ISSN No. 0976-5697

Volume 4, No. 4, March-April 2013

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A Model for Inferring Trust for an Unknown User on Social Networking Sites

Amrita Srivastava[*1], Ruchi Verma[2], Ram Chandra Yadav[3] and S. Venkatesan[4]
Division of MBA & Master of Science in Cyber Laws & Information Security
Indian Institute of Information Technology Allahabad
Allahabad, U.P. India-211012
amritasrivastava24@gmail.com[*1], ruchi487@gmail.com[2], ramchandrayadav412@gmail.com[3]

***Abstract*:-** The advent of technology has given birth to new ways of socializing in the form of web based social networks. The increase in the number of users being associated with these networks like Facebook, LinkedIn or Orkut has drawn attention of many researchers in the past few years. The reason is not the increasing number, but the way users share their personal and sensitive information without being aware of its potential to be misused. If a malicious user or a non-trusted person gets access to our social network, he might cause a lot of damage to various aspects of our personal life. Hence, allowing only trusted people to social networks is an important matter of concern. This gives rise to the concept of trust, what trust is and on what basis one can trust another person on a web based social network. Previous studies have already shown how trust ratings given by others can influence the overall trust rating. A lot of work has already been done in the field of inferring trust score or trust value for users based on which another user can make decisions, but most of them have primarily focused on the trust ratings given to a salable product for computing the trust score .This paper will be focusing on another model for inferring trust value between unknown members, by first finding trusted paths between them and then using the trust ratings given by the immediate neighbors of the sink, that are also a part of those trusted paths.

***Keywords*:** trust rating; trust propagation; social networking sites; trust score; transitivity reduction;

I. INTRODUCTION

Today, social networking has become a major topic of interest over the web comprising of millions of users sharing their information with each other. It requires awareness among the users about sharing their information with the correct set of people. This introduces the idea of whom to trust and whom to distrust. In a study, it is stated that however, trust has a wide definition and depends on various factors like past experiences with a person, history and background information but on a social network trust value is computed based on information supplied by the user and opinions given by his associates [1]. Since, inferring trust score from the supplied information itself may not be sufficient, hence public opinions must also be considered while computing trust score. Our paper proposes a model for deducing trust based on trust ratings individually given by users and a propagation strategy between two nodes that are not directly connected. The users give these ratings as either trusted or non-trusted based on their interactions as a whole. Since trust is a personal affair and depends on a number of psychological factors where scaling can be difficult to comprehend, we are allowing users to give a binary rating to their connections.

II. RELATED WORK

A lot of research work has already been done in the area of computing trust score. We were inspired by a study, given by Golbeck which gave a concept of binary rating where users could be grouped in either of the classes as trusted and distrusted, but this was only from any user's perspective, however, it is obvious that a trusted person from one's perspective may not be trusted by others since trust is asymmetric[2]. Yarden in his paper, discussed on prioritizing the default logics by coupling between the trust computing method and the prioritized default logics [3].[1] has presented two variations of algorithms based on binary ratings used by the users to rate each other. Most of the earlier work is based on the principle of transitivity which is a general notion, where it is obvious that if A trusts B, B trusts C then by the principle stated above A can trust C. According to Gambetta D. the definition of trust itself talks about the probability of the trustee to perform any action in favor or benefit of the truster [4].A research has estimated a comparative study of various trust inference models like social trust ,Bayesian trust inference mechanism, matrix factorization and many others, pin pointing on their benefits but most of them have dropped some or the other factors[5]. The reason behind this is considering all factors in predicting trust is not feasible because trust in itself is abstract in nature. A simple trust inference algorithm named Tidal Trust Algorithm which is based on averaging model has been given[6]. In this algorithm, the rating given by source to sink is computed by taking a weighted average of the ratings given by source's neighbors to the sink. Trust between peers has been studied by computing cumulative trust scores between them, based on reliability of information received by both of them from each other, social opinions and similarity between the profiles of the peers [7].

The paper has been divided into certain sections. Section I was the introduction that discussed the background work that has been done by various researchers. Section II deals with the various concepts that have been applied in the model. Section III describes the various computational steps required to be used with this approach. Section IV details the initial analysis using data. Eventually, the future work and conclusion is stated.

