



Estimating Ground Shaking Regions with Social Media Propagation Trees

Marcelo Mendoza¹(✉), Bárbara Poblete², and Ignacio Valderrama²

¹ Millennium Institute for Foundational Research on Data,
Universidad Técnica Federico Santa María, Valparaíso, Chile
`marcelo.mendoza@usm.cl`

² Millennium Institute for Foundational Research on Data,
Department of Computer Science, Universidad de Chile, Santiago, Chile
`{bpoblete,ivalderr}@dcc.uchile.cl`

Abstract. The Mercalli scale of quake damages is based on perceived effects and it has a strong dependence on observers. Recently, we proposed a method for ground shaking intensity estimation based on lexical features extracted from tweets, showing good performance in terms of mean absolute error (MAE). One of the flaws of that method is the detection of the region of interest, i.e., the area of a country where the quake was felt. Our previous results showed enough recall in terms of municipality recovery but a poor performance in terms of accuracy. One of the reasons that help to explain this effect is the presence of data noise as many people comment or confirm a quake in areas where the event was unperceived. This happens because people get awareness of an event by watching news or by word-of-mouth propagation. To alleviate this problem in our earthquake detection system we study how propagation features behave in a region of interest estimation task. The intuition behind our study is that the patterns that characterize a word-of-mouth propagation differ from the patterns that characterize a perceived event. If this intuition is true, we expect to separate both kinds of propagation modes. We do this by computing a number of features to represent propagation trees. Then, we trained a learning algorithm using our features in the specific task of region of interest estimation. Our results show that propagation features behave well in this task, outperforming lexical features in terms of accuracy.

Keywords: Social networks · Disaster management ·
Mercalli intensity · Social media during emergencies · Propagation trees

1 Introduction

Richter and Mercalli scales measure the level of impact of an earthquake in a given region. Whilst Richter measures the energy released during an earthquake, Mercalli represents the level of damages produced during an earthquake. Both scales are related but may differ due to several factors as the quality of the

buildings, the type of ground where the quake happens or the depth of the epicenter, i.e., the distance of the epicenter to the ground surface.

Mercalli reports are prepared by observers who record the effects of an earthquake on humans and man-made structures. However, these reports may be released even hours or days after an earthquake, as the strong dependence on local observers makes difficult to provide fresh information. Recently we proposed a method for fast estimation of Mercalli intensities using social media [9]. Our method is based on the observation of Twitter and it computes lexical features on a set of messages related to the event. We showed that there are lexical features that are useful for Mercalli intensity estimation. However, one of the difficulties found during the study was the estimation of the ground shaking region. As many people get awareness of an event watching news or by word-of-mouth, these comments are mixed with comments of observers who are placed in the region of interest, introducing noise during the region of interest estimation step of our method. With a good recall but a poor accuracy in the region estimation step, our method shows enough room for improvement.

In this paper we study how propagation features can be used to mitigate the effect of noise during the region of interest estimation process. We extend our method providing better features for ground shaking region estimation. To do this we compute eight propagation features showing how useful they are to alleviate the effect of noise in our system.

Main Contribution of the Paper: In this paper we address the problem of ground shaking region estimation using social media propagation features. Our method starts extracting propagation trees from propagation graphs, detecting seeds and measuring a number of features that characterize spreading patterns. To the best of our knowledge, this is the first work that addresses the problem of ground shaking region estimation using propagation network features. Our intuition indicates that there are measurable differences in propagation patterns between perceived and unperceived events. We sustain this intuition in our previous findings on rumor detection [8]. The intuition behind our study is that the patterns that characterize a word-of-mouth propagation differ from the patterns that characterize a perceived event. If this intuition is true, we expect to separate both kinds of propagation modes. To study this hypothesis we compute a number of features to represent propagation trees. We will show that our features are useful for ground shaking region estimation, giving support to our hypothesis.

This paper is organized as follows. Related work is discussed in Sect. 2. Preliminaries are discussed in Sect. 3. Ground shaking region estimation based on propagation features is introduced in Sect. 4. Experiments are discussed in Sect. 5. Finally, we conclude in Sect. 6.

