A community based algorithm for deriving users' profiles from egocentrics networks: experiment on Facebook and DBLP

Dieudonné Tchuente, Marie-Françoise Canut, Nadine Jessel, André Péninou, Florence Sèdes

A community-based algorithm for deriving users' profiles from egocentrics networks: experiment on Facebook and DBLP

Dieudonné Tchuente · Marie-Francoise Canut · Nadine Jessel · André Peninou · Florence Sèdes

Abstract Nowadays, social networks are more and more widely used as a solution for enriching users' profiles in systems such as recommender systems or personalized systems. For an unknown user's interest, the user's social network can be a meaningful data source for deriving that interest. However, in the literature very few techniques are designed to meet this solution. Existing techniques usually focus on people individually selected in the user's social network and strongly depend on each author's objective. To improve these techniques, we propose using a community-based algorithm that is applied to a part of the user's social network (egocentric network) and that derives a user social profile that can be reused for any purpose (e.g., personalization, recommendation). We compute weighted user's interests from these communities by considering their semantics (interests related to communities) and their structural measures (e.g., centrality measures) in the egocentric network graph. A first experiment conducted in Facebook demonstrates the usefulness of this technique compared to individual-based techniques and the influence of structural measures (related to communities) on the quality of derived profiles. A second experiment on DBLP and the author's social network Mendeley confirms the results obtained on Facebook and shows the influence of the density of egocentrics network on the quality of results.

Keywords User profile · Social network · Egocentric network · Social profiling · Facebook · DBLP

D. Tchuente (✉) · M.-F. Canut · N. Jessel · A. Peninou · F. Sèdes
IRIT, University of Toulouse, 118, route de narbonne, 31062 Toulouse, France
e-mail: tchuente@irit.fr

1 Introduction

The development of users' profiles is central for mechanisms such as recommendation or personalization of information that correspond to the specific needs of the user. In an information system, a user profile is usually built and enriched in an iterative way from the user's behavior (e.g., rating purchased products, annotating resources, publishing scientific papers) (Gao et al. 2010). The user's profile is usually represented through weighted interests in one or several domains (e.g., culture, sports) (Gauch et al. 2007). The user's interests can also vary according to contextual information (e.g., time, location) (Tchuente et al. 2010). However, the user profile does not always contain all the interests that can be useful for a mechanism of personalization or recommendation. These situations are quite common for new users in the system (their profiles are empty) and for users who are not too active (their profiles do not contain enough interests) (Massa and Avesani 2007). To solve these problems and enrich the user profile when needed, other people's behaviors are usually used to derive interests that could be relevant for the profiled user.

The central issue is: how to choose people from which the user's profile will be derived? The first way to answer this question is to use "similar people" (collaborative filtering techniques) (Massa and Avesani 2007; Esllimani et al. 2011). However, this technique cannot be applied to new users because their profile is empty and thus there is no way to find similar people. Moreover, this technique is very time consuming because the user has to be compared to all other people in the systems, implying the storage and use of huge sparse matrices (Massa and Avesani 2007). To improve collaborative filtering techniques, more and more authors use the user's social network (Kautz et al. 1997; Cabanac 2011; Carmel et al. 2009; Bonhard et al. 2006). This helps to reduce the number of potential people who can be relevant to

the user and can also solve the problem for users with empty profiles (when the social network is known).

In this paper, we are interested in this latter solution. Existing approaches based on this solution can be summarized in two points (Kautz et al. 1997; Cabanac 2011; Carmel et al. 2009; Bonhard et al. 2006; Bender et al. 2008): (1) people used in the user's social network are selected individually, (2) each approach strongly depends on the underlying mechanism that uses generated profiles (e.g., personalization, recommendation) and on each application domain (e.g., search engines, products recommendation). Instead of considering that only some individually selected people in the user's social network are significant to describe the user, we rather consider that the user will be better described by communities of people around him, as already demonstrated in social sciences (Goffman 1959). Thus, we propose a community-based algorithm to derive weighted user's interests from a part of his social network (egocentric network). This algorithm considers the semantics of communities (interests related to a community) and structural measures (centrality measures related to a community) of communities in the egocentric network graph. Additionally, we choose an approach that consists in separating the user profile into two dimensions: the user dimension and the social dimension. These dimensions are independent and can be used by any mechanism or application domain. A first experiment conducted in Facebook demonstrates the usefulness of this two-dimensional representation, the relevance of the proposed algorithm compared to existing algorithms and the influence of structural measures (related to communities) on the quality of the derived profiles. A second experiment with more data on DBLP and the author's social network Mendeley confirms results obtained on Facebook, and shows the influence of the density of egocentric network on the quality of results.

The rest of this paper is structured as follows: in the next section, we present related works. In the third section, we present our methodology and the two-dimensional profile representation. In Sect. 4, we present and describe the proposed community-based algorithm and individual-based ones. In Sect. 5, we present and comment on the results of our first experiment in Facebook. Section 6 presents and comments on our second experiment on DBLP and Mendeley. Section 7 concludes and presents the perspectives of our work.

2 Related works

Recently, some authors have proposed techniques based on the user's social network to improve mechanisms of personalization (Carmel et al. 2009; Bender et al. 2008) or recommendation (Cabanac 2011; Bonhard et al. 2006). The conclusion from all these works is clear: integrating the user's social network in these mechanisms has improved their performances (compared to the case where only the user behavior is used). Thus, to go a step forward and have better results, it is most important to find the best way to derive the user's interests from his social network. That is why we study related works, particularly at the level where they compute interests from the user's social network. For this, we focus on two issues: which part of the social network graph is relevant? How do we choose people who will best describe the user from this part of the graph?

For the first issue, sociology studies (Sinha and Swearingen 2001) as well as some automatic experiments (Bhattacharyya et al. 2011) show that the user's direct relationships (direct neighbors) are more similar to the user than other people in the social network.

For the second issue, studies (Cabanac 2011; Carmel et al. 2009) usually use people individually selected in the user's social network to derive items that could be relevant to the user. For instance, (Carmel et al. 2009) uses the user's social network to improve queries results of a search engine. For each user (u) submitting a query (q), the relevance of each document (d) is computed with respect to: (1) a nonpersonalized score $S_{np}(q,d)$, (2) a personalized score that takes into account the user profile $S_p(u,d)$ and (3) a personalized score that takes into account the user's social network $S_p[N(u),d]$, where $N(u)$ is the list of all people directly connected to the user u in the social network. Similarly, (Cabanac 2011) proposes a social recommender system (in bibliometry) that uses a graph of co-author and a graph of venue (in conferences) to recommend relevant authors to a researcher. Even in this work, people are selected individually on the basis of their topical similarity, their proximity and their connectivity in the co-author graph, and finally their meeting opportunities (number of shared venues) in the graph of venues. In a similar context, (Zeng et al. 2009) and (Ren et al. 2010) propose a social information retrieval system of scientific papers on DBLP. They compute the interests of each author on DBLP (self-retained interests) by using titles of the author's publication. They also compute a second set of author's interests from his co-authors (co-authors interests) by using titles of publication of each author's co-author. They build a DBLP search support engine (DBLP-SSE) which can personalize a search query of an author based on his self-retained interests or co-author's interests. They show that personalized results based on co-author's interests can really improve the personalized results based on self-retained interests.

On the basis of these three examples and other works (Kautz et al. 1997; Bonhard et al. 2006; Bender et al. 2008), we find that:

- Techniques used to exploit the user's social network strongly depends on each author's objective (e.g. application context, personalization, recommendation).
- These techniques rely usually only on individuals selected in the user's social network.

In this paper, we propose a technique to derive some user's interests from his social network that can be reused in any application context and for any mechanism (e.g., personalization, recommendation). Even if techniques based on individual people selected in the user's social network give satisfactory results, we propose an alternative that can give better results by using communities in the user's social network.

3 Methodology and concepts

As stated in the last section, we are interested in techniques for deriving user's interests from his social network independently of the mechanism that can use them. Because existing works (Sinha and Swearingen 2001; Bhattacharyya et al. 2011) show that people directly connected to the user in the social network are most similar to the user, we consider only these people in this paper. As we are interested in communities around the user, for each user (u) we consider the non-oriented graph $G = (V, E)$ where V is the set of people directly connected to the user (user u is not in V) and E is the set of relationships between people in V. This graph V for a user (u) is the egocentric network of this user as already studied in sociology (Masrden 2002). With respect to this graph, the user u is called ego. We are interested in this graph because if a community detection algorithm is executed on this graph, extracted communities will represent groups of people with particular affinities with the user (e.g., family, sports club) (Cazabet et al. 2010). Figure 1 shows a sample egocentric network; tags (e.g., sport club) in this figure have been added manually by the user.

For each domain (e.g., sports, culture), we consider a user profile as composed of a vector of weighted user's interests. We also modelize each user profile as being composed of two dimensions $<P(u), S(u)>$:

- the user dimension $P(u)$ which contains the user's interests in one or more domains and which is computed by using only the user behavior;
- the social dimension $S(u) = P(G)$ which contains the user's interests in one or more domains and which is computed by using the behavior of people in the user's egocentric network (G). This social dimension can be computed either by using people selected individually in G (Cabanac 2011; Carmel et al. 2009; Bonhard et al. 2006; Bender et al. 2008) or rather by using communities in G (as we proposed in this paper).

