CrossMark

A graph-based model to improve social trust and influence for social recommendation

Gourav Bathla[1] · **Himanshu Aggarwal**[1] ·
Rinkle Rani[2]

Abstract Social big data is large scale of data due to exponential popularity of social network and social media. Researchers can use social big data and social network for their observations if they analyse those in an intelligent manner. The target of intelligent decision is to find the most credible user in social network, who has the highest influence. A very large number of users are connected in social networks, implicitly friends-of-friends or explicitly mutual friends. They are able to communicate with each other and share their likes or dislikes on different topics. If users want to analyse any topic or purchase product like movie, book, they are populated with a lot of choices. Information overload due to large number of choices available to users limits effective product selection and hence results in reduced users' satisfaction. Recommendation models are solution for providing better suggestion to users. Product's recommendation at Amazon, friend's recommendation at Facebook and music recommendation at iTunes are some of the popular examples of suggestions made on the basis of user's interests. Recommendation models ease the user by reducing search space in social network graph. The main purpose of this paper is to improve social recommendations so that better and more appropriate choices are available for users. In this paper, an efficient technique for social recommendations using hyperedge and transitive closure is proposed. Social big data is processed and analysed in the form

✉ Gourav Bathla
gouravbathla@gmail.com

Himanshu Aggarwal
himanshu.pup@gmail.com

Rinkle Rani
raggarwal@thapar.edu

[1] Department of Computer Engineering, Punjabi University, Patiala, India

[2] Department of Computer Science and Engineering, Thapar University, Patiala, India

🖄 Springer

of social graphs. User–user and user–item connections are represented in the form of matrices. We have exploited homophily so that large number of connected users have trust on each other. Our model provides better recommendation to users by leveraging increased trust. The proposed model overcomes the limitations of traditional recommender systems like sparsity, cold start. It also improves prediction accuracy. The proposed model is evaluated through different metrics like MAE, precision, recall and RMSE. Empirical analysis shows significant improvement in recommendations. We have used Mahout library for improving recommendation accuracy and also handling large volume of data. SNAP library is also used for analysis of social big graphs. The proposed recommendation model is evaluated using Epinions and FilmTrust datasets. These datasets contain user's ratings for various products in the scale of 1–5. Through analysis it is verified that the proposed model boosts the performance significantly. We have formulated recommendation model using manipulated social graph as per our proposed technique. This manipulated graph is mentioned as influence product graph (IPG) throughout this paper. IPG increases social trust value between connected users and this effect in recommending products in an effective and efficient manner.

Keywords Social big data · Social recommendation · Social trust · Social influence · Mahout · SNAP · IPG

1 Introduction

Users communicate and share information on social networking sites like Facebook, Twitter which results in large scale of data. Big data with main focus on social media, i.e. data from multiple Internet sources, is social big data. Analysis of social big data is an active research topic nowadays. Many researchers are analysing social big data, and a lot of researches are being carried out in opinion mining, machine learning, natural language processing, sentimental analysis and social recommender systems [1–4]. One of the most popular applications of social big data is recommendation. If a user wants to buy any product using social networking sites, these sites return a lot of choices, which is like noise for a common user. These choices are very difficult to analyse in order to buy a product of particular interest for user. Recommendation system provides selected options to users based on their interest, i.e. personalized recommendation [5]. Many social networking sites use social recommendation systems to provide users with choices which are of high relevance [6]. Users are connected to each other by means of trust or friends. If a user gets suggestion from social connected friends and the user finds the suggestion to be good, trust is increased between these users. Social recommendation systems use social trust to further improve the accuracy of recommendations. Traditional recommender systems assume each user an independent entity [7], whereas social recommender systems overcome this limitation, as they consider social relation amongst users [8]. Traditional recommender systems use content-based and collaborative filtering techniques for recommendation [7,9–12]. In content-based approach, users are provided with recommendations based on their past interest or behaviour. The limitations of this approach are user's privacy [9] and also assuming users as independent. In collaborative filtering, if a user provides ratings to any product or topic and other user also provide same rating to this product or topic,

there is high probability that these users also like and provide same rating to other product in future [7]. The limitation of this technique is that users are assumed as independent.

User–item matrix signifies details about rating of items by different users [13]. There are usually few entries for this matrix as the product dataset is large and user provides rating to only few products of his interest. When entries are very less, the resulting matrix is a sparse matrix due to few ratings. When users are new to recommender systems, no entries are made by him. This scenario is known as cold start, as the new user does not have any entry for his like or dislike for any product or topic. Our proposed technique IPG overcomes both of these drawbacks of traditional recommender system—sparsity and cold start [14,15].

IPG represents social network connections in the form of graph. Different users are represented as nodes in graph, and user–user matrix entries are filled corresponding to connected nodes, i.e. edges in graph. When users provide ratings to some products, user–item matrix is designed accordingly [16]. A user–user matrix has information about a user's trust on another user. IPG improves social trust density, i.e. number of users who trust each other. Improving social trust directly improves recommendations to any user. Experimental section verifies this by using standard evaluation measures like MAE and RMSE.

Social recommendation systems consist of two types of systems: (1) systems based only on user–item interaction and (2) systems based only on user–user interaction, i.e. users' trust [17,18]. The limitation of 'only user–item-based' systems is not considering social relation between users. Drawbacks of 'only trust-based' model are (i) trust matrix is sparse, as there are very few entries to analyse and (ii) user–item ratings matrix is missing. Information about user preferences is scattered in user–user as well as user–item interactions. IPG considers trust as well as item ratings. Previous techniques which are investigated by many researchers revealed that only immediate connected users have influence on a user only [2,19–21]. In IPG, the social network graph uses hyperedge and transitive closure methodology to connect a user to friends-of-friends also.

