Link Prediction Using Kalman Filter

Ajay Kumar Singh Kushwah[1], Prof. Amit Kumar Manjhvar[2]
Department of CSE/IT
[1, 2] Madhav Institute of Technology and science Gwalior, M.P.,India

Abstract- *Link prediction in social networks refers to predicting the manifestation of future relations between nodes. It is measured one of the significant tasks in a variety of data mining applications for recommendation systems, bioinformatics, world wide web and it has concerned a great pact of consideration recently. There are several studies on link prediction based on stationary topological likeness metrics and stationary graph representation without taking into account the temporal growth of link occurrences. Most of the previous methods for link prediction in developing networks use the exisiting associations in the network to predict new ones. Here, I have proposed a novel method, called Multivariate Time Series Link Prediction, for link prediction in evolving networks using Kalman Filtering based approach that integrates (1) temporal evolution of the network (2) node similarities (3) node connectivity information. The proposed method is based on a Kalman filter prediction Model for Multivariate Time Series forecasting which enables to represent time information over a combination of node similarities and node connectivities. The projected method is experienced on co-authorship networks. It is shown that integrating time information with node similarities and node connectivities improves the link prediction performance to a large extent.*

Keywords *- Data Mining, Link Mining, Link Prediction, Kalman Prediction, Kalman Filter.*

I. INTRODUCTION

A social network is defined as a linked social structure which is composed of individuals, groups or organizations also called nodes and relations (edges) among the nodes. Social network is a general way to represent the exchanges among the people in a grouping or society and social network analysis has gained significant attention in the last few years. Social networks are dynamic structures and they evolve over time through the addition and removal of nodes and edges. Thus, considerate the mechanisms that utilize the evolution of social network is an essential task and it forms the motivation of our work.

As the increase in online social network and information platform, people show more interest in complicated network. Despite that there are already numerous studies on snapshots of single networks, but these studies lack dynamic information of evolving network [1-2]. Extracting dynamic information which changes over time by analyzing the increase or decrease in amount of nodes in network is necessary to solve the problem of link prediction.

With the rapid growth of internet expertise, the amount of information in general networks increases significantly. While accessing valuable information from social networks has become more and more difficult [3]. Social networks contain large number of potential useful information that is valuable for people's daily lives and social business [4]. Therefore, social network analysis (SNA) has become a research focus to mine underlying useful information from massive social network data. As part of this research, how to accurately predict a potential link in a real network is an important and challenging problem in many domains, such as recommender systems, decision making and criminal investigations. For example, we can predict a potential relationship between two persons to recommend new relationships in the Facebook network. In general, we call the above problem as link prediction [5].

As a subset of link mining [6], link prediction aims to compute the existence probabilities of the missing or future links among vertices in a network [7, 8]. There are two main difficulties in the link prediction problem: (1) huge amount of data, which requires the prediction approaches to have low complexity. (2) prediction accuracy, which requires the prediction approaches to have high prediction accuracy. However, traditional data mining approaches can't solve the link prediction problem well because they do not consider the relationships between entities, but the links between entities in a social network are interrelated.

To overcome the above two difficulties and meet the practical requirements, many similarity-based methods have been proposed.

In this paper, we propose a new prediction technique in link mining. They consider the different roles of nodes and upon application remarkably outperform the existing Methods.

1.1 Contributions

The contributions of this paper consist of the following three aspects:

(1) We propose a kalman prediction based approach for multivariate temporal prediction.
(2) The measurement factors are considered so that the approach proves correct towards the final results.
(3) Makes the prediction better by using the kalman prediction.
(3) We use the adamic/adar method for finding the prediction in a better way.
(4) We use the clustering information that is important information for predicting links, which can improve the prediction accuracy.

Experimental evaluation demonstrates our approaches outperform other methods in terms of accuracy and complexity. Our approaches are very suitable for large-scale sparse networks.

