Memory Based Temporal Network Prediction

Zou, Li; Wang, An; Wang, Huijuan

DOI
10.1007/978-3-031-21131-7_51

Publication date
2023

Document Version
Final published version

Published in
Complex Networks and Their Applications XI - Proceedings of The 11th International Conference on Complex Networks and Their Applications

Memory Based Temporal Network Prediction

Li Zou, An Wang, and Huijuan Wang

Abstract Temporal networks are networks like physical contact networks whose topology changes over time. Predicting future temporal network is crucial e.g., to forecast and mitigate the spread of epidemics and misinformation on the network. Most existing methods for temporal network prediction are based on machine learning algorithms, at the expense of high computational costs and limited interpretation of the underlying mechanisms that form the networks. This motivates us to develop network-based models to predict the temporal network at the next time step based on the network observed in the past. Firstly, we investigate temporal network properties to motivate our network prediction models and to explain how the performance of these models depends on the temporal networks. We explore the similarity between the network topology (snapshot) at any two time steps with a given time lag/interval. We find that the similarity is relatively high when the time lag is small and decreases as the time lag increases. Inspired by such time-decaying memory of temporal networks and recent advances, we propose two models that predict a link's future activity (i.e., connected or not), based on the past activities of the link itself or also of neighboring links, respectively. Via seven real-world physical contact networks, we find that our models outperform in both prediction quality and computational complexity, and predict better in networks that have a stronger memory. Beyond, our model also reveals how different types of neighboring links contribute to the prediction of a given link's future activity, again depending on properties of temporal networks.

Keywords Temporal network prediction · Network-Based prediction · Temporal network property

L. Zou · H. Wang (✉)
Faculty of Electrical Engineering, Mathematics and Computer Science,
Delft University of Technology, Delft, The Netherlands
e-mail: h.wang@tudelft.nl

A. Wang
Department of Chemistry, University of Warwick, Coventry, UK

H. Cherifi et al. (eds.), *Complex Networks and Their Applications XI*,
Studies in Computational Intelligence 1078,
https://doi.org/10.1007/978-3-031-21131-7_51

1 Introduction

Complex systems can be represented as networks, where nodes represent the components of a system and links denote the interaction or relation between the components. The interactions are, in many cases, not continuously active. For example, individuals connect via email, phone call or physical contact at specific times instead of constantly. Temporal networks [1, 2] could represent these systems more realistically with time-varying network topology. It has been shown that temporal network properties such as community structure, the degree distribution in the aggregated network, inter-event time and the non-Markovian evolution influence dynamic processes on the temporal network [3–6].

Temporal network prediction is a task of predicting temporal contacts at the next time step based on the temporal network topology observed in the previous L steps. Predicting the temporal network such as a physical contact network is essential to forecast and mitigate the spread of epidemics and misinformation on the network. The temporal network prediction problem is also equivalent to problems in recommender systems, e.g., predicting which user will purchase which product, which individuals will become acquaintance [7, 8].

Recently, machine learning algorithms have been developed to predict temporal networks. Examples include temporal network embedding [9–11], restricted Boltzmann machine (RBM) based methods [12] and Graph neural networks [13]. These methods, however, are at the expense of high computational costs and limited in providing insights regarding which network mechanisms are used for network prediction thus could possibly form temporal networks. Few network-based methods have been proposed to predict new links, i.e., the node pairs that will have contact in the future but have not had any contact in the past, instead of predicting all contacts at a future time step. These network-based methods consider a network property, also called similarity, of a node pair as the tendency that a new link will appear between the node pair [14–16]. Initial network-based methods for temporal network prediction have been explored recently, assuming that a link is more likely to have a contact (be active) in the future if it has contacts recently, depending possibly on the the previous contacts of other neighboring links [17].

