

Analysis of Major Events in Egypt Based on Twitter

Hana Anber^{1*}, Akram Salah², A. A. Abd El-Aziz³

¹ Computer and Information Sciences Department, Institute of Statistical Studies and Research, Cairo University, Giza, Egypt.

² Computer Science Department, Faculty of Computers and Information, Cairo University, Giza, Egypt.

³ Information System and Technology Department, Institute of Statistical Studies and Research, Cairo University, Giza, Egypt.

* Corresponding author. Tel: +2 01062628474; email: hana.anber@gmail.com

Manuscript submitted March 10, 2017; accepted June 20, 2017.

doi: 10.17706/jcp.13.3.327-336

Abstract: The number of users on Twitter has increased in the Middle East since the Arab Spring in 2011. Because Egypt has become a changing country in the last few years, it faced a number of different events; that made Twitter the most appropriate virtual environment to monitor and track those events. In this paper, we study the life cycle for each event to identify the peak time and to investigate what are the most events affected on Twitter users. We tracked the most major hashtags occurred in Egypt within the period 2014–2015 and classified each one according to its impact. We have found that the peak time of the Planned-Events was on the day before and on the event day. Most of the Sudden-Events' peak time was on the first day of the event. We also have found that users engaged more in political and public opinion events than other events.

Key words: Data analysis, Egypt, event life cycle, twitter.

1. Introduction

The growing phenomena of social media, such as: Twitter, Facebook, and Instagram, with each one has its own characteristics and its usages, are constantly affecting our societies. For example, Facebook is considered as a social network where everyone in the network has a reciprocated relationship with another one in the same network. The relationship in this case is undirected. Conversely, in Twitter everyone in the network does not necessarily have a reciprocated relationship with others. In this case, the relationship is either directed or undirected.

We focus on Twitter for data analysis, where Twitter is an online networking service that enables users to send and read short 140-character messages called "tweets" [1]. In addition to its publicity, Twitter is accessible for unregistered users to read and monitor most tweets, unlike Facebook where users can control the privacy of their profiles. Twitter is also considered as a source of information. Since [2], and [3] agreed that Twitter is a source of information than a social networking.

In the last five years since the Arab Spring in 2011, the number of users on Twitter has been increased. Egypt has become a changing country in the last three years since 30th June 2013 revolution against Muslim Brotherhood which faced a number of different events along this period; that enthused us to track and monitor different events.

In this paper, we need to track the major events occurred in Egypt in order to identify the peak time for each event. We assume that each event has a starting point until the frequency of data reaches the peak "The

highest density of tweets over the time". To investigate that assumption firstly, we need to classify each hashtag under a certain category in order to enhance the analysis by categorizing each one and compare between them to examine the most affected event; secondly to employ a statistical analysis to identify the life cycle for each hashtag.

