We thus apply the following two sampling frames: an interest-based sample and a random sample.

Interest-based sampling

Our first sampling frame targets at users with a focused aspect of interest. We chose "movie" as our target aspect because *i)* many trending topics are related to movies according to the "Top Twitter Topics" by Mashable[3], and *ii)* people discuss a lot on movie related topics in real life.

In particular, we select the set of users from the STREAM dataset using the following snowball sampling procedure:

1) Construct the retweet network from the complete collection of the STREAM dataset.

2) Select top 3 most popular Twitter users from Twellow[4] in category "movies" as seeding users.

3) Following the *out-links* in the retweet network, we collect all the first-order and second-order neighbors of the seedling users. With such a process, we have selected users who have either retweeted the three seedling users, or have retweeted another user who have retweeted them.

Combing the seedling users, their first-order and second-order out-neighbors, a pool of 178,323 Twitter users is constructed. A further process selects a set of users who have posted at least one hashtag within 10 days before Δ_1. In

[3]http://mashable.com/tag/top-twitter-topics

[4]http://twellow.com

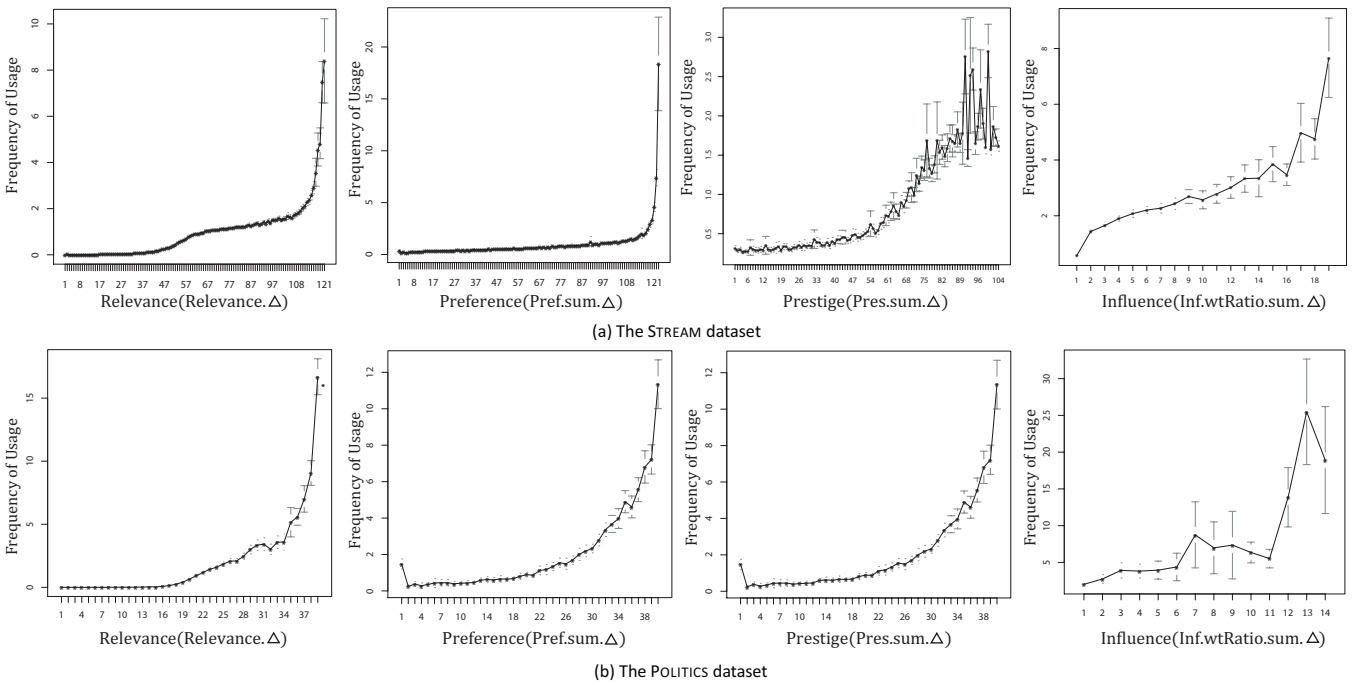

Figure 2: The degree of interest in hashtags correlates to the role-specific factors.

the end, we obtain 8,029 users for this group, and we name the group as the MOVIE group.