2 Related Work

The relation between physical events and its correspondence in Twitter has been an active research area during the last years [6]. These efforts have shown

interesting results. For instance, a research found that during the Tohoku earthquake in 2011 there were a number of high correlations between the amount of tweets and the intensity of the disaster in some locations [4]. Recently, Poblete *et al.* [10] provided a system for the early detection of earthquakes using social media features. The system dubbed “Twicalli”¹ detects worldwide earthquakes in real time, illustrating the consonance between physical events and social media trends.

There are more quake alert systems based on social media around the world. Systems as the ones placed in Australia [11] or Italy [1] use burst detection algorithms to report earthquakes, where a burst is defined as a large number of occurrences of tweets within a short time window [13]. In addition to the detection of an event, the estimation of the intensity of a quake has also aroused interest. Sakaki *et al.* [12] showed that it is possible to estimate the epicenter of an earthquake event using only information recovered from Twitter as tweets counts and tweets rates. Burks *et al.* [2] proposed an approach to estimate the Mercalli intensity of an earthquake performing a cross match between seismological recording stations and tweets that mention the word ‘earthquake’. Computing a number of lexical features in each areal disc centered around each seismograph, the authors studied the correlation of these features with the Mercalli intensity. Using linear regression models, the authors showed good results in terms of accuracy for Mercalli intensity estimation tasks.

The estimation of the maximum intensity of an earthquake using Twitter was studied by Cresci *et al.* [3]. Using linear regression models over a huge collection of aggregated features (45 features were tested in that proposal), the authors showed that Twitter has enough predictive power to infer the maximum intensity of an earthquake in the Mercalli scale. Recently, we showed that it is possible to provide an early estimation of the maximum intensity of an earthquake (just 30 min after the event) using only 12 lexical features, performing well in this specific task [9]. However, one of the limitations of that work relies on the poor accuracy achieved during the estimation of the ground shaking region. As many people get awareness of an event watching news or by word-of-mouth, these comments are mixed with comments of observers who are placed in the region of interest, introducing noise during the estimation process.

To alleviate the effect of noise during the ground shaking estimation process, we study the effectiveness of eight features that characterize the propagation of the event across the network. Propagation features has succeeded in predictive tasks as rumor detection [8] and research output forecasting [7]. The intuition behind this study is to check if there is a consonance between the impact of a perceived event and propagation traces. If this intuition is true, we expect to measure and use the correspondence of the event in the network improving the accuracy on the region of interest estimation task.

The estimation of the ground shaking region of an earthquake using social media has gained attention in last years. Systems based on crowd-sourcing tools²

¹ <http://www.twicalli.cl>.

² “Did you feel it?” website located at <https://earthquake.usgs.gov/data/dyfi/>.

or based on geolocated tweets as TwiFelt [5], have revealed the interest of government agencies as the US Geological Survey (USGS) on the use of social media for these tasks. In this paper we will show that propagation features are key event descriptors of earthquakes to address this challenging task.

3 Preliminaries

We proposed a method for the early estimation of the intensity of an earthquake in the Mercalli scale [9]. In that method, we used information gathered from Twitter. Our method works in a tandem with Twicalli [10], the system for detection of earthquakes based on Twitter. Once an earthquake is detected by Twicalli, the event is characterized at municipality level, the finer level of geolocation considered in our system. Then we conduct a regression process to infer the region of interest of a given earthquake. Finally, our method takes the collection of point estimates to infer the maximum intensity in the Mercalli scale for a given quake.

In our system, posts are collected to extract features of the event that characterize the social perception of the earthquake. Each perceived event is characterized at a level of aggregation that describes the perception of the earthquake in a municipality. For each municipality batch, a set of features is computed to describe the earthquake.

Municipality batches are built as follows. After each earthquake, a set of tweets that matches the keywords “quake”, “earthquake” or “seismic” are retrieved from Twitter. The time considered to collect the data is a parameter of our system, with a window length of 30 min by default. Shorter windows can be considered but at the cost of less accurate Mercalli predictions. Tweets that are mapped to municipalities are aggregated into municipality batches.