2. Related Work

Table 1. Hashtag Tracking Period

Date		Hashtag	Definition
From	To		
Nov 23, 2014	Nov 30, 2014	#28نوفمبر	"28 November", Egypt's ultra-conservative Islamist "Salafist Front" called on their allies to protest on 28 November to impose the Islamic identity without disguise.
Feb 8, 2015	Feb 9, 2015	#شهداء_الزمالك	"Zamalek martyrs", This catastrophic incident occurred on 8 th Feb 2015, when twenty Zamalek football fans were killed in stampede at stadium.
Feb 14, 2015	Feb 16, 2016	#شهداء_مصر_في_ليبيا	"Egyptians martyrs in Libya", This incident occurred when the Coptic Egyptians killed by ISIS (Islamic State of Iraq and Syria) in Libya.
Jul 22, 2015	Aug 11, 2015	#قناة_السويس_الجديدة	"The New Suez Canal", is the name of a waterway project in Egypt, to expand the capacity of the existing Suez Canal [13]. That was lunched on 6 th August 2015 by president Abdel Fattah el-Sisi.
Sep 2, 2015	Sep 3, 2015	#اكتشاف_أكبر_حقل_غاز_بمصر	"Discovery of the largest natural gas field in Egypt", that hashtag became a trend after Italy's ENI announced that it has discovered the largest offshore natural gas field in the Mediterranean off the Egyptian coast.
Sep 8, 2015	Sep 9, 2015	#وزير_الزراعة	"The Agriculture Minister", that hashtag became a trend after the corruption case of the Egyptian Agriculture Minister, Salah Helal.
Sep 12, 2015	Sep 14, 2015	#استقالة_حكومة_محلل	"Mahlab cabinet resignation", that hashtag became a trend after the prime minister; Ibrahim Mahlab announced his resignation, and after the pervious event showed up.
Sep 13, 2015	Sep 14, 2015	#الوحدات	"El- Wahat", this accident occurred on 13 th Sep 2015, when Egyptian forces hunted militants mistakenly attacked the Mexican tourists' convey in Egypt's western desert.
Oct 5, 2015	Oct 7, 2015	#انقذوا_الانترنت	"Save the Internet", that hashtag became a trend after the Ministry of Communication decision to block the voice chat of the most popular mobile applications, such as Whatsapp, Skype, Viber and the online games as well.
Oct 29, 2015	Oct 31, 2015	#ريهام_سعيد	"Riham Saeed", that hashtag became a trend on 30 th Oct 2015, after the TV presenter, Riham Saeed abused a victim on air.
Nov 2, 2015	Nov 3, 2015	#قتلوا_المصري	"They killed the Egyptian", that hashtag became a trend after the accident occurred in Kuwait where Kuwaiti nationals ran over an Egyptian citizen by their vehicles.

Analyzing social media data specifically Twitter has become an active research area. Most studies focused on the topological characteristics of Twitter. In [4], five different hashtags were demonstrated and analyzed by tracking the most uprising political events, such as #FreeIran, #FreeVenezuela, #Jan25, #SpanishRevolution, and #OccupyWallSt, to investigate the correlation between retweets, mentions, followers, following and volume of messages by studying the network topology.

What makes the life cycle of each hashtag persist is the propagation of the messages, since studies [4], [5] examined the message propagation by studying the first-network topology and by the retweet rate, where [6] studied the message content. According to the study in [4], it is found that bloggers spread information more than other categories like celebrities, media, or organizations.

The period of time for each hashtag should be consistent. For example, when crawling a specific hashtag

each one should be measured yearly, monthly, weekly, daily, or hourly. Unlike [4], the analysis of the five political hashtags was not consistently measured. The time breaks have different time measures.

Since Twitter is a source of information, [7] found that users reply in breaking news messages than ordinary messages, that is, users discuss and share information and ideas towards a specific topic more than engaging in conversations. That resulted in increasing the network of users in breaking news events.

According to our previous survey in [8], most studies focused on the topological characteristics in Twitter, as well as focusing on the political and disastrous hashtags. In this paper, we need to track different hashtag categories, and we need to study the behavior of life cycle for each hashtag in order to examine the peak time for each one. We also need to analyze the activity of users towards such events according to the frequency of tweets within a given period of time.

3. Data Collection

There are different types of Twitter data, such as user profile data and tweet messages. The former is considered static, while the latter is dynamic. Tweets could be textual, images, videos, URL, or spam tweets [4]. In this paper, we focused on tweet messages within certain hashtags. Since one of the most popular problems when retrieving Twitter data is the rate limit, which is limited to 180 queries per 15-minute per window [9].

The huge amounts of data on social media become a challenging topic. To tackle this problem, we focused on certain Twitter hashtags that is related to the major events occurred in Egypt from two to twenty-one consecutive days. The retrieved data comes in JSON format [10], which is considered as semi-structured data.

We employed Twitter Data Crawler which uses Twitter Streaming-API to collect the required data in real-time [11]. We implemented a Python Listener to collect those data using the Tweepy library [12].

We crawled different hashtags within the period of 2014 and 2015 and we collected data for eleven hashtags. Table 1 shows the collected hashtags with the tracking period and the definition for each hashtag. Regarding the difficulty to track hashtags for a long period of time, most hashtags were tracked for two or three consecutive days, while others for eight and more.

4. Analysis of Twitter Data Feeds

Accessing eleven hashtags provides an opportunity to run a statistical analysis to identify the life cycle of each event, and to investigate the highest frequency of tweets over the given period. To do such analysis, we employed the well-known Statistical Analysis Tool and Language, R [14]. We need to investigate the peak time for each event, whether it is before the event, in the same day, or after the event day? Before answering these questions we need to classify each event according to its impact.