Random sampling

Another sampling strategy is to randomly select users from the entire dataset. We gather all users that appear in the STREAM dataset, and randomly select a set of users who have posted at least one hashtag within 10 days prior to Δ_1. In the end, we collected 15,038 users for this group, which is named as the RANDOM group.

5.1.2 Sample Negative Examples

Given a set of users, it is straightforward to generate data points (i.e., user-hashtag pairs) with a positive outcome of hashtag adoption. The data points with a negative outcome (i.e., the event that a hashtag is not adopted), however, is trickier. The simplest way is to create a negative example for each existing hashtag (i.e., within 10 days before the held-out period Δ_1) that a user did not use during Δ_1. However, the massive number of hashtags simply makes the negative examples dominating the collection, diluting all useful signal. Thus, for each user, we randomly sample negative examples proportional to the number of positive examples (i.e., number of hashtags adopted by this user).

5.2 Results of Regression Analysis

We employ *logistic* regression to predict the adoption of hashtags in the held-out time period Δ_1 from features computed based on Δ_0 prior to Δ_1. One instantiation of each measure is included. We set the time window Δ_0 as the period of 30 days prior to Δ_1. To comprehensively understand the predictive power of the role-specific measures, we also incorporate baseline features into the regression analysis. Table 2 presents the results of the regression analysis.

Apparently, all the four role-specific measures have a significant and positive predictive power of hashtag adoption, even when merged together with all baseline measures. One baseline measure, the variable *Popularity* in the MOVIE and RANDOM group, yields a negative coefficient in the regression. This is because the popularity of a hashtag is highly correlated with the sum of prestige of hashtag users (e.g., with a correlation over 0.9196 in the MOVIE group). However, such correlation in the POLITICS group is lower, thus yielding a positive coefficient for *Popularity*. Indeed, when we remove the *Prestige* variable from the regression, the coefficient of *Popularity* becomes positive in the two groups (and remain significant). The *Length* of hashtags presents a negative relationship with hashtag adoption, possibly because people tend to adopt short and concise hashtags.

The results of the regression analysis thus provide a much stronger evidence that both the content role and the community role affect hashtag adoption. All four role-specific factors we presented have a significant positive predictive power to the adoption of hashtags.

6. PREDICTION ANALYSIS

The regression analysis has proved that all the role-specific measures are predictive to the adoption of hashtags. This reassures our hypothesis that the dual role of a hashtag affects the adoption. We are, however, moving forward to investigate the feasibility of constructing an effective prediction and recommender systems. Unfortunately, the regression analysis doesn't tell how an effective prediction system can be constructed based on these measures, or how hashtag recommendation can be done. To test the feasibility of hashtag recommendation, we proceed with building a machine learning model to predict the adoption of new hashtags in the future. Compared to the regression analysis that serves as a white box to interpret the effect of the measures, this analysis provides a black box that aims at optimizing the accuracy of of predictions.

Table 2: Regression Results.

Group	Feature abbr.	β	Significance
Politics	Popularity.Δ_0	2.911e-04	
	Indegree.Δ_0	-1.096e-01	
	Outdegree.Δ_0	-9.604e-02	
	Length	-4.345e-02	***
	N.uniTag.Δ_0	-2.831e-03	
	Inf.num.Δ_0	1.991e-00	***
	Pref.sum.Δ_0	1.211e-02	***
	Relevance.Δ_0	2.262e+01	***
	Pres.sum.Δ_0	8.466e+01	***
	Sample size = 3,753		
Movie	Popularity.Δ_0	-6.955e-05	***
	Indegree.Δ_0	-1.181e-03	***
	Outdegree.Δ_0	-1.975e-03	**
	Length	-4.613e-02	***
	N.uniTag.Δ_0	1.200e-03	***
	Inf.num.Δ_0	1.030e+00	***
	Pref.sum.Δ_0	2.472e-02	***
	Relevance.Δ_0	3.962e+01	***
	Pres.sum.Δ_0	8.086e+03	***
	Sample size = 26,188		
Random	Popularity.Δ_0	-3.512e-05	***
	Indegree.Δ_0	-3.777e-03	***
	Outdegree.Δ_0	-7.536e-04	***
	Length	-9.904e-02	***
	N.uniTag.Δ_0	1.107e-03	***
	Inf.num.Δ_0	1.720e+00	***
	Pref.sum.Δ_0	2.514e-02	***
	Relevance.Δ_0	5.186e+01	***
	Pres.sum.Δ_0	7.478e+03	***
	Sample size = 27,878		

