Temporal Link Prediction Using Node Centrality and Time Series

Ting Zhang, Kun Zhang, Laishui Lv, and Xun Li

Abstract—**Link prediction is an important task in the area of complex networks. Some networks can be better modeled by temporal networks where the patterns of link appearance and disappearance varying with time. However, most of the previous link prediction researches ignore the temporal behaviors of links. The temporal link prediction needs to predict future links via a known network, considering the temporal relationship of node pairs. We propose a method combining the node centrality with time series. We distinguish the contributions of common neighbors to link generation by their centralities. Compared with benchmark approaches in several temporal networks, the proposed method can improve the accuracy of temporal link prediction efficiently.**

Index Terms—**Temporal link prediction, eigenvector centrality, time series, temporal networks.**

I. INTRODUCTION

Link prediction aims to estimate the link occurrence probability from available information in complex networks [1]. The link prediction has received lots of attention due to its important role in network analysis and link mining [2]. A link prediction model based on topological information was proposed by Liben-Nowell and Kleinberg [3]. The commonly similarity link prediction methods [4]-[9] allocate scores to node pairs and compute the likelihood of link generation by such scores. In order to improve the predict performance, Lü *et al*. adopted weak ties from social theory [10]. Considering the different roles of common neighbors, Liu *et al*. [11] combined node centrality with weak ties. Li *et al*. [12] proposed a node-centrality based link prediction method. They improved Salton index and random walk by using degree and betweenness centralities to measure the influence of nodes. Several methods consider community information in link prediction [13], [14]. They assumed that nodes have more probability to connect with other nodes in the same community.

Many networks have dynamic structures, which means that links may appear and disappear over time. A temporal network can be represented as a sequence of snapshots $\{G_1, G_2, ..., G_T\}$. Given a temporal network, link prediction task aims to predict future links in G_{T+1}, considering the temporal patterns in previous snapshots.

Time series models are employed to capture the temporal pattern of networks. In [15], time series of the non-connected nodes pair are constructed via the similarity scores at past timestamps. The final link prediction scores are computed according to such model. In [16], the time series for each node pairs is calculated by the frequency of link occurrence between adjacent snapshot. Then, the likelihood of links are estimated based on the time series model. In [17], a collection of time-series to model are adopted to produce similarity scores over time.

Several methods try to utilize the time-evolving patterns of temporal networks. O'Madadhain *et al*. [18] regarded the historic events and node attributes as the input of a logistic regression classifier and decided the future links according to the output. In [19], the local probabilistic model is extended to include time awareness.

We present a node centrality-based temporal link prediction approach by the extendence of Salton. The main contributions of this paper are as follows:
1) We assess the influence of common neighbors in temporal network by the eigenvector centrality of nodes. In order to describe the influence of the previous timestamps on the later ones, we use time series model to capture the evolution of node centrality over time. Then, the future node centralities are produced.
2) We regard the centrality of common neighbors as their contribution to the connection likelihood. We extend the definition of global neighbors from multilayer networks to temporal networks.
3) We execute experiments on several real temporal networks. The results show that the performance of the proposed approach outperforms other benchmark methods.

II. RELATED WORK

A. Similarity-Based Methods

A lot of similarity-based link prediction methods are depicted in the literature. We present several approaches briefly. For a node x, let $\Gamma(x)$ be the neighborhood set of x and k_x be the degree of x.

Common Neighbors (CN) [4] is one of the basic methods for link prediction. The similarity of node x and y is measured by the number of their common neighbors. The definition is written as:

$$S^{CN} = |\Gamma(x) \bigcap \Gamma(y)| \tag{1}$$

For Adamic-Adar (AA) [5], the node pairs that share fewer common nodes are weighted more heavily. The AA score is defined as:

$$S^{AA} = \sum_{z \in \Gamma(x) \bigcap \Gamma(y)} \frac{1}{\log(k_z)}$$
(2)

For node pairs without direct connection, Jaccard Index [6] produces higher score value when they have more common neighbors in proportion to their total number of neighbors. It is defined as:

$$S^{Jaccard} = \frac{|\Gamma(x) \bigcap \Gamma(y)|}{|\Gamma(x) \bigcup \Gamma(y)|}$$
(3)

Preferential Attachment (PA) [7] gives the score according to the degrees of node neighbors, and it is defined as:

$$S^{PA} = k_x \times k_y$$
(4)

Analogous to AA, Resource-Allocation (RA) [8] assigns more weights to the nodes with lower-degree neighbors. The RA score is defined as follows:

$$S^{RA} = \sum_{z \in \Gamma(x) \bigcap \Gamma(y)} \frac{1}{k_z}$$
(5)

Salton index [9] is based on cosine similarity, and it is defined as follows:

$$S^{Salton} = \frac{|\Gamma(x) \bigcap \Gamma(y)|}{\sqrt{k_x \times k_y}}$$
(6)

B. Eigenvector Centrality

In complex network, eigenvector centrality can measure the influential nodes. The eigenvector centrality is related to the number of node neighbors and the influence of each neighbor. The eigenvector centrality of node x can be calculated as follows:

$$c_x = \frac{1}{\lambda} \sum_{y \in \Gamma(x)} c_y$$
(7)

where λ is a constant and $\Gamma(x)$ is the neighbor set of node x.

