



Delft University of Technology

Prediction and Intervention Strategy Design on Temporal Networks

Zou, L.

DOI

[10.4233/uuid:7ada0861-6bc3-4561-9d48-84a8191fee26](https://doi.org/10.4233/uuid:7ada0861-6bc3-4561-9d48-84a8191fee26)

Publication date

2024

Document Version

Final published version

Citation (APA)

Zou, L. (2024). *Prediction and Intervention Strategy Design on Temporal Networks*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:7ada0861-6bc3-4561-9d48-84a8191fee26>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

PREDICTION AND INTERVENTION STRATEGY DESIGN ON TEMPORAL NETWORKS

PREDICTION AND INTERVENTION STRATEGY DESIGN ON TEMPORAL NETWORKS

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 24 May 2024 om 10:00 uur

door

Li Zou

Master of Science in Systems Theory,
Beijing Normal University, Beijing, China,
geboren te Sichuan, China.

Dit proefschrift is goedgekeurd door de

promotor: Prof. dr. A. Hanjalic

promotor: Dr. ir. H. Wang

Samenstelling promotiecommissie:

Rector Magnificus,

voorzitter

Prof. dr. A. Hanjalic,

Technische Universiteit Delft

Dr. ir. H. Wang,

Technische Universiteit Delft

Onafhankelijke leden:

Prof. dr. ir. R.E. Kooij

Technische Universiteit Delft

Prof. dr. ir. A. Bozzon

Technische Universiteit Delft

Prof. dr. F. De Pellegrini

University of Avignon

Dr. K. Avrachenkov

INRIA Sophia Antipolis

Dr. M.A. Kitsak

Technische Universiteit Delft



Keywords: Temporal Network, Intervention, Prediction

Printed by: ProefschriftMaken

Copyright © 2024 by L. Zou

ISBN 978-94-6366-872-9

An electronic version of this dissertation is available at

<http://repository.tudelft.nl/>.

To my family

CONTENTS

Summary	ix
1 Introduction	1
1.1 Background	2
1.2 Research challenges	2
1.3 Thesis scope and contribution	4
1.4 Publication related to this thesis	5
1.5 How to read this thesis	6
References	6
2 Temporal Network Prediction and Interpretation	9
2.1 Introduction	10
2.2 Temporal network representation	11
2.3 Empirical data sets	12
2.4 Temporal network prediction methods	12
2.4.1 Lasso regression and random forest model.	13
2.4.2 Training and test data	15
2.4.3 Baseline models	15
2.5 Model evaluation	15
2.6 Model interpretation	17
2.6.1 Construction of the prediction backbone network using influence coefficients	17
2.6.2 Backbone network in relation to time series	19
2.6.3 The backbone network in relation to the aggregated network	21
2.7 Long-term effect	24
2.8 Conclusion	25
2.9 Appendix	27
2.9.1 Data description	27
2.9.2 Supplementary figures and tables	27
References	31
3 Short- and Long-term Temporal Network Prediction Based on Network Memory	35
3.1 Introduction	36
3.2 Empirical data sets	39
3.3 Memory in temporal networks	39
3.4 Short-term network prediction methods	40
3.4.1 Our network-based models	40
3.4.2 Baseline models	42

3.5	Performance analysis in short-term prediction	44
3.5.1	Model evaluation	44
3.5.2	Model interpretation	46
3.6	Long-term prediction methods	48
3.6.1	Recursive long-term prediction	48
3.6.2	Repeated long-term prediction	49
3.7	Performance analysis in long-term prediction	49
3.7.1	Network snapshot	50
3.7.2	Aggregated network	52
3.7.3	The distribution of inter-event time	54
3.8	Conclusion	55
3.9	Appendix	56
	References	60
4	Intervention in Information Transport on Temporal Networks	65
4.1	Introduction	66
4.2	Temporal network	68
4.3	Empirical data sets	68
4.4	Link-removal strategies	70
4.4.1	Transport efficiency of the two end nodes of a link	70
4.4.2	Aggregated network-based properties	70
4.4.3	Path-based properties	72
4.5	Results	74
4.5.1	Evaluation	74
4.5.2	Explanation	76
4.6	Conclusion	79
	References	80
5	Conclusion	83
5.1	Main contribution and reflections	83
5.2	Future work	85
	Acknowledgements	87
	Curriculum Vitæ	89
	List of Publications	91

SUMMARY

Communications networks such as opportunistic mobile networks, vehicle networks, and social contact networks can be represented as temporal networks, where two cars or individuals are connected at a discrete time step if they are close to each other or interact with each other at that time step. These networks facilitate the propagation of information or epidemics where a piece of information or a virus is transmitted from one node to another node through their time-varying contacts. Since the spread of misinformation or disease can lead to disruptions in social stability and hinder economic development, effective mitigation strategies can serve as a safeguard against potential losses. In this thesis, we pursue the challenge of enabling an effective intervention in spreading processes on temporal networks.

The effectiveness of existing mitigation strategies critically depends on the insight into the effect of the intervention on the dynamic process in the future, which depends on the future dynamics of a complex system itself. However, in reality, a future topology of the temporal networks underlying those spreading processes is unknown. Hence, we first design methods to predict temporal networks in the future and also provide insightful interpretation regarding which inherent properties of temporal networks or mechanisms of network formation to consider when generating these predictions. For this prediction problem, we focus on network-based prediction models due to their general advantages, namely low computational complexity and analytic insights into the mechanisms that influence the evolution of the temporal network. Given that the development of such methods needs an understanding of patterns underlying the temporal networks that possibly facilitate the prediction, we start with investigating two interpretable learning algorithms i.e., Lasso Regression and Random Forest for temporal network prediction. We find that the previous activity of the target link itself contributes more than that of the neighboring links in predicting the target link's future activity. Equipped with such insights, we then design network-based methods to predict temporal networks. Additionally, predicting temporal networks like physical contact networks in the long-term future beyond the short-term i.e., one step ahead is crucial to forecast and mitigate the spread of epidemics and misinformation on the network. This long-term prediction problem has been seldom explored. Therefore, we propose basic methods that adapt each aforementioned prediction model to address classic short-term network prediction problems for long-term network prediction tasks.

Assuming the temporal network in the future is perfectly predicted or known, we further design advanced strategies beyond the current state-of-the-art to effectively mitigate the spreading process on temporal networks. We investigate the intervention in information transport on a temporal network to link removal. The objective is to understand the removal of which types of links negatively influence the efficiency of information transport the most. Identifying such critical links via their properties will enable better protection (intervention) to facilitate (prohibit) the spread of (mis)information.

To identify the types of links whose removal decreases the efficiency of the network the most, we propose link-removal strategies based on transport efficiency between two end nodes of each link, and based on properties/measures of links in the aggregated network and in the fastest time-respecting paths. We find that our proposed strategies can reduce the transport efficiency of the network more significantly than random removal in seven physical contact networks and six virtual networks. Most importantly, the most effective strategy for identifying critical links is the one based on time-respecting paths. We find that the links whose contacts appear more often and occur early in the fastest time-respecting paths are critical for the spread of information.

The last chapter summarizes and reflects upon the insights from our research and provides suggestions for possible future directions.

1

INTRODUCTION

*If you want to live a happy life,
tie it to a goal,
not to people or things.*

Albert Einstein

1.1. BACKGROUND

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 [1]. The interactions are, in many cases, not continuously active. For example, two cars or individuals in a communications network (such as an opportunistic mobile network, vehicle network, and social contact network) are connected (or have a contact) at a discrete time step if they are close to each other or interact with each other at that time step. The time-varying nature of such systems makes it less suitable to represent them using static network topologies, requiring *temporal network* [2], [3] to represent these systems more realistically. In temporal networks, links activate or deactivate over time. Temporal networks facilitate the propagation of information or epidemics on them, where a piece of information or a virus is transmitted from one node to another node through their time specific contacts (or activation of links) [4], [5].

One of the fundamental objectives in the field of complex systems is *to devise strategies to intervene in the dynamic processes on temporal networks*, like epidemic spreading and information diffusion. Given that the spread of misinformation or disease can lead to disruptions in social stability and hinder economic development, effective mitigation strategies can serve as a safeguard against potential losses. Progress has been made in prohibiting the spreading of information/epidemics starting from one seed node by selecting and removing a given number of links or contacts based on their properties. Here, links are node pairs that have at least one contact in a temporal network. The effectiveness of these strategies critically depends on the insight into the effect of the intervention on the dynamic process in the future, which again depends on the future dynamics of a complex system itself. In reality, however, a future topology of the temporal networks underlying those spreading processes is unknown.

In this thesis, we pursue the challenge of enabling an effective intervention in spreading processes without knowing future temporal network topology. We design methods to *predict the development of temporal networks in the future* and also provide insightful interpretation regarding which inherent properties of temporal networks or mechanisms of network formation to consider when generating these predictions. Both the predicted future temporal network and the detected underlying mechanisms of network formation may facilitate the development of mitigating strategies. Following this contribution, we also make a first step towards devising *advanced mitigation strategies*. For this, we assume the temporal network in the future is known or perfectly predicted and investigate how to select links to block based not only on classic network properties of links, but also on the new type of link properties/centralities in spreading trajectories.

1.2. RESEARCH CHALLENGES

Given that dynamic processes on temporal networks are heavily influenced by the evolution of the networks, the prediction of temporal networks plays a key role in both designing and testing mitigation strategies. Recently, *machine learning methods* have been developed to predict temporal networks. Embedding algorithms embed each node in a low-dimensional space based on the network observed. If the learned representations of two nodes are closer in the vector space, it is more likely to have a contact between

this node pair one time step ahead [6]–[10]. Restricted Boltzmann machine (RBM) based methods [11] and graph neural networks [12]–[14] have also been developed for this prediction task and they can achieve relatively high prediction accuracy. These methods, however, are at the expense of high computational costs and only provide limited insight (if any) regarding which mechanisms enable the prediction and thus could possibly influence the evolution of temporal networks in the future. Another category of prediction methods is *network-based methods* that exploit a network property, also called similarity, of a node pair to model whether a new link will appear between the node pair [15], [16] or not. The main advantage of network-based methods is low computational complexity and an analytic insight into the mechanisms that form temporal networks. However, these methods have been proposed mainly to predict new links, i.e., the node pairs that will have contact(s) in the future but have not had any contact in the past, instead of predicting all contacts at a specific future time step. These methods cannot predict temporal networks in the future well since past activity of each target link and patterns underlying the temporal networks have not been systematically utilized.

In this thesis, we focus on network-based models to predict future temporal networks due to the abovementioned general advantages. Given that the development of such methods needs an understanding of patterns underlying the temporal networks that possibly facilitate the prediction, we start with investigating two interpretable learning algorithms i.e., Lasso Regression and Random Forest for temporal network prediction. We find that the previous activity of the target link itself contributes more than that of the neighboring links. Equipped with such insights, we then design network-based methods to predict temporal networks. Additionally, predicting temporal networks like physical contact networks in the long-term future beyond the short-term i.e., one step ahead is crucial to forecast and mitigate the spread of epidemics and misinformation on the network. This long-term prediction problem has been seldom explored. Therefore, we propose basic methods that adapt each aforementioned prediction model to address classic short-term network prediction problems for long-term network prediction tasks.

Moving on to address the development of mitigation strategies, recent efforts made to intervene in the dynamic processes on temporal networks can be classified into two main categories. The first line of research has explored the mitigation of epidemic spread through node-level approaches. Génois et al. [17] have demonstrated that effectively preventing epidemic outbreaks can be achieved by removing individuals who serve as bridges between communities in the time-aggregated network. Gemmetto et al. [18] investigated how to remove a sub-group of nodes in a temporal network to mitigate infectious diseases in school environments. Apart from node blocking, strategies based on link blocking [19] and contact blocking [20] in temporal networks have also been proposed. In these works, a given number of links or contacts are selected and removed based on their properties mainly in the aggregated network. Temporal properties of temporal networks have been found to have a profound effect on the dynamic processes deployed on them. For example, inter-event time influences the diffusion processes on temporal networks [21], [22]. Consequently, this research aims to conduct a thorough exploration to determine whether temporal information associated with links can be leveraged to devise advanced strategies for more effective mitigation of the spreading process.

1

1.3. THESIS SCOPE AND CONTRIBUTION

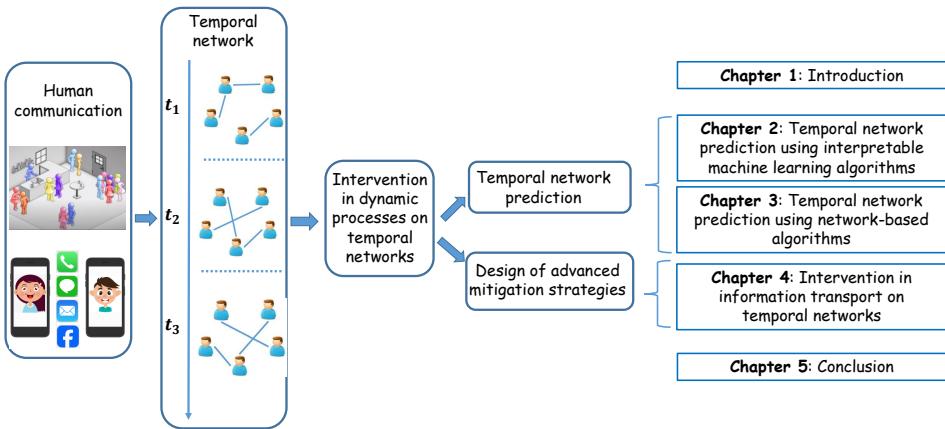


Figure 1.1: The structure of the thesis.

In Chapters 2 and 3, we address the challenge of predicting a temporal network in the future based on the network observed in the past of a given duration, and with the goal in mind to enable effective intervening in the processes on the network. The temporal network prediction problem is more challenging than the static network prediction problem, which aims to predict the missing links based on the links observed and assuming that the network topology is not changing over time. As explained in the previous section, we focus on network-based methods. These methods reveal directly which mechanisms or properties are used for the prediction, and also they have low computational complexity. Given that the development of such methods needs an understanding of patterns underlying the temporal networks that possibly facilitate the prediction, we start with investigating two interpretable learning algorithms i.e., Lasso Regression and Random Forest for temporal network prediction in Chapter 2. From the coefficients learned from each algorithm, we construct the prediction backbone network that presents the influence of all links in determining each link's future activity. Analysis of the backbone, and its relation to the link activity time series and to the time aggregated network reflects which properties of temporal networks are captured by the learning algorithms. Via six real-world contact networks, we find that the next step activity of a particular link is mainly influenced by (a) its current activity and (b) links strongly correlated in the time series to that particular link and close in distance (in hops) in the aggregated network.

Equipped with insights into the mechanisms and network properties that enable the prediction, we then design network-based methods to predict temporal networks in Chapter 3. Specifically, we propose methods to predict the connection of each node pair one step ahead based on the connections of this node pair itself and of node pairs that share a common node with this target node pair in the past. The concrete design of our two prediction models is based on the analysis of the memory property of real-world

physical networks, i.e., to what extent two snapshots of a network at different times are similar in topology (or overlap). State-of-the-art prediction methods that allow interpretation are considered as baseline models. In seven real-world physical contact networks, our methods are shown to outperform the baselines in both prediction accuracy and computational complexity. They perform better in networks with stronger memory. Importantly, our models reveal how the connections of different types of node pairs in the past contribute to the connection estimation of a target node pair.

To develop effective strategies to mitigate the epidemic or information spreading on the network, it is essential to predict the contact network (activities of all links) in the long-term future, instead of only one step ahead. This problem remains challenging as it has yet been insufficiently explored. In Chapter 3, we pursue this challenge by adapting our models for short-term network prediction to solve the long-term network prediction problem. The prediction quality of all adapted models is evaluated via the accuracy in predicting each network snapshot and in reproducing key network properties. We find that the prediction based on one of our models tends to have the highest accuracy and lowest computational complexity.

In Chapter 4, we assume the future temporal networks are given or perfectly predicted, and develop advanced strategies to mitigate an information diffusion process. For this purpose, we investigate the intervention in information transport on temporal networks. The objective is to understand the removal of which types of links deteriorates the efficiency/speed of information transport the most. Identifying such critical links will enable better intervention to facilitate/prohibit the spread of (mis)information. To identify critical links, we propose link-removal strategies based on transport efficiency between two end nodes of each link, properties of links in the aggregated network, and in routing paths respectively. Each strategy ranks links according to their corresponding properties and removes links with the highest measures. Strategies are evaluated via the relative change in transport efficiency after link removal in real-world networks. We find that the path-based strategy performs the best: links that appear more often and occur early in the fastest time-respecting paths tend to be critical. Via comprehensive analysis, we explain this strategy's out-performance and its dependency on network properties.

1.4. PUBLICATION RELATED TO THIS THESIS

The following papers are completed by the author of this thesis while pursuing the Ph.D. degree at the Delft University of Technology.

1. **L. Zou**, X.-X. Zhan, J. Sun, A. Hanjalic, and H. Wang, Temporal Network Prediction and Interpretation, IEEE Transactions on Network Science and Engineering 9, 1215-1224 (2022). [[Chapter 2](#)]
2. **L. Zou**, A. Wang, and H. Wang, Memory Based Temporal Network Prediction, In Conference on Complex Networks and Their Applications XI, Springer, Cham, 661-673 (2023). [[Chapter 3](#)]
3. **L. Zou**, C. Ceria, and H. Wang, Short- and long-term temporal network prediction based on network memory, Applied Network Science 18, 639 (2023). [[Chapter 3](#)]

4. **L. Zou**, and H. Wang, Intervention in information transport on temporal networks, submitted to IEEE Transactions on Network Science and Engineering [[Chapter 4](#)]

1.5. HOW TO READ THIS THESIS

Chapters 2, 3, and 4 of this thesis comprise original publications. The corresponding publication references can be found in the footnote at the beginning of each chapter. Each chapter functions as a stand-alone work, capable of being read without reference to previous chapters. The length and depth of chapters may vary in different chapters as we publish the papers in different scientific journals or conferences.