Significant at the: *** 0.01, ** 0.05, or * 0.1 level. Bold: role-specific measures.

Table 3: Statistics of training and test datasets.

	Movie		Random		Politics	
All	Train	Test	Train	Test	Train	Test
# of (+)	13,071	11,932	13,969	12,393	1,886	1,086
# of (-)	13,117	11,884	13,909	12,464	1,867	1,071
NonRTs	Train	Test	Train	Test	Train	Test
# of (+)	6,348	5,912	8,093	7,233	1,612	928
# of (-)	6,358	5,842	8,106	7,272	1,600	931
RTs	Train	Test	Train	Test	Train	Test
# of (+)	7,332	6,550	6,368	5,536	335	207
# of (-)	7,397	6,618	6,346	5,472	334	208

NonRTs = Non-retweets, RTs = Retweets

Table 4: Accuracy of the SVM predictor.

Group	Measures	Accuracy (%)		
		All Tweets	Non-RTs	Retweets
Politics	(B)aseline	68.15	66.97	65.54
	B+Rel.	75.29 ***	74.23 ***	72.53 ***
	B+Pref.	70.84 ***	71.17 ***	67.23 ***
	B+Inf.	69.31 ***	68.42 ***	67.23 ***
	B+Pres.	75.52 ***	74.88 ***	71.32 ***
	All	**78.25** ***	**78.32** ***	**74.93** ***
Movie	(B)aseline	75.98	74.43	77.10
	B+Rel.	80.42 ***	78.93 ***	81.66 **
	B+Pref.	79.63 ***	77.66 ***	80.62 ***
	B+Inf.	79.93 ***	76.89 ***	81.04 ***
	B+Pres.	74.09 ***	71.57 ***	74.12 ***
	All	**80.64** ***	**79.13** ***	**82.80** ***
Random	(B)aseline	74.66	73.30	75.41
	B+Rel.	83.19 ***	82.64 ***	84.50 ***
	B+Pref.	81.39 ***	79.97 ***	83.39 ***
	B+Inf.	77.42 ***	75.56	80.18 ***
	B+Pres.	74.37 ***	73.39 ***	75.72 ***
	All	**84.03** ***	**82.45** ***	**85.64** ***

Significant at the: *** 0.01, ** 0.05, or * 0.1 level, paired t-test. Rel.: Relevance; Pref.: Preference; Inf.: Influence; Pres.: Prestige

6.1 Experiment Setup

Consistent to the regression analysis, we setup three experiments over three groups of users, and sample negative examples accordingly. Held-out time periods are employed to surrogate the "future." The prediction problem is cast as a binary classification task: whether or not a user will adopt a new hashtag in a held-out period. The performance of such a classifier is evaluated by the accuracy of the predicted classes. Strictly, the behavior of a user to adopt a hashtag in her own tweets is quite different from adopting a hashtag by retweeting others. The former is spontaneous and the latter is passive. We further differentiate the tasks of predicting hashtag adoption in all tweets, retweets, and user-composed tweets (i.e., non-retweets) respectively.