For the adjacent matrix A, the eigenvector c associated with the largest eigenvalue of A is obtained as follows:

$$Ac = \lambda c$$
(8)

The centrality of node x is represented by the x th element of eigenvector c.

C. Time Series

The Auto-Regressive Integrated Moving Average (ARIMA) includes autoregressive and moving average models. It is a combination model for time series data. The ARIMA model is defined as follows:

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$
(9)

Here, X_t represents the estimated variable at t th snapshot, and $\phi_1, ..., \phi_p$ and $\theta_1, ..., \theta_p$ denote parameters, and

$\varepsilon_1, ..., \varepsilon_{t-q}$ denote error terms.

III. THE TSALTON APPROACH

Many activities are closely related to more influential nodes. Node centrality is regarded as the weight that represents the contribution of nodes to link generation. Therefore, the key issue is how to distinguish the nodes in important positions in the future when predicting links. Our proposed approach is referred to as the extension of Salton index [9] for temporal link prediction (TSalton).

A. The Proposed Model

In the TSalton method, we utilize the eigenvector centrality of each node to distinguish their roles to the likelihood of links. We use the ARIMA model to capture the changes of node centrality with time as follows:

$$C_t = \sum_{i=1}^{p} \phi_i C_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$
(10)

where C_t represents the estimated node centrality at t th snapshot. We calculates the eigenvector centrality at each timestamp. Then, we forecast the future node centrality by ARIMA.

Inspired by the node-centrality-based method in [12], we proposed the TSalton which is eigenvector-centrality-based Salton. The proposed model is read:

$$S_{xy}^{TSalton} = \frac{\sum_{z \in \Gamma_{GLN}(x) \bigcap \Gamma_{GLN}(y)} C_t(y)}{\sqrt{C_t(x) \times C_t(y)}}$$
(11)

where z denotes the global common neighbors of node x and y in the temporal network, and $C_t(x)$ means the centrality of node x produced by Eq.(10).

We extend the neighbors to temporal neighbors according to the definition of neighbors in multilayer network [20]. The global neighbors of node x represents the unique neighbors over time:

$$\Gamma_{GLN}(x) = \{ y \in V_t : e_{xy} \in E_{t-1} \bigcup E_t \}$$
(12)

where e_{xy} is the connections between node x and y, E_t is the link set at the t th snapshot.

B. The Proposed Algorithm

The TSalton algorithm consists of three main components: node centrality calculation, centrality-based score construction and temporal link prediction. Framework of TSalton algorithm is presented as follows:

Algorithm 1: TSalton algorithm

Input: A_t : Temporal network;

Output: A_{T+1} : Temporal network at time $T+1$
1 : Compute the node centrality of each snapshot according to Eq.(8);

2 : Construct the node centrality at time $T+1$ according to Eq.(10);

3 : Compute the scores according to Eq.(11);

4 : Predict temporal links according to the TSalton scores.

5 : return A_{T+1}.

IV. RESULTS

A. Evaluation Metrics

We adopt two standard metrics, AUC [21] and precision [22] to estimate the accuracy of temporal link prediction algorithms.

The AUC score is defined as $AUC = \dfrac{n' + 0.5n''}{n}$. Suppose nodes x and y are connected in the next timestamp. Randomly select other nodes x' and y' which don't exist link at the same timestamp. n is the total comparison times of S_{xy} and $S_{x'y'}$. n' is the number of times that S_{xy} has greater scores. n'' is the number of times that S_{xy} and $S_{x'y'}$ have the same scores.

The precision is defined as $precision = \dfrac{L_r}{L}$. The precision value is the ratio of relevant links to the number of selected links. If we predict L links to be connected and among them L_r links are correct.

B. Settings

TABLE I: THE STATISTICS OF TEMPORAL NETWORKS

Dataset	Vertices	Edges	|T|
Workplace	92	2,548	11 days
EnronEmail	151	50,572	3 years
Manufacture	167	82,927	9 months
CollegeMessage	1899	61,735	7 months

TABLE II. COMPARISON IN TERMS OF AUC

Dataset	Workplace	EnronEmail	Manufacture	CollegeMessage
TSalton	0.7782	0.9134	0.8722	0.9507
AA	0.6764	0.6884	0.6767	0.6769
CN	0.6750	0.6923	0.6771	0.6756
Jaccard	0.6688	0.6538	0.7985	0.6731
PA	0.6794	0.6896	0.8047	0.9017
RA	0.6761	0.6919	0.6787	0.6771

The temporal networks applied in this paper includes Workplace [23], EnronEmail [24], Manufacture [25] and CollegeMessage [24]. Table I reports the important statistics of the real-world network.