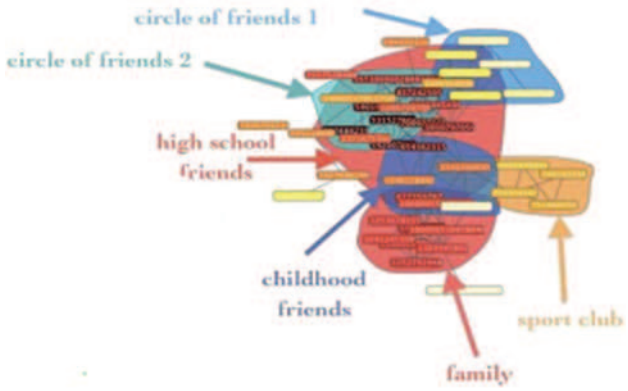

Fig. 1 Example of an egocentric network after executing a community detection algorithm (Cazabet et al. 2010; Ren et al. 2010)

Our goal is to evaluate individuals versus communities techniques for computing the social dimension $S(u)$, so that this dimension will best describe the user dimension $P(u)$. To evaluate this approach for each domain D (e.g., sports, culture), we will compute the cosine similarity between the vector $P_D(u)$ and the vector $S_D(u)$. The technique that will always return the highest value of cosine will be considered as the best technique. Besides the cosine similarity, other measures such as precision and recall can be used to analyze the extent to which the computed social dimension of the user profile can predict the user dimension.

This methodology is independent of any application context or mechanism (e.g., personalization, recommendation). Given the user dimension and the social dimension of a user profile, each mechanism (such as (Carmel et al. 2009) described in related works) can evaluate one or many ways to use these two dimensions to improve their results. For instance, if the user's dimension is empty, only the social dimension can be used. If both dimensions are not empty, they can be combined to improve mechanisms (Carmel et al. 2009). Here, we are interested in finding the best way to derive the social dimension $S(u)$ of the user profile.

4 Algorithms and semantic profile's representation

Given a user u, our aim is to evaluate techniques (or algorithms) for deriving a social dimension $S(u)$ that will best describe the user dimension $P(u)$ of the user's profile. We first present the proposed community-based algorithm (called CoBSP: community-based social profile). Then we present an individual-based algorithm (called IBSP1: individual-based social profile 1) that is similar to existing techniques (Cabanac 2011; Carmel et al. 2009). We finally present a trivial individual-based algorithm (noted IBSP2: individual-based social profile 2). At the end of this section, we present the semantic profile representation that will be useful in evaluations.

4.1 Community-based algorithm (CoBSP)

This algorithm is based on the assumption that the user is better described by communities around him (egocentric network) than by individuals in this network. Sociology study such as (Goffman 1959) has already demonstrated this assumption. It can also be natural to think that if a community in the user egocentric network is characterized by an interest (e.g., sports club in Fig. 1), by affinity this denotes that the user (ego) is certainly related to sports items that characterize this community. In contrast, it is more likely to find an individual strongly interested in "sports" in the user egocentric network, but for which no interests in sports are related to the user's interests in "sports". Thus, we hypothesize that the affinity of the user (ego) with strongly connected users (a community) in his egocentric network is more important than the affinity of this user with a single user in his egocentric network.

Given a user u, with an egocentric network G, the weight $W(i, S(u))$ of each interest i in the social dimension $S(u)$ of the user's profile is computed by the algorithm in Fig. 2.

This algorithm performs in three major steps: (I) community detection in the egocentric network, (II) profiling of each community found in the first step and (III) deriving the social dimension of the user's profile by combining communities' profiles computed in the second step.

(I) The first instruction and first step of the algorithm (line a) consists in finding overlapping communities in the user's egocentric network G. Many algorithms in social network analysis are interested in detecting communities in social networks by using edges between individuals (Cazabet et al. 2010). Some of them can detect overlapping communities (a node can be a member of several detected communities). As communities usually overlap in real egocentric networks (Cazabet et al. 2010), an overlapping algorithm must be used here. The quality of detected communities is usually measured by their modularity (this is based on the proportions of edges internal to communities and the proportion of edges linked to communities) or by their social cohesion (Friggeri et al. 2011). In our case, we used the iLCD algorithm (Cazabet et al. 2010, 2012b) which performs very well with overlap and better than many other algorithms particularly for egocentric networks (Cazabet et al. 2012a). Additionally, the iLCD algorithm is a dynamical one; this means that once communities are detected, when the user adds a new member in his egocentric network, this member is automatically classified into existing communities or new communities. This avoids the overload of re-computing communities when any change appears in the structure on the user's egocentric network. After this first step, the parameter C contains all communities detected by this algorithm.

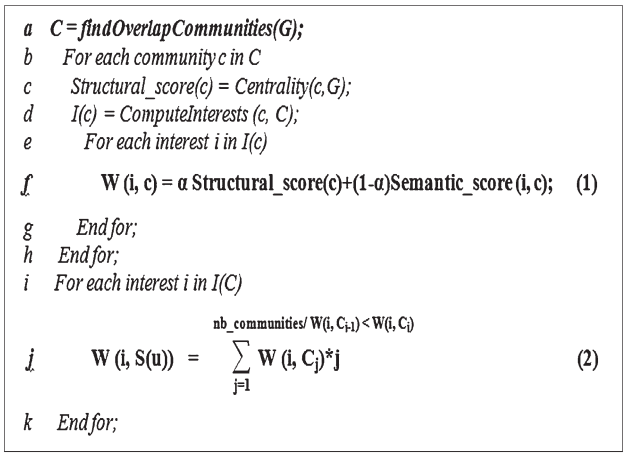

Fig. 2 Community-based algorithm (CoBSP) to derive interests in the social dimension $S(u)$ of the user's profile

(II) The second step of the algorithm (line b to h) consists in computing 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. The set $I(c)$ contains all the community's interests. The weight of an interest i in a community c (called $W(i, c)$) depends on two scores (structural score and semantic score) by a parameter α (formula 1, Fig. 2).

- The structural score of a community c in the ego network G is a centrality measure (e.g., degree, proximity) of this community in graph G (line c). It is important to consider this score because all communities in the user's egocentric network do not probably have the same relevance for the user. A parameter such as the size of the community or the position of the community with respect to other nodes of the graph is important when studying the behavior of communities in social network analysis (Everett and Borgatti 1999). Everett and Borgatti (1999) propose extensions of usual individual-based centrality measures to groups and classes based centrality measures in social networks. For instance, the degree centrality of a community c in a graph $G(V, E)$ is defined as in formula (3): that is, the number of people not in c who are connected to at least one member of c ($|N(c)|$) divided by the number of people not in c ($|V|-|c|$). The lower this measure, the more isolated is the community c in the network. The impact of this kind of structural measure has to be evaluated in the social dimension of the user's profile by the parameter α comprised in [0, 1].

$$\text{Centrality degree}(c, G) = |N(c)| / (|V| - |c|) \quad (3)$$

- The semantic score of an interest i in a community c depends on the weight of this interest for all members of this community. For instance, if interests are computed by analyzing textual information related to users, the weight of an interest can be measured as tf or tf-idf scores (Salton and Waldstein 1978). For a community c, the semantic score of an interest i will be the average of weight of this interest for all members of this community.

(III) The third and final step consists in computing the weight of each interest i in the social dimension $S(u)$ of the user's profile (called $W(i, S(u))$ in formula 2, Fig. 2). From formula 1, an interest i may have a weight in different communities in the user's egocentric network. Due to the assumption explained at the beginning of this section, each weight of the interest i in a community c represents the level of affinity of this community with the user (ego) for this interest. The question now is how to combine these weights to obtain a single weight for the interest i in the social dimension of the user's profile. This combination should take into account the fact that if only one community has a high weight for an interest i, the combination for all communities should return a high weight for this interest. This choice is logical because the more specific a community is concerning any interest, the more this interest can be the affinity between the user and this community. In Fig. 1 for instance, to derive the sports interests of the user (ego), it is logical to focus more on the sports interests of the sports club community in the user's egocentric network. To combine the weight of interests in communities, we use a variant of the function CombMNZ (Fox and Shaw 1994). This function is usually used in information retrieval to solve a problem similar to ours. It is used to merge many search engines by combining scores they each give to a document. When the combined search engine is set to return a high score for a document when at least one search engine has returned a high score for this document, a variant of the CombMNZ function can be used (Hubert et al. 2007). We use this variant in our case by making these two analogies: (a) documents are seen as users' interests, (b) search engines are seen as communities of the user's egocentric network. Thus, we compute the combined weight of the interest in the social dimension of the user profile, $W(i, S(u))$, as the linear combination in formula (2). In this formula, $W_i(C_j)$ is the weight of the interest i in the community C_j as in formula (1). To compute $W(i, S(u))$, communities are ordered increasingly $[W(i, C_{j-1}) < W(i, C_j)]$ according to their weights for this interest. Thus, in the linear combination, if n communities (nb_communities) have been detected in the first step (line a), the weight of this interest in the community which has the lowest weight is not privileged and is multiplied by 1, the second lower weight is multiplied by 2, …, the second higher weight is multiplied by $n-1$ and the highest weight is privileged and multiplied by n.