2 Related work

Several research works have been carried out in analysing social big data specifically for social recommendation. Social big data is of deep interest of researchers and is an active research area. In [1] recent research work is focused on social big data analysis. This work demonstrates that combination of big data technologies and machine learning algorithms provides solution for social big data analysis. Social big data analysis is interdisciplinary area analysis such as information retrieval, data mining, machine learning, big data computing and natural language processing. In this work, network analysis based on graph theory, text analysis and social application like marketing with the use of big data technologies are implemented with examples. In the research work by [9], content-based and collaborative filtering recommendation techniques are differentiated. Hybrid recommender system for digital library (FAB) is proposed in this paper. Data mining techniques specifically for recommender systems

are explained by [22]. Many techniques of data mining like clustering and classification are used to improve recommender systems in this work. Dimensionality reduction and sampling are used to improve recommender system as they reduce search space for user. Authors have used association rules to train data for experimental purpose. In the research by [19] triple (sender, receiver, item) social recommendation is compared with pair (user, item). The use of triple improves recommendation because of information available about sender and receiver interest. The factors that are used for social activity are receiver interest, item quality and interpersonal influence. In [2] topical affinity propagation (TAP) to model social influence based on topic is proposed. In this work, authors have used node-specific and network structures for social influence. Scaling improvement is also demonstrated with the use of map reduce technique. Data sparsity problem is alleviated using social contextual in [20]. Many users do not rate any product; entries are not filled in user–item matrix [23]. The entries in user–item matrix are filled by using social contextual information. Probabilistic matrix factorization is used for fusing user–item matrix. Social trust is used for entries in user–item matrix. Metrics used for experiment analysis in this research work are MAE (mean absolute error) and RMSE (root-mean-square error). In [24] social recommender system is improved by using similarity, trust and reputation. Metrics for each of these parameters are calculated and shown to perform better. If a user i trusts user j, it does not mean that user j also trusts i. This assumption is used by authors to modify trust matrix. Similarity value, trust value and reputation value are included in recommender algorithm. In [17] matrix factorization is used for improving user–user matrix. N-dimensional space is analysed for users who have highest social trust. Large-scale dataset can be used by using the approach proposed (SoRec) in this paper. Comparison of proposed approach with state-of-the-art approaches proves that significant improvements are reflected to users as recommendation. Complexity analyses prove that this approach scales linearly with the increase in number of dataset. Bedi et al. [18] used ontologies for recommending product to peer agents. In this approach, trust network is used for generating recommendations. Tourist recommender system is used in this paper for checking the validity of proposed approach. Factorization of graph is visualized by the use of bipartite graph in [25]. TrustSVD is proposed in [5] which used trust-based matrix factorization technique. Multiple types of information are collected and integrated in recommendation model to overcome data sparsity and cold start. In this paper, four datasets are used for experiment purpose and proved that proposed technique performs better than state-of-the-art approaches. Yin et al. [26] proposed LinkRec framework to find the link relevance of connected friends on social network. The global as well as local influence of attributes is used in this framework. Link relevance is calculated by the use of assigning different weighting criteria for nodes. IMDB and DBLP datasets are used in this paper and shown to outperform state-of-the-art link recommendation frameworks. Roy and Ravindran [27] defined hypergraph as a graph which has hyperedge between multiple vertices. In this paper, also importance of hyperedge in co-citation and co-author in social graph is highlighted.

SoRec [17] and SoReg [28] are recommendation models that are based on trust. SocialMF [29] and TrustMF [30] are based on matrix factorization approach which

Table 1 Comparison of various existing and proposed technique

Research work	Cold start	Sparsity	Scalability	Social trust	Prediction accuracy
Tang [2]	×	×	✓	✓	✓
Ma [20]	×	✓	✓	×	✓
Neto [32]	✓	×	×	×	✓
Than [24]	×	✓	✓	✓	✓
Hu [17]	✓	×	×	✓	✓
Guo [5]	✓	✓	×	×	✓
Proposed technique (IPG)	✓	✓	✓	✓	✓

also uses social trust and influence of social network. TrustSVD [31] regularizes recommendation model with user trust and item ratings.

Table 1 represents the feature-wise comparison of existing recommendation techniques and proposed technique IPG. In the existing techniques researchers have tried to improve social recommendation by the use of matrix factorization or social regularization or improving trust and item rating matrices. Standard measures like cold start, sparsity, social trust and prediction accuracy are used as basis for comparison. It is clearly demonstrated in Table 1 that none of the techniques are able to perform well for all measures. Some techniques are able to solve cold start problem, and others are able to improve social trust, but not all measures are solved in one technique. It is mentioned in table that our proposed technique IPG is able to successfully deal with the problems of cold start and sparsity and improves social trust and prediction accuracy.

3 Social recommendation using social influence

Recommendation provides better suggestion for any product to individual user or group of users. Traditional recommender systems exploit interest networks (user–item ratings). Content-based technique and collaborative filtering are the techniques of traditional recommender model integrating only interest networks. Content-based approach is used for analysing user behaviour and history. Products which were liked by a user in the past are analysed, and based on this analysis, other products are recommended to that user. The limitation of content-based technique is violation of user privacy. Collaborative filtering technique is based on user–item similarity. Users having similar taste for any product are likely to prefer another product of similar taste. When a user provides rating to any product, similarity is calculated to those other users who have rated that product with the same rating. User–user matrix and user–item matrix entries are filled based on same ratings for a product by a pair of users. User–User similarity is calculated by analysing rating provided by users. Similarity between a pair of users is calculated by Pearson correlation coefficient [33].