In this paper, we propose a novel method for prediction of new link formations and repeated occurrences of existing links which utilizes the network temporal information along with modeling the combination of topological metrics and link occurrences information. Multivariate Time Series based method is employed in order to represent time information and predict future new connections between nodes in a social network. The novelty of our approach lies in the method by which we apply Multivariate Time Series model to combine topological metrics and temporal evolutions of link occurrences information simultaneously which enables to predict repeated link occurrences as well as to identify new link occurrences. The proposed model is one of the most successful models for the analysis of multivariate time series. It is an extension of the univariate autoregressive model to multivariate time series and it often provides better forecasts than univariate time series models. The model has been shown to be mainly applicable for describing the dynamic behavior of time series and for forecasting. In order to assess the effectiveness of the proposed method, we performed experiments on the DBLP bibliographic co-authorship networks.

In general, the experiments showed that our Multivariate Time Series approach outperforms both the static approach and the univariate time series techniques aforementioned.

1.2. Organization

The rest of this paper is organized as follows: Section II. provides the overview of the related works of link prediction. Some preliminaries are briefly introduced in Section III. Section IV. presents the idea of our approaches, and gives their complexity analysis. Experimental study is presented in Section V. Section VI. concludes this paper and the future work.

II. RELATED WORK

The link prediction is the problem of estimating the chance or probability of the future happening of a connection in a network. There are different techniques which have been useful in the domain of link prediction problem. These procedures can be classified keen on two substantial types such as (1) static topological/structural techniques and (2) temporal techniques. The static structural techniques consider local features which are limited to the direct neighbors of nodes and global attributes which cover the entire network. In the structural techniques, the couples of non-connected nodes are ranked according to a selected metric such as the number of common neighbours [10], Adamic/Adar [11], preferential attachment [12], Katz coefficient [13], resource allocation [14], Jaccard's coefficient [15], and so on. Then, the peak ranked non-coupled nodes are predicted to be correlated in the future. Liben-Nowell and Kleinberg (2003) planned one of the most primitive link prediction models for social networks [1]. They used the ranking on the similarity scores to predict the link between non-connected nodes. Then, Hasan et. al. (2006) extended this work and used different similarity metrics as features for several supervised learning techniques [3]. In these approaches, link prediction is considered as a binary classification task. Lichtenwalter et. al. (2010) introduced a link prediction method called PropFlow in social networks. It corresponds to the probability that a restricted random walk ends in l steps or fewer and uses the link weights as transition probabilities. PropFlow method selects the links based on their weights and estimates the likelihood of new links [16]. Most of the prior work on link prediction relies on using static structural techniques whereas there are few studies examining the role of temporal data in link prediction task. The dynamic structure of social networks and the effects of this structure on link prediction task indicate that the time information should be used for obtaining better results [17]. Basically, the temporal approach uses time based structural information such as the link building time. Tylenda et. al. (2009) proposed a method for time-aware Link Prediction. They extended the local probabilistic model to comprise period awareness [5]. Munasinghe and Ichise (2012) introduced a new period aware key which is called Time Score to use the link building time and temporality of link strengths [18]. Time series models are also employed to forecast connection occurrence probabilities at a scrupulous time taking into contemplation sequential evolutions of connection occurrences. Huang and Lin (2009) proposed an AutoRegressive Integrated Moving Average (ARIMA) model for link prediction in which a time sequence for each couple of nodes is built by computing the frequency

of the occurrence of links between the nodes using dissimilar pictures of the network along time. Then, the time sequence forecasts modeled by ARIMA were used to estimate the likelihood of future link occurrences [7]. Soares and Prudencio (2012) also proposed a univariate time series forecasting method for link prediction problem. They applied forecasting models such as Moving Average, Random Walk, Linear Regression and Simple Exponential Smoothing. They divide the network into several times sliced snapshots that indicated the network configuration at different times in the past and calculated the similarity scores based on dissimilar similarity metrics for each brace of non-connected nodes. Then, they built univariate time series models for each brace by using the similarity scores in different snapshots and obtained the next proximity value. These values are used as an input to any of the supervised or unsupervised learning algorithms [8]. Most previous studies using temporal network data focus on the basic univariate time series model to estimate the likelihood of future link occurrences. However, the limitation of such approaches is that they consider only the past occurrence of a link and cannot predict the occurence of a link that has not observed previously. The main contribution of this article is the use of Multivariate Time Series models that can integrate covariance structures which exploit the correlations among similarity metrics and link occurrences information simultaneously for accurate prediction of both new links and repeated links.

