Chen, Lin, Nayak, Richi, & Xu, Yue (2010) *Improving matching process in social network.* In: IEEE International Conference on Data Mining Workshops, 14 - 17 Dec, Sydney.

Improving Matching Process in Social Network

Lin Chen
Computer Science Discipline
Queensland University of Technology
Brisbane, Australia
l33.chen@student.qut.edu.au

Richi Nayak
Computer Science Discipline
Queensland University of Technology
Brisbane, Australia
r.nayak@qut.edu.au

Yue Xu
Computer Science Discipline
Queensland University of Technology
Brisbane, Australia
xu.yue@qut.edu.au

Abstract— **Online dating networks, a type of social network, are gaining popularity. With many people joining and being available in the network, users are overwhelmed with choices when choosing their ideal partners. This problem can be overcome by utilizing recommendation methods. However, traditional recommendation methods are ineffective and inefficient for online dating networks where the dataset is sparse and/or large and two-way matching is required. We propose a methodology by using clustering, SimRank to recommend matching candidates to users in an online dating network. Data from a live online dating network is used in evaluation. The success rate of recommendation obtained using the proposed method is compared with baseline success rate of the network and the performance is improved by double.**

Keywords-online dating; clustering; SimRank

I. INTRODUCTION

Personalization plays an important role in helping users identify the right products of their choice automatically. This reduces the user efforts to select the product manually among the massive choices available online. Traditional recommendation methods, content-based, collaborative or hybrid, consider the product detail, the user ratings or these two combined respectively [2]. With the advent of Web 2.0, online social networks have gained popularity in the last decade [4]. Additional to user ratings and product details, information such as users' behavior on the networks and their social relationships have also become available. This information is quite useful for making recommendations especially to users of the social networks.

Online dating is one of the many types of online social networks and is expanding rapidly with many people joining such dating networks. Different from traditional recommendation which is usually items to users recommendation, online dating recommendation requires two way matching; that is both users need to be interested in each other in order to start proper communication. Another challenge is how to efficiently find the matches for a user considering the number of online dating network's members is in millions.

An online dating network usually allows users to set up an account and join. It also allows members to create a profile which holds their own information and their preferences of a potential partner. These networks provide searching and limited matching mechanisms so that users can find potential partners. They usually require members to buy "stamps" in order to communicate with other members by secure means. This means of communication include sending a pre-typed message to show user's interest (called as "kiss" in this research), emails, instant message and chatting.

The online dating network is selected for the purpose of studying social recommendation systems because of its rich social connections and users activity. Pair to pair recommendation considering all members in the network is time consuming; therefore, the proposed method improves the recommendation efficiency by assigning users to groups. The SimRank method [8] is used for finding the similar users.

The proposed method is evaluated using the dataset collected from a live online dating website. Accuracy of the proposed method is measured as the "kiss" success rate and compared with the underlying network without using the proposed method. The success rate improves from 13.9% to 32.16%. The SimRank with in-link and out-link as the input information performs better than SimRank with in-link only and out-link only in terms of success rate and recall.

The rest of the paper is organized as follows. In section 2, related work is mentioned. Analysis of online dating social network is conducted in section 3. In section 4, the proposed method is defined and discussed. Experiment setup and results are pressed in section 5. Lastly, conclusions are drawn in section 6.

II. RELATED WORK

Since this paper combines the fields of online networks and recommendation systems, we briefly discuss research in the following three related areas.

A. *Traditional Recommendation System*

Content-based and collaborative-based recommendation systems are the most commonly implemented recommender systems [2]. Content based recommendation system learns the correlation of users and items either from user history or machine learning methods. However, the main drawback of this approach is the recommended items are too similar to the item that users have viewed before [3]. Collaborative-based recommender system collects all users' ratings of items, and identifies the similarity between users. Then it generates the ratings for unrated items based on similar users' ratings. One problem of this type of system is the high users' entry cost because of the requirement of explicit user feedback in form of ratings [7]. The rating information is used in collaborative approach. However, other information such as user demographic information, user relations in social networks is not integrated into the existing systems [19].

