Popularity and Geospatial Spread of Trends on Twitter: A Middle Eastern Case Study

Nabeel Albishry[1,3*], Tom Crick[2], Teleem Fagade[1], and Theo Tryfonas[1]

[1] Faculty of Engineering, University of Bristol, UK
{n.albishry,tesleem.fagade,theo.tryfonas}@bristol.ac.uk
[2] Department of Computer Science, Swansea University, UK
thomas.crick@swansea.ac.uk
[3] Faculty of Computing & IT, King Abdulaziz University, Jeddah, Saudi Arabia
nalbishry@kau.edu.sa

Abstract. Thousands of topics trend on Twitter throughout the day, making it difficult to provide in-depth analysis of current issues, topics and themes being discussed across various locations and jurisdictions. There is thus a need for simple, generic and extensible approaches to provide deeper insight into these trends and how they propagate across locales. This paper represents one of the first studies to look at geospatial spread of trends on Twitter, presenting various techniques to provide a better understanding of how trends on social networks spread across various regions and nations. It is based on a year-long data collection (N=2,307,163) and analysis between 2016-2017 of seven Middle Eastern countries (Bahrain, Egypt, Kuwait, Lebanon, Qatar, Saudi Arabia, and the United Arab Emirates). Using this year-long dataset, the project investigates the popularity and geospatial spread of trends, focusing on trend information but not processing individual topics, with the findings showing that likelihood of trends spreading to other locales is to a large extent influenced by the place in which it first appeared. Alongside our network graph approach, we make some observations on how this work could be extended further for real-time trend analysis.

Keywords: Trends, topic spread, popularity, network graphs, Twitter

1 Introduction

With the huge daily volume of generated content on Twitter – c.500 million per day – trending topics serve as valuable sources of information on highlighting what is going on in the world, or in specific locations. Apart from the "official" trend lists provided by the platform (on the website or through API endpoints), generating insight from trends and topics detection has been receiving significant attention from across a variety of big social data-driven research domains. In

* This work has been supported by a doctoral research scholarship for Nabeel Albishry from King Abdulaziz University, Kingdom of Saudi Arabia.

health for example, monitoring and analysis of trending topics through social media has been adopted to measure public health issues, such as the spread of influenza [1–3]. Furthermore, in the marketing and business domain, topic detection and classification are important approaches to extract knowledge and insight on public opinions from posts on social media [4, 5], as well as in analysing voting intentions and political view of users [6].

With the increasing popularity and use of social networks, the impact of trends on public opinion and perceptions has placed them centrally in many social media campaigns and public relations strategy. This has made trends a valuable target for manipulation [7], stuffing [8], spamming [9, 10], and hijacking [11]. For example, the study in [12] explored a trend hijacking case and suggested that increasing social media engagement may not always be beneficial for public relations strategies.

A common approach in analysing Twitter trends is by clustering and classification of trending topics based on content [13–16]. The study in [17] presented a content-independent method to model trends progression through the dynamics of users interactions; other studies have also attempted to provide real-time classification or detection of trends [18, 19]. With the increasing demand for trends analysis across various domains, customisable clustering tools that can be used by non-technical users have started to emerge, such as the recent example introduced in [20].

2 Methodology

2.1 Context

Seven Middle Eastern countries were selected for this study: Bahrain, Egypt, Kuwait, Lebanon, Qatar, Saudi Arabia, and the United Arab Emirates (UAE). The selection includes countries with relatively large population (e.g. Egypt: 97,553,000) and relatively small populations (e.g. Bahrain: 1,493,000) [21]. Kuwait is reported to have the most active daily users on Twitter [22]; as of March 2016, Saudi Arabia and Egypt generated 33% and 20% of the tweets in the Middle Eastern region. Bahrain is the most balanced location in terms of gender breakdown of active users. Interestingly, between March 2014 and March 2016, Lebanon was the only location in the Middle Eastern states that has not seen growth in active users, while UAE increased by 60%. The Gulf Cooperation Council countries – Bahrain, Kuwait, Qatar, UAE, and Saudi Arabia – were reported to have the highest penetration rates [22].