4.2 Individual-based algorithm 1 (IBSP1)

Individual-based algorithms (Cabanac 2011; Carmel et al. 2009) use individual people (rather than communities) selected in the user's social network. Individual people are usually selected according to the strength of their tie with the user (if this strength is known, of course) (Carmel et al. 2009) or to their centrality values (Cabanac 2011). It is not always easy to define or compute the effective strength of ties in a social network. That is why we choose to use centrality values as the relevance of each individual in the user egocentric network. However, the algorithm can be easily extended to take into account the strength of ties if they are known. Thus, algorithm I1 can be defined as a particular case of the community-based algorithm by considering that each individual (ind) in the user's egocentric network $G(V, E)$, represents a community (Fig. 3). So, the first step of computing communities in algorithm C is not needed here. The structural score is a centrality value of individuals in the egocentric network (e.g., centrality degree of users). The semantic score of an interest i for an individual v, $W(i, v)$, will be the weight of the interest i in the user dimension of the profile of the individual v. The scores combination is made in the same way as in algorithm C.

4.3 Individual-based algorithm (IBSP2)

The second individual-based algorithm considered here is the most trivial one. If V is the set of individuals directly connected to the user and $I(V)$ the set of interests of all users in V, the weight of an interest i in the social dimension $S(u)$ is simply computed by summing the semantic score of this interest for each individual in V (Fig. 4). No structural score is considered here.

```
b   For each individual ind in V
c     Structural_score(ind) = Centrality(v, G);
d     I(ind) = ComputeInterests (ind, V);
e       For each interest i in I(ind)
f         W (i, u) = α Structural_score(ind)+(1-α) Semantic_score(i, ind);  (1)
g       End for;
h   End for;
i   For each interest i in I(V)
                                  nb_individuals/ W(i, Ind_{j-1}) < W(i, Ind_j)
j         W (i, S(u))  =          Σ_{j=1}  W (i, Ind_j)*j                      (2)
k   End for;
```

Fig. 3 Individual-based algorithm (IBSP1) for deriving interests in the social dimension $S(u)$ of the user's profile

4.4 Profile's representation

A user profile is usually represented as a vector of weighted user's interests per domain (e.g., sports, culture) (Gauch et al. 2007). We adopt this representation for both the user dimension and the social dimension of a user profile. We represent the user and social dimension in the same manner to make them similar, so that they are comparable. In our specific context, we choose to organize the domains of a user profile as taxonomy (XML document) such as the one in Fig. 5, for two major reasons:

- Firstly, we choose to represent each user profile with three attributes, because each of them can characterize very particular communities in the user's egocentric network: (1) static attributes (e.g., gender, name) that never change over time. Static attributes can help to detect "static communities" such as family; (2) acquired attributes (e.g., work history, attended schools) that the user acquired at some point and remain unchanged from this point. Acquired attributes can help to detect "acquired communities" such as colleagues; (3) evolutionary attributes (e.g., sports, culture) which are users' interests that vary over time based on the user's behavior. Evolutionary attributes can help to detect "communities of interests" such as sports club (Fig. 1).

For each interest i in I(V)

$$W(i, S(u)) = \sum_{ind}^{V} Semantic\_score(i, ind) \quad (4)$$

End for;

Fig. 4 Individual-based algorithm (IBSP2) for deriving interests in the social dimension $S(u)$ of the user's profile

- Secondly, because we want to build generic profiles that can be used for any mechanism (e.g., personalization, recommendation), it is important to build profiles with many granularity levels. For instance, a mechanism can be interested in having the general user's interests about "sports" at a given time and a specific user's interests about "football" at another. Thus, we represent each attribute of the user profile with the taxonomy of domains (e.g., Fig. 5). The user and the social dimension of each user profile are represented with the same taxonomy. The structure of the taxonomy must exist and be defined by domain specialists before building profiles. When building the user profile, interests are computed on the leaves of the taxonomy. Then they are automatically reported on the top of the taxonomy over the parents of elements as in Fig. 6. In this figure, interests in tennis and football domains (which are leaves of the taxonomy) are computed as presented by algorithms in this section. Interests in the sports domain are automatically computed by summing the weight of each interest in all the children of this domain. If a same interest is found in many children of a node, the weight of this interest in the parent node will be the average of children's weights for this interest. This process is repeated for all the parents until the root of the taxonomy is reached.

In the next two sections, we present two experiments (in Facebook and DBLP) conducted to search for the optimal algorithm (among those presented in the last sections) for deriving the social dimension of a user profile, represented as a taxonomy like the one shown in this section.

5 Experiment on Facebook

We have made a first experiment in Facebook by studying the egocentric networks of 15 very active Facebook users. In this section, we will present the dataset used, our process

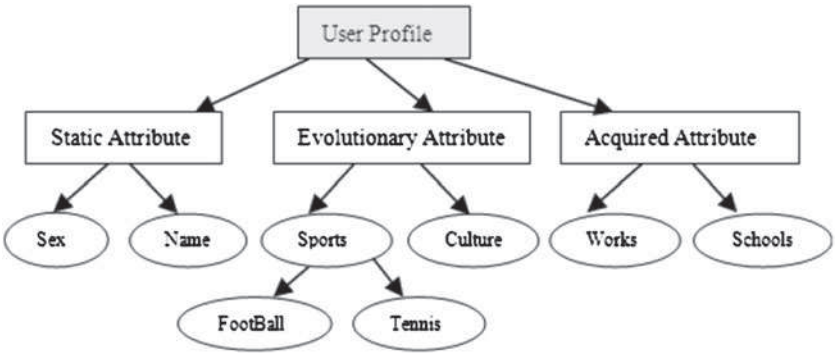

Fig. 5 Example of taxonomy structure used to represent a user or social dimension of a user profile

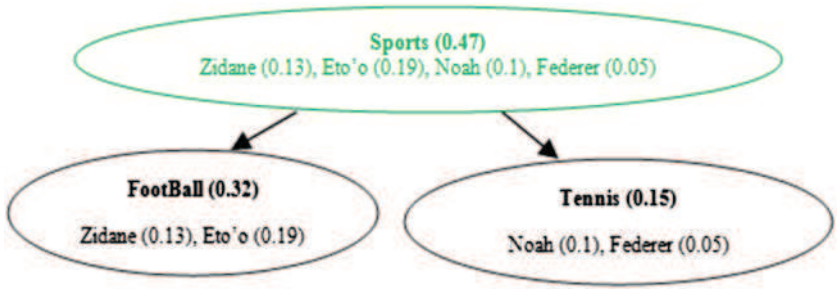

Fig. 6 Example of reporting interests from children domains to a parent domain in the taxonomy

for building profiles in Facebook and the main results of our experiment.

5.1 Dataset

To build the user and social dimension of each user profile, we use users' activities on Facebook (Tchuente et al. 2010, 2012). We use a third-party application (with Facebook API) to access data about users from Facebook. For this, the user must agree to install a third-party application on his Facebook profile. Depending on the data we can use for the evaluation (Table 1), when a user installs a third party in his Facebook profile, the third-party application accesses two categories of data in the user's profile and in his friends' profiles: data accessed automatically and data accessed with the user's explicit authorization (Table 1). As seen in Table 1, we can automatically access the user's egocentric network without his explicit authorization. However, to access all types of attributes (static, acquired, evolutionary) needed to compute the user and social dimension of his profile (Fig. 3), it is mandatory to ask explicit authorization of the user to access further information (specially acquired and evolutionary attributes) from his profile.

Thus, we develop a specific third-party application (https://apps.facebook.com/egoaccess/) dedicated to volunteers who can give us this special authorization. Because the aim of our study is not to break users' privacy, all data are anonymized. The only exception to this rule will be for some of our users who also accept to explicitly validate the relevance of the user and social dimension of their profile built by the process. All the attributes that we really use for building profiles in our experiment are the ones in italics Table 1.

- *Static attributes* We used only "gender" and "explicit interests" provided by users when they registered in the Facebook platform. We did not use attributes such as name because of privacy reasons.
- *Acquired attributes* We used three attributes: the list of occupations held by users, the list of schools attended and the list of their hometown locations.
- *Evolutionary attributes* These attributes are extracted from users' activities such as status published, links published, joining Facebook applications such as "fan pages", "groups" or "events". We only use the activities of users that consist in joining these three Facebook applications for two reasons. Firstly, because the action of joining a "fan page" for instance is more relevant to deduce an interest for the user in the content of this "fan page" than a user's action that can consist in publishing a status or a comment. Secondly, "fan page", "groups" and "events" are already categorized into domains in Facebook. This helps us to reuse an existing taxonomy even if we have rearranged a lot of redundant categories in this taxonomy.