III. PRELIMINARIES

This paper presents a model for determining trust and propagating trust along trusted paths from the user's perspective. It is a general notion that we always want to add a person to our network ,if we trust that person .But social networking sites like Facebook only allows you to see the public profile information that too is decided by the profile owner. Hence it becomes difficult to judge an unknown member's authenticity. Facebook does not provide any such feature of giving profile ratings to users so as to help other users in judging the trustworthiness of some unknown profile. All the researches have considered the social network as a highly connected directed graph, consisting of nodes and edges. The various concepts that form the basis of our research, are as follows:

A. Transitivity:

Computing trust has been a matter of concern for long and much research work has been done. Golbeck stated that trust is transitive in nature, may be, not completely but partially[2].It has been indicated in most of the researches that highly trusted neighbors give the most correct information which forms the backbone of our research. It is generally seen that people normally believe on things and ideas believed by their trusted friends. The same concept has been used in this paper. We are assuming that users give a rating to their friends that they either trust or distrust them. Distrust does not mean complete distrust. These ratings will help the users for assessing the trustworthiness of an unknown profile.

B. Social Feedback:

Social opinion is another way of knowing about a person where majority votes counts. In real world, it is generally noticed that people seek others' views before taking decisions. This concept has also been incorporated for giving a probabilistic trust rating to a person using a precise Binomial test, which is used for predicting truth from assertions made. Binomial test tests the means of two groups in a sufficiently but not very large sample, to check whether they are statically different or not.Since dichotomous data is involved, binomial test is performed on the sample as in equation (1)

$$Z = \frac{\left(\frac{x}{n}\right) - p}{\sqrt{\frac{pq}{n}}} \quad (1)$$

Where,

z = z-score

x= number of 1's received by user from n friends, where '1' means trusted and '0' means distrusted

n = number of friends

p = probability of success in each trial.

q = probability of failure in each trial.

Binomial test is based on hypotheses[8]. The null hypothesis considers that the proportion of friends that trust the node is not significant than those that are not trusting. The alternate hypothesis asserts that the node is trustworthy. The computed score is compared against the tabulated score. Based on the comparison, the node is given trust rating. Here we are not only considering the individual opinion about the trustworthiness of a person, but the social feedback is also considered for inferring the same.

C. Transitivity Reduction:

There are chances that many paths may lead to the sink which may have many redundant edges that may add to network and time complexity. Transitivity reduction gives a minimal representation of the formed graph, consisting of all the paths from source to sink which works accordingly as shown in figure 1:

```
for( x=1; x<=V; x++)
 { for( y=1; y<=V; y++)
   { for( z=1; z<=V; z++)
     { If (x ,z)!=(x, y) &&(x, z)!=(y, z)
       Delete edge(x,z), if (x, y)&&(y,z) exists}}}
```

where, V is a set of all vertices, E is a set of all edges of a graph G{V,E}, (x, y),(y, z),(x, z) ∈ E & x,y,z ∈ V

Figure 1: Transitivity reduction steps

The time complexity associated with removing an edge is O (n), since graphs are normally implemented using linked lists and it must be ensured before deleting an edge that the edges must be ordered and arranged again. The effectiveness of transitivity reduction in removing redundancy in directed graphs has already been proved experimentally [9].A research paper has determined that the time complexity for computing the transitive reduction of any directed graph is equal to the time complexity for determining the transitive closure or for Boolean matrix multiplication because in each of these cases the graphs can be represented as an adjacency matrix [10].

IV. TRUST INFERENCE MODEL

In this section, we apply the discussed concepts to our model of trust. We are using an example of a small network as shown in figure 2.for explaining the concept. If we assume the social networks as a graph G(V,E)comprising of a set of vertices V and a set of edges E, then the vertices represent the users and the edges represent their connections. Here, we are assuming that each user gives a rating of trust and distrust to each of its connections. According to Tidal trust algorithm, it is a proved fact that the trusted connections are supposed to be more strong than the untrusted ones. We are also using a similar concept as was used in Tidal Trust algorithm, where polling was done to determine the path from source to sink with the help of trusted neighbours. The neighbours continue the polling process for the sink until a path is obtained. Once a path is found, the polling stops and the trust results are returned back to the source [6].