REFERENCES

- [1] M. Newman, *Networks: An Introduction*. OUP Oxford, 2010, ISBN: 9780191500701.
- [2] P. Holme and J. Saramäki, “Temporal networks,” *Phys. Rep.*, vol. 519, no. 3, pp. 97–125, 2012, ISSN: 0370-1573.
- [3] N. Masuda and R. Lambiotte, *A Guide to Temporal Networks* (complexity science). World scientific (Europe), 2016, vol. 4.
- [4] M. Karsai, N. Perra, and A. Vespignani, “Time varying networks and the weakness of strong ties,” *Scientific Reports*, vol. 4, p. 4001, 2014.
- [5] J.-P. Eckmann, E. Moses, and D. Sergi, “Entropy of dialogues creates coherent structures in e-mail traffic,” *Proceedings of the National Academy of Sciences*, vol. 101, no. 40, pp. 14 333–14 337, 2004.
- [6] S. Kumar, X. Zhang, and J. Leskovec, “Predicting dynamic embedding trajectory in temporal interaction networks,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD ’19, Anchorage, AK, USA: Association for Computing Machinery, 2019, pp. 1269–1278, ISBN: 9781450362016.
- [7] S. M. Kazemi, R. Goel, K. Jain, *et al.*, “Representation learning for dynamic graphs: A survey,” *J. Mach. Learn. Res.*, vol. 21, no. 1, Jan. 2020, ISSN: 1532-4435.
- [8] L. Zhou, Y. Yang, X. Ren, F. Wu, and Y. Zhuang, “Dynamic network embedding by modeling triadic closure process,” in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, ser. AAAI’18/IAAI’18/EAAI’18, New Orleans, Louisiana, USA: AAAI Press, 2018, ISBN: 978-1-57735-800-8.
- [9] Y. Wang, Y.-Y. Chang, Y. Liu, J. Leskovec, and P. Li, “Inductive representation learning in temporal networks via causal anonymous walks,” *ArXiv*, vol. abs/2101.05974, 2021.
- [10] M. Rahman, T. K. Saha, M. A. Hasan, K. S. Xu, and C. K. Reddy, *Dylink2vec: Effective feature representation for link prediction in dynamic networks*, 2018. arXiv: [1804.05755](#).

- [11] H. Zhang, "A deep learning approach to dynamic interbank network link prediction," *International Journal of Financial Studies*, vol. 10, no. 3, 2022, ISSN: 2227-7072.
- [12] A. Pareja, G. Domeniconi, J. Chen, *et al.*, "Evolvegcn: Evolving graph convolutional networks for dynamic graphs," *ArXiv*, vol. abs/1902.10191, 2019.
- [13] L. Wu, P. Cui, J. Pei, L. Zhao, and X. Guo, "Graph neural networks: Foundation, frontiers and applications," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, ser. KDD '23, , Long Beach, CA, USA, Association for Computing Machinery, 2023, pp. 5831–5832, ISBN: 9798400701030.
- [14] Y. Ma and J. Tang, "Graph embedding," in *Deep Learning on Graphs*. Cambridge University Press, 2021, pp. 75–106.
- [15] 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 Orleans, LA, USA: Association for Computing Machinery, 2003, pp. 556–559, ISBN: 1581137230.
- [16] N. M. Ahmed, L. Chen, Y. Wang, B. Li, Y. Li, and W. Liu, "Sampling-based algorithm for link prediction in temporal networks," *Information Sciences*, vol. 374, pp. 1–14, 2016, ISSN: 0020-0255.
- [17] M. GÉNOIS, C. L. VESTERGAARD, J. FOURNET, A. PANISSON, I. BONMARIN, and A. BARRAT, "Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers," *Network Science*, vol. 3, no. 3, pp. 326–347, 2015.
- [18] "Mitigation of infectious disease at school: targeted class closure vs school closure..," en, *BMC infectious diseases*, vol. 14, no. 1, p. 695, Dec. 2014, ISSN: 1471-2334.
- [19] X.-X. Zhan, A. Hanjalic, and H. Wang, "Suppressing information diffusion via link blocking in temporal networks," in *Complex Networks and Their Applications VIII*, H. Cherifi, S. Gaito, J. F. Mendes, E. Moro, and L. M. Rocha, Eds., Cham: Springer International Publishing, 2020, pp. 448–458, ISBN: 978-3-030-36687-2.
- [20] S. Zhang, X. Zhao, and H. Wang, "Mitigate sir epidemic spreading via contact blocking in temporal networks," *Applied Network Science*, vol. 7, p. 2, 2022.
- [21] B. Min, K.-I. Goh, and A. Vazquez, "Spreading dynamics following bursty human activity patterns," *Phys. Rev. E*, vol. 83, p. 036102, 2011.
- [22] G. Grinstein and R. Linsker, "Power-law and exponential tails in a stochastic priority-based model queuens," *Phys. Rev. E*, vol. 77, p. 012101, 2008.

2

TEMPORAL NETWORK PREDICTION AND INTERPRETATION

Temporal networks refer to networks like physical contact networks whose topology changes over time. Predicting future temporal networks is crucial e.g., to forecast epidemics. Existing prediction methods are either relatively accurate but black-box, or white-box but less accurate. The lack of interpretable and accurate prediction methods motivates us to explore what intrinsic properties/mechanisms facilitate the prediction of temporal networks. We use interpretable learning algorithms, Lasso Regression and Random Forest, to predict, based on the current activities (i.e., connected or not) of all links, the activity of each link at the next time step. From the coefficients learned from each algorithm, we construct the prediction backbone network that presents the influence of all links in determining each link's future activity. Analysis of the backbone, its relation to the link activity time series and to the time aggregated network reflects which properties of temporal networks are captured by the learning algorithms. Via six real-world contact networks, we find that the next step activity of a particular link is mainly influenced by (a) its current activity and (b) links strongly correlated in the time series to that particular link and close in distance (in hops) in the aggregated network.

2.1. INTRODUCTION

Real-world systems can be represented as complex networks, where nodes denote the components and links denote relations or interaction between these components. In many cases, however, the interactions are not continuously active. For example, individuals connect via email, text message, phone call or physical contact at specific time stamps instead of constantly. Temporal networks [1]–[3] could represent these systems more realistically with time-varying network topology. A temporal network can be regarded as a static network where each link is further associated with a time series specifying whether an interaction (contact) occurs or not at each time step. Temporal networks display non-trivial properties, which may have profound effect on the dynamic processes deployed on them. For example, the inter-event (contact, activation) time between a node pair has been found to follow a heavy-tail or power-law distribution in many temporal networks [4]–[6]. It has been shown that temporal network properties such as community structure, the degree distribution in the aggregated network, and inter-event time influence the diffusion processes on the temporal network [7]–[15].

Temporal network prediction is a task of predicting temporal interactions/contacts at a future time step based on the temporal network topology observed in the past. Predicting the temporal network such as a physical contact network in the future is essential to forecast performance of a process upon the network like the prevalence of epidemic spreading. 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 [16]–[18].

Existing prediction methods are either relatively accurate but black-box, or white-box but less accurate, although progress has also been made recently in evaluating the predictability of a temporal network [19]. Markovian Methods and machine learning algorithms have been developed to predict temporal network in short term, i.e., at the next time step based on the network observed so far within a given time window. Markovian models [20] can be developed by considering the time series or activity of each link and predict a link's future activity based on its previous activities. Markovian models have

also been built by regarding the temporal network or the link activated at each time step as the state [21], [22]. Deep learning methods have been further developed to improve the temporal link prediction. Examples include temporal network embedding [23]–[25], restricted Boltzmann machine (RBM) based methods [26], [27] and Graph neural networks [28]–[31]. These methods, however, do not allow for insightful interpretation regarding which inherent property or mechanism of the temporal networks could these methods capture when predicting temporal networks.

In this work, we address the problem of temporal network prediction, and its interpretation with respect to what underlying properties of temporal networks a prediction algorithm possibly captures or utilizes. We confine ourselves to the problem of predicting the activity of each link at a given time step based on the activities of all the links at the previous step. A statistical learning algorithm, i.e., Lasso Regression and a basic machine learning algorithm, i.e., Random Forest have been used for network prediction because of their interpretability. We further construct the prediction backbone network using the coefficients learned from the algorithms. The weighted backbone network suggests the influence of every link in determining a given's activity. Characterizing the backbone network in relation to the time series of all the links and the aggregated network unveils other patterns underlying the temporal networks that possibly facilitate the prediction. We find that a link's current state is largely determined by its own activity but also influenced by the activities of other links, at the previous time step. Links tend to influence each other more if they have a shorter and/or more shortest paths in the aggregated network and are more strongly correlated in their time series.

These findings, when combined with modern deep learning techniques can potentially lead to interpretable yet accurate prediction models. They may also inspire the development of temporal network models and strategies to mitigate epidemic spreading on physical contact networks.

2.2. TEMPORAL NETWORK REPRESENTATION

A temporal network, denoted as G , can be represented as a sequence of network snapshots: $G = \{G_1, G_2, \dots, G_T\}$, 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 $i \in [1, M]$. 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 . The activity of all links can be captured by a $M \times T$ dimensional matrix \mathbf{X} with its element $X(i, t) = x_i(t)$ where $t \in [1, T]$ and $i \in [1, M]$.

2.3. EMPIRICAL DATA SETS

Most real-world temporal network data sets available are contact networks. Without losing the generality, we choose six empirical networks that range from physical and virtual human contact networks to animal contact networks: Hypertext 2009 [32], [33], High-school [34], Call [35], Sms [35], Baboons [36], and Ant [37]. Basic description is given in section 2.9.1 in Supplementary and properties of these data sets are given in Table 2.1. Note that the time steps at which there is no contact in the whole network have been deleted. Basic description of how each temporal network is measured and constructed explains to some extent the difference of these networks in, for example, the average number of contacts per link.

We report the distribution of inter-event(contact) time in Figure 2.1, i.e., the interval between two consecutive contacts between a node pair. As it is often the case for human dynamics, the distributions of inter-contact time are heterogeneous. All our six systems show a heavy-tail distribution. It means the networks we consider exhibit burstiness which corresponds to frequent activities over a short period of time followed by a long period of inactivity [38]–[40].

Table 2.1: 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 sec), the average number of contacts within the observation time window per link (η) and the type of contacts in each empirical network.

Network	N	M	T	δ	η	Type
Hypertext 2009	113	2196	5246	20	9.5	Human contact (conference)
Highschool	312	2242	899	20	12.8	Human contact (high school)
Call	75	270	8597	1	34	Human contact (phone call)
Sms	110	210	60932	1	291	Human contact (message)
Baboons	26	303	10072	5	1401	Animal contact (Baboons)
Ant	89	649	993	0.5	2.8	Animal contact (ants)

2.4. TEMPORAL NETWORK PREDICTION METHODS

In this section, we propose our methodology which allows not only temporal network prediction but also the deduction of the relationship between links in the aggregated network G_w in influencing each other's activity, i.e., the dynamic of link activities. Specifically, we aim to understand to what extend a link's activity (active/having contact or not) at a given time step is determined by the other links' and its own activity at the previous time step.

Firstly, we introduce a statistical learning algorithm, i.e., Lasso Regression and a basic

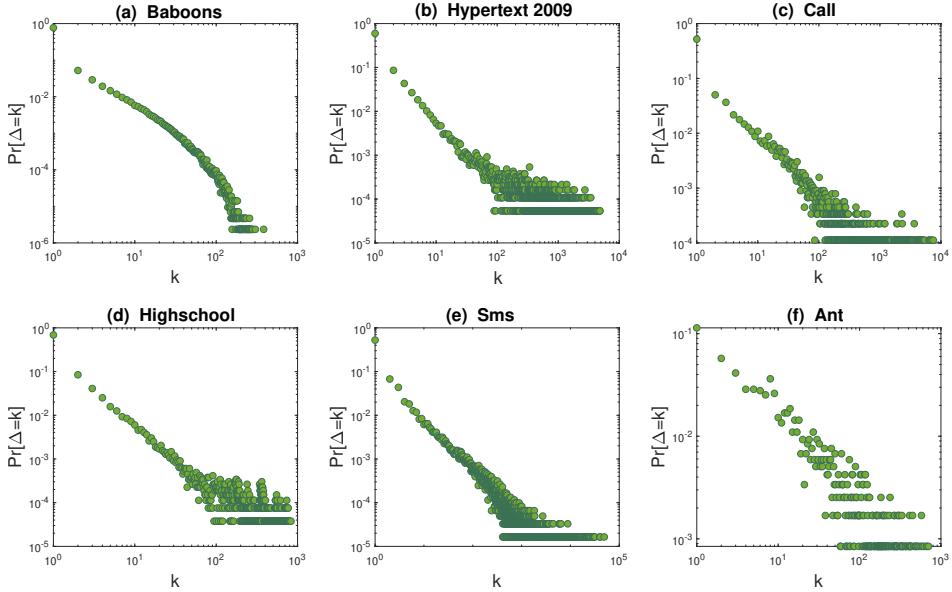


Figure 2.1: The probability distribution $Pr[\Delta = k]$ of the inter-event time Δ in number of time steps in log-log scale for (a) Baboons, (b) Hypertext 2009, (c) Call, (d) Highschool, (e) Sms and (f) Ant.

machine learning algorithm, i.e., Random Forest, to predict temporal networks. In view of the heavy-tail distribution of inter-event time, thus the possibility that the activity of a link remains the same within a short period, we introduce two baseline models that assume the activity of a link is determined only by its own activity at the previous time step. These four models predict the activity of a link at a given time based on its and/or other links' activities at the previous time step.

Afterwards, we illustrate how to deduce the influence between links in activities via applying these proposed models to real-world temporal network data. This requires the calibration of the coefficients of the models and entails the setup of training and test data sets.

2.4.1. LASSO REGRESSION AND RANDOM FOREST MODEL

Our method applies to a generic temporal network with N nodes and M links (node pairs that have at least one contact) whose activities are recorded within a time window $[1, T]$. The activities of the M links are recorded by a $M * T$ matrix X . The state or activity of link i at time $t + 1$ is $x_i(t + 1)$ ($t \in [p, p + L - 1]$), which equals 1 when link i is active, and equals 0 otherwise. We assume that the activity of link i at time $t + 1$ is a function of the activities of all the links at time t , i.e.,

$$x_i(t + 1) = f_i(x_1(t), x_2(t), \dots, x_M(t)). \quad (2.1)$$

The mapping function f_i is unknown and link-specific. It can be learned from the activities of all links, i.e., $[x_i(p), x_i(p+1), \dots, x_i(p+L)]$ where $i \in [1, M]$ within a time window $[p, p+L]$, and denoted as $f_i^{p,L}$. We construct in total L training data samples for each link i based on the temporal network observed within $[p, p+L]$: we use link i 's state at each time step $t+1 \in [p+1, p+L]$ as target and the corresponding features are the states of all links at time step t . The training data samples for node pair i is expressed as a set $\mathcal{D}_i(p, L)$:

$$\mathcal{D}_i(p, L) = \{x_i(t+1); x_1(t), x_2(t), \dots, x_M(t)\}_{t=p}^{p+L-1}. \quad (2.2)$$

A learning algorithm assumes a given function $f_i^{p,L}$, whose coefficients can be learned from a training set $\mathcal{D}_i(p, L)$. The learned function $f_i^{p,L}$ tells us to what extent $x_i(t+1)$ can be estimated by the activity of each link at t respectively.

We explore a statistical learning (Lasso Regression) and a machine learning algorithm (Random Forest) to learn $f_i^{p,L}$.

Lasso Regression assumes f_i to be a linear function [41], [42]

$$x_i(t+1) = \sum_{j=1}^M x_j(t) \beta_{ij} + c_i. \quad (2.3)$$

The objective is

$$\min_{\beta_i} \left\{ \sum_{t=p}^{p+L-1} (x_i(t+1) - \sum_{j=1}^M x_j(t) \beta_{ij} - c_i)^2 + \alpha \sum_{j=1}^M |\beta_{ij}| \right\}. \quad (2.4)$$

where L is the number of training samples, M is the number of features as well as the number of links, c_i is the constant coefficient and $\beta_i = \{\beta_{i1}, \beta_{i2}, \dots, \beta_{iM}\}$ are the regression coefficients of all the features for link i . A large coefficient β_{ij} indicates that feature $x_j(t)$ influences or determines significantly the target $x_i(t+1)$.

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). If α is zero, Lasso Regression reduces to the classical linear regression algorithm. Given a training data set $\mathcal{D}_i(p, L) = \{x_i(t+1); x_1(t), x_2(t), \dots, x_M(t)\}_{t=p}^{p+L-1}$, the coefficients $\beta_i(p, L)$ of the Lasso Regression model for each node i can be learned. The optimal α that achieves the best prediction is chosen from 50 logarithmically spaced points within $[10^{-4}, 10]$.

Random Forest is a non-linear ensemble learning algorithm for tasks such as classification [43], [44]. A large number of decision trees can be constructed from a training set. A decision tree is a flowchart-like structure in which each internal node represents a “test” on a feature, each branch represents the outcome of the test, and each leaf node represents a class label. The paths from root to leaf represent classification rules. Each tree is grown based on each training set $\mathcal{D}_i(p, L) = \{x_i(t+1); x_1(t), x_2(t), \dots, x_M(t)\}_{t=p}^{p+L-1}$ as follows: 1) choose randomly a set of m ($m \ll M$) features out of the M features as

the nodes in the tree 2) collect from each training sample the m features and the corresponding target 3) construct the decision tree based on the data collected from 2). The optimal m that leads to the highest prediction precision is chosen.

Random Forest could rank the importance of the features in estimating the target in a nonlinear way. Considering training set $\mathcal{D}_i(p, L) = \{x_i(t+1); x_1(t), x_2(t), \dots, x_M(t)\}_{t=p}^{p+L-1}$, the value of the j th feature in the first sample is $x_j(p)$, the value in the second sample is $x_j(p+1)$. The values for the j th feature are ordered as $\{x_j(p), x_j(p+1), \dots, x_j(p+L-1)\}$ from the first sample to the $(p+L)$ th sample. To measure the importance of the j th feature, its values $\{x_j(p), x_j(p+1), \dots, x_j(p+L-1)\}$ are randomized/permuted. Random Forest model is then trained by the original training set and the permuted training set respectively. The importance of a feature is reflected by the difference between the prediction errors of the model learned from the original and permuted training set respectively. The coefficient β_{ij} is obtained as the normalized difference in prediction error. A larger difference in prediction error means a larger contribution of the feature to the target prediction. We use TreeBagger implementation in Matlab with 1000 trees and use default values for other parameters.

2.4.2. TRAINING AND TEST DATA

The temporal network observed in each sub-window $[p, p+L]$ where $p \in [1, T-L-1]$ is considered as a training set and the learned model function will be tested in predicting the temporal network observed at $p+L+1$, using the temporal network observed at $p+L$. For each learning algorithm, the coefficients $\{\beta_i(p, L)\}$, $i = 1, 2, \dots, M$, learned from each training set $D_i(p, L)$ will be used to predict the activity of the links in the test set $Q_i(p, L) = \{x_i(p+L+1); x_1(p+L), x_2(p+L), \dots, x_M(p+L)\}$. In total, $T-L-1$ training sets, together with their corresponding test sets, will be considered for each temporal network.

