Instagrammers report about the deadly wildfires of East Attica, 2018, Greece: An introductory analytic assessment for disaster management purposes

Stathis G. Arapostathis
Geographer, PhD graduate
Harokopio University
e.arapostathis@gmail.com

ABSTRACT
This article contributes to identifying the capabilities of Instagram when utilized as a source of Volunteered Geographic Information (VGI) for disaster management (DM) purposes. The geographic focus of this research is in the Mediterranean area. As case study, the fire event of East Attica 2018, Greece, was chosen. This major fire occurred on the 23rd of July 2018 and caused the death of 100 people, the injury of additional 164 while the total burnt area was about 1275,9ha. It is the deadliest in modern Greece's history and the second deadliest at a global level, within the 21st century. About 15000 related photos along with the corresponding captions and timestamps were crawled from Instagram. An initial sample of about 1100, was analyzed, by using a certain methodology divided in certain steps, the most important of which include the classification of the information to certain categories, geo-referencing and the creation of graphs and maps that visualize the processed data.

Keywords
volunteered geographic information (VGI), Instagram, social media, fires, disasters, disaster management

INTRODUCTION

There is over a decade passed since the notion of Volunteered Geographic Information (VGI) emerged from Goodchild (2007) and many scientific teams are now at the of stage researching the potentials of this phenomenon in a variety of sciences. VGI is the act of having normal citizens, without any related scientific background, produce geographic information or information with spatial extension, either intentionally or unintentionally. One of the most important applications of VGI focus on disaster management (DM). A plethora of research articles have analyzed various aspects of the potentials of VGI in disaster management, some of them are mentioned indicatively in the following paragraphs.

The work of McDougall (2011) focuses on the effectiveness of social media for providing additional insights in the flood events of Queensland floods in Australia. Those major floods affected a huge geographic area consisted of 30 cities and agricultural areas while, in terms of budget, the restoration cost was about five billion dollars. The most important conclusion of the researcher was related to the immediacy and depth of information that is provided through social media, conclusion that is also verified by Yin et al. (2012) who stated that by using this information, the relevant authorities can improve their awareness and as a result the response in emergency situations. Yin et al. (2012) also developed a system for manipulating the extracted information. Regarding fire events, previous research demonstrates some recent work about the social media network twitter and the fire event of North-East Attica that occurred in Greece during the summer of 2017 (Arapostathis and Karantzia 2019).

VGI research is not only limited on processing data of existing VGI sources but also include various participatory activities like the approach presented by Rahman et al. (2018). That approach is related to the development of a community was able to identify spots of the Dhaka city in Bangladesh, that were vulnerable to earthquakes. The community handled the development of strategies for reducing the vulnerability along with the execution of these related actions. A similar approach is presented by De Brito et al. (2018) and is related to the development of an expert community who was involved in defining the criteria that emerge a geographic area as vulnerable in the

event of a flood.

Especially regarding social media, there is a logical trend of combining various social media sources or even social media sources with satellite imagery in order to extract a more complete output (de Albuquerque et al. 2015; Morstatter et al. 2013). Another nice combination of data source is presented in the work of Smith et al. (2012) who combined social media data and data gathered from a graphics processing unit (GPU) in the case study of the Tyne and Wear floods that occurred during June and August 2012 in the United Kingdom. Those approaches are rather the most effective ways towards perfection. However, futher research is vital to be performed for each social media network exclusively, using case studies that vary geographically in order to have the most comprehensive assessment regarding the comparative potentials of each social media source when is utilized as a VGI source for DM purposes. Especially regarding geographic variations, one of the most important characteristics of user generated content is the ubiquitous and anarchic production in terms of volume, volunteers and geographic areas (Girres and Touya 2010; Haklay et al. 2010).

According to author's opinion the main open challenges, regarding the manipulation of social media datasets, are clustered in four certain categories, all transpired by the significant notion of quality. The first cluster is related to defining the appropriate classification structure for the processed information, suitable for each disaster event category according to the semantics and data properties of each VGI source. There are many articles of various respectful researchers that propose various structures either while researching certain case studies or by creating various conceptual frameworks (Albuquerque et al. 2015; Fahrer et al. 2018; Arapostathis et al. 2018; Dashti et al. 2013). However, as research in the field is constantly evolving, there should be a common definition regarding the most appropriate one.

