

TipMaster: A Knowledge Base of Authoritative Local News Sources on Social Media

Xin Shuai, Xiaomo Liu, Armineh Nourbakhsh, Sameena Shah, Tonya Custis

Research and Development
Thomson Reuters

{xin.shuai, xiaomo.liu, armineh.nourbakhsh, sameena.shah, tonya.custis}@thomsonreuters.com

Abstract

Twitter has become an important online source for real-time news dissemination. Especially, official accounts of local government and media outlets have provided newsworthy and authoritative information, revealing local trends and breaking news. In this paper, we describe *TipMaster*, an automatically constructed knowledge base of Twitter accounts that are likely to report local news, from government agencies to local media outlets. First, we implement classifiers for detecting these accounts by integrating heterogeneous information from the accounts' textual metadata, profile images, and their tweet messages. Next, we demonstrate two use cases for *TipMaster*: 1) as a platform that monitors real-time social media messages for local breaking news, and 2) as an authoritative source for verifying nascent rumors. Experimental results show that our account classification algorithms achieve both high precision and recall (around 90%). The demonstrated case studies prove that our platform is able to detect local breaking news or debunk emergent rumors faster than mainstream media sources.

Introduction

Social media has become a prominent source for spreading and consumption of information and news (Kwak et al. 2010). It often provides faster access to breaking news than traditional news agencies (Laird 2012). Nowadays, it is common practice of journalists to monitor social media, especially Twitter in identifying breaking news stories (Vis 2013). However, the large and ever-changing amount of noisy and imprecise content on Twitter makes it difficult for journalists to sift through all the irrelevant tweets to find valuable news stories, especially since most journalists rely on manually selected user accounts and keyword searches to filter for newsworthy content (Nourbakhsh et al. 2017). As one of the largest media companies in the world, Thomson Reuters invests in new social mining technologies that assist journalists in identifying local breaking news.

In this paper, we propose an automated method for identifying the accounts on Twitter that are most likely to notify journalists of local breaking news in the fastest way. We conducted a survey among Reuters journalists as to which types

of Twitter accounts are most likely to originate local breaking news. The results show that two types of local accounts are favored by most journalists. One type is *media* accounts associated with news agencies, TV channels and radio stations etc., which focus on reporting local news and events (e.g., @KTLA). The other type is *government* accounts associated with local government organizations, which mainly announce news related to social and public life within a jurisdiction (e.g., @NYPD). Generally, both media and government accounts belong to *organization* accounts, which are typically operated by an organization (or a person representing the organization). By contrast, *personal* accounts are usually operated by individuals and used for personal status updates and communications.

Figure 1 shows an example in which a local government account broke a news story ahead of traditional media. When the San Bernardino shooting happened at 10:59am PST on Dec. 2, 2015, the Fire Department of San Bernardino responded immediately and notified the public through Twitter at 11:14am PST. This was the first public piece of information about the shooting, earlier than any news media. After 5 minutes, a local news channel (KTLA) picked it up and produced the first news report about it. The tragedy was then broadcast country-wide and worldwide by Reuters and CNN, at 11:28am and 11:32am respectively.

In this paper, we present *TipMaster*, a knowledge base consisting of local government and media Twitter accounts. *TipMaster* derives its name from "tip," a word used by journalists to describe early indicators of an event. The system *tips* the journalists about breaking stories by focusing on accounts that are likely to break news at the local level.

We describe the algorithms that power *TipMaster*, and demonstrate its capability for detecting local breaking news and debunking emergent rumors. Specifically, we propose a three-step classification pipeline to integrate information from both Twitter accounts and external data resources (e.g. Census index and Facebook page). Furthermore, we demonstrate how *TipMaster* works as a real-time monitoring platform of breaking news, dynamic local trends and other related information. Our contributions include: (1) a systematic methodology consisting of information retrieval and classification to automatically distinguish government and media Twitter accounts by leveraging both account features and external data resources; (2) a news monitoring platform



Figure 1: Example of a local government account reporting local breaking news earlier than traditional media.

