ISSN 2039-7941
Anno 2013_Numero 254

rapporti tecnici INGV

TwiFelt: real-time mapping of earthquake perception areas through the analysis of Twitter streams

254

Istituto Nazionale di
Geofisica e Vulcanologia

ISSN 2039-7941
Anno 2013_Numero 254

rapporti tecnici INGV

TWIFELT: REAL-TIME MAPPING OF EARTHQUAKE PERCEPTION AREAS THROUGH THE ANALYSIS OF TWITTER STREAMS

Luca D'Auria, Flora Giudicepietro

INGV (Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Napoli - Osservatorio Vesuviano)

254

Table of contents

Introduction 5
1. Twitter stream parsing 5
2. Database structure 7
3. Data mining 7
4. Web interfaces 10
5. Map rendering 12
6. Case histories 13
7. Future developments 16
Acknowledgements 16
References 16
Appendix: Useful links 17

Introduction

Twitter is one of the most used social networks and its specific features makes it well suited for the real-time analysis of geographic trends of a specific topic. Earle et al. (2011) have shown how the analysis of Twitter streams can provide a useful tool for the early detection of earthquakes at a global scale. They proved that data mining of social networks could provide useful information in Seismology. Here we present a software system named TwiFelt, aimed at providing real-time earthquake perception maps from the analysis of Twitter streams. The system is based on the collection of geotagged tweets (i.e. tweets having a geographic reference) containing selected keywords, its statistical interpretation and its interactive graphical representation. Figure 1 represents a schematic overview of the system. The most important parts of TwiFelt are: a Twitter stream parser (see sec.1), a database (see sec.2) and an interactive web interface (sec.4). The web interface provides maps showing, using a shaded overlay, an indication of the area where an earthquake has been felt.

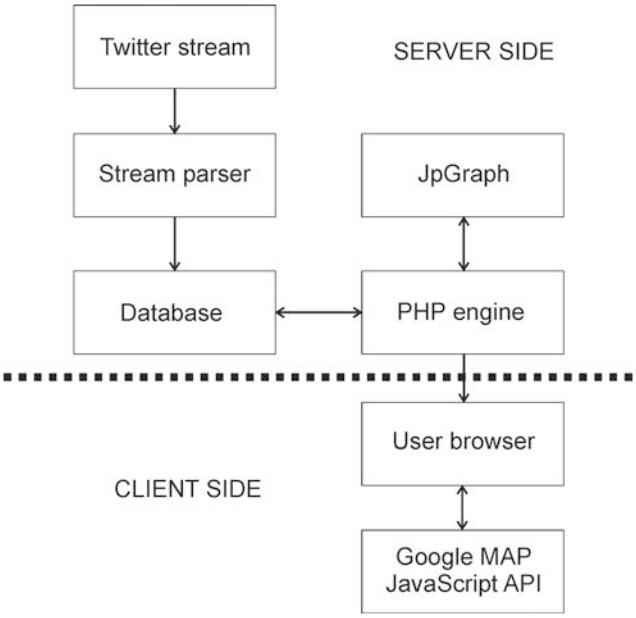


Figure 1. Schematic overview of the TwiFelt architecture.

In the following we describe the system components, showing also some preliminary results. Since most of the references to the mentioned technologies are currently only available online, the relevant links are listed in a separate Appendix.

1. Twitter stream parsing

Twitter is one of the most used social networking and microblogging systems. It is based on short (140 characters) messages called **tweets**. The Twitter platform API (Application Programming Interface) provides access to tweets in three different ways: Search, REST and Streaming. The Search interface allows retrieving recent tweets (not older than about a week) matching a specific query. The REST interface can be used to access specific elements such as the timeline of a user. The **Streaming** interface gives access in real-time to tweets matching a specific query. It has been designed to give low latency times to the global stream of tweets and for this reason is the best suited for tracking spatial and temporal trends. There are three different stream types: User (following a specific user), Site (following simultaneously different users) and Public (accessing all the public tweets). **Public streams** could, in principle, to access all the public tweets (i.e. the tweets from users having a public account). To restrict the search to a subset of tweets, the API allows the use of various filters. TwiFelt uses the "**track**" filter that allows the selection of tweets matching specific keywords. It searches for tweets containing keywords as: “terremoto” and “scossa” (which in Italian mean respectively earthquake and shock) as well as their English counterparts as: “earthquake”. These keywords are the most commonly used by Twitter users in their tweets when they believe they have felt an earthquake, using sentences like:

- “*C'è appena stato un* ***terremoto***” (There's just been an earthquake)
- “*Ho sentito una* ***scossa*** *davvero forte*” (I felt a very strong shock)

As it will be discussed in sec.3 these keywords could also include false positive as:

- “*Dopo questa notizia mi sento* ***scossa***” (After this news I feel shocked)
- “*Stasera allo stadio sarà un* ***terremoto***” (Tonight at the stadium it will be like an earthquake)
- “*Ho sentito che c'è stato un* ***terremoto*** *in Giappone*” (I heard there has been an earthquake in Japan)

The download protocol of the Twitter Stream is HTTPS. Data can be downloaded from an HTTPS stream, using the GET method to select the keywords. The Stream API allows a basic http authentication. TwiFelt uses the GNU **wget** software to download the stream. wget is a command line utility available on different platforms. It can be easily embedded in scripts performing various tasks. For example TwiFelt attaches to the stream using a command line as:

```
wget --user=X --password=Y --no-check-certificate -O - https://stream.twitter.com/1/statuses/filter.json?track=terremoto,scossa
```

The previous command downloads a real-time stream of tweets containing one or both the keywords “terremoto” and “scossa” (case insensitive). The user and the password of a valid Twitter account are required to perform this task. This command line output can be redirected and processed by appropriate software. To ensure the continuous operation of the code, this command line is embedded within an infinite loop, which restarts it in the case of crash.

