PAPER

Time Score: A New Feature for Link Prediction in Social Networks*

Lankeshwara MUNASINGHE[†a)], *Nonmember and* **Ryutaro ICHISE**[†b)], *Member*

SUMMARY Link prediction in social networks, such as friendship networks and coauthorship networks, has recently attracted a great deal of attention. There have been numerous attempts to address the problem of link prediction through diverse approaches. In the present paper, we focus on the temporal behavior of the link strength, particularly the relationship between the time stamps of interactions or links and the temporal behavior of link strength and how link strength affects future link evolution. Most previous studies have not sufficiently discussed either the impact of time stamps of the interactions or time stamps of the links on link evolution. The gap between the current time and the time stamps of the interactions or links is also important to link evolution. In the present paper, we introduce a new time-aware feature, referred to as *time score*, that captures the important aspects of time stamps of interactions and the temporality of the link strengths. We also analyze the effectiveness of *time score* with different parameter settings for different network data sets. The results of the analysis revealed that the *time score* was sensitive to different networks and different time measures. We applied *time score* to two social network data sets, namely, Facebook friendship network data set and a coauthorship network data set. The results revealed a significant improvement in predicting future links.

key words: *link prediction, time stamps, temporal behavior, social networks*

1. Introduction

Link prediction [13] was introduced as a way to infer which new links are likely to occur in the near future in a given network. If we are presented with a snapshot of a network at time t_c, the goal is to predict links that are likely to occur at a future time t_f. The information of the structure of the given network and the features of nodes and edges can be used to predict future links.

Link prediction in social networks has become an important task in network science because it offers great benefits to the users of social networking services as well as to various organizations and researchers. For example, online social networking services, such as Facebook, can provide their users with more accurate service and more precise recommendations or suggestions. Therefore, users of these services can efficiently find their friends, colleagues, or people whom they wish to meet [14]. Organizations such as security agencies and business organizations will be able to find more accurate information regarding unseen relationships among people or organizations and so may operate more effectively. Researchers can find other individuals in the same research field, experts, and research organizations [8], [9], [21], [24], [25], [32], [33]. However, highly structured massive real-world networks involving heterogeneous entities with complex associations have added new challenges to link prediction research. Supervised and unsupervised learning methods [4], [11] have been used in previous studies with different frameworks for link prediction, but machine learning approaches remain an immense challenge [5], [27]. Machine learning methods are difficult to apply because of the complexity and size of the networks as well as the temporal behaviors of the links in the networks.

The temporality of links can be caused by various factors depending on the nature of the network. The factors that cause the temporal behavior of the links and how these factors can be effectively used for link prediction in networks must be determined. To our knowledge, this scenario has not been discussed sufficiently in the context of link prediction. The links are strong for a certain period of time but then become weaker and fade. Such link behavior increases the complexity of link prediction because stronger links have a greater influence over link evolution than weaker links. The main contribution of the present study is determining the impact of the relationship between the time stamps of the interactions and the link strength for future links. Therefore, we introduce a new feature to incorporate the impact of the time stamps of the interactions and the gap between the current time and the time stamps. In addition, the present paper discusses the correlation between the measurement unit of time and the parameters of the new feature as an extension of a previous study [16]. We use the new feature in conjunction with supervised machine learning methods in order to predict links in network data sets.

The remainder of the present paper is organized as follows. In Sect. 2, we discuss related studies and the importance of time awareness for link prediction. In Sect. 3, we introduce a method of link prediction and the newly proposed feature, *time score*. Experimental results are presented in Sect. 4, and a general discussion is presented in Sect. 5. Section 6 presents our conclusions and discusses future research.

2. Related Research

In this section, we review research related to link prediction

Manuscript received August 16, 2011.
Manuscript revised November 21, 2011.

[†]The authors are with National Institute of Informatics, Tokyo, 101–8430 Japan.

*This is an extended paper based on research presented at DaWaK 2011, Toulouse, France.

a) E-mail: lankesh@nii.ac.jp
b) E-mail: ichise@nii.ac.jp

DOI: 10.1587/transinf.E95.D.821

as well as background information on link prediction. The increase in the number of studies related to link prediction in the recent literature reveals a growing interest in link prediction. Diverse approaches, including machine learning approaches and probabilistic approaches, have been proposed in order to address the problem of link prediction.

Link prediction is a type of link mining, which is a newly emerging research field in the realm of data mining, and presents new challenges to machine learning technologies [6]. Feature construction and collective classification using a learned model is a prominent feature of machine learning. A support vector machine (SVM) was used in combination with the structural features of networks introduced in [13] for link prediction in coauthorship networks [8], [21]. Later, the introduction of features such as keyword match count for paper topics and abstracts [24], [32], in combination with decision trees provided more accurate link predictions in coauthorship networks. These previous studies have proven the consistency and effectiveness of decision trees and the SVM [3] in link prediction. However, sparse real-world networks have presented additional difficulties in machine learning approaches due to the huge imbalance between possible links and actual links observed in these networks. The authors of a previous study [14] interpreted the problem of link prediction as a problem in class imbalance between possible links and actual links. They used SMOT [2], which is a widely accepted sampling strategy to overcome imbalance.

