Weather events identification in social media streams: tools to detect their evidence in Twitter

Valentina Grasso[1,2], Imad Zaza[3], Federica Zabini[1,2], Gianni Pantaleo[3], Paolo Nesi[3], and Alfonso Crisci[1]

[1]IBIMET CNR, Institute of Biometeorology, Italian National Research Council, Florence, Italy
[2]LAMMA Consortium, Tuscany Region-CNR, Sesto Fiorentino, Italy
[3]DISIT Lab, Distributed Technologies Lab - Dep. of Information Engineering (DINFO), University of Florence, Florence, Italy

ABSTRACT

Severe weather impact identification and monitoring through social media data is a good challenge for data science. In last years we assisted to an increase of weather related disasters, also due to climatic changes. Many works showed that during such events people tend to share messages by means of social media platforms, especially Twitter. Not only they contribute to *"situational"* awareness, improving the dissemination of information during emergency, but can be used to assess social impact of crisis events. We present in this work preliminary findings concerning how temporal distribution of weather related messages may help the identification of severe events that impacted a community. Severe weather events are recognizable by observing the synchronization of Twitter activity volumes across keywords and hashtags, including geo-names. Impacting events present a recognizable visual pattern recalling a "Half Onion Shape", where Twitter activity across keywords is synchronized. In reason of these interesting indications, it's becoming fundamental to have a suite of reliable tools to monitor social media data. For Twitter data a comprehensive suite of tools is presented: the DISIT-Twitter Vigilance Platform for Twitter data retrieve, management and visualization.

Keywords: weather event identification, social media data, Twitter

INTRODUCTION

Any weather event, severe or not, is bounded in time and space. When these events affect citizens and urban environments they engender great public attention especially through social media conversations, messages and users interaction. Furthermore in recent years climate change increased public concerning on this topic and weather and climate have been experienced more often as threats. During severe weather events social media data represents a novel source to quantify the impact of the phenomena and their temporal evolution. Many researches investigated already how the public increasingly rely on social media during disasters and natural hazards (Palen et al., 2010; Vieweg et al., 2010; Giglietto et al., 2013; Hughes et al., 2014; Mendoza et al., 2010; Kongthon et al., 2012; Bruns and Burgess, 2014; Sutton et al., 2014; BonnanWhite et al., 2014; Starbird and Palen, 2010). Social media have become a primary source of information during emergencies where emergency managers, authorities and citizens may interact with each others, providing and receiving useful information as the event unfold. As research (Vieweg et al., 2010) shows, the information shared on social media, particularly on Twitter may improve *situational awareness* and help people to collect useful information for decision making . Because users cannot look at millions of messages at a time during a given event, they usually rely on hashtags to coordinate conversations about events or topics. Hashtag is a word prefixed with a hash symbol (#) used to categorize the Tweets. They first emerged on Twitter during the 2007 California wildfires as a way to track relevant information about the natural disaster by labeling content so that it could be filtered and shared. Hashtagging practices are thus becoming very important for public and private organizations that wish to deliver important information to the public during a crisis. In order to increase messages retrieval and coordinate conversations public bodies and organization are proposing *codified hashtags* to be used during crisis events; some government, like the Philippine, published a ”grammar” to help citizens to create proper ones in case of emergency. Also in Italy a proposal for codified hashtags for weather warning has been proposed (*http://capitanachab.tumblr.com/post/74053317969/20-hashtag-per-una-protezione-civile-partecipata*) and used since 2014. Even if not all users use hashtags it surely exists a crisis lexicon (Olteanu et al., 2014) as confirmed by some systematic review and collection of the more used terms (Temnikova et al., 2015). Considering the amount of Twitter users (at beginning of 2016 there were around 6,4 millions of monthly active users in Italy as reported by *http://vincos.it/2016/04/01/social-media-in-italia-analisi-dei-flussi-di-utilizzo-del-2015/*), the analysis of Twitter posts, where different class of users find and share information, give a good chance for a fast social recognition of weather impacts. Furthermore there is a link between spatial population density and the use of social media (Botta et al., 2015) and during an eventual environmental disaster this is recognizable Toepke and Starsman (2015). Population density is considered one of the key-factor for vulnerability assessment (Beccari, 2016). Furthermore the use of geo-hashtag (Lachlan et al., 2014) or codified hashtag containing geographical indications represent a reliable option to share geo-localized information. The information coming from the *interconnected world of techno-social systems* (Vespignani, 2009), where social media could be considered a data interface requires novel frameworks to verify work’s hypotheses and mostly reliable tools to facilitate extraction and analysis. The amount of data potentially retrievable from social media as Twitter is very huge and tools must be addressed for an effective data refinement and filtering especially if social media analysis is aimed at event’s identification. Some experience on weather, but also on more general topic, already exists. About Twitter streams, considered not only a mere media amplifier (Petrovic et al., 2013), generally the methodologies proposed are based on evaluating the time and geography dimension of the streams by finding shifts in the inverse document frequency, in order to capture trending terms (Boettcher and Lee, 2012; Weiler et al., 2013). The theoretical basis of these methodologies started from the seminal works belonging to information science (Sparck Jones, 1972). Generally, for social media event detection, the time comparison between abundance of term-related streams and the evaluation of shifts in document term frequency are methods widely used. The critical step is a good choice of terms themselves and their relative semantic differences for each kind of events investigated, that in our case are represented by the ones linked to severe weather and its impacts. The availability of tools that allow multiple and connected term queries on social media data is a point of strength, because it helps an effective Twitter monitoring.

