Semantic link-based Model for User Recommendation in Online community

Abeer ElKorany

Computer Science Department , Faculty of compyuters and information, Cairo University

a.korani@fci-cu.edu.eg

ABSTRACT

Recommendation systems have been widely used to overcome the problem of information overloading and help people to make the right decision of needed items, such as: movies, books, products, or even people. This paper proposes a semantic model for people recommendation in online community. This model predicts significant links that would be established between community' members even they do not know each other using cascaded collaborative filtering (CF). Two major steps in this type of recommendation model are (i) the method used to compute similarity between people, and (ii) the method used to combine these similarities in order to compute the overall similarity between target member and others. By utilizing local attributes of nodes and links, similarity between members is calculated. Semantic relatedness between members is delivered from connection strength and trust score in order to identify the closeness between them. Extensive experimental of the proposed model using real dataset of scientific community was applied to recommend authors as possible coworkers for a target researcher. Experimental results on real dataset from publication network show that the proposed model for people recommendation outperforms other known techniques in ranking recommended collaborators.

Keywords

Recommendtion systems, Link prediction, People recommendation, Similarity measurement.

1 INTRODUCTION

Social collaboration refers to processes that help multiple people interact, share information or knowledge in order to to achieve any common goal. Social collaboration media can be classified into two main types: Social networking and online communities. A social network is a social structure between actors, indicates the ways in which they are connected through various social familiarities ranging from casual acquaintance to close familiar bonds [9]. Online communities form a fundamental part of the web today where a large portion of the Internet's traffic is driven by and through them [1]. They are fundamentally defined by the repetition of interactions among a set of nodes (representing actors) over time known as *knowledge workers* who used to share content, seek support, and socialize. In these communities, people are not aware of each other, but people tend to communicate with their peers, whom they trust, to get advice. Accordingly, people to people recommendation has recently become an important task in many online communities where it is significant to find the "right" actor who can interact with or work together. A recommendation service in such communities must not suggest just anybody who is possibly relevant, but has to check up the trust score of such member in order to be able to give advice. Thus, it is more rational to deliver recommendations within an informal community of users using their interest and social context. In research community for example, people tend to interact with others who have similar research interests [10]. Ttherefore, it is significant to consider social structure of user community as well as users' profiles and social behaviors, as an additional source of information in building people recommendation service.

There are two prevalent approaches for building recommender systems: content-based (CB) [20] and collaborative filtering (CF) [8]. Traditional content- based RS recommends items similar to those the user preferred in the past. It also consider feature of items preferred by users in order to get similar items. While collaborative filtering recommends items that other users with similar preferences have liked in the past [13]. Therefore, it is necessary to define the community (similar people) in order to identify their preference. Hybrid approaches combine content-based and collaborative filtering methods in several different ways [21] in order to achieve a better performance of filtering, and take the advantages of each.

This research work aims to help finding appropriate member(s) to communicate with in collaborative environment such as online communities where users are not familiar with each other's. Thus, we propose a cascaded collaborative filtering model for people recommendation (CCFPRE) that predicst candidate users to communicate with in the future. In people recommendation, "items" are also users who actively participate via community interactions which differ from traditional CF. Mutual nature of such interactions is utilized to recommend, for a given user, other users they may like to contact. CCFPRE is based on calculating similarity between target user and other users using semantic local features of nodes within a community. It starts by modeling information about community members (nodes); identify link-based features that is used in calculating the tie strength, and trust score between members. Next, similarity engines are used to calculate similarity between target user and others using different similarity measurement techniques appropriate to user's feature(s). A new promising spread similarity technique is proposed here which consider the length and weight of connection among members in online community. The main contribution of this research is providing an advanced recommendation model that utilizes link features of users in online community to identify the expected reciprocity. Section2 describes different recomemnation techniques that could be applied to people recommendadtyion. Section3 illustraetes the proposed CF model for people recommendation. Link-based recommendadtion engine is described in section4. Experiment applied on real world online community for reseraccher is explained in section5. Experiments show that the proposed model outperform previous recomemndadtion techniques. Finally, conclusion and future work is described in section6.

2 RELATED WORK

Recommendations systems have been applied in diverse fields, most of it focus on user-product recommendations emphasizing on the one-way relationship between user and the products they are interested in buying. These recommender systems cannot be directly applied for people recommendations due to the existence of two-way relationship between people as well as highly sparse of dataset. People-to-people matching systems were initially used for dating services, where users join with the objective of meeting other users with the common need. Peaple are matched based on metric that include some personal information such as education and professional background, personal interests, hobbies, etc. Some real-world examples of these systems are employer-employee (in job search networks), mentor-student (in university social networks), consume-to-consumer (in marketplaces) and male-female (in an online dating network)[14]. In social networks, users are "flooded" with information from feed reader and many other resources, thus recommendation of other people that a member may like has recently become an important task in many online social networks. Several systems have been proposed for people recommendation in social network such as CollabNet[6], which uses gradient descent to learn the relative contributions of similar users or items to the ranking of recommendations produced by a recommender system. By using weights to represent the contributions of similar users for each active user in social networks, similarity among people is measured which is then used to identify and ranking people. SocialCollab [5] has adapted traditional collaborative filtering algorithm to predict, for a given user in social network, other users they may like to contact. User similarity is calculated based on both attractiveness and taste. Users' taste is defined through s their favorites when they actively make decisions selecting other users. While attractiveness is measured based on their involvement in interactions by being selected by other users. Another proposed a people to people recommendation system was proposed in [15] which utilized tensor models that have the ability to correlate and find latent relationships between similar users based on both information, user interactions and user attributes, in order to generate recommendations. People recommendation could also be considered as a link prediction problem in heterogeneous and reciprocal networks. The system developed in [4] used structural features and structural collaborative information about people, and properties of links between people to measure positive and negative signs through its path. Recently,

