*January-February 2020*
*ISSN: 0193-4120 Page No. 9918 - 9924*

# An Egocentric Algorithms of Extractions of the Interests of the users from the RSN

**Yassir Aadil, Ibtissam El Achkar, Hind Ouzif, Mohamed Sillare, and El Houssine Labriji**

Faculty of Sciences Ben M'sick, University Hassan II, Casablanca, Morocco y
Laboratory of Technological Information and Modeling

***Article Info***
***Volume 82***
***Page Number: 9918 – 9924***
***Publication Issue:***
***January-February 2020***

***Abstract:***
The search for information (RI) personalized tends mainly to model the user according to a profile and then to integrate it into the chain of access to information, in order to better meet his specific needs.

The user profile is a central element in the information adaptation systems. We are interested in the process of enriching the user profile from his social network, which represents a very rich source of information about the user, but the problem that arises is that the user profile may not contain all the interests and information that may be useful for a given mechanism, especially for new users of the system and those who are not very active. In order to solve these problems we will use a mechanism to detect users similar to this user in the system, and analyze their interests, using similarity techniques, and consequently use the CoBSP algorithms by improving its performance in a very remarkable way by extending the user's 1-egocentric network, which allows to add more nodes from different social networks to cover all the interests of the user.

So our proposal is to combine the user's (friend's) relationships on different types of social networks (here Facebook, Twitter and LinkedIn) in a single social egocentric graph before the application of the CoBSP algorithm whose goal is to obtain a more complete User Profile that brings together the majority of the user's interests.

***Article History***
***Article Received****: 18 May 2019*
***Revised****: 14 July 2019*
***Accepted****: 22 December 2019*
***Publication****: 17 February 2020*

***Keywords****: User Profile, egocentric algorithm, individual algorithm, social network, Information System, information retrieval system.*

## 1 INTRODUCTION

Who says Web says data, and who says data says treatment. Especially since the latter does not cease to develop, and consequently the number of digital data increases exponentially, and this increase in data is due to several factors such as the advent of Web 2.0 then 3.0 and the appearance of several programming languages, which have facilitated the creation of thousands of dynamic sites (1 billion and 900 million sites www.internetlivestats.com), various sources of data sharing, in addition to the use of multiple interfaces and IT tools and the emergence of cloud computing with mass data storage techniques (Big data). So publishing content on the web has become a very easy task, but access to this content has become difficult for users, given the diversity and the enormous amount of information that is supposed to interest them.

Faced with this data, the user faces problems of overload and disorientation, and finds it difficult to find the information that best meets his needs, for example the case of a nutritionist who uses the 'Apple' query to find information about the benefits of the apple, but it receives a mix of information about Apple computer and apple fruit, which increases the response time and find the appropriate information for the user. This kind of problem, it is found in a lot of computerized domain such as E-commerce sites, which must manage a large number of customers that is increasing daily, and a huge database of products or services they offer and so

January-February 2020
ISSN: 0193-4120 Page No. 9918 - 9924

they must find a better method to improve their customer services, and know which product or service for which customer whose purpose is to increase their sales and make adequate decisions in real time to meet their needs. Clients. The researchers proposed as a solution, information retrieval systems based on the "U" process of information retrieval, in order to improve conventional information retrieval systems, and overcome the main problematic studied by (J. Budzik , K. Hammond, 2000) which resides in the fact that the classical search engines are based on a generalist approach which considers that the need for information of the user is completely represented by his request and then deliver results taking into account only the selection criteria by content and the availability of information sources, and as a result they return the same list of results for the same query submitted by users being in different search contexts and therefore having different information needs.

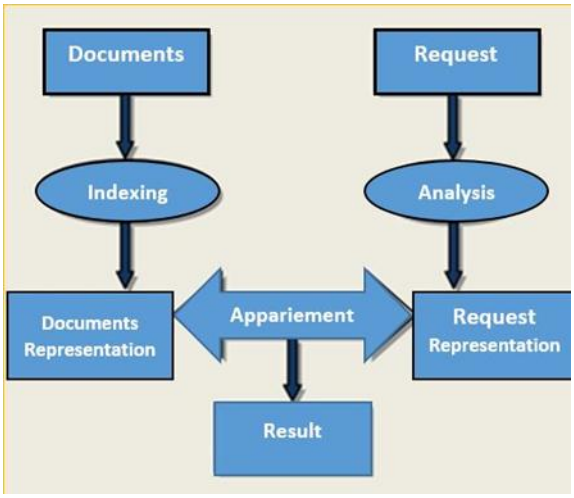



*Figure 1 : U-process of Information Retrieval.*

According to(M. DAOUD, 2009) there are 3 main techniques developed in adaptive IR:

**1) Customized selection of information**

**2) Reformulation or expansion of the request**

**3) Reordering the results**

The purpose of these techniques is to improve the search results of the user by exploiting in addition to its request additional information extracted from the user's interactions with the system.