However, we still lack deep understanding of how to design network-based methods to predict temporal networks and how the performance of prediction methods depend on properties of temporal networks. Hence, this work aims to explore basic temporal network properties, to motivate the design of network-based temporal network prediction methods and to explain these methods' network dependent performance. Firstly, we explore the similarity between the activity (i.e., connected or not) of a link at any two time steps with a given time lag/interval and the similarity between the network topology (snapshot) at any two time steps with a given time lag. Intuitively, if such similarity is relatively high, thus there exists memory in temporal networks, we may predict a temporal network in the future based on the network observed in the past. We find both similarities, or memories, decay as the time lag increases and they are relatively high when the time lag is small. Correspondingly, we propose two temporal network prediction models utilizing the observed time decaying memory in temporal networks.

Our first model, called the self-driven (SD) model, assumes that a link's future state (connected or not) is only influenced by it's past states and the influence of its earlier state is smaller than that of more recent states. Specifically, it assumes that the tendency for link i being connected or active at time step $t+1$ is given by $w_i(t+1) = \sum_{k=t-L+1}^{k=t} e^{-\tau(t-k)} x_i(k)$, where $x_i(k)$ is the state of link i at time step k and τ is the decay factor controlling the contribution from each past state. This definition and concept is not new, and has been used in [17, 18]. The SD model is emphasized as one model here because we will explore in depth the decay factor and its implications and it is the basis to built our SCD model. We find the SD model performs well in network prediction when the decay factor is chosen arbitrarily in the broad range $\tau \in [0.5, 5]$ in each of the seven real-world physical contact temporal networks. This implies that our real-world physical contact networks measured in the context of school, hospital, workplace etc. may be formed by a universal class of time decaying memory. This common range of the decay factor $\tau \in [0.5, 5]$ suggests that the state of a link is mainly determined by the link's states in few recent steps.

Furthermore, we generalize the SD model to a self- and cross-driven (SCD) model that predicts a link's next step activity by using the SD connection tendency of the link itself and also of the other neighbor links. We find that SCD outperforms SD and both SCD and SD perform better than the baseline models such as linear regression. Both models perform better in networks with a stronger memory. The SCD model also reveals how different types of neighboring links contribute to the prediction of a given link's future activity, which we find also depend on temporal network properties.

We will introduce the presentation of temporal networks (Sect. 2), real-world temporal networks to be considered (Sect. 3), analyze key temporal networks (Sect. 4) to motivate our temporal network prediction models (Sect. 5). The proposed models will be evaluated and interpreted in Sects. 6 and 7 respectively.

2 Temporal Network Representation

A temporal network can be represented as a sequence of network snapshots $G = \{G_1, G_2, ..., G_T\}$, where T is duration of the observation window, $G_t = (V; E_t)$ is the snapshot at time step t with V and E_t being the set of nodes and contacts, respectively. If node j and k have a contact at time step t, $(j, k) \in E_t$. Here, we assume all snapshots share the same set of nodes, i.e., V. The links in the aggregated network G^w are defined as $E = \cup_{t=1}^{T} E_t$. That is, a pair of nodes is connected with a link in the aggregated network if at least one contact occurs between them in the temporal network. Hence, the link set E in the aggregated network contains all the node pairs that have contact(s) in the temporal network and the total number of links is $M = |E|$. We give each link in the aggregated network an index i, where

Table 1 The number of nodes ($N = |V|$), the number of node pairs that have contact(s) (M), the length of the observation time window (T), time resolution (δ s), the type of contacts and the location where the data is collected

Network	N	M	T	δ	Type	Location
Hospital	75	1139	9453	20	Physical	Hospital
Hypertext2009	113	2196	5246	20	Physical	Conference
Workplace	92	755	7104	20	Physical	Office
LH10	73	1381	12605	20	Physical	Hospital
HighSchool	327	5818	7375	20	Physical	School
PrimarySchool	242	8317	3100	20	Physical	School
SFHH	403	9565	3509	20	Physical	Conference

$i \in [1, M]$. The temporal connection or activity of link i over time could then be represented by a T-dimension vector $\mathbf{x_i}$ whose element is $x_i(t)$, where $t \in [1, T]$, $x_i(t) = 1$ when node pair i has a contact at time t and $x_i(t) = 0$ if no contact occurs at t.

