HAL open science

# Toward a combinatorial analysis and parametric study to build time-aware social profiles.

Sirinya On-At, André Péninou, Marie-Françoise Canut, Florence Sèdes

## To cite this version:

Sirinya On-At, André Péninou, Marie-Françoise Canut, Florence Sèdes. Toward a combinatorial analysis and parametric study to build time-aware social profiles.. 9th International Conference on Management of Digital EcoSystems (MEDES'17), Nov 2017, Bangkok, Thailand. pp.113-120. hal-03624136

**HAL Id: hal-03624136**

**https://hal.science/hal-03624136v1**

Submitted on 30 Mar 2022

**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 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's version published in: https://oatao.univ-toulouse.fr/22102

**Official URL :**
https://doi.org/10.1145/3167020.3167038

**To cite this version:**

On-At, Sirinya and Péninou, André and Canut, Marie-Françoise and Sèdes, Florence *Toward a combinatorial analysis and parametric study to build time-aware social profiles.* (2017) In: 9th International Conference on Management of Digital EcoSystems (MEDES'17)Conference on Management of Emergent Digital EcoSystems (MEDES), 7 November 2017 - 9 November 2017 (Bangkok, Thailand).

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

# Toward a combinatorial analysis and parametric study to build time-aware social profile

**Sirinya On-at, André Péninou, Marie-Françoise Canut, Florence Sèdes**

IRIT, Toulouse Institute of Computer Science Research
University of Toulouse, CNRS, INPT, UPS, UT1, UT2J

Toulouse, France

{sirinya.on-at, andre.peninou, marie-françoise.canut, florence.sedes}@irit.fr

## ABSTRACT

Research has shown the effectiveness of inferring user interests from social neighbors, also called "social profiling". However, the evolution in the social profile is not widely taken into consideration. To overcome this drawback, we propose a time-aware social profiling method that considers the temporal factors of the information and the relationships between the user and his/her social neighbors. This method aims at weighting user interests in the social profile, by applying a time decay function. The temporal score of a given interest is computed by combining the temporal score of information used to extract the interests with the temporal score of individuals who share the information in the network. The experiments conducted on a co-authorship network, DBLP showed that the time-aware social profiling process applying our proposed time-aware method outperforms the existing time-agnostic social profiling process. The combinatorial analysis and the parametric study led us to observe that in the context of co-authorship network, the individual temporal score has more influence than the information temporal score. As this kind of network does not exhibit a rapid evolution of information and relationships, to obtain a relevant social profile, the information should be damped slowly.

## CSS CONCEPT

• **Information systems** → Information systems application systems and tools

## Keywords

User profiling; Social profiling; Social network evolution.

https://doi.org/10.1145/3167020.3167038

## 1 Introduction

The Web 2.0 drives users in a virtual world composed of web services and online social networks. Such online applications provide different features allowing them to access to digital services and information, to contribute and share information and to exchange with each other. This leads to the emergence of intelligent environments namely "web2.0 ecosystems" composed of multiple entities such as individuals, services, user-generated content (UGC), which are related by interactions and inter-relationships among them. Web 2.0 ecosystem becomes a valuable resource to extract user knowledge. Towards this goal, several works have studied the extraction of user interests and preferences from users' online activities such as posts, comments, interactions among users in social network. This promises benefits for various applications, in particular, for those that rely on information about their users. This work focuses primarily on exploiting web 2.0 ecosystems to extract user interests and build a user profile, which is a key element for personalized systems (e.g. recommendation, personalized information retrieval …). Personalized mechanisms make use of the user profile to propose relevant content according to the user specific needs.

Rather than gathering the user's information or activities to extract his/her interests, we are interested in the social-network based user profiling, which consists in using the information shared by the neighbors in his/her social network to extract his/her (social) interests. In this work, we called this approach, "social profiling" and the profile built within this approach, is known as the "social profile". The underlying idea is based on the homophily and social influence theories that explain the fact that user tend to influence each other via the link in their social network and connect with those who share the common interests. The social profile can be used as an additional profile to the classical user profile to complete the missed information about the user or to propose new content to the user.

Several techniques have been proposed to ensure the quality of interests inferred from user's social neighbors. However, the problem of user interest evolution is not widely taken into consideration. In online social networks (OSNs), user behavior evolves quickly over time. The information and relationships in the user's social network can rapidly evolve and some of them can become obsolete for him/her, as time goes by. It's unlikely that a

person would consider all information of all his/her social neighbors especially those with whom he/she does not regularly and recently interact. Thus, we cannot consider all social neighbors and information shared in user social network as relevant information to reflect the user interests for a period. Building a social profile from user's social network without considering its evolving characteristic may lead to misinterpretation of user long-term interests.