Probabilistic approaches basically estimate the likelihood of future possible links [10]. Among recent studies, a local probabilistic model was used in [31] to estimate the cooccurrence probability of a node with other nodes within the local proximity of the node. The local proximity is defined on the path length, and the path length is defined in terms of the number of links. However, the temporal behavior of the links within the defined proximity has not been taken into account. It would be more effective if the impact of older links and more recent links were taken into account in defining local proximity, rather than considering only the path length. The probabilistic graph created using the structural features introduced in [13] has been used in [12] to estimate the probabilities of future links in a network. Here, the time stamps of the links were used to compute the difference in joining time for groups. However, the temporal behavior of the links in link prediction phase was not considered.

Besides machine learning and probabilistic approaches, other different approaches can be seen in the literature. Link prediction models were built using statistical relational learning and properties of relational data [22]. Relational markov network model was used in [28] to define joint probabilistic model over entire network. Then the model was used for link prediction in entire network. A matrix alignment method was used to determine the most predictive features of a link structure by aligning adjacency matrix of a network with weighted similarity matrices [26]. The weighted similarity matrices computed from node attributes and neighborhood topological features and the weights were learned by minimizing an objective function.

There are many time-evolving network models can be found in the literature. Epidemic models [17] are one of the popular time-dependent models which are generally designed to study the disease outbreaks over the human networks. Baràbasi-Albert model [34] is one of the best known generative network model which has been use to study the evolution of networks. Exponential random graph models [35] are widely use to estimate probabilistic models for small-world networks. However, most of the above network models are focused on a particular type of networks. In contrast to that, our method is adaptable for any type of networks because it can easily use with machine learning methods.

Some of the above worthy studies considered the temporal behaviors of the links in the networks but most others are not. For example, when matching semantic similarities, matching abstract keywords [24], would be more effective if higher weights were assigned to keywords in more recent publications. The random walk [14] would be more effective if the random walker were to choose its path according to not only the path weight but also using the link strength, which varies over time. Recently, the time-aware maximum entropy [29] was introduced in order to assign higher weights to more recent collaborations, as compared to older collaborations, in coauthorship networks. Although the impact of the time stamps on the temporality of the links was discussed, the impact of all coocurrences and time stamps of all interactions is not taken into account when assigning a score to a node. These observations inspired us to investigate the temporal behavior of the links. Therefore, we focused on finding a relationship between the time stamps of interactions or links and the temporal behaviors of the links and how this relationship affects future link evolution.

3. Supervised Learning Method for Link Prediction

As discussed above, link prediction deals with predicting future possible links in a given network. Most of the approaches discussed in Sect. 2 use structural features of networks and the features of the nodes and edges for link prediction. In coauthorship networks, the nodes are authors, and the edges are the publications by these authors, whereas in friendship networks, such as Facebook, the nodes are users, and the links are the relationships between users. In both cases, similarities between nodes and the structural features of the networks can be used to predict future links. For example, the number of common neighbors of a node pair and Jaccard's coefficient [15] can be computed. Once these features are calculated for a particular node pair, we have a vector of values referred to as a *feature vector* [21], which may be correlated with the future possible link between that node pair.

In a supervised learning approach, we use the feature vectors of each node pair to learn a model that can then be used to predict the appearance of future links. Once we com-

Table 1 Feature listing.

Feature	Formula	BC†	TSC††
Adamic/Adar	$\sum_{v_k \epsilon \Gamma(v_i) \cap \Gamma(v_j)} \frac{1}{log\lvert\Gamma(v_k)\rvert}$	✓	✓
Common neighbors	$\lvert\Gamma(v_i) \bigcap \Gamma(v_j)\rvert$	✓	✓
Jaccard's coefficient	$\frac{\lvert\Gamma(v_i) \bigcap \Gamma(v_j)\rvert}{\lvert\Gamma(v_i) \bigcup \Gamma(v_j)\rvert}$	✓	✓
Preferential attachment	$\lvert\Gamma(v_i)\rvert\lvert\Gamma(v_j)\rvert$	✓	✓
Time score	$\sum_n \frac{H_{m_n}\beta^{k_n}}{\lvert t_{1_n} - t_{2_n}\rvert + 1}$	-	✓

†**Baseline combination** ††**Time score combination**

pute the feature vectors for each node pair, we obtain a set of feature vectors for node pairs that are already linked and another set of feature vectors for node pairs that are not linked. The goal is to find a model that predicts unlinked node pairs that are likely to be linked in the future using feature vectors of already linked node pairs. To this end, we train the supervised machine learning method using the set of feature vectors to find unlinked node pairs which are likely to become linked in the future.