4.1. Event “Hashtag” Classification

Table 2. Classification of Events

Planned- Events	Sudden- Events
28 نوفمبر "28 November".	#شهداء_الزمالك "Zamalek martyrs".
#قناة_السويس_الجديدة "New Suez Canal"	#شهداء_مصر_في_ليبيا "Egyptian martyrs in Libya".
	#اكتشاف_أكبر_حقل_غاز_بمصر "Discovery of natural gas in Egypt".
	#وزير_الزراعة "Agriculture Minister".
	#استقالة_حكومة_محلل "Mahlab cabinet resignation".
	#الواحات, "El- Wahat".
	#انقذوا_الانترنت "Save the Internet".
	#ريهام_سعيد "Riham Saed".
	#قتلوا_المصري "They killed the Egyptian".

Table 3. Classification of the Planned-Events

Event "Hashtag"	Category
28#_نوفمبر "28 November".	Political
القناة_السويس_الجديدة Canal" #قناة_السويس_الجديدة	Economical

Table 4. Classification of the Sudden-Events

Event "Hashtag"	Category
شهداء_مصر_في_ليبيا martyrs in Libya". الواحات, "El- Wahat". #وزير_الزراعة Minister". #استقالة_حكومة_محلِب cabinet resignation".	Political
اكتشاف_أكبر_حقل_غاز_بمصر "Discovery of natural gas in Egypt".	Economical
شهداء_الزمالك martyrs". #انقذوا_الانترنت Internet". #ريهام_سعيد "Riham Saed". #قتلوا_المصري Egyptian.	Public -Opinion

According to Twitter, users use the hashtag symbol # before a relevant keyword or phrase in their tweets to categorize those tweets and help them show more easily in Twitter search. In [6], and [15], the authors found that using hashtags in tweets improves the accuracy and the performance of the analysis.

In Table 2 we classified those hashtags according to their impact to two main categories: the "Planned-Events" and the "Sudden-Events", the first one is intended to occur at a certain day. The second one occurs abruptly without prior intentions. Regarding the various types of those hashtags, we classified the main categories into two and more sub-categories. As shown in Table 3 and Table 4.

4.2. Event Life Cycle

In this analysis, we are interested in measuring the frequency distribution of tweets over a period of time in order to identify the peak time for each event. For such, we analyzed the dated tweets for each hashtag. To run the analysis, we employed a graphical representation to discover and show the underlying frequency distribution, which is called the Histogram.

To contrast a Histogram from continuous data, first we split the data into intervals; called bins "Time" with each bin representing an hourly period. Each bin contains the number of tweets in the hashtag. The results of the analysis are shown in Fig. 1.

By analyzing hashtags for The Planned-Events, we make a few observations. We observe from the above figure. In hashtags 28#_نوفمبر (28 Nov) and #قناة_السويس_الجديدة (new Suez Canal) both events reached the peak on the day before and on the event day. We can also see that the frequency starts to diminish afterward. In #قناة_السويس_الجديدة (new Suez Canal) there was a noticeably increase in the frequency of tweets on July 25; we investigated that increase by monitoring the tweets, since Egypt began the first trial of the new Suez Canal on Saturday 25th July with three containers ships beginning to traverse the channel. We can conclude

from the two previous events that the frequency distribution of the Planned-Events reaches their peak one day before and on the event day. We can also see that users were more concerned with the economical event than with the Islamist political event, since users were more active in *#قناة_السويس_الجديده* (new Suez Canal) than in *#نوفمبر_28* (28 Nov).