To overcome this issue, we propose a time-aware social profiling method that considers the temporal factors of the information and the relationships between the user and his/her social neighbors. To do so, we extend our existing social profiling processes (individual-based and community-based approach) and integrate the proposed "time-aware method" in the interest extraction step. The method consists in assigning a temporal score to each interest of the social profile according the relevance of the information sources (information and individuals).

This paper is structured as follows. In the second section, we present the related works on social profiling. The third section presents the existing techniques to deal with evolution in social profiling. In the fourth section, we present our time-aware social profiling method. In the fifth section, we describe the experimentation conducted on DBLP. The last section concludes and presents some future works.

## 2 RELATED WORKS: SOCIAL PROFILING

In the social profiling context: user interests are extracted by using the information shared by the neighbors in the user's social network (also called, social neighbors). The social profiling technique and social profile representation can be based on the classical user profiling [1]. The main difference between the social profiling and user profiling is the source of information. In the case of user profile, user interests are extracted from the user's own activities (shared information, search log …). For the social profile, user interests are extracted from the information gathered from the members of the user's social network. Note that the social profile considered in our context is represented by a vector of weighted keywords.

Research has shown that a user's interest may be inferred from his/her social neighbors [2]–[5]. It is, however, challenging to obtain consistently high quality results in inferring user interests from social neighbors. This paper focuses on the evolution of user interests. In the social profiling context, user (social) interests are related to the activities (interactions and shared information) in the user's social network. As social networks evolve over time, it becomes necessary to consider this evolution in the user's social profile. This issue becomes very important in the OSNs context where information and users' relationships can rapidly evolve. In term of information evolution, social events or viral marketing (buzz) increase information sharing, which in turn increase online social content sharing. Thus, we cannot consider all information shared in user's social network as relevant information to reflect the user interests for a period of time. In term of relationships evolution, since links are quickly established in OSNs, two users creating a relationship are not required to know each other in real life. As a result, the links persistency is not always maintained. Thus, the existence of a link between the user and an individual is not always an evidence of the real relationship between them in real life. Building a social profile from user's social network without considering its evolving characteristic may lead to misinterpretation of user long-term interests. However, this issue has not been widely taken into consideration in the social profiling context. For the best of our knowledge, [5] is the only work that considers the evolution of information in the interests weighting step. However, the evolution of relationships between users and his social neighbors are not taken into consideration.

This paper focuses on this issue and tries to answer the following question: how to deal with the social network evolution in social profile building process? In the next section, we provide an overview about existing techniques that deal with the user interests evolution.

## 3 EXISTING TECHNIQUES TO DEAL WITH EVOLUTION IN SOCIAL PROFILING

We can distinguish two main axes for incorporating the evolution of user interests. The first technique is based on a time window or time forgetting technique which selects only the information from the latest time periods [6]. In this approach, outdated information outside the selected time window is completely ignored. However, in some cases, the ignored information could eventually be valuable [7]. The second technique, called instance weighting, consists in weighting different time periods or instances according to their relevance. To do so, time-decay functions, which assign the higher weights to the most recent information, are widely used [8], [9]. We describe below, three temporal functions presented in the literature.

- Linear temporal function: a simple way to calculate a temporal score of a piece of information is to apply a linear function proportional to its publication date.

$$f_{lin}(t) = \frac{1}{t+1} \qquad (1)$$

$t \in \mathbb{N}$ is the distance between the publication date of the information and the current date (at the time of calculation). For example, if the calculation date is 2016, for the information published in 2016 $t$ is equal to 0, for those published in 2015 $t$ is equal to 1, and so on. In other words, $t$ represents the freshness of the information in relation to the current date. The smaller the value of $t$, the more recent the information is.

- Exponential temporal function [10]: the exponential temporal function is widely used in many applications in which it is desirable to gradually decay the history of past behavior as time goes by. The exponential temporal function is as follows:

$$f_{exp}(t) = e^{-\lambda t} \qquad (2)$$

As in the linear temporal function, $t$ represents the information freshness, $\lambda \in [0,1]$ represents the time decay rate. The higher $\lambda$ is, the less important the old information is.

- Polynomial temporal function [11]: for some applications in which exponential decay is too fast and a slower decay is required, we can apply the polynomial decay as follows:

$$f_{poli}(t) = (t+1)^{-\lambda} \quad (3)$$

# 4 PROPOSITION: TIME AWARE METHOD

We have seen the effectiveness of social profiling technique in the existing works. The aim of this work is not to rebuild a novel social profiling process but to improve the existing ones by considering the evolution in user social network. We decide to adopt two social profiling processes introduced in the existing work of our team [4]; the individual based social profiling process (IBSP) and the community based social profiling processes (CoBSP).