The output format of Twitter streams is JSON (JavaScript Object Notation). This open-standard format allows the definition of complex data structures and associative arrays. It is generally used for lightweight data-interchange between applications. JSON is an ASCII, human-readable format and consists of a single basic object definition:

```
{
        string1 : value1,
        string2: value2,
        ...
}
```

The values could also be nested objects or ordered arrays. Many open-source libraries are available for the parsing of the JSON streams. TwiFelt uses a C++ code implemented “ad hoc” for this purpose and based on the open-source library: **libjson**. libjson provides a C++ API for parsing and accessing JSON data structures. The C++ code developed to parse Twitter JSON streams performs various tasks. The most important are:

- Discarding useless tweets: having no geotagging, being a retweet (i.e. a forwarding of another tweet), located outside the Italian region (defined as longitude between 6° and 20° E and latitude between 35° and 48° N). This last check is needed because the word “terremoto” is the same also in Spanish. For this reason the majority of tweets containing this keyword comes from Latin America countries.
- Extracting from the tweet information relevant for the database structure (see sec.2)
- Insertion in the database

The code processes sequentially the incoming Twitter JSON stream data structures performing the aforementioned tasks. A detailed description of the tweets data structure can be found at https://dev.twitter.com/docs/platform-objects/tweets.

The background tweet rate is around 6000 per day. Of these an average of less than 0.01% is accepted and inserted in the database. Most of the tweets are discarded because of the lack of geotagging, making them useless for the purposes of TwiFelt. The geotagging of tweets can be of two types:

- **Point**: this is the geotagging provided by GPS devices (mobile phones, tablets) giving a point estimate of the user position when the tweet has ben posted. It is defined using the geoJSON format.

- **Place**: this kind of geotagging is attached to tweets associated (but not necessarily originating from!) to a given place. The place could be generically associated to a city or a region or specifically to a street address. The structure characterizing the geographic location is a bounding box. This is a georeferentiated polygon delimiting the area associated to the "place" tag. TwiFelt computes the barycentre of the polygon and uses it as geographic location of the tweet.

These two types of geotagging are kept distinguished in the database and in the web interface. The "Place" geotags are less reliable, but the experience has shown that they can contribute positively in the definition of earthquake perception area (see sec.6).

Together with the geotagging coordinates TwiFelt stores also the place name retrieved from the "name" tag that contains a short human-readable version of the city name (e.g. Naples, Modena, etc.)

2. Database structure

Once parsed and selected, the relevant tweets information are inserted in a relational database. This database has been implemented on a MySQL platform and consists of three tables:

- **tweets**: contains all the information about the parsed tweets
- **blacklist**: contains a list of blacklisted users
- **whitelist**: not yet used, reserved for future developments

Table 1. Structure of the **tweets** table. The first two shaded rows indicate the primary key.

Field	Type	Description
user	string	user name
date	datetime	date and time of the tweet
pcoord	boolean	type of geotagging (point/place)
lat	float	Latitude
lon	float	Longitude
place	string	Place name
msg	string	Tweet message

The structure of the **tweets** table is shown in table 1. The first two rows are the primary key. The blacklist table structure consists in a single column containing the user name. The **blacklist** is used to exclude from the analysis tweets coming from specific users. They fall mostly in two categories: amateur seismologists tweeting the occurrence of an earthquake and user prone to false positive (i.e. users using to often the keywords "*terremoto*" and "*scossa*"). Users are periodically checked, to identify those who should be blacklisted, through a SQL query like:

*select user, count(*) as n from tweets where user not in (select * from blacklist) group by user order by n desc;*

Notice how a subquery has been used to exclude users already in the blacklist. Users exceeding 10 false positives are inserted in the blacklist that currently contains about 30 elements.

The **whitelist** is reserved for future development. It will contain a list of trusted users whose tweets will have a higher weight in the mapping of the earthquake perception area.

3. Data mining

To exploit the spatial and temporal distribution of earthquake related tweets, we first need to consider the effect of the background "noise" which are tweets not related to earthquake perception. As discussed in sec.1 false positives could come from sentences in which the words "*terremoto*" (earthquake) and "*scossa*" (shock) are used metaphorically or they refer to an earthquake that has not been directly felt by the reporter. Because of the intrinsic feature of tweets (e.g. highly variable sentence structure, poor grammar) a semantic analysis, aimed at discriminating tweets, would be a difficult task, which we reserve for future implementations.

We have opted for a simpler approach based on statistical modelling of the dataset. In practice we define a background noise pattern using a simple distribution like:

$$N(lat, lon, time) = A(lat, lon)B(time) \text{, (eq. 1)}$$

where time is the number of seconds since the midnight of each day. The integration of the function N over a given area and a given time interval would return the expect number of "background" tweets.

The functions A and B have been retrieved by analysing the spatial and temporal distribution of the observed background tweets. Figure 2 shows the spatial distribution of "noisy" tweets (tweets not related to earthquake perception). They have been selected by manually removing from the complete tweet list those related to recognized events (see table 2). This dataset covers the period 05/11/2012-26/01/2013 and consists of 1430 tweets, 1014 of them being geotagged by a point and 416 by a polygon (see sec.1). In the same period the Twitter account @INGVterremoti has posted 485 earthquake notifications.

In figure 2 it is evident the concentration of tweets around major Italian urban areas (Milan, Rome, Naples) and around the Emilia region. This region was affected by a destructive earthquake sequence in May-June 2012 with the strongest events having M=5.9 on 20/05 and M=5.8 on 29/05. After this event, topics concerning earthquakes were still popular, even if they were not usually related to actual earthquake perception. For instance most of the tweets were related to controversies regarding the distribution of funds for the post-earthquake reconstruction. In figure 3 the tweet distribution is represented by a spatial density map, computed over 1x1° areas. This distribution has been used as an empirical representation of the A function in eq.1. In practice the function A is calculated numerically by using a bilinear interpolation over a regular grid of points spaced of 1° apart.

The temporal distribution (in local time) of tweets is shown in figure 4. The plot shows a marked temporal pattern, with two maxima, one between 9 and 11 and another between 19 and 23. The minimum in the tweet percentage is between 4 and 7. This distribution has been used as an empirical estimator of the function B in eq.1. This function too is calculated numerically by using a linear interpolation of the data shown in Fig.4.

