

# A Parametric Study to Construct Time-Aware Social Profiles

Sirinya On-at, Arnaud Quirin, André Péninou, Nadine Baptiste-Jessel, Marie-Françoise Canut, and Florence Sèdes

## 1 Introduction

In information systems, user profile is an essential element for adaptive information mechanisms (e.g., personalization, information access, recommendation). These mechanisms rely on users' profiles to propose relevant content according to the user specific needs. Various models of user profile have been proposed (e.g., vector of weighted keywords, vector of weighted topics, semantic network, etc.), as presented in [1]. In this work, we model the user profile by using the keyword-based approach. Each keyword represents user's interest to which we associate a numerical weight according to its importance in the profile.

Users' profiles can be built by using the user's own information or by gathering information from external sources such as user's social network. In this paper, we build user profile by extracting user's interests from the information and relationships of the user social network, so-called social profile. We focus on user "egocentric network" which only considers the user's directed contacts. Social network-based user profiling is useful, on one hand, to provide additional information that can improve the performance of existing user profile and, on the other hand, to overcome the cold start problem or to complete non-existing/missing profile of inactive users.

We observe that in Online Social Networks (OSN), users are encouraged to contribute, broadcast contents, and connect to those who share the same interests and/or activities. As the behavior of online user evolves quickly over time, the

---

S. On-at (✉) • A. Quirin • A. Péninou (✉) • N. Baptiste-Jessel • M.-F. Canut • F. Sèdes  
Toulouse Institute of Computer Science Research (IRIT), University of Toulouse, CNRS, INPT,  
UPS, UT1, UT2J, 118 route de narbonne, 31062 Toulouse, Cedex 9, France  
e-mail: [sirinya.on-at@irit.fr](mailto:sirinya.on-at@irit.fr); [andre.peninou@irit.fr](mailto:andre.peninou@irit.fr)

information and relationships in an OSN can rapidly become obsolete [2, 3]. Thus, building a social profile from this kind of data without taking into account its evolving characteristic may lead to misinterpretation of user long-term interests.

As a motivating example, given a user  $u$  is not football fan but supports occasionally his national team during the world cup, the well-known sport event that takes place every 4 years. He follows the official account of his favorite team and/or players in social media (e.g., Facebook, Twitter), to follow updated information about this team and the players. After the world cup, the information shared from these accounts become less meaningful for him. If this fact is not considered, we may find “football,” “world cup,” “France national team” as interests in his profile for a long time. As a result, the adaptive mechanisms that exploit this profile (e.g., personalized news recommendation) will continue to propose him the up-to-date news about football while this information becomes actually useless to him. This example can show that: for a user, his/her relationships between his/her social networks friends may drift overtime. The link between two users can be relevant for a period of time and may become less meaningful for a later period of time. Furthermore, information sharing behavior in social network has a dynamic and non-persistent characteristic. Some information shared by a user may not always reflect his/her interests for a long-time period.

To overcome this problem, we focus on taking into account the evolution of user’s interests in social network-based user profiling process in order to build the more relevant and up-to-date user profile. We try to answer to the following questions: how to select the relevant individuals in the user social network as meaningful information sources? And then how to select relevant information from these selected individuals to extract the relevant and fresh interests to build the user social profile?

In our previous work [4], published in Asonam’ 2015, we applied a time-aware method to improve a time-agnostic approach previously proposed in our team. We integrated a “temporal score” to each interest according to its relevance and freshness. The time-aware weight of each interest is computed by combining two time-aware weights: (1) the time-aware individual weight that weights the relevance of the information sources (individuals), computed by applying a time-aware link prediction technique; (2) the time-aware semantic weight that weights the relevance of information used to extract the user’s interests. This weight is computed by applying a term frequency method and a time-weighted function.

In this paper, we extend this work in many ways. We conduct first a parametric study in order to find out the best value and the impact of different parameters used to calculate the temporal score. The parametric study shows the impact of the individual score compared to the information score in the temporal score calculation. We describe the improvements made in calculating the temporal score compared to the corresponding one of our previous work. At last, we validate our new method using (DBLP/Mendeley) dataset with providing new experiments showing the relevance and the compatibility of two data sources.

The rest of the paper is organized as follows. Section 2 describes the related work with presenting our previous user profiling method and existing techniques to

mine user interests' evolution. In Sect. 3, we present our time-aware social profiling method. In Sect. 4, we describe the results of various experiments conducted on two co-authorships networks: DBLP and Mendeley. Section 5 concludes and presents some future works.

## 2 Related Works

### 2.1 User Profile Building Process

In the literature, different user profiling techniques are discussed. According to [1], user's interests can be extracted from information explicitly given by users or explicitly gathered when they interact with the system. Several works [5–7] also proposed to extract the user's interests from their social networks in order to build his/her social profile. In that case, user's interests are extracted by using the information shared by his/her social network members. We are particularly interested by extracting user's interests from their egocentric network. An egocentric network is a specific social network which takes into account only direct connections to a user called "ego." An egocentric network is thus composed of individuals having a direct relationship with the "ego" user and relationships between these individuals. This decision is motivated by the work of [8] which assumes that a user tends to connect with the people who share common interests with him/her. To build a social profile, user's interests can be extracted either from individual people, as adopted in [5, 6], or by extracting this information from their communities [7]. The effectiveness of social profiles has been proved with empirical results [5–7]. Tchuente et al. [7] also showed the effectiveness of the proposed community-based approach compared to an individual-based approach. However, we observe that user's interests and social networks evolution have not been widely taken into consideration in the proposed social profile building process. So, in the next section, we provide an overview about existing techniques that deal with the user's interests evolution.

### 2.2 Incorporating Dynamic Interests in the Profile

User's interests evolution is considered as one of the concept drift examples [9–11]. "In supervised learning context, concept drift refers to an online supervised learning scenario when the relation between the input data and the target variable changes over time. The real concept drift refers to changes in the conditional distribution of the output (i.e., target variable) given the input (input features), while the distribution of the input may stay unchanged. A typical example of the real concept drift is a change in user's interests when following an online news stream. Whilst the distribution of the incoming news documents often remains the same, the conditional distribution of the interesting (and thus not interesting) news" [9].

In our work context, we are interested in extracting user's interests from his/her social network to build his/her social profile. Dynamic social network properties allow us to consider that user's relations and shared information within this kind of network cannot be considered relevant for a long-time period. To build an effective user social dimension in our work context, it is necessary to take into account the evolution characteristic of online social network. It means that we should prevent user's interests drift regarding the relevance of social relations in user social network and also the relevance of shared information in user social network.

We can distinguish two main axes for incorporating dynamic user interests. The first axis consists in taking into account dynamic characteristic of the information sources during the interest extraction process. The second axis consists in updating an existing profile after completing the construction process [12, 13]. In this paper, we focus on the first axis only to build a relevant and up-to-date social profile. Then, as a future work, we can extend the work to include the study of user profile updating process to maintain the relevant of the profile over time.

To deal with the user's interest evolution, we can apply different techniques used to handle the concept drift issue [10, 14]. The first technique is based on a time window or time forgetting technique which selects only the information from the latest time periods [15]. In the same approach, outdated information outside a time window is completely ignored. In the information retrieval literature, the same principle is applied to build a short-term user profile [16]. However, in some cases, the ignored information could eventually be valuable [17]. Thus, this might lead to the loss of useful knowledge.

The second technique, called instance weighting, consists in weighting different time periods according to their relevance. To perform this computation, time decay functions, which assign the higher weights to the most recent information, are widely used [18, 19]. This technique enables the use of all available information in a restricted way. In fact, the relevant information is selected by weighting their importance using some temporal factors. To extract user interests for a recommendation system, [20] apply an exponential temporal function to score the tags before using them. This concept is also adopted in personalized information retrieval literature. For example, [21] applied a temporal function to weight the user interests according to the freshness of the information sources.

The third technique relies on ensemble learning to generate a family of predictors [14]. Each predictor is weighted by its relevance according to the present time point (e.g., predictors that were more successful on recent instances get higher weights). Koren [14] modeled the evolution of user behavior during the whole time period for collaborative filtering, with showing the effectiveness of his contribution for movies recommendation. Unfortunately, this kind of techniques requires a training dataset to achieve relevant families of predictors.

In our case, user's interests are extracted from the information shared by the connected individuals in the user social network. To build an effective social profile, it is important to take into account the dynamic of social networks.

### 2.3 *Social Network Evolution*

In 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) [22, 23].

In the OSNs context, the advent of social media incites a rapid evolution of social networks in terms of network structure (users' relationships) and information flow [24]. Since links are quickly established in OSNs, two users creating a relationship are not required to know each other in real life. Thus, the links persistency is not always maintained in this case. Thus, the links persistency is not always an evidence of a real relationship. Arnaboldi et al. [3] analyzed the records of a Twitter communication to study the dynamics that govern the preservation of online social relationships and discovered that Twitter users tend to keep weak social relationships rather than strong ones, with a high turnover of contacts in their networks. In terms of information flow, social events or viral marketing (buzz) increase information sharing, which in turn enhance online social content sharing. Often, this social phenomenon occurs for a short period, then disappears and may reappear in another period. For example, during the World Cup 2014, 672 million tweets were posted related to the tournament. The Ice Bucket Challenge campaign that became viral on social media during July–August 2014 generated more than 2.2 million hashtag mentions on Twitter [25]. Thus, for a user, the relationships and information in his/her social network can drift overtime. Information can be relevant for him/her for some time and becomes obsolete later.

According to [26], the dynamic characteristic of social networks depends on users' behaviors. Some users are very active and tend to frequently share information. In this case, the relationships and/or information in their social networks quickly evolve. Conversely, some users are less active and their social networks change slowly. Moreover, as previously mentioned, users of social networks can also evolve chaotically regarding social events or buzz. Thus, to apply a time-aware approach, we should choose a technique which can fit to different types of social networks characteristics.