TWITTER VIGILANCE PLATFORM

DISIT Twitter Vigilance (TV) platform, available at website *http://www.disit.org/tv/* is a multipurpose comprehensive dashboard that provides different tasks suitable for Twitter streams monitoring. The architecture of the platform is described in Figure 1. In particular, main tasks performed by the platform are: (i) a continuous data extraction by using Twitter Search API; (ii) feeding a desktop dashboard where it is possible to easily configure queries on Twitter API search. Twitter queries proposed by the Twitter Vigilance may be a single users (user), a simple word or an #hashtag, or a combination of any of the above. Every single query term is stored by the platform; users may perform flexible data extraction through appropriate queries. Semantically oriented combination of queries is defined as ”channel”. DISIT TV harvests Twitter messages; data are then easily visualized and plotted against time through a graphical interface. The channel and/or search metrics continuously displayed by dashboard are: the number of Tweets, the number of Re-Tweets and the number of users. More information on DISIT Twitter Vigilance Platform is available at this web-link *http://www.disit.org/6793*. Single channel reports and time series visualization could be done by using the platform services and inserting the name of channel in this way as final argument: *http://www.disit.org/tv/index.php?p=chart_singlechannel&canale=NAME_OF_CHANNEL*. Not all channels on Twitter Vigilance are public, some of the channel’s owner prefer to keep them private.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2241v1 | CC BY 4.0 Open Access | rec: 5 Jul 2016, publ: 5 Jul 2016

METHODS

TV platform was used to analyze Twitter streams related to severe weather events. Several channels were created to query Twitter for messages containing hashtags and simple words semantically related to severe weather or containing names of places recently hit by natural hazards in Italy. Channels were created to respond to different purposes: monitoring use of Italian codified hashtags for weather warning over time, track activity of several Twitter account of weather forecasting services, analysis of specific severe weather events, like flash-floods. Tweets monitoring made it possible to evaluate the efficacy of different hashtags to retrieve information during natural hazards. Through the TV platform was possible to monitor the evolution of Twitter activity across codified or simple hashtags, semantically or geographically related to severe weather events (like for instance #allertameteoTOS; #maltempo; #temporale; #nubifragio; #alluvione; #Firenze; #Toscana; #Sardegna; #Olbia; #Rossano; #Calabria). Temporal evolution of Twitter activity referred to keyword and places proved to be a valuable tool to rapidly assess whether or no severe weather impacted a community.

RESULTS AND DISCUSSION

The following table shows some of the channels actually active in the platform.