The best way to estimate if a set of tweet is significantly above the background noise would be the statistical comparison of the observed distribution with the background noise distribution (i.e. using tests like χ^2, Kolmogorov-Smirnov etc.). The development of procedures aimed at this goal is under development. Currently TwiFelt uses a heuristic simplified approach in which a weight W_i is assigned to each event:

$$W_i = N^{-\frac{1}{2}}(lat_i, lon_i, time_i) \text{. (eq. 2)}$$

These weights can be used by the map processor to realize shaded maps (see sec.5). From eq.2 it is evident that the weight is inversely proportional to the number of events expected. The power ½ has been chosen by a trial and error approach. This value gives the best delimitation of perception areas mitigating, at the same time, the effect of isolated spurious spikes.

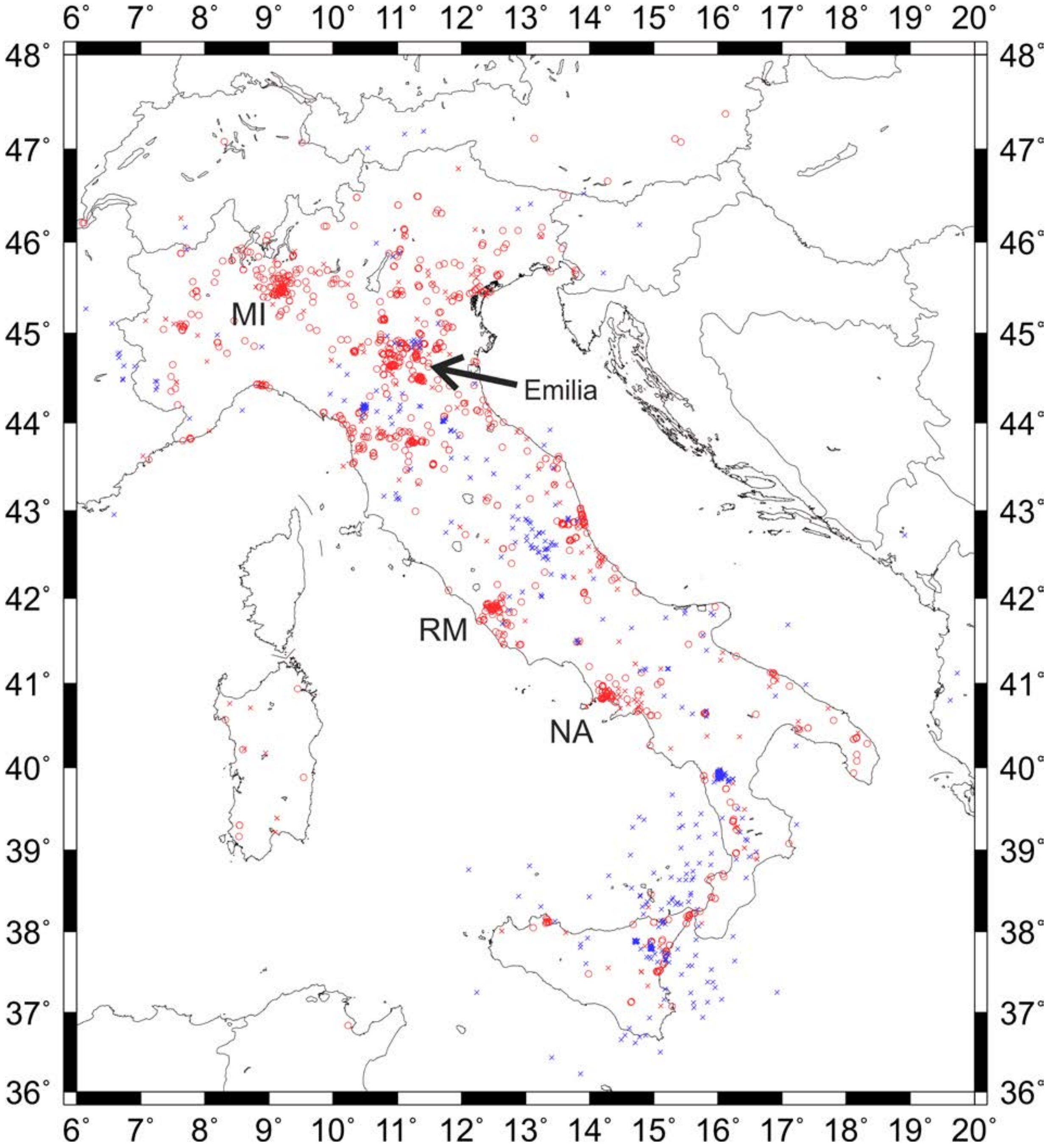


Figure 2. Distribution of "noisy" tweets in the period 05/11/2012 - 26/01/2013. Red circles indicate tweets geotagged by points, while red crosses those geotagged by a place. Earthquake epicentres (provided by @INGVterremoti Twitter account) are indicated by blue crosses.

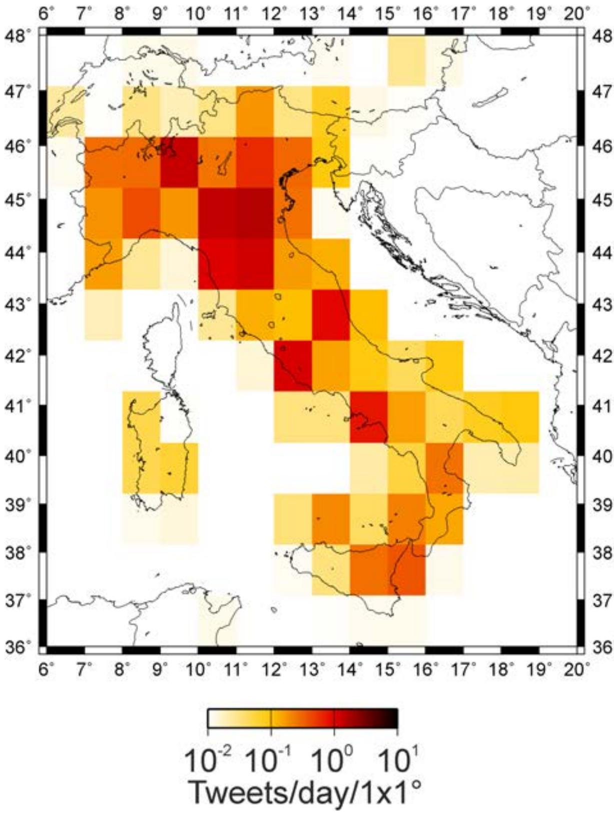


Figure 3. Density distribution of "noisy" tweets in the period 05/11/2012 - 26/01/2013. The plot represents the average daily number of "noisy" tweets for areas of 1x1°.