C. Evaluation of Link Prediction

In these experiments, we compare the temporal link prediction approaches on real-world networks. The comparative approaches are similarity-based methods in Section II. Experiments have been performed adopting the networks in Table I. We take snapshots of $G = \{G_1, G_2, ..., G_T\}$ for each temporal network. We evaluate the prediction results by comparing with the actual G_{T+1}.

The AUC results of these methods are given in Table II and the related ROC (Receiver operator characteristic) curves are depicted in Fig. 1. We can see that the TSalton outperforms all the other approaches.

(a) Workplace (b) EnronEmail
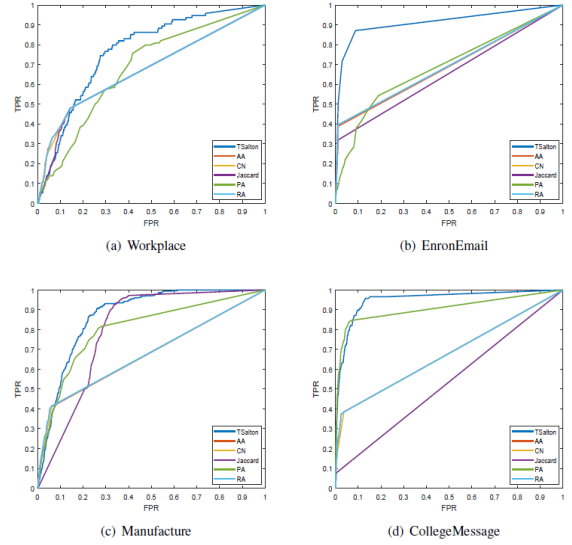
(c) Manufacture (d) CollegeMessage

Fig. 1. ROC curves of compared methods in each network.

The similarity-based methods ignore the temporal information of network and only use local topological information, so they perform poorly on temporal link prediction. Meanwhile, TSalton takes the temporal information into account, and the time series model can reflect the influential nodes more accurately. Let's take the example of EnronEmail. TSalton achieves the highest AUC 0.9134 and is about 20% higher than CN which is in second place.

We list the precision results in Table III. TSalton yields higher precision than similarity-based methods. As for EnronEmail, TSalton is about 0.4% higher than PA in second place. This result shows the advantages of time influence in link prediction.

TABLE III: COMPARISON IN TERMS OF PRECISION

Dataset	Workplace	EnronEmail	Manufacture	CollegeMessage
TSalton	0.0638	0.0964	0.0844	0.0689
AA	0.0267	0.0908	0.0489	0.0617
CN	0.0263	0.0853	0.0490	0.0649
Jaccard	0.0524	0.0910	0.0491	0.0622
PA	0.0298	0.0923	0.0380	0.0654
RA	0.0048	0.0846	0.0490	0.0611

In general, we can see that TSalton performs the best among the baseline approaches in term of AUC and precision on real networks. This results verify that TSalton is suitable for temporal link prediction. The possible reasons are from two aspects :(1) TSalton employs the time series model of eigenvector centrality which can represent the influential nodes more accurately in temporal networks. (2) TSalton considers the contributions of each neighbor to link generation via Salton scores to adjust such contribution.

V. CONCLUSION

The paper proposes a novel algorithm named TSalton. It

can obtain the topology information, which considers the influential nodes over time. The TSalton scores the likelihood of links by taking the eigenvector centralities of common neighbors into account. We have considered link prediction methods in terms of AUC and precision. Experiments in real networks show that, TSalton performs better than the methods without using temporal information. As a future direction, we aim to consider other information on each snapshot included in the model, such as the context and the community.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Ting Zhang conducted the research and wrote the paper. Kun Zhang analyzed the data. All authors had approved the final version.

ACKNOWLEDGMENT

REFERENCES

[1] L. Lü and T. Zhou, "Link prediction in complex networks: A survey," *Physica A Statistical Mechanics & Its Applications,* vol. 390, no. 6, pp. 1150-1170, 2011.

[2] L. Getoor and P. C. Diehl, "Link mining: A survey," *ACM Sigkdd Explorations Newsletter,* vol. 7, no. 2, pp. 3-12, 2005.

[3] D. Liben-Nowell and J. Kleinberg, "The link prediction problem for social networks," *Journal of the American Society for Information Science and Technology,* vol. 58, no. 7, pp. 1019-1031, 2007.

[4] M. E. Newman, "Clustering and preferential attachment in growing networks," *Physical Review E,* vol. 64, no. 2, 025102, 2001.