A total of 64 users have been volunteers and have installed our application with explicit authorization to access all data needed in our experiment. However, only 15 users were considered sufficiently active (because they are connected to at least 250 pages, groups or events), to build consistent user dimension of their profile that can be later compared with the social dimension built by each of the three algorithms presented in Sect. 4. These 15 users have an average of 235 friends in their egocentric network and are connected to an average of 235 pages, groups or events. Through these 15 users and their friends, we have access ed and analyzed a total of 3,525 Facebook profiles (Table 2).

For evolutionary attributes, only domains where each user (ego) has at least ten connections were used, because we consider that only these domains will be relevant (consistent) to have a realistic interest in the user's dimension of the user's profile. For our 15 profiled users (egos), these domains are sports, literature, education, music, geography and medias. All these domains are directly children of evolutionary attributes in the taxonomy used here (Fig. 5). They also have subdomains, but we are not interested in the entire taxonomy in this experiment. We will only consider direct children of static, acquired and evolutionary attributes (see Fig. 5).

5.2 Process of building profiles

The process for building profiles (user dimension or social dimension) consists of four steps (Fig. 7).

Step 1 consists in extracting the category and title of all groups, pages and events corresponding to the user (user

Table 1 Data accessed in a Facebook profile by a third-party application

	Data accessed automatically	Data accessed with explicit user's authorization
User		
Egocentric network	Accessed	Accessed
Static attributes	e.g., name, gender	e.g., interests explicitly given by the user
Acquired attribute	Nothing	e.g., work history, schools attended, hometown location
Evolutionary attributes	Nothing	e.g., status, links, notes, photos, videos, groups, pages, events
Friends		
Egocentric network	Nothing	Nothing
Static attributes	e.g., name, gender	e.g., interests explicitly given by friends
Acquired attributes	Nothing	e.g., work history, schools attended, hometown location
Evolutionary attributes	Nothing	e.g., status, links, notes, photos, videos, groups, pages, events

Table 2 Some statistics on data used in the Facebook experiment

Number of egocentric networks analyzed	Average number of people in an egocentric network	Number of Facebook profiles accessed in this experiment	Average number of connections to pages, groups and events per user
15	235	3,525	285

dimension) or each user's friend or communities in the user's egocentric network (social dimension). Each item (here we call item a group, a page or an event) category is matched (projected) to an existing domain on the leaves of the profile taxonomy as described in the previous section. The corresponding leaf will be updated with interests computed from the item title (steps 2 and 3).

Step 2 consists in detecting interests and computing their weight. Interests are detected by mining texts that appear in the title of each item. This approach is similar to building authors' interests in a bibliometric field by mining the titles of all papers published by the author (Cabanac 2011). In this experiment, a text of an item's title is decomposed into semantic units (distinct words) by text separators (e.g., comma, semicolon) (Tchuente et al. 2012). The extracted semantic units pass through a text-mining engine that uses dictionaries/thesaurus (to merge similar semantic units) and filters (to remove empty words with a stop wordlist) to retain only consistent semantic units (Tchuente et al. 2012). These consistent semantic units are considered as interests. The semantic score of each interest is computed by its tf or tf–idf measure (Salton and Waldstein 1978). The structural score (only for the social dimension) is derived from the egocentric network. This can be a centrality measure such as degree, proximity or betweenness (Everett and Borgatti 1999). For this first experiment, we use the degree

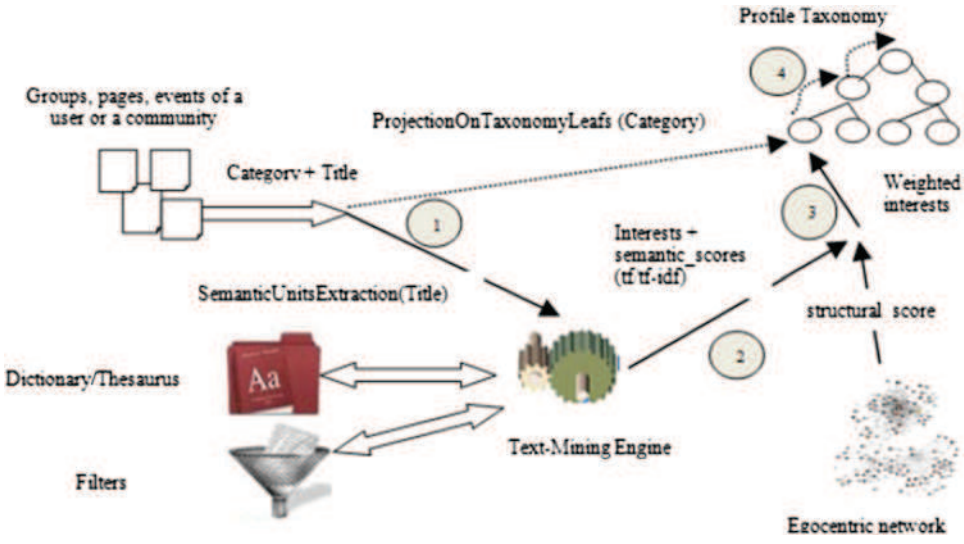

Fig. 7 Process for building profiles in Facebook

centrality which is computed as formula (3) for the community-based algorithm or as formula (4) for the individual-based algorithm.

Step 3 consists in computing the final weight of each interest by merging structural and semantic scores as shown by algorithms presented in Sect. 4. Of course, this step is only used for the social dimension of the user profile. When building the user dimension (by the user connection to groups, pages and events), only the semantic score (tf score here) is reported for use in the next step.

Step 4 consists in reporting interests weights from leaves to the root of the taxonomy as described in Sect. 4 (Fig. 6).

5.3 Evaluation and use of built profiles

As stated in Sect. 4, the user and social dimension are represented with the same predefined taxonomy. For each algorithm, community based (algorithm CoBSP) and individual based (algorithm IBSP1, algorithm IBSP2), a social dimension of the user profile is built and represented. Each social dimension is compared to the user dimension for every domain in the taxonomy, with a top, intermediate or low granularity level (Fig. 8).

Since each domain in the taxonomy is a vector of weighted interests (Fig. 4), we compare the similarity between the user and social dimension by computing the cosine of the angle between these vectors. The higher this cosine value, the smaller is the angle between these vectors. Thus, the algorithm (algorithm CoBSP, IBSP1 and IBSP2) that will build a social dimension that has the highest cosine value with the user's dimension will be the best algorithm.

5.4 Results and comments

For the most active user studied in this experiment, Fig. 9a, b represents tag clouds describing, respectively, the user and social dimension (with community-based algorithm) in the sports domain. We find that all major interests in the user dimension (e.g., basketball, NBA, judo, France, Limoges) are also present and relevant in the social dimension. So, if we consider that the user dimension was unknown, we see that the social dimension derived here would be relevant to the user. The social dimension (Fig. 9b) contains some interests that are not in the user dimension (e.g., football, rugby); thus, these interests can be used to enrich the user dimension (if the user dimension is already known as in Fig. 9a).

The ego in this experiment confirms the relevance of all these interests in both dimensions. Thus, if the user dimension was unknown, for instance, the social dimension computed here can be reported in the user dimension.

Figure 10 presents the comparisons of cosine values between the user dimension and each of the three social dimensions computed by algorithms CoBSP, IBSP1 and IBSP2 (Sect. 4). The cosines values plotted are the average of cosines values from all the 15 user's egocentric network studied in this experiment. These comparisons depend on the parameter α (formula 1, Fig. 1) which allows us to evaluate the relevance of the structural (centrality) measures when deriving the social dimension. Here, the semantic measure used is the tf measure and the structural (centrality) measure used is the degree centrality (e.g., formula 3). Whatever the type of attribute (static, acquired or evolutionary), we observe the same tendencies. The curve representing algorithm IBSP2 is linear because this algorithm does not depend on the parameter α. The two other algorithms (CoBSP and IBSP1) give better results when α is low (between [0, 0.2]) and poor results for other values (generally when >0.2) of this parameter. This indicates that the structural measure is also relevant when computing the social dimension; however, its participation in computing weight of interests should be nearly a fifth (or

Fig. 8 Comparing user and social dimension of a user profile at many granularity levels

Fig. 9 **a** Tag cloud representing the user's profile. **b** Tag cloud representing the social dimension (sport domain) of the user's profile with the community-based algorithm (Algorithm CoBSP)

less) of the semantic measure. Globally, the community-based algorithm CoBSP (proposed in this paper) gives the best social dimension when α is in [0, 0.1]. Additionally, this algorithm is also the most accurate (α in [0, 0.1]), as we can see for the relevance of the corresponding curve (blue) for the top ten interests (a, b, c). For the top 20 and top 50 interests, this algorithm remains the best, however with less relevance than for the top 10 interests. This implies that the social dimension derived from the community-based algorithm cannot sometimes contain most interests (case of Fig. 10a, d, f, g) that are in the user dimension. But what is really relevant is that this algorithm is the one that returns the most important user's interests (Top 10), because in a personalized or recommender system these interests will be relevant.

