

Indirect Weighted Association Rules Mining for Academic Network Collaboration Recommendations

Yun Sing Koh¹Gillian Dobbie²

Department of Computer Science
University of Auckland, New Zealand,
Email: ykoh@aut.ac.nz¹, rpears@aut.ac.nz²

Abstract

Collaborative research is increasingly important and popular in academic circles. However for young researchers identifying new research collaborators to form joint research and analyzing the level of cooperation of the current partners can be a very complex task. Thus recommendation of new collaborations would be important for young researchers. This paper presents a new approach to recommend collaborators in an academic social network using the co-authorship network. We propose a weighted indirect rule mining approach using a novel weighting mechanism called sociability.

Keywords: Social Network, Indirect Weighted Association Rule Mining, Sociability

1 Introduction

Rapid growth and exponential use of social digital media has led to an increase in popularity of social networks and the emergence of social network mining which combines data mining with social computing. As social networks are generally made of social entities that are linked by some specific type of interdependency such as friendship. Social networks represents social relationships in terms of nodes and links. Nodes are the individual actors within the networks, and links are the relationships between the actors. Social Network Analysis (SNA) analyses the importance relationships between actors, and is a central point to the evaluation and the analysis of social interactions.

Nowadays, this type of network is commonly used, and each network connects millions of users. Social network mining aims to discover implicit, previously unknown and potentially useful knowledge from a vast pool of data residing in the social networking sites such as Twitter (Weng et al. 2010, Ghosh et al. 2012), Facebook (Fan & Yeung 2010), Google+ (Leenes 2011), LinkedIn.

An example of the social network application is the Co-authorship Social Network which represents a scientific collaboration network (Huang et al. 2008). Increasing research collaboration amongst researchers can bring together different points of view to address a particular research issue. Furthermore, studies have shown that scholars with higher levels of collaboration tend to be more productive (Lotka 1926). Thus, it would be beneficial for new emerging researchers

Copyright ©2012, Australian Computer Society, Inc. This paper appeared at the 10th Australasian Data Mining Conference (AusDM 2012), Sydney, Australia, December 2012. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 134, Yanchang Zhao, Jiuyong Li, Paul Kennedy, and Peter Christen, Ed. Reproduction for academic, not-for profit purposes permitted provided this text is included.

to find potential successful collaborators. Yet traditional digital libraries and search engines focus on discovering relevant documents which does not make it straightforward to search for people who share similar research interests (Chen et al. 2011). There have been a few digital library platforms, such as ArnetMiner (Tang et al. 2008) and Microsoft Academic Search (Microsoft Academic Search 2011) which returns a list of experts given a particular domain. However, the list only provides a limited set of names does not consider the implicit social networks of the experts.

To help in efficiently discovering potential collaborators, we present a new approach that considers social network structure based on reachability, and sociability of a researcher as a recommendation tool for potential collaborators. Our approach weights researchers based on a sociability factor, which tries to capture how often they work with a different researcher. Using these weights we are able to generate rules to describe the connection between a researcher and a collaborator. We then use these rules to generate recommendations to other researchers who are indirectly associated to them and may be possible collaborators. In our experiments we used the collaborative network from the digital community DBLP.

The paper is organized as follows. In Section 2 we look at related work in the area of recommendations for social network. In Section 3 we present our weighted indirect association rule mining approach. In Section 4 we discuss our experimental results. Finally, Section 5 concludes the paper.

2 Related Work

Ever since the proliferation of social network research, there has been a considerable amount of research carried out to build recommendations for social networks (Ogata et al. 2001, Quercia & Capra 2009, Karagianis & Vojnovic 2009, Chen et al. 2009, Weng & Chang 2008, Cheong & Corbitt 2009, Roth et al. 2010). The related work presented in this section aims to use social networks in the context of recommendation systems for an academic network.

Aleman-Meza et al. (2006) proposed a solution to solve the conflict of interest problem using social networks. The main objective was to detect relationships of conflict of interest amongst authors of scientific papers and potential reviewers of those papers based on public sources such as DBLP and the Friend of a Friend project.

Kautz et al. (1997) proposed the ReferralWeb system to identify experts in searches by keywords and generate a path of social relationships between a user and the recommended expert. The proposed solution models and extracts existing relationships among people in the area of Computer Science using public data available on Web documents. McDonald (2003) pro-

posed an evaluation of two different social networks that can be used in a system to recommend individuals for possible collaborations. The system matches individuals looking for expertise within people that could have this expertise.

Zaiane et al. (2007) proposed a technique which explored a social network based on the DBLP database by using a new random walk approach to find interesting information about the research community and then recommended collaborations. The approach aims at helping the user in the process of searching for relevant conferences, similar authors and interesting research topics.