$$\text{pcc}(U_a, U_b) = \frac{\sum_{i=1}^{n} (R(a)_i - \overline{R(a)})(R(b)_i - \overline{R(b)})}{\sqrt{\sum_{i=1}^{n} (R(a)_i - \overline{R(a)})^2 \sum_{i=1}^{n} (R(b)_i - \overline{R(b)})^2}} \tag{1}$$
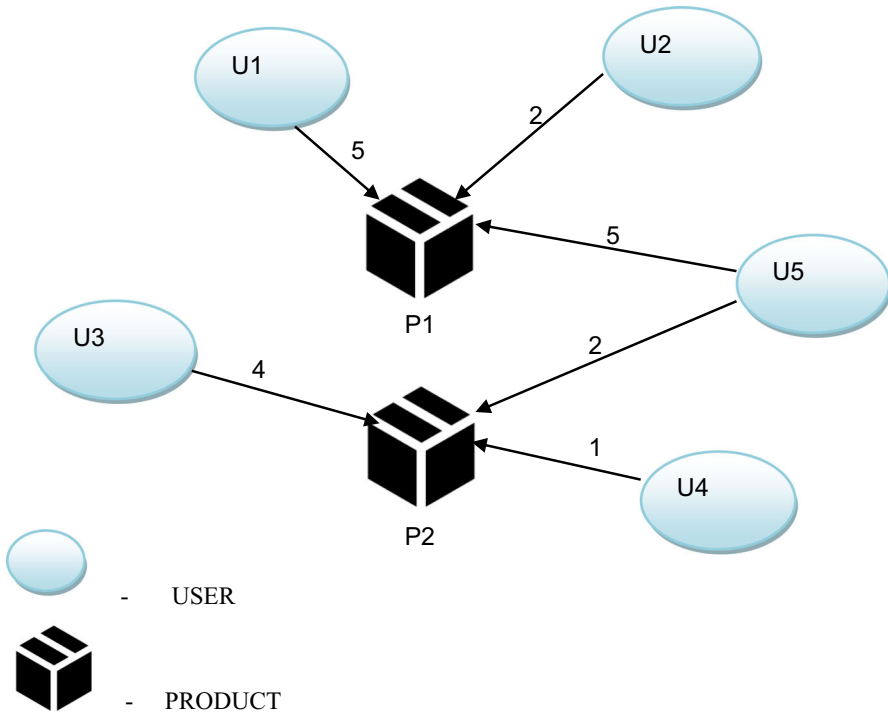
Fig. 1 User product rating

where $R(a)_i$ is rating provided by user a for product i and $\overline{R(a)}$ is average of ratings provided by user a for products, $R(b)_i$ is rating provided by User b for product i and $\overline{R(b)}$ is average of ratings provided by user b for products. Similarity is calculated for products $1\ldots n$ between users U_a and U_b. The value of correlation is -1 to 1.

The main limitations of collaborative filtering are sparsity and cold start. Various researchers have proposed improvements in collaborative filtering to improve recommendation. Traditional recommender systems assume that users are independent, which is main limitation in providing better suggestions to user.

Figure 1 explains user–item rating in collaborative filtering. Same ratings for any product from users *x* and *y* can conclude that these two users have same liking on that product. Thus, in future there is high probability that these two users will have same liking on some other product also. Users provide ratings to products based on their experience. We have used ratings in the range 1–5 for this example. User 1 rates product P1 with 5, and also User 5 rates P1 with 5. These users are similar according to collaborative filtering technique, as they have provided same product with same rating. It can be observed that these users are independent; no user is socially connected to other user; hence, there is no social influence.

Content-based and collaborative filtering-based approaches used only user–item ratings for recommendation. These techniques do not consider social relation of users. Social recommendation system gives suggestions based on user's connections on social

networking sites. Several definitions are being articulated by researchers for social recommendation [2, 19, 30, 32]. Users are connected to each other using social graph by social friendship or social trust [28]. When a user u_i likes the product or topic which is liked by other user u_j, then there is increase in user u_i trust on u_i. Social trust is unidirectional as u_i trust on u_j does not indicate that u_j also trusts u_i. On the contrary, in social friendship u_i and u_j are having same liking and it is bidirectional. This trust results in social influence of one connected user to another connected user. Social influence can be calculated by the use of social trust as well as ratings to a particular product.

Social trust is pivotal for predicting ratings for a product. According to [5] Social trust and ratings are correlated with each other. Interest networks (user–item interactions) and friendship networks (user–user connection) are highly correlated and also helps mutually for finding better user–item preference and user–user interactions [34] . In this manner their social trust improves, and their influence on each other also improves. Many research works improved social recommendation by the use of matrix factorization [17, 24, 35] and regularization [5, 35]. Some other studies integrate social trust in recommendation model by using link prediction. In these research works, random walk is used on social graph to calculate the similarity between vertex or users [34, 36]. Transition probability is calculated by using Eq. 2. It signifies similarity between user u_j and user u_i.

$$P\left(j|i\right) = w_{ij}/d_i \tag{2}$$

where w_{ij} is weight between connected edges between u_i and u_j. Here, d_i is degree of vertex i. The limitation of random walk is not consideration of unidirectional property of social trust. To the best of our knowledge none of research studies have improved trust and influence both by the use of social graph, transitive closure and unidirectional property of social trust.

4 Social big data

Structured, unstructured and semistructured data are different varieties of big data [5]. Simple data which are organized like row–column data are structured, log Internet files are semistructured, and images are examples of semistructured data. Big data is defined by 3 Vs model—volume, velocity and variety. It is a collection of high volume—large scale of data collected daily from different sources, high velocity—data collection speed is very high due to frequent sharing of data, i.e. streaming data, high variety—different types of data from social networking sites, blogs, networks, etc. Different varieties of data are collected under large scale of data [37]. Users share a lot of data on social networking sites like Twitter in the form of social data [38]. Combination of social network data and big data is termed as social big data. It is clear from Fig. 2 that social big data is combination of social data, like Facebook, Twitter; big data (large volume of data); and technologies for processing big data like Hadoop, Mahout etc.
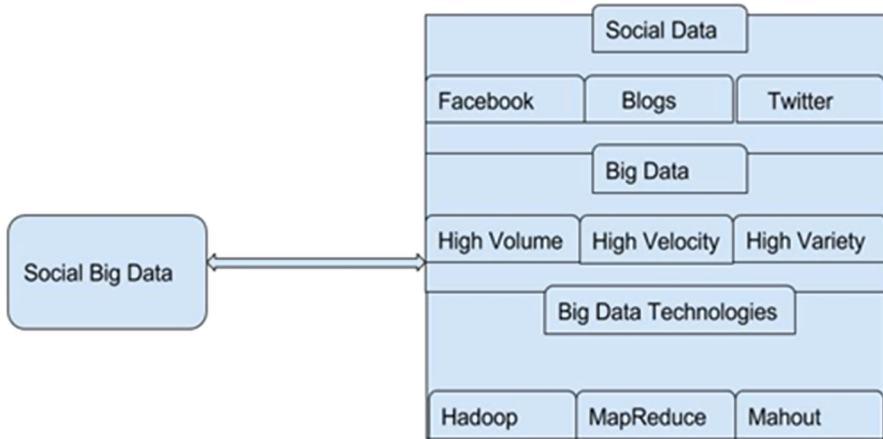
Fig. 2 Social big data

Various research works are being carried out for social big data like data processing, user behaviour analysis, social recommendation, opinion mining and sentiment analysis [1]. Social media is classified into various categories like blogging, social search, crowd sourcing, social gaming and collaboration [8]. Social big data is collection of heterogeneous data, but data related to social media are main point of interest for a better recommendation.