An analysis made by(Omar Hasan ; Benjamin Habegger ; Lionel Brunie ; Nadia Bennani ; Ernesto Damiani, 2013) on the major European EEXCESS project (www.eexcess.eu) shows that the best additional information that can be integrated in information retrieval processes is the use of the profile of user.

In the rest of this article, we will present the user profile, its main representations, models, data sources, and the reason for the emergence of the use of user data on its social networks to build and enrich its profile. we will also present two main types of user profile enrichment algorithms from social networks as well as a comparative study between these types, and will end with our contribution to develop one of these algorithms while by citing the methodology when will follow and the challenges or obstacles that must be overcome.

## 2 THE ENRICHMENT OF THE PROFILE OF THE USER

The user profile may not contain all the interests and information that may be useful for a given mechanism, especially for new users of the system and those who are not very active. In order to solve these problems and enrich the user profile as needed, these mechanisms tend to detect users similar to this user in the system, and analyze their interests, using similarity techniques, and as a result to use algorithms and processes specific to the need of each mechanism to enrich the user profile and predict its behavior from those users who like them, except that this technique is not applicable in the case of a user's cold start, and for low-power users, more than that this kind of technique requires a lot of time because it compares each user to all users of the system, and it involves storage techniques and the use of huge sparse matrices.These issues have led researchers to analyze the data of users and their networks of friends on its social networks, because the individuals of the social network of the user are the most similar to him. This has given rise to a new approach to filtering information: social filtering, the

January-February 2020
ISSN: 0193-4120 Page No. 9918 - 9924

latter is particularly interested in individuals of the social network of the user who would not only be similar to him, but who would also be the most likely to influencing and influencing one's interests and ultimately one's profile.

The huge number of users that increases, and by tracking the number of data, reinforces this type of filtering, the following figure shows statistics taken by www.statista.com on the evolution of the number of users on social networks.

So in the rest of this chapter we will present some concepts and enrichment work of the user profile from these social networks, and some enrichment algorithms, and we will make a comparative study between them to introduce our contribution.

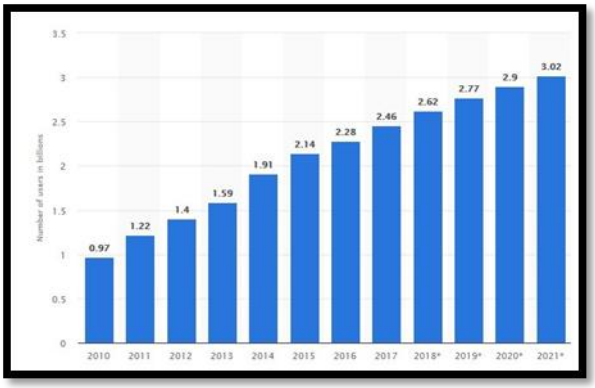



*Figure 2: Number of social media users worldwide from 2010 to 2018 (in trillions)*

## 2.1 The use of social networks in the enrichment of user profiles

The enrichment of the profile of the user of social data is based on the concept of devising the profile of the user in two dimensions:

- **The user dimension PD (u)**: is the most important of the profile because it contains the elements of information and interactions of the user with the system. However, if, for example, information is missing in this dimension, then the social dimension can be exploited by the mechanism.
- **The social dimension SD (u)**: is complementary to the user dimension, it contains the elements built from the information and interactions of the communities of the social network of the user, and can be exploited according to the needs of each mechanism.

These two dimensions can be exploited according to the needs of each mechanism: For example, if the dimension of the user is empty, only the social dimension can be used. If the two dimensions are not empty, they can be combined to improve the mechanisms(CARMEL D., ZWERDLING N., GUY I., OFEK-KOIFMAN S., HAR'EL N., RONEN I., UZIEL EREL., YOGEV S., CHERNOV S, 2009).

