



**HAL**  
open science

## Taking into Account the Evolution of Users Social Profile: Experiments on Twitter and some Learned Lessons

Sirinya On-At, Arnaud Quirin, André Péninou, Nadine Jessel,  
Marie-Françoise Canut, Florence Sèdes

### ► To cite this version:

Sirinya On-At, Arnaud Quirin, André Péninou, Nadine Jessel, Marie-Françoise Canut, et al.. Taking into Account the Evolution of Users Social Profile: Experiments on Twitter and some Learned Lessons. 10th IEEE International Conference on Research Challenges in Information Science (RCIS 2016), Jun 2016, Grenoble, France. pp.87-98, 10.1109/RCIS.2016.7549325 . hal-01514634

**HAL Id: hal-01514634**

**<https://hal.science/hal-01514634v1>**

Submitted on 26 Apr 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>  
Eprints ID : 17020

The contribution was presented at RCIS 2016 :  
<http://www.sense-brighton.eu/rcis2016/>

**To cite this version** : On-At, Sirinya and Quirin, Arnaud and Péninou, André and Jessel, Nadine and Canut, Marie-Françoise and Sèdes, Florence *Taking into Account the Evolution of Users Social Profile: Experiments on Twitter and some Learned Lessons*. (2016) In: 10th IEEE International Conference on Research Challenges in Information Science (IEEE RCIS 2016), 1 June 2016 - 3 June 2016 (Grenoble, France).

Any correspondence concerning this service should be sent to the repository administrator: [staff-oatao@listes-diff.inp-toulouse.fr](mailto:staff-oatao@listes-diff.inp-toulouse.fr)

# Taking Into Account The Evolution Of Users Social Profile: Experiments On Twitter And Some Learned Lessons

Sirinya On-at, Arnaud Quirin, André Péninou, Nadine Baptiste-Jessel, Marie-Françoise Canut, Florence Sèdes  
Toulouse Institute of Computer Science Research, University of Toulouse, CNRS, INPT, UPS, UT1, UT2J  
TOULOUSE, France

{sirinya.on-at, andre.peninou, nadine.baptiste, marie-francoise.canut, florence.sedes}@irit.fr,  
aquirin@gmail.com

**Abstract**— Incorporating user interests evolution over time is a crucial problem in user profiling. We particularly focus on social profiling process that uses information shared on user social network to extract his/her interests. In this work, we apply our existing time-aware social profiling method on Twitter. The aim of this study is to measure the effectiveness of our approach on this kind of social network platform, which has different characteristics from those of other social networking sites. Although the improvement compared to the time-agnostic baseline method is still low, the experiments using a parametric study showed us the benefit of applying a time-aware social profiling process on Twitter. We also found that our method performs well on sparse networks and that the information dynamic influences more the quality of our proposed time-aware method than the relationships dynamic while building the social profile on Twitter. This observation will lead us to a more complex study to find out meaningful factors to incorporate user interests evolution on social profiling process in such a network.

**Keywords**— *Social profile; User profile; Egocentric network; Time-aware method*

## I. INTRODUCTION

In Online Social Networks (OSNs), users are encouraged to contribute, broadcast contents and connect to those who share the same interests and/or activities. This behavior is the source of valuable resources to infer user personal interests and build a user profile, which is an important element in adaptive systems. Adaptive mechanisms (e.g. recommendation, personalization, ...) make use of the user profile to identify user interests in order to propose relevant content according to the user specific needs. In our work, rather than using the user behavior, we focus on using information shared by the user neighbors to extract the user interests and build a profile, so-called “social profile”. In this context, the social profile can be used as an additional profile to the classical user profile in adaptive systems.

In OSNs setting, user online behavior evolves quickly over time. The information and relationships in an OSN can rapidly become obsolete. This characteristic reflects the importance of taking into account the evolution of user interests in our

context. To overcome this issue, we proposed a time-aware social profiling method in our previous work [1] in order to embody the evolution of user interests in the social profile building process. In [1], based on our existing community based social profiling approach [2], we proposed to integrate a “temporal score” to each interest of the social profile according to its relevance and freshness. This temporal score reflects the relevance of the information sources (information and individuals).

Different types of social media platforms, which vary in terms of their scopes and functionalities, are living together today [3]. They implement various functionalities (e.g. social networking sites such as Facebook or Myspace, bookmarking sites such as Delicious, microblogging sites such as Twitter, Sina or Weibo, ...), and have very different evolution patterns. The social network characteristics (e.g. density, type of social relationships, frequency of contact, ...) of each platform are also different [4], [5]. We suggest that the evolution of a given social network can have a fundamental impact over social profiling process.

In this paper, we are interested in applying our time-aware social profiling method [1] on Twitter, a popular microblogging platform that has a huge growth rate. The aim of this study is to measure the effectiveness of our approach on this kind of social networks, which have different characteristics than that of the co-authorship networks previously studied. In fact, Twitter is more global (less focused) in terms of users and shared information. Thus, we can extract various domains of user interests from this platform. Twitter is considered as an information sharing site rather than a strict social networking site [6]. Compared to other social network platforms, Twitter users tend to follow or share information rather than establishing connections. Furthermore, on Twitter, the information evolves more quickly, so the life-cycle of a tweet is shorter compared to posts in other platforms.

We hypothesize that the characteristics of Twitter may have an impact over our time-aware method to obtain a relevant social profile. Firstly, the influence of information dynamic

compared to the influence of relationship dynamic may be different from that observed on the co-authorships networks. Secondly, the evolution rate of the information and user relationships in Twitter may be also different. Finally, we also checked if extracting communities using the relationships information on Twitter is relevant. To summarize, we try to answer the following questions:

- Will our time aware method perform equally well on Twitter than on co-authorships networks (DPLP/Mendeley)?
- Will the influence of information dynamic compared to relationships dynamic be different on Twitter?

This paper is structured as follows. In the two next sections, we present the related work on user profiling and social profiling. In the third section, we present our time-aware social profiling method. In the fourth section, we describe the experimentation case study conducted on Twitter. The last section concludes and presents some future works.

## II. SOCIAL PROFILE

In this paper, we use the term “social profile” to designate a specific user profile built by using the information from his/her social network members. Social profile is useful, on the one hand, to provide additional information that can improve the relevance of existing user profile and, on the other hand, to complete non-existing/missing profiles for new or less active users. According to [7], various models of user profiles have been proposed. We consider the user profile (personal or social) as a vector of weighted user interests represented by keywords.

We are particularly interested in extracting user interests from his/her egocentric network. An egocentric network is a specific social network that takes into account only the user’s direct connections. This decision is motivated by the assumption that a user tends to connect with the people who share common interests with him/her [8]. An egocentric network is composed of the individuals (alters) having a direct relationship with the user (ego) and the relationships between these individuals. The egocentric network of a user is defined as follows: for each user ( $u$ ) we consider the undirected graph  $G(u) = (V, E)$  where  $V$  is the set of nodes directly connected to  $u$ , and  $E$  is the set of relationships between each node pair of  $V$ . We emphasize that  $u$  is not included in  $V$ .

In the literature, we can distinguish two social profiling approaches. The individual based approach extracts user interests and computes the score of each interest according to the characteristics of each individual [9], [10]. Conversely, the communities based approach extracts the communities from user egocentric network and generates the user interests according to the characteristics of each community [2]. The effectiveness of social profiles has been proved with empirical results [2], [9], [10]. [2] also showed the effectiveness of the community based approach compared to the individual based approach applied on co-authorship networks.

As social networks evolve over time, it is necessary to take into consideration user interests evolution in his/her social profile. However, this issue has not been widely taken into

consideration in the proposed social profile building approaches.

Based on the social network analysis literature, social network evolution is related to network structural dynamics (existence, creation and persistence of social relationships among social actors) and information flows (information sharing and diffusion between social actors) [11], [12].

We suggest that the interests evolution in a user social profile is related to the behavior of user social network members (alters). Hence the evolution of user interests is related to the evolution of information shared on user social network and the evolution of relationships between the user (ego) and his/her alters. In the literature, user interests evolution has been taken into consideration in different works [13]–[18]. However, the user interests are extracted by using only the user information/behavior so the relationships evolution is not taken into consideration in these works.