The second cluster of challenges is related to the geo-referencing of information. Accurate and precise geo-referenced information can lead to the rapidness of the actions taken regarding response during a disaster management situation. However, this is not always feasible. The most desired outcome would be to have a precision level of few meters or even of few centimeters. Some of the DM stakeholders though, may be interested in having information displayed in other scales.

The third category of challenges is related to developing the appropriate techniques for visualizing information that makes sense to all DM participants. From the decision makers, up to simple volunteers which are trying to contribute to the way they can. Taking this into consideration, the graphs and maps used should be very simple and easily readable in few seconds despite of the educational background of each individual.

Finally, the fourth cluster of challenges is related to automation of the methodologies producing the desired output in real time. Analyzing data coming from VGI sources, such as social media, is a time-consuming process as many researchers conclude (Imran et al. 2013, Zhong et al. 2016, Grunder-Fahrer et al. 2018).

Current status of this research aspires to make a step further towards the first three categories.

CASE STUDY

The wildfires of East Attica 2018, Greece, started within the 23rd of July 2018, at 18:41 local time. Main affected areas were the areas of Mati, Marathonas, Rafina, Nea Makri, and the east side of the Penteli mountain. A hundred people died while 164 were injured. According to local witnesses, the speed of the fire was so high that burnt two young people on a scooter while they were trying to run out of the burning area. 26 people were burnt within a taverna that was located near the coastal part of Mati, while hundreds of citizens stayed into the sea for many hours. 1275 hectares were burnt in east Attica according to official measurements of the Copernicus emergency service while more than a thousand buildings were destroyed by the fire. The authorities, apart from East Attica had to deal a fire event that occurred in West Attica (region of Kineta, mountains of Geraneia) few hours before. The combination of those two fire events really struggled the antifire mechanism of the country.

This fire event, until today, is the country's deadliest since the start of modern Greece, and the world's second deadliest of the 21st century, after the Victorian fires.

DATA AND MATERIAL USED

The data were collected by using a crawler developed by Andrea Tarquini (https://github.com/h4t0n/). 15011 photographs, along with the corresponding timestamps and captions, were crawled. All of data contained at least one of the following hashtags: #athensfire, #prayforathens, #prayforgreece and #matifire. This amount of photographs is about the 38% of the total photos available in Instagram, containing those certain hashtags. The author chose not to use a certain bounding box as the news about this event were spread at a global level. The dataset consisted of 4 json datasets which included the timestamp information, the online url and the caption of each image. All images were downloaded by using a wget command in a Linux terminal.

Apart from the crawler, the author also used the Calc of LibreOffice software, for some basic edits on the dataset, R-studio for converting json files to csv and for creating scatterplots, and finally ArcGIS Pro software for creating the final maps.

METHODOLOGY

The methodology applied, was consisted on five basic steps (Figure I). The first step was related to crawling the appropriate volume of Instagram posts according to the most representative and descriptive to the fire event hashtags, while the second step was linked to the data preparation; the conversion of the files that contain the crawled information in an spreadsheet-friendly form along with the selection of the data posted within certain time periods. In the third step, the selected posts were classified into certain defined categories considering each photo and each photo's caption. In this initial investigation, 4 certain categories were defined: a. fire identification, b. disaster management info, c. consequence score values assessed from each image's caption and d. consequence score values assessed from each image. The consequence score values score board was presented in previous research (Arapostathis and Karatzia 2019) and is associated to indicating a certain numeric value that represents effectively the negative impact of each consequence as it is described in each post. The score range is defined from I to V; value I represents the simple fire event identification while value V is strictly associated to the human loss (table I). In cases of different consequence scores between the photo and the caption the highest value was selected as a default rule.

Table I: Description of the conceptual consequence score value range

CONSEQUENCE SCORE VALUES	DESCRIPTION
I	Simple fire event identification
II	Small areas consisted of burnt woodland
III	Emergency situation, large areas are burnt, property in danger
IV	Danger of human life, burnt property, extremely large areas are burnt
V	Human loss

The next step of the methodology was linked to geo-referencing. In this stage, the geo-referencing was performed by considering only the detection of geographic coordinates that exist within each photo's caption. The geo-referencing method applied, was presented in previous research for the social media network twitter (Arapostathis et al. 2018). In brief, the number of the geographic entities that are contained within the text of each caption was noted and each caption was replicated N times where N was the number of detected entities. Afterwards, the corresponding x, y coordinates were added for each caption in a way that each geographic entity of a group of replicated posts was represented once. Moreover, a geographic precision score value according to the precision level of each mentioned geolocation was indicated. Those score values range also from I to V, where I refers to the most precise geolocations (a POI, or a certain street name and number) and V is the most general, related to the administrative level of a prefecture/provincia (table II).

