

2020

Twitter Analysis based on Damage Detection and Geoparsing for Event Mapping Management

yasmeen ali

Helwan University, Faculty of Commerce and Business Administration, Business Information System Department, Cairo, Egypt., dr.ali.yasmeen@gmail.com

Khaled Bahnasy

Professor, Information Systems Department Faculty of Computer & Information Sciences Ain Shams University, Abbasia, Cairo, Egypt., Khaled.bahnasy@cis.asu.edu.eg

Adel El-Mahdy

Professor, Department of Economics and Foreign Trade, Faculty of Commerce and Business Administration, Helwan University, Cairo, Egypt, Adelmahdy@link.net

Follow this and additional works at: <https://digitalcommons.aaru.edu.jo/fcij>



Part of the [Business Analytics Commons](#), [Business Intelligence Commons](#), [Computer and Systems Architecture Commons](#), [Management Information Systems Commons](#), and the [Systems and Communications Commons](#)

Recommended Citation

ali, yasmeen; Bahnasy, Khaled; and El-Mahdy, Adel (2020) "Twitter Analysis based on Damage Detection and Geoparsing for Event Mapping Management," *Future Computing and Informatics Journal*. Vol. 5 : Iss. 1 , Article 1.

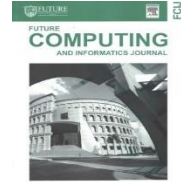
Available at: <https://digitalcommons.aaru.edu.jo/fcij/vol5/iss1/1>

This Article is brought to you for free and open access by Arab Journals Platform. It has been accepted for inclusion in Future Computing and Informatics Journal by an authorized editor. The journal is hosted on [Digital Commons](#), an Elsevier platform. For more information, please contact rakan@aarj.edu.jo, marah@aarj.edu.jo, dr_ahmad@aarj.edu.jo.

Twitter Analysis based on Damage Detection and Geoparsing for Event Mapping Management

Cover Page Footnote

Yasmeen Ali Ameen, LL.M., in the management information system, from Sadat Academy for management science. Researcher at doctoral studies programme at the University of Helwan, Faculty of Commerce, Business Information System Department. Currently, she is employed as a Teaching Assistant at the El- Gazeera Institute for Computer and Management Information Systems. Cairo, Egypt. The author can be contacted at dr.yasmeen@commerce.helwan.edu.eg Khaled El-Bahnasy, Ph.D., Professor, Information Systems Department Faculty of Computer & Information Sciences Ain Shams University, Abbasia, Cairo, Egypt. Teaches courses Information Systems Department Faculty of Computer & Information Sciences Ain Shams University. Dean of the Higher Obour Institute for Management, Computers and Management Information Systems from August 1 until the date. The author can be contacted at Khaled.bahnasy@cis.asu.edu.eg Adel El-Mahdy, Ph.D., Professor, Professor, Department of Economics and Foreign Trade, Faculty of Commerce and Business Administration, Helwan University, Egypt. Teaches courses Economics and Foreign Trade, Faculty of Commerce, Business information system department, Helwan University, Egypt. Dean of the Higher Institute for Administrative Sciences, Al-Kattameya. The author can be contacted at Adelmahdy@link.net



Twitter Analysis based on Damage Detection and Geoparsing for Event Mapping Management

Yasmeen Ali Ameen^{1,a}, Khaled Bahnasy^{2,b}, Adel Elmahdy^{1c}

¹Helwan University, Faculty of Commerce and Business Administration, Business Information System Department, Cairo, Egypt.

²Ain Shams University, Faculty of Computer and Information, Computer Science Department, Cairo, Egypt.

^adr.yasmeen@commerce.helwan.edu.eg, ^bKhaled.bahnasy@cis.asu.edu.eg,

^cmaelmahdy@yahoo.com

Abstract

Background: Early event detection, monitor, and response can significantly decrease the impact of disasters. Lately, the usage of social media for detecting events has displayed hopeful results. **Objectives:** for event detection and mapping; the tweets will locate and monitor them on a map. This new approach uses grouped geoparsing then scoring for each tweet based on three spatial indicators. **Method/Approach:** Our approach uses a geoparsing technique to match a location in tweets to geographic locations of multiple-events tweets in Egypt country, administrative subdivision. Thus, additional geographic information acquired from the tweet itself to detect the actual locations that the user mentioned in the tweet. **Results:** The approach was developed from a large pool of tweets related to various crisis events over one year. Only all (*very specific*) tweets that were plotted on a crisis map to monitor these events. The tweets were analyzed through predefined geo-graphical displays, message content filters (damage, casualties). **Conclusion:** A method was implemented to predict the effective start of any crisis event and an inequity condition is applied to determine the end of the event. Results indicate that our automated filtering of information provides valuable information for operational response and crisis communication.

Keywords: Twitter; Geoparsing; event detection; mapping; crisis response; twitter scoring.

1 Introduction

Lately, social media, and twitter, have a novel source of information on emergency events. The tweets that are carried by millions of users around the globe hold huge potential in disaster management. When analyzed, they can contribute valuable information about the impacts of ongoing disaster events. Twitter enables users to automatically attach their current GPS location to a tweet, defining their position at the moment a tweet is posted [1]. Nevertheless, because this feature is switched off by default, only 0.9% of the tweets have geo-coordinate information attached [2].

