

Finding and Tracking Local Twitter Users for News Detection

Hong Wei
Department of Computer Science
University of Maryland
College Park, Maryland 20742
hyw@cs.umd.edu

Jagan Sankaranarayanan
UMIACS
University of Maryland
College Park, Maryland 20742
jagan@umiacs.umd.edu

Hanan Samet
Department of Computer Science
University of Maryland
College Park, Maryland 20742
hjs@cs.umd.edu

ABSTRACT

The popular micro-blogging service, Twitter, hides invaluable information about real world events. Extracting such information, especially local news, exerts a measure of influence over various applications such as situation awareness and disaster recovery. Detecting small, local news is very challenging, however, due to the sparsity of publicly available Tweets on a specific area. In this work, we present a Twitter user-based method to mitigate the data sparsity. In essence, for a given geographical area, we aim to find as many Twitter users as possible from it and then track their posts to monitor the news happening in that area. However, only a small fraction of Twitter users provide information about their location, making the location information for most of Twitter users not available. Therefore, we utilize a geotagging procedure to estimate location for unknown-location Twitter users thereby finding more Twitter users in a given geographical area. Tracking the post updates of such Twitter users yields a real-time local live tweet stream and thus alleviates the paucity of local tweet data. On the real-time collected tweet stream, we perform online clustering to group tweets together to report potential news. The evaluation shows that in so doing, our method detects hundreds of more local news in comparison with solely utilizing the existing Twitter's publicly available tweet stream.

CCS CONCEPTS

•Information systems →Data streaming; Clustering;

KEYWORDS

Twitter, Live Tweet Stream, News Detection, Local News, Geotagging, Spark

ACM Reference format:

Hong Wei, Jagan Sankaranarayanan, and Hanan Samet. 2017. Finding and Tracking Local Twitter Users for News Detection. In *Proceedings of ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Redondo Beach, CA, November 7-10 2017 (SIGSPATIAL '17)*, 4 pages.

DOI: 10.1145/3139958.3141797

1 INTRODUCTION

TwitterStand [1–3] is a news tweet processing system that collects tweets and classifies them as news or non-news. It is based on our earlier work on browsing spatial [4] and news [5] data. It aggregates the news tweets into clusters, determines their geographical focus, displays them on a map, and makes them accessible via a map

Permission to make digital or hard copies 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 bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGSPATIAL '17, Redondo Beach, CA

© 2017 Copyright held by the owner/author(s). 978-1-4503-5490-5/17/11...\$15.00

DOI: 10.1145/3139958.3141797

query interface. However, its dependence on the data source Twitter Sample API stream¹—which is claimed to sample around 10% of all tweets but now is only 1%, greatly limits TwitterStand's ability to detect local news because small, local events may only span a very limited number of tweets. For example, Figure 1 shows an event about “Women’s Networking in West Roxbury” that happened in the morning of Oct 20, 2016 at West Roxbury, MA. By the time we took the screenshot, which is 4 hours later, we only find 10 tweets (including retweets) in total about this event by searching for keywords “women networking roxbury” in Twitter to find related tweets. Although the search function provided by Twitter doesn't guarantee finding all matching tweets, this strongly illustrates how sparse the tweets are. To make the matter worse, none of the related tweets appear in the access to Twitter Sample API. In this case, TwitterStand might not be able to capture enough tweets to detect this news event. In this work, we extend TwitterStand by enabling it to detect local news by alleviating the sparsity of tweet data.



Figure 1: Local news in West Roxbury, Boston on Oct 20th, 2016

Data sparsity is pervasive in Twitter's publicly accessible tweets and compromises the researchers' opportunity to identify events with smaller scales of tweets. Twitter provides three live stream endpoints to access public tweets: statuses/firehose², statuses/sample³ and statuses/filter⁴. Firehose has access to all tweets but costs too much to obtain according to an article in TIME BUSINESS⁵ and thus is not available to the public. Since the access to statuses/sample has a very low rate of tweets, we resort to exploiting statuses/filter. Statuses/filter provides 3 predicate parameters to specify: follow, track, and locations. Providing a list of Twitter users (at most 5,000),

¹<https://dev.twitter.com/streaming/reference/get/statuses/sample>

²<https://dev.twitter.com/streaming/public>

³<https://dev.twitter.com/streaming/reference/get/statuses/sample>

⁴<https://dev.twitter.com/streaming/reference/post/status/filter>

⁵<http://business.time.com/2013/10/08/twitter-is-selling-access-to-your-tweets-for-millions/>