approaches based on analysis of both content of micro-blogs to detect users' interests and in the exploration of the topology of the network to find candidate users for recommendation has been proposed [2,3]

3 PEOPLE RECOMMENDATION MODEL IN ONLINE COMMUNITY

Collaborative filtering (CF) methods recommend items based on aggregated user preferences of those items. The main principal of CF technique is that when people historically select the same set of items, this indicated that they have the same interest and therefore, are likely to be interested in other items in the future. However, people recommendation is different since a user would have a dual role as both "users" and "items", and the assumption of active users and passive items in traditional CF cannot be applied here. Thus, people recommendation is based on defining similarity between users in terms of the mutual properties such as: attractiveness, interest, and taste. Therefore, the proposed CF people recommendation model involves two important challenges which are illustrated in this section. First, accurately identifying the user preferences, this is based on utilizing link-based attributes that distinguish users within the community. Second, assign different similarity measurement mechanism for each type of attributes which is used to rank the unified set of recommended users.

3.1 Semantic User Modeling

User model contains information that the system knows about the user concentrating on modeling of user knowledge, plans, and preferences in a domain [10]. Modeling the user would help in identifying the taste of each active user using information collected from other connecting users as well as the user' activities. In online community, this information is either explicitly defined or implicitly inferred. Explicit aspect provides us with description about the user's regular information which leads to form information network while the implicit knowledge enables us to add meta knowledge (semantic attribute) about users. Attributes used for modeling of user in online community could be classified into: content-based and link-based each is related to set of features of member. Content-based attributes such as interest, emotional, and taste are extracted from post, comments, documents, tags, published by user. While, link-based attributes such as: connectivity, behavior and trust are identified using type and frequency of interaction between users, activities, and community pattern. In this research we focus on only link-based attribute and classify link-based attributes into direct and indirect (semantic) attribute. Direct attributes are found when two member of the community explicitly communicate with each other through sending message or comment on the each other posts. While, semantic attributes represent hidden relationships extracted from indirect relationships. Number of common friends, frequency of joining in the same type of events or activities, and having the same social pattern are example of such semantic attributes. Therefore, the proposed user model would contain both types of attributes to be used in measuring he strength of ties between users as well as trust score which is used in the proposed model to enhance the coverage of recommendations process.

3.2 User Similarity Measurement

In traditional CF, the most important step is to define similarity. Two items are similar if both are selected together by the same set of users. Alternatively, two users are similar if they both select the same set of items (i.e., they have similar taste). In online communities, "items" as the recipients of actions are also *users* who are contributing in network interactions. Therefore, traditional CF cannot be directly applied to people recommendation, since we don't have an explicit rank from user to other users. Thus, similarity among members is calculated based on connectivity, behavior and trust. Motivated by the homophily [18], users in online community are more easily influenced by the friends they regularly communicating with and trust. Strength of bonds between members and trust score are key issues to associate similarity value between members. The proposed recommendation model extend the traditional CF to handle the specific nature of online community such that it would recommend users to a target $user_x$ through the following assumptions:

If people similar to target $user_x$ are connected to $user_y$, then $user_x$ could also be connected to $user_y$,

If people similar to target $user_x$ would trust $user_y$, then $user_x$ could also trust $user_y$

Therefore, the proposed framework works as follows: it identify similar people for a candidate u_x based on connective attributes WA_x using spread similarity technique based on. Next, HIT algorithm is applied to calculate trust similarity score between people to u_x based on trust-based attributes TS_x. Motivated by the idea of ranking the items in the right order to get the Top-N recommendation list[7,11], potential recommendations list for user u_x is assembled and re-ranked according to the intersections between the sets $LW_x \cap TS_x$.

4 LINK-BASED COLLABORATIVE FILTERING RECOMMENDER ENGINE

Real-world recommendation systems are typically hybrid systems that combine multiple recommendation engines to improve recommendadtions. Current research in social network considers recommendations based on links analysis as a *Link Prediction* problem. Link prediction in social networks, tries to infer which new interactions among members are likely to occur in the near future. Generally, there are two main approaches [23] that handle the *link prediction* problem. The first approach considers local features of a network, focusing mainly on the nodes structure such as the number of common friends that two users share as well as mutual interaction. While the second approach consider the global features, and thus detection of the overall path structure in a network being computationally expensive for huge-size social networks. In this section, a cascaded link-based people recommender engine is illustrated which consider the local features of the node in social network. The main idea of recommender engine is to utilize link-based attributes of users that cover both connectivity and trust which are used to estimate strength of ties as well as trust score between nodes (users) and accordingly calculate the similarity between users. During phase1, strength of links between nodes in online community is

calculated based on user's interaction features. While phase2 cacaulate trust scores using distinguished potential relations in order to converge ranking of recommended users generated from phase1.

