Neurocomputing 163 (2015) 137–150

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Semantic query suggestion using Twitter Entities

Ioannis Anagnostopoulos [a,*], Gerasimos Razis [a], Phivos Mylonas [b], Christos-Nikolaos Anagnostopoulos [c]

[a] Computer Science and Biomedical Informatics Department, University of Thessaly, Greece
[b] Department of Informatics, Ionian University, Greece
[c] Cultural Technology and Communication Department, University of the Aegean, Greece

ARTICLE INFO

Article history:
Received 2 February 2014
Received in revised form
22 December 2014
Accepted 25 December 2014
Available online 21 March 2015

Keywords:
Query suggestion
Microblogging
Viral content
Twitter

ABSTRACT

There are many web information management methods and techniques that help search engines and news services to provide useful suggestions with respect to queries, thus facilitating the users' search. However, the penetration of microblogging services in our daily life demands to also consider social sphere as far as query suggestion is concerned. Towards this direction, we introduce an algorithmic approach capable of creating a dynamic query suggestion set, which consist of the most viral and trendy Twitter Entities (that is hashtags, user mentions and URLs) with respect to a user's query. For evaluation purposes, we firstly compare the results derived from two case studies, against the suggestions of popular services like Google News, Yahoo! News, Bing News, and Reuters. In addition we further evaluate our approach with subjective user ratings against Google Trends service. Finally, we provide comparative results that clearly show that our proposal outperforms other methods and baselines in the respective literature.

1. Introduction – motivation and research contribution

Microblogging – a “light”, rather live, version of blogging – is considered to be one of the most recent social raising issues on the Internet, being one of the key concepts that brought Social Web to the broad public. The main characteristic of microblogging is the fact that posts are produced almost in real-time and are strictly limited to a specific and rather small amount of characters, such as short sentences, term concatenation or tinyURLs[1] that point to hyperlinks with web and/or even multimedia content. It comprises many very brief updates that are presented to the microblog's readers in reverse-chronological order. Motivated by its increasing popularity, among many microblogging services we focus herein on the Twitter[2] social network, where microblogs are known as ***tweets***.

In web information retrieval, the effectiveness of search engines strongly depends on whether users can express their information needs through the terms they submit. However, submitting queries is not an easy task. Queries are short, not written in natural language, and – mostly – their terms are ambiguous. Many proposed methods offer meaningful query suggestions, usually by employing knowledge extraction methods from browsing history records or search logs. However, very few consider time as an important parameter related to the actual meaning of a query term. Thus, in this work, we do not tackle query suggestion in the traditional way, but we provide time-aware suggestions according to the most viral terms that appear in Twitter along with the user's query.

In other words, the main contribution of this work is the effective suggestion of microblogging social content (called hereafter as *Twitter Entity/Entities* – TE/TEs) that manage to become viral in time, given a user query; the more viral the social content is, the more relevant are the suggestions. Our ultimate goal is to provide users a way to enter any type of query and retrieve accurate, relevant and popular (viral) Twitter Entities suggestions that would semantically “fit” to their information needs. In order to measure virality, we extend the capture–recapture methodology, which is mainly used for estimating population properties (e.g., birth/survival rates) in real-life biological experimentations; in our work, the concept of virality in social content is considered equal to survivability in animal populations under study. The concept of social content on the other hand is directly related to Twitter Entities (hashtags (#), user mentions (@) and URLs) and should not be related with named entities or Wikipedia concepts as considered in

* Corresponding author.
E-mail addresses: janag@dib.uth.gr (I. Anagnostopoulos), razis@dib.uth.gr (G. Razis), fmylonas@ionio.gr (P. Mylonas), canag@ct.aegean.gr (C.-N. Anagnostopoulos).
URL: http://www.anagnostopoulos.name (I. Anagnostopoulos).

[1] http://en.wikipedia.org/wiki/TinyURL.
[2] http://www.twitter.com.

http://dx.doi.org/10.1016/j.neucom.2014.12.090
0925-2312/

most papers in the related literature of text mining and information retrieval[3] (e.g., see the work described in [20]).

It is rather true that research works on microblog posts analysis and extraction of meaningful information from them in a (semi-) automated manner have been considered recently in the literature, yet we have reason to believe these approaches are quite different to ours. As the interested reader will see within next Section 2 of this manuscript, related research on query suggestion [15] is highly related with query expansion [14] query substitution, query recommendation or query refinement tasks. In this work, we deviate from the traditional query suggestion proposals in a sense that users have their queries expanded directly from Twittersphere,[4] and without having their queries or browsing history processed by search engines. In addition, another important difference against related query suggestion techniques, focused on web search and the real-time variant of the problem at hand, is the narrow time frame considered herein in which suggestions have maximal impact. For this work we were extra motivated by the facts that (a) microblogging social content annotation is provided directly in real-time by users worldwide and (b) the more this annotation becomes important or so-called "*viral*", the more semantically related it becomes with a recent trend, top news, thematic categorization, etc.

The rest of this manuscript is organized as follows. In the next section, we provide an overview over the literature within the query suggestion field, emphasizing on related works within the social sphere. Section 3 provides an overview of the methodology we use, as well as the basic steps of our proposed query suggestion method. In Section 4 – and in order to clearly show how our query suggestion expansion mechanism works – we describe the results of two real-life scenarios (case studies) that lasted more than a week within January 2014. In addition, we evaluate our results against four famous Internet commercial news services (Google News, Yahoo! News, Bing News, and Reuters). In Section 5 we further evaluate our approach by subjective comparisons with respect to Google Hot Trends service, as well as against a cluster labeling and a microblog retrieval task for a time span of one month (May 9, 2014–June 9, 2014), providing comparative results. Finally, in Section 6 we conclude this work by highlighting its main outcomes, and describing in parallel our future directions based on the experiences we faced.

2. Works on related information search and retrieval tasks

Given the fact that microblogging is increasingly popular, several research methods for organizing and providing access to microblog data have been emerged on this topic since the last few years. In this section, we provide an overview over microblog-related information retrieval research works focusing on the field of query suggestion.

2.1. Information search and retrieval in microblogs

In general, the social sphere consists of the so-called tweets or microblogging posts [5], where the large amount of real-time tweets per day is highly attractive for information retrieval research. Within the *social sphere* context query suggestions must be in real-time, i.e., results need to be temporally relevant and timely [15]. Now, microblogs form a rather special category of user-generated data: they typically contain two major characteristics that seriously affect linguistic analysis techniques, namely: (a) they contain strong vernacular (acronyms, spelling changes, etc.) and (b) they do not include any memorable repetition of words. More specifically, Massoudi et al. in [14] studied a Twitter-based retrieval model by considering the model with textual quality and Twitter specific quality indicators. Naveed et al. [16] combined document length normalization in a retrieval model to resolve short texts sparsity in the case of tweets. Motivated by the observation that in a typical microblog user tends to retrieve meaningful information through queries formulation, researchers focus on each post's characteristic features [8], whose quantitative evaluation could potentially affect the way in which the relevance between the user query and its returned results may be calculated. Even TREC 2011 introduced the Microblog Track which addressed one single pilot task, entitled "real-time search task", where the user wished to see the most recent but relevant information to the query (e.g., [17]). A first step towards this direction is discussed in [22], where Tao et al. identify two feature categories, i.e., features related or not to the user query.

The fact that microblog posts contain *hashtags* is also exploited in the literature in the direction of acquiring information that the user "is not aware of" and formulate queries that the user "does not know how to express" (e.g., [4]). In a representative approach [5] and given a query, Efron attempts to statistically identify a number of hashtags relevant to the given query, that may be used to expand it and lead to better results. Even in our own previous work [2], we proposed the utilization of hashtags as the main source of information acquisition, by searching the specific query terms within microblog posts under the condition that the former need to appear as hashtags; then, we calculated the most common hashtags that co-occur together with the original query, and, thus, expanded the query with the new hashtags. Last but not least, the observation that microblog posts are created during an actual event and contain comments and/or information directly related to it leads to event detection research efforts [18] based on posts and/or hashtags.

2.2. Query manipulation works

Typical microblog query manipulation research problems include both *query analysis and expansion* and *query suggestion* approaches. Still, there are also some distinctive differences between the two. A *query expansion* task is typically used transparently to the end-user and internally within a search engine mechanism, whereas a query suggestion is exposed to its end-users and therefore can use additional explicit information to its aid. In this manner, Bandyopadhyay et al. [3] attempt to improve weak ad-hoc queries through a process they call "web assistance", by exploring standard query expansion approaches and utilizing external corpora as a source for the query expansion terms, namely pages derived from the Web and their titles. Efron [5] showed that for a Twitter microblog collection, hashtags may be predicted using query expansion techniques; he proposed restricting the added query terms to those candidates that are hashtags, stripping candidates of their leading "#". In another more recent approach, Kumar and Carterette [10] take into account the fact that most existing models for Information Retrieval do not take the very important time aspect into account and focus on Twitter search models; they utilize time-based feedback and a simple query expansion by using highly frequent terms in top tweets as their expanded terms. In another detailed approach by Massoudi et al. [14], authors propose an efficient dynamic query expansion model for microblog post-retrieval, utilizing a language modeling approach to search microblog posts by incorporating query expansion and certain "quality indicators" during matching. The latter is very interesting since several typical microblog characteristics may be exploited as quality indicators, such as temporal [12] or topological ones.

In the case of actual *query suggestion* tasks though, the problem at hand becomes slightly different and its complexity increases as all current major web-search engines and most proposed methods that suggest queries rely solely on search engine query logs to determine

[3] Thus, it should be clear that whenever we mention the term "entity" in the manuscript, we refer to Twitter Entity/Entities (TE/TEs), unless otherwise explicitly stated.

[4] http://www.oxforddictionaries.com/definition/english/Twittersphere.

their possible query suggestions. Although there are some research works on the topic in general, the consideration of the very important temporal parameter is rarely tackled, due to the fact that it is considered much more difficult to effectively suggest relevant queries to a recent search query, which has absolutely none or very few historical evidences in the aforementioned type of query logs. In this manner, Li et al. [13] introduce the notion of *fresh queries*, trying to offer an effective query suggestion methodology for fresh search queries, but they utilize word frequency statistics to extract a set of ordered candidate words for suggestions and not the most common Twitter Entities (namely: hashtags (#), mentions (@) and links) appeared in tweets, as we propose in this work. Moreover, other recent attempts like [7] implement empirical evaluations on a selected Twitter dataset in comparison to crawled hot queries published by Google Trends for a given period of time, in a manner similar to the herein proposed approach. Finally, Mishne et al. [15] present the architecture behind Twitter's real-time related query suggestion and spelling correction service, as a case study illustrating the challenges of real-time data processing in the era of "big data" and argue that query expansion terms may be considered explicitly controlled by the user in an early form of query suggestion.