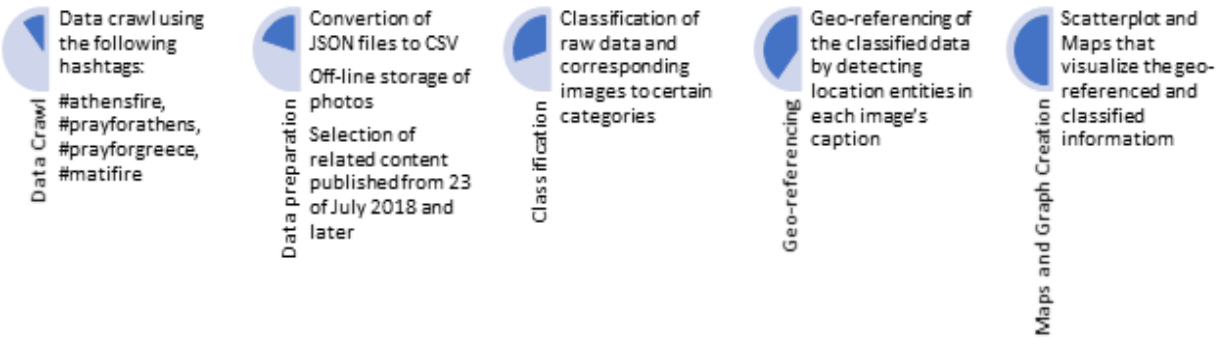
Table II: Description of the geographic precision score value range

GEOGRAPHIC PRECISION SCORE VALUES	DESCRIPTION
I	Street name and number or Specific POIS
II	Street name
III	Neighborhood or hamlet
IV	Municipality
V	Prefecture and above

Finally, the spatial distribution of all the pairs of x, y coordinates were randomized within the polygon that represents each detected geolocation. All the randomized coordinates were stored in a table which also contains the classified information. Even there is previous research related to the development of an R-script that automates

this geo-referencing technique (Arapostathis 2018) due to the initial stage of the analysis it was preferred to check a data sample, manually, checking each post in an 1 by 1 basis in order to gain experience and explore hidden potentials of Instagram, regarding the content of each post. The final step of the methodology was consisted on creating various maps and graphs that were suitable for an initial analysis of the data; In specific scatterplots were created that visualize the volume of information in certain time periods. The starting point of the time periods is the 23rd of July 2018, at 18:00 (Greek summer zone time). Moreover, three maps were generated containing information regarding a. the quantity of posts in each geographic area, b. the geo-referenced information regarding disaster management and c. the geo-referenced consequence score values, respectively. (Maps I, II, III). The total number of processed Instagram posts was 1070.