In order to incorporate interests evolution in social profiling process and to take into account both relationships and information dynamics, we have proposed a time-aware social profile building method using the community based approach [1]. The detail of this method is described in the next section.

## III. TIME AWARE SOCIAL PROFILE CONSTRUCTION

Before detailing our time-aware method that deals with user interests evolution, we first describe the main steps of the entire social profile building process. A global view of the profiling process is presented as follows:

- For a given user(ego), for whom we want to build social profile, we use his egocentric networks as the information sources to extract the interests and build his social profile: we collect the information shared by each member of user’s egocentric network.
- Then, we extract the keywords from all collected information and aggregate them by using a scoring function. This step is customizable so we can apply additional features or techniques to calculate the score of each extracted element.
- We derive all calculated elements (interests) to the social profile according to their score. Finally, the derived social profile is represented in the form of a vector of weighted user interests.

In the social profile construction process, we call  $u$  the user (ego) for whom we desire to build social profile. We use the term individuals ( $SetIndiv$ ) to represent the set of user’s egocentric network members. For each individual  $Indiv \in SetIndiv$ , his information is called  $Info_{Indiv}$ . Each  $Info_{Indiv}$  contains elements called  $e$ , extracted by using classical text analysis techniques [2]. Note that in our context, a user interest is represented by an element, more precisely a keyword.

The main steps of the entire social profile building process are as follows (see Fig. 1). The first step is to extract the information  $Info_{Indiv}$  of each  $Indiv$  in the user egocentric network. In the second step, we extract the keywords ( $e$ ) from all  $Info_{Indiv}$  and we aggregate them by using a scoring function (detailed below and see also section IV.C). This step is

customizable so we can apply additional features or techniques to calculate the score of each extracted element. Each element is modeled by a pair  $(e, score)$ . In the last step we derive all calculated elements (interests) to the social profile according to their score.

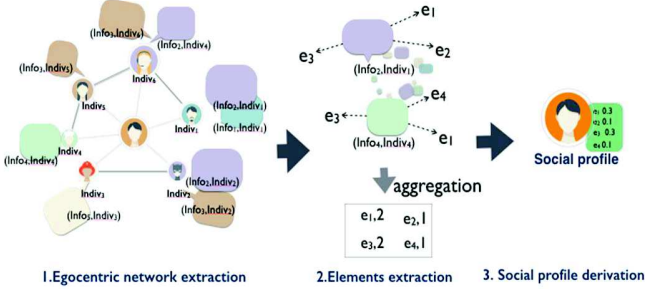


Fig. 1. Social profiling process

We now focus on the second step of the process (see Fig. 1) and propose to integrate a temporal score (TempScore), to each extracted element  $e$ .

A temporal score of an element  $e$  is computed by using, on the one hand, the temporal relevance of the information  $Info_{Indiv}$  used to extract  $e$  and, on the other hand, the relevance of the individual  $Indiv$  that is the source of this information. The relevance of information  $Info_{Indiv}$  is computed regarding its freshness. The relevance of an individual  $Indiv$  is computed regarding his/her relationship strength with the central user (*ego*). Next sections detail the computation of these score and the combination of these two scores to obtain the final temporal score (TempScore).

### 1) Information temporal score

To weight any information  $Info_{Indiv}$  using temporal factors, we adopt a time exponential function (1) proposed in [19].

$$f(t) = e^{-\lambda t} \quad (1)$$

The value of  $t$  represents the information freshness for each  $t=i$  ( $i \in N$ ), represented by the elapsed time between the information timestamp and the given timestamp. Thus,  $t=0$  represents the freshness value of the most recent period (ex:  $t=0$  for the current date,  $t=1$  for yesterday, and so on).  $\lambda \in [0,1]$  represents the time decay rate. The higher  $\lambda$  is, the less important the old information is.

The information temporal score is computed by applying the temporal function  $f(t)$  of formula (1). In our case, the value of  $t$  is not a simple freshness calculated from a reference timestamp. In fact, we suggest to take into account and to combine two kinds of freshness information: the freshness based on the last interaction between  $Indiv$  and the central user  $u$ , called  $t_{interaction}$  and the freshness based on the timestamp of  $Info_{Indiv}$ , called  $t_{timestamp}$ .

The value  $t_{timestamp}$  of information  $Info_{Indiv}$  is computed by considering the temporal distance between the  $Info_{Indiv}$  timestamp and the current time, as shown in the formula (2).

$$t_{timestamp}(Info_{Indiv}) = |current\ date - timestamp(Info_{Indiv})| \quad (2)$$

The value of  $t_{interaction}$  is computed by considering the timestamp of the last interaction between  $u$  and  $Indiv$  as shown in the formula (3). The relevance of information shared by  $Indiv$  before and after the interaction is reduced according to their temporal distance from the last interaction between  $Indiv$  and  $u$ . For example, if the last interaction between  $Indiv$  and  $u$  is in 2014,  $t_{interaction}$  of the information shared in 2013 and 2015 is equal to 1 and  $t_{interaction}$  of the information shared in 2012 and 2016 is equal to 2 (see figure 2).

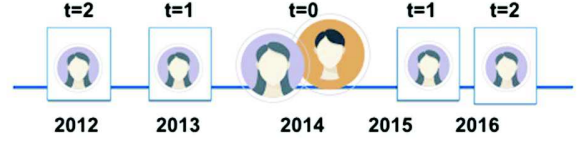


Fig. 2. The temporal relevance of information based on the last interaction of the user and a given individual

$$t_{interaction}(Info_{Indiv}) = \left| \left( \text{last interaction time stamp}(Indiv, u) \right) - \text{timestamp}(Info_{Indiv}) \right| \quad (3)$$

Finally, the information temporal score of a given information  $Info$  shared by an individual  $Indiv$  is assigned to each element  $e$  extracted from  $Info$  and is computed for each  $e$  using the formula (4):

$$P_{Temp}^{Info, Indiv}(e) = f(t_{timestamp}(Info_{Indiv})) + f(t_{interaction}(Info_{Indiv})) \quad (4)$$

### 2) Individual temporal score

In order to compute a relationships strength between an individual  $Indiv$  and the user  $u$ , we propose to apply a link prediction metric to compute a similarity score between each given  $Indiv$  and  $u$ . This similarity score will represent their relationship persistence and approximate their relationship strength.

To take into account the temporal factor, we adopt a time-aware link prediction technique introduced by [20] which proposes the integration of the temporal score into the existing Adamic/Adar metric [21]. Based on this work, we use the following ‘‘temporal Adamic/Adar’’ metric:

$$\text{TemporalAdamicAdar}(x, y) = \sum_{z \in \{\Gamma(x) \cap \Gamma(y)\}} \frac{w(x, z) \cdot w(z, y)}{\log|\Gamma(z)|} \quad (5)$$

where  $\Gamma(x)$  represents the set of the neighbors of  $x$ . The  $w(x, y)$  function represents the temporal relevance score of two given nodes  $x$  and  $y$ .

In our work, we have used the temporal function from the formula (1) to modify the function  $w()$  as follows (6):

$$w(x_1, x_2) = f(\text{current timestamp} - \text{last interaction timestamp}(x_1, x_2)) \quad (6)$$

where  $x1$  and  $x2$  are two given individuals.

Finally, the individual temporal relevance score of an individual  $Indiv$  is assigned to each element  $e$  extracted from any information  $Info_{Indiv}$  as follows (7):

$$P_{Temp}^{Indiv}(e) = \text{TemporalAdamicAdar}(u, Indiv) \quad (7)$$

We note that this temporal score is the same for all elements  $e$  extracted from any  $Info_{Indiv}$  shared by the individual  $Indiv$ , i.e. the  $Info_{Indiv}$  is not considered here as a parameter.

