User Profile Enrichment from His Social Networks

Elachkar Ibtissam, Rachik Zineb, M'rhaouarh ibtissam, Labriji Elhoussine

Abstract—Recently information systems have started to develop several mechanisms for improving the user experience on the web, the majority of these mechanisms is based on the use of users profiles, and since these profiles changes from one moment " t " to another, they require regular updates and a permanent enrichment, which leads these systems to enrich the profile of each user from the profiles and information of users who resemble him in the same system, but this technique is not applicable to all users of the system, especially in the case of cold start problem and for the users less active, since their profiles are empty we cannot detect profiles similar to them, which leads us to enrich these profiles from information provided by social networks about the users, which gave rise to algorithms for enriching user profiles from social networks. These enrichment algorithms are divided into two main categories: individual algorithms and community algorithms. In this work, we will first address the notion of the user profile, secondly the types of enrichment algorithms on social networks and their main strengths and weaknesses, and finally we propose a new approach to improve one of the famous community-based algorithms for user profile enrichment from his social egocentric network.

Index Terms— Community algorithm, data science, egocentric network, individual algorithm, information system, user profile, social network.

1 Introduction

Since the creation of the web, the number of digital data increases exponentially because of several factors such as the advent of web 2.0 then 3.0 as well as the appearance of multiple programming languages (Java Python C # Scala JavaScript ...) that contributed to the creation of thousands of dynamic websites (1 billion and 900 million websites[1]), in addition to the use of multiple interfaces and IT tools and the emergence of cloud computing with Big data techniques, Insofar as it has become easier to publish content on the web but access to a specific content for each user has become more difficult for him, given the large number and diversity of information likely to interest him, generally this poses problems of cognitive overload and disorientation [2] to the user who have more and more difficulty finding the information corresponding to his expectations. For example: a developer uses the request "java" in intention to find the new API (s) and technologies related to this language, and by executing this request, the search engine provides him results about the Parisian Java dance, or the Indonesian island Java. Also we can find the case of a farmer who uses the request "Apple" to find information about the growth of delicious apples, but he receives a mix of information about Apple computer and apple fruit. To solve these information access issues, the researchers proposed information adaptive systems then contextual system [3] to the user based on the U-process of searching information. in order to improve classical information retrieval systems, and to overcome the main problem studied by [4] which lies in the fact that classical search engines are based on a generalist approach which considers that the need of information for the user is completely represented by his request and consequently deliver results taking into account only the criteria of selection by content and the availability of sources of information, 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.

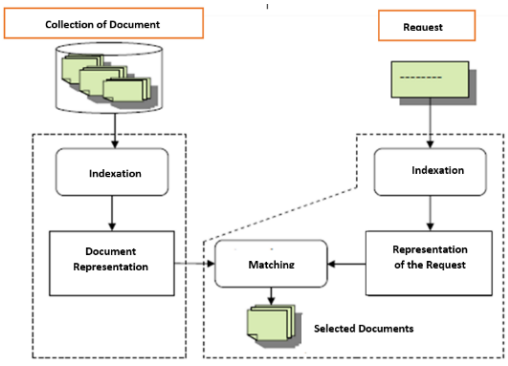


Fig. 1. U-Process of Information Retrieval.

There are 3 main techniques developed in personalization Information research system: Customized selection of information, reformulation or expansion of the request and reordering the results. The fundamental purpose of these techniques is to exploit in addition to the request of the user,

- *Elachkar Ibtissam is currently a phd student in Laboratory of Technological Information, and Modeling, University Hassan II, Casablanca, Morocco, E-mail: elachkar.ibtissam@gmail.com*
- *Rachik Zineb is a professor at Hassania School of Public Works EHTP, Morocco,E-mail: zineb.rachik@gmail.com*
- *M'rhaough Ibtissam is currently a phd student in Laboratory of Technological Information, and Modeling, University Hassan II, Casablanca, Morocco, E-mail: mrhaourhibtissam12@gmail.com*
- *Labriji El houssine is a professor in University Hassan II, Casablanca, Morocco, E-mail: labriji@yahoo.fr*

[1] www.internetlivestats.com

additional information (browsing history, profiles on social networks, information entered into forms, user ratings ....), extracted from the interactions of the user with the system, in order to improve the results of his research. An analysis by [5] and the study of the EEXCESS European project [6] shown that the best additional information, which we can integrate during information retrieval processes is the use of the user profile. So, in the rest of our work, we will present the user profile, his sources of construction and enrichment, and more particularly the reason for using the user's data on his social networks to build and enrich his 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 we will end with our contribution to develop one of these algorithms while citing the methodology that we will follow.