2.2 Data Collection

Trending topic lists in the seven countries were monitored for a year between October 2016 and October 2017. Every hour, trending lists were collected through the Twitter REST API, which resulted in 7,948 hour's worth of records for all the countries, accumulating 2,307,163 trend records. It is important to note that the Twitter API does not necessarily provide trends data for every request; for example, it is possible to receive no information for tweet volume. For each location, the list of available trending topic is returned. From this list, four pieces of information are extracted from each trend record:

- *woeid*: the Yahoo! Where On Earth ID (WOEID) of the location;
- *name*: text of trending topic (e.g '`#Call_For_Action`');
- *as_of*: recorded timestamp of the trend;
- *tweet_volume*: volume of tweets over the past 24 hours, if available.

While the Twitter API returns a list of trending topics for a specific *woeid* location, the tweet volumes do not provide a comprehensive measure of the tweeting activity in that location. Rather, the tweet volume refers to the overall number of tweets containing the trend, regardless of their location. Although the Twitter documentation[4] does not provide the necessary detail on this, it was apparent after observing trends that showed up in various locations. Trends were found with the same tweet volume across all locations and, hence, participation volume of each location was not possible to be accurately measured. Therefore, the context of this study does not include any reference to this volume entity.

2.3 Graph Construction

The study is based on generating graph structures and conducting analyses of their properties. The analysis approach involves constructing two graphs; the *temporal base graph* that captures the structure of trend raw data, and the *weighted aggregated graph* which is generated from the base graph to further explore its structure to provide additional insights. Figure 1 illustrates these graphs; the nodes and direction orientation of edges are the same in both graphs. Thus, nodes with zero indegree identify places, while trend nodes feature zero outdegree.

The *temporal base graph* is a directed graph that consists of three trend entities: *place*, *trend* and *timestamp*. Nodes represent place and trends, and edges are labelled with timestamps to indicate the time at which the trend appeared in a location. This graph is used to examine temporal properties, such as spread.

The *weighted aggregated graph* is a graph that combines temporal edges between two nodes (in the base graph) into a single weighted edge. The feature of weighted edges in this graph is used to measure the popularity of trends, repetition rate, participation of countries, and the volume of the engagement.

3 Results

Observation of the weighted graph provided an overall evaluation of activity for trends and places. In total there were 76,266 distinct trends that trended

[4] `https://developer.twitter.com/en/docs/trends/trends-for-location/api-reference/get-trends-place`

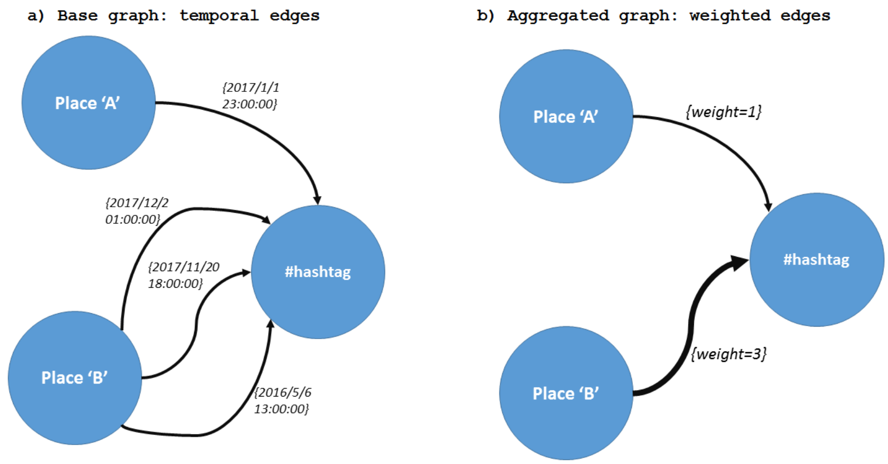

Fig. 1. Graphs used in analyses

2,307,163 times across all locations; this suggests that trends may appear repeatedly over time. The overall repetition ratio in the dataset was 97%, and ranged from 80% to 98% for individual locations, with Saudi Arabia scoring lowest and Qatar scoring highest rate. *Indegree*, *outdegree*, and *edges* were used to conduct subsequent results, with further explanation to follow in the relevant sections.

3.1 Commonality and Popularity of Trends

The node indegree indicates the number of locations at which the trend showed (*commonality*), and the weighted indegree is used to measure the total number of times a trend showed (*popularity*). Indegree was used to group trends to, while weighted indegree was used to analyse activity in generated groups, as shown in Table 1. Although 83% of trends have appeared in one location only, their total weighted indegrees was c.40%; in other words, there were less common trends amongst locations, but their popularity was higher than isolated trends[5] – this implies that trends showing across location does not necessarily imply the prominance or importance of activity or topic.