3.1 Features Used for Link Prediction

Table 1 lists the details of the features used in the present study. We used two different combinations of features in the proposed machine learning approach for link prediction. One set was used as the *baseline combination*, and the other set is the *time score combination*, which includes the *time score* introduced herein. The existing features are described below.

Adamic/Adar [1] This measure indicates that if a node pair has a common neighbor that is not common to several other nodes, then the similarity of that particular node pair is higher than that of node pairs having neighbors that are common to several other nodes. This measure assigns higher weights to common neighbors that are not common to several other nodes.

Common neighbors Number of common neighbors of a node pair.

Jaccard's coefficient [15] Normalized measure of common neighbors.

Preferential attachment [18] This measure indicates that new links are more likely to be formed with nodes of higher degree, or nodes that are popular in the network.

In the formulas in Table 1, v_i, v_j, and v_k denote nodes, and $\Gamma(v_i)$ and $\Gamma(v_j)$ denote the sets of neighbors of v_i and v_j, respectively. In Sect. 3.2, we discuss the new feature called *time score* introduced herein.

3.2 Time Score

We introduced a new feature to incorporate the effectiveness of common neighbors and their temporality. The features discussed in Sect. 3.1 are based solely on common neighbors, but do not consider the temporal behavior of the common neighbors. The strengths of links with common neighbors vary over time. In the context of social networks, the

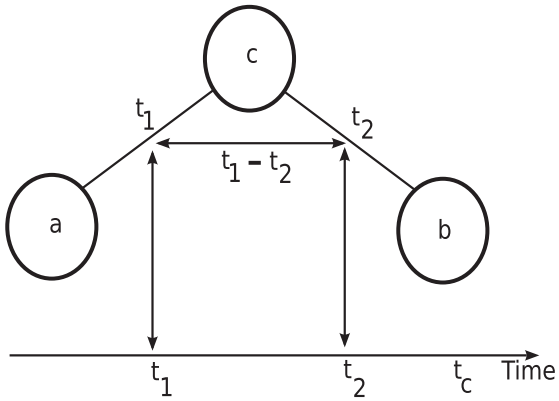


Fig. 1 Nodes a and b have common neighbor c. Here, t_1 is the most recent time stamp of the interactions between a and c, and t_2 is the most recent time stamp of the interactions between b and c. The current time is denoted as t_c.

effectiveness of the common neighbors depends not only on the cooccurrence frequency, or number of common neighbors, but also on how long the neighbors have been in contact. The time stamps of the interactions are useful in finding such information. This information provides a far better view of the importance of common neighbors than considering only the number of common neighbors. To this end, we designed a new feature based on the following concepts.

1. The strength of a link varies over time. If the nodes at the ends of a link have not interacted with each other for a long time with respect to the current time, then the link becomes weaker.
2. We assigned a higher score to node pairs which have interacted with their common neighbors within a closer proximity of time. In other words, if the difference between the time stamps of the most recent interactions of common neighbors having the node pair is small, then this difference has a greater effect on future links.

Combining the above considerations, we introduced a new feature, *time score (TS)*, to take into account the time awareness for link prediction. *Time score* for the node pair a and b that has n common neighbors is defined as follows:

$$TS(a,b) = \sum_n \frac{H_{m_n}\beta^{k_n}}{|t_{1_n} - t_{2_n}| + 1} \tag{1}$$

This concept is illustrated in Fig. 1. Nodes a and b have common neighbor c. Here, t_1 is the most recent time stamp of the interactions between a and c, and t_2 is the most recent time stamp of the interactions between b and c. In addition, H_m is the harmonic mean of the cooccurrence frequencies of a and b with the common neighbor c. β is a damping factor ($0 < \beta < 1$). k is the difference between current time t_c and the most recent time stamp from t_1 and t_2, and k is defined as follows:

$$k = t_c - max(t_1, t_2) \tag{2}$$

The number of interactions or cooccurrences, referred to as

link value, of a node pair is also important in determining the link strength. Therefore, we used the harmonic mean of the link values of each node in a node pair with their common neighbor. The harmonic mean, H_m, of numbers $x_1, \ldots, x_j$ is defined as follows:

$$H_m = \frac{1}{\frac{1}{j}\sum_{i=1}^{j}\frac{1}{x_i}} \tag{3}$$

Typically, the harmonic mean is appropriate for situations in which an average of rates is desired. In Eq. (3), x_i ($i = 1, \ldots, j$) denote the rates. In the present case, $j = 2$ because we used the link values of each node in a node pair with their common neighbors as the rates.

In Eq. (1), the term β^{k_n} increases as k_n decreases. We use the reciprocal of the term $|t_{1_n} - t_{2_n}| + 1$, where t_{1_n} and t_{2_n} are the time stamps of the most recent interactions of the node pair with the common neighbor. This term becomes larger when the difference between t_{1_n} and t_{2_n} becomes larger. The addition of one in the term is in order to avoid the *time score* from becoming infinite when the two time stamps are equal.