BACKGROUND

In this section, we describe the topological metrics for weighted and unweighted networks and the forecasting models adopted in our work. In this study, we consider undirected graphs and w(x, y) denotes the link weight between nodes xand y.

III. SIMILARITY METRICS

A. Local Similarity Indices

Common neighbors

Common neighbor is a technique based on node neighborhood. The extent of common neighborhood of two nodes x and y can be defined as

S_xy^CN= ⟦| ┌⟧_x∩ ┌_(y)|(1)

Equation (1) represents the number of neighbors that x and you have in common. This technique is based on the intuition that if there is a node that is connected to x as well as you, then there is a high probability that vertex x be connected to vertex y. Thus, as the number of common neighbors grow up higher, the possibility that x and y have associations between them increases. Kossinets and Watts [16] job to consider a large-scale social network resembling to Facebook. In their work, they suggest that two individuals having many common friends are very probable to be friend in the future.

B. Salton Index

Salton index [17] is defined as

$$S_{xy}^{Salton} = \frac{\Gamma_x \cap \Gamma_y}{\{\sqrt{Kx} * \sqrt{Ky}\}} \quad (2)$$

Where & k_y represent the degree of node x and node y. This index is also called the cosine similarity.

a. Sorensen Index

Sorensen Index is defined as

$$S_{xy}^{Sorensen} = 2\,\frac{|\Gamma_x \cap \Gamma_y|}{kx + ky} \quad (3)$$

This index [18] is used primarily for biological community data.

d. Hub Promoted Index (HPI)

This index is offered for enumerating the topological overlap of pairs of substrates in metabolic networks, and is defined as

$$S_{xy}^{HPI} = \frac{|\Gamma_x \cap \Gamma_y|}{min\{kx, ky\}} \quad (4)$$

In this measurement, the links adjacent to hubs are likely to be assigned high scores since the denominator is decided by the lower degree only.

e.Hub Depressed Index (HDI)

Similar to the HPI, the HDI also considers a measurement with the opposite effect on hubs. It is defined as

$$S_{xy}^{HDI} = \frac{|\Gamma_x \cap \Gamma_y|}{max\{kx, ky\}} \quad (5)$$

f. Leicht-Holme-Newman Index (LHN1)

This index consigns high similarity to node couples that have many common neighbors associated not to the possible maximum, but to the expected number of such neighbors**.** It is defined as

$$S_{xy}^{LHN1} = \frac{\Gamma_x \cap \Gamma_y}{\{kx * ky\}} \quad (6)$$

Where the denominator $kx * ky$ is proportional to the likely number of common neighbors of nodes x and y in the configuration model [19].

g. Jaccard's Coefficient

Paul Jaccard introduces Jaccard coefficient over hundred years ago, which isbasically used to determine the association between two words. The Jaccard coefficient [20] is also

known as the Jaccard similarity coefficient. Jaccard index is a name frequentlyrecycled for comparing distance, similarity and dissimilarity of the data set. To measure the Jaccard similarity coefficient between two data sets is defined as

$$J(x,y) = \frac{\Gamma_x \cap \Gamma_y}{\{\tau(x) * \tau(y)\}} \quad (7)$$

Jaccard distance is non-similar measurement between data sets. It can be resolute by the converse of the Jaccardcoefficient, which is obtained by removing the Jaccard similarity from (7). It is equal to a number of features that are all, minus by a number of features that are common to all divided by the number of features as presented below.

$$j\delta(A,B) = 1 - Jx,y \quad (8)$$

This is the similarity of a symmetric binary attributes.

h. Adamic Adar

N by N similarity matrix that contains the Adamic Adar similarity between every two nodes in the data sets. This technique was firstly proposed for the metric of similarity between two web pages. It calculates the likelihood when two particular homepages are strongly connected. It computes features that are common among nodes and then describes the similarity among them. For this major the features of the pages are calculated and then the similarities are defined.

$$Score\ (x,y) = \sum_{z \in\ \Gamma(x) \cap \Gamma(y)} \frac{1}{\log|\Gamma(z)|} \quad (9)$$

i. Shortest Path

In graph theory, the shortest path problem is the problem of finding a route between two vertices or nodes in a graph such that the sum of the weights of its constituent edges is reduced. The shortest route problem can be defined for graphs whether these are directed, undirected or mixed.

$$S_{xy}^{SP} = -dx,y \quad (10)$$

j. Clustering

One might look forward to improve on the quality of a predictor by deleting the more unconvincing edges in $Gcollab$through a clustering procedure, and then running the predictor on the resulting cleaned-up sub-graph. Consider a measure, calculating values for Score (x,y) for all pairs (x, y) on this sub-group.In this way we determine node proximities using only edges for which the proximity measure itself has the most self-assurance.