The time window technique can lead to lose meaningful information in case of gradually evolving networks. The ensemble learning technique shows good performance in terms of accuracy. However, this technique requires a training dataset as well as it is not appropriate for new or less active users who have poor or empty relationships or information in the network. Consequently, we adopt in our contribution an instance weighting technique which enables us to use all available information in a differentiated manner by applying temporal factors. We describe this contribution in the next section.

### 3 Proposition: Temporal Scores to Construct Social Profiles

In this section, we present our proposal of time-aware social profile building process. We first present the agnostic social profile building process (CoBSP) developed in our team [7]. Then, we introduce the temporal score that we integrate to the CoBSP in order to take into account user interests evolution. Finally, the algorithm is described showing the whole process.

A global view of the profiling process is presented as follows:

For a given user (ego), for whom we desire to build social profile, we use his egocentric networks as information sources to extract the interests and build then 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.

Based on the existing social dimension derivation process (CoBSP) proposed by Tchuente et al. [7], we integrate a temporal score to the interests weighting step of the social profile construction process. We first describe hereafter the CoBSP process followed by our proposal time-aware social profile construction process called CoBSPT.<sup>1</sup>

#### 3.1 Notations

In the social profile construction process, we denote  $u$  as the user (ego) for whom we desire to build social profile. 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  $V$ . We use the term individuals (SetIndiv) to represent the set of user's egocentric network members. For each individual  $\text{Indiv} \in \text{SetIndiv}$ , his shared information is called  $\text{Info}_{\text{Indiv}}$ . Each  $\text{Info}_{\text{Indiv}}$  contains elements called  $e$ , extracted by using classical text analysis techniques [7]. Note that an element  $e$  is a keyword that represents a user interest.

---

<sup>1</sup> $T$  for "Temporal".

### 3.2 Community-Based Social Profile Construction Process

The social dimension derivation process (CoBSP) proposed by [7] takes the following as inputs: (1) a given user  $u$  (ego), (2) the egocentric network of  $u$ , and (3) all information shared by all individuals in the egocentric network of  $u$ . The output is a social profile of  $u$  in the form of a vector of weighted elements (keywords). The algorithm consists of four steps described as follows:

- Step 1: Extraction of the communities from the user egocentric network. This step uses the iLCD algorithm [27] which performs very well with overlapping communities. The choice of algorithm is motivated by [7] where the iLCD is compared to the InfoMap algorithm [28], the CFinder algorithm [29], and the social cohesion-based algorithm [30].
- Step 2: Building the profile of each community found in the first step. The profile of a community is computed by analyzing the behavior of all members of this community. For each community  $c_i$ , we extract the elements  $e$  from the Info of all members. We use the term frequency tf measure to compute the score of the extracted elements, called  $S_{tf}$ , standing for tf score. The score of an element found in the community  $c_i$  is represented by  $S_{tf}(e, c_i)$ .
- Step 3: Computing the score of each element found in the profile of each community. The score of an element  $e$  from a community  $c_i$  is the combination of the structural score ( $S_{struct}$ ) of  $c_i$  and the semantic ( $S_{sem}$ ) score of  $e$ . The structural score (Eq. (1)) applied to an element  $e$  is the centrality value of  $c_i$  in the egocentric network compared to other communities.

$$S_{struct}(c_i) = \text{centrality\_score}(c_i, C) \quad (1)$$

The semantic score of an element  $e$  of a community  $c_i$  depends on the weight of this element for all members of this community. This score can be computed by using tf or tf-idf measure [31].

The combination of the structural and the semantic scores is performed using Eq. (2).

$$S_{Sem,Struct}(e, c_i) = \text{combination}(S_{struct}(e, c_i), S_{sem}(e, c_i), \alpha) \quad (2)$$

The combination is a linear function:  $\text{combination}(A, B, p) = p \times A + (1 - p) \times B$ , where  $p \in [0, 1]$  represents the proportion between  $A$  and  $B$ . In Eq. (3),  $\alpha$  is the proportion which varies the importance of the structural score compared to the semantic one. We describe in Sect. 5 the parametric study which enabled us to find out the fittest values of  $\alpha$ .

- Step 4: Deriving the extracted interests for each community into the social dimension (social profile) according to the weights computed in the third step. At the end of step 3, a same element  $e$  may have different weights in different

communities in the user egocentric network. In order to obtain a single weight for the element  $e$ , the function CombMNZ, proposed by [32], is adopted to combine different weights of this element from different communities.

### 3.3 Community-Based Social Profile Construction Process with Temporal Score

The main difference between the proposed CoBSPT and the existing CoBSP algorithms lies in the integration of temporal factors to the community profiling (step 2 of CoBSP). In the step 2, we propose to assign a temporal score to each element from each community. The input of step 2 is the information shared by the individuals in the user's egocentric network. While introducing temporal factors in our context, it is important to note that we are in an egocentric network where people know and interact with each other, in particular with the user (*ego*). These relationships exist because ego and these individuals share some common interests. We use this specific characteristic in the interests weight calculation with assuming that the temporal factors are also related to the interactions between individuals and *ego*.

#### 3.3.1 Temporal Score Calculation

When considering any individual *Indiv* of the user's egocentric network and the information *Info* shared by *indiv*, we introduce two different temporal factors for the information.

First, as described in Sect. 2.2, we use a time decay function in order to evaluate the freshness of information. This allows us to give a temporal score to each information in a classical way. Nevertheless, rather than considering the freshness with respect to a particular timestamp (current date), we consider the freshness with respect to the timestamp of the last interaction between *Indiv* and *ego*. We thus assume that the information shared by *Indiv* at a timestamp close to their last interactions with *ego* is more relevant for *ego* than a more distant one.

Second, we assume that *ego* may be more influenced by the individuals having a strong relationship with him. So, another way of computing the temporal weight of the information  $\text{Info}_{\text{Indiv}}$  extracted from the individual *Indiv* is to assume that it is proportional to the relationship strength between *Indiv* and *ego*.

Finally, these two temporal factors will be combined and their possible combinations will be studied to decide which element is more important to compute the social profile: information weight only, relationship weight only, or some combination of the two weights. In the following section, we explain in more details how to compute the two score and their combination method.

### Existing Temporal Function

To weight any information *Info* using temporal factors, we are interested in the time exponential function, which is widely used in many applications to gradually decay the importance of past behavior as time goes by. Based on the demonstrated performance of the time exponential function (3) proposed in [33], we decided to adopt this function to calculate the information temporal score.

$$f(t) = e^{-\lambda t} \quad (3)$$

The value of  $t$  represents the elapsed time (e.g., days, weeks, months, years, etc.) between the information timestamp and a given timestamp. Thus,  $t = 0$  represents the value of the most recent period. For example, if we use day as a time unit,  $t = 0$  for the current date,  $t = 1$  for yesterday, and so on.  $\lambda \in [0,1]$  represents the time decay rate. The higher the  $\lambda$  is, the less important the old information is. The value  $t$  of an information *Info* is computed by considering the temporal distance between the *Info* timestamp and the current time, as shown in Eq. (4).

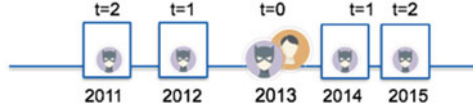
$$t = |\text{Current date} - \text{timestamp}(\text{info})| \quad (4)$$

### Information Temporal Score Calculation (calculInfoTempScore)

We present in this section, the `calculInfoTempScore` function of the CoBSPT algorithm. In this function, we assign a temporal score to each information  $\text{Info}_{\text{Indiv}}$  shared by an individual *Indiv* belonging to the user  $u$  egocentric network. The information temporal score is computed by applying the temporal function  $f(t)$  of Eq. (3).

In our case, we do not use the value of  $t$  as a simple elapsed time between the information timestamp and the current time. We suggest that the *Info* shared close to the last interaction between *Indiv* and the central user  $u$  is considered more relevant to  $u$  than the distant one. So, we modified Eq. (4) by changing the current date by the timestamp of the last interaction between  $u$  and *Indiv* as shown in Eq. (5). 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 2013,  $t_{\text{interaction}}$  of the information shared in 2012 or 2014 is equal to 1, and  $t$  of the information shared in 2011 or 2015 is equal to 2 and so on (see Fig. 1).

$$t_{\text{Info}_{\text{Indiv}}} = |(\text{last interaction timestamp}(\text{Indiv}, u)) - (\text{Info}_{\text{Indiv}} \text{ timestamp})| \quad (5)$$



**Fig. 1** The temporal relevance of information based on the last interaction of the user and a given individual

Finally, the information temporal score of a given information *Info* shared by an individual *Indiv* that will be assigned to each element *e* extracted from *Info* by using Eq. (6):

$$S_{\text{Temp}}^{\text{InfoIndiv}} = \text{calculInfoTempScore}(\text{Info}_{\text{Indiv}}) = f(t_{\text{InfoIndiv}}) \quad (6)$$

### Relationships Strength Techniques

There are several techniques to compute the relationship strength between two individuals. The interaction frequency is one of the widely used techniques. In the bibliometric field, we can use the co-publication frequency to identify the relationship strength between co-authors. We can also apply an advanced method by assigning the higher weight to the publication that has fewest co-authors [5]. In the case of social networks, we can use topology-based metrics (e.g., common neighbors, shortest paths, etc.) to compute the relationship strength between two nodes. These metrics are also applied for link prediction [34]. In this work, we propose to apply a link prediction metric to compute a similarity score between two connected nodes. This similarity score will represent the relationship persistency between two nodes and approximate their relationship strength. We assume that evaluating the possible persistency of a relationship in a future is more relevant than evaluating its current relationship strength.