Many previous studies clarified this problem such as Middleton et al. [4] explained that named entity matching NEM performs more reliable than NER (Name Entity Recognition) on tweets. This approach divides tweets into tokens and matches these tokens first to cities, the streets, and finally, regions, while excluding matched tokens to avoid double matches. Zhang et al. [5] analyzed various spatial indicators, as the time area, the user place field, and other textual evidence, to get a more reliable assessment of a particular tweet's location. Their results exposed that tweet geoparsing results can be developed using these methods, but only for those tweets with available spatial indicators. As spatial information is not always available, this approach cannot be simply applied to all tweets. Furthermore, even when this information is available, it does not regularly match the place mentioned by the user.

Most research performs event detection first, sometimes followed by geoparsing. For example, Sakaki et al. [1] represent a system that accumulates tweets, filters them using a support vector machine SVM. Then, they detect emergency events using a system that measures the probability of a special number of sensors reporting an emergency event

within a critical time interval. Once an event is detected, localization of the event is done using Kalman and filters based on the tweets' GPS-coordinates and their users' registered locations. Sarmiento et al. [6] propose a system that first obtains location mentions from the textual tweet and then detects events based on irregularities in the activity related to geographic locations within fixed time windows. Jongman et al. [7] explained that an increase in geographical tweets occurs within flood events, which in some cases, provides for quicker flood detection than using other methods. Nevertheless, no event detection step is employed. Rossi et al. detect flood events in Italy using flood-related tweets test utilizing fixed time windows. Yet, we claim that it is problematic to apply fixed-length time windows, particularly in data-poor environments. Due to the lack of data, methods would require using a long-time window in these regions to detect mild or slow-onset events. Though, this also means that the detection of a fast onset critical event takes much longer. On the other hand, a flexible time window method can detect events at the time enough data is available irrelevant to a fixed time window.

The purpose of our study is to develop a geoparsing approach for tweets without assuming a priori information about an event. So, for event detection and mapping; the tweets will locate and monitor them on a map. This new approach uses grouped geoparsing then scoring for each tweet based on three spatial indicators. Plotting of tweets is reliably found by determining locations, events, and their timeframe.

the study is outlined to apply the development of our research and show its methodology and phases using approximately one year of locally sourced tweets with multiple event-related keywords in Egypt, collected between October 18, 2019, and July 14, 2020.

2 Methodology

For the event monitoring and mapping them, the challenge thus includes the monitoring of many synchronous events that can have a gradual or sudden onset across different regions. A solution is to first use geoparsing on the text of tweets and then make event detection. The geoparsing method is employed to extract location mentions within the textual tweet. Therefore, an approach has been designed for locating and detecting crisis events on a local scale and present the resulting real-time database. Afterward, these locations are employed to detect sudden rushes in the number of event-related tweets linked to governorates in Egypt and their administrative subdivisions, such as “Ahya” and “marakez”. GeoNames database 1 . In this work, such a rush is determined by the number of event-related tweets within a region during a specific time range.

Accordingly, a database including known geographic places (a gazetteer) was applied to match a text tweet to one or more real locations. Thus, additional geographic information acquired from the tweet itself to detect the actual locations that the user mentioned in the tweet.

The GeoNames database has been used to build our gazetteer. it is a geographical database containing over 4 million administrative divisions, cities, towns, and villages. it has also including their administrative parent area, population, and coordinates locations. the dataset has alternative names, like abbreviations, and dialect terms, (e.g., for "مصر الجديدة", it includes, for example, Misr al Jadidah, msr aljdydt), and the language of each alternative name. The Arabic tweets text and their metadata (as GPS coordinates of the machine from which the text tweet was sent) were

collected via the Twitter streaming API using many keywords as (سيول, امطار, منخفض التينين, حريق, كورونا) about Various types of events, like a flood, fire, and coronavirus covering a considerable part of Egypt. For mapping the tweets events and ongoing monitoring of events, a dataset consists of 297,150 tweets, posted between October 18, 2019, and July 14, 2020; is used.

The methodology consists of five main phases, namely: filtering, geoparsing, classifying, grouping, and finally scoring.

Figure2-1 explains the procedure followed by the new approach. First, the tweets are collected during specific day hours within an event and each tweet are analyzed. All tweets are grouped according to the three spatial indicators in which the (toponym groups) are the most important (Grouping phase). each of the text tweet locations was determined by a score showing how well it matched the tweet’s spatial information. Then, the total score is computed for each tweet by determining the event type, toponym, and damage during the timeframe (Scoring phase). In addition, a toponym table was made to store the toponyms, their coordinates locations, and their scores for each tweet. This table is later utilized to geoparse tweets in real-time. Once locations had been detected in the tweets, the same phases were applied, which contained new incoming tweets.

2.1 Filtering

Ali et al. [2] approach is applied for filtering tweets such that if that tweet contains an Arabic text in addition digits, URLs, non-Arabic letters, this tweet will be filtered, the Arabic text only will be saved, and the other parts will be removed. The result was a list of extracted locations for each toponym mentioned in the tweet text.