B. Global Similarity Indices

a. Katz Index

Katz index [21] is based on the joint of all paths, which straight sums over the collection of paths and is exponentially damped by length to provide the shorter paths further weights. The scientific expression is defined as

$$S_{xy}^{Katz} = \sum_{l=1}^{\infty} \beta^l * \left|paths_{xy}^{<l>}\right| \quad (11)$$

Where $paths_{xy}^{<l>}$Is the set of all paths with length l connecting x and y and β is a unrestrictedconstraint (i.e.,the damping factor) monitoring the path weights. A very small β yields a measurement close to common neighbor, because the long paths contribute very little. The similarity environment can be defined as

$$S_{xy}^{Katz} = I - \beta A - 1 - l \quad (12)$$

Here β must be lower than the reciprocal of the largest Eigen value of the matrix A to ensure the convergence of Eq. 12.

b. Simrank

If two neighbors are so secure to each other that they should be joined by an edge**.** Numerically, this is specified by defining similarity (x, x) =1 and $similarity\ (x.y) = \gamma * \frac{\sum_{a \in \Gamma(x)} \sum_{b \in \Gamma(y)} similarity\ \ (a,b)}{|\Gamma(x)||\Gamma(y)|}$ (13)

For some γ∈ [0,1] is the decay factor. They finally stated Score (x, y) = similarity (x, y). Simrank can also be interpreted by the random walk on the collaboration graph.

MULTIVARIATE TIME SERIES LINK PREDICTION

In this paper, we study the problem of both new and repeated link prediction in evolving social networks by exploring the evolution of topological metrics and link occurrences. First of all, we divide our temporal network dataset into time-sliced snapshots from time t0 to tk+1. Each snapshot represents the state of the network at different timesteps as illustrated in Figure 1. The example illustrated in Figure 1 summarizes the consecutive static snapshots from time t0 to tk+1 which is denoted by [Gt0 ,Gt1 ,…..Gtk+1] as a sequence of graphs.

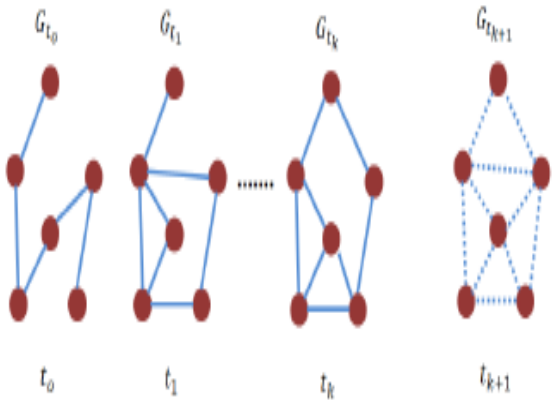


Fig. 1: Evolving Network example

as a graph Gtk = (V;Etk), with a set of nodes V and a set of edges Etk . The graphs at time points tk and tk+1 have the same set of nodes but different set of links. Letting j V j be the number of nodes in the network, the graph Gtk can be represented by a |V|* |V| adjacency matrix denoted by Atk . If two nodes i and j are connected, then the value of A(i, j) = 1; otherwise, it is 0 for the unweighted network.

In this paper, A(i, j) = wtk (i, j) weight values indicate the number of papers co-authored by author i and j at time tk.

