Sammer A. Qader
Computer Sciences
University Of Technology
Baghdad, Iraq
111817@student.uotechnology.edu.iq

Ayad R. Abbas
Computer Sciences
University Of Technology
Baghdad, Iraq
110010@uotechnology.edu.iq

Survey of User to User Recommendation System in Online Social Networks

Abstract- The widespread use of online social networks (OSN) and their applications by users lead to the lack of knowledge identification of their needs across the vast amount of data, which made the need to create systems that help people to solve the problems and make decisions with more accuracy, an example of these systems is the Recommendation system (RS), which helps users to make decision and save time in search on a commercial or personal level, one of the most critical types of recommendation systems is the friends recommendation system (FRS) . In this survey, several studies have been suggested to solve the problem of FRS and its mechanism, techniques, and algorithms used to create them *Also, the RS types and techniques, a variety of dataset that deals with a specific system, are explained. Moreover, the challenges they face to determine the needs of people in terms of the choice of items or at the level of social networks are included.*

Keywords- *Recommendation System, Friends Recommendation System, Collaborative Filtering, Content-based Filtering, Social dataset.*

1. Introduction

With the presence of the Internet, using a social network is more popular nowadays and increased rapidly to make life easier, by helping people to know and get used to different cultures, also to connect quickly all around the world, saving time, distance and effort. With the expanded use of these social networks, it produced an increase in data flow generating so-called, (Big data). Moreover, facing the era of data explosion makes data analysis and mining more serviceable and accessible. As a result, the network turned into a sharing platform, and it is progressively more challenging to find what kind of data users want in such infinite amounts of data [1].

However, the amount of data that users now deal with is significantly large. Many users use social applications with no specific goal to identify their real needs or interests. Such as creating new friendships that have similar trends and terms of requirements, or at the level of choice of Items. Where create a recommendation system presents a great way to help users and deals in speed and accuracy for decision-making according to the user's historical behavior and information gathered by these systems. Currently, the RS has successfully implemented in the area of E-commerce and various websites such as (movies, music), for instance, Amazon and Netflix for recommending Items to the users besides diversity of OSN, such as microblog and forums. All these websites use an application, which can suggest making friendship such as Facebook and

Twitter, while to the end users they both recommend information. In the social network, the RS differs much from the one that implemented in E-Commerce, compared with the friend's suggestion. The general recommendation application owns the characters that have less attributes rarely been a change in items and that the factors that influence the users and items are comparably simple, for example, characteristics like genre, writer, singer, and publication date in the music. However, if we are studying the connections between user and user, many factors are correlated [2].

Firstly, the items described are more efficiently than the person's behavior description. Secondly, the user's behavior always changes over time. Finally, the connection between people in the social network not only based on a similar user's behavior such as interests, but it depends on a geographical site, study, or work specialization. In this case, the factors above influence the FRS that the user may be interested. Importantly, potential FRS is essential nowadays as well as searching on the social network and finding suitable friends method, because similar users met within social networks are much relevant for themselves [2]. Furthermore, Many of FRS on social network sites based on People You May Know (PYMK), this method is proposed by MySpace using people's relation and profile information, or the (friend of a friend) technique [2, 3]. The same strategy used in Renner, twitter, Facebook, and other OSN. This method depends

on recommending people that are similar according to age, location, and study information. This survey paper, explains the RS types, practices, and challenges referred to as a set of studies that have been proposed to solve the problem of FRS and their algorithms that used to create them.

2. Literature Surveys

Several kinds of literature decide different types of RS applications. One of these systems that play an active role in the current studies is FRS that studies and analyzes the user's relationship on a social network and recommending friends based on the user's context information and behaviors.

I. Graph-based Friend suggestion framework

In [4], proposed a FRS by utilizing Artificial Intelligence Bee Colony Algorithm (ABC), which recommends new links among users inside the network. This method is based on the structural features and topological characteristics of OSN and suggested a new connection to the root user depending on calculating four parameters within the generated subgraph by studying the connected users individually. Then, a Bee colony algorithm was used to optimize the closest weights for each parameter to create an efficient friend recommendation.

In [5], suggested an FRS by using a Genetic algorithm (GA) to optimize the indicators obtained from the complex network. A filtering method was used by extracting the sub-graph then using an algorithm to analyze the sub-graph, which is made by a client and all the other clients, is connected using a three-degree partition in the system to investigate the relationship of the connection. After the filtering method in the ordering step, the three indexes were calculated utilizing the information extracted from the sub-graph then using the genetic algorithm to optimize these three indexes, and utilizing the optimized value within the recommendation framework.

