🏆 *reviewed paper*

Social Media Geographic Information: Current Developments and Opportunities in Urban and Regional Planning

Pierangelo Massa, Michele Campagna

(Michele Campagna, University of Cagliari, Via Marengo 2, 09123 Cagliari, Italy, campagna@unica.it)
(Pierangelo Massa, University of Cagliari, Via Marengo 2, 09123 Cagliari, Italy, pmassa@unica.it)

1 ABSTRACT

This paper deals with the convergence of Social Media and Geographic Information and discusses its potential as useful source of knowledge in spatial planning. With the underlying assumption of the acknowledgement of the innovation that digital geographic information- including Spatial Data Infrastructures (SDI) and Volunteered Geographic Information (VGI)- is already bringing to urban and regional planning, the authors argue Social Media may also play an important role due to both their pervasiveness in content exchange and their emerging spatial convergence. To support this thesis, a review of best practice examples in different domain is presented in order to understand what tools are currently available and what kind of knowledge can be extracted from Social Media. On the base of this analysis, the paper present an original user-friendly tool developed by the authors to extract information from Social Media and to perform Spatial-Temporal Textual (STTx) analysis. The paper ends with some brief conclusions on the opportunities for the application of STTx analysis in urban and regional planning.

2 INTRODUCTION

In the last decade an unprecedented wealth of digital (geographic) information has been made available to planners to support spatial analysis, design and decision-making. On the one hand, development in Spatial Data Infrastructures (SDI) (i.e. INSPIRE Directive in Europe) is fostering the access and reuse of public authorities' (Authoritative) Geographic Information (A-GI) according to common data, technology, and policy standards (Campagna and Craglia, 2012). On the other hand, developments in geobrowsers and mobile technologies enabled citizens acting as volunteer sensors (Goodchild, 2007) to crowdsource GI real-time in a bottom-up fashion. These trends represent enormous opportunities to enhance the knowledge base for supporting informed spatial decision-making in urban and regional planning. Both expert knowledge from planning professionals and individual volunteers and experiential knowledge from local communities can be now easily collected in digital form, analysed, understood, and eventually used to support decision-making according to a transactive approach (Friedman, 1973). These opportunities may become an important building block for implementing sustainable and democratic communicative planning processes (Innes, 1995). More technically, the unprecedented wealth of GI, which has been made available to planners to support analysis, design and decision-making, is fostering wider innovation in urban and regional planning methodologies, as claimed by a growing number of geodesign evangelists (Miller, 2012; Steinitz, 2012). More recently, widespread diffusion of social media has favoured the diffusion of the geo-referenced multimedia (Sui and Goodchild, 2011), or Social Media Geographic Information (SMGI). The latter can be considered an innovative Big Data source (Caverlee, 2010), inasmuch as traditional spatial analysis methods and techniques may be not fully suitable for eliciting its full knowledge potential. This potential may be considered still limited for the public users both in terms of accessibility and of available analytical apparatus, and new methods and tools should be developed accordingly.

Focusing on these trends with a closer look, the implementation of the Directive 2007/02/CE, establishing a shared INfrastructure for SPatial InfoRmation in Europe, has fostered developments in Spatial Data Infrastructures (SDI) in Member States and regions. SDIs enable the access and reuse of authoritative Geographic Information (GI) according to common data, technology, and policy standards, with beneficial impacts for public administration, developers and planning practitioners (Campagna and Craglia, 2012). In Italy for example, in regions such as Lombardy and Sardinia, spatial planning regulations established the local SDIs as the technical platform supplying data and services for developing regional and local plan-making processes. Representations and design models change format in the plan, and a delay in the adoption of such digital uptake would imply an unbearable loss of competiveness for practitioners.

A second current trend with minor, but still relevant impact in spatial planning, follows developments in geobrowsers and mobile technologies, which are overtaking past limits in data acquisition and communication. Citizens act as volunteer sensors (Goodchild, 2007) to crowdsource and broadcast GI in

real-time with an innovative bottom-up fashion. In some countries worldwide, the use of this kind of technology is easing and fostering the diffusion of participatory processes, and has became a main source of information in domains such as planning emergency response, and in local planning especially in those countries where authoritative data sources are lacking at large scale. As a matter of facts, the concept of citizens obeservatories for environmental protection is an issue of interest in the EU Framework Programme for Research and Innovation Horizon 2020 (i.e. Call SC5-17-2015).

More recently, progress in technology and connectivity, and diffusion of social media disclosed innovative opportunities for disseminating and gathering geographic information. The widespread diffusion in social media applications, especially location-based social networks, fostered the media convergence with GIS (Sui and Goodchild, 2011), spreading the geo-referenced multimedia, or Social Media Geographic Information (SMGI) among million of users over the internet. This trend may also offer opportunities for innovation in spatial planning since users are offering through social media petabytes of freely and publicly available information, which can be detected and used to perform further analysis, aiming to extract useful knowledge for decision-making (Zin et Al, 2013). The integration of SMGI with A-GI can lead towards innovative analysis opportunities in spatial planning, with regards not only to measures of geographic facts but also to users perceptions and feelings of places and localities (Campagna et Al, 2013). However, one major issue to extract useful knowledge from these innovative sources is to find an efficient way to manage this great deal of information. The management issues for information from Big Data sources (Caverlee, 2010) gave rise to the emergence of a new field of research called computational social science (Lazer et Al, 2009). In literature several applications have been found, that propose different approaches for management and analysis of social media contents in a variety of domains, aiming to elicit useful knowledge.