This first experiment shows the relevance of the proposed algorithm; however, two aspects of this experiment can be improved to have more significant results:

- The number of egocentrics networks studied: because of the difficulty to automatically collect activities data about a user and his friends in his egocentric network on Facebook, the number of egocentrics networks studied here is quite low (15). We think that an experiment with more egocentrics networks can be done to confirm the results obtained.
- The validation methodology: the user dimension and the social dimension in this experiment have been built from the same type of activities data about the user and members of his egocentric network (evolutionary attributes). Because the use of applications such as groups, pages and events spread virally in an online social network such as Facebook, when a user use this kind of application, his friends are more likely to use the same application because of the appearance of the flow of activity of this application on their profile. Thus, when building the user dimension and the social dimension by users' activities in the same platform, a bias can be induced because data used to build these dimensions can derive from a social influence phenomenon. This will not be a problem inside social platforms because social influence is a normal phenomenon. However, in our experiments, we want to show that the social dimension of the user profile is representative of the user dimension independently of an influence phenomenon forced by tools available to users to socialize online. Thus, to have more accurate results,

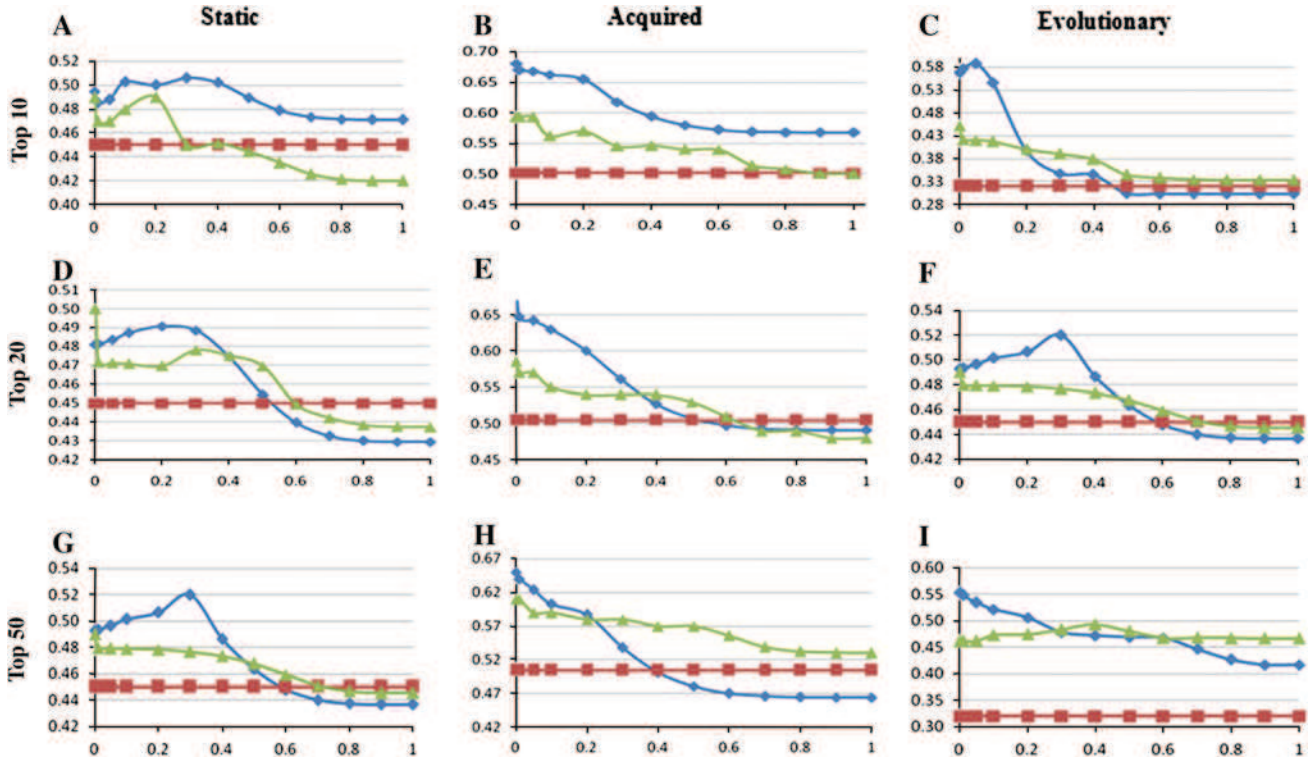

Fig. 10 Graphics comparing the cosine (Y axis) of users' dimensions with socials' dimensions of the 15 egocentrics networks studied (*blue diamonds* community-based algorithm CoBSP, *green triangles* individual-based algorithm IBSP1, *red squares* individual-based algorithm IBSP2) depending on the values of the parameter α (X axis). In the profile taxonomy (Fig. 3), we compared the first three domains defined for all profiles: static (**a**, **d**, **g**), acquired (**b**, **e**, **h**) and evolutionary (**c**, **f**, **i**).To measure the relevance of the best interests computed by each algorithm, we used in the cosine comparison only either the best 10 interests computed (top 10: **a**, **b**, **c**)or the best 20 (top 20: **d**, **e**, **f**)or the best 50 (**g**, **h**, **i**)

we think that we can build the user dimension and the social dimension from two distinct data sources and compare the relevance of our proposed algorithm with individual-based algorithm by these independently built dimensions. Another option would be to ask users themselves to validate their social dimension built from each of the algorithms presented in this paper. However, this option would be time consuming and require an important design time for a correct user interface for testing.

To take into account the two previous aspects and have more consistent and accurate results, we conducted a second experiment from co-authors network in the DBLP database.

6 Experiment on DBLP

In the DBLP bibliography, we use the co-author network for this second experiment. An author's egocentric network is composed of his co-authors and the set of relationship between these author's co-authors. We were interested in the DBLP bibliography because we can address the two aspects mentioned as drawback of the Facebook experiment:

- This database is publicly available; thus, we can analyze a more important number of egocentrics networks (Ley 2009).
- An author's profile can be easily built by analyzing keywords from the titles of his publications (Cabanac 2011; Zeng et al. 2009; Ren et al. 2010). We can thus easily build substantial user dimension of an author's profile with a large number of publications (e.g., 200 publications). However, for testing our algorithms and avoid the bias of using the same data (authors' publications here) to build the user and the social dimension, we choose to build the social dimension with co-author's publications and find another data source where we can access realistic author's interests (user dimension) independently of author's publications in DBLP. This led us to use Mendeley,[1] a scientific authors' online social network. From Mendeley, we can extract interests that authors explicitly fill in their profile. Thus, we integrate two distinct data sources to build the user and the social dimension of an author's profile. These dimensions will then be compared to evaluate the three algorithms explained in this paper.

For explaining this second experiment and results with more details, the next sub-sections consist of presenting the type of datasets used, presenting the process of building and evaluating authors' profiles in DBLP and Mendeley and presenting and commenting on the results obtained.

6.1 Dataset from DBLP and Mendeley

DBLP data are publicly available by the XML API described by Ley et al. (2009). Figure 11 shows three samples of XML files returned when looking for the list of co-authors of the author Dieudonné Tchuente (Fig. 11a), the list of publications of this author (Fig. 11b) and the details about a publication of this author (Fig. 11c).

We built an author's egocentric network by looking for relationships between co-authors of this author. The social dimension of the author's profile is built by mining keywords of publication titles from his co-author's publications (see next section for more details).

The user dimension of an author's profile is built by mining keywords in the list of interests he explicitly gives in the Mendeley social network (see next section for more details). Figure 12 shows a sample of an author profile with his explicit interests (surrounded in the figure) in the Mendeley social network.

Fig. 11 Sample of XML files returned by the DBLP XML API for the author Dieudonné Tchuente. **a** List of co-authors; **b** list of publications; **c** details of a publication

[1] http://www.mendeley.com

Now, the next question is how to integrate these two data sources for our experiment: i.e., for one author, find his egocentric in DBLP and his own profile in Mendeley. In this case, the only same attribute that can help identify the same author in DBLP and in Mendeley is the author's name. So, it is possible to use string matching features such as those used in the semantic Web to match several data sources on the Web. The Mendeley social network has an API (similar to the Facebook API) that a developer can use to automatically extract data on authors' profiles. However, Mendeley allows this data extraction only after each author's explicit authorization. Thus, it is practically impossible to extract automatically the name and interests of all author's profiles of Mendeley. Because of this constraint, we have adopted manual data source integration by identifying manually author names matching in DBLP and Mendeley. We think that analyzing about 100 authors' profiles can be sufficient to have consistent results in our experiment (see next section).

6.2 Building and evaluating authors' profiles from DBLP and Mendeley

The methodology process for building profiles is similar to the one used for Facebook (Fig. 13) except that there is no taxonomy in this case. Research areas of each author are much restricted compared to the diversity of user actions on an online social network such as Facebook. Each dimension of an author profile is thereby composed here with a single weighted vector of author's interest. Thus, this process contains only the first three steps of the Facebook process explained in Fig. 7, with the only difference that textual features used to extract interests came from authors' publication titles from DBLP (for the social dimension) or from the author's list of explicit interests indicated in Mendeley (for the user dimension). Of course, no structural score is used to build the user dimension; so, step 3 is not necessary when building this dimension and only the firsts two steps are necessary (arrows with dashes).

The social dimension built by each of the three algorithms studied in this paper is validated by the validation process described in Fig. 14. This validation process can be described in four steps.