We map tweets to municipalities using the user location field. We were forced to use this field as only a very small fraction of the tweets in our country is geolocated. In order to geolocate tweets we use the following steps: (1) if available, we extract the exact GPS coordinates from the tweet’s *location* field, (2) if the location field was not provided by the user in their tweet, we then process the tweet’s textual content. This is, we analyze the message’s text (e.g., “*Earthquake in Valparaiso!!!*”) to extract, using a fuzzy string matching procedure, any location mentions, or (3) if all else fails, we apply the same procedure as in (2) but this time on the text provided by the user in their profile information.

Our method starts detecting the region of interest from where municipality data batches will be used to infer Mercalli intensities. This step of the method separate municipalities into two classes. We do this using a 0/1 classifier trained over municipality-seismic data batches pairs. These data batches were labeled according to the actual Mercalli intensity reported into two disjoint classes. The 0 class represents an earthquake that was not perceived (not reported in the Mercalli scale) and the 1 class represents an earthquake that was effectively perceived by people with an intensity value in the Mercalli scale. Each data batch is represented by a vector of features. Once the 0/1 classifier was trained,

our method is ready to detect the region of interest on new earthquakes at county level.

After the estimation of the region of interest, our method estimate the maximum intensity of the event. Further details of this process are provided in Mendoza *et al.* [9].

4 Region Estimation Based on Propagation Features

4.1 Features

Eleven lexical features are considered at this level of aggregation as is shown in Table 1. In addition, eight propagation features are computed for this task. We also consider the inclusion of the municipality population as a feature. These features are calculated in each municipality data batch, characterizing the set of tweets mapped to each specific county for a given seism.

Table 1. Features used in our study.

	Feature	Description
LEXICAL	NUMBER OF TWEETS	Number of tweets in the data batch
	TWEETS NORM	Fraction of tweets over county population
	AVERAGE WORDS	Average length of tweets in number of words
	AVERAGE LENGTH	Average length of tweets in number of chars
	QUESTION MARKS	Fraction of tweets with question marks
	EXCLAMATION MARKS	Fraction of tweets with exclamation marks
	UPPER WORDS	Fraction of tweets with uppercase words
	HASHTAG SYMBOLS	Fraction of tweets containing the hashtag symbol
	MENTION SYMBOLS	Fraction of tweets containing the mention symbol
	RT SYMBOLS	Fraction of tweets containing the “RT” symbol
	CONTAINS EARTHQUAKE	Fraction of tweets containing the word <i>earthquake</i>
	POPULATION	Number of inhabitants in the county
PROPAGATION	NUMBER OF SEEDS	Number of seeds in the data batch
	NUMBER OF TREES	Number of trees in the data batch
	AVG USERS IN TREES	Average number of users across trees
	AVG USERS IN TREES (NI)	Average number of users across not isolated seeds
	BIGGEST TREE SIZE (U)	Size of the biggest tree in number of users
	BIGGEST TREE SIZE (I)	Size of the biggest tree in number of interactions
	AVG TREE SIZE	Average number of interactions across trees
	AVG TREE SIZE (NI)	Average number of interactions across ni seeds

To compute propagation features we need to process the propagation graph recovered for each event. The propagation graph is a graph of message sharing and replaying. In a propagation graph, each node represents a post. Each post can be read by the followers of the post owner. If a follower decides to share (to retweet in Twitter jargon), reply or mention a post, a new node is recorded in the graph, linking both nodes with an arc. Original posts (posts that are

not retweets, replies or mentions) are seeds of claims. If a seed post is shared in the network, the propagation graph records an information cascade. As each interaction with the original post produces a new message, the cascade is cycle-free and it compounds a tree.