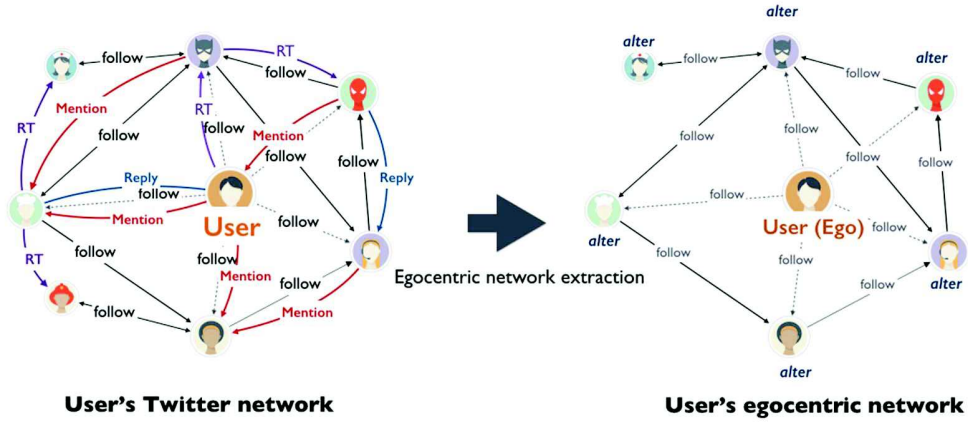


Fig. 3 Egocentric network extraction based on the following relation

### 3) Temporal score calculation and integration

The final temporal score of an element  $e$  extracted from the information  $Info$  shared by an individual  $Indiv$  is computed by combining the information temporal score and the individual temporal score as follows (8):

$$P_{Temp}^{Comb}(e) = combination(P_{Temp}^{Indiv}, P_{Temp}^{Info, Indiv}, \gamma) | e \in Info \quad (8)$$

The combination function of (8) is a linear function:  $combination(A, B, p) = p*A + (1-p)*B$ , where  $p \in [0,1]$  represents a proportion between  $A$  and  $B$ . In the formula (8),  $\gamma$  is the proportion which varies the importance of the individual score compared to the information one.

Once the temporal score of all elements are computed, in the next step, the elements are aggregated. Finally, the aggregated elements can be derived to the social profile as weighed user interests.

Based on the experiments results on co-authorship networks conducted in our previous work [1], our proposed time-aware method outperforms the time-agnostic method on the community-based social profiling approach. We also found that the relationships dynamic plays an important role compared to that of information dynamic to generate the most relevant social profile.

In this paper, we suggest that the different characteristics of OSNs platforms (particularly in terms of functionalities and evolution patterns) can play a fundamental role in our time-aware social profiling method. Factors that influence the effectiveness of the social profile could be different according to the kind of networks. In information sharing sites, we can hypothesize that user interests evolution is impacted more by information dynamic than by the relationships dynamic. We will apply the previous approach on a new kind of networks (Twitter) in order to check our hypothesis and if our technique is generic or not.

## IV. TWITTER BASED SOCIAL PROFILING

### A. Twitter as a data source

In this paper, we particularly focus on conducting our proposed approach on Twitter which is one of the most popular microblogging sites [14], [22].

Twitter is well-known both as a micro-blogging service provider and a social network platform. A message posted by a user is called a tweet, limited to 140 characters. The tweets can contain short texts, hyperlinks to other multi-media (image, video) and also hashtags constructed by prefixing a word with a '#' character (for example, #prayForParis, #Euro2016). A hashtag is an explicit topic categorization used to create and to follow threads of discussion.

In Twitter, any user can establish connections with other users. The connection "follow" allows a user to receive the tweets that others post on their timeline. The people who follow a user are called "followers" and the people followed by a user are called "following". Note that the relationships of following and followers require no reciprocation (see Fig. 3). Users can also create their own lists (of users) or subscribe to lists created by others. Viewing a list timeline will enumerate a stream of tweets from only the users on that list. In term of user interactions, users can: i) post tweets, ii) "reply" to a tweet of another user by placing a @user reference in it, or iii) forward a tweet to their followers, which is called "retweet". The @user reference can also be used to "mention" (refer to) a particular user in the tweet content.

[22] studied the characteristics of Twitter data, and found that trending topics on Twitter evolved quickly and can come back over time. The life cycle of a tweet is shorter than the posts on most other OSNs (a week or shorter). They also found that the average number of tweets against the number of followers per user is always above the median. In our point of view, micro-blogging sites are more considered as information sharing sites rather than social networking sites.

### B. Extracting user egocentric network from Twitter

We remind that we are interested in using the information from user egocentric network to extract his/her interests since the user tends to establish a directed connection with the

individuals who share some common interests or characteristics with him/her.

Among its features, Twitter is a multi-relation social network (i.e. follow, reply to, mention, retweet the post). In this paper, we suppose that a user tends to follow other users that share information concerning his/her interests. So, we use the “following” relations to extract the user egocentric network.

In this context, we consider that:

- the egocentric network of a user  $u$  corresponds to the set of his/her followings.
- the “following” relations represent the relationships between the alters.
- the reply, retweet and mention relations are “local references representing the interactions between those members. We consider that the users  $u_1$  and  $u_2$  interact with each other as soon as  $u_1$  retweets a post of  $u_2$ ,  $u_1$  replies to a post of  $u_2$  or  $u_1$  mentions a post of  $u_2$ . In this definition, as soon as one user retweets, replies or mentions another user, the two are considered as interacting with each other.

Various works propose to identify user’s interests using tags content and prove the relevance of this approach [23]. We propose in this paper to extract the interests from this kind of content (i.e. hashtags) to validate the effectiveness of our approach on Twitter.

### C. Twitter based social profiling process

We crawled Twitter users using the Application Programming Interface (API) provided by the official platform. We started by collecting members of some selected lists concerning different topics (social network analysis, technology, video game, sport, travel). Note that almost all the users in each list are the most popular users for the corresponding topic. So, we can extract meaningful information from their posts and relations (avoiding spam accounts).

After analyzing the egocentric networks of all crawled users, we observed that the alters of the users are not widely connected between them, in particular if we compare to our previous experiments on co-authorship networks. In our previous work, we adopted a community based approach to build the social profile. The adopted community extraction algorithm was only based on the relationships between the individuals in the network. In the context of microblogging, according to [15], [24], social tie information may not be very helpful for users who use microblogging for information purpose, since users with similar topic interests may not be explicitly connected. Furthermore, in this kind of network weak ties can often provide access to new information field.

Based on this observation, we suppose that the community based social profiling process used in our previous work, may generate a less relevant social profile since the extracted community could be less meaningful on this kind of network.

Therefore, in this paper, we propose to apply our time-aware method on both the individual based approach and the

community based approach to ensure the effectiveness of our method regarding the adopted approach. We will compare our results with classical time-agnostic approaches. In the following subsections, we present respectively the time aware individual based social profiling process (IBSPT) and the time aware community based social profiling process (CoBSPT).

Note that, in this context, the individuals  $Indiv$  are the followings of the principal user  $u$ . The information used is the tweet and the extracted elements  $e$  are the hashtags contained in tweets. We consider “day” as the time granularity for our temporal score calculation. To compute the information temporal score (see section III), we use the post timestamp to infer the information temporal (freshness) of each tag. To compute the individual temporal score, we consider the timestamp of “retweet”, “reply” or “mention” between two given users as their interaction timestamps.

#### 1) Time aware individual based social profiling process (IBSPT)

Based on the individual based social profiling approach (IBSP), we present the time-aware individual based social profiling process (IBSPT) for a given user  $u$  as follows.

Step 1: Extracting the followings of the user  $u$  ( $SetIndiv$ ) and the posts on their timeline.

Step 2: Computing the profile of each  $Indiv$  by extracting elements  $e$  from his/her tweet ( $e, Indiv$ ). We apply our time-aware method presented in the section III to compute the weight of each extracted element  $e$ . Once the temporal score  $P_{Temp}^{Comb}(e)$  of each element  $e$  is calculated by using the formula (8), we aggregate them by computing their weights compared to the total weight of all founded elements in the profile of  $Indiv$  using the formula (9).

$$P_{Temp}^{Comb}(e, Indiv) = \frac{\sum P_{Temp}^{Comb}(e)}{\sum_{f \in E(Indiv)} P_{Temp}^{Comb}(f)} \quad (9)$$

$E(Indiv)$  is the set of extracted elements from the tags shared by the individual  $Indiv$ .