With the above premises, the remainder of the paper is organised as follows. In section 3 recent trends on social media are analysed in more details, in order to better frame discussion on the convergence of GIS and social media. In section 4 the authors present a literature review of advanced social media analysis case studies and propose an assessment framework, with the aims of putting light on current development opportunities and identifying research issues contributing to further innovation. In section 5 an original methodology encoded in the tool Spatext by the authors is proposed for the development of spatial-textual analysis of SMGI in GIS. In section 6 the authors draw some conclusions, summarising the discussion on the opportunities offered by SMGI as analysis, design and decision-making support in urban and regional planning.

3 SOCIAL MEDIA AND INFORMATION RELIABILITY

Widespread diffusion of web and mobile social media applications is producing an unprecedented availability of information over the internet. A measure of the information avalanche phenomenon can be easily given by available statistics on major social platforms (100 Social Networking Statistics & Facts, 2012): Facebook holds over 1 billion of registered users, of which 552 million are daily active, spending over 6 hours monthly on the social platform; Twitter, Google+, Linkedin, Instagram, Pinterest, Flickr, YouTube, to name few of the major, show smaller values in terms of users but considerably high values in terms of contents production. Every day 340 million tweets are sent, 300 million pictures are uploaded via Instagram, every minute 2 million queries are submitted to Google, and 48 hour of videos are uploaded on YouTube, getting 600 million views per day. Contents about any topic could be readily found through available internet services, fostering new opportunities for analysis and research. Furthermore, new features in social networks enable users to embed geographic location into their own media contents, bringing geography into people daily routines so favoring the convergence of GIS and social media.

As such, arguably social networks could be considered as potentially boundless and affordable sources for information about not only on facts, of which accuracy represents a major issue, but also on opinions and feelings of users featuring a geographical component. However, the latter kind of information faces major issues in finding ways for both managing huge data volumes and eliciting reliable knowledge from these sources. In general the social platforms offer diverse ways for content management and extraction, affecting the degree of suitability for knowledge processing. Unlike traditional data, SMGI refers to dynamic processes and requires new kind of tools to support real-time monitoring and decision-making. The quality and reliability of SMGI for research and practices is actually being discussed, and further investigations are required to establish the extent of credibility of this information. Moreover, several issues arise in finding

suitable practices and procedures for the integration of crowd-sourced with authoritative data. Advanced Big Data analysis may be a suitable solution for extracting and managing social media information. Several analysis could lead toward a direct extraction of the content (what?) rather than the causality (why?) from data (Pohl et Pohl, 2013), so avoiding data volume issues and shifting the traditional scientific method by means of current computing capabilities. In several domains, advanced Big Data analysis approaches and tools have been proposed and explored to manage this wealth of digital information for different analysis purposes, as explained in the next section.

4 ADVANCED SOCIAL MEDIA ANALYSIS AND ASSESSMENT FRAMEWORK

The wealth of information available from social media about facts, opinions and feelings of users can affect several domains of interest where subjectivity of observation is relevant for expressing the views, the needs, the call of individuals and communities, such as: disaster events, political events, media events, social studies and last but not least urban planning, to name few. In literature several relevant applications have been found in these domains, which promote methodologies and innovative tools to provide solutions in dealing with social media contents. Although the applications provide specific procedures and features according to the different design purposes and domains of interest, several commonalities can be found both in analysis and in technology. In order to develop a suitable assessment framework for these innovative methodologies and analysis, a review of applications for each domain is presented in this section. The aim of the framework, which will be presented later in the following section, is to define current opportunities for development, considering both technology and analysis of applications, and offering a set of tools to choose from, according to specific different working settings and capabilities. The view offered by the assessment framework may also contribute to define open issues to shape the research agenda.

4.1 Disaster events

In the domain of disaster events four applications have been found, which offer different methodologies and procedures to deal with SMGI in eliciting useful knowledge for decision-making. As first example, an application developed by Liu et Al. (2008) to retrieve photos of disasters from Flickr is reviewed. The application proposes a methodology based upon a qualitative study to investigate if and how users activity on Flickr can evolve in case of notable disasters. Flickr is a social platform, that allows users to store, search and share photos publicly or into thematic groups. Photographs and activities of groups report features of the disaster over time, exposing the online social convergence related to the event. Through a qualitative study, the collected images regarding the hazard, the post-impact and the online convergence may be significant to disaster response efforts. Especially the photos of latter kind, combining data from different sources, have been found useful to create new overviews and map mash-ups, providing spatial information on the event location and the potential response resources. The application proposes an approach to capture SMGI, demonstrating how these data could be useful for disaster response and recovery issues. However, the methodology does not provide advanced technology and requires a manual intervention for recognition and extraction of data in order to develop a qualitative analysis for suitable results.

