# Who Will Retweet This? Detecting Strangers from Twitter to Retweet Information

KYUMIN LEE, Utah State University
JALAL MAHMUD, IBM Research – Almaden
JILIN CHEN, Google
MICHELLE ZHOU, IBM Research – Almaden
JEFFREY NICHOLS, Google

There has been much effort on studying how social media sites, such as Twitter, help propagate information in different situations, including spreading alerts and SOS messages in an emergency. However, existing work has not addressed how to actively identify and engage the right strangers at the right time on social media to help effectively propagate intended information within a desired time frame. To address this problem, we have developed three models: (i) a feature-based model that leverages peoples' exhibited social behavior, including the content of their tweets and social interactions, to characterize their willingness and readiness to propagate information on Twitter via the act of retweeting; (ii) a wait-time model based on a user's previous retweeting wait times to predict her next retweeting time when asked; and (iii) a subset selection model which automatically selects a subset of people from a set of available people using probabilities predicted by the feature-based model, and maximize retweeting rate. Based on these three models, we build a recommender system that predicts the likelihood of a stranger to retweet information when asked, within a specific time window, and recommends the top-N qualified strangers to engage with. Our experiments, including live studies in the real world, demonstrate the effectiveness of our work.

Categories and Subject Descriptors: **H.5.2 [Information Interfaces and Presentation]**: User Interfaces

General Terms: Design, Algorithms, Experimentation

Additional Key Words and Phrases: Twitter, Retweet, Social Media, Willingness, Personality

**ACM Reference Format**:

Kyumin Lee, Jalal Mahmud, Jilin Chen, Michelle Zhou, and Jeffrey Nichols, 2014. Who Will Retweet This? Detecting Strangers from Twitter to Retweet Information. *ACM Trans. Intell. Syst. Technol. 0, 0, Article 0* (2014), 25 pages.
DOI:http://dx.doi.org/10.1145/0000000.0000000

## 1. INTRODUCTION

With the widespread use of social media sites, like Twitter and Facebook, and the ever growing number of users, there has been much effort on understanding and modeling information propagation on social media [Agarwal et al. 2008; Bakshy et al.

---

An early version of this manuscript appeared in the 2014 ACM Proceedings of the IUI conference [Lee et al. 2014].

Author's addresses: K. Lee, Department of Computer Science, Utah State University, Logan, UT; email: kyumin.lee@usu.edu; J. Mahmud, and M. Zhou, IBM Research – Almaden, San Jose, CA; email: {jumahmud, mzhou}@us.ibm.com; J. Chen, and J. Nichols, Google, Mountain View, CA; email: jilinc@acm.org, and jeff@jeffreynichols.com. The work was done while J. Chen and J. Nichols were at IBM.

Permission to make digital or hardcopies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credits permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.
© 2014 ACM 1539-9087/2010/03-ART39 $15.00
DOI:http://dx.doi.org/10.1145/0000000.0000000

2011; Cha et al. 2010; Goyal et al. 2010; Huang et al. 2012; Romero et al. 2011; Singer 2012; Ver Steeg and Galstyan 2012; Weng et al. 2010].

Most of the work assumes that information is propagated by a small number of influential volunteers, who possess certain qualities, such as having a large number of followers, which make them extremely effective in propagating information [Starbird and Palen 2010]. For example, these users can help spread emergency alerts, such as fire hazard or SOS messages like requesting blood donations, to reach more people faster.

However, prior research efforts ignore several critical factors in influencer-driven information propagation. First, influential users may be unwilling to help propagate the intended information for various reasons. For example, they may not know the truthfulness of a piece of information, and thus are unwilling to risk their reputation to spread the information. Second, an influential user may be unavailable to help propagate information when needed. For example, influential users may not be online to help propagate SOS messages when a disaster strikes.

Since everyone is potentially an influencer on social media and is capable of spreading information [Bakshy et al. 2011], our work aims to identify and engage the right people at the right time on social media to help propagate information when needed. We refer to these people as *information propagators*. Since not everyone on social media is willing or ready to help propagate information, our goal is to model the characteristics of information propagators based on their social media behavior. We can then use the established model to predict the likelihood of a person on social media as an information propagator. As the first step, we focus on modeling *domain-independent* traits of information propagators, specifically, their *willingness* and *readiness* to spread information.

In many situations including emergency or disastrous situations, information propagation must be done within a certain time frame to optimize its effect. To satisfy such a time constraint, we thus also develop a wait-time model based on a user's previous retweeting wait times to predict the user's next retweeting time when asked.

For the sake of concreteness, in this manuscript we focus on Twitter users, although our core technology can be easily applied to other social media platforms. On Twitter, the most common method for propagating information is retweeting[1], which is to repost others' tweets in your own content stream. Our work is thus reduced to the problem of finding strangers on Twitter who will retweet a message when asked.

To model one's willingness and readiness to retweet information, we first identify a rich set of features to characterize the candidate, including derived personality traits, social network information, social media activity, and previous retweeting behavior. Unlike existing work, which often uses only social network properties, our feature set includes *personality traits* that may influence one's retweeting behavior. For example, when asked by a stranger in an emergency, a person with a high level of altruism may be more responsive and willing to retweet. Similarly, a more active user who frequently posts status updates or reposts others' tweets may be more likely to retweet when asked. Our features capture a variety of characteristics that are likely to influence one's retweeting behavior.

To predict one's likelihood to retweet when asked, we train statistical models to infer the weights of each feature, which are then used to predict one's likelihood to

[1] We use the term "repost", "retweet" and "propagate" interchangeably

retweet. Based on the prediction models, we also build a real-time recommender system that can rank and recommend the top-$N$ candidates (*retweeters*) to engage with on Twitter.

To demonstrate the effectiveness of our work, we have conducted extensive experiments, including live studies in the real world. Compared to two baselines, our approach significantly improves the *retweeting rate*[2]: the ratio between the number of people who retweeted and the number of people asked. To the best of our knowledge, our work is the first to address how to *actively* identify and engage strangers on Twitter to help retweet information. As a result, our work offers three unique contributions:

- A feature-based model including one's personality traits for predicting the likelihood of a stranger on Twitter to retweet a particular message when asked.
- A wait-time model based on a person's previous retweeting wait times to estimate her next retweeting wait time when asked.
- A subset selection model which *automatically* selects a subset of people from a set of available people using probabilities predicted by the feature-based model and maximize retweeting rate.
- A retweeter recommender system that uses the three models mentioned above to effectively select the right set of strangers on Twitter to engage with in real time.

## 2. RELATED WORK

Our work is most closely related to the recent efforts on actively engaging strangers on social media for accomplishing certain tasks [Mahmud et al. 2013; Nichols and Kang 2012]. However, ours is the first on modeling and engaging strangers on social media to aid information propagation within a given time window.

Our work is also related to the effort on characterizing retweeters and their retweeting behavior [Macskassy and Michelson 2011]. However, the existing work does not include personality features as our model does. More importantly, unlike the existing model focusing on *voluntary* retweeting behavior, ours examines a person's retweeting behavior at the request of a stranger.

Some researchers studied recommending personalized tweets. For example, Chen et al. [Chen et al. 2012] make use of different information to help recommendation, including the user's own tweet history, retweet history and social relations between users. Their method of tweet recommendation makes use of collaborative ranking to capture personal interests. Feng and Wang [Feng and Wang 2013] developed a predictive model to rank the tweets according to their probability of being retweeted. Such ranked list of tweets is recommended to the user for retweeting. In contrast to the work which recommends tweets from users' followees on Twitter, we focus on recommending retweeters. Some of our features such as personality, past retweeting rate, and readiness based features are not used in the papers. Furthermore, our work also address different objectives such as maximizing retweeting rate or information reach which were not the focus of the above work.

There are many efforts on modeling influential behavior in social media. Such work finds influential users by their social network properties [Bakshy et al. 2011; Cha et al. 2010; Goyal et al. 2010; Huang et al. 2012; Singer 2012; Weng et al. 2010], content of posts [Agarwal et al. 2008], information forwarding/propagating activity [Romero et al. 2011], and information flow [Ver Steeg and Galstyan 2012]. In

[2] We use the term "information propagation rate", "information repost rate" and "retweeting rate" interchangeably

comparison, our work focuses on an individual's characteristics that influence their willingness and readiness to retweet at a stranger's request. Some of these characteristics, such as personality and readiness to retweet, have not been studied before.

As our goal is to support effective information diffusion, our work is related to efforts in this space. Bakshy et al. [Bakshy et al. 2012] examine the role of the social network and the effects of tie strength in information diffusion. Hodas and Lerman [Hodas and Lerman 2014] show that the position of exposing messages on the user-interface strongly affects social contagion. Chaoji et al. [Chaoji et al. 2012] show how to maximize content propagation in one's own social network. In contrast, our approach aims at selecting a right set of *strangers* on social media to help spread information. Budak et al. [Budak et al. 2011] have studied a different type of information diffusion, which spreads messages to counter malicious influences, and hence minimize the influence of such campaigns. They proposed to identify a subset of individuals to start a counter campaign based on a set of viral diffusion features, including user virality and susceptibility, and item virality [Hoang and Lim 2012]. These features are complementary to the features that we use, such as personality, messaging activity, and past retweeting activity. Moreover, there is little work on automatically identifying and engaging the *right* strangers at the *right* time on social media to aid information propagation as ours does.