In our existing works, [12] proposed a time-aware user egocentric network based user profile building method using the community based approach. This work adopts the instance weighting technique to calculate the temporal score of extracted interests by applying temporal factors and considering both relationships and information evolution. The authors show results of intensive experiments conducted on scientific publication networks DBLP, which represent the effectiveness of the time-aware method against the time agnostic one. [13] proposed to apply on Twitter, the existing time-aware social profiling method proposed in [12]. The obtained results on Twitter are different than those obtained on co-authorship networks, explained by their different social network characteristics.

This paper presents the improved version of our proposed time-aware method along with different temporal score calculation techniques and the combinatorial analysis and parametric study, which allow to study and determine the best combination(s) of techniques and parameters that provide(s) the most relevant social profile for a given social network.

## 4.1 Notation

Throughout this section, we denote $u$ as the focal user (ego) for whom we desire to build social profile. We use the term individuals INDIVS to represent the social neighbors (alters) of $u$. For each individual $indiv \in INDIVS$, the set of information shared by *indiv* is denoted $INFO_{indiv}$. Each piece of information $info_{indiv}$ contains a set of terms that we call here "elements", denoted $E$. The term "time-agnostic social profile" is used to refer to the profile built by using the existing social profiling processes. The term "time-aware social profile" is used to refer to the profile built by applying our proposed time-aware method.

## 4.2 Social profile construction process

We consider two social profiling approaches: individual-based approach (supported by the IBSP) and community-based approach (supported by the CoBSP). Both social profiling processes have the same calculation principle, only the level of calculation is different (individual or community level). We summarize the global view of both social profiling processes for a given $u$, as follows:

- Step1: collect the information from user's social neighbors: for each indiv $\in$ INDIVS we collect the $\mathrm{INFO_{indiv}}$.
- Step2: elements extraction and weighting: for each $\mathrm{info_{indiv}}$, we extract the set of element E and aggregate them at the individual level (for IBSP) or community level (for CoBSP) by using a weighting function. The weight of each element e is a combination of structural score and semantic score. The structural score of e from an indiv (resp. community c), is the centrality value of indiv (resp. c) in the egocentric network. The semantic score of $e$ depends on the time-frequency of $e$. The score combination is made by a linear combination function:

$$score(e) = \alpha * (score_{struct}(e)) + (1-\alpha) * (score_{sem}) \quad (4)$$

$\alpha \in [0,1]$ is the proportion between the structural score and the semantic score.

- Step3: elements combination and social profile derivation: we combine the weighted elements from every individual (for IBSP) or community (for CoBSP) by using linear combination method Lin_ComMNZ [4]. Finally, we derive all aggregated elements considered as user interests to the social profile according to their score. The derived social profile is represented in the form of a vector of weighted user interests.

## 4.3 Time-aware method

To consider the evolution of user's social network, we consider two mains factors related to the social network evolution: one the one hand, the evolution related to the dynamics of the relationships between the user and the individuals in his/her social network and in the other hand, the evolution related to the dynamics of the information shared by the individuals in his social network. The challenge is to study how to integrate these factors into the social profiling process. Integrating the relationships evolution consists in qualifying the most relevant individuals for the user. The "relevant individuals" refers to those having the significant and up to date relationships with the user. Integrating the information evolution consists in selecting only the most significant information for the user from the information shared by the selected individuals. The "relevant information" refers to the significant and up-to date information.

We propose to apply the instance weighting technique (cf. section 3) to select the relevant individuals and information from user's social network, based on their temporal criteria: (i) the timestamp of the interactions between the user and the individuals of his social network and (ii) the publishing timestamp of information

The proposed method is mainly applied to the step 2 of the general social profiling process. Unlike the existing social profiling process (time-agnostic), we calculate the temporal score of each extracted element instead of using the term frequency (*tf) score*, before combining them and deriving the social profile.

The temporal score of an element $e$ extracted from $info_{indiv}$ shared by *indiv* is computed on one the hand, from the temporal score of $info_{indiv}$ , noted as $TS_{inf}(info_{indiv})$ and on the other hand, from the temporal score of *indiv*, noted as, $TS_{ind}(indiv)$ . In a social network, the shared information can be more important than the relations between the individuals and vice-versa. To ensure the genericity of our method for different social networks, we apply a combination technique to combine and adjust the importance of the two weights with regards to the social network characteristics: $TS_{inf}(info_{indiv})$ only, $TS_{ind}(indiv)$ only, or the combination of the two weights with a specific proportion.

Let us note that the techniques used to compute the individual temporal ($Compute_TS_{ind}$ $(indiv)$) and those used to compute the information temporal score ($Compute_TS_{inf}(info_{indiv})$) are customizable. In this work, we apply various techniques to compute these scores. This will lead to study the efficiency of each technique and to choose the most adapted one for a given social network.