B. SimRank

Measuring "Similarity" is the key in the recommender systems. In general, similarity is the users' rating similarity or items' rating similarity in recommendation systems. Different from similarity measures in recommender systems, SimRank is a similarity based method that focuses on measuring the structural context in which objects occur, based on their relationships with other objects in a network [8]. There have been many variations proposed since its conception. Most of these works aim to improve the scalability and efficiency of the algorithm. In one paper [13], a technique is developed to estimate the finite number of iterations needed to converge the similarity score. P-Rank updates the SimRank algorithm by joining both "In" and "Out" link relations into computation [21]. However, only a minority of researchers have focused on applying SimRank in different applications. For example, work in [18] applied SimRank notion to heterogeneous data objects such as web page and user query context. Work in [20] proposed SimRank to measure similarities among academic papers based on references. To our best knowledge, SimRank has not yet been employed in recommendation systems.

C. Online Dating Recommendation

Only a handful work has been done related to online dating recommendation. The author [1] utilizes the existing collaboration recommendation method to data from online dating website. In this method, ratings are the only parameter which affects the match making algorithm. However, many factors such as age, job, ethnicity, education etc play an important role in the match making process. Work in [10] claims that not only the users' interests and demographic data need to be considered, but also their activities and relationships with other users need to be considered. This work gives a comprehensive thought over how a matching algorithm should be implemented. However, this work is at a theoretical level and there are no experiments carried out to prove the effectiveness of this theory. This work uses a weighted linear combination of various factors, which may negatively influence it being an effective algorithm. Another problem with this method is computation complexity due to the need of pair wise matching computation between all members in the network.

III. Online Dating Social Network Analysis

A. Online Dating Social Network

Users join an online dating social network[1] in order to communicate with potential partners and eventually set up a start of good relations. The user is usually asked to provide his/her profile and partner's preference during the registration. Profile includes information on demographic, fixed-choice responses on physical, identity, lifestyle, career, education, politics and religion and other attributes, free-text responses to various interests such as sports, music. Users are also allowed to upload photograph of them. Partners' preference includes the same type of information that users like the ideal partners to have. Users can have multiple choices for a preference attribute. If the registration is successful, the user can view other users' profile. Following up, the user can choose to initiate a kiss to other users. A kiss is usually a pre-typed message up to 150 characters to show user's interest. The recipient users can respond to the kiss with another kiss to show the nature of their response. The receiving users can also choose not to respond to a kiss. Usually, sending kiss is the first step to start the communication. Positive kiss reply encourages the follow-up communication type-email or chat. Users are considered in online relationship when they contact each other frequently via email or chat and before they move on to face-to-face contact.

There are many mode of communication, however, pre-typed short messages called as "kiss" are considered as the activity and measure of relationship between two users in this research. This communication mode is an effective way to show the distinct interests between two potential matches. Analysis of the dataset shows that positive kiss replies have a strong correlation with stamp purchase behavior on the network. Thus, positive kiss reply indicates not only the member's interest but also a good sign for the network revenue. Therefore the number of positive kisses is used in testing the proposed social recommendation system.

B. Small World Network Or Not

A social network exhibits the small world phenomena if any two individuals in the network are likely to be connected through a short sequence of intermediate acquaintance [11]. For example Web and YouTube have small world properties [14].

A small world network can be characterized with short average path length, small diameter, and high clustering coefficient [14]. Average path length is simply the average path of all-pairs-shortest paths on social network. Eccentricity is the maximal shortest path distance between a node and any other node (a path of the node to furthest other). The diameter is the maximum eccentricity across all vertices. Clustering coefficient is a measure of degree to which nodes in a graph tend to cluster together.