This large scale of social data is represented and analysed using formal methods and tools which are largely centric towards graph theory [39–41]. Social big data can also be analysed using formal methods based on fuzzy set [42]. Users are interconnected through sharing of information; these sharing can be represented as edges. Various latest technologies are used for social big data analytics like GraphLab, SNAP, Hadoop, MapReduce and Mahout Library. This analysis is used for social recommendation, social influence and user behaviour [8].

5 Proposed technique

Recommender system works on set of users U and set of items A. Recommendation is provided to target users based on user–item ratings.

$$f : U \times A \rightarrow R \qquad (3)$$

User U_i rates item A_i, and other user U_j rates item A_i. Based on this rating, recommendation R_i is provided to user U_i and R_j is provided to user U_j. In recommender system, it is assumed that if users have rated same item on the same scale, there is higher probability that same set of users will like other item A_k for any random k value.

Traditional recommendation systems focus only on user–item rating. Users are assumed as independent, and no trust between users is considered. Other kind of

recommendation systems works only for trust-based model [17,18]. Many research studies have proved that interest network (user–item interactions) and social trust (user–user interactions) are correlated [5,34]. In our work, we have incorporated both of these features in our model to improve prediction accuracy. Social influence is computed using the concept of hyperedge of proposed social graph. A hyperedge connects multiple adjacent vertices in a social graph [27]. In social network graph, every node is influenced by other node in some sense. Social influence changes the decision-making of a user due to effect of immediate neighbours [29]. These nodes or users can provide recommendation for a friend or product based on this influence. In previous studies on social influence, only immediate node is analysed for providing recommendation to connected nodes. In our technique, nodes are also influenced by multiple connected nodes where these multiple users are connected through hyperedge and transitive closure. Transitive closure of a graph is the binary relation on a graph to check the reachability of a node. A relation R on x, y and z is transitive; if $x R y$ and $y R z$, then there must exist $x R z$. This can be represented in the form of a matrix which contains information about the number of hops to reach from x to z. In this paper, Warshall algorithm is used for finding the transitive closure of social network graph.

Input : Set of users, Set of items and user-user, user-item matrix

Output : Improved social trust and recommendation

Algorithm IPG (U_i , I_j , M_{ij})

1. Initialize social trust relations by using user-user matrix and interest network by user-item matrix.

2. Check social trust relationship.

> **Let M ={m_{ij}} is a social trust matrix of users u_i and u_j.**

> **Value of m_{ij} is w_{ij}**

> **(weight or trust between users u_i and u_j) if there is trust between users otherwise m_{ij} is 0.**

$$m_{ij} = \begin{cases} w_{ij} \ if \ u_i \xrightarrow{trust} u_j \\ \\ 0 \ otherwise \end{cases}$$

3. Create directed edge graph G= {V,E} from matrix m_{ij}

> **where V is set of vertices and E is set of edges.**

> **Vertex v_i is connected to v_j by edge e_{ij} having weight w_{ij}.**

4. Calculate MAE

$$M_{ae} = \sum_{i=0}^{n} \big(P(u,i) - p(u,i)\big)/n$$

> **and F Measure**

$$F_{m=} \frac{2. \ precision. \ recall}{precision + recall}$$

5. Calculate transitive closure of modified user-item matrix

$$M_m = M^1 \cup M^2 \ … \cup M^n$$

6. Manipulate graph with the use of hyperedge theory proposed for user- item recommendation.

> **Modified graph G_m = (V_m, E_m) use modified social trust matrixM_m.**

7. Using SNAP library, create graph Gm of proposed technique

8. Calculate modified MAE M_{aem} , Precision P_{rm} and Recall R_{em} and F Measure F_{mm} of proposed technique

9. Compare existing technique metrics with proposed technique metrics

10. Exit

The above algorithm uses social graph G_r which represents connections of users on a social network. Social trust matrix is represented for graph G_r with edge value 1 for social connected users and value 0 for not connected users. Mean absolute error (M_{ae}), precision (P_r) and F measure (F_m) are calculated for graph G_r. Transitive closure is calculated with the use of hyperedge theory, and as a result modified graph G_m is designed. Again MAE (M_{aem}), precision (P_m) and F measure (F_m) for the modified graph G_m are calculated.

$$M = U^{m*n} \tag{4}$$

User–user matrix M is populated with entries of user U_m trust value for U_n. We have used binary trust values, i.e. 0 or 1. Using proposed technique with hyperedge, matrix entries are improved with the use of transitive closure. M^1 is matrix with single connected edges, M^2 is matrix with 2 edges between nodes, and hence, M^n is matrix with n edges between nodes. Union of these matrices is connected nodes with n edges.

$$M = M^1 \cup M^2 \cdots \cup M^n \tag{5}$$

Using Eq. 5, data sparsity is removed by using entries of friends-of-friends. User A is connected to user B, i.e. A → B, and user B is connected to user C, i.e. B → C. Using hyperedge, it is concluded that A → C. If all transitive closure values are considered, then recommendation accuracy is degraded. In this work, only values of those users are considered who have strong influence on the target node. Our assumption is that more close neighbour nodes have more influence than other nodes. By incorporating this assumption in our experimental work, only nodes connected with 2 edges are considered as strongly influenced nodes.