Step 1 consists in identifying relevant authors for our experiment. An author is relevant to our experiment if he has indicated as many as possible interests in Mendeley (so that we build the most realistic user dimension profile) and if he also has enough co-authors in DBLP (so that realistic communities can be found in his egocentric network by the community detection algorithm). In this experiment, we selected only the authors who had indicated at least six interests in Mendeley and who had at least 50 co-authors in DBLP. We manually identified 105 authors who met these two conditions. The average number of interests indicated by these authors in Mendeley was 11 and they had an average of 98 co-authors in DBLP. By analyzing the egocentrics networks of these authors, we reached a total of 10,008 authors in DBLP. Figure 15 represents the number of authors for each number of co-authors. For instance, we can see that only one author had 500 co-authors among the 105 authors studied in this experiment. This figure shows that many of the studied authors have a low number of co-authors (between 50 and 150 co-authors) and only a few have a great number of co-authors (more than 150 co-authors). Thus, this distribution follows a power law distribution as the same distribution for all authors in DBLP (Zeng et al. 2009).

Step 2 consists of building the user dimension of the author's profile (process in Fig. 13).

Step 3 consists of building the social dimension of the author's profile from his egocentric network in DBLP (process in Fig. 13) by each of the three algorithms presented in this paper.

Step 4 consists of evaluating how author's interests built by each algorithm for the social dimension predict the realistic author's interests in the user dimension. The weight computed for interests in the user dimension of an author's profile is not necessarily significant in this experiment. For instance, if an author is more interested in social networks than in data mining, he can express in Mendeley "social network, data mining" as the list of his interests and, only by this list, these two interests will have

Fig. 12 Screenshot of an author's explicit interests indicated in his profile on the Mendeley social network

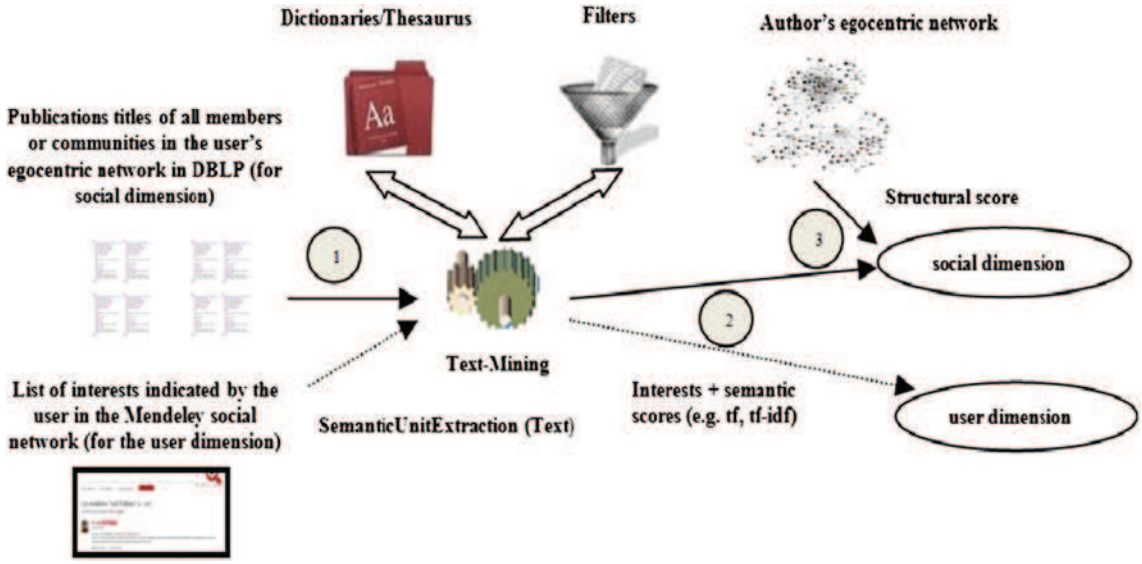

Fig. 13 Process for building profiles from DBLP to Mendeley

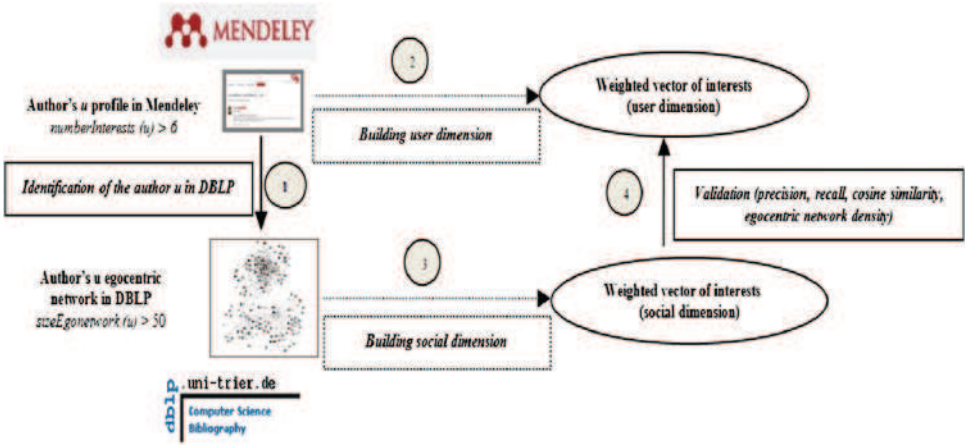

Fig. 14 Validation process in DBLP and Mendeley

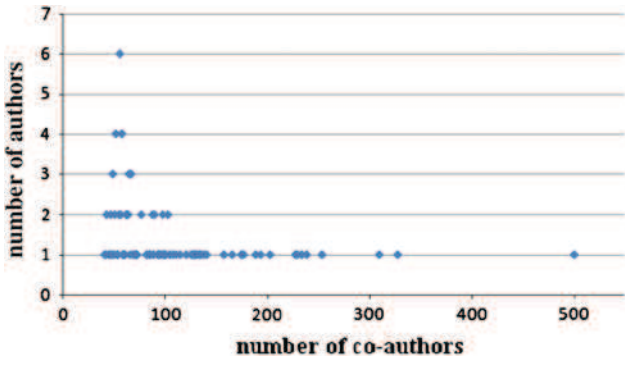

Fig. 15 Distribution of the number of authors for each co-author (for the 105 DBLP and Mendeley authors studied in this experiment)

the same weight in his user profile computed here. Thus, using only the cosine similarity to evaluate the effectiveness of each social dimension is not necessarily sufficient in this experiment. Another way to evaluate the relevance of each social dimension would be to measure simply the percentage of computed interests in the social dimension which are present or not in the user dimension. In this case, we use precision (formula 5) and recall (formula 6) to do that.

If we denote:

$N(I_{su})$: number of interests in the social dimension which are present in the user dimension (number of true positive).

$N(I_s)$: total number of interests in the social dimension.

$N(I_u)$: total number of interests in the user dimension.

Precision and recall are computed as:

$$\text{Precision} = N(I_{su}) / N(I_s) \quad (5)$$

$$\text{Recall} = N(I_{su}) / N(I_u) \quad (6)$$

In this experiment, the number of interests in the user dimension $N(I_u)$ is a finite number (the average is 11 as

indicated above). However, we can have too many interests computed in the social dimension by using relevant semantic units derived from publications of all the author's co-authors. To compute precision and recall, we only consider the $N(I_s) = \text{top } N(I_u) + m$ first interests obtained after building the social dimension ($m = 5$ in this experiment). For instance, if the user dimension of an author's profile contains ten interests ($N(I_u) = 10$), we will consider the social dimension as only the top 15 ($N(I_s) = 15$) first interests computed in the social dimension.

Unlike online social networks where many friends of a user are usually also connected (homophily phenomena) (Aiello et al. 2010), co-authors networks such as DBLP can be relatively less connected. We measure the quantity of relationships between co-authors of an author by the density of the author's egocentric network ($D_{\text{egonetwork}}$). If we denote:

$N(R_{\text{Co}})$: number of relationships between author's co-authors and

$N(PR_{\text{Co}})$: total number of possible relationships between the author's co-authors,

the density of the author's egocentric network is computed as $N(R_{\text{Co}})$ divided by $N(PR_{\text{Co}})$. If an author has n co-authors, the value of $N(PR_{\text{Co}})$ is evaluated as $n \times (n-1)/2$, and thus the density of the author's egocentric network is evaluated as:

$$D_{\text{egonetwork}} = 2 \times N(R_{\text{Co}}) / n \times (n-1) \quad (7)$$

The density of an author's egocentric network can have an impact in the relevance of results of the community-based algorithm (CoBSP). In fact, if an egocentric network is too sparse, we think that the community detection algorithm will tend to discover too many small communities (communities of 1, 2 or 3 users for instance). Thus, the community-based algorithm in this case will tend to give results more similar to individual-based algorithms. In this experiment, we also want to analyze the effect of the user egocentric network density on the quality of results obtained by the community-based algorithm. The density distribution of the 105 authors studied in this experiment is presented in Fig. 16 (each author is represented by a number between 1 and 105). The average density of these authors' egocentric network is about 0.1 (10 %). This average is also the median density because there is almost an equal repartition of author's egocentric network density above and below 0.1.

In the next section on the results of this experiment, we will also analyze the impact of author's egocentric network on the relevance of studied algorithms.