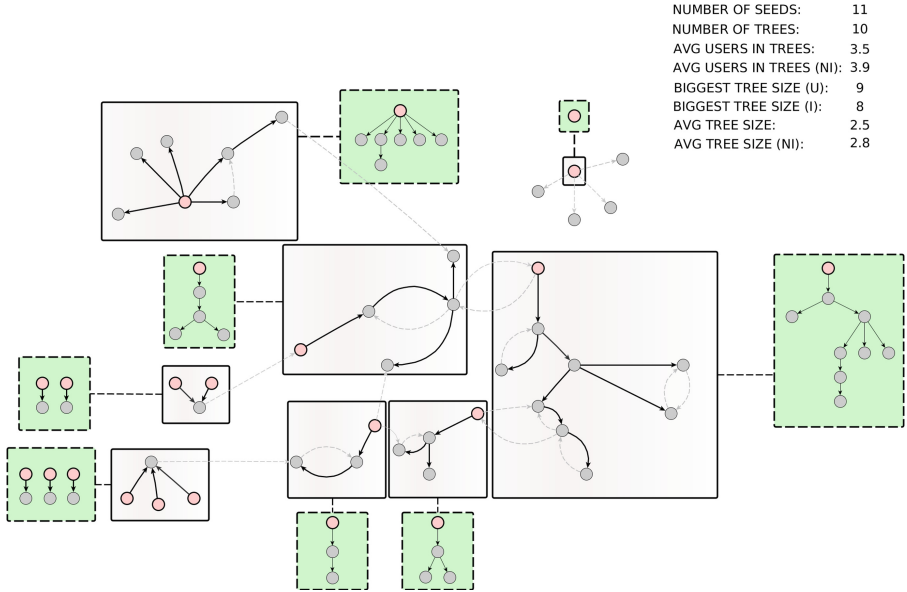


Fig. 1. How propagation features are computed. Graphs inside grey boxes represent the original propagation graph. Each node represent a message. Black edges represent RTs or mention posts. Grey edges represent inactive following links. Inferred trees are depicted in green boxes. Seeds are depicted with pink nodes. At the top of the figure we show the eight propagation features that correspond to this example. (Color figure online)

We show in Fig. 1 how propagation features are computed. Black edges show message sharing between posts. Gray edges show followers/followees relationships that do now share a message during the claim. Note that the propagation graph is a subgraph of the social network graph. Each propagation tree is boxed by a grey shaded rectangle. Inferred propagation trees are bounded in green boxes. The example shows eleven seeds (shaded in pink) and ten trees (note that the example shows an isolated seed).

4.2 Estimation of a Region of Interest

The next stage of our approach is estimating which municipalities were affected by the earthquake. We refer to these municipalities as the *region of interest* or

ground shaking region of an earthquake. To estimate the geographical subdivisions that were affected by the seismic event, we use a supervised classification model. This model separates municipalities into two classes: *unaffected* by the earthquake and *affected* by the earthquake.

To create this model we used a 0/1 classification algorithm, which we trained using municipality-level data modeled as feature vectors (using the features shown in Table 1). The labels that we used for each municipality were class “0” if the earthquake was not perceived by the population (i.e., the municipality had no official Mercalli intensity value associated to it), and class “1” if the earthquake was perceived by the population (i.e., the municipality had an official Mercalli value associated to it). The Mercalli intensity values that we used to label the municipality-level data corresponded to values in official earthquake reports. More details on the technical and empirical aspects of the model creation are presented in Sect. 5.

5 Experiments

5.1 Dataset

A collection of 825310 tweets was retrieved from Twitter. These tweets were collected using keywords as “quake”, “earthquake” and “seismic movement” (in Spanish). The collection comprises a year and a half of Twitter data, matching the keywords during 2016 and the first semester of 2017. From these tweets, only 2200 include the geolocation field, representing only the 0.26% of the data. The collection was posted by 309749 users where 207015 records a location field in their profiles, representing the 66.8% of the users recorded in the data. From the set of 207015 users with user location in our dataset, 57546 matched Chile in the country field. Then we used approximate matching to associate this field with a Chilean municipality using *Fuzzy wuzzy*³. Using an 80% of fuzzy confidence level, a total of 41885 Chilean users were mapped to Chilean counties. These users record in the dataset a total of 190249 tweets mapped to the 345 different counties in Chile.