There can be large scale of values for which user has no trust on other users. This results in data sparsity. In this paper, improvement for social recommendation is given by factor $f(m)$. $f(m)$ value is 1 if user U_m is connected to user U_n by single edge or through transitive closure, and 0 when there is no trust.

$$M = \sum_{i=1}^{m} f(m) * U^{m*n} \tag{6}$$

Connections and interactions between users on social networks are demonstrated by considering small set of users in Fig. 3. Eight users are connected by directed edges. Single edge between two nodes indicates that users are trusted friends and directly connected. Directed edges are drawn for users who are trusted friends. Trust is not symmetric, i.e. if a user A trusts user B, it is not necessary that user B also trusts user A. We have used homophily feature which is connection of users with same attributes and used for relevant link prediction [26].

The limitation of social graph is that nodes connected through single edge, i.e. direct neighbour nodes, are considered as trusted friends only. In our proposed technique IPG, nodes which are even connected through more than single edge can be given the privilege of trusted friends. Node 7 is connected to node 3 by direct single edge. Node 7 is having influence of node 3, i.e. trusted by node 3. Node 2 trusts node 3, and node 3 have social trust on node 7. By using transitive property of graph, i.e. 2 → 3 → 7

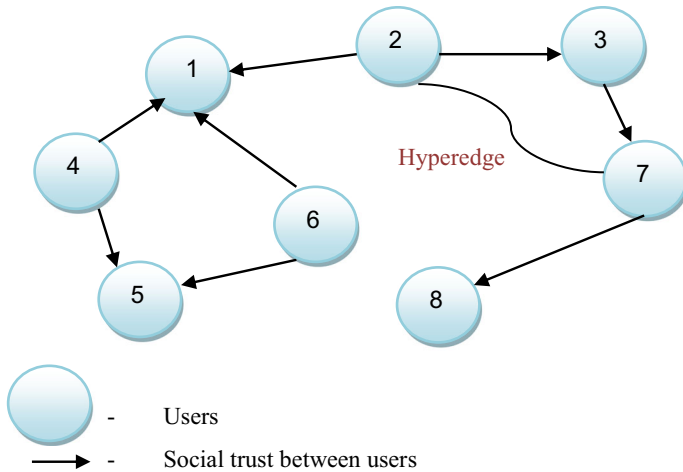**Fig. 3** Social trust improvement using hyperedge for users

Table 2 User–user matrix using traditional recommendation

Users	u1	u2	u3	u4	u5	u6	u7	u8
u1								
u2	1		1					
u3							1	
u4	1				1			
u5								
u6	1				1			
u7								1
u8								

implies 2→7, node 7 has gained social trust and influence from node 2. From Fig. 3, it is clear that node 7 is now provided with better recommendation as it has gained trust from nodes 2 and 3 directly.

Table 2 demonstrates trust between direct neighbour nodes. Node 2 is directly connected by single edge to node 3. So, trust from node 2 to node 3 is set to 1 as we have considered only binary trust values, i.e. 0 or 1.

Table 3 shows improved trust by using IPG as now node 2 trusts nodes 3 and 7. Recommendations from new trusted friends will help node 2 for making better decision for any product or topic. It is also proved by this example that sparsity and cold start problems are eliminated. This small improvement on only 8 users' connection scenario will be clearly visible in Experimental section where we have deployed IPG on Epinions and FilmTrust dataset. It is clear from modified matrix that social trust is improved between users. Many research works have used social trust already built between users. We have improved social trust significantly, and in conclusion, rating prediction is improved.

Table 3 User–user matrix using proposed technique

Users	u1	u2	u3	u4	u5	u6	u7	u8
u1								
u2			1				1	
u3							1	1
u4	1				1			
u5								
u6					1			
u7								1
u8								

$$r_{i,j} = \frac{\sum_{k=0}^{n} r_{k,j}}{n} \qquad (7)$$

Equation 7 calculates ratings for user i for item j. $r_{k,j}$ is ratings of trusted user k for item j, and n is total numbers of trusted users for user i. Predicted rating for user i is average of all trusted users' ratings. Prediction accuracy is improved by our technique as we are only considering strongly connected and trusted users. This is verified from Experiment section that prediction accuracy is better as compared to state-of-the-art recommendation models.

The drawbacks of previous studies for recommending products are sparsity and cold start. This proposed technique overcomes these drawbacks of previous studies. Sparsity is when user–user matrix entries are very less. This is due to the fact that many users do not rate products or participate in sharing likes and ideas. This results in sparse user–user and user–item matrix. In our proposed technique, users are connected as graph hyperedge. If a user does not provide rating to products or share ideas, then user entry in user–user matrix and user–item matrix can be filled by using users' connections on social network graph. If any user is new in recommender system, entries in user–user and user–item matrix can be referred from trusted friends rating. This helps in alleviating cold start which is main limitations of traditional recommender system. Precision and recall are also improved significantly as compared to existing techniques which is explained in next section. In Experiment section, 2 datasets (FilmTrust and Epinions) are used which are having data with different scale. The reason for implementing our approach on these 2 datasets is that when IPG recommendation accuracy is proved on FilmTrust, then it will also be proved for large-scale dataset, i.e. Epinions dataset. Improved scalability is proved in Experiment section.

6 Performance evaluation

Standard evaluation measures, which are popularly used for observing recommendation systems accuracy, are MAE, RMSE, F measure, precision and recall. Two most common evaluation measures—MAE and RMSE, are explained in following subsections:

6.1 Mean absolute error

This metric is used by many research works for evaluation of recommender system [43,44]. Good recommender system should reduce this error as much as possible. Experiments prove that our proposed technique reduces error significantly and provides good accuracy.

$$\text{MAE} = \sum_{i=0}^{n} (P(u,i) - p(u,i)) / n \tag{8}$$

Mean absolute error is calculated by the average of difference between $P(u,i)$ which is prediction by recommender system and $p(u,i)$ which is actual prediction for n products. Bell et al. [45] explained that even small improvement in value of MAE improves quality of social recommendation significantly. Proposed technique IPG is compared with state-of-the-art approaches for computing MAE.