The basic univariate time series link prediction problem is defined as to predict the probability of link occurrence at time tk+1 according to the exact probability of occurrence for each link in the given consecutive graph snapshots at instants t0, …,tk [14]. However, the limitation of this approach is that it considers only the past occurrence of a link and cannot predict the occurence of a link that has not observed previously. In this paper, a method called Multivariate Time Series Link Prediction is developed to exploit the correlations among similarity metrics and link occurrences information simultaneously using multivariate time series analysis across different past time periods. Then, a multivariate time series model is built for each pair of nodes in consecutive snapshots. Finally, the combination of similarity scores and the link occurrences information for each pair are used as an input to the model and future score values are estimated.

VAR Models for Multivariate Time Series Link Prediction In this section, we introduce the Vector AutoRegression (VAR) model that integrates multiple parameters simultaneously. The VAR model is one of the most effective approaches for the analysis of multivariate time series. It is an extension of the univariate autoregressive model to the multivariate case. The VAR model has been shown to be appropriate mainly for describing the dynamic behavior of time series and for forecasting. It takes into account the correlations among all parameters simultaneously which results in better forecasts to those obtained by using univariate time series models [9]. 1) Vector Autoregression of Order p: Let Yt(t = 1,…. T) be a multivariate time series with T observations, the pth order vector autoregression, written as VAR(p), be a process that evolves as shown in Eq. 1:

$$Yt^{\wedge} = C + \Pi^{1} Y_{t-1} + \Pi^{2} Y_{t-2} + :::+ \Pi^{p}{}_{t-p} + \epsilon_t, \quad t = 1.....T \quad (1)$$

In this equation, $Yt^{\wedge} = (y1t, y2t.....ynt)T$ is a vector of dimension n consisting of the estimated values of the variables in hand at time t; C is a vector of dimension n of intercepts; Πj , j=1,...,p are nXn coefficient matrices; and ϵt is a vector with n dimensions of errors following a multivariate white noise process which has zero mean, constant variance, and finite covariance and it is uncorrelated with its past values [20], and n denotes the number of variables.

Akaike Information Criterion (AIC) is used to compute the model parameters and the lag length for the VAR(p) model is determined by using AIC. The lag values of a time series variable are the values of the specified variables occuring prior to the current observation

IV. PROPOSED APPROACH

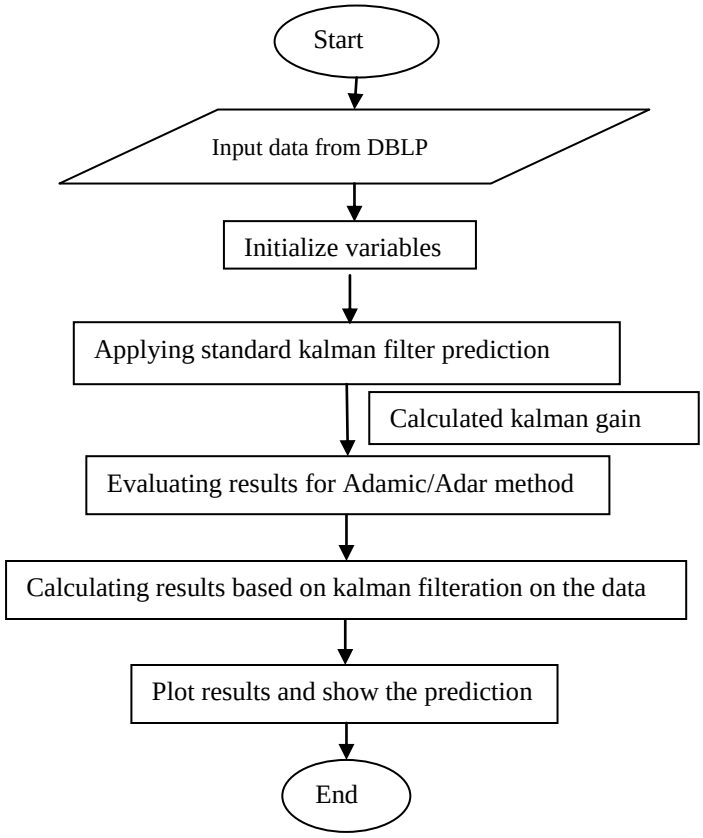
Kalman Filter Prediction

Kalman filter is an algorithm allowing accurate inference in a linear dynamical system, which is a Bayesian model similar to a hidden Markov model but where the position space of the underlying variables is continuous and where all underlying and observed variables have a Gaussian distribution(often a multivariate Gaussiandistribution).