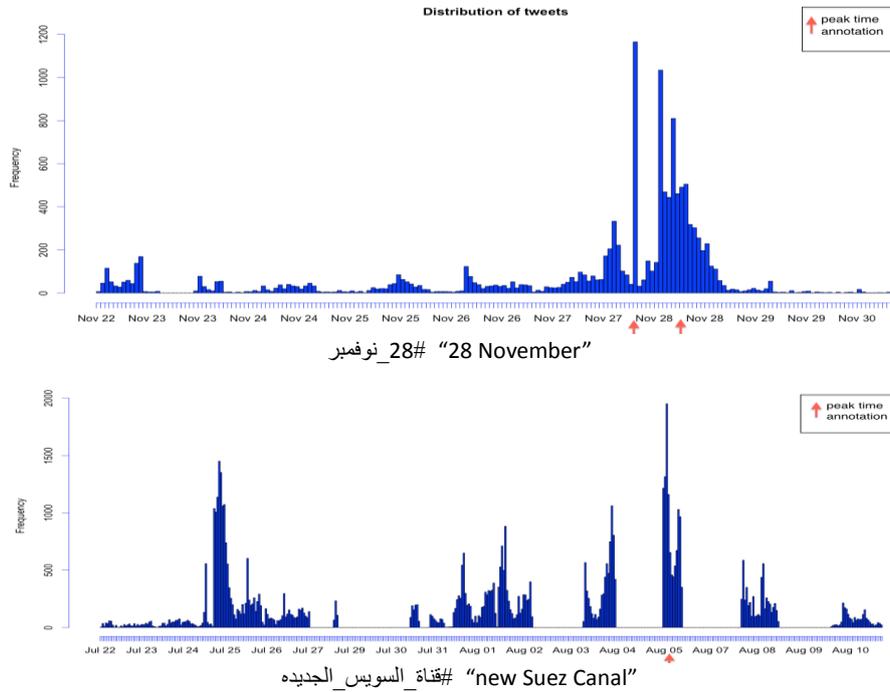


Fig. 1. Planned-events frequency distribution.

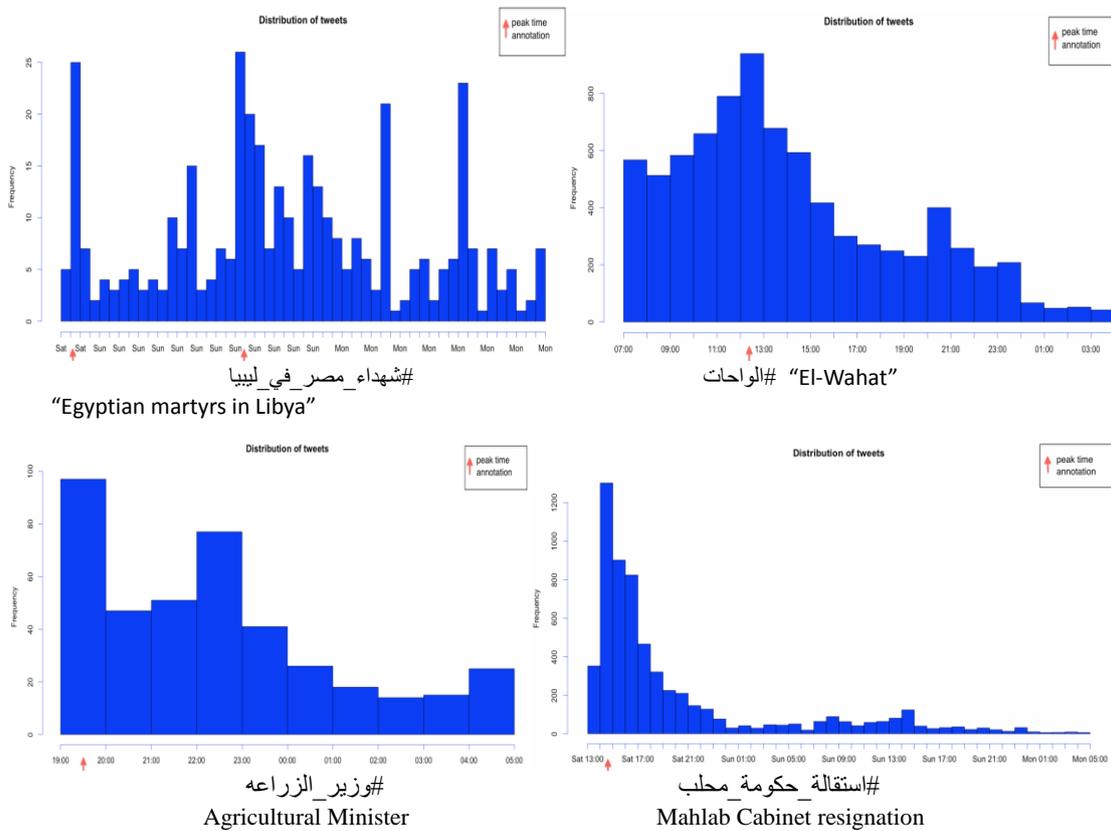


Fig. 2. Political sudden-events frequency distribution.