6.2 RMSE

Root-mean-square error metric is used for checking the recommendation accuracy. Several research studies have pointed out that RMSE is more informative than MAE in terms of signifying accuracy of recommendation.

$$\text{RMSE} = \sqrt{\sum_{i=0}^{n} (P(u,i) - p(u,i))^2 / n} \tag{9}$$

From Eq. 8, it is clear that square of error is calculated. More difference between RMSE and MAE signifies degraded prediction accuracy results. RMSE is calculated by the square root of average of square of difference between $P(u,i)$ which is prediction by recommender system and $p(u,i)$ which is actual prediction for n products .

In Experiment Setup and Findings section, only RMSE is used for proving better recommendation accuracy because results by using RMSE metrics signified better prediction errors.

6.3 Dataset description

We have used Epinions and FilmTrust dataset for validating our proposed technique IPG. Epinions.com is a site that maintains users' reviews and also trust and distrust between users. Users register on this site to give feedback and review for any product and also provide trust information for other users explicitly. This dataset forms Web of Trust, i.e. who trusts who. So, this dataset is standard set for evaluation of social recommendation. In this dataset, 40,163 nodes and 6,64,824 edges are used for assigning values for social trust. This dataset has sparse matrix because there are very few users who have social trust relationship. The details of Epinions dataset are given in Table 4.

Table 4 Epinions social trust dataset statistics

Dataset statistics	
Users	40,163
Items	1,39,738
Ratings	6,64,824
Density	0.051%
Social trust statistics	
Trusters	33,960
Trustees	49,288
Trusts	4,87,183
Density	0.029%

Table 5 FilmTrust social trust dataset statistics

Dataset statistics	
Users	1508
Items	2071
Ratings	35,497
Density	1.14%
Social trust statistics	
Trusters	609
Trustees	732
Trusts	1853
Density	0.42%

This can be observed from social trust statistics that density of users trust is 0.029% which is very less. By using IPG this trust is improved to 0.12% which is significant improvement for providing better recommendation to users.

In our proposed technique IPG, FilmTrust dataset is also used for checking the accuracy of proposed work. In FilmTrust dataset, users rate products in the scale of 1–5. List of co-purchased products are also summarized in this dataset. For each product title, sales rank, list of similar products and product reviews are mentioned. The details of FilmTrust dataset are given in Table 5.

It can be observed from Table 5 that density of social trust is 0.42% which is very less. Our proposed technique IPG improves this trust of connected users from 0.42 to 0.73%. When social trust is improved, it means there are more entries in user–user matrix and user–item matrix. More social trust of users provides better recommendation based on product ratings.

6.4 Experiment setup and findings

SNAP library is used for analysis of social network represented in the form of graphs. We have modified user–user matrix by using hyperedge and transitive closure on SNAP. Fivefold is used for experimental analysis. Mahout library is also used for
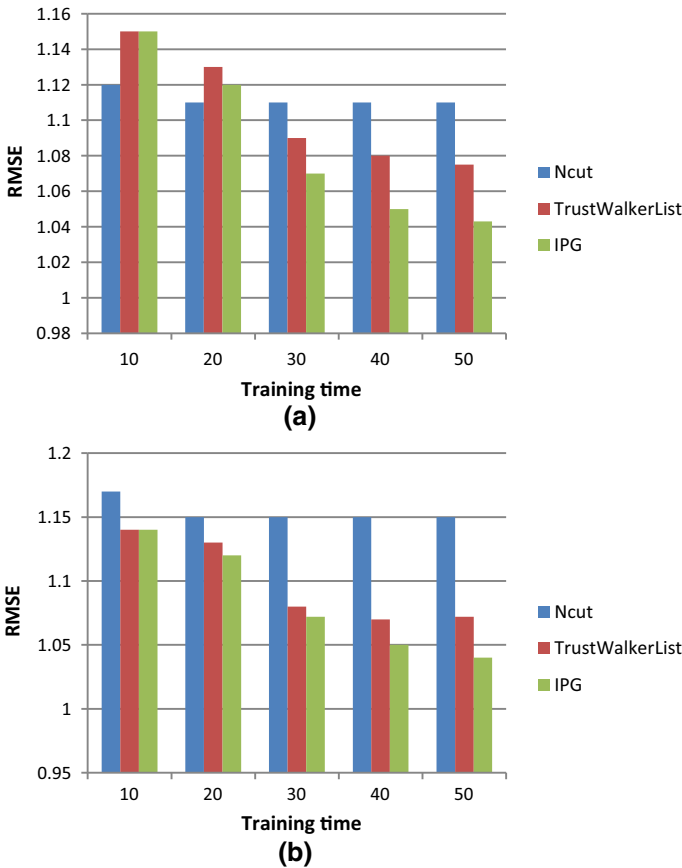
Fig. 4 **a** RMSE for Epinions dataset. **b** RMSE for FilmTrust dataset

implementing our approach on large scale of data. Social trust is analysed between immediate neighbours, and using dynamic update of trust based on IPG, we have trained dataset for improving trust.

Different recommendation models are compared with proposed IPG technique. TrustWalkerList finds top *n* recommendations by using random walk on graph of trust network [46]. Normalized cut for neighbour selection is based on collaborative filtering, and it predicts ratings with graph partitioning [47]. We have selected in random 40% data from Epinions as well as from FilmTrust dataset. Out of this selected data, 80% is used for training and 20% for testing purpose.