## 3. CREATING GROUND-TRUTH DATASETS

Since there is no publicly available ground-truth data with which we can train and build our predictive models, we collected two real-world datasets. We created a total of 17 Twitter accounts and our system automatically sent retweeting requests to 3,761 strangers on Twitter. Our first data set examines *location-based targeting*, where people who live in a particular location were asked to retweet information relevant to that location. The second examines *topic-based targeting*, where people interested in a certain topic were asked to retweet information relevant to that topic.

We hypothesize that information relevance influences a person's retweeting behavior especially at the request of a stranger. For example, people might be more likely to retweet news about public safety in an area where they live or work rather than for other locations. Similarly, a person might be more willing to retweet information on a topic in which s/he is interested.

Our dataset for location-based targeting (named *"public safety"*) and the dataset for topic-based targeting (named *"bird flu"*) are intended to examine how different types of information (location vs. topic) may impact retweeting behavior.

**Public Safety Data Collection**: For location-based targeting, we chose the San Francisco bay area as the location and sent tweets about local public safety news to people whom we identified as living or staying in that area. First, we created 9 accounts on Twitter. All accounts had the same profile name, "Public Safety News" and the same description (Figure 1).

Note that we created multiple accounts to send a few messages per hour from each account in order to create a reasonable pretense of human behavior. Furthermore, previous studies have shown that if not careful, target strangers would silently flag an account as a spam to cause the suspension of the account by Twitter [Mahmud et al. 2013; Nichols and Kang 2012]. Creating multiple accounts helped us avoid this possibility, and thus increased the number of users that we could reasonably contact per hour (each user received only one message).


**Fig. 1. An example Twitter account created for Public Safety data collection.**

Creating multiple accounts for research purposes is a commonly used methodology [Lee et al. 2010; Lee et al. 2011]. To make these accounts appear to be genuine, all accounts followed 4~10 users and had 19 followers. We also created the following and follower accounts, and some were also followed by the original accounts. We posted 11 public safety messages using each of the 9 accounts before we contacted anyone on Twitter. We identified 34,920 bay area Twitter users using the Twitter Streaming API[3] with a geo-location filter corresponding to the bay area in June 2012. This stream retrieved only tweets that were marked as being sent within a bounding box equivalent to the bay area determined by using the Google Geocoding API[4]. We filtered out non-English tweets in this stream, and created a list of unique users whose tweets were in the stream.

Among all the identified Twitter users, we randomly selected 1,902 people. From our public safety accounts, our system automatically sent messages to those people using the Twitter API and ensured that each person received only one message to avoid overburdening the person. Here is an example message sent:

*@ SFtargetuser "A man was killed and three others were wounded in a shooting... http://bit.ly/KOl2sC" Plz RT this safety news"*

Each message contained the target person's screen name, the title of a news article obtained from a local news media site, a link to the article, and a phrase asking the person to retweet the message. The original link URL was shortened with the bit.ly URL shortening service to allow us to track user clicks on the link. Per our requests, 52 of the 1,902 (2.8%) people retweeted our message, which reached a total of 18,670 followers of theirs.

**Bird Flu Data Collection**: for topic-based targeting, we chose people who tweeted about "bird flu", a topic commonly being discussed at the time of our study. First, we created 8 accounts on Twitter (Figure 2). All accounts followed 2~5 users and had 19 followers. The following and followers accounts were created using the same method as in the public safety scenario. We then collected 13,110 people's profiles using the Twitter Search API and the queries "bird flu", "H5N1" and "avian influenza" in June

[3] http://dev.twitter.com/pages/streaming_api
[4] https://developers.google.com/maps/documentation/geocoding/

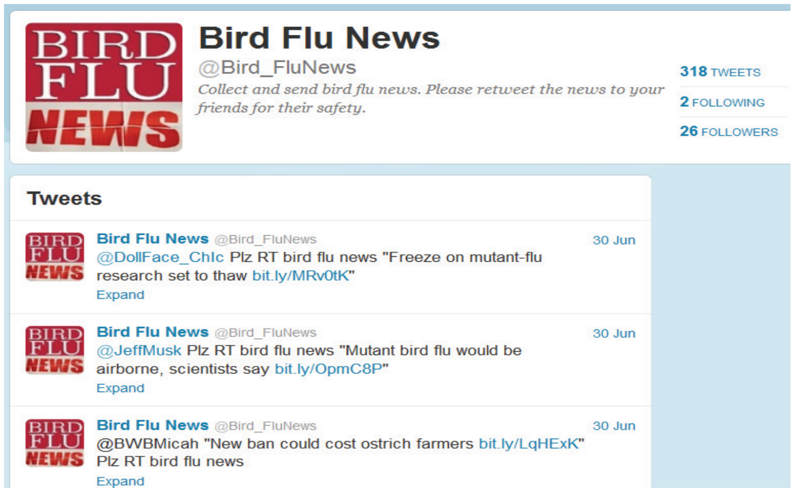


**Fig. 2. An example Twitter account created for Bird Flu data collection.**

2012. We excluded non-English tweets and randomly selected 1,859 users. A message was then automatically sent to each selected person. Here is an example message sent:

*@birdflutargetuser Plz RT bird flu news "Bird Flu viruses could evolve in nature http://bit.ly/MQBASY"*

As in the public safety study, the news articles were obtained from the news media sites. 155 of the 1,859 users (8.4%) retweeted our messages, which reached their 184,325 followers.

For both datasets, through the Twitter API we collected publicly available information of each person whom we asked to retweet. This included their profile, people they followed, followers, up to 200 of their most recently posted messages, and whether they retweeted our message (the ground truth).

## 4. FEATURE EXTRACTION

To model a person's likelihood to retweet, we have identified six categories of features, as described below.

### 4.1 Profile Features

Profile features are extracted from a user's Twitter profile and consist of: *longevity (age) of an account*, *length of screen name*, *whether the user profile has a description*, *length of the description*, and *whether the user profile has a URL*. Our hypothesis behind the use of these features is that a user with a richer profile or a longer account history may be more knowledgeable in using advanced social media features, such as retweeting. Hence, when asked, they are more likely to retweet than those who have just opened an account recently or have little information in their profile.

### 4.2 Social Network Features

We use the following features to characterize a user's social network: *number of users following (friends)*, *number of followers*, and the *ratio of number of friends to number of followers*. These features indicate the "socialness" of a person. Intuitively, the more social a person is (e.g., a good number of followers), the more likely the person may be willing to retweet. These features may also signal potential motivations for retweeting (e.g., an act of friendship and to gain followers) [Boyd et al. 2010]. However, a person (e.g., a celebrity) with an extraordinary number of followers may be unwilling to retweet per a stranger's request.

### 4.3 Personality Features

Researchers have found that word usage in one's writings, such as blogs and essays, are related to one's personality [Fast and Funder 2008; Gill et al. 2009; Pennebaker et al. 2001]. In particular, Linguistic Inquiry and Word Count (LIWC) is used to analyze text statistically and find psychologically-meaningful categories [Pennebaker et al. 2001]. Inspired by the existing work, we used the LIWC-2001 dictionary to compute one's personality features. LIWC-2001 defines 68 different categories, each of which contains several dozens to hundreds of words. For each person, we computed his/her LIWC-based personality feature in each category as follows:

Let $g$ be a LIWC category, $N_g$ denotes the number of occurrences of words in that category in one's tweets and $N$ denotes the total number of words in his/her tweets. A score for category $g$ is then: $N_g/N$.

Psychologists have developed several models of human personality. One of the more accepted models is the Big Five framework of personality traits [Costa and McCrae 1992], which proposes five key traits: *neuroticism*, *extraversion*, *openness*, *agreeableness*, and *conscientiousness*. Previous works, such as [Fast and Funder 2008; Gill et al. 2009], reveal correlations between the Big5 personality traits and the LIWC-category-based features extracted from text, such as blogs and essays. More recently, Yarkoni [Yarkoni 2010] showed that correlations exist among LIWC features and lower-level facets of Big5. Motivated by the previous findings, we used the Big5 lower-level facets as well as the Big5 traits themselves as additional personality features in our model.

| Top-20 Correlations of Friendliness with LIWC categories |
|---|
| Friends (0.23), Leisure (0.22), 1st Person Pl. (0.22), Family (0.2), Other Refs. (0.18), Up (0.18), Social Processes (0.17), Positive Emotions (0.17), Sexual (0.16), Space (0.16), Physical States (0.15), Home (0.15), Sports (0.15), Motion (0.14), Music (0.14), Inclusive (0.14), Eating (0.14), Time (0.13), Optimism (0.13), Causation (−0.13) |

**Table I. Correlations of a Big5 Facet with LIWC Categories**

All previous works use the results of personality tests taken by their participants to determine the values of the Big5 features. However, their approach requires that users take a personality test, which is not practical in our situation. To derive personality scores for each of the Big5 dimensions and their lower-level facets, we use an alternative approach. We use the coefficients of correlation between Big5 lower-level facets and LIWC categories found by Yarkoni to compute those facet-level feature values. For example, Table I shows example correlations for a Big5 facet feature. To derive a feature value for a lower-level facet, we use a linear combination of LIWC categories (for which correlation was found statistically significant by Yarkoni), where correlation coefficients are used as weights. Yarkoni also reports correlation values between LIWC category features and Big5 traits. We use such correlations as weights for deriving Big5-feature values from LIWC-category-level features.