3. Proposed query suggestion mechanism

Having discussed most of related research works in the field, in this section we present the basic aspects of our proposed methodology. The microblogging service used in this work is Twitter. Thus, herein discussed Query Suggestion is based on the most common Twitter Entities – TEs, namely hashtags (#), mentions (@), as well as the links appeared in Tweets. Hashtags are unique or concatenated terms (or even phrases) prefixed with the symbol "#". This type of information is widely used in Twitter when users want either to emphasize a term/phrase or intend to semantically annotate the tweeted information. Thus, if a hashtag becomes extremely popular, it will appear in the "Trending Topics" area of a user's account. There are no strict syntactic rules (apart from the concatenated form in subsequent terms), so users may post any term that best represents a concept according to their opinion, knowledge or way of expression. A mention (@) points to an active Twitter user (or active Twitter account), thus considered as an entity that describes content referring to a person, an organization, a community etc., who or which is able to act as an individual actor inside the Twittersphere. Finally, links as Twitter Entities provide to users the ability to relate their tweet with a web page, a photo or even with a short video message. In order to keep the restrictions of 140 characters per Tweet, Twitter uses a specific service that shortens the hyperlinks to 22 characters (tiny URLs).

Our proposed Query Suggestion mechanism is two-fold. At first, we measure the virality (or survivability in our model) of the suggested TEs within a specific time and given a user's query, thus forming a cluster of candidate suggested terms for the Query Suggestion Set (Section 3.1). Then, among the suggested TEs, we calculate their ranking order (Section 3.2). Upon describing these issues, we also provide a specific example in order to better illustrate the way time-based virality is calculated.

3.1. Survivability factor – clustering suggested terms

Prior to describing our methodology, we need to introduce the framework of capture–recapture experiments,[5] which are used mainly in wildlife biological studies [9,19,21]. In these experiments, animals, birds, fish or insects (subjects of investigation) are captured, marked and then released. If a marked individual is captured on a subsequent trapping occasion, then it is mentioned as a "recaptured" instance. The amount of the marked and recaptured individuals, can lead to an estimation of the total population size, as well as the birth, death and survival rate of each species under study. In our methodology we specifically employ the Pollock's Robust Design model for clustering the candidate suggested terms (TEs in our case) that consist of the Query Suggestion Set. The Pollock's Robust Design model helps us to calculate the survivability factor (also called survival probability – φ), which creates the cluster of the candidate suggested terms. In our paradigm, social content dynamics are considered analogous to the above population dynamics. More specifically, a birth is the appearance of a new Twitter Entity, while high survivability rates in these entities reflect high levels of virality.

To further elaborate on this aspect, the methodology of capture–recapture in real-life experimentations is briefly presented in the following: the sampling process is divided into k primary sampling periods, each of them consisting of l secondary sampling periods. At this point we have the distinction of the "open" and "close" model. In the first case, we assume that we can have births, deaths and/or migration incidents within the population under study, while in the latter the population and its evolution sustains constant. In our model, we consider the "open" model among primary sampling periods and the "close" model among secondary sampling periods [9,19,21]. The basic measurements are conducted during a secondary sampling period, where a set of different individuals is trapped. Then these individuals are marked – keeping in parallel a history record of them – and then are released back to their environment. After a specific time interval, the second secondary sampling period occurs and so forth until the end of the last l secondary sampling period. Secondary periods are near and quite short in time, while trapping occasions are instantaneous for assuming that the population under study is closed. However, longer time intervals between primary sampling periods are desirable so evolution events can occur (e.g., survival, movement, and growth).

Now, in our paradigm, the trapping occasions correspond to the query term (seed) we want to extend. Primary sampling periods consist of 16 secondary sampling occasions. In each of these 16 distinct samplings we capture and mark some entities under a probability value p. This probability value is the proportion of marked or total marked and unmarked Twitter Entities that are captured during a sampling occasion, thus ensuring that all secondary samplings are conducted in a "close" pool of instances and under the basic principle of the Pollock's model. Then by investigating the recaptured instances, we calculate the survival probability of the examined entity according to Eq. (1), where $(M_i - m_i)$ defines the marked entities not captured during the ith sampling period, while R_i are the entities captured at the ith period, marked, and then released for possible recapture in future samplings. Moreover, M_i is the number of marked entities in the population at the time where the ith sample is collected ($i = 1, 2, \ldots, k$; $M_1 = 0$) and m_i stands as the number of the marked Twitter Entities captured in the same sample.

$$\tilde{\varphi}_i = \frac{\tilde{M}_{i+1}}{\tilde{M}_i - m_i + R_i} \tag{1}$$

Fig. 1 highlights the basic structure of the capture–recapture model we follow. More specifically, the red boxes correspond to a primary sampling period, which is divided in 16 secondary sampling periods (blue boxes) and each one last for 1 min. As mentioned above, secondary sampling periods are quite important for our methodology since we can measure how viral a TE is. Now, in order to get a clear insight on how the algorithm works, let us see all involved steps with an example. We assume that the user wants to have some suggestions next to the initial query term "Schumacher".

As the flowchart of Fig. 2 depicts, during every separate minute after the system receives the query "Schumacher" (also called seed term), we

[5] http://en.wikipedia.org/wiki/Mark_and_recapture.

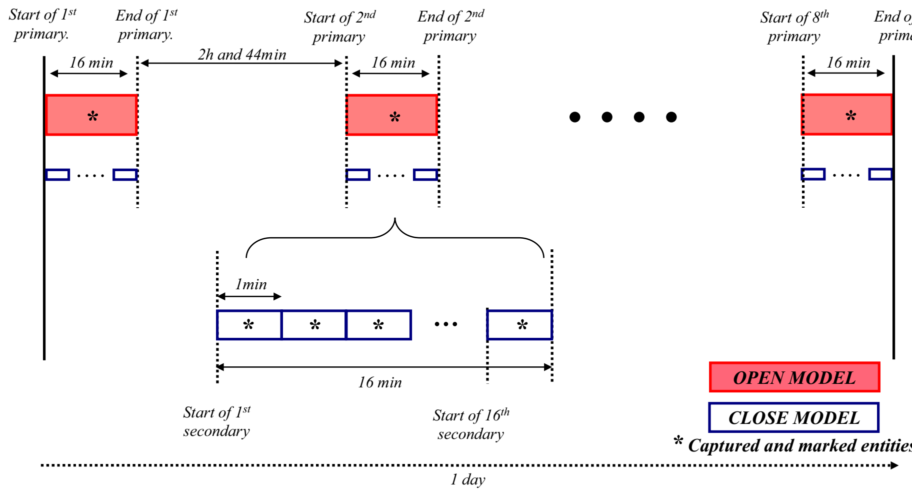


Fig. 1. Structure of the conducted capture–recapture experiments (primary/secondary sampling periods).

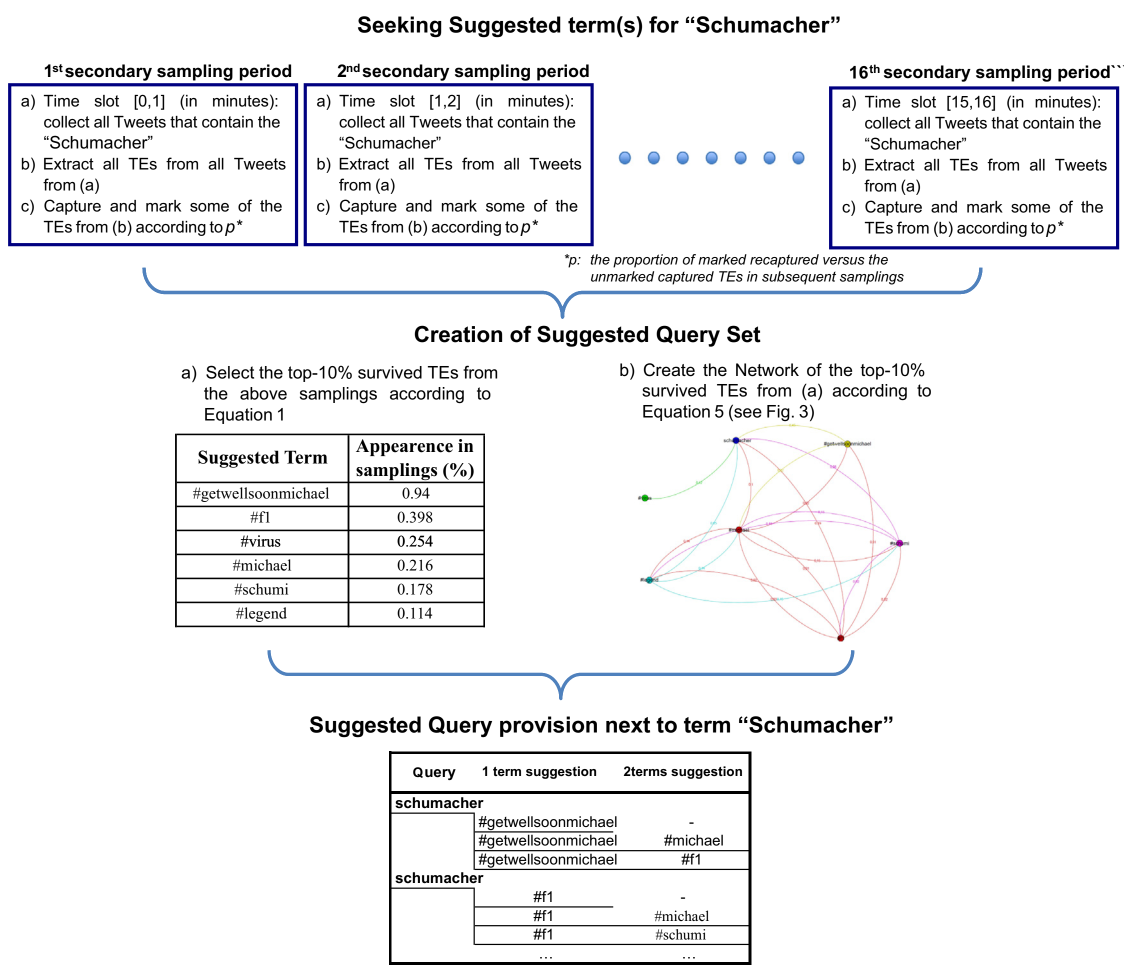


Fig. 2. Flowchart of our proposed algorithm towards Query Suggestion provision with an example.