RMSE of IPG is compared with TrustWalkerList and NCut recommendation models. Initial trust between users is improved by IPG after training the data for a certain amount of time. When training time is increased, direct trust and indirect trust between users build more strongly and user–user matrix entries are improved significantly by using IPG. Comparison is on the basis of increasing training times as shown in Fig. 4a, b.
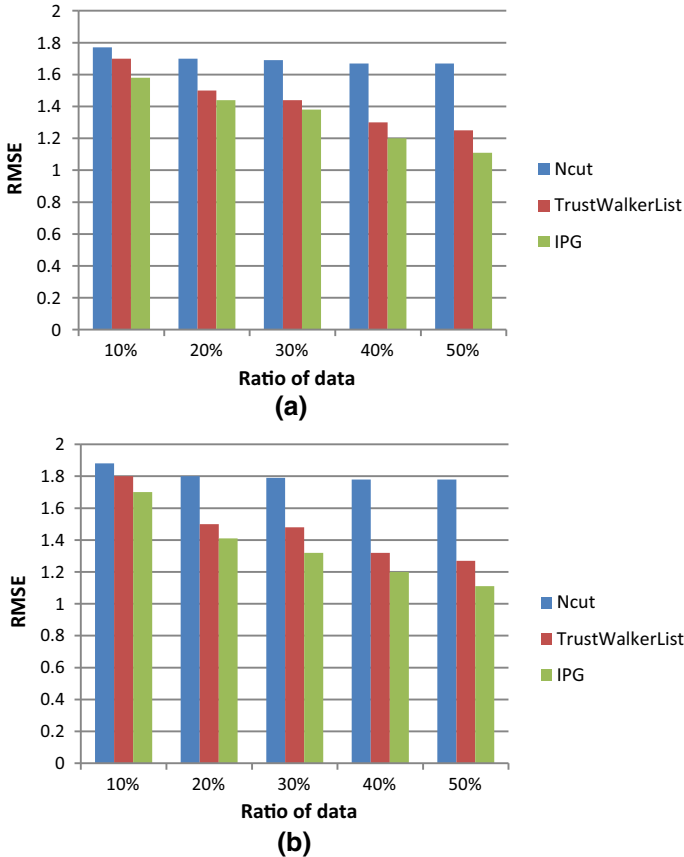
Fig. 5 **a** RMSE for Epinions sparse dataset. **b** RMSE for FilmTrust sparse dataset

It is clear from Fig. 4a, b that during initial training, i.e. when time is 10–20, there is no much difference in performance of NCut, TrustWalkerList and IPG. This is due to non-improvement in trust in such short time, but gradually trust between users' increases and performance is improved for trust-based models, i.e. TrustWalkerList and IPG. Limitation of TrustWalkerList is random walk for trusted users. The focus of IPG is only on users who have strong trust and influence. RMSE is improved in conclusion of this, as demonstrated in Fig. 4a, b.

Figure 5a, b demonstrates the effect of sparse data on recommendation accuracy. Our technique is advantageous for sparse data as trust can be built between friends-of-friends by the use of transitive closure. IPG performs well when ratio of data is 40–50% as can be visible in Fig. 5a, b.

6.5 Practical implications of experiment

Empirical results prove that density of user trust on other user (friends-of-friends) improves significantly. If trust improves, it directly provides better recommendation

to users as every user has high probability of liking items from trusted friends. Density of trust improves from 0.029 to 0.12% in Epinions dataset and from 0.42 to 0.73% in FilmTrust dataset. It is also validated from experimental analysis that MAE and RMSE are improved as compared to other recommendation models. From this comprehensive analysis, it is concluded that even smaller improvement in trust results in significant improvement in the accuracy of recommendation.

7 Conclusion and future directions

Our proposed technique improves social recommendation accuracy by using hyper-edge theory of social graph. We have represented user– user connections by directional social graph. Traditional recommender systems use content-based and collaborative filtering techniques. In traditional recommendation system, it is assumed that users are independent. But, the main limitations in using this approach are sparsity and cold start. Our proposed technique IPG overcomes these limitations. Interest networks and trust networks are represented using user–item matrix and user–user matrix, respectively. By increasing trust between users, more numbers of entries are filled in user–user and user–item matrix. This results in eliminating sparsity and cold start problem. Using social influence and trust, more recommendations are provided to the user in an effective manner. We have used Mahout library for large scale of data, hence improving recommendation accuracy. In our proposed technique IPG, SNAP library is used for analysis of social network graph. Epinions and FilmTrust real datasets are used for the experimental purpose. IPG is compared with state-of-the-art approaches. Through experiments it is proved that the proposed technique outperforms state-of-the-art recommender systems. Accuracy of proposed social recommender system is measured by metrics MAE and RMSE. It is proved that RMSE values are comparatively better even for large scale of data. In future, the concepts of social tagging and social contextual information will be used in IPG for enhancing the prediction accuracy. Social tags appear attractive as they will improve the social trust between users by providing extra information in the form of tags. Further improvements in this technique are possible if we assign weights to each node, which can be considered for computing social influence and trust of that particular node. Improved threshold can also be set for social trust, in order to consider nodes having social trust value above threshold only.

References

1. Bello-Orgaz G, Jung J, Camacho D (2016) Social big data: recent achievements and new challenges. Inf Fusion 28:45–59
2. Tang J, Sun J, Wang C, Yang Z (2009) Social influence analysis in large-scale networks. In: Proceedings of Knowledge Discovery and Data Mining, pp 807–816
3. Huang Z, Chung W, Chen H (2004) A graph model for E-commerce recommender systems. J Am Soc Inf Sci Technol 55(3):259–274
4. Pan W, Zhong H, Xu C, Ming Z (2015) Adaptive Bayesian personalized ranking for heterogeneous implicit feedbacks. Knowl Based Syst 73:173–180
5. Guo G, Zhang J, Smith N (2016) A novel recommendation model regularized with user trust and item ratings. IEEE Trans Knowl Data Eng 28(7):1607–1620

6. Leskovec J, Singh A, Kleinberg J (2006) Patterns of Influence in a Recommendation Network. In: Ng WK, Kitsuregawa M, Li J, Chang K (eds) Advances in knowledge discovery and data mining. PAKDD 2006. Lecture notes in Computer Science, vol 3918. Springer, Berlin, Heidelberg

7. Su X, Khoshgoftaar T (2009) A survey of collaborative filtering techniques. Adv Artif Intell 2009:421425

8. Ishikawa H (2015) Social big data mining. CRC Press, Boca Raton

9. Balabonovic M, Shoham Y (1997) Fab : content-based, collaborative recommendation. Commun ACM 40(3):66–72

10. Resnick P, Varian H (1997) Recommender systems. Commun ACM 40(3):56–58

11. Koren Y (2010) Factor in the neighbors: scalable and accurate collaborative filtering. ACM Trans Knowl Discov Data 4(1):1