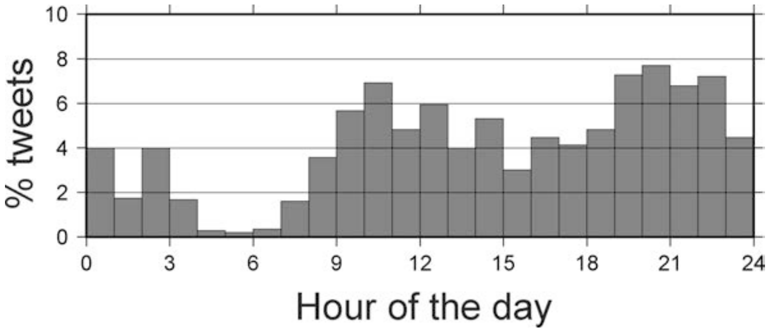


Figure 4. Temporal distribution of "noisy" tweets in the period 05/11/2012 - 26/01/2013. The histogram represents the percentage of tweets for each hour.

4. Web interfaces

The TwiFelt interface has been implemented on a LAMP (Linux-Apache-MySQL-PHP) platform. Currently TwiFelt has two different web interfaces. The first is addressed to the general public and can be accessed through the URL: http://twifelt.ov.ingv.it. This URL links to a dynamic web page that is automatically refreshed every 5 minutes. The web application contains both server-side elements, realized using PHP technology, and client-side elements, realized in JavaScript. Currently the web interface is not optimized for mobile devices, even if it can be viewed on them with sufficient clarity.

The PHP scripts of this interface executes the following tasks:

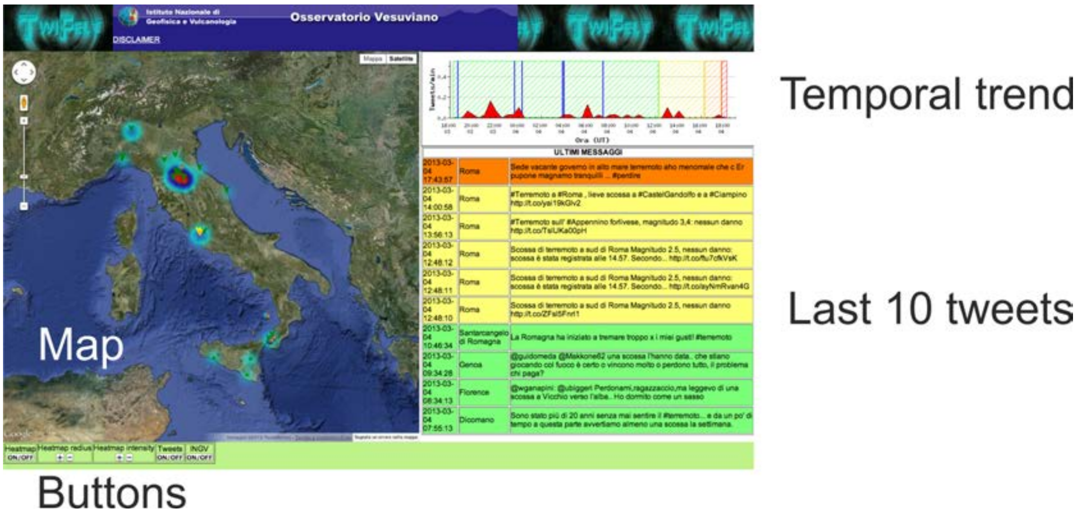
- Querying the database to retrieve tweets posted in the last 24 hours;

- Creating the plot with the temporal pattern of tweets using JpGraph PHP libraries;
- Creating an HTML table showing messages from the last 10 tweets;
- Determining the weights of each event (see sec.3) to create the map with the perception area using Google Maps API (see sec.5);

The JavaScript is used to provide interactivity to the map using Google Maps JavaScript API (sec.5).

The web page layout is structured in 5 elements using CSS formatting (Fig.5). In detail:

- **Header:** contains logos and links to a brief description of TwiFelt and to the disclaimer.
- **Map:** with the representation of the individual tweets with different symbols and colours. Arrows are used to mark tweets geotagged with a point, while circles are used for tweets geotagged with a place. The colour is indicative of the occurrence time. *Red* is for the last 30 min, *orange* for tweets in the last 2 hours, *yellow* for the last 6 hours and *green* for the remaining. The map also contains the epicentres of earthquakes located in the last 24h, determined using the tweets from the account @INGVterremoti. Small INGV icons indicate them. Each symbol is attached to an InfoWindow that is a small window where a message, written in HTML, can be displayed. InfoWindows of user tweets contain the date, the place and the associated message, while the InfoWindows associated to event epicentres contains the basic information about the earthquake. The size of the map is determined dynamically on the basis of the browser window size. The header and the footer of the page have a fixed height of 100 and 60 pixels respectively while the width of the right panel is fixed to 700 pixels. The remaining part of the available windows is filled with the map.
- **Temporal trend:** this plot is realized dynamically using the free version of JpGraph PHP libraries that allow the creation of many types of graphics. This plot shows, with a red area, the tweet rate (tweets/min) computed on intervals of 30 minutes. Vertical blue bars indicate the occurrence of earthquakes. The background is coloured using the same colour scale used for the time intervals used for the markers on the map. The plot is created dynamically at each page reload with a name containing a random part to avoid conflict between different sessions. Figure 7 shows an example of this graph for the M=4.8 Garfagnana earthquake.
- **Table:** it shows the last 10 tweet messages with the same colour scale used for the markers on the map.
- **Buttons**: The buttons are linked to JavaScript functions that allow the masking/unmasking of map elements (heatmap, tweets markers, epicentres) and the changing of heatmap parameters (radius and intensity, see sec.5 for details).