Step 3: For each  $Indiv$  profile, compute the structural-semantic-temporal weight of each extracted element  $e$ . The weight of an element  $e$  is the combination of its structural score and its semantic score. The structural score applied to an element  $e$  is the centrality value of  $Indiv$  in the egocentric network.

$$P_{struct}(e, Indiv) = centrality\_score(Indiv, SetIndiv) \quad (10)$$

$SetIndiv$  represents the set of individuals of the egocentric network of  $u$ .

The semantic score of an element  $e$  is computed by dividing the score  $P_{Temp}^{Comb}(e, Indiv)$  of  $e$ , previously calculated in the second step by the formula (9), by the overall temporal score of  $e$  found in profile of others individuals ( $\{P_{Temp}^{Comb}(e, indiv_i)\} | indiv_i \in SetIndiv$ ). We computed this weight using the formula (11).

$$P_{sem}(e, Indiv) = \frac{\sum P_{Temp}^{Comb}(e, Indiv)}{\sum_{a \in SetIndiv} P_{Temp}^{Comb}(e, a)} \quad (11)$$

Then, we linearly combine the structural and the semantic score with a parameter  $\alpha$  as follows (12):

$$P_{SemStruct}(e, Indiv) = combination(P_{struct}(e), P_{sem}(e, Indiv), \alpha) \quad (12)$$

Step 4: Derive the extracted elements (interests) from each individual *Indiv* profile into the social profile of the user. In this step, we combine the extracted elements from all *Indiv* profiles according to the weights computed in the step 3. At the end of step 3, a given element  $e$  may have different weights for different *Indiv* profiles in the user egocentric network. In order to obtain a single weight for the element  $e$ , the function CombMN (proposed by [25]), is adopted to combine the different weights of  $e$  from different individuals *Indiv*. Finally, we return the social profile as a vector of weighted user interests.

## 2) Time aware community based social profiling process(CoBSPT)

As done for IBSPT, we incorporate our time aware method in the interest calculation of the community based approach, CoBSP [2]. The difference of the two approaches lies on how we apply the time score in the process. In fact, in CoBSPT, we apply the time score on the profile of each extracted community instead of each *Indiv* profile as done in IBSPT. The CoBSPT consists in four steps.

Step 1: Extract the communities from the user egocentric network. This step applies the iLCD algorithm [26], which performs very well with overlapping communities.

Step 2: Compute the profile of each extracted community. For each community  $C_i$ , we extract the elements  $e$  from the tags of all members. We apply the temporal score calculated by using the formula (8) to compute the score of the extracted elements using the same principle as used in the step 2 of the IBSPT process (section IV.C.1). When all elements are computed, we aggregate them by dividing their weights by the total weight of all founded elements in the profile of the community  $C_i$ .

$$P_{Temp}^{Comb}(e, C_i) = \frac{\sum_{f \in E(C_i)} P_{Temp}^{Comb}(e, f)}{\sum_{f \in E(C_i)} P_{Temp}^{Comb}(f)} \quad (13)$$

where  $E(C_i)$  is the set of extracted elements from the information shared by all individuals in the community  $C_i$ .

Step 3: Compute the weight of each element found in the profile of each community. The weight of an element  $e$  from a community  $C_i$  is the combination of the structural score of  $C_i$  and the semantic score of  $e$ . The structural score applied to an element  $e$  is the centrality value of  $C_i$  in the egocentric network compared to other communities.

$$P_{Struct}(e, C_i) = centrality\_score(C_i, C) \quad (14)$$

The semantic score of an element  $e$  from a community  $C_i$  is the importance weights of the score  $P_{Temp}(e, C_i)$  regarding the weight of the same element of the other communities (15):

$$P_{sem}(e, C_i) = \frac{P_{Temp}^{Comb}(e, C_i)}{\sum_{C_j \in C} P_{Temp}^{Comb}(e, C_j)} \quad (15)$$

The combination of the structural and the semantic scores is performed using the formula (16).

$$P_{SemStruct}(e, C_i) = combination(P_{struct}(e, C_i), P_{sem}(e, C_i), \alpha) \quad (16)$$

Step 4: Derive the extracted interests for each community into the social dimension (social profile) according to the weights computed in the third step. In order to obtain a single weight for the element  $e$ , we adopt the CombMN (proposed by [23]) to combine different weights from different communities.

## V. EXPERIMENTS

To validate our proposal, we compare the relevance of our time-aware method against that of the existing time-agnostic technique in both CoBSP and IBSPT as presented in figure 4.

The social profiles generated by CoBSPT and IBSPT represent respectively the community and individual time-aware construction process. The strategy is to find out which one provides a social profile the closest to the real user profile which is computed by using the user own tags.

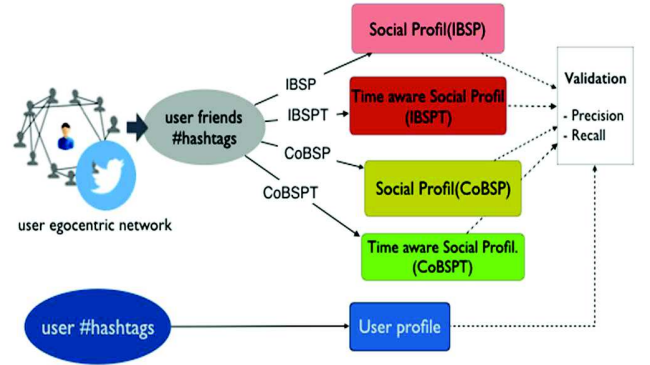


Fig.4 Validation process

### A. Dataset description

Our dataset contains 90 users. As previously mentioned, almost all of the crawled users are among the most popular accounts from different topics (social network analysis, technology, video game, sport, and travel). The users have between 80 and 200 followings<sup>1</sup>. The average following number of the studied user is 135. Overall, we crawled 10087 twitter users. Due to the rate limit of the API, we can crawl up to 3200 most recent tweets of each user. The collected tweets from user timelines are dated between February 2009 and January 2016. Finally, we extracted from all crawled tweets, 53089 tweets that contain at least a hashtag.

### B. Case study

#### 1) Ground truth: extraction of the real user profiles

To build user profiles as a ground truth (that is individual user profiles), we analyze keywords in the list of hashtags posted by the user. We extract the interests from the collected hashtags using a tag-mining engine: we first use dictionaries/thesaurus to merge tags having the same meaning.

Then, we compute the interests weights using a time-sensitive strategy proposed by [14] that prove their

<sup>1</sup> Our dataset will be made available online and can be requested by others researchers at wish.



effectiveness compared to a non-time-sensitive baseline strategy in the case of Twitter.

In the work of [14], the profile of a user  $u$  is a set of weighted concepts where a concept may be represented via a named entity or a hashtag. In this paper, we consider only the hashtags as user interests. The profile of a user  $u$  is represented as follows:

$$P(u, time) = \{(h, w(h, time, T_{tweets,u})) | h \in CH\} \quad (17)$$

$w(h, time, T_{tweets,u})$  is a time weight function that computes the weight associated with the hashtag for the given user  $u$  based on the messages published by  $u$  and the given timestamp.  $T_{tweets,u,h}$  denotes the set of tweets that have been published by  $u$  and refer to the concept  $h$ .  $CH$  represents the set of hashtags. This function dampens the occurrence frequency according to the temporal distance between the hashtag occurrence time and the given timestamp based on the normalized time calculation in the formula (18):

$$normalized_{time}(tw, h) = \frac{|time - time(tw)|}{Max_{time} - Min_{time}} \quad (18)$$

For a given tweet  $tw$  containing the hashtag  $h$ ,  $time(tw)$  represents its timestamp.  $max_{time}$  and  $min_{time}$  denote the highest (most recent) and lowest (earliest) timestamp of a tweet in  $T_{tweets,u,h}$ . The final time-weight score of a hashtag is computed as follows:

$$w(h, time, T_{tweets,u}) = \sum_{t \in T_{tweets,u,h}} (1 - normalized_{time}(tw, h))^d \quad (19)$$

The parameter  $d$  is used to adjust the influence of the temporal distance. The higher  $d$  is set, the less important the old hashtags is. The value of  $d$  is set to 4 in the work of [14].