To test whether the underlying online dating social network is small world or not, 1000 random users who are active during the selected 6 months period are extracted from the network. Due to computation complexity, it is not feasible to consider all the nodes in the network. Analysis of the selected network shows that the network diameter is 14 and the average path length is 4.923. The Web, on the other hand, has a diameter of 16.12 and an average path length of 905 [17]. Compared to the Web, the online dating social network has smaller diameter and shorter average path length. However, the clustering coefficient is 0 for these 1000 nodes. The reason can be explained by this social network structure. In online dating social network, 97% of links are between males and females. The number of links exist in the same gender group are rare. The neighborhood of a male user only has female users and female users are rarely directly linked, similarly, the neighborhood of a female user only has male users.

[1] Due to privacy reasons the details of this network are not given.

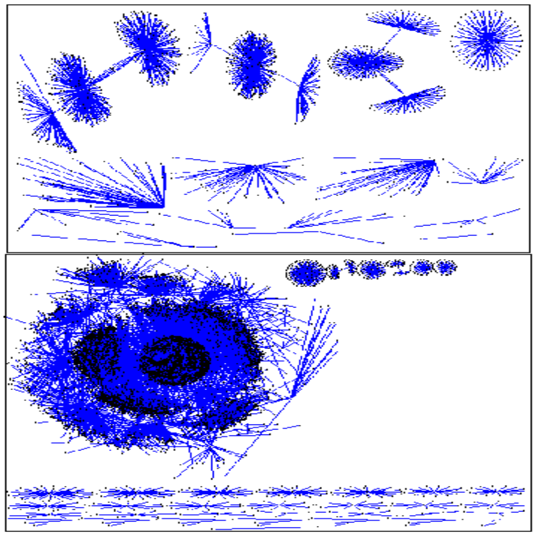

Figure 1. Selected Nodes Visualization

Top graph in Figure 1 is the graph of 10 out of 1000 random users and their links. Bottom graph is 100 random users and their links. It shows that only minority of users link to lots of other users and majority of users only link to a few users.

C. *Reachability*

Reachability is defined as the ability of a node to pass to another node in the network. If every node in the network can directly connect to the majority of all the other nodes, then the network is well-connected. Breadth-first-search (BFS) is implemented to test the reachability of the underlying online dating social network. BFS on a directed graph starts with a node *u* in the graph. It then counts the number of nodes reachable from *u* in a series of layers which are disjoint. The first layer has all nodes that are pointed to by links from *u*. A layer *k* consists of links which connected to nodes from layer *k-1* excluding those in any earlier layer. For the analysis purpose, we randomly selected 300 users who have logged in the dating network at any time during the defined six months as the starting nodes. Their communication records are observed for this experiment purpose and the direction of communication is the forward direction which means the layer k users are the initiators of the communications to layer k+1 users.

As a result, lots of nodes die out. These nodes are connected to few other nodes which also have few links or no links to other nodes. A small amount of nodes explode quickly after a few layers. Figure 2 shows the results for 2 to 5 layers. Noticed from Figure 2, 221 starting nodes out of 300 starting nodes are only linked to 1 other node or do not link to any other nodes where the second layer links have limited or no connections to other nodes. The left 79 nodes' reachability grows quickly. For example, the node with the maximum reachability in this test can reach 100 nodes at the second layer, 10^4 at third layer, $10^{5.2}$ at fourth layer and $10^{5.5}$ at fifth layer. This experiment shows that around 73% of nodes are linked to a few nodes, and only around 26% of nodes are able to connect to lots of nodes (more than 10,000). These experiments ascertain that a method, if it requires walking through the graph, should need to control the number of layers in the walk. Otherwise, it would become computationally untraceable to load the whole graph especially when a walk involves millions of nodes in this network