ULTIMI MESSAGGI		
2013-03-04 17:43:57	Roma	Sede vacante governo in alto mare terremoto aho menomale che c Er pupone magnamo tranquilli ... #perdire
2013-03-04 14:00:58	Roma	#Terremoto a #Roma , lieve scossa a #CastelGandolfo e a #Ciampino http://t.co/yai19kGlv2
2013-03-04 13:56:13	Roma	#Terremoto sull' #Appennino forlivese, magnitudo 3,4: nessun danno http://t.co/TsIUKa00pH
2013-03-04 12:48:12	Roma	Scossa di terremoto a sud di Roma Magnitudo 2.5, nessun danno: scossa è stata registrata alle 14.57. Secondo... http://t.co/fu7cfkVsK
2013-03-04 12:48:11	Roma	Scossa di terremoto a sud di Roma Magnitudo 2.5, nessun danno: scossa è stata registrata alle 14.57. Secondo... http://t.co/ayNmRvan4G
2013-03-04 12:48:10	Roma	Scossa di terremoto a sud di Roma Magnitudo 2.5, nessun danno http://t.co/ZFsl5Fnrl1
2013-03-04 10:46:34	Santarcangelo di Romagna	La Romagna ha iniziato a tremare troppo x i miei gusti! #terremoto
2013-03-04 09:34:28	Genoa	@guidomeda @Makkone62 una scossa l'hanno data.. che stiano giocando col fuoco è certo o vincono molto o perdono tutto, il problema chi paga?
2013-03-04 08:34:13	Florence	@wganapini: @ubiggeri Perdonami,ragazzaccio,ma leggevo di una scossa a Vicchio verso l'alba.. Ho dormito come un sasso
2013-03-04 07:55:13	Dicomano	Sono stato più di 20 anni senza mai sentire il #terremoto... e da un po' di tempo a questa parte avvertiamo almeno una scossa la settimana.

Figure 5. Public web interface of TwiFelt with annotated elements. This page has been captured on 04/03/2013 at 19:00 local time.

Data : 2013-03-04 07:55:13
Luogo : Dicomano
Messaggio : Sono stato più di 20 anni senza mai sentire il #terremoto... e da un po' di tempo a questa parte avvertiamo almeno una scossa la settimana.

INGVterremoti
#terremoto Ml:3.4 2013-03-04 03:53::15 UTC Lat=44.04 Lon=11.57 Prof=9.1Km Prov=FIRENZE

Figure 6. InfoWindows associated to markers on the map. On the left InfoWindow of an user tweet while on the right of a seismic event. The spot shown on the map refers to the M=3.4 event of 04/03/2012 (see table 2).

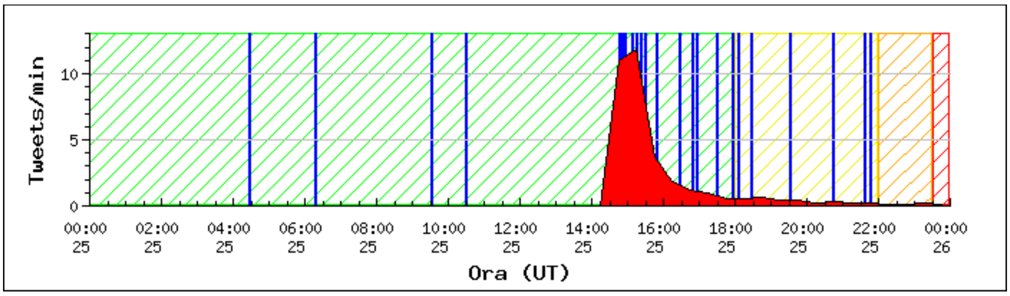


Figure 7. Tweet rate after the 25/01/2013 M=4.8 Garfagnana earthquake.

There is another web interface, reserved to administrators, which includes two additional features:

- Selection of an arbitrary intervals to realize the map and the other elements of the page.
- Identification of users both on markers and on the table with a link to insert them in the blacklist.

5. Map rendering

The rendering of interactive maps is realized using the Google Maps JavaScript API. This technology allows the dynamic creation of interactive maps containing various elements. The background represents a satellite image with a minimal control set allowing the resizing of the map. The tweet and the epicentre markers are realized using custom marker images (PNG files) linked to a Google Maps InfoWindow. InfoWindows allow a rapid check of the tweet messages, to determine if they are or not related to a felt earthquake.

The implementation of colour shading overlays showing the likely perception area has been realized using the Google Maps API Heatmap. These are client-side JavaScript libraries allowing the visualization of density maps overlaying background maps. They allow the customization of the palette and of the appearance of the maps. The construction of a heatmap starts from a discrete set of geolocated and weighted points. A Google Maps heatmap is characterized by 5 parameters:

- **maxIntensity**: if this value is defined it sets the value associated with the topmost colour of the palette. Otherwise it is scaled automatically. TwiFelt fix this value to 200 but also allows the user to change it using the intensity buttons below the map.
- **radius**: this is the radius of influence of a single marker in pixels. TwiFelt uses a value of 0.5, allowing the user to change it through the radius buttons.
- **opacity**: sets the transparency of the map (between 0 and 1). The default value of 0.6 is used.

- **gradient**: specifies the palette of the heatmap through a JavaScript array. TwiFelt uses a custom palette, which goes from cyan to red through blue shading.
- **dissipating**: boolean flag which specifies if the heatmap should dissipate on zoom. It is set to false in TwiFelt.

The heatmap palette has been designed to be quickly interpreted by visual inspection. Cyan shadings indicate that an event has been possibly felt; blue shadings indicate a higher probability while red indicates that an earthquake has been likely felt in that area. The colours of the heatmap shading can be easily distinguished from the marker colours.