We have slightly modified the equation (19) to apply it on our context. In fact, as we consider only the hashtags to build user profiles, there can be hashtags that occur only once in the user timeline. In this case, the range between  $max_{time}$  and  $min_{time}(max_{time} - min_{time})$  becomes 0, which returns an infinite value. Furthermore, in case of hashtags posted in short time periods, the range between  $max_{time}$  and  $min_{time}$  can be very low. This value could also be lower than the range between time and  $time(t)$  ( $time - time(t)$ ). So, the value of  $normalized_{time}$  can be  $> 1$ . Hence,  $1 - normalized_{time}$  can be negative. Consequently, the final score would make no sense. So, we modify the  $normalized_{time}$  calculation as follows (20):

$$normalized'_{time} = \frac{|time - time(t)| + 1}{(Max_{time} - Min_{time}) + 1} \quad (20)$$

We also normalized the value of the calculated  $normalized'_{time}$  between 0 and 1, as follows (21):

$$w(h, time, T_{tweets,u}) = \sum_{t \in T_{tweets,u,h}} (1 - \frac{normalized'_{time}}{Max(NORM)})^d \quad (21)$$

$NORM$  represents the set of  $normalized'_{time}$  score of all hashtags found on the user tweets.

## 2) Social profiles construction and parametric study

To build social profiles for each approach, interests are detected by mining hashtags of individuals or communities according to the algorithm (CoBSP or IBSP) used to derive the social profile presented in the section IV. We gather the followings of each user to build his/her egocentric network.

For each alter, we analyze the collected hashtags to extract the meaningful keywords using the text-mining engine as presented in the previous section. We apply the time aware method to both individual and community based approaches as previously described in the section IV.C.1 and IV.C.2.

For both individual and community based approaches, we build the social profiles with a parametric study in order to infer the suitable values for parameters in the formulas used in the section III and IV. We remind that:

- $\lambda$  represents the time decay rate of the temporal function presented on the formula (1).
- $\gamma$ , presented in the formula (8), represents the proportion of the individual temporal score compared to the information temporal score when computing a temporal score of an element  $e$  for an *Indiv*.
- $\alpha$  represents the proportion of the structural score compared to the semantic score as presented in the formula (12) and (16).

Furthermore, our aim is to find out the impact of each parameter to the time-aware social profile building process.

## 3) Evaluation

The experiments are made for different combinations of values of each parameter between 0 and 1:  $\lambda, \gamma, \alpha \in \{0.0, 0.001, 0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 0.95, 1.0\}$ . Note that we can build the social profiles of the time-agnostic approaches (CoBSP, IBSP) by fixing the value of  $\gamma = 0.0$  and  $\lambda = 0.0$ , and the social profiles of our time-aware approach with the different combination of the parameters.

To evaluate the relevance of each social profile regarding the user profile (ground truth, see section V.B), we use the precision and the recall measures. The precision assesses the proportion between the relevant found interests and the total number of found interests. The precision formula is presented as follows (22):

$$precision = \frac{Nb.Interests(Social Profile \cap User Profile)}{Nb.Interests(Social Profile)} \quad (22)$$

The recall assesses the proportion of relevant founded interests compared to the total number of real relevant interests (user profile). In our experimentation context, the recall formula is presented as follows (23):

$$recall = \frac{Nb.Interests(Social Profile \cap User Profile)}{Nb.Interests(User Profile)} \quad (23)$$

## C. Results

This section presents the results of our experiments by comparing the effectiveness of the existing time-agnostic social profiling processes CoBSP and IBSP with the proposed time-aware social profiling processes CoBSPT and IBSP. The use of hashtags may lead to very long lists of used hashtags (often more than 200) meanwhile only some very relevant interests are useful for user profiling and adaptation purposes. Therefore, we only consider the most relevant interests in the social profile to compute the precision and recall (the  $n$  top of all interests sorted in descending order). Note that all following results shown in this section are computed by using the top 5 interests.

In all results presented hereafter, we will compare CoBSP, IBSP against our CoBSPT and IBSPT results with respect to  $\lambda$ ,  $\gamma$  and  $\alpha$ . Nevertheless,  $\lambda$ ,  $\gamma$  are not implied in CoBSP and IBSP ( $\lambda=\gamma=0$ ). Thus, CoBSP and IBSP results will never vary whatever the values of  $\lambda$  and  $\gamma$ . Furthermore, we note that for the top 5 interests, there is a very low difference between the results of precision and recall. Therefore, we present the results solely in term of precision since we are interested in top relevant interests.

1) *Global results*

We first present the results of our parametric study for all 90 users. The global results for each approach are presented by the average of the precision for all users.

a) *Social profiling process comparison*

The figure 5 shown below presents the comparison of the best precision of the social profiles built by the CoBSP, CoBSPT, IBSP and IBSPT processes.

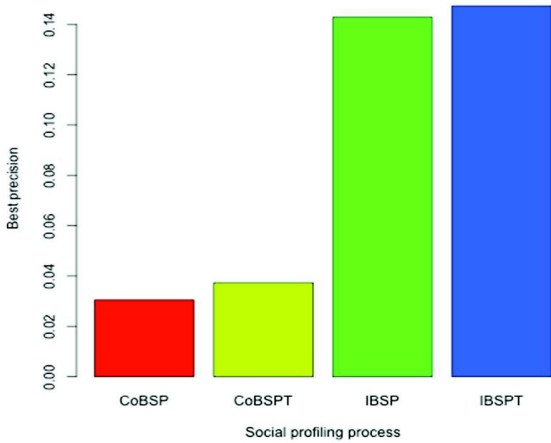


Fig.5 Comparison of the best precision results from each technique (CoBSP, CoBSPT, IBSP and IBSPT)

For the CoBSP, the best precision (0,0305) is observed when  $\alpha$  equals to 0.95. For our algorithm CoBSPT, the best precision (0,0373) is observed when the parameters  $\alpha$  equals to 0.95,  $\gamma \in \{0.1,0.25,0.5,0.75\}$  and  $\lambda$  equals to 0.01.

For the IBSP algorithm, the best precision (0,1428) is observed when  $\alpha$  equals to 0.95. For our algorithm IBSPT, the best precision (0,1473) is observed when the parameters  $\alpha \in \{0.1,0.25\}$ ,  $\gamma \in [0; 0.25]$  and  $\lambda$  equals to 0.01.

For the community based approach, we found that the best result of CoBSPT improves the result of CoBSP of 22% in terms of precision and recall. This improvement demonstrates the effectiveness of our temporal method compared to the time-agnostic method on the community based user profiling process. However, the precision and recall of the two algorithms are very low compared to precision and recall of IBSP and IBSPT algorithms. As shown in the figure 5, the results of both individual based algorithms clearly improve the results of the community based algorithms in terms of precision and recall whatever the value of  $\alpha$ ,  $\gamma$ ,  $\lambda$  is.

With regard to these results, we can conclude that the individual based approach performs better than the community based approach on Twitter. This can be explained by the fact that in the community based approach, the algorithm adopted to extract the communities from user egocentric network is based on the relationships between the alters. However, following relation is not often considered as a relevant factor to extract communities on Twitter, as argued in [15], [24]. This can lead to irrelevant extracted communities and a misinterpretation of the interests on user social profile.

Thus, hereafter, we mainly focus on the results of the individual based approach algorithms (IBSP and IBSPT).

b) *Individual based social profiling processes comparison*

We first studied the improvement of our time aware process IBSPT against the time-agnostic process IBSP.