fetch through the Twitter API all related tweets that contain this specific term. Then, from these tweets, we extract all Twitter Entities (hashtags, mentions and links) and finally we select some of them according to the proportion of marked recaptured versus the unmarked captured Twitter Entities during subsequent sampling occasions (in this example the first value between the 1st and the 2nd secondary period was measured close to $p=6.25\%$. As a result, the more a marked TE manages to appear again and again in the 16 subsequent samplings, the more viral is considered and becomes a strong candidate for suggested term. On the contrary, the fewer times a TE appears within the short-time subsequent sampling occasions, the less viral and important is considered and eventually will be ignored as suggested term. Finally, upon completion of the first primary sampling period, we select the top-$k\%$ TEs (in this example $k=10$) that managed to appear in most of the 16

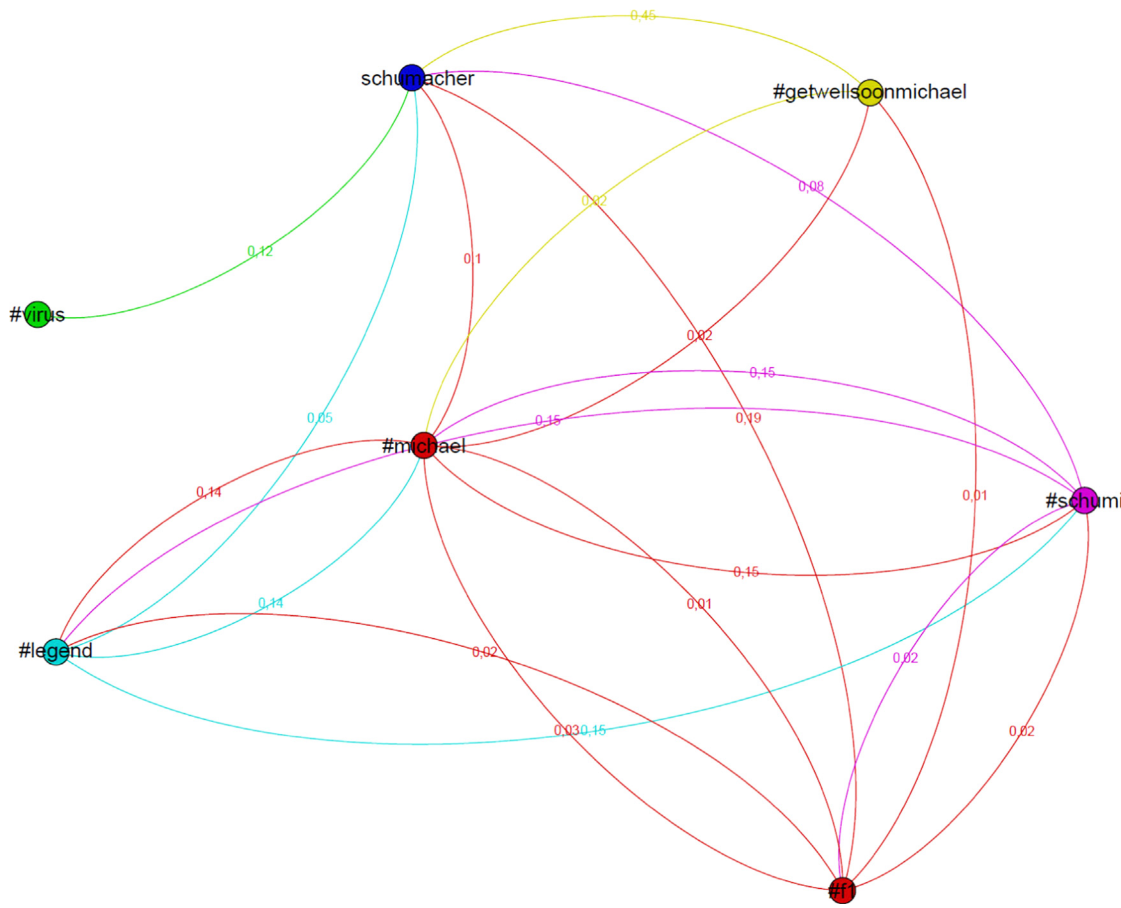


Fig. 3. Network created from the top-10% survived TEs for the example provided in Section 3.

separate secondary sampling periods of the first primary sampling period. These entities reflect a significant trend behavior with respect to the seed term. Especially the hashtag #getwellsoonmichael presented the higher survival rate reaching 94% frequency of appearances (appeared in 15 out of the total 16 samplings of the primary period). The rest top appearances were measured for #f1, #virus, #michael, #schumi and #legend, all related with the famous formula 1 driver who had a serious ski accident on December 29 of 2013 in the French Alpine resort of Meribel.

We want to stress out here that the capture–recapture paradigm helps us to not only suggest popular TE, but also to suggest TE that that remain popular in time. This practically means that we are not so interested in TE that appear suddenly and then die, but we are interested in TE that appear and re-appear in many subsequent sampling occasions. Thus, the concept of virality in this work is strongly related to survivability as estimated in real-life capture–recapture experiments.

3.2. Weighting factor – ranking suggested terms

The main scope of the weighting factor is to calculate the weights of the most trending entities provided from the survivability factor (above section) and then provide their ranking position towards query suggestion provision. After having the top-k% most frequent (survived) entities according to the survivability factor of Eq. (1), we further calculate their relation of co-appearance within the secondary sampling periods. This is performed by calculating their Twitter Semantic Weight (TSW) score according to Eq. (2), where $ER(e_x, e_y)$ defines the frequency of co-appearance for entities x and y. This actually provide us a ranking order (higher to lower TSW values) of the coupled Twitter Entities x and y.

$$TSW(e_x, e_y) = \frac{\|\phi_i(e_x)\| + \|ER(e_x, e_y)\|}{2} \tag{2}$$

Now, as far as our example is concerned, and as depicted in the flowchart of Fig. 2, the network that consists of the top-10% survived TEs according to Eq. (2) is illustrated in Fig. 3. The most frequently appeared and viral entity is the hashtag "#getwellsoonmichael"

Table 1
Top 10%-survived TEs among 16 subsequent samplings of January 13, 2014 – (a) Egypt, (b) Syria.

Query Suggestion Set	Metrics	
	TE type	φ_i *(top-10%)*
(a) Seed {Egypt}		
egypt_now111	@	0.881
anticoup	#	0.855
kuwait	#	0.615
saudi	#	0.602
morsi	#	0.539
uae	#	0.340
sta	#	0.168
(b) Seed {Syria}		
freethe7	#	0.830
iran	#	0.659
iraq	#	0.621
Free_Media_Hub	@	0.335
egypt	#	0.181
assad	#	0.160
un	#	0.143
isis	#	0.072

Table 2
TSW parameters between the top-3 and the rest survived entities (TEs) of the query suggestion set for January 13, 2014 – (a) Egypt, (b) Syria.

Entity(x)	Entity(y)	Freq.(x,y)	Freq.(seed,x)	Freq.(seed,y)
(a) Seed: Egypt				
@egypt_now111	#saudi	0	854	584
@egypt_now111	#kuwait	0	854	596
@egypt_now111	#morsi	0	854	523
@egypt_now111	#uae	0	854	330
@egypt_now111	#anticoup	0	854	829
@egypt_now111	#sta	0	854	163
#anticoup	#saudi	0	829	584
#anticoup	#kuwait	0	829	596
#anticoup	@egypt_now111	0	829	854
#anticoup	#morsi	32	829	523
#anticoup	#uae	0	829	330
#anticoup	#sta	0	829	163
#kuwait	#saudi	560	596	584
#kuwait	@egypt_now111	0	596	854
#kuwait	#morsi	324	596	523
#kuwait	#uae	227	596	330
#kuwait	#anticoup	0	596	829
#kuwait	#sta	163	596	163
(b) Seed: Syria				
#freethe7	#isis	0	545	47
#freethe7	#egypt	15	545	119
#freethe7	@Free_Media_Hub	0	545	220
#freethe7	#iraq	124	545	408
#freethe7	#iran	220	545	433
#freethe7	#un	183	545	94
#freethe7	#assad	0	545	105
#iran	#isis	1	433	47
#iran	#egypt	9	433	119
#iran	@Free_Media_Hub	0	433	220
#iran	#iraq	142	433	408
#iran	#freethe7	220	433	545
#iran	#un	198	433	94
#iran	#assad	0	433	105
#iraq	#isis	6	408	47
#iraq	#egypt	8	408	119
#iraq	@Free_Media_Hub	0	408	220
#iraq	#iran	142	408	433
#iraq	#freethe7	124	408	545
#iraq	#un	141	408	94
#iraq	#assad	6	408	105

(TSW=0.448) and as a result it is proposed as the first suggestion term next to query “Schumacher”, while the second one term suggestion is “#f1”. Regarding the best two-term suggestion with respect to the same seed, these are “#getwellsoonmichael_#michael”, as well as “#getwellsoonmichael_#f1”, having TSW values equal to 0.181 and 0.177, respectively, followed by “#f1_#michael”, “#f1_#schumi” with TSW equal to 0.084 and 0.082, respectively. The whole network that shows the relations and the respective weights between survived entities in the query suggestion set of the Schumacher case, as well as the query suggestions, are shown in Fig. 3.

We would like to note here that tokenization, topic/word segmentation, as well as further lexical analysis procedures that deal with breaking a stream of text up into words/phrases are not considered in this paper and are left for future work; they would be very useful, mainly for the hashtag entities.

4. Case studies: gaining more insights in the suggested results

In this section, we provide the query suggestion results we received with respect to two parallel one-term queries we considered, namely “egypt”, and “syria”. We divide the obtained results in two separate case studies. Experiments lasted for 8 days (January 8, 2014–January 15, 2014). For each day we conducted 8 primary sampling periods, while each primary period consisted of 16 secondary sampling periods, according to the modified capture–recapture model we followed (see Fig. 1 and the rationale of the flowchart of Fig. 2). In order to provide a thorough representation of the results, we analyze in detail the results of only one day (January 13th) for all tested cases, and we then summarized them. Finally, we compare the results derived from our case studies with respect to the query suggestions of well-known search engines, as well as to a heavily visited mainstream web media service.

4.1. Query suggestion provision over the case studies

As mentioned before, we aim at providing query suggestions to the user's submitted term(s) under only the knowledge disseminated publicly in Twitter, and without having any other access or use of query logs. As case studies in our experiments, we consider the political situations in Egypt and Syria, which have a constant interest for many years worldwide. Taking into account the knowledge we earned from two other previous work of ours [1,2], we initiated the results procedure by using “Egypt”, and “Syria” as seeds that correspond to our case studies.