Figure *I: Methodology*



RESULTS

Figure II: Scatterplots of Instagram posts within certain time periods

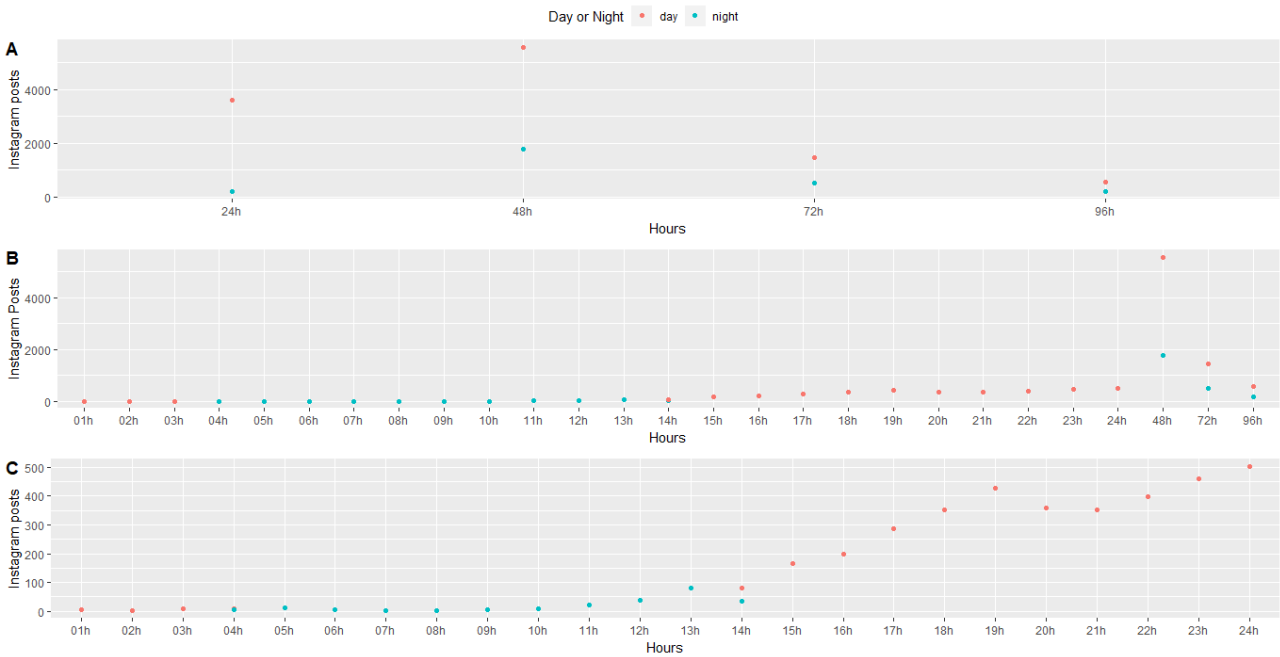
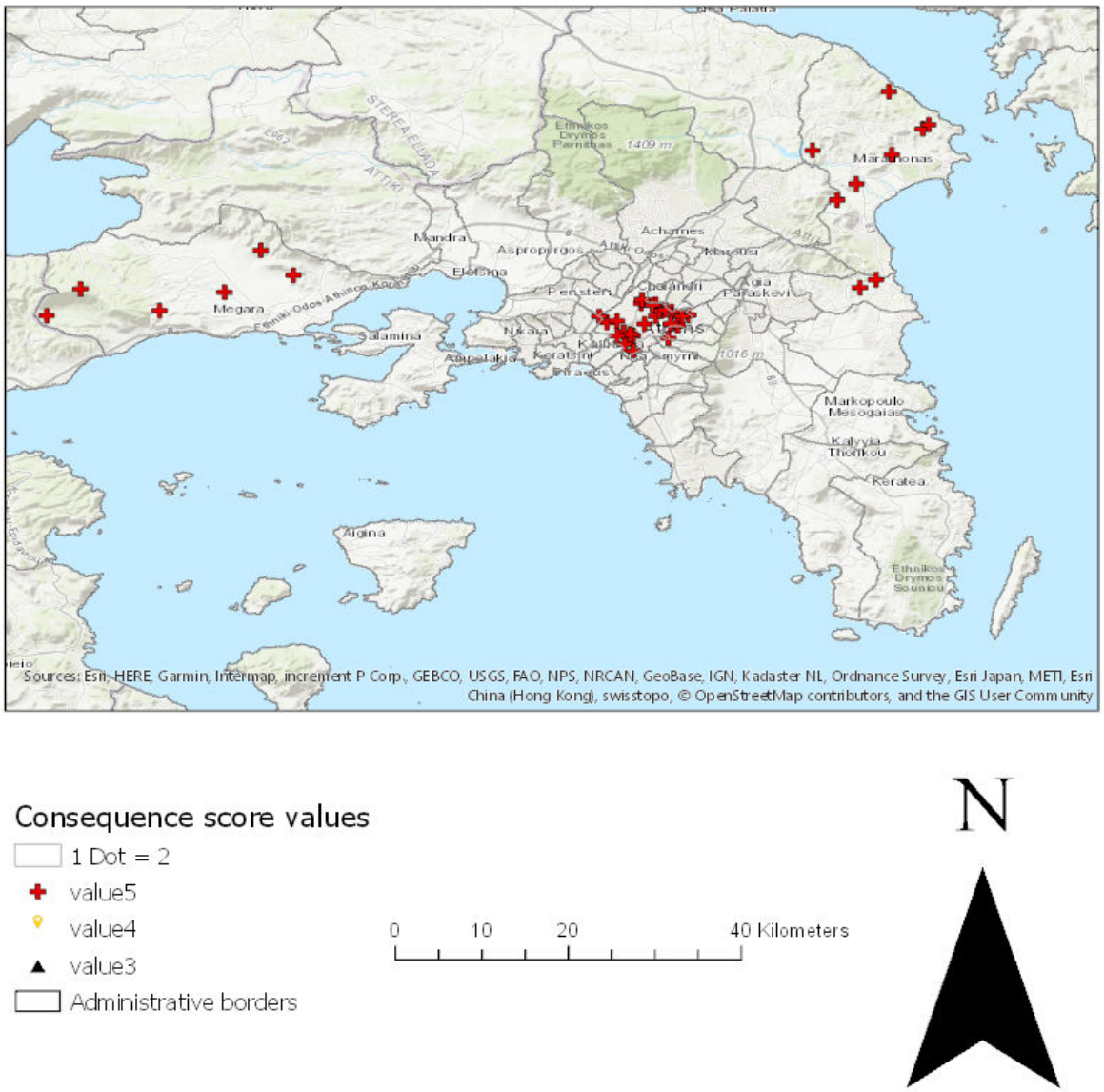


