ORIGINAL RESEARCH

CrossMark

Discovery and tracking of temporal topics of interest based on belief-function and aging theories

Mondher Sendi[1] · Mohamed Nazih Omri[1] · Mourad Abed[2]

Abstract

Content-based microblogging networks like Twitter have a significant and fast evolution of the information amount. In addition to textual content and member activities, analyzing structural information concerning user neighborhood and communities is an opportunity for several applications such as user interest modeling. In this context, this work proposes a new approach for temporal topics of interest derivation and tracking in the Twitter micro-blogging network. The challenge of this approach is to take into consideration the different dimensions of the user's interest such as personal, social and temporal aspect. We propose a new definition of an ego network that incorporates friendship reliability and dependency between members. The belief function theory that allows combining evidence from dependent or independent sources in an uncertain context is introduced. This theory has been widely utilized in various decision-making systems. To model the topics of interest evolution, we introduce the aging theory suggested mainly for event life-cycle detection and modeling in a micro-blogging network. We present an experimental study performed on the actual Twitter users, validating our approach. This model is compared to other methods using the same database.

Keywords Belief function · Aging theory · Topics · Interest discovery · Temporal topics · Social medias mining · Social filtering

1 Introduction

Micro-blogging networks actually, like Twitter, Weibo, and Tumblr, are the most popular social platforms for interaction and discussion in real-time between members. Social activities proposed in these platforms (e.g. like, dislike, replies, …) enhance greatly the interactions between members. Furthermore, social services such as message sharing, following other members and entering in new communities form the booster elements of the micro-blogging network dynamics. Due to social services and activities, interests, behaviors and opinions are formed in important nodes and gradually spread in the network from one user to another through interactions. This phenomenon is linked to several other facts such as influence between users, dependency, and friendship degree. Subsequently, with huge reachable and diversified information, online the social medias constitute a major opportunity for capturing and analyzing user profiles.

Discovering users' trends in social media and their evolution has attracted much research attention in recent years. It represents the fundamental step for important applications like recommendations (Dai et al. 2018), personalizations (Maleki-Dizaji et al. 2014; Karidi et al. 2017), marketing spreading analysis (Rechy-Ramirez et al. 2017) and user's activitie prediction (Boukhechba et al. 2017). In fact, discovering emerging topics of interest in the network has been treated lately with different technics. The incremental classification (Boschee et al. 2013), the LDA model (Latent Dirichlet Allocation) with matrix factorization (Wang et al. 2012), the pLSA model (Probabilistic Latent Semantic Analysis) (Chen et al. 2015) and others tools have been introduced to discover popular and trending topics over sharing publications. These works are particularly interested in the

✉ Mondher Sendi
sendi.mondher@gmail.com

Mohamed Nazih Omri
mohamednazih.omri@fsm.rnu.tn

Mourad Abed
mourad.abed@univ-valenciennes.fr

[1] MARS Research Lab LR 17ES05, University of Sousse, Sousse, Tunisia

[2] LAMIH, UMR CNRS 8201, University of Valenciennes and Hainaut-Cambresis, Valenciennes, France

global information in the network. In most cases, popular topics do not concern all profiles and communities and do not allow recognizing the real image of a network. Subsequently, these approaches are not sufficient for discovering and tracking specific and personal topics of interest in the network.

Combining structural and textual information for discovering topics of interest has been the subject of several recent works. Models based on the principle of homophily such as ego-centric networks (On-At et al. 2016), user communities (Mislove et al. 2010) and user neighborhood (Wen and Lin 2010), have been used to infer topics of interest. These approaches differentiate between personal and social user interests (On-At et al. 2016). Personal interest is the topic related only to a given user and can be discovered from their publications and activities. These interests are formed by some users in the network but can be transmitted to neighborhood users with the information dissemination phenomenon in the social media (TalebiFard and Leung 2014). An inferred topic of interest from the neighborhood is called a social topic of interest. In fact, homophily in the social network approves that members tend to create relationships with similar members in terms of age, interest and location, …. Utilizing this principle, we can infer topics of interest of the least active users. Thereby, we can exploit the user's locations in the network to explain the formation of users' interests.

Personal and social topics of interest have a temporal aspect in the microblogging networks. Indeed, nothing is stable in these platforms. Users daily publishe messages, interact with each other and create new relations according to their tastes. Some works have been interested in detecting and modeling temporal and dynamic elements like temporal-community detection (Pietilänen and Diot 2012), trend identification (Mathioudakis and Koudas 2010), event extraction and lifecycle modeling (Chen et al. 2003; Ferrari et al. 2014). Topics in the microblogging network emerge in users publications. Their popularity increases and finally, they disappear. Recent approaches have taken into account the temporal aspect of topics of interest (Cheng et al. 2008; Li et al. 2014; Abel et al. 2011). In most of these approaches, time is divided into fixed intervals. At the end of each time interval, an update is made to determine the remaining topics, the ones that disappear and the new emerging one. Users' history is analyzed to discover the last emerging topics of interest. Different decay functions have been introduced to follow the evolution and regression of the interest in topics.

Despite these proposals, other problems related to temporal interest tracking in the microblogging network have not been addressed. Indeed, with the significant and fast evolution of the amount of information (350,000 tweets per minute[1]), centralized algorithms become non-sufficient for analyzing large-scale microblogging networks and streamed content. Thus, the random feature of the appearance and disappearance of topics remains a major difficulty. We propose in this work an approach for the discovery of temporal topics of interest and their tracking over time. The two main contributions of our approach are the following:

– Proposal of a model to infer topics of interest from ego-networks.
– Proposal of a model to tracking the temporal evolution of topics of interest.

In the first contribution, we are using the ego-network to model links between each user and his neighbors. Several factors will have to be taken into account like the reliability of the links in an ego-network, the dependency and the similarity between users. The social topics of interest of each user are inferred from his personal network. For this task, we introduce the belief function theory to combine the evidence from different neighbors. Unlike the previous approaches based on belief function, the combination rule that we use in this approach takes into account the dependency between sources. It allows identifying the inter-dependence (dependence among features of a system) and the outer-dependence (dependence among the evidence sources). This strategy minimizes the redundancy when we combine evidence.

The second contribution consists of modeling the topics of interest evolution. For this task, we introduce the aging theory that allows for tracking the event or topic evolution in social media. Most of the previous approaches use the decay function to model the temporal topics of interest evolution. Unlike decay function, the aging theory allows modeling the different steps in the life cycle of topics of interest. Thus, it allows measuring the importance of the topics of interest at each time interval by taking into account the latest information in the network.

The rest of this paper is organized as follows: We survey the related works in Sect. 2 and we present the theoretical framework in Sect. 3. Section 4 presents the different steps and the architecture of the proposed approach. We present the experimental results and interpretations of our model in Sect. 5. Section 6 concludes our work.

---

[1] http://www.internetlivestats.com/twitter-statistics/

2 Related works

Our approach combines the derivation of users' interests from personal networks and the tracking of their evolution. These issues have been addressed in several works. We present in this section the most ones similar to our work.

2.1 Inferring users' interests

The derivation of users' interests from personal networks or user communities is mainly based on the homophily concept. The approaches for inferring users' interests assumed that the users' interests are strongly related to the interests of their neighbors. Several techniques have been introduced to this task. In the work (Zheleva and Getoor 2009), topics of interest were inferred from the users' community. Each hidden interest, called a sensitive interest, had a set of possible values from the interest of the community. The probability of each topic was determined based on the number of users in the community interested in this topic. In the approach proposed by Mislove et al. (2010), users' interests were inferred from global and local communities. Each interest had an affinity value in the community determined by the number of links sharing this interest. Topics that had important affinities were more susceptible to be the interests of the user.

Wen and Lin (2010) combined the LDA model with matrix factorization to derive users' interests. Two matrixes were utilized: The first one represented the relationships between members and topics, and the second one represented the influence between users. The distance in the network and the strength of communications between each user and other members were utilized to select the most important profiles for a given user. The work (Bhattacharya et al. 2014) was based on the idea that topics of interest of each user were driven by the interests of experts followed by this user. It discovered experts' interests from the list specified by users (each user could specify an interest list for experts) and then inferred users interests.

Other approaches are interested in identifying interests using social filtering technique (De Maio et al. 2018; Vinagre et al. 2015). In the approach (Li et al. 2013), the list of user interests is identified from the analysis of historical social application usage records. This list is then enriched by the interests of the user's neighbors. The approach proposed by Zarrinkalam et al. (2016) used a graph of emerging topics and a graph of followership to infer the topics of interest. In the approach sugested by Tchuente et al. (2013), each user profile had a personal and social dimension. Topics of interest of the second dimension were inferred from communities around the user. Community interests were determined by the combination of the interests of its members. The social interest of a given users was determined utilizing the combined interest of communities. Wang et al. (2014) put forward an approach to infer users' interests from neighbors. He represented a social network using a bipartite graph that contained hubs and authority nodes. The weight of each link in the network represented the strength of relationships and the influence between both ends. The similarity between nodes from the same class was taken into account in the interest inference process.

The work of Han et al. (2015) gave an experimental study that showed the homophile of interest in the online social networks. Thus, he proposed an approach to infer users interests from neighbors' based on features like geographic distance, friend similarity, and interest entropy. Cosine similarity was introduced to determine the affinity of interests between two users.