Here we are interested in the best way to derive the social dimension SD (u) from the user's profile. So the question is how to select in a relevant way in the social graph of the user the individuals on which we will rely to create the social dimension of the user to enrich his profile.

## 2.2 Types of graph analysis in social networks

In the literature the graph analysis categories in the most used social networks are:

**- Egocentric analyzes**: example of (ZENG Y., YAO Y.Y., ZHONG N, 2009) which usually focus on individuals located at distance 1,(called alters) from the user (called ego).

*Figure 3:The K-egocentric network*

**- Socio-centric analyzes** : Analyzes the entire graph, which is unacceptable for the type of system that require profiles built in real time, and up to date, because the calculation time of centrality measures of individuals throughout the graph is very high. Works such as(CABANAC, 2011) demonstrate that

January-February 2020
ISSN: 0193-4120 Page No. 9918 - 9924

an individual may only be characterized by egocentric network analysis, especially since the computing times in this type of network will remain much more reasonable for the analysis of a network. Ego, A dozen or even hundreds of nodes only. This may also justify the fact that the majority of the works are interested only in the close neighborhood (distance 1 most often) of the user.

### 2.3 Egocentric algorithms for enriching the user profile

The question that arises is how to exploit this egocentric network, some authors consider all the alters, others only subsets deemed more relevant (those with the strongest links between the ego and users). It is obvious that the alters of an ego do not have the same importance for him, so they should not be considered in the same way in the analysis of the ego. Indeed, the weak links of the user are considered as vector of new information for the ego. In the same way, for the user's profile, his weak links would also be relevant to derive new information potentially useful for him. In the end, relying solely on the strong links of the user is not enough, we should also take into account the weak links to be more optimal.

This gave rise to the two types of egocentric algorithms: individual algorithms and community algorithms depending on how it handles the alters of the user's social graph.

#### 2.3.2 Egocentric algorithms based on individuals

This type of algorithms is divided into two main categories as we have seen previously:

- The algorithms that use and analyze the data of all the individuals of the
- egocentric network of the user.
- The algorithms that analyze only the individuals of the egocentric network of the user who have a stronger link with the user (ego).

Among the most well-known individual algorithms are:

**_Individual-Based Algorithm1 IBSP1** :This algorithm analyzes only the data of the individuals of the egocentric network of the user who have a stronger link with him.

*For each individual ind in V*
*Structural_score(ind) = Centrality(v, G);*
*I(ind) = ComputeInterests (ind, V);*
*For each interest i in I(ind)*

$$W(i, u) = \alpha\ \text{Structural\_score(ind)} + (1-\alpha)\ \text{Semantic\_score(i, ind)};$$

*End for;*
*End for;*
*For each interest i in I(V)*

$$W(i, S(u)) = \sum_{j=1}^{nb\_individuals/\ W(i, Ind_{j-1}) < W(i, Ind_j)} W(i, Ind_j) * j$$

*End for;*

The structural score is a value of centrality of the individuals in the egocentric network (e.g., degree of centrality of the users). W (i, u) will be the weight of interest i in the user dimension of the profile of the individual u.

**_ Individual-Based Algorithm 2 :IBSP2** :
This algorithm analyzes the data of all the individuals of the egocentric network of the user whatever the nature of the links between the ego and these individuals (weak or strong)
Individual-based algorithm: If V is the set of individuals directly connected to the user and I (V) the set of interests of all users of V, the weight of an interest i in the social dimension S (u) is simply calculated by summing the semantic score of this interest for each individual in V (Fig. 2).

*For each interest i in I(V)*

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

*End for:*

January-February 2020
ISSN: 0193-4120 Page No. 9918 - 9924

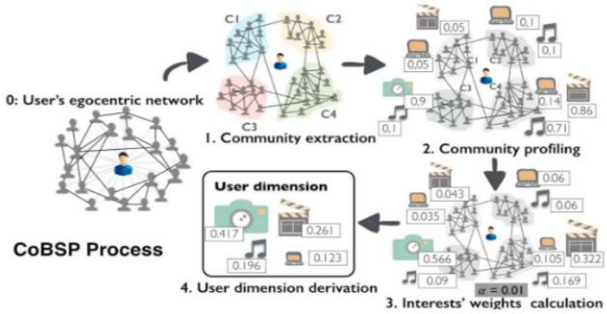
### 2.3.3 Self-centered algorithms based on communities