Table 1 presents the entities with the top-k% survivability rate between 16 subsequent secondary samplings appeared in January 13, 2014, thus highlighting the entities that reflect a significant trend behavior with respect to the seed term during that day ($k=10$). Unfortunately, none URL appeared as survived entity among the top-10% population, while only two mention entities appeared each one for the case of “egypt” and “syria”, respectively. The reader can see some URLs and mention entities in Appendix for the cases of Egypt, and Syria, where k is equal to 30% and 20%, respectively. Entities are depicted in alphabetical order and are colored differently according to their type (blue-highlighted: mentions, green-highlighted: hashtags, red-highlighted: URLs). For example, the entity @egypt_now111 presented the higher survivability during that day, since it was captured in nearly 88% of all sampling occasions. Similarly, hashtags #freethe7 presented the higher survival rates for the Syria case, having a value equal to 83%. All entities that are highlighted in Table 1a as well as Table 1b form the Query Suggestion Sets for the two cases and derived as a result of the survivability factor φ_i as described in Section 3.1.

In conjunction to Table 1, Table 2 depicts some metrics between the top-3 survived entities and all others that belong to the query suggestion set. These values are taken for all three cases within the 16 subsequent sampling periods of January 13, 2014. In the third, fourth and fifth column of Table 2, we can see the frequency of co-appearances between entities x and y, as well between the seed term and x, y, respectively. For example, in the case of Egypt (Table 2a) hashtags #kuwait and #morsi appeared together in 324 captured tweets, while they co-appeared with the seed term (egypt) in 596 and 523 tweets, respectively.

Taking into account ER(e_x, e_y), the algorithm dynamically calculates the weights according to Eq. (2) and generates the network of related entities for further query suggestions. Fig. 4a and b depicts these networks with respect to the cases we investigate. Nodes correspond to the entities of the query suggestion set (survived entities from the capture–recapture experiments), while curving edges indicate clockwise the direction from a source to the target node and the respective weight. Networks (layout type: Fruchterman Reingold) are created with the open graph visualization Gephi tool.

In Appendix, we provide as many results can practically be depicted for these two cases. As we have mentioned earlier, the reader can see the query suggestion sets with all entities which have the top-k% survival rates within the daily capture–recapture experiments (“egypt”: $k=30$%, “syria”: $k=20$%).

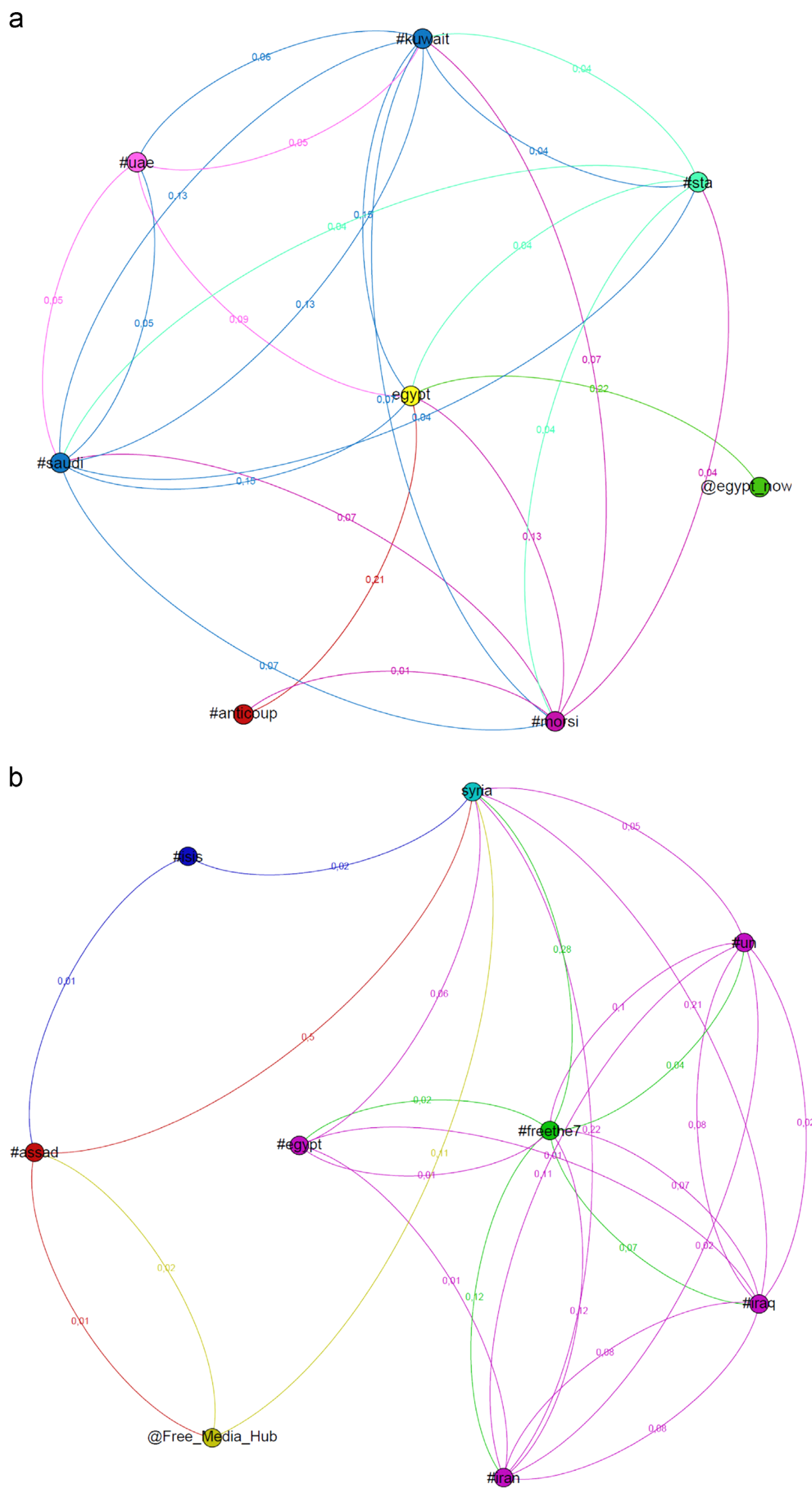


Fig. 4. (a) Network of survived Twitter Entities in Query Suggestion Sets – Egypt. (b) Network of survived Twitter Entities in Query Suggestion Sets – Syria.

Table 3
Query suggestions provided by Google, Yahoo!, Bing and Reuters in respect to our case studies (January 8–January 15 2014).

Google News			Yahoo! News		
rank	egypt 8-15 Jan 2014	syria 8-15 Jan 2014	rank	egypt 8-15 Jan 2014	syria 8-15 Jan 2014
1	snow	news	1	protests*	news
2	news	chemical weapons	2	news*	war 2013
3	constitution	rebels	3	shark attack	chemical attack
4	turkey	kurds	4	pyramids warning	chemical weapons
5	protest	nuns	5	elections	protests
6	economy	aleppo	6	locusts	israel
7	ghana	chemical	7	crisis 2011	uprising
8	russia	war	8	__'s president	russia
9	referendum	children	9	antiquities	turkey
10	clashes	fighting	10	israel	rebels

Bing News			Reuters		
rank	egypt 8-15 Jan 2014	syria 8-15 Jan 2014	rank	egypt 8-15 Jan 2014	syria 8-15 Jan 2014
1	news	news	1	news	the syrian front*
2	newspaper	latest news	2	economy*	news*
3	snow (14 Jan)		3	reuters*	chemical weapons
4	air	chemical weapons*	4	iran revolution	the snipers of
5	sherrod	map		qatar+ (10 Jan)	latest news+ (12 Jan)
6	map*	now*		egyptian+ (11 Jan)	syrian war+ (13 Jan)
7	chaos	war		constitution+ (14 Jan)	gas attacks+ (14 Jan)
8	facts*	tv*	* terms that their rank interchanges		
9	pyramids	tube	+ new entry (date)		

4.2. Evaluation of results against major search services

In this section we evaluate the query suggestions provided by our approach. Moreover, we compare the query recommendations with respect to the recommendations of Google News,[6] Yahoo! News,[7] Bing,[8] as well as from Reuters portal[9] for the two cases (seeds: "egypt", "syria"). We should note here that our intention is not only to evaluate the accuracy of the provided suggestions, but also to investigate how quickly the suggestions reflect their trends on a daily-basis. Our query suggestion set consists of user-generated content (#hashtags), main influencers and web content, which is disseminated and semantically annotated, in the Twittersphere (Twitter users, mentions, URIs). The innovation of our method relies in network analysis of a large ecosystem that involves users, semantics, and web content and not from query logs. In other words, suggestions are not driven from past user searches, but nearly on-the-fly and directly from the Twitter API. According to our knowledge, Google's predicting algorithm used for query suggestion displays search queries based on other users' search activities and the contents of Web pages indexed by Google.[10] In addition, Google users might also see search queries from their previous related searches. We suppose that the rest information services involved for comparison (Yahoo! News, Bing and Reuters) practically work under the same concepts. Still, we should note at this point, that if the search service uses a search results based approach, query suggestion depends mainly on a specific number of the top-N results for that query. Yet, if the service uses logs, query suggestion may be provided upon other relevant user query terms, or even other user personalized behavior-based search pattern.

Prior to starting our evaluation and discussion regarding our results, we introduce Table 3, which presents the query suggestions provided by Google, Yahoo!, Bing and Reuters within our testing period between the 8th and the 15th of January 2014 for the two case studies. This table is divided in four parts that reflect the results of the above-mentioned news services with respect to the tested seed terms. We can notice that Google and Yahoo! provided 10 suggestions per seed, while Bing and Reuters 9 and 4, respectively. In the first column of each part we have the ranking position of the suggestion. For example, this means that "constitution" was ranked as the third suggested term (seed: egypt) from Google, and "protests" as the fifth suggested term (seed: syria) from Yahoo!, respectively. In addition, blue highlighted terms followed by an asterisk, denote terms which their rank interchanged during the testing period. For example, in Bing results and for the seed "egypt", terms "map" and "fact" firstly appeared as the 6th and the 8th result, respectively, but later on their ranking positions was switched. Finally, green highlighted terms followed by "+", denote new terms that appeared on the position of an already suggested one. This means that in such cases, suggested terms are updated and the list is refreshed. For the period of our tests, such refresh activity appeared mainly from Reuters and less for Bing. For example, "iran revolution" was the fourth suggested term from Reuters with respect to the seed "egypt" for January 8 and January 9. Then, the next day (January 10th), it was replaced in the same ranking position by the term "qatar", then by the term "egyptian" (January 11) and then by the term "constitution" for the rest days (January 14–15). Similarly for the seed "syria", new suggestions over "the snipers of" appeared on the fourth ranking position during January 12, 13 and 14 with the two-term suggestions "latest news", "Syrian war" and "gas attacks", respectively. The only refresh activity from the other three search engines appeared for Bing, where the term "newspaper" was replaced by the term "snow" on January 14th.