Compiling all, the new feature *time score* can be used as a feature, which is used for predicting future possible links. In order to show how to calculate *time score*, let us assume that two authors, a and b, have common neighbor author c. If a and c published two papers in 2005 and 2006 and authors b and c published one paper in 2008, then the harmonic mean of two publications and one publication is obtained as follows:

$$H_m = \frac{1}{\frac{1}{2}(\frac{1}{2} + \frac{1}{1})} = 1.3333 \tag{4}$$

If the current year is assumed to be 2011, then the *time score* for a future possible link between a and b can be calculated as follows:

$$TS(a, b) = \left(\frac{1.3333 * 0.5^3}{|2008 - 2006| + 1}\right) \approx 0.05555 \tag{5}$$

In this case, $k = 2011 - 2008 = 3$, because the latest time stamp is 2008, and the current year is 2011. The number of common neighbors, n, is 1, and we assume that $\beta = 0.5$.

4. Experimental Evaluation

In order to test the effectiveness of the proposed method, we performed two experiments using two real-world social network data sets. The first experiment tested the correlation between β and the unit of k. The purpose of this experiment was to provide guidelines for choosing values of the damping factor β, particularly for different time units k and different data. The second experiment tested the effectiveness of *time score* for link prediction.

We used two data sets in these experiments. The first data set was Facebook friendship network data from [30], which were collected from the regional Facebook network of New Orleans. The Facebook data was collected for 60,290 users who are connected by 1,545,686 links. We extracted a snapshot of the data from October 2007 to January 2009. The second data set is a coauthorship data set extracted from 66,791 publications on condensed matter physics from 1997 to 2005 in the *cond-mat archive* [†]. This data set contains data for 79,208 authors who are connected by 641,796 links.

In the experiments, we used J48 weka implementation [7] of C4.5 decision tree algorithm [23] and SMOT oversampling algorithm [2] with default parameters. Supervised machine learning algorithms required training data to train the learner. Therefore, we used user interactions (wall postings) within three consecutive months to predict the links of the following month because social networks such as Facebook show drastic changes within short periods of time. In order to predict the links that emerged during January 2009, we trained the decision tree algorithm using the data from September 2008 to December 2008. Features were computed using the network data from September 2008 to November 2008, and the links that emerged during December 2008 were considered to be the positive examples for training data. The trained model was applied for the features calculated for the data from October 2008 to December 2008 in order to predict the links that emerged during January 2009.

For the coauthorship data, we used data for three consecutive years to predict the links of following year, and the unit of k is years. For example, in order to predict the set of links that emerged in 2010, features for the training set were calculated using the coauthorship data from 2006 to 2008, and links that emerged in the year 2009 were considered to be positive examples for training data.

Tables 2 and 3 show the statistics of the two network data sets used in the experiments. The real-world networks we used for our experiments are very sparse, and so the rate of positive examples is very low. On average, the percentages of positive examples in the Facebook data and the coauthorship data were 0.05% and 0.08%, respectively. In order to solve this problem, we used the SMOT oversampling algorithm [2] in these experiments. After oversampling, the percentages of positive examples in the Facebook data and the coauthorship data were 0.3% and 0.5%, respectively.

4.1 Correlation between Damping Factor and Time Unit

In the first experiment, we test the correlation between β and the unit of k. We analyzed the precision, recall, and F-measure of the predictions for each data set by varying β between 0.1 and 0.9. The purpose of this analysis is to provide a guideline for selecting β according to the unit of k. The range of k depends on the time unit.

Figure 2 shows the variation of average precision, recall, and F-measure for each β value for Facebook data. The average precision, recall, and F-measure increase as β in-

[†]http://arxiv.org/archive/cond-mat/

Table 2 Statistics of the Facebook data.

Prediction month	Training data		Test data	
	Nodes	Edges	Nodes	Edges
2008 Feb	13,733	50,248	13,732	47,986
2008 Mar	13,732	47,986	13,998	48,238
2008 Apr	13,998	48,238	14,762	50,732
2008 May	14,762	50,732	15,705	56,014
2008 Jun	15,705	56,014	16,381	58,546
2008 Jul	16,381	58,546	17,268	60,718
2008 Aug	17,268	60,718	18,339	63,392
2008 Sep	18,339	63,392	20,476	71,792
2008 Oct	20,476	71,792	22,732	80,848
2008 Nov	22,732	80,848	25,427	92,990
2008 Dec	25,427	92,990	28,370	106,106
2009 Jan	28,370	106,106	31,832	123,650

Table 3 Statistics of coauthorship data.