Overall, we computed 103 personality features from one's tweets: 68 LIWC features (e.g., word categories such as "sadness"), and 5 Big5 dimensions (e.g., *agreeableness* and *conscientiousness*) with their 30 sub-dimensions. These features may signal potential motivations for retweeting (e.g., an act of altruism and to gain followers) [Boyd et al. 2010].

### 4.4 Activity Features

This feature category captures people's social activities. Similar to the reasons stated earlier, our hypothesis is that the more active people are, the more likely they would retweet when asked by a stranger. Moreover, new Twitter users or those who rarely tweet may not be familiar with the retweeting feature and be less likely to reweet. To evaluate this hypothesis, we use the following features:

- *Number of status messages*
- *Number of direct mentions (e.g., @johny) per status message*
- *Number of URLs per status message*
- *Number of hashtags per status message*
- *Number of status messages per day during her entire account life (= total number of posted status messages / longevity)*
- *Number of status messages per day during last one month*
- *Number of direct mentions per day during last one month*
- *Number of URLs per day during last one month*
- *Number of hashtags per day during last one month*

These features also help us distinguish "sporadic" vs. "steady" activeness. We hypothesize that "steady" users are more dependable and are more likely to retweet when asked. For each person, we computed these features based on their 200 most recent tweets, as our experiments have shown that 200 tweets are a good representative sample for deriving one's features.

### 4.5 Past Retweeting Features

We capture retweeting behavior with these features:

- *Number of retweets per status message: $R/N$*
- *Average number of retweets per day*
- *Fraction of retweets for which original messages are posted by strangers who are not in her social network*

Here $R$ is the total number of retweets and $N$ is the total number of status messages. We hypothesize that frequent retweeters are more likely to retweet in the future. The third feature measures how often a person retweets a message originated outside of the person's social network. We hypothesize that people who have done so are more likely to retweet per a stranger's request to do so.

| Readiness Features | Computation |
|---|---|
| Tweeting Likelihood of the Day | $T_D/N$, where $T_D$ is the number of tweets sent by the user on day D and N is the total number of tweets. |
| Tweeting Likelihood of the Hour | $T_H/N$, where $T_H$ is the number of tweets sent by the user on hour H and N is the total number of tweets. |
| Tweeting Likelihood of the Day (Entropy) | Entropy of tweeting likelihood of the day ($T_D/N$) |
| Tweeting Likelihood of the Hour (Entropy) | Entropy of tweeting likelihood of the hour ($T_H/N$) |
| Tweeting Steadiness | $1/\sigma$, where $\sigma$ is the standard deviation of the elapsed time between consecutive tweets of users, computed from users' most recent K tweets (where K is set, for example, to 20). |
| Tweeting Inactivity | $T_R$ - $T_L$, where $T_R$ is the time the request was sent and $T_L$ is the time the user last tweeted. |

**Table II. Readiness Features and their Computations**

### 4.6 Readiness Features

Even if a person is willing to retweet per a request, he may not be ready to do so at the time of the request due to various reasons, such as being busy or not being connected to the Internet. Since such a context could be quite diverse, it is difficult to model one's readiness precisely. We thus use the following features (listed in Table II) to approximate readiness based on one's previous activity:

- *Tweeting Likelihood of the Day*
- *Tweeting Likelihood of the Hour*
- *Tweeting Likelihood of the Day (Entropy)*
- *Tweeting Likelihood of the Hour (Entropy)*
- *Tweeting Steadiness*
- *Tweeting Inactivity*

The first two features are computed as the ratio of the number of tweets sent by the person on a given day/hour and the total number of tweets. The third and fourth features measure entropy of tweeting likelihood of the day and the hour, respectively [Shannon 1948]. Below is a person's ($u$) entropy of tweeting likelihood of the hour $P(x_1)$, $P(x_2)$, $P(x_3)$ ... $P(x_n)$:

$$Entropy(u) = -\sum_{i=1}^{n} P(x_i)\log P(x_i)$$

In the above equation, $n$ is 24 to estimate the daily likelihood to tweet. The tweeting steadiness feature is computed as 1/σ, where σ is the standard deviation of the elapsed time between consecutive tweets, computed from the most recent $K$ tweets (where $K$ is set to 20). The tweeting inactivity feature is the difference between the time when a retweeting request is sent and the time when user last tweeted.

Our rationale of choosing this set of features is two-fold. First, these features are good indicators of one's readiness from a particular aspect. For example, the value of *Tweeting Inactivity* may hint at one's availability, as a larger value may indicate either that the person is busy and hence uninterruptible, or that s/he is out of reach. Second, these features are easy and fast to compute based on one's past tweeting activity instead of the tweet content.

## 5. PREDICTING RETWEETERS

Based on the features described above, we train a model to predict a user's likelihood to be a retweeter.

**Training and Testing Sets.** First we randomly split each dataset (public safety and bird flu) into training (containing 2/3 data) and testing sets (containing 1/3 data). The two sets were stratified, and contained the same ratio of retweeters and non-retweeters. Finally for public safety, the training set had 35 retweeters and 1,233 non-retweeters; and the testing set had 17 retweeters and 617 non-retweeters. For bird flu, the training set had 103 retweeters and 1136 non-retweeters; the test data had 52 retweeters and 568 non-retweeters. For each person in the sets, we computed all the features described previously.

**Predictive Models.** We compared the performance of five popular models: Random Forest, Naïve Bayes, Logistic Regression, SMO (SVM), and AdaboostM1 (with random forest as the base learner). We used WEKA [Hall et al. 2009] implementation of these algorithms and trained these models to predict the probability of a person to retweet and classify a person as a retweeter or non-retweeter. If a person's retweeting probability is greater than 0.5, the person will be classified as a retweeter.

**Handling Class Imbalance**. Both our datasets have an imbalanced class distribution: only 52 out of 1,902 users (2.8%) in the public safety dataset and 155 out of 1,859 users (8.4%) in the bird flu dataset were retweeters. Imbalanced class distribution in a training set hinders the learning of representative sample instances, especially the minority class instances, and prevents a model from correctly predicting an instance label in a testing set. The class imbalance problem has appeared in a large number of domains, such as medical diagnosis and fraud detection. There are several approaches to the problem, including over-sampling minority class instances, under-sampling majority class instances, and adjusting the weights of instances. Currently, we used both over-sampling and weighting approaches to our class imbalance problem. For over-sampling, we used the SMOTE [Chawla et al. 2002] algorithm. For weighting, we used a cost-sensitive approach of adding more weight to the minority class instances [Liu and Zhou 2006].

| Feature Group | Public Safety Features |
|---|---|
| Profile | the longevity of the account |
| Social-network | \|following\| ratio of number of friends to number of followers |
| Activity | **\|URLs\| per day**<br>**\|direct mentions\| per day**<br>**\|hashtags\| per day**<br>\|status messages\|<br>\|status messages\| per day during entire account life<br>\|status messages\| per day during last one month |
| Past Retweeting | **\|retweets\| per status message**<br>**\|retweets\| per day** |
| Readiness | Tweeting Likelihood of the Day<br>Tweeting Likelihood of the Day (Entropy) |
| Personality | 7 LIWC features: **Inclusive**, Achievement, Humans, Time, Sadness, Articles, Nonfluencies<br>1 Facet feature: Modesty |

**Table III. 21 Features Selected by $\chi^2$ in Public Safety Dataset**

**Feature Analysis.** To improve the performance of our models, we analyzed the significance of our features using the training set. We computed the $\chi^2$ value for each feature to determine its discriminative power [Yang and Pedersen 1997], and eliminated the features that do not contribute significantly to the result. Our analyses found 21 and 46 significant features for the two data sets, respectively (Tables III and IV). Note that these features consistently had positive discriminative power under 3 fold cross-validation setting.

Several feature groups have more significant power distinguishing between retweeters and non-retweeters: *activity*, *personality*, *readiness*, and *past retweeting*. Although our two datasets are quite different, we found six significant features common to both sets (bolded in Tables III and IV). This suggests that it is possible to build *domain-independent* models to predict retweeters.