3.2 Location Participation

The node outdegree reflects how many unique trends a location is connected to (*diversity*), and weighted outdegree measures the ability of the location to generate trends (*activity*). The outdegree descriptive statistics, presented in Table 2,

[5] Isolated trends are those that have trended in one place, i.e. their indegree equal 1.

Table 1. Trends indegree groups

Indegree	No. trends	Ratio	Total W.	W. Ratio	Max.	Mean	Std
1	62,959	0.826	936,959	0.406	2,146	14.88	43.26
2	7,338	0.096	335,073	0.145	1,957	45.66	77.45
3	2,840	0.037	220,805	0.096	1,842	77.75	96.29
4	1,538	0.020	216,524	0.094	2,797	140.78	201.03
5	850	0.011	184,127	0.080	3,604	216.62	297.36
6	463	0.006	192,968	0.084	3,998	416.78	581.76
7	278	0.004	220,707	0.096	5,367	793.91	994.66

shows that Saudi Arabia came at the top of the list, with 42% of outgoing edges and weighing 20% of the total weight of the graph. Closeness in the table shows how close a location node is to all other trend nodes; it shows that Saudi Arabia has connections to 56% of trends in the graph. Nevertheless, Saudi Arabia was found lowest in terms of maximum edge weight, mean and standard deviation; Qatar was found to have the reverse values. This can be interpreted as the trends activity in Saudi Arabia was more diverse in total, but more consistent. In contrast, Qatar is connected to a limited number of trends with more focused activity. Also, Qatar's outdegree is just 60% of Bahrain's, although its weighted degree was 1.9 higher.

Table 2. Location outdegree descriptive statistics

Location	Outdegree	Out. ratio	W. Out.	W. Ratio	Closeness	Max.	Mean	Std
Bahrain	7,424	0.07	133,069	0.06	0.10	1,949	17.92	78.48
Egypt	14,282	0.14	383,830	0.17	0.19	1,408	26.88	58.68
Kuwait	13,891	0.14	397,960	0.17	0.18	1,400	28.65	51.67
Lebanon	6,044	0.06	294,761	0.13	0.08	2,146	48.77	133.64
Qatar	4,484	0.04	248,003	0.11	0.06	2,173	55.31	146.79
Saudi Arabia	42,767	0.42	468,081	0.20	0.56	1,175	10.95	17.43
UAE	12,389	0.12	381,459	0.17	0.16	1,655	30.79	70.75

3.3 Edges Properties

Edge weights in the graph were utilised to evaluate location activity in indegree groups, as shown in Figure 2. Overall, most of location activity went to common trends. Although Saudi Arabia was the highest in terms of total activity, the majority of its activity (61.26%) was identified as isolated trends. Moreover, observing originating locations for isolated trends shows that c.30% of inbound edges came from Saudi Arabia, as shown in Figure 3. Egypt contributed the most in 2 and 3 indegree trend groups, United Arab Emirate in 4, 5 and 6 indegree trends, and for 7 indegree trends most of in edges originated from the Lebanon.