2.2 Modelling of temporal interests

Approaches for temporal interest discovery consist in detecting the emergence of new topics, tracking their evolution and identifying the disappearance moments of these topics. Different techniques have been introduced for the topics' emergence detection. Contrariwise, most of these approaches have used the decay function to track the evolution of interests.

The method suggested by Cheng et al. (2008) was based on two assumptions: (1) human interests would decrease over time, and (2) forgetting speed would slow down gradually and the accumulated interests would become more stable. Two classes have been used to represent temporal interests: long-term and short-term. The forgetting function was introduced to track the interest evolution. The same principle is used by Li et al. (2014). He put forward an approach for discovering the temporal interest based on a time-sensitive interest weighting scheme. The LDA model was introduced to discover the topics of interest. Then the profile history analysis permitted distinguishing between durable and temporary interests. The approach (Ahmed et al. 2011) also used the LDA model to discover temporal topics of interest. Long-term interests were generated from the history of users, and short-term interests were generated based on the last user publications. The decay interest function was utilized to compute the interest lifetime. The time-varying hierarchical interest model derived by content popularity enabled discovering the evolution over time.

Abel et al. (2011) gave an experimental study of content adoption and the evolution on their popularity in twitter microblogging. The lifetime of users interest depends on the timestamp of content. The interest decay function was determined according to the temporal distance between the

topic occurrence time and the given timestamp. Bao et al. (2013) proposed to model temporal topics for predicting users future interests in microblogging networks. The matrix factorization technique was introduced to learn users and topics features. The fusion of social structure and the history of users' interests allow controlling the interest propagation between nodes and then predicting the future topics of interest.

Time window technique is also used for modeling temporal evolution of generated content in microblogging networks (De Maio et al. 2018; Vinagre et al. 2015). In the approach suggested by Budak et al. (2014), the time was divided into the fixed intervals. The probability that the user was interested in a given topic of interest at any time period is determined utilizing the probabilistic model. Previous interests, user activities and neighbors' influences were taken into consideration in each time step. On-At et al. (2016) used the neighbor's in the egocentric network to derive users' interests. Each derived interest had a temporal weight that would determine its lifetime. This value was computed utilizing word frequencies and the link strength between the user and the source of interest (neighbors).

The temporal approach we propose in this work is different from the previous works in two principal points:

– The treatment of the temporal aspect
– The methodology for discovering users' interests.

For the treatment of the temporal aspect of topics' of interest, previous approaches have used two principle ways. The first one is the division of time into the fixed interval. The interest in each time interval is determined utilizing the previous interests and accumulated content up to the current state. At the end of each interval, the interest list is updated. Contrariwise, the emergence of the topic can be triggered at any time. Thus, the topic can disappear at any time. The system must always be able to identify new topics and update the list of user interests. The second way to model the temporal aspect is the use of the decay function which computes the lifetime of topics. This method does not take into consideration the nutrition that feeds the existence of a topic. In our approach, the temporal aspect is modeled utilizing the aging theory. This method takes into account the impulsive appearance of interests, nutrition and decay factors.

Some previous approaches of topics of interest discovery do not consider the structural information in the social network. In fact, users' topics of interest are not only in personal publications. A user can be influenced by the topics created or adopted by his neighbors. The location of a given node in the social graph can contribute to the decision of its interests. Other approaches exploited this structural information to infer topics' of interest from neighbors, user communities, experts profiles, …. Nevertheless, these approaches did not take much importance on the modeling of dependencies and the links reliability between users. In this approach, we propose to model these elements through egocentric networks. We exploit the activities of social media members to determine the dependencies and the links reliability between the members of these networks.

3 Theoretical framework

The belief function theory proposed by Shafer (1976) was a formal framework for computation and reasoning with partial (uncertain, imprecise) information. The belief theory permitted the fusion of information from different dependent or non-dependent sources and provided the tools needed for final decision making. Uncertain and imprecise information was modeled with mass functions that were subsequently merged using a combination rule. We present in this paper the basic elements of the Dempster theory. For more details, the reader can take a look at the following works (Dubois and Yager 1992; Cattaneo 2003; Destercke and Dubois 2011; Shafer 1976; Hajlaoui et al. 2017b).

A frame of discernment $\Omega = \{w_1, ..., w_n\}$ is a finite domain containing all possible events for a given problem. Any assumptions given by an agent must necessarily belong to the set $2^{\Omega} = \{A_i \mid A_i \subseteq \Omega\}$ containing all possible combinations of the elements of Ω. A hypothesis A_i given by an agent v_j is accompanied by its degree of certainty in the form of a basic belief assignment $m_j(A_i)$. The belief basic assignment (BBA) is a mapping from 2^{Ω} into [0,1] satisfying the following condition in (Eq. 1):

$$\sum_{A_i \subseteq 2^{\Omega}} m_j(A_i) = 1 \tag{1}$$

when $m_j(A_i) > 0$, the subset A_i is called a focal set of m_j. A belief function is called a Bayesian BBA if all the focal elements are singletons. In this case, it is similar to a probabilistic distribution over the frame of discernment Ω. The BBA is said to be consonant if the focal elements are nested and we can then determine the measure of possibility and necessity for this BBA. Other facets of the same information enable making other interpretations, such as belief (Eq. 2), plausibility (Eq. 3), and pignistic probability (Eq. 4). Details of other facets can be found in Denœux (2008). The three facets utilized in this work are the following:

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \tag{2}$$

$$pl(A) = \sum_{\emptyset \neq B \cap A} m(B) \tag{3}$$

$$BetP(A) = \sum_{\emptyset \neq B \subseteq 2^{\Omega}} \frac{|A \cap B|}{|B|} * \frac{m(B)}{1 - m(\emptyset)} \tag{4}$$

Table 1 Principal notations

Variable	Description
Fr (u, v)	Friendship degree for directed link (u,v)
f_{k_i, P_u}	Frequency of keyword k_i in publications of user u
IPN(u)	Influential personal network of user u
δ_u^j	Dependency degree of user v_i in personal network of user u
$m_j(A_i)$	Belief assignment for hypothesis A_i from source j
Fr (u,v)	Friendship degree between user u and v
$An(a_i, a_j)$	Number of times that a_i annotate a_j in their publications
$Rt(a_i, a_j)$	Number of times that a_i retweet publications of user a_j
$Rp(a_i, a_j)$	Number of times that a_i reply to publications of user a_j
$N_{(t_a)}(z_i)$	Nutrition of topic z_i at instant t_a
$nE(z_i)$	New energy for topic z_i
$rE(z_i)$	Remaining energy for topic z_i
β	Decay nutrition factor
D_u	Dependency matrix between members in personal network of user u
M_t	Matrix contains relations between the keywords and the user's

where:

- bel(A) is the degree of belief in A,
- pl (A) is the plausibility measure and the upper bound of the belief degree,
- BetP (A) is the pignistic probability and the compromise between the two previous facets; bel(A)< BetP(A)< pl(A).

When the focal elements of *m* are nested, *pl* represents the distribution of possibility and *bel* represents the necessity on the set $\{A_k \mid m_j(A_k) > 0\}$. Generally, a possibility distribution over a finite set can be seen as a consonant BBA.

The combination of different BBAs is the main step in the belief function theory. It allows synthesizing the received information in order to make a more reliable decision. The choice of the combination rules depends on certain initial assumptions such as the independence of the sources of information, the management of conflicts between sources and the reliability of sources. The first proposed rule is the Dempster-Shafer rule (Shafer 1976) used for the conjunction or disjunction operator (Eq. 5):

$$m'(A) = \oplus m_{i=1,\dots,n}(A) = \frac{1}{1-k} \sum_{X_1 \cap \dots \cap X_n = A} \left(\prod_{i=1}^n m_i(X_i) \right) \tag{5}$$

where $A, X_i \in 2^a$ and $K = \sum_{X_1 \cap \dots \cap X_n = A} (\prod_{i=1}^n m_i(X_i))$ represent the conflict degree between sources. It can be defined as the contradiction among the mass functions of several agents for a given hypothesis. If the sources are in strong conflict (K is large), the Dempster rule can lead to erroneous results (Dubois and Yager 1992). There are different sources of conflict such as the unreliability of sources or observation

of different phenomena. The discounting of the BBA given by the source is suggested to solve the problem. Assuming that α_j is the degree of reliability and the discount rate for a source v_i, a BBA can be transformed into another BBA prior to the merger (Eq. 6) as follows:

$$\begin{cases} m_\alpha^j(A) = \alpha_j \, m_j(A), & \forall A \in 2^\Omega \\ m_\alpha^j(\Omega) = (1 - \alpha_j) \, (1 - m_j(\Omega)) \end{cases} \tag{6}$$

Dependency among sources is an important factor that cannot be ignored when merging information. The Dempster combination rule supposes the independence between sources of information but this is not the case in practice. Other rules are proposed to remedy this problem. In most of these methods, dependence is considered by modifying the combination rule or by changing the belief function. The method put forward by Cattaneo (2003) favored the maximum consistency, instead of independence, to minimize the conflict between the sources. The combination rule suggested by Destercke and Dubois (2011) adapted the minimal operator proposed in the theory of possibility for combining different BBAs. The work (Basir et al. 2005) opted for treating dependence by learning to minimize conflicts between sources. Moreover, Su et al. (2015) proposed a method for combining belief functions by identifying the inter-dependence(dependence among features of a system) and the outer-dependence (dependence among the evidence sources) (Table 1).