12. Linden G, Smith B, York J (2003) Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Comput 7(1):76–80

13. Deshpande M, Karypis G (2004) Item-based top-N recommendation. ACM Trans Inf Syst 22(1):143–177

14. Jiang M, Cui P, Chen X, Wang F, Zhu W, Yang S (2015) Social recommendation with cross-domain transferable knowledge. IEEE Trans Knowl Data Eng 27(11):3084–3097

15. Chen H, Li X, Huang Z (2005) Link prediction approach to collaborative filtering. In: Proceedings of 5th ACM/IEEE-CS Joint conference on Digital Libraries, pp 141–142

16. Herlocker J, Konstan L, Terveen L, Riedl J (2004) Evaluating collaborative filtering recommender systems. ACM Trans Inf Syst 22(1):5–53

17. Ma H, Yang H, Lyu M, King I (2008) Sorec: social recommendation using probabilistic matrix factorization. In: Proceedings of the 17th ACM Conference on Information and Knowledge Management, pp 931–940

18. Bedi P, Kaur H, Marwaha S (2007) Trust based recommender system for semantic web. In: Proceedings of International Joint Conference on Artificial Intelligence, pp 2677–2682

19. Huang J, Cheng X, Guo J, Shen H, Yang K (2010) Social recommendation with interpersonal influence. In: Proceedings of European Conference on Artificial Intelligence. pp 601–606

20. Ma H, Zhou T, Lyu M, King I (2011) Improving recommender systems by incorporating social contextual information. ACM Trans Inf Syst 29(2):9

21. Guy I, Zwerdling N, Carmel D, Ronen I, Uziel E, Yogev S, Ofek-Koifman S (2009) Personalized recommendation of social software items based on social relations. In: Proceedings of Third ACM Conference on Recommender Systems, pp 53–60

22. Amatriain X, Jaimes A, Oliver N, Pujol J (2011) Data mining methods for recommender system . Recommender Systems Handbook Springer pp 39–71

23. Ma H, King I, Lyu M (2007) Effective missing data prediction for collaborative filtering. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp 39–46

24. Than C, Han S (2014) Improving recommender systems by incorporating similarity, trust and reputation. J Internet Serv Inf Sec 4(1):64–76

25. Kschischang F, Frey B, Loeliger H (2001) Factor graphs and the sum-product algorithm. IEEE Trans Inf Theory 47(2):498–519

26. Yin Z, Gupta M, Weninger T, Han J (2010) LINKREC: a unified framework for link recommendation with user attributes and graph structure. In: Proceedings of 19th International Conference on World Wide Web, ACM, pp 1211–1212

27. Roy S, Ravindran B (2015) Measuring network centrality using hypergraphs. In: Proceedings of the Second ACM IKDD Conference on Data Sciences, pp 59–68

28. Ma H, Zhou D, Liu C, Lyu M, King I (2011) Recommender systems with social regularization. In: Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM), pp 287–296

29. Jamali M, Ester M (2010) A matrix factorization technique with trust propagation for recommendation in social networks. In: Proceedings of the ACM Conference on Recommender Systems(RecSys), pp 135–142

30. Yang B, Lei Y, Liu D, Liu J (2013) Social collaborative filtering by trust. In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence, pp 2747–2753

31. Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans Knowl Data Eng 17(6):734–749

32. Mello Neto W, Nowe A (2012) Insights on social recommender system. In: Beyond RMSE ACM RecSys, International Workshop on Recommendation Utility Evaluation
33. Resnic P, Iacovou N, Suchak M, Bergstrom P, Riedl J (1994) Grouplens: an open architecture for collaborative filering of netnews. In: Proceedings of ACM Conference on Computer Supported Cooperative Work
34. Yang S, Long B, Smola A, Sadagopan N, Zheng Z, Zha H (2011) Like like alike: joint friendship and interest propagation in social networks. In: Proceedings on International Conference on World Wide Web, ACM, pp 537–546
35. Yuan Q, Chen L, Zhao S (2011) Factorization vs regularization: fusing heterogeneous social relationships in top-n recommendation. In: Proceedings of the 5th ACM Conference on Recommender Systems, pp 245–252
36. Jamali M, Ester M (2009) TrustWalker: a random walk model for combining trust-based and item-based recommendations. In: Proceedings of Knowledge Discovery Data Mining, pp 397–406
37. Wu X, Zhu X, Wu G, Ding W (2014) Data mining with big data. IEEE Trans Knowl Data Eng 26(1):97–107
38. Lin J, Ryaboy D (2013) Scaling big data mining infrastructure: the twitter experience. Proc SIGKDD 14(2):6–19
39. Elser B, Montresor A (2013) An evaluation study of big data frameworks for graph processing. In: Proceedings of IEEE International Conference on Big Data, pp 60–67
40. Seo S, Yoon EJ, Kim J, Jin S, Kim JS, Maeng S (2010) Hama: an efficient matrix computation with the mapreduce framework. In: Proceedings of the Second International Conference of Cloud Computing Technology and Science IEEE, pp 721–726
41. Wang G, Shen Y, Ouyang M (2008) A vector partitioning approach to detecting community structure in complex networks. Comput Math Appl 55(12):2746–2752
42. Mukkamala R, Hussain A, Vatrapu R (2014) Towards a set theoretical approach to big data analytics. In: Proceedings of IEEE International Congress on Big Data, pp 629–636
43. Herlocker J, Konstan J, Borchers A, Riedl J (1999) An algorithmic framework for performing collaborative filtering. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp 230–237
44. Zhao Z, Lu H, Cai D, He X, Zhuang Y (2016) User preference learning for online social recommendation. IEEE Trans Knowl Eng 28(9):2522–2534
45. Bell R, Koren Y, Volinsky C (2007) Modeling relationships at multiple scales to improve accuracy of large recommender systems. In: Proceedings of International SIGKDD Conference on Knowledge Discovery and Data Mining, pp 95–104
46. Jamali M, Ester M (2009) Using a trust network to improve top-N recommendation. In: Proceedings of the Third ACM Conference on Recommender Systems, pp 181–188
47. Bellogin A, Parapar J (2012) Using graph partitioning techniques for neighbour selection in user-based collaborative filtering. In: Proceedings of the Sixth ACM Conference on Recommender Systems, pp 213–216