The Kalman filter [15] (and its variants such as the extended Kalman filter [16] and unscented Kalman filter [17]) is one of the most eminent and popular data union algorithms in the field of information dispensation.

The Kalman filter is usually consequent with vector algebra as a least mean squared estimator [18], an approach appropriate for students positive in mathematics but not one that is easy to grasp for students in disciplines that do not require strong mathematics. The Kalman filter is derived here from first philosophy considering an easy physical paradigm exploiting a key property of the Gaussian distribution—specially the property that the artifact of two Gaussian divisions is another Gaussian division.

PROPOSED WORK FLOW CHART



Proposed Pseudo-code

1. Initialize m=1, n=160, mlen = 30, n=15.
2. Define xmin and xmax as output of z.
3. [x_kf KF] = StandardKalmanFilter(z,MAlen,N)
   i. Calculate x_apriori, covariance.
   ii. Kalman_gain = Zeros(DIM,DIM,n)
   iii. Update apriori estimate = AACovEst(z,smoothed,I,N,R,Q)
   iv. AACovEst(z,smoothed,I,N,R,Q)
      a. [R Q] = StandardCovEst(z,smoothed_z,i,N)
      b. smoothed_z = smoothed_z(i-(N-1):i)
      c. Q(i,j) = sum((smoothed_z(i,:)-mean1i).*(smoothed_z(j,:)-mean1j))/(N-1)
      d. Qnew = smoothed_z(i) - mean(smoothed_z((i-(N-1)):i))
   v. Compute Kalman Gain = KalmanGainCalc(P_apriori(:,:,i),R(:,:,i))
   vi. Update aposteriori state estimate
   vii. Update aposteriori error covariance estimate
4. result=flipud(result).
5. AUC_KF = 0;
6. for i=1:13
   AUC_KF = AUC_KF + result(i);
   end
7. AUC_KF = AUC_KF/(0.5*max(result))*rand
8. p1 = plot(x((MAlen+N-1):end)-2,z(i,(MAlen+N-1):end)-2,'r+')
9. p2=plot(result,aa,'g')
10. p3 = plot(x((MAlen+N-1):end)-2,x_kf(i,(MAlen+N-1):end)-2,'b')

V. DATA AND METHODS

We report all results on a single, relatively small publicly available longitudinal data set. The data set, to which we will henceforth refer as condmat, is constructed by moving through a sequence of collaboration events in the condensed matter physics community. Each association of k individuals forms an undirected k-clique with weights in converse linear proportion to k. The system is thus weighted and undirected. We illustrate our points using results for the usefulness of only three prediction methods, but each method represents a different modeling approach. The preferential attachment predictor [19] uses measure product and represents forecasters on the basis of node statistics. The Adamic/Adar forecaster [20] signifys the family of common neighbors predictors. The PropFlow forecaster represents the family of predictors based on paths and random walks.

We emphasize here that the point of this work is not to illustrate the superiority of one method of link prediction over another. It is instead to point out that

the described effects and arguments have real impacts on performance and evaluation.

If we show that the effects pertain in at least one network, then they may exist in others and must be considered. We also selected condmat because its small size makes exploring some of these problems computationally easier. Larger and sparser networks will typically exhibit the patterns we describe to a greater degree rather than to a lesser degree.