Based on the link prediction metrics comparative study [34], we decide to adopt the Adamic/Adar metric which, despite its simplicity, outperforms the other metrics in terms of accuracy. With this metric, a pair of nodes having a common neighbor that is not common to several other nodes is considered more important. The similarity score is computed by the following Eq. (7).

$$\text{AdamicAdar}(x, y) = \sum_{z \in \{\Gamma(x) \cap \Gamma(y)\}} \frac{1}{\log |\Gamma(z)|} \quad (7)$$

where  $\Gamma(x)$  represents the set of the neighbors of  $x$ .

To take into account the temporal factor, we adopt a time-aware link prediction technique introduced by [35] which proposes the integration of the temporal score into the existing Adamic/Adar metric. Based on this work, we use the following “temporal Adamic/Adar” metric:

$$\text{TemporalAdamicAdar}(x, y) = \sum_{z \in \{\Gamma(x) \cap \Gamma(y)\}} \frac{w(x, z) \bullet w(z, y)}{\log |\Gamma(z)|} \quad (8)$$

where  $\Gamma(x)$  represents the set of the neighbors of  $x$ . The  $w(x_1, x_2)$  function represents the temporal relevance score of two given nodes  $x_1$  and  $x_2$ . This function can be customized according to the required techniques to compute the temporal score for a given dataset.

#### Temporal Individuals Score Calculation (calculIndivTempScore)

We present in this section the function `calculIndivTempScore` of the CoBSPT algorithm. We adopt the temporal Adamic/Adar metric from Eq. (8) to compute the relationship strength between the central user  $u$  and a given individual *Indiv*. We suggest that individuals that have the most recent relationships with  $u$  are the ones that have the highest probability to share up-to-date common interests with him/her. In order to introduce it in Adamic/Adar metric, we apply the temporal function from Eq. (3) to compute the temporal relevance score (for Adamic/Adar metric) between two individuals  $x_1$  and  $x_2$  as follows (9):

$$w'(x_1, x_2) = f(\text{CurrentTimeStamp} - \text{LastInteractionTimeStamp}(x_1, x_2)) \quad (9)$$

Finally, the temporal individual relevance score of *Indiv* that will be assigned to each element  $e$  extracted from any information *Info* shared by *Indiv* is computed by using the temporal relevance score (Eq. (9)) for two individuals as follows (10):

$$\begin{aligned} S_{\text{Temp}}^{\text{Indiv}} &= \text{calculIndivTempScore}(\text{Indiv}) = \text{TempAdamicAdar}(u, \text{Indiv}) \\ &= \sum_{z \in \{\Gamma(u) \cap \Gamma(\text{indiv})\}} \frac{w'(u, z) \bullet w(z, \text{indiv})}{\log |\Gamma(z)|} \end{aligned} \quad (10)$$

We note that this temporal score is the same for each element  $e$  extracted from any *Info* shared by the individual *Indiv*.

#### Final Temporal Score Calculation

The final temporal score of an element  $e$  extracted from the information  $\text{Info}_{\text{Indiv}}$  shared by an individual *Indiv* is computed by linearly combining the temporal information weight and the temporal individual weight with a parameter  $\gamma$  as follows (11):

$$P_{\text{Temp}}^{\text{Indiv}, \text{Info}}(e) = \text{combination} \left( P_{\text{Temp}}^{\text{Info}, \text{Indiv}}, P_{\text{Temp}}^{\text{Indiv}}, \gamma \right) \Big| e \in \text{Info} \quad (11)$$

We present in Sect. 5, the parametric study to determine the best value of  $\gamma$ .

### 3.3.2 Temporal Score Integration

Once the temporal scores ( $S_{Temp}^{Comb}(e)$ ) of all elements for each community are computed, we aggregate them by computing their weight compared to the total weight of all elements found in the community (the AggregationScore function of the CoBSPT algorithm, computed by Eq. (12)).

$$S_{Temp}^{C_i}(e) = \text{AggregationScore} \left( S_{Temp}^{comb}(e, c_i) \right) = \frac{\sum S_{Temp}^{Comb}(e)}{\sum_{f \in E(C_i)} S_{Temp}^{Comb}(f)} \quad (12)$$

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

In step 3, we replace the term frequency ( $S_{tf}$ ) used in CoBSP by the temporal score ( $S_{Temp}^{C_i}$ ) calculated from the second step to compute their semantic score as follows (13):

$$S_{sem}(e, c_i) = S_{Temp}^{C_i}(e) \quad (13)$$

We keep the same step 4 as the CoBSP algorithm to derive the social profile. Finally, we return the social profile as a vector of weighted user interests.

We present below the proposed time-aware social profile building process based on the CoBSP algorithm.

*Used notations:*

$u$  is the given user (ego) for whom we desire to compute social profile

$k$  is the egocentric network of the user  $u$ .

$c_i$  is a community extracted from the egocentric network.

$C = \{ c_i \}$  is the set of communities extracted from the egocentric network.

$e$  is an element (term or keyword) extracted from a piece of information.

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

Notations of the form  $S_{Temp}^{xxx}$  are temporal weights calculations detailed in previous sections.

*Algorithm CoBSPT input:  $u$ ,  $k$  output: Social profile of user  $u$*

Begin

// 1<sup>st</sup> step: detect communities from the egocentric network  $k$  with iLCD algorithm

$C := iLCD(k)$

// 2<sup>nd</sup> step: community profiling

For each community  $c_i \in C$  do

For each individual  $Indiv \in c_i$  do

// formula (10), section 3.3.1.4

$S_{Temp}^{Indiv} = \text{calculIndivTempScore}(Indiv);$

For each information  $Info$  shared by  $Indiv$  do

//formula (6) section 3.3.1.2

$S_{Temp}^{InfoIndiv} = \text{calculInfoTempScore} (Info_{Indiv}) ;$   
 //formula (11) section 3.3.1.5

$$S_{Temp}^{Comb} (e) = \text{combination} \left( S_{Temp}^{Indiv}, S_{Temp}^{Info,Indiv}, \gamma \right)$$

End For

End For

// Update the score of each element  $e$  found in the community  $c_i$  by aggregating  
 $S_{Temp}^{Comb} (e)$  from the set of elements found in the community

For each element  $e \in E(c_i)$  do

//formula (12), section 3.3.2

$$S_{Temp}^{C_i} (e) = \text{AggregationScore} \left( S_{Temp}^{Comb} (e), c_i \right) ;$$

End For

End For

// 3<sup>rd</sup> step: community structural and semantic score calculation

For each community  $c_i \in C$  do

$$S_{Struct} (c_i) = \text{calculCentralityScore} (c_i) ;$$

For each  $e \in E(c_i)$  do

//formula (13) section 3.3.2

$$S_{sem} (e, c_i) = S_{Temp}^{C_i} (e) ;$$

//formula (2) section 3.2

$$S_{Sem, Struct} (e, c_i) = \text{combination}(S_{Struct}(c_i), S_{sem}(e, c_i), \alpha) ;$$

End For

End For

// 4<sup>th</sup> step: Social profile derivation

For each element  $e \in E(c_i)$ ,  $\forall c_i \in C$ , do

$$S_{Social} (e) = \text{CombMNZ} \left( S_{sem,struct} (e, c_i) \right), \forall c_i \in C ;$$

End For

Return social profile of  $u$ ;

End

## 4 Experiments

To validate our proposition, we compare the relevance of our time-aware social profile building technique (CoBSPT) against the existing time-agnostic technique (CoBSP) as explained in Fig. 2. The strategy is to find out which one can provide social profiles close to the real user profile. We use the precision and recall metrics to assess the comparison. Note that [7] yet showed the effectiveness of the community-based approach (CoBSP) compared to a classical individual-based approach.