Evaluation results with respect to these cases and our method are depicted in Table 4, where we can see on the left the initial query terms (seeds), followed by suggested entities in two levels that correspond to entity(x) and entity(y). These levels resemble to the

[6] https://news.google.com/.
[7] http://news.yahoo.com/.
[8] http://www.bing.com/news/.
[9] http://www.reuters.com/.
[10] http://support.google.com/websearch/bin/answer.py?hl=en&answer=106230.

Table 4
Query suggestions according to Twitter Semantic Weighting (January 13, 2014) – (a) Egypt, (b) Syria.

Seed	1st level suggestion Entity(x)	2nd level suggestion Entity(y)	*TSW weights*	
(a) Egypt				
	@egypt_now111		*0.220*	
	#anticoup		*0.214*	
		#morsi		*0.136*
	#kuwait		*0.154*	
		#saudi		*0.117*
		#morsi		*0.108*
	#saudi		*0.151*	
		#kuwait		*0.115*
		#morsi		*0.106*
	#morsi		*0.135*	
		#saudi		*0.097*
		#kuwait		*0.097*
	#uae		*0.085*	
		#kuwait		*0.063*
		#saudi		*0.062*
	#sta		*0.042*	
		#kuwait		*0.033*
		#saudi		*0.033*
		#morsi		0.033
(b) Syria				
	#freethe7		*0.277*	
		#iran		*0.147*
		#un		*0.139*
	#iran		*0.220*	
		#freethe7		*0.126*
		#un		*0.121*
	#iraq		*0.207*	
		#iran		*0.105*
		#freethe7		*0.105*
	@Free_Media_Hub		*0.112*	
		#assad		*0.046*
	#egypt		*0.060*	
		#freethe7		*0.029*
		#iraq		*0.026*
		#iran		*0.026*
	#assad		*0.053*	
		@Free_Media_Hub		*0.026*
		#isis		*0.023*
	#un		*0.048*	
		#freethe7		*0.032*
		#iraq		*0.025*
		#iran		*0.024*
	#isis		*0.024*	

automatic recommendation provided by several search engines based on the already submitted user term(s). Finally, according to Eq. (2) column "TSW weights" corresponds to the weighting values for the query suggestions "{seed}- _entity(x)" and "{seed}_entity(x)_entity (y)". For simplicity reasons, Table 4 holds only the TSW values of the first two second level entities y that suggests entity x.

4.2.1. *Egypt case*

Moreover, for the case of Egypt political situation, the entity that presented the higher TSW during the 13th of January 2014 (equal to 0.220) was the Twitter user @egypt_now111. This Twitter account disseminates breaking news regarding the unstable political situation in Egypt, having nearly 170.000 followers and 19.000 tweets. It is worth noticing that Al Jazeera's Twitter account has less than the half of followers and tweets (nearly 65.000 and 8.000, respectively). Even thought there were many other Twitter Entities captured and marked along with @egypt_now111, there is no second level suggestion since most of them are written in Arabic. That is why in Fig. 4a there is no other edge from @egypt_now111 node to the other nodes of the query suggestion set. The entity with the second most higher TSW (0.214) is the hashtag "#anticoup", which obviously comes from combining the Greek term "anti-" (expressing opposing to or against to something/someone) and "coup". This proposal was highly relevant and quite trendy with respect to the political status within the testing period, since there were remonstrations in many Egypt cities and villages condemning coup crimes. This trend was not captured by the other services as we can see from Table 3. However, the suggestion "protest (s)" that actually reflects similar actions over the coup in Egypt, was suggested by Yahoo! and Google in the first and fifth place accordingly. As a second level suggestion related to "#anticoup" is "#morsi", having a TSW equal to 0.136. This means that in case we want to find two-term suggestions (entities in this paper) for our seed, then the sequence "#anticoup #morsi" is the most frequent among others.

Then, as the third, fourth and fifth recommendation with respect to the seed "egypt", we have the hashtags "#kuwait", "#saudi" and "#morsi" which are strongly related with each other, since everyone is suggested by the other two in a second level. For instance, "#saudi" and "#morsi" suggest "#kuwait", "#kuwait" and "#morsi" suggest "#saudi", as well as "#saudi" and "#kuwait" suggest "#morsi" according to a descending weighting order. This is also highlighted in Fig. 4a where we can notice that the corresponding entity nodes are networked one-by-one. This strong network relation between these three entities is justified, since during the period of our experiments the trial of ex-president of Egypt (Morsi) was supposed to take place at January 8 (eventually it was postponed). In addition, a day prior to this trial there was a political declaration in favor of the new Egyptian government and against Morsi's believers by Saudi Arabia, the United Arab Emirates, and Kuwait. This was also noticed with the "#uae" as the sixth first level suggestion, where even though the TSW values were lower, yet "#uae" was suggested by "#kuwait" and "#saudi" in a second level. The whole network that shows the relations and the respective TSWs in the query suggestion set of seed "egypt" case, as well as the proposed query suggestions, are shown in Fig. 4a and Table 4a, respectively.

4.2.2. *Syria case*

Results with respect to "syria" as a seed term, revealed strong relations between entities "#freethe7", "#iran", "#iraq", and "#egypt". Similarly with the above explanation, the top-3 suggestions included "#freethe7", "#iran", "#iraq" with TSW values 0.277, 0.22 and 0.207, respectively. As far as the top two-term suggestions with respect to the seed, these were "syria_#freethe7_#iran", "syria_#freethe7_#un" with TSW values equal to 0.147 and 0.139, respectively, followed by "syria_#iran_#freethe7" and "syria_#iran_#un" with TSW values 0.126, 0.121, respectively. The trending behavior of these entities is justified not only due to the political positions and status between those countries over the last years, but also because a couple of days before capturing these trends, UN experts urged Iraq to establish once again the fate and whereabouts of the seven residents of Camp Ashraf, who were allegedly abducted on September of 2013 after an attack in which more than 50 persons were killed. The network that shows all relations and their respective weights between proposed entities in this case, as well as the query suggestions, is shown in Fig. 4b and Table 4b, respectively.

4.2.3. *Comparison with other services*

Now seeing the query suggestions provided by the other search services (see Table 3), it surely worth discussing the performance of Google, Yahoo!, Bing and Reuters in terms of query suggestion freshness (how often these services update their suggestions after a query). Google News surprised us negatively since its query suggestions were identical and static for all evaluation period (from January 8, 2014 up to January 15, 2014). Yahoo! News had also static suggestions with respect to the tested terms. There were only some re-rankings for the first two proposed suggestions in the case of Egypt, but no new entries. On the other hand, Bing presented more freshness activity mostly as far the re-ranking of its suggestions. In addition,

Table 5
The five-point Likert scale used for user ratings.

1	2	3	4	5
Strongly disagree (Totally irrelevant suggestion)	Disagree (Not so good suggestion)	Neither agree nor disagree (Nearly same suggestion)	Agree (Potentially better)	Strongly agree (Surely better)

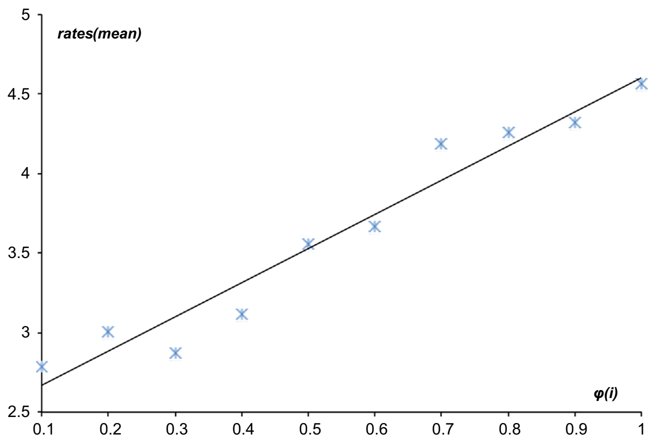


Fig. 5. Rates (in mean values) vs. proposed Twitter Entities from Query Suggestion sets.

nearly the end of our evaluation period (January 14), Bing replaced the suggested term "newspaper" with the term "snow" in the second position (seed "egypt"). However, this replacement was quite outdated since it was related with a snowfall in Northern Egypt territories, but nearly more than a month ago (in mid-December of 2013).

Among all, Reuters provided the most dynamic results in terms of freshness and position re-ranking. Despite the fact that Reuters returns fewer suggestions, these are updated quite often. From Table 3 we notice that within an eight-day period and for the cases of Egypt and Syria, the last suggestions were replaced by newer ones at least three times (10th, 11th, 14th and 12th, 13th, 14th of January, respectively). According to our research in query suggestion for web news services, this is the second time where Reuters receives the best comments with respect to trendy proposed suggestions near a user's query (see the evaluation described in [2]). Finally, in an attempt to directly compare if our approach "captures" the new replacements made by Reuters, we checked whether these suggestions appeared in our records. So, as it can be seen in Appendix, the entities "#qatar", "#egyconstitution" (obviously concatenation of egypt and constitution), as well as the entity "#gasattacks" appeared among the top-30% and top-20% survived entities with respect to the cases of Egypt and Syria.

5. Evaluation – discussion

The overall evaluation of our proposed methodology is twofold. At first, we describe a generic evaluation, which involves subjective user ratings for results obtained from our approach and from Google Hot Trends. Then we provide comparative evaluations with respect to two similar baseline methods of the literature, namely a cluster labeling and a microblog retrieval task, over traditional information retrieval metrics.

5.1. Evaluation against user ratings

In order to evaluate whether query suggestions returned from our method satisfy the user's information needs we engaged 17 postgraduate students from an MBA course class at National Technical University of Athens. Their task was to subjectively rate the suggested queries against the Google Hot Trends service.[11] Each student was asked to select 3 different events from Google Hot Trends for a specific testing period. Furthermore, each individual had to explicitly rate the suggested entities as these were derived by our query suggestion method, against their selected events as appear in Google Hot Trends. Google Hot Trends displays several top fastest rising searches (and search-terms) by day in the U.S.A. The rating performed upon a five-point Likert scale (see Table 5).