We first analyzed the results of the IBSPT process with respect to the value of  $\gamma$ . We can observe that, to obtain the best results, the value of  $\gamma$  must be low ( $\gamma \in [0; 0.25]$ ). Hence, the value of  $\gamma$  has quite weak impact over the relevance of our time-aware method for the individual based approach. Based on this observation, we hypothesize that the relationships dynamic has no or not much impact over the social profiling process on Twitter. Thus, hereafter, we analyze all results by fixing the value of  $\gamma$  to 0.0.

We then analyze the results with respect to the value of  $\lambda$ . The figure 6 presents the comparison of the best precision of the social profiles built by the IBSP and IBSPT processes.

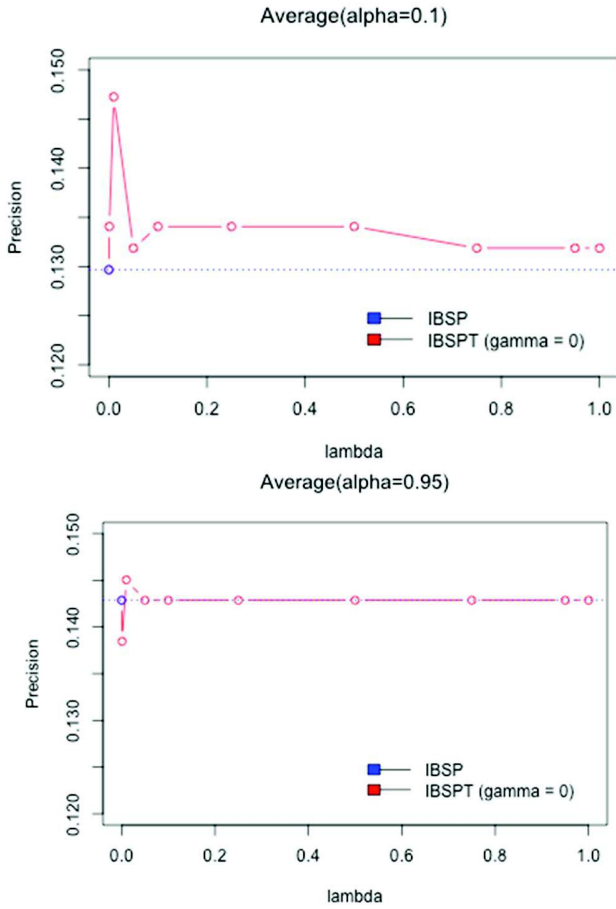


Fig. 6 (Top) overall comparison of the IBSPT and IBSP with  $\alpha = 0.1$ . (Bottom) overall comparison of the IBPT and IBSP with  $\alpha = 0.95$ . The value of  $\gamma$  is set to 0.00 for all graphs.

In the figure 6, the red curve represents the best precision of the IBSPT process using different value of  $\lambda$ . The blue curve represents the best precision of the IBSP process and thus does not vary whatever the values of  $\lambda$ .

For the top graph, the precision of the IBSPT process is computed by fixing  $\gamma = 0.0$  and  $\alpha = 0.1$  since we obtain the best precision for the IBSPT with this combination. For the bottom graph, the precision of the IBSPT process is computed by  $\alpha = 0.95$ .

We can see that the result of IBSPT improve the result of the IBSP algorithm whatever the value of  $\lambda$  when  $\alpha = 0.1$ . For  $\alpha = 0.95$ , the IBSPT outperforms the IBSP when  $\lambda = 0.01$ . The best results can always be found when  $\lambda = 0.01$

The IBSPT method improves the IBSP of respectively 6.45% and 1.5% when  $\alpha = 0.1$  and 0.95. This improvement demonstrates the effectiveness of our temporal method compared to the time-agnostic method, on the individual based user profiling process. However, the improvements of 6,45 % and 1.5% remain low.

We observed that the best results of the IBSP process is obtained when the value of  $\alpha$  is very high (0.95) and the best results of the IBSPT process is obtained and can improve the result of the IBSP when the value of  $\alpha$  is quite low ( $\{0.1, 0.25\}$ ). This observation presented in the figure 7 shows the comparison of the best precision of the social profiles built by the IBSP and IBSPT varying with different values of  $\alpha$ . The precision of the IBSPT process is computed by fixing  $\gamma = 0.0$  and  $\lambda = 0.01$  (as we obtain the best precision for the IBSPT with this combination).

We remind that  $\alpha$  is the ratio between the structural score (centrality score of individuals), which does not depend on the temporal factor and the semantical score, calculated for IBSPT using the temporal score. The higher the value of  $\alpha$  is, the higher the structural score becomes.

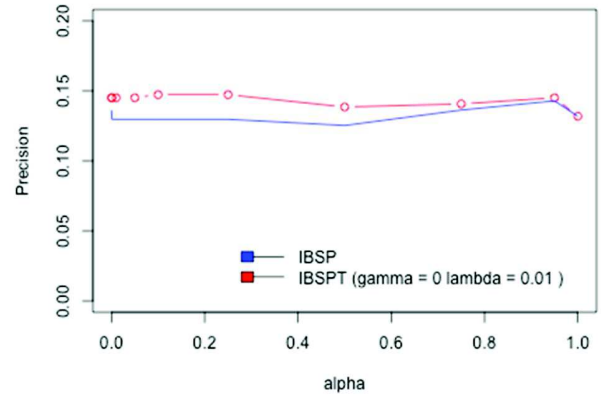


Fig. 7 Overall comparison of the IBSPT and IBSP, varying with  $\alpha$  ( $\gamma = 0.0$  and  $\lambda = 0.01$  for the IBSPT)

This observation shows that, in our context, the temporal score is not the only factor of the social profiles performance. The results depend also on the centrality score of the individuals. This reinforces the interest of applying the proposed time-aware method against the time-agnostic method when the centrality score of the individuals has not much impact on the social profiling process. This may be helpful in case of sparse egocentric networks in which the members have very few connections between them.

## 2) Results by network density

The interaction of the users in Twitter has a low density and this characteristic could impact the performance of our method and its improvement on the time-agnostic method. In fact, our method relies on the timestamp of the interaction between the users and his/her alters (followings) but also on the timestamp interaction among his/her followings (cf. formula 3-6). We remind that in our context, we consider replies, mentions and retweets as user interactions. Thus, our time-aware method may not have much impact on the users whose egocentric network contains few interactions between the nodes and vice versa. Furthermore, as previously mentioned, the social network density may have also some impact over the importance of the structural score and temporal score on the social profiling process.

We then study the results of the two techniques regarding activities and links density on user egocentric network. We divided our dataset into three different corpus based on the user egocentric network density. The density of the egocentric network  $G$  of a given user  $u$  is computed by using the formula (24) as follows:

$$density(G, u) = \frac{nb_{relationship}(G)}{nb_{friends}(u) * (nb_{friends}(u) - 1)} \quad (24)$$

$nb_{relationship}$  represents the number of relationships between alters in  $G$ .  $nb_{friends}$  represents the number of alters (neighbors of the user) in  $G$ .

*a) Results for the egocentric networks having 5% of density*

The first corpus contains 41 users having less than 5% of density in their egocentric network. For this corpus, the best results (0.160) of the IBSP can be observed when  $\alpha$  is 0.95. The best results (0.175) of the IBSPT can be observed when  $\alpha \in [0.0; 0.25]$ ,  $\gamma \in [0.0; 0.5]$  and  $\lambda = 0.01$ . With this setting, the IBSPT outperforms the IBSP with 9,09% of improvement. We can see that on this corpus, we can improve the results of the time-agnostic method when the value of  $\alpha$  is low. The figure 8 presents the comparison of the best precision of the social profiles built by the IBSP and IBSPT processes for this corpus. In this figure, the blue and red curves represent respectively the best precision of the IBSP and IBSPT processes using different values of  $\alpha$ . The precision of the IBSPT process is computed by fixing  $\gamma = 0.0$  and  $\lambda = 0.01$  (as we obtain the best precision for the IBSPT with this combination).