4.1 Problem Definition

The proposed recommender engine is generally applicable for any online community. The main challenge is to identify link-based attributes and to classify them to be either related to connectivity or trust features. In this paper, we use academic information network which stores publication and citation data in order infer authors' future collaboration with others in the same community. In publication network[23], direct relationship between member are in form of co-authorship, direct citation, and being working in the same research institution. While, indirect relations are co-citation, bibliographic coupling, and regularly attending the same conferences. Link recommender engine computes similarity between nodes by taking into account both the "connection", and "trust" strength between members. Accordingly, our model starts by forming a graph using direct relationship between members. The graph contains two types of nodes: authors, and papers and two types of edges representing the above mentioned direct relationships. An edge is introduced between two authors if they if they coauthored at least one paper. Another edge is introduced between an author and a paper if the author citied this paper as shown in figure1. Thus, we deduce different metric that are used to quantify the "similarity score" between authors by considering both connectivity and trust measurements. The recommendation process works as follow: First, a starting node (author) is selected for which we will recommend the "right" authors to interact or collaborate with in the future. Similarities between target author and others are calculated using both spread similarity and HITS algorithm. Those scores are then used to obtain lists of compatible people to be recommended to the target author. Next, semantic features are used to re-ranked obtained results and converge recommendation. Finally, disjoint set of top-k similarity values are recommended to the selected target author.
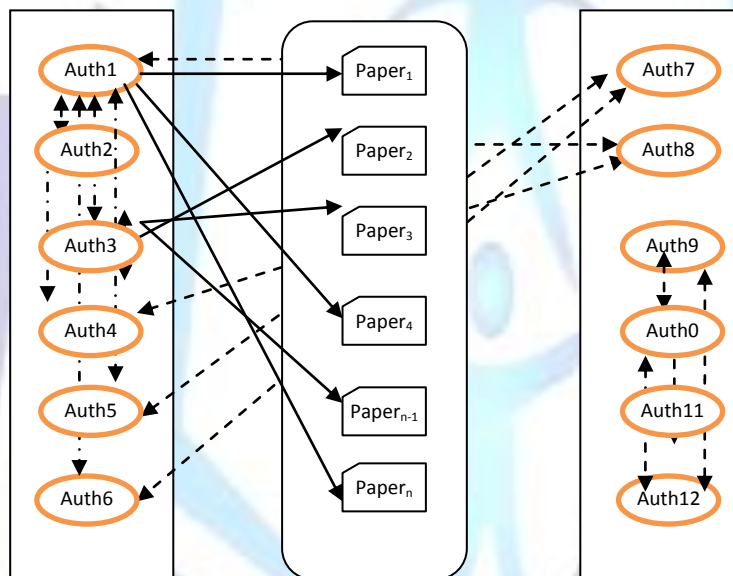

Figure1: Publication information networks

4.2 Spread similarity calculation using connective features

Using publications graph representation, connective similarity between authors is calculated. We adapt the FriendTNS algorithm [22]which is based on the degree of closeness between the neighbor nodes that form the pathways of maximum length of 2 between two nodes. This algorithm has been applied on real life social networks and over performed other local link approaches in terms of accuracy and time . It recommends friend based on the number of common friends each candidate friend has with the target user. In order to illustrate the effectiveness of proposed spread similarity technique, let us consider the case shown in figure2, which represents eight connected authors through their common publication and number appear on the links between nodes represent the frequency of interaction between users (the number of co-authored papers between two authors). When applying FriendTNS algorithm to recommend people to Au_2 which is connected to both author Au_4, Au_1, it yields to generate the following list Au_3, Au_5, Au_6, Au_8. However, this list is not ordered yet. Therefore, FriendTNS algorithm applies the following formula to compute similarity between neighborhood nodes based on the inverse sum of node degree in order to rank the above list of candidate authors.

$$Sim(x,y) = \frac{1}{neighbour(x) + neighbour(y) - 1} \qquad Equ(1)$$

However, this formula considers only the binary relationships between people (the existence of relation between two people) and accordingly similarity values between all non-neighbor nodes in a graph is set to zero. Furthermore, this formula does not take into account the semantic of the relationship among people in the same community which is measured through the frequency of interaction (generally it could be number of posts which two people sent to each other,
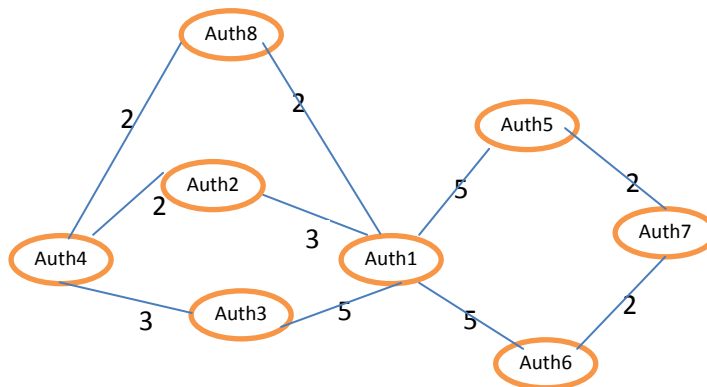
Figure2: Graph representation of author in publication network