2.4.3. BASELINE MODELS

We introduce two baseline models that predict a link's future activity based on its current activity. The probability that a link has the same state at two consecutive time steps is high, above 0.93 in each network. Hence, the baseline model 1 predicts the activity of a link at the next time step equal to the link's own activity at the current step, i.e., $x_i(t+1) = x_i(t)$. If link i is active (inactive) at time step $t-1$, then its state at t is predicted to active (inactive) in baseline model 1.

Baseline model 2 is the corresponding Lasso Regression, $x_i(t+1) = \beta_{i,i} x_i(t) + c_i$ where a link's current activity is a linear function its own previous activity. The same training and test sets have been used as introduced in section 2.4.1.

2.5. MODEL EVALUATION

For a given length L of the training sets, we evaluate each model via its average quality in predicting links' activities in a test set $Q_i(p, L)$ using the coefficients $\{\beta_i(p, L)\}$ learned from the corresponding training set $D_i(p, L)$, where $i = 1, 2, \dots, M$. The average is over all test sets, i.e., $p \in [1, T-L-1]$.

The prediction quality in a test set is measured via the area under the *ROC* curve

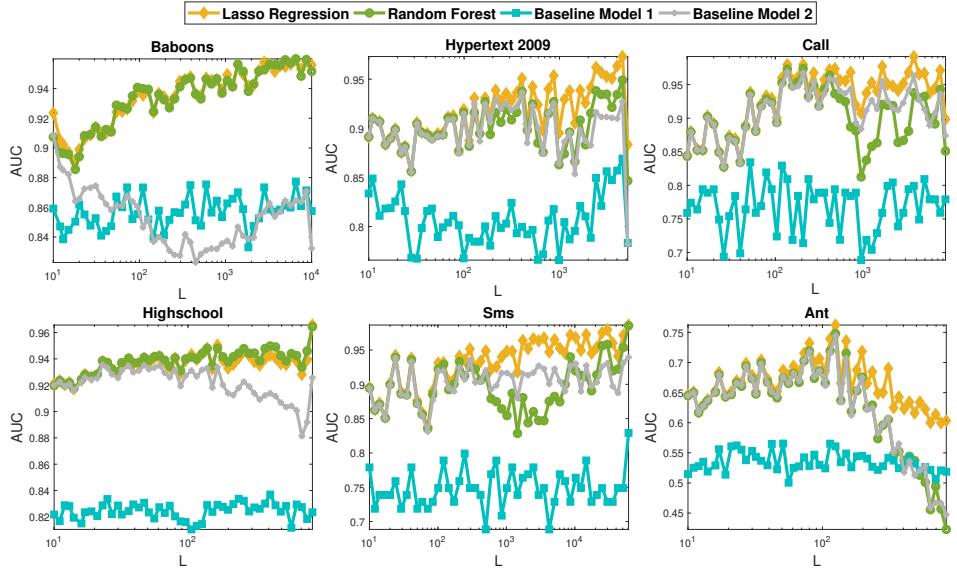


Figure 2.2: The prediction quality AUC for Lasso Regression, Random Forest and Baseline model 1 and 2 respectively in six temporal networks at different training set lengths L .

(AUC) [45], [46]. AUC provides an aggregate measure of performance across all possible classification thresholds. It ranges in value from 0 to 1. A high AUC implies high prediction quality.

Different lengths $L \in [1, T - 50]$ of the training set are considered when evaluating the performance of each model. The maximum $L_{max} = T - 50$ ensures a minimum of 50 training/test sets for each temporal network.

Figure 2.2 shows that the training set length L indeed affects the prediction quality AUC in all the networks. A relatively good performance tends to be obtained by a medium training length L , e.g., $L \sim 100$. A small length, e.g., $L \sim 10$ is insufficient for a model to learn the coefficients that to have reasonable prediction quality. A model with a large length may not capture the change of network dynamics over time, if there is. When the length is extremely large, e.g., $L \rightarrow T - 50$, the number of training set $T - (L + 1)$ is small and the corresponding test sets lie mainly at the end of the observation window $[0, T]$. Such boundary effect leads to low robustness of the model against abrupt change in data at the end of the observation window. For example, Supplementary Figure 2.11 shows an abrupt change in the total number of contacts per time step at the end of the time window in network Call. Correspondingly, the prediction quality AUC changes sharply when L is around $T - 50$.

Almost all the AUC values are larger than 0.5, which corresponds to the performance of random guessing. This suggests that all the models including the simple baseline models perform better than random guessing.

Lasso Regression performs the best in Hypertext 2009, Call, Sms and Ant networks.

And for Baboons and Highschool networks, Lasso regression and Random Forest perform comparably, better than the baseline models. Random Forest does not perform evidently better than the baseline model2 in Call, Sms and Ant, which have a lower number of contacts per step on average than the other networks (see Supplementary Figure 2.10). In general, the linear relationship of Lasso Regression models the link activity dynamic in temporal networks the best. In contrast, the baseline models that predict a link activity based on the link's own activity in the previous time step, gain a smaller *AUC*. Hence, the activities of other links contribute to the prediction of a given link's activity. Both Lasso Regression and Random Forest could achieve a reasonably good prediction quality via the choice of the training length L .

The area under the precision recall curve *AUPR* [47] is also considered to measure the prediction quality. It is considered as a more suitable measure for imbalanced classification problems. A larger *AUPR* suggests a better prediction quality. Similar results are obtained when *AUPR* is used to measure the link prediction quality for model evaluation (see Supplementary Figure 2.10). The prediction quality is the lowest in network Ant, which is possibly due to its lowest average number of contacts per link observed within the observation time window.

2.6. MODEL INTERPRETATION

The relatively good performance of the two models motivates us to further explore which links' activities influence a given link's activity more via the coefficients learned from the two models. We firstly introduce how to construct the prediction backbone network using the coefficients learned from a model. The backbone is a directed weighted network where nodes are the links in the aggregated network and weight B_{ij} , $i, j \in [1, M]$ represents the influence of link j in the aggregated network on link i in predicting link i 's activity. Furthermore, we unravel which links' activities influence a given link's activity more via analyzing properties of the backbone as well as its relation to the aggregated network and the time series of the links.

2.6.1. CONSTRUCTION OF THE PREDICTION BACKBONE NETWORK USING INFLUENCE COEFFICIENTS

The coefficients of each algorithm can be derived as follows. From a training set $D_i(p, L)$, where $i = 1, 2, \dots, M$, we can obtain the coefficients or coefficient matrix $\{\beta_{ij}(p, L)\}_{i,j=1}^M$ for each learning algorithm, either Lasso Regression or Random Forest. Each element $\beta_{ij}(p, L)$ indicates the contribution or influence of the activity of link j at a time $t - 1$ in determining the activity of link i at t , where $t \in [p + 1, L + p]$.

For each network, we consider from now on the training set length L at which the Lasso Regression obtains the maximal *AUC* value. Furthermore, we randomly choose 50 out of $T - (L + 1)$ training sets. We consider the coefficient matrices obtained from these 50 training sets via Lasso Regression and Random Forest, respectively, as samples to understand the influence between links.

We find a positive correlation between the coefficients $\beta_{ij}(p, L)$ obtained from the two algorithms respectively. Their Pearson correlation coefficients is higher than 0.5 in all networks and is higher than 0.8 in network Baboons, Hypertext2009, Call and Sms.

It indicates that the coefficients, i.e., the influence between links, obtained by these two learning algorithms are consistent with each other. Hence, we will focus on the coefficients and the corresponding prediction backbone of Lasso Regression since now on, which performs the best in link prediction.

2

The prediction backbone network can be constructed as follows. The nodes of the backbone correspond to the M links in the aggregated network. The backbone is a directed and weighted complete network with self-loops. The weight $B_{ij} = E[\beta_{ij}(p, L)]$ where $i, j \in [1, M]$ is the average of the coefficient over the 50 samples, representing the influence of link j in the aggregated network on link i in determining link i 's activity. The coefficient $\beta_{ij}(p, L)$ where $i, j \in [1, M]$ derived from a sample, possibly positive or negative, represents to what extent the contact of link j leads to the contact of link i at the next step. Among the 50 samples, the coefficients of any two samples are positively correlated on average. The average Pearson correlation of the coefficients from two random samples is 0.77, 0.72, 0.45, 0.74, 0.88 and 0.07 for Baboons, Hypertext 2009, Call, High-school, Sms, and Ant respectively. Hence, the weight $B_{ij} = E[\beta_{ij}(p, L)]$ in the backbone suggests the average influence of link j on link i in activity.

We evaluate to what extent a link's activity is influenced by the activity of its own and of the other links. The probability density function¹ $f_{B_{ij}}(x)$ where $i = j$ of the influence of a link on its own activity and $f_{B_{ij}}(x)$ where $i \neq j$ of the influence of a different link are given in Figure 2.3. The influence of the link itself $B_{i=j}$ tends to be larger than the influence of another link $B_{i \neq j}$ on link i 's activity in most networks except for Hypertext 2009 and Ant, where the self-influence $B_{i=j}$ can be negative. This suggests that Lasso's out-performance than the baseline model is because Lasso considers other links' influence on a given link's activity and Lasso considers a link's its own influence differently from the baseline model.

To have a better understanding of our backbone networks which are weighted directed complete graphs with self-loops, we visualize the sub-network of a backbone network. The sub-network includes only the none self-loop links in the backbone network that have the highest weights/influence and the corresponding end nodes of these links, such that the average degree of the sub-network is 2. We take the Sms backbone as an example, since it has the smallest number of nodes among all data sets, and visualize the sub-network of its backbone in Figure 2.4. Since the backbone is a directed network, a node pair in the sub-network may have none, one or two unidirectional links. Figure 2.4 shows that few node pair is connected by two unidirectional links or a bidirectional link, which is represented by two green links, whereas most node pairs are connected by an unidirectional link, which is colored in red. This suggests that a high weight B_{ij} of link from i to j in the backbone does not imply a high weight of B_{ji} . This observation is in line with the weak correction $\rho(B_{ij}, B_{ji})$ between the weight of the two reciprocal links of a node pair in the original un-sampled backbone network, as shown in in Table 2.2. Furthermore, node size and node color in Figure 2.4 are proportional to the node's in-strength and out-strength in the sub-graph respectively. A dark blue (white) color of a node represents a large (small) out-strength. We find that a node with a large in-

¹The probability density function $f_{B_{ij}}(x)$ of a continuous variable B_{ij} is defined as $f_{B_{ij}}(x) = \lim_{\Delta x \rightarrow 0} \frac{Pr[x < B_{ij} \leq x + \Delta x]}{\Delta x}$, the probability that the variable is within each range or bin $(x, x + \Delta x]$ normalized by the size of the bin Δx .

strength in the sub-network does not necessarily have a large out-strength. This finding via visualization is consistent with the weak Pearson correlation $\rho(S_{in}, S_{out})$ between the in-strength and out-strength of a node in the original un-sampled backbone network, as given in Table 2.2.

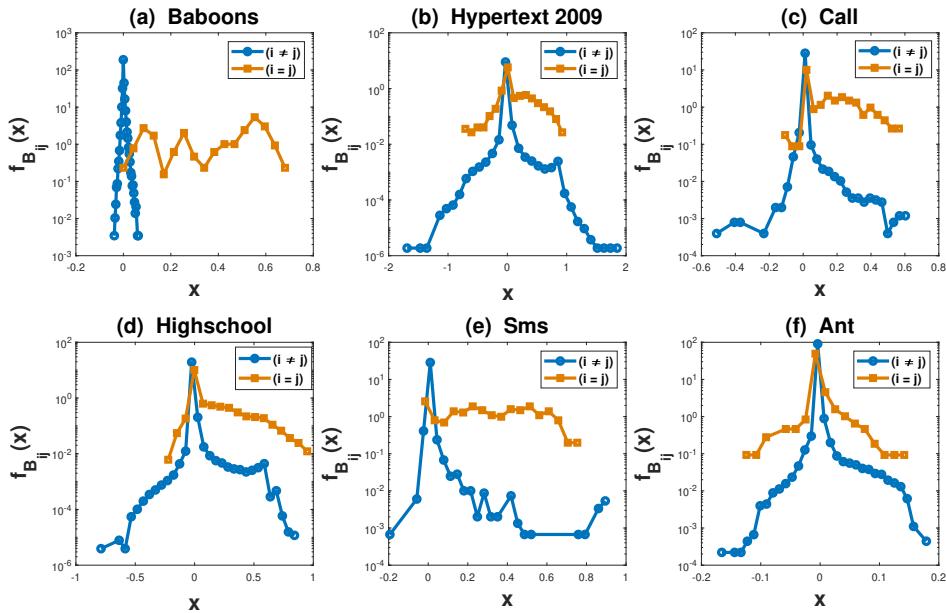


Figure 2.3: The probability density function $f_{B_{ij}}(x)$ of the weight B_{ij} in the backbone network, when $i = j$ and $i \neq j$ respectively.

2.6.2. BACKBONE NETWORK IN RELATION TO TIME SERIES

To understand which kind of links influence a given link i 's activity more, we explore the relation between the weight B_{ij} in the backbone and the correlation of the time series corresponding to link i and j .

Table 2.2: The Pearson correlation coefficient $\rho(S_{in}, S_{out})$ between in-strength and out-strength of node, and $\rho(B_{ij}, B_{ji})$ between the weight $B(i, j)$ and $B(j, i)$ of the two reciprocal links of a node pair in the backbone network.

Network	$\rho(S_{in}, S_{out})$	$\rho(B_{ij}, B_{ji})$
Baboons	0.30	0.27
Hypertext 2009	-0.02	-0.01
Call	0.01	0.01
Highschool	-0.22	0.00
Sms	-0.40	0.08
Ant	0.19	-0.01

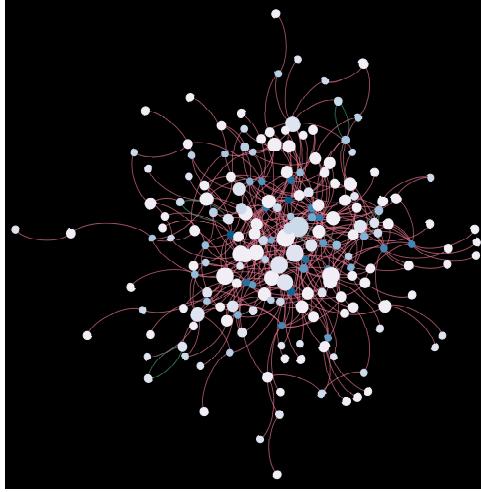


Figure 2.4: Visualization of a sub-network of the SMS backbone network. The sub-network is composed of only the non self-loop links in the backbone network that have the highest weights and their corresponding end nodes, such that average degree of the sub-networks is 2. When a node pair is connected by one (two) unidirectional links, the connection is represented by a link in red (two links in green). Node size and node color are proportional to the node's in-strength and out-strength in the sub-graph respectively. A dark blue (white) color of a node represents a large (small) out-strength.

We first explore whether the relatively high coefficients B_{ii} can be explained by the auto-correlation of a link i 's time series $\{x_i(t)\}_{t=1,2,\dots,T}$. Auto-correlation describes the degree of similarity between a given time series and its lagged version. It measures the correlation between current value of a time series and its past value. Our models use links' activities at the previous time step to predict a link's activity at current time step. Hence, we compute, for each link i , the Pearson correlation coefficient $R_{x_i x_i}(t, t - 1)$ between $\{x_i(t)\}_{t=1,2,\dots,T-1}$ and $\{x_i(t)\}_{t=2,3,\dots,T}$ as its auto-correlation coefficient. The distribution of the auto-correlation coefficient of a link in each empirical temporal network is shown in Figure 2.5. In networks, such as Baboons (Hypertext 2009 and Ant) where the average auto-correlation coefficient is high (low), the self-influence $B_{i=j}$ tends (not) to be evidently larger than the influence of another link $B_{i \neq j}$ on the given link i 's activity. Moreover, we find that the ranking of these networks in the average number of contacts within the observation time window per link (see Table 2.1) is the same as their ranking in the average auto-correlation (see Figure 2.5). Hence, a network with a large average number of contacts per link tends to have a high auto-correlation. Correspondingly, its self-influence $B_{i=j}$ in the backbone tends to be evidently larger than the influence of another link $B_{i \neq j}$ (see Figure 2.3).

Similarly, we study further the relation between the coefficient B_{ij} and the Pearson correlation coefficient $R_{x_i x_j}(t, t - 1)$ between $\{x_i(t)\}_{t=2,3,\dots,T}$ and $\{x_j(t)\}_{t=1,2,\dots,T-1}$, where $i \in [1, M]$ and $i \neq j$. This aims to understand whether the influence B_{ij} of another link j on i can be explained by the cross correlation $R_{x_i x_j}(t, t - 1)$ between the two links' activity

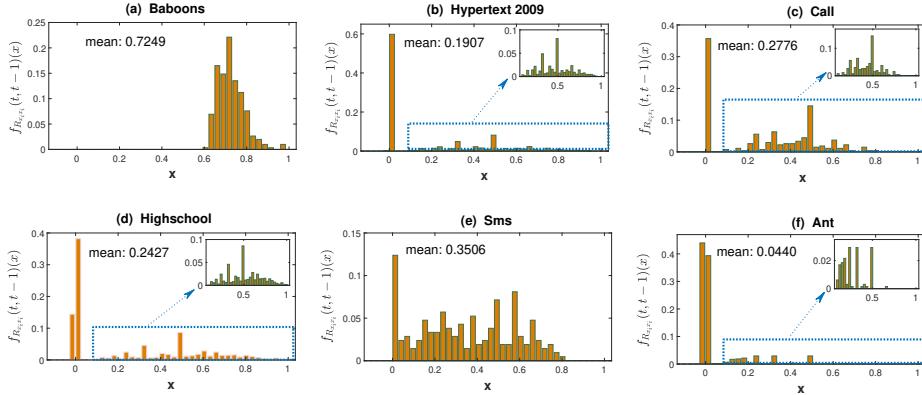


Figure 2.5: The probability density function $f_{R_{x_i x_i}(t, t-1)}(x)$ of the auto-correlation coefficient $R_{x_i x_i}(t, t-1)$ in each of the six networks.

series. The cross correlation is non-reciprocal $R_{x_i x_j}(t, t-1) \neq R_{x_j x_i}(t, t-1)$. The Pearson correlation coefficients between $R_{x_i x_j}$ and $R_{x_j x_i}$ are 0.99, 0.29, 0.25, 0.42, 0.68 and 0.05 for Baboons, Hypertext 2009, Call, Highschool, Sms and Ant respectively. Specifically, we compute the Pearson correlation coefficient at each node i between the influence B_{ij} and cross correlation $R_{x_i x_j}(t, t-1)$ where $i \in [1, M]$ and $i \neq j$. The probability density function of this Pearson correlation coefficient at a random node i in Figure 2.6 shows that the influence B_{ij} tends to be positively correlated with thus can be partly explained by the cross correlation coefficient $R_{x_i x_j}(t, t-1)$. The correlation $R_{x_i x_j}(t, t-1)$ between the activities of link i and j allows our models to predict the activity of link i at t using the activity of link j at $t-1$. The Pearson coefficient BC between B_{ij} and cross-correlation $R_{x_i x_j}$ is the strongest in Sms. The skewed probability density function $f_{R_{x_i x_j}(x)}$ of the cross-correlation in Sms (see Supplementary Figure 2.12) explains the skewed distribution $f_{B_{ij}}(x)$ where $i \neq j$ of the influence of another link in the backbone (see Figure 2.3).