[5] L. A. Adamic and E. Adar, "Friends and neighbors on the web," *Social Networks,* vol. 25, no. 3, pp. 211-230, 2003.

[6] J. Paul, "A comparative study on the distribution of flowers in the Alpis and Rula areas," *Bulletin of the Wozhou Natural Science Society,* vol. 37, pp. 547-579, 1901.

[7] M. Mitzenmacher, "A brief history of generative models for power law and lognormal distributions," *Internet Mathematics,* vol. 1, no. 2, pp. 226-251, 2004.

[8] L. Lü, T. Zhou, and Y. C. Zhang, "Predicting missing links via local information," *European Physical Journal B,* vol. 71, no. 4, pp. 623-630, 2009.

[9] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*, mcgraw-hill, 1983.

[10] L. Lü and T. Zhou, "Role of weak ties in link prediction of complex networks," in *Proc. the 1st ACM international Workshop on Complex Networks Meet Information & Knowledge Management,* 2009, pp. 55-58.

[11] H. Liu, Z. Hu, H. Haddadi, and H. Tian, "Hidden link prediction based on node centrality and weak ties," *Epl,* vol. 101, no. 1, 2013.

[12] L. Li, X. Liu, N. Chen, and H. Tian, "Link prediction based on node centrality," in *Proc. the International Conference on Information Technology and Electrical Engineering*, 2018, pp. 1-6.

[13] J. Valverde-Rebaza and A. Andrade Lopes, "Exploiting behaviors of communities of twitter users for link prediction," *Social Network Analysis and Mining,* vol. 3, no. 4, pp. 1063-1074, 2013.

[14] A. Biswas and B. Biswas, "Community-based link prediction," *Multimedia Tools and Applications,* vol. 76, no. 18, pp. 18619-18639, 2017.

[15] R. P. S. Soares and R. B. C. Prudencio, "Time series based link prediction," in *Proc. The 2012 international joint conference on neural networks (IJCNN), 2012,* pp. 1-7.

[16] Z. Huang and D. K. J. Lin, "The time-series link prediction problem with applications in communication surveillance," *INFORMS Journal on Computing,* vol. 21, no. 2, 2009.

[17] Í. Güneş¸ S. Gündüz Ö˘güdücü, and Z. Cataltepe, "Link prediction using time series of neighborhood-based node similarity scores," *Data Mining and Knowledge Discovery,* vol. 30, no. 1, pp. 147-180, 2016.

[18] J. O'Madadhain, J. Hutchins, and P. Smyth, "Prediction and ranking algorithms for event-based network data," *ACM Sigkdd Explorations Newsletter,* vol. 7, no. 2, pp. 23-30, 2005.

[19] T. Tylenda, R. Angelova, and S. Bedathur, "Towards time-aware link prediction in evolving social networks," in *Proc. the 3rd Workshop on social Network Mining and Analysis*, 2009, pp. 1-9.

[20] D. Hristova, A. Noulas, C. Brown, M. Musolesi, and C. Mascolo, "A multilayer approach to multiplexity and link prediction in online geo-social networks," *EPJ Data Science,* vol. 5, no. 1, pp. 24, 2016.

[21] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters,* vol. 27, no. 8, pp. 861-874, 2005.

[22] J. Herlocker, J. Konstan, L. Terveen, and J. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems,* vol. 22, no. 1, pp. 5-53, 2004.

[23] M. Génois, C. L. Vestergaard, J. Fournet, A. Panisson, I. Bonmarin, and A. Barrat, "Data on face-to-face contacts in an office building suggests a low-cost vaccination strategy based on community linkers," *Network Science,* vol. 3, no. 3, pp. 326-347, 2015.

[24] R. Rossi and N. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Proc. Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015, pp. 4292-4293.

[25] R. Michalski, S. Palus, and P. Kazienko, "Matching organizational structure and social network extracted from email communication," in *Proc. International Conference on Business Information Systems*, 2011, pp. 197-206.

Ting Zhang received the B.S and the M.S from Harbin Engineering University. She is currently pursuing the Ph.D. degree with Nanjing University of Science and Technology, Nanjing, China. Her research interests are focused on complex networks and artificial intelligence.

Kun Zhang received the B.S., M.S., and Ph.D. from Nanjing University and Science and Technology. She is currently a professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology. Her research interests are focused on big data, pattern recognition, medical engineering, complex networks, network security and trusted computing.

Laishui Lv received the B.S and M.S from Gannan Normal University. He is currently pursuing the Ph.D. degree with Nanjing University of Science and Technology, Nanjing, China. His research interests are focused on complex networks, statistical analysis and nonlinear optimization

Xun Li received the M.S from Shenyang Ligong University. She is currently pursuing the Ph.D. degree with Nanjing University of Science and Technology, Nanjing, China. Her research interests are focused on machine learning and complex networks.