Table 1: Summarization of features and models used in account classification studies.

Related Work	Model Features					Learning Model
	Tweet Time	Tweet Content	User Interaction	Profile Description	Followers(ee) Network	
(Oentaryo, Low, and Lim 2015)		✓	✓			Gradient Boosting Generative Model Linear SVM K-nearest Neighbor Linear SVM Naive Bayes Linear SVM
(Yin et al. 2014)	✓	✓				
(Yan, Ma, and Yoshikawa 2013)				✓	✓	
(De Choudhury, Diakopoulos, and Naaman 2012)		✓	✓		✓	
(McCorriston, Jurgens, and Ruths 2015)		✓	✓			
(Tavares and Faisal 2013)	✓					
(Silva and Riloff 2014)		✓				

that is able to track local breaking news and may further reveal local trends and relationships between entities related to the news. To the best of our knowledge, this is the first such solution based on the Twitter platform. Experimental results and a case study show that our classification algorithms achieve both high precision and recall (around 90%). In addition, *TipMaster* is able to detect local breaking news or debunk emergent rumors faster than other mainstream media sources.

Related Work

Account Classification

Social media provides a platform not only for individuals to communicate with each other, but also for organizations to disseminate the most recent information to the public. To distinguish these two types of accounts is very useful for human behavioral studies (Tavares and Faisal 2013), as well as for targeted recommendation services (Takemura and Tajima 2016). There are several previous studies about Twitter account classification, and the main features and models used are summarized in Table 1. The key process of constructing the knowledge base in our paper is to detect government and media accounts from Twitter, which is in essence, an account classification problem. Compared to the listed previous studies, there are two main differences: First, all of them classify person vs. organization accounts, while our goal is to further distinguish local government and media accounts from already detected organization accounts. Sec-

ond, the pool of local government and media accounts will eventually be leveraged to build a knowledge base for automatic local news finding, through monitoring, mining and analyzing their real-time tweets.

Local Event Detection

A huge amount of information from social media is related to real-life events, which makes it an ideal source for real-time events detection. Most previous studies follow a similar procedure for event detection on social media: collecting real-time content associated with geo-location metadata, such as geo-tagged photos (Xia et al. 2014; Xie et al. 2013) and geo-tagged tweets (Cheng and Wicks 2014), clustering similar content into different topic groups (Ishikawa et al. 2012; Cheng and Wicks 2014) or detecting the bursts of certain features from those groups within a certain period of time (e.g. words (Hu, Farnham, and Monroy-Hernández 2013), number of photos (Xia et al. 2014)), and finally defining events based on bursty features or related content. All of the above studies have practical value, but the process of data collection, pre-processing and clustering is relatively complicated and time-consuming. In addition, the newsworthiness of the detected events is not easily verifiable. Instead, we propose to solve the problem from a different angle. By automatically curating a knowledge base of Twitter accounts (i.e. local government and media accounts) that are widely recognized as trustworthy by news professionals and providing a monitoring & analytics platform, journalists are able to

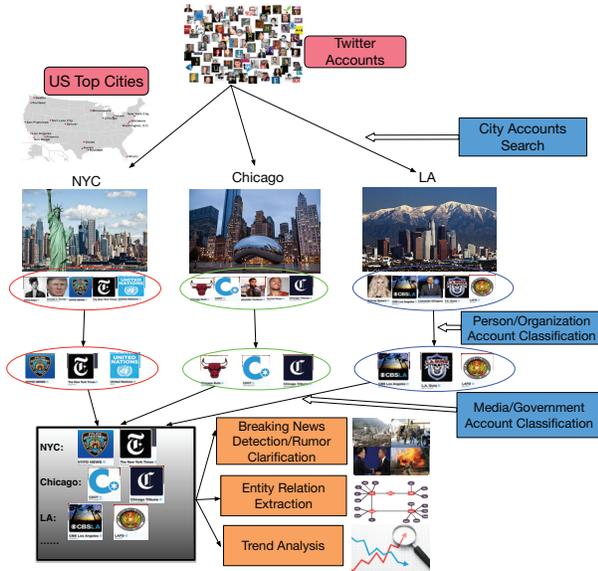


Figure 2: Pipeline of govt. & media account detection.

capture real-time local breaking news without having to process a large volume of tweets.