We can see from the analysis, the results of the political Sudden-Events in Fig. 2. We observe that the peak time for the three hashtags (#الواحات “El-Wahat”, #وزير_الزراعة “Agriculture Minister” and #استقالة_حكومة_محلِب “Mahlab cabinet resignation”) were on the first day of each event until the frequency distribution diminished gradually the days after. In #شهداء_مصر_في_ليبيا (Egyptian martyrs in Libya) there was no specific peak time as the frequency of tweets fluctuated in the three tracked days.

Fig. 3 shows the frequency distribution of the economical Sudden-Event, discovery of the largest natural gas field in Egypt. We can see the frequency distribution starts with a low frequency until it reaches the peak at the end of the first day. The frequency of tweets was higher in the first day than the second day.

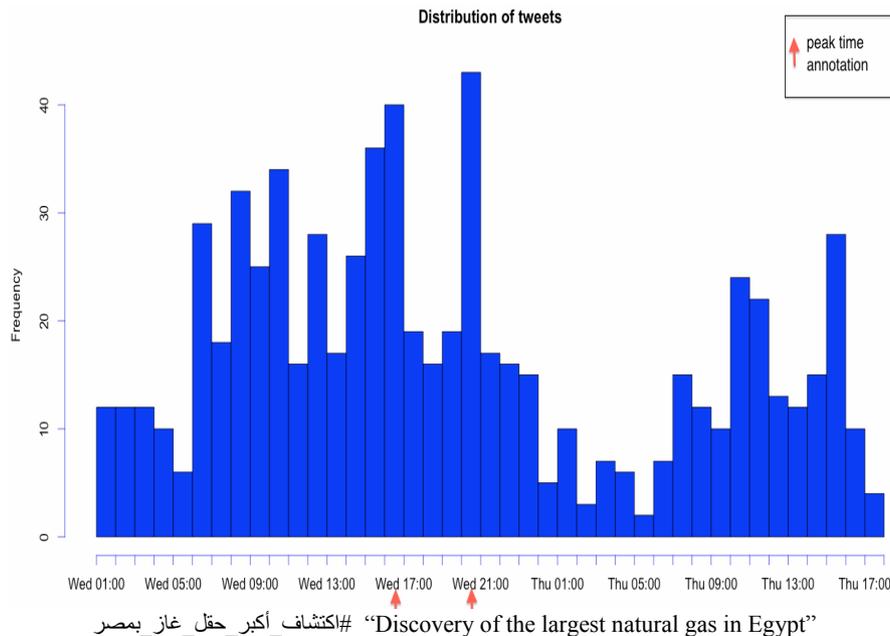


Fig. 3. Economical sudden-events frequency distribution.

Fig. 4 shows the frequency distribution of the four public opinion Sudden-Events. In #شهداء_الزمالك “Zamalek martyrs” the highest frequency of tweets was at the beginning of the event and on the second day. In #انقذوا_الانترنت “Save the internet” and #قتلوا_المصري “They killed the Egyptian” the highest frequency of tweets was on the first day of the event until the frequency distribution decreases gradually for both events, but in the latter event the users were more active than in the former event. In #ريهام_سعيد “Riham Saed” the highest frequency of tweets was on the second day conversely to the previous events. We investigated the reason beyond that, which is on Friday Al- Nahar TV, the network that carries Reham Saeed’s program, addressed the controversy by suspending her program and apologized to viewers. In turn, that decision affected in increasing the frequency of tweets.