Chen et al. (2011) proposed CollabSeer which is an open system to recommend potential research collaborators for scholars and scientists. The proposed approach discovers collaborators based on the structure of the coauthor network and a users research interests. Currently, three different network structure analysis methods that use vertex similarity are supported in CollabSeer: Jaccard similarity, cosine similarity, and the relation strength similarity measure.

There has been some research in using frequent pattern mining in finding interesting patterns in an academic network (Adnan et al. 2009, Nohuddin et al. 2012). In this paper we propose a new approach, in finding recommendations for an academic network, using an indirect frequent mining approach.

3 Mining Weighted Indirect Association Rules

In this section we describe our proposed weighted indirect rule mining approach. In Section 3.1 we discuss the weighted association rule mining approach, and our weighting mechanism called sociability weight. In Section 3.2 and Section 3.3 we discuss the combination of weighted association rule mining and indirect association rule mining.

3.1 Weighted Association Rules Based On Sociability

Association rule mining is an important data mining task that discovers relationships among items in a transaction database. Most approaches to association rule mining assume that all items within a dataset have a uniform distribution with respect to support. Therefore, weighted association rule mining was introduced to provide a notion of importance to individual items.

Given a set of items, $I = \{i_1, i_2, \dots, i_n\}$, a transaction may be defined as a subset of I and a dataset as a set D of transactions. A set X of items is called an itemset. The support of X , $\text{sup}(X)$, is the proportion of transactions containing X in the dataset. An *association rule* is an implication of the form $X \rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \emptyset$. The rule $X \rightarrow Y$ has *support* of s in the transaction set D , if $s = \text{sup}(XY)$. The rule $X \rightarrow Y$ holds in the transaction set D with *confidence* c where $c = \text{conf}(X \rightarrow Y) = \text{sup}(XY)/\text{sup}(X)$. The association rules are also known as a direct association rules. Given a transaction database D , a support threshold minsup and a confidence threshold minconf , the task of association rule mining is to generate all association rules that have support and confidence above the user-specified thresholds.

In weighted association rule mining a weight w_i is assigned to each item i , reflecting the relative importance of an item over other items that it is associated with. The weighted support of an item i is $w_i \text{sup}(i)$. Similar to traditional association rule mining, a weighted support threshold and a confidence

threshold is assigned to measure the strength of the association rules produced. The weight of a k -itemset, X , is given by:

$$\left(\sum_{i \in X} w_i\right) \text{sup}(X) \tag{1}$$

Here a k -itemset, X , is considered a frequent itemset if the weighted support of this itemset is greater than the user-defined minimum weighted support ($w\text{minsup}$) threshold.

$$\left(\sum_{i \in X} w_i\right) \text{sup}(X) \geq w\text{minsup} \tag{2}$$

The weighted support of a rule $X \rightarrow Y$ is:

$$\left(\sum_{i \in X \cup Y} w_i\right) \text{sup}(XY) \tag{3}$$

In our approach we proposed a new sociability weight as the weighting mechanism.

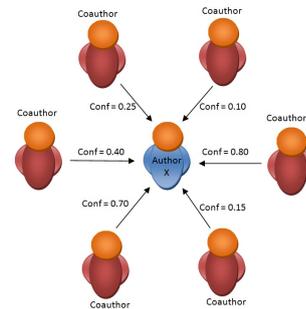


Figure 1: Author-Coauthor Graph

Definition 1 (Sociability Weight). The sociability weight is defined based on the coauthors, i , an expert (author), k , has and the confidence of the coauthors towards the author. Given an author, k , which is connected to a set of n coauthors, the sociability weight, soc_k is defined as:

$$\text{soc}_k = \sum_i^n \frac{\text{sup}(i, k)}{\text{sup}(i)} \tag{4}$$

which is equalvant to

$$\text{soc}_k = \sum_i^n \text{conf}(i \rightarrow k)$$

The reasoning behind this is that we are interested in promoting an expert (author) which works with a range of other researchers (coauthors). In turn the other researchers must also have a high confidence towards the author, which means that they have published frequently with the same expert.

Figure 1 shows an author-coauthors relationship. The arrows represent the rules formed between a coauthor and author X . In this example the sociability weight for the author X is $0.25 + 0.40 + 0.70 + 0.15 + 0.80 + 0.10 = 2.65$. The weights are used to float experts which are deemed to be important to the top.

Here we discuss weighted direct rules in collaboration recommendation.

Definition 2 (Weighted Direct Rule). Let D be a dataset. A weighted direct association rule between two authors is the relationship between an author X and its coauthor Y , where $X \rightarrow Y$, where $X \subseteq D$, $Y \subseteq D$, and $X \cap Y = \emptyset$. The $w\text{sup}(X \rightarrow Y) \geq w\text{minsup}$ and $\text{conf}(X \rightarrow Y) \geq \text{minconf}$.