2.6.3. THE BACKBONE NETWORK IN RELATION TO THE AGGREGATED NETWORK

The activities of other links have been shown to contribute to a better prediction of a link's activity. We would like to explore which kind of other links have more influence B_{ij} on a link. Would links that are close in the aggregated network tend to have a high influence on each other in activity prediction? Firstly, we define the topological distance or hopcount between two links in the aggregated network. The topological distance, also called hopcount, between two nodes on a static network is the number of links along the shortest path between the two nodes. The distance H_{ij} between two links i and j is defined on the aggregated network. The distance H_{ii} between the same link i is 0. The distance H_{ij} between two different links i and j is defined as the minimal hopcount between one end node of link i and one end node of link j plus one. Hence, two links that share one end node in common have a hopcount 1. The line graph, e.g., G_w^* of an aggregated network G_w can be constructed by considering each link in G_w as a node, and

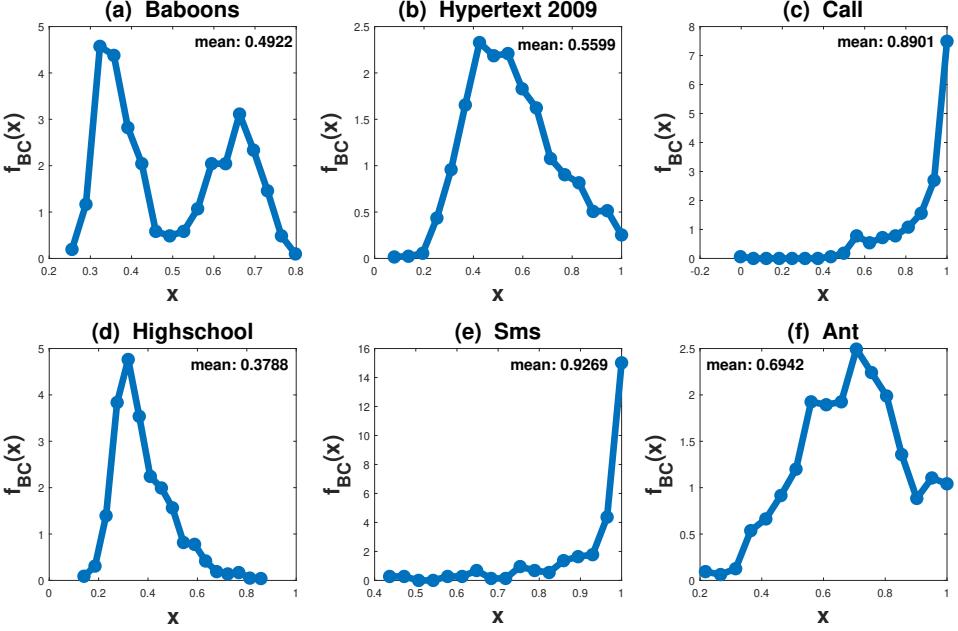


Figure 2.6: The probability density distribution of Pearson coefficient BC between B_{ij} and cross-correlation coefficient $R_{x_i x_j}$.

two nodes are connected in G_w^* if the two corresponding links in G_w share a common end node. The distance between two links in G_w equals the hopcount between their corresponding nodes in the line graph G_w^* .

Figure 2.7 shows the distribution $Pr[H = k]$ of link distance for all the six networks. The distance is in general small, except that few nodes are isolated from the largest connected component, leading to an infinite distance to the other nodes. Networks that are measured within a small (large) spatial space like Ant, Baboons, Hypertext (Call, Sms, Highschool) tend to have a small (large) average hopcount and a large (small) link density in the aggregated network, which is the number of links M normalized by $N(N-1)/2$. Of course, the average hopcount of a network could be also influenced by the number N of nodes.

Ref. [48] has shown that nodes that are closer to a target node have more influence on the target node's state. So here we further explore whether links that are closer in topological distance tend to have more influence on each other's activity. Figure 2.8 illustrates the influence between two links $E[B_{ij}|H_{ij} = k]$ given their topological distance. Still, we focus on Lasso Regression. The influence $E[B_{ij}|H_{ij} = k]$ in general decreases with k . This is also supported by the negative Pearson correlation between the B_{ij} and the distance H_{ij} of two links for Baboons (-0.32), Hypertext 2009 (-0.13), Call (-0.014), Highschool (-0.068), Sms (-0.035) and Ant (-0.027). In network call and Ant, the probability that two links have a distance 4 or larger is small. The average influence $E[B_{ij}|H_{ij} = k]$ when $k \leq 4$ is thus derived from few link pairs and can be noisy. Since B_{ij} can be nega-

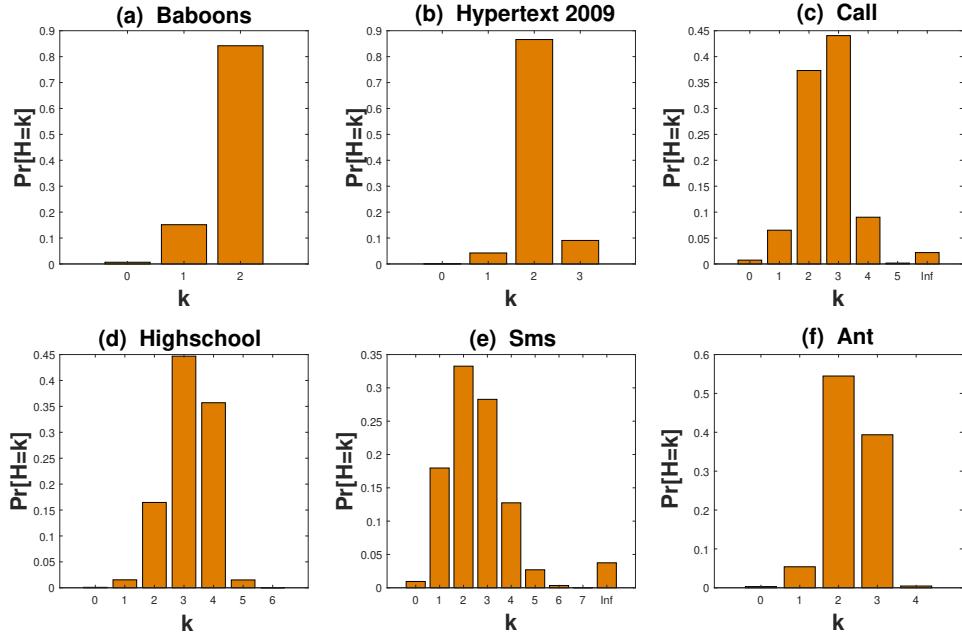


Figure 2.7: The probability distribution $Pr[H = k]$ of the topological distance H of two random links in the aggregated network. All six real-world networks are considered.

tive, $E[B_{ij}|H_{ij} = k]$ represents to what extent on average the contact of link j could lead to the contact of link i given that i and j have a distance k . The average strength of influence $E[|B_{ij}| | H_{ij} = k]$ also tends to decrease with k , as shown in Supplementary Figure 2.13.

It is interesting to notice that the average weight $E[B_{ij}|H_{ij} = \text{inf}]$ is none zero when the distance is infinity in network Call and Sms. Normally, it is assumed that links can influence each other at least when they are connected. The none zero coefficient $E[B_{ij}|H_{ij} = \text{inf}]$ may suggest that the two links are possibly connected in the aggregated network, but not observable in the current aggregated network or the temporal network measured within the short time window $[1, T]$.

Figure 2.8 shows that $E[B_{ij}|H_{ij} = k]$ is relatively large when $k = 1, 2$. We consider further the link pairs that have a distance $H_{ij} = 1$ and $H_{ij} = 2$ respectively. Among those link pairs, we explore whether link pairs that are well connected tend to have a high influence. How well link i and j are connected can be measured by the number of common neighbors n_{ij} of their corresponding nodes in the line graph G_w^* . Figure 2.9 demonstrates the influence between two links $E[B_{ij}|H_{ij} = 1, n_{ij} = k]$ and $E[B_{ij}|H_{ij} = 2, n_{ij} = k]$ respectively as a function k of the number of common neighbors. We find no clear relation between $E[B_{ij}|H_{ij} = 1, n_{ij} = k]$ and k . The distance $H_{ij} = 1$ means that node i and j are connected by a link in the line graph G_w^* . In this case, the number of common neighbors, which equals the number of two-hop paths, paths that are longer than the shortest path is not correlated with the influence. However, $E[B_{ij}|H_{ij} = 2, n_{ij} = k]$ does increase with

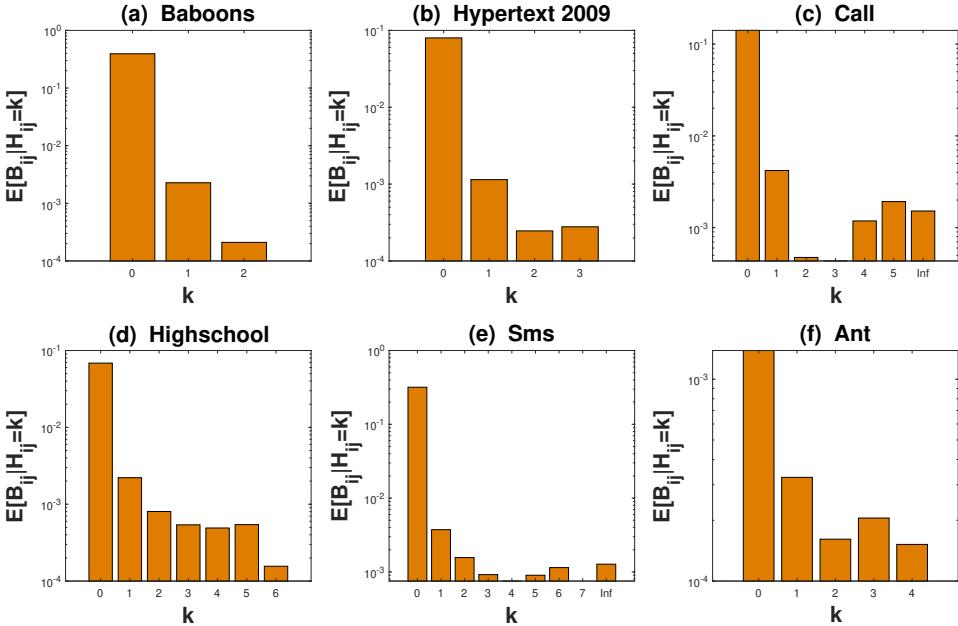


Figure 2.8: The average weight in Lasso Regression $E[B_{ij} | H_{ij} = k]$ given the distance of the two links in each of the six networks.

k. Node pairs that have more shortest paths, i.e., two hop paths tend to have a higher influence. Similarly the average influence strength $E[|B_{ij}| | H_{ij} = 2, n_{ij} = k]$ also tends to increase with k , as shown in Supplementary Figure 2.14. Links that have a short distance and are connected by many shortest paths tend to influence each other strongly.

2.7. LONG-TERM EFFECT

The Lasso regression assumes that a link's current state is a linear function of the activities of all the links at the previous step. In this section, we generalize this assumption as:

$$x_i(t) = \sum_{j=1}^M \{x_j(t-1)\beta_{ij}^1 + x_j(t-2)\beta_{ij}^2 + \dots + x_j(t-p)\beta_{ij}^p + c_i\}. \quad (2.5)$$

where a link's current state is a linear function of the activities of all the links in the previous p steps. The Lasso regression (see Eq. 3.3) that we have investigated in the previous sections corresponds to the case when $p = 1$. The prediction quality *AUC* is shown in Table 2.3 for different choices of p . In general, considering the activity of a longer influence period, i.e., $p > 1$ can hardly improve the link prediction quality. The same is observed when the prediction quality is evaluated via *AUPR* (see Supplementary Table 2.4). The current activity of a link is mainly influenced by the activities of links at the previous time step.

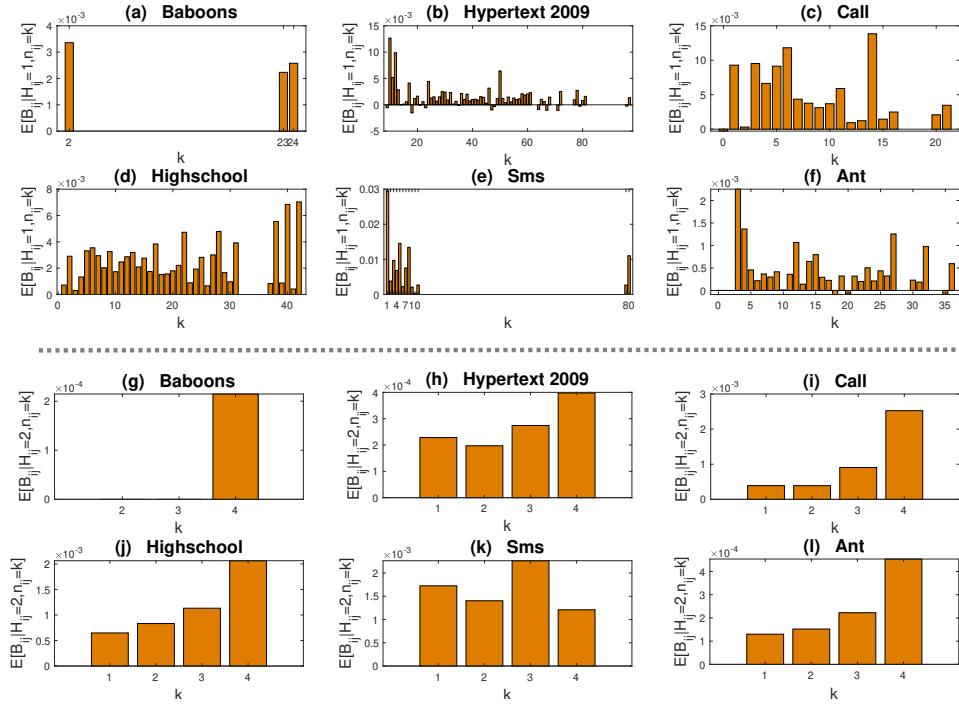


Figure 2.9: The average weight $E[B_{ij} | H_{ij} = 1, n_{ij} = k]$ and $E[B_{ij} | H_{ij} = 2, n_{ij} = k]$ in Lasso Regression coefficient given the distance H_{ij} and number of common neighbors n_{ij} between link i and j .

The small memory length p that we have considered is insufficient to capture periodic or pseudo-periodic behavior, in view the time resolution or duration of a time step, which is in the order of seconds. This choice is limited by two factors: the observation window of most real-world temporal networks is short and the computational complexity of Lasso Regression [49] is high: $O((Mp)^3 + (Mp)^2L)$, where M and L are number of nodes in the backbone and the length of training set, respectively. We deem it as an important future work to explore how underlying periodic behavior in the temporal network is captured by a learning model and influences the link prediction. Our finding with regard to which kind of links in the backbone tend to have a high weight may shed light on the selection of model features to reduce the computational complexity.

2.8. CONCLUSION

In this work, we illustrate our method that enables interpretable temporal network prediction. Interpretable learning algorithm Lasso Regression and Random Forest are employed to predict the activity (connected or not) of each link at the next time step based on the current activities of all links. The coefficients learned from each algorithm are further used to construct the prediction backbone network, presenting the influence or

Table 2.3: The prediction quality AUC for Lasso Regression for different influence period p .

Network	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
Hypertext 2009	0.97	0.97	0.97	0.93	0.93
Highschool	0.93	0.94	0.93	0.93	0.93
Call	0.97	0.97	0.96	0.96	0.96
Sms	0.96	0.96	0.96	0.96	0.96
Baboons	0.96	0.96	0.96	0.96	0.96
Ant	0.76	0.77	0.76	0.76	0.76

contribution of all links in determining each link's activity. Via exploring the properties of the backbone network and its relation to the activity time series of links and its relation to the aggregated network, we find the following in six real-world physical and virtual contact networks. A link's next step activity is mainly influenced by the current activity of the link itself and of other links that are better connected with the link. Two links are better connected if they have shorter and/or more shortest paths in the aggregated network. The influence between two links tend to be large if their corresponding activity time series are strongly correlated. Hence, the learning algorithm also captures the underlying network properties and correlation in activity time series, which are usually utilized by network property based prediction methods. Finally, both algorithms' performance can be hardly improved by considering the activities of more than one time steps in the past. The physical contact networks considered differ in the average number of contacts per link observed due the nature of these networks and the methods they are measured or defined. A low average number of contacts per link, e.g., in Ant, may suggest a low prediction quality, a low auto-correlation of the time series in the network and a less evident self-influence in the backbone.

These findings, when combined with modern deep learning techniques, can potentially lead to interpretable and more accurate prediction. Our findings may also shed lights on the modeling of the long-term temporal network evolution, in contrast to short-term network prediction. Such models are crucial to forecast the long-term performance of e.g. epidemic/information spreading on the network. The linear regression assumed by Lasso could be one elementary mechanism to model temporal networks. The influence patterns that we have discovered can be further used to adapt other dynamic processes to model temporal networks. Our findings of the backbone network and its association with other network properties may also inspire the solution of network classification and optimization problems. For example, the spread of epidemic/information can be mitigated by blocking the temporal interactions of selected links [50], [51]. The influence of a link on and by the other links could possibly help with the selection of the links to block.

High-order models have been explored recently to account for various types of high-order dependencies in data on complex systems [52]. It has been found, for example, high-order models of paths in temporal networks, could improve node ranking [53], [54] and community detection [55]. Our Lasso Regression and Random Forest Model are high-order in broad sense: the state of a node pair depends on the states of many node pairs in the past over a period. These models together with the interpretations may shed

Table 2.4: The prediction quality AUPR of Lasso Regression for different influence period p .

Network	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
Hypertext 2009	0.23	0.25	0.26	0.08	0.09
Highschool	0.59	0.59	0.59	0.59	0.59
Call	0.11	0.13	0.14	0.14	0.14
Sms	0.14	0.15	0.13	0.07	0.07
Baboons	0.71	0.71	0.72	0.72	0.72
Ant	0.04	0.05	0.03	0.04	0.04

light on the possible mechanism by which the temporal network may emerge and patterns in paths [52] may emerge.