6. Case histories

Since November 2012, TwiFelt has been able to show perception maps for about 25 events. Such events were associated to at least 5 tweets, allowing the creation of a reliable heatmap. In table 2 we report their hypocentral parameters (source INGV-CNT). The events have magnitudes ranging from 4.8 to 2.5. There are many other smaller magnitude events that gave rise to only 1-2 tweets. In the following we compare the TwiFelt maps with the macroseismic intensity map compiled by INGV (Sbarra et al., 2010). We show significant results for three selected events: the first M=4.8 affected an area close to Florence, the second M=3.1 affected Northern Italy and the third M=2.5 the urban area of Rome.

Table 2. Events having at least 5 related tweets on TwiFelt (events belonging to the same sequence are reported within the same row).

Date	Time (UT)	Lat (°N)	Lon (°E)	Depth (km)	Mag	Place
2012/11/13	15:09:29	44.93	11.364	2.3	3.0	Pianura padana emiliana
2012/11/16	02:37:12	45.843	10.928	10.2	3.0	Zona Lago di Garda
2012/11/20	10:32:13	44.974	8.23	29.7	3.3	Monferrato
2012/11/22	09:10:41	37.801	14.95	26	3.9	Etna
	09:21:59	37.803	14.977	33	3.0	
	11:25:52	37.800	14.950	26	3.9	
	11:28:55	37.800	14.950	26	3.5	
2012/11/26	19:18:55	44.129	10.676	16.7	3.2	Appennino pistoiese
2012/11/30	00:02:38	44.041	11.753	7.0	3.1	Appennino forlivese
	00:35:06	44.030	11.725	7.7	3.1	
	02:47:19	44.007	11.707	6.6	3.1	
2012/12/03	18:13:56	40.467	15.796	20.5	3.3	Appennino lucano
2012/12/05	01:18:19	42.911	13.66	26.8	4.0	Zona Ascoli Piceno
2013/01/04	07:50:06	37.873	14.722	10.1	4.3	Monti Nebrodi
2013/01/05	21:26:55	45.110	11.637	5.0	3.0	Pianura padana veneta
2013/01/25	14:48:18	44.168	10.454	15.5	4.8	Garfagnana
2013/01/29	20:49:27	44.173	8.562	9.7	3.1	Mar Ligure
	21:17:42	44.158	8.573	9.1	3.1	
2013/01/30	23:42:01	44.129	10.489	10.8	3.3	Garfagnana
2013/02/06	01:36:47	44.275	11.737	53.7	3.5	Zona Forlì
2013/02/12	18:12:43	46.311	12.567	9.7	3.8	Prealpi venete
2013/02/16	21:16:09	41.714	13.576	10.7	4.8	Monti Ernici-Simbruini
2013/02/17	01:00:07	42.464	13.454	16.6	3.7	Gran Sasso
2013/02/25	01:01:41	45.342	7.42	15.9	3.3	Alpi Graie
2013/02/25	22:19:25	45.685	10.068	2.8	3.1	Prealpi lombarde
2013/03/03	13:57:46	41.818	12.499	10.5	2.5	Roma
2013/03/03	23:39:13	38.126	15.821	7.8	3.3	Aspromonte
2013/03/04	03:53:15	44.037	11.566	9.1	3.4	Appennino forlivese

25/1/2013 M=4.8 Garfagnana

Up to now this is one of the strongest events since the operation of TwiFelt (Nov. 2012). After this event, it was observed a tweet rate of about 12 tweets/min, on a 30 minutes average (the highest since Nov. 2012), with a peak of about 50 tweets/min around 14:51 (Fig.7). The mainshock origin time is 14:48:18, while the first tweet related to the event had a time tag at 14:49:05, with a difference of only 47 s. This event alone generated about 3000 tweets (60% of the current TwiFelt database!). The peak rate (Fig.7) shows a marked onset and an exponential-like coda lasting for about 7 hours. The comparison between the TwiFelt map and the macroseismic web-based survey (Fig.8) shows a good agreement. The map in figure 8 has been realized only 12 minutes after the origin time.

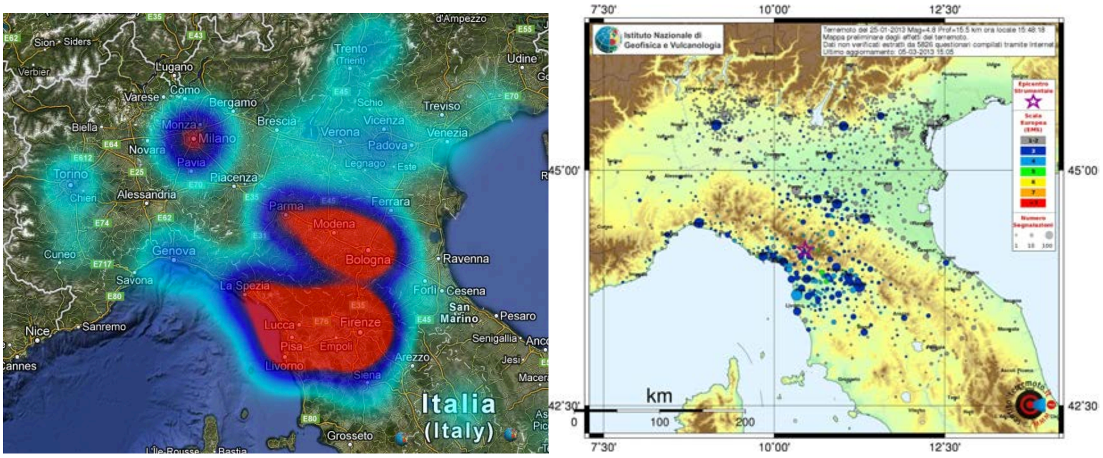


Figure 8. Comparison of the TwiFelt map (left) (snapshot taken at 15:00 of 25/01) and the map with the EMS intensity for the 25/01/2013 M=4.8 Garfagnana earthquake. For sake of clarity the map on the left shows only the heatmap layer.

