Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Analyzing the spread of tweets in response to Paris attacks

Sreten Cvetojevic*, Hartwig H. Hochmair

Fort Lauderdale Research and Education Center, University of Florida, 3205 College Avenue, Davie, FL 33314, USA

ARTICLE INFO

Keywords:
Twitter
Information spread
Social media
Paris
Terror attacks

ABSTRACT

Twitter is a widely used social media platform that provides a rich source of geotagged posts along with extended content information, such as hashtags or images. Since tweets are frequently used for the instant sharing of news, thoughts, and ideas, they reflect to some extent the effects of critical events, such as natural disasters or political riots, on the society, and how the society perceives such events. This paper analyzes, using exploratory methods and regression models, the spatio-temporal patterns of geo-located tweets that were posted in response to the November 2015 Paris terrorist attacks. It determines how information about this event spread around the world in the Twitter network and assesses how content category, tweet format, and profession of the user influence tweet popularity, measured by the number of retweets. The paper contributes to a better understanding of how the Twitter community reacts to unexpected events and how information about such an event propagates geographically.

1. Introduction

Over the past decade, the number of social media and crowd-sourced data-sharing platforms has grown substantially and opened a new era of information collection and analysis. Understanding the dynamics of social networks is crucial for tracking of opinions (e.g. political trends), management of crises (e.g. environmental natural hazards or diseases), optimization of business performance (e.g. marketing campaigns), or the detection of popular topics (Guille, Hacid, Favre, & Zighed, 2013). Twitter provides a prominent platform to study communication patterns among people and the information flow between them, although, unlike many other social media platforms, Twitter does not enforce reciprocal sharing (Lotan et al., 2011). The (non-spatial) spread of information through the Twitter network has been analyzed in numerous studies (Ferguson et al., 2014; Lerman & Ghosh, 2010; Pei, Muchnik, Andrade, Jr., Zheng, & Makse, 2014; Romero, Meeder, & Kleinberg, 2011), which complements another major thread of Twitter-related analysis, namely that of human mobility patterns (Hawelka et al., 2014; Hochmair & Cvetojevic, 2014; Hübl, Cvetojevic, Hochmair, & Paulus, 2017; Jurdak et al., 2015; Lenormand, Tugores, Colet, & Ramasco, 2014, Lenormand, Gonçalves, Tugores, and Ramasco, 2015; Li, Li, & Shan, 2017; Steiger, Ellersiek, Resch, & Zipf, 2011; Valle et al., 2017). Although several studies addressed the connection between geographic and social space when analyzing community interaction in social media platforms (Gründemann & Burghardt, 2016; Takhteyev, Gruzd, & Wellman, 2012) most information diffusion models operate exclusively within the social space, focusing, for instance, on information promotion (Achananuparp, Lim, Jiang, & Hoang, 2012), or the effects of repeated exposure to hashtags on hashtag adoption (Romero et al., 2011). To better understand the information spread across the physical world, there is a need to integrate spatial components into diffusion models.

As a step in this direction, we selected a series of six attacks (including suicide bombings and mass shootings) that occurred in Paris on the night of November 13th, 2015, and analyzed the diffusion of tweets that contain information pertaining to this event around the globe. Related tweets were divided based on format and content. Included formats are tweets with images, tweets with hashtags and tweets with keywords. Related images posted through tweets were visually inspected to identify dominant content categories. This led to two distinct content categories, namely, tweets related to the attacks and those expressing sympathy or support. Diffusion characteristics were then analyzed for each of these two classes separately. This two-class content distinction is in line with an earlier study (Seo, 2014) which analyzed images posted to the November 2012 Gaza conflict. It found that Israeli images primarily featured the analytical propaganda theme, which included images relating to attacks and destruction, whereas the emotional propaganda theme, e.g., raising sympathy towards their own people, was dominant in Hamas images. Our paper identifies several factors that influence tweet popularity (measured by the number of retweets), including content category (attacks vs. support related), tweet format (keywords vs. hashtags vs. images), and Twitter user profession (journalist vs. non-journalist). Using these categories, various exploratory spatial methods, such as Kernel density maps, are

* Corresponding author.
E-mail addresses: scvetojevic@ufl.edu (S. Cvetojevic), hhhochmair@ufl.edu (H.H. Hochmair).

https://doi.org/10.1016/j.compenvurbsys.2018.03.010
Received 4 April 2017; Received in revised form 12 February 2018; Accepted 27 March 2018

applied to assess the global spread of event-related information through tweets. This is followed by a spatiotemporal negative binomial regression model, which uses tweets with event-related hashtags to identify significant predictors of information spread around the world. In summary, this study addresses the following research objectives:

- determine the effect of tweet content category, tweet format, and user profession on the popularity of tweets that are posted in connection with the Paris attacks;
- explore the geographic spread of event-related tweets over time;
- use of tweets with hashtags that relate to the Paris attacks to identify factors contributing to the information spread around the world within a spatiotemporal regression model.

The remainder of the paper is structured as follows. Section 2 reviews previous work on information diffusion through Twitter. This is followed by a description of the study setup in Section 3. Section 4 provides results of the tweet popularity analysis, followed by results of exploratory analysis methods and a spatiotemporal regression model for twitter related information diffusion in Section 5. Section 6 discusses findings and the utilized analysis methods, which is followed by conclusions and directions for future work.