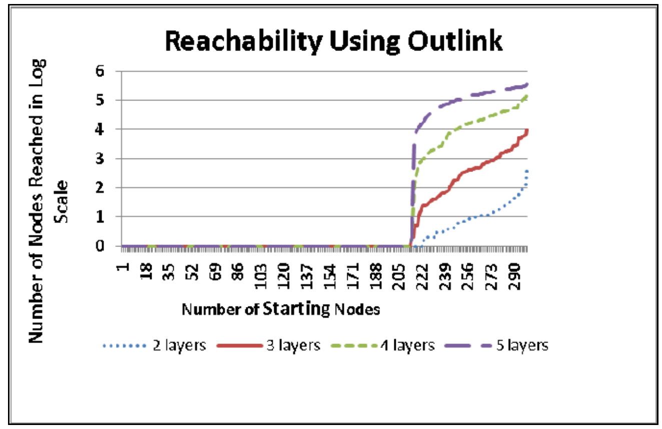

Figure 2. Reachability

D. *Density of Network*

The density of a network is the proportion of links actually present in the network. It is the ratio of number of links present to the maximum possible links [16]. In online dating networks, the majority of connections are between males and females. The number of connections for male-to-male or female-to-female is very small and, therefore, these connections are ignored in this research. The maximum numbers of possible links including in-links and out-links is $C_F \times C_M \times 2$, where C_F is the number of females in the network and C_M is the number of males in the network. The multiple of 2 is there to show the group counts for in-links as well as for out-links. Density of the underlying network is deemed to be sparse with a score of 0.00029. Work in [15] defines a sparse graph as $|E| = O(|V|^k)$ with *1<k<2,* where $|E|$ denotes the number of edges, $|V|$ denotes the number nodes and O is the big O notation. For our dataset, *k* is calculated and is found to be equal to 1.278. From both tests, it is shown that the network is a sparse network.

Collaborative recommendation systems are not able to deal with sparse network data. In this social network there are 2 million distinct users. The majority of these users only send around 20 kisses to potential partners (other users). Thus in most cases, for a given user there is only 20 other users who have been rated out of the total 2 million users in the network or around 210,000 active users during 3 months period. This means the similarity score between two users will be 0 in most cases, due to the lack of overlap in the few kisses being sent/received. It is hard to populate the rating data even with matrix decomposition methods. Therefore, a collaborative method alone cannot be applied to this network.

E. Variation between explicit and implicit information

Online dating networks require users to set up profiles in order to join as members. The user profile includes two types of information: (1) a list of personal profile attributes (2) a list of user's ideal partner profile attributes. Many times, however, users do not follow their preferences when they contact other users. Usually there remains inconsistency between the given user profile preferences and who they actually contact on the network. According to our experiment conducted, the sender does not follow his/her preference in searching and finding their interested users for 90% of time. This indicates that the preference only information should not be taken into the consideration as the input for clustering. A combined profile information and link information should be applied to overcome the shortcomings of using a single type of information as the source data.

IV. THE PROPOSED FRAMEWORK

A. Framework

Figure 3 shows the flow chart of the proposed method. Users are divided into a female group U_F and a male group U_M initially. A clustering algorithm is then applied to U_M and U_F each to divide the male and female users into smaller groups according to their explicit information. The next task is to find the similarity between each user *u* in a cluster with other members in the cluster. This task allows the user find the nearest neighbors in the cluster. SimRank algorithm is utilized to find the nearest neighbors. To compute the original SimRank score between members of a cluster, a graph which carries linked node information is generated and a similarity measure is employed. Finally, the system utilizes the collaborative filtering and recommends the *Top-n* potential partners to a cluster member that his/her nearest neighbors have contacted. The premise of this recommendation is that users who are similar would prefer similar partners. So the partners of two similar users can be recommended to each other.

B. Clustering

The purpose of clustering is to divide the large user base into smaller groups by identifying similar users in the network. The assumption is that by clustering similar users, the partners these users are interested in should be quite similar and be of potential interest to other users in the cluster. Users are clustered based on explicit information including profile and/or preference attributes. A combination of profile and preference information, or the profile information only, or the preference information only is used as an input for clustering. The Profile and preference combined information as the input for the clustering enhances the similarity between users in the cluster. As similar people are not only similar in their nature (how they describe themselves) but also their choice of ideal dating partners are similar. The preference only information as the clustering input is based on the assumption that people searching for similar type of partner contact similar candidates in reality. As mentioned before, data analysis in the network shows that users do not follow their preference to search their ideal partner. Therefore the profile only information is also considered as an input for clustering.