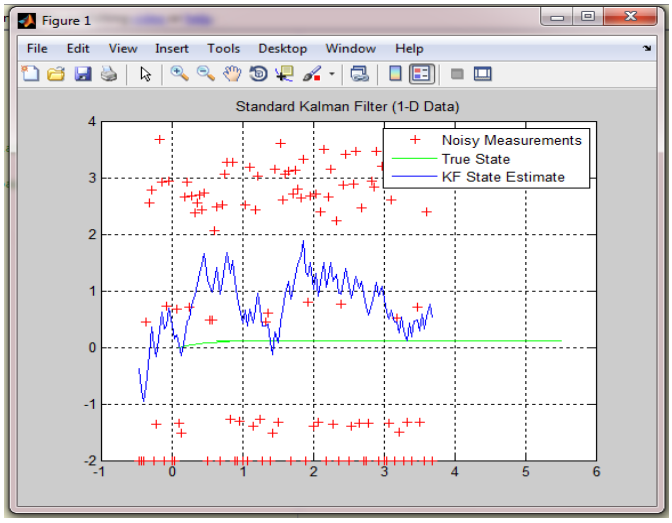
A.Dataset

The DBLP bibliographic database of computer science articles is used in the experiments1. The DBLP dataset presents a wide ranging directory of research papers in computer science. The complete dataset contains 1; 086; 418 papers, 960; 669 authors, 23; 643 conferences, and 79; 002 citations. This dataset has been widely used in many studies on modeling of social networks and co-authorship link prediction studies [21].

B. RESULT ANALYSIS

In this section, we first describe the univariate time series models adopted in the link prediction task to compare with our proposed model in the experiments. We compare our model with alternative model which have been used in several works. To evaluate the link prediction performance, we use Area Under Curve (AUC) measure since it is a significant performance measure which has been used in imbalanced classification problems such as link prediction.

The AUC was computed for each combination of similarity metrics and forecasting models. Table I&II show the AUC values, for repeated and new link prediction task on unweighted and weighted networks respectively. Our results showed that the proposed multivariate time series method outperforms other univariate time series models and static graph methods by 5-10% in terms of AUC.



AUC measures for AA method comes out to be:

TABLE I: Repeated-New Link Prediction Task AUC Measures on unweighted DBLP Data

	AA
VAR	0.7516
KF	0.8289

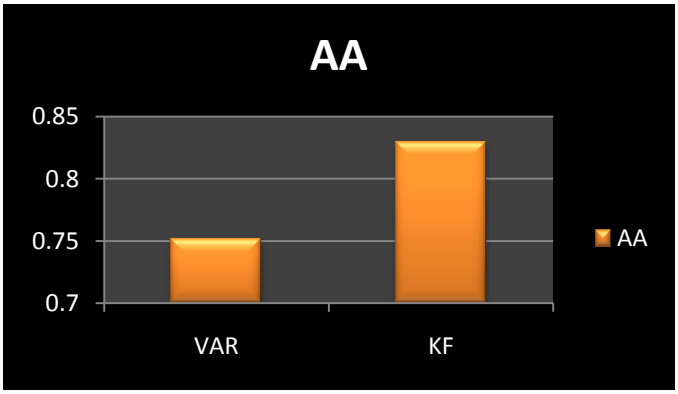
TABLE II: Repeated-New Link Prediction Task AUC Measures on weighted DBLP Data

	AA
VAR	0.90
KF	0.95

Experimental results showed that the proposed KF model significantly outperformed the VAR model. The KF model utilizing link occurrences information and each different similarity metric was the best forecasting model for all metrics.

The results for the new link prediction task indicate that the KF model had the overall best performance where it achieved an average AUC measure of 0.8289.

The comparison can be shown in the following way:



VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed a novel method based onMultivariate Time Series forecasting which considers the temporal correlations among similarity metrics and evolutions of link occurrences simultaneously in order to predict future new connections between nodes in a social network. Differently from the static approach that analyzes the network statically in a given time period without considering when links have been created and Univariate Time Series Model that uses the connectivity information of nodes as an input, our method takes the correlation between the similarity metrics and temporal evolutions of link occurrences information into account by using Autoregression Model for Multivariate Time Series forecasting. Repeated and new link prediction task experiments were performed on DBLP bibliographic coauthorship networks in order to evaluate the robustness of the proposed method. The experimental results showed that our Multivariate Time Series approach that can integrate covariance structures achieved better results for link prediction in temporal social networks than both the static approach and the univariate time series techniques aforementioned. In terms of AUC performance, the proposed technique results in an increase of 5-10% compared to the univariate time series models. In the future examine we will aim to research our proposed multivariate time series models in other social networks. And also, we will plan to extend our proposed model by using global topological metrics which analyze the whole topological information.

REFERENCES

[1] D. Liben-Nowell and J. Kleinberg, "The link prediction problem for social networks," in Proceedings of the Twelfth International Conference on Information and Knowledge Management, ser. CIKM '03. New York, NY, USA: ACM, 2003, pp. 556–559.