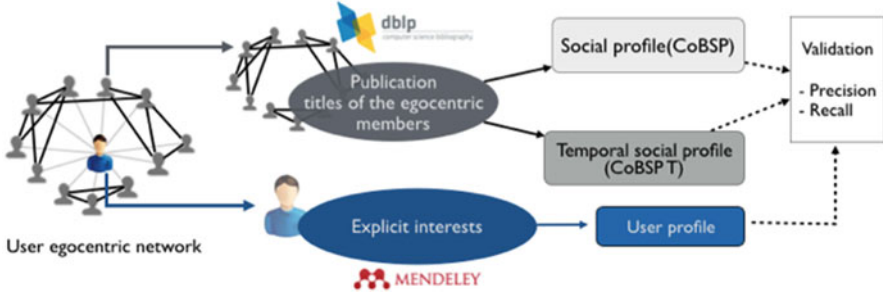


Fig. 2 User profile and social profile building process

### 4.1 Dataset Description

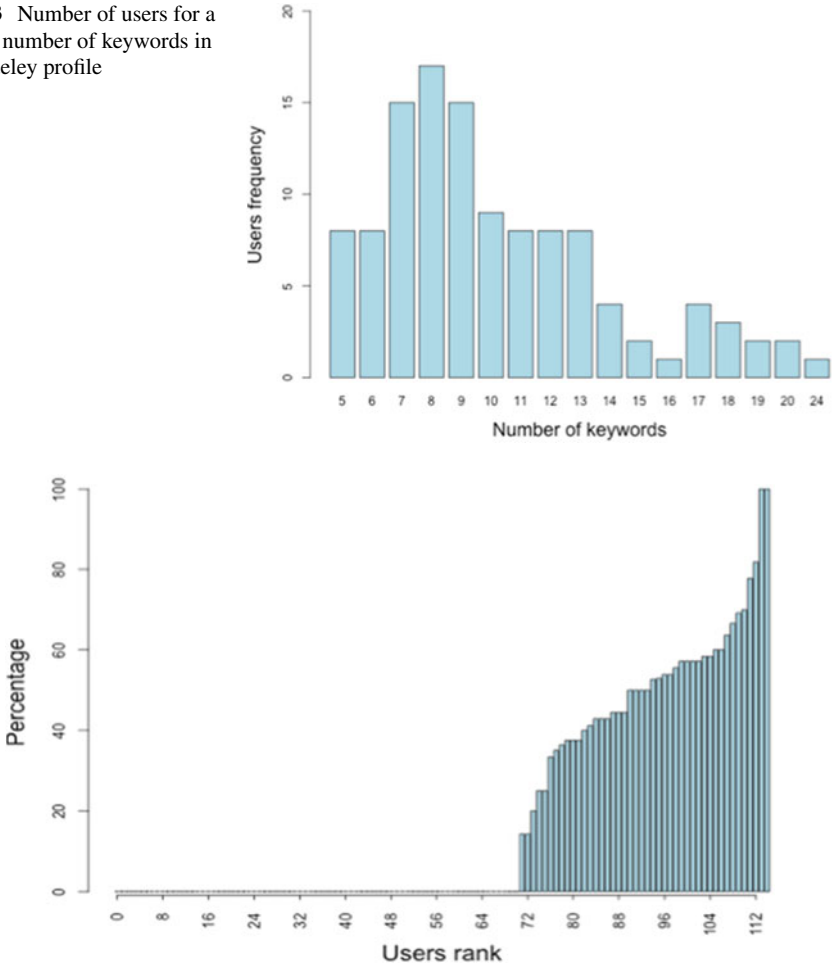
We conducted the experiments on two distinct and not connected co-authorships networks: DBLP and Mendeley. In the DBLP network, the nodes represent the authors. Two authors are connected if they published together at least one article. Two authors are connected as often as they published together and each link is labeled by the publication date. We used DBLP to build the social profile using the two approaches (CoBSP and CoBSPT). Using this data, we extract the user's interests from the publication titles. On the other hand, the real profile is built by using the interests indicated explicitly by the user in his/her Mendeley profile. This allows to use two distinct data sources and thus avoiding many biases in the results obtained and associated interpretations.

In our experiment, we consider the authors who exist in both DBLP and Mendeley databases and have enough interests in their Mendeley profile to significant and valuable results. Subsequently, we included authors that have at least 5 interests in their Mendeley profile. Our dataset contains 112 users (authors) distributed between 24 and 495 co-authors, with in average 85 of co-authors.

### 4.2 Analysis of Common Keywords Between DBLP and Mendeley

As we used distinct data sources to build both user and social profiles, we analyzed the number of common keywords found on the Mendeley user's profile and the publication titles of the target user in DBLP. We aimed to demonstrate the relevance and the compatibility of the two data sources and even to dismiss incompatible data. Figure 3 presents the number of users for a given number of keywords in the Mendeley profile. The 112 users of our dataset have between 5 and 24 interests in Mendeley.

**Fig. 3** Number of users for a given number of keywords in Mendeley profile



**Fig. 4** Percentage of Mendeley keywords found in the DBLP publication titles

Figure 4 shows the percentage of Mendeley keywords found in the DBLP publication titles for all users, sorted by this percentage. The average percentage of common keywords for all users is 19.3%. In the sub-group of 42 users that have at least one keyword of Mendeley in their DBLP publication titles, we observe that they have between 5 and 20 keywords in their Mendeley profile. For this sub-group, the average percentage of common keywords is 50.5%, the minimum value is 14.3%, and the maximum value is 100%.

### 4.3 Case Study

#### 4.3.1 Ground Truth: Extraction of the Real User Profile from Mendeley

To build the real user profile as a ground truth, we use the interests indicated explicitly by the users in their Mendeley profile. First, for each user, we collect the keywords from his/her Mendeley profile. Then we extract interests from the collected keywords using text-mining classical tools: we used dictionaries and thesaurus to merge keywords having the same meaning and we removed empty words using filters, in order to keep only consistent interests. Finally, we compute the weight of each extracted interest using the term frequency (tf) measure that represents the term frequency of each interest in the set of all found interests [7].

#### 4.3.2 Social Profiles Construction and Parametric Study

The first step of the social profile building process consists in gathering the co-authors of each user to build his/her egocentric network. Then, we extract the communities from the user egocentric networks. For each community, the publication titles of the members are collected. Finally, we analyze the collected publication titles to extract the meaningful keywords using the text-mining engine as presented in Sect. 4.3.1. We apply the temporal score to each keyword according to Eq. (11) before computing their semantic score.

We build the social profiles with a parametric study in order to infer the suitable values for parameters  $\alpha$ ,  $\gamma$ , and  $\lambda$ , represented in Eqs. (2), (3), and (11), that give the most accurate results. Furthermore, our aim is to find out the impact of each parameter to the time-aware social profile building process. The value of each parameter is ranged between 0 and 1 ( $\lambda, \gamma, \alpha \in \{0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 1.0\}$ ). Note that we can build both the social profile of the time-agnostic approach (CoBSP) 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.

### 4.4 Results

This section presents the results of our experiments by comparing the effectiveness of the existing time-agnostic approach CoBSP and our proposed time-aware approach CoBSPT. We only consider the most relevant interests in the social profile to compute the precision and recall (the top  $n$  of all interests sorted in descending order). Note that all following results shown in this section are computed by using the top 5 interests. Results with top 10 interests are comparable but precision is less meaningful with real user profiles having less than ten keywords.

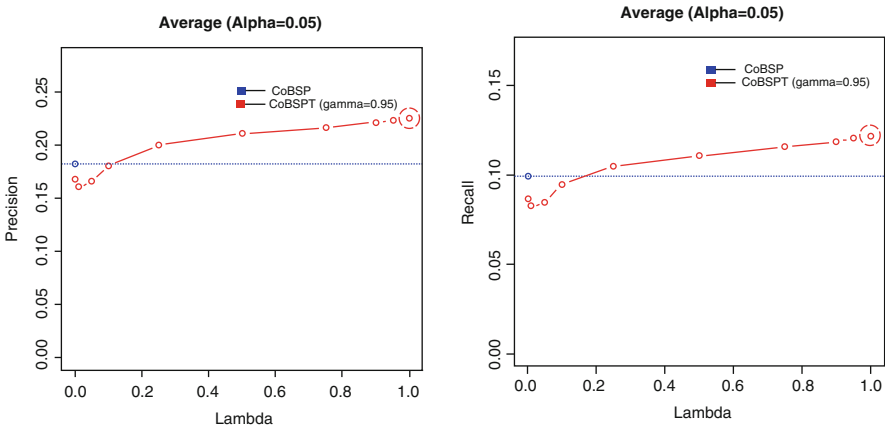
It is important to mention that, all results presented hereafter, we compare CoBSP and our CoBSPT results with respect to  $\lambda$ ,  $\gamma$ , and  $\alpha$ . Nevertheless,  $\lambda$ ,  $\gamma$  are not implied in CoBSP and thus CoBSP results will never vary whatever the values of  $\lambda$  and  $\gamma$  (the representation is a straight line).

#### 4.4.1 All Users Results

We first present the results of our parametric study for all users (112). The global results of each approach are presented by the average of the precision and recall for all users.

For the CoBSP approach, the best precision (0.1821) is observed when  $\alpha \in \{0.01, 0.05, 0.1, 0.25, 0.5\}$ . The best recall (0.099) is observed when  $\alpha = 0.05$ . For our proposed approach CoBSPT, the best precision (0.225) is observed when the parameters  $(\alpha, \gamma, \lambda)$  equal to  $(0.05, 0.95, 1)$  and  $(0.01, 0.01, 0.01)$ . The best recall (0.1218) is observed when  $\alpha = 0.05$ ,  $\gamma = 0.95$ , and  $\lambda = 1$ . We present a complete parametric study on this dataset in terms of precision in Fig. 11 in the appendix. The results in terms of recall are presented in Fig. 12 in the appendix.

Figure 5 presents the comparison of the best precision and recall of the social profiles built by the CoBSP and CoBSPT approaches. As the best results for both approaches are obtained with  $\alpha = 0.05$ , we fixed this value for the precision graph (see left side of Fig. 6). The red curve represents the precision of the CoBSPT approach for different value of  $\lambda$ . The precision is computed by fixing  $\gamma = 0.95$  (as we obtain the best precision by this value and when  $\alpha = 0.05$ ). The blue dot represents the best precision of the CoBSP approach and thus it does not vary whatever the values of  $\lambda$  and  $\gamma$ . The recall graph (right part of Fig. 5) represents



**Fig. 5** *Left:* comparison of the average precision for all users with the best parameters for each approach for  $\alpha = 0.05$  ( $\gamma = 0.95$  for CoBSPT). *Right:* comparison of the average recall for all users with the best parameters for each approach for  $\alpha = 0.05$  ( $\gamma = 0.95$  for CoBSPT)

the same information as for the precision graph with the fixed value of  $\gamma = 0.95$  (as we obtain the best recall with this value when  $\alpha = 0.05$ ).

We found that the best results of CoBSPT outperform the obtained ones by CoBSP of 23.5% and 23% in terms of precision and recall, respectively. This improvement shows the effectiveness of our temporal method compared to the time-agnostic one. Nevertheless, any improvement is achieved only if  $\lambda > 0.2$ .