Problem Definition

We use A to denote all Twitter accounts. For each account $a \in A$, there are multiple pieces of metadata associated with it: Name (N_a), Location (L_a), Profile Image (I_a), Profile URL (U_a) and Profile textual description (D_a). In addition, in the account’s timeline, $S_a(n)$ represents the n most recent tweets. Given a city c , we use $A_c = \{a | L_a = c\}$ to denote all accounts associated with c and $T(a)$ to indicate the account type of a , where $T(a) \in \{O, P\}$ with O and P representing *organization* and *person* accounts. In addition, $O \in \{G, M, Ot\}$, with M , G and Ot representing *media*, *government* and other types. Moreover, we use $A_c^t = \{a | T(a) = t\}$ to represent all accounts in city c belonging to type t , $t \in \{G, M, Ot, P\}$.

We want to find a feature mapping function of the metadata set $F(N_a, I_a, U_a, D_a, S_a(n))$ and a classification function of $C(F(\cdot))$, which can correctly predict $T(a)$:

$$F(N_a, I_a, U_a, D_a, S_a(n)) \xrightarrow{C(F)} T(a) \quad (1)$$

Consequently, the classification will result in a set of local government and media accounts, i.e. $A_c^* = A_c^G \cup A_c^M$.

System Overview

Figure 2 illustrates the pipeline of our knowledge base construction process as well as the news monitoring platform. First, we retrieved Twitter accounts linked to a list of specific city names (e.g. NYC, Chicago). Second, we designed a classifier to distinguish organizational accounts from personal accounts and discard the latter. Third, we designed a topic-based classifier to select government and media accounts from all organization accounts. Finally, we built a platform to monitor local breaking news or verify rumors.

Curating the *TipMaster* Knowledge Base

We automatically construct the knowledge base of local government agencies and media outlets by collecting a dataset of potentially relevant Twitter accounts, and iteratively classifying them into corresponding subtypes.

City accounts search. The first step is to retrieve A_c from A . After determining the top 500 most populous cities in the US¹, we queried an external Twitter account search engine, Followerwonk², using multiple metadata fields of an account (e.g. location, URL, etc.). Given a city name c , we created and sent the following search query to Followerwonk: `Location:c AND URL:http AND min_followers:100 AND min_tweets:100`. We set these restrictions in the query to filter out a large amount of inactive and non-newsworthy accounts, which also greatly accelerated and improved downstream tasks.

Organization vs. person accounts classification. The second step is to assign a type label O or P to each account $a \in A_c$. The most important issue is to find features of a that help differentiate organization accounts from person accounts, i.e. to determine $F(N_a, I_a, U_a, D_a, S_a(n))$ in Equation 1. The following features proved to be useful:

- $NE(N_a)$: Named entity type of N_a , i.e. person, organization, or other. The named entity type is indicative of the corresponding account type.
- $Tag(I_a)$: Image tag type of I_a , i.e. person, or others. Personal accounts are more likely to use facial profile pictures. A face recognition algorithm can generate this tag.
- $MatchP(D_a, KW)$: Special keywords (KW) match in D_a , including occupation titles, biography words and first-person singular (FPS) and plural (FPP) pronouns (see Table 2). Intuitively, occupation titles, biography words or FPS occur more frequently in person accounts; while FPP occur more frequently in organization accounts.
- $More(S_a(n), FPS, FPP)$: Binary relation of comparative frequency of FPS vs. FPP in $S_a(n)$. Personal accounts are likelier to have more frequent occurrences of FPS.

To obtain a relatively complete list of occupation keywords, we downloaded Census 2010 Occupation Index³, and manually normalized and cleaned occupation titles. After extracting all features, we trained a classifier to divide A_c into two exclusive subsets, A_c^O and A_c^P .