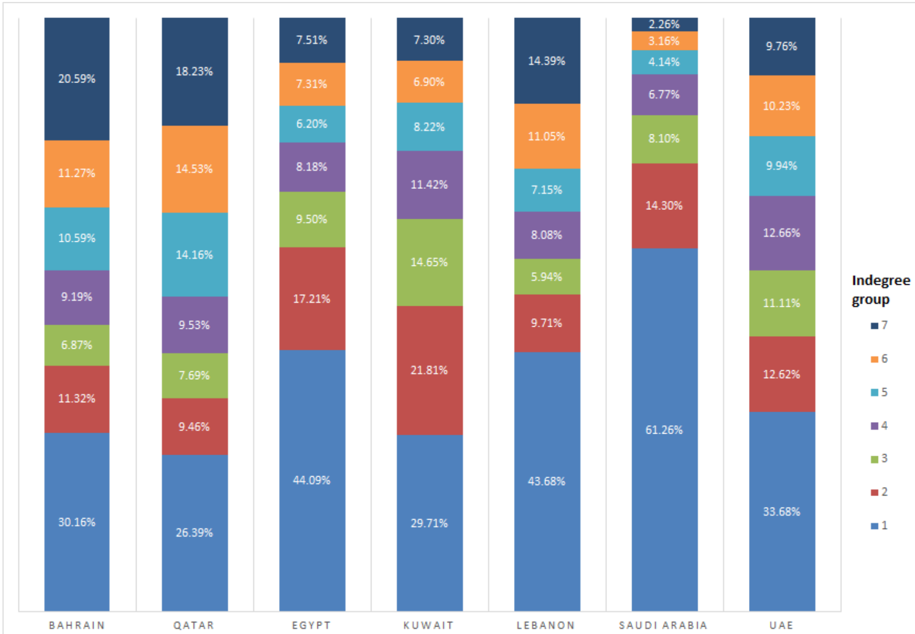

Fig. 2. Trends Indegree group distributions across countries

3.4 Temporal Spread and Reach

As shown in Table 1, c.60% of weighted indegree was identified as common trends. To further examine temporal changes on those trends, timestamps on in-edges of trend nodes in the temporal graph were observed. Those timestamps were used to measure temporal order of locations for trend, as shown in Table 3. For instance, about 42% of first appearance of trends was in Saudi Arabia, while 36% of 7th trend appearance was in Bahrain.

Table 3. Distribution of temporal orders of location for multi-indegree trends

Location	1st	2nd	3rd	4th	5th	6th	7th
Bahrain	2.3	3.0	3.5	6.3	14.9	26.0	36.0
Egypt	16.3	13.2	21.3	22.8	16.8	11.7	2.5
Kuwait	16.5	29.5	23.8	15.1	13.2	11.2	8.3
Lebanon	5.7	5.4	6.2	7.0	8.7	11.3	23.7
Qatar	4.1	4.9	8.9	16.1	25.1	25.6	12.6
Saudi Arabia	42.2	28.7	10.2	7.1	7.3	8.1	14.4
UAE	12.9	15.2	26.0	25.4	14.0	5.9	2.5
Total	100	100	100	100	100	100	100

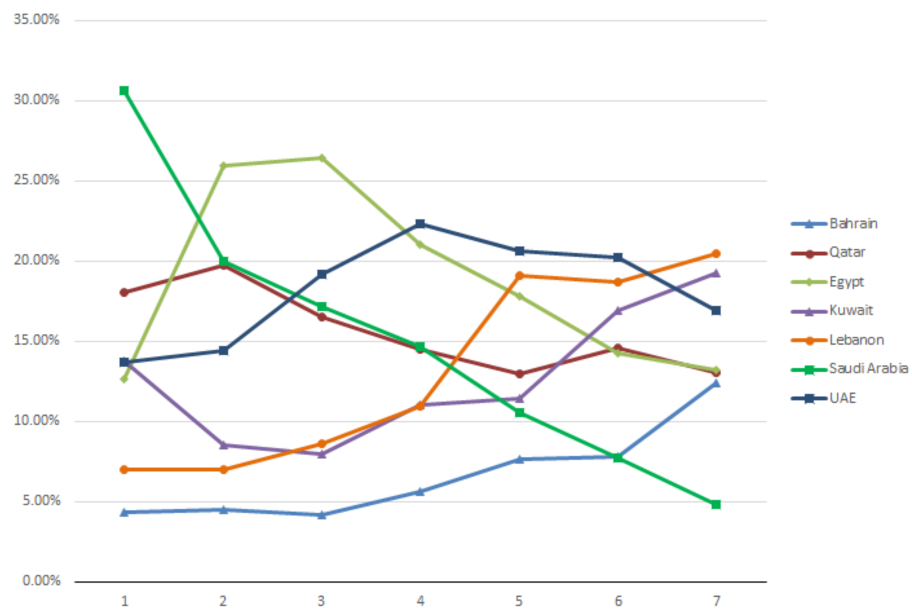

Fig. 3. Weighted contribution of countries toward trends indegree groups

A similar observation was made on the outdegree measures, with timestamps on in-edges of place nodes in the temporal graph observed; the results are presented in Table 4. The highest portion of activities in Saudi Arabia, Egypt and Lebanon made them 1st locations for trends to appear in. However, Bahrain, UAE, Qatar, and Kuwait were more active with trends that have appeared previously.