We used data collected by the National Seismological Center of Chile, comprising 331 records of earthquakes in Chile during the observation period, ranging magnitudes in Richter from 2.2Mw to 7.6Mw. The cross match between our tweet collection and the Mercalli earthquake records was conducted over the municipality field. Only municipality batches that record tweets until 30 min after an earthquake were studied, accounting for a total of 6790 municipality-Mercalli pairs with Twitter activity. A total amount of 6548 municipality batches unmatched a Mercalli report, indicating the presence of tweets that mention earthquake keywords in counties where it was unperceived. In summary, our Twitter-Mercalli dataset comprises 331 earthquakes with 187317 tweets

³ Fuzzy wuzzy is a Python string matching library that uses the Levenshtein Distance to calculate differences between string sequences. It is available in: <https://github.com/seatgeek/fuzzywuzzy>.

distributed over 345 Chilean counties during 18 months of Twitter activity, with county-earthquake pairs separated into 6790/6548 perceived/not-perceived earthquake data batches.

From the total amount of 331 earthquakes, 264 were selected for training and exploratory issues, reserving the remaining 68 earthquakes for testing and validation tasks, representing a training/testing split of 80/20%. The training/testing splitting process was conducted using stratified sampling over earthquakes according to each Mercalli level. Training/testing proportions of instances according to the maximum Mercalli intensity report of each earthquake are shown in Table 2. Data and its description are available at <https://doi.org/10.6084/m9.figshare.c.4206689>.

Table 2. Training/testing instance partitions according to the maximum Mercalli intensity of each quake

Partition	II	III	IV	V	VI	VII
Training	11	105	103	39	4	2
Testing	3	26	26	10	2	1
Overall	14	131	129	49	6	3

5.2 Exploratory Analysis

We first performed a data exploration process to analyze the relationship between municipality-level features and Mercalli values. We studied the existence of correlations, which are shown in Table 3.

Table 3 shows correlations in terms of the Spearman coefficient, as the variables studied are skew. All the coefficients found are statistically significant with p-values equal to $2.2e-16$. The correlation between propagation features is strong. Note that the correlation between `MERCALLI` and the other variables is not as strong. The table shows a strong correlation between size features. Interestingly, the correlation between `NUMBER OF SEEDS` and `NUMBER OF TREES` is not as strong, showing that there are a number of isolated seeds that do not achieve a spread in the network.

A strong correlation was also detected between some lexical features as `NUMBER OF TWEETS` and `TWEETS NORM`, `AVERAGE WORDS` and `AVERAGE LENGTH` and `MENTION SYMBOLS` and `RT SYMBOLS`. In general, the correlation between lexical features was weak, except for the indicated cases. A more detailed analysis of the correlation between lexical features can be checked in [9].

5.3 Estimating the Region of Interest

Training/testing municipality data batches accounts for 10491/2847 instances at municipality level. To study the problem of perceived/not-perceived earthquakes

Table 5. Testing accuracy by class using lexical features

Class	FP Rate	Precision	Recall	F-measure	ROC area
0 (unreported)	0.184	0.765	0.517	0.617	0.666
1 (reported)	0.483	0.593	0.816	0.687	0.666
Weighted avg.	0.323	0.685	0.655	0.649	0.666

Tables 4 and 5 show a good performance in terms of recall for the class of interest but a poor performance in terms of precision. Accordingly, the F-measure has a performance around 68% for the class of interest on testing data, achieving a ROC value around 0.666.

Table 6. Training accuracy by class using propagation features

Class	FP Rate	Precision	Recall	F-measure	ROC area
0 (unreported)	0.195	0.763	0.683	0.721	0.744
1 (reported)	0.317	0.734	0.805	0.768	0.744
Weighted avg.	0.259	0.748	0.746	0.745	0.744

Table 7. Testing accuracy by class using propagation features

Class	FP Rate	Precision	Recall	F-measure	ROC area
0 (unreported)	0.207	0.771	0.603	0.677	0.698
1 (reported)	0.397	0.633	0.793	0.704	0.698
Weighted avg.	0.295	0.707	0.691	0.690	0.698

The classifier based on propagation features performs better than the one based on lexical features, as it is shown in Tables 6 and 7. These results show that the classifier achieves a precision around 63% on testing data and a F-measure over the 70%, as well as a ROC value near 0.7. These improvements show that the use of propagation features is helpful for this task.