2 Related Works

According to [6] a user profile represents a collection of personal data associated with a specific user or a group of users that describes a set of attributes, these attributes may include geographic location, academic and professional experience, goals (short term and long term), behavior, interests (entertainment, commercial products ...), etc.

Generally a user profile can be built according to two methods: either by the user himself what is called explicit profile, or automatically from data of interactions between the user and the system, in this case it is called implicit profile. This last method is the most common, as the manual input of parameters (preferences, interests ...) by the user can be a tiring task for him and may take a long time to express his needs.

2.1 User Profile Modeling

After collecting and preparing data, comes the phase of data mining, several algorithms can be used depending on the type of profile model to build, there is three types of modelization:

- behavior modeling;
- interests Modeling ;
- intentions Modeling

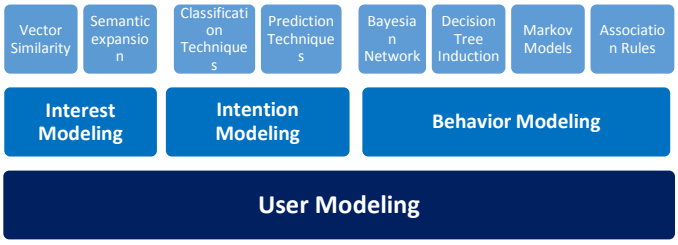


Fig. 2. User Profile Modeling

_ **Interest modeling:** This type of modeling is characterized by the definition of the user interests by a function fct (i) which gives the degree of interest or disinterest of a user for an item i by analyzing his previous behavior.

_ **Intention modeling:** An intention here is the purpose for which the user uses the information system. The modeling of user's intentions is to build a model that will identify the purpose of each user of the information system. For example, customers of an e-commerce website can be divided into two groups: those who really aim to buy and those who don't aim to buy. Intention modeling is largely based on classification techniques with predefined categories.

_ **Behavior modeling**: Consists of analyzing user behaviors via browsing histories or transactions they perform on Web servers [10], in order to determine recurring navigation courses, to validate marketing strategies or to check the relevance of marketing campaigns. Among the methods of this type of analysis we cite: The Markov chains to predict future URLs that will be visited by the user.

Interest modeling is usually prior to intention modeling and behavioral modeling.

2.2 User Profile Enrichment

Since the implicit user profile is built and enriched as the user interacts with a given system (Purchased Products, Browsing History, etc.) his profile is still incomplete, and therefore it may not contain all interests and information that may be useful for this mechanism, especially for the new users of the system and for users less active. To solve this problem and enrich the user profile as needed, these systems tend to analyze the interests of users similar to this user in the system, using similarity techniques [7]-[8]-[9]-[10] to detect these people and after that using specific algorithms and processes according to the needs of each system to enrich the user's profile and predict his behavior and interests from these users who are similar to him, but this technique is not applicable for all users of the system, especially in the case of cold start problem [1], this problem is very widespread in the recommendation systems based on collaborative filtering in e-commerce, and for users less active, since their profiles are empty, we cannot apply the similarity methods to detect the users who are similes to them. In addition to this, this kind of technique requires a lot of time because it compares each user to all the users of the system, and it involves storage techniques and the use of huge sparse matrices. As a solution to this problem, works are starting to develop multi-application user profiles [11] which is characterized by the use of multiple platforms and data sources to build and enrich the user's profile, also, they tends to use techniques based in semantic approaches. This leads researchers to collect users' data from their social networks and analyze them, in order to improve these information systems, because the individuals of users' social network are the most similar to him. Which leads to a new approach of information filtering: Social filtering, which is particularly interested in the analysis of user's individuals in the social network (e.g. friends), in order to understand his behaviors and to know his interests, since these individuals are the most similar to him and can even influence his interests and finally his profile [17][18].

This type of filtering is reinforced by the number of users in social networks that exponentially increases day by day, and consequently their data rise, which makes these sources very important to enrich the user profiles. The following figure shows a statistics taken by statista.com about the evolution of the number of users on this social network.