“follow” returns their statuses in the stream; providing a list of keywords, “track” delivers the tweets matching one, some, or all of the keywords in the stream; while with specified geographical areas, “location” tracks the tweets falling in the given bounding boxes either according to the tweet’s embedded GPS coordinate values or place names. Since tweets having embedded geographical information only make up a tiny part of all tweet data [6], we explore Twitter Streaming API statuses/filter with parameter “follow” to obtain dramatically more tweets from an area. The problem is, whom to follow? Because the publicly-known location, which is provided in a Twitter user’s profile description—which we term *profile-location* in the rest of this paper, is only available for less than 20% of Twitter users [6].

To solve this problem, we implement an online social network-based Twitter user geotagging approach, which approximates the location of a Twitter user by examining publicly-known locations of his online friends (neighbors). With the help of this scheme, we try to find as many active Twitter users from a given area as possible, and putting their posting statuses (tweets) to the local live tweet stream to largely increase its number of local tweets. While real-time collecting tweets from a local place, a TF-IDF feature based online clustering procedure is performed to timely group tweets into potential news events.

The rest of this paper is organized as follows. Section 2 summarizes the related work. Section 3 describes the method of geotagging Twitter users to infer their home location. Section 4 specifies different sources used to collect local tweets. Section 5 outlines the clustering method to detect news. Section 6 describes the experimental evaluation of our methods. Section 7 contains concluding remarks and directions for future work.

2 RELATED WORK

There is a large body of related work that deals with extracting useful patterns (e.g., news, events) from social media, Twitter in particular. We found that clustering-based detection is the most popular technique, although different features are exploited to perform clustering such as bursty keywords [7–10], segments [11, 12] (e.g., one or more consecutive words in a tweet), tweet signals [13] (e.g., the occurrence of keywords over time can be viewed as signals) and sometimes a tweet’s bag-of-words model [14–17]. Two recent surveys by Atefeh and Khreich [18], and Abdelhaq [19] provide an excellent description of different techniques. Below, we review some of the related work that deals specifically with the problem of detecting local events.

There exist some methods working mainly on the geotagged tweets (i.e., containing embedded geographical [latitude, longitude] values) such as [20–22]. One common strategy of dealing with geotagged tweets is to treat geographical information as an additional variable to the existing models. For example, in calculating similarity between documents while performing a clustering algorithm, geographical distance between tweets can be incorporated in the clustering algorithm to form potential events [21, 23, 24]. Benson et al. [25], Hong et al. [26], Zhou and Chen [27], Wei et al. [28] treat geographical regions as latent variables in their generative topic detection model. Another popular strategy is to first tessellate the space into small, non-overlapping cells, and then determine the existence of news in one or adjacent cells based on if there is an abnormal abrupt number of tweets [19, 20, 22, 29–36].

Another set of methods do not necessarily require working on geotagged tweets as which are very rare in Twitter accounting for less than 1% [33, 37–39]. The strategy, after identifying events (e.g., by clustering similar tweets), is to utilize an additional step of spatial analysis to determine the location where the events are happening [1, 40, 41]. For example, TwitterStand [1], after clustering tweets to identify events, estimates each news cluster’s geographical focus by making use of both geographical information in the content of the tweet and by the source location of the users.

However, even though all possible event related documents are exploited (not just the ones with location information) by the methods described in previous paragraph, their data sources are still suffering from sparsity to detect small, local events. For example, TwitterStand’s data source, Gardenhose, even with a sampling rate of 10% of tweets, is still too few for small-scale events that might only span 3 ~ 5 tweets in total.

Thereafter, realizing it is the local data sparsity that undermines the opportunities for researchers to discover small-scale events in Twitter, we propose to enhance the public local live tweet stream in an area by i) finding as many Twitter users as possible that are from that area and then ii) tracking the tweets that they publish. Since the location information for most of Twitter users is not available, finding Twitter users from a local place relies on determining the location for unknown Twitter users, which might be estimated through social network based procedures [6, 42–45] or tweet content-based methods [46–50].