25/2/2013 M=3.1 Prealpi lombarde

This earthquake was associated to about 25 tweets. In this case too there is a good agreement between the TwiFelt map and the macroseismic data (Fig.9a). In this case the highest peak rate was 0.6 tweets/min (on a 30 minutes average). Posting of tweets related to this event lasted for about 2 hours (Fig.9b)

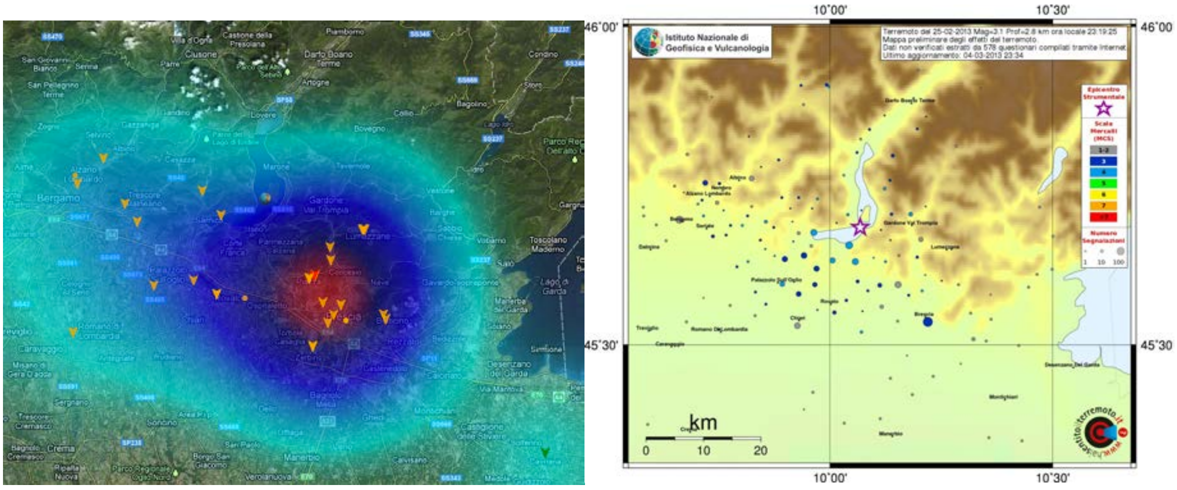


Figure 9.a Comparison of the TwiFelt map (left) (snapshot taken at 00:00 of 26/02) and the map with the EMS intensity for the 25/02/2013 Prealpi lombarde earthquake.

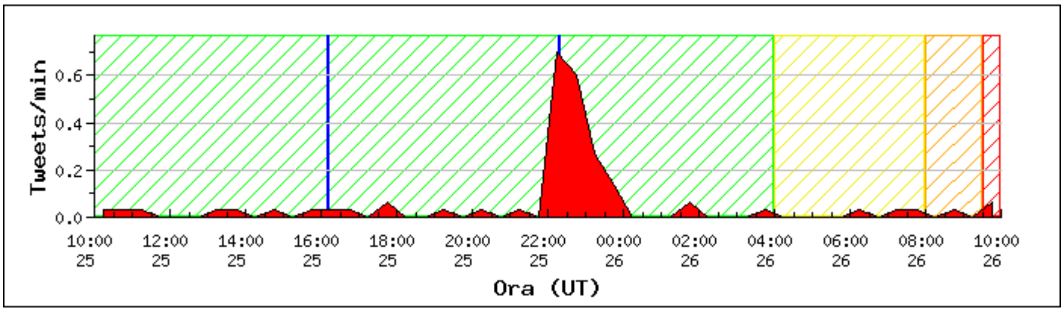


Figure 9.b Tweet rate after the 25/02/2013 Prealpi lombarde earthquake.

3/3/2013 M=2.5 Rome

This small magnitude earthquake was felt through the urban area of Rome (Fig.10a). It ws related to 9 tweets. In this case the TwiFelt map gives a good estimate of the spatial estension of the perception area (Fig.10a) compared to the macroseismic map where only the city of Rome is indicated. From the map it can be seen that also tweets having a place geotagging contributed positively to the map. Figure 10b shows the temporal pattern of the tweet rate.

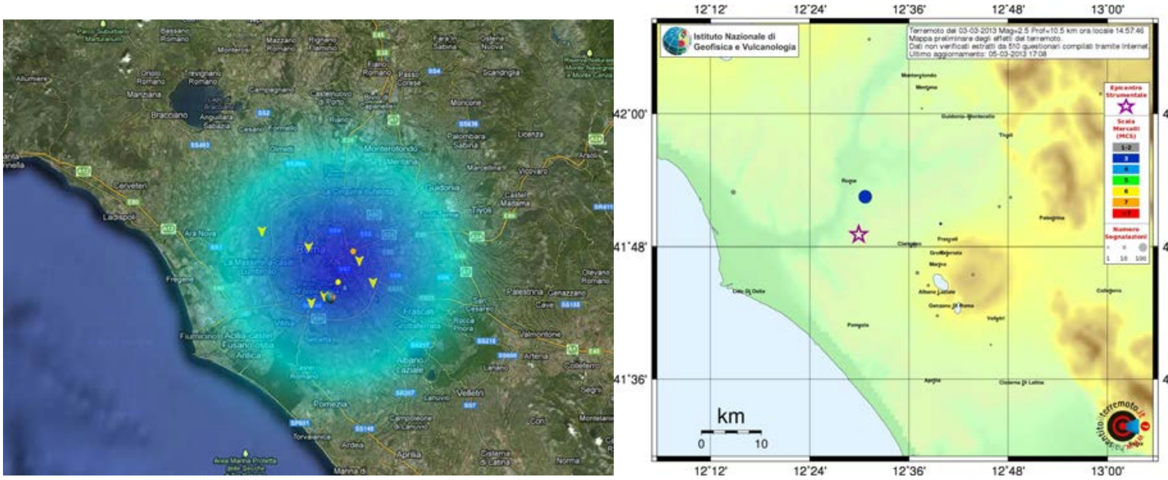


Figure 10.a Comparison of the TwiFelt map (left) (snapshot taken at 19:00 of 03/03) and the map with the EMS intensity for the 03/03/2013 Roma earthquake.