Figure 2: Direct Rule

Weighted direct association rules represent regularities discovered from a large dataset based on the weighting scheme. The problem of mining association rules is to extract rules that are strong enough and have the weighted support ($wsup$) and confidence value greater than given thresholds: minimum weighted direct support ($wminsup$) and minimum direct confidence ($minconf$).

We use the rules generated from weighted association rules in this section to form indirect rules which we will discuss in the next section.

3.2 Indirect Association Rules in Social Networking

In a classical sense, an indirect association pattern refers to a pair of items that rarely occur together but highly depend on the presence of a mediator itemset (Tan & Kumar 2002). Indirect association has been used extensively to build web recommendation systems (Kazienko 2009, Tan & Kumar 2002). In this research, we propose to use a weighted indirect association rule mining approach for collaboration recommendation in an academic social network.

Let us consider another type of associations: indirect association rules.

Definition 3 (Weighted Indirect Itemset). *An itemset (pair of researchers) $\{X, Y\}$ is indirectly associated via a mediator M , if $sup(X, Y) < wminsup$, $sup(X, M) \geq wminsup$, and $sup(Y, M) \geq wminsup$.*

Definition 4 (Weighted Indirect Rule). *Let D be a dataset. A weighted indirect association rule $X \rightarrow^{M\#} Z$ is the indirect relationship from X to Z with respect to M , for which two direct weighted association rules exist: $X \rightarrow M$ and $M \rightarrow Y$, where $X \subseteq D$, $M \subseteq D$, and $Y \subseteq D$; $X \neq M \neq Y$; and $conf^I(X \rightarrow^{M\#} Y) \geq minconf^I$.*

Each weighted indirect association rule $X \rightarrow^{M\#} Y$ has an indirect confidence $conf^I$ value which can be defined as follows:

$$conf^I(X \rightarrow^{M\#} Y) = conf(X \rightarrow M).conf(M \rightarrow Y)$$

For example given there are two rules $X \rightarrow M$ with $conf = 0.90$ and $M \rightarrow Y$ with $conf = 0.80$. Thus, $conf^I(X \rightarrow^{M\#} Y) = 0.90 \times 0.80 = 0.72$. There are two types of weighted indirect rule: partial indirect and complete indirect.

Definition 5 (Weighted Partial Indirect Rule). *Let D be a dataset. A weighted partial indirect association rule $X \rightarrow^{M\#} Y$ is the indirect relationship from X to Y with respect to M , for which two direct weighted association rules exist: $X \rightarrow M$ and $M \rightarrow Y$, where $X \subseteq D$, $M \subseteq D$, and $Y \subseteq D$; $X \neq M \neq Y$; and $conf^I(X \rightarrow^{M\#} Y) \geq minconf^I$; and $X \cap Y \neq \emptyset$.*

A weighted partial indirect rule $X \rightarrow^{M\#} Y$ reflects one indirect association existing between X and Y , with no direct association $X \rightarrow Y$, even though X occurs together with Y (shown in Figure 3). In Figure 3 the solid line between X and Y represents that both the authors are co-authors but $sup(X, Y) < wminsup$, thus no weighted direct rule between these two authors are generated.

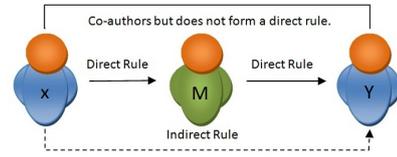


Figure 3: Partial Indirect Rule

Definition 6 (Weighted Complete Indirect Rule). *Let D be a dataset. A weighted complete indirect association rule $X \rightarrow^{M\#} Y$ is the indirect relationship from X to Y with respect to M , for which two direct weighted association rules exist: $X \rightarrow M$ and $M \rightarrow Y$, where $X \subseteq D$, $M \subseteq D$, and $Y \subseteq D$; $X \neq M \neq Y$; $conf^I(X \rightarrow^{M\#} Y) \geq minconf^I$; and $X \cap Y = \emptyset$.*

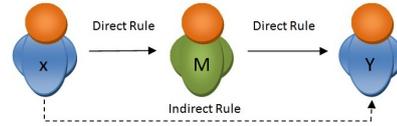


Figure 4: Complete Indirect Rule

A weighted complete indirect rule $X \rightarrow^{M\#} Y$ reflects one indirect association existing between X and Y , with no direct association $X \rightarrow Y$, and X does not occur with Y (shown in Figure 4).

3.3 Weighted Indirect Association Rule Mining

In our algorithm, we focus on finding weighted indirect rules by combining the Sociability weight in Section 3.1 and the indirect rule mining approach described in Section 3.2. In this section we describe this combined approach. The algorithm is divided into two major phases. In Phase 1 we generate all weighted frequent itemsets (Algorithm 1) using a sociability weight function shown in Algorithm 2. In Phase 2, we find all indirect associations (Algorithm 3).