Figure II provides an interesting insight regarding the published quantity of Instagram posts through time. Axis X demonstrates the time periods values in hours (h); time starts since the occurrence of the fire event. In Axis Y the volume of instagram posts is displayed while the different colors represent the day-night condition. The most

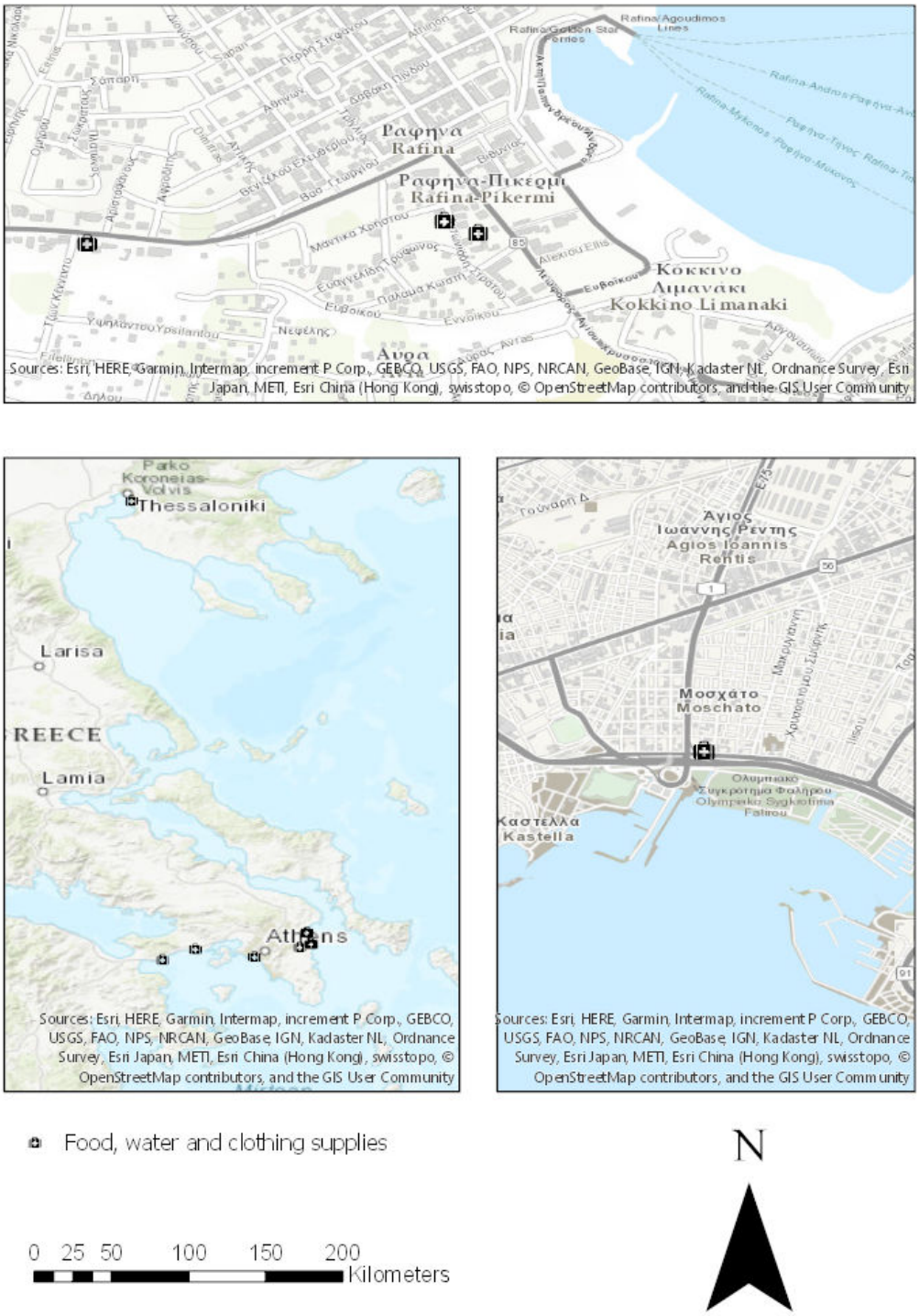
interesting part is that within the first hours, since the start of the fire event, the volume of related published posts is impressively low. The volume starts to grow significantly just after the first 13 hours while the total production is almost doubled after the first day.