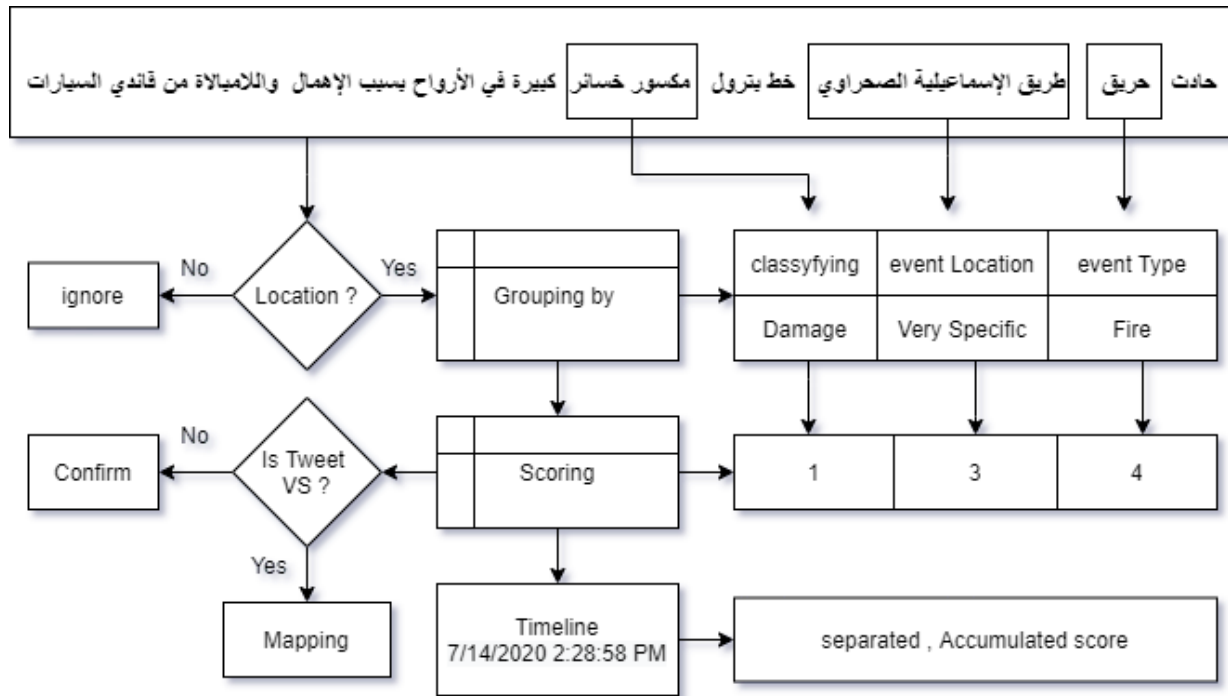


Figure2-1: Overall Model Procedure for Location and Timeframe.

2.2 Geoparsing

In this phase, the text of each tweet is split into individual words, a process also referred to as “tokenization”. then, these tokens, such as “Al-Maadi”, as well as consecutive tokens, such as “Cairo” are matched to a set of geographical locations extracted from the GeoNames database. To recognize extracted locations for a tweet, a tweet’s text was matched to the gazetteer. Tweets are often written in Egyptian dialect.

further filtration of the results is applied to obtain the extracted locations by extraction of the mentions of locations higher (toponym parent) or lower (toponym child) in the hierarchy. For example, "Al-Orouba Tunnel" is a geographical child of "misr al-Jadida hay", which is a geographical child for Cairo governorate.

2.3 Classifying

The word “flood” and its translations are often used figuratively (e.g., “سيول الدموع”) or in transferred sense (e.g., “a flood of tears”). In both cases, the tweets do not always refer

to current ongoing flood (i.e., disaster) events. on the contrary, they can refer to not only a current event, but also to a future (forecasted events) or to the past (historic events). To classify the tweets, we can use natural language processing algorithms [8]. Here, we used the pre-trained model to classify tweets in two categories [2] : related to an ongoing event and have damage (i.e., “damage”) and related to the ongoing event and have not damage (i.e., “no damage”). this algorithm is a machine learning- based natural language processing (NLP) model that learns relationships between words and sub-words in a text (i.e., word embeddings) and utilizes these to encode the text, to make predictions.

2.4 Grouping

The internal steps of grouping were defined as the following: First, the tweet was grouped by event type as (fire, flood, coronavirus). then, the event type tweet was grouped by the toponym location. last, the toponym tweet was grouped by damage or not (Table 2-1).

All tweets were grouped by a (toponym group) pipeline as follows:

- Very Specific (VS): the tweet which contains a toponym for a very specific location that can be drawn immediately on the map like "Al Orouba tunnel", and it is considered the most important tweet.
- Specific (S): the tweet that includes a toponym for a specific location as "hay", "Markaz" or village e.g. "misr al-jadida hay", and it is considered an important tweet.
- Governorate (G): the tweet that contains a toponym for a location as the city governorate like Cairo, and it is considered a less important tweet.
- Noise (N): the tweet does not contain a toponym, and it is excluded.