Weng and Lee [51] similarly track a number of users in Singapore to detect news but only at a small scale, i.e., 1K Twitter users. In contrast, we identify and track 176K users in Boston. Our work is also different from Albakour et al. [52], which directly chooses several areas in London to collect tweet data, and tries to detect events for each of these areas separately [52]. Their method doesn’t solve the problem of local data sparsity by using Twitter’s Streaming API, i.e., statuses/filter with parameter “locations”, which we will show in our experiment, is still very sparse and thus makes a very limited contribution to local news detection.

3 FINDING LOCAL TWITTER USER

The profile-location information for specifying where a Twitter user comes from is only sparsely available in public data. For example, Compton et al. [6] showed that the proportion of Twitter users with a reliable profile-location is below 20%. Therefore, inspired by studies that online social friendships are often formed over short geographic distances [53, 54], a social network-based Twitter user geotagging method which approximates the location of a Twitter user by examining publicly-known locations of the user’s online friends (neighbors)⁶ is proposed [6]. We implement an online version of their work [6] using Spark [55, 56] and address the social network-based geotagging problem from the view of solving an optimization problem. To be specific, inferring user locations is solved by finding

$$\min_{\mathbf{f}} \|\nabla \mathbf{f}\| \quad \text{s.t. } f_i = l_i, \forall i \in L \quad (1)$$

where $\mathbf{f} = (f_1, f_2, f_3, \dots, f_n)$ represents a location estimation for each user $1 \dots n$, L denotes the set of users who opt to make their locations l_i public, and $\|\nabla \mathbf{f}\|$, called the total variation is defined as,

$$\|\nabla \mathbf{f}\| = \sum_{ij} w_{ij} d(f_i, f_j) \quad (2)$$

⁶We will be using the terms *online friends* and *online neighbors* interchangeably.

where $d(\cdot, \cdot)$ measures geodesic distance via Vincenty’s formulae [57], and w_{ij} weighs the friendship between user i and user j , which essentially reflects how many times user i interacts with j such as retweeting, mentioning.

The above minimization problem could be solved by calculating, for each user, a geographical median point from their neighbors’ locations. One representative of geographical median calculation is *L1-multivariate median* [58], which essentially finds a point that minimizes the sum of its distances to a given list of points, and acts as a user’s estimated (geotagged) location.

4 TRACKING LOCAL TWITTER USERS

Given an area to collect tweets, we exploit three types of Twitter APIs as described in Section 1: statuses/filter w/ “follow”, statuses/filter w/ “locations” and statuses/sample. For statuses/filter w/ “follow”, we feed a list of Twitter users whose locations fall in the given area after performing a geotagging procedure in Section 3 to real-time track their posts; for statuses/filter w/ “location”, the input parameter is just the given area (e.g., a geographical bounding box) to collect tweets whose embedded coordinate values or place names fall inside the given area; for statuses/sample, as it samples on the general tweets, we only keep the ones whose publisher fall in the given area. Our experiments show that the source of statuses/filter w/ “follow” dramatically increases the number of local tweets.

5 ONLINE CLUSTERING TO DETECT NEWS

Grouping news tweets together is done by utilizing an online clustering procedure [1] which assigns an input news tweet to the closest candidate among a list of active clusters according to their TF-IDF based cosine similarity, calculated as follows:

$$\delta(t, c) = \frac{\overrightarrow{TFV}_t \cdot \overrightarrow{TFV}_c}{|\overrightarrow{TFV}_t| |\overrightarrow{TFV}_c|} \cdot e^{-\frac{(T_t - T_c)^2}{2(\sigma)^2}} \quad (3)$$

where $\overrightarrow{TFV}_t, \overrightarrow{TFV}_c$ are the TF-IDF feature vectors of a tweet t and a candidate cluster c , respectively. The exponential factor in the above equation is a Gaussian attenuator with a variance of σ on the cosine distance to account for closeness in the temporal dimension in clustering, in which T_t is the publication time of tweet t and T_c is cluster c ’s time centroid.

The main goal of clustering is to automatically group news tweets so that each cluster contains tweets pertaining to a specific topic. We adopt a one-shot process, meaning that once a tweet is added to a cluster, it remains there forever.

6 EXPERIMENTS

We choose Jan 16, 2017 in the Boston metropolitan area to evaluate our method by comparing the number of news events we can detect with that when not tracking a set of local Twitter users.