Using all the alters of the egocentric network or just using the alters that represent the strongest links, when analyzing its social graph, is a problem that has led to a new approach to community-based algorithms in order to consider both the links that would be strong or weak. This type of algorithm is based on the assumption that the user is best described by the surrounding communities rather than by the individuals in his network. It may also be natural to think that if a community of the egocentric network of the user is characterized by an interest (for example, photography, by affinity, this means that the user (ego) is certainly interested in photography. On the other hand,

It is more likely to find a person who is strongly interested in photography in the egocentric network of the user, but for whom no interest in photography is linked to the interests of the user. strongly use the communities (rather than individuals) around the user to understand his behavior, as well as the functioning of his social network.

In the rest of this part, we will deal with the main egocentric community-based algorithms and their strong and weak points in order to introduce our contribution

_ Community-Based Algorithm:CoBSP




This algorithm is summarized in four essential steps:

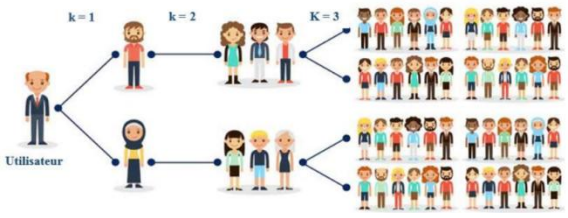
1. Step 1: In this phase we apply the iLCD algorithm, to extract communities from the egocentric network of the user.
2. Step 2: In this step we analyze the behavior of all the members of the communities, in order to calculate the profile of each community.
3. Step 3: after we calculate the weight of each interest in the social dimension of the profile of the user.
4. Step 4: allows to derive the social dimension of the user.

The CoBSP algorithm has proven its performance and shown that it gives better results compared to individual algorithms, but the constraint here is that this algorithm gives better results in dense networks as opposed to small networks, so Community algorithm in this case tends to give results more similar to individual algorithms.

Hence the proposal of CoBSPL, which suggests increasing the size and density of the user's egocentric network, by adding new important nodes to the user's egocentric network before applying CoBSP.

Community-Based Algorithm: CoBSPL (Snowball Sampling, Friend of a Friend, Link Prediction)

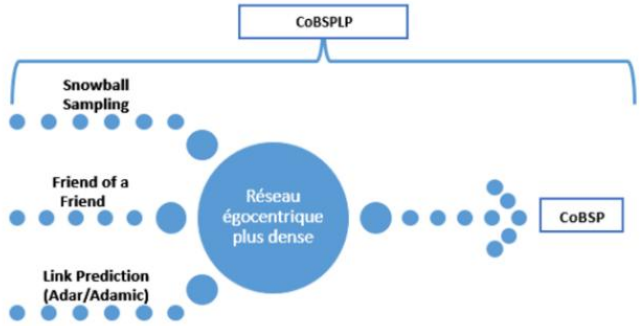
The CoBSPkL algorithm adopts the technique of Snowball Sampling: it is a sampling technique mainly used in the search to locate other information Figure 7: the processes of the CoBSP algorithm to increase the number of samples research. In our case this technique works as follows:




And after we filter the population of the sample found using a property called "Friend of a Friend", and like that we get a larger network, where we can apply the algorithm CoBSP.But we have no idea of the actual distribution of the population and the sample. Hence the addition of the Adamic / Adar Adamic prediction technique to sampling and filtering techniques, which consists in predicting, from a given network, the pairs of nodes that are

January-February 2020
ISSN: 0193-4120 Page No. 9918 - 9924

likely to bind. in the future, that is to say people who can become friends of the user in the near future.