This corpus represents a very sparse egocentric network. Since alters have very few relationships between them, the range of centrality degree between each alter can be low. Hence, the structural score calculated based on the centrality degree may have less impact over the profiling process. The semantical score based on the temporal score becomes more important and influences the social profiling process. Consequently, in network having such low density, our time-aware method can improve the results of the time-agnostic one.

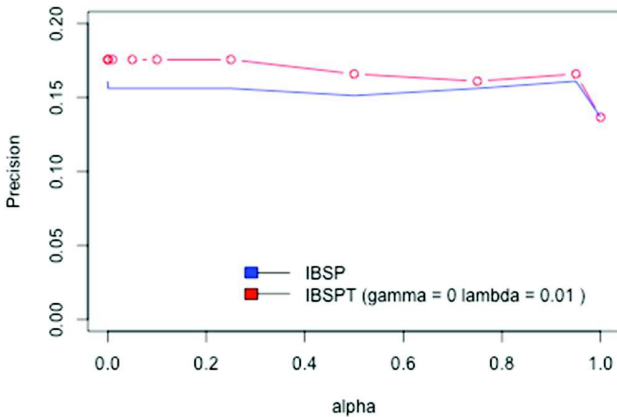


Fig. 8 Comparison of the IBSPT and IBSP for the egocentric network having less than 5% of density, varying with  $\alpha$  ( $\gamma = 0.0$  and  $\lambda = 0.01$  for the IBSPT)

*b) Results for the egocentric networks having between 5% and 10% of density*

The second corpus contains 33 users having between 5% and 10% of density in their egocentric network. The best result of IBSP (0.1211) can be observed when  $\alpha \in \{0.75, 0.95, 1\}$ . The best results (0.1211) of the IBSPT can be observed for all value of  $\gamma$  and  $\lambda$  when  $\alpha \in \{0.75, 0.95, 1\}$ .

The figure 9 shown below presents the comparison of the best precision of the social profiles built by the IBSP and IBSPT processes for this corpus. In this figure, the blue and red curves represent respectively the best precision of the IBSP and IBSPT processes using different values of  $\alpha$ . The precision of the IBSPT process is computed by fixing  $\gamma = 0.0$  and  $\lambda = 0.01$  (as we obtain the best precision for the IBSPT with this combination).

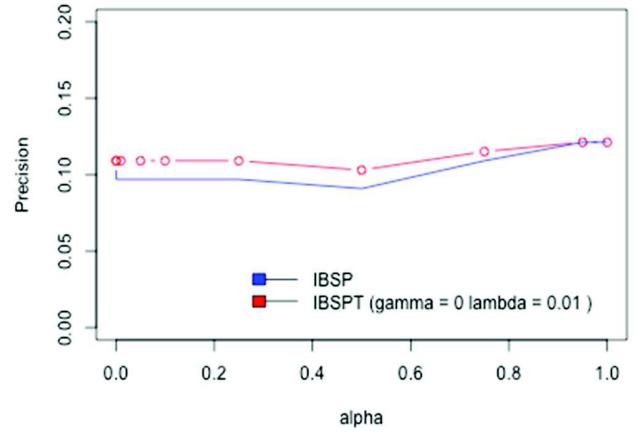


Fig. 9 Comparison of the IBSPT and IBSP for the egocentric network having between 5% and 10% of density, varying with  $\alpha$  ( $\gamma = 0.0$  and  $\lambda = 0.01$  for the IBSPT)

When we observe the results regarding the value of  $\alpha$ , our method outperforms the time-agnostic method when  $\alpha$  is lower than 0.75. The IBSPT gives the same best result as the IBSP when  $\alpha$  is higher than 0.75. We can see that the best results of both techniques are obtained when  $\alpha$  is set to be very high. For such setting, the temporal score cannot help to improve the relevance of the social profile. That means that on this kind of network, the relevance of the social profile is more impacted by the centrality score of the individuals than by the temporal score of information.

*c) Results for the egocentric networks having between 10% and 22% of density*

The third corpus contains the 17 users having between 10% and 22% of density in their egocentric network. The best results (0.1411) of the IBSP is observed when  $\alpha \in \{0.75, 0.95, 1\}$ . The best results (0.153) of the IBSPT can be observed for the set of ( $\alpha \in \{0.1, 0.25\}$ ,  $\lambda = 0.01$ ,  $\gamma \in \{0.0, 0.001, 0.01, 0.05, 0.1\}$ ) and ( $\alpha = 0.75$ ,  $\lambda \in \{0.75, 0.95, 1\}$ ,  $\gamma \in \{0.0, 0.001, 0.01, 0.05\}$ ). With both settings, the IBSPT outperforms the IBSP with 8,3% of improvement. We can see that the best results for the IBSP can still be observed when the value of  $\alpha$  is set to a high value. To obtain the best results for

the IBSPT, the value of  $\alpha$  and  $\lambda$  must be either very high or very low. The value of  $\gamma$  must be very low.

The figure 10 presents the comparison of the best precision of the social profiles built by the IBSP and IBSPT processes for this corpus.

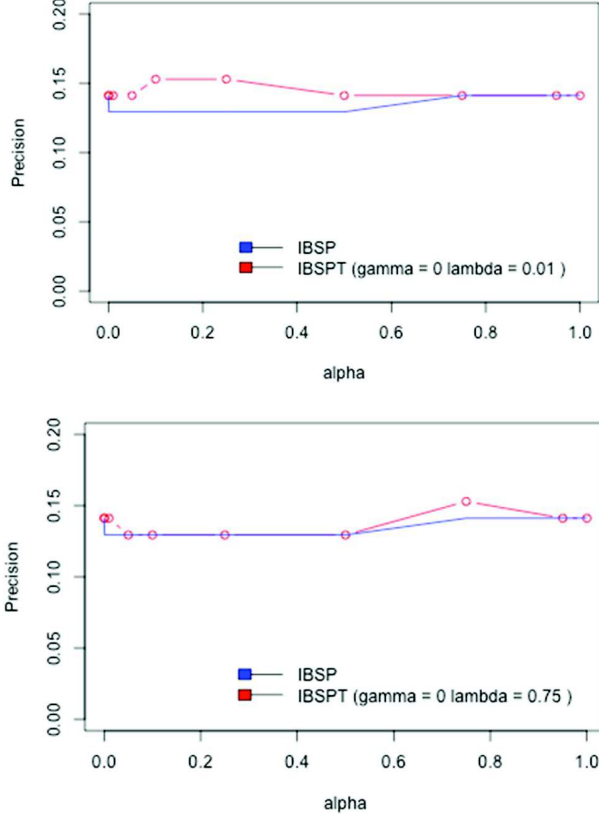


Fig. 10 Comparison of the IBSPT and IBSP for the egocentric network having between 10% and 22% of density, varying with  $\alpha$ . (Top): the results obtained by fixing  $\gamma = 0.0$  and  $\lambda = 0.01$  for the IBSPT. (Bottom): the results obtained by fixing  $\gamma = 0.0$  and  $\lambda = 0.75$  for the IBSPT.

In the figure 10, the blue and red curves represent respectively the best precision of the IBSP and IBSPT processes using different values of  $\alpha$ . Since we obtain the best results by fixing two different intervals of  $\lambda$ , we present here two figures for the results of this corpus. In the figure 10 (top), the precision of the IBSPT process is computed by fixing  $\gamma = 0.0$  and  $\lambda = 0.01$ . In the figure 10 (bottom) the precision of the IBSPT process is computed by fixing  $\gamma = 0.0$  and  $\lambda = 0.75$ .

This corpus represents a rather dense egocentric network (for twitter context) in which the alters have more relation between them. The centrality range in such network can be high. So, the structural score based on the centrality degree may have more impact over the profiling process. However, there is more chance that the alters interact between them and have more of interactions and activities on the network. So the temporal score can still influence the profiling process although the centrality degree can have a lot of impact in such network.

We have shown that the relevance of our proposed method is impacted by the network density. Our method performs well on a very sparse network, which represents a general characteristic of user egocentric networks on Twitter. As Twitter has a very specific characteristic and contains various users and information characteristic, there can also be the other factors that can influence the effectiveness of the social profiling process. We need more studies to infer the different factors and the best combination of these factors in order to obtain a more effective social profiling process.