The input Boston metropolitan area is set to a geographical latitude/longitude bounding box, i.e., a rectangle area [-73.424434, 41.342222, -69.656400, 43.132453]. The three Twitter APIs described in Section 4 are initialized to collect local live tweets as follows: 1) statuses/filter w/ “location” simply starts by taking the Boston bounding box as input; 2) to start statuses/filter w/ “follow”, we first obtain 176K Twitter users for the input Boston bounding box (which is almost one-sixth of the population of Boston City) by utilizing the Twitter user geotagging procedure in Section 3, and provide this list of Twitter users to statuses/filter w/ “follow” to track

their real-time updates; 3) the tweets coming from statuses/sample are also kept as long as their publishers are among the list of Twitter users specified in statuses/filter w/ “follow”.

Table 1: Contributions of Different Local Live Tweet Stream

Source	# of tweets		# of news	
	Local tweets	News tweets	Involved	Exclusive
sample API	6,182	360	21	9
filter API w/ loc.	76,983	709	30	16
filter API w/ fol.	4,717,180	35,587	405	341
Total	4,800,345	35,695	409	

The collected local live tweets are then clustered to form news events, for which we analyze the contributions of different local live tweet stream sources to detecting news. The results are presented in Table 1, which first shows how many tweets each source contributes to collecting the local live tweet stream, and also to the news tweets (i.e., the tweets composing of detected news). We collected 4,800,345 tweets in total from the Boston area. Among which, Twitter Streaming API statuses/filter with parameter “follow” contributes the most, by making up of 98.3% of all the tweets. This is almost 100 times of another Twitter Streaming API statuses/filter with parameter “locations”. In contrast, statuses/sample only outputs a slight portion of local tweets for Boston area. Similarly, regarding the tweets comprising the events we detect, 99.6% are contributed by the Twitter Streaming API statuses/filter w/ “follow”. This reinforces the importance of tracking local Twitter users to detect news.

In addition, Table 1 also lists the number of “Involved” news (i.e., how many news a source’s tweets have participated in forming) and the number of “Exclusive” news (i.e., how many news a source’s tweets have exclusively formed, in other words, these news are formed by tweets only from this source). The result shows that the majority of news events are generated using the tweets in Streaming API statuses/filter w/ “follow”, indicating that by tracking local Twitter users, our method is able find much more news than solely using the Twitter’s publicly available tweet streams.

7 CONCLUSION AND FUTURE WORK

In this paper, we presented an approach to detect as much local news as possible in a geographical area. In order to deal with the infamous sparsity problem in publicly available Twitter data, we first tried to collect as many tweets as possible from a local place. To accomplish this, we first identified a large body of Twitter users from that area to track their posts in real-time. Since most of Twitter users don’t specify their home locations, we utilized an online Twitter user geotagging procedure to estimate a Twitter user’s location based on the location information of his neighbors. The experiments showed that doing so it dramatically increased the number of tweets generated from that area. In addition, the results also showed that this augmented real-time tweet stream helps to boost the number of detected local news in comparison to only using the existing Twitter’s API.

Although tracking local Twitter users helped us to detect more news events than before, there is much more news in the collected local live tweets that we failed to discover due to the current TF-IDF feature based similarity metric being susceptible to micro local news because the number of related tweets is so limited that the words used in the tweets to describe the events are not able to yield significant TF-IDF scores, which leads to failures in grouping them together or falsely putting them together with noisy tweets. In addition,

the geotagging procedure described in Section 3 works well with relatively dense social networks (e.g., Twitter users have higher numbers of online friends and their locations are geographically close), this is not necessarily the case for Twitter users who have weak evidence of the locations of their online friends (e.g., only few online friends provide location) as suggested in [43] and therefore needs improvement. We leave these questions to future work.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grant IIS-13-20791 and Grant CNS-1405688.