Government & media accounts classification. The third step is to further divide A_c^O into three exclusive subsets: A_c^G (*government*), A_c^M (*media*) and A_c^{Ot} (*others*). Similarly, feature selection is the most important issue here; useful features are listed as follows:

- $MatchGM(D_a, KW)$: Special keywords match in D_a (see Table 2). We generated two keyword sets, one for government and another for media.

¹<http://www.census.gov/popest/data/cities/totals/2014/index.html>

²<https://moz.com/followerwonk/bio>

³<https://www.census.gov/people/io/methodology/indexes.html>

Table 2: Keywords sets defined in account classification.

Category	Keywords
Occupation title	author, reporter, economist, CEO, musician, lawyer, etc.
FPS	I, me, my, mine, myself
FPP	we, us, our, ours, ourselves
Biography	mom, dad, husband, wife, born, married, son, daughter
Government	governor, mayor, sheriff, police, attorney general, city attorney, city clerk, fire department, fire dept, fire fighters, emergency, transportation authority, city hall, inspector general
Media	news, radio, channel, television, TV

- $MatchG(U_a, KU)$: URL pattern match in U_a , mainly for govt. accounts. The URLs of govt. accounts are likely to contain “.gov” or “.gov.” In addition, each city has an official “City of XXX” website and the URLs of many local govt. accounts are likely to contain the website’s root path. For instance, the website of “City of LA” is `lacity.org`, and the pages for Transportation Dept., and Personnel Dept. of LA are located at `ladot.lacity.org`, and `per.lacity.org`.
- $FB(N_a)$: Facebook page category of N_a . A Twitter account generally has a corresponding Facebook page labeled with category name, which contains important information about the account type, such as *Government Organization, News/Media/Publishing*. We match the extracted Facebook categories with those keywords defined for government and media in Table 2.
- $Auth(S_a)$: Authority score of an account based on the retweet rates⁴, an important indicator of the account’s influence. Generally speaking, a government/media account has higher authority than other organization accounts.

Our preliminary study shows that government and media accounts only constitute a very small portion of total organization accounts. Therefore, we first apply keywords and URL patterns to filter A_c^O to a candidate set. and then leverage the features of Facebook category and authority score to further select A_c^G and A_c^M from the candidate set.

Applications of *TipMaster*

On top of the knowledge base, we developed a web UI that utilizes its local sources to monitor news and verify rumors.

News Monitoring. The final output of the above classification process is an account set $A_c^* = A_c^G \cup A_c^M$, based on which we can build a news monitoring platform. The basic functionality of this platform is to monitor local breaking news released from A_c^* in real-time. Specifically, we tracked the occurrences of a list of keywords in a real-time tweet stream, which are curated by journalists and indicative of breaking news, such as *shooting, killed, fire, accident, etc.*

Rumors Verification Social media is just as much a hotbed of false rumors as it is a platform for breaking real news (Liu et al. 2015). The knowledge base consisting of local official government and media accounts can effectively help us to clarify those suspect stories initiated and spread on social media. Especially, the verification process uses tweets containing both negative-sense keywords (e.g. “not,” “no,” “unfounded,” “false,” “denies,” etc.) and incident-related

⁴<https://moz.com/followerwonk/social-authority>

Table 3: Classification metrics of organization/person accounts.

	precision	recall	f1-score
$N_a + D_a$	0.72	0.85	0.78
$N_a + D_a + S_a$	0.87	0.87	0.87
All features	0.90	0.90	0.90

keywords (e.g. “gunshot,” “fire,” “killed,” “death,” etc.) are captured by the knowledge base.

All candidate informative tweets, either revealing a breaking story or debunking a hot rumor, are sent to a trained classifier. This classifier categorizes those candidate tweets into three types: breaking news event, rumor verification report or others (Liu et al. 2016), and the first two types of tweets are sent to the journalists. After that, more advanced analysis will be conducted on the knowledge base, like extracting entities and their relations behind the breaking news (Nourbakhsh et al. 2017), as well as mining trending topics (Li et al. 2016b), users, sentiment (Li et al. 2016a), the novelty and other semantic dimensions of news (Li et al. 2017), by leveraging a series of information retrieval, natural language processing, and machine learning tools. Since the focus of this article are to illustrate the construction of *TipMaster* and to demonstrate its core functionalities, other, more advanced functionalities won’t be detailed here.