number of replies, or number of co-author papers as we considered here). In online community, frequency of online interactions between a pair of users directly impacts the degree of relationship strength. The stronger the relationship, the higher likelihood that a certain type of interaction will take place between the pair of users[24]. Accordingly, node similarity is essentially emphasized the notation relationship strength between nodes. Therefore, we multiply the above formula by the frequency of interactions to calculate the node similarity. We use the following formula to calculate node similarity between two connected nodes.

$$Sim(x,y)= \frac{frequency\ of\ online\ interactions}{neighbour(x)+neighbour(y)-1} \qquad Equ(2)$$

Equation2 is used to calculate basic node similarity and thus similarity matrix is created between users as shown in table1. This matrix contains weight associated between connected nodes (authors). However, the proposed link-based recommender model aims to form a similarity index to weight (predict) unconnected nodes in order to recommend suitable authors to interact with.

Table1: Basic node similarity matrix between authors

	Auth1	Auth2	Auth3	Auth4	Auth5	Auth6	Auth7	Auth8
Auth1	0	0.5	0.833	0	0.833	0.833	0	0.334
Auth2	0.5	0	0	0.334	0	0	0	0
Auth3	0.833	0	0	0.75	0	0	0	0
Auth4	0	0.334	0.75	0	0	0	0	0.667
Auth5	0.833	0	0	0	0	0	0.667	0
Auth6	0.833	0	0	0	0	0	0.667	0
Auth7	0	0	0	0	0.667	0.667	0	0
Auth8	0.334	0	0	0.667	0	0	0	0

Thus, after calculating basic node similarity matrix, link-based recommender engine identifies the set of neighboring nodes to a target user$_x$. Then through a progressive process, other connected nodes to this set are collected with their predefined basic similarity. Next, spread similarity is calculated for these unconnected set for target user using the product of the basic similarities between the nodes of the shortest path from user$_x$, to other nodes up to two successive nodes in the collected set. Finally, the top-k users are recommended to the target user$_x$. To explain this process let us recall the case of Au_2 is which is connected to both user Au_4, Au_1. Au_4 is connected with 2 other nodes (Au_3 and Au_8), while Au_1 is connected with 4 other nodes (Au_3, $Au5$, $Au6$, $Au8$). Spread similarity is then calculated for both sets connected to Au_4, Au_1 in order to find the set of recommended users. As seen from that example Au_3 appears in both recommended list of Au_1 and Au_4. Therefore, it is significant to consider this intersection of recommended users as well. Accordingly, spread similarity (Ssim) equation is defined as:

$$Ssim(X,\ Y) = \prod_{i=1}^{k-1} sim(X,\ X+1) \qquad Equ(3)$$

Applying equation3, spread similarity between nodes Au_2 and Au_3 using equation (3) equals:

$Ssim($Au 2, Au3$) = (\sum sim(Au_2,\ Au_4) \cdot sim\ (Au_4,\ Au_3)+\ sim(Au_2,\ Au_1) \cdot sim\ (Au_1,\ Au_3)$

$\qquad\qquad = (\quad 0.\ 0.334 * 0.75 \qquad\quad +\ 0.\ 499* 0.\ 0.835)$

$\qquad\qquad = 0.\ 876$

Table2: Spread similarity matrix between authors

	Auth1	Auth2	Auth3	Auth4	Auth5	Auth6	Auth7	Auth8
Auth1	0	0.5	0.833	1.014	0.833	0.833	1.101	0.334
Auth2	0.5	0	0. 876	0.334	0.833	0.833	0	0.389
Auth3	0.835	0.876	0	0.75	0.697	0.697	0	0.787
Auth4	1.014	0.334	0.75	0	0	0	0	0.667
Auth5	0.835	0.4166	0.697	0	0	1.001	0.667	0.277
Auth6	0.835	0.4166	0.697	0	1.1421	0	0.667	0.277
Auth7	1.101	0	0	0	0.667	0.667	0	0
Auth8	0.334	0.389	0.787	0.667	0.277	0.277	0	0

When applying spread similaty, it would recommend the following order of authors to Au_2: (Au_3, Au5, Au6, Au8) as shown in table2. Then, semantic connective features such as attending the same conferences are used to converge (re-rank) the set of top ranked authors. The number of common attended conference between target author and each author in the recommended list would increase the score of that author.