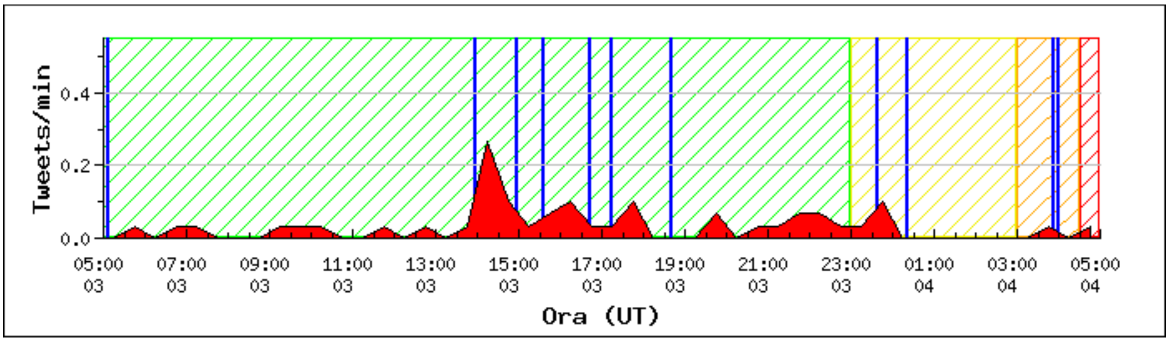


Figure 10.b Tweet rate after the 03/03/2013 Roma earthquake. The secondary peak around 23:00 is related to the Aspromonte earthquake (see Table 2).

7. Future developments

The current system, even if in a prototypical stage, is already able to provide in real-time a map of the earthquake perception area with a certain degree of reliability. The full validation of this approach would require an accurate comparison with macroseismic surveys and with shake maps. A further improvement would be provided by a more sophisticated statistical approach to estimate the deviation of the observed spatial and temporal tweet pattern from the background noise.

It should be carefully evaluated if the web interface should be made public only to experts or to a wider public. In the latter case there is a tangible risk that some "enthusiastic" users could post tweets only with the aim of having it posted on the web page, leading to systematic biases in the maps.

TwiFelt would benefit from the use of a whitelist of trusted users that, using a given keyword, would provide highly reliable information about earthquake perception. Tweets receives from those users would receive a higher weight compared to other users, improving the reliability of maps.

Acknowledgements

We would like to thank Emanuele Casarotti, for his valuable suggestions. Furthermore we warmly thank the organizers of the 38th Workshop of the International School of Geophysics in Erice "*Global challenges for seismological data analysis*" since this work has been inspired by the lectures held there.

References

Earle P. S., Bowden D. C., Guy M. (2011). *Twitter earthquake detection: earthquake monitoring in a social world.* Ann. Geoph, 54/6, 563-581.

Sbarra P., Tosi P., De Rubeis V. (2010). *Web-based macroseismic survey in Italy: method validation and results*. Natural Hazards, 54/2, 708-715.

Appendix: Useful links

In the following we list a set of links to the documentation and specification of technologies mentioned through the text and to data sources. Links are grouped on the basis of the section where they are first mentioned.

1

Twitter API: https://dev.twitter.com/docs/faq

Twitter Streaming API: https://dev.twitter.com/docs/streaming-apis

Twitter search filters: https://dev.twitter.com/docs/api/1.1/post/statuses/filter

Search keywords: https://dev.twitter.com/docs/streaming-apis/parameters#track

wget documentation: http://www.gnu.org/software/wget

JSON specification: http://json.org

libjson documentation: http://sourceforge.net/projects/libjson

Tweets specification: https://dev.twitter.com/docs/platform-objects/tweets

geoJSON specification: http://www.geojson.org

Place definition in tweet structures: https://dev.twitter.com/docs/platform-objects/places

2

MySQL documentation: http://dev.mysql.com/doc

MySQL subqueries: http://dev.mysql.com/tech-resources/articles/subqueries_part_1.html

3

Twitter account @INGVterremoti: https://twitter.com/INGVterremoti

4

Apache documentation: http://httpd.apache.org/docs

PHP documentation: http://php.net/docs.php

CSS specification: http://www.w3.org/Style/CSS

5

API introduction: https://developers.google.com/maps/documentation/javascript/tutorial

Map controls: https://developers.google.com/maps/documentation/javascript/controls

Custom markers: https://developers.google.com/maps/documentation/javascript/overlays#Icons

InfoWindows: https://developers.google.com/maps/documentation/javascript/overlays#InfoWindows

Heatmaps: https://developers.google.com/maps/documentation/javascript/layers#JSHeatMaps

Heatmap options: https://developers.google.com/maps/documentation/javascript/reference#HeatmapLayerOptions

JPgraph documentation: http://jpgraph.net

6

INGV earthquake database ISIDE: http://iside.rm.ingv.it/iside/standard/index.jsp

INGV web-based macroseismic data: http://www.haisentitoilterremoto.it

25/01/2013 M=4.8 Garfagnana earthquake: http://terremoti.ingv.it/it/ultimi-eventi/912-evento-sismico-in-provincia-di-lucca.html

25/01/2013 M=4.8 Garfagnana earthquake macroseismic data: http://www.haisentitoilterremoto.it/repository/7226520880/index.html

25/02/2013 M=3.1 Prealpi lombarde earthquake: http://cnt.rm.ingv.it/data_id/7226971790/event.html

25/02/2013 M=3.1 Prealpi lombarde earthquake macroseismic data: http://www.haisentitoilterremoto.it/repository/7226971790/index.html

03/03/2013 M=2.5 Roma earthquake: http://cnt.rm.ingv.it/data_id/7227053170/event.html

03/03/2013 M=2.5 Roma earthquake macroseismic data: http://www.haisentitoilterremoto.it/repository/7227053170/index.html

Coordinamento editoriale e impaginazione
Centro Editoriale Nazionale | INGV

Progetto grafico e redazionale
Daniela Riposati | Laboratorio Grafica e Immagini | INGV

Via di Vigna Murata, 605
00143 Roma
Tel. +39 06518601 Fax +39 065041181

http://www.ingv.it

Istituto Nazionale di Geofisica e Vulcanologia