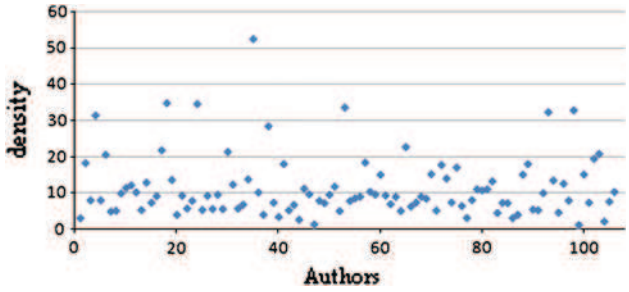

Fig. 16 Egonetwork density distribution of 105 studied authors (each author is represented by a number between 1 and 105)

6.3 Results and comments

Figure 17 shows the comparative curves (precision, recall, cosine) of the three algorithms presented in this paper for all the 105 authors' egocentric networks studied in this experiment. All the curves have the same tendencies as those observe in the Facebook experiment (Fig. 10). Algorithms tend to build best social dimensions (which best predict interests in the user dimension) for small values of the parameter alpha. However, we find that the community-based algorithm (CoBSP) is not absolutely the best algorithm even if it gives best results in terms of precision for values of alpha in [0, 0.1]. The comparison using the cosine similarity tends to give less important results for the CoBSP algorithm with respect to precision and recall. This can be explained by the fact that weights computed for interests in the user dimension are not necessarily relevant as indicated in the last sections. Thus, we think that precision and recall give more accurate results than the cosine similarity in this experiment.

As explained in the last section, we think that the community-based algorithm should give better results for authors with more dense egocentric network. For analyzing the impact of author's egocentric network density, Fig. 18 presents the same comparative curves but only for authors with an egocentric network density $\geq$0.1 (10 %).

Unlike the comparison with all studied authors, we clearly see that the community-based algorithm outperforms individual-based algorithms. This is particularly important in terms of precision and recall. We still observe best results when the alpha parameter is in [0, 0.1], and the cosine similarity still gives less good results (certainly for the reasons already mentioned above in this paper). This result is consistent with our assumption that the more an egocentric network is dense, the more the community-based algorithm will give best results. To go further in the confirmation of this assumption, we did a third comparison by using only authors with an egocentric network density $\geq$0.2 (20 %) (Fig. 19).

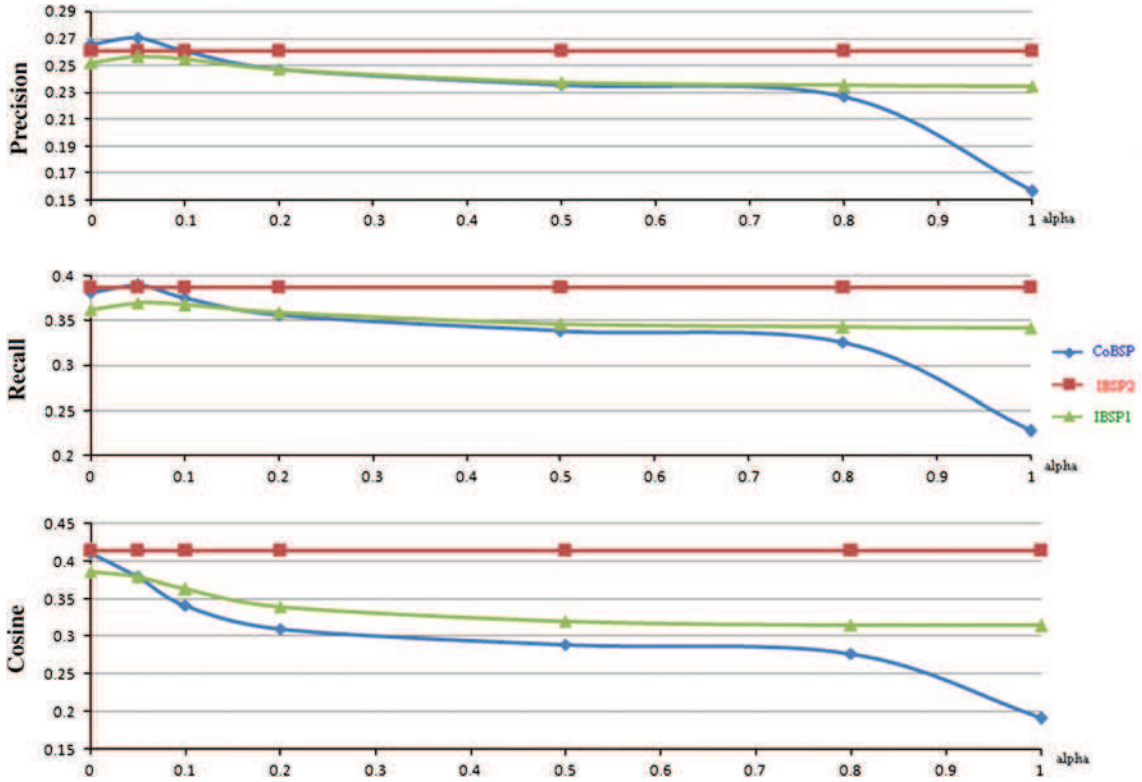

Fig. 17 Comparison (precision, recall and cosine) of the user dimension with the social dimensions built by algorithms CoBSP, IBSP1 and IBSP2 for all the 105 studied authors

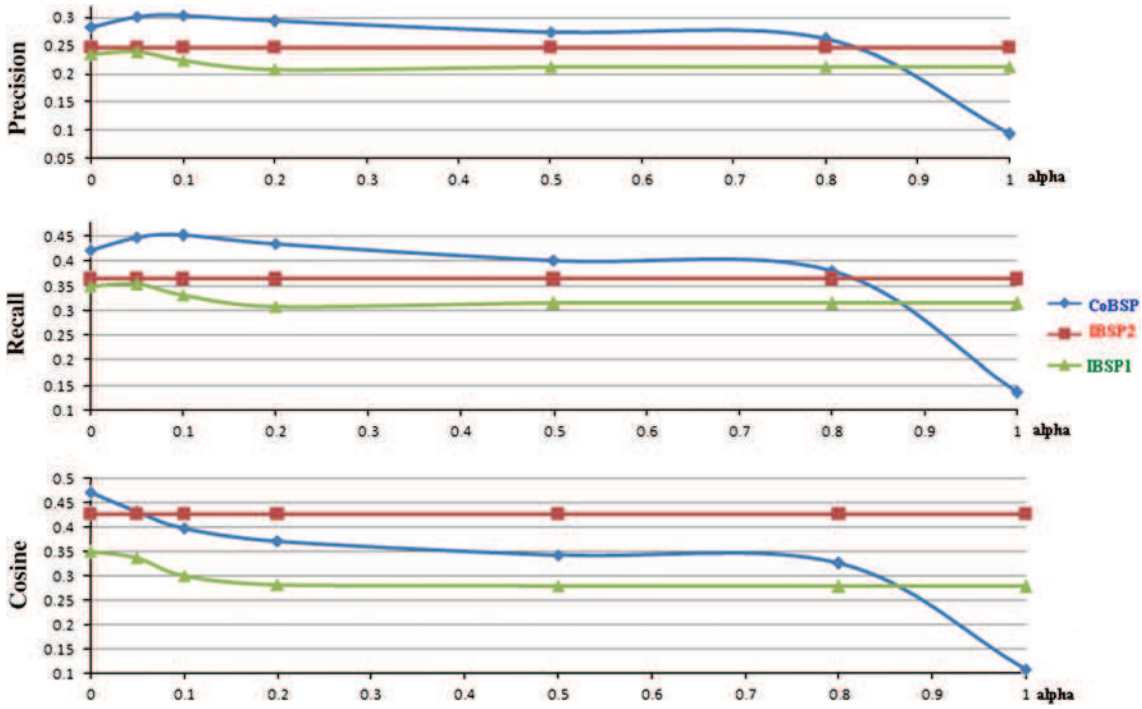

Fig. 18 Comparison (precision, recall and cosine) of the user dimension with the social dimensions built by algorithms CoBSP, IBSP1 and IBSP2 for only authors with egocentric network density $\geq$0.1 (10 %), 51 authors