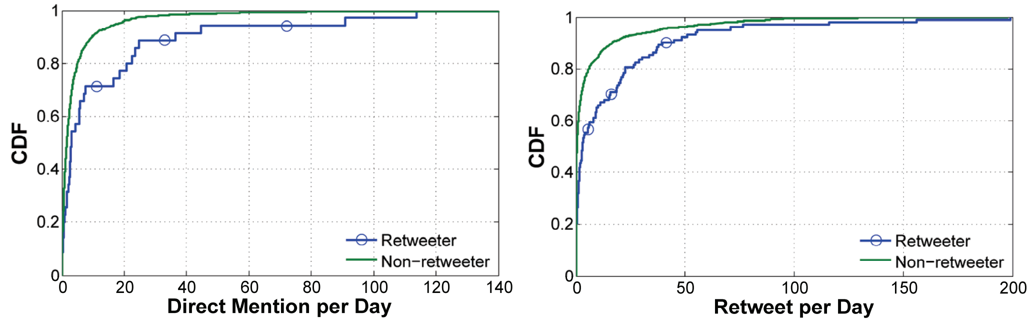


**Figure 3. CDF Plots of "Direct Mention per Day" Feature in Public Safety Training Set (left figure) and "Retweet per Day" Feature in Bird Flu Training Set (right figure)**

Next, we have analyzed how feature values of the significant features are different between retweeters and non-retweeters. Because of the limited space, we only show the cumulative distribution function (CDF) plots of two features which were commonly significant across both datasets. Figure 3 shows CDFs of |direct mention| per day feature in public safety training set and |retweets| per day in bird flu training set. We observe that retweeters have posted more number of direct mention per day and more retweet messages per day than non-retweeters. Overall, our analysis suggests that retweeters are more advanced Twitter users, since they use advanced features more frequently (e.g., inclusion of URLs and hashtags in their tweets).

| Feature Group | Bird Flu Features |
|---|---|
| Profile | the length of description<br>has description in profile |
| Activity | **\|URLs\| per day**<br>**\|direct mentions\| per day**<br>**\|hashtags\| per day**<br>\|URLs\| per status message<br>\|direct mentions\| per status message<br>\|hashtags\| per status message |
| Past Retweeting | **\|retweets\| per status message**<br>**\|retweets\| per day**<br>\|URLs\| per retweet message |
| Readiness | Tweeting Likelihood of the Hour (Entropy) |
| Personality | 34 LIWC features: **Inclusive**, Total Pronouns, 1st Person Plural, 2nd Person, 3rd Person, Social Processes, Positive Emotions, Numbers, Other References, Occupation, Affect, School, Anxiety, Hearing, Certainty, Sensory Processes, Death, Body States, Positive Feelings, Leisure, Optimism, Negation, Physical States, Communication<br>8 Facet features: Liberalism, Assertiveness, Achievement Striving, Self-Discipline, Gregariousness, Cheerfulness, Activity Level, Intellect<br>2 Big5 features: Conscientiousness, Openness |

**Table IV. 46 Features Selected by $\chi^2$ in Bird Flu Dataset**

### 5.1 Incorporating Time Constraints

While our predictive models compute a person's likelihood to retweet upon request, it does not predict when that person will retweet. Some situations may require important messages to be spread quickly, such as emergency alerts and SOS messages, so we also explore how to predict when a person will act on the retweeting request. To do this, we examine the person's previous temporal behavior and use this information for prediction.

In the simplest case, our model estimates the wait time for a person to respond to a retweeting request. We further assume that retweeting events follow a poisson process during which each retweeting occurs continuously and independently at a constant average rate. We thus use an exponential distribution model to estimate a user's retweeting wait time with a probability. The cumulative distribution function (CDF) of an exponential distribution is:

$$f(x;\lambda) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0 & , \ x < 0. \end{cases}$$

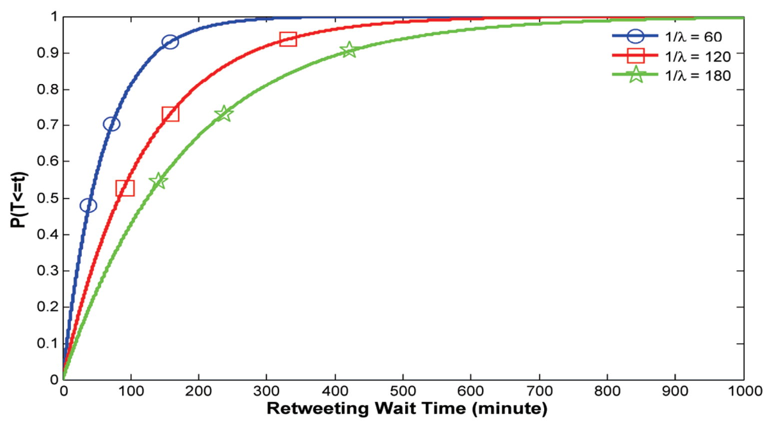


**Fig. 4. Three examples of the exponential distribution.**

The distribution is on the interval from zero to infinite. We measure $\frac{1}{\lambda}$ which is the average wait time for a user based on prior retweeting wait time. For a user's specific retweeting wait time $t$, our model can predict the probability of the user's next retweeting $P(t)$ within that wait time. Figure 4 shows our model with three examples. The green line with stars indicates that a person's average wait time is 180 minutes based on past retweeting behavior. The retweeting probability within 200 minutes is larger than 0.6. The lower a person's average retweeting wait time $t$ is, the higher probability of her retweeting is within time $t$.

**Algorithm 1** Retweeter Identification under a Time Constraint

Given a user set $U$, a time constraint $t$, cut-off probability $c$,
**for** $u \in U$ **do**
  **if** *classify(u)* = "retweeter" **then**
    **if** *CumulativeProbOfWaitTime(u, t)* >= $c$ **then**
      ask $u$ to retweet a message
    **end if**
  **end if**
**end for**

In practice, given a specific time constraint $t$, we select a *cut-off probability c* that is then used to select people whose probability of retweeting within time $t$ is greater than or equal to $c$. For example, with the cut-off probability of 0.7, our model will select only those who have at least 70% chance to retweet within the given time constraint. Incorporating the time estimation with our prediction models, we contact only people who are likely to retweet and whose cumulative probability of the retweeting wait time is greater than or equal to the *cut-off probability c* as described in Algorithm 1.

### 5.2 Incorporating Benefit and Cost

We have also explored the trade-offs between the cost of contacting a user and the benefit of a re-tweet. We assume the benefit is the number of people who are directly exposed to the message as a result of the re-tweets, which is the total number of followers of the retweeter. Using this assumption, if our system contacts $N$ users and $K$ retweet, the total benefit is then the sum of all followers of the $K$ users. Assuming a unit cost per contact, the total cost is then $N$. We normalize the total benefit by total cost to compute *unit-info-reach-per-person*:

$$\textit{unit-info-reach-per-person} = \frac{\sum_{1}^{K} \textit{followers}\,(i)}{N}$$

To address the case that the same person follows multiple retweeters, we count just the number of *distinct* followers for each retweeter.

Note that measuring effectiveness of a re-tweet in multiple hops (i.e., the followers of the followers to be as distinct as possible, or quantifying the influence of a node in a social network based on a recursive formulation that further investigates the influence of the followers) can be an alternative numerator of the above evaluation metrics. But, It requires collecting each follower's follower list, and even collecting additional information of followers in multiple hops. Since Twitter has changed their API limits, this approach takes a long time and may be unrealistic. Instead, we used the unit-info-reach-per-person to measure how many users were directly exposed by each re-tweet in the first hop.

### 5.3 Experiments

We designed and conducted an extensive set of experiments to measure the performance of various prediction models. We also compared the effectiveness of our approach with two base lines in various conditions.

#### *5.3.1 Evaluating Retweeter Prediction*

To evaluate the performance of our prediction models, we used only the significant features found by our feature analysis (Tables III and IV) in our experiments.

**Accuracy Metrics.** We use three metrics to assess prediction accuracy: Area under the ROC Curve (AUC), F1, and F1 of the retweeter class. We use AUC as our primary performance measure, since a higher AUC means that a model is good at correctly predicting both class instances regardless of class imbalance [Fawcett 2006]. We report an overall F1 score as a reference measure, and F1 of the retweeter class on the performance of predicting minority class instances.

| Classifier | AUC | F1 | F1 of Retweeter |
|---|---|---|---|
| Basic | | | |
| Random Forest | 0.638 | 0.958 | 0 |
| Naïve Bayes | 0.619 | 0.939 | 0.172 |
| Logistic | 0.640 | 0.958 | 0 |
| SMO | 0.500 | 0.96 | 0 |
| AdaBoostM1 | 0.548 | 0.962 | 0.1 |
| SMOTE | | | |
| Random Forest | 0.606 | 0.916 | 0.119 |
| Naïve Bayes | 0.637 | 0.923 | 0.132 |
| Logistic | 0.664 | 0.833 | 0.091 |
| SMO | 0.626 | 0.813 | 0.091 |
| AdaBoostM1 | 0.633 | 0.933 | 0.129 |
| Cost-Sensitive (Weighting) | | | |
| Random Forest | **0.692** | 0.954 | 0.125 |
| Naïve Bayes | 0.619 | 0.93 | 0.147 |
| Logistic | 0.623 | 0.938 | 0.042 |
| SMO | 0.633 | 0.892 | 0.123 |
| AdaBoostM1 | 0.678 | 0.956 | 0.133 |

**Table V. Prediction Accuracy (Public Safety)**

**Settings.** We ran all five prediction models under three settings: basic, SMOTE, and cost-sensitive. The *basic* setting did not handle class imbalance. SMOTE was an over-sampling approach in which we over-sampled the minority class instances in the training set such that there was an equal number of majority and minority class instances. Under the *cost-sensitive* setting, we used a weighting scheme that weighted the minority class instances higher than the majority class instances. In our experiments, we tried five different weight ratios from 10:1 through 50:1 at intervals of 10. With five prediction models under three settings, we ran a total of 35 experiments: 5 in the basic setting, 5 in the SMOTE setting, and 25 using the cost-sensitive setting (5 models by 5 weight ratios).