#### *4.3.1 Individual temporal score calculation*

As previously mentioned, the temporal score of a given *indiv* depends on the relevance of the relationship between *indiv* and *u*. It can be defined by the strength of relationship between *indiv* and *u*. We assume that the user may be more influenced by the individuals having a strong relationship with him/her. In the literature, there are several measures to calculate the strength of relationship between a pair of individuals in the social network (frequency of interaction, last interaction timestamp …). We can also use topology-based metrics (e.g. number common neighbors, shortest paths, etc.), which are well known and widely applied for link prediction context [14].

Based on these measures, we propose 3 different techniques to compute the strength of relationship between a given *indiv* and *u,* while considering the temporal criteria of their relationships. The three techniques are presented as follows:

*4.3.1.1 Individual score calculation using temporal link prediction*

Rather than using the link prediction to predict the future relationship between two non-connected nodes, we propose to use this technique to compute the persistency of the relationship between two connected nodes, which we consider significant to represent their relationships strength. The two nodes having a persistent relationship have more chance to stay in contact as time goes by. Thus, for a given *indiv* we propose to compute his/her individual temporal score by calculating the link prediction score of *indiv* and *u*. The higher the predicted score is, the more *indiv* is significant for *u*. Based on the link prediction metrics comparative study proposed in [12], we decide to adopt the Adamic/Adar metric, noted as *AA*, which, despite its simplicity, outperforms the other metrics in term 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. To take into account the temporal factor, we adapt a time-aware link prediction technique introduced by [15], which proposes the integration of the temporal score into the existing Adamic/Adar metric. Based on this work, we define the "*temporal Adamic/Adar*" metric, noted as *AATemp*:

$$AATemp(u, indiv) = \sum_{z\epsilon\{\Gamma(u)\cap\Gamma(indiv)\}} \frac{w(u,z)\cdot w(z,indiv)}{log|\Gamma(z)|} \quad (5)$$

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*. We propose to compute the w(x,y) by applying the 3 temporal functions, previously presented in the section 3. This is to study the performance of each function and select the most appropriate function. Let us note $f_{func,}$ a temporal function $func \mid f_{func} \in f_{lin}, f_{exp}, f_{poly}$.

The calculation of the temporal score of a given pair of nodes *x* and *y* is as follows:

$$w(x, y) = f_{func}\left(timestamp_{actual} - timestamp_{last_{interaction(x,y)}}\right) \quad (6)$$

*4.3.1.2 Individual score calculation using link prediction based on the direct link*

With the previous measure, the relations strength between *u* and *indiv* does not depend directly on the links of *u* and *indiv* but on the links between u, *indiv* and their common neighbors. The relationship strength of *u* and *indiv* will not be valued if they have very few common neighbors. This discrimination could be irrelevant for some individuals who interact frequently with the user for a specific topic but does not have a lot of common neighbors.

We proposed a new technique which is based on link prediction technique while relying directly on the links between ***u*** and ***indiv***. We consider that the individuals having the recent interactions with the user should have a greater probability to share the up-to-date interests with the user. The proposed technique consists in assigning the weight to the link between ***u*** and ***indiv*** based on the timestamp of their last interaction, as presented in the formula (7).

$$Direct_{AATemp(u,indiv)} = w(u, indiv) \cdot \sum_{z\epsilon\{\Gamma(u)\,\cap\,\Gamma(indiv)\}} \frac{1}{log|\Gamma(z)|} \quad (7)$$

*w(u, indiv)* represents the temporal score of *u* and *indiv*, calculated by using the same principle as presented in the previous technique.

*4.3.1.3 Individual score calculation using the sum of interaction weight between u and indiv*

We propose in this section an individual score calculation technique based on the number of interactions between *u* and *indiv*. The more *u* interacts with *indiv*, the higher the weight of *indiv* is. To consider the temporal factors, we suggest calculating the temporal score of each interaction: the more recent the interaction is, the more it is considered important and significant. This technique leads to give more weight to the individuals who interact frequently and recently with the user. The calculation of the temporal score of an individual is presented in the following formula.

$$SITemp(u, indiv) = \sum_{interact_i \in INTERACTS} f_{func}\begin{pmatrix} timestamp_{actual} \\ -timestamp_{interact_i} \end{pmatrix} \quad (8)$$

*INTERACTS* represents the set of interactions between ***indiv*** and ***u***.

#### *4.3.2 Information temporal score calculation*

To weight any piece of information $info_{indiv}$ using temporal factors, we propose to apply a temporal function (1), (2) or (3), presented in the section 3. Note that we modify the existing temporal function (1), (2) or (3) by considering the value t in two ways:

- $t_{curr}$: distance between the publishing timestamp of $\boldsymbol{info_{indiv}}$ and the current date.

$$t_{curr}(Info_{Indiv}) = \left|timestamp_{actual} - timestamp_{pub}(info_{Indiv})\right| \quad (9)$$

- $t_{interact}$: distance between the publishing timestamp of the information and the timestamp of interaction between *u* and *indiv*.