3 Empirical Data Sets

To design and evaluate temporal network prediction methods, we consider seven empirical physical contact networks: Hospital, Workplace, PrimarySchool, High-School, LH10 [19], SFHH [20] and Hypertext2009 [21]. Basic properties of these data sets are given in Table 1. The time steps at which there is no contact in the whole network have been deleted.

4 Memory in Temporal Networks

In this section, we aim to understand whether a temporal network at different times shares certain similarity or has memory.

Auto-correlation Firstly, we explore the correlation of the activity of a link at two times with a given interval Δ, called time lag, via the auto-correlation of the activity series of each link. The auto-correlation of a time series is the Pearson correlation between the given time series and its lagged version. We compute, for each link i, the Pearson correlation coefficient $R_{x_i x_i}(\Delta)$ between $\{x_i(t)\}_{t=1,2,\ldots,T-\Delta}$ and $\{x_i(t)\}_{t=\Delta+1,\Delta+2,\ldots,T}$ as its auto-correlation coefficient. Figure 1a shows that the average auto-correlation coefficient over all links decays with the time lag Δ in each of the seven data sets. The average auto-correlation decays slower as the time lag increases.

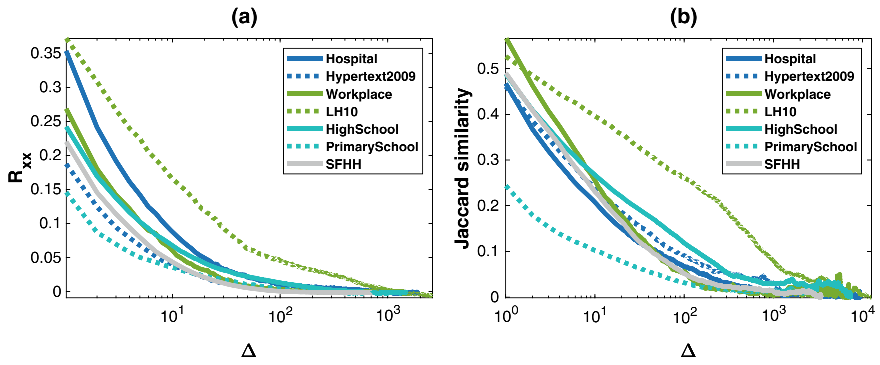

Fig. 1 **a** The average auto-correlation coefficient R_{xx} over all links as a function of the time lag Δ and **b** the average Jaccard similarity of two snapshots of a temporal network with a given time lag Δ in each of the seven data sets

Jaccard similarity Furthermore, the similarity of the network at two times with a given time lag Δ is examined via Jaccard similarity (JS). JS measures how similar two sets are by considering the percentage of shared elements between them. For two snapshots of a temporal network G_t and $G_{t+\Delta}$, their Jaccard similarity is defined as the size of their intersection in contacts divided by the size of the union of their contact sets, that is, $JS(G_t, G_{t+\Delta}) = \frac{E_t \cap E_{t+\Delta}}{E_t \cup E_{t+\Delta}}$. Large JS means large overlap/similarity between the two snapshots of the temporal network. Figure 1b shows the average Jaccard similarity over all possible pairs of temporal network snapshots that have a time lag Δ. Similar to auto correlation in link activity, the correlation between temporal snapshots decays with their time lag in all empirical data sets, manifesting the time decaying memory of real-world temporal networks.

5 Temporal Link Prediction Methods

Inspired by the time decaying memory of temporal networks, we propose two temporal link prediction models. Our previous work on Lasso Regression [22], a statistical learning model has found that a link's state at the next step is largely determined by the current state of the link itself and the neighboring links that share a common node with the link. Hence, our two network-based models in this work will predict a link's future activity based on the past activities of the link itself, and also of the neighboring links respectively by taking the memory effect into account.

5.1 Self-Driven (SD) Model

The self-driven (SD) model defines the tendency $w_i(t+1)$ for link i being active at time $t+1$ as:

$$w_i(t+1) = \sum_{k=t-L+1}^{k=t} e^{-\tau(t-k)} x_i(k). \tag{1}$$

where the decay factor τ controls the rate of the memory decay and $x_i(k)$ is the state of link i at time step k. A large τ corresponds a fast decay of memory, such that a small number of previous states affect the tendency of connection. When $\tau = 0$, all past states have equal influence on the future connection tendency and $w_i(t+1)$ reduces to the contact number of link i during the past L steps. Such exponential decay has also been considered in [17, 23]. In Sect. 6, we will show that the SD model performs well for a common wide range of the decay factor τ among all real-world networks considered and we do not need to learn τ from the temporal network observed in the past.