As some studies [4], [15] found that the political hashtags persisted more period of time than other ones, which means a higher frequency of tweets over a period of time. We observe in Sudden-Events the users were interacted more in #استقالة_حكومة_محلِب (Mahlab cabinet resignation) and #ريهام_سعيد (Riham Saed) since the peak time for the latter event reached 2000 frequency in the second day where the former one reached more than 1200 frequency in the first day. We can conclude that not only political hashtags have the highest frequency of tweets but also the public opinion hashtags.

Table 5 summarizes the peak time and the highest frequency of tweets for each hashtag. From Table 5 we can conclude that the peak time for the Planned-Events were on the event day and one day before. And for most Sudden-Events, the peak time was on the first day of event.

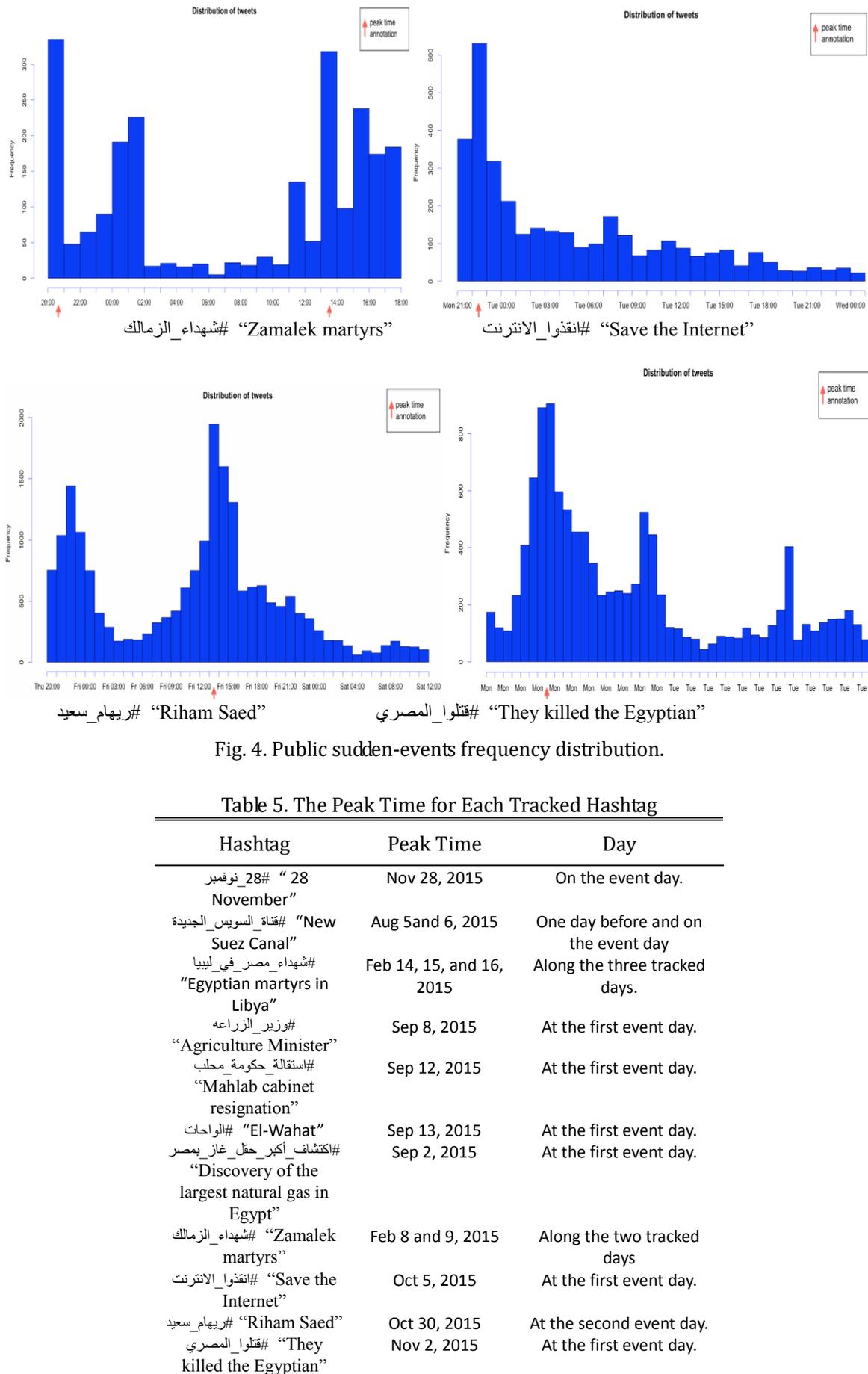


Fig. 4. Public sudden-events frequency distribution.