$$t_{interaction}(Info_{Indiv}) = \left| \begin{matrix} (timestamps_{lastInteraction}(indiv, u)) \\ -timestamp_{pub}(info_{Indiv}) \end{matrix} \right| \quad (10)$$

Finally, the information temporal score of a given piece of information $\text{info}_{\text{indiv}}$ shared by an indiv is calculated as follows

$$TS_{inf}(info_{indiv}) = f_{func}(t_{curr}(info_{indiv}) + f_{func}(t_{interact}(info_{indiv})) \quad (11)$$

#### *4.3.3 Final temporal score calculation.*

The (final) temporal score of an element *e* extracted from $info_{indiv}$ shared by *indiv* is computed by combining the individual and information temporal scores using the combination function:

$$TS(e) = \gamma * (TS_{inf}) + (1 - \gamma) * (info_{indiv}) | \, e \in info_{indiv} \quad (12)$$

$\gamma$ is the proportion between the individual temporal score and the information temporal score.

We consider the temporal score to weight each element instead of using the *tf* score as done by the IBSP and CoBSP. To combine the calculated elements from all individual/communities and to derive the social profile, we keep using the same technique as presented in the step 3 of the IBSP or CoBSP (cf. section 4.2).

We use the term "temporal social profile" to refer to the social profiling process that apply the proposed time-aware method: IBSPT[1] for the individual based social profiling process and CoBSPT for the community based social profiling process.

## 4.4 Combinatorial analysis and parametric study

In this work, different factors can be customized to build a temporal social profile for a given user:

- Social profiling approach: individual-based (IBSPT) and community-based (CoBSPT).
- Temporal score calculation techniques: AAtemp (5), Direct_AAtemp (7) and SITemp (8).
- Temporal function: $f_{lin}$ (1), $f_{exp}$ (2), $f_{poly}$ (3).

We can also vary the values of involved parameters:

- $\gamma$, presented in the formula (12), represents the proportion between the individual temporal score and the information temporal score while computing the final score of a given element.
- $\lambda$ represents the time decay rate of the temporal exponential and temporal polynomial function, presented in the formula (2) and (3).
- α, presented in the formula (4), represents the proportion between the structural score and the semantic score.

We realize a combinatorial study consisting in combining the three mentioned factors. Let *P* = {IBSPT, CoBSPT} denotes the set of social profiling processes, *T* = {*AAtemp*, *Direct_AAtemp, SITemp*} denotes the set of temporal score calculation techniques and *F* = {$f_{lin}$, $f_{exp}$, $f_{poly}$} denotes the set of temporal functions. The set of combinations of factors is the Cartesian product of *P*, *T* and *F*, noted as $(P \times T \times F)$. We denote $pro_{tech,func}$, the social profile built with a given combination $(pro, tech, f_{func}) \mid pro \in P, tech \in T, f_{func} \in F$. For example, $IBSPT_{AAtemp,Exp}$ represents the social profile built by using the temporal individual social profiling process that applies the *AATemp* technique and the temporal exponential function to calculate the temporal score.

For each combination of factors, we also realize a parametric study consisting in varying the value of the parameters (λ, γ, α) involved in our calculation: the value of each parameter is ranged between 0 and 1 Note that for the linear temporal function, there is no variation for λ.

The aim of the combinatorial analysis and parametric study is twofold: i) to find out the best combination(s) of factors and parameters that provide(s) the most relevant social profile for the user, ii) to study in detail, the effectiveness of each factor and the variation of each parameter for a given social network.

# 5 EXPERIMENTS

The experiments are conducts on the co-authorship network DBLP. In the DBLP network, the nodes represent the authors. Two authors are connected if they published together at least one article. We consider the publication titles as information to extract the element considered as user interests. We use the year of publication as time granularity.

## 5.1 Evaluation protocol

To validate our proposition, we apply the proposed time-aware method to build the social profile by applying the combinatorial analysis and parametric study. Note that within this parametric study, we can build the social profile of the time-agnostic approach (IBSP, CoBSP) by setting the values γ = 0.0 and λ = 0.0. Then we compare the relevance of the profile built by applying different combination and parameters. The strategy is to find out which one provides the social profile the closest to the user profile containing the user's explicit interests. To evaluate the relevance of each social profile we use the precision and the recall measure.