For simplicity reasons and without any loss of generality students were asked to rate only hashtags as extended entities (terms) and considered only the survivability factor (as described in Section 3.1). After processing the one-week results we ended up with 87 unique related terms (as these were provided by Google Hot Trends) in 31 distinct events (20 out of the 51 events were identical). The average amount of suggested terms per tested event was 2.81, which practically means that nearly 3 terms in average suggest the basic term that describes a specific event. The inter-annotator agreement was the Fleiss' kappa statistical measure for assessing the reliability of agreement between a fixed number of raters. In our evaluation we had 17 individuals (raters) for assigning 87 unique related terms (subjects) to a total of 31 distinct events. Value of kappa statistical measure was measured at nearly 0.37, which is an almost fair agreement according to the literature [6]. At Fig. 5, we can see some points that indicate the average evaluator rating for suggested entities, as derived from our proposal. In addition, Table 6 summarizes all mean rate values with respect to the survival rate, as well as other parameters taken into consideration for this evaluation.

It is worth noticing that the larger the survival rate (φ) is, the higher the mean subjective rate appears in the five-point Likert scale. This proves that in this way the query suggestion set formed, consists of more trendy entities related to each other. More specifically, we noticed that entities belonged in query suggestion sets that had survivability rate above 0.7, were subjectively evaluated as more relevant, thus presenting nearly one-point higher level in the Likert scale. This practically proves that through subsequent samplings in Twitter, the most viral entities are trendier in comparison to a related query term resides in Google's log. This was somehow expected, since we performed short-term trend analysis rather than long-term log analysis, yet it is an indicative assumption that our query suggestion method is in the right direction. We can also notice that the majority of subjective rates in average values (more than 60%) were close or slightly higher in the third level (point 3) in the Likert scale, thus indicating a "nearly same suggestion" in comparison to the compared Google search service.

5.2. Evaluation against a cluster labeling and a microblog retrieval task

Towards a more comprehensive evaluation, we compared our approach with respect to two similar baseline approaches, namely a cluster labeling and a microblog retrieval task.

5.2.1. Comparative evaluation against a cluster labeling task

During the first comparison approach we consider as baseline the evaluations taken into consideration in the work described in [7].

[11] http://www.google.com/trends/hottrends.

Table 6
Evaluation metrics against user ratings.

$\phi(i)$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Rates (mean values)	2.72	3.05	2.79	3.12	3.55	3.61	4.21	4.34	4.41	4.48
# of individuals (raters): *17*										
Unique related terms (subjects): *87*										
Distinct events: *31*										
Inter-annotator agreement: *Fleiss' kappa statistical measure k=0.37*										

Table 7
Results from methods that integrate the BOW and our method (cluster labeling task).

Method/metric	F1-score	Deviation over BOW (±%)	Accuracy	Deviation over BOW (±%)
BOW	0.468	N/A	0.510	N/A
WordNet_Method	0.476	(+1.735)	0.520	(+1.919)
Wiki_Method	0.499	(+6.687)	0.541	(+6.023)
SemKnow_Method	0.500	(+6.929)	0.543	(+6.477)
Our Method	0.532	(+13.624)	0.596	(+16.825)

Table 8
Comparing results (ranking problem).

Method/metric	nDCC5	Deviation over Kphrase (±%)
Kphrase	0.438	N/A
WordNet_Method	0.448	(+2.281)
Wiki_Method	0.520	(+18.761)
Our Method	0.576	(+31.484)

In this work, the authors propose methods to aggregate related microblogging messages into clusters and automatically assign them semantically meaningful labels. They use hot queries of diverse topics selected from Google Hot Trends, where each query is considered to be a trending topic, while they consider the top-five query suggestions from Google as subtopics of this topic. In order to be able to compare our approach on a common basis, we considered the suggested TEs for us to form a cluster similar to the ones in [7]. This actually means that the generation of relevant labels around a topic/subtopic is considered similar to selecting appropriate terms in the Query Suggestion Set of our method. So, for collecting the pool of our data, we systematically crawled the hot queries published by Google Hot Trends between May 9 and June 9 of 2014, having in mind 20 hot queries from miscellaneous fields of interest, for which Google provided some relevant/similar queries. Each selected hot query was considered as a different topic. Then, for each topic we further crawled the top-three query suggestions, thus forming totally 60 separate subtopics. The returned suggestions formed a cluster, while subtopics were considered associated with the cluster label. For each subtopic (query suggestion), we harvested exactly the last 200 tweets from Twitter. As a result we collected an amount of nearly 12.000 tweets with all of their TEs.

The next step included comparison of our clusters with the clusters formed by the 3 methods mentioned in [7], namely the WordNet_Method (WNT), the Wiki_Method (WK) and the SemKnow_Method (SMK). We used the harmonic mean of precision and recall (F1-score) and Accuracy metrics to evaluate the performance of the compared methods. Table 7 depicts the values of these metrics for all three methods, as well as ours. Since the clusters in our method dynamically change every 3 h (time interval between subsequent primary periods), the respective values in the last row of Table 7 correspond to their average values across all testing period. We noticed that all approaches improve the classical BOW model, both in terms of F1-score and accuracy. Still, the best improvement appeared for our method and it is 13.6% better with respect to the BOW model and 6.5% better compared (wrt absolute F1-score values) to the second best method (SMK). In terms of Accuracy, our method presented a 16.8% and 9.8% (wrt absolute accuracy values) improvement having BOW and SMK as baselines, respectively. Similarly to the work described in [7] we noticed that SMK, WK and WNT increase (in that particular order) the level of the used metrics.

Table 9
Evaluation metrics for BS+*R*, RM2 and our method.

Method/metric	MAP {Deviation over BS+*R* (±%)}	MRR {Deviation over BS+*R* (±%)}	P@5 {Deviation over BS+*R* (±%)}
BS+R	0.362 {N/A}	0.723 {N/A}	0.468 {N/A}
RM2	0.394 {+8.831}	0.615 {−14.938}	0.493 {+5.341}
Our Method	0.477 {+31.768}	0.923 {+27.663}	0.897 {+91.667}

Now, in order to enhance the aforementioned evaluation, we considered the suggested top-5 TEs generated by our weighting factor Eq. (2) to be the best labels per subtopic for our method. This allowed us to directly compare our method with WNT, WK and SMK, treating the cluster-labeling task as a problem that ranks all concepts from Wikipedia and the best matched label for a cluster of microblogging messages. As ground truth for cluster labeling, we considered the subtopics used for crawling microblogging messages. Table 8 summarizes the results based on the normalized Discounted Cumulative Gain up to the fifth position of the ranked results (nDCG5). Similarly to the previous described clustering evaluation, values in the last row of Table 8 (our method), correspond to averaged values across all testing period. As we can see, our method presents the best normalized Discounted Cumulative Gain value in comparison with the other baselines. We strongly believe that this is due to the fact that our method provide quite up-to-date and "fresh" query suggestion in terms of TE (Hashtags, User mentions, URIs) that derive directly from users' intelligence and their capability to describe content.

5.2.2. Comparative evaluation against a microblog retrieval task

A second evaluation procedure was followed against two other baselines described within [14], where authors present a model for retrieving microblog posts enhanced with textual and microblog quality indicators, as well as with a dynamic query expansion model. In particular, we wanted to test the ability of our query suggestion mechanism in terms of viral terms recommendation given a trending topic. So, for the same period utilized in our previous evaluation (May 9, 2014–June 9, 2014), we selected some trending topics as proposed by Twitter, thus forming 20 different queries. Working similarly to [14], we harvested all tweets that have been posted between the very last day the topic was announced as trending and three days before that day, ending up with nearly 28.400 tweets. We then followed a simple procedure that required retrieval experiments with respect to the top-5 results for all trending topics that fall within the three-day time window. We should note here that if a topic presented a trending behavior for more than one day (and there were many such cases!), the experimentation run only for the first three-day time window, just before the day the topic appeared in Twitter Trends for the first time.

Results were judged as relevant or not. The inter-annotator agreement was once more the Fleiss' kappa statistical measure for assessing the reliability of agreement between the raters, who in this case were 4. The value of kappa statistical measure was measured at the level of 0.74, while the evaluation metrics used were the Mean Reciprocal Rank (MRR), the Mean Average Precision (MAP), and the Precision at the fifth position (P@5). The baseline was a Boolean search method, strongly biased towards newer results. That means that newer tweets were ranked in higher position. This baseline is called "*Boolean search with recency features*" (BS+R). Joint to this method, a classical relevance model (RM2) was also employed [11].

Table 9 depicts the metric values we considered for evaluation purposes, as well as the deviation over the baseline. We observe that RM2 improves the Mean Average Precision as well as the Precision at the fifth position by nearly 8.8% and 5.3%, respectively. For the same method, the Mean Reciprocal Rank was measured in nearly 15% lower level values with respect to BS+R. Now as far as our method is concerned, it performed significantly better than the rest, considering all the above metrics; the Mean Reciprocal Rank was improved by 27.7% and nearly 50% (in terms of absolute MRR values) compared to the baseline and RM2, respectively. Similar improvements were achieved for the Mean Average Precision values at the levels of 31.8% and nearly 21% (in terms of absolute MAP values), respectively. However, the most impressive improvement was measured for the P@5 metric. Our approach returned more than 91% higher precision level in comparison to BS+R (reaching the value of 0.9), while it also outperformed the traditional RM2 method by leveraging the P@5 value up to nearly 82% (in terms of absolute P@5 values). This actually proves how significant it is to provide viral suggested terms (in our case TEs) through a query suggestion method and not only rely on recency criteria. In addition to that, this kind of evaluation revealed that tokens with numeric or non-alphabetic characters, which are usually eliminated by traditional information retrieval methods, are of great importance towards query suggestion in microblog post search.