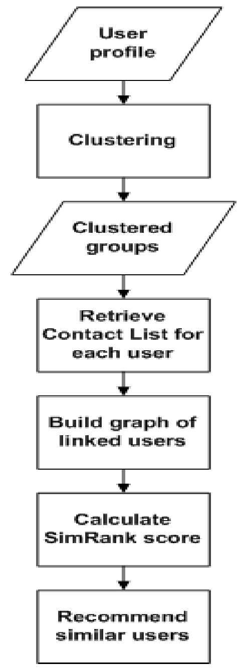

Figure 3. Proposed Framwork

C. SimRank

The clustering process identifies the smaller but similar groups, however, the similarity between group members is yet to be found. The SimRank score is calculated to measure the similarity between each pair of members in a cluster. The basic theory of SimRank [8] is that two objects are similar if they are related to similar objects. We have applied this SimRank theory to the dating network scenario assuming that two users are similar if they contact similar users. Let a, b denote as users in the cluster group C and $a \in U_M$, $b \in U_M$. Let $s(a,b)$ denote the SimRank score between two male users in C. $R_0(a,b)$ is the SimRank score of a, b at iteration 1 and is calculated using Equation 1. R_{k+1} is derived from R_k and to compute R_{k+1}, Equation 2 is used. $|O(a)|$, $|O(b)|$ are the number of out-link neighbors users a *and* b have. To compute R_{k+1} , iterate all out-link neighbors of $O_i(a), O_j(b)$ and sum up the similarity of $R_k(O_i(a), O_j(b))$ pairs. If SimRank calculation is based on in-links, $|I(a)|$ and $|I(b)|$ replace $|O(a)|$, $|O(b)|$ in equation

(2). $|I(a)|$, $|I(b)|$ are the number of in-link neighbors users a, b have. Finally, c is the damping factor which indicates the weakening confidence in two users being similar as the links (kisses) joining two users grow larger. As k increases, the SimRank score will converge. That is $\lim_{k\to\infty} R_k(a,b) = s(a,b)$. If both in-links and out-links are used, Equation 3 can be applied. The SimRank score calculation for a female group is computed analogously.

$$R_0(a,b) = \begin{cases} 0 & if \quad a \neq b \\ 1 & if \quad a = b \end{cases} \tag{1}$$

$$R_{k+1}(a,b) = \frac{c}{|O(a)||O(b)|} \sum_{i=1}^{|O(a)|} \sum_{j=1}^{|O(b)|} R_k(O_i(a), O_j(b)) \tag{2}$$

$$s(a,b) = \lambda \times \frac{c}{|I(a)||I(b)|} \sum_{i=1}^{|I(a)|} \sum_{j=1}^{|I(b)|} s(I_i(a), I_j(b)) + (1-\lambda) \frac{c}{|O(a)||O(b)|} \sum_{i=1}^{|O(a)|} \sum_{j=1}^{|O(b)|} s(O_i(a), O_j(b)) \tag{3}$$

D. Recommendation

Suppose cluster group C includes k number of male users represented as $C = \{u_{M_1}, u_{M_2}, ..., u_{M_k})$. A male user u_{M_i} in cluster C has either $\varnothing$ contact list or a set of contacts $L_{M_i} = \{u_{F_{i1}}, u_{F_{i2}}, ...u_{F_{in}}\}$. The SimRank scores can be computed for members in the cluster and obtained through the process previously described. Neighbors of a user in the cluster can be generated by taking SimRank score into consideration.