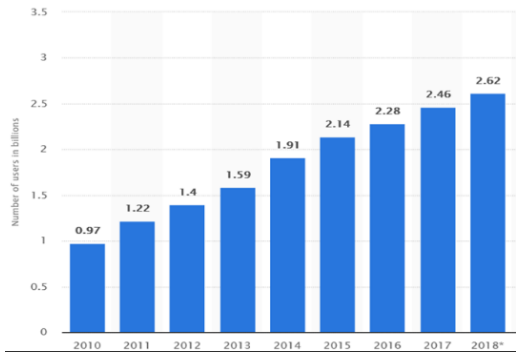


Fig. 3. Number of social media users worldwide from 2010 to 2018 (in billions)

2.3 The Use of Social Networks in the User Profiles Enrichment

The user profile enrichment from data on social networks is based on the concept of devising the user profile into two dimensions <PD(u),SD(u)> :

The User Dimension PD(u) : is a dimension that contains the elements built from the information and interactions of the user with the system. This is the most important dimension of the profile that should be used as a priority by such an information retrieval system. However, if, for example, the information system needs information that does not exist in this dimension, then the social dimension can be exploited by the mechanism.

The Social Dimension SD(u) : is a dimension that contains the elements built from the information and interactions of the user community of his social network. The information presented in this dimension is complementary information to the information of the user dimension, which can be exploited according to the needs of each mechanism. 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 [12].

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

2.4 Types of Graph Analysis in Social Networks

The most used types of graph analysis in social networks are: *Egocentric analyzes*: e.g. [12]-[13]. Which usually focus on individuals located at a distance of 1 (called alters) of the user (called ego).

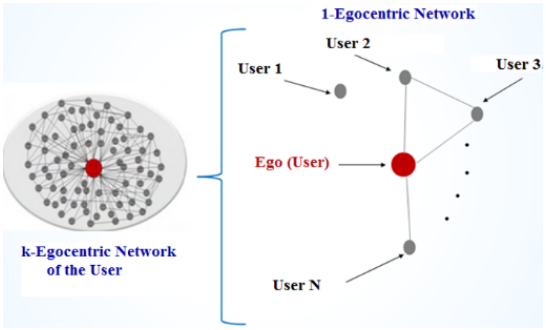


Fig. 4. The K-egocentric network

Socio-centered analysis: e.g. [14] that analyzes the entire social graph. This type of analysis is very time-consuming for systems requiring profiles built in real time, and regularly updated, (short-term profiles, for example for the customization of queries in a search engine). Because of the computation of individuals' centrality measures in the entire graph, and because of that this kind of analyzes is unacceptable for this type of system. [19] Demonstrate that an individual may be characterized only through the analysis of his egocentric network, especially as the computing times in this type of network will remain much more reasonable for the analysis of an 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. So in the rest of our work we will focus only on user 'profile enrichment algorithms from data of his social egocentric graph.

3 Egocentric Algorithms for Enriching the User Profile

Even if it makes more sense to focus on egocentric networks, the question of how to exploit them also arises. Some authors analyze the data of all alters (users), others only subsets deemed more relevant those with the strongest links between the ego and users. It seems obvious that for an ego, all the alters of his egocentric network don't have the same importance for him, and logically shouldn't be considered in the same way when analyzing the ego. Works that relies only on the strong links of the user such as [14] are certainly interesting, but it is also clear that a lot of potentially useful information for the user from his weak links are ignored in these works. Indeed, the weak links of the user may be a vector of new information about him. To better explain this point, consider the following example:

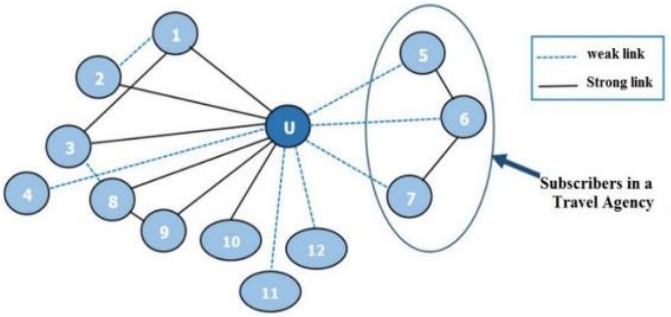


Fig. 5. The social graph of the user "U"

A user "U" is interested in traveling, and this interest is translated by his subscription to a travel agency. Suppose that an information system doesn't know the interest of «U" for travelling, but knows a number of interests of the individuals in the egocentric network of "U", as well as the relationships between these individuals. It turns out that in reality, the user "U" meets from time to time during his visit to the travel agency the individuals 5, 6, 7 who are also subscribers to this agency, and know each other, but each has weak links with the user "U" in his egocentric network.