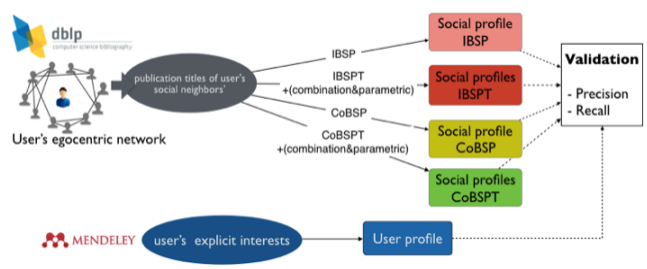


**Figure 1 Evaluation process**

#### *5.1.1 Ground truth extracted from Mendeley profile*

To build the user profile as a ground truth, we use the interests indicated explicitly by the user in his/her Mendeley profile, a reference manager and an academic social network. This allows to use two distinct data sources and to avoid many biases in the results obtained and associated interpretations. First, for each user, we collect the keywords from his/her Mendeley profile. Then we extract interests from the collected keywords using text-processing classical tools: we used dictionaries and thesaurus to merge keywords having the same meaning and we removed empty words using filters, to keep only consistent interests. Finally, we compute the weight of each extracted interest using the *tf* measure that

[1] T for temporal

represents the term frequency of each interest in the set of all found interests [4].

#### *5.1.3 Dataset description*

In our experiment, we consider authors who exist in both DBLP and Mendeley databases and have enough interests in their Mendeley profile to calculate significant and valuable results. Our dataset contains 236 users (authors) having between 5 and 495 co-authors. In total, we collect 10105 scientific authors and 522132 publication titles dated from 1959 to 2017. The experiments were conducted on the platform OSIRIM, available in IRIT laboratory, by using 64 cores, 128 GB of RAM and 265 MB of disk space.

## 5.2 Results

In this section, we present the results of our evaluations. Let us note that to calculate the precision and the recall, we only consider the top N first interest for all built social profiles. In fact, as the built social profiles can contain a lot of interests (> 100), using all extracted interests in the profile in personalization mechanisms could reduce the precision of the personalization mechanism.

We select only the N first interests to calculate the precision and recall. The results presented in this section are calculated by considering the top 5 of interests. Note that the results for the top 10 and more are also satisfying. However, in this case, the results in term of precision can be less significant for users having less than 10 interests indicated in their Mendeley profile: the number of interests found in social profile is increased whereas the number of interests found in the user profile remains less than 10. The results are presented by the average precision (resp. recall) of all users, for each combination of factors and parameters.

#### *5.2.1 Results by setting $\alpha = 0$*

To study the impact of our proposed time-aware method without considering other factors, we propose to set $\alpha = 0.0$. In fact, $\alpha$ is a parameter already integrated in the existing social profiling processes. The Figure 2 represents the results in terms of average precision and average recall according to social profiling approach (IBSP(T), CoBSP(T)), temporal score calculation technique (*AAtemp*, *Direct_AAtemp, SITemp*) and temporal function ($f_{lin}$, $f_{exp}$, $f_{poly}$). The precision (resp. recall) presented in the graph is calculated by applying the optimum value of $\gamma$ and $\lambda$. We observe that, in general, the social profile obtained by applying the proposed time-aware method outperforms that of the time-agnostic method, regardless of the social profiling approach, the temporal score calculation technique and the temporal function.

For the Individual based approach, the best precision and recall are observed by applying the combination $IBSPT_{SITemp,Exp}$. The best precision 0.35993 is obtained by setting $\lambda = 0.01$ and $\gamma = 0.95$. The best recall 0.20421 is observed by setting $\lambda = 0.01$ and $\gamma = 0.95$. We found that the best results of IBSPT outperform the IBSP of 10.45 % and 10.95 % in terms of precision and recall, respectively. For the community-based approach, the best precision and recall are observed by applying the combination $CoBSPT_{Direct_AATemp,Poly}$. The best precision 0.33758 is obtained by setting $\lambda = 0.5$ and $\gamma = 0.95$. The best recall 0.1913 is observed by setting $\lambda = 0.0$ and $\gamma = 0.95$. We found that the best results of the CoBSPT outperform the CoBSP of 13.419% and 14.455% in terms of precision and recall, respectively. One can be surprised by the low scoring results and low obtained gain. We emphasize that the results are compared to optional data from the user's Mendeley profile which therefore might be partly incompatible and/or outdated.

The combinatorial analysis and parametric study allowed us to observe different points:

- Comparing results when considering the social profile construction process, we can find that globally the *IBSP* outperforms the *CoBSP*. These results seem conflicting with those obtained in the work of [4] that demonstrate the efficiency of the *CoBSP* compared to the IBSP. However, we emphasize that the efficient results showed in [4] are calculated from the users having more than 70 co-author and having between 10% and 30% of density in their egocentric network. In this work, the studied users have between 10 and 250 co-authors and the density of the egocentric network is distributed between 5% and 55 %. Thus, it is obvious that the IBSP outperforms the *CoBSP* in this work. The results according to the size and density of user egocentric networks will be studied in our future work.
- Comparing results when considering the temporal score calculation technique, we found that the technique AATemp generally provide the less accurate results. By comparing the Direct_AATemp and SITemp, we found that the optimum technique varies depending on the applied social profiling process. For the IBSPT, the most accurate technique is the SITemp. For the CoBSPT, the most accurate technique is the Direct_AATemp. This can be explained by the fact that for the IBSP which is the individual based-approach, the relevance of a given individual is based on the strength of his/her relations with the user. Thus, the SITemp technique, which is based on the directed interactions between the individual and the user, is efficient to calculate the individual temporal score. For the CoBSPT which is the community-based approach, the communities are extracted based on the interactions and interconnection between the individuals. The members of each extracted community are supposed to be well connected and hence have a lot of common neighbors. In this case, the Direct_AATemp technique which is based on the number of common neighbors becomes efficient to calculate the individual temporal score.
- Comparing results when considering the temporal functions, we observed that by applying the temporal polynomial and temporal exponential function, the results are quite similar. Overall, the linear function provides the less accurate results. In the worst