Implementation & Evaluation

We extracted a total of 2.5M Twitter accounts from top 500 US cities using Followerwonk. We removed accounts associated with more than one city in their profiles, since local organization accounts of one city are unlikely to show more than one city name in their profile. Next, we sent those accounts through our person/organization account classifier, leaving 737K organization accounts. We used the Twitter API⁵ to extract each user’s 20 most recent tweets excluding re-tweets. In addition, we used the Stanford NLP tool⁶ for NER task and OpenCV⁷ for image tagging. Finally, we sent those organization accounts to the government/media account classifier and obtained 1,458 news accounts and 6,198 government accounts. We leveraged the Facebook Graph API to extract page categories⁸ as a feature of this classifier.

To evaluate the classification performance of person and organization accounts, we randomly selected 1,500 accounts from three cities: New York (a large city in the eastern US), Minneapolis (a medium-sized city in the central US) and San Bernardino (a small city in the western US). We manually labeled those accounts and filtered out 140 unknown accounts, which resulted in 427 organization accounts and 933 person accounts. We trained a Random Forest classifier with 10 trees based on labeled data, and our 3-fold cross

⁵https://dev.twitter.com/rest/reference/get/statuses/user_timeline

⁶<http://nlp.stanford.edu/software/CRF-NER.shtml>

⁷<http://opencv.org>

⁸<https://developers.facebook.com/docs/graph-api>

Table 4: Classification metrics of media/govt. accounts.

	precision	recall	f1-score
government	0.90	0.90	0.90
media	0.88	0.95	0.91

Table 5: Examples of breaking news detected by *TipMaster* (faster than Reuters news).

Event	Time (EST)	Account in KB	KB Type	Ahead of Reuters
San Bernardino Shooting	13:59 12/02/15	@SBCityFire	gov.	14min
Shooting in Pike County	09:15 04/22 /16	@Local12	media	45min
Baltimore police shot a boy	16:15 04/27/16	@baltimoresun	media	104min
Gas Well explosion in WPXI	08:30 04/29/16	@WPXI	media	46min

validation results are listed in Table 3. Clearly, precision and recall both peak when all features are combined together.

As is mentioned above, the percentages of government and media accounts are both very small in all organization type accounts. If we use the 427 organization accounts detected from the last step of organization and person accounts classification, the total amount of government and media accounts will be too small (i.e. less than 50). Therefore, To evaluate the classification performance of media and government accounts, we first generated two candidate sets from the above three cities (i.e. New York, Minneapolis, and San Bernardino) through keywords and URL patterns match, i.e., 422 candidate government accounts and 320 candidate media accounts. Then we labeled the two sets of accounts and obtained 239 government and 188 media accounts. After applying all features to the Random Forest classifiers trained for both government and media accounts classification, the evaluation results are listed in Table 4. The evaluation shows that *TipMaster* contains a large number of authentic local government and media Twitter accounts.

Case Studies

In this section, we will use two case studies to demonstrate the core functionalities built into our knowledge base: breaking news detection and rumors verification.

Breaking News Detection

Table 5 lists typical local breaking news detected through *TipMaster*. Compared with Reuters (using traditional news monitoring methods), our platform is able to detect local news almost one hour earlier on average. Furthermore, local governments tend to release follow-up stories with regards to the newsworthy events, which are captured by the tool. Our evaluation shows that this platform can empower Reuters journalists to keep track of local events in nearly real-time. Figure 3 shows the UI of our news tracking platform. Local government and media accounts are divided by cities and listed on the left column. Once breaking news is detected, a message window pops up and shows the tweet content. Additionally, other functionalities will be triggered as well, including: (1) detecting event location from tweet content or geotag information; (2) extracting key entities and concepts from all tweets related to the detected news (i.e.