Map I: Consequence Value Scores according to Instagram data (after analyzing 1100 posts)



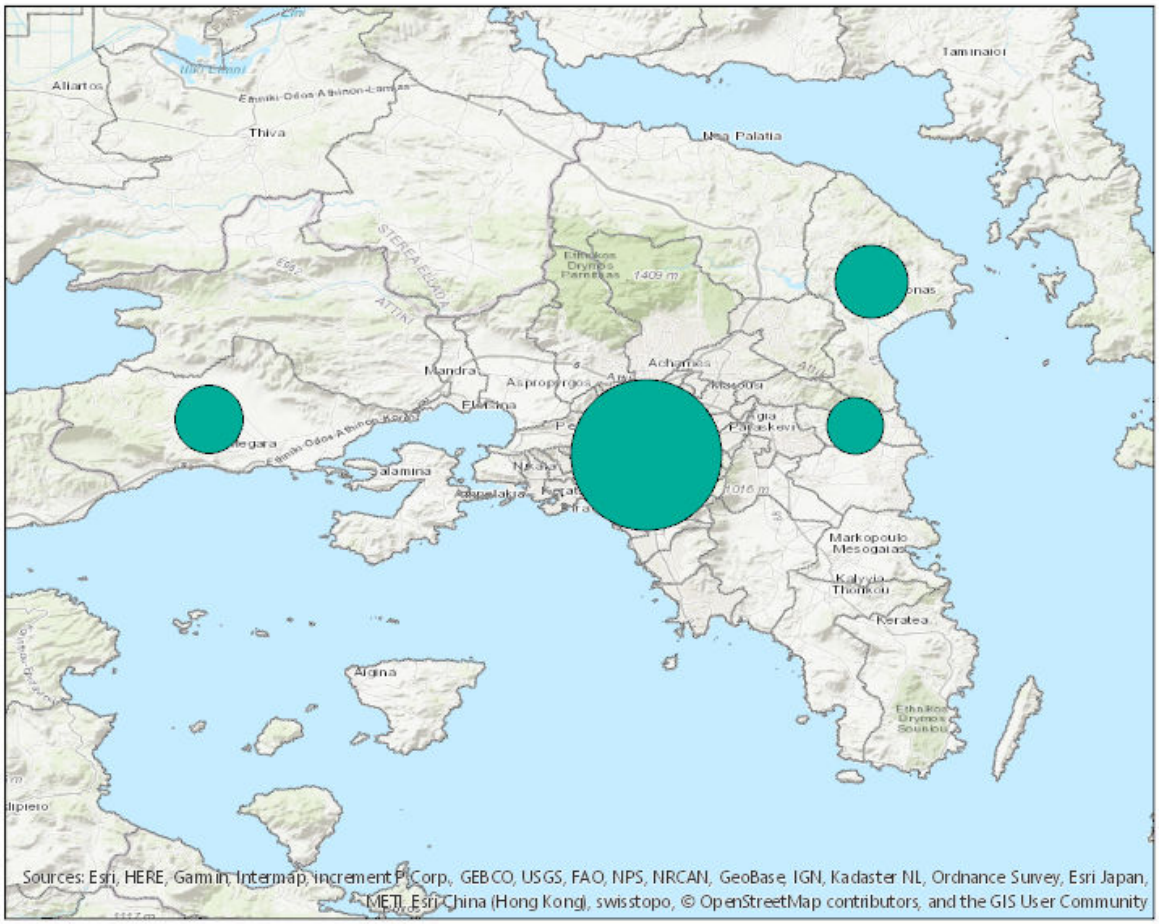
In Map I the geo-referenced consequence score values are vizualized. The outcome was a result of applying the methodology presented in previous sections. As already stated, value 5 (symbol of red cross) is linked to the human loss while each symbol represents two Instagram posts that indicate the same consequence score value. In reality, the human losses were located only at the east part of Attica. The crosses located in the municipality of Athens (central part of the map) appear due to the high values of the geographic precision that many posts received. A lot of people used the word Athens, while posting, in order to describe the whole region that is displayed in Map I (known as Tsipras syndrome). That region is the prefecture of Attica. The red crosses at the west appear due to a fire event that occurred in that area, few hours before the East Attica fire event. As a result, many Instagram posts included information for both fires along with the information that there are dead people, without clarifying the certain area in which the people died.