Different from the regression analysis, we train the prediction model with a training dataset and assess the effectiveness with a separate test dataset. To do this, we select different pairs of "history" (Δ_0) and "future" (Δ_1) time windows. We then use some "history-future" pairs to train the prediction model, and use other time window pairs to test the model. From each pair of time windows, we sample a collection of user-hashtag pairs as data examples. For each data example, we then compute **all** the features in Table 1 based on Δ_0, and identify the label (positive or negative) based on whether the hashtag is adopted by the user during Δ_1. The number of positive and negative samples in our training and test datasets are presented in Table 3.

6.2 Results and Discussion

We first employ an SVM classifier with all baseline features in Table 1. The RBF kernel is adopted, and 5-fold cross-validation is used to select parameters. The performance of the SVM predictor is presented in Table 4. Such a baseline model performs reasonably well - with 68% accuracy on the Politics group and around 75% on both the Movie group and the Random group, comparing to a 50% accuracy of random guess. We then add the all features instantiating each of the four role-specific measures into the classifier, to test the four families of measures one by one, and then altogether. The inclusion of each of the four families of measures has improved the prediction performance significantly, with a few exceptions of the Prestige measures. In all three groups, the mixture of all four families of features has further improved the prediction accuracy.

The only exception is the *prestige* measure on the Movie and Random group. Surprisingly, the inclusion of the prestige measure even decreases the prediction performance when all tweets are considered. While on the Politics group, the prestige features performed significantly well. This is possibly due to the way prestige is computed. Of both the Movie group and the Random group, prestige is computed based on the global retweet network of millions of users, most of which are not part of the group. Therefore, the prestige is global and unspecific to the community of the group. Of the Politics group, however, the retweet network is specific to the community. The prestige is thus relatively "localized" and more specific to the user group [4]. Therefore, we further conducted an experiment to remove the prestige measures from the Movie and the Random group. The accuracy of

the prediction task over MOVIE group increased from 80.64% to 82.00%, and the accuracy over RANDOM group increased from 84.03% to 84.35%.

Interestingly, predicting the hashtag adoption in retweets seems to be easier than predicting the behavior in spontaneous tweets (MOVIE and RANDOM). The exception is the POLITICS group, where the candidates prefer to tweet rather than to retweet (Table 3). In general, it is reasonable that the prediction of content in retweets is easier than the prediction of spontaneous tweeting behavior.

With all features included, the best prediction model achieves an accuracy around 80% among all datasets. This is a promising result, suggesting the feasibility of building effective recommender systems of hashtags in Twitter.

7. CONCLUSION

We presented a formal empirical analysis to test how the dual role of hashtags in Twitter affects hashtag adoption. Results of a correlation analysis, a regression analysis, and a prediction analysis all suggest that a hashtag serves as both a tag of content and a symbol of membership of a community. The measures we propose to quantify the factors of tagging content or joining communities all present significant predictive power to the adoption of hashtags. The prediction analysis using a SVM predictor provides a feasibility study of hashtag recommender systems, which suggests a promising future direction of research.

8. ACKNOWLEDGMENTS

This work is supported by the National Science Foundation under grants no. IIS-0968489 and IIS-1054199.

9. REFERENCES

[1] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In *KDD '06*, pages 44–54, 2006.

[2] N. J. Belkin and W. B. Croft. Information filtering and information retrieval: two sides of the same coin? *Commun. ACM*, 35:29–38, 1992.

[3] C. Castillo, M. Mendoza, and B. Poblete. Information credibility on twitter. In *WWW '11*, pages 675–684, 2011.

[4] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring user influence in twitter: the million follower fallacy. In *ICWSM '10*, 2010.

[5] H.-C. Chang. A new perspective on twitter hashtag use: diffusion of innovation theory. In *ASIS&T '10*, pages 85:1–85:4, 2010.

[6] J. Chen, R. Nairn, and E. Chi. Speak little and well: recommending conversations in online social streams. In *CHI '11*, pages 217–226, 2011.

[7] J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi. Short and tweet: experiments on recommending content from information streams. In *CHI '10*, pages 1185–1194, 2010.

[8] D. Davidov, O. Tsur, and A. Rappoport. Semi-supervised recognition of sarcastic sentences in twitter and amazon. In *CoNLL '10*, pages 107–116, 2010.