Let the ranked list of neighbors for male user u_{M_i} be denoted as $A_{M_i} = \{u_{M_a}, u_{M_b}, ..., u_{M_q}\}$ where $|A_{M_i}| = k-1$ an $u_{M_i} \notin A_{M_i}$ .Suppose $s(u_{M_i}, u_{M_a}) > \forall s(u_{M_i}, u_{M_p})$ where *p=(1, 2…k)* and $p \neq i$ *and* $p \neq a$. The top-1 recommendation for u_{M_i} is $L_{M_a} - (L_{M_i} \cap L_{M_a})$ where L_{M_a} is the contact list of user u_{M_a}. The Top-n recommendations for u_{M_i} are $(L_{M_a} \cup L_{M_b} ... \cup L_{M_x}) - ((L_{M_a} \cap L_{M_i}) \cup (L_{M_b} \cap L_{M_i}) ... \cup (L_{M_x} \cap L_{M_i}))$. For female users' recommendations, the process can be done accordingly.

V. Experiments and Disscussion

A. Dataset

Data is obtained from a live online dating website. There are around 2 million members who have joined this dating network. The dataset for this work contains 87,304 male users who are active during the selected 6 months period. A user is called active user if they have logged in at least once during this period. In the experiments, positive kisses are used as an indicator to determine whether the recommended user is suitable. If the user sends a kiss to our recommended user, and the recommended user replies to the kiss sender with positive message, then the recommendation is identified as being successful. There are 1,310,551 unique kisses in the selected dataset that have been sent by the 87,304 male users in this period. Among the sent kisses, 182,169 kisses are identified as being successful for which the users have received the positive responses from the partners. The dataset shows that 124,062 unique female users who may appear before or after chosen six months period are contacted by the 87,304 male users.

B. Experiment Setup

The Cluto software [9] is used to cluster the 87,304 male users into approximately 1,000 groups. Three sources of input data are used in clustering: (1) the user profiles combined with preference, (2) the user preference alone, and (3) the user profile alone.

As suggested by previous SimRank work [8], 5 iterations are sufficient enough to stabilize the score and thus in this work, 5 iterations are applied. Once the similarity amongst all users of a cluster is calculated, we test two approaches to recommend potential partners to a user *u*. In the first approach (labeled as Top-n all matched users), the system recommends to user *u* all users who were contacted by users U_{TOP}, where U_{TOP} represents the Top-*n* most similar users to *u*. In the second approach (labeled as Top-n successful matched users), the system only recommends to user *u* those users contacted by users U_{TOP} who replied positively (e.g., a successful kiss between a user in U_{TOP} and user U_r being considered for recommendation). If the user being considered for recommendation did not reply positively to a user in U_{TOP} then they are not recommended to user *u*. We can then compare the performance of these two approaches. The methods used in comparison are shown in Table I.

TABLE I. METHOD ACRONYMS

Acronym	Method
CSOS	combined profile with preference + cosine similarity + SimRank with out-links only
CSIOS	combined profile with preference + cosine similarity + SimRank with in-links and out-links
CSIS	combined profile with preference + cosine similarity + SimRank with in-links only
CDOS	combined profile with preference + distance similarity + SimRank with out-links only
CDIOS	combined profile with preference + distance similarity + SimRank with in-links and out-links
CDIS	combined profile with preference + distance similarity + SimRank with in-links only
BSR	$BSR\ (U_M)$ Baseline SR

C. Evaluation Metric

The evaluation metric for this experiment is based on deciding whether the recommended users to a given user *u* will be successful. So one of the metrics to evaluate the performance is success rate (SR). $SR\ (u_M)$ as defined in Equation 5 is to be compared with baseline success rate $BSR\ (u_M)$. $BSR\ (u_M)$ is the success rate of current online dating network without using the proposed recommendation approach. Another metric is recall which is to measure the ratio of correctly identified matches from the proposed recommendation approach to the number of matches in the dataset.

$$SR(U_M) = \frac{Number\ of\ unique\ successful\ kisses\ initiated\ by\ u_M}{Number\ of\ unique\ kisses\ initated\ by\ u_M} \quad (5)$$

$$\mathrm{Re}call(U_M) = \frac{Number\ of (Kissed\ partners \cap \mathrm{Re}commended\ partners)}{Number\ of (Kissed\ Partners)} \quad (6)$$

D. *Overall Performance*

In terms of success rate performance, recommending Top-n successful matched users is a better method than recommending the Top-n all matched users (Table II & III). Most of time, the Top-n successful matched users success rate gives double the performance over Top-n all matched users. From Table II we can see the CSIS method produces the best performance in Top-n all matched users experiment, followed by CDIS. In-link based SimRank is better performing than both in-link & out-link based and out-link based SimRank for the Top-n all matched users. The reason is that in-link based SimRank retrieves the positive kiss information when a user receives a positive kiss back from the potential partner. In & out SimRank performance is lowered by having out-link information.