5.2 Self- and Cross-Driven (SCD) Model

Furthermore, we generalize the SD model to a self- and cross-driven (SCD) model that predicts a link's next step activity by using the SD connection tendency defined in Eq. (1) of the link itself and also of neighboring links that share a node with the link in the aggregated network. The union of the target link and its neighboring links is also called the ego-network centered at the target link, exemplified in Fig. 2. Furthermore, we differentiate three types of links in an ego-network, colored in differently in Fig. 2: the target link itself, links that form a triangle with the target link and the rest links. We believe the previous states of these three types of links may contribute differently to the estimation of the target link's next step activity, motivated by the finding of the Lasso Regression in temporal network prediction [22], the common neighbor similarity method in static network prediction and temporal motifs (e.g., three contacts that happen within a short duration and with a specific ordering in time, and form a triangle in topology) that have been widely observed in temporal networks [24, 25].

Hence, our SCD model assumes that the SCD tendency $h_i(t+1)$ for link i to be active at time step $t+1$ is a linear function

$$h_i(t+1) = \beta_0^* + \beta_1^* w_i(t+1) + \beta_2^* u_i(t+1) + \beta_3^* f_i(t+1). \tag{2}$$

of the contributions of the link itself $w_i(t+1)$ as defined in Eq. (1), the neighboring links that form a triangle with the target link $u_i(t+1)$ and the other neighboring links $f_i(t+1)$. The latter two factors $u_i(t+1)$ and $f_i(t+1)$ will be defined soon as a function of the SD tendency at $t+1$ of all the links in the ego-network.

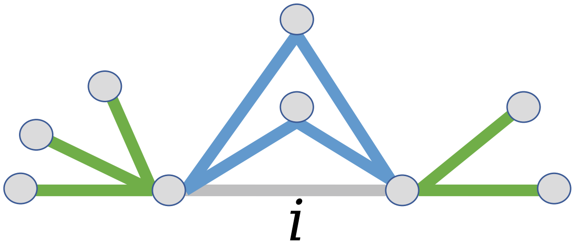

Fig. 2 An illustrative example of an ego-network centered at a target link i. The three types of links, the target link itself, links that form a triangle with the target link and the other links, are colored in grey, blue and green respectively

The contribution $u_i(t+1)$ of the neighboring links that form a triangle with the target link i is defined as follows. For each pair of neighboring links j and k that form a triangle with the target link i, the geometric mean $\sqrt{w_j(t+1) \cdot w_k(t+1)}$ suggests the strength that the two end nodes of link i interact with the corresponding common neighbor. We define $u_i(t+1)$ as the average geometric mean over all link pairs that form a triangle with the target link, a weighted version of common neighbor similarity. The contribution of the other links $f_i(t+1)$ in the ego-network is defined as the average SD tendency of connection. The coefficients β_0^*, β_1^*, β_2^*, and β_3^* in Eq.(2) will be learned through Lasso Regression from the temporal observed in the past L steps for each link. Using previous states of neighboring links to predict the future connection of a link has been explored in [17]. The design of SCD model in e.g., $u_i(t+1)$ aims to capture the weighted version of common neighbor similarity, which enables us later to discover the relation between model performance and the clustering coefficient of the aggregated network.

5.3 Baseline Models

Here, we introduce two baseline models.

Common neighbor similarity (CN). We generalize the common neighbor similarity method from static network prediction to the temporal network prediction problem. The number of common neighbors [14] of a node pair can be computed for each of the previous L snapshots. The sum of the number of common neighbors over the past L snapshots, are used to estimate this node pair's tendency of connection at the next time step.