As second example of this review, a case study proposed by Vieweg et Al (2010) to enhance the situational awareness (SA) through analysis of Twitter posts during a disaster, is analysed . The aim of the application is the harvesting of real-time contents during a crisis event according to information production and consumption lifecycle in microblogging. Two disaster events occurred in the 2009 in USA have been analyzed, relying on the Twitter API for data collection. The research of information among tweets is based on well-defined terms related to each event, oriented by the results of an initial investigation of the Twitter public stream. The geographic component is obtained through a manual analysis of users locations to achieve a manageable dataset for each event. The analysis shows different behaviours between people in the warning phase (anticipatory awareness) of a disaster and those in the impact phase (real-time awareness). Moreover, results exhibit a higher percentage of tweets with georeferenced information during the impact phase in order to supply useful information to users. The case study proposes several procedures and analysis for managing SMGI of Twitter, relying on advanced instruments and technology. The manual intervention is strongly reduced both in data harvesting and analysis.

The third case study presents an application which allows to detect events related to disasters by Twitter contents. The Twitter Event Detection Analysis System (TEDAS) proposed by Li et Al (2012) aims at

ISBN: 978-3-9503110-6-8 (CD-ROM); ISBN: 978-3-9503110-7-5 (Print)
Editors: Manfred SCHRENK, Vasily V. POPOVICH, Peter ZEILE, Pietro ELISEI

detecting and ranking new events according to their importance, generating spatial and temporal patterns of the extracted data. The system relies on Java, PHP and APIs (Twitter API and Google Maps API) to collect tweets and location of users, according to well-defined search terms. The results of the query offer a real-time overview of spatial and temporal patterns of the detected events. In this case technologically advanced instruments are used, and no manual intervention is required.

The last case study example in disaster management is a recent application developed by Zin et Al (2013) to extract information about disaster events from social media. The application extracts visual and textual data respectively from YouTube and Twitter to describe the situation awareness related to disasters. An approach composed by several steps is proposed to analyze SMGI, focusing on location, network, contents and aggregating the relevant data into hourly temporal groups. Statistical operations are then carried out on collected data to rank the detected events according to their importance. Nevertheless, the application requires an empirical procedure to correctly manage data, and the results shows differences for the detection of events between textual SMGI and visual SMGI.

4.2 Political events

A case study was found in the field of politics which proposes several analysis on SMGI. The event of French Presidential election in 2012 has been analyzed from the Twitter stream by an application named Pytolab (Luce, 2012). The application relies on the Twitter API to perform several real-time analysis on contents related to well-defined search terms. The results show a report on textual analysis of relationships among keywords in data and several statistical reports about spatial and temporal patterns of the event. Moreover, the results show that less than 1% of processed SMGI of Twitter supplies geographical information, causing a potential loss of information. In this case several advanced instruments and procedures are fostered to elicit knowledge from SMGI.

4.3 Media events

Further case studies have been found in the domain of media events, of which two applications have been reviewed here, introducing several methods and procedures for obtaining useful information from social media data. The first one enables the real-time exploration of Twitter contents in search of media events. The TwitInfo platform developed by Marcus et Al (2012) aims at identifying media events according to well-defined terms of research, offering results in a timeline with an aggregate sentimental analysis of contents. The data volume is processed in real-time, offering users the capability to further explore the detected events. The application also provides analysis on spatial and sentimental pattern of the event. In this case both advanced technology and analysis have been used to deal with SMGI retrieval and analysis. The second case study, proposed by Mathioudakis and Koudas (2010) enables the analysts to detect in real-time social media trends from the Twitter stream . The TwitterMonitor application can automatically detect high-rate keywords in tweets (bursty keywords), relying on the Twitter API. The results are then processed to extract further information about the detected event or topic. In this case study advanced technology and instruments have been also applied to perform a trend analysis on social media data.

4.4 Social studies

In the domain of social studies two case study were reviewed, which propose the use of SMGI for different research purposes: semantic analysis and social graph analysis. The first case study offers several methods for the semantic analysis of the Flickr tags, with the aim of determine whether the tag refers to a place or an event (Rattenbury et Al, 2007). The system relies on several statistical methods to achieve the result of information extraction from the unstructured knowledge of tags. However, manual intervention is required in the data extraction and processing. In addition, a technical paper has been found which presents a methodology to visualise social graphs from the Twitter contents. The proposed approach relies on the Twitter API to perform an automatic extraction of data and on a graphing software to achieve the social graph (Bertrand, 2012). In this case, available data of social media are processed and used as input for the creation of a social graph representation.

4.5 Urban planning

Two case studies have also been found with relevance for urban and regional planning, which introduce different methodologies and procedures in an attempt of extracting knowledge from SMGI for supporting design and decision-making.