**Prediction Results.** Table V shows the results for the public safety dataset. Overall, the cost-sensitive setting (weighting) yielded better performance than SMOTE for both AUC and F1 of the retweeter class. Both random forest and AdaBoostM1 performed particularly well under the cost-sensitive setting. We found the similar results using the bird flu dataset (Table VI). The class imbalance problem can be observed in the poor results under the basic setting. For example, SMO completely failed to predict retweeter instances (F1 of retweeter is 0). Although both SMOTE and the cost-sensitive settings outperformed the basic one, we did not observe any clear advantage of one over the other. Note that we also ran the same experiments under 3 fold cross-validation setting, we got consistent results (i.e., SMOTE and the cost-sensitive settings outperformed the basic one).

In summary, we have found prediction configurations that produced good results by the measures of AUC and F1. Since Random Forest in the cost-sensitive setting performed the best, we used it in the rest of our experiments.

| Classifier | AUC | F1 | F1 of Retweeter |
|---|---|---|---|
| Basic | | | |
| Random Forest | 0.707 | 0.877 | 0.066 |
| Naïve Bayes | 0.670 | 0.834 | 0.222 |
| Logistic | 0.751 | 0.878 | 0.067 |
| SMO | 0.500 | 0.876 | 0 |
| AdaBoostM1 | 0.627 | 0.878 | 0.067 |
| SMOTE | | | |
| Random Forest | 0.707 | 0.819 | 0.236 |
| Naïve Bayes | 0.679 | 0.724 | 0.231 |
| Logistic | 0.76 | 0.733 | 0.258 |
| SMO | 0.729 | 0.712 | 0.278 |
| AdaBoostM1 | 0.709 | 0.837 | 0.292 |
| Cost-Sensitive (Weighting, showing the best results in each model) | | | |
| Random Forest | **0.785** | 0.815 | 0.296 |
| Naïve Bayes | 0.670 | 0.767 | 0.24 |
| Logistic | 0.735 | 0.742 | 0.243 |
| SMO | 0.676 | 0.738 | 0.256 |
| AdaBoostM1 | 0.669 | 0.87 | 0.031 |

**Table VI. Prediction accuracy (Bird Flu).**

### *5.3.2 Comparison with Two Baselines*

To validate how well our prediction approach helps improve retweeting rate in practice, we compared the retweeting rates produced by our approach with those of two baselines: random people contact and popular people contact.

The *random people contact* approach randomly selects and asks a sub-set of qualified candidates on Twitter (e.g., people living in San Francisco or tweeted about bird flu) to retweet a message. This is precisely the approach that we used during our data collection to obtain the retweeting rates for both data sets. The *popular people contact* approach first sorts candidates in our testing set by their follower count in the descending order. It then selects and contacts "popular" candidates whose follower count is greater than a threshold. In our experiment, we chose 100 as the threshold since a recent study reported that more than 87% of Twitter users have less than 100 followers[5]. We also considered other threshold values (e.g., 50, 500, 1000) and found that their retweeting rates were comparable.

| Approach | Retweeting Rate in Testing Set | |
|---|---|---|
| | Public Safety | Bird flu |
| Random People Contact | 2.6% | 8.3% |
| Popular People Contact | 3.1% | 8.5% |
| Our Prediction Approach | **13.3%** | **19.7%** |

**Table VII. Comparison of retweeting rates.**

Table VII shows the comparison of retweeting rates produced by our approach against the two base lines. Overall, our approach produced a significantly higher

[5] http://www.beevolve.com/twitter-statistics/

retweeting rate than both baselines. Specifically, ours increases the average retweeting rate of two baselines by 375% (13.3% vs. 2.8%) in the public safety domain, and by 135% (19.7% vs. 8.4%) in the bird flu scenario.

**Adding Wait Time Constraint.** We also tested our wait-time model that predicts when a person would retweet after receiving a request. We compared the retweeting rate obtained using our approach with the wait-time model with that of three settings: (a) random user contact, (b) popular user contact, and (c) our approach without the use of the wait time model. In this experiment, the retweeting rate was the ratio of the people who retweeted our messages *within* the allotted time and the total number of people whom we contacted. In other words, if a person retweeted a requested message after the allotted time (e.g., 24 hours), s/he would be considered a non-retweeter as s/he did not meet the time constraint.

| **Approach** | **Average Retweeting Rate in Testing Set under Time Constraints** | |
|---|---|---|
| | Public Safety | Bird flu |
| Random People Contact | 2.2% | 6.5% |
| Popular People Contact | 2.7% | 6.4% |
| Our Prediction Approach | **13.3%** | **13.6%** |
| Our Prediction Approach + Wait-Time Model | **19.3%** | **14.7%** |

**Table VIII. Comparison of retweeting rates with time constraints.**

In our approach with the wait-time model, we set the cut-off probability at 0.7. As described previously, we first selected a subset of people who were predicted as retweeters and then eliminated those whose estimated probability to retweet within the given time window was smaller than the cut-off probability. We experimented with different time windows, such as 6, 12, 18 or 24 hours. Table VIII shows our experimental results with the averaged retweeting rates obtained for both of our data sets. Overall, our approach with the wait-time model outperformed the other three approaches in both data sets, achieving a 19.3% and 14.7% retweeting rate, respectively. Specifically, our model with wait time constraint increases the average retweeting rate of two baselines by 680% (19.3% vs. 2.45%) in the public safety domain, and by 130% (14.7% vs. 6.45%) in the bird flu scenario. This is also an improvement of 45% (19.3% vs. 13.3%) in the public safety domain and 8% (14.7% vs. 13.6%) in the bird flu domain over our own algorithm when wait time model was not used. In summary, the *combined approach* of using our prediction model and wait-time estimation further improved retweeting rates.

| **Approach** | **Unit-Info-Reach-Per-Person** | |
|---|---|---|
| | Public Safety | Bird flu |
| Random People Contact | 6 | 85 |
| Popular People Contact | 11 | 116 |
| Our Prediction Approach | **106** | **135** |
| Our Prediction Approach + Wait-Time Model | **153** | **155** |

**Table IX. Comparison of information reach.**

**Effects of Benefit and Cost.** As described previously, another method of evaluating the performance of our work is via a benefit-cost analysis using the notion of information reach. We compared the results obtained during data collection with the results of our best prediction results on the testing set. Table IX shows the comparison of random user contact, popular user contact, and our approach without or with the wait-time model. The results show that our approach with/without the wait-time model achieved higher unit-info-reach per person than the two baselines. In particular, our approach with the wait-time model increased the average information reach of two baselines by 1,700% (153 vs. 8.5 = avg (6, 11)) in public safety and 54% (155 vs. 100.5 = avg (85, 116)) in bird flu case, respectively.

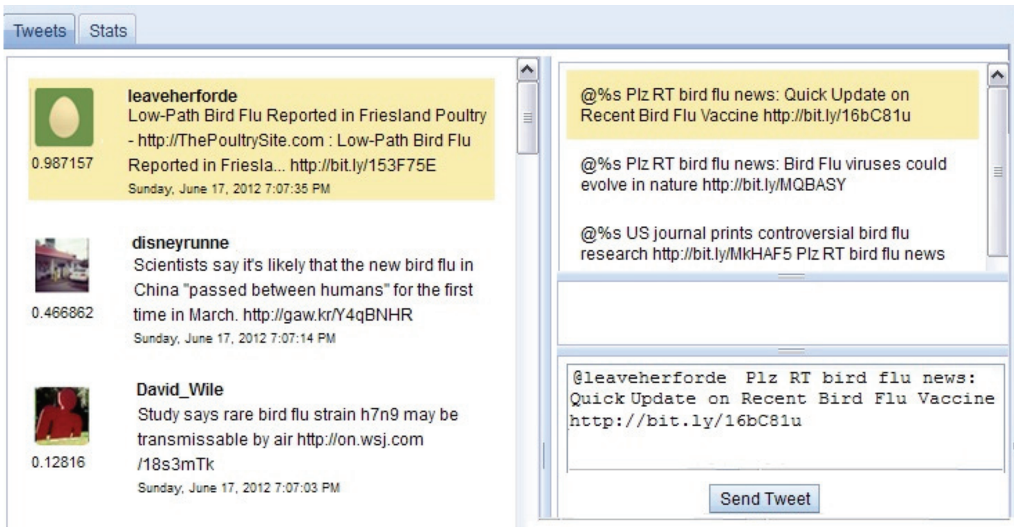


**Fig. 5. The interface of our retweeter recommendation system: (a) left panel: system-recommended candidates, and (b) right panel: a user can edit and compose a retweeting request.**

## 6. REAL-TIME RETWEETER RECOMMENDATION

As mentioned earlier, our goal is to automatically identify and engage the right strangers at the right time on social media to help spread intended messages within a given time window. We thus have developed an interactive recommender system that uses our prediction model and the wait-time estimation model in *real time* to recommend the right candidates to whom retweeting requests will be sent. Figure 5 shows the interface of our system. Our system monitors the Twitter live stream and identifies a set of candidates who have posted content relevant to the topic of a retweet request (e.g., “bird flu” alerts). Such content filtering can be done by using the approaches detailed in [Chen et al. 2013]. Based on the identified candidates, our system uses the prediction model to compute the candidates’ likelihood of retweeting and their probability of retweeting within the given time window $t$. It then recommends the top-$N$ ranked candidates whose probability of retweeting within $t$ is also greater than or equal to the cut-off probability $c$ (Figure 5a). A user (e.g., an emergency worker) of our system can interactively examine and select the recommended candidates, and control the engagement process, including editing and sending the retweeting request (Figure 5b).