We assumed that many tweets that mentioned the same toponym within a given event timeframe referred to the same location. For instance, when a fire occurred in "Cairo Ismailia road", we expected that all users mentioning "fire" and "Cairo" were referencing the Ismailia and Cairo governorate in Egypt. Then, all text tweets mentioned the same toponym was grouped together. For that, the higher number of tweets mentioning a location, the greater was the related group. Since tweets could contain many toponyms, a problem could be encountered as each tweet could belong to more than one group. Therefore, this problem can be resolved as follows: (a) if each tweet includes more than one toponym for (specific group) as "misr al-Jadida" and another toponym for (governorate group) as "Cairo". The (specific group) would be taken into consideration because it has the most important in the scoring and the most specific of the event location. (b) when each tweet contains more than one toponym of (very specific group) as " Al Orouba tunnel" and a "Cairo airport Lounge". Each toponym would be inserted in a separate pipeline to scoring and mapping them on the map.

2.5 Scoring

For each tweet, for which we found extracted locations, as illustrated in the geoparsing detection, its spatial scores were matched to each extracted location. We use these scores to explain the importance of each tweet in (toponym group) clarification. To understand our approach, two questions must be answered: First, What are the mapping tweets? Just all (VS) tweets that were plotted on a crisis map as shown in Figure2-1. Second, what are confirmation tweets? all Tweets that are in the same event area and have the same timeframe. In addition, these tweets confirm the spatial information for the (VS) Tweet. In this phase, the event type scores were created according to their intensity, as follows: Coronavirus of the second degree, a flood of the third degree, the fire of the fourth degree. e.g. 2, 3, 4 Respectively, where score 4 is considered the most important and intensity. Assigning this typical score will be useful for determining the level of priority in handling crises.

Table 2-1 Overview of Grouping and Scoring

toponym	group	score
Cairo	G	2
Misr al Jadidah	S	3
Al Orouba Tunnel	VS	4
No location	N	0

For each tweet, a location score was determined according to its importance as in (grouping phase) Shown in section 2.4. Then, the total score of the tweet was calculated for three spatial indicators as in (Table 3-1). A higher total score for each tweet means higher confidence that an extracted location is correct for the event presence. also, these tweet scores confirm the spatial information for the (VS) tweet.

3 Spatial information indicator

The scoring system is used to confirm (VS) tweets group and mapping these tweets on the crisis map during timeframe of the event by scoring the other tweets groups referenced shown in section 2.5. A summary of the scores toward each of the three spatial indicators (for the toponym group, event type, and damage) is presented in Table 3-1.

Table 3-1 Special Indicators.

indicator	term	score
event	fire	4
Toponym group	Very specific	4
damage	Broken, losses	1

These scores are summed to obtain the total score (maximum of 9) per tweet, which indicates the mapping of tweets. The following points describe the process for each of the metadata and textual spatial information:

- User can determine their hometown in their user profile. However, many differences are probable, including fantasy places, various locations, and incomplete data entries. for instance, a user who lives in Alexandria might enter Cairo in the location field. Therefore, this spatial information is not considered.
- The Tweets sent from the same device IP (PC, Mobile) within the same timeframe for the event are excluded to protect the security and credibility of the mapping tweets and confirmation.
- geolocation field (metadata) is utilized if available to obtain an accurate mapping. but this additional information is not often available because it statistically represents 1% for location coordinates in all tweet's dataset.

4 Updating toponym table

In addition, the toponyms are saved in a toponym table as introduced in section 2.

That table showed the location with the highest score per tweet (for the toponym group, event, and damage). This toponym table is continuously refreshed and used to geoparse new incoming tweets.

5 Event detection

To detect events from social media, an event detection step is required. i.e. to predict the beginning and the end of the event. Various approaches have been developed for detecting Twitter activity [6] [7] [9]. These approaches involve sequentially checking whether the time between consecutive tweets is shorter than a certain time threshold. In this work, an algorithm is developed to detect event on each incoming tweet rather than at the end of a fixed time. thus, it works both for unexpected events and that developed in minutes and hours. The approach is expanded in two ways:

1. By calculating the fluctuations in numbers of tweets with spatial scores according to time as shown in Figure 5-1. In other words, Score per tweet S_t (sum score of three indicators) was calculated as shown in section 3. then, using equation 1, Separated score per minute Sep_m (total score per minute) was calculated.

$$Sep_m = \frac{\sum_{t=1}^n S_t}{60} \quad (1)$$

Last, using equation 2, accumulated score overtime until the end of event Sac_m was calculated which t tweet number; m minutes number (see Figure 5-2).

$$Sac_m = \frac{\sum_{t=1}^{n_m} S_t}{m \times 60} \quad (2)$$

2. By employing a threshold for the detection of an event.

A “disaster event” state with Twitter activity is defined as the time zone between a

specific start and end of event during which the disaster event occurs, also known as an event area. A “natural state” is defined as the time zone through which no disaster event occurs. The approach keeps track of the scores in time passed between consecutive tweets defined to each area. By analyzing the score and time variations between scores these tweets and comparing them to a disaster area-specific threshold. The state of the area is considered a "disaster event" state during higher tweets intensity and a "natural event" state when tweets intensity reverts to natural levels of activity.

The rate of tweet scores is computed every minute and the difference between accumulated scores is calculated. If the difference between the tweet scores at every consecutive minute is higher than the threshold value θ_S , the algorithm will assign "TRUE TWEET". The first "TRUE TWEET" will be considered as the start of the event. If the rate between the tweet scores is lower than the threshold values, the algorithm will assign "FALSE TWEET". The event is considered to end if "FALSE TWEET" were found through 15 consecutive minutes θ_E .