REFERENCES

- [1] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling. TwitterStand: News in Tweets. SIGSPATIAL '09.
- [2] N. Gramsky and H. Samet. Seeder Finder: Identifying Additional Needles in the Twitter Haystack. LBSN '13.
- [3] A. Jackoway, H. Samet, and J. Sankaranarayanan. Identification of Live News Events Using Twitter. LBSN '11.
- [4] H. Samet, H. Alborzi, F. Brabec, C. Esperança, G. R. Hjaltason, F. Morgan, and E. Tanin. Use of the SAND Spatial Browser for Digital Government Applications. *CACM*, 46(1):61–64, January 2003.
- [5] H. Samet, J. Sankaranarayanan, M. D. Lieberman, M. D. Adelfio, B. C. Fruin, J. M. Lotkowski, D. Panozzo, J. Sperling, and B. E. Teitler. Reading News with Maps by Exploiting Spatial Synonyms. *CACM*, 57(10):64–77, September 2014.
- [6] R. Compton, D. Jurgens, and D. Allen. Geotagging One Hundred Million Twitter Accounts with Total Variation Minimization. *BigData '14*.
- [7] M. Mathioudakis and N. Koudas. TwitterMonitor: Trend Detection over the Twitter Stream. SIGMOD '10.
- [8] M. Cataldi, L. Di Caro, and C. Schifanella. Emerging Topic Detection on Twitter Based on Temporal and Social Terms Evaluation. MDMKDD '10.
- [9] R. Parikh and K. Karlapalem. ET: Events from Tweets. WWW '13 Companion.
- [10] H. Sayyadi, M. Hurst, and A. Maykov. Event detection and tracking in social streams. ICWSM '09.
- [11] Y. Qin, Y. Zhang, M. Zhang, and D. Zheng. Feature-Rich Segment-Based News Event Detection on Twitter. IJCNLP '13.
- [12] C. Li, A. Sun, and A. Datta. Twevent: Segment-based Event Detection from Tweets. CIKM '12.
- [13] D. T. Nguyen and J. J. Jung. Real-time Event Detection on Social Data Stream. *Mobile Networks and Applications*, 20(4):475–486, 2015.
- [14] D. Shan, W. X. Zhao, R. Chen, B. Shu, Z. Wang, J. Yao, H. Yan, and X. Li. EventSearch: A System for Event Discovery and Retrieval on Multi-type Historical Data. KDD '12.
- [15] C. C. Aggarwal and K. Subbian. *Event Detection in Social Streams*. SDM '12.
- [16] S. Petrović, M. Osborne, and V. Lavrenko. Streaming First Story Detection with Application to Twitter. HLT '10.
- [17] H. Becker, M. Naaman, and L. Gravano. Beyond Trending Topics: Real-World Event Identification on Twitter. ICWSM '11.
- [18] F. Atefeh and W. Khreich. A Survey of Techniques for Event Detection in Twitter. *Computational Intelligence*, 31(1):132–164, February 2015.
- [19] H. Abdelhaq. *Localized Events in Social Media Streams: Detection, Tracking, and Recommendation*. PhD thesis, Heidelberg University, November 2015.
- [20] J. Krumm and E. Horvitz. Eyewitness: Identifying Local Events via Space-time Signals in Twitter Feeds. SIGSPATIAL '15.
- [21] C. Zhang, G. Zhou, Q. Yuan, H. Zhuang, Y. Zheng, L. Kaplan, S. Wang, and J. Han. GeoBurst: Real-Time Local Event Detection in Geo-Tagged Tweet Streams. SIGIR '16.
- [22] A. Magdy, A. Aly, M. Mokbel, S. Elnikety, Y. He, S. Nath, and W. Aref. *GeoTrend: Spatial trending queries on real-time Microblogs*. SIGSPATIAL '16.
- [23] M. Walther and M. Kaiser. Geo-spatial Event Detection in the Twitter Stream. ECIR '13.
- [24] A. Boettcher and D. Lee. EventRadar: A Real-Time Local Event Detection Scheme Using Twitter Stream. GreenCom '12.
- [25] E. Benson, A. Haghighi, and R. Barzilay. Event Discovery in Social Media Feeds. HLT '11.
- [26] L. Hong, A. Ahmed, S. Gurumurthy, A. J. Smola, and K. Tsioutsoulis. Discovering Geographical Topics in the Twitter Stream. WWW '12.
- [27] X. Zhou and L. Chen. Event Detection over Twitter Social Media Streams. *VLDB*, 23(3):381–400, June 2014.
- [28] W. Wei, K. Joseph, W. Lo, and K. Carley. A Bayesian Graphical Model to Discover Latent Events from Twitter. ICWSM '15.
- [29] A. Skovsgaard, D. Sidlauskas, and C. S. Jensen. Scalable top-k spatio-temporal term querying. ICDE '14.
- [30] H. Abdelhaq and M. Gertz. On the Locality of Keywords in Twitter Streams. IWGS '14.