## VI. CONCLUSION AND PERSPECTIVES

In this paper, we proposed to apply our existing time-aware social profiling method on Twitter. With the particular social network characteristics of Twitter, we obtain different results than those obtained on co-authorship networks. We found that the community based social profiling approach, which relies on the relationships between egocentric members, does not perform well in this kind of social networks. Therefore, in a short term, we are going to consider other communities extraction techniques that are based on other kinds of relationships. We will particularly focus on the relationships that are based on sharing information behavior (e.g. sharing common hashtags, retweeting the same post, etc.).

For the individual based social profiling approach, our method can slightly improve the time-agnostic profiling method. This demonstrates the interest of taking into account the proposed time-aware method on the social profiling process. We found that the relationships dynamic has weak impact over the social profiling process on Twitter compared to the information dynamic. However, to obtain more improvement compared to the time-agnostic method in this kind of network, we need to improve our time-aware method in a future work to find out algorithms or temporal score calculation techniques that fit more with this specific kind of network. We will also consider other link prediction algorithms and other time-weight functions to enhance the performance of our approach.

Furthermore, we observed that our method gives relevant results in terms of precision and recall on very sparse networks or rather dense networks. This observation needs also more studies to find out better explanations for these results. We also want to find out other factors that can influence the social profiling process to improve it.

Finally, we have shown in this paper that our time-aware social profiling method performs differently regarding the characteristics of the studied egocentric network. Therefore, our long-term perspective consists in the proposal of a generic platform that extracts the information and builds the user social profile according to the type and the specific characteristics of the underlying social network. Such a platform would be parameterized by the characteristics of the targeted social network using a machine learning approach.

## ACKNOWLEDGEMENTS

This presentation was subsidized by the Pyrenean Working Community and the Region Midi-Pyrenées (Toulouse, France).



## REFERENCES

- [1] M.-F. Canut, S. On-At, A. Péninou, and F. Sèdes, "Time-aware Egocentric Network-based User Profiling," in Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, New York, NY, USA, 2015, pp. 569–572.
- [2] D. Tchunte, M.-F. Canut, N. Jessel, A. Peninou, and F. Sèdes, "A community-based algorithm for deriving users' profiles from egocentric networks: experiment on Facebook and DBLP," *Soc. Netw. Anal. Min.*, vol. 3, no. 3, pp. 667–683, Sep. 2013.
- [3] J. H. Kietzmann, K. Hermkens, I. P. McCarthy, and B. S. Silvestre, "Social media? Get serious! Understanding the functional building blocks of social media," *Bus. Horiz.*, vol. 54, no. 3, pp. 241–251, May 2011.
- [4] J. Yang and J. Leskovec, "Patterns of Temporal Variation in Online Media," in Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, New York, NY, USA, 2011, pp. 177–186.
- [5] E. Gilbert and K. Karahalios, "Predicting Tie Strength with Social Media," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2009, pp. 211–220.
- [6] J. Chen, R. Nairn, and E. Chi, "Speak Little and Well: Recommending Conversations in Online Social Streams," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2011, pp. 217–226.
- [7] S. Gauch, M. Speretta, A. Chandramouli, and A. Micarelli, "User Profiles for Personalized Information Access," in *The Adaptive Web*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Springer Berlin Heidelberg, 2007, pp. 54–89.
- [8] S. Aral and D. Walker, "Tie Strength, Embeddedness & Social Influence: Evidence from a Large Scale Networked Experiment," *Social Science Research Network*, Rochester, NY, SSRN Scholarly Paper ID 2197972, Jan. 2013.
- [9] G. Cabanac, "Accuracy of Inter-researcher Similarity Measures Based on Topical and Social Clues," *Scientometrics*, vol. 87, no. 3, pp. 597–620, Jun. 2011.
- [10] D. Carmel, N. Zwerdling, I. Guy, S. Ofek-Koifman, N. Har'el, I. Ronen, E. Uziel, S. Yogev, and S. Chernov, "Personalized Social Search Based on the User's Social Network," in Proceedings of the 18th ACM Conference on Information and Knowledge Management, New York, NY, USA, 2009, pp. 1227–1236.
- [11] E. Stattner, M. Collard, and N. Vidot, "D2SNet: Dynamics of diffusion and dynamic human behaviour in social networks," *Comput. Hum. Behav.*, vol. 29, no. 2, pp. 496–509, Mar. 2013.
- [12] L. Weng, J. Ratkiewicz, N. Perra, B. Gonçalves, C. Castillo, F. Bonchi, R. Schifanella, F. Menczer, and A. Flammini, "The Role of Information Diffusion in the Evolution of Social Networks," in Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2013, pp. 356–364.
- [13] F. Abel, Q. Gao, G.-J. Houben, and K. Tao, "Twitter-based User Modeling for News Recommendations," in Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, Beijing, China, 2013, pp. 2962–2966.
- [14] F. Abel, Q. Gao, G.-J. Houben, and K. Tao, "Analyzing Temporal Dynamics in Twitter Profiles for Personalized Recommendations in the Social Web," in Proceedings of the 3rd International Web Science Conference, New York, NY, USA, 2011, pp. 2:1–2:8.
- [15] H. Liang, Y. Xu, D. Tjondronegoro, and P. Christen, "Time-aware Topic Recommendation Based on Micro-blogs," in Proceedings of the 21st ACM International Conference on Information and Knowledge Management, New York, NY, USA, 2012, pp. 1657–1661.
- [16] N. Zheng and Q. Li, "A recommender system based on tag and time information for social tagging systems," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 4575–4587, Apr. 2011.
- [17] Y. Cheng, G. Qiu, J. Bu, K. Liu, Y. Han, C. Wang, and C. Chen, "Model Bloggers' Interests Based on Forgetting Mechanism," in Proceedings of the 17th International Conference on World Wide Web, New York, NY, USA, 2008, pp. 1129–1130.
- [18] L. Li, L. Zheng, F. Yang, and T. Li, "Modeling and broadening temporal user interest in personalized news recommendation," *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3168–3177, Jun. 2014.
- [19] Y. Ding, E. K. Jacob, J. Caverlee, M. Fried, and Z. Zhang, "Profiling social networks: A social tagging perspective," *-Lib Mag.*, vol. 15, no. 3/4, Mar. 2009.
- [20] T. Tylenda, R. Angelova, and S. Bedathur, "Towards Time-aware Link Prediction in Evolving Social Networks," in Proceedings of the 3rd Workshop on Social Network Mining and Analysis, New York, NY, USA, 2009, pp. 9:1–9:10.
- [21] D. Liben-Nowell and J. Kleinberg, "The Link Prediction Problem for Social Networks," in Proceedings of the Twelfth International Conference on Information and Knowledge Management, New York, NY, USA, 2003, pp. 556–559.
- [22] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a Social Network or a News Media?," in Proceedings of the 19th International Conference on World Wide Web, New York, NY, USA, 2010, pp. 591–600.
- [23] M. Mezghani, C. A. Zayani, I. Amous, and F. Gargouri, "A User Profile Modelling Using Social Annotations: A Survey," in Proceedings of the 21st International Conference on World Wide Web, New York, NY, USA, 2012, pp. 969–976.
- [24] E. A. Leicht and M. E. J. Newman, "Community structure in directed networks," *Phys. Rev. Lett.*, vol. 100, no. 11, p. 118703, Mar. 2008.
- [25] J. A. Shaw, E. A. Fox, J. A. Shaw, and E. A. Fox, "Combination of Multiple Searches," in *The Second Text Retrieval Conference (TREC-2)*, 1994, pp. 243–252.
- [26] R. Cazabet, M. Leguistin, and F. Amblard, "Automated Community Detection on Social Networks: Useful? Efficient? Asking the Users," in Proceedings of the 4th International Workshop on Web Intelligence; Communities, New York, NY, USA, 2012, pp. 6:1–6:8.