Prediction year	Training data		Test data	
	Nodes	Edges	Nodes	Edges
2001	23,411	135,798	27,349	167,180
2002	27,349	167,180	31,662	209,632
2003	31,662	209,632	34,860	237,346
2004	34,860	237,346	38,039	266,236
2005	38,039	266,236	41,213	288,796

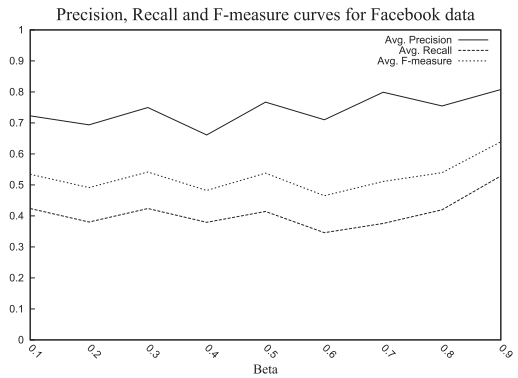


Fig. 2 Variation of performance metrics with β for Facebook data.

creases. A notable increase occurs at $\beta = 0.9$. We conducted a Grubbs' test to determine the significance of the difference between the F-measure at $\beta = 0.9$ and the F-measures at $\beta = 0.1$ to 0.8. The results of the Grubbs' test indicate that $\beta = 0.9$ is an outlier with a significance level of 5%. This indicates that the performance of the *time score* at $\beta = 0.9$ is significantly higher than for other β values and is thus a good parameter for Facebook data. The approximate range of k for the Facebook data is 0 to 90 days. Therefore, we can recommend higher β values as more appropriate when k takes a wide range of values.

Figure 3 shows the variation of average precision, recall, and F-measure for each β for coauthorship data. Better performance is obtained for lower β values, as indicated by the slight decrease in performance when β is greater than 0.5. The range of k is small ($0 \leq k \leq 2$) for the coauthorship data because we used the data of three consecutive years to predict links in the following year. The term β^k can take a higher value, even when β is small. Since the range of k is

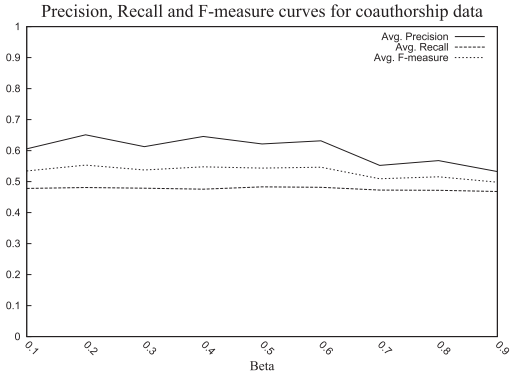


Fig. 3 Variation of performance metrics with β for coauthorship data.

small for coauthorship data, the term β^k takes closer values for higher β values, and *time score* becomes less effective for the learning algorithm. Therefore, lower β values should be used to compute *time score* when k has a small range.

The unit of k can be years, days, or hours, depending on the data set. We need to set β according to the unit of k in order assign higher scores to interactions that have occurred more recently. When k increases, β^k decreases. For higher values of k and lower values of β, the term β^k is approximately 0. For example, when $k = 10$ and $\beta = 0.5$, β^k is approximately 0.00098. In order to obtain a meaningful value for β^k when k has a wide range, we must use a higher β. In a network such as Facebook, interactions that occurred ten days ago have more of an effect on future links than interactions that occurred ten years ago in the coauthorship network. Therefore, higher β values are better when k has a wide range, and lower β values are better when k has a small range.

4.2 Effectiveness of Time Score

In this section we discuss the experiments carried out to test the effectiveness of *time score* for link prediction. We used the β values corresponding to the highest F-measure for each data set in the first experiment to compute *time score*. We compared the performance metrics for *baseline combination* (BC), which combines the existing features, and *time score combination* (TSC), which combines the new feature *time score* with the existing features.

4.2.1 Experiment Using Facebook Data

In the Facebook data, the frequency of the wall postings between users is considered to be the link value of each node pair that is already connected. Time stamps of the links are created using the time stamps of the wall postings so that the time stamp of a link represents the day of the most recent interaction between two users. We set β to be 0.9. The unit of k is days.

The performance metrics for Facebook data are compared in Fig. 4. The performance metrics show a notable improvement for *time score combination*, as compared to