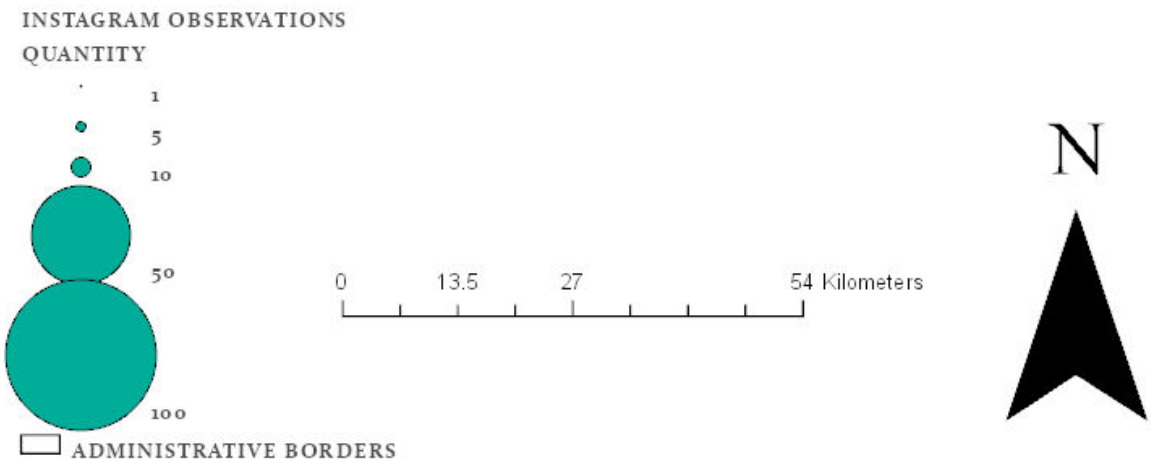
Map II: Disaster Management Maps: Locations for collecting resources for the fire event victims



Map II is more close to the desired final output of disaster management maps. Contains the most precise Instagram observations in terms of geography. Those spots that appear are at precision level I (POI, or street name and number) and visualize various places for collecting food and water supplies along with clothes or any other material that would be useful for the victims of the fire event.

Map III: Proportional map: Quantity of Posts in certain areas





In Map III the spatial distribution of the volume of geo-referenced posts is displayed. The map uses proportional symbols in order to visualize the quantities. The displayed symbols at the east are within the main affected areas from the fire event that was analyzed in current research, while the symbol at the west identifies a fire event that occurred in West Attica few hours before. Finally, the big circle in the center of the map is linked to the geographic precision issues (the so-called Tsipras syndrom) of the information as many Instagram posts disseminated the information that there is a fire event in the geographically general area of Athens.

DISCUSSION – CONCLUSIONS

Instagram is a powerful social media network that was surely not designed for disaster management purposes. However, major disaster events seem to concern Instagram users, who post at a high volume at least, in catastrophic events. One of the most interesting output of this research is the latency between the start of the fire event and the post of related information. This characteristic of Instagram is rather different to other social media networks, such as twitter, the dissemination of information through which, is performed almost instantly. This time lag makes Instagram rather not useful during the initial response phase of the disaster management procedures. The plethora of information though gives a nice insight regarding event-tracking purposes and for displaying some important information about disaster management procedures that are performed after the initial response. The appropriate geo-referencing of the information is still under development. Although some of the Instagram Posts may contain the x y coordinates of the spot that a photo is published, the credibility of that information is ambiguous as according to Huiji and Barbier (2011) and logical assumptions, a post could be published from another geographic place than the place that the post refers to. However, a more detailed analysis regarding that part, needs to be performed in the future in order to investigate all the potential aspects of the embedded geographic information. Moreover, the geo-referencing method used in this research extracts many coordinates. There seems to be a mess though when people are generalizing the areas in which they refer to.