In [6], proposed a FRS using the same algorithm (GA) to optimize the relationship of inclinations in FRS that is exhausted two-step sifting preparation utilizing (FOF) and degree centrality. The strategy and outcomes in this paper present initial findings to a potentially robust method of providing friend recommendations in social networks while additionally picking up insights into how friendships are built up.

II. Social Friend Recommendation based on Trust-aware system

In [7], proposed a framework system "trust-aware system", which recommended friends in OSN based on the collaborative reputation mechanism that determines the implicit and explicit plus analyze the semantics and dynamics connection between the network members relation, also the element of time is inserted in order to handle the social network dynamics. The system process produced three phases' semantics among the created members, reputation ratings and personalized users recommendations. Moreover, it gives positive or negative suggestions that can utilize for making functional trust/distrust associations in OSN.

III. Social Friend Recommendation Based on Network Correlation

In [8], suggested a two-stage friend recommendation process. Use multimedia information, flicker tags features, and friendship network in stage one to produce a friend list based on the relation of different OSN. Then, apply a co-clustering algorithm on the user, tag, and image information in the second stage to create groups, and build more precise RS to improve the recommendation outcomes in stage one by assuming that friend's - friends, also be likely to be friends. Finally, the system outcome shows that the proposed framework obtains excellent performance, and each stage participates in the RS. Figure 1 shows the system process.
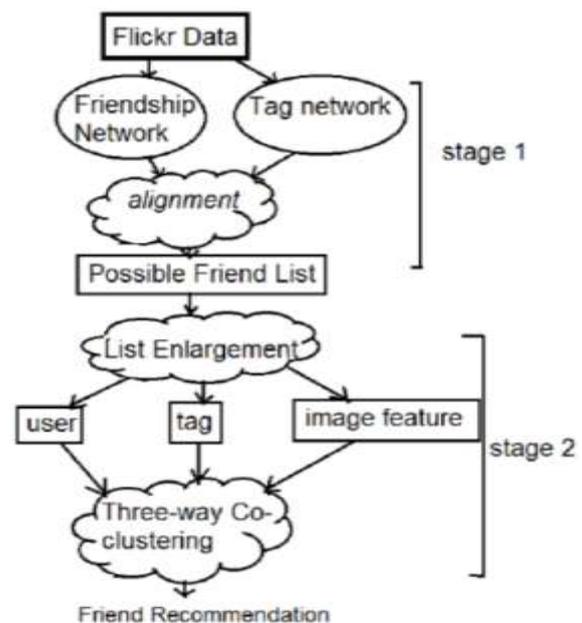

Figure 1: System illustration form [14]

IV. Social FRS Based on user behavior

In [9], proposed a novel friend system which utilizing the individuals behavior on the social network site, and authors measured the recurrence of the activities which done by the clients and upgrading the dataset according to the activities. Then applying client's behavior classification strategy by utilizing association rule (FP-Growth) algorithm to create the dataset frequent pattern rule. Finally, finds out the required behavior (ordinary and exceptional), and multilayer threshold for FRS is used. In [10], suggested FRS that based on multiple social behavior features. First, the system used three types of behavior characteristics in a social network, i.e., social rating, social content, and relation features, and extracts the three features from portraying the association between two users in the microblog information. Then, it utilized these features extraction, in a double classifier algorithm (SVM and Naïve Bays), after information preparing to determine whether the second client in each pair should prescribed to the primary one.

V. Social Friend Recommendation Based on Location

In [11], proposed a system for FRS based on the location preference, in which the temporal, spatial and social relationships are considered, and suggested friends with a similar area inclination for location-based social network (LBSN) users. The approach is primarily isolated to the following steps: First, the strategy of arbitrary local walk based on the Markov chain to calculate the user's friendship similitude on a social network. Second, the user's area inclination similitude within the real world based on calculating the history check-in information. Third, recommend friends to users by building a blended user preferences model. Finally, soundness and adequacy of the algorithm were confirmed by employing a real client check-in dataset. The modern approach suggested friends with both similar companionship and area inclination to clients within the large-scale.