_ In the case that we analyze just the information extracted of the strong links between the user and the individuals of his egocentric network: the interest "traveling" will not appear in the profile of "U" derived from his social network, because all the links between the user "U"and his friends subscribed to the travel agency are weak.

_ In the case that we consider all the links between the user U and the individuals of his egocentric network are equivalent, and we analyse informations of these links, the interest " traveling " will appear but may be drowned in a lot of interests derived from other alters, especially if the user has a very large number of alters.

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

3.1 Egocentric Individuals-based Algorithms

Individual-based algorithms [16] use individual people selected in the user's social network. Individual people are usually selected according to the strength of their tie with the user or to their centrality values.

This type of algorithms is divided into two main categories as we have seen previously: 1) the algorithms that use and analyze the data of all the individuals of the egocentric network of the user.2) the algorithms that analyze only those individuals in the user's egocentric network who have a stronger connection with the user (ego).

Among the most well-known individual algorithms, we cite:

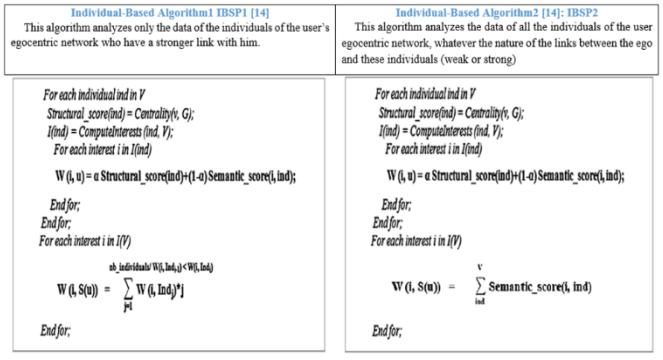


Fig. 6. Individual-Based Algorithme IBSP1 and IBSP2

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

3.2 Egocentric Community-Based Algorithms

In the example that we mentioned in Figure 5, we discussed the problem of choosing to analyse data of all the individuals of the user's egocentric network when analyzing their social graph, or to just analyse data of the individuals who have the strongest links with the user. This kind of conflict leads to a new approach of community-based algorithms, in order to consider both strong and weak links between the user and the individuals of his egocentric network. Community-based algorithms are based on the assumption that the user is best described by the surrounding communities rather than the individuals in his network. It may also be logical to think that if an egocentric network community of the user is characterized by an interest e.g. programming, by affinity, it means that the user (ego) is certainly interested in programming. On the other hand, finding a person who is very interested in programming in the egocentric network of the user does not imply that the user is interested in programming. The most marked algorithm of this type is: **Community-Based Algorithm: CoBSP [12],** its process is summarized in four essential steps, as shown in the figure below:

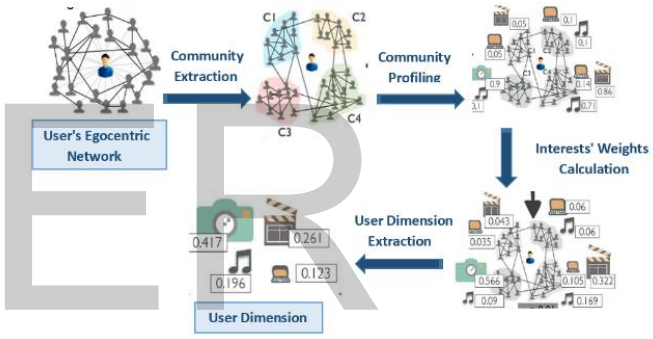


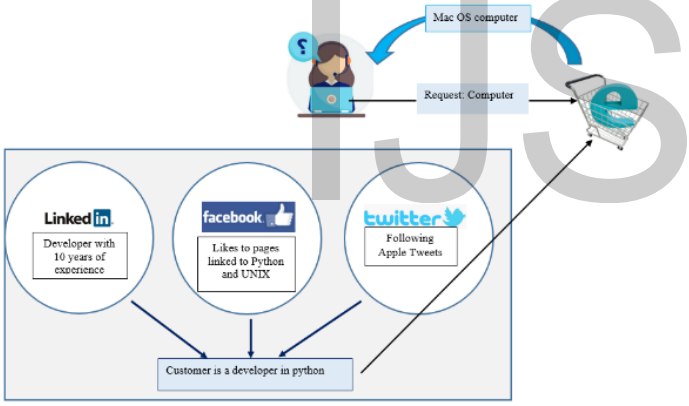
Fig. 7. CoBSP Process

1. Extracting communities from the user's egocentric network. This phase is performed by applying the iLCD algorithm [20];
2. Extracting the profile of each community found in the first step, the profile of a community is extracted by analyzing the interests of all members of this community
3. Calculation of the weight of each interest in the social dimension of the user's profile (community profile);
4. Finally, deriving the social dimension of the user according to the weights calculated in the third step.

4 Discussion and our Contribution

The performance of CoBSP algorithm has been proven with empirical results on Facebook and the bibliographic site DBLP which have shown that it gives better results compared to individual algorithms [16]. But the constraint here is that this algorithm gives better results just in dense networks unlike small networks, since if an egocentric network is small or scattered, the community detection algorithm extract too many small communities (communities of 1, 2 users, for example). Thus, the CoBSP algorithm in this case tends to give us results

more similar to the individual algorithms. Hence our approach to increase the size and density of the user's egocentric network by adding new important nodes to the user's egocentric network before applying CoBSP. So, our intervention is in step 0 of the CoBSP process i.e. the preparation of the egocentric network before the detection of the communities. So we plan to combine the user's relationships from different types of social networks (here Facebook, Twitter and LinkedIn) in a single social egocentric graph before applying the CoBSP algorithm, whose purpose to obtain a profile more complete who gathers the majority of user interests. To better clarify the importance of our proposal here is an example : a customer access to an e-commerce website in the intention of purchasing a Mac OS computer, so he types the request 'Computer' to the search engine of this website, if the system of this website has the user profile based on our proposal it will know that this customer is a developer in python and he is looking for a very powerful computer of type Mac OS, because the customer is following Apple Tweets in Twitter, and he has made likes to pages linked to Python and Unix in Facebook and he has marked that he is a developer with 10 years of experience in LinkedIn, as well as his friends on Facebook have computers of Mac OS type , and consequently the engine of research of this website will provide him with the products that meet his needs very quickly.



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

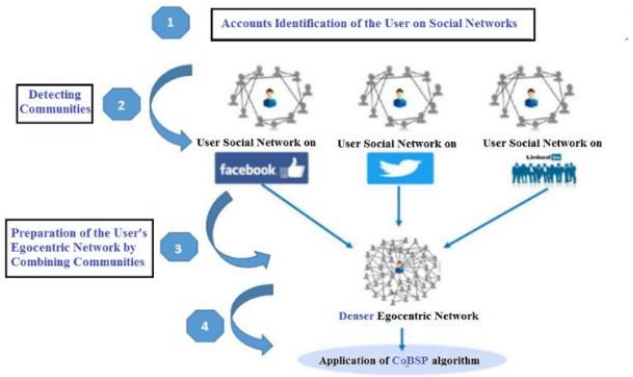


Fig. 8. Methodology of our approach

This makes it possible to link user data from different sources and group them into a single user profile. In addition, information from different sources can be correlated to validate or invalidate information discovered from a source, in order to obtain a more complete user profile, which contains the maximum number of information and interests about the user, useful for any information system.

4 Conclusion

The user profile is a very important tool in several fields such as recommendation systems, customization systems etc., it is used to narrow the number of data or results provided for a specific user, also to minimize the cost and the time of processing of multiple systems. Whatever the user profile model used, it's updating and enrichment is a very essential step in the information research process in order to obtain more interesting and satisfactory results, which lead the information systems to develop several techniques aiming to enrich them, thus we have presented the main user profile enrichment algorithms based on the user's egocentric social network as well as discussing our approach to developing one of these algorithms.

References

[1] N Silva, D Carvalho, ACM.Pereira, F Mourão, L Rocha, "*The Pure Cold-Start Problem: A Deep Study About How To Conquer First-Time Users In Recommendations Domains*," Information Systems, vol. 80, pp. 1-12, September 2018.