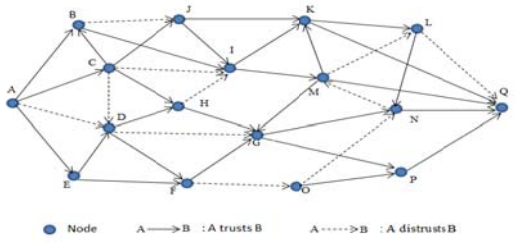

Figure 2: An illustration of the social network indicating trusted and untrusted nodes.

Let's say if A wants to connect to Q then A will first assess the authenticity of Q by computing trust rating of Q. This comprises of the following steps:

a. A will first compute the trustworthiness of its immediate friends B, C, D and E by performing a binomial test for each of its friends. Binomial test will first presume a null hypothesis and an alternate hypothesis for inferring the trust ratings by considering the ratings given by the friends of the friends.

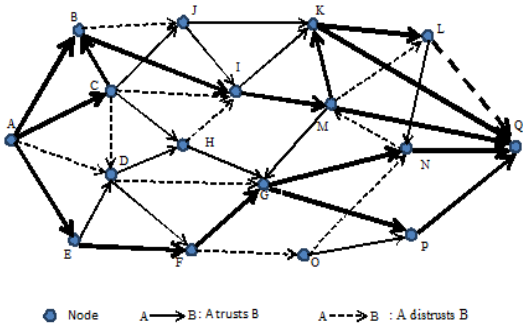

Figure 3: Different paths from source to sink.

b. Based on the acceptability of the hypothesis, the nodes B, C, D& E are given the trust ratings. If the null hypothesis is accepted, a distrust rating will be given and if alternate hypothesis is accepted, trust rating will be given. The node A will consider the trusted nodes for answering the question whether they know and trust Q. Now this query is propagated through the trusted nodes to find out a path from source node A to sink node Q. If a non-trusted node is encountered in between the path, the path is no more considered in computation. Similarly, a number of paths are found from source to destination as shown in figure 3.

Suppose B, C and E are inferred to be trusted Let's say the bold colored edges indicate the trusted paths from A to Q found after the trust computation. We can see that a number of paths are found from A to Q. Let X= {i, ii, iii, IV….n} where X is a set of paths so found from source to destination, n is the number of paths and n(X) be the set cardinality.

i. A→B→I→M→Q
ii. A→C→B→I→M→Q
iii. A→B→I→M→K→L→Q
iv. A→C→B→I→M→K→Q
v. A→C→B→I→M→K→L→Q
vi. A→B→I→M→K→Q
vii. A→E→F→G→N→Q
viii. A→E→F→G→P→Q

c. In order to remove redundancy in the found network, the transitive reduction is applied to obtain a minimal representation of the paths. Trust propagation depends upon the property of transitivity and composability where trust ratings are passed back to the source through the intermediate nodes who had queried, the source collects and performs probabilistic binomial test on the observed ratings to determine an inferred trust rating for the sink.

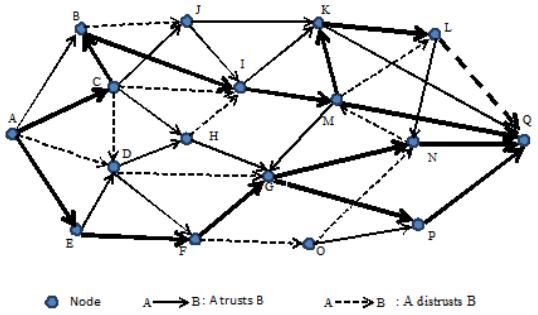

Figure 4:Reduced number of paths after transitivity reduction.

In figure 4, we can see the reduced number of paths after applying transitivity reduction. The edges A→B and K→Q have not been considered which has led to the removal of three paths i, ii and iii from the above set of edges. Let the minimal set be represented by Y= {i, ii, m} and n(Y) be the set cardinality. The given example clearly indicates that n(Y) < n(X). Hence, we can say that transitivity reduction will reduce the computational complexity. Now, binomial test can be applied in order to infer the trust rating of Q from the received ratings.