#### 4.4.2 Results for Selected Users

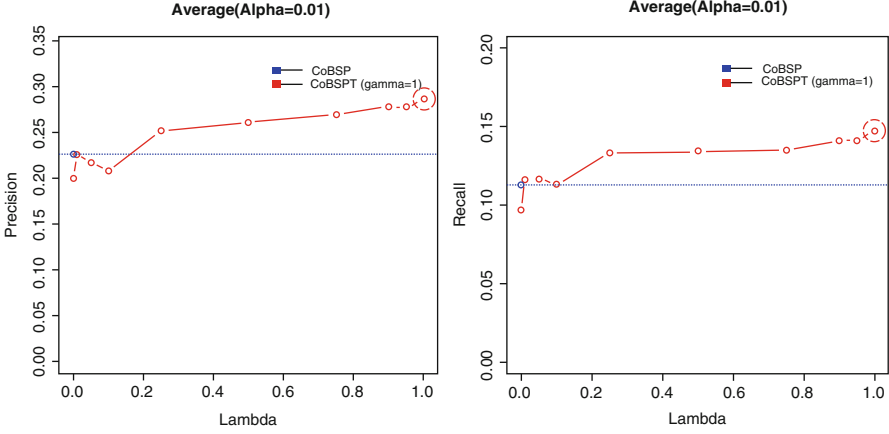
Based on the analysis of common keywords between DBLP and Mendeley, previously explained in the Sect. 4.2, we observe that only 42 users have at least a common keyword between their Mendeley profile and their publication titles in DBLP. This means that Mendeley keywords have very few chances to be retrieved by CoBSP or CoBSPT algorithms, and thus can affect the experimentation results. Thus, we performed another study by taking into account only those users who have more than 50% of common keywords between DBLP and Mendeley. This new dataset contains 25 users.

For the CoBSP approach, the best precision (0.226) is observed when  $\alpha \in \{0.01, 0.05, 0.1, 0.25, 0.5\}$ . The best recall (0.1154) is observed when  $\alpha = 0.95$ . For our proposed approach CoBSPT, the best precision (0.2869) is observed when  $\alpha = 0.01$ ,  $\gamma = 1.0$ , and  $\lambda = 1.0$ . The best recall (0.1472) is observed when  $\alpha = 0.01$ ,  $\gamma = 1.0$ , and  $\lambda = 1.0$ . We present a complete parametric study on this dataset in terms of precision shown in Fig. 13 in the appendix. The results in terms of recall are presented in Fig. 14 in the appendix.

Figure 6 presents the comparison of the best precision and recall of the social profiles built by the CoBSP and CoBSPT approaches for the selected users. As the best results for both approaches are obtained when  $\alpha = 0.01$ , we fixed this value for the precision graph (see left side of Fig. 6). The red curve represents the precision of the CoBSPT approach for different values of  $\lambda$ . As we obtained the best precision when  $\alpha = 0.01$  and  $\gamma = 1.0$ , we fixed these values. The blue dot represents the best precision for the CoBSP approach. Note that the precision of CoBSP is only computed when  $\lambda = 0.0$  and  $\gamma = 0.0$ .

The recall graph (see right side of Fig. 6) presents the same information as the precision graph when the values of  $\gamma = 1.0$  and  $\alpha = 0.01$ .

We can observe that the best result of the CoBSPT algorithm also outperforms the result of CoBSP of 23.89 and 27.8% in terms of precision and recall, respectively. Furthermore, the precision and recall of both approaches are better than those compared to the results for all users. Nevertheless, any improvement is now achieved only if  $\lambda > 0.2$ .



**Fig. 6** *Left:* comparison of the average precision for selected users with the best parameters for each approach with  $\alpha = 0.01$  ( $\gamma = 1$  for CoBSPT). *Right:* comparison of the average recall for selected users with the best parameters for each approach with  $\alpha = 0.01$  ( $\gamma = 1$  for CoBSPT)

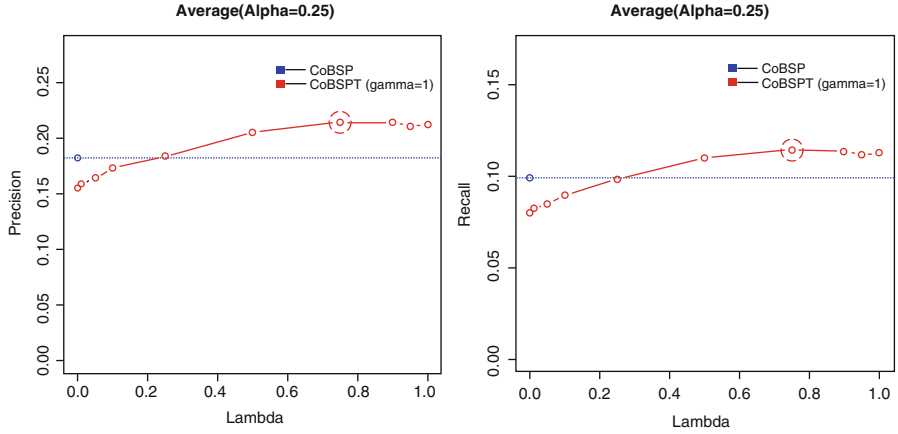
#### 4.4.3 Different Time Decay Rate for the Relationships and the Information

In CoBSPT algorithm, to compute the temporal score of each interest, we apply a time decay function for the individual and information temporal score calculation. This decay function uses a time decay rate ( $\lambda$ ) set in previous experiments to the same values for both calculations. With this technique, we obtain better results than the existing approach as shown in Sects. 4.4.1 and 4.4.2.

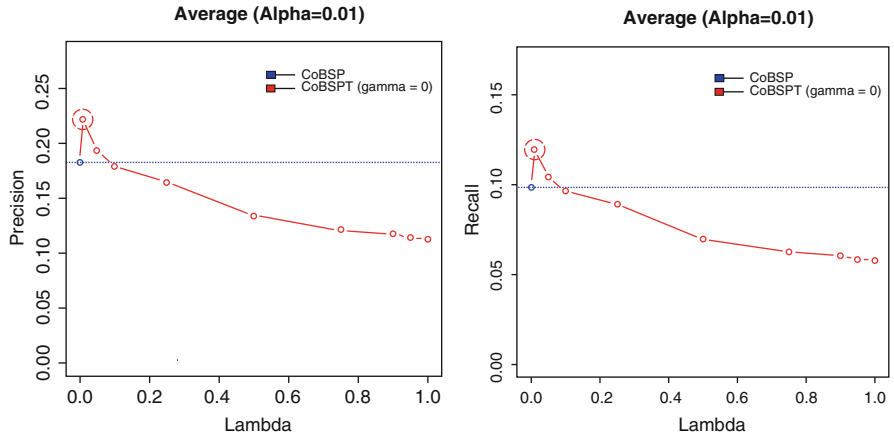
In this section, we hypothesis that the time decay rate of the relationships and the information could be different. In fact, it could be possible that in a social network, to have people tend to create/share information more than creating new contacts or interact with each other, and vice versa. Thus, in this section we study the impact of the relationship dynamic (individual temporal score) compared to the information dynamic (information temporal score). Note that we use the “all users” dataset of Sect. 4.4.1 for this experiment.

In order to study the impact of the individual temporal score, we focus on the precision and recall values for  $\gamma = 1$ . In fact, this value represents the results of the social profile computed by considering only the individual temporal score. We obtained the best precision (0.214) when  $\lambda = 0.75$  and  $\alpha \in \{0.25, 0.5\}$ . For the recall, we got the best value (0.114) when  $\lambda = 0.75$  and  $\alpha = 0.25$ . With these settings, the result of CoBSPT outperforms the result of CoBSP of 17.5 and 23.9% in terms of precision and recall (see Fig. 7). These values ensure the benefits of using a temporal link selection method in building the user social profile.

Similarly, we focus on the precision and recall values computed when fixing  $\gamma = 0$  to study the impact of the information temporal score (ignoring individual temporal score). We obtained the best precision (0.221) when  $\lambda = 0.01$  and

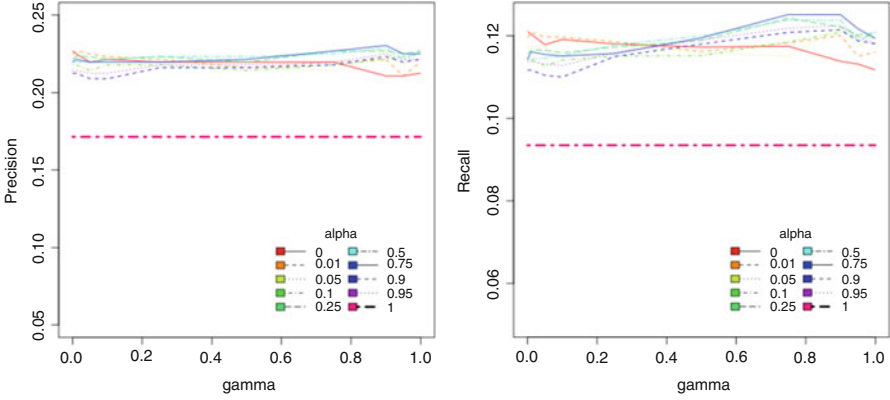


**Fig. 7** Comparison of the average precision and recall for all users with the best parameters for each approach for  $\alpha = 0.25$  ( $\gamma = 1$  for CoBSPT)



**Fig. 8** Comparison of the average precision recall average for all users with the best parameters for each approach with  $\alpha = 0.01$  ( $\gamma = 0$  for CoBSPT)

$\alpha = 0.01$ . For the recall, we got the best value (0.1195) when setting  $\lambda = 0.01$  and  $\alpha = 0.01$ . With these settings, the result of CoBSPT outperforms the result of CoBSP of 21.3 and 29.89% in terms of precision and recall (see Fig. 8). These results show the benefit of exploiting the temporal characteristic of the information to extract meaningful user's interests to build relevant user social profile. This ensures also the fact that the diffusion of information plays also an important role in the evolution of the user's interests in online social networks, as the social network structure does. However, we observe that the value  $\lambda = 0.01$ , corresponding to the best precision and recall for the CoBSPT, is different from the individual temporal score which is  $\lambda = 0.75$ .