A general weighted association rule mining algorithm is shown in Algorithm 1. The algorithm requires a weighted minimum support to be provided. In this algorithm L_k represents the weighted frequent itemsets and C_k represents the candidate itemsets. Candidate itemsets whose weighted support exceeds the weighted minimum support are considered large itemsets and will be included in the rule generation phase.

Algorithm 1 Weighted candidate generation algorithm

Input: Transaction database D , $wminsup$ value, universe of items I
 Output: Weighted frequent itemsets, L_k
 $k \leftarrow 1$
 $L_k \leftarrow \{\{i\} | i \in I, soc(i) * sup(i) \geq wminsup\}$
 while $L_k \neq \emptyset$ do
 $k \leftarrow k + 1$
 $C_k \leftarrow \{x \cup y | x, y \in L_{k-1}, |x \cap y| = k - 2\}$
 $L_k \leftarrow \{c | c \in C_k, soc(c) * sup(c) \geq wminsup\}$
 end while
 return $\bigcup_{t=2}^{k-1} L_t$

Algorithm 2 Sociability Weight, $\text{soc}(i)$

Input: Item i , universe of items I
Output: Sociability Weight
{Find the neighbourhood of item i .}
 $N_i \leftarrow \{\{j\} | j \in I, \text{sup}(i, j) > 0\}$
return $\sum_{j \in N_i} \frac{\text{sup}(j, i)}{\text{sup}(j)}$

Frequent itemset L_k is used to generate candidate indirect associations P . Each candidate in P is a triplet $\langle x, y, M \rangle$, where x and y are the items which are indirectly associated by mediator M . P is generated joining the frequent itemsets in L_k . During the join, a pair of frequent itemsets $\{x_1, x_2, \dots, x_k\}$ and $\{y_1, y_2, \dots, y_k\}$ are joinable if the two itemsets have exactly $k - 1$ items in common. If so, they generate a candidate indirect association $\langle x, y, M \rangle$, where x and y are the different items, one from each k -itemset, and M is the set of common items.

Algorithm 3 Indirect rule mining algorithm

Input: Itemset L_k , minconf^I value
Output: Indirect Rules
 $R \leftarrow \emptyset$
 $P \leftarrow \{x \cup y | x, y \in L_k, |x \cap y| = k - 1\}$
for $\langle x, y, M \rangle \in P$ do
 if $\text{conf}(x \rightarrow M) \cdot \text{conf}(M \rightarrow y) > \text{minconf}^I$
 then
 $R \leftarrow \{\langle x, y, M \rangle | x \in P, y \in P\}$
 end if
end for
return R

For example, two itemsets $\{a, y\}$ and $\{a, z\}$ can be joined together to generate a candidate indirect association $\langle y, z, \{a\} \rangle$. Since the candidate indirect associations are generated by joining two frequent itemsets, they certainly satisfy the mediator confidence condition, minconf^I . In this example, $\{a\}$ is the mediator.

4 Results and Evaluation

In this section we evaluate the performance of mining indirect weighted association rules for collaboration recommendations. To the best of our knowledge, our technique is the first to suggest recommendation using indirect rule mining in a social media context. However there has been some work carried out based on other techniques.

We compare our results to strength vertex similarity method used in Collabseer (Chen et al. 2011). The relation strength of two adjacent authors is proportional to the number of their coauthored articles. If author A has n_A publications, author B has n_B publications, author A and author B coauthored n_{AB} articles. The relation strength from author A to author B is defined as follows.

$$R(A, B) = \frac{n_{AB}}{n_A}$$

For two non-collaborator authors A and C , if A could reach C only through author B , then how close author A is to author C should be proportional to the relation strength of author A to author B and the relation strength of author B to author C . We define indirect relation strength from author A to author C as:

$$R'(A, C) = R(A, B) \cdot R(B, C)$$

which can be written as:

$$R'(A, C) = \frac{n_{AB}}{n_A} \cdot \frac{n_{BC}}{n_B} = \frac{\text{sup}(AB)}{\text{sup}(A)} \cdot \frac{\text{sup}(BC)}{\text{sup}(B)}.$$

Thus,

$$R'(A, C) = \text{conf}(A \rightarrow B) \cdot \text{conf}(B \rightarrow C)$$

which is similar to a standard indirect rule conf^I measure for the indirect rule $A \rightarrow^{B\#} C$ with B as the mediator where is not included.

In their approach all authors are given equal weighting. We believe some authors are more active and more likely to form collaboration. Thus we used the sociability weights to promote these collaborations.