4.3 Trust-based node similarity calculation

Traditional recommender systems ignore trust relations among users [16]. In real world, users can be easily influenced by friends trust. Trust is the measure of willingness to believe in a user based on its competence and behavior within a specific context at a given time [25]. Trust-based recommendation methods assume additional heuristics knowledge about the trust network among users[17]. This knowledge is related to the asset of trust value between individuals. Trust score between nodes is computed based on nepotistic relationships [19] which are either explicitly or implicitly defined. Users in online community are able to assign trust values for each other through voting or reputation scores. On the other hand, trust score can also be inferred through tracing who comment back a positive review such as in on eBay, following a new follower or re-tweet in Twitter, etc. In publication network, citation is considered as a key attributes in trust measurement which assume that authors tend to trust the work she/he cited to. Thus, citation matrix is first created which relate author to cited paper, next HITS ("Hypertext induced topic selection") algorithm [12] is used to identify candidate list of authors to be recommended to a target user. Essentially, HITS algorithm is based on setting a hub and authority values for each web page and create adjacency matrix to be used to rank pages with respect to a specific query. HITS uses an iterative approach, it assigns two scores to each node: a hub score and an authority score. Through transferring between the nodes, interactively change the values for hub and authority and rank the nodes according to the values. Similarly, in the proposed model, authors citations are used to create citation matrix as shown in table3.

Table3: citation matrix between authors

	Auth1	Auth2	Auth3	Auth4	Auth5	Auth6	Auth7
Auth1	15	8	3	7	4	3	3
Auth2	4	113	0	12	13	1	5
Auth3	1	1	15	1	1	1	0
Auth4	7	17	0	260	149	39	4
Auth5	4	10	0	86	304	23	7
Auth6	1	2	0	24	38	25	1
Auth7	3	8	1	4	7	1	29

Next, authority score is caculated for those authors identified from phase. This score converges after a few iterations (we use the normal 20 iterations). Thus, authors are ranked using their "authoritieness" value which provides a measure of importance of an author (target) with respect to other neighboring nodes. A good authority is a user who obtained more citation from her/his colleagues which is measure of her/his trusts score while good hub is a user who cites to her/his colleagues. Thus, authority measure is used by trust-based recommender engine as an indicator for author trust score. In the above example, authority scores for those authors are {0.020447, 0.063971, 0.000349, 0.590895, 0.795742,0.112491,0.021734} respectively. Co-citation and bibliographic coupling have become standard measurements in scientometrics for detecting author similarity [10]. Thus, co-citation is considered here as sematic feature which is used to increase trust score among users such that the increasing the number of co-cited papers, the oscillation of trust score.

5   EXPERIMENTAL PROTOCOL AND EVALUATION METRICS

The proposed framework is general and can be applied to any online community. In order to describe the effectiveness of the proposed model, we apply a set of experiments on large publication datasets which are publicly available.[1] The basic idea of experiments is to validate correctives, accuracy, and effectiveness of proposed model. This is achieved through applying a set of experiments which are explained in this section.

5.1   Experimental Setup and dataset

We perform our experiments on real-world datasets containing academic coauthors network and papers citation network. Both are extracted from academic search system Arnetminer[2]. The coauthor data set consists of 1,036,990 authors and 1,554,643 coauthor relations, while the citation data set contains 1,632,442 papers and 12,710,347 citations between these papers[23]. Furthermore, this dataset contains 2166 pairs of advisor-advisee coauthors as well as 3937 pairs of coauthors as colleagues. Also for each paper, a reference to the conference's name in which this paper has been published is mentioned (which we use to identify authors who attended the same conference). We use the following attribute to model the user connectivity feature: author ID, list of co-authors associated with the number of co-author papers, list of advisor (students), list of colleagues, and list of authors who attend the same conferences. The last three attributes are considered as semantic attribute which are used to enhance strength of ties among authors. Moreover, we use the following attributes to model the user trust feature: author ID, list of cited authors associated with the number of citation papers, list of co-citation authors, and list of bibliographic coupling. The last two attributes are considered as semantic attributes which are used to enhance the trust score among authors. The experiment setting was defined as follows. First, in order to validating the correctness of model , we divide the list of co-author for each user into two separate lists: original and test sets with ratio 80, 20 % respectively such that information in the test set is not allowed to be used for recommendation. Thus, the list of recommended author obtained using original set only is compared with test set in order to estimate the relevance of the recommended authors with those co-authors in the test set. Furthermore, a comparison between the proposed model with other recommendation mechanisms is applied which shows superior of CCFPRE. Finally, we compare the results obtained using the two similarity measure: spread similarity and trust score with the cascaded version.

5.2   Evaluation methodology

We use the classic precision performance measure for people recommendations. Accuracy of the proposed recommender system is evaluated by its ability to retrieve relevant co-authors among the top-N recommended authors. For a target $user_x$ receiving a list of *k* recommended authors (top-*k* list), precision is defined as follows:

Precision is the ratio of the number of relevant authors in the top-*k* list (i.e., those in the top-*k* list that belong in the test set of co-author of the target author) to *k*.

$$precision\ (RE) = \frac{co-authoer\ test\ \cap recommended\ author}{REcommended\ author}$$