Table 4. Distribution of appearance orders of locations

Order	Bahrain	Saudi	Egypt	UAE	Lebanon	Qatar	Kuwait
1st	18.4	53.6	34.6	26.9	32.2	19.1	26.3
2nd	24.2	36.4	28.0	31.9	30.5	22.9	47.1
3rd	12.8	5.8	20.3	24.4	15.8	18.6	17.1
4th	12.1	2.1	11.4	12.5	9.3	17.6	5.7
5th	14.5	1.1	4.3	3.5	5.9	13.9	2.5
6th	11.8	0.6	1.4	0.7	3.6	6.6	1.0
7th	6.1	0.4	0.1	0.1	2.8	1.2	0.3
Total	100	100	100	100	100	100	100

Additionally, the reach of trend was measured to examine how many other locations a trend is likely to reach based on the location in which it first appeared.

Therefore, edges and related nodes relating to the first column in Table 3 were used. The results presented in Table 5 show that 62.3% of trends that first appeared in Kuwait have also appeared in one other location, and 5.1% of those that first appeared in Qatar have also appear in six more locations.

Table 5. Further reach of trends per locations

Reach	Bahrain	Egypt	UAE	Lebanon	Qatar	Kuwait	Saudi
1	54.8	59.2	54.2	50.1	41.2	62.3	53.1
2	19.3	22.7	21.2	21.3	24.3	18.9	21.7
3	8.3	9.8	11.1	10.4	12.6	10.1	13.2
4	8.3	4.4	6.4	8.1	12.0	4.1	7.2
5	4.7	2.7	4.3	6.5	4.7	2.6	3.3
6	4.7	1.2	2.8	3.7	5.1	2.1	1.6
Total	100	100	100	100	100	100	100

Finally, Table 6 shows the origin of common trends grouped by their indegree (connected places). As can be seen, 31.7.1% of trends that appeared in seven locations have originated from Saudi Arabia, and 5% from Bahrain. Nevertheless, Bahrain was better in terms of further reach.

Table 6. Origin of trends per reached locations

Location	2	3	4	5	6	7
Bahrain	2.2	2.0	1.6	2.9	3.0	5.0
Egypt	17.5	17.3	13.8	11.3	12.5	9.4
Kuwait	18.7	14.6	14.4	10.5	12.3	16.5
Lebanon	5.2	5.7	5.1	7.2	10.6	10.1
Qatar	3.1	4.7	4.5	7.8	5.6	10.0
Saudi Arabia	40.7	42.9	48.1	47.4	40.2	31.7
UAE	12.7	12.8	12.4	12.9	15.8	17.3
Total	100	100	100	100	100	100

4 Discussion and Conclusions

From the results, we can see that isolated trends were found to be most common across countries, although the study includes countries with a high proportion of active users and high tweet generation rate, such as Saudi Arabia and Egypt [22]. As previously mentioned, the number of trends returned by the Twitter API does not accurately reflect the true activity of the location. Low trending topics may indicate low consensus on these discussed topics and does not necessarily reflect tweeting activity. Also, the number of trending topics is very likely to include repeated ones, and therefore a high number of trends does not necessarily imply more new topics. Furthermore, the number of trends was not found to correlate with the tendency of location to participate in common trends. For example, Saudi Arabia was found to be to connected to 56% of trends, however 61% them were isolated trends i.e. trends that only appeared in Saudi Arabia. Meanwhile, most of Qatar's trends (73.6%) were common ones, although it had edges with 6% of trends; this indicates that the activity of certain location is more focused on internal issues and concerns.

Also, the further reach of trends (i.e. appearing in other locations) was observed for each location. Although a specific location may do well in reaching other locations, the number of trends it generates may affect the total reach. For example, Qatar was highest in reaching other locations, however it was the 5th in being the origin of trends that reach all locations. This was certainly clear in the case of Saudi Arabia: its scores in reaching other locations were not comparable to its scores in being the origin of common trends, as shown in Tables 5 and 6.

In conclusion, the study has presented an approach to analysing trends data using graphs and their properties. It demonstrates the importance of graph construction techniques to capture raw trends data, resulting in the temporal base graph. Then, it presented how aggregated weighted graph can be generated from the base graph. The temporal graph was used to measure temporal properties such as spread and reach; the weighted graph was used to measures overall activities, such as commonality and popularity of trends, and diversity and activity of locations.

The presented approach showed how trends data can be used to evaluate topics and location activity without the need to crawl individual topics. Also, it shows how to measure spread of trends and reach based on their historical records as well as the originating location. This approach can be used and extended to identify trends of important features; for example, to extract high spread trends, or how likely it is for a trend to reach a specific location from another one.