The first case study deals with an application developed by Frias-Martinez et Al (2012) for investigating the human interaction with the urban environment. The methodology proposes an interesting approach to determine the urban land uses according to human dynamics (i.e. where people are at what time?) as detected by Twitter . The approach is applied on the case study of the urban area of Manhattan, relying on the Twitter API to extract geotagged tweets. In this methodology several spatial clustering operations are performed, according to the neural networks theory, to obtain meaningfull maps from SMGI. Moreover, statistical operations on temporal frequency of SMGI have been used to identify the different land uses on maps. The results of the analyses shows how the approach can accurately detect the commercial, leisure and residential areas, as well as new urban land uses. This methodology relies on advanced instruments and technology to automatically perform innovative analysis. The results demonstrate how SMGI can supply meaningfull information for urban analysis and planning processes.

A second methodology was also found which enables to investigate people movement and landmarks preferences in urban environment. The approach proposed by Jankowski et Al (2010) aims at applying geovisual analytics techniques on photographs of urban landscapes and their related tags, to detect new potential landmarks in the city. The approach has been applied on the case study of Seattle, relying on the Flickr API for performing an exploratory spatio-temporal analysis on both locations and periods of the data volume. The procedural approach for data processing involves three main steps: in the first one, several spatial clustering operations are performed on SMGI to obtain maps, while time periods are divided in weekly intervals; in the second step, statistical operations are computed on data to identify potential interesting landmarks, according to given ranking criteria; eventually, further investigations are performed in the third step to validate the results. In this case advanced instruments and innovative spatio-temporal analysis have been applied on SMGI to elicit knowledge. The results of this methodology also contribute to demonstrate how SMGI can be used as affordable source for the collection of geographic information relevant for urban studies and planning.

4.6 Assessment framework

In the light of above review of advanced social media analysis methods and tools, an assessment framework is presented for synthesising recent advances and opportunities. The aim of the framework is to provide a usable set of parameters for the classification of case studies with regards to technology and analysis. Central to the framework is a two-dimensional matrix, whose variables are “Technology” and “Analysis”. In the Technology-Analysis Matrix (TAM) each field represents an unique combination of technological level and analysis level. The reviewed applications and methods have been evaluated by means of suggested matrix, with the aim of defining the current state of development. The TAM is shown in Table 1, highlighting the most interesting advances.

TECHNOLOGY LEVEL / ANALYSIS LEVEL	*MANUAL TASKS (T1)*	*AUTOMATIC TASKS (T2)*
VISUALISATION (A1)	*case study: 1* retrieve Flickr photos for response efforts	*case study: 9* visualise social graph from Twitter contents
ADVANCED PROCESSING (A2)	*case study: 2,4,8* harvest real-time tweets on crisis event analyse video and tweets to rank events perform a semantic analysis on Flickr tag	*case study: 3,5,6,7,10,11* detect disasters events real-time from Twitter Twitter statistical, spatial, temporal reports detect media event real-time from Twitter detect real-time trends on Twitter investigate urban land uses from Twitter investigate favorite landmarks from Flickr

Table 1: Matrix Technology-Analysis for 11 reviewed case studies.

The Technology Level relies on the following two categories:

ISBN: 978-3-9503110-6-8 (CD-ROM); ISBN: 978-3-9503110-7-5 (Print)
Editors: Manfred SCHRENK, Vasily V. POPOVICH, Peter ZEILE, Pietro ELISEI

level T1 – a strong manual intervention is required to perform data collection or analysis;

level T2 – advanced programming procedures perform automatic operations.

The Analysis Level relies on the following two categories:

level A1 – analysis results offer a descriptive visualisation of collected data;

level A2 – analysis on data elicit further knowledge from data.

The classification of the methodologies provide interesting information about advanced social media analysis methods and tools. All considered methodologies include data visualization. While in 36.4% of the case studies a manual intervention is required, in the remaining 63.6% automatic procedures for data processing are available. Altogheter 54.5% of cases include both automatic procedures and advanced analysis functions. The field T2–A2 can be considered the space for advanced systems both in terms of technology and analysis, including most powerfull tools encoded in user-friendly applications.

The application of the TAM for categorizing and assessing the case studies put lighs on opportunities and issues for SMGI analysis. The results of this assessment were used by the authors as a guide for the design of a novel Spatial-Temporal Textual (STTx) analysis tool presented in the next section.

5 A METHODOLOGICAL APPROACH FOR SMGI IN GIS

The current state of development for advanced social media analysis methods and tools demonstrates how the use of SMGI is spreading in several domains. Nevertheless, further advances and innovation, both in technology and analysis, may lead toward more reliable and user-friendly methods for knowledge extraction from SMGI. In order to contribute to this challenge, an original methodology is proposed below for the STTx analysis of SMGI. The method has been applied to a case study example using a tool called Spatext, that features several functionalities to analyse SMGI in GIS environment.