Figure 5-1 presents the fluctuations in numbers of tweets with spatial scores according to time on the graph between separated score and accumulated score.

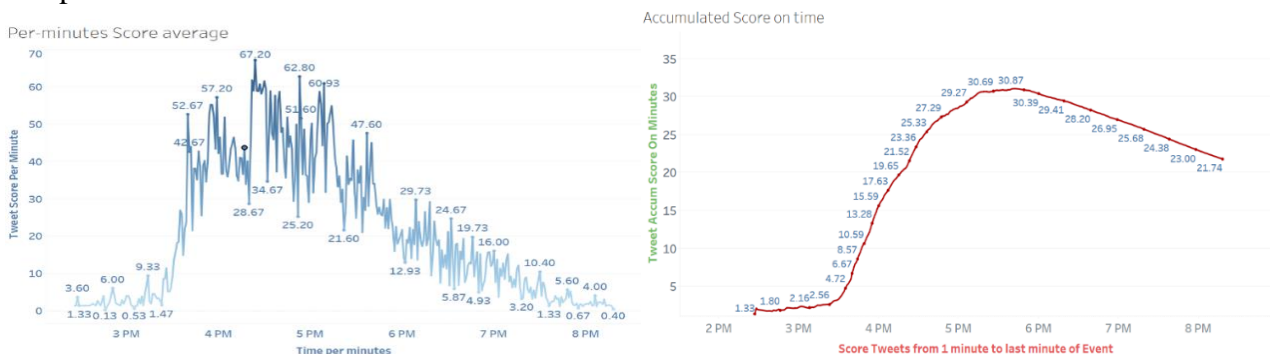


Figure 5-1: Fluctuations in Numbers of Tweets with Spatial Scores According to Time.

To get the threshold value θ_S , four case studies are analyzed at only the first day of the event:

- Case 1: flood in the “Al-Orouba tunnel, Cairo governorate” on 22 October 2019.
- Case 2: Coronavirus in “Belqas markaz, the Dakahlia Governorate” on 12 March 2020.
- Case 3: flood in “Zarayeb Helwan, Helwan Governorate” on 13 March 2020.
- Case 4: Fire in “Cairo Ismailia Road” on 14 July 2020.

The target is to find the average maximum slope between two consecutive accumulated scores (every minute). The following

- (c) zarayeb Helwan flood
- (d) Ismailia road fire

Figure 5-3 shows the accumulated scores for every case consequently.

Table 5-1 indicates the maximum slope between two consecutive accumulated scores for every case. *The threshold value is taken as the average of these values.*

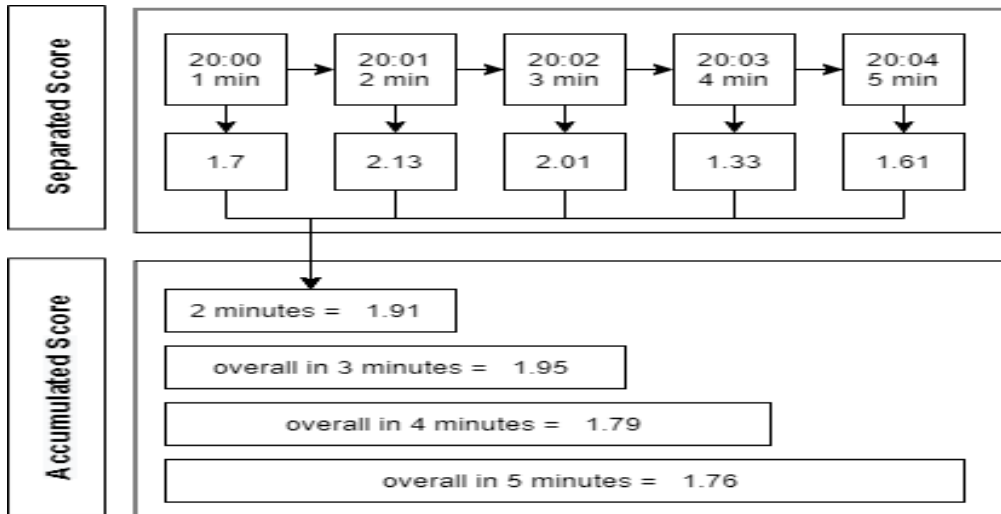
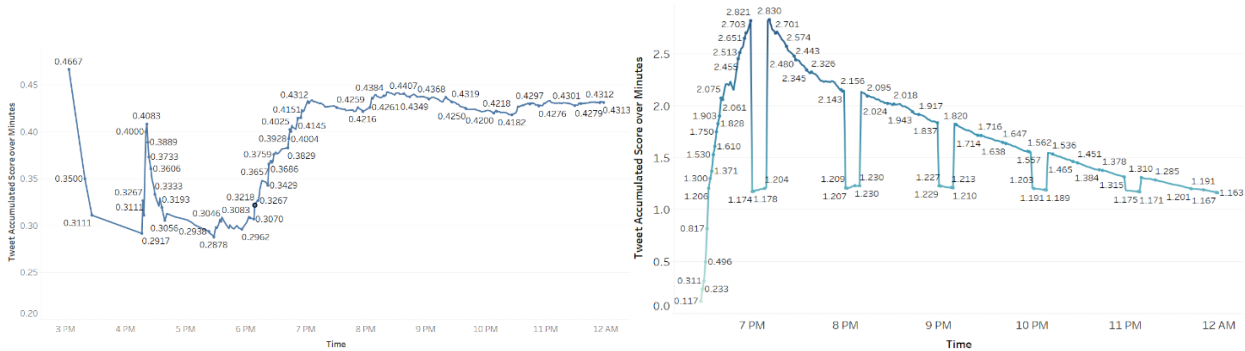
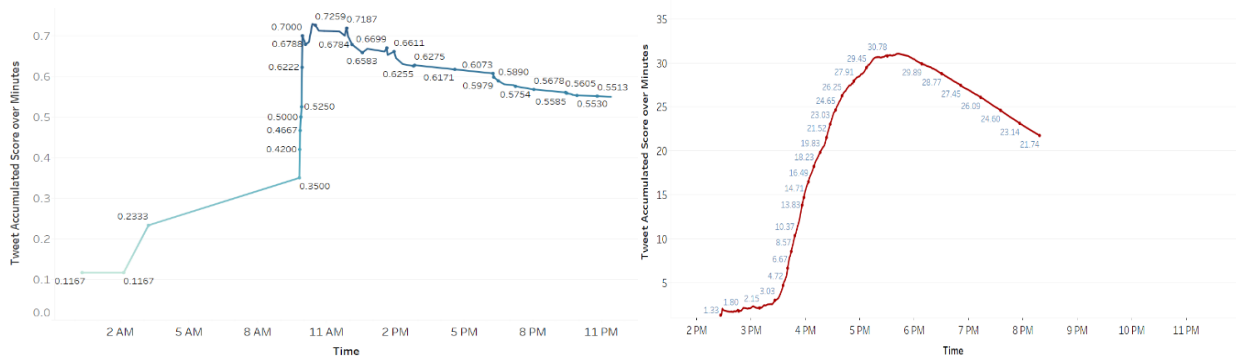