- [31] H. Abdelhaq, C. Sengstock, and M. Gertz. EvenTweet: Online Localized Event Detection from Twitter. *PVLDB*, 6(12):1326–1329, August 2013.
- [32] K. Y. Kamath, J. Caverlee, K. Lee, and Z. Cheng. Spatio-temporal Dynamics of Online Memes: A Study of Geo-tagged Tweets. WWW '13.
- [33] K. Watanabe, M. Ochi, M. Okabe, and R. Onai. Jasmine: A Real-time Local-event Detection System Based on Geolocation Information Propagated to Microblogs. CIKM '11.
- [34] Y. Zhu and S. Newsam. Spatio-temporal Sentiment Hotspot Detection Using Geotagged Photos. SIGSPATIAL '16.
- [35] A. Magdy, M. F. Mokbel, S. Elnikety, S. Nath, and Y. He. Mercury: A Memory-Constrained Spatio-temporal Real-time Search on Microblogs. ICDE '14.
- [36] R. Lee and K. Sumiya. Measuring Geographical Regularities of Crowd Behaviors for Twitter-based Geo-social Event Detection. LBSN '10.
- [37] C. Budak, T. Georgiou, D. Agrawal, and A. El Abbadi. GeoScope: Online Detection of Geo-correlated Information Trends in Social Networks. *PVLDB*, 7(4):229–240, December 2013.
- [38] C.-H. Lee. Mining Spatio-temporal Information on Microblogging Streams Using a Density-based Online Clustering Method. *Expert Syst. Appl.*, 39(10):9623–9641, August 2012.
- [39] Z. Liu, Y. Huang, and J. R. Trampier. LEDS: Local Event Discovery and Summarization from Tweets. SIGSPATIAL '16.
- [40] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. WWW '10.
- [41] B. De Longueville, R. S. Smith, and G. Luraschi. "OMG, from Here, I Can See the Flames!": A Use Case of Mining Location Based Social Networks to Acquire Spatio-temporal Data on Forest Fires. LBSN '09.
- [42] R. Li, K. H. Lei, R. Khadiwala, and K. C.-C. Chang. TEDAS: A Twitter-based Event Detection and Analysis System. ICDE '12.
- [43] Y. Yamaguchi, T. Amagasa, and H. Kitagawa. Landmark-based User Location Inference in Social Media. COSN '13.
- [44] C. A. Davis Jr., G. L. Pappa, D. R. de Oliveira, and F. de L. Arcaño. Inferring the Location of Twitter Messages Based on User Relationships. *Trans. GIS*, 15(6):735–751, 2011.
- [45] A. Sadilek, H. Kautz, and J. P. Bigham. Finding Your Friends and Following Them to Where You Are. WSDM '12.
- [46] Y. Chen, J. Zhao, X. Hu, X. Zhang, Z. Li, and T.-S. Chua. From Interest to Function: Location Estimation in Social Media. AAAI '13.
- [47] Z. Cheng, J. Caverlee, and K. Lee. You Are Where You Tweet: A Content-based Approach to Geo-locating Twitter Users. CIKM '10.
- [48] K. Ryo and S. Moon. Inferring Twitter User Locations with 10 Km Accuracy. WWW '14 Companion.
- [49] B. Han, P. Cook, and T. Baldwin. Text-based Twitter User Geolocation Prediction. *J. Artif. Int. Res.*, 49(1):451–500, January 2014.
- [50] N. Dalvi, R. Kumar, and B. Pang. Object Matching in Tweets with Spatial Models. WSDM '12.
- [51] J. Weng and B.-S. Lee. Event Detection in Twitter. ICWSM '11.
- [52] M.-D. Albakour, C. Macdonald, and I. Ounis. Identifying Local Events by Using Microblogs As Social Sensors. OAIR '13.
- [53] Geography of Twitter networks. *Social Networks*, 34(1):73 – 81, 2012.
- [54] D. Mok, B. Wellman, and J. Carrasco. Does Distance Matter in the Age of the Internet? *Urban Studies*, 47(13):2747–2783, 2010.
- [55] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M. J. Franklin, S. Shenker, and I. Stoica. Resilient Distributed Datasets: A Fault-tolerant Abstraction for In-memory Cluster Computing. NSDI '12.
- [56] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica. Spark: Cluster Computing with Working Sets. HotCloud '10.
- [57] T. Vincenty. Geodetic inverse solution between antipodal points. *Richard Rapp Geodetic Science Ohio State University*, 1975.
- [58] Y. Vardi and C.-H. Zhang. The multivariate L1-median and associated data depth. *Proceedings of the National Academy of Sciences*, 97(4):1423–1426, 2000.