V. DATA ANALYSIS

This model will work where social network is highly connected. The better the nodes are connected the better the trust computation shall take place and accuracy will increase eventually. By segregating the nodes as trusted and distrusted, we have reduced the unnecessary involvement of the untrusted nodes in trust computation as well as trust is propagated using trusted paths only. This ensures high chances of getting a more accurate trust inference for an unknown node. This model may not prove very fruitful for very sparse networks where the nodes are scattered because paths may not exist in such cases for every other node. For analysis purposes we could only get the dataset of Epinions site that was found suitable for our model.www.Epinions.com is a website where users can write reviews and rate about different products. The special part is that the users can rate other users as trusted or untrusted. The dataset consisted of trust ratings by 131,828 users in the form of 841,372 trust and distrust ratings given to each other [11]. We have chosen a random sample of data for analyzing our case due to computational complexity.

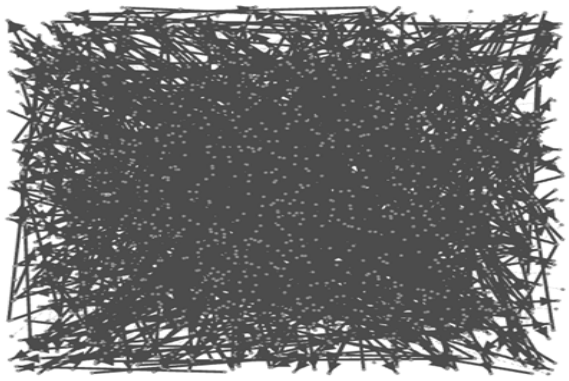

Figure 5: Dense network connectivity based on ratings.

We analyzed the data using a graph visualization tool named Gephi 0.8.2 beta version in order to understand the nature of data. Due to heavy size of data, we considered a random 1389 nodes and 999 edges for assessing the social network connectivity. Figure 5 shows how densely or sparsely the users are connected.

In order to see the distribution of trust and distrust ratings, we checked the first 30 profiles which is indicated in figure: 6, which clearly indicates that there is wide variation in the trust ratings received by users. Another observation that was made during the analysis of data was variation in z-score, which indicated that for every node trust rating could not be inferred using binomial test. Many profiles were found which did not receive any trust rating from their immediate neighbors. Here we are using level of significance as 5% corresponding to which the critical z-score is 1.65 as obtained from the z table. The following conditions must be considered:

a. If sample size n is greater than 30 then only this concept will be used.
b. If calculated z-score is less than critical z-score that is assumed to be 1.65, null hypothesis will be accepted and the node or user is assumed to be distrusted.
c. If calculated z-score is greater than critical z-score then the alternative hypothesis is accepted and the node is assumed to be trusted.

This indicates that the minimum size of sample should be 30 or above for deducing the trust rating of a particular node. If the z-score is more than 1.65 alternate hypothesis is accepted else null hypothesis is accepted. Figure 7 gives the computed z-score using equation (1) for 30 random profiles using the trust ratings received by them from their immediate friends.

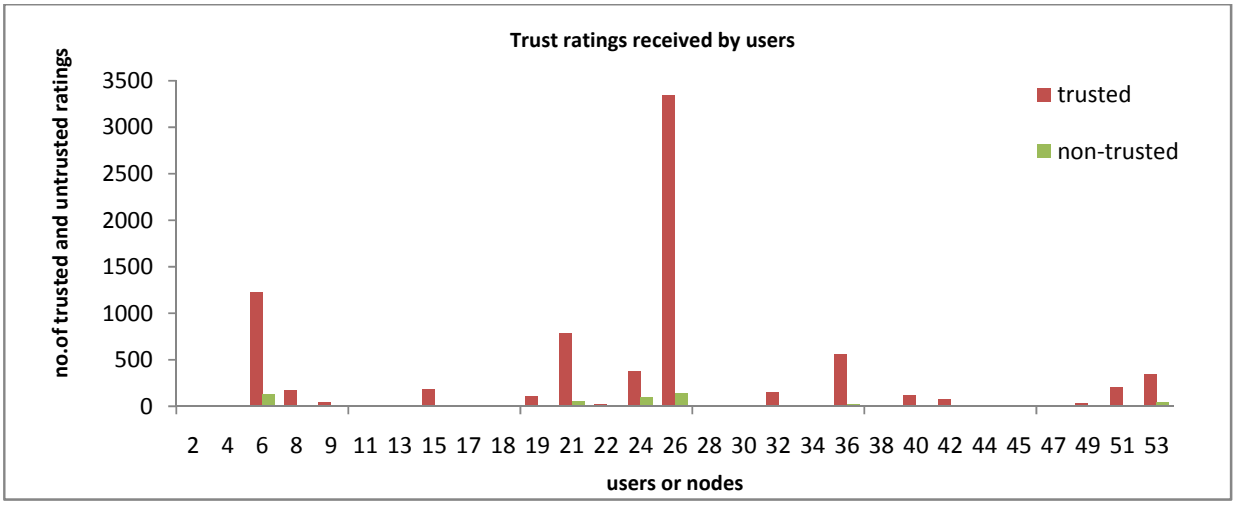

Figure 6: Variation in the trust and distrust ratings received by 30 profiles.