Lasso Regression [26] assumes that the activity of link i at time $t+1$ is a linear function of the activities of all the links at time t, i.e.,

$$x_i(t+1) = \sum_{j=1}^{M} x_j(t)\beta_{ij} + c_i. \tag{3}$$

The objective is

$$\min_{\beta_i}\{\sum_{t=1}(x_i(t+1)-\sum_{j=1}^{M}x_j(t)\beta_{ij}-c_i)^2+\alpha\sum_{j=1}^{M}|\beta_{ij}|\}. \tag{4}$$

where M is the number of features as well as the number of links, c_i is the constant coefficient and $\beta_i = \{\beta_{i1}, \beta_{i2}, \ldots, \beta_{iM}\}$ are the regression coefficients of all the features for link i. The coefficients will be learned from the temporal network observed in the past L steps for each link. We use $L1$ regularization, which adds a penalty to the sum of the magnitude of coefficients $\sum_{j=1}^{M}|\beta_{ij}|$. The parameter α controls the penalty strength. The regularization forces some of the coefficients to be zero and thus lead to models with few non-zero coefficients (relevant features). The optimal α that achieves the best prediction is chosen from 50 logarithmically spaced points within $[10^{-4}, 10]$.

6 Model Evaluation

In this section, we firstly introduce the method to evaluate the models in link prediction quality. Secondly, we explore how to choose the decay factor in the SD model. Thirdly, we compare the link prediction quality of all the models.

6.1 Link Prediction Quality

Each model predicts the link activity at time step $t+1$ based on the temporal network observed in the past L steps. The number of contacts at each time step shows periodic behaviour, i.e., large number of contact recurs at regular intervals. In order to capture such potential periodic patterns of a temporal network, we consider $L = T/2$, i.e., half of the length of a real-world temporal network's time window. The prediction step $t+1$ is sampled 1000 times from $[T/2+1, T]$ with equal space. The average proportion of the M links that are active at a time step is lower than 1% in all the real-world networks we considered. The classification labels (the number of active links and inactive links per time step) are imbalanced. Hence, we evaluate the prediction quality via the area under the precision-recall curve (AUPR) [27]. AUPR provides an aggregate measure of performance across all possible classification thresholds. The average AUPR of a model over the 1000 prediction snapshots quantifies the prediction quality of the model.

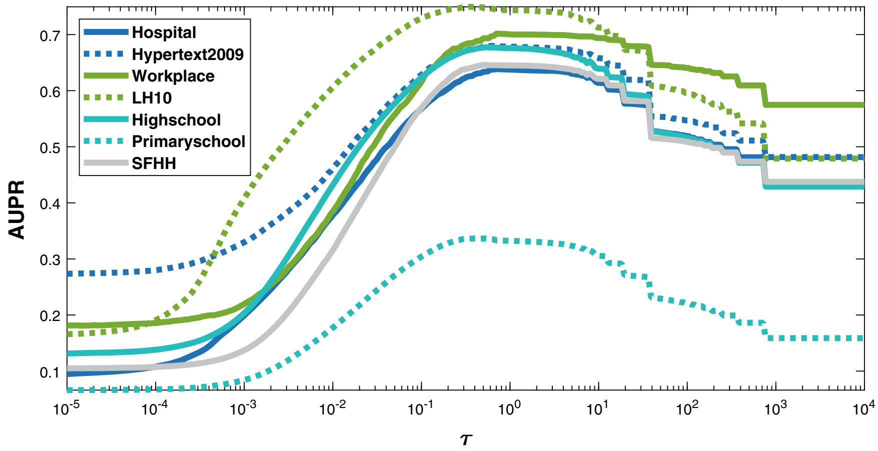

Fig. 3 Link prediction quality AUPR of the SD model as a function of the decay factor τ in seven data sets

6.2 *Choice of Decay Factor*

How to chose the decay factor τ will be motivated by comparing two possibilities. We first consider a simple case where τ is a control parameter and does not vary over time, i.e., remaining the same for the 1000 samples of the prediction steps. For a given τ, the tendency $w_i(t+1)$ ($i \in [1, 2, ..., M]$) is obtained at each prediction step $t+1$ based on Eq.(1). Figure 3 shows that the decay factor τ indeed affects the prediction quality AUPR of the SD model. A universal pattern is that the optimal performance is obtained by a common and relatively broad range of $\tau \in [0.5, 5]$ in all networks. This implies that our real-world physical contact networks measured at school, hospital, workplace etc. may be formed by a universal class of time decaying memory. Hence, τ can be chosen arbitrarily within [0.5, 5].