When lexical and propagation features are combined in a single classifier, the results get worse. As Tables 8 and 9 show, the 0/1 classifier increases the presence of false positives, and as a consequence, it decreases its performance in terms of precision and F-measure. These results show that it is better to address this specific task using propagation features, confirming the intuition behind the consonance between propagation patterns and the physical coverage of earthquakes.

The results show that each region of interest is over-estimated as the low precision for class 1 shows but achieving a good coverage of the actual region as

Table 8. Training accuracy by class using lexical and propagation features

Class	FP Rate	Precision	Recall	F-measure	ROC area
0 (unreported)	0.187	0.744	0.591	0.659	0.702
1 (reported)	0.409	0.684	0.813	0.743	0.702
Weighted avg.	0.303	0.713	0.707	0.703	0.702

Table 9. Testing accuracy by class using lexical and propagation features

Class	FP Rate	Precision	Recall	F-measure	ROC area
0 (unreported)	0.180	0.769	0.519	0.619	0.669
1 (reported)	0.481	0.595	0.820	0.690	0.669
Weighted avg.	0.320	0.689	0.658	0.652	0.669

its high recall shows. To better understand how the 0/1 classifier behaves, we disaggregate matching/mismatching testing instances according to the actual level of Mercalli intensity.

Table 10. Matching/mismatching instances according to the actual Mercalli intensity using lexical features

Actual	Predicted	-	I	II	III	IV	V	VI	VII
0	0	790	-	-	-	-	-	-	-
0	1	737	-	-	-	-	-	-	-
1	0	-	66	85	62	25	5	-	-
1	1	-	130	234	351	198	65	95	4
Instances		1527	196	319	413	223	70	95	4
Error rate		0.48	0.33	0.26	0.15	0.11	0.07	-	-

As Tables 10, 11 and 12 show, the false negative rate is very low, and as long as the intensity of the earthquake increases, the error rate decreases. High intensity earthquakes (V to up) show an almost perfect performance. The thick part of this error occurs in low intensity earthquakes (III to down), which is natural for this kind of phenomena as in this part of the Mercalli scale many people do not recognize the event as an earthquake, being felt only under very favorable conditions (for instance, on upper floors of buildings). When the classifier based on propagation features is used for this task, the error in level IV events decreases and it achieves a perfect performance in level V earthquakes. The global error rate using propagation features goes to 0.39 points, almost 10 points below the error rate achieved using lexical features. When both types of features are used,

Table 11. Matching/mismatching instances according to the actual Mercalli intensity using propagation features

Actual	Predicted	-	I	II	III	IV	V	VI	VII
0	0	921	-	-	-	-	-	-	-
0	1	606	-	-	-	-	-	-	-
1	0	-	83	85	65	40	-	-	-
1	1	-	113	234	348	183	70	95	4
Instances		1527	196	319	413	223	70	95	4
Error rate		0.39	0.42	0.26	0.15	0.17	-	-	-

Table 12. Matching/mismatching instances according to the actual Mercalli intensity using lexical and propagation features

Actual	Predicted	-	I	II	III	IV	V	VI	VII
0	0	792	-	-	-	-	-	-	-
0	1	735	-	-	-	-	-	-	-
1	0	-	71	88	50	25	4	-	-
1	1	-	125	231	363	198	66	95	4
Instances		1527	196	319	413	223	70	95	4
Error rate		0.48	0.36	0.27	0.12	0.11	0.05	-	-

as it is shown in Table 12, the performance get worse, confirming that the use of lexical features in this specific task introduces noise during the estimation process.