In Table III, it is shown that CSIOS performs the best and achieves a success rate of 32.16% for Top-1 successful matched users. The in & out link based method outperforms in-link based only and out-link based only methods. Positive kiss information is known in this experiment when the potential partners who have returned a positive kiss are recommended. Therefore, methods containing in-link information only do not benefit.

Top-n all matched users approaches offer more potential partners for recommendation than Top-n successful matched users approaches. In terms of recall, Top-n matched users (Table IV) method outperforms Top-n successful matched users (Table V) method.

In most cases, the success rate decreases as *n* increases in Top-n (all/successful) matched users. But in some cases, the success rate increases as *n* increases. For example, for CSOS in Table III the success rate increases initially. The reason for this is that Top-1 recommendation is recommending the most similar user's contacted partners to the user. The number of contacted partners varies. The Top-1 most similar user may have a huge number of contacted partners. The chance of getting high success rate is less than those similar users who have a smaller number of contacted partners. Top-3 users, the success rate could be averaged out if one user's success rate does not perform well. Recall increases as *n* increases in Top-n (all/successful) matched users.

TABLE II. SUCCESS RATE OF TOP-N ALL MATCHED USERS

	Top-1	**Top-3**	**Top-5**	**Top-10**
CSOS	16.15%	14.4%	13.58%	12.56%
CSIOS	16.24%	13.97%	13.0%	13.0%
CSIS	22.06%	18.62%	17.27%	16.01%
CDOS	15.11%	13.36%	12.7%	11.87%
CDIOS	15.02%	12.87%	12.31%	11.73%
CDIS	19.89%	17.21%	16.16%	15.19%
BSR	13.9%			

TABLE III. SUCCESS RATE OF TOP-N SUCCESFUL MATCHED USERS

	Top-1	**Top-3**	**Top-5**	**Top-10**
CSOS	23.58%	24.07%	24.16%	23.9%
CSIOS	32.16%	27.04%	25.27%	24.14%
CSIS	31.37%	28.18%	25.87%	25.10%
CDOS	23.45%	23.85%	23.94%	23.74%
CDIOS	30.08%	25.9%	24.89%	24.03%
CDIS	29.58%	26.5%	25.4%	24.84%
BSR	13.9%			

TABLE IV. RECALL OF TOP-N ALL MATCHED USERS

	Top-1	**Top-3**	**Top-5**	**Top-10**
CSOS	3.08%	7.04%	9.23%	11.47%
CSIOS	3.89%	7.90%	9.75%	11.46%
CSIS	2.82%	5.88%	7.22%	8.53%
CDOS	1.92%	4.03%	4.98%	5.81%
CDIOS	2.35%	4.43%	5.23%	5.91%
CDIS	1.74%	3.28%	3.90%	5.26%

TABLE V. RECALL OF SUCCESSFUL MATCHED USERS

	Top-1	**Top-3**	**Top-5**	**Top-10**
CSOS	0.71%	1.11%	1.30%	1.44%
CSIOS	0.66%	1.14%	1.31%	1.42%
CSIS	0.64%	1.11%	1.27%	1.38%
CDOS	0.34%	0.56%	0.64%	0.69%
CDIOS	0.36%	0.59%	0.65%	0.70%
CDIS	0.35%	0.57%	0.64%	0.67%

VI. CONCLUSION

Online dating social networks are expanding quickly with many people joining, all requiring personalized recommendation. Traditional recommendation methods are inefficient and ineffective due to the existence of large and sparse databases and the need of handling two way matching. The proposed method clusters users into groups to reduce the computation time and complexity. The link based SimRank algorithm applied after the clustering.