Table 1. Weather Channel active on DISIT TV platform

Channel	Total	Tweets(%)	Re-tweets (%)	Period
Allertameteo TOSCANA	1051131	59.64%	40.36%	2009-12-04
MeteoUSER	58385	39.69%	60.31%	2012-08-07
rossano	117213	50.64%	49.36%	2013-04-06
protezione civile toscana	33501	19.95%	80.05%	2014-06-21
Codified Hashtags Allerta	29431	32.2%	67.8%	2014-11-07
LaMMA	10860	31.46%	68.54%	2012-12-14
CALDO	2815412	55.98%	44.02%	2009-10-23

From these and in particular from the ”Codified Hashtags Allerta” channel, three severe weather events occurred in 2015 were identified and highlighted: (i) Olbia flooding event (Date: 01-10-2015) in north-eastern Sardinia 3; (ii) the flash-flood event of Rossano Calabro situated in Calabria 4 (Date:12-08-2015) ; (iii) severe weather episode of Florence located in northern Tuscany 5 (Date:01-08-2015). These events were characterized by short time bounding in time and space: daily horizon and more localized event. Meteorological observation networks not ever are able to detect the magnitude of these particular class of events. Social media monitoring working as ”social radar” can be very useful. The event identification is more clear in the all presented figures. Twitter activity across different keywords, hashtags and geographic names showed that although over time they show differences in volumes and trends, during impacting event they all synchronize reaching their relative peaks. A visual pattern of simultaneous peaks in tweets activity resembling to a ”half onion shape” is recognizable, where higher volumes of tweets are reached by local geo-names and generic hashtags and smaller one by codified hashtags. These may give interesting insights on how to improve Twitter communication and monitoring by emergency managers and institutions during severe weather events.

ACKNOWLEDGMENTS

This study was supported by the European Project Culture And RISk management in Man-made And Natural Disasters (CARISMAND) - H2020-DRS-2014 - Grant agreement no: 65374.

REFERENCES

Beccari, B. (2016). A Comparative Analysis of Disaster Risk, Vulnerability and Resilience Composite Indicators. *PLoS Currents.*, (8).

Boettcher, A. and Lee, D. (2012). Eventradar: A real-time local event detection scheme using twitter stream. In *Green Computing and Communications (GreenCom), 2012 IEEE International Conference on*, pages 358–367. IEEE.

BonnanWhite, J., Shulman, J., and Bielecke, A. (2014). Snow tweets: emergency information dissemination in a us county during 2014 winter storms. *PLoS currents*, 6.

Botta, F., Moat, H. S., and Preis, T. (2015). Quantifying crowd size with mobile phone and twitter data. *Royal Society open science*, 2(5):150162.

Bruns, A. and Burgess, J. (2014). Crisis Communication in Natural Disasters: The Queensland Floods and Christchurch Earthquakes. In Weller, Katrin, Bruns, Axel, Burgess, Jean, Mahrt, Merja, P., editor, *Twitter and Society*, pages 373–384. Cornelius (EDS) Peter Lang, New York.

Giglietto, F., Carlo, U., and Lovari, A. (2013). Amministrazioni pubbliche e gestione degli eventi critici attraverso i social media : il caso di # firenzeneve. 2013:98–116.

Hughes, A. L., St. Denis, L. a. a., Palen, L., and Anderson, K. M. (2014). Online public communications by police & fire services during the 2012 Hurricane Sandy. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pages 1505–1514.

Kongthon, A., Haruechaiyasak, C., Pailai, J., and Kongyoung, S. (2012). The Role of Twitter during a Natural Disaster : Case Study of 2011 Thai Flood. *Technology Management for Emerging Technologies*, pages 2227–2232.

Lachlan, K. A., Spence, P. R., Lin, X., Najarian, K. M., and Greco, M. D. (2014). Twitter Use During a Weather Event: Comparing Content Associated with Localized and Nonlocalized Hashtags. *Communication Studies*, 65(5):519–534.

Mendoza, M., Poblete, B., and Castillo, C. (2010). Twitter Under Crisis : Can we trust what we RT ?