Thus, precision is a measurement of the percentage of overlap between generated recommendation list and the user actual test set of co-authors. Precision is measured at different points in the ranked list of suggested collaborators. Thus, precision at rank *k* (P@*k*) is defined as the proportion of recommended authors that were relevant, i.e. were in the target user test set, between K recommended authors. In the following experiments, we evaluate precision for values of *k* equal to 5, 10, 20 respectively. Moreover, we use the Average Reciprocal Hit-Rank measure (*ARHR*) [7] to qualify the accuracy of recommendation model. This measurement is used to rewards each hit based on where it occurred in the *top-N* list. The number of hits is the number of co-author in the test set that were also present in the *top-N* recommended items returned for each user. Therefore, If *h* is the number of hits that occurred at positions *p*1, *p*2, . . . , *ph* within the *top-N* lists (i.e., 1 ≤ *pi* ≤ *N*), then the ARHR is equal to:

$$ARHR = \frac{1}{n}\sum_{i=1}^{h}\frac{1}{p_i}$$

This measure indicates that authors who occur earlier in the *top-N* recommendation lists are weighted higher than authors occur later in the list. The highest value of ARHR is equal to the hit-rate and occurs when all the hits occur in the first position, whereas the lowest value of the ARHR is equal to *hit-rate/N* when all the hits occur in the last position in the list of the *top-N* recommendations.

5.3   Experimental results

The first experiment targets the correctives and accuracy of the proposed model. Thus, CCFPRE is applied over a set of 10 different authors, lists of top ranked 20 candidate authors for ercah of them are then retrieved, and precision and ARHR are calculated at different level in order to measure the accuracy of model.

---

[1] `arnetminer.org`

Table4: Precision of Proposed model

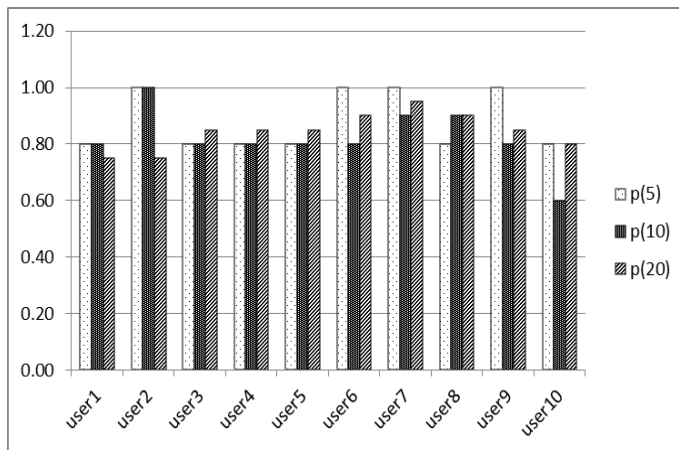|	p(5)	p(10)	p(20)
user1	0.80	0.80	0.75
user2	1.00	1.00	0.75
user3	0.80	0.80	0.85
user4	0.80	0.80	0.85
user5	0.80	0.80	0.85
user6	1.00	0.80	0.90
user7	1.00	0.90	0.95
user8	0.80	0.90	0.90
user9	1.00	0.80	0.85
user10	0.80	0.60	0.80

Figure3: precision of different 10 users

Tables 4,5 and Fig.3,4 show the results in terms of precision and ARHR which rewards each hit based on where it occurs in the top-*N* authors. It could be observed that the precision diminishes for longer recommendation lists, but have promising values for shorter lists as shown in first column in table4, where it reach the maximum for user2,6,7,9 respectively. Furthermore, the value of ARHR shown in table5 and figure4 rewards each hit based on where it occurs in the top-*N* collaborators that were recommended and provide a promising at

Table5: ARHR of Proposed model

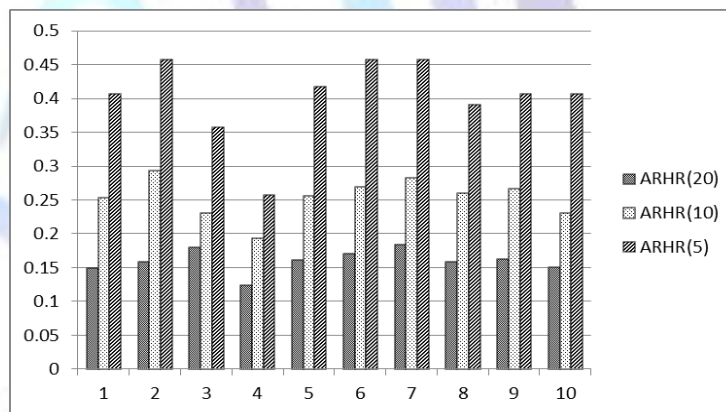|	ARHR(20)	ARHR(10)	ARHR(5)
user1	0.150	0.254	0.407
user2	0.158	0.293	0.457
user3	0.180	0.230	0.357
user4	0.124	0.193	0.257
user5	0.162	0.256	0.417
user6	0.170	0.269	0.457
user7	0.184	0.283	0.457
user8	0.159	0.260	0.390
user9	0.162	0.266	0.407
user10	0.151	0.230	0.407

Figure4: ARHR of different 10 authors using CCFPRE model

Other experiments have been applied in order to compare the accuracy of each proposed similarity measurement techniques with other state-of-the-art similarity measurement such as friend-TSN and HITs. Thus, we first compare spread similarity with Friend-TSN using both precision and ARHR values at different levels. According to Table6,7 which present those comparisons, where numbers in bold indicates the improvement of precision value when using spread similarity. It is significant to mention that the P(5), ARHR(5) obtained for user2 using friend-TSN is zero which means that none of recommended authors exist in the original test set while spread similarity has got a higher value indicating its effectiveness in recommendation. Furthermore, for all the results presented so far, the value of ARHR improved at all levels using spread similarity.