Table 5. The Peak Time for Each Tracked Hashtag

Hashtag	Peak Time	Day
#28_نوفمبر "28 November"	Nov 28, 2015	On the event day.
#القناة_السويس_الجديدة "New Suez Canal"	Aug 5 and 6, 2015	One day before and on the event day
#شهداء_مصر_في_ليبيا "Egyptian martyrs in Libya"	Feb 14, 15, and 16, 2015	Along the three tracked days.
#وزير_الزراعة "Agriculture Minister"	Sep 8, 2015	At the first event day.
#استقالة_حكومة_محلِب "Mahlab cabinet resignation"	Sep 12, 2015	At the first event day.
#الواحات "El-Wahat"	Sep 13, 2015	At the first event day.
#اكتشاف_أكبر_حقل_غاز_بمصر "Discovery of the largest natural gas in Egypt"	Sep 2, 2015	At the first event day.
#شهداء_الزمالك "Zamalek martyrs"	Feb 8 and 9, 2015	Along the two tracked days
#انقذوا_الانترنت "Save the Internet"	Oct 5, 2015	At the first event day.
#ريهام_سعيد "Riham Saed"	Oct 30, 2015	At the second event day.
#قتلوا_المصري "They killed the Egyptian"	Nov 2, 2015	At the first event day.

5. Discussion

In our experiment, we used various events on Twitter and we investigated their peak time within the given period by employing a statistical analysis. From our analysis we can deduce that users are more affected by the political and public opinion events according to their activity on the hashtags. Some sort of events affects politics, and public opinion, as in the following events:

#انقذوا_الانترنت “Save the Internet”, this hashtag reaches to the trending topics after the Ministry of Communication’s decision to block the voice chat for the most popular mobile applications, such as: Whatsapp, Viber, and Skype, that decision affected the public opinion on Twitter when users showed their dissatisfaction. Previous analysis showed that the highest frequency was on the first event day and it decreased on the second day due to the denial of officials regarding the decision.

#قتلوا_المصري “They killed the Egyptian”, this accident took place in Kuwait; the hashtag became a trending topic after broadcasting the video of the accident on TV shows and on online channels. In turn, the highest frequency of tweets was on the first day of event.

#الواحات “El-Wahat”, this tragic accident made a huge stir in the international media, particularly in Mexico, where they diffused a hashtag #Egipto. Moreover, it became a trending topic in Egypt and Mexico as well. In turn, the Egyptian government officials compensate all the families of the victims and the injured tourists.

#وزير_الزراعة “Agriculture Minister”, this event caused outrage in media and on social media which affected the political decisions, which in turn the prime minister announced the resignation. That made hashtag #استقالة_حكومة_محلِب (Mahlab cabinet resignation) became a trending topic as well. For both events, users interacted more in the latter event than the former event since the resignation of the cabinet was one of their demands.

#ريهام_سعيد “Riham Saed”, this event irritated Twitter users, as they diffused that hashtag to stop her to show on TV, in turn the network carrier suspended her program. We can state that the effect of social media may change the political decisions and the public opinion as well.

We can infer from the analysis that the most active events which users interacted more were #قناة_السويس_الجديده “New Suez Canal” for the Planned-Events, and #ريهام_سعيد “Riham Saed”, #استقالة_حكومة_محلِب “Mahlab cabinet resignation” for the Sudden-Events since those three events have the highest frequency of tweets.

We deduce from that research that the effect of events on Twitter is more effective than ordinary media channels such as TV media. News on social media is delivered more widely than other ordinary media, besides the social interaction between users makes the influence of event more effective.