5.1 Spatext tool

The Spatext STTx suite is implemented as add-in for ESRI ArcGIS©. Spatext includes nine tools, which can be used to achieve three main goals: harvesting social media data from Twitter, data geocoding, and tag clouding. The functionalities rely on three open Python modules, namely tweetersearch for data harvesting, geopy for the geocoding, and pytagcloud for the tag clouding. The functionalities introduced by Spatext and their coupling with GIS spatial analysis tools eases the integration of SMGI with authorative data, for analysis, design and decision-making support in urban and regional planning. In the remainder of the paper, the Spatext functionalities for SMGI processing are synthetically described using the illustrative case study of the cyclone Cleopatra in Sardinia (Italy). Collected SMGI in the case study encloses perceptions, opinions and needs from the local communities during the occurrence of cyclone Cleopatra in November 2013. The integration of authorative geographic information with opinions and needs from the local communities provided by SMGI could offer meaningful information both for design and decision-making. The results of the data harvesting, the spatial-temporal analysis and the textual analysis on the case study are proposed to demonstrate the main features of the suite.

5.2 Data harvesting

In Spatext the Twitter data harvesting is executed by the “Tweet Extractor” tool, which retreives all relevant data according to given keywords (several languages can be set for the query). In the case study the dataset has been collected setting the keywords “Sardinia floods” in English language to avoid disturbance for harvesting because of too generic keywords. The setting allowed the extraction of 399 different tweets strictly related to the topic, approximately 16 and 24 hours after the start of event (2013.11.19 from 8:30 to 9:30 and from 17:30 to 18:00). The tool automatically stored the public tweets and their metadata in a data table including text, user name, user location, user registration data, tweet creation time, and if available, country, place name, the geographic coordinates. Only 0.01% of extracted tweets presented geographic information. The value is smaller than 0.77% that statistically refers to percentage of tweets with geographic component (Semiocast2012 – Twitter Stats), but this may be due to time and language constraints. Spatext has been also designed for metadata processing, in order to address the lack of geographic information, so providing suitable results for spatial-temporal analysis.

5.3 Spatial-temporal analysis

The lack of available geographic information in the dataset could prevent the development of spatial analysis on SMGI. A potential solution provided by Spatext to deal with this issue is the use of tweets' metadata, that is user location for populating the location fields. This approach could introduce positional uncertainty because the geographical information is being shifted from the tweet to the user location, but at the same time it could provide the spatial distribution of the topic, according to the spatial distribution of users. In the case study example, the tool has been used to generate a point feature dataset containing the locations of the 399 collected tweets, offering as a result opportunity for developing further spatial analysis in combination with other spatial data layers. A check on the dataset has pointed out how 42.6% of tweets has provided wrong places for user location, showing reluctance by users for disseminating personal information.

The spatial analysis has been conducted studying the spatial distribution of the topic among worldwide countries with the aim of discovering dissimilarities and potential spatial patterns. Several countries have been chosen because are the most representative of the spatial distribution of the topic: the United States, the United Kingdom, Italy, Nigeria and India. According to the spatial distribution, the dataset has exposed as percentage of tweets a value of 22.5% for the USA, 6.5% for the UK, 6.0% for Italy, 2.5% for Nigeria and 1.7% for India. These values could be considered as a representation of the different degree of interest of users about the topic among these countries. An assessment of this hypothesis could be addressed through the analysis of the different percentages. The USA percentage has exhibited notably higher value (> 300%) than the UK, Italy, Nigeria and India, but this peak of interest could be explained considering that the USA are ranked 1st in Twitter top countries classification with over 140M accounts. Similarly, the percentage of the UK could be explained considering the 4th position in Twitter rank with over 30M accounts. The italian percentage (6.0%) raises interesting questions about the spatial influence of the topic. In fact, the percentage is both affected by language constraint in data harvesting and by the less importance of Italy in Twitter top countries (over 5M accounts). An analysis on values shows the 1.2 value for the ratio percentage / accounts for Italy, exposing the higher level of interest among the countries. This peak of interest in Italy could be explained considering the geographical extent, indeed the topic concerns Sardinia. In addition, the percentages of Nigeria and India raise further questions related to the spatial distribution of the topic on Twitter. In fact, both countries are affected by language constraint and do not expose specific spatial relationships with the topic. These values have been investigated through temporal and textual analysis to better understand potential hidden dynamics.

The analysis on the spatial distribution of dataset has been conducted through standard spatial analysis tools in GIS environment. The dataset used as input for the analysis has been developed directly by Spatext, demonstrating how the tool could be suitable for an integration of SMGI from Twitter in GIS environment. In figure 1 the spatial distribution of the collected tweets about the topic is provided worldwide. Several tweets are depicted with different colours and symbol, because they are related to the relevant countries. The map of spatial distribution is useful to better explain the results of assessment on spatial distribution. Moreover, the map may suggest hints on the temporal distribution of SMGI, that are discussed below.