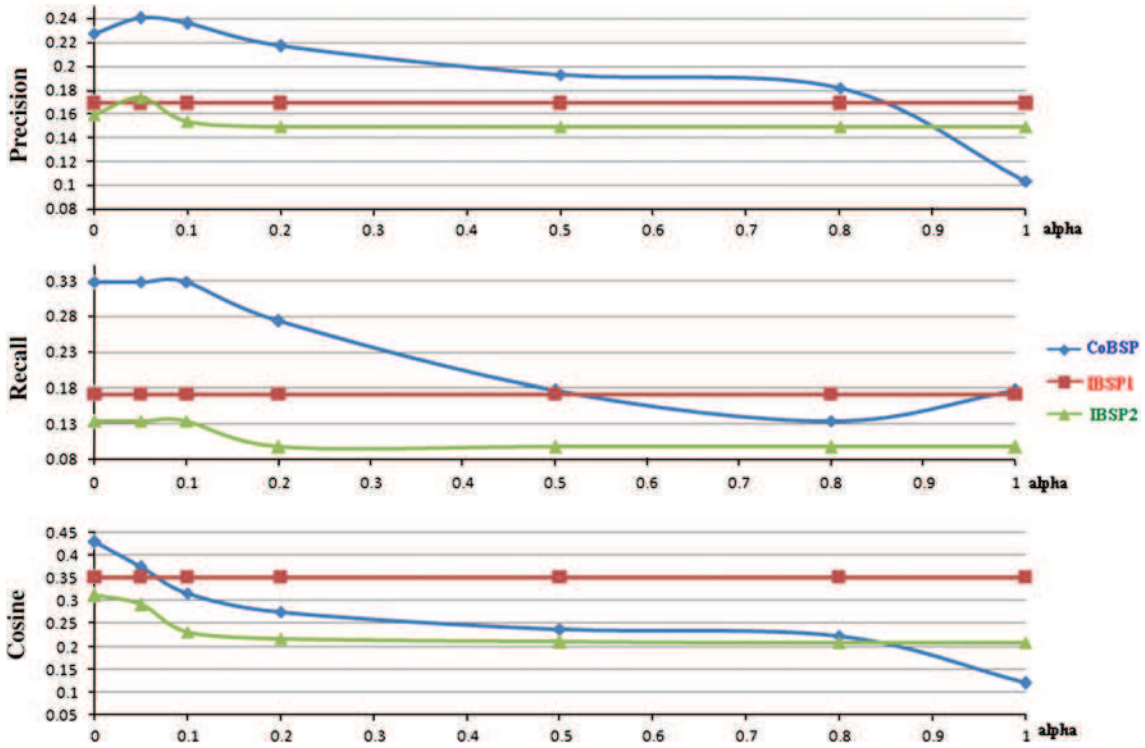

Fig. 19 Comparison (precision, recall and cosine) of the user dimension with the social dimensions built by algorithms CoBSP, IBSP1 and IBSP2 for only authors with egocentric network density ≥0.1 (20 %), 22 authors

Once more, it is clear that the community-based algorithm proposed in this paper gives better results when the author's egocentric network in denser. The gap between the community-based algorithm curves (CoBSP) and individual-based algorithms (IBSP1, IBSP2) is much more important.

Finally, this second experiment with more data in DBLP and Mendeley allows us to deduce three conclusions:

- The community-based algorithm (CoBSP) gives in general better results than individual-based algorithms.
- These results are more relevant when the user's egocentric network is more dense, particularly when this density becomes greater than 0.1 (10 %).
- Structural scores (centrality degree of community in this case) can help the improvement of results when the parameter alpha is smaller than 0.1.

7 Conclusion and perspectives

In this paper, we have presented a community-based algorithm for deriving a social dimension of a user profile that can be relevant to enrich the user dimension. The social dimension is computed by analyzing the user's egocentric network behavior. The user dimension is computed by analyzing the user's own behavior The built profiles are generic, structured with high and low granularity levels, so that they can be used for any mechanism (e.g., personalization, recommendation).

A first experiment done by analyzing 15 egocentrics networks in Facebook gives us many indicators that show that the proposed community-based algorithm outperforms individual ones. This experiment also shows that using community centrality measure (degree centrality of communities in this case) when building weights of interests in the social dimension of the user profile can improve results. These results also show that in platforms such as online social networks, the user's privacy can be really protected only if the user can make not only his profile information, but also his friends' list private.

To confirm these results with a bigger dataset, we performed a second experiment on DBLP which is a public and open database. The process of this second experiment was based on two distinct data sources: the Mendeley social network was used to extract explicit authors' interests (user dimension) and the DBLP database was used to build the social dimension of authors' profiles by using publications of their co-authors. Because we can access more data on this experiment, we also analyzed the relevance of the proposed algorithm with respect to the density of authors' egocentric networks. The results obtained in this experiment confirm the first results obtained from Facebook. Furthermore, analysis according to densities of

author's egocentrics networks shows that the more an egocentric network is dense, the more the community-based algorithm gives best results.

There are many perspectives for this work. The most important are the evaluation of the proposed algorithm on other data sources such as Twitter data and the evaluation of the impact of other communities centrality scores (e.g., proximity, betweenness). Finally, since the derivate social dimension of the user profile by the community-based algorithm is more relevant, it will be interesting to integrate the user and social dimension of the user profile in personalized and recommender systems.

References

Aiello LM, Barrat A, Cattuto C, Ruffo G, Schifanella R (2010) Link Creation and Profile Alignment in the aNobii Social Network. SocialCom/PASSAT pp 249–256

Bender M, Crecelius T, Kacimi M, Michel S, Neumann T, Parreira JX, Schenkel R, Weikum G (2008) "Exploiting social relations for query expansion and result ranking" In: IEEE 24th International Conference on Data Engineering Workshop, ICDEW 2008. vol no 7–12 pp 501–506

Bhattacharyya P, Garg A, Felix WuS (2011) Analysis of user keyword similarity in online social networks. Soc Netw Anal Min 1(3):143–158 Springer

Bonhard P, Sasse MA (2006) "Knowing me, Knowing you – using profiles and social networking to improve recommender systems", BT Technology Journal, vol 24 No 3

Cabanac G (2011) Accuracy of inter-researcher similarity measures based on topical and social clues. Scientometrics 87(3):597–620

Carmel D, Zwerdling N, Guy I, Ofek-Koifman S, Har'el N, Ronen I, Uziel E, Yogev S, Chernov S (2009) "Personalized social search based on the user's social network", In: 18th ACM conference on Information and knowledge management (CIKM '09). ACM, New York, NY, USA, pp 1227–1236

Cazabet R, Amblard F, Hanachi C (2010) "Detection of overlapping communities in dynamical social networks" In: IEEE second international conference on social computing (social com), 2010 vol no. 20–22, pp 309–314

Cazabet R, Maud L, Amblard F (2012) Automatic community detection in online social networks: useful? efficient? asking the users. In: The 4th international workshop on web intelligence and communities, Lyon, pp 6.1–6.8

Cazabet R, Takeda H, Hamasaki M, Amblard F (2012b) Using dynamic community detection to identify trends in user-generated content. Soc Netw Anal Min 2(4):361–371 Springer

Esllimani I, Brun A, Boyer A (2011) Densifying a behavioral recommender systems by social networks link prediction methods. Soc Netw Anal Min 1(3):159–172 Springer

Everett MG, Borgatti SP (1999) The centrality of groups and classes. J Math Sociol 23(3):181–201

Fox EA, Shaw JA (1994) "Combination of multiple searches, the 2nd text retrieval conference (TREC-2)", NIST Special Publication 500–215, pp 243–252

Friggeri A, Chelius G, Fleury E (2011) Triangles to capture social cohesion CoRR abs/1107.323

Gao M, Liu K, Wu Z (2010) Personalisation in web computing and informatics: theories, techniques, applications, and future research. Inf Syst Frontiers 12(5):607–629

Gauch S, Mirco S, Aravind C, Alessandro M (2007) "User profiles for personalized information access". In: The adaptive web, vol. 4321, pp 54–89

Goffman E (1959) "The presentation of self in everyday life" Garden City, NY, 2002

Hubert G, Loiseau Y, Mothe J (2007) Etude de différentes fonctions de fusion de systèmes de recherche d'information, CIDE 10: Nancy, 02/07/2007-04/07/2007, EUROPIA, pp 199–207

Kautz H, Selman B, Shah M (1997) Referral web: combining social networks and collaborative filtering. Commun ACM 40:3

Ley M (2009) DBLP—some lessons learned. PVLDB 2(2):1493–1500

Masrden PV (2002) "Egocentric and sociocentric measures of network centrality", Social networks, Vol 24, No 4 pp 407–422

Massa P, Avesani P (2007) "Trust-aware recommender systems". In: Proceedings of the 2007 ACM conference on recommender systems (RecSys '07). ACM, New York, NY, USA, pp 17–24

Ren X, Zeng Y, Qin Y, Zhong N, Huang Z, Wang Y, Wang C (2010) Social relation based search refinement: let your friends help you!, International conferences on active media technology, AMT 2010 pp 475–485

Salton G, Waldstein RK (1978) Term relevance weights in on-line information retrieval. Inf Process Manage 14(1):29–35

Sinha R, Swearingen K (2001) "Comparing recommendations made by online systems and friends" In: DELOS-NSF workshop on personalization and recommender systems in digital libraries

Tchuente D, Canut CMF, Jessel NB, Péninou A, Haddadi AE (2010) "Visualizing the evolution of users' profiles from online social networks". In: ASONAM 2010, Odense–Denmark pp 370–374

Tchuente D, Canut CMF, Jessel NB, Péninou A, Sedes F (2012) "Visualizing the relevance of social ties in user profile modeling". WIAS (Web Intelligence and Agent Systems) An international journal 10(2):261–274

Zeng Y, Yao YY, Zhong N (2009) DBLP-sse: A DBLP search support engine. In: Proceedings of the 2009 IEEE/WIC/ACM international conference on web intelligence, pp 626–630