Figure 5-2: calculating the separated score and accumulated score in the timeframe.



(a) Al-Orouba flood

(b) belkas coronavirus



(c) zarayeb Helwan flood

(d) Ismailia road fire

Figure 5-3: Accumulated scores of tweets for: (a) case 1, (b) case 2, (c) case 3, (d) case 4.

Using equation (3), the decision is calculated according to the following empirical inequity condition:

Table 5-1 Maximum Slope Value

Event	Max slope
Case 1	0.466
Case 2	0.388
Case 3	0.116
Case 4	1.333
Avg.	$\theta_s = 0.576$

$$\Delta Sac_m > (1 + \theta_s) \times \Delta Sac_{m-1} \quad (3)$$

Where ΔSac_m and ΔSac_{m-1} are the deference between two consecutive accumulated scores.

6 Experimental results

For each group, the number of geoparsed (very specific) tweets based on their calculated scores between October 2019 and July 2020 was plotted (

Figure 6-1) against population density 2 and urban regions over this period. Also presents the four cases mentioned above. This provides an impression of how twitter reporting correlates to the population density and urban regions. The data made clear that in high population density (Oruba Tunnel and Cairo Ismailia Road) urban regions, there were about six orders of magnitude more tweets than in low population density (Zarayeb Helwan and Belqas) non-urban regions.

On the other hand, when the Coronavirus pandemic gained attention and fear of people around the world, not only in Egypt, it was noticed that the volume of tweets mentioned in "Belqas" is relatively high, although it is considered from non-

urban areas with a low population density. As the tweets mentioned in the "Belqas" were sent from all governorates of Egypt, not from it only, which clarify the spread of the Coronavirus in it. This has happened at the same timeframe in which the Coronavirus infections and deaths have spread at a rapid rate since the beginning of its appearance at Egypt in February 2020.

We have applied our approach to the 297,150 tweets in a historical dataset, applying as if the data were prepared in real-time. Of this dataset tweets, we found that 43.2% mentioned governorate locations, 23.8% (very specific) tweets, 24.4% (specific) tweets, and 8.4% (noise) tweets (see Table 6-1).

Table 6-1 Results of the Automated Geoparsing of Dataset tweets.

	No. of tweet	%
total	297,150	----
governorate	128,419	43.2%
specific	72,570	24.4%
Very specific	70,968	23.8%
noise	25,193	8.4%

In addition, when distinguishing between administrative levels, roughly 50% of the locations mentioned refer to governorates, while cities and the lower administrative level locations for each account; roughly 25% of the mentions. These findings suggest that not only the density population or the twitter user base but also the events are responsible for the high number of tweets during the studied timeframe. The points in

Figure 6-1 demonstrate that in general, more event tweets seemed to be linked to greater levels of event damage over the study period. These relations are influenced by

several other factors, including variations in the extent of twitter usage per region, the urban regions, and rural regions.

This explains why in regions that suffered from disastrous fire or flood events that caused meaningful damage, the high numbers of tweets about these disasters were generated.