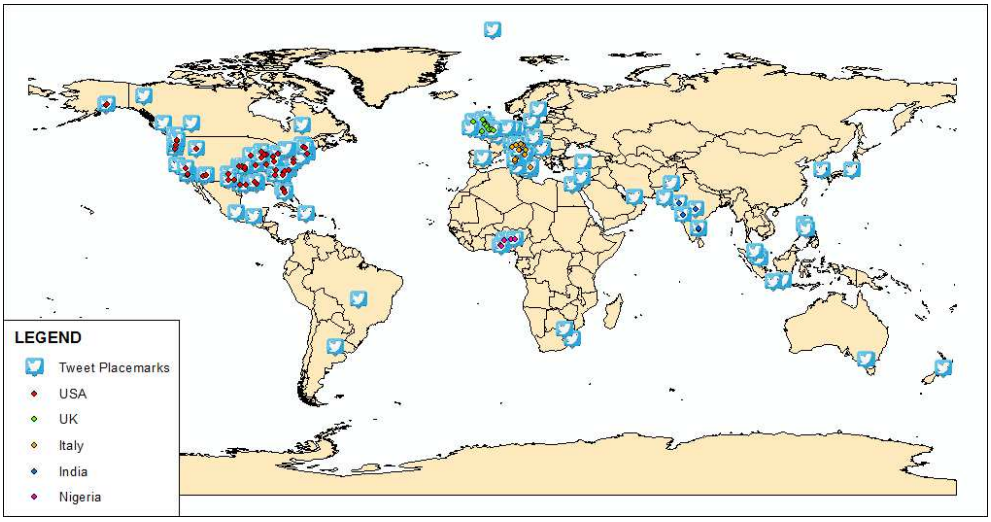

Fig. 1: Spatial distribution. Results of analysis on most representative countries.

ISBN: 978-3-9503110-6-8 (CD-ROM); ISBN: 978-3-9503110-7-5 (Print)
Editors: Manfred SCHRENK, Vasily V. POPOVICH, Peter ZEILE, Pietro ELISEI

The introduction of temporal distribution analysis has been possible using the metadata of tweets creation, which are stored in the dataset by Spatext. The temporal component has been investigated in search of meaningful patterns about the topic according to two different approaches. For the temporal distribution the collected tweets have been processed with the aim of inquiring the geographic extents through the associated temporal component. The first approach considered the spatial distribution of the first 30 created tweets, but the results have pointed out a heterogeneous spatial distribution, so dismissing any meaningful spatial pattern. The second approach considered the spatial distribution of tweets for the two extraction periods (from 8:30 to 9:30 and from 17:40 to 18:05 2013 Nov 19) and then for 17 periods of 5 minutes. The result of analysis on extraction periods has exposed 199 tweets for the first period and 200 for the second, suggesting how the interest was increasing along time. The analysis on 5 minute periods has been computed to search for anomalies in temporal distribution and to show the trend in the time-series graph.

The results showed peaks of interest in the 9th and 13th period that were further investigated for validation. The validation of a peak of interest was based on the contemporary fulfilment of two criteria in order to avoid false positives. The first criterion for the validation of a peak requires a value for the period >=100% than the specific value of linear regression in the same period, meanwhile the second criterion requires a value for the period =>100% than the average value. The identified peaks satisfy both criteria and therefore were validated as peaks of interest in the trend. The graph of temporal analysis is provided in figure 2, which shows the time-series and the peaks of interest. The identified peaks have been further investigated in the textual analysis.

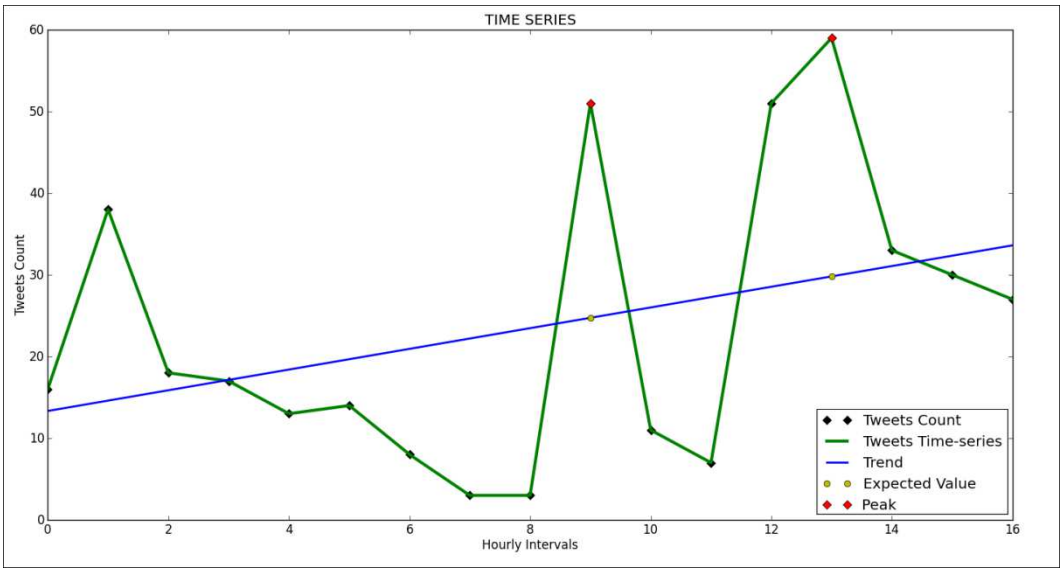

Fig. 2: Temporal distribution. Results of analysis on tweets.