3. Recommendation System

A recommendation engine helps users to discover what he/she desires from a vast collection of entities belonging to the same categories. A well-designed of RS can progress the users' information of websites or other administration [12], RS utilize distinctive sources of data for giving clients with forecast and proposals of items [13]. The authors attempted to adjust the variables like exactness, oddity, disparity, and

stability within the suggestions [14]. Moreover, RS has been broadly utilized for different purposes, such as media proposal [15] and friend proposal [16, 17, 18].

4. Applications of Recommendation System

Examples of new recommendation applications include Amazon website item recommendations, Movie Lens and Netflix for movie recommendations, and Google for webpage recommendations. In addition, Facebook and, Twitter provide features to friend's recommendation of "people you may know". LinkedIn, provides a job recommendation, e-learning, and e-society, etc.

5. Techniques of Recommendation System and Challenges

Many recent works of literature evaluate different types of RS technologies as seen in Figure 2, and their applications introduce and classify RS into three main techniques [12, 19]:

a) Content-based recommendation system techniques

These techniques based on a data description of an item that recommended to the user, the similarity of recommended this item is figured and established on the features associated with compared items. Content-based filtering use the two methods of classification, and nearest-neighbor methods[20].

b) Collaborative filtering

Collaborative filtering technique in many kinds of literature is called "people-to-people correlation" [12]. This is the most popular technique [21]. It based on the use of a large amount of data on people behavior, preferences and activities, and that predicts what people will like to depend on their comparison to other people based on the similarity. This technique consists of two methods:

• Memory-based CF methods: It finds comparative clients (or items) for the dynamic client or item utilizing the likeness estimation strategy and total the evaluation of these neighbors as, the expectation [22]. Memory-based CF contains two well-known strategies, user-based CF and item-based CF.

• Model-based CF methods: It provides an adequate strategy Building of the dataset, allows off-line processing for the foremost through likeness calculations. A standard model-based collaborative filtering algorithm uses k-means to cluster comparable clients into a group [23].

c) Hybrid recommendation system

The hybrid technique is a combination of CF, and Content-based techniques that based on using the advantages of one technique to fix the disadvantages of the other technique [24].
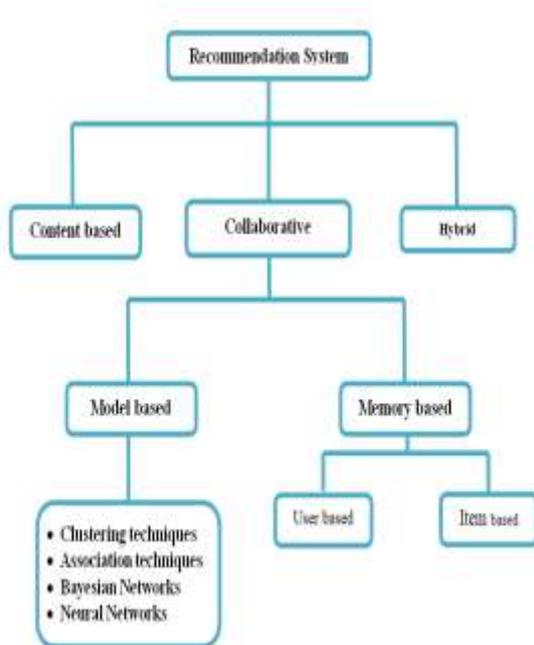

Figure 2: Recommendation System Techniques

There are various challenges and boundaries with such recommendation systems [19]. Moreover, these could be reviewed in different general classes as presented below:

• New User problem: The system of recommendation does not have any knowledge to obtain these recommendations regarding the new user. This can be referred to as a cold start problem.

• Sparsity Problem: As a result of a large amount of both people and items, there is no doubt that people may rate on limited items, which are seen to have relation to themselves. This case will lead to the lack of rating to a vast amount of different items, or maybe they will not have social touches by the user.

• Over-Specialization: Throughout this issue, the system has the capability to propose only the items that have already been seen by the users or those items that possess high scores. In addition, the user will be limited to be recommended to the items, which are considered as the same as the ones that are rated earlier [25].

• Limited Content Analysis Basically, such an issue is similar to the cold start issue, and there is no sufficient information of the items at various times.

Acording to the previously mentioned articles, Table 1 presents the comparison as a survey of

many recommendation methods. More details are illustrated in [25].