Fig. 1 Influential personal network construction: Ego network contains three types of links between the ego node and the alters node, (1) reciprocal links, (2) non-reciprocal links from ego to alters, and (3) non-reciprocal links from alters to ego. In the Influential personal network, the third type links are eliminated (from alters to ego)

Ego network   Influential personal network
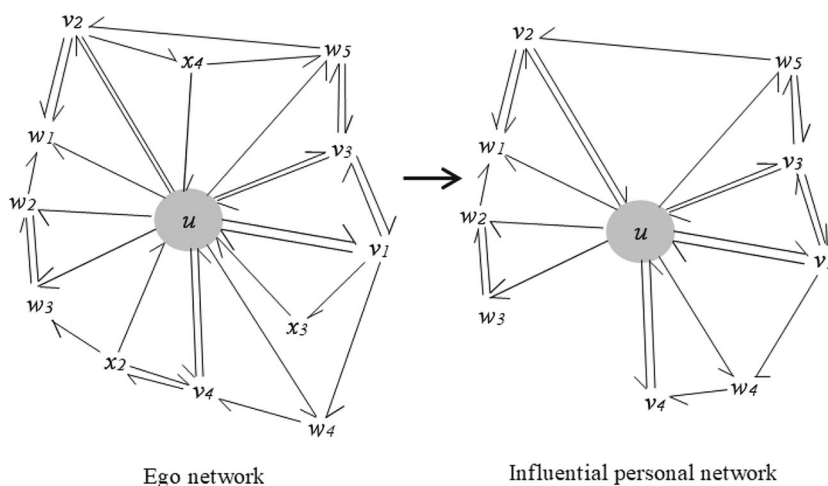
4 Motivation and problem formulation

4.1 Content-based microblogging networks

Users' content on microblogging networks is diversified. In fact, members in the virtual social space interact with each other with several services. The principal communication way is the sharing of publications in the personal network. In most cases, users integrate hashtags on the shared message to clarify and summarize the subject of publication. This activity has a great success in social media (75% of people on social media use hashtags[2]). Actually, each emerging event and topic in microblogging networks is represented by hashtags. Furthermore, users on social media interacts also utilizing annotation. They have the possibility to join the identifiers of other users in the publication, say that a message interests such a person. If a user is interested in a message shared by one of their following, he reacts utilizing retweet (share) or reply and gives their opinion on the subject of publication.

The structural information of social network is used in our approach. The structural indices (e.g. centrality of nodes) are used in several approaches for social network analysis and content recommendation on social media. In this work we are interested more particularly in the number of common followings and the number of triadic closures. These two pieces of information will be used to determine the friendship degree among users.

The first objective of our approach is the discovery of topics of interest in the microblogging network. Personal topics of interest will be discovered using single-pass classification. For social topics of interest, we exploit ego networks to infer the relevant interests for ego user. We utilize in this step the belief function theory. The second objective is to track the

evolution of users' interests towards the interests change in their ego network. We introduce the aging theory for this task. Indeed, the belief and aging theories well adapt to the problematic of our work. On the one hand, the belief theory allows combining evidence given by several sources and takes into account the uncertainty of the provided information. We combine in our case the beliefs given by members of the ego network. On the other hand, the aging theory is among the few models proposed for the modeling of information life cycle. It has been used beforehand for modeling the life cycles of events and topics.

4.2 Problem formulation

In this section, we introduce the formulation of the temporal topics of interest discovery problem with definitions and formal concepts.

4.2.1 Influential personal network

Social media like Twitter and Tumblr contain a social network where links between users are directed. It can be modeled with a directed graph G = (U, E) where U is the users set and E the set of directed links (Fig. 2). In our model, the directed network is composed of a set of directed and connected ego-networks. Each ego network has an ego node (focal node) and nodes connected directly to an ego called "alter" (On-At et al. 2016; Mcauley and Leskovec 2014). Each node in the network has his/her own ego network, and all ego networks interlock to form the social network.We invent the concept of Influential Personal Network (IPN) extracted from the ego network (Fig. 1). It only contains nodes that have the ability to influence the ego node. For a given ego node u, we define the sets of nodes that constitute the IPN of u.

Fig. 2 Overview of our approach : We combined the structural information (in the IPN), the textual information (user publications) and the users' activities to construct the belief function model
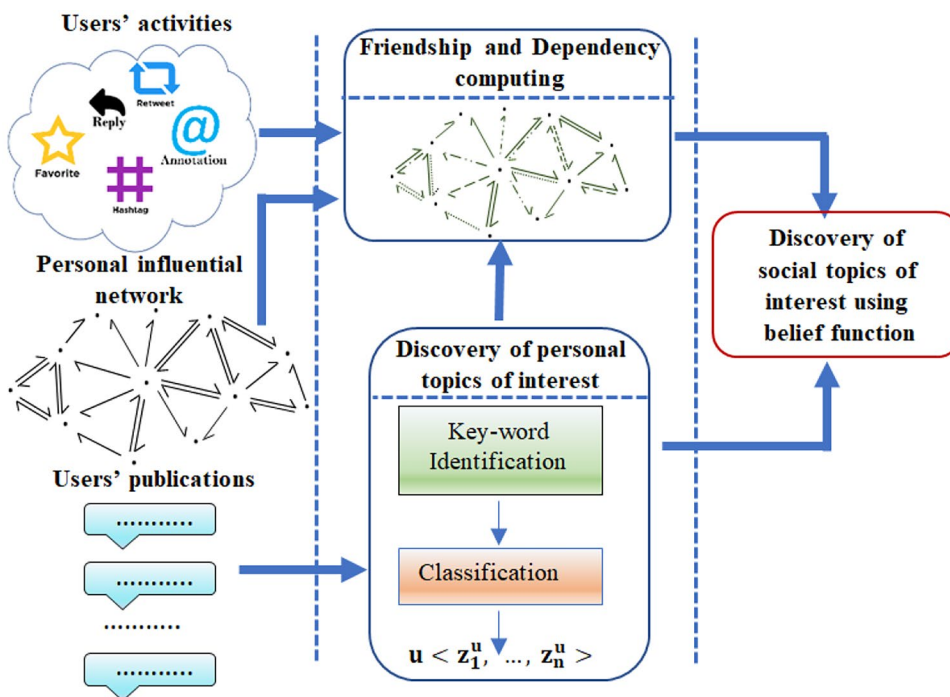
Definition 1 The IPN for a user u is IPN (u) = $(W_u \cup V_u$, A) = (U \ $X_u \cup Y_u$, A) where :

- $V_u = \{v_i | v_i \in U, (u, v_i) \in E \ and \ (v_i, u) \in E\}$ users that have reciprocal links with user u.
- $W_u = \{w_i | w_i \in U, (u, w_i) \in E \ and \ (w_i, u) \notin E\}$ users followed by the user u.
- $X_u = \{x_i | x_i \in U, (u, x_i) \notin E \ and \ (x_i, u) \in E\}$ containing the user that follows u.
- $Y_u = \{y_i | y_i \notin U, (u, y_i) \notin E \ and \ (y_i, u) \notin E\}$ users not linked to u.
- A is the set of directed links between the users in $W_u \cup V_u$.

This new network contains only the nodes that are likely to influence the focal node: nodes representing users followed by u. We assume that W_u is the most significant set for the user u. Indeed, users can create several types of links, essentially to sources of information or to personal friends (Aiello et al. 2014). The links between personal friends are mostly reciprocal but do not necessarily imply that their interests are close (V_u). Another type of links in the Twitter microblogging is the follow back: The same users follow back to their followers as an act of courtesy (Takemura et al. 2015; Imamori and Tajima 2016). Links to information sources are more relevant than links to personal friends. In fact, if a user follows an information source, the publications of this source will influence the topics of interest of the user. We will use this assumption to weight links according to their types in the IPN of each user (Sect. 4.2.2). The nodes X_u cannot involve directly in the inference of topics

of interest of u but they can be involved through their followers. If a user v is followed by u and follows a user x, he can be influenced by the interests of x, and subsequently his interests can be transmitted to u.

4.2.2 Topic of interest

Active users in microblogging networks publish tweets by inventing certain topics. Thus, they read and react with publications of other users in their ego networks. When reacting with the publications of its neighbors, a user may be interested in existing topics either with agreement or disagreement. A topic of interest can be defined by a set of keywords over users' publications. We propose the following common definition for a personal and social topic of interest.

Definition 2 A topic of interest is a set of trending keywords related to a specific subject $z = \{k_1, \ldots, k_n\}$ where each k_i is a keyword or tag. The keywords are identified based on their TF-IDF degree. A topic can be the interest of one or more users in the microblogging network. It can be a personal or social interest.