## 7. MAXIMIZING RETWEETING RATE

In the previous sections, we have described classifiers we trained to model the likelihood to retweet. Such classifiers can classify a user as either retweeter or non-retweeter. Based on the classification output, our system can contact people who are classified as retweeters. However, depending on the applications such people selection is often associated with certain objectives, such as maximizing the retweet rate.

Since a classifier is not always perfect (it may not be 100% accurate), selecting people using classification output may not fulfill such objective. For example, a classifier may predict user A as a retweeter with 60% retweeting probability, but she may not actually retweet a message and thus a non-retweeter. In such scenarios, where classification result may be inaccurate, selecting people based on classification output may not fulfill our objective.

Selecting people based on their probability of retweet instead of merely classification result is another possibility. For example, the people selection approach could select top-$K$ percent users based on their probability of retweet or it could select users for whom the probability of retweet is above a certain threshold (e.g., more than 60%). However, in such approach, the value of $K$ or threshold probability has to be manually selected which is adhoc and not generalizable. In addition, due to the inherent imperfections (prediction errors) in any classification models, in reality, the predicted top-$K$ people may not necessarily be the best choices in terms of maximizing the overall retweet rate. To maximize the overall retweet rate for a given set of available people, we thus have implemented the following algorithm.

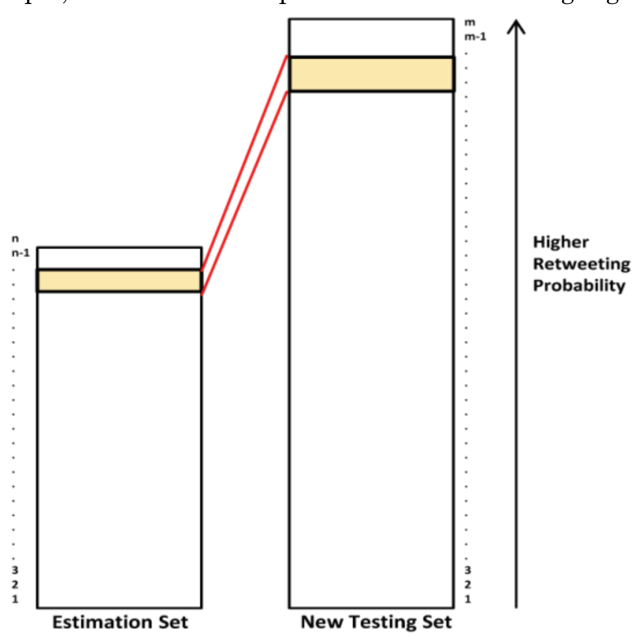


**Fig. 6. Optimal Interval Returning Maximal Retweet Rate in the Estimation Set and New Testing Set (Colored Rectangle is the Optimal Interval)**

In this section, we propose a *subset selection* based algorithm which *automatically* selects a subset of people from a set of available people using probabilities predicted by the classifier and an estimation set from which the best interval for people selection is estimated. Our algorithm is based on the assumption that the retweeting probability of a user predicted by the classifier has high correlation with the user's actual retweeting behavior. Thus a set of users may be sorted according to their retweeting probabilities, and an optimal interval from the sorted set corresponding to the maximum retweeting rate can be estimated. Such an optimal interval is applied to any arbitrary set of available people to select a subset of them to contact. In the experimental section, we demonstrate that our algorithm is effective in maximizing the retweet rate.

Specifically, our approach for maximizing the retweet rate is shown in Figure 6: first, we divide the testing set into an estimation set (1/3 data of the original testing set) and new testing set (2/3 data of the original testing set). Next, we compute retweeting probability of each user in the estimation set using our trained classifier. We sort the estimation set using the computed probabilities.

Let $\{\boldsymbol{p}_1,\ldots,\boldsymbol{p}_n\}$ denotes the sorted list of people in the estimation set. Next, we find an interval $[i, ,j]$ ($1\leq i < j \leq n$) from this set, where the corresponding interval subset $\{\boldsymbol{p}_i,\ldots,\boldsymbol{p}_j\}$ has a maximal retweeting rate among all interval subsets. We have also tried a slight variant of this approach which searches for a restricted choice of intervals; only those that extend to the top, i.e., of the form $[i, n]$. However, that produced sub optimal result. The best subinterval $[i_r, j_r]$ in the estimation set defines a corresponding subinterval $[i_s, j_s]$ in the new testing set, based on percentiles. That is, if $m$ is the cardinality of the new testing set, then $i_s = [(i_r \cdot m)/n]$ and $j_s = [(j_r \cdot m)/n]$. For example, the estimation set consists of 100 users ($n$=100), new testing set consists of 200 users ($m$=200). The goal is to find the optimal interval returning maximal retweet rate. Our algorithm measures retweet of every $k$% interval in the estimation set, starting from top ($n$ in the figure) and moving down 1% in each iteration. Say $k$ is 10%. The first interval in the estimation set is [100, 91], the next one is [99, 90], and so on. The algorithm finds the interval returning the highest retweeting rate in the estimation set. Then, it selects users in the same interval in the new testing set. Say, optimal interval is [98, 89] in the estimation set. We select [196, 177] in the new testing set (again, $m$=200) and contact the users in the interval. The example shows the optimal interval with 10% interval size. However, we can apply this method for other interval sizes (say from 1% to 30%) in order to find the best interval size returning the optimal interval with maximal retweet rate.

We can also incorporate additional constraints in our optimal interval selection. For example, we can specify the exact size of the interval, minimum, or maximum size of the interval as constraints. For example, if a minimum size of the interval is specified, our method will ignore intervals that are smaller than the specified minimum.

### 7.1 Experiments

We have experimentally found some reasonable classification settings for classifying a user as retweeter or non-retweeter. As described in the previous sections, we use the probabilities computed by the trained classifier (cost-sensitive with Random Forest) with our algorithm for maximizing the retweeting rate. We keep the training set (containing 2/3 data), and divide the original testing set into an estimation set (containing 1/3 of the original testing set) and new testing set (containing 2/3 of the original testing set).

To measure the performance of our algorithm, we computed retweeting rate by increasing an interval size from 1% to 5% by 1% within top 15% percentiles. Tables X and XI present the optimal retweeting rate of different interval size in public safety and bird flu datasets. The average retweeting rate of our algorithm in new testing set of public safety and bird flu datasets are 16% and 54%, respectively.

Next, we compare the average retweeting rates of our classification and subset selection based prediction approach with two baselines (random people contact and popular people contact) and our classification based prediction approach. Table XII shows the comparison of retweeting rates in each of the approaches. The baselines resulted in the lowest retweeting rate in both datasets. Our prediction approach (classification + subset selection) produced a significantly higher retweeting rate than both baselines. Specifically, ours increase the average retweeting rate of two baselines by 460% (16% vs. 2.8%) in the public safety domain, and by 540% (54% vs. 8.4%) in the bird flu scenario. This is also an improvement of 45% (16% vs. 11.1%) in the public safety domain and 170% (54% vs. 20%) in the bird flu domain over our own algorithm (classification). In summary, our classification and subset selection based prediction approach improved retweeting rates.

| Interval Size | Optimal Interval | Retweeting Rate in Estimation Set | Retweeting Rate in New Testing Set |
|---|---|---|---|
| 1% | top 0 ~ 1% | 50% | 25% |
| 2% | top 0 ~ 2% | 25% | 13% |
| 3% | top 0 ~ 3% | 17% | 8% |
| 4% | top 0 ~ 4% | 25% | 19% |
| 5% | top 0 ~ 5% | 20% | 14% |
| Average | | 27% | **16%** |

**Table X. Variations of Optimal Interval, Estimation and New Testing Set Retweeting Rate with Increasing Interval Size (Public Safety Dataset)**

| Interval Size | Optimal Interval | Retweeting Rate in Estimation Set | Retweeting Rate in New Testing Set |
|---|---|---|---|
| 1% | top 8 ~ 9% | 100% | 100% |
| 2% | top 7 ~ 9% | 75% | 56% |
| 3% | top 6 ~ 9% | 50% | 39% |
| 4% | top 5 ~ 9% | 38% | 41% |
| 5% | top 4 ~ 9% | 30% | 33% |
| Average | | 58% | **54%** |

**Table XI. Variations of Optimal Interval, Estimation and New Testing Set Retweeting Rate with Increasing Interval Size (Bird Flu Dataset)**

**Adding Wait Time Constraint.** In order to predict when a user retweets information after she received a message, we use an exponential distribution model as we described in Section 5.1.