Recommendation System		
Methods	**Heuristic-Based**	**Model-based**
Conten-based	• Term frequency — inverse document frequency (TF-IDF) Information retrieval • Clustering	• Classifier technique (Bayesian) • Clustering • D. trees • ANN
Collaborative	• Distance metrics (Nearest neighbor) • Grouping a set of objects (Clustering) • Relations between objects (Graph theory)	• Bayesian networks • Clustering • ANN • Probabilistic methods • Regression analysis (Linear)
Hybrid	• Linear combination • Different voting designs • Combining one an ingredient as a piece of the model for the other	• Combining one an ingredient as a piece of the model for the other • Standard construction model

Table 1: survey of Ttechniques

6. Importance of Recommendation System in Social Network

It has agreed upon the fact that the use of social network based predictions has been discussed by a large number of researchers, but the system of recommendation in OSN area remains in its initial stages that offer a clear description of non-OSN based RS, for instance, Collaborative Filtering (CF) as a traditional methodology. Moreover, the drawbacks and weak points that happen in it are discussed [19]. Additionally, Collaborative Filtering (CF) procedures take a significant role in the recommendation despite the fact that they have been used with the other filtering techniques, such as content-based, knowledge-based or even social ones, [14]. Meanwhile, the amount of information that is being produced and shared can be nearly increased every year, it will make the internet content evolution unexampled. To give more explanation, such explosive growth and variability create difficulties in front of the users to get particular information that is justifying their interests. Therefore, significant attention has directed by the researchers towards the Recommender System (RS), and they are attempting to handle the information load by the way of personalization of the results to fit the needs of the user [26]. The solutions to the problem recognized in the current RS could be improved by applying OSN data in

recommendation framework. Integration of social networks can theoretically enhance the performance of current recommender systems, since the OSN based RS gives the community-based recommendation method more robust and useful [19].

7. Friend recommendation system (FRS)

Importantly, FRS is seen to be a deep-seated issue in the online social network. The purpose behind, is recommended different users or links for every user and this may be seen acceptable by the user. As well as, recommending a suitable and respectable friend is considered one of the preferable challenges in the system. To illustrate the social networks, Twitter, Facebook, and LinkedIn are regarded as the most significant sites that are visited by people on the internet nowadays [4]. As it has been mentioned previously in the introduction, many FRS strategies are using friend–of–friend (FOF) method and the mostly tradition strategy used in the user suggestion, such as Facebook, My Space, Twitter, etc. [3]. Also, for friend recommendation, different methods have been used, such as graph-based FRS [4, 6, 7]. Multimedia and image feature extraction [8]. All these methods have been used to evaluate and analyze the people relationship on the social network and recommend friends based on users context information and behavior. There are common algorithms that are used in RS such as, fuzzy c-means clustering, KNN, hierarchical clustering, K-means clustering, matrix factorization, association rules, TF-IDF, deep learning, similarity-based recommendations, and weight-based recommendations

8. Dataset Survey

In this section, several datasets have been introduced. In other words, these datasets can be used in friend's recommendation systems to experiments and learning models.

I. YouTube [4], Dataset

This dataset is collected from YouTube, it consists of a video-sharing website that uploaded, shared and viewed by users. Furthermore, it contains (1,138,499) users and (2,990,443) of friendship edge. For example, this data can be used for friend recommendation test.

II. The Social circles: Facebook [27], Dataset

This Social dataset is collected from using the Facebook app by survey participants; the data consists of 'circles' or 'friends lists' from Facebook. It is including 4039 nodes (profiles), circles, and 88234 edges with ego networks.

III. The Social circles for Twitter[27], Dataset

It has been noticed that the dataset contains circles or what is so-called lists from Twitter. To put it another way, such data has been crawled from different sources such as the public one. Furthermore, the dataset contains node information or profiles, circles, in addition to ego networks, the details of the dataset are shown in Table 2.

Table 2: Twitter network details

Dataset Statistics	
Nodes	81306
Edges	1768149
Nodes in Lagest WCC	81306(1.000)
Edgesnin Largest WCC	1768149(1.000)
Nodes in Lagest SCC	68413(0.841)
Edgesnin Largest SCC	1685163(0.953)
Average clustring coefficient	0.5653

IV. Epinions social network [28], Dataset

This dataset is collected from a customer review of Epinions.com members based on who-trust-whom in the online social network. There is no doubt that most of the trusted relationships can interact and make the trusted Web; it consists of 75879 Node and 508837 edges.