Users in the microblogging network can be represented by their personal topics of interest extracted from their publications $Z_u = < z_u^1, \ldots, z_u^k >$. This method is limited for two reasons. Firstly, users who do not have publications in their timelines do not have topics of interest and we have no information about them. Secondly, it is difficult to follow people who rarely share publications in their timelines. Inferring social topics of

$$M_t = \begin{array}{c} \\ p_1 \\ \vdots \\ p_i \\ \vdots \\ p_n \end{array} \overset{\displaystyle k_1 \cdots\cdots k_j \cdots\cdots k_m}{\left(\begin{array}{cccccc} 1 & \vdots & \vdots & \vdots & \vdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \vdots & \vdots & \vdots & \vdots & 1 \end{array} \right)}$$

Fig. 3 The matrix M_t contains the relation between the keywords and the user's publications. Starting with k_1 word, the first topic contains only this word, say T_1. At this point, T_1 contains only one item, k_1, so the centroid of k_1 is simply the vector for k_1 ; $T_1 = <1, 0, 1, 0>$. For $1 < i < m$, if $SIM(k_i, T_j) >$threshold and $k_2 \in T_j$, then the centroid of T_j is the average vector for T_1 and T_2. Else construct the new topic $T_s = k_i$

Fig. 4 Triadic closure and common following

a. Non-symmetric reliability

b. Triadic closure

c. Common following
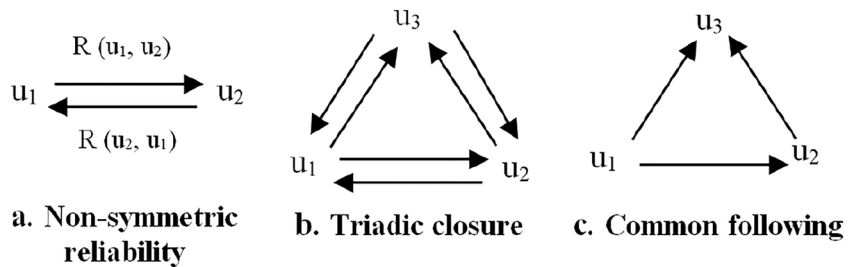
interest from the IPN is the solution that we propose in this work to remedy the two previous difficulties. Let u be user where $Z_u = <z_u^1, \ldots, z_u^k>$ and $Z_a = <z_a^1, \ldots, z_a^k>$ are respectively the personal topics of interest of user u and member $a \in W_u \cup V_u$. Given a list of the topics of interest of personal network members, we will be able to infer the social interests of u and track their evolution.

5 Proposed approach

5.1 Model Preparation

5.1.1 Identification and classification of keywords

Based on the timeline of each user, we rank the most utilized words according to their frequency. When a new word appears with a high frequency, the classifier assigns it to a class (topic) or creates a new class. A word is recognized as important if its TF-IDF (Eq. 7) is greater than a threshold as follows:

$$TF - IDF(k_i, u) = \frac{f_{k_i, P_u}}{max_{k_i}(f_{k_i, P_u})} * log\left(\frac{|P_u|}{|p_i \in P_u : k_i \in p_i|}\right) \tag{7}$$

where P_u is the set of publications shared by the user u and f_{k_j, P_u} is the frequency of the word k_i in P_u.

Based on the co-occurrence between words, the single-pass algorithm determines the new topics (Fig. 3). The matrix M_t contains the relationship between the identified words and the user publications. M_t (i,j) = 1 if the word k_j appears in a p_i publication, and 0 otherwise. After this classification, users who have shared publications must have a set of personal topics of interest that represents their profile before the inference process.

5.1.2 Link reliability

An IPN contains links among an ego node, their followers and followings. Not all links are important for a given user. Thus, the topics of interest can change over time and the links can persist, but they have no meaning (e.g. one user interested in the election will be interested in a politician profiles, but he becomes disinterested after a certain period). For this reason, we propose a link reliability measure based on friendship degree and semantic similarity. It allows determining whether a link between two users is important or not.

Friendship degree The degree of friendship between two users can be determined from the communication flow between them. In the work of Cheng et al. (2011), the friendship degree was determined by the number of messages sent between two users. In the work of Hopcroft et al. (2011), two users would be considered as friend if each one followed the other. In our approach, we put forward another definition for the degree of friendship utilizing the structural network features. More precisely, we introduce triadic closures and common followings. The number of triadic closures in a non-oriented social network is considered as the number of common friends between two nodes (Hajlaoui et al. 2017c). This measure is used in several approaches for the prediction of the new links in the network. The triadic closure can be introduced to weight links in the IPN. In fact, the three nodes forming a triadic closure have higher trust to each other. It gives the incentive to follow updates for each other. If the number of triadic closures increases between two users, we can be more certain that the relationship between them is important. Since the links in the microblogging network are oriented, we utilize for each reciprocal link $< u, v >$ two

measures representing the friendship weight of (u, v) and (v, u). We use the number of triadic closure to determine the strength of the relationship between them. If the user u only follows the user v, we count the common followings (Fig. 4) without checking whether there is a triadic or not. These followings are users followed at the same time by both users.

Definition 3 The friendship degree for a link (u, v) is important if and only if the number of triadic closures or the number of the common-following users is is greater than a given threshold. The friendship degree is computed as follows (Eq. 8):

$$
Fr(u, v) = \begin{cases} \frac{Tr(u,v)}{\sum_a Tr(u,a)} & \text{if } (u, v) \text{ is reciprocal.} \\ \\ \frac{|IPN(u) \cap IPN(v)|}{|IPN(u) \cup IPN(v)|} & \text{elsewhere.} \end{cases} \tag{8}
$$

Semantic similarity The second measure that we use to compute the link reliability is the semantic similarity between the topics of interest of two users. Each user has a list of interests in each time interval $u < z_u^1, \ldots, z_u^1 >$. In the first step of our approach, this list is initialized by the set of personal topics of interest extracted from the publications of each user. The similarity between two profiles is determined utilizing the dot product similarity (Eq. 9):

$$
\begin{cases} sim(u, v) = max_{z_u^i \in Z_u, z_v^j \in Z_v}(sim(z_u^i, z_v^j)) \\ sim(z_i, z_j) = \frac{\sum_{w_k \in z_i \cap z_j} w_k^2}{\sqrt{\sum_{w_s \in z_i}(w_s^2)} * \sqrt{\sum_{w_l \in z_i}(w_l^2)}} \end{cases} \tag{9}
$$

After computing the semantic similarity between two users, we compute the link reliability using the combination of two previous formulas (Eq. 10):

$$
R(u, v) = \lambda \quad Fr(u, v) + (1 - \lambda) \, Sim(u, v) \tag{10}
$$

where $\lambda \in]0, 1[$ is a control variable.

The function R is not symmetric and "R(u, v) ≠ R(v, u)". Hence, in some cases, one of two functions is null (when the relation $\}\} < u, v >''$ is not reciprocal).

5.1.3 Dependency in influential personal network

In addition to the structural information of the network, we utilize the activities of users to determine the dependency degree. More precisely, we use the retweets, replies, and annotation between users. The retweet is the activity of sharing tweets of other users. The replies are the comments writing to response for some publications and the annotations are the markings of one user by another in a publication or comment. Using these activities, we propose a definition for the dependency between users.

Definition 4 Two users in a microblogging network are dependent on each other if the number of retweets, replies and annotations among them are greater than a given threshold.

We utilize the method suggested by Su et al. (2015) integrating dependencies between sources in the combination rule (outer-dependence). The dependence in our case is not symmetric. If a user u_1 depends on another user u_2, it is not necessary that u_2 depends on user u_1. The dependency matrix D_u is determined utilizing the users' activities as follows (Eq. 11):

$$
D_u = \begin{bmatrix} d_{11} & \ldots & \ldots & d_{1n} \\ & \ldots\ldots\ldots\ldots & \\ & \ldots\ldots\ldots\ldots & \\ d_{n1} & \ldots & \ldots & d_{nn} \end{bmatrix} \tag{11}
$$
$$
d_{ij} = \frac{A(a_i, a_j)}{max_{a_s \in IPN(a_i)}(a_i, a_s)}
$$

where:

- $A(a_i, a_j) = Rt(a_i, a_j) + Rp(a_i, a_j) + An(a_i, a_j)$
- $a_i, a_j \in IPN(u)$
- $Rt(a_i, a_j)$ and $Rp(a_i, a_j)$ are the number of times that the user a_i retweets and replies to the publications of a_j.
- $An(a_i, a_j)$ is the number of time that the user a_i annotates the user a_j.

The total dependence degree (Eq. 11) for each user a_i in the personal network of u can be computed utilizing the matrix D as follows (Eq. 12):

$$
Td_i = \sum_{j=1, j \neq i}^{n} d_{ij} \tag{12}
$$

The weight of dependency (Eq. 13) of each user is determined as follows:

$$
\delta_u^i = \frac{\frac{1}{Td_i}}{\sum_j \frac{1}{Td_j}} \tag{13}
$$

5.2 Inferring topics of interest using belief theory

Each profile in the microblogging network is represented with a vector of topics of interest $Z_u = < z_u^1, \ldots, z_u^m >$ where each topic contains a list of keywords $z_u^i = < k_1, \ldots, k_s >$. The topics