Table 1: Characteristics of Datasets

Dataset	Trans	Items	Avg Len
DBLP Data Mining	34215	2117	2.71
DBLP Artificial Intelligence	35380	6817	2.57
DBLP Software Engineering	21628	1591	2.58
DBLP Database	11931	2922	2.59
T10I4D100K	100000	870	10.1

In our experiments we used the DBLP Computer Science Bibliography dataset (<http://dblp.uni-trier.de/xml/>), and a frequent mining dataset that is available from the Frequent Itemset Mining Implementations (FIMI) repository (<http://fimi.ua.ac.be/>). When we use the frequent mining dataset, we map each unique item as an author and the set of items it co-occurs with as their collaborators. Note that the transactions in the above dataset share similar characteristics as those in a academic collaboration network. T10I4D100K is a dataset with a large number of items and transactions. The lengths of transactions within these datasets are relatively short. These datasets represent scenarios of a social network which comprises of many people with a small group of people that they interact with. This is similar to that of a collaboration network. From the DBLP dataset, we extracted papers written from 2000-2009. From the selected papers we extracted, we partitioned the datasets into different research areas based on the publication venue. We chose the datasets from four different research areas: databases, data mining, artificial intelligence, and software engineering. Table 1 summarizes the characteristics of the datasets used in the following experiments. For each dataset, we show the number of transactions, number of items, and average length of the transactions.

4.1 Number of Indirect Rules

In the first experiment we compare the number of recommendations generated by our algorithm and the existing algorithm. The number of indirect rules represents the number of recommendations found. The results are shown in Figure 5 and Figure 6.

Figure 5 shows the complete indirect rules generated, whereas, Figure 6 shows the number of partial indirect rules generated. In all the experiments we used a $w\text{minsup}$ of 0.001. We varied the minConf^I from 0.20 to 0.50. The number of recommendations or rules generated are inversely proportional to the minConf^I threshold. When the minConf^I threshold decreases, the number of recommendations produced increases.