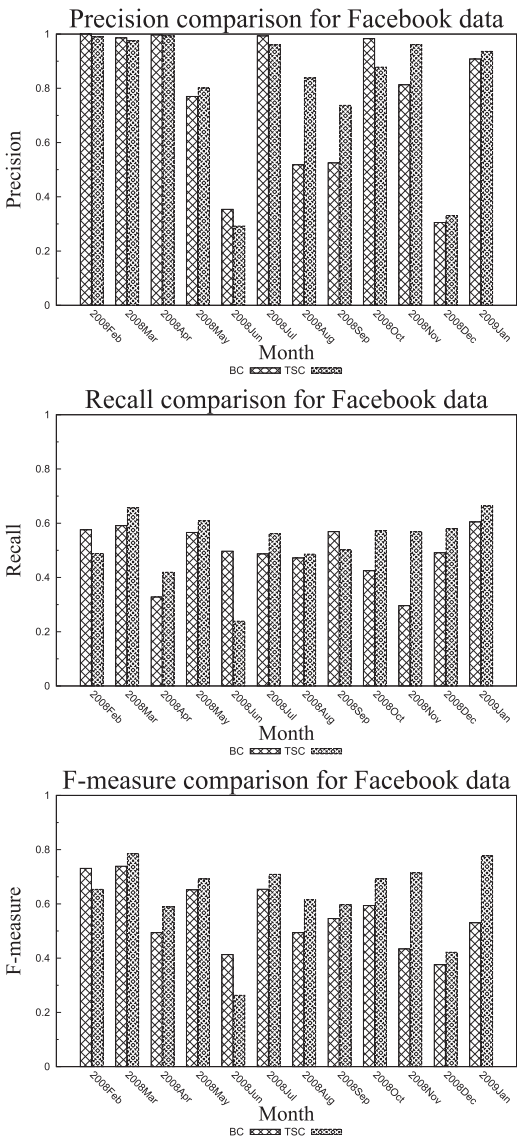


Fig. 4 Comparison of performance metrics for Facebook data.

the *baseline combination*. On average, the use of *time score* increased the precision, recall, and F-measure by 4%, 3%, and 7%, respectively.

According to the wall post data shown in Fig. 5, and as stated in [30], the number of wall posts increases rapidly from July 2008 to January 2009. This makes the network more active, and most of the existing links become stronger. The stronger links have a greater influence on future link evolution. Therefore, the use of *time score* is more effective and yields better results. This observation further indicates that *time score* is more sensitive to the temporal behavior of user interactions. However, in February 2008 and June 2008, there is a decrease in the number of wall posts. Thus, the network becomes less active, and the strengths of the links do not exhibit temporal variations in behavior in the network during this period. Therefore, the performance metrics exhibit slightly lower values for *time score combination* than for *baseline combination*. Except for the results of February 2008 and June 2008, the t-test at the 5% significance level indicates significant improvements. Therefore, we can conclude that *time score* is more effective for rapidly evolving networks.

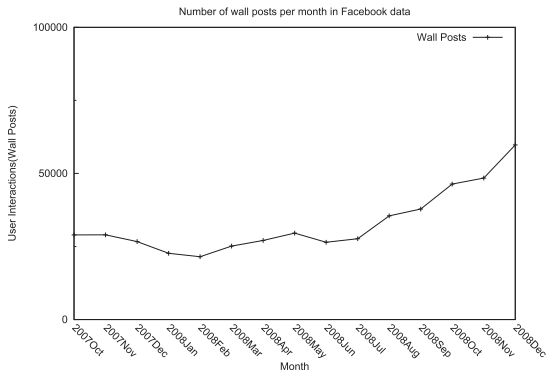


Fig. 5 Variation of the number of wall posts in Facebook data.

4.2.2 Experiment Using Coauthorship Data

In the coauthorship data set, the unit of k is years. The time stamp for the interaction between a pair of authors represents the year of publication of the coauthored paper. Hence, the time stamp of a link represents the year of most recent publication by a pair of authors. A damping factor of $\beta = 0.2$ was used in this experiment.

The performance metrics of this experiment are compared in Fig. 6. The improvements in precision, recall, and F-measure indicate the impact of *time score* for link prediction in the coauthorship network that evolves primarily over recent collaborations. In the graph comparing precision, with the exception of 2001, the results obtained using *time score combination* are better than the results obtained using *baseline combination*. All three performance metrics indicate significant improvements according to the t-test at the 5% significance level. The average improvements in precision, recall, and F-measure are 14%, 11%, and 13%, respectively.

5. Discussion

In the Facebook friendship network, the friends of a user can view the wall posts of that user if the user shares the wall posts with his/her friends. Thus, users who have that particular user as a common neighbor, while having no other relationship, can become friends through each other's postings. Burst of wall postings indicates that more people are interacting with each other and become friends. Therefore, recent interactions happen in closer proximity of time have a greater influence on link evolution. Besides the factors we investigated in our experiments, the link evolution could be depend on other temporal factors such as duration of data collection and geographical region of the network. In particular, the Facebook network exhibits different patterns de-