Fig. 5 We use three combinations of belief functions: f_1 is for the fusion of beliefs that come from friends of u $(V = v_1, \ldots, v_s)$, f_2 for the fusion of beliefs that come from followers of u $(W = w_1, \ldots, w_r)$ and f_3 for the fusion of beliefs from the personal network of u $(W \cup V)$ and their own beliefs. The result is an ordered list of interesting topics for the user u
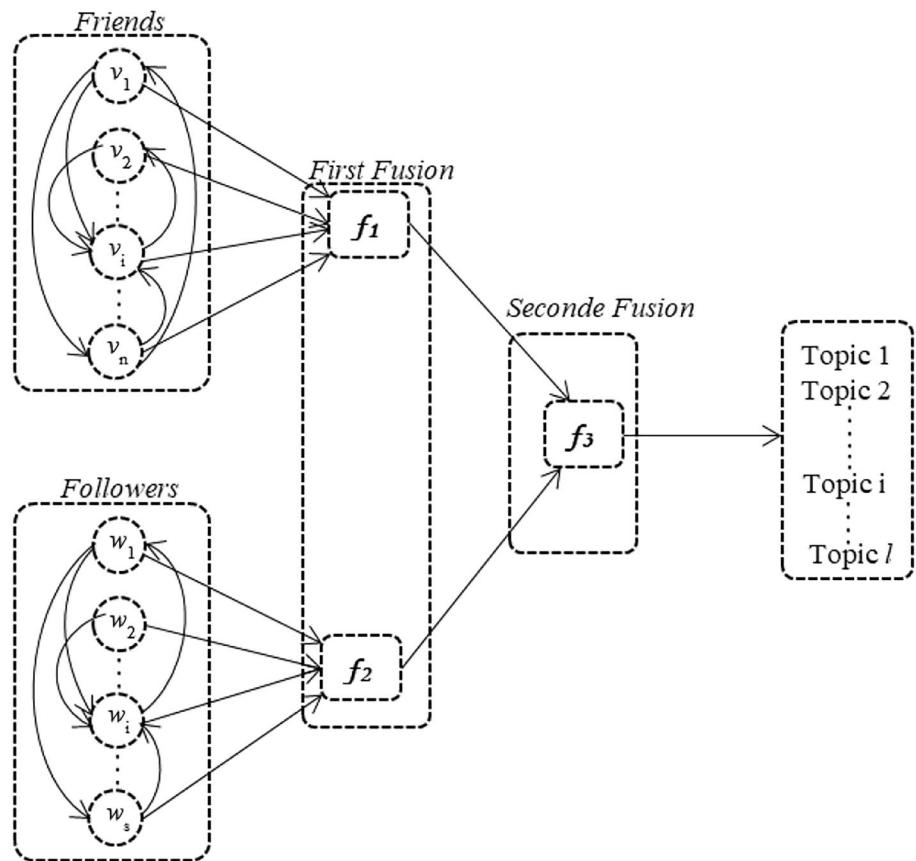
list can be empty $(|Z_u| = 0)$ if the user u did not share any publication. We use the belief function theory to infer the social topics of interest of each user. The frame of discernment in our problem is the set $\Omega_u = k_1, \ldots, k_n$ containing all the keywords identified in the *IPN(u)*. The set $2^{\Omega_u} = A_i | A_i \subseteq \Omega_u$ contains all possible combinations of the elements of Ω_u; all possible topics that may interest the members of the IPN(u). The belief of $h \in IPN(u)$ on its personal topic of interest z_i is computed from the importance of keywords composing this topic (Eq. 14):

$$m_h(z_i) = \frac{\sum_{k_j \in z_i} TF - IDF(k_j)}{|Z_j|} \qquad (14)$$

The combination of BBAs from the following and friends of the user u does not happen in the same way. We suppose that the followings are more reliable than friends. The conjunction operator is utilized with the Dempster rule for the combination of beliefs from friends and the disjunctive combination for the beliefs from followings (Eq. 15):

$$m_\theta(z_i) = \frac{1}{1 - k} \sum_{b_1 \Lambda \ldots \Lambda b_s = z_i} \left(\prod_{i=1}^{n} R(u, a_i) \cdot \delta_u^i \cdot m_i(b_i) \right) \quad (15)$$

where $R(u, a_i)$ is the reliability of the links (u, a_i) defined in Sect. 4.2.2, $b_i \in 2^{\Omega_u}$, Λ is the conjunction operator if $a_i \in V_u$ or the disjunction operator if $a_i \in W_u$ and $\theta \in \{V_u, W_u\}$.

The third fusion is the combination of the results found by (Eq. 15) and the BBAs given by u. We suppose that we have three sources of belief: w (beliefs from following), v (beliefs from friends) and u. For each topic z_k resulting from the first or second combination (f_1 or f_2, we use the conjunction operator to combine the beliefs as follows (Eq. 16):

$$m_c(z_k) = \frac{1}{1 - k} \cdot \sum_{b_1 \cup b_2 \cup b_3 = z_k} m_V(b_1) \cdot m_W(b_2) \cdot m_u(b_3) \quad (16)$$

After the fusion of the BBAs utilizing the three steps, we can make the decision to identify the most related topics of interest for a given user. We use the measure in (Eq. 2) defined for the belief degree. The list of topics is sorted from the most to the least "important" (Fig. 5).

Algorithm 1 Inferring social topics of interest from IPN.

Require:
  IPN (u).
Ensure:
  Z_u': list of social topics of interest.
begin
  $Z_v \leftarrow$ personal topics of friends of u.
  $Z_w \leftarrow$ personal topics of following of u.
  $Z \leftarrow \emptyset$.
  for each $z_i \in Z_v$ **do**
    $S_v(z_i) \leftarrow \{\{a_1 \cap \cdots \cap a_n\} \mid \{a_1 \cap \cdots \cap a_n\} = z_i, z_i \in Z_v\}$
    Compute $m_v(z_i)$ using eq. 15.
    if $m_v(z_i) > 0$ **then**
      Add z_i to Z.
    end if
  end for
  for each $z_j \in Z_w$ **do**
    $S_w(z_j) \leftarrow \{\{b_1 \cup \cdots \cup b_m\} \mid \{b_1 \cup \cdots \cup b_m\} = z_j, z_j \in Z_w\}$
    Compute $m_w(z_j)$ using eq. 15.
    if $m_w(z_j) > 0$ **then**
      Add z_j to Z.
    end if
  end for
  for each $z_k \in Z$ **do**
    $S(z_k) \leftarrow \{\{a \cup b\} \mid \{a \cup b\} = z_k, z_k \in Z\}$
    Compute $m_c(z_k)$ using eq. 16.
    Compute $bel(z_k)$ using eq. 2.
    if $bel(z_k) > 0$ **then**
      Add z_j to Z_u'.
    end if
  end for
  Return Z_u'.
end.

5.3 Tracking of topics of interest

Topics of interest can be evolving over time for several reasons. They may be due to the emergence of new events, the creation of new friendship relationships or following new information sources. Whenever the interests of a given member change, updating the interests of their friends is necessary to keep the correct image of the network.

The lifetime of a topic of interest begins with the emergence step when the keywords describing this topic appear. The popularity of the topic increases after its adoption by several users, and finally it disappears. Feeding a topic so that it can stay in the network consists in publishing tweets similar to it. The decrease in the number of publications related to the topic implies the decline in its popularity. This assumption has been used in previous works to model the life cycle of events (Chen et al. 2003) and subjects (Bao et al. 2015; Chen et al. 2007). Our modeling of the life cycle of topics of interest is inspired by these approaches. The aging theory can be utilized to model the temporal aspect of the interest in the microblogging network.

Like events, the importance of topics falls regressively in the network. To discover if a user is still interested in a given topic, we analyze the activities of the other users in their personal networks. If the users in the IPN(u) react again

about this topic, it is more likely that the user u is still interested in this topic. We define the new nutrition that permits estimating whether the topic $z_i = <k_1, \ldots, k_n>$ will remain in the user profile or no at each instant t_a as follows (Eq. 17):

$$\begin{cases} N_{ta}(z_i) = \dfrac{\sum_{1 < j < |z_i|} N_{ta}(k_j)}{|z_i|} \\ N_{ta}(k_j) = \dfrac{f(k_j)}{max_{t_h < t_a}(f(k_s))} \cdot e^{\left(\dfrac{f_{ipn}(k_j)}{max_{t_h < t_a} f_{ipn}(k_j)}\right)} \end{cases} \quad (17)$$

where:

– $f_{ipn}(k_j)$ is the frequency of the keyword k_j in the $IPN(u)$,
– $f(k_j)$ is the frequency of the keyword k_j in the network.

When the frequencies $f_{ipn}(k_j)$ and $f(k_j)$ go up, we recompute $N_{t_a}(k_j)$ to determine the new lifespan. If these frequencies remain unchanged in a time interval, we suppose that the lifespan of this topic is finished. The new energy $nE(z_i)$ for the topic of interest z_i is utilized to estimate the remaining topic life. The remaining energy $rE(z_i)$ for the topic of interest must be greater than a given constant β [decay nutrition factor (Chen et al. 2003)] in order to remain in the list of interests (Eq. 18):

$$\begin{cases} nE_{t_a}(z_i) = \dfrac{N_{t_a}(z_i)}{max_{t_h < t_a}(N_{t_a}(z_j))} \\ rE_{t_a}(z_i) = rE_{t_a - \Delta t}(z_i) + nE_{t_a}(z_i) - \beta \end{cases} \quad (18)$$

If $rE_{t_a}(z_i) \leq 0$, the topic of interest z_i disappears.

A topic of interest then has a lifespan that can be extended if other related publications on this subject appear. This model enables following the evolution of topics in the network. Indeed, whenever a topic appears in a node u, this node must inform other nodes representing the people it follows to update their lists of interests. When members of the IPN (u) receive the update of u, they determine the relevance of this new topic with their profile using the belief function model given in Sect. 3. The disappearance of a topic of interest from a given profile is determined utilizing the decay factor. If $rE_{t_a}(z_i) < \beta$, the topic must be deleted.