In the second method of choosing τ, a $\tau(t+1)$ for each prediction step $t+1$ is learned from the network observed in the past L steps. The $\tau(t+1)$ is choose as the one that allows the SD model to best predict the temporal network at t based on the network observed in the past $L-1$ steps just before t. The prediction quality from the first (second) method of choosing τ are 0.63 (0.61), 0.68 (0.67), 0.69 (0.63), 0.75 (0.74), 0.68 (0.67), 0.34 (0.33) and 0.65 (0.63), for the seven data sets, respectively.

Hence, τ could be chosen arbitrarily from [0.5, 5], which has lower computational complexity and better prediction quality than learning τ dynamically over time. We consider $\tau = 0.5$ to derive the SD tendency and SCD tendency in the rest analysis.

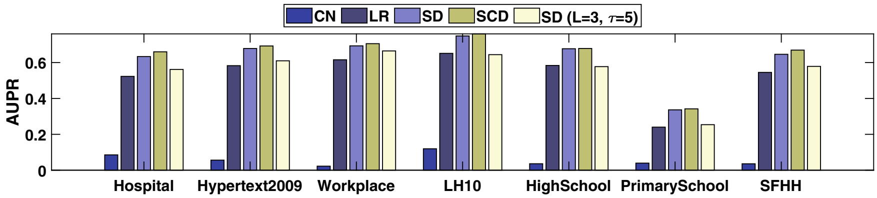

Fig. 4 Temporal network prediction quality AUPR of baseline CN method, baseline Lasso Regression (LR), SD model and SCD model. All methods consider $L = T/2$ and $\tau = 0.5$ except for SD ($\tau = 5$, $L = 3$), which is needed only for Sect. 7.2

6.3 Comparison of Models

We further compare the prediction quality of our SD model, SCD model and baseline models in Fig. 4. We find that both SD and SCD models perform better than the baselines. The SCD model, which predicts a link's connection utilizing SD tendency of the neighboring links and of the link itself, indeed performs better than the SD model that uses only the SD tendency of the link itself. Moreover, the SD and SCD models perform the best (worst) in LH10 (PrimarySchool), in line with the strongest(weakest) memory/similarity of LH10(PrimarySchool) observed in Fig. 1.

7 Model Interpretation

In this section, we interpret firstly the SCD model, to understand how the past states of different types of links in the ego-network (neighborhood) contribute to the prediction of the center link of the ego-network and interpret afterwards the common range $\tau \in [0.5, 5]$.

7.1 Interpretation of SCD Model

As defined in Eq. (2), SCD model predicts a link's future connection, based on the SD tendency of the link itself, links that form a triangle with the link and the rest links that share a common node with the link. The contribution of these three types of links are reflected in the learned coefficients in Eq. (2). The average coefficients over all prediction steps are given in Table 2. We can see a link's next step activity is mainly influenced by the past activities of the link itself, and slightly influenced by the neighboring links that form a triangle with the target link. The other neighboring links have very limited impact on target link's activity. The predictive power of neighboring links that form a triangle with the target link may come from the nature

Table 2 The learned coefficient β_1^*, β_2^* and β_3^* in SCD model averaged over 1000 prediction steps are also provided

Network	β_1^*	β_2^*	β_3^*	cc
Hospital	0.31	0.07	0.00	0.37
Hypertext2009	0.32	−0.02	0.00	0.32
Workplace	0.32	0.00	0.00	0.28
LH10	0.32	0.21	0.00	0.41
HighSchool	0.33	0.04	0.00	0.38
PrimarySchool	0.24	0.48	−0.02	0.54
SFHH	0.32	0.03	0.01	0.21

And the clustering coefficient (cc) of the aggregated network in each empirical network

of physical contact networks: contacts are often determined by physical proximity and two people that are close to a third but not yet close to each other are likely to already be in relatively close proximity.