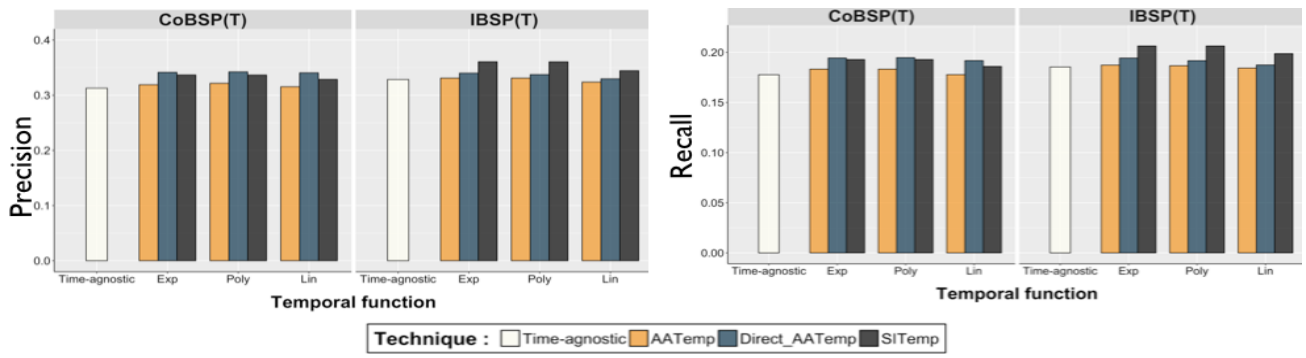


**Figure 2 Global results in terms of average precision and recall, with the optimum value of γ and λ, by setting α = 0**

case, we observed the loss of 1.943 % and 0.666 % in terms of precision and recall respectively, while applying the IBSPAATemp_Lin combination. The temporal linear function gives the less accurate results but is the simplest one. Using this temporal function can help avoiding fixing the value of λ and hence reduce the computation time. However, when the more precise results are required, the exponential or polynomial function is the best and recommended ones.

### *5.2.2 Results according to γ and λ*

In this section, we analyze the variation of the results according to the value of γ and λ, by setting α = 0. The Figure 3 presents the results according to γ (horizontal axis) and λ (different curves) in terms of average precision.

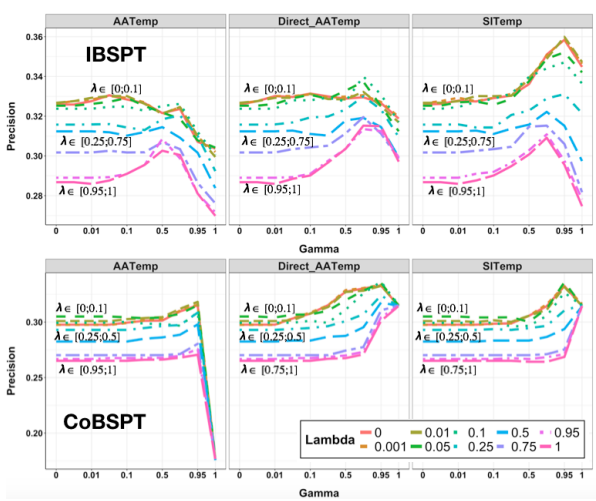


**Figure 3 Results in term of average precision according to γ and λ, by setting α = 0**

- Regarding the results in term of γ, which represents the proportion between the temporal individual score and the temporal information score, we found that the best results are generally obtained when γ is set quite high [0.5; 0.95], regardless the social profile constriction approached and the temporal score calculation technique. This shows the importance of the individual temporal score compared to that of the information, in the context of co-authorship networks. These observations can be explained by the fact that in the co-authorship network, the research field of researchers generally remains stable for a long period whereas the co-authorships can evolve over time. It is often when the authors change the co-authors that they change the domain of interests.

- Regarding the results in term of λ, which represents the time decay rate of the information and relationships, we found that the best results are obtained when λ is very small [0; 0.1]. However, the best result is obtained when $\lambda \neq 0.0$, which means that to obtain the best social profile, we the information and relations should be slowly damped according to their freshness. These observations can be related with the characteristic of the co-authorships networks: as previously described, for the scientific authors, their research field generally remains the same for a long period. This kind of network do not exhibit a rapid evolution of information compared to other OSNs.