## 3 DISCUSSION AND OUR CONTRIBUTION

In this state of the art, we have put the spotlight on the importance of user profiles in improving the results of information systems, as well as the emergence of these systems towards the use of user data. from its social networks, the latter adopt two main types of algorithms for enriching the user profile: the individual algorithms and the community algorithms that we discussed in the previous part, in this 2 nd type of algorithm we quoted the CoBSP algorithm, this weak-end algorithm that resides in the fact that it gives better results just in egocentric user networks or there are many nodes (user friends) hence the proposal for the CoBSPL algorithm which improves its performance in a very remarkable way in small and / or scattered networks by the addition of new nodes using a few e XTensions; On the one hand the addition of these extensions proves the importance of integrating more individuals into the egocentric network of the user before the application of the CoBSP algorithm process. On the other hand studies have shown that to have best results by analyzing the egocentric network of the user should not exceed k = 1 (1-egocentric), ie. in our case we have to analyze just the data of the user's friends in question, not the data of friends of his friends, on top of that this kind of treatment in 2-egocentric networks can take a lot of time. Hence the proposal of our contribution to improve the CoBSP algorithm by extending the user's 1-egocentric network by adding more nodes from different social networks to cover all the interests of the user. proposal is instead of using the techniques of Snow-Ball Sampling, Friends of Friends and Link Predection to increase the density of the egocentric network of the user, it is to combine the relations (friends) of the user on different types of social networks (here Facebook, Twitter and LinkedIn) in a single social egocentric graph before applying the CoBSP algorithm whose goal is to obtain a more complete User Profile that brings together the majority of the user's interests. So our intervention is at step 0 of the CoBSP Community algorithm process.

### 3.1 Methodology of Work:

The realization of our work is summarized in 4 main stages:

1. Detection of user accounts that belong to the same person on different social networks. (for our case on Twitter, Facebook and LinkedIn);
2. Detection of the user's communities on these different social networks.
3. The preparation of the egocentric network of the user by the combination of the communities detected in these three social networks in a single egocentric graph of (so we will have a denser network)
4. The application of the algorithm CoBSP :

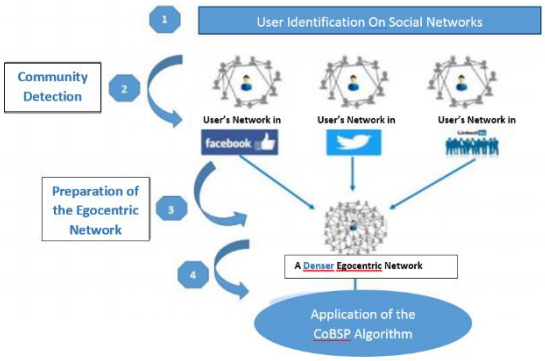



*Figure 4 : Methodology of Work*

## 4 CONCLUSIONS AND FUTURE WORK

Our work is about the construction of the user profile, an essential element in a system of information, the user profile is a central element in the information adaptation systems. We are

*January-February 2020*
*ISSN: 0193-4120 Page No. 9918 - 9924*

interested in the process of enriching the user profile from his social network, which represents a very rich source of information about the user.

We propose to take into account the dynamic characteristics of social networks in the step of extracting the interests of the process of building the social profile.

We used a mechanism to detect users similar to this user in the system, and analyze their interests, using similarity techniques, and consequently use the CoBSPalgorithms,by combining the user's relationships on different types of social networks in a single social egocentric graph, with the goal of getting a more complete User Profile that brings together the majority of the user's interests.

In the short term, we plan to apply this approach to carry out an evaluation of our proposal on a larger scale. We will also study our approach according to different characteristics of the network (size and density).

REFERENCES

1. CABANAC, G., 2011. Accuracy of inter-researcher similarity measures based on topical and social clues. Scientometrics, Volume 87, no. 3, pp. 597-620.
2. CARMEL D., ZWERDLING N., GUY I., OFEK-KOIFMAN S., HAR'EL N., RONEN I., UZIEL EREL., YOGEV S., CHERNOV S, 2009. Personalized social search based on the user's social network. New York,USA, s.n.
3. Damian Fijałkowski ,RadosławZatoka, 2011. An architecture of a Web recommender system using social network user profiles for e-commerce. s.l., s.n.
4. J. Budzik , K. Hammond, 2000. User interactions with every applications as context for just-in-time information access. s.l., s.n.
5. M. DAOUD, 2009. Accèspersonnalisé à l'information :approchebaséesurl'utilisation d'un profilutilisateursémantiquedérivéd'uneontologie de domaines à travers l'historique des sessions de recherche. Institut de recherche en informatique de TOULOUSE.
6. Omar Hasan ; Benjamin Habegger ; Lionel Brunie ; Nadia Bennani ; Ernesto Damiani, 2013. A Discussion of Privacy Challenges in User Profiling with Big Data Techniques: The EEXCESS Use Case. Santa Clara, CA, USA.
7. ZENG Y., YAO Y.Y., ZHONG N, 2009. A dblp search support engine. s.l., s.n.