V. The Trust-Aware Recommender(Epinions, Flixster) [29] Dataset

In the current dataset, there are double datasets: the first one is Epinions dataset, whereas the second is called the Flixster dataset. The former is for a products review website. While, Flixster is considered as a kind of social networking service to enable the users to rate ads friends and movies, the details of the dataset are shown in Table 3.

Table 3: Epinions and Flixster dataset details

Dataset type	Epinions	Flixster
Users	49.290	104.9445
Items	139.738	492.359
Rating	664.824	8.238.597
Trust Relation	487.181	26.771.123

9. Discussion and Summarizing

Generally speaking, various kinds of literature have examined the structural topology of OSN to design and implement their recommendation system. Moreover, using topological and structural features has been demonstrated to work

well on a variety of datasets sizes and algorithms. The element and features are observed to be successful, especially with evolutionary algorithms, such as Bee colony and Genetic algorithms. Evolutionary-based friend recommendation systems have yielded significant results over traditional friend-of-friend methods on large datasets with some settle differences in the methodology. As seen, some techniques, such as AI Bee Colony and Genetic algorithms are needed to incorporate the information derived from clustering algorithms to obtain a higher performance in an extensive scale network. The concept of "trust-aware" as seen has shown significant results. Using the attracting-repelling mechanism that reflects the friend-foe relationship in a social network makes it convenient to be used in product-centered social networks such as Amazon's IMDB. In these types of system, pushing a user away from dissimilar users (foes) is as equally important as bringing them together with similar users (friends). However, requiring more metadata about users to fetch more features might be a disadvantage of such a system. So, this system is recommended to be used when abundant information is exists on the users, and the type of recommendation service is multifaceted.

Features extracted from users images and photos may be an indication of users' preference and personality. Using the co-clustering algorithm in combination with features obtained from users' images has been successful, but only when used on informative datasets. In real life scenarios, it is challenging to obtain consistent photos of users such that they reflect specific information. For example, to decide whether or not the users have powerful interests from their user profile or images, a large number of photos related to sports must be obtained, and this is most unlikely to be the case. Besides, the efforts required to de-bias and remove the noisy information from users photo make it infeasible. In short, in the light of the studies reviewed in this study, extracting features from users photos is neither successful, nor practical, so they are recommended to be used only when necessary. While collaborative filtering has been the most frequently used method in item-based friend recommendation systems, and not deemed success when applied on the user-to-user recommendation. Treating the latter as a supervised classification task has shown promising results in the studies reviewed in this work. Different classifiers were used with various data and situations, and they mostly yield outstanding performance. For example, both Support Vector Machine and Naïve Bayes algorithms were trained on data collected from blogs with straightforward features result in high precision and recall. However, one drawback of these systems was that when applied on large datasets, it were observed that computationally expensive in training time. Still, such a disadvantage is traded off for performance.

10. Conclusion

In this study, several recommendations are concluded. First, the Genetic Algorithm along with clustering features, and AI Bee colony ABC algorithm seemed to be a generally optimal solution when used on a large dataset. Second, if the type of recommendation system requires repelling users from each other's, the trust-aware framework should be incorporated. Third, features based on time and location are more affluent in information and much easier to work with than those extracted from user images. Finally, in cases of user-to-user recommendation it is recommended to implement a classification-based machine learning pipeline that includes one of the state-of-the-art algorithms, such as SVM or NB.

References

[1] F. Jiang, C.K. Leung, and A.G.M. Pazdor, "Big data mining of social networks for friend recommendation," IEEE ACM International Conference on Advances in Social Networks Analysis mining pp. 921-922, 2016.

[2] Z. Deng, B. He, C. Yu, and Y. Chen, "Personalized Friend Recommendation in Social Network Based on Clustering Method," Communications in computer and Information science, no. 316, pp. 84-91, 2012.

[3] M. Moricz, Y. Dosbayev, M. Berlyant, "PYMK: Friend recommendation at MySpace," Conf. Manage. Data Proceedings of the ACM, pp. 999-1002, 2010.

[4] F. Akbari, A. H. Tajfar, and A. F. Nejad, "Graph-Based Friend Recommendation in Social Networks Using Artificial Bee Colony," IEEE 11th International Conference on Dependable, Autonomic Secure, Computing, pp. 464-468, 2013.

[5] N.B. Silva, I.-R. Tsang, G.D.C. Cavalcanti, and I.-J. Tsang, "A graph-based friend recommendation system using Genetic Algorithm," IEEE Congress on Evolutionary Computation, pp. 1-7, 2010.