6. Conclusions – future work

In this paper, we introduced a query suggestion method based on a social network derived by related trendy entities that become viral in Twitter worldwide. The innovation in this work stands in the fact that we use the users' intelligence and capability to describe information (e.g., through the hashtags), as well as the power that social media have to validate it, enhance it, or modify it in real-time. In comparison to other query suggestion methods, the added value of our proposal is two-fold, since we achieve better freshness and trendiness rates for our query suggestion set. We witnessed many cases, in which suggestions proposed by our method were not appeared in the lists of other known web search services, as well as many cases where the suggestions of commercial services like Google, Yahoo! and Bing were actually obsolete and rather outdated. In addition, our suggestions are based on common appearances of Twitter Entities (e.g., hashtags, mentions/replies to others users, web content via tinyURLs) in human annotated content (tweets). Such a content is rapidly disseminated, while if it is of great interest, it is maintained and reproduced many times, reused with other Twitter Entities, thus forming a dynamic network of resilient content capable of creating trends, top news, thematic categorizations, top-influencers, etc. We demonstrated in this work how the most viral part of this network can be used in order to suggest related terms with respect to a user query. Another positive experience we had, showed us that our suggestions might also protect users from phishing attacks, malware material and other cybercrimes. Finally, we ended up with some quite promising evaluations. The first enrolled human raters and subjective comparisons of suggested results, with respect to related terms provided by Google for similar news events. The second evaluation provided us comparative results that clearly show that our work does not exactly belong solely to generic query-suggestion research category, since it provides a broader, semantic-based view on it and most importantly it additionally utilizes the popularity/virality of Twitter Entities in the process. The high-level TEs that we aim to identify and suggest, are characterized as "semantic entities carrying meaningful information", rather than "meaningless pieces of information".

Thus, the main contributions towards the construction of the proposed methodology can be identified with respect to five different aspects; each one of them illustrates in an increasing manner the contribution of our research efforts. More specifically our proposed approach:

a. is based on a novel methodology, since it introduces the benefits of viral social content for the query suggestion problem,
b. is based on a solid research foundation, since it was evaluated against famous search services, user subjective ratings, as well as against similar baseline approaches and traditional information retrieval metrics,
c. successfully incorporates multiple types of information knowledge (e.g., temporal, textual and social content),
d. advances typical query suggestion methodologies by taking into account the concept of collective intelligence, as well as
e. further exploits the notion of Twitter Entities with respect to the query analysis research task.

Having a first experience from this work, in the future we aim to further investigate the benefits of our proposal, as a resource/social suggestion mechanism. Towards the first goal, we are working on a service where given a query (seed) the system will recommend you possible URIs and multimedia content from well-known and popular social networks, such as ImgUr,[12] YouTube[13] and others. With respect to social suggestion, we are working in a similar way to suggest high-influencers through social sphere (e.g., follow a user, account), or to provide recommendations for joining a whole group and/or community. In parallel to the above, we plan to compare the provided query suggestions with other approaches that consider time as a virality factor when generating suggestion terms, along with the recency factor evaluated in this work.

In addition, we intent to further investigate the feasibility of defining an adaptive mechanism capable of selecting the top-k% viral Twitter Entities when forming the Query Selection Set, given how viral a user's query is. Also it will be very interesting to investigate the impact of this value (k), as a trade-off parameter between the system's latency (towards fast query provision) and classical information retrieval metrics (e.g., Mean Average Precision, etc.). Also, given the fact that query suggestion resembles with the cluster labeling problem, our work can be used for enhancing unstructured microblogging messages similarly to the work of [17] and others that deal with same issues.

Finally, future work includes issues such as an extension towards the semantification of the query suggestion mechanism. This practically means, the association of related or synonymous hashtags for future queries, the hierarchical expression of types and relationships between suggested terms, as well as the involvement of well-known semantic vocabularies (e.g., FOAF protocol, Dublin Core Metadata Initiative (DCMI), etc.).

Appendix

See Tables A1 and A2.

[12] http://imgur.com/.
[13] http://www.youtube.com/.

Table A1
Viral Twitter Entities with top-30% survival rates for January 13, 2014 – seed: “egypt”.

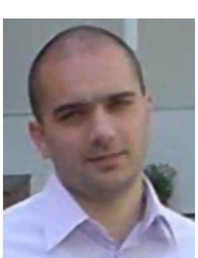
1/8/14	1/9/14	1/10/14	1/11/14	1/12/14	1/13/14	1/14/14	1/15/14
@diet_hawaa	@Morsi_RT	@EgyAntiCoup	@asmaamo83850399	@dignity4theold	@Adamitv	@borzou	@devotion4countr
@el_balad	@youm7	@egypt_now111	@7asnaad1g	@el_balad	@ArabNewsRt7	@dodomoeen	@egypt_now111
@KGirls66	#_	@freedom2mankind	@baladtv	@KGirls66	@arabresistance	@egypt_now111	@el_balad
#_	#abudhabi	@NewIranFree	@CBC_EGY	@Morsi_RT	@Beltrew	@elisferre	@fumi210500576
#abudhabi	#anticoup	@NseejNews	@el_balad	@twahodvoice	@CBC_EGY	@esfahanhanim	@fumikop500579
#android	#bahrain	@rightnowio_feed	@NadaaHesh	@TweetEgyptian	@CBCExtra	@Hazem_Azim	@RiicardoCB
#anticoup	#cairo	@syria_now111	@shamsalhurryah	@youm7	@EgyBloodBank	@IslamRahman	@Shabace1
#ara	#dubai	#abudhabi	@twahodvoice	#_	@egypt_now111	@KarlreMarks	@shadihamid
#arab	#egyconstitution	#anticoup	#_	#bahrain	@el_balad	@kelo3adi	@SonsOfEgypt
#bahrain	#ff	#egyconstitution	#abudhabi	#cairo	@Morsi_RT	@Misr25TV	@syria_now111
#balad	#freethe7	#ff	#ara	#egypt	@TweetEgyptian	@Morsi_RT	@TheBigA7a
#cairo	#iran	#freethe7	#bahrain	#ff	@youm7	@TheMiinz	@youm7
#dubai	#iraq	#iran	#egypt	#freethe7	#_	#_	#_
#egyconstitution	#ksa	#iraq	#ff	#iran	#anticoup	#anticoup	#abudhabi
#ff	#kuwait	#ksa	#freethe7	#iraq	#ara	#bbc	#anticoup
#freethe7	#lebanon	#kuwait	#iran	#ksa	#bahrain	#cairo	#ara
#iran	#morsi	#morsi	#iraq	#kuwait	#balad	#coup	#bahrain
#iraq	#qatar	#saudi	#ksa	#lebanon	#dubai	#coupstitution	#balad
#ksa	#saudi	#sex	#kuwait	#morsi	#egypt	#dubai	#cairo
#kuwait	#sex	#sta	#lebanon	#q8	#ff	#egyconstitution	#dubai
#lebanon	#sisi	#syria	#morsi	#qatar	#freethe7	#egypt	#egypt
#morsi	#syria	#uae	#nowplaying	#rt	#iran	#ff	#ff
#nowplaying	#uae	http://t.co/CQhAHCy4ch	#q8	#saudi	#iraq	#freethe7	#freethe7
#oman	http://t.co/RDEhHMn6tm	http://t.co/D00QkdoG9z	#qatar	#sex	#ksa	#iran	#iran
#q8	http://t.co/sckuvUKUtJ	http://t.co/DWeu2E7E00	#rt	#sta	#kuwait	#iraq	#iraq
#qatar	-	http://t.co/N1GFYmdquR	#saudi	#staracademy	#morsi	#jan25	#ksa
#saudi	-	http://t.co/sckuvUKUtJ	#sta	#syria	#news	#ksa	#kuwait
#sex	-	http://t.co/WG2aABrX8g	#staracademy	#uae	#rt	#kuwait	#lebanon
#syria	-	-	#syria	http://t.co/3eoTyXd1rB	#saudi	#morsi	#qatar
#uae	-	-	#uae	http://t.co/D00QkdoG9z	#sharon	#news	#referendum
http://t.co/DWeu2E7E00	-	-	http://t.co/3eoTyXd1rB	http://t.co/DWeu2E7E00	#sta	#obama	#saudi
http://t.co/J9xh2TYLNt	-	-	http://t.co/DWeu2E7E00	http://t.co/J9xh2TYLNt	#syria	#r4bia	#sta
http://t.co/lKh2rCWL0V	-	-	http://t.co/H0EqkoVsG2	http://t.co/KswIwWWaoI	#uae	#referendum	#syria
-	-	-	http://t.co/J9xh2TYLNt	-	#un	#saudi	#uae
-	-	-	http://t.co/RDEhHMn6tm	-	#usa	#syria	#un
-	-	-	-	-	http://t.co/sckuvUKUtJ	#uae	#yemen
-	-	-	-	-	http://t.co/yb4c67OkOJ	http://t.co/sckuvUKUtJ	http://t.co/lKh2rCWL0V
-	-	-	-	-		http://t.co/yb4c67OkOJ	http://t.co/ORk0I6FSl2
-	-	-	-	-	-	http://t.co/ZdhnOToith	http://t.co/RDEhHMn6tm
-	-	-	-	-	-	-	-

Table A2
Viral Twitter Entities with top-20% survival rates for January 13, 2014 – seed: “syria”.