One exception is the PrimarySchool, where $\beta_2^* > \beta_1^*$. The aggregated network of PrimarySchool has the largest clustering coefficient[1] in the aggregated network as shown in Table 2. In general, we find the contribution of links that form a triangle with the target link tends to be more significant in temporal networks whose aggregated network has more triangles.

7.2 Decay Factor

Finally, we interpret the common range of the decay factor $\tau \in [0.5, 5]$ where the SD performs optimally. According to the definition of SD tendency of connection in Eq.(1), only the coefficients/contributions $y = e^{-\tau(t-k)}$ of the previous 24 steps (3 steps) are larger than 10^{-5} when $\tau = 0.5$ ($\tau = 5$), out of $L = T/2 > 1000$ previous steps observed. We wonder whether considering only few previous steps instead of $L = T/2$ steps would be sufficient for a good prediction. The prediction quality of our SD model when $L = 3$ and $\tau = 5$ is given in Fig. 4. It is worse than the performance of SD model when $L = 50\%T$ and $\tau = 0.5$. This suggests that although the contribution of each early state of a link is small, the accumulated contribution of many early states improves the prediction quality. The prediction quality of SD model when $L = 3$ and $\tau = 5$, whose computational complexity is extremely low, is still better or similar to that of Linear Regression, reflecting the prediction power of recent states of a link.

[1] The clustering coefficient of network is the probability that two neighbors of node are connected.

8 Conclusion

In this work, we propose two network-based temporal network prediction models motivated by the observed time decaying similarities/memory in temporal networks. The proposed self-driven (SD) model and self- and cross-driven (SCD) model predict a link's future activity based on the past activities of the link itself, and also of the neighboring links, respectively. Both models perform better than the baseline models and the SCD outperforms SD model. Interestingly, we find that both models perform better in temporal networks with a stronger memory (similarity over time). The SCD model reveals that a link's future activity is mainly determined by (the past activities of) the link itself, moderately by neighboring links that form a triangle with the target link, and hardly by other neighboring links. However, if the temporal network has a high clustering coefficient in its aggregated, the contribution of the neighboring links that form a triangle with the target link tends to be significant and possibly dominant.

Our work is a starting point to explore network-based temporal network prediction methods, especially how methods could be designed based on network proprieties and how their performance could be explained again by network properties. It is interesting to evaluate network-based prediction methods more systematically in comparison with learning-based methods and explore the integration of both types of methods.

Acknowledgements The authors would like to thank the support of the Netherlands Organisation for Scientific Research NWO (TOP Grant no. 612.001.802) and also thank the support of the China Scholarship Council.

References

1. Holme, P., Saramäki, J.: Temporal networks. Phys. Rep. **519**, 97–125 (2012)
2. Masuda, N., Lambiotte, R.: A guide to temporal networks. In: Series on complexity science, vol. 4, pp. 252. World scientific, Europe (2016)
3. Masuda, N., Klemm, K., Eguíluz, V.M.: Temporal networks: slowing down diffusion by long lasting interactions. Phys. Rev. Lett. **111**, 188701 (2013)
4. Delvenne, J.-C., Lambiotte, R., Rocha, L.E.C.: Diffusion on networked systems is a question of time or structure. Nat. Commun. **6**, 7366 (2015)
5. Peixoto, T., Rosvall, M.: Modelling sequences and temporal networks with dynamic community structures. Nat. Commun. **8**, 582 (2017)
6. Zhan, X.-X., Hanjalic, A., Wang, H.: Information diffusion backbones in temporal networks. Sci. Rep. **9**, 6798 (2019)
7. Lü, L., Medo, M., Yeung, C.H., Zhang, Y.-C., Zhang, Z.-K., Zhou, T.: Recommender systems. Phys. Rep. **519**, 1–49 (2012)
8. Aleta, A., Tuninetti, M., Paolotti, D., Moreno, Y., Starnini, M.: Link prediction in multiplex networks via triadic closure. Phys. Rev. Res. **2**, 042029 (2020)
9. Zhou, L., Yang, Y., Ren, X., Wu, F., Zhuang, Y.: Dynamic network embedding by modeling triadic closure process. In: 32nd AAAI Conference on Artificial Intelligence, pp. 571-578. AAAI Press, California USA (2018)
10. Rahman, M. Saha, T.K., Hasan, M.A., Xu, K.S., Reddy, C.K.: DyLink2Vec: Effective Feature Representation for Link Prediction in Dynamic Networks (2018)