It is also noticed that in Densely populated urban areas with high internet spread, a large number of people refers to the (toponym parent) level and (e.g., “سيول القاهرة تسببت في غرق وخسائر كبيرة”) , while in rural areas with low population density where internet spread is much lower, the event is frequently referred to the (toponym child) level. (e.g., “غرق منطقة زرايب حلوان ووفاة اشخاص”).

6.1 Mapping tweets

Of all the 297,150 tweets for the three events, only very specific tweets were mapped with

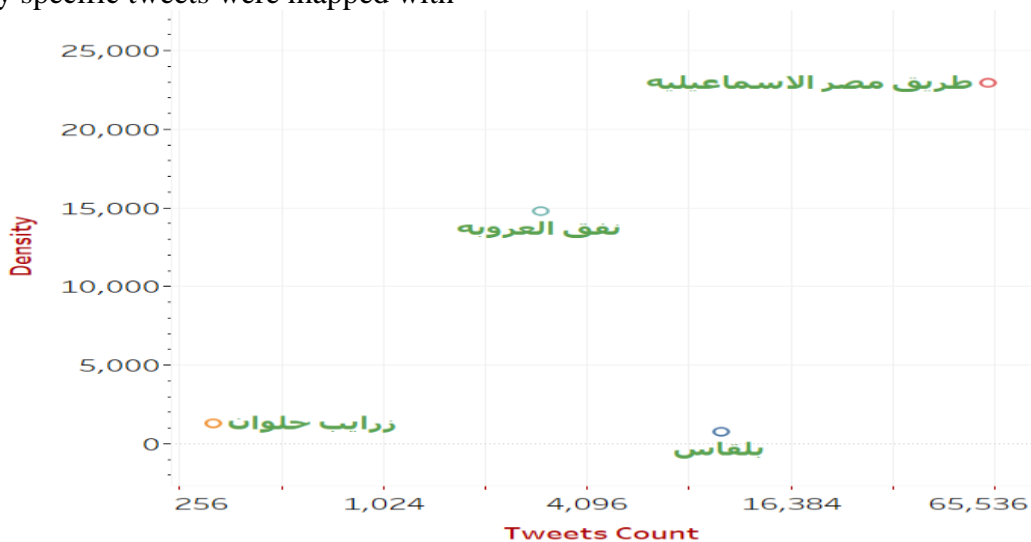


Figure 6-1: The Number of Geoparsed Tweets in Four Locations Relative to their Population Density for Three Various Events.

a geo-location toponym in the event timeframe. Figure 6-2 shows the distribution of these tweets on an Egypt map for the four selected regions. "Al-Oruba tunnel flood " (2981 tweets), "Belkas coronavirus " (10,183 tweets), "Cairo Ismailia Road fire " (62,384 tweets), and "Zarayb Helwan flood " (323 tweets). The details were shown in Figure 6-2 (a), (b), (c), and (d) respectively using Google Data Studio tools.

The map plots the tweets published in the hours following the start of the event. From the toponym table that created in section 4. These maps show that as time elapses, the news spreads across the Cairo governorate (and beyond). In the first quarter after each event, 77% of the (mapping tweets) originated from the event location. This percentage reduces to 15% and 4% for the second and third periods, respectively.



Figure 6-2: Mapping a Very Specific Tweets for Four Regions.

6.2 Visual Analysis

The number of tweets identified by the search query per minute was presented. Figure 6-3 presents these tweets six hours after the fire event started. The figure illustrates the sudden rises in tweets which are an obviously visible sign that something is worthy monitoring has attracted the user's attention. Figure 6-3 shows a small number of 9 tweets

per minute at 2:26 PM, and the growth soon after the start event at 3:30 PM. The growth reveals an erratic pattern and peaks at 4:24 PM (504 tweets per minute), then follows a descending trend. At 8:18 PM the number of tweets has fallen near-zero, but at a midnight, new tweets appear despite the reduced rate after the event ends.

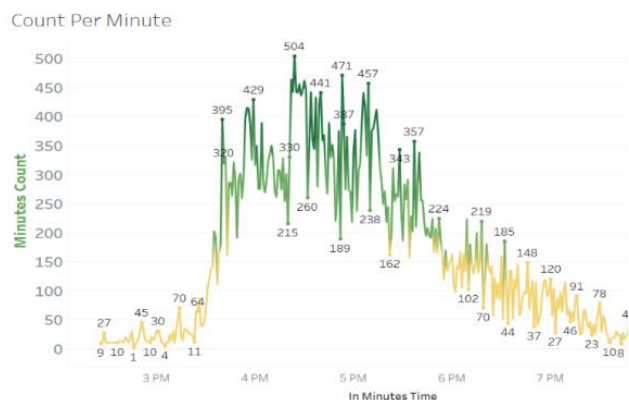


Figure 6-3: Number of Tweets Per Minute During Fire Event.

The content analysis of tweets at frame time was presented. The set included all 38,228 tweets published during and after the event (between 3:40 PM and 5:37 PM). Clearly, the analysis of tweets that were published after the event focuses on the filtering of damage and casualties related to tweets. Also, a sudden increase in tweets during those two hours was noticed because the fire happened in the region on the main highway between Cairo and Ismailia (Misr-Ismailia Road). the time workers have finished their jobs and used this road. so, the tweets' interaction was greater after people attended the fire event. The reason was the increased Population density and traffic density in this period.