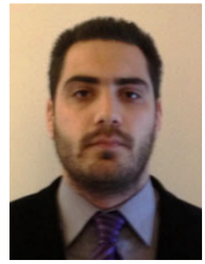
1/8/14	1/9/14	1/10/14	1/11/14	1/12/14	1/13/14	1/14/14	1/15/14
@Free_Media_Hub	@GSolaimani	@AM000Z	@Free_Media_Hub	@30mehr	@Free_Media_Hub	@Free_Media_Hub	@melissarfleming
@mxxxx_mmmmm7430	@tintin1957	@epaulnet	@GeredaV	@Free_Media_Hub	#abc	@IAC4RC	@nasrinforiran
@nasrinforiran	#_	@greyfoxguy	@GSolaimani	@GSolaimani	#afp	@nasrinforiran	@SyriaTwitte
@Shareif	#aleppo	@mxxxx_mxmxm7422	@i_magpie	@SIranChange	#aljazeera	@Refugees	@UN
@tintin1957	#aljazeera	@nasrinforiran	@NcrIran	@tintin1957	#ap	#_	#_
#_	#assad	@no2censorship	@RevolutionSyria	#_	#assad	#abc	#afp
#aleppo	#bahrain	@TheIslamicUmmah	@tintin1957	#abc	#campashraf	#afp	#aleppo
#aljazeera	#campashraf	@tintin1957	#abc	#afp	#campliberty	#aljazeera	#aljazeera
#assad	#cbs	#abc	#aljazeera	#aljazeera	#cnn	#assad	#assad
#bahrain	#foxnews	#ap	#ap	#bahrain	#egypt	#breaking	#cnn
#breaking	#freethe7	#assad	#assad	#breaking	#freethe7	#campashraf	#egypt
#campashraf	#google	#bahrain	#assadcrimes	#damascus	#gasattacks	#campliberty	#euronews
#cbs	#hama	#breaking	#bahrain	#fox	#geneva2	#cnn	#foxnews
#damascus	#homs	#campashraf	#belgium	#freethe7	#google	#damascus	#freethe7
#egypt	#iran	#cnn	#campashraf	#health	#iran	#egypt	#google
#freethe7	#iraq	#egypt	#campliberty	#iran	#iraq	#euronews	#health
#fsa	#isis	#fox	#damascus	#iraq	#isis	#fox	#iran
#gasattacks	#kuwait	#freethe7	#euronews	#isis	#jordan	#freethe7	#iraq
#google	#lebanon	#gasattacks	#fox	#islam	#lebanon	#google	#kuwait
#iran	#nbc	#homs	#freethe7	#nbc	#nbc	#iran	#nbc
#iraq	#news	#iran	#google	#news	#pmoi	#iraq	#pmoi
#ksa	#politics	#isis	#homs	#sms	#politics	#lemonde	#rajavi
#lebanon	#qatar	#london	#iran	#sun	#reuters	#libyan	#reuters
#nbc	#rajavi	#london	#ksa	#sydney	#russia	#nbc	#saudi
#news	#reuters	#news	#kuwait	#syria	#seckerry	#pmoi	#sms
#oman	#russia	#pmoi	#lebanon	#un	#syria	#rajavi	#sun
#politics	#sms	#rajavi	#lemonde	#unami	#uae	#reuters	#sydney
#sms	#syria	#reuters	#unami	#usa	#un	#syria	#syria
#sun	#uae	#rt	#unhcr	http://t.co/5Jsm2IRg0C	#unami	#un	#syriacrisis
#syria	#uk	#sms	#upi	http://t.co/vUZtxXoomC	#upi	#unhcr	#uae
#un	#un	#sun	#usa	-	#world	#world	#un
#unami	#unami	#syria	#world	-	http://t.co/01tSxI2klt	http://t.co/1J2mrY70Fq	http://t.co/SoBQgWybzJ
#unhcr	#upi	#us	http://t.co/8KabhVygS9	-	http://t.co/6JEZvw5HyR	http://t.co/d31Qw0p0uH	http://t.co/YZvwPljb7y
#usa	#world	#usa	http://t.co/krvCS7rs2d	-	http://t.co/83SrzqVlpn	http://t.co/r21jSS4Eah	-
#world	http://t.co/i8QegFximG	http://t.co/1E9UxXC3I9	-	-	http://t.co/jhGROR01IV	-	-
http://t.co/8nhfTJO9iz	http://t.co/KssepuNIfV	http://t.co/8N1W4UhsD0	-	-	-	-	-
http://t.co/F79vjGXToI	-	http://t.co/bTYwjRs393	-	-	-	-	-
	-		-	-	-	-	

References

[1] I. Anagnostopoulos, G. Razis, Ph. Mylonas, C.N. Anagnostopoulos, Query expansion with a little help from Twitter, engineering applications of neural networks, Commun. Comput. Inf. Sci. 384 (2013) 20–29.
[2] I. Anagnostopoulos, V. Kolias, Ph. Mylonas, Socio-semantic query expansion using Twitter hashtags, in: Proceedings of the 2012 7th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP '12), Luxembourg, 2012, pp. 29–34.
[3] A. Bandyopadhyay, K. Ghosh, P. Majumder, M. Mitra, Query expansion for microblog retrieval, Int. J. Web Sci. (2012) 368–380.
[4] C. Biancalana, F. Gasparetti, A. Micarelli, G. Sansonetti, Social semantic query expansion, ACM Trans. Intell. Syst. Technol. 4 (4) (2013) 1–43.
[5] M. Efron, Hashtag retrieval in a microblogging environment, in: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, Geneva, 2010, pp. 787–788.
[6] J. Geertzen, Inter-rater agreement with multiple raters and variables. (Retrieved June, Dec 11, 2013), ⟨https://mlnl.net/jg/software/ira/⟩, 2012.
[7] X. Hu, L. Tang, H. Liu, Enhancing accessibility of microblogging messages using semantic knowledge, in: Bettina Berendt, Arjen de Vries, Wenfei Fan, Craig Macdonald, Iadh Ounis, Ian Ruthven (Eds.), Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM '11), ACM, New York, NY, USA, 2011, pp. 2465–2468.
[8] B.A. Huberman, Daniel M. Romero, Fang Wu, Social networks that matter: Twitter under the microscope, First Monday 14 (1) (2009).
[9] G. Jolly, Explicit estimates from capture–recapture data with both death and immigration stochastic model, Biometrika 52 (1965) 225–247.
[10] N. Kumar, B. Carterette, Time-based feedback and query expansion for twitter search, in: Pavel Serdyukov, Pavel Braslavski, Sergei O. Kuznetsov, Jaap Kamps, Stefan Rüger (Eds.), Proceedings of the 35th European Conference on Advances in Information Retrieval (ECIR'13), Springer-Verlag, Berlin, Heidelberg, 2013, pp. 734–737.
[11] V. Lavrenko, W.B. Croft, Relevance based language models, in: Proceedings of the 24th ACM SIGIR, 2001, pp. 120–127.
[12] C. Lee, H. Kwak, H. Park, S. Moon, Finding influential based on the temporal order of information adoption in twitter, in: Proceedings of the WWW 2010, 2010.
[13] L.Li, X. Chen, G. Xu, Suggestions for fresh search queries by mining mircoblog topics, behavior and social computing, Lect. Notes Comput Sci. 8178 (2013) 214–222.
[14] K. Massoudi, Manos Tsagkias, Maarten de Rijke, Wouter Weerkamp, Incorporating query expansion and quality indicators in searching microblog posts, Adv. Inf. Retr. 6611 (2011) 362–367.
[15] G. Mishne, J. Dalton, Z. Li, A. Sharma, J. Lin, Fast data in the era of big data: Twitter's real-time related query suggestion architecture, in: Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data (SIGMOD '13), ACM, New York, NY, USA, 2013, pp. 1147–1158.
[16] N. Naveed, T. Gottron, J. Kunegis, A. Alhadi, Searching microblogs: coping with sparsity and document quality, in: Proceedings of the 20th ACM International Conference on Information and Knowledge, Management, ACM, 2011, pp. 183–188.
[17] I. Ounis, C. Macdonald, J. Lin, I. Soboroff, Overview of the TREC 2011 Microblog Track, In TREC, 2012.
[18] H.S. Packer, S. Samangooei, J.S. Hare, N. Gibbins, P.H. Lewis, Event detection using Twitter and structured semantic query expansion, in: Proceedings of the 1st International Workshop on Multimodal Crowd Sensing (CrowdSens '12), Sheraton, Maui Hawaii, 2012, pp. 7–14.
[19] K.H. Pollock, J.D. Nichols, C. Brownie, J.E. Hines, Statistical inference for capture–recapture experiments, Wildl. Monogr. 107 (1990) 3–97.
[20] D. Spina, E. Meij, M. de Rijke, A. Oghina, M. Thuong Bui, M. Breuss, Identifying entity aspects in microblog posts, in: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '12), ACM, New York, NY, USA, 2012, pp. 1089–1090.
[21] C. Schwarz, W. Stobo, Estimating temporary migration using the robust design, Biometrics 53 (1997) 178–194.
[22] K. Tao, Fabian Abel, Claudia Hau, Geert-Jan Houben, Twinder: a search engine for Twitter streams, in: Proceedings of the 12th International Conference on Web Engineering (ICWE 2012), Berlin, Germany, 2012, pp. 153–168.



Ioannis E. Anagnostopoulos was born in Athens, Greece in 1975. He received his diploma from the Department of Electrical Engineering and Computer Technology, University of Patras, Greece, in 1998 and his Ph.D. from the School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), Greece, in 2004. From 2004 until 2010 he was with the University of the Aegean, Department of Information and Communication Systems Engineering, serving as Lecturer. Currently, he is with the University of Thessaly at the Department of Computer Science and Bioinformatics, serving as an Assistant Professor. His research interests include Internet and Web Technologies, Search and Retrieval Software Methodologies, E-Commerce, Telecommunication Networks and Intelligent Information Systems. He has more than 100 papers published in International journals and conference proceedings.



Gerasimos Razis was born in Athens, Greece, in 1987. He received his B.Sc. degree in Computer Science, specialized in Network and Computing Systems, from the Department of Informatics, University of Piraeus, Greece, in 2011 and his M.Sc. degree in Advanced Computer Science, specialized in Advanced Web Technologies, from the School of Computer Science, University of Manchester, UK, in 2012. From 2008 until 2010 he was with NeuroSoft S.A., serving as Software Engineer. Currently he is a Ph.D. candidate in the Department of Computer Science and Biomedical Informatics, University of Thessaly, Greece. His research interests include Web Information Management, Information Retrieval and Advanced Web Technologies.

Phivos Mylonas was born in Athens in 1978. He received his Diploma in Electrical and Computer Engineering from the National Technical University of Athens (NTUA) in 2001, his Master of Science (M.Sc.) in Advanced Information Systems from the National & Kapodestrian University of Athens (UoA) in 2003 and earned his Ph.D. degree (with distinction) at the former University (NTUA) in 2008. He is currently an Assistant Professor by the Department of Informatics of the Ionian University, Greece and collaborates as a Senior Researcher with the Image, Video and Multimedia Laboratory, School of Electrical and Computer Engineering, Department of Computer Science of the National Technical University of Athens, Greece. His research interests include content-based information retrieval, visual context representation and analysis, knowledge-assisted multimedia analysis, issues related to multimedia personalization, user adaptation, user modeling and profiling. He has published articles in 29 international journals and book chapters, he is the author of 55 papers in international conferences and workshops, he has edited 9 books and is a guest editor of 6 international journals, he is a reviewer for 20 international journals and has been actively involved in the organization of 42 international conferences and workshops. He is a member of the Technical Chamber of Greece since 2001, a member of the Hellenic Association of Mechanical & Electrical Engineers since 2002 and a member of W3C since 2009, whereas he is a past member of IEEE (1999–2010) and ACM (2001–2010).

Christos-Nikolaos Anagnostopoulos was born in Athens, Greece in 1975. He received his Mechanical Engineering Diploma from the National Technical University of Athens (NTUA) in 1998, and the Ph.D. degree from the Electrical and Computer Engineering Department, NTUA in 2002. From 2008, he serves the University of the Aegean as Assistant Professor in the Cultural Technology and Communication Department. He is a member of the Greek Chamber of Engineers and member of IEEE. His research interests include image processing, computer vision, neural networks and artificial intelligence. He has published more than 120 papers in journals and conferences, in the above subjects as well as other related fields in informatics.