Olteanu, A., Castillo, C., Diaz, F., and Vieweg, S. (2014). Crisislex: A lexicon for collecting and filtering microblogged communications in crises. In *ICWSM*.

Palen, L., Anderson, K. M., Mark, G., Martin, J., Sicker, D., Palmer, M., and Grunwald, D. (2010). A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters. In *Proceedings of the 2010 ACM-BCS visions of computer science conference*, page 8. British Computer Society.

Petrovic, S., Osborne, M., McCreadie, R., Macdonald, C., and Ounis, I. (2013). Can twitter replace newswire for breaking news?

Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21.

Starbird, K. and Palen, L. (2010). Pass it on?: Retweeting in mass emergency. *Proceedings of the 7th International ISCRAM Conference*, (December 2004):1–10.

Sutton, J., Spiro, E. S., Johnson, B., Fitzhugh, S., Gibson, B., and Butts, C. T. (2014). Warning tweets: serial transmission of messages during the warning phase of a disaster event. *Information, Communication & Society*, 17(6):765–787.

Temnikova, I., Castillo, C., and Vieweg, S. (2015). Emterms 1. 0: a terminological resource for crisis tweets. In *ISCRAM 2015 proceedings of the 12th international conference on information systems for crisis response and management*.

Toepke, S. L. and Starsman, R. S. (2015). Population distribution estimation of an urban area using crowd sourced data for disaster response. In *Proceedings of the Twelfth International Conference on Information Systems for Crisis Response and Management, Kristiansand, Norway*.

Vespignani, A. (2009). Predicting the behavior of techno-social systems. *Science*, 325(5939):425–428.

Vieweg, S., Hughes, A. L., Starbird, K., and Palen, L. (2010). Microblogging during two natural hazards events. *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 1079.

Weiler, A., Scholl, M. H., Wanner, F., and Rohrdantz, C. (2013). Event identification for local areas using social media streaming data. In *Proceedings of the ACM SIGMOD Workshop on Databases and Social Networks*, pages 1–6. ACM.

FIGURES

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2241v1 | CC BY 4.0 Open Access | rec: 5 Jul 2016, publ: 5 Jul 2016