2. Related work

The geospatial data component that comes from social media content and from crowd-sourcing applications used for communication, navigation, or sharing travel experiences, is primarily generated by passive, often unaware, contributions means, and therefore sometimes referred to as Involuntary Geographic Information (iVGI) (Fischer, 2012). Although georeferenced tweets fall into the same category, and the sharing of one's location is not the main purpose of tweets, Twitter position information has been frequently used to assess the spatiotemporal dimension of emergency situations, such as earthquakes, floods, forest fires, or terrorist attacks (De Longueville & Smith, 2009; Hung, Kalantari, & Rajabifard, 2016; Li & Goodchild, 2010; MacEachren et al., 2011), to predict the spread of diseases (Brennan, Sadilek, & Kautz, 2013; Signorini, Segre, & Polgreen, 2011), and to model human mobility patterns in the case of unexpected events (Shelton, Poorthuis, Graham, & Zook, 2014).

Twitter is used by over 300 million users every month and therefore provides a significant data source for studying communication patterns and information flows among people (Lotan et al., 2011; Pei et al., 2014). However, it suffers from user sampling bias (Duggan, Ellison, Lampe, Lenhart, & Madden, 2015), and geographical bias through its concentration on certain countries (Hawelka et al., 2014). Furthermore, only about 1% of all tweets are geo-tagged (Graham, Hale, & Gaffney, 2014). This means that results of Twitter studies are not necessarily representative of the general population or even of all Twitter users. To compensate for the scarcity of geo-tagged tweets, various studies have explored methods to geo-locate tweets (Cheng, Caverlee, & Lee, 2010; Zahra, Ostermann, & Purves, 2017) or Twitter users (Jurgens, 2013; Kotzias, Lappas, & Gunopulos, 2014) through other sources of information in the tweet post or in the user profile, such as geographic references in the tweet text and the social network structure. Though these geo-positioning methods are consistently improving, they add a level of positional uncertainty to any subsequent analysis, and often require manual checks for reliable results. Therefore, for the presented study only geo-tagged tweets were used.

Modeling of information diffusion in the Twitter network was often approached through the analysis of retweet patterns (Guille et al., 2013), where a retweet is an action taken by a Twitter user to share someone else's tweets without alteration (Compston, 2014). For example, Cha, Haddadi, Benevenuto, & Gummadi (2010) compared three measures of user influence on others, namely the number of followers, the number of retweets, and the number of user mentions. Results showed that popular users with a high number of followers do not necessarily have more retweets and mentions, but that it is more influential to have an active audience that retweets or mentions the user. Another study showed that tweets that contain interesting URLs (as rated by others), and are posted by users with many followers were likely to be more widely spread (Bakshy, Hofman, Mason, & Watts, 2011). Similarly, Pei et al. (2014) used several network topology measures, including degree, PageRank, and k-core, to detect influential spreaders of information in online social media platforms Twitter, Facebook, and Livejournal. Based on a diffusion network model Yang and Counts (2010) predicted the speed, scale, and range of information diffusion on Twitter using a variety of user and tweet related predictors, including a user's activity level, the presence of URL in a tweet, or the stage of topic lifespan when a tweet was posted. Achananuparp et al. (2012) introduced the notion of weak retweets in their information propagation model. This concept describes a user posting a tweet that mentioned a relevant item, such as a URL or hashtag, from an earlier tweet posted by another user.

Besides retweet patterns, hashtags have often been used to observe content trends and to track topical information propagation. A Twitter hashtag is a string of characters preceded by the hash (#) character, and is generated by users as a method to categorize content and to highlight topics. A recent study extracted sentiments and topics from tweets that contained the #prayforparis hashtag and that were sent four days after the Paris attacks (Chong, 2016). The topics were extracted using latent semantic analysis (LSA) (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Evangelopoulos, Ashton, Winson-Geideman, & Roulac, 2015; Landauer & Dumais, 1997) and included among others a tribute to the victims of the Paris attack during the soccer game between England and France. Lotan et al. (2011) analyzed Twitter information flows during the 2011 revolutions in Egypt and Tunisia for mainstream media organizations, journalists, and bloggers using tweets with hashtags, such as #sidibouzid or #jan25. The study concluded that Twitter accounts of organizations have substantially higher retweet rates than accounts of individuals, but that news on Twitter is being co-constructed by bloggers and activists alongside journalists. Tsur and Rappoport (2012) showed that a post's content (e.g. length of a hashtag) and context (e.g. cognitive categories), as well as the topology of the social graph (e.g. number of followers) and global temporal features (e.g. peak hours) are important predictors of the popularity of hashtags over time. Another study found that the spread of hashtags varies by topic and that, especially for political hashtags, repeated exposure leads to frequent hashtag adoption by followers (Romero et al., 2011). Chang (2010) proposed a Diffusion of Innovation Theory that examines a trend of hashtag adoption during certain time periods after the user has been exposed to hashtag information.

Regarding news topicality Kwak, Lee, Park, and Moon (2010) compared the occurrence of headlines between Twitter and CNN and found that some events, such as accidents and sporting events, broke out on Twitter first. A comparative analysis of the relative importance of social media for news in six European countries, Japan, and the U.S. revealed that television is still the most widely used and most important source of news (Nielsen & Schrøder, 2014).

Several studies examined the ties between spatial and social network structure on twitter. For example, it was found that smaller Twitter networks are more socially clustered and extend over a smaller physical distance than larger ones, suggesting that network and physical distances are related (Stephens & Poorthuis, 2014). Similarly, Takhteyev et al. (2012) showed that a substantial share of Twitter ties lies within the same metropolitan region and that distance related variables, such as language, country, and the number of flights affect Twitter ties between regional clusters.

Overall, the literature review reveals that the spatial and geographic aspects in current diffusion network models of social media platforms are largely neglected. To narrow this research gap, the role of distance and spatial hierarchy will be explored in the context of information