6. Conclusion and Future Work

The various events occurred in Egypt in the last three years and the public nature of Twitter have allowed us exploiting Twitter in analysis. Every hashtag has a life cycle, which starts at some point until it diminishes. To start analyze different hashtags we classified those hashtags according to their impact. We also implemented a statistical analysis on the data to investigate the peak time for each event and to investigate the most events affected in users. We found that for the Planned-Events category, the peak time occurs on the event day and the day before; and for the Sudden-Events category, most hashtags’ peak time was on the first day of the event. After employing a comparison between various hashtags, we found that users engaged more in economical, public opinion, and political hashtags. In future work we will examine the peak time for each tracked day to identify the most tweeting time over the day for various hashtags. Also we will analyze time zone for each hashtag to identify the location of tweets.

References

- [1] Twitter. Retrieved from <https://en.wikipedia.org/wiki/Twitter>
- [2] Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is twitter, a social network or a news media. *Proceedings of the 19th International Conference on World Wide Web* (pp. 591-600).
- [3] Wu, S., Hofman, J., Mason, W., & Watts, D. (2011). Who says what to whom on twitter. *Proceedings of the 20th International Conference on World Wide Web* (pp. 705-714).
- [4] Bastos, M., Travitzki, R., & Raimundo, R. (2012). *Tweeting political dissent: Retweets as pamphlets in freeiran, freevenezuela. Spanish Revolution and Occupy Wallst.*
- [5] Bastos, M., Travitzki, R., & Puschmann, C. (2012). What sticks with whom? Twitter follower-followee networks and news classification. *Proceedings of the Potential of Social Media Tools and Data for Journalists in the News Media Industry-Sixth International AAAI Conference on Weblogs and Social Media*. Ireland: Dublin.
- [6] Tsur, O., & Rappoport, A. (2012). What's in a hashtag: Content based prediction of the spread of ideas in microblogging communities. *Proceedings of the 5th ACM International Conference on Web Search and Data Mining* (pp. 643-652).
- [7] Shaozhi, Y., & Felix, W. (2013). Measuring message propagation and social influence on Twitter. com. *International Journal of Communication Networks and Distributed Systems*, 11(1), 59-76.
- [8] Hana, A., Akram, S., & Abd El-Aziz, A. (2016). *A literature review on twitter data analysis. International Journal of Computer and Electrical Engineering*, 8(3), 241-249.
- [9] Twitter rate limiting. Retrieved from <https://dev.twitter.com/rest/public/rate-limiting>
- [10] JSON. Retrieved from <http://www.json.org>
- [11] Twitter Streaminh-API. Retrieved from <https://dev.twitter.com/streaming/overview>
- [12] Tweepy. Retrieved from <http://www.tweepy.org>
- [13] New Suez Canal. Retrieved from https://en.wikipedia.org/wiki/New_Suez_Canal
- [14] R programming language and software environment. Retrieved from <https://www.r-project.org>
- [15] Romero, D., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. *Proceedings of the 20th International Conference on World Wide Web* (pp. 695-704).



Hana Anber obtained a high graduate studies degree in computer science in 2013 from the Institute of Statistical Studies and Research, Cairo University, Egypt. Currently, she is a master's student. Her current research is centered on enhancing the analysis of twitter data for specific events to detect the life cycle and to measure the effect as well as the behavior of users towards different events categories. Hana's research interest is in social network analysis, machine learning, and big data.



Salah graduated from Mechanical Engineering and worked in computer programming for 7 years before he got his M.Sc. and Ph.D. from University of Alabama at Birmingham, USA in 1986 in computer and information sciences.

He taught in the American University in Cairo, Michigan State University, Cairo University, before he joined North Dakota State University where he designed and started a graduate program that offers Ph.D. and M.Sc. in software engineering. Salah's research interest is in data knowledge, and software engineering. He has over than 100 published papers. Currently, he is a professor in Faculty of Computer and Information, Cairo University. His current research is in knowledge engineering, ontology, semantics, and semantic web.



A. A. Abd El-Aziz has completed Ph.D. degree in June 2014 in information science & technology from Anna University, Chennai-25, India. He has received B.Sc., and master of computer science in 1995 and 2006 respectively from Faculty of Science, Cairo University. Now, He is an assistant professor in the ISSR, Cairo University, Egypt. He has 12 years experiences in teaching at Cairo University, Egypt. His research interests include database system, database security, XML security, cloud computing, big data, data mining, and social network analysis. He has published about 30 research papers in international

conferences and journals.