[9] L. D'Monte. Swine flu's tweet tweet causes online flutter. *Business Standard*, 2011.

[10] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

[11] M. Ebner, H. Mühlburger, and et al. Getting granular on twitter : tweets from a conference and their limited usefulness for non-participants. *KCKS '10*, pages 102–113, 2010.

[12] G. Golovchinsky and M. Efron. Making sense of twitter search. In *CHI '10 Workshop on Microblogging*, 2010.

[13] M. Graves. The 2010 world cup: a global conversation. *Twitter Blog*, 2010.

[14] Z. Guan, J. Bu, Q. Mei, C. Chen, and C. Wang. Personalized tag recommendation using graph-based ranking on multi-type interrelated objects. In *SIGIR '09*, pages 540–547, 2009.

[15] M. Gupta, R. Li, Z. Yin, and J. Han. Survey on social tagging techniques. *SIGKDD Explor. 10*, 12:58–72, 2010.

[16] J. Hannon, M. Bennett, and B. Smyth. Recommending twitter users to follow using content and collaborative filtering approaches. In *RecSys '10*, pages 199–206, 2010.

[17] B. Hecht, L. Hong, B. Suh, and E. H. Chi. Tweets from justin bieber's heart: the dynamics of the location field in user profiles. In *CHI '11*, pages 237–246, 2011.

[18] P. Heymann, D. Ramage, and H. Garcia-Molina. Social tag prediction. In *SIGIR '08*, pages 531–538, 2008.

[19] J. Huang, K. M. Thornton, and E. N. Efthimiadis. Conversational tagging in twitter. In *HT '10*, pages 173–178, 2010.

[20] R. Jäschke, L. Marinho, A. Hotho, S.-T. Lars, and S. Gerd. Tag recommendations in folksonomies. In *PKDD '07*, pages 506–514, 2007.

[21] Y. Jiang, C. X. Lin, and Q. Mei. Context comparison of bursty events in web search and online media. In *EMNLP '10*, pages 1077–1087, 2010.

[22] J. Letierce, A. Passant, J. Breslin, and S. Decker. Understanding how twitter is used to widely spread scientific messages. In *WWW '10 Workshop WebSci10*, 2010.

[23] D. Liben-Nowell and J. Kleinberg. The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.*, 58(7):1019–1031, 2007.

[24] A. Livne, M. Simmons, E. Adar, and L. Adamic. The party is over here: structure and content in the 2010 election. In *ICWSM '11*, 2011.

[25] C. Marlow, M. Naaman, D. Boyd, and M. Davis. Ht06, tagging paper, taxonomy, flickr, academic article, to read. In *HT '06*, pages 31–40, 2006.

[26] Q. Mei, D. Zhang, and C. Zhai. A general optimization framework for smoothing language models on graph structures. In *SIGIR '08*, pages 611–618, 2008.

[27] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[28] R. Noon and H. Ulmer. Analyzing conferences in twitter with social aviary. *Stanford University CS 322*, 2009.

[29] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, 1999.

[30] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei. Rumor has it: identifying misinformation in microblogs. In *EMNLP '11*, pages 1589–1599, 2011.

[31] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *CSCW '94*, pages 175–186, 1994.

[32] E. M. Rogers. *Diffusion of Innovations, 5th Edition*. Free Press, 5th edition, 2003.

[33] S. Sen, S. K. Lam, A. M. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, F. M. Harper, and J. Riedl. Tagging, communities, vocabulary, evolution. In *CSCW '06*, pages 181–190, 2006.

[34] K. Starbird and L. Palen. "voluntweeters": self-organizing by digital volunteers in times of crisis. In *CHI '11*, pages 1071–1080, 2011.

[35] C. Tan, J. Tang, J. Sun, Q. Lin, and F. Wang. Social action tracking via noise tolerant time-varying factor graphs. In *KDD '10*, pages 1049–1058, 2010.

[36] C. White. Reaching 200 million accounts: twitter's explosive growth. *Mashable*, 2011.