

# SBP-BRIMS 2016 Tutorial: Exploring Public Events Using Social Media Data: Challenges in Extraction, Modeling and Analyzing

Fred Morstatter  
Arizona State University  
[fred.morstatter@asu.edu](mailto:fred.morstatter@asu.edu)

Kenneth Joseph  
Carnegie Mellon University  
[kjoseph@cs.cmu.edu](mailto:kjoseph@cs.cmu.edu)

Sumeet Kumar  
Carnegie Mellon University

Social media contains a significant amount of information about its users and their actions and opinions both online and off.

In this tutorial, we provide an overview of a variety of important topics relating to social media data and how it can be used to study public events.

We first will provide an overview of how social media data is collected. Tools such as TweetTracker facilitate large-scale collection of social media data. TweetTracker is a powerful tool from Arizona State University that can help you track, analyze, and understand activity on Twitter, Instagram, Youtube, VKontakte, and Yik Yak. The system also allows a variety of visualizations based on this information, including streaming geospatial maps, tag cloud summarizations, post-event investigations in pseudo real-time, automatic translation of Non-English tweets, and keyword trending and comparison. TweetTracker is used to assist Humanitarian Assistance and Disaster Relief by working NGOs and other agencies. In this tutorial we will demonstrate how this tool can be applied in other areas of research. In addition to the research applications of social media data, we will discuss the implications of bias that is present in this dataset.

After discussions about data collection, we will move on to consider methodologies for extracting events from social media data, specifically from Twitter. We will consider a variety of examples of state-of-the-art approaches for this task, as well as discussing where open research questions in the field exist. Finally, we will close with a similarly structured discussion of the field of sentiment analysis on Twitter, along with a case study showing how sentiment analysis of Twitter data can provide indicators for various kinds of offline events.

**Target Audience:** Researchers and graduate students with a background in social computing, data mining, natural language processing and machine learning.

## About the Presenters:

Fred Morstatter is a PhD student in computer science at Arizona State University in Tempe, Arizona. Fred won the Dean's Fellowship for outstanding leadership and scholarship during his time at ASU. Among his publications is an ICWSM paper that investigates the representativeness of Twitter's Streaming API, a WWW Web Science paper that seek to find periods of bias automatically in streaming Twitter data, 2 KDD demo papers, an article in IEEE Intelligent Systems, and a book: Twitter Data Analytics. He has served as a PC member of ICWSM 2014 and 2016, IEEE/CIC ICC 2014 Symposium on Social Networks and Big Data, and has been a co-chair of the Social Computing, Behavioral-Cultural Modeling and Prediction Conference's Grand Challenge organizing committee in 2014, 2015, and 2016. He has been a Visiting Scholar at Carnegie Mellon University as well as a Research Intern at Microsoft Research. He is the Principal Architect for TweetXplorer, an advanced visual analytic system for Twitter data. A full list of publications can be found at <http://www.public.asu.edu/~fmorstat>. Contact him at [fred.morstatter@asu.edu](mailto:fred.morstatter@asu.edu).

Kenny Joseph is a graduate student in the Societal Computing program at Carnegie Mellon University. His research focuses at the intersection of machine learning and the social sciences, and his work has been published at various reputed conferences including WWW, ICWSM, ASONAM and the Journal of Mathematical Sociology. His webpage is <http://cs.cmu.edu/~kjoseph>.

Sumeet Kumar is a PhD student in the Electrical and Computer Engineering department at Carnegie Mellon University, He is advised by Prof. Kathleen M. Carley. His research interests include cyber-security, data mining and high dimensional networks. Before joining CMU, Sumeet worked at Symantec Corporation, Talentica Software, and Intergraph Corporation.