Figure 3: UI of *TipMaster*'s news monitoring platform.

retweeting/mentioning) and further quantifying and visualizing their relations through statistical analysis; (3) using IR, NLP, and ML techniques to mine popular tweets, users, hashtags, and to analyze trends and collective sentiments.

Rumor Verification

In this section, we demonstrate the capacity of *TipMaster* as a verification platform through a case study. On Aug 9, 2017, Washington DC Police received a gunshot report near the headquarter of U.S. Environmental Protection Agency (EPA) before 5pm. Figure 4 illustrates the timeline of the storyline. The initial rumor was reported by a user (@afabbri) who posted a video at 4:15pm, and alleged that there was a gunshot attack in the EPA building according to what she heard from EPA employees. Several journalists on Twitter immediately began to communicate with her and asked for permission to share the news. About half an hour later, the official Twitter account of EPA (@EPA) responded to the gunshot report, based upon the investigation results from from DC police and Fed Protective Services. Only two mins later, local media agency WJLA also posted the most recent updates about the investigation to its Twitter account (@ABC7News). The important tweets from both @EPA and @ABC7News were captured by *TipMaster*. By contrast, *Reuters* published the report about the false gunshot event almost 20 mins after the official announcement from @EPA; New York Post published a similar report more than an hour later.

The above two case studies demonstrate the practical value of *TipMaster* in detecting break news events or debunking emergent rumors from Twitter, which is usually much faster than most well-known news agencies.

Conclusions

In this paper, we proposed a method for automated detection and classification of government/media accounts on Twitter, that help break news at a local level. After retrieving Twitter accounts linked with the top 500 US cities, we categorized them into organization and person accounts. Our classifier integrates heterogenous features from Twitter accounts with external data resources. Then, we further detected a set

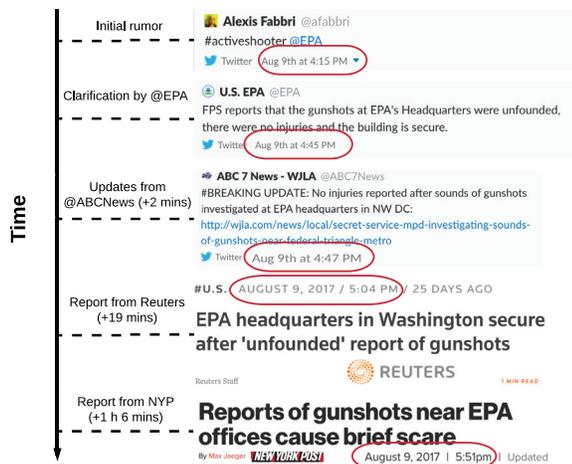


Figure 4: UI of *TipMaster*'s news verification platform.

of government and media accounts using similar classification strategies, based on which we implemented *TipMaster*, a platform capable of detecting and reporting local breaking news. Two case studies demonstrated the efficiency of *TipMaster* for the purpose of news detection and verification in real-time. Our future plan is to further expand the functionalities of the system in order to provide richer insights to journalists, and to perform a larger scale evaluation on the functional performance and timeliness of the tool.