The proposed method has been evaluated on an online dating network dataset. The best performing method has improved the success rate from 13.9% to 32.16%.

REFERENCES

[1] L. Brozovsk and V. Petricek. (2007, December 10th). *Recommender System for Online Dating Service.* Available: www.occamslab.com/petricek/papers/dating/brozovsky07recommender.pdf

[2] R. Burke, "Hybrid Recommender Systems: Survey and Experiments," *User Modeling and User-Adapted Interaction,* vol. 12, pp. 331-370, 2002.

[3] W. Chu and S. T. Park, "Personalized Recommendation on Dynamic Content Using Predictive Bilinear Models," presented at the WWW2009, Madrid, ACM, pp. 691-700, 2009.

[4] A. Colley, "Upload surge increasing net access costs," in *The Australian*, ed, 2007, p. 31.

[5] R. Fagin*, et al.*, "Comparing top k lists," presented at the ACM Symposium on Discrete Algorithm, Society for Industrial and Applied Mathematics, pp. 134-160, 2003.

[6] A. T. Fiore*, et al.*, "Assessing Attractiveness in Online Dating Profiles," presented at the SIGCHI, Florence, ACM, pp. 797-806, 2008.

[7] I. Guy*, et al.*, "Personalized Recommendation of Social Software Items Based on Social Relations," presented at the RecSys'09, New York, ACM, pp. 53-60, 2009.

[8] G. Jeh and J. Widom, "SimRank: A Measure of Structural-Context Simiarity," presented at the KDD'02, ACM, pp. 538-543, 2002.

[9] karypis, "Cluto: A Clutering Toolkit," Minnesota University, 2003.

[10] P. Kazienko and K. Musial, "Recommendation FrameWork for Online Social Networks," presented at the 4th Atlantic Web Intelligence Conference (AWIC'06), Washington D.C., Springer, pp.111-120, 2006.

[11] Kleihberg, "The small world phenomenon: An algorithm perspective," ACM Symposium on Theory of Computing, pp. 163-170, 2000.

[12] R. Kreider and J. Fields, "Number, timing and duration of Marriages and divorces," Household Economic Studies, pp. 70-97, US Census Bureau. Washington DC., 2002.

[13] D. Lizorkin*, et al.*, "Accuracy estimate and optimization techniques for SimRank computation," *VLDB,* vol. 19, pp. 45-66, 2010.

[14] A. Mislove*, et al.*, "A measurement and anlysis of online social network," presented at the Internet Measurement Conference, ACM, pp. 29-42, 2007.

[15] B. R. Preiss, Data structures and algorithms with object oriented design patterns in Java: John Wiley & Sons, 1998.

[16] S. Wasserman and K. Faust, *Social Network Analysis Method and Applications*: Cambridge University Press, 1994.

[17] C. Wilson*, et al.*, "User interaction in social networks and implications," presented at the EuroSys'09, Germany, ACM, 2009.

[18] W. S. Xi*, et al.*, "SimFusion: Measuring Similarity using Unified Relationship Matrix," presented at the SIGIR'05, ACM, pp. 130-137, 2005.

[19] X. Xin*, et al.*, "A Social Recommendation Framework Based on Multi-Scale Continuous Conditional Random Fields," presented at the CIKM'09, HongKong, ACM, pp. 1247-1256, 2009.

[20] S. H. Yoon*, et al.*, "Link based similarity measure for scientific literature," presented at the WWW, ACM, pp. 1213-1214, 2010.

[21] P. X. Zhao*, et al.*, "P-Rank: A Comprehensive Structural Similarity Measure over Information Networks," presented at the CIKM'09, HongKong, ACM, pp. 553-556, 2009.

[22] G. Karypis and V. Kumar, "Multilevel k-way Hypergraph Partitioning", presented at 36^{th} Design Automation Conference, ACM, pp. 343-348.

[23] J. W. Han and M. Kamber, Data Mining Concepts and Techniques: Elsevier Inc.,2006.