2.9. APPENDIX

2.9.1. DATA DESCRIPTION

Here, we describe the data we used.

Baboons: 10 testing booths are connected with an outdoor enclosure ($100m^2$) where the baboons live. At each time step, a temporal contact is recorded for two baboons if they are present in two adjacent booths.

Ant: At each time step of half a second, the location and body angle of every ant is recorded. Two ants are considered to have an interaction if the angle between their bodies is greater than 70° and if there is more than 1 second in a position where at least one ant could reach the body of the other. Any interaction that lasts between 1 and 10 seconds is recorded as a contact happening at the first time step that this interaction has been observed. An interaction that lasts 20 seconds, for example, will be recorded as two contacts separated by 10 seconds in time.

Sms (Call): At each time step, two nodes have a contact if one sends a message (make a call) to the other at that time step. We consider the undirected temporal network.

Highschool (Hypertext 2009) records the face-to-face proximity of students (conference participants) via radio badges they wear. At each time step, two individuals are considered to have a contact if their sensors exchanged at least one packet during that time step.

2.9.2. SUPPLEMENTARY FIGURES AND TABLES

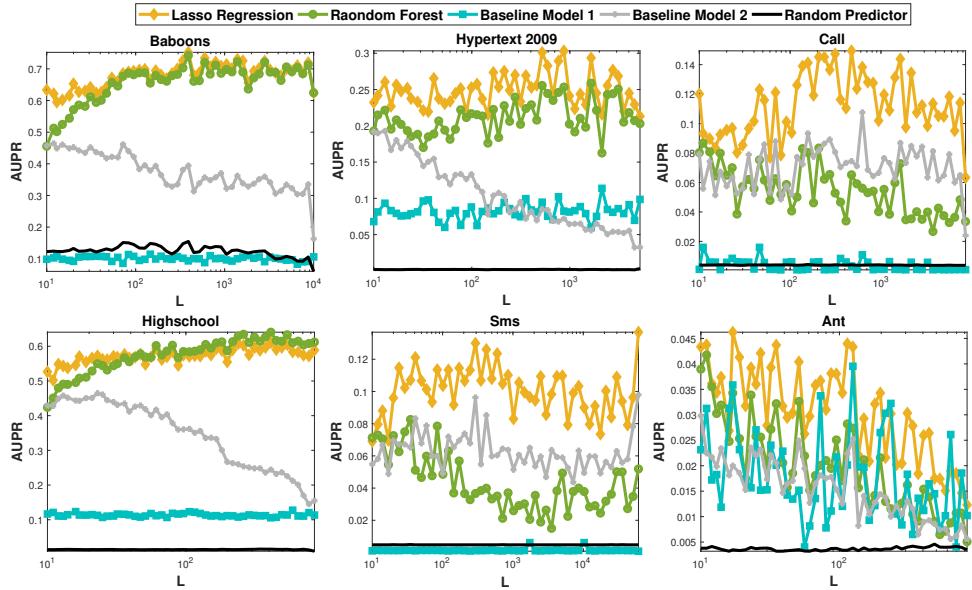


Figure 2.10: The prediction quality $AUPR$ of Lasso Regression, Random Forest and Baseline model respectively in six temporal networks at different training set length L .

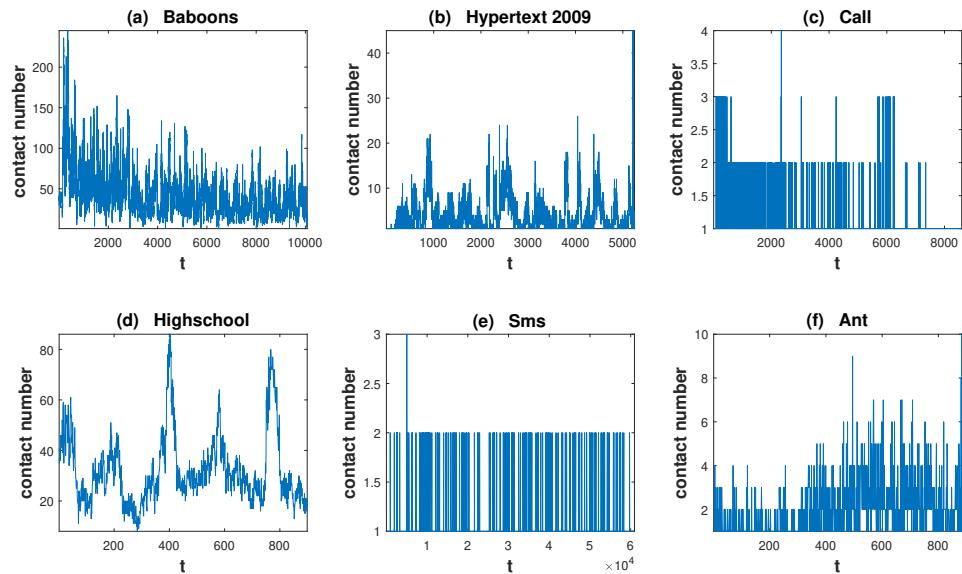


Figure 2.11: The total number of contacts per time step in each of the six temporal networks.

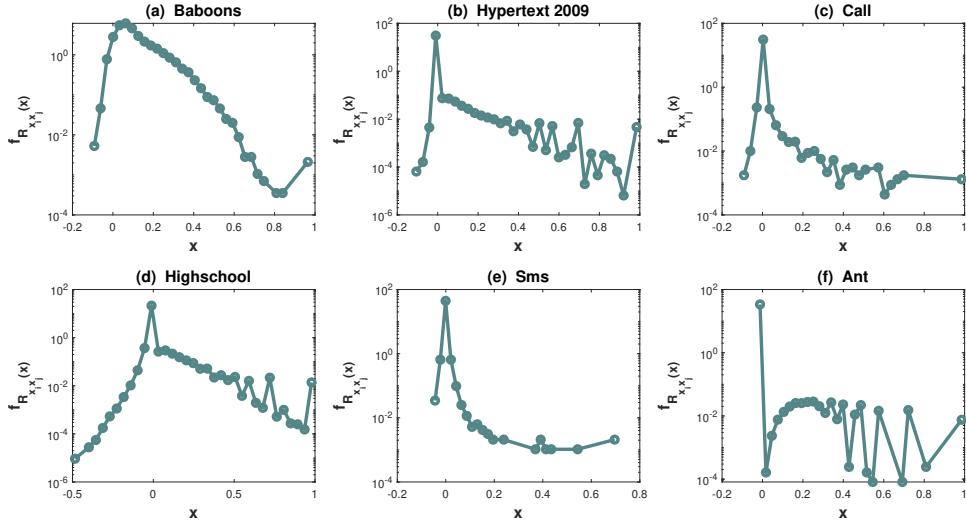


Figure 2.12: Probability density function $f_{R_{x_i x_j}}(x)$ of the cross-correlation between two random link's activity time series.

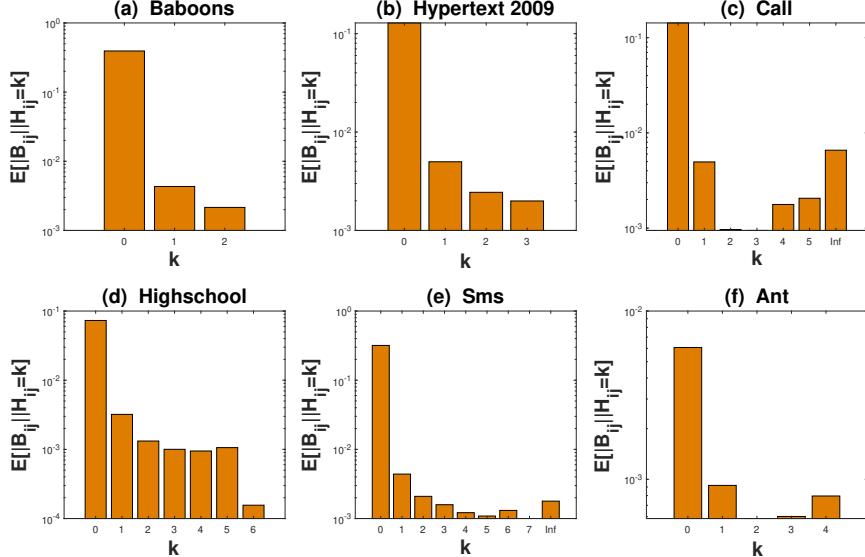


Figure 2.13: The average influence strength in Lasso Regression $E[|B_{ij}| | H_{ij} = k]$ given the distance of the two links k in each of the six networks.

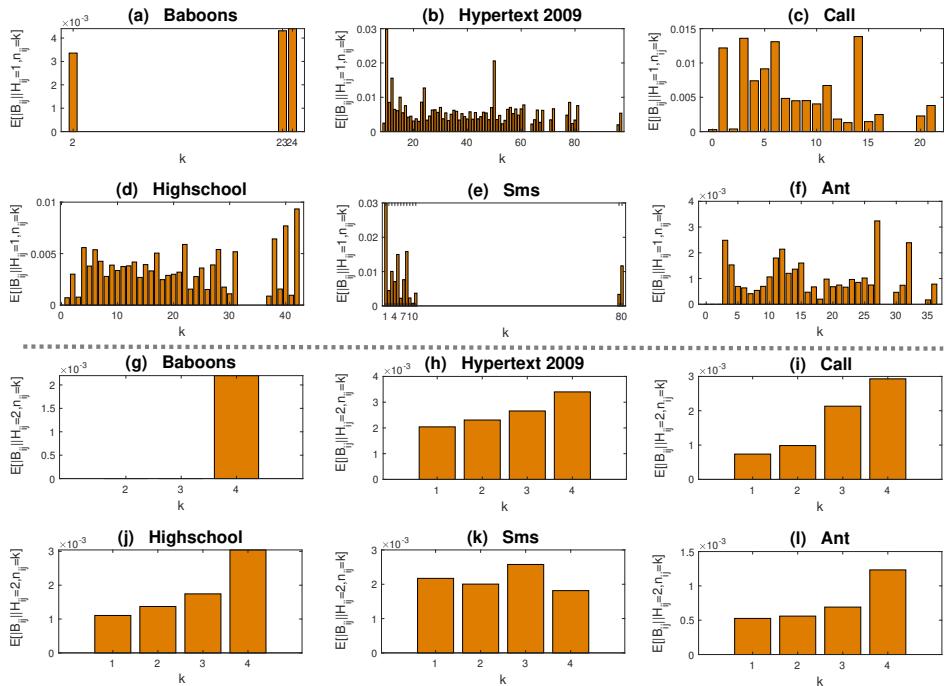


Figure 2.14: The average weight $E[|B_{ij}| | H_{ij} = 1, n_{ij} = k]$ and $E[|B_{ij}| | H_{ij} = 2, n_{ij} = k]$ in Lasso Regression given the distance H_{ij} and number of common neighbors n_{ij} between link i and j .

REFERENCES

- [1] P. Holme and J. Saramäki, "Temporal networks," *Phys. Rep.*, vol. 519, no. 3, pp. 97–125, 2012, ISSN: 0370-1573.
- [2] N. Masuda and R. Lambiotte, *A Guide to Temporal Networks* (complexity science). World scientific (Europe), 2016, vol. 4.
- [3] P. Holme and J. Saramäki, *Temporal Network Theory* (Computational Social Sciences), 1st ed. Springer International Publishing, 2019.
- [4] B. Min, K.-I. Goh, and A. Vazquez, "Spreading dynamics following bursty human activity patterns," *Phys. Rev. E*, vol. 83, p. 036102, 2011.
- [5] G. Grinstein and R. Linsker, "Power-law and exponential tails in a stochastic priority-based model queuens," *Phys. Rev. E*, vol. 77, p. 012101, 2008.
- [6] H.-H. Jo, M. Karsai, J. Kertész, and K. Kaski, "Circadian pattern and burstiness in mobile phone communication," *New J. Phys.*, vol. 14, p. 013055, 2012, ISSN: 1367-2630.
- [7] M. Morris and M. Kretzschmar, "Concurrent partnerships and transmission dynamics in networks," *Soc. Networks*, vol. 17, pp. 299–318, 1995.
- [8] S. Bansal, J. Read, B. Pourbohloul, and L. A. Meyers, "The dynamic nature of contact networks in infectious disease epidemiology," *J. Biol. Dyn.*, vol. 4, no. 5, pp. 478–489, 2010.
- [9] N. Masuda, K. Klemm, and V. M. Eguiluz, "Temporal networks: Slowing down diffusion by long lasting interactions," *Phys. Rev. Lett.*, vol. 111, no. 18, p. 188701, 2013.
- [10] B. Ribeiro, N. Perra, and A. Baronchelli, "Quantifying the effect of temporal resolution on time-varying networks," *Sci. Rep.*, vol. 3, no. 1, p. 3006, 2013.
- [11] J. L. Iribarren and E. Moro, "Impact of human activity patterns on the dynamics of information diffusion," *Phys. Rev. Lett.*, vol. 103, no. 3, p. 038702, 2009.
- [12] M. Karsai, M. Kivelä, R. K. Pan, *et al.*, "Small but slow world: How network topology and burstiness slow down spreading," *Phys. Rev. E*, vol. 83, no. 2, pp. 1550–2376, 2011.
- [13] T. Takaguchi, N. Masuda, and P. Holme, "Bursty communication patterns facilitate spreading in a threshold-based epidemic dynamics," *Plos ONE*, vol. 8, e68629, 2013.
- [14] X.-X. Zhan, A. Hanjalic, and H. Wang, "Information diffusion backbones in temporal networks," *Sci. Rep.*, vol. 9, p. 6798, 2019, ISSN: 1.
- [15] X.-X. Zhan, Z. Li, N. Masuda, P. Holme, and H. Wang, "Susceptible-infected-spreading-based network embedding in static and temporal networks," *EPJ Data Science*, vol. 9, p. 30, 2020, ISSN: 1.
- [16] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, "Recommender systems," *Phys. Rep.*, vol. 519, no. 1, pp. 1–49, 2012, Recommender Systems, ISSN: 0370-1573.

- 2
- [17] A. Aleta, M. Tuninetti, D. Paolotti, Y. Moreno, and M. Starnini, “Link prediction in multiplex networks via triadic closure,” *Phys. Rev. Res.*, vol. 2, 042029(R), 2020.
 - [18] C. Ma, T. Zhou, and H.-F. Zhang, “Playing the role of weak clique property in link prediction: A friend recommendation model,” *Sci. Rep.*, vol. 6, p. 30098, 2016.
 - [19] D. Tang, W. Du, L. Shekhtman, *et al.*, “Predictability of real temporal networks,” *Natl. Sci. Rev.*, vol. 7, no. 5, pp. 929–937, 2020.
 - [20] J. G. Kemeny and J. L. Snell, *Markov Chains*. New York: Springer-Verlag, 1976.
 - [21] S. Das and S. K. Das, “A probabilistic link prediction model in time-varying social networks,” in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6.
 - [22] R. R. Sarukkai, “Link prediction and path analysis using markov chains,” *Computer Networks*, vol. 33, pp. 377–386, 2000.
 - [23] L. Zhou, Y. Yang, X. Ren, F. Wu, and Y. Zhuang, “Dynamic network embedding by modeling triadic closure process,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, Apr. 2018.
 - [24] G. H. Nguyen, J. B. Lee, R. A. Rossi, N. K. Ahmed, E. Koh, and S. Kim, “Continuous-time dynamic network embeddings,” in *Companion Proceedings of the The Web Conference 2018*, ser. WWW ’18, Lyon, France: International World Wide Web Conferences Steering Committee, 2018, pp. 969–976.
 - [25] M. Rahman, T. K. Saha, M. A. Hasan, K. S. Xu, and C. K. Reddy, *Dylink2vec: Effective feature representation for link prediction in dynamic networks*, 2018. arXiv: [1804.05755](https://arxiv.org/abs/1804.05755).
 - [26] X. Yu and T. Chu, “Dynamic link prediction using restricted boltzmann machine,” in *2017 Chinese Automation Congress (CAC)*, 2017, pp. 4089–4092.
 - [27] X. Li, N. Du, H. Li, K. Li, J. Gao, and A. Zhang, “A deep learning approach to link prediction in dynamic networks,” in *Proceedings of the 2014 SIAM International Conference on Data Mining (SDM)*, pp. 289–297.
 - [28] J. Chen, X. Xu, Y. Wu, and H. Zheng, “Gc-lstm: Graph convolution embedded lstm for dynamic link prediction,” *CoRR*, 2018. arXiv: [1812.04206v1](https://arxiv.org/abs/1812.04206v1).
 - [29] A. Pareja, G. Domeniconi, J. Chen, *et al.*, “Evolvegcn: Evolving graph convolutional networks for dynamic graphs,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, pp. 5363–5370, Apr. 2020.
 - [30] J. Li, J. Peng, S. Liu, and C. L. L. Weng, “Tsam: Temporal link prediction in directed networks based on self-attention mechanism,” *CoRR*, 2021. arXiv: [2008.10021v1](https://arxiv.org/abs/2008.10021v1).
 - [31] L. Qu, H. Zhu, Q. Duan, and Y. Shi, “Continuous-time link prediction via temporal dependent graph neural network,” *Proceedings of The Web Conference 2020*, 2020.
 - [32] *Hypertext 2009 network dataset-konect*, <http://konect.uni-koblenz.de/networks/sociopatterns-hypertext>.
 - [33] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. V. den Broeck, “What’s in a crowd? analysis of face-to-face behavioral networks,” *J. Theor. Biol.*, vol. 271, pp. 166–180, 2011, ISSN: 0022-5193.