### *5.2.3 Results according to α*

In this section, we analyze the variation of the results according to the value of α. The presented results are calculated by using the optimum value of γ and λ. The Figure 4 represents the best results in terms of average precision.

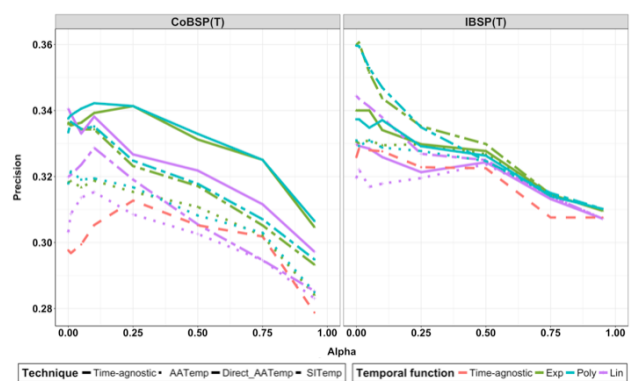


**Figure 4 Results in term of average precision according to α**

We observed that for all value of α the proposed time-aware method outperforms the time-agnostic social profiling process, represented by the orange curve except for those applied the temporal linear function. The best results are obtained when α is

set very low, which correspond with the optimum value obtain in the existing work ([4]).

## 6 CONCLUSION

Considering the evolution of social network, we proposed in this work, a so-called time-aware method that consists in integrating the temporal score of a given user interest in our existing social profiling processes (Individual-based and community-based). The temporal score calculation is based, on the one hand, on the temporal criteria of the piece of information used to extract the interest and on the other hand, on the temporal criteria of the individual who shared this piece of information with the user. The experiments results demonstrate the effectiveness of our proposition and prove so how much is beneficial of leveraging the evolution the social network evolution (information and relationships) in the social profile building process. The combinatorial analysis and the parametric study led us to observe that the optimum temporal score calculation technique or temporal functions can vary according to the adopted social profiling process and may also depend on the type and characteristic of social network. We also observe that the individual temporal score has a larger preference over the information temporal score, in the co-authorship network context. The weight of information and relationships should be slowly damped according to their freshness, as this kind of network does not exhibit a rapid evolution of information.

We plan to apply our approach on other social networks that exhibit a more rapid evolution of information (Twitter, Facebook, Reddit…). We also plan to improve the temporal score calculation technique to enhance the performance of our approach. 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 by using a machine learning approach.

## REFERENCES


[1] 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.

[2] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You Are Who You Know: Inferring User Profiles in Online Social Networks," in *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, New York, NY, USA, 2010, pp. 251–260.

[3] Z. Wen and C.-Y. Lin, "On the Quality of Inferring Interests from Social Neighbors," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2010, pp. 373–382.

[4] D. Tchuente, M.-F. Canut, N. Jessel, A. Peninou, and F. Sèdes, "A community-based algorithm for deriving users' profiles from egocentrics networks: experiment on Facebook and DBLP," *Soc. Netw. Anal. Min.*, vol. 3, no. 3, pp. 667–683, Sep. 2013.

[5] Y. Zeng, Y. Yao, and N. Zhong, "DBLP-SSE: A DBLP Search Support Engine," in *IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT '09*, 2009, vol. 1, pp. 626–630.

[6] M. A. Maloof and R. S. Michalski, "Selecting Examples for Partial Memory Learning," *Mach. Learn.*, vol. 41, no. 1, pp. 27–52, Oct. 2000.

[7] B. Tan, X. Shen, and C. Zhai, "Mining Long-term Search History to Improve Search Accuracy," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2006, pp. 718–723.

[8] D. Li, P. Cao, Y. Guo, and M. Lei, "Time Weight Update Model Based on the Memory Principle in Collaborative Filtering," *J. Comput.*, vol. 8, no. 11, pp. 2763–2767, Nov. 2013.

[9] 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.

[10] Y. Ding and X. Li, "Time Weight Collaborative Filtering," in *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, New York, NY, USA, 2005, pp. 485–492.

[11] G. Cormode, V. Shkapenyuk, D. Srivastava, and B. Xu, "Forward Decay: A Practical Time Decay Model for Streaming Systems," in *2009 IEEE 25th International Conference on Data Engineering*, 2009, pp. 138–149.

[12] 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.

[13] S. On-at, A. Quirin, A. Péninou, N. Baptiste-Jessel, M. F. Canut, and F. Sèdes, "Taking into account the evolution of users social profile: Experiments on Twitter and some learned lessons," in *2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS)*, 2016, pp. 1–12.

[14] 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.

[15] 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, p. 9:1–9:10.