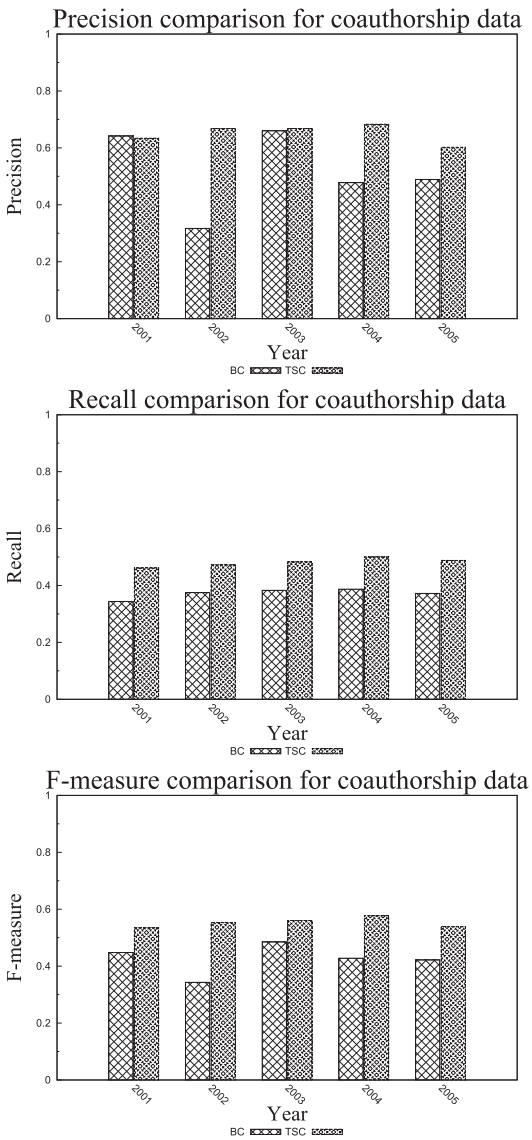


Fig. 6 Comparisons of performance metrics for coauthorship data.

pending on the time, the major events that occur during the period of data collection, and the geographical region of the network. Such kind of factors are to be explored in our future works.

Previous studies have extensively investigated the evolution of scientific collaboration networks using network statistics [19], [20]. However, scientific collaborations are time-sensitive. Researchers prefer to explore evolving topics through new collaborations. To this end, researchers tend to find associates or experts through their most recent collaborations. This increases the temporality of the links among the researchers. On the other hand, the temporality in coauthorship networks has several causes. For example, researchers tend to change research fields according to current research trends and occasionally change institutes or universities. In such situations, the geographical locations of the researchers and current research trends become important factors in predicting links in coauthorship networks.

6. Conclusion and Future Research

The main contribution of the present study is the introduction of a new time-aware feature, called *time score*, for link prediction in social networks using supervised machine learning methods. The primary focus of the present study was the impact of the relationship between the temporal behavior of link strength and the time stamps of interactions and links for link evolution, which had not previously been discussed sufficiently. We found that the time stamps of interactions are crucial factors for link evolution. In particular, we focused on the temporal behavior of common neighbors in terms of link strength. We examined the proposed method using two real-world data sets. The improvements in performance metrics indicated by the experimental results verify the effectiveness of *time score* for link prediction in social networks. Therefore, we can obtain better predictions using the newly proposed *time score* feature.

The present study was limited to node pairs having common neighbors. In the future, we intend to extend the proposed method to any node pair in a network. Exploring other factors of temporal behaviors of networks is one of the primary goals of our future research. Some of these factors are network specific. Therefore, the use of temporal behaviors for link prediction is a challenging task. Furthermore, we intend to demonstrate that the proposed method is applicable to a wide range of algorithms that have been used for link prediction, such as flow-based algorithms and statistical modeling approaches.

References

[1] L.A. Adamic and E. Adar, "Friends and neighbors on the web," Social Networks, vol.25, pp.211–230, 2003.
[2] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," J. Artificial Intelligence Research, vol.16, pp.321–357, 2002.
[3] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol.20, pp.273–297, 1995.
[4] S. Lin and H. Chalupsky, "Unsupervised link discovery in multi-relational data via rarity analysis," Proc. 3rd IEEE International Conference on Data Mining, pp.171–178, 2003.
[5] L. Getoor, "Link mining: A new data mining challenge," SIGKDD Explor., vol.5, no.1, pp.84–89, 2003.
[6] L. Getoor and C.P. Diehl, "Link mining: A survey," SIGKDD Explor., vol.7, no.2, pp.3–12, 2005.
[7] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten, "The weka data mining software: An update," SIGKDD Explor. Newsl., vol.11, no.1, pp.10–18, 2009.
[8] M.A. Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link prediction using supervised learning," Proc. SDM 06 Workshop on Link Analysis, Counterterrorism and Security, 2006.
[9] Z. Huang, X. Li, and H. Chen, "Link prediction approach to collaborative filtering," Proc. 5th ACM/IEEE-CS Joint Conference on Digital Libraries, pp.141–142, 2005.
[10] H. Kashima and N. Abe, "A parameterized probabilistic model of network evolution for supervised link prediction," Proc. 6th International Conference on Data Mining, pp.340–349, 2006.
[11] H. Kashima, T. Kato, Y. Yamanishi, M. Sugiyama, and K. Tsuda,

"Link propagation: A fast semi-supervised learning algorithm for link prediction," Proc. Sesure Data Management, pp.1099–1110, 2009.