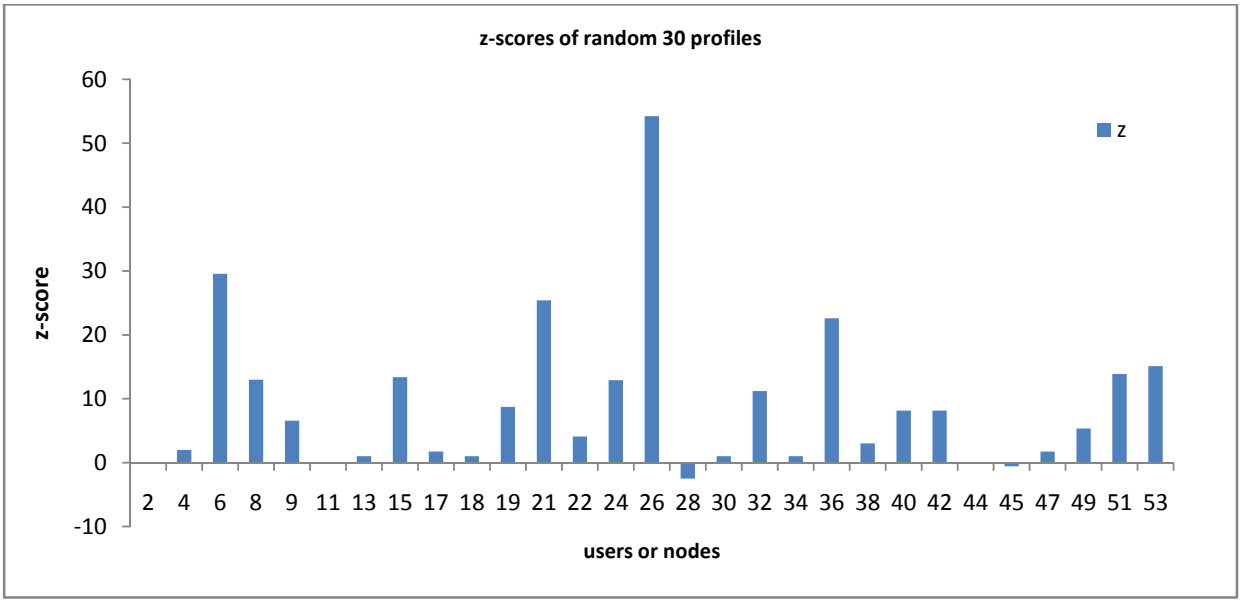

Figure 7: Computed z-scores for 30 profiles based on ratings received by them.