One of the most important advantages of Instagram is the ability to publish a photo, as those are very helpful for classifying purposes. Although a photo has the power to instantly describe a whole situation, it is logically assumed that in disaster management situations not all photos are published. For instance, as Instagram is designed mostly for fun purposes, it is rather rare to find pictures of dead people, information that could be vital during DM procedures. That certain information though is published, in a high rhythm, within each photo's caption. Concluding, it seems that Instagram can act as a supplementary source, by giving very good insights in certain parts of the DM process. However, the final assumptions need further research.

The future steps of this introductory analysis include the processing of the total dataset, the application of SVM techniques for classifying images to the appropriate categories in almost real time, the development of more meaningful to the DM stakeholders and participants, graphs and maps. Finally, the automation of the whole steps of the methodology in a way that it could be executed at once and in real time is vital for the applicability in DM situations in live time.

References

Arapostathis S. G. (2018), "Automated Methods for Effective Geo – Referencing of Tweets Related to Disaster Management", International Geomatics Appliations Conference – Book of abstracts (online), ISSN: 2653-9098, page 62.

Arapostathis, S. G., Karantzia, M. (2019). Mapping Information of Fire Events, from VGI Source (Twitter), for Effective Disaster Management (in Greece); The Fire of North-East Attica, August 2017 (Greece) Case Study. In Advances in Remote Sensing and Geo Informatics Applications (pp. 257-260). Springer, Cham.

Arapostathis, S. G., Lekkas, E., Kalabokidis, K., Xanthopoulos, G., Drakatos, G., Spirou, N., Kalogeras, I (2018).: Developing seismic intensity maps from twitter data; the case study of Lesvos, Greece 2017 earthquake: Assessments, improvements and enrichments on the methodology, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-3/W4, 59-66, https://doi.org/10.5194/isprs-archives-XLII-3-W4-59-2018, 2018.

Dashti S., Palen L., Heris M. P., Anderson K. M., Anderson S., Anderson T. J. (2013) Supporting disaster reconnaissance with social media data: a design-oriented case study of the 2013 Colorado Floods. In Proceedings of the 11th International ISCRAM Conference – University Park, Pennsylvania, USA .

Goodchild MF (2007) Citizens as sensors: the world of volunteered geography. Geojournal 69(4):211-221.

De Albuquerque, J. P., Herfort, B., Brenning, A., Zipf, A. (2015). A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. International Journal of Geographical Information Science, 29(4), 667-689.

De Brito M., Evers, M., Almoradie ADS (2018) Participatory flood vulnerability assessment: a multi-criteria approach. Published by Copernicus Publications on behalf of the European Geosciences Union.

Girres, J. F., Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, *14*(4), 435-459.

Grunder-Fahrer S., Schlaf A., Wustmann S. (2018) How social media text analysis can inform disaster management GSCL 2017, LNAI 10713, pp. 199-207. https://doi.org/10.1007/978-3-319-73706-5_17

Haklay, M., Basiouka, S., Antoniou, V., Ather, A. (2010). How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *The Cartographic Journal*, *47*(4), 315-322.

Huiji, G., and Barbier, G. 2011. Harnessing the Crowdsourcing Power of Social Media for Disaster Relief. IEEE

Intelligent Systems 26 (3): 1541-672.

Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., Meier, P. (2013, May). Extracting information nuggets from disaster-related messages in social media. In Iscram.

McDougall K. (2011) Using Volunteered Information to Map the Queensland Floods In: Proceedings of the Surveying Spatial Sciences Biennial Conference 2011. 21 – 25 of November 2011, Wellington, New Zealand .

Morstatter F., Pfeffer J., Liu H., Carley K. M. (2013). Is the sample good enough? Comparing data from twitter's streaming APIwith Twitter's firehose. In Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013 (pp. 400-408). AAAI press.

Rahman M. M., BaruaU., Khatun F., Islam I., Rafiq R. (2018) Participatory vulnerability Reduction (PVR): an urban community-based approach for earthquake management. Natural Hazards. https://doi.org/10.1007/s11069-018-3364-2.

Smith L., Liang Q., James P., Lin W. (2017) Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. Journal of Flood Risk Management 10:370-380

Yin J., Lampert A., Cameron M., Robinson B., Power R. (2012) Using Social Media to Enhance Emergency Situation Awareness IEEE Computer Society pp 1541-1672.

Zhong, X., Duckham, M., Chong, D., Tolhurst, K. (2016). Real-time estimation of wildfire perimeters from curated crowdsourcing. Scientific reports, 6, 24206.