Table6: Comparison between semantic spread similarity and friend-TSN algorithm

	Friend-TSN p(5)	Spread similarity p(5)	Friend-TSN p(10)	Spread similarity p(10)	Friend-TSN p(20)	Spread similarity p(20)
user1	0.8	**1**	0.9	**1**	0.7	**0.75**
user2	0	**0.8**	0.1	**0.8**	0.5	**0.85**
user3	1	1	0.8	0.8	0.85	**0.9**
user4	0.6	**1**	0.8	**0.9**	0.9	**0.95**
user5	0.6	**0.8**	0.6	0.6	0.65	**0.8**

Table7: Comparison between spread similarity and friend-TSN algorithm

	FriendTNS-ARHR(5)	Spread-ARHR(5)	ARHR(10)	spread-ARHR(10)	ARHR(20)	spread-ARHR(20)
user1	0.357	**0.457**	0.243	**0.293**	0.136	**0.158**
user2	0.000	**0.417**	0.010	**0.208**	0.041	**0.124**
user3	0.457	0.457	0.266	**0.269**	0.162	**0.170**
user4	0.190	**0.457**	0.160	**0.283**	0.113	**0.184**
use5	0.367	**0.407**	0.217	**0.230**	0.132	**0.151**

The third experiment was applied to measure the effectiveness of the cascaded model. Thus, we compare the precision and *ARHR* at each level of cascaded model with other previous similarity measurement techniques such as FriendTNS and HITS. A random author was selected and precision and *ARHR* is calculated at each level of recommendation model and the result is shown in table8.

Table8: Comparison between P(K)and ARHR(K) at different levels

	p(5)	p(10)	p(20)	ARHR(5)	ARHR(10)	ARHR(20)
FriendTNS	0.00	0.10	0.50	0.00	0.01	0.04
spread similarity	0.80	0.80	0.80	0.42	0.21	0.12
semantic connective	0.80	0.80	0.80	0.42	0.21	0.13
Hit algorithm	0.80	0.40	0.20	0.41	0.20	0.10
semantic hit	0.85	0.45	0.25	0.42	0.21	0.10
Cascaded Filtering	0.80	0.70	0.65	**0.57**	**0.29**	**0.17**

According to table8, figure 5, which presents the precision and ARHR values at each level of cascaded model. It is significant to mention that the ARHR values has been improved for all levels which proof that CCFPRE outperforms other techniques in ranking recommended authors, and hence the effectiveness of cascaded model.
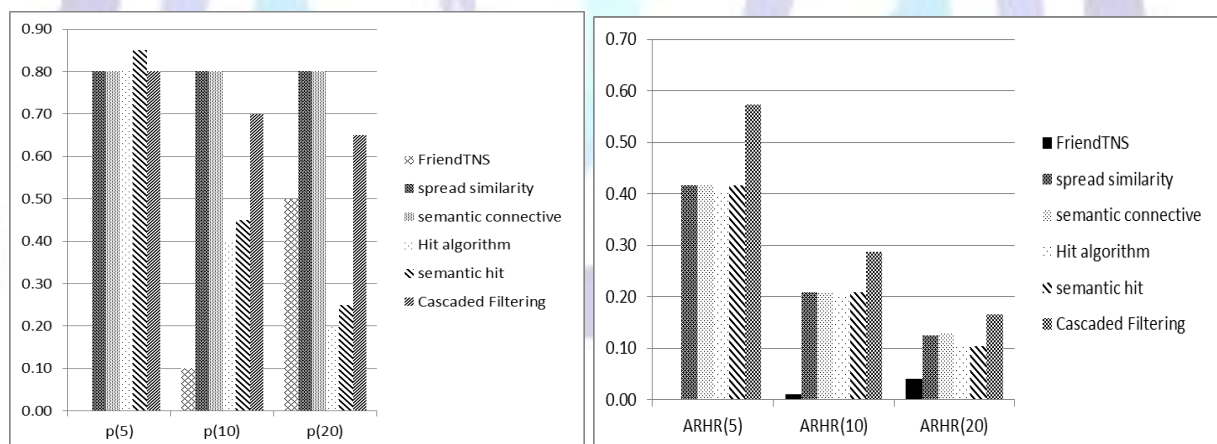

Figure5: Comparison between proposed algorithms and others stat of art

6 Conclusion

This paper presents a novel and effective model for recommending collaborators in co-workers environment. This model utilizes link-based attributes of members in online community to predict appropriate actor(s) to communicate with. The model depends on effectively identify suitable feature to model user connectives and trust attribute. Local attributes of links present relationships between people in community in order to handle the mutual nature of their interactions. Then, different similarity measurements techniquesa are utilized to identify for a target user, candidate recommendations and then sorts these candidates according different semantic weighting features. The results reported here from several experiments are promising especially for spread similarity technique that has been proposed here. Future work intent to consider the content published by users as an indication of her/his interest.