References

1. Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.H., Liu, B.: Predicting Flu Trends using Twitter data. In: Proc. of 2011 IEEE Conference on Computer Communications Workshops. (2011) 702–707
2. Parker, J., Wei, Y., Yates, A., Frieder, O., Goharian, N.: A framework for detecting public health trends with Twitter. In: Proc. of 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'13). (2013) 556–563
3. Parker, J., Yates, A., Goharian, N., Frieder, O.: Health-related hypothesis generation using social media data. Social Network Analysis and Mining **5**(7) (2015)
4. Blamey, B., Crick, T., Oatley, G.: R U :-) or :-(? Character- vs. Word-Gram Feature Selection for Sentiment Classification of OSN Corpora. In: Research and Development in Intelligent Systems XXIX. Springer (2012) 207–212

5. Bello, G., Menéndez, H., Okazaki, S., Camacho, D.: Extracting Collective Trends from Twitter Using Social-Based Data Mining. In: Computational Collective Intelligence. Volume 8083 of LNCS. Springer (2013) 622–630
6. Fang, A., Ounis, I., Habel, P., Macdonald, C., Limsopatham, N.: Topic-centric Classification of Twitter User's Political Orientation. In: Proc. of 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. (2015) 791–794
7. Zhang, Y., Ruan, X., Wang, H., Wang, H., He, S.: Twitter Trends Manipulation: A First Look Inside the Security of Twitter Trending. IEEE Transactions on Information Forensics and Security **12**(1) (2017) 144–156
8. Irani, D., Webb, S., Pu, C., Drive, F., Gsrc, B.: Study of Trend-Stuffing on Twitter through Text Classification. In: Proc. of 7th Annual Collaboration, Electronic Messaging, AntiAbuse and Spam Conference (CEAS 2010). (2010)
9. Sedhai, S., Sun, A.: HSpam14: A Collection of 14 Million Tweets for Hashtag-Oriented Spam Research. In: Proc. of 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. (2015) 223–232
10. Chu, Z., Widjaja, I., Wang, H.: Detecting Social Spam Campaigns on Twitter. In: Applied Cryptography and Network Security. Volume 7341 of LNCS. Springer (2012) 455–472
11. VanDam, C., Tan, P.N.: Detecting hashtag hijacking from Twitter. In: Proc. of 8th ACM Conference on Web Science (WebSci'16). (2016) 370–371
12. Sanderson, J., Barnes, K., Williamson, C., Kian, E.T.: 'How could anyone have predicted that #AskJameis would go horribly wrong?' public relations, social media, and hashtag hijacking. Public Relations Review **42**(1) (2016) 31–37
13. Zubiaga, A., Spina, D., Fresno, V., Martínez, R.: Classifying trending topics: a typology of conversation triggers on Twitter. In: Proc. of 20th ACM International Conference on Information and Knowledge Management. (2011) 2461–2464
14. Benhardus, J., Kalita, J.: Streaming trend detection in Twitter. International Journal of Web Based Communities **9**(1) (2013)
15. Ferragina, P., Piccinno, F., Santoro, R.: On Analyzing Hashtags in Twitter. In: Proc. of 9th International AAAI Conference on Web and Social Media (ICWSM'15). 110–119
16. Albishry, N., Crick, T., Tryfonas, T.: "*Come Together!*": Interactions of Language Networks and Multilingual Communities on Twitter. In: Computational Collective Intelligence. Volume 10449 of LNCS. Springer (2017)
17. Thij, M., Bhulai, S.: Modelling Trend Progression Through an Extension of the Polya Urn Process. In: Proc. of 12th International Conference and School on Advances in Network Science. Volume 9564., Springer (2016) 57–67
18. Mathioudakis, M., Koudas, N.: TwitterMonitor: Trend detection over the twitter stream. In: Proc. of 2010 ACM SIGMOD International Conference on Management of Data. (2010) 1155–1158
19. Zubiaga, A., Spina, D., Martínez, R., Fresno, V.: Real-Time Classification of Twitter Trends. Journal of the Association for Information Science and Technology **66**(3) (2015) 462–473
20. Arin, I., Erpam, M.K., Saygn, Y.: I-TWEC: Interactive clustering tool for Twitter. Expert Systems with Applications **96** (2018) 1–13
21. United Nations: World Population Prospects 2017. Technical report, Department of Economic and Social Affairs, Population Division (2017)
22. Salem, F.: Social Media and the Internet of Things towards Data-Driven Policy-making in the Arab World: Potential, Limits and Concerns. Technical report, he Arab Social Media Report, Dubai: MBR School of Government (February 2017)