- [34] *Sociopatterns*, <http://www.sociopatterns.org/datasets/>.
- [35] P. Sapiezynski, A. Stopczynski, L. D. D., and L. S., “Interaction data from the copenhagen networks study,” *Sci. Data*, vol. 6, p. 315, 2019.
- [36] V. Gelardi, J. Fagot, A. Barrat, and N. Claidière, “Detecting social (in)stability in primates from their temporal co-presence network,” *Animal Behaviour*, vol. 157, pp. 239–254, 2019.
- [37] D. P. Mersch, A. Crespi, and L. Keller, “Tracking individuals shows spatial fidelity is a key regulator of ant social organization,” *Science*, vol. 340, pp. 1090–1093, 2013.
- [38] A.-L. Barabási, “The origin of bursts and heavy tails in human dynamics,” *Nature*, vol. 435, pp. 207–211, 2005.
- [39] A. Moinet, M. Starnini, and R. Pastor-Satorras, “Burstiness and aging in social temporal networks,” *Physical review letters*, vol. 114 10, p. 108701, 2015.
- [40] B. Min and K. I. Goh, “Burstiness: Measures, models, and dynamic consequences,” English, in *Temporal Networks* (Understanding Complex Systems), Understanding Complex Systems. Springer Verlag, 2013, pp. 41–64, ISBN: 9783642364600. DOI: [10.1007/978-3-642-36461-7_3](https://doi.org/10.1007/978-3-642-36461-7_3).
- [41] F. Santosa and W. W. Symes, “Linear inversion of band-limited reflection seismograms,” *SIAM J. Sci. and Stat. Comput.*, vol. 7, no. 4, pp. 1307–1330, 1986.
- [42] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *J. R. Stat. Soc. Series B*, vol. 58, no. 1, pp. 267–88, 1996.
- [43] T. K. Ho, “The random subspace method for constructing decision forests,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.
- [44] E. M. Kleinberg, “Stochastic discrimination,” *Ann. Math. Artif. Intell.*, vol. 1, pp. 207–239, 1990.
- [45] F. Tom, “An introduction to roc analysis,” *Pattern Recognition Letters*, vol. 27, pp. 861–874, 2006.
- [46] L. Zou, C. Wang, A. Zeng, Y. Fan, and Z. Di, “Link prediction in growing networks with aging,” *Social Networks*, vol. 65, pp. 1–7, 2021.
- [47] M. Kitsak, I. Voitalov, and D. Krioukov, “Link prediction with hyperbolic geometry,” *Phys. Rev. Res.*, vol. 2, p. 043113, 4 Oct. 2020.
- [48] J. Sun, F. Quevedo, and E. M. Boltt, “Data fusion reconstruction of spatially embedded complex networks,” *preprint at arXiv:1707.00731v1*, 2017. arXiv: [1707.00731](https://arxiv.org/abs/1707.00731).
- [49] B. Efron, T. Hastie, and I. Johnstone, “Least angle regression,” *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [50] X.-X. Zhan, A. Hanjalic, and H. Wang, “Suppressing information diffusion via link blocking in temporal networks,” in *Complex Networks and Their Applications VIII*, H. Cherifi, S. Gaito, J. F. Mendes, E. Moro, and L. M. Rocha, Eds., Cham: Springer International Publishing, 2020, pp. 448–458, ISBN: 978-3-030-36687-2.

- 2
- [51] A. Arenas, W. Cota, J. Gómez-Gardeñes, *et al.*, “Modeling the spatiotemporal epidemic spreading of covid-19 and the impact of mobility and social distancing interventions,” *Phys. Rev. X*, vol. 10, p. 041 055, 4 2020.
 - [52] R. Lambiotte, M. Rosvall, and I. Scholtes, “From networks to optimal higher-order models of complex systems,” *Nature Physics*, vol. 15, pp. 313–320, 2019.
 - [53] I. Scholtes, N. Wider, and A. Garas, “Higher-order aggregate networks in the analysis of temporal networks: Path structures and centralities,” *Eur. Phys. J. B*, vol. 89, p. 61, 2016.
 - [54] M. Rosvall, A. V. Esquivel, A. Lancichinetti, and J. D. West, “Memory in network flows and its effects on spreading dynamics and community detection,” *Nat. Commun.*, vol. 5, p. 5630, 2014.
 - [55] A. R. Benson, D. F. Gleich, and J. Leskovec, “Higher-order organization of complex networks,” *Science*, vol. 353, pp. 163–166, 2016.

3

SHORT- AND LONG-TERM TEMPORAL NETWORK PREDICTION BASED ON NETWORK MEMORY

Temporal networks are networks whose topology changes over time. Two nodes in a temporal network are connected at a discrete time step only if they have a contact/interaction at that time. The classic temporal network prediction problem aims to predict the temporal network one time step ahead based on the network observed in the past of a given duration. This problem has been addressed mostly via machine learning algorithms, at the expense of high computational costs and limited interpretation of the underlying mechanisms that form the networks. Hence, we propose to predict the connection of each node pair one step ahead based on the connections of this node pair itself and of node pairs that share a common node with this target node pair in the past. The concrete design of our two prediction models is based on the analysis of the memory property of real-world physical networks, i.e., to what extent two snapshots of a network at different times are similar in topology (or overlap). State-of-the-art prediction methods that allow interpretation are considered as baseline models. In seven real-world physical contact networks, our methods are shown to outperform the baselines in both prediction accuracy and computational complexity. They perform better in networks with stronger memory. Importantly, our models reveal how the connections of different types of node pairs in the past contribute to the connection estimation of a target node pair. Predicting temporal networks like physical contact networks in the long-term future beyond short-term i.e., one-step ahead is crucial to forecast and mitigate the spread of epidemics and misinformation on the network. This long-term prediction problem has been seldom explored. Therefore, we propose basic methods that adapt each aforementioned prediction model to address classic short-term network prediction problem for long-term network prediction task. The prediction quality of all adapted models is evaluated via the accuracy in predicting each network snapshot and in reproducing key network properties. The prediction based on one of our models tends to have the highest accuracy and lowest computational complexity.

3.1. 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]–[3] could represent these systems more realistically with time-varying network topology.

The classic temporal network prediction problem aims to predict the interactions (or equivalently the network) one step ahead based on the network observed in the previous L steps. This problem is also equivalent to problems in recommender systems, e.g., predicting which user will purchase which product, which individuals will become acquaintances at the next time step [4], [5]. The temporal network prediction problem is more challenging than the static network prediction problem, which aims to predict the missing links or future links based on the links observed [6]–[10]. Recently, machine learning algorithms have been developed to predict temporal networks. Embedding algorithms embed each node in a low-dimensional space based on the network observed. If the learned representations of two nodes are closer in the vector space, it is more likely to have a contact between this node pair one-time step ahead [4], [11]–[15]. Restricted

Boltzmann machine (RBM) based methods [16] and Graph neural networks [17]–[19] have also been developed for this prediction task and they can achieve high prediction accuracy. These methods, however, are at the expense of high computational costs and are limited in providing insights regarding which mechanisms enable the prediction and thus could possibly form temporal networks.

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 specific 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 [20]–[22]. Network-based methods tell directly which mechanisms or properties are used for the prediction and tend to have a low computational complexity. Recently, network properties of a node pair have been combined with learning algorithms to address the classic temporal network prediction problem [23].

In this work, we aim to design network-based methods to solve the classic temporal network prediction problem and to unravel which mechanisms and network properties enable the prediction.

A temporal network measured at discrete times can be represented as a sequence of network snapshots $G = \{G_1, G_2, \dots, G_T\}$, where T is the 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 , $(j, k) \in E_t$. We assume all snapshots share the same set V of nodes. The time aggregated network G^w contains the same set V of nodes and set of links $E = \cup_{t=1}^T E_t$. 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. We give each link in the aggregated network an index i , where $i \in [1, M]$ and $M = |E|$ is the total number of links. 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 link i has a contact at time t and $x_i(t) = 0$ if no contact occurs at t . A temporal network can be thus equivalently represented by its aggregated network, where each link i is further associated with its activity time series \mathbf{x}_i .

Specifically, the prediction task is to predict the activation/connection tendency of each link i at the next time step $t + 1$ based on the network observed in the previous L steps $[t - L + 1, t]$, where $1 \leq t - L + 1 < t \leq T$. The aggregated network G^w is assumed to be known in the prediction problem because it represents social relationships and varies relatively slowly in time compared to contacts. The prediction accuracy is evaluated via the area under the precision-recall curve (AUPR), which compares the predicted activity tendency and the ground-truth connection of each link at the prediction time step.

Firstly, we explore the structural similarity between two network snapshots at any two time steps with a given time lag. We find the similarity or so-called network memory is relatively high when the time lag is small and decays as the time lag increases in the seven real-world physical contact networks considered. Based on this observed time-decaying memory in temporal networks, we design two network-based temporal network prediction models.

The self-driven (SD) model assumes that the activation tendency of a link at a pre-

diction step is solely influenced by its past activity states, with a stronger influence from more recent states. This concept is not new [23], [24]. The SD model is emphasized as one model here because we will explore in depth the choice and interpretation of its parameter and it is the basis to build our self- and cross-driven (SCD) model [25]. In the SCD model, the activity tendency is firstly derived for each link at the prediction time step based on SD model. SCD assumes the connection tendency of a target link at a prediction step depends not only on the SD activity tendency of the link itself at the same prediction step but also of the neighboring links (links share a common end node with the target link) in the aggregated network. State-of-the-art models that allow interpretation are considered as baselines: Common Neighbor, Lasso Regression, Correlated Discrete Auto-regression model, and the Markov model. In several real-world contact networks, we find that SCD outperforms SD model and both SCD and SD models perform better than the baselines. Both SD and SCD perform better in networks with a stronger memory. Additionally, the SCD model allows us to understand how different types of neighboring links (depending on whether they form a triangle with the target link or not) contribute to the prediction of a target link's future activity.

It is essential to predict the contact network in the long-term future, instead of one step ahead, in order to develop strategies to mitigate the epidemic or information spreading on the network. However, this long-term prediction problem on temporal networks remains unexplored. Hence, we further propose basic methods that adapt the aforementioned models for short-term network prediction to solve the long-term network prediction problem. Specifically, the long-term temporal network prediction problem is to predict the network (activities of all links) at each time step within the prediction period $[t + 1, t + L^*]$ based on the network observed within $[t - L + 1, t]$. Moreover, the aggregated network G^w and the total number of contacts at each time step within the prediction period are assumed to be given. The latter assumption aims to simplify the problem, also because the number of contacts can be influenced by factors like weather and policy other than the network observed in the past. The prediction quality is evaluated via whether the predicted network within the prediction period is precise and could reproduce key network properties. We find in general, the adapted SD model performs the best among all models in all data sets. Its prediction accuracy decays as the prediction step is further ahead in time and this decay speed is positively correlated with the decay speed of network memory. Finally, networks predicted by various models respectively within the prediction period have a heterogeneous distribution of inter-event time of contacts along a link, similar to real-world networks.

The rest of the paper is organized as follows. We will introduce real-world temporal networks to be used to design and evaluate temporal network prediction methods in Section 4.3. Key temporal network properties will be analyzed in Section 3.3 to motivate our network-based models (Section 3.4) for the classic short-term network prediction problem. The proposed models will be evaluated and interpreted in Section 3.5. Finally, our network-based models and baseline models will be further developed and evaluated for the long-term prediction problem in Section 3.6 and Section 3.7 respectively.

3.2. EMPIRICAL DATA SETS

To design and evaluate temporal network prediction methods, we consider seven empirical physical contact networks: Hospital [26], Workplace [27], PrimarySchool [28], HighSchool [29], LH10 [27], SFHH [30] and Hypertext2009 [31]. The basic properties of these data sets are given in Table 3.1. The time steps at which there is no contact in the whole network have been deleted.

Table 3.1: 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 sec), 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

3.3. MEMORY IN TEMPORAL NETWORKS

In this section, we explore whether a temporal network has memory, i.e., the network observed at different times share certain similarity. Such memory property may inspire the design of network-based temporal network prediction methods and influence prediction quality.

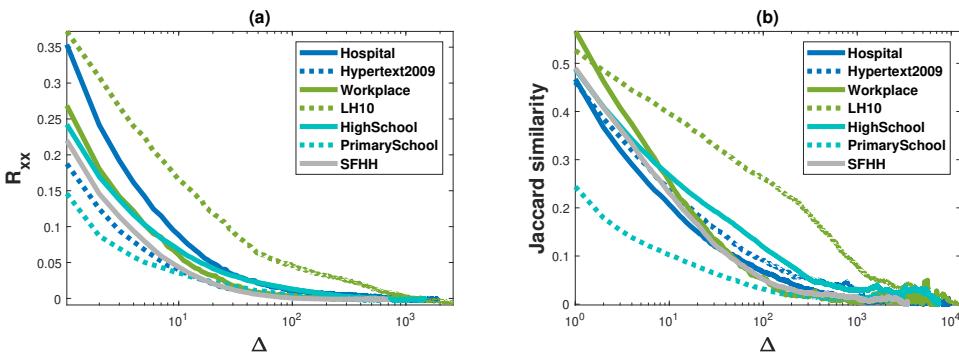


Figure 3.1: (a) The average auto-correlation coefficient R_{xx} over all links as a function of the time lag Δ and (b) the average Jaccard similarity of two snapshots of a temporal network with a given time lag Δ in each of the seven data sets.

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 $\{x_i(t)\}_{t=1,2,\dots,T-\Delta}$ and $\{x_i(t)\}_{t=\Delta+1,\Delta+2,\dots,T}$ as its auto-correlation coefficient. Figure 3.1 (a) shows that the average auto-correlation coefficient over all links decays with the time lag Δ in every real-world network. The average auto-correlation decays slower as the time lag increases.

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. Given 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 a large overlap/similarity between the two snapshots of the temporal network. Figure 3.1 (b) 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 similarity between temporal snapshots decays with their time lag in all empirical data sets, manifesting the time-decaying memory of real-world temporal networks.

3

3.4. SHORT-TERM NETWORK PREDICTION METHODS

In this subsection, we will propose two network-based prediction models and four baseline models for the classic short-term temporal prediction problem, that is, predicting the activation tendency of each link in the aggregated network G^w (G^w is given) at the next time step $t + 1$ based on the network observed in the previous L steps within $[t - L + 1, t]$.

3.4.1. OUR NETWORK-BASED MODELS

Inspired by the time-decaying memory of temporal networks, we propose two network-based temporal link prediction models. In our previous work [32] that uses Lasso Regression for short-term prediction explained in section 3.4.2, it has been found that a link's state at the next step is largely determined by the current state of the link itself and of the neighboring links that share a common node with the target link in the aggregated network. Hence, our two network-based models will estimate a link's activity tendency one step ahead based on the past activities of the link itself and of its neighboring links respectively by taking the memory effect into account.

SELF-DRIVEN (SD) MODEL

The self-driven (SD) model predicts the tendency $w_i(t + 1)$ of the link i to be active at the prediction time $t + 1$ as:

$$w_i(t + 1) = \sum_{k=t-L+1}^{k=t} e^{-\tau(t-k)} x_i(k). \quad (3.1)$$

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 to 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

total number of contacts of link i during the past L steps. Such exponential decay has also been considered in [23], [33]. In Section 3.5.1, 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.

SELF- AND CROSS-DRIVEN (SCD) MODEL

Furthermore, we generalize the SD model to a self- and cross-driven (SCD) model. The SCD model assumes that the activity tendency of a target link one step ahead depends on the SD connection tendency defined in Eq. (3.1) of the link itself and also of neighboring links that share an end-node node with the target 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 Figure 3.2. Furthermore, we differentiate three types of links in an ego-network, colored differently in Figure 3.2: the target link itself (in grey in Figure 3.2), links that form a triangle with the target link (in blue), and the remaining links (in green). We assume that the previous states of these three types of links may contribute differently to the connection tendency of the target link. This is motivated by a) our finding that, when Lasso Regression is used to estimate connection tendency [32], the previous activity of the target link itself contributes more than that of the neighboring links, b) the common neighbor similarity method in static network prediction and c) the observation of temporal motifs (e.g., three contacts that happen within a short duration with a specific ordering in time, and form a triangle in topology) in temporal networks [34], [35].

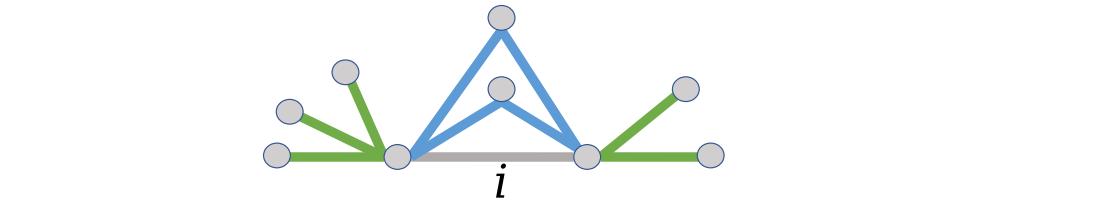


Figure 3.2: An illustrative example of an ego-network centered at a targeted link i . The target link itself, links that form a triangle with the target link, and the other neighboring links, are colored in grey, blue and green respectively.

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

$$h_i(t+1) = \beta_0^* + \beta_1^* w_i(t+1) + \beta_2^* u_i(t+1) + \beta_3^* g_i(t+1). \quad (3.2)$$

of the contributions of the link itself $w_i(t+1)$ as defined in Eq. (3.1), the neighboring links that form a triangle with the target link $u_i(t+1)$ and the other neighboring links $g_i(t+1)$. The latter two factors $u_i(t+1)$ and $g_i(t+1)$ will be defined soon as a function of the SD tendency at $t+1$ of links in the ego-network.

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. This design of $u_i(t+1)$ aims to capture the weighted version of common neighbor similarity. The contribution of the other links $g_i(t+1)$ in the ego-network is defined as the average of their SD activity tendency. For each prediction time step $t+1$, a set of coefficients β_0^* , β_1^* , β_2^* , and β_3^* in Eq.(3.2) will be learned through Lasso Regression from the temporal network observed in the past L steps for all possible target links.

3

3.4.2. BASELINE MODELS

Our goal is to develop network-based models for predicting temporal networks. This is because they usually have low computation complexity and allow us to understand the underlying mechanism that enables the prediction, thus mechanism that potentially forms temporal networks. Hence, as baselines, we introduce four models that are relatively interpretable in their mechanisms of prediction.

COMMON NEIGHBOR

We generalize the common neighbor method from static network prediction [20] to the temporal network prediction problem. The number of common neighbors of a target node pair can be computed for each of the previous L snapshots. The total number of common neighbors (CN) over the past L snapshots is used as the target node pair's tendency of connection at the prediction time step $t+1$.

Scholz et al. [36] and Tsugawa et al. [37] have used the number of common neighbors (CN_{agg}) of a target node pair in the unweighted aggregated network over the past L snapshots to estimate if there will be a new link between this node pair at the prediction time step. Later, we will show that the CN_{agg} method performs overall worse than the CN method in all real-world networks.

LASSO REGRESSION

Lasso Regression [38] 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.,

$$x_i(t+1) = \sum_{j=1}^M x_j(t) \beta_{ij} + c_i. \quad (3.3)$$

The objective is

$$\min_{\beta_i} \left\{ \sum_{t=1}^M (x_i(t+1) - \sum_{j=1}^M x_j(t) \beta_{ij} - c_i)^2 + \alpha \sum_{j=1}^M |\beta_{ij}| \right\}. \quad (3.4)$$

where M is the number of features, as well as the number of links in the aggregated network, c_i is the constant coefficient, and $\beta_i = \{\beta_{i1}, \beta_{i2}, \dots, \beta_{iM}\}$ 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 leads to models with few non-zero coefficients (relevant features). The optimal α that achieves the best prediction is chosen by searching 50 logarithmically spaced points within $[10^{-4}, 10]$.