**Fig. 9** Comparison of the precision average of all users on the basis of  $\gamma$  and  $\alpha$  approach by fixing  $\lambda$  to 0.75 for the individual temporal score and by  $\lambda$  to 0.01 for the information temporal score

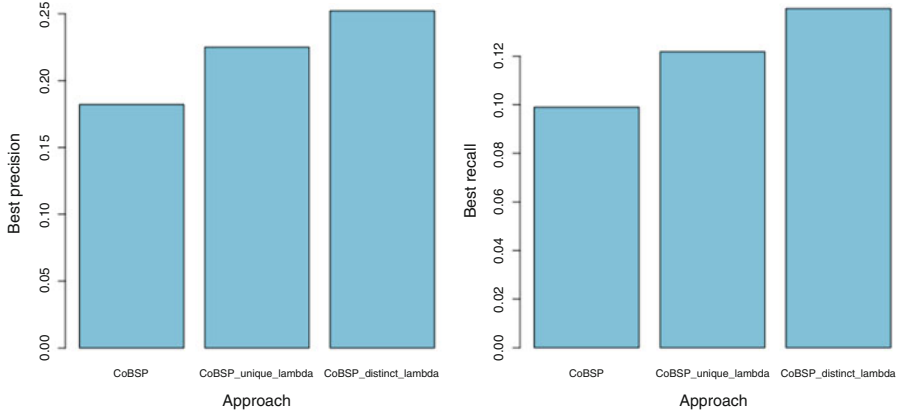
We then build the social profile using CoBSPT approach with fixing the parameters  $\lambda$  to 0.75 of the individual temporal score, and  $\lambda$  to 0.01 of the information temporal score. We check the results on the basis of  $\gamma$  and  $\alpha$ . According to the left part of Fig. 9, we got the best precision (0.23) when  $\gamma = 0.9$  and  $\alpha = 0.75$ . For the recall on the right part of Fig. 8, we obtained the best value (0.125) when  $\gamma = 0.75$  and  $\alpha = 0.75$ .

Based on Fig. 9, we observe that the curves are almost flat even if temporal information and the individual scores clearly have different properties (as underlined by Figs. 7 and 8). This behavior can be interpreted in two ways: (1) the two scores are correlated, and therefore the same set of profiles is correctly recognized when varying  $\gamma$  is; or (2) the two scores are not correlated, but when changing  $\gamma$  a certain amount of profiles is not correctly recognized anymore and thus a comparable amount of profiles is now correctly recognized (the loss is compensated by the gain).

We obtained a better improvement compared to the resulted one by the previous approach, in which we fixed the same value for the temporal individual score and the temporal information score (see Fig. 10). The CoBSPT algorithm outperforms the CoBSP with an improvement of 26.5% and 26.3% in terms of precision and recall, respectively.

#### 4.4.4 Discussion

Obviously, the experiments shown the effectiveness of our proposition compared to the existing time-agnostic egocentric network-based user profiling process in terms of precision and recall. This proves how much is beneficial of leveraging the evolution of user's interests (by considering the social network evolution) in the social network-based user profiling process. One can be surprised and skeptical



**Fig. 10** Comparison of the best result from each technique (CoBSP, CoBSP\_unique\_lambda), obtained by using the same value of  $\lambda$  for the individual temporal score and the information temporal score (CoBSP\_distinct\_lambda), obtained by fixing distinct values of  $\lambda$  ( $\lambda = 0.75$  for the individual temporal score and  $\lambda = 0.01$  for the information temporal score)

considering the low scoring results. We emphasize that the results are compared to optional data from users' Mendeley profiles which therefore might be incompatible and/or outdated.

We have assumed that a time-aware social profiling method proposed answers to the two research questions stated in the introduction (how to select the relevant individuals and how to select the relevant information). Choosing both relevant information and individuals is a crucial issue. The selection criteria and features can vary according to the social network type and users' behaviors. Our time-aware method is thus more generic and it seems to give interesting results.

The results have been improved on the 25 selected users having more than 50% of common keywords between DBLP. This demonstrates that the matching step of keywords has also an impact on the precision and recall of the two algorithms in the evaluation process. We suggest that the keywords-based approach can lead to have wide range of words having the same meaning (synonym) and sometimes the keywords are too specific to the users. For example, a user may be interested in machine learning in general but when we extract keywords from publication titles of his friends we can found "svm" which is a specific machine learning method. So, it is not obvious to link "svm" to "machine learning." As a result, the extracted interests are too specific to the user and we can miss the fact that the user is also interested in other machine learning methods. This problem can be more crucial on other types of social network which are more general in terms of users and shared information. For example, on Twitter or Instagram, users tend to share information about their daily life or social events, buzz, etc. In this case, it would be appropriate to apply other techniques to extract the user interests or other models to represent the social profile. For example, we can adopt the concept-based model that represents concepts rather than words or bags of words and the concepts can be also set on

different hierarchical level. We can also apply the semantic network model that represents the profile by a weighted semantic network in which each node represents a concept [1].

To obtain the best results, the time decay rates applied to relationships and information must be defined differently. The best decay rate applied to relationships and used to compute the individual temporal score is 0.75 and that of the information is 0.01. We can conclude that the relationships between users evolve more quickly than the information in the co-author network context. Thus, taking into account the freshness of relationships is very important meanwhile taking into account the freshness of information is less important. Furthermore, the best proportion between the individual temporal score and that of the temporal information score is 0.9. This shows the importance of the individual temporal score compared to that of the information in the context of co-author networks. These observations can be related with the characteristic of our dataset. We can observe that co-authorships networks do not exhibit a rapid evolution of information compared to other OSNs (Facebook, Twitter, Reddit). The authors can collaborate with different co-authors overtime, but their research field generally remains the same for a long period. So, their publication titles could remain related to the same topic domains whereas the relationships with their co-authors change more rapidly over time. Thus, we should take into consideration the dynamic characteristic of the relationships between users compared to the information dynamics on co-authorships networks.

## 5 Conclusion and Future Works

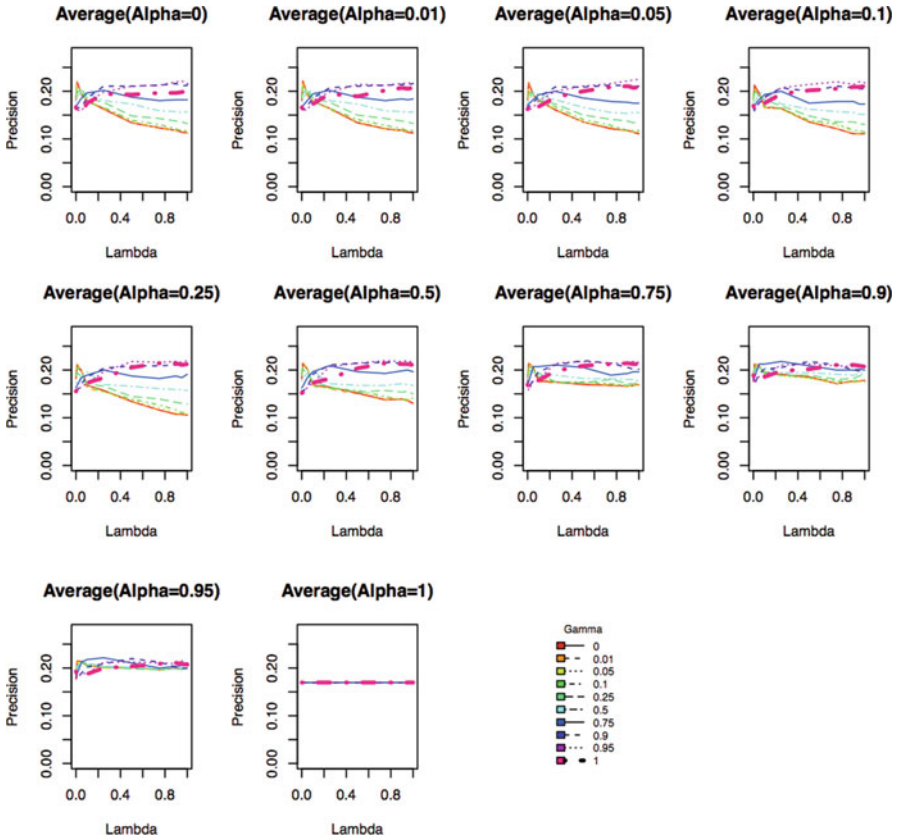
In this paper, we proposed to take into account the social network dynamic in user social profile building process. Considering the importance of the individuals of user social networks and their information, we defined temporal criteria (temporal scores) to weight both individuals and information in order to extract relevant and up-to-date user interests. The experiments shown the effectiveness of this method compared to the existing time-agnostic one. This demonstrates the benefit of taking into account the evolution of user interests (by considering the social network evolution) in the social network-based user profiling process. We observed that in the co-authorship network context, the time decay rate of user relationship is required to be higher than the one of the shared information. We also observed that the individual temporal score has a larger preference over the information temporal score in our final temporal score calculation.

We plan to apply our approach on other social networks with even a higher dynamic characteristic than co-authorships networks in order to evaluate our proposition for large-scale data and also to study the impact of link dynamics and information dynamics on these social networks (such as Facebook, Twitter, or Reddit). We plan also to use other link prediction algorithms and other time-weight functions to enhance the performance of our approach. Other models of social profile will be also studied and adopted for our future work to overcome the limit of granularity and synonym of the bags of words model.