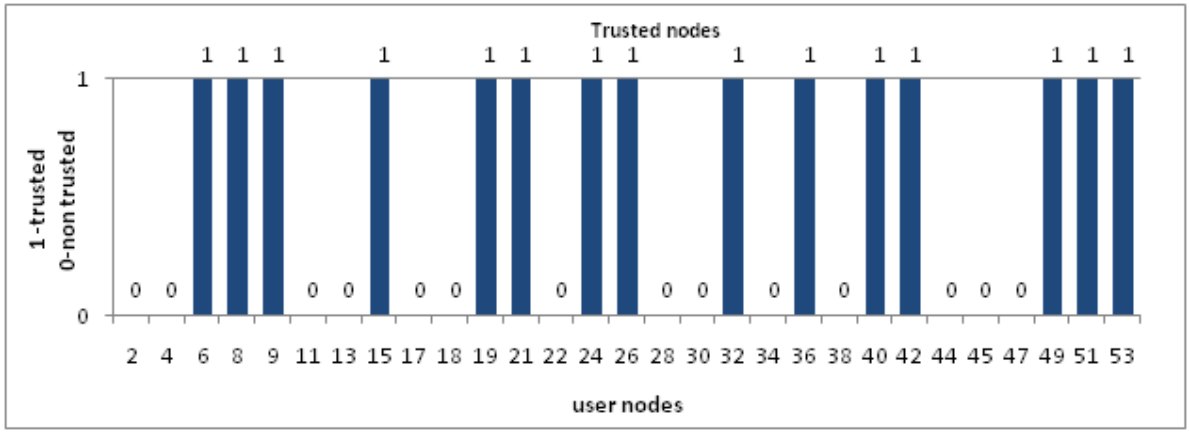

Figure 8: Inferred trust ratings of 30 profiles based on the z-scores.

The trusted nodes found after comparing the computed z-score with the critical z-score and labeling the outliers as distrusted are shown in figure: 8.

This simply proved that binomial test is able to deduce the probabilistic trust rating of the users based on the ratings received by their immediate neighbors. Hence, it can be used for predicting the trustworthiness of the source's neighbors which can further propagate the same to find out their trustworthy neighbors forming a chain of trust. Using this chain of trust, the source can find out the ratings received by the sink which can be finally used for deducing the trust rating of the sink as explained in section III.

VI. FUTURE WORK

The quality of the proposed model can be empirically analyzed by applying the above model in real time social networks and integrating with applications. This model uses a simple approach of predicting trust using the ratings given by a node's peers because we assume that immediate peers are the ones with whom the maximum interaction takes place. The peers have access to all the information associated with the node and hence can give more accurate rating to the concerned node. This information can be used for predicting trust by other unknown nodes. As for our future work, we will be working on identifying a simple and better probabilistic approach for computing trust of the immediate peers through which trust is propagated.

VII. CONCLUSION

We discussed in this paper how trust and distrust ratings given by known neighbors of a user can be used for deducing overall trust rating of that user, using statistical tests. This paper has also used the concept of trust chains by extending transitivity where A trusts B ,B trusts C,C trusts D and so on. This model might help in preventing users from allowing any doubtful profile into their network which will ultimately prevent his or her personal information from being accessed by any malicious user. This model will allow any user to test the authenticity of any other user using trust ratings given by other users known to that unknown user. A drawback of this model would be that it is useful only when paths exist between two unknown nodes on the network. However; this model is complex but we still believe that further drilling down using experimental analysis, this model shall prove its potential.

VIII.REFERENCES

[1]. J. Golbeck and J. Hendler.Inferring binary trust relationships in web-based social networks. ACM Trans.Internet Technol., 6(4):497–529, 2006.

[2]. Golbeck, J. 2005.Computing and Applying Trust in Web-based Social Networks. Ph.D. Dissertation, University of Maryland, College Park.

[3]. YardenKatz,JenniferGolbeck,Social Network-based Trust in Prioritized Default Logic, Proceedings of The Twenty-First National Conference on Artificial Intelligence (AAAI-06),Boston, Massachusetts,July 2006.

[4]. Gambetta, D., Can We Trust Trust? In D. Gambetta, editor, Trust: Making and Breaking Cooperative Relations, pp. 213-237 Oxford, 2000.

[5]. Zeinab Dehghan, Yahya AlMurtadha, Lai NganKuen and Zailan Abdul Salam, Current trust inference mechanisms in social networks, Journal of Computer Science, August, 2012, pp 1496-1500.

[6]. Jennifer Golbeck and James Handler, Trust and nuanced profile similarity in online social networks,ACM Transactions on the Web (TWEB) vol. 3 issue 4 article 12, September 2009.

[7]. Justin Zhan and Xing Fang ,A novel trust computing system in online networks, The Third IEEE International Conference on Social Computing, Socialcom, MIT, Boston, USA, October, 2011, pp. 1284-1289

[8]. http://www.elderlab.yorku.ca/~aaron/Stats2022/BinomialTest.htm

[9]. Michael. L. Case, Online Algorithms To Maintain A Transitive Reduction, CS 294-8 class project, fall 2006.

[10]. A. V. Aho, M. R. Garey and J. D. Ullman, The Transitive Reduction Of A Directed Graph, SIAM Journal On Computing,June,1972,pp 131-137.

[11]. http://konect.uni-koblenz.de/networks/epinions