6 Conclusion

In this paper we have studied the performance of propagation features in a ground shaking region estimation task. Our results show that the use of propagation features is useful for this task outperforming classifiers based on lexical features. The intuition behind this finding sustains that lexical features are unable to hand noise during the inference process, as many observers comment unperceived events getting awareness of earthquakes watching news of by word-of-mouth propagation effects. The use of propagation features allows building robust classifiers for ground shaking region estimation tasks, corroborating the presence of a consonance between how actual events spread in social media and how physical events are perceived in the physical world.

Currently, we are extending our method to work with more features. The inclusion of time-based features helps to characterize the tweet stream (e.g. tweet interval rate), a valuable source of information for earthquake detection task. We think that these features will also be helpful in the elaboration of spatial intensity reports.

At last but not least, the design of a system for early tracking of earthquake damages is the next step of this project. How to efficiently use our method to provide spatial real-time damage reports is one of our most challenging tasks in the near future. The pursuit of this goal involves efforts in data integration and visualization, among other challenging tasks for our group.

Acknowledgements. This work was supported by the Millennium Institute for Foundational Research on Data. M. Mendoza was also funded by Conicyt PIA/Basal FB0821.

References

1. Avvenuti, M., Cresci, S., Marchetti, A., Meletti, C., Tesconi, M.: EARS (earthquake alert and report system): a real time decision support system for earthquake crisis management. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1749–1758. ACM (2014)
2. Burks, L., Miller, M., Zadeh, R.: Rapid estimate of ground shaking intensity by combining simple earthquake characteristics with tweets. In: 10th US National Conference on Earthquake Engineering Frontiers of Earthquake Engineering, Anchorage, AK, USA, 21–25 July 2014 (2014)
3. Cresci, S., La Polla, M., Marchetti, A., Meletti, C., Tesconi, M.: Towards a timely prediction of earthquake intensity with social media. IIT TR-12/2014. Technical report. IIT: Istituto di Informatica e Telematica, CNR (2014)
4. Doan, S., Vo, B.-K.H., Collier, N.: An analysis of Twitter messages in the 2011 Tohoku earthquake. In: Kostkova, P., Szomszor, M., Fowler, D. (eds.) eHealth 2011. LNCS, vol. 91, pp. 58–66. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29262-0_8
5. D’Auria, L., Convertito, V.: Real-time mapping of earthquake perception areas in the Italian region from Twitter streams analysis. In: D’Amico, S. (ed.) Earthquakes and Their Impact on Society. SNH, pp. 619–630. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-21753-6_26
6. Earle, P., et al.: OMG earthquake! Can Twitter improve earthquake response? *Seism. Res. Lett.* **81**(2), 246–251 (2010)
7. Guevara, M.R., Hartmann, D., Aristarán, M., Mendoza, M., Hidalgo, C.A.: The research space: using career paths to predict the evolution of the research output of individuals, institutions, and nations. *Scientometrics* **109**(3), 1695–1709 (2016)
8. Mendoza, M., Poblete, B., Castillo, C.: Twitter under crisis: can we trust what we RT? In: Proceedings of the First Workshop on Social Media Analytics, SOMA 2010, pp. 71–79. ACM (2010)
9. Mendoza, M., Poblete, B., Valderrama, I.: Early tracking of people’s reaction in Twitter for fast reporting of damages in the mercalli scale. In: Meiselwitz, G. (ed.) SCSM 2018. LNCS, vol. 10914, pp. 247–257. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91485-5_19
10. Poblete, B., Guzman, J., Flores, J.A.M., Tobar, F.A.: Robust detection of extreme events using Twitter: worldwide earthquake monitoring. *IEEE Trans. Multimed.* **20**(10), 2551–2561 (2018)
11. Robinson, B., Power, R., Cameron, M.: A sensitive twitter earthquake detector. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 999–1002. ACM (2013)

12. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th International Conference on World Wide Web, WWW 2010, pp. 851–860. ACM (2010)
13. Zhang, X., Shasha, D.: Better burst detection. In: 2006 Proceedings of the 22nd International Conference on Data Engineering, ICDE 2006, pp. 146–146. IEEE (2006)