5.4 Textual analysis

The peaks of interest, detected by the temporal analysis, oriented the analysis toward the study of further aspects on the SMGI dataset. A textual analysis has been conducted on the dataset to discover further knowledge in order to find potential answers to peaks. Tag-cloud functionalities in Spatext have been used to discover the most used words in the dataset. The tag-cloud is the visualisation of word frequency in a weighted list and graphical form, and it is suitable to depict the most used words in a text and recognize possible hidden information. The textual analysis has been conducted both on worldwide tweets and on tweets from Italy, investigating possible differences in contents. Furthermore, a search was conducted on tweets from Nigeria to discover potential causes for spatial distribution in this country.

The results of tag cloud on worldwide tweets identifies words used for the data harvesting (Sardinia and floods), words strictly related to the subject (storm, cyclone, cleopatra, rain, italy), several terms related to the conseguences of the event (apocalyptic, dead, bridges) and several unrelated terms (wall, journal, bbc, http). The results of tag cloud on Italy tweets have exposed the same related terms, while several different unrelated terms were existent in the results (allertameteosar, yesterday, http). A further investigation has been conducted on the unrelated words worldwide, disclosing how they were provoked by a viral sharing phenomenon of tweets posted by the Wall Street Journal and BBC World. These tweets have caused the

identified peaks of interest and the tag ‘http’ by means of their diffusion and their url-related contents. Results of analysis on Italy showed the use of the word ‘allertametosar’ as tag in several tweets, while the word ‘yesterday’ was related to tweets with a temporal reference to start of the hazard event. Finally, an investigation on nigerian tweets has disclosed how the interest on the topic was caused by an high sharing phenomenon among users leaded by Breaking News Nigeria, which offers overviews of worldwide events. In conclusion, the provided results have explained the causes of the peaks and demonstrate how the textual analysis could enhance the awareness on contents of SMGI. These results also strengthen both the concept of media as GIS and the current convergence of GIS and social media. The resulting tag cloud visualisations are provided in the next figure (fig.3) respectively for italian and worldwide tweets.

6 CONCLUSION

In this paper the authors presented a review of advanced social media analysis method and tools, and an original methodology for the spatial-temporal textual analysis of SMGI. Advanced methodologies and analysis have been reviewed in order to discover potential suitable approaches for the integration of the SMGI spatial analysis as a support in urban and regional planning. Early examples of analysis methods and tools were found in several domains such as disaster events, media events, political events, social studies and urban planning in facts. An assessment framework was proposed to define the current advances and opportunities for SMGI analysis. The proposed framework sets the context for the proposal of an original methodology developed by the author for the Spatial-Temporal Textual analysis of SMGI. The method relies on a tool called Spatext in order to demonstrate the opportunities and the potentialities for the collection and analysis of SMGI. A simple example case study is provided as a demonstrator of how the SMGI could be directly used into GIS environment, disclosing new opportunities for spatial analysis. These results arguably show new opportunities for enriching authoritative data with information about perception, opinions and needs from the local communities. In the current time of convergence of GIS with pervasive social media, the users perceptions and feelings of facts and places enclosed in SMGI may eventually offer meaningful information for planners. As such language and power-relational barriers between expert professionals, stakeholders, and local communities may blur in pluralist representation models and databases. The early examples proposed in the paper aims at demonstrating how it is now possible to analysis what, when and where people know, feel, appreaciate, need with regards to places, facts, and processes. If properly (and ethically) used, this new pluralistic knowldge might eventally change decision-making dynamics and affect the discourse in the urban and regional planning process. In conclusion, the knowledge of SMGI if proficiently elicited might be used to discover and expose the will of users and could be a valid support for design, analysis and decision-making in urban and regional planning. Further research is definitely needed and SMGI analysis methods and technology should be applied from within real-life urban and regional planning process to proof the full efficacy. Nevertheless, early results are promising and the research agenda challenging.

7 ACKNOWLEDGEMENTS

The work presented in this paper was developed by the authors within the research project “Efficacia ed efficienza della governance paesaggistica e territoriale in Sardegna: il ruolo della VAS e delle IDT” [Efficacy and efficiency of landscape and environmental management in Sardinia: the role of SEA and of SDI] CUP: J81J11001420007 funded by the Autonomous Region of Sardinia under the Regional Law n° 7/2007 "Promozione della ricerca scientifica e dell'innovazione tecnologica in Sardegna".

8 REFERENCES

100 SOCIAL NETWORKING STATISTICS & FACTS FOR 2012. In website available from: http://visual.ly/100-social-networking-statistics-facts-2012 [Accessed 2013 Nov 13]

BERTRAND F.: Visualizing the Twitter social graph, Part 1: Getting the data. In: Technical writeup on Recollect Engineering. 2012. Available from: http://code.recollect.com/post/20476037331/visualizing-twitter-social-graph-pt1 [Accessed 2013 Nov 13]

CAMPAGNA M., Craglia M.: The socioeconomic impact of the spatial data infrastructure of Lombardy. In: Environment and Planning B, Vol. 39, Issue 6, pp. 1069-1083. 2012.