[6] J. Naruchitparames, M.H. Gunes, S.J. Louis, "Friend recommendations in social networks using genetic algorithms and network topology," pp. 2207-2214, 2011.

[7] M. Eirinaki, M.D. Louta, and I. Varlamis, "A trust-aware system for personalized user recommendations in social networks," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 44, no. 4, pp. 409-421, 2014.

[8] S. Huang, J. Zhang, S. Lu, and X.-S. Hua, "Social Friend Recommendation Based on Network Correlation

and Feature Co-Clustering," presented at the Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, Shanghai, China, 2015.

[9] M.M. Hasan, N.H. Shaon, A.A. Marouf, M.K. Hasan, H. Mahmud, and M. M. Khan, "Friend recommendation framework for social networking sites using user's online behavior," in 2015 18th International Conference on Computer and Information Technology (ICCIT), 2015, pp. 539-543.

[10] H. Siyao and X. Yan, "Friend recommendation of microblog in classification framework: Using multiple social behavior features," International Conference on Behavior, Economic Social, Computing, pp. 1-6, 2014.

[11] M. Wu, Z. Wang, H. Sun, and H. Hu, "Friend Recommendation Algorithm for Online Social Networks Based on Location Preference," International Conference on Information, Science Control, Engineering, pp. 379-385, 2016.

[12] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, "Recommender Systems Handbook," ed. Boston, MA: Springer Science+Business Media, LLC, 2011.

[13] H. Ma, H. Yang, M. Lyu, and I. King, "SoRec: Social recommendation using probabilistic matrix factorization" pp. 931-940, 2008.

[14] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," KNOSYS Knowledge-Based Systems, vol. 46, pp. 109-132, 2013.

[15] S. Deb Roy, T. Mei, W. Zeng, and S. Li, "Social Transfer: Cross-domain transfer learning from social streams for media applications, ", pp. 649-658, 2012.

[16] J. Hannon, M. Bennett, and B. Smyth, "Recommending Twitter users to follow using content and collaborative filtering approaches," pp. 199-206, 2010.

[17] N. Li and G. Chen, "Multi-layered Friendship Modeling for Location-based Mobile Social Networks," in 2009 6th Annual International Mobile and Ubiquitous Systems: Networking & Services, MobiQuitous, pp. 1-10, 2009.

[18] X. Xie, "Potential Friend Recommendation in Online Social Network," IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing pp. 831-835, 2010.

[19] C. Song, S. Owusu, and L. Zhou, "Social Network Based Recommendation Systems: A Short Survey," International Conference on Social, Computing, pp. 882-885, 2013.

[20] X. Ning, C. Desrosiers, and G. Karypis, "A Comprehensive Survey of Neighborhood-Based Recommendation Methods," 2015.

[21] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative Filtering Recommender Systems," Lecture notes in computer science. no. 4321, pp. 291-324, 2007.

[22] N. VAIDYA and K. A.R, "Survey on Types and Techniques Used in Recommender systems" International Journal of Advanced Computational Engineering and Networking, vol. 5, no. 6, 2017.

[23] B. Mobasher, R. Burke, and J. J. Sandvig, "Model-based collaborative filtering as a defense against profile injection attacks," presented at the proceedings of the 21st national conference on Artificial intelligence - Volume 2, Boston, Massachusetts, 2006.

[24] R. Burke, "Hybrid Web Recommender Systems," in The Adaptive Web: Methods and Strategies of Web Personalization, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 377-408.

[25] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Transaction on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734-749, 2005.

[26] C. Amri, M. Bambia, and R. Faiz, "Behavior-Based Approach for User Interests Prediction," in IEEE ACS the International Conference on Computer Systems Applications, pp. 541-548, 2017.

[27] J. McAuley and J. Leskovec, "Learning to discover social circles in ego networks," Presented at the Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, Lake Tahoe, Nevada, 2012.

[28] M. Richardson, R. Agrawal, and P. Domingos, "Trust Management for the Semantic Web," In the Semantic Web - ISWC 2003, Springer Berlin, Heidelberg, , pp. 351-368, 2003.

[29] S. Deng, L. Huang, G. Xu, X. Wu, and Z. Wu, "On Deep Learning for Trust-Aware Recommendations in Social Networks," IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 5, pp. 1164-1177, 2017.