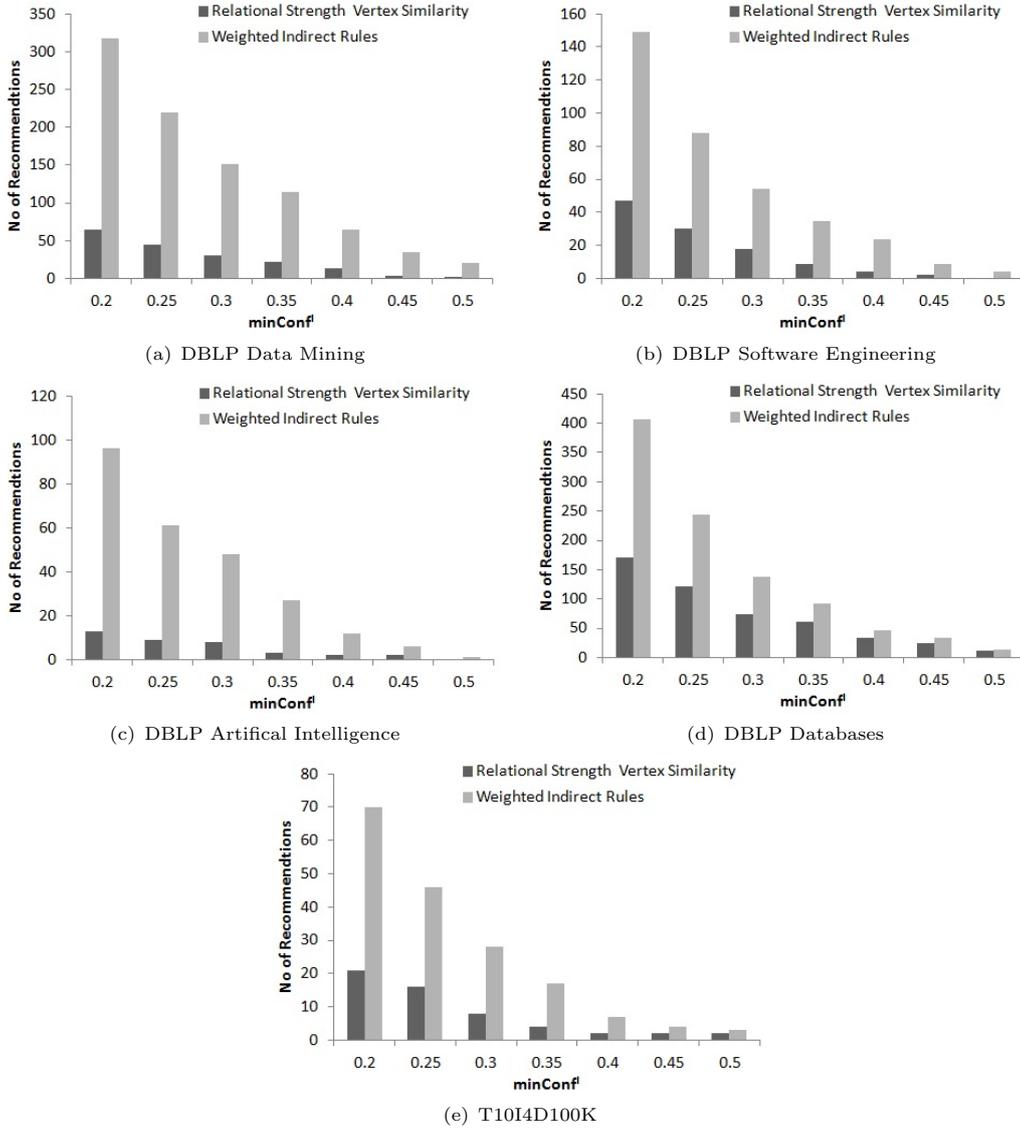


Figure 5: Number of Complete Indirect Rules

Overall the number of recommendations found by our technique is higher than the relation strength vertex similarity algorithm. The number of recommendations found by our technique is between 1 to 7 times more than the relation strength vertex similarity algorithm.

4.2 Lift Analysis

To evaluate the strength of the recommendations (rules) produced we use the lift measure. Lift is a well-known statistical measure that can be used to rank rules in IBMs Intelligent Miner (Bayardo & Agrawal 1999):

$$lift(X \rightarrow Y) = \frac{\sup(XY)}{\sup(X).\sup(Y)}$$

Note that if the occurrence of A and B are perfectly independent, the $lift(X \rightarrow Y) = 1$. If X and Y appear together more often than we would expect under independence, the lift is greater than 1, and otherwise it is less than one.

In the similar way indirect confidence, $conf^I$, is defined for indirect rule, we adapt the lift measure to an indirect lift measure, $lift^I$, for $X \rightarrow^{M\#} Y$ as:

$$lift^I(X \rightarrow^{M\#} Y) = lift(X \rightarrow M).lift(M \rightarrow Y)$$

Table 2: Average Lift Values

Dataset	Weighted Indirect Rules		
	Complete	Partial	Total
DBLP Data Mining	505033.3	2923.5	507956.7
DBLP Artificial Intelligence	142485.7	114	142599.6
DBLP Software Engineering	303300.3	180.7	303481.1
DBLP Databases	502548	107.9	502655.9
T10I4D100K	4578.4	4.5	4582.9
Dataset	Rel. Strength Vertex Similarity		
	Complete	Partial	Total
DBLP Data Mining	116442.2	5640.2	122082.4
DBLP Artificial Intelligence	121360.7	361.3	121722.0
DBLP Software Engineering	139179.6	373	139552.6
DBLP Databases	525566.7	242.3	525809.0
T10I4D100K	3495.9	178.9	3674.7

Table 2 shows the average lift values produced by our algorithm compared to the strength vertex similarity method. Overall our algorithm consistently produced rules which had a higher lift value. In this experiment, the $minConf^I$ set at 0.20. We chose a low $minConf^I$ value as it produces the most recommendations for both algorithms. If a higher $minConf^I$ threshold is selected, the set of recommendations would be a subset of the recommendations generated whilst using the lower $minConf^I$ thresh-