[2] L. L¨u and T. Zhou, "Link prediction in complex networks: A survey," Physica A: Statistical Mechanics and its Applications, vol. 390, no. 6, pp. 1150–1170, 2011.

[3] M. A. Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link prediction using supervised learning," in In Proc. of SDM 06 workshop on Link Analysis, Counterterrorism and Security, 2006.

[4] L. Getoor and C. P. Diehl, "Link mining: A survey," SIGKDD Explor. Newsl., vol. 7, no. 2, pp. 3–12, Dec. 2005.

[5] T. Tylenda, R. Angelova, and S. Bedathur, "Towards time-aware link prediction in evolving social networks," in Proceedings of the 3rd Workshop on Social Network Mining and Analysis, ser. SNA-KDD '09. New York, NY, USA: ACM, 2009, pp. 9:1–9:10.

[6] P. R. S. Soares and R. B. C. Prudˆ eNcio, "Proximity measures for link prediction based on temporal events," Expert Syst. Appl., vol. 40, no. 16, pp. 6652–6660, Nov. 2013.

[7] Z. Huang and D. K. J. Lin, "The time-series link prediction problem with applications in communication surveillance," INFORMS J. on Computing, vol. 21, no. 2, pp. 286–303, Apr. 2009.

[8] P. R. da Silva Soares and R. Bastos Cavalcante Prudencio, "Time series based link prediction," in Neural Networks (IJCNN), The 2012 International Joint Conference on. IEEE, 2012, pp. 1–7.

[9] P. D. Gilbert, "Combining var estimation and state space model reduction for simple good predictions," Journal of Forecasting, vol. 14, no. 3, pp. 229–250, 1995.

[10] M. Newman, "Clustering and preferential attachment in growing networks," Proceedings of the National Academy of Sciences, vol. 98, no. 2, pp. 025–102, 2001.

[11] L. A. Adamic and E. Adar, "Friends and neighbors on the web," Social networks, vol. 25, no. 3, pp. 211–230, 2003.

[12] M. E. Newman, "The structure of scientific collaboration networks," Proceedings of the National Academy of Sciences, vol. 98, no. 2, pp. 404–409, 2001.

[13] L. Katz, "A new status index derived from sociometric analysis," Psychometrika, vol. 18, no. 1, pp. 39–43, 1953.

[14] T. Zhou, L. L¨u, and Y.-C. Zhang, "Predicting missing links via local information," The European Physical Journal B-Condensed Matter and Complex Systems, vol. 71, no. 4, pp. 623–630, 2009.

[15] I. Mogotsi, "Christopher d. manning, prabhakar raghavan, and hinrich sch¨utze: Introduction to information retrieval," Information Retrieval, vol. 13, no. 2, pp. 192–195, 2010.

[16] R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla, "New perspectives and methods in link prediction," in Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '10. New York, NY, USA: ACM, 2010, pp. 243–252.

[17] Y. Dhote, N. Mishra, and S. Sharma, "Survey and analysis of temporal link prediction in online social networks," in Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference
on, Aug 2013, pp. 1178–1183.

[18] L. Munasinghe and R. Ichise, "Time score: A new feature for link prediction in social networks," IEICE TRANSACTIONS on Information and Systems, vol. 95, no. 3, pp. 821–828, 2012.

[19] A.-L. Barabˆasi, H. Jeong, Z. N´eda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations," Physica A: Statistical mechanics and its applications, vol. 311, no. 3, pp. 590–614, 2002.

[20] R. S. Tsay, Analysis of financial time series. John Wiley & Sons, 2005, vol. 543.

[21] Y. Sun, R. Barber, M. Gupta, C. C. Aggarwal, and J. Han, "Co-author relationship prediction in heterogeneous bibliographic networks [14] Z. Huang and D. K. J. Lin, "The time-series link prediction problem with applications in communication surveillance," INFORMS J. on Computing, vol. 21, no. 2, pp. 286–303, Apr. 2009.

[22] R. E. Kalman, "A new approach to linear filtering and prediction problems," J. Basic Eng., vol. 82, no. 1, pp. 35–45, Mar. 1960.

[23] J. Bibby and H. Toutenburg, Prediction and Improved Estimation in Linear Models. New York: Wiley, 1977.