11. Zhan, X.-X., Li, Z., Masuda, N., Holme, P., Wang, H.: Susceptible-infected-spreading-based network embedding in static and temporal networks. EPJ Data Sci. **9**, 30 (2020)
12. Li, X., Du, N., Li, H., Li, K., Gao, J., Zhang, A.: A deep learning approach to link prediction in dynamic networks. In: 2014 SIAM International Conference on Data Mining, pp. 289-297 (2014)
13. Pareja, A., Domeniconi, G., Chen, J., Ma, T., Suzumura, T., Kanezashi, H., Kaler, T., Schardl, T.B., Leiserson, C.E.: EvolveGCN: evolving graph convolutional networks for dynamic graphs. In.: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 5363–5370. AAAI Press, California, USA (2020)
14. Liben-Nowell, D., Kleinberg, J.: The link prediction problem for social networks. In: The Twelfth International Conference on Information and Knowledge Management, vol. 4, pp. 556–559. Association for Computing Machinery, New York (2003)
15. Ahmed, N.M., Chen, L., Wang, Y., Li, B., Li, Y., Liu, W.: Sampling-based algorithm for link prediction in temporal networks. Inf. Sci. **374**, 1–14 (2016)
16. Xu, H.H., Zhang, L.J.: Application of link prediction in temporal networks. Adv. Mater. Res. **2231**, 756–759 (2013)
17. Li, X., Liang, W., Zhang, X., Liu, X., Wu, W.: A universal method based on structure subgraph feature for link prediction over dynamic networks. In: 39th International Conference on Distributed Computing Systems, pp. 1210–1220. IEEE, Dallas, USA (2019)
18. Jo, H.-H., Perotti, J.I., Kaski, K., Kerteéz, J.: Correlated bursts and the role of memory range. Phys. Rev. E. **92**, 022814 (2015)
19. Génois, M., Barrat, A.: Can co-location be used as a proxy for face-to-face contacts? EPJ Data Sci. **7**, 11 (2018)
20. Rossi, R.A., Ahmed, K.: The network data repository with interactive graph analytics and visualization. In: The Twenty-Ninth AAAI Conference on Artificial Intelligence, pp. 4292–4293. AAAI Press, Palo Alto, California (2015)
21. Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J.-F., den Broeck, W.V.: What's in a crowd? Analysis of face-to-face behavioral networks. J. Theor. Biol. **271**, 166–180 (2011)
22. Zou, L., Zhan, X.-X., Sun, J., Hanjalic, A., Wang, H.: Temporal network prediction and interpretation. IEEE Trans. Netw. Sci. Eng. **9**, 1215–1224 (2022)
23. Yu, W., Cheng, W., Aggarwal, C.C., Chen, H., Wang, W.: Link prediction with spatial and temporal consistency in dynamic networks. In: The Twenty-Sixth International Joint Conference on Artificial Intelligence, pp. 3343–3349. International Joint Conferences on Artificial Intelligence, Melbourne, Australia (2017)
24. Paranjape, A., Benson, A.R., Leskovec, J.: Motifs in temporal networks. In: Tenth ACM International Conference on Web Search and Data Mining, pp. 601–610. Association for Computing Machinery, Cambridge, United Kingdom (2017)
25. Saramäki, J., Moro, E.: From seconds to months: an overview of multi-scale dynamics of mobile telephone calls. Eur. Phys. J. B. **88**, 164 (2015)
26. Tibshirni, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Series B. **58**, 267–88 (1996)
27. Davis, J., Goadrich, M.: The relationship between precision-recall and ROC curves. In: The 23rd International Conference on Machine Learning, vol. 8, pp. 233–240. Association for Computing Machinery, New York, USA (2006)