Using the method in section 5 to the event detection, we resulted in the threshold for the start of an event and the condition of the end of an event with respect to the fire on "misr-Ismailia road". It was found that the score difference was higher than the estimated threshold value $(1+\theta_5)$ at 14:28 and the condition θ_E was accomplished at 19:03. Thus, the action and reporting decision is taken from the third minute when the fire occurs.

6.3 Damage information During and after the event

The tweets were analyzed by topic filters of the damage information. this topic is critical because damage and casualty information can assist indicators for the impact of an event. So, information can support crisis managers to allocate resources based on determining request.

The "damage tweets" were identified by filtering tweets on keywords: 'ضرر', 'انهيار', 'غرق', 'ضحايا', 'خسائر', and conjugations of these terms. Figure 6-3 shows that the size of "damage tweets" per minute increased in the 30 minutes after started the fire event, peaked at 4:24 PM when damage and casualty information was officially advertised by national news media, and subsequently decreased. Furthermore, (51%) of the tweets about damage contained hyperlinks to other social media and news media websites. Tweets that referred to news media (e.g. <https://www.idsc.gov.eg/>) clarifying what properties were damaged.

7 Discussion

Many of the "damage tweets" included links to uploaded pictures and videos, leaving no doubt about the credibility of the information about the damage. Therefore, only when the rumors about deaths were confirmed by official news media, the number of tweets increased significantly, many of which were retweets. This infers that social norms on Twitter restrict the propagation of uncertain information about sensitive topics. Social media also contribute clear opportunities for emergency communication between media, authorities, and citizens. For example, on Twitter crisis managers can interact with citizens and media immediately by confirming or denying rumors. For instance, by plotting the tweets on a map, crisis administrators can determine potential 'hot spots'; i.e. locations where people are concerned about.

Second, analyses showed that classifying tweets on ‘damage’ would have given crisis managers with graphic evidence of the situation on the ground. Because Twitter gives a unique source of original and real crisis event.

Third, tweets that especially get several retweets attract care and are a piece of evidence that twitter users trust the information presented in the original tweet. Peaks in the tweet intensity usually contain several retweets on a ‘hot event’, such as news about casualties and damage. These retweets may have significant implications for crisis management.

8 Conclusions and Future work

There are many methods for future research. First, the preparation of a multi-agent-based prototype for crisis management is important. The purpose of the future work is to enable the administrators to manage fire or flood by contributing advice to support the decision process and response. The proposed framework oriented towered firefighting and suppression. This idea that is assisted by a fire control based expert system agents, GIS agent, and other sophisticated web-based services are presented. This prototype is to give firefighters with effective management of resources and activities that take place in handling fires and civil defense men in flooding.

second, early detection of risks and rumors are relevant topics for improving crisis communication. The previous studies showed that only a small proportion of tweets included information about the coming storm and rumors about deaths. Thus, the approach can be developed to enable early detection of risks and rumors. It seems that retweets play an important role in how twitter users infer the source credibility of a tweet. For example, during the Red River flood (local) media were retweeted relative frequently [10]. Developing indicators for real-time source

credibility of Twitter user accounts during crises would be a value for the performance of crisis communication [11].

References

- [1] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors," 2010, pp. 851-860.
- [2] y. A. Ameen, "Classification of Arabic Tweets for Damage Event Detection " International Journal of Scientific & Engineering Research, vol. Volume 11,, no. Issue 4,, April-2020, p. 160:166, 2020.
- [3] K. Lee, R. Ganti, M. Srivatsa, and P. Mohapatra, "Spatio-temporal provenance: Identifying location information from unstructured text," 2013: IEEE, pp. 499-504.
- [4] S. E. Middleton and V. Krivcovs, "Geoparsing and geosemantics for social media: Spatiotemporal grounding of content propagating rumors to support trust and veracity analysis during breaking news," ACM Transactions on Information Systems (TOIS), vol. 34, no. 3, pp. 1-26, 2016.
- [5] W. Zhang and J. Gelernter, "Geocoding location expressions in Twitter messages: A preference learning method," Journal of Spatial Information Science, vol. 2014, no. 9, pp. 37-70, 2014.
- [6] H. Sarmiento, B. Poblete, and J. Campos, "Domain-Independent detection of emergency situations based on social activity related to geolocations," 2018, pp. 245-254.
- [7] B. Jongman et al., "Declining vulnerability to river floods and the global benefits of adaptation," Proceedings of the National Academy of

Sciences, vol. 112, no. 18, pp. E2271-E2280, 2015.

- [8] F. Atefeh and W. Khreich, "A survey of techniques for event detection in twitter," *Computational Intelligence*, vol. 31, no. 1, pp. 132-164, 2015.
- [9] W. J. Riley, "Algorithms for frequency jump detection," *Metrologia*, vol. 45, no. 6, p. S154, 2008.
- [10] K. Starbird and L. Palen, *Pass it on?: Retweeting in mass emergency. International Community on Information Systems for Crisis Response and ...*, 2010.
- [11] K. Starbird, L. Palen, A. L. Hughes, and S. Vieweg, "Chatter on the red: what hazards threat reveals about the social life of microblogged information," 2010, pp. 241-250.