CDARN MODEL

The correlated Discrete Auto-Regression Network (CDARN) model has been shown to be able to capture the non-Markovian evolution of temporal networks and also the correlation between links in their activities [39]. It assumes that the state of a link at each time step t is either a copy of a previous state of the link itself or another link or is a Bernoulli random variable. The dynamics of each link i is governed by the process:

$$x_i(t) = Q_i(t)x_{C_i(t)}(t - Z_i(t)) + (1 - Q_i(t))Y_i(t). \quad (3.5)$$

where $Q_i(t) \sim \mathcal{B}(q_1)$ is a Bernoulli variable and $Q_i(t) = 1$ (or $Q_i(t) = 0$) with probability q_1 (or $1 - q_1$), the current state $x_i(t)$ of a link i is equal to the state of link $C_i(t)$ at a past time $t - Z_i(t)$ if $Q_i(t) = 1$, and $Y_i(t) \sim \mathcal{B}(q_2)$ is a Bernoulli variable with average q_2 controlling the density of the network.

$Z_i(t)$ is a discrete random variable that is uniformly distributed within $\{1, 2, \dots, P\}$ and it means that states of previous P steps have equal probability to be chosen as $x_i(t)$. This random variable $C_i(t)$ encodes which link's state would be copied by link i and distributed as

$$Pr[C_i(t) = j] = \begin{cases} 1 - c, & \text{if } j = i; \\ c \frac{1}{|\Gamma_i|}, & \text{if } j \in \Gamma_i; \\ 0, & \text{otherwise.} \end{cases} \quad (3.6)$$

Where Γ_i is the set of neighboring links of link i in the aggregated network G^w . Two links in the aggregated network are neighboring links if they share a common end node.

At any time t , $Q_i(t)$ is an independent and identically distributed random variable. The same holds for $C_i(t)$, $Z_i(t)$ and $Y_i(t)$. The parameters q_1 , q_2 , and c can be estimated via Maximum Likelihood Estimation as described in [39] based on the network topology observed in the previous L steps. The same as Lasso Regression models, we confine ourselves to the CDARN model with memory length $P = 1$, where a link's current state is determined probabilistically by the states of the link itself or its neighboring links at the previous time step. This choice is also motivated by the high computational cost of CDARN.

Based on the estimated parameters (q_1 , q_2 and c), the CDARN tendency for each link i to be active at $t + 1$ has been derived in [39] for link prediction task, as

$$S_i(t + 1) = q_1((1 - c)\tilde{D}_i(t + 1) + c\tilde{C}_i(t + 1)) + (1 - q_1)q_2. \quad (3.7)$$

with

$$\tilde{D}_i(t + 1) = \delta(1, x_i(t)), \quad (3.8)$$

$$\tilde{C}_i(t + 1) = \sum_{j \in \Gamma_i} \frac{1}{|\Gamma_i|} \delta(1, x_j(t)) \quad (3.9)$$

where $\delta(a, b)$ is the Kronecker delta, equal to 1 if $a = b$, otherwise 0, and Γ_i is the set of neighboring links of link i in the aggregated network. The term $\tilde{C}_i(t+1)$ represents the fraction of active links among all the neighboring links of link i at t . The term $(1 - c)\tilde{D}_i(t+1) + c\tilde{C}_i(t+1)$ interprets the probability that the state of link i at $t+1$ is active given it is a copy of a previous state of the link itself or its neighboring links.

MARKOV MODEL

3

Markov model [40], [41] assumes that the activity or time series of a link in a temporal network is independent of that of other links and a link's activity at the current time step depends only on its state at the previous time step. For each link, we can obtain a 2×2 transition matrix, where each element represents the transition probability from each possible state (either 0 or 1) at any time step to each possible state at the next consecutive time step, based on the states of the link observed in the last L steps. The Markov tendency for each link being active at $t+1$ is the transition probability from its state at t to an active state.

3.5. PERFORMANCE ANALYSIS IN SHORT-TERM PREDICTION

In this section, we will evaluate and interpret the performance of these short-term network prediction models in the aforementioned set of real-world physical contact networks.

3.5.1. MODEL EVALUATION

We first introduce the method to evaluate the prediction accuracy of a model. Secondly, we explore how to choose the decay factor in the SD model. Thirdly, we compare the performance of all the models.

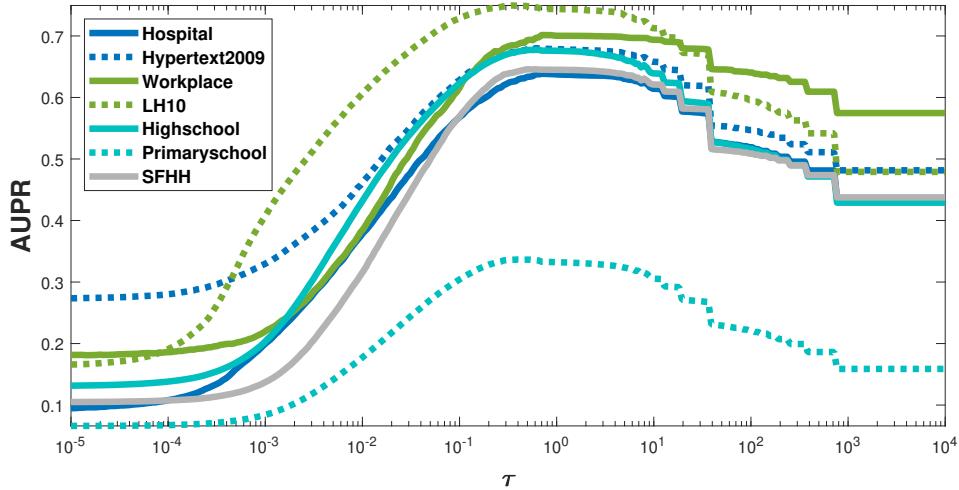
TEMPORAL NETWORK PREDICTION ACCURACY OF SHORT-TERM PREDICTION

Each model predicts the activation tendency of each link at time step $t+1$ based on the temporal network observed in the past L steps. 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 accuracy via the area under the precision-recall curve (AUPR) [42]. An AUPR can be derived for the prediction of each network snapshot, using the connection tendency of each link derived by a given model and the actual network snapshot. AUPR provides an aggregated accuracy across all possible classification thresholds. The average AUPR of a model over the 1000 prediction snapshots quantifies the prediction accuracy of the model. A high AUPR means high prediction accuracy.

CHOICE OF DECAY FACTOR

How to choose 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 time step $t+1$. Given a τ , the tendency $w_i(t+1)$ ($i \in [1, 2, \dots, M]$) can be obtained at each prediction step based on



3

Figure 3.3: Link prediction accuracy AUPR of the SD model as a function of the decay factor τ in seven data sets.

Eq.(3.1). Figure 3.3 shows that the decay factor τ indeed affects the prediction accuracy 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 chosen 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 steps. The prediction accuracy achieved by 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 accuracy than learning τ dynamically over time. We consider $\tau = 0.5$ to derive the SD tendency and SCD tendency in the rest analysis.

COMPARISON OF MODELS

We further compare the prediction accuracy of all models. As shown in Figure 3.4, 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 Figure 3.1.

Previous studies have shown that the number of common neighbors (CN_{agg}) in the

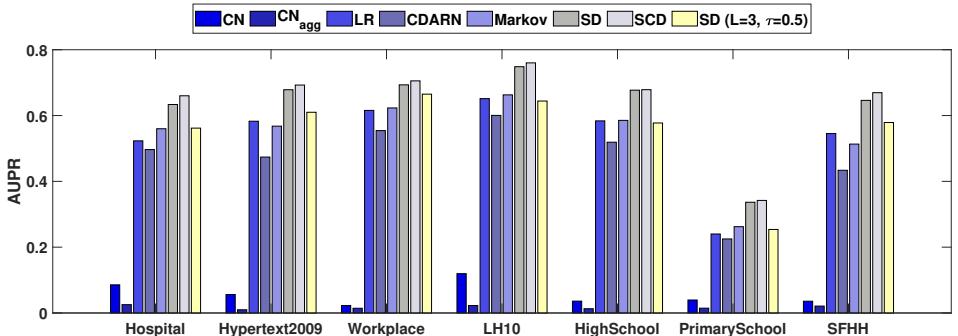


Figure 3.4: Temporal network prediction accuracy AUPR of Common Neighbor model (CN and CN_{agg}), Lasso Regression (LR), CDARN model, Markov model, SD model, and SCD model. All methods consider $L = T/2$ and $\tau = 0.5$ except for SD ($\tau = 0.5$, $L = 3$), which is needed only for Section 3.5.2.

unweighted aggregated network over the past observation period could relatively accurately predict new links to appear in the aggregated network [36], [37]. However, our CN method, though performs better than the CN_{agg} , performs poorly in the short-term network prediction problem. This is likely because when the neighboring links that form a triangle with the target have contacts is crucial for the short-term network prediction problem, but largely ignored by CN and CN_{agg} methods. The SCD model, in contrast, weighs events that happen earlier in time less and estimates implicitly the chance those two neighboring links have contacts at the same time, giving rise to its superior performance. CN method uses the sum of the number of common neighbors over the past L snapshots to estimate a target node pair's tendency of connection at the prediction time step. In this case, every two contacts of a node with the target node pair respectively at the same time contribute to the connection tendency. The CN method taking into account more time information of contacts than CN_{agg} , performs better than CN_{agg} .

3.5.2. MODEL INTERPRETATION

In this subsection, we interpret firstly the SCD model, to understand how the past states of different types of links in the ego-network (neighborhood) of a target link contribute to the activation tendency of the target link. Afterwards, we interpret the decay factor τ and the duration L of past observation to understand how past contacts over time contribute to the prediction.

INTERPRETATION OF SCD MODEL

As defined in Eq. (3.2), 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 of the links that share a common node with the link. The contributions of these three types of links are reflected in the learned coefficients in Eq. (3.2). The average of each coefficient over all prediction steps is given in Table 3.2. In all networks except for Primary School, $|\beta_1^*| > |\beta_2^*| > |\beta_3^*| \approx 0$. This means that the activity of a target link in the future is mainly

influenced by the past activity of the link itself, slightly influenced by the activity of the neighboring links that form a triangle with the target link, and seldom affected by the activity of the other neighboring links. The predictive power of neighboring links that form a triangle with the target link may come from the nature of physical contact networks: contacts are often determined by physical proximity; 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 $\beta_2^* > \beta_1^*$. Table 3.2 shows the aggregated network of PrimarySchool has the largest clustering coefficient¹ in the aggregated network as shown in Table 3.2. In general, we find the contribution β_2^* of links that form a triangle with the target link tends to be more significant in temporal networks with a larger clustering coefficient cc.

Table 3.2: The learned coefficient β_1^* , β_2^* , and β_3^* in SCD model averaged over 1000 prediction steps and the clustering coefficient (cc) of the aggregated network in each empirical network.

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

DURATION L OF PAST OBSERVATION

According to the definition of SD tendency of connection in Eq.(3.1), only the coefficients/contributions $e^{-\tau(t-k)}$ 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 a few previous steps instead of $L = T/2$ steps would be sufficient for a good prediction. As shown in Figure 3.4, the prediction accuracy of SD model when $L = 3$ and $\tau = 0.5$ is worse than that when $L = T/2$ and $\tau = 0.5$. This suggests that although the contribution of each early state of a target link is small, the accumulated contribution of many early states improves the prediction accuracy. The prediction accuracy of the SD model when $L = 3$ and $\tau = 0.5$, whose computational complexity is extremely low, is still better or similar to that of Lasso Regression, reflecting the prediction power of recent states of a link.

The choice of L may influence the prediction accuracy of all models. Hence, we compare further the prediction accuracy of all models when $L = T/4$ in Figure 3.5. We find the same conclusion holds as that when $L = T/2$: both SD and SCD models perform better than the baselines, and SCD performs better than SD model. Additionally, we observe a significant decrease in prediction accuracy when L decreases from $L = T/2$ to $L = T/4$ for both Lasso regression and CDARN models, likely because these learning models need a sufficiently long period of observation for training. In contrast, the prediction accuracy

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

remains relatively stable for the SD, SCD, Makrov, and CN models, despite the change in L . This suggests that the SD, SCD, Markov, and CN models are more resilient to variations in L compared to Lasso regression and CDARN models.

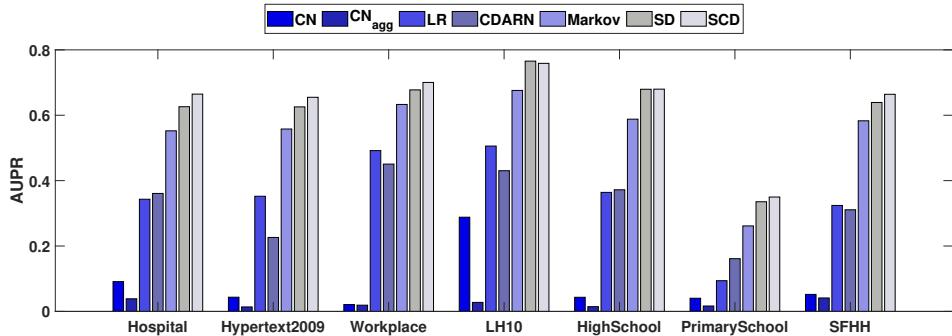


Figure 3.5: Link prediction accuracy AUPR of all models when $L = T/4$, and $\tau = 0.5$ in seven data sets. The prediction accuracy is averaged over 1000 prediction snapshots for all models except that the accuracy of the CDARN model is averaged over 100 prediction snapshots due to its computational complexity.

3.6. LONG-TERM PREDICTION METHODS

Strategies to mitigate epidemics or information spreading are supposed to be carried out for a relatively long period instead of only one time step. Hence, predicting the temporal network in the long-term future is essential for the development of mitigating strategies.

The long-term prediction problem is to predict the temporal network in the long-term future within $[t+1, t+L^*]$ based on the network topology observed in the past L time steps within $[t-L+1, t]$. The number of contacts $m(t+\Delta t)$ at each prediction step and the aggregated network over the whole time window $[1, T]$ of each data set are known. We introduce two basic methods that adapt each short-term network prediction model for the long-term prediction task: recursive long-term prediction and repeated long-term prediction. The common neighbor model is not considered in view of its low performance in short-term prediction.

3.6.1. RECURSIVE LONG-TERM PREDICTION

In short-term prediction, the SD model differs from all the other models in the sense that SD model has only one parameter τ , which can be chosen arbitrarily from $[0.5, 5]$ to achieve approximately the optimal performance, whereas parameters for the other models (Eq. 3.2 for SCD model, Eq. 3.3 for Lasso regression, Eq. 3.7 for CDARN model, and transition matrix for Markov model) need to be trained from the network observed in the past. Hence, we will explain how to make recursive long-term predictions using these two kinds of models respectively.

For SD model, the SD connection tendency for each link at $t+1$ can be obtained according to Eq. 3.1 based on the network observed in the past $[t-L+1, t]$ and $\tau = 0.5$.

Since the number of contacts $m(t+1)$ is known, we predict the temporal network by considering the $m(t+1)$ links with the highest connection tendency as contacts. The predicted network G'_{t+1} at $t+1$ could be represented by the predicted state of each link $\{x'_1(t+1), x'_2(t+1), \dots, x'_M(t+1)\}$ at $t+1$. The predicted network G'_{t+1} at $t+1$ and the network observed within $[t-L+2, t]$ will be used to compute the SD connection tendency of each link at $t+2$ and to derive further the predicted network G'_{t+2} at $t+2$, equivalently the $m(t+2)$ contacts. The connection tendency of each link at each future step $t+\Delta t$, where $2 \leq \Delta t \leq L^*$ is derived recursively using the network observed in $[t+\Delta t-1-L, t]$ and network predicted in $[t+1, t+\Delta t-1]$ according to

$$w_i(t+\Delta t) = \sum_{k=t+\Delta t-1-L}^{k=t} e^{-\tau(t-k)} x_i(k) + \sum_{k=t+1}^{k=t+\Delta t-1} e^{-\tau(t-k)} x'_i(k). \quad (3.10)$$

The SD connection tendency of each link and the given number of contacts at each future step $t+\Delta t$ are used to predict the temporal network at that time step.

For SCD model, Lasso regression, CDARN model, and Markov model, we train each model only once based on the network observed in the past L steps within $[t-L+1, t]$ to obtain its parameters. Each trained model will be used to derive the activation tendency of each target link at $t+1$ using the network observed at t . Then we predict the temporal network at $t+1$ by considering the $m(t+1)$ links with the highest connection tendency to be active. The same trained model will be applied recursively to derive the activation tendency at $t+\Delta t$ using the network $G'_{t+\Delta t-1}$ predicted at $t+\Delta t-1$ and predict the network $G'_{t+\Delta t}$ as the set of contacts along the $m(t+\Delta t)$ links with the highest connection tendency.

3.6.2. REPEATED LONG-TERM PREDICTION

The repeated long-term prediction based on each of the aforementioned models is defined as follows. We firstly derive the connection tendency of each link at $t+1$ in the same way as in short-term prediction based on the network observed within $[t-L, t]$ using a given model. Then the connection tendency of each link at any prediction step $t+\Delta t$ where $\Delta t \in [2, L^*]$ is assumed to be the same as the connection tendency of that link at $t+1$. Given the connection tendency of each link at any prediction step $t+\Delta t$ where $\Delta t \in [1, L^*]$, we predict the network $G'_{t+\Delta t}$ by considering the $m(t+\Delta t)$ links with the highest connection tendency to be active.

Our short-term prediction models can be applied thus either recursively or repeatedly to predict the network in the long-term future. The repeated long-term prediction assumes that the connection tendency of each link remains the same over the long-term prediction period. In contrast, the recursive long-term prediction uses both the observed network and predicted network to predict the network further in time. Hence, it possibly captures the evolving nature of the network over time but leads to accumulative prediction error over time.

3.7. PERFORMANCE ANALYSIS IN LONG-TERM PREDICTION

In this section, we explore the performance of different models applied either repeatedly or recursively in long-term prediction and its relation with the memory property of

temporal networks.

The prediction quality of any method is evaluated via the accuracy in 1) predicting the network at each prediction step within the prediction period $[t+1, t+L^*]$, 2) predicting the weighted aggregated network over the prediction period and 3) reproducing the inter-event time distribution of contacts along a link within the prediction period. The accuracy in these three perspectives is, in general, desirable for long-term network prediction since the network per snapshot, the aggregated network, and the distribution of inter-event time of contacts along a link affect evidently spreading processes unfolding on the network [43]–[46].