[12] V. Leroy, B.B. Cambazoglu, and F. Bonchi, "Cold start link prediction," Proc. 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.393–402, 2010.

[13] D. Liben-Nowell and J. Kleinberg, "The link prediction problem for social networks," Proc. 12th International Conference on Information and Knowledge Management, pp.556–559, 2003.

[14] R.N. Lichtenwalter, J.T. Lussier, and N.V. Chawla, "New perspectives and methods in link prediction," Proc. 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.243–252, 2010.

[15] C.D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.

[16] L. Munasinghe and R. Ichise, "Time aware index for link prediction in social networks," Proc. 13th International Conference on Data Warehousing and Knowledge Discovery, pp.342–353, LNCS 6862, 2011.

[17] M.E.J. Newman, Networks: An Introduction, Oxford University Press, 2011.

[18] M.E.J. Newman, "Clustering and preferential attachment in growing networks," Phys. Rev. E, vol.64, no.2, 025102, July 2001.

[19] M.E.J. Newman, "Scientific collaboration networks. II, shortest paths, weighted networks, and centrality," Phys. Rev. E, vol.64, no.1, 016132, June 2001.

[20] M.E.J. Newman, "The structure of scientific collaboration networks," Proc. National Academy of Sciences of the United States of America, vol.98, no.2, pp.404–409, Jan. 2001.

[21] M. Pavlov and R. Ichise, "Finding experts by link prediction in coauthorship networks," Proc. Workshop on Finding Experts on the Web with Semantics, pp.42–55, Nov. 2007.

[22] A. Popescul, R. Popescul, and L.H. Ungar, "Statistical relational learning for link prediction," Inf. Sciences, pp.149–172, 2003.

[23] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, 1993.

[24] M. Sachan and R. Ichise, "Using abstract information and community alignment information for link prediction," Proc. 2nd International Conference on Machine Learning and Computing, pp.61–65, 2010.

[25] R. Schifanella, A. Barrat, C. Cattuto, B. Markines, and F. Menczer, "Folks in folksonomies: Social link prediction from shared metadata," Proc. 3rd ACM International Conference on Web Search and Data Mining, 2010.

[26] J. Scripps, P.-N. Tan, F. Chen, and A.-H. Esfahanian, "A matrix alignment approach for link prediction," Proc. 19th International Conference on Pattern Recognition, pp.1–4, 2008.

[27] T.E. Senator, "Link mining applications: Progress and challenges," SIGKDD Explor. Newsl., vol.7, no.2, pp.76–83, 2005.

[28] B. Taskar, M.-F. Wong, P. Abbeel, and D. Koller, "Label and Link prediction in relational data," Working Notes of the IJCAI-2003 Workshop on Learning Statistical Models from Relational Data, pp.145–152, 2003.

[29] T. Tylenda, R. Angelova, and S. Bedathur, "Towards time-aware link prediction in evolving social networks," Proc. 3rd Workshop on Social Network Mining and Analysis, pp.1–10, 2009.

[30] B. Viswanath, A. Mislove, M. Cha, and K.P. Gummadi, "On the evolution of user interaction in facebook," Proc. 2nd ACM SIGCOMM Workshop on Social Networks, Aug. 2009.

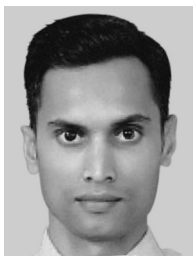
[31] C. Wang, V. Satuluri, and S. Parthasarathy, "Local probabilistic models for link prediction," Proc. 7th IEEE International Conference on Data Mining, pp.322–331, 2007.

[32] T. Wohlfarth and R. Ichise, "Semantic and event-based approach for link prediction," Proc. 7th International Conference on Practical Aspects of Knowledge Management, pp.50–61, 2008.

[33] E. Zheleva, L. Getoor, J. Golbeck, and U. Kuter, "Using friendship ties and family circles for link prediction," Proc. 2nd ACM SIGKDD Workshop on Social Network Mining and Analysis, 2008.

[34] A.-L. Bàrabasi and R. Albert, "Emergence of scaling in random networks," Science, vol.286, pp.509–512, 1999.

[35] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, "An introduction to exponential random graph (p^*) models for social networks," Social Networks, vol.29, no.2, pp.173–191, 2007.



Lankeshwara Munasinghe received his B.Sc. degree in Statistics and Computing from University of Kelaniya, Srilanka in 2002. He is currently a Ph.D. candidate at the National Institute of Informatics in Japan. His research interests include machine learning and data mining.

Ryutaro Ichise received his Ph.D. degree in computer science from Tokyo Institute of Technology, Tokyo, Japan, in 2000. From 2001 to 2002, he was a visiting scholar at Stanford University. He is currently an associate professor in the Principles of Informatics Research Division at the National Institute of Informatics in Japan. His research interests include machine learning, semantic web, and data mining.