REFERENCES

[1] Angeletou, S., Rowe, M., & Alani, H. (2011). Modeling and analysis of user behavior in online communities. In *The Semantic Web–ISWC 2011* (pp. 35-50). Springer Berlin Heidelberg.

[2] Armentano, M. G., Godoy, D., & Amandi, A. A. (2013). Followee recommendation based on text analysis of micro-blogging activity. *Information Systems*.

[3] Armentano, M. G., Godoy, D., & Amandi, A. A. (2011, July). Recommending information sources to information seekers in Twitter. In *International Workshop on Social Web Mining*.

[4] Cai, X., Bain, M., Krzywicki, A., Wobcke, W., Kim, Y. S., Compton, P., & Mahidadia, A. (2012). Reciprocal and heterogeneous link prediction in social networks. In *Advances in Knowledge Discovery and Data Mining* (pp. 193-204). Springer Berlin Heidelberg.

[5] Cai, X., Bain, M., Krzywicki, A., Wobcke, W., Kim, Y. S., Compton, P., & Mahidadia, A. (2011). Collaborative filtering for people to people recommendation in social networks. In *AI 2010: Advances in Artificial Intelligence* (pp. 476-485). Springer Berlin Heidelberg.

[6] Cai, X., Bain, M., Krzywicki, A., Wobcke, W., Kim, Y. S., Compton, P., & Mahidadia, A. (2010, December). Learning collaborative filtering and its application to people to people recommendation in social networks. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on* (pp. 743-748). IEEE.

[7] Deshpande, M., & Karypis, G. (2004). Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems (TOIS)*, *22*(1), 143-177.

[8] Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. 1992. Using Collaborative Filtering to Weave an Information Tapestry. *Commun. ACM* 35, 12 (Dec. 1992), pp 61-70.

[9] Hanneman, R. A., & Riddle, M. (2005). Introduction to social network methods.

[10] Heck, T., Peters, I., & Stock, W. G. (2011). Testing collaborative filtering against co-citation analysis and bibliographic coupling for academic author recommendation. In *Proceedings of the 3rd ACM RecSys' 11 Workshop on Recommender Systems and the Social Web* (pp. 16-23).

[11] Kabbur, S., Ning, X., & Karypis, G. (2013). FISM: Factored Item Similarity Models for Top-N Recommender Systems, KDD'13, August 11–14, 2013, Chicago, Illinois, USA.

[12] Kleinberg, J. M. (1999). Hubs, authorities, and communities. *ACM Computing Surveys (CSUR)*, *31*(4es).

[13] Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., Riedl, J.: GroupLens: Applying Collaborative Filtering to Usenet News. Communications of the ACM 40(3), 77–87 (1997)

[14] Kutty, S., Nayak, R., & Chen, L. (2013). A people-to-people matching system using graph mining techniques. *World Wide Web*, 1-39.

[15] Kutty, S., Chen, L., & Nayak, R. (2012, March). A people-to-people recommendation system using tensor space models. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing* (pp. 187-192). ACM.

[16] Ma, H., King, I., & Lyu, M. R. (2009, July). Learning to recommend with social trust ensemble. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (pp. 203-210). ACM

[17] Massa, P., & Avesani, P. (2007, October). Trust-aware recommender systems. In *Proceedings of the 2007 ACM conference on Recommender systems* (pp. 17-24). ACM.

[18] McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 415-444.

[19] Ortega, F. J., Troyano, J. A., Cruz, F. L., Vallejo, C. G., & Enríquez, F. (2012). Propagation of trust and distrust for the detection of trolls in a social network. *Computer Networks*, *56*(12), 2884-2895.

[20] Pazzani, M.J., & Billsus D. 2007. Content-based recommendation systems. *The Adaptive Web*, 325-341

[21] Shih, Y. Y., & Liu, D. R. (2005, January). Hybrid recommendation approaches: collaborative filtering via valuable content information. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on* System Sciences IEEE.

[22] Symeonidis, P., & Tiakas, E. (2013). Transitive node similarity: predicting and recommending links in signed social networks. *World Wide Web*, 1-34.

[23] Tang, W., Zhuang, H., & Tang, J. (2011). Learning to infer social ties in large networks. In *Machine Learning and Knowledge Discovery in Databases* (pp. 381-397). Springer Berlin Heidelberg.

[24] Xiang, R., Neville, J., & Rogati, M. (2010, April). Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World wide web* (pp. 981-990). ACM.

[25] Yuan, W., Guan, D., Lee, Y. K., Lee, S., & Hur, S. J. (2010). Improved trust-aware recommender system using small-worldness of trust networks. *Knowledge-Based Systems*, *23*(3), 232-238.

Author' biography with Photo

Abeer El_korany received Ph.D. in Electronics and Communications Engineering, 2002, M.Sc. in Electronics and Communications Engineering 1996, B.S. in EE '92 from faculty of engineering, Cairo University, Egypt with honor degree. She was a key researcher in the expert system development at from 93-2003 at center laboratory of agriculture expert system (CLAES) before joining department of computer science,faculty of computers and information , Cairo university. Her research interests include quality assurance of knowledge based system, ontology, knowledge management, machine learning, social network, recommendation systems