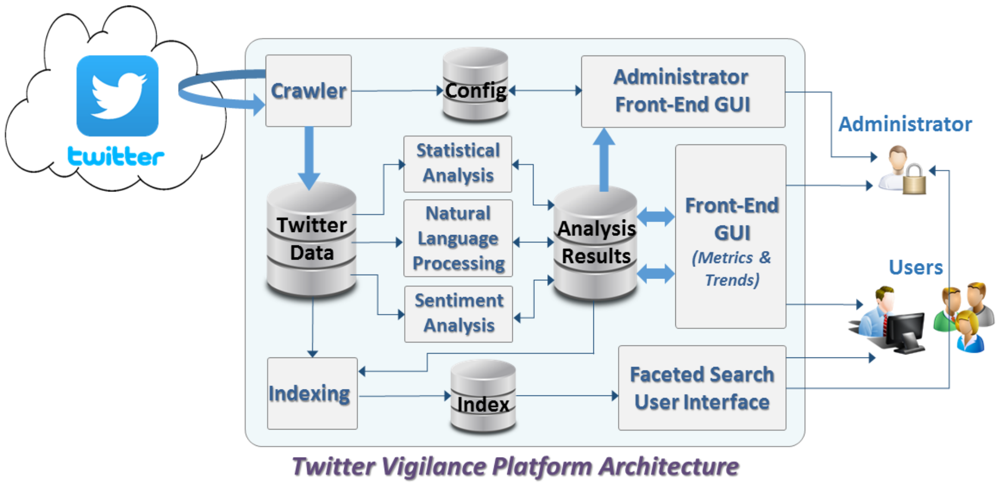

Figure 1. DISIT Twitter Vigilance Platform architecture

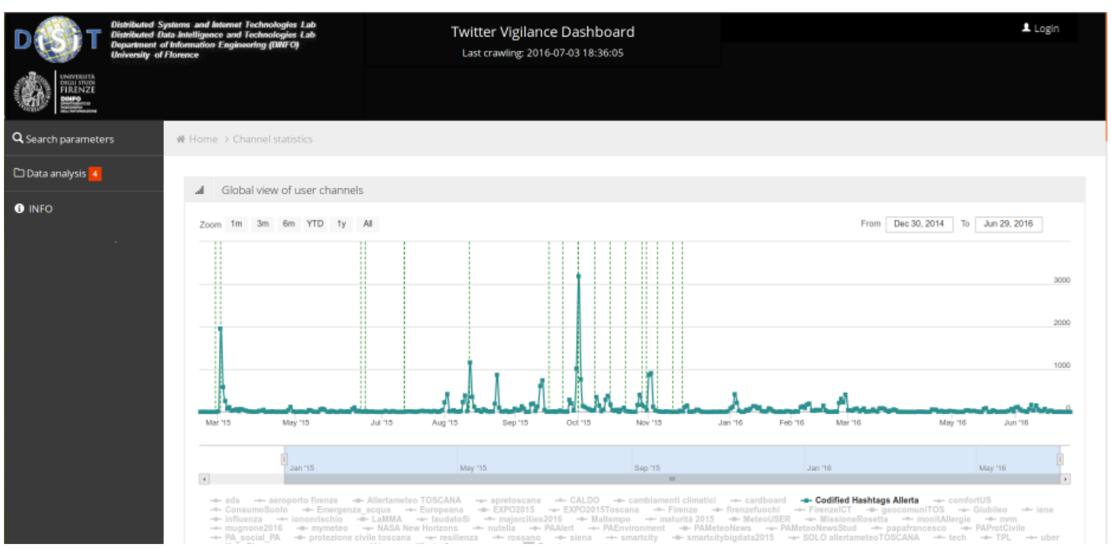

Figure 2. DISIT Twitter Vigilance interface: Codified Hashtags channel

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2241v1 | CC BY 4.0 Open Access | rec: 5 Jul 2016, publ: 5 Jul 2016