3

The prediction length L^* is chosen as $10\%T$, and the starting point $t+1$ of each prediction period $[t+1, t+L^*]$ is sampled 1000 times from $[T/2+1, 90\%T]$ with equal space, to illustrate our method.

3.7.1. NETWORK SNAPSHOT

MODEL EVALUATION

Since the number of contacts $m(t + \Delta t)$ in each prediction step $t + \Delta t \in [t + 1, t + L^*]$ is given, the number of contacts in the network predicted $G'_{t+\Delta t}$ at time step $t + \Delta t$ is the same that of the real-world network (ground-truth) $G_{t+\Delta t}$. Hence, we evaluate the accuracy of the network predicted at each time step $t + \Delta t \in [t + 1, t + L^*]$ via recall, the number of contacts that exist both in the predicted network snapshot $G'_{t+\Delta t}$ and the real-world network $G_{t+\Delta t}$ divided by $m(t + \Delta t)$.

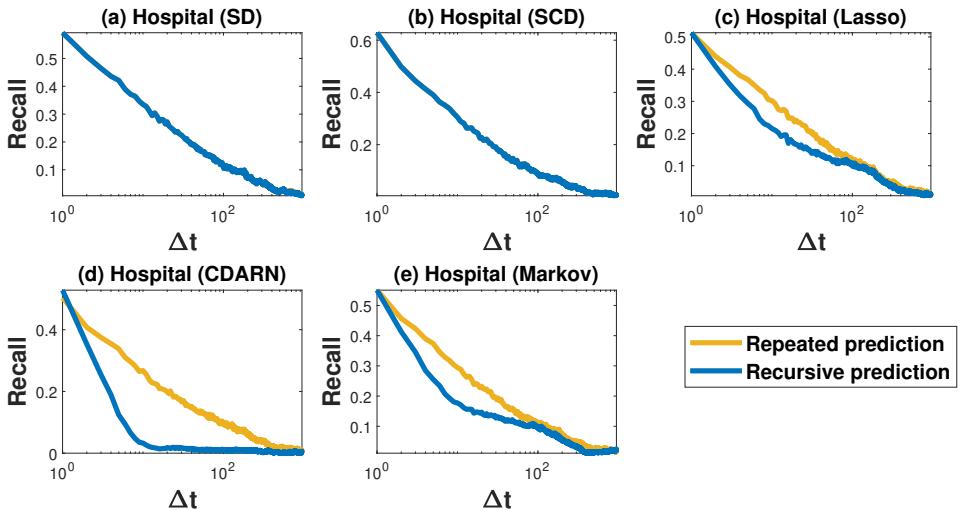


Figure 3.6: The prediction accuracy, Recall, for SD, SCD, Lasso Regression, CDARN, and Markov model, respectively applied recursively or repeatedly, in Hospital at each Δt step ahead.

Firstly, we compare the prediction accuracy recall of each model using the recursive and repeated prediction methods respectively, as a function of prediction time gap Δt , the number of time steps that the prediction step $t + \Delta t$ is ahead of the observation win-

dow $[t - L + 1, t]$. For each $\Delta t \in [1, L^*]$, the prediction accuracy recall is averaged over the 1000 samples of the prediction period. From Figure 3.6 (for network Hospital) and Figure 3.11 (for other networks) in Appendix, we could not recognize any difference between the repeated and recursive methods for SD and SCD models, while Lasso Regression, CDARN, and Markov model tend to perform better using the repeated prediction method.

The similar performance of the SD model when it is applied recursively and repeatedly can be explained by the following. When SD model is applied recursively to predict the network at step $t + 1$, the $m(t+1)$ links with the highest connection tendency are predicted to have contacts, which in return makes their connection tendency at $t + 2$ higher than the other links. In this way, the ranking of links in connection tendency at each prediction time step within $[t + 1, t + L^*]$ remains nearly the same, as in the SD model applied repeatedly. At any prediction step, it is the ranking of links in connection tendency that decides the predicted network.

The prediction accuracy of Lasso Regression, CDARN model, and Markov model is low in short-term prediction. When we use the network predicted by any of these models at $t + \Delta t$ to predict a network at $t + \Delta t + 1$ using the recursive prediction method, the prediction error is accumulated. This is likely why these three models tend to perform better using the repeated prediction method. We consider the repeated prediction method in the rest analysis of this section.

Secondly, we compare the performance of all models using the repeated prediction method in each data set. Figure 3.7 shows the average Recall decreases as the prediction gap Δt increases for all models in all temporal networks. In general, SCD performs the best among all models in all data sets when $\Delta t = 1$, as observed in the short-term prediction in section 3.5.1. When $\Delta t > 2$, SD achieves roughly the best prediction accuracy.

PREDICTION ACCURACY IN RELATION TO NETWORK MEMORY

As SD achieves roughly the best prediction accuracy among all models, we further explore the relation between the prediction accuracy of SD model and the memory property of temporals, aiming to understand in which kind of temporal networks the SD model predicts better.

The prediction accuracy recall of SD model in general decreases as the prediction time gap Δt increases. The decrease is faster when Δt is smaller, as shown in Figure 3.7. We will focus on the prediction gap within $[1, 1\%T]$ since the prediction accuracy is too low when $\Delta t > 1\%T$.

Intuitively, it's probably difficult to predict a temporal network if the network has a weak memory, i.e., the network observed at different times shares low similarity, especially for models like SD and SCD that utilize network memory in network prediction. Hence, we explore the relation between the prediction accuracy of the SD model and the memory property of the temporal network. Figure 3.8 (a) shows the recall of SD model as a function of the normalized prediction time gap $\frac{\Delta t}{1\%T}$. Figure 3.8 (b) illustrates the average Jaccard similarity of two snapshots of a temporal network when their time lag equals $\frac{\Delta t}{1\%T}$. We observe that approximately the prediction accuracy tends to be better in networks with stronger memory (Jaccard similarity). For example, the recall is the largest (smallest) in LH10 (PrimarySchool, Workplace, and Hospital), whose Jaccard similarity

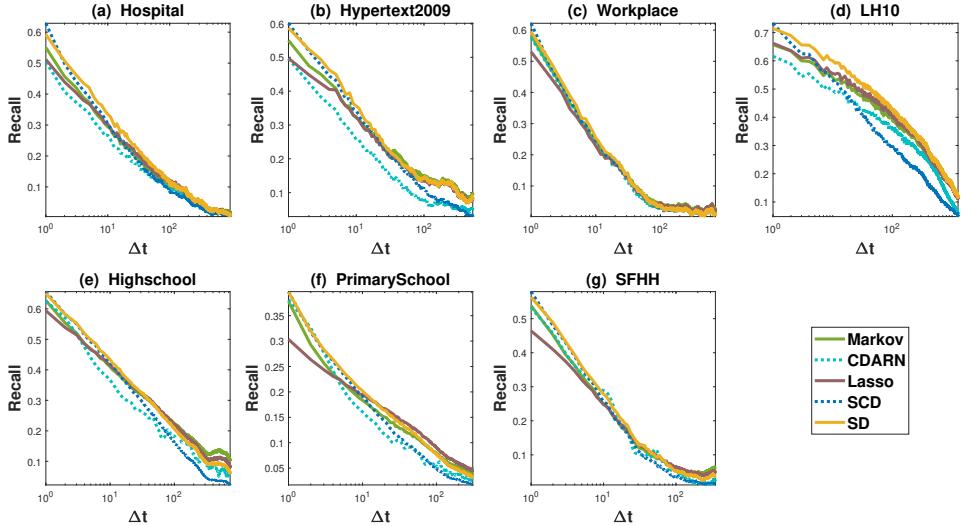


Figure 3.7: The prediction accuracy, Recall, for SD, SCD, Lasso Regression, CDARN, and Markov model applied repeatedly, respectively in seven temporal networks at each prediction step $t + \Delta t$. In the Random model, the $m(t + \Delta t)$ predicted contacts are randomly chosen at the prediction step $t + \Delta t$.

is also the largest (smallest). Furthermore, the decay rate of the prediction accuracy with prediction time gap $\frac{\Delta t}{1\%T}$ in Figure 3.8 seems to be related to the decay rate of Jaccard similarity with the time lag $\frac{\Delta t}{1\%T}$ in Figure 3.8 (b). The decay rate of recall within the interval $[\frac{1}{1\%T}, \frac{\Delta t}{1\%T}]$ is defined as $\frac{\text{Recall}(\frac{\Delta t}{1\%T}) - \text{Recall}(\frac{1}{1\%T})}{\frac{\Delta t}{1\%T} - \frac{1}{1\%T}}$, and the same definition holds for the decay rate of Jaccard similarity. Figure 3.8 (c) and (d) show the decay rate of recall and JS respectively within $[\frac{1}{1\%T}, \frac{\Delta t}{1\%T}]$ as a function of $\frac{\Delta t}{1\%T}$. We find the ranking of the real-world networks in the decay rate of recall approximates that in decay rate of JS at any $\frac{\Delta t}{1\%T}$. This means that prediction accuracy of SD model decays fast in networks with fast decaying memory.

3.7.2. AGGREGATED NETWORK

We evaluate further the precision of the predicted aggregated network $G'_w(t+1, t+L^*)$, which is the network predicted per time step aggregated within the prediction period $[t+1, t+L^*]$. The aggregated network $G_w(t+1, t+L^*)$ of the real-world network is constructed as follows. Two nodes are connected by a link in $G_w(t+1, t+L^*)$, if the two nodes have at least a contact in the real-world network within $[t+1, t+L^*]$. Moreover, the weight of each link is defined as the number of contacts along the link within $[t+1, t+L^*]$. The weighted aggregated network $G_w(t+1, t+L^*)$ could be represented by a weighted adjacency matrix $A_{G_w(t+1, t+L^*)}$ whose element in row i and column j is the number of contacts between node i and node j within $[t+1, t+L^*]$. Similarly, we can construct the predicted aggregated network $G'_w(t+1, t+L^*)$, a weighted network, based on the net-

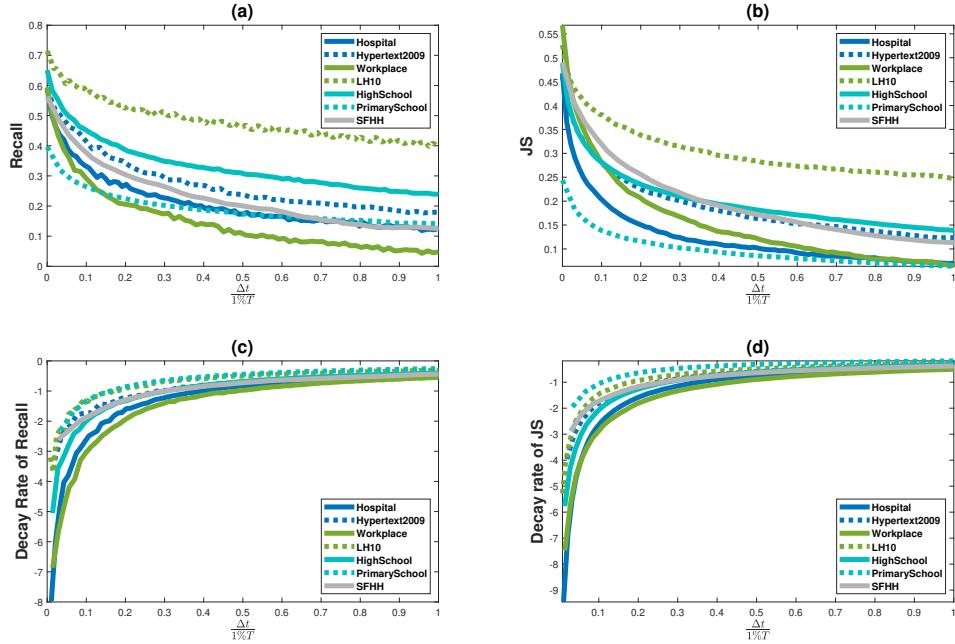


Figure 3.8: (a) The prediction accuracy, recall, of SD model applied repeatedly, (b) the Jaccard similarity (JS), (c) the decay rate of recall and (d) the decay rate of JS as a function of normalized prediction time gap $\frac{\Delta t}{1\%T}$.

work predicted within $[t + 1, t + L^*]$ and represent it by the weighted adjacency matrix $A'_{G_w(t+1,t+L^*)}$. Note that the number of contacts in the predicted network is the same as that in the real-world network at any prediction step.

We evaluate the accuracy in predicting the aggregated network via the generalized recall ($Recall_{wei}$):

$$\frac{\sum_{i \neq j} \min\{A'_{G_w(t+1,t+L^*)}(i, j), A_{G_w(t+1,t+L^*)}(i, j)\}}{\sum_{i \neq j} A_{G_w(t+1,t+L^*)}(i, j)}, \quad (3.11)$$

which measures the extent that the two weighted aggregated networks $G'_w(t + 1, t + L^*)$ and $G_w(t + 1, t + L^*)$ overlap.

We first compare the generalized recall of each model using the recursive and repeated prediction methods respectively, as a function of the prediction period L^* . Instead of considering $L^* = 10\%T$ as in the previous sections, we consider the general scenario where L^* is a variable $L^* \in [1, 10\%T]$. We have observed the same when evaluating the prediction accuracy per snapshot and per aggregated network. We could not recognize any difference in prediction accuracy between the repeated and recursive methods for both SD and SCD models, while Lasso Regression, CDARN, and Markov model tend to perform better using the repeated prediction method (see Figure 3.12 in Appendix).

When all models are applied repeatedly, SD achieves roughly the best prediction accuracy (see Figure 3.13 in Appendix). The SD model tends to predict better in networks

with a strong memory, and the prediction accuracy of SD model decays fast in networks with fast-decaying memory as shown in Figure 3.14 in Appendix.

3.7.3. THE DISTRIBUTION OF INTER-EVENT TIME

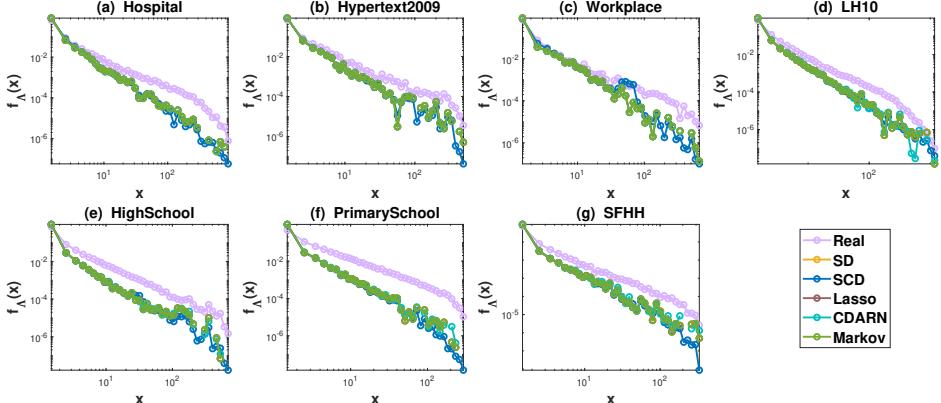


Figure 3.9: The probability density function of inter-event time (Λ) in each real network and network predicted by various models applied repeatedly.

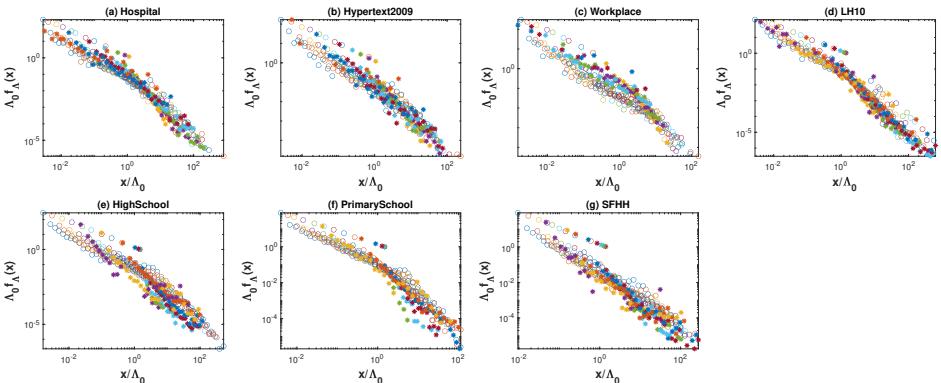


Figure 3.10: The scaled probability density function $\Lambda_0 f_\Lambda(x)$ of the inter-event times derived from each group of links as a function of x/Λ_0 , where Λ_0 is the average inter-event time of the same group, in each real temporal network (circle) and network predicted by SD model (asterisk). Different colors indicate the distribution derived from different groups of links. Links are sorted into 10 groups with a logarithmically increasing width based on their number of contacts.

The inter-event time (Λ) is the time between two consecutive contacts of a link. Firstly, we derive the inter-event distribution of a real-world (predicted) temporal network within $[t+1, t+L^*]$ from the inter-event times collected from all links that have at least two contacts within $[t+1, t+L^*]$. The objective is to explore whether the predicted

network and corresponding real-world network during the prediction period have a similar inter-event distribution. Figure 3.9 (Figure 3.15 in Appendix) shows the inter-event time distribution in each real-world network and the corresponding predicted networks when each model is applied repeatedly (recursively). In each data set, the networks predicted by various models possess almost the same heterogeneous inter-event time distribution, which can be explained as follows.

For repeated prediction, the tendency for each link to be active at each prediction time step $t + \Delta t$ ($\Delta t \in [2, L^*]$) is the same as its activation tendency at $t + 1$, and $m(t + \Delta t)$ links with the highest connection tendency are considered to have contacts at $t + \Delta t$. When no links have the same rank in tendency, the total number of contacts in the network at each time step over time decides the distribution of inter-event time. The link whose connection tendency is the r^{th} largest among all links will be active at a prediction step if the total number of contacts at that prediction step is no less than r . Links with high (low) link tendency are likely to be active (inactive) at each time step, leading to many (few) small (large) inter-event times, which leads to a heterogeneous distribution of inter-event time. The distribution of inter-event time observed in predicted networks approximates roughly the distribution in the corresponding real-world network.

Still, this does not mean that predicted networks have the burstiness of inter-event time as observed in real-world networks: contacts between a pair of nodes usually occur in bursts of many contacts close in time followed by a long period of inactivity. To systematically explore the burstiness property of inter-event time along a link, we group links based on their total number of contacts within $[t + 1, t + L^*]$ as in the method of [47] and derive the probability density function $f_\Lambda(x)$ of inter-event times (Λ) collected from all links in each group. Figure 3.10 shows the scaled probability density function $\Lambda_0 f_\Lambda(x)$ for each group of links