6 Experimentation

6.1 Data collection and experimentation process

6.1.1 Crawling data from twitter

The database that we use for the evaluation of our approach is collected from the Twitter site. Diversified data are available from this micro-blogging. In our model we have interested in users' activities (replies, retweets, and mentions), publications, and the social network
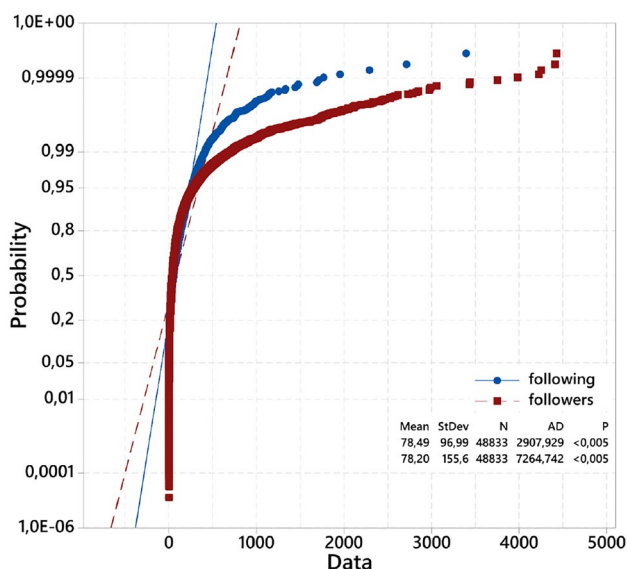
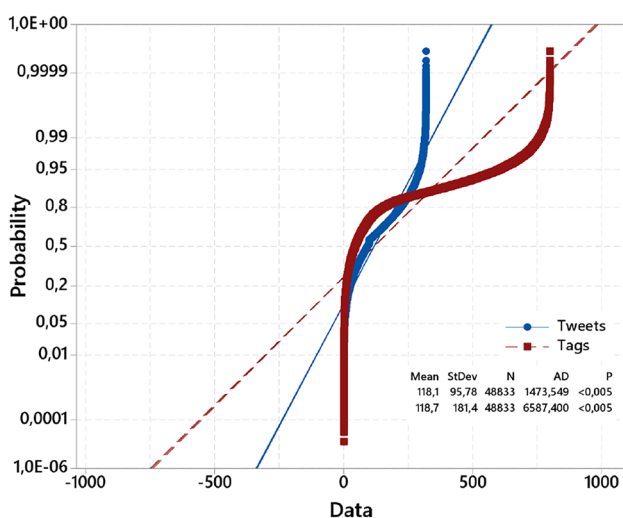**Fig. 6** Distribution of followers and followings

Fig. 7 Distribution of tweets and tags

containing directed links. We collect the data generated between January 1st, 2017 and April 31st 2017.

Over than 3 million tweets (Table 2) are collected in this period to extract and track the topics of interest. This database contains active users that publish and react frequently in the network and less active users. We crawl over than 48,000 profiles on Twitter including users' activities and directed links (followers and followings) to construct the ego network for each user. Users' activities, such as mentions, replies and retweets, are collected utilizing the Twitter query. For every two users linked on Twitter, we search the activities carried out each other. Figures 6 and 7

Table 2 Data base characteristics

Data type	Data size
Users	48,833
Links	1,193,254
Tweets	3,303,270
Key-words	8,405,867
Retweets	3,627,453
Mentions	751,384
Replies	962,463

illustrate the distribution of followers, followings, tweets, and tags. These graphics show that our data proportion is normal.

6.1.2 Preparation and procedure of experiments

The main preprocessing steps in the experimentation study are the following:

– StopWords elimination
– TF-IDF computing using Eq. 5 to identify keywords
– Keywords classification into topics for each profile

Tweets are separated into two sets. The first one contains tweets shared between January 1st, 2017 and February 15th, 2017 and the second one contains tweets shared between February 15th, 2017 and 30th March.

The first set is utilized to learn personal topics of interest for the members of the network. Topics are identified using the single-pass classification algorithm after identifying keywords with the TF-IDF measure. Users that have no keywords (subsequently have no personal topics of interest) in the first period represent 25% of the network. All profiles will be enriched utilizing our inferred topics of interest from the IPN.

The second set is utilized for the evaluation of discovered and tracked temporal topics of interest. When a keyword is captured, and if it already exists in one topic, the keyword frequency will be increased. Else, we compute the similarity between new keywords and topics. If none of them is close to it, we create a new topic contains only the last keyword. When list of personal topics of interest is modified for a given user, their neighbors must update their list of interests. This change may be due to the appearance of a new topic or the disappearance of another if its remaining energy becomes insufficient (verified with the aging theory). In this second period, 324,032 new topics are captured and the disappearance of 462,134 topics is registered. These topics of interest are used for tracking the evolution of the profiles in the network.

Table 3 Example of discovered topics

Order number	Topics	Cs
1	Named, beauty, scientists, discovered, experiment, oddly, particles, lhc	0.351
2	Government, executive, voting, bill, blame, majority, fbi, republican, proof, republican, democrats, nsa, cia, bully, cowardice, legislative	0.562
3	Leadership, protect, economics, pollution, environmentalism, disconnected	0.673
4	Heart, victims, families, london, terror, shake, strength, resilience, british, people, students, attack, attacks, die, hurt, school, westminster	0.544
5	Free, signed, disaster,,victims, workers, relief, putting, flood, generously, evacuees	0.362
6	Big, science, intelligence, risk, web, opening, finance, lead, front, cloud, data, technologies, developer, architect, analytics, scala, java	0.456
7	Interesting, tomorrow, visit, svvr, worlds, familys, lab, creator, technologies, greatly, blind, robots, drones	0.285
8	Order, ban, acting, general, attorney, called, illegal, muslim, fires, enforce, refuses, uncons	0.332
9	Social, super, control, signals, inflammation, bacteria, bacterial, spinal, supple, neuronal, cords, bendy, nanowires, viruses	0.347
10	Military, trump, cover, worried, north, parade, russians, korean, leakers, democracykilling, pagesamericas	0.473

6.1.3 Evaluation metrics

For evaluating the precision and recall of our approach, a recommendation system is used. For each user, relevant tweets are chosen (containing user keywords). All tweets are combined in a single data corpus to return relevant tweets for each user. This model allows recognizing the capacity of our approach to return relevant tweets for each user after learning user profiles with discovered topics of interest (Sendi et al. 2017). We utilize precision P@5, P@10, P@20 and P@30 respectively for the first 5, 10, 20 and 30 returned tweets. We use the F-measure that combines precision and recall to compare our approach with others.

$$P@X = \frac{\sum_{i=1}^{n} P_i @X}{n}, \quad P_i @X = \frac{\sum_{j=1}^{x} relevance(E_j^i)}{X} \quad (19)$$

$$R = \frac{\sum_{i=1}^{n} R_i}{n}, \quad R_i = \frac{|relevant\ elements\ returned|}{|relvant\ elements|} \quad (20)$$

$$F = \frac{2*(P*R)}{(P+R)}, \quad P = \frac{\sum P@X}{4} \quad (21)$$

where p_j^i is a publication returned by the recommendation system.

6.2 Results and discussion

6.2.1 Consistency of topics

Table 3 contains 10 topics discovered for different users in the three time slices. It shows that most topics containing less keywords are more coherent. For example, in the topics T_1, only words w_1 and w_2 do not have the same meaning as the other keywords. To highlight this consistency, we compute the internal consistency of topics of interest $Cs(z_i)$. This measure is based on the co-occurrence between keywords in the same topic (Eq. 18):

$$\begin{cases} Cs(Z_i) = \frac{2*\sum_{w_s \neq w_g} cs(w_s, w_g)}{|z_i^2| - |z_i|} \\ cs(w_s, w_g) = \frac{c(w_s, w_g)}{max_{r \neq h} c(w_r, w_h)} \end{cases} \quad (22)$$

where $c(w_s, w_g)$ is the number of co-occurrences between keywords such that $w_s, w_g \in z_i$. We have two possibilities to compute the consistency among keywords. The first one consists of using an external resource such as WordNet to determine semantic relations between keywords (Hajlaoui et al. 2017a). The second way is the computation of the co-occurrence of keywords in the corpus of tweets. The problem in the first method is the non-existence of important numbers of keywords in external resources like abbreviations and names. Therefore, we opt for utilizing the second method and we propose the following measure.

When we compute the consistency of each topic, this value can be important only if most of the topic's keywords co-occur in several tweets. For this, in the classification process, we limit the size of topics between 6 and 22 keywords in order to find meaningful results.