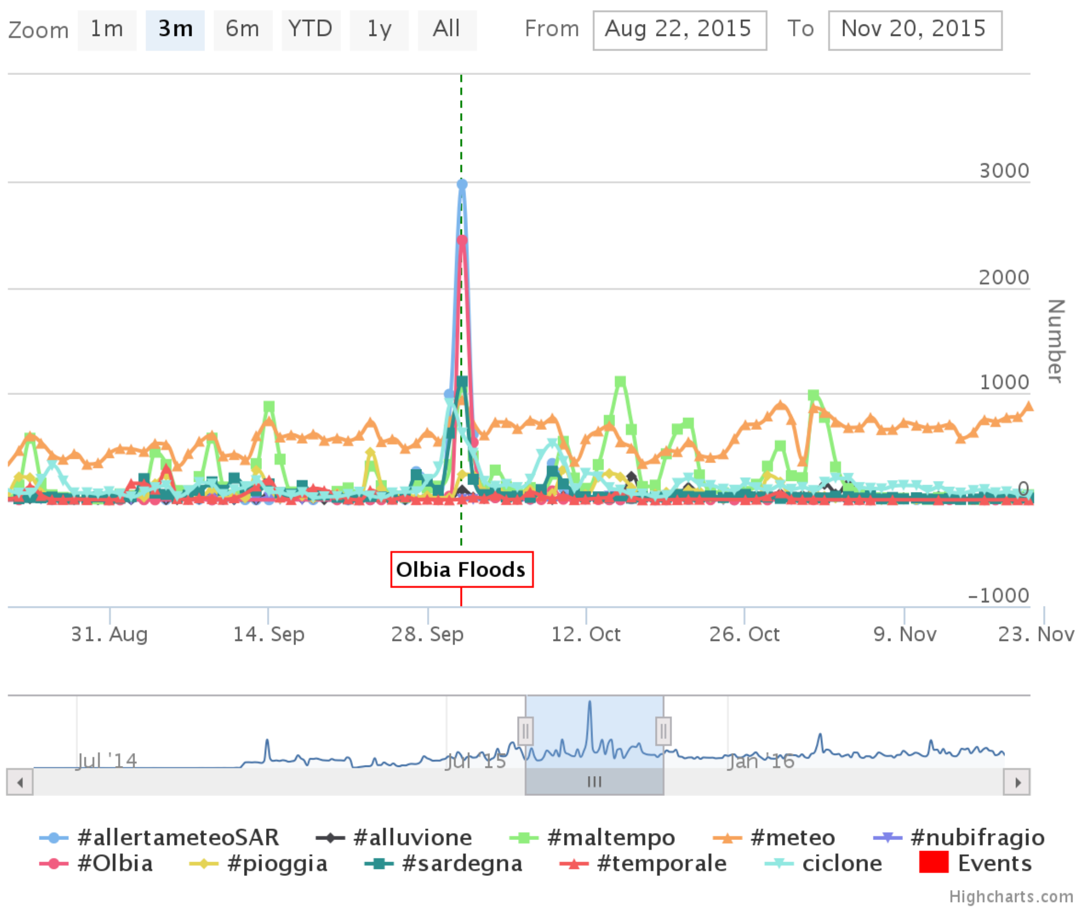

Figure 3. Olbia flood, October 1st, 2015

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2241v1 | CC BY 4.0 Open Access | rec: 5 Jul 2016, publ: 5 Jul 2016