References

- Cheng, T., and Wicks, T. 2014. Event detection using twitter: A spatio-temporal approach. *PLOS ONE* 9(6):1–10.
- De Choudhury, M.; Diakopoulos, N.; and Naaman, M. 2012. Unfolding the event landscape on twitter: Classification and exploration of user categories. In *Proceedings of the ACM 2012 CSCW, CSCW '12*, 241–244.
- Hu, Y.; Farnham, S. D.; and Monroy-Hernández, A. 2013. Whoo.ly: Facilitating information seeking for hyperlocal communities using social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13*, 3481–3490. New York, NY, USA: ACM.
- Ishikawa, S.; Arakawa, Y.; Tagashira, S.; and Fukuda, A. 2012. Hot topic detection in local areas using twitter and wikipedia. In *ARCS 2012*, 1 – 5.
- Kwak, H.; Lee, C.; Park, H.; and Moon, S. 2010. What is twitter, a social network or a news media? In *Proceedings of WWW, WWW '10*, 591–600. New York, NY, USA: ACM.
- Laird, S. 2012. How social media is taking over the news industry. <http://mashable.com/2012/04/18/social-media-and-the-news/>.
- Li, Q.; Shah, S.; Fang, R.; Nourbakhsh, A.; and Liu, X. 2016a. Tweet sentiment analysis by incorporating sentiment-specific word embedding and weighted text features. In *2016 IEEE/WIC/ACM WI*, 568–571.
- Li, Q.; Shah, S.; Liu, X.; Nourbakhsh, A.; and Fang, R. 2016b. Tweet topic classification using distributed language representations. In *2016 IEEE/WIC/ACM WI*, 81–88.
- Li, Q.; Nourbakhsh, A.; Shah, S.; and Liu, X. 2017. Real-time novel event detection from social media. In *2017 IEEE 33rd ICDE*, 1129–1139.
- Liu, X.; Nourbakhsh, A.; Li, Q.; Fang, R.; and Shah, S. 2015. Real-time rumor debunking on twitter. In *Proceedings of the 24th ACM CIKM*, CIKM '15, 1867–1870. New York, NY, USA: ACM.
- Liu, X.; Li, Q.; Nourbakhsh, A.; Fang, R.; Thomas, M.; Anderson, K.; Kociuba, R.; Vedder, M.; Pomerville, S.; Wudali, R.; et al. 2016. Reuters tracer: A large scale system of detecting & verifying real-time news events from twitter. In *CIKM*, 207–216. ACM.
- McCorriston, J.; Jurgens, D.; and Ruths, D. 2015. Organizations are users too: Characterizing and detecting the presence of organizations on twitter. In *Proceedings of ICWSM 2015*, 650–653.
- Nourbakhsh, A.; Li, Q.; Liu, X.; and Shah, S. 2017. "breaking" disasters: Predicting and characterizing the global news value of natural and man-made disasters. In *KDD workshop on Data Science + Journalism*.
- Oentaryo, R. J.; Low, J.-W.; and Lim, E.-P. 2015. *Chalk and Cheese in Twitter: Discriminating Personal and Organization Accounts*. Cham: Springer International Publishing, 465–476.
- Silva, L. D., and Riloff, E. 2014. User type classification of tweets with implications for event recognition. In *Proceeding of the Joint Workshop on Social Dynamics and Personal Attributes in Social Media*, 98 – 108.
- Takemura, H., and Tajima, K. 2016. Classification of twitter accounts into targeting accounts and non-targeting accounts. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media*, 291–296. New York, NY, USA: ACM.
- Tavares, G., and Faisal, A. 2013. Scaling-laws of human broadcast communication enable distinction between human, corporate and robot twitter users. *PLoS ONE* 8(7):1–11.
- Vis, F. 2013. Twitter as a reporting tool for breaking news. *Digital Journalism* 1(1):27–47.
- Xia, C.; Schwartz, R.; Xie, K.; Krebs, A.; Langdon, A.; Ting, J.; and Naaman, M. 2014. Citybeat: Real-time social media visualization of hyper-local city data. In *Proceedings of WWW, WWW '14 Companion*, 167–170.
- Xie, K.; Xia, C.; Grinberg, N.; Schwartz, R.; and Naaman, M. 2013. Robust detection of hyper-local events from geotagged social media data. In *Proceedings of the Thirteenth MDMKDD*, 2:1–2:9. New York, NY, USA: ACM.
- Yan, L.; Ma, Q.; and Yoshikawa, M. 2013. *Classifying Twitter Users Based on User Profile and Followers Distribution*. Berlin, Heidelberg: Springer Berlin Heidelberg, 396–403.
- Yin, P.; Ram, N.; Lee, W.-C.; Tucker, C.; Khandelwal, S.; and Salathé, M. 2014. *Advances in Knowledge Discovery and Data Mining: PAKDD 2014*. chapter Two Sides of a Coin: Separating Personal Communication and Public Dissemination Accounts in Twitter, 163–175.