6.2.2 Decay nutrition factor

In the aging theory, the role of the decay nutrition factor (β) is the control of the destruction and regeneration of topics of interest in the networks. In other approaches, this value is learnt using the aging theory (Chen et al.
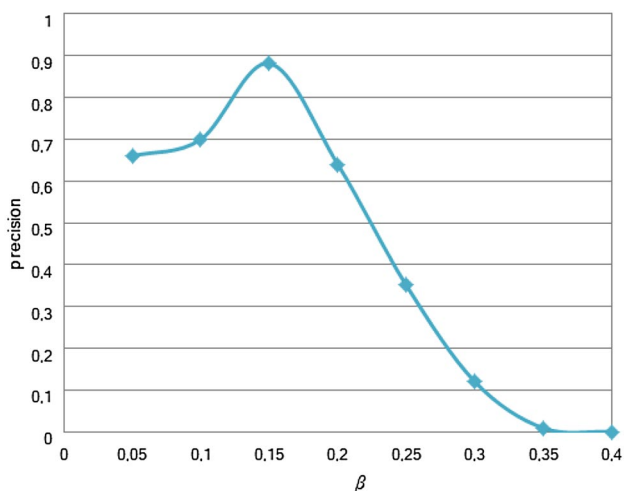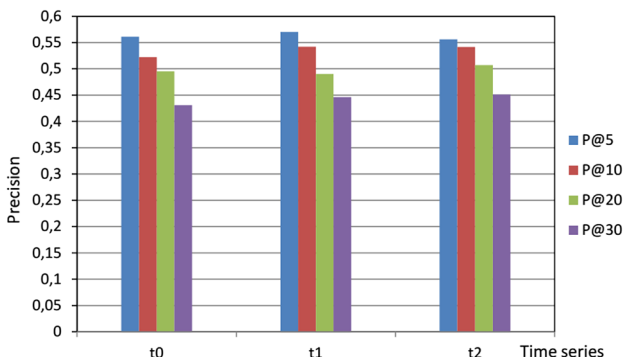
Fig. 8 Decay nutrition factor

Fig. 9 Precision variation in three times slices

2003). To learn this parameter, we use 100 ego networks, 3,200 topics and we compute precision's for five-time series. The results are in Fig. 8. Precision is maximal with β between 0.13 and 0.17. When β increases, the topic destruction becomes more severe. In the rest of this experimentation, we consider that β = 0.13 for tolerating topics with average nutrition.

6.2.3 Precision in times series

Fig. 9 contains the variation of precision in three time slices. It is slightly decreased or increased between the three time series. For example, precision P@5 varies between first and second time's series with + 0.005 and with − 0.003 in the second interval. The results show the stability of our approach when we update topics of interest.
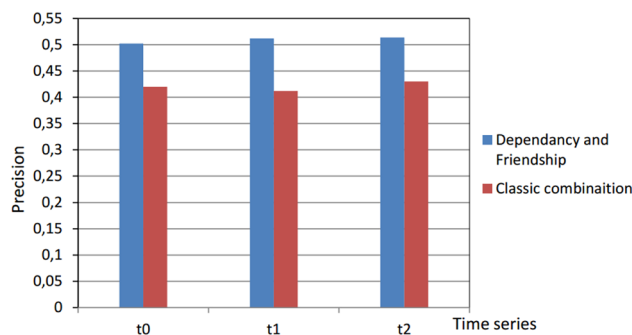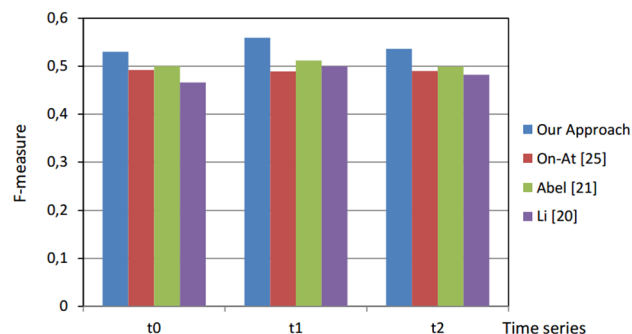

Fig. 10 Dependency and friendship result

Fig. 11 Comparison between our approach and other approaches using F-measure

6.2.4 Comparison

In Fig. 10, we compare the results found when we use dependency and friendship degree in the Dempster rule combination. The precision will decrease significantly when we do not consider dependency and friendship degree. The friendship degree reflects the relation between two users in the network. This degree is computed utilizing the number of common triadic closures, common followings and semantic similarity between two . When we infer topics of interest from neighbors, it is more likely that two users are interested in same topics of interest if the friendship degree is important. Therefore, this value may contribute to the combination of belief functions. Furthermore, the consideration of dependency between users in the same ego network must eliminate the redundancy of the assigned beliefs to the topics.

Figures 11 and 12 show the comparison between our approach and two different approaches. Our model has the best measure of the three-time slices (Table 4). In the work proposed by Li et al. (2014), the temporal aspect was modeled using the decay function $e^{(-\alpha t)}$. This function gave the same period for each interest and did not take into account the topic nutrition factor. In the work suggested by Abel
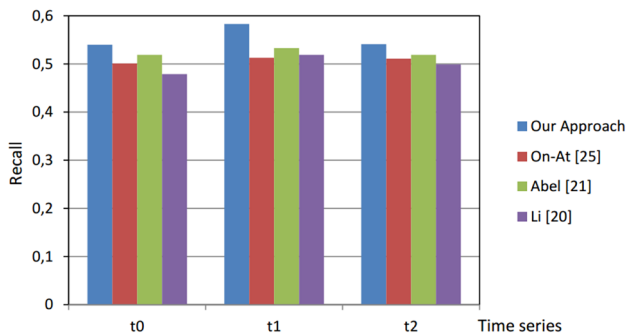
Fig. 12 Comparison between our approach and other approaches using Recall

Table 4 Comparison between our approach and other approaches using Precision

	Our approach (%)	On-At (2016) (%)	Abel (2011) (%)	Li (2014) (%)
t_0	52	48	48	45
t_1	53	46	49	48
t_2	53	47	48	46

et al. (2011), time-stamps for topics of interest were given by the user. A topic of interest was represented by the set of concepts. Each profile was represented by the popular interest in a given time stamp. Thus, in these two approaches, topics were discovered only from user publications in each time slice. If a user does not post messages, their interests remain unchanged. Contrariwise, in our approach, we determine the interests of a given user from their ego network. If the topic of interest in the ego network changes, it will be more likely that the interests of the user will change too. The aging theory models the temporal aspect and takes into consideration the keywords that appear for the nutrition of interests. Each topic has its own lifetime.

The approach proposed by On-At et al. (2016) also used the exponential decay function. Like our approach, this method is based on the principle of social filtering. The social interests of each user are identified from his egocentric network. Nevertheless, this method does not manage the dependency between the users. Thus, all members of an egocentric network are considered when inferring social interests whereas only influential members are considered in our approach.

6.3 Discussion

The proposed approach in this work has shown important results compared with previous approaches. It is designed to run on a real-world application such as the current social

media platforms by modeling the latent social network and identifying users' interests. However, some modifications must be made to the model before deployment. Firstly, the determination of the importance of links between the users must be carried out from time to time to keep a correct image on the egocentric networks. In fact, the frequency of activities between members changes over time. Users interested in specific profiles for a period of time may change their interests and they will communicate with new members. This change implies the change of influence and the dependency in the network. Afterward, the temporal topics of interest identification system must be able to discover at each time interval the nature of the relationships between the members.

The second point to consider when deploying our approach is the updates propagation of in the network. In fact, each node in the network updates his topics of interest list according to the information submitted from his neighbors. Then, this node must inform his neighborhood of the change made. The update can be transmitted gradually to reach a large number of nodes of the network which may alter the performances of the system. The solution that we propose to solve this problem consists of determining the influence between members of the network. An algorithm like PageRank (1999) can be used for this task. Then, a filtering step must be performed on the updates transmitted between the nodes. Each node will only take into account updates sent by important neighbors.

7 Conclusion and perspectives

We have proposed in this work an approach for the discovery and tracking of temporal topics of interest in the Twitter. This model is based on belief function and aging theories. The belief theory is introduced for combining evidence from different alters in the user ego network. We utilize only the direct followings and friends from alter users. We have shown that these nodes are the most informative for inferring topics of interest of an ego user. Different factors are taken into account when combining the evidence mainly dependency, friendships, and similarity. Dependency is introduced in the combination step for eliminating evidence redundancy. Friendship and similarity degrees between users are introduced to compute the reliability of links. It permits to raise evidence from friends who are close to an ego user and decrease other evidence. The aging theory is introduced for tracking the temporal evolution of topics of interest. It is based on two measures: the nutrition function and the decay factor. In our case, the first measure is computed using the frequency of keywords in the ego and the global network. The nutrition informs us a topic is still an interest or not. The decay factor is determined utilizing the experimental

analysis. This combination between the belief function and aging theories allows both treating the uncertain information during user profile derivation and tracking the evolution of topics of interest over time.

The suggested algorithm in this work can be executed on a distributed platform with tracking interest transitions and capturing real network images. A node in the microblogging network is an ego node. Each node analyzes its generated content, determines the degree of friendship with its alters and receives the evidence from different participants to infer its interests. If a topic appears in a profile, the node inform its neighbors to update their interest lists.

Our proposed approach has given important results for user profile derivation and tracking the evolution of interests. In the future, we aim to exploit the proposed model for other potential applications. In particular, we will focus on the predictions of social and personal interests. It will be interesting to incorporate other social information and factors into the model.