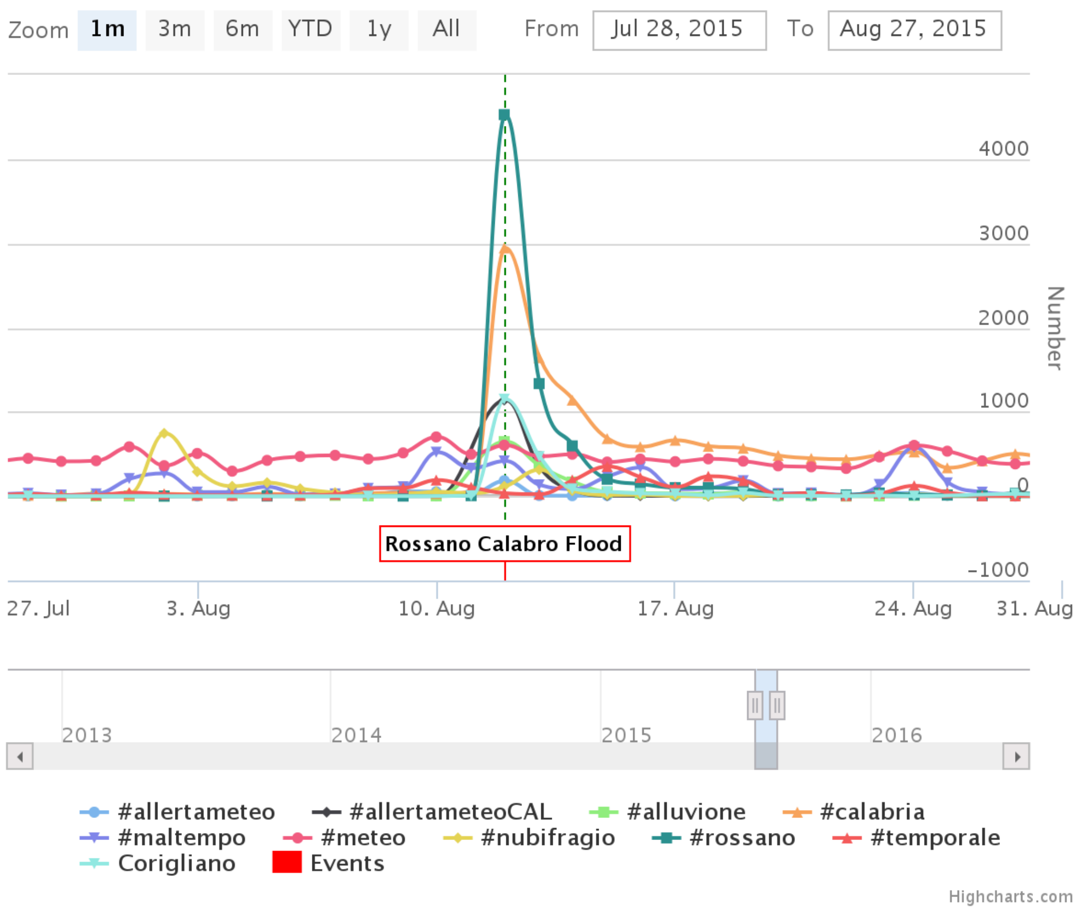

Figure 4. Rossano Calabro flood, August 12th 2015

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2241v1 | CC BY 4.0 Open Access | rec: 5 Jul 2016, publ: 5 Jul 2016

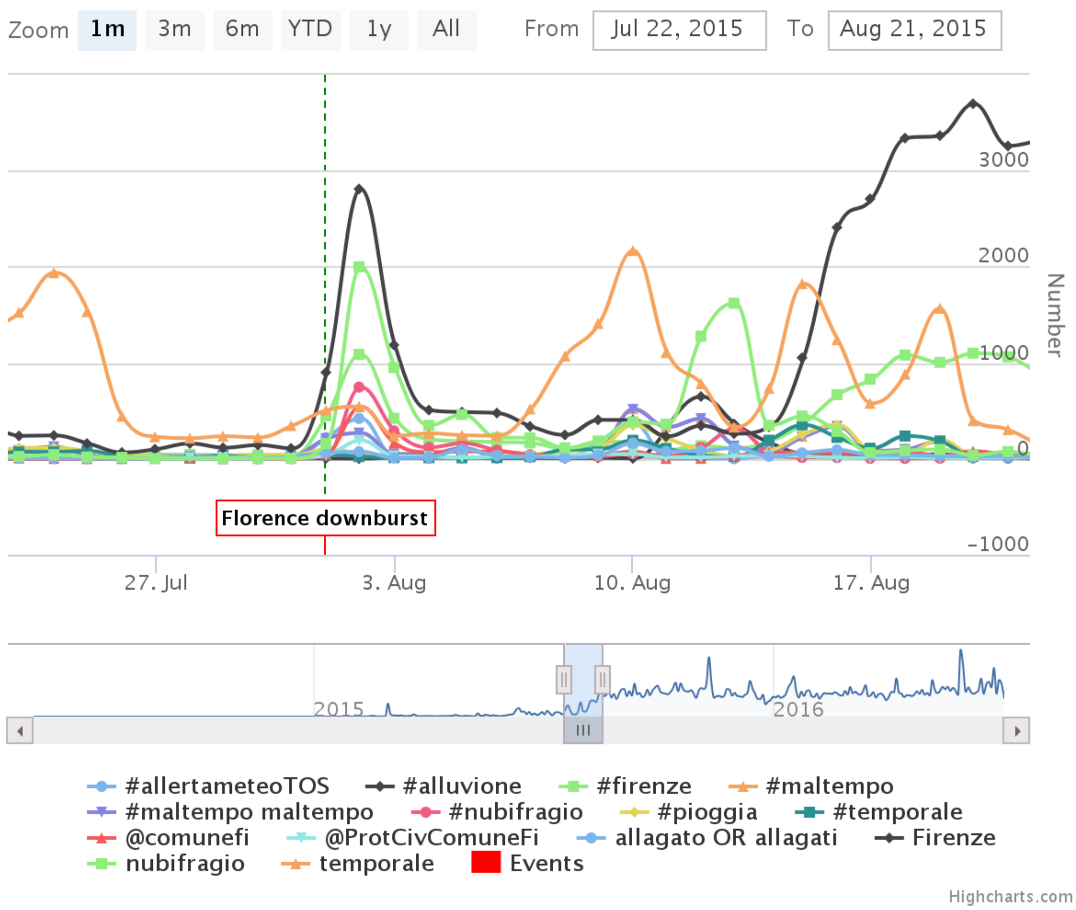

Figure 5. Florence downburst event, August 01st, 2015

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2241v1 | CC BY 4.0 Open Access | rec: 5 Jul 2016, publ: 5 Jul 2016