[2] J Conklin, M Begeman, "*GIBIS A Hypertext Tool For Team Design Deliberation*," pp. 247-251, 1987

[3] G Salton, "*The SMART Retrieval System - Experiments In Automatic Document Processing*", New Jersey: Englewood Cliffs, Prentice-Hall, 1971.

[4] J. Budzik, K. Hammond, "*User Interactions with Every Applications as Context for Just-In-Time Information Access*," in the 5th international conference on intelligent user interfaces, 2000.

[5] D Fijałkowski, R Zatoka, "*An Architecture of A Web Recommender* System Using Social Network User Profiles For E-Commerce," in the Federated Conference on Computer Science and Information Systems, 2011.

[6] O Hasan, B Habegger, L Brunie, N Bennani, E Damiani, "*A Discussion Of Privacy Challenges In User Profiling With Big Data Techniques: The EEXCESS Use Case*" in 2013 IEEE International Congress on Big Data , Santa Clara, CA, USA, 2013

[7] N Reddahi, A Labriji,I Abdelbaki, EH Ben Lahmar, EH Labriji," *A New Approach That Improves TF-IDF Weighting Measure*", International Journal of Information and Communication Technology Research Volume 5 No. 10, October 2015

[8] A Labriji, S Charkaoui, I Abdelbaki, A Namir, EH Labriji," *Similarity Measure Of Graphs*" International Journal of Recent Contributions from Enginner Science & IT, Vol 5 No 2 ,2017

[9] I El Achkar, A Labriji, F Ghanimi, EH Labriji, "*Ontology-Based Approach For Extracting User Interests*" in SCA '18 Proceedings of the 3rd International Conference on Smart City Applications Tetouan ,Morocco, October 2018

[10]I. El Achkar, H. Ouzif, H. Labriji , « *A Structural Similarity Measure Of Users Profiles Based On A Weighted Bipartite Graphs* »The Fifth International Conference On Smart City Applications, 2020

[11] M Viviani, N Bennani, E Zsigmond, "*Survey On User Modeling In Multi-Application Environments*," in The Third International Conference on Advances in Human-oriented and Personalized Mechanisms Technologies and Services CENTRIC'10, Nice, France, 2010.

[12]D Tchuente, MF Canut, N Jessel, A Peninou, F Sedes, "*A Community-Based Algorithm For Deriving Users' Profiles From Egocentrics Networks: Experiment On Facebook And DBLP*," Social Network Analysis and Mining, Springer, vol. 3, no. 3, p. 667–683, 2013.

[13] I El achkar, A Labriji, EH Labriji,"*A Semantic Method To Extract The User Interest Center* In "Smart Cities Applications ,SCA2018,Lecture Notes in Intelligent Transportation and Infrastructure,Springer,2019

[14] D Carmel, N Zwerdling, I Guy.,SO Koifman, N Har'el, I Ronen,U Erel.,S Yogev, S Chernov , "*Personalized Social Search Based On The User's Social Network*," in Proceedings of the 18th ACM conference on Information and knowledge management (CIKM), New York,USA, 2009.

[15] Y Zeng, YY Yao, N Zhong, "*A Dblp Search Support Engine*," in Proceedings of the 2009 IEEE/WIC/ACM International Conference on Web Intelligence, 2009

[16] MF Canut, M Mezghani, S On-At, A Péninou, F Sèdes, "*A Comparative Study Of Two Egocentric-Based User Profiling Algorithms Experiment In Delicious*," in ICEIS 2015, Barcelone, Spain., 2015.

[17] U. Sonja, "*Show Me Your Friends And I Will Tell You What Type Of Person You Are: How One's Profile, Number Of Friends, And Type Of Friends Influence Impression Formation On Social Network Sites*," 2020

[18] A. Nicolas, "*The New Personal Influence: How Our Facebook Friends Influence the News We Read*,"2017.Political Communication volume 34 Issue 4.

[19] L. Zhaoyi, X. Fei, W. Ximeng, C. Hongshu, and X. Xi, "*Topological Influence-Aware Recommendation On Social Networks*," 2019

[20] R. Cazabet, F. Amblard and C. Hanachi, "*Detection Of Overlapping Communities In Dynamical Social Networks*," 2010 IEEE Second International Conference on Social Computing, Minneapolis,2010, pp. 309-314.