References

Abel F, Gao Q, Houben GJ, Tao K (2011) Analyzing temporal dynamics in Twitter profiles for personalized recommendations in the social web. In: Proceedings of the 3rd international web science conference (WebSci '11). ACM, New York, NY, USA, Article 2 , p 8

Ahmed A, Low Y, Aly M, Josifovski V, Smola AJ (2011) Scalable distributed inference of dynamic user interests for behavioral targeting. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining (KDD '11)

Aiello LM, Schifanella R (2014) Reading the source code of social ties. In: Proceedings of the 2014 ACM conference on Web science (WebSci '14). ACM, New York, NY, USA, pp 139–148

Bao H, Li Q, Liao SS, Song S, Gao H (2013) A new temporal and social pmf-based method to predict users' interests in microblogging. Decis Support Syst 55(3):698–709

Bao B-K, Xu C, Min W, Hossain MS (2015) Cross-platform emerging topic detection and elaboration from multimedia streams. ACM Trans Multimed Comput Commun Appl (TOMM) 11(4):54

Basir O, Karray F, Zhu H (2005) Connectionist-based dempster-shafer evidential reasoning for data fusion. IEEE Trans Neural Netw 16(6):1513–1530

Bhattacharya P, Zafar MB, Ganguly N, Ghosh S, Gummadi KP (2014) Inferring user interests in the Twitter social network. In: Proceedings of the 8th ACM Conference on Recommender systems (RecSys'14). ACM, New York, NY, USA, pp 357–360

Boschee E, Natarajan P, Weischedel R (2013) Automatic extraction of events from open source text for predictive forecasting. In: Subrahmanian V (ed) Handbook of computational approaches to counterterrorism. Springer, New York, pp 51–67

Boukhechba M, Bouzouane A, Gaboury S, Gouin-Vallerand C, Giroux S, Bouchard B (2017) Prediction of next destinations from irregular patterns. J Ambient Intell Hum Comput 9(5):1345–1357

Budak C, Kannan A, Agrawal R, Pedersen J (2014) Inferring user interests from microblogs. AAAI ICWSM

Cattaneo M E (2003) Combining belief functions issued from dependent sources. In: Research report/seminar für statistik,

eidgenössische technische hochschule (eth), vol 117, Seminar für Statistik, Eidgenössische Technische Hochschule (ETH), Zürich

Chen Y-W, Wang J-L, Cai Y-Q, Du J-X (2015) A method for chinese text classification based on apparent semantics and latent aspects. J Ambient Intell Hum Comput 6(4):473–480

Chen CC, Chen Y-T, Sun Y, Chen MC (2003) Life cycle modeling of news events using aging theory. European conference on machine learning. Springer, Berlin, Heidelberg, pp 47–59

Chen K-Y, Luesukprasert L, Seng-cho TC (2007) Hot topic extraction based on timeline analysis and multidimensional sentence modeling. IEEE Trans Knowl Data Eng 19(8):1016–1025

Cheng J, Romero D M, Meeder B, Kleinberg J (2011) Predicting reciprocity in social networks. In: IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, Boston, pp 49–56

Cheng Y, Qiu G, Bu J, Liu K, Han Y, Wang C, Chen C (2008) Model bloggers' interests based on forgetting mechanism. In: Proceedings of the 17th international conference on world wide web, pp 1129–1130

Dai T, Zhu L, Cai X, Pan S, Yuan S (2018) Explore semantic topics and author communities for citation recommendation in bipartite bibliographic network. J Ambient Intell Hum Comput 9(4):957–975

De Maio C, Fenza G, Gallo M, Loia V, Parente M (2018) Social media marketing through time-aware collaborative filtering. Concurr Computat Pract Exp 30(1):e4098

Denœux T (2008) Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. Artif Intell 172(2–3):234–264

Destercke S, Dubois D (2011) Idempotent conjunctive combination of belief functions: Extending the minimum rule of possibility theory. Inform Sci 181(18):3925–3945

Dubois D, Yager RR (1992) Fuzzy set connectives as combinations of belief structures. Inform Sci 66(3):245–276

Ferrari L, Mamei M, Colonna M (2014) Discovering events in the city via mobile network analysis. J Ambient Intell Hum Comput 5(3):265–277

Hajlaoui JE, Omri MN, Benslimane D (2017a) Multi-tenancy aware configurable service discovery approach in cloud computing. In: IEEE 26th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), Poznan, pp 232–237

Hajlaoui JE, Omri MN, Benslimane D (2017b) A qos-aware approach for discovering and selecting configurable iaas cloud services. Comput Syst Sci Eng 32(4)

Hajlaoui JE, Omri MN, Benslimane D, Barhamgi M (2017c) Qos based framework for configurable iaas cloud services discovery. In: IEEE International Conference on Web Services (ICWS), Honolulu, HI, pp 460–467

Han X, Wang L, Crespi N, Park S, Cuevas Á (2015) Alike people, alike interests? inferring interest similarity in online social networks. Decis Support Syst 69:92–106

Hopcroft J, Lou T, Tang J (2011) Who will follow you back?: reciprocal relationship prediction. In: Proceedings of the 20th ACM international conference on information and knowledge management, Glasgow, Scotland, UK, pp 1137–1146

Imamori D, Tajima K (2016) Predicting popularity of twitter accounts through the discovery of link-propagating early adopters. In: Proceedings of the 25th ACM international on conference on information and knowledge management, ACM, New York, NY, USA, pp 639–648

Karidi DP, Stavrakas Y, Vassiliou Y (2017) Tweet and followee personalized recommendations based on knowledge graphs. J Ambient Intell Hum Comput 1–15

Li Y-M, Hsiao H-W, Lee Y-L (2013) Recommending social network applications via social filtering mechanisms. Inform Sci 239:18–30

Li L, Zheng L, Yang F, Li T (2014) Modeling and broadening temporal user interest in personalized news recommendation. Expert Syst Appl 41(7):3168–3177

Maleki-Dizaji S, Siddiqi J, Soltan-Zadeh Y, Rahman F (2014) Adaptive information retrieval system via modelling user behaviour. J Ambient Intell Hum Comput 5(1):105–110

Mathioudakis M, Koudas N (2010) Twittermonitor: trend detection over the twitter stream. In: Proceedings of the 2010 ACM sigmod international conference on management of data, (SIGMOD '10). ACM, New York, NY, USA, pp 1155–1158

Mcauley J, Leskovec J (2014) Discovering social circles in ego networks. ACM Trans Knowl Discov Data (TKDD) 8(1):4

Mislove A, Viswanath B, Gummadi K P, Druschel P (2010) You are who you know: inferring user profiles in online social networks. In: Proceedings of the third ACM international conference on web search and data mining, (WSDM '10). ACM, New York, NY, USA, pp 251–260

On-At S, Quirin A, Péninou A, Baptiste-Jessel N, Canut M-F, Sèdes F (2016) Taking into account the evolution of users social profile: Experiments on twitter and some learned lessons. In: Research challenges in information science (rcis), 2016 IEEE tenth international conference on research challenges in information science (RCIS), Grenoble, pp 1–12

Page L, Brin S, Motwani R, Winograd T (1999) The pagerank citation ranking: bringing order to the web. Technical Report, Stanford InfoLab

Pietilänen A-K, Diot C (2012) Dissemination in opportunistic social networks: The role of temporal communities. In: Proceedings of the thirteenth acm international symposium on mobile ad hoc networking and computing. ACM, New York, NY, USA, pp 165–174

Rechy-Ramirez EJ, Marin-Hernandez A, Rios-Figueroa HV (2017) Impact of commercial sensors in human computer interaction: a review. J Ambient Intelli Hum Comput 9(5):1479–1496

Sendi M, Omri MN, Abed M (2017) Possibilistic interest discovery from uncertain information in social networks. Intell Data Anal 21(6):1425–1442

Shafer G (1976) A mathematical theory of evidence, vol 42. Princeton university press, Princeton

Su X, Mahadevan S, Xu P, Deng Y (2015) Handling of dependence in dempster-shafer theory. Int J Intell Syst 30(4):441–467

Takemura H, Tanaka A, Tajima K (2015) Classification of twitter follow links based on the followers' intention. In: Proceedings of the 30th annual ACM symposium on applied computing (SAC '15). ACM, New York, NY, USA, pp 1174–1180

TalebiFard P, Leung VC (2014) Context-aware dissemination of information and services in heterogeneous network environments. J Ambient Intell Hum Comput 5(6):775–787

Tchuente D, Canut M-F, Jessel N, Péninou A, Sèdes F (2013) A community-based algorithm for deriving users profiles from egocentrics networks: experiment on facebook and dblp. Soc Netw Anal Mining 3(3):667–683

Vinagre J, Jorge AM, Gama J (2015) An overview on the exploitation of time in collaborative filtering. Wiley Interdiscip Rev Data Mining Knowl Discov 5(5):195–215

Wang J, Zhao WX, He Y, Li X (2014) Infer user interests via link structure regularization. ACM Trans Intell Syst Technol (TIST) 5(2):23

Wang Y, Agichtein E, Benzi M (2012) Tm-lda: efficient online modeling of latent topic transitions in social media. In: Proceedings of the 18th ACM sigkdd international conference on knowledge discovery and data mining (SAC '15). ACM, New York, NY, US, pp 123–131

Wen Z, Lin C-Y (2010) On the quality of inferring interests from social neighbors. In: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining (KDD '10). ACM, New York, NY, USA, pp 373–382

Zarrinkalam F, Fani H, Bagheri E, Kahani M (2016) Inferring implicit topical interests on twitter. In: Advances in Information Retrieval, Springer International Publishing, pp 479–491

Zheleva E, Getoor L (2009) To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In: Proceedings of the 18th international conference on world wide web (WWW '09). ACM, New York, NY, USA, pp 531–540