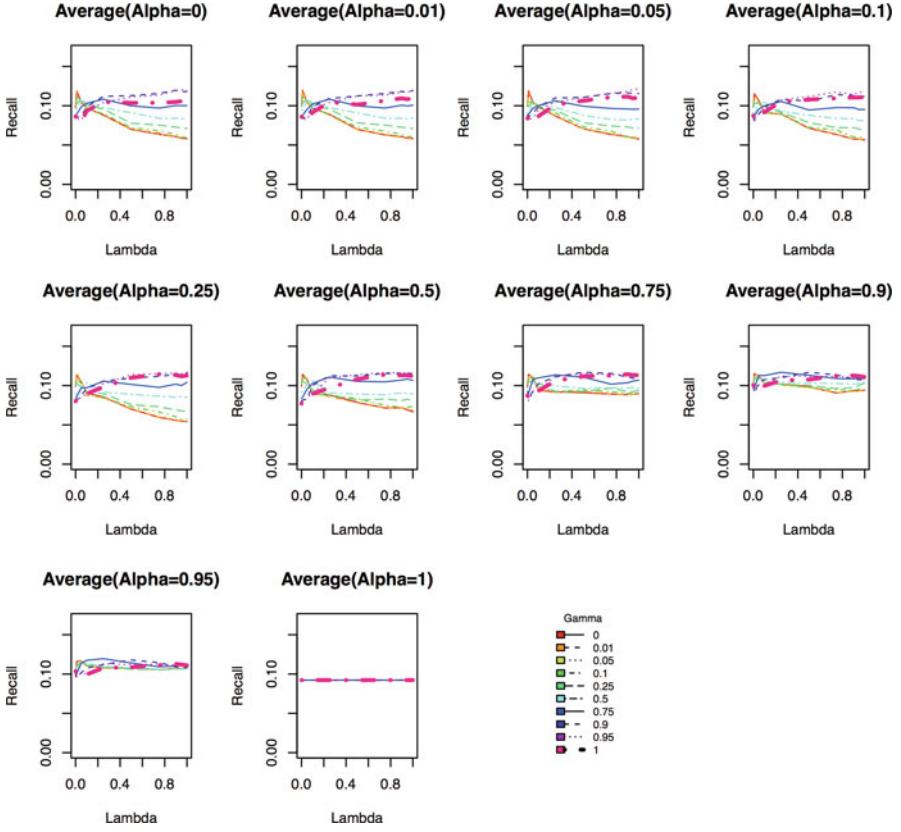
Our long-term perspective consists in the proposition of a generic platform that extracts the information and builds the user social profile according to the type and the specific characteristics of the underlying social network. Such a platform would be parameterized by the characteristics of the targeted social network using a machine learning approach.

## Appendix

Figure 11 represents the result of the parametric study in terms of precision for all users. We plot a graph for each value of the parameter  $\alpha$ . For a given graph, each curve represents the precision for a given value of the parameter  $\gamma$ , for all values of the parameter  $\lambda$ , shown in the X-axis. Figure 12 represents the same information in terms of recall. We recall that  $\alpha$  represents the proportion of the structural score

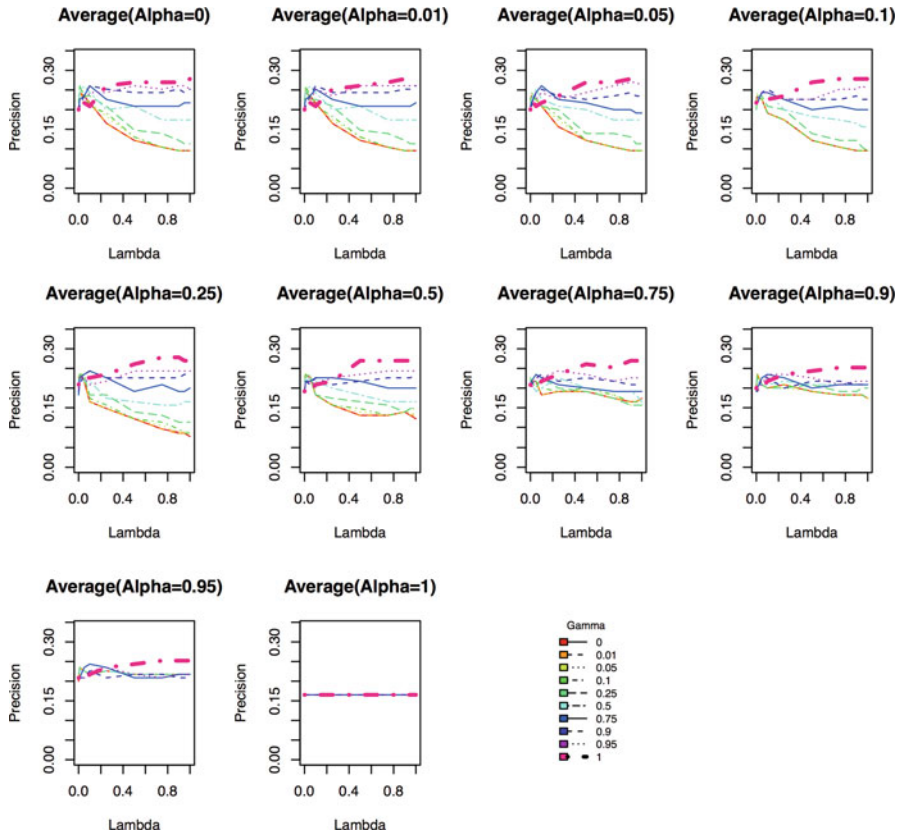


**Fig. 11** The average precision for all users with the parametric study for parameters  $\alpha$ ,  $\gamma$ , and  $\lambda$

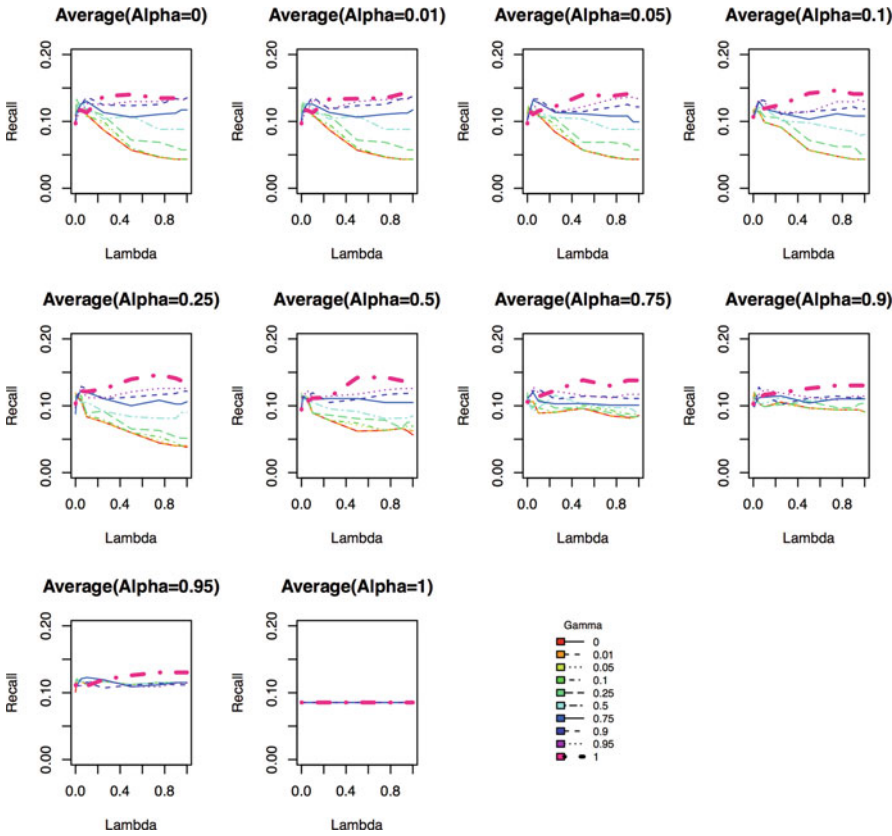


**Fig. 12** The average recall for all users with the parametric study for parameters  $\alpha$ ,  $\gamma$ , and  $\lambda$

compared to the semantic score as presented in Eq. (1),  $\lambda$  represents the time decay rate as presented in Eq. (2), and  $\gamma$  represents the proportion of the individual temporal score compared to the information temporal score as presented in Eq. (9). As noted above, the set of points corresponding to  $\lambda = 0.0$  and  $\gamma = 0.0$  represents the CoBSP results. Figures 13 and 14 represent the result of the parametric study in terms of precision and recall for 25 selected users that have at least 50% of common keyword between Mendeley and DBLP.



**Fig. 13** The average precision for selected users with the parametric study for parameters  $\alpha$ ,  $\gamma$ , and  $\lambda$



**Fig. 14** The recall average for all users with the parametric study for parameters  $\alpha$ ,  $\gamma$ , and  $\lambda$