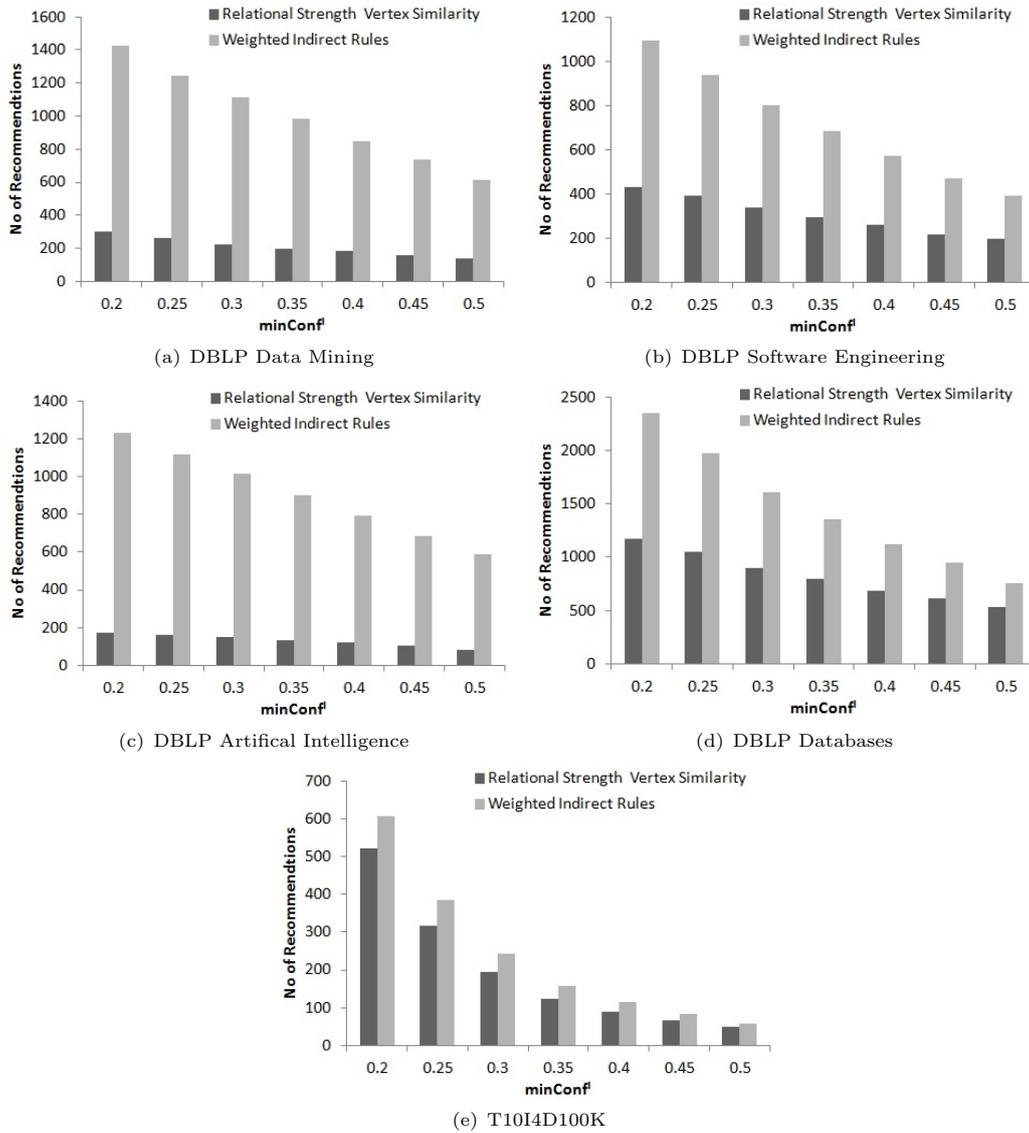


Figure 6: Number of Partial Indirect Rules

old. Thus by choosing a lower $minConf^I$ we are evaluating the superset of the recommendations generated.

4.3 Runtime Analysis

Here we compare the execution time of the two algorithm. Figure 7 shows the results of the experiment. Overall the number of recommendations influence the

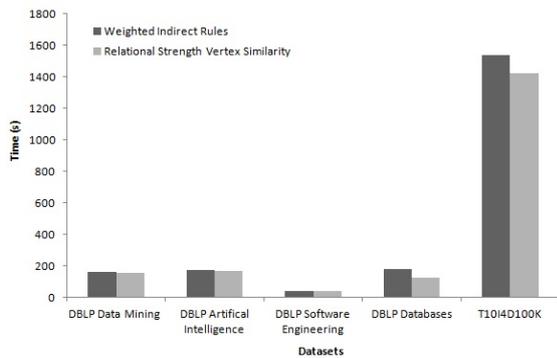


Figure 7: Runtime Analysis

runtime of the algorithms. Despite the additional

recommendations generated by the weighted indirect rule algorithm, the runtime is still comparable to the relation strength vertex similarity algorithm.

5 Conclusions

In this paper we proposed a novel algorithm to mine social networks for collaboration recommendation. Our proposed technique uses a weighted mechanism called sociability weight and combined it with indirect association rule mining. Overall our technique generated more recommendations as compared to a previous approach, relation strength vertex similarity algorithm and the additional recommendation are considered strong.

In the future we may consider other features such as citations or latent semantic analysis (abstracts or keywords for example), which better spans across academic domains.

References

Adnan, M., Alhadjj, R. & Rokne, J. (2009), Identifying social communities by frequent pattern mining, *in* 'Information Visualisation, 2009 13th International Conference', pp. 413 -418.

- Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Ding, L., Kolari, P., Sheth, A. P., Arpinar, I. B., Joshi, A. & Finin, T. (2006), Semantic analytics on social networks: experiences in addressing the problem of conflict of interest detection, in 'Proceedings of the 15th international conference on World Wide Web', WWW '06, ACM, New York, NY, USA, pp. 407–416.
- Bayardo, Jr., R. J. & Agrawal, R. (1999), Mining the most interesting rules, in 'Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining', KDD '99, ACM, New York, NY, USA, pp. 145–154.
URL: <http://doi.acm.org/10.1145/312129.312219>
- Chen, H.-H., Gou, L., Zhang, X. & Giles, C. L. (2011), Collabseer: a search engine for collaboration discovery, in 'Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries', JCDL '11, ACM, New York, NY, USA, pp. 231–240.
- Chen, J., Geyer, W., Dugan, C., Muller, M. & Guy, I. (2009), Make new friends, but keep the old: recommending people on social networking sites, in 'Proceedings of the 27th international conference on Human factors in computing systems', CHI '09, ACM, New York, NY, USA, pp. 201–210.
- Cheong, F. & Corbitt, B. J. (2009), A social network analysis of the co-authorship network of the australasian conference of information systems from 1990 to 2006, in '17th European Conference on Information Systems (ECIS 2009)', pp. 292–303.
- Fan, W. & Yeung, K. (2010), Virus propagation modeling in facebook, in 'Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on', pp. 331–335.
- Ghosh, S., Viswanath, B., Kooti, F., Sharma, N. K., Korlam, G., Benevenuto, F., Ganguly, N. & Gumadi, K. P. (2012), Understanding and combating link farming in the twitter social network, in 'Proceedings of the 21st international conference on World Wide Web', WWW '12, ACM, New York, NY, USA, pp. 61–70.
URL: <http://doi.acm.org/10.1145/2187836.2187846>
- Huang, J., Zhuang, Z., Li, J. & Giles, C. L. (2008), Collaboration over time: characterizing and modeling network evolution, in 'Proceedings of the international conference on Web search and web data mining', WSDM '08, ACM, New York, NY, USA, pp. 107–116.
- Karagiannis, T. & Vojnovic, M. (2009), Behavioral profiles for advanced email features, in 'Proceedings of the 18th international conference on World wide web', WWW '09, ACM, New York, NY, USA, pp. 711–720.
- Kautz, H., Selman, B. & Shah, M. (1997), 'Referral web: combining social networks and collaborative filtering', *Commun. ACM* **40**, 63–65.
- Kazienko, P. (2009), 'Mining indirect association rules for web recommendation', *Int. J. Appl. Math. Comput. Sci.* **19**, 165–186.
- Leenes, R. (2011), Who needs facebook or google+ anyway: privacy and sociality in social network sites, in 'Proceedings of the 7th ACM workshop on Digital identity management', DIM '11, ACM, New York, NY, USA, pp. 31–32.
URL: <http://doi.acm.org/10.1145/2046642.2046650>
- Lotka, A. J. (1926), 'The frequency distribution of scientific productivity', *Journal of the Washington Academy of Sciences* **16**(12), 317–323.
- McDonald, D. W. (2003), Recommending collaboration with social networks: a comparative evaluation, in 'Proceedings of the SIGCHI conference on Human factors in computing systems', CHI '03, ACM, New York, NY, USA, pp. 593–600.
- Microsoft Academic Search (2011), <http://academic.research.microsoft.com/>.
- Nohuddin, P. N., Coenen, F., Christley, R., Setzkorn, C., Patel, Y. & Williams, S. (2012), 'Finding interesting trends in social networks using frequent pattern mining and self organizing maps', *Knowledge-Based Systems* **29**(0), 104–113.
- Ogata, H., Yano, Y., Furugori, N. & Jin, Q. (2001), 'Computer supported social networking for augmenting cooperation', *Comput. Supported Coop. Work* **10**, 189–209.
- Quercia, D. & Capra, L. (2009), Friendsensing: recommending friends using mobile phones, in 'Proceedings of the third ACM conference on Recommender systems', RecSys '09, ACM, New York, NY, USA, pp. 273–276.
- Roth, M., Ben-David, A., Deutscher, D., Flysher, G., Horn, I., Leichtberg, A., Leiser, N., Matias, Y. & Merom, R. (2010), Suggesting friends using the implicit social graph, in 'Proceedings of the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining'.
- Tan, P.-N. & Kumar, V. (2002), Mining indirect associations in web data, in R. Kohavi, B. M. Masand, M. Spiliopoulou & J. Srivastava, eds, 'WEBKDD', Vol. 2356 of *Lecture Notes in Computer Science*, Springer, pp. 145–166.
- Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L. & Su, Z. (2008), Arnetminer: extraction and mining of academic social networks, in 'Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining', KDD '08, ACM, New York, NY, USA, pp. 990–998.
URL: <http://doi.acm.org/10.1145/1401890.1402008>
- Weng, J., Lim, E.-P., Jiang, J. & He, Q. (2010), Twitrank: finding topic-sensitive influential twitterers, in 'Proceedings of the third ACM international conference on Web search and data mining', WSDM '10, ACM, New York, NY, USA, pp. 261–270.
URL: <http://doi.acm.org/10.1145/1718487.1718520>
- Weng, S.-S. & Chang, H.-L. (2008), 'Using ontology network analysis for research document recommendation', *Expert Syst. Appl.* **34**, 1857–1869.
- Zaiane, O. R., Chen, J. & Goebel, R. (2007), Dbconnect: mining research community on dblp data, in 'Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis', WebKDD/SNA-KDD '07, ACM, New York, NY, USA, pp. 74–81.