CAMPAGNA M., Kudinov A., Ivanov K., Falqui R., & Anastacia G.: PLACE, I CARE! CROWDSOURCING PLANNING INFORMATION (p. 18). In: presented at the AESOP-ACSP Joint Congress, Dublin, Ireland, 2013.

CAVERLEE J.: A few thoughts on the computational perspective. In: Paper presented during the Specialist Meeting on Spatio-Temporal Constraints on Social Networks, December 2010, Santa Barbara, CA. 2010. Available from: http://www.ncgia.ucsb.edu/projects/spatio-temporal/docs/Caverlee-presentation.pdf [Accessed 2014 Feb 07]

FRIAS-MARTINEZ V., Soto V., Hohwald H., & Frias-Martinez E.: Characterizing Urban Landscapes using Geolocated Tweets (p. 10). In: presented at the 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust, Amsterdam, Holland, 2012.

FRIEDMANN J.: Retracking America: A Theory of Transactive Planning. Garden City, NY, 1973. Doubleday/Anchor

GOODCHILD M.F.: Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. In: International Journal of Spatial Data Infrastructures Research, Vol. 2, pp. 24-32. 2007.

INNES J.: Planning Theory's Emerging Paradigm: Communicative Action and Interactive Practice. In: Journal of Planning Education and Research, 14(3) pp. 183-189. 1995.

JANKOWSKI P., Andrienko N., Andrienko G., & Kisilevich S.: Discovering Landmark Preferences and Movement Patterns from Photo Postings. In: Transaction in GIS, 14(6), 833–852. 2010. doi:10.1111/j.1467-9671.2010.01235.x

LAZER D., Pentland A., Adamic L., Aral S., Barabasi A. L., Brewer D., Van Alstyne M.: Life in the network: the coming age of computational social science. In: Science, (323), 721–723. 2009. doi:10.1126/science.1167742

LI R., Lei K. H., Khadiwala R., & Chang K. C.-C.: TEDAS: a Twitter Based Event Detection and Analysis System. In: IEEE 28th International Conference on Data Engineering (p. 3). 2012.

LIU S. B., Palen L., Sutton J., Hughes A. L., & Vieweg S.: In Search of the Bigger Picture: The Emergent Role of On-Line Photo Sharing in Times of Disaster. In: proceeding of the Information Systems for Crisis Response and Management Conference (p. 10). Washington, DC, USA, 2008.

LUCE L.: Python, Twitter statistics and the 2012 French presidential election. In: Technical writeup. Available from: http://www.laurentluce.com/posts/python-twitter-statistics-and-the-2012-french-presidential-election/ [Accessed 2013 Nov 14]

MARCUS A., Bernstein M. S., Badar O., Karger D. R., Madden S., & Miller R. C.: TwitInfo: Aggregating and Visualizing Microblogs for Event Exploration. In: proceedings of the 2011 annual conference on Human factors in computing systems (CHI'11) (pp. 227– 236). New York, NY, USA, 2011.Association for Computing Machinery (ACM). doi:10.1145/1978942.1978975

MATHIOUDAKIS M., & Koudas N.: TwitterMonitor: Trend Detection over the Twitter Stream. In: proceeding of the 2010 ACM SIGMOD International Conference on Management of Data (pp. 1155–1558). Indianapolis, Indiana, USA, 2010.

MILLER W. R.: Introducing Geodesign: The Concept. ESRI Press, Redlands California, USA, 2012.

POHL J., & Pohl K. J.: Big Data: Immediate Opportunities and Longer Term Challenges (p. 12). In: presented at the InterSymp-2013, Germany, 2013.

RATTENBURY T., Good N., & Naaman M.: Towards Automatic Extraction of Event and Place Semantics from Flickr Tags. In: proceeding of the 30th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 103–110). New York, NY, USA, 2007.

SEMIOCAST: Twitter reaches half a billion accounts More than 140 millions in the U.S. In: website available from: http://semiocast.com/en/publications/2012_07_30_Twitter_reaches_half_a_billion_accounts_140m_in_the_US [Accessed 2013 Nov 14]

SUI D., Goodchild M.F.: The convergence of GIS and social media: challenges for GIScience. In: International Journal of Geographical Information Science, Vol. 25, Issue 11, pp. 1737-1748. 2011.

STEINITZ C.: A Framework for Geodesign: Changing Geography by Design. ESRI Press, Redlands California, USA, 2012. ISBN: 9781589483330

VIEWEG S., Hughes A. L., Starbird K., & Palen L.: Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. In: proceedings of the 2010 annual conference on Human factors in computing systems (CHI '10) (pp. 1079–1088). Atlanta, GA, USA, 2010.

ZIN T. T., Pyke T., Hiromitsu H., & Takashi T.: Knowledge based Social Network Applications to Disaster Event Analysis. In: proceedings of the International MultiConference 2013 of Engineers and Computer Scientists IMECS (Vol. 1, p. 6). Hong Kong, 2013.