## References

1. Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A.: User profiles for personalized information access. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) *The Adaptive Web*. Lecture Notes in Computer Science, vol. 4321, pp. 54–89. Springer, Berlin (2007). [http://link.springer.com/chapter/10.1007/978-3-540-72079-9\\_2](http://link.springer.com/chapter/10.1007/978-3-540-72079-9_2)
2. Abel, F., Gao, Q., Houben, G.-J., Tao, K.: Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web. In: *Proceedings of the 3rd International Web Science Conference*, pp. 2:1–2:8, WebSci '11, ACM, New York, NY (2011). doi:[10.1145/2527031.2527040](https://doi.org/10.1145/2527031.2527040)
3. Arnaboldi, V., Conti, M., Passarella, A., Dunbar, R.: 2013. Dynamics of personal social relationships in online social networks: a study on Twitter. In: *Proceedings of the First ACM Conference on Online Social Networks*, pp. 15–26. COSN '13, ACM, New York, NY. doi:[10.1145/2512938.2512949](https://doi.org/10.1145/2512938.2512949)
4. Canut, M.-F., On-At, S., Péninou, A., Sèdes, F.: 2015. 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*, pp. 569–572, ASONAM '15, ACM, New York, NY. doi:[10.1145/2808797.2809415](https://doi.org/10.1145/2808797.2809415)
5. Cabanac, G.: Accuracy of inter-researcher similarity measures based on topical and social clues. *Scientometrics*. **87**(3), 597–620 (2011). doi:[10.1007/s11192-011-0358-1](https://doi.org/10.1007/s11192-011-0358-1)
6. David, C., Zwerdling, N., Guy, I., Ofek-Koifman, S., N. Har'el, I. Ronen, E. Uziel, S. Yogev, S. Chernov. 2009. Personalized social search based on the user's social network. In: *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pp. 1227–1236, CIKM '09, ACM, New York, NY. doi:[10.1145/1645953.1646109](https://doi.org/10.1145/1645953.1646109)
7. Tchuente, D., Canut, M.-F., Jessel, N., Peninou, A., Sèdes, F.: A community-based algorithm for deriving users' profiles from egocentric networks: experiment on Facebook and DBLP. *Soc. Network Anal. Min.* **3**(3), 667–683 (2013). doi:[10.1007/s13278-013-0113-0](https://doi.org/10.1007/s13278-013-0113-0)
8. Aral, S., Walker, D.: Tie strength, embeddedness, and social influence: a large-scale networked experiment. *Manage. Sci.* **60**(6), 1352–1370 (2014). doi:[10.1287/mnsc.2014.1936](https://doi.org/10.1287/mnsc.2014.1936)
9. Gama, J., I. Žliobaitė, A. Bifet, M. Pechenizkiy, A. Bouchachia. 2014. A survey on concept drift adaptation." *ACM Comput. Surv.* **46** (4): 44:1–44:37. doi:[10.1145/2523813](https://doi.org/10.1145/2523813).
10. Tsymbal, A.. 2004. The problem of concept drift: definitions and related work. Technical Report, Department of Computer Science, Trinity College, Dublin
11. Widmer, G., M. Kubat. 1993. Effective learning in dynamic environments by explicit context tracking. In: P.B. Brazdil (ed.) *Machine Learning: ECML-93*. Lecture Notes in Computer Science, vol. 667, pp. 227–243, Springer, Berlin. [http://link.springer.com/chapter/10.1007/3-540-56602-3\\_139](http://link.springer.com/chapter/10.1007/3-540-56602-3_139)
12. Mianowska, B., Nguyen, N.T.: Tuning user profiles based on analyzing dynamic preference in document retrieval systems. *Multimedia Tools Appl.* **65**(1), 93–118 (2013). doi:[10.1007/s11042-012-1145-6](https://doi.org/10.1007/s11042-012-1145-6)
13. Sugiyama, K., K. Hatano, M. Yoshikawa. 2004. Adaptive web search based on user profile constructed without any effort from users. In: *Proceedings of the 13th International Conference on World Wide Web*, pp. 675–684, WWW '04, ACM, New York, NY. doi:[10.1145/988672.988764](https://doi.org/10.1145/988672.988764)
14. Koren, Y.. 2009. Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 447–456, KDD '09, ACM, New York, NY. doi:[10.1145/1557019.1557072](https://doi.org/10.1145/1557019.1557072)
15. Maloof, M.A., Michalski, R.S.: Selecting examples for partial memory learning. *Mach. Learn.* **41**(1), 27–52 (2000). doi:[10.1023/A:1007661119649](https://doi.org/10.1023/A:1007661119649)
16. Bennett, P.N., R.W. White, W. Chu, S.T. Dumais, P. Bailey, F. Borisyuk, X. Cui. 2012. Modeling the impact of short- and long-term behavior on search personalization. In: *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 185–194, SIGIR '12, ACM, New York, NY. doi:[10.1145/2348283.2348312](https://doi.org/10.1145/2348283.2348312)

17. Tan, B., X. Shen, and C. Zhai. 2006. Mining long-term search history to improve search accuracy. In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 718–723, KDD '06, ACM, New York, NY. doi:[10.1145/1150402.1150493](https://doi.org/10.1145/1150402.1150493)
18. Li, D., Cao, P., Guo, Y., Lei, M.: Time weight update model based on the memory principle in collaborative filtering. *J. Comput.* **8**(11), 2763–2767 (2013). doi:[10.4304/jcp.8.11.2763-2767](https://doi.org/10.4304/jcp.8.11.2763-2767)
19. Li, L., Zheng, L., Yang, F., Li, T.: Modeling and broadening temporal user interest in personalized news recommendation. *Expert Syst. Appl.* **41**(7), 3168–3177 (2014). doi:[10.1016/j.eswa.2013.11.020](https://doi.org/10.1016/j.eswa.2013.11.020)
20. Zheng, N., Li, Q.: A recommender system based on tag and time information for social tagging systems. *Expert Syst. Appl.* **38**(4), 4575–4587 (2011). doi:[10.1016/j.eswa.2010.09.131](https://doi.org/10.1016/j.eswa.2010.09.131)
21. Kacem, A., M. Boughanem, R. Faiz. 2014. Time-sensitive user profile for optimizing search personalization. In: V. Dimitrova, T. Kuflik, D. Chin, F. Ricci, P. Dolog, G.-J. Houben (eds.) *User eModeling, Adaptation, and Personalization. Lecture Notes in Computer Science*, vol. 8538, pp. 111–121, Springer International Publishing, Basel. [http://link.springer.com/chapter/10.1007/978-3-319-08786-3\\_10](http://link.springer.com/chapter/10.1007/978-3-319-08786-3_10)
22. Stattner, E., Collard, M., Vidot, N.: D2SNet: dynamics of diffusion and dynamic human behaviour in social networks. *Comput. Hum. Behav. Adv. Hum. Comput. Interact.* **29**(2), 496–509 (2013). doi:[10.1016/j.chb.2012.06.004](https://doi.org/10.1016/j.chb.2012.06.004)
23. Weng, L., J. Ratkiewicz, N. Perra, B. Gonçalves, C. Castillo, F. Bonchi, R. Schifanella, F. Menczer, A. Flammini. 2013. 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*, pp. 356–364, KDD '13, ACM, New York, NY. doi:[10.1145/2487575.2487607](https://doi.org/10.1145/2487575.2487607)
24. Gomez Rodriguez M., J. Leskovec, B. Schölkopf. 2013. Structure and dynamics of information pathways in online media. In: *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, pp. 23–32, WSDM '13, ACM, New York, NY. doi:[10.1145/2433396.2433402](https://doi.org/10.1145/2433396.2433402)
25. Koohy, H., Koohy, B.: A lesson from the ice bucket challenge: using social networks to publicize science. *Front. Genet.* **5**, (2014). doi:[10.3389/fgene.2014.00430](https://doi.org/10.3389/fgene.2014.00430)
26. Kumar, R., J. Novak, A. Tomkins. 2006. Structure and evolution of online social networks. In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 611–617, KDD '06, ACM, New York, NY. doi:[10.1145/1150402.1150476](https://doi.org/10.1145/1150402.1150476)
27. Cazabet, R., M. Leguistin, F. Amblard. 2012. Automated community detection on social networks: useful?Efficient?Asking the users. In: *Proceedings of the 4th International Workshop on Web Intelligence; Communities*, pp. 6:1–6:8, ACM, New York, NY. doi:[10.1145/2189736.2189745](https://doi.org/10.1145/2189736.2189745)
28. Rosvall, M., Bergstrom, C.T.: An information-theoretic framework for resolving community structure in complex networks. *Proc. Natl. Acad. Sci. U. S. A.* **104**(18), 7327–7331 (2007). doi:[10.1073/pnas.0611034104](https://doi.org/10.1073/pnas.0611034104)
29. Pollner, P., Palla, G., Vicsek, T.: Parallel clustering with cfinder. *Parallel Process. Lett.* **22**(1), 1240001 (2012). doi:[10.1142/S0129626412400014](https://doi.org/10.1142/S0129626412400014)
30. Friggeri, A., G. Chelius, E. Fleury. 2011. Triangles to capture social cohesion. arXiv:1107.3231 [Physics]. <http://arxiv.org/abs/1107.3231>
31. Salton, G., Waldstein, R.K.: Term relevance weights in on-line information retrieval. *Info. Process. Manage.* **14**(1), 29–35 (1978). doi:[10.1016/0306-4573\(78\)90055-9](https://doi.org/10.1016/0306-4573(78)90055-9)
32. Shaw, J.A., E.A. Fox, J.A. Shaw, E.A. Fox. 1994. Combination of multiple searches. In: *The Second Text REtrieval Conference (TREC-2)*, pp. 243–252
33. Ding, Y., Jacob, E.K., Caverlee, J., Fried, M., Zhang, Z.: Profiling social networks: a social tagging perspective. *D-Lib Mag.* **15**(3/4), (2009). doi:[10.1045/march2009-ding](https://doi.org/10.1045/march2009-ding)

34. Liben-Nowell, D., J. Kleinberg. 2003. The link prediction problem for social networks. In: *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, pp. 556–559, CIKM '03, ACM, New York, NY. doi:[10.1145/956863.956972](https://doi.org/10.1145/956863.956972)
35. Tylenda, T., R. Angelova, S. Bedathur. 2009. Towards time-aware link prediction in evolving social networks. In: *Proceedings of the 3rd Workshop on Social Network Mining and Analysis*, pp. 9:1–9:10, SNA-KDD '09, ACM, New York, NY. doi:[10.1145/1731011.1731020](https://doi.org/10.1145/1731011.1731020).

Trends in Social Network Analysis

Information Propagation, User Behavior Modeling,  
Forecasting, and Vulnerability Assessment

Missaoui, R.; Abdessalem, T.; Latapy, M. (Eds.)

2017, XIII, 255 p. 90 illus., 68 illus. in color., Hardcover

ISBN: 978-3-319-53419-0