Like our previous experiments, retweeting rate obtained using our prediction approach (classification + subset selection) with a time-constraint is compared against the baselines and our classification approach. For the baselines and classification approach, retweeting rate is computed as the ratio of the users who actually retweeted our messages within that time and the total number of users we actually contacted.

| Approach | Retweeting Rate in New Testing Set | |
|---|---|---|
| | Public Safety | Bird flu |
| Random People Contact | 2.6% | 8.4% |
| Popular People Contact | 3.1% | 8.5% |
| Our Prediction Approach (Classification) | 11.1% | 20% |
| Our Prediction Approach (Classification + Subset Selection) | **16%** | **54%** |

**Table XII. Comparison of Retweeting Rates**

For our prediction approach (classification + subset selection) with wait time model, we set up the cut-off probability in the exponential distribution model to 0.7. We eliminated the users for whom estimated wait time for next retweeting was smaller than this cut-off probability from both of our estimation and new testing set.

We experimented with different time windows, such as 6, 12, 18 or 24 hours. Table XIII shows our experimental results with the averaged retweeting rates obtained for both of our data sets. Overall, our prediction approach (classification +

subset selection) with wait time model outperformed the other three approaches in both data sets, achieving a 19.3% and 19.8% retweeting rate, respectively. Specifically, our model with wait time constraint increases the average retweeting rate of two baselines by 640% (19.3% vs. 2.6%) in the public safety domain, and by 230% (19.8% vs. 5.9%) in the bird flu scenario. This is also an improvement of 74% (19.3% vs. 11.1%) in the public safety domain and 61% (19.8% vs. 12.3%) in the bird flu domain over our own prediction approach (classification) when wait time model was not used. In summary, the *combined approach* of using our prediction model (classification + subset selection) and wait-time estimation further improved retweeting rates.

| **Approach** | **Average Retweeting Rate in New Testing Set under Time Constraints** | |
|---|---|---|
| | Public Safety | Bird flu |
| Random People Contact | 2.2% | 6.2% |
| Popular People Contact | 2.9% | 5.6% |
| Our Prediction Approach (Classification) | 11.1% | 12.3% |
| Our Prediction Approach (Classification + Subset Selection) + Wait time model | **19.3%** | **19.8%** |

**Table XIII. Comparison of Retweeting Rates under Time Constraints**

**Effect of Benefit and Cost.** As described in Section 5.2, we have investigated the use of benefit of retweeting and the cost of contacting a user using the notion of information reach. Thus, we compared information reach using our prediction approach (classification + subset selection), our classification approach and the baselines. Table XIV shows the comparisons of random user contact, popular user contact, classification approach and our prediction approach (classification + subset selection). The results show that our prediction approach (classification + subset selection) achieved higher unit-info-reach per person than the two baselines. In particular, our approach with the subset selection increased the average information reach of two baselines by 1,300% (162 vs. 11.5 = avg (9,14)) in public safety and 110% (233 vs. 109 = avg (91,127)) in bird flu case, respectively. This is also an improvement of 7% (162 vs. 152) in the public safety domain and 110% (233 vs. 111) in the bird flu domain over our own prediction approach (classification).

| **Approach** | **Unit-Info-Reach per User in New Testing Set** | |
|---|---|---|
| | Public Safety | Bird flu |
| Random People Contact | 9 | 91 |
| Popular People Contact | 14 | 127 |
| Our Prediction Approach (Classification) | 152 | 111 |
| Our Prediction Approach (Classification + Subset Selection) | **162** | **233** |

**Table XIV. Comparison of Information Reach**

To summarize, as long as we have increased retweeting rate or found the best range giving us the highest retweeting rate, we can get the higher unit-info-reach per user than the baselines and the classification approach.

## 8. LIVE EXPERIMENTS

To validate the effectiveness of our approaches, the classification approach and classification + subset selection approach, in a *live* setting, we used our recommender system, which was developed in Section 5, 6 and 7, to test our approaches against the

| Approach | Retweeting Rate |
|---|---|
| Random People Contact | 4% |
| Popular People Contact | 9% |
| Our Prediction Approach (Classification) | 19% |
| Our Prediction Approach (Classification + 24% Subset Selection) | 18% |
| Our Prediction Approach (Classification + 5% Subset Selection) | 38% |

**Table XV. Comparison of retweeting rates in live experiment.**

| Approach | Average Retweeting Rate |
|---|---|
| Random People Contact | 4% |
| Popular People Contact | 8.7% |
| Our Prediction Approach (Classification) | 18% |
| Our Prediction Approach (Classification) + Wait time model | 18.5% |
| Our Prediction Approach (Classification + 24% Subset Selection) + Wait time model | 21.3% |
| Our Prediction Approach (Classification + 5% Subset Selection) + Wait time model | 34.6% |

**Table XVI. Comparison of retweeting rates in live experiment (with time constraints).**

two baselines (random people contact and popular people contact). First, we randomly selected 426 candidates who had recently tweeted about "bird flu" during July 2013. We then used each approach to select 100 users among the candidates. The popular people contact and our classification approach selected the top 100 candidates based on their popularity (number of followers) rank and our prediction rank, respectively. Our classification + subset selection approach also selected the best 24% range (100 candidates) among 426 candidates sorted by our prediction rank. We selected the 24% range giving us the highest retweeting rate in the estimation set as we described in Section 7. We also selected the best 5% range to see whether the best 5% range by our classification + subset selection approach would achieve higher retweeting rate than the 24% range. If a person happened to be selected by more than one approach, we contacted the person only once to avoid overburdening the person. Overall, we contacted a total of 236 unique people. Table XV shows the comparison of retweeting rates for each approach. Our approaches outperformed two baselines in a live setting significantly. Specifically, our classification approach and classification + 24% subset selection approach increase the average retweeting rate of two baselines by more than 190% (19% vs. 6.5%) and 175% (18% vs. 6.5%), respectively. Interestingly, classification + 5% subset selection approach even increases the average retweeting rate of two baselines by more than 480% (38% vs. 6.5%). We checked the social graph of the retweeters (those who retweeted our message). They were not connected at all. Thus, our result was unlikely to be affected by their social relationship.

We also wanted to investigate the effectiveness of our approach with time constraints. Thus, we repeated the above experiment with different time windows, such as 6, 12, 18 or 24 hours as we did in Sections 5.3 and 7.1. Table XVI shows the comparison of retweeting rates for each approach. Again, our approaches with our wait time model outperformed all other three approaches. Specifically, our classification approach with the wait time model increases the average retweeting rate of two baselines by more than 190% (18.5% vs. 6.35%). Our classification + 24% subset selection approach and classification + 5% subset selection approach with the wait time model increase the average retweeting rate of two baselines by more than 235% (21.3% vs. 6.35%) and 440% (34.6% vs. 6.35%), respectively. Our approaches with wait time model outperform our classification approach when the wait time model was not used. In summary, this result confirms that our approaches consistently outperformed others in a live setting by a large margin.

## 9. DISCUSSION

Here we discuss several of observations during our investigation and the limitations of our current work.

### 9.1 Why People Retweet at a Stranger's Request

Although previous studies discuss various reasons why people retweet in general [Boyd et al. 2010; Starbird and Palen 2010], they focus on people's voluntary retweeting behavior. We were curious to find out why people retweet upon the request of a stranger. We randomly selected 50 people who retweeted per our request and asked them why they chose to retweet. 33 out of 50 replied to us. Their responses revealed several reasons why people accept our retweeting requests. One reason was the trustworthiness of the content to be spread: *"Because it contained a link to a significant report from a reputable media news source"*. Another reason is content relevance, e.g., messages about their own local area: *"Because it happened in my neighborhood*". Interestingly, several mentioned that they retweeted because the message contained valuable information and was helpful to society: *"my followers should know this or they may think this info is valuable"*. Some of other reasons, such as to spread tweets to new audience or to entertain a specific audience, were discussed by others [Boyd et al. 2010], however not mentioned in our context. In future, it would be interesting to study whether including the rationale in a retweeting request would help motivate the target strangers and affect the retweeting rate.

### 9.2 Retweeting with Modification

We have observed that some people retweeted our messages with modifications (e.g., adding hashtags to clarify the message or their own opinion to the original message):

***#publichealth*** *news: The Evolution of Bird Flu, and the Race to Keep Up http://nyti.ms/Qf6zsM @nytimesscience*

***what a shame + waste of tax $$*** *"@BayPublicSafety: @esavestheworld "Hacker created fake Sierra LaMar posting http://bit.ly/Leaojo" Plz RT"*

Such behavior suggests that the target information propagators may augment/alter the original message with additional information including their personal opinions, especially if they strongly agree/disagree with the intended information. Based on this observation, it would be interesting to investigate the additional gains and risks that a potential information propagator might bring when asked to spread the message. For example, the added hashtag (#publichealth) in the re-tweet above would help propagate the message not only to the followers but also those who follow the hashtag. On the opposite, a propagator's negative opinions may affect the spread and perception of the intended message.

### 9.3 Generalizability

We wanted to examine how well our findings can be generalized across topics. We ran an experiment where we combined the training and testing sets of public safety and bird flu. We trained prediction models on the combined training set using the significant features identified for the combined set. AUC in this experiment was 0.736, better than the original public safety result (0.692), but lower than the original bird flu result (0.785). The resulted retweeting rate was 12.5%, better than the random user contact (5.5%) and popular user contact (6%) for the combined set, but lower than the rates achieved in public safety (13.3%) and bird flu alone (19.7%). Our results suggest that it is feasible to build a domain-independent prediction model, if we have sufficient training-data from different domains. We are investigating the applicability of our models to new domains, e.g., new topics that our model is not trained on.

### 9.4 Maximizing Information Reach

In the previous section, we have presented an experiment that incorporated cost/benefit and showed that our algorithm which maximizes retweet rate achieved higher information reach than the classification approach and baselines. Another objective could be to design an algorithm which can maximize the information reach. We attempted to develop such algorithm using similar approach as our algorithm for maximizing retweet rate. In particular, we tried two variants of our heuristic when finding optimal interval from estimation set: 1) sorting each user by the product of retweeting probability and number of followers, and finding optimal range based on the unit-info-reach per user; (2) sorting each user by only retweeting probability and finding optimal range based on the unit-info-reach per user. However, in practice we discovered that these heuristics produced sub-optimal result in comparison to our original heuristic (sorting by retweeting probability and finding interval that maximizes retweet rate) for maximizing information reach. One reason why these heuristics produced sub-optimal result is both of them could be affected by users with large number of followers. Such users may have low retweet probability (making first heuristic ineffective) or may not appear uniformly in both estimation and testing set (making second heuristic ineffective). We plan to further investigate such issues, and develop solutions for maximizing information reach. Furthermore, we plan to experiment with information reach scenarios when benefit of a retweet is computed not only from followers of those who retweeted, but also from users who are indirectly exposed to the message (such as followers' followers) as a result of a retweet.

### 9.5 Optimizing Multiple Information Spreading Objectives

Currently, our work focuses on maximizing the retweeting rate in information diffusion. However, in practice, there may be multiple objectives to be satisfied, such as maximizing the expected net benefit or minimizing the reach time. We thus are investigating a model that can optimize multiple objectives at the same time. However, this is non-trivial as satisfying one objective may influence the other especially in a real world situation, where many of these objectives may be dynamically changing (e.g., the availability of retweeting candidates and the required time frame for a message to reach a certain audience).

### 10. CONCLUSIONS

In this manuscript, we have presented a feature-based prediction model that can automatically identify the right individuals at the right time on Twitter who are likely to help propagate messages per a stranger's request. We have also described a time estimation model that predicts the probability of a person to retweet the requested message within a given time window. In addition, we have developed a subset selection model to maximize the rate of retweeting and demonstrated how such model can work under different constraints. Based on these three models, we build an interactive retweeter recommender system that allows a user to identify and engage strangers on Twitter who are most likely to help spread a message. To train and test our approaches, we collected two ground-truth datasets by *actively* engaging 3,761 people on Twitter on two topics: public safety and bird flu. Through an extensive set of experiments, we found that our approaches were able to at least *double* the retweeting rates over two baselines. With our time estimation model, our approach also outperformed other approaches significantly by achieving a much higher retweeting rate within a given time window. Furthermore, our approach has achieved a higher unit-information-reach per person than the baselines. In a live setting, our approach consistently outperformed the two baselines by almost doubling their retweeting rates. Overall, our approach effectively identifies qualified candidates for retweeting a message within a given time window.

### ACKNOWLEGEMENT

Research was sponsored by the U.S. Defense Advanced Research Projects Agency (DARPA) under the Social Media in Strategic Communication (SMISC) program, Agreement Number W911NF-12-C-0028. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Defense Advanced Research Projects Agency or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

### REFERENCES

Nitin Agarwal, Huan Liu, Lei Tang, and Philip S. Yu. 2008. Identifying the Influential Bloggers in a Community. In WSDM.

Eytan Bakshy, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. 2011. Everyone's an Influencer: Quantifying Influence on Twitter. In WSDM.

Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. 2012. The Role of Social Networks in Information Diffusion. In WWW.

Danah Boyd, Scott Golder, and Gilad Lotan. 2010. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In HICSS.

Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. 2011. Limiting the Spread of Misinformation in Social Networks. In WWW.

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

Vineet Chaoji, Sayan Ranu, Rajeev Rastogi, and Rushi Bhatt. 2012. Recommendations to Boost Content Spread in Social Networks. In WWW.

Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. J. Artif. Int. Res. 16, 1 (June 2002), 321–357.

Jilin Chen, Allen Cypher, Clemens Drews, and Jeffrey Nichols. 2013. CrowdE: Filtering Tweets for Direct Customer Engagements. In ICWSM.

Kailong Chen, Tianqi Chen, Guoqing Zheng, Ou Jin, Enpeng Yao, and Yong Yu. 2012. Collaborative personalized tweet recommendation. In SIGIR.

P.T. Costa and R.R. McCrae. 1992. Revised NEO Personality Inventory (NEO PI-R) and NEP Five-factor Inventory (NEO-FFI): Professional Manual. Psychological Assessment Resources.

Lisa A Fast and David C Funder. 2008. Personality as manifest in word use: correlations with self-report,

acquaintance report, and behavior. Journal of personality and social psychology 94, 2 (2008), 334.
Tom Fawcett. 2006. An Introduction to ROC Analysis. Pattern Recogn. Lett. 27, 8 (June 2006), 861–874.
Wei Feng, and Jianyong Wang. 2013. Retweet or not?: personalized tweet re-ranking. In WSDM.
Alastair J. Gill, Scott Nowson, and Jon Oberlander. 2009. What Are They Blogging About? Personality, Topic and Motivation in Blogs. In ICWSM.
Amit Goyal, Francesco Bonchi, and Laks V.S. Lakshmanan. 2010. Learning Influence Probabilities in Social Networks. In WSDM.
Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA Data Mining Software: An Update. SIGKDD Explor. Newsl. 11, 1 (Nov. 2009), 10–18.
Tuan-Anh Hoang and Ee-Peng Lim. 2012. Virality and Susceptibility in Information Diffusions. In ICWSM.
Nathan O. Hodas, and Kristina Lerman. 2014. The simple rules of social contagion. Scientific reports 4 (2014).
Junming Huang, Xue-Qi Cheng, Hua-Wei Shen, Tao Zhou, and Xiaolong Jin. 2012. Exploring Social Influence via Posterior Effect of Word-of-mouth Recommendations. In WSDM.
Kyumin Lee, James Caverlee, and Steve Webb. 2010. Uncovering Social Spammers: Social Honeypots + Machine Learning. In SIGIR.
Kyumin Lee, Brian David Eoff, and James Caverlee. 2011. Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter.. In ICWSM.
Kyumin Lee, Jalal Mahmud, Jilin Chen, Michelle Zhou, and Jeffrey Nichols. 2014. Who Will Retweet This? Automatically Identifying and Engaging Strangers on Twitter to Spread Information. In IUI.
Xu-Ying Liu and Zhi-Hua Zhou. 2006. The Influence of Class Imbalance on Cost-Sensitive Learning: An Empirical Study. In ICDM.
Sofus A. Macskassy and Matthew Michelson. 2011. Why do People Retweet? Anti-HomophilyWins the Day!. In ICWSM.
Jalal Mahmud, Michelle X. Zhou, Nimrod Megiddo, Jeffrey Nichols, and Clemens Drews. 2013. Recommending Targeted Strangers from Whom to Solicit Information on Social Media. In IUI.
Jeffrey Nichols and Jeon-Hyung Kang. 2012. Asking Questions of Targeted Strangers on Social Networks. In CSCW.
James W. Pennebaker, Martha E. Francis, and Roger J. Booth. 2001. Linguistic Inquiry and Word Count. Lawerence Erlbaum Associates.
Daniel M. Romero, Wojciech Galuba, Sitaram Asur, and Bernardo A. Huberman. 2011. Influence and Passivity in Social Media. In ECML/PKDD.
Claude Shannon. 1948. A Mathematical Theory of Communication. Bell System Technical Journal 27 (July, October 1948), 379–423, 623–656.
Yaron Singer. 2012. How to Win Friends and Influence People, Truthfully: Influence Maximization Mechanisms for Social Networks. In WSDM.
K. Starbird and L. Palen. 2010. Pass it On?: Retweeting in Mass Emergency. International Community on Information Systems for Crisis Response and Management.
Greg Ver Steeg and Aram Galstyan. 2012. Information Transfer in Social Media. In WWW.
Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. 2010. TwitterRank: Finding Topic-sensitive Influential Twitterers. In WSDM.
Yiming Yang and Jan O. Pedersen. 1997. A Comparative Study on Feature Selection in Text Categorization. In ICML.
Tal Yarkoni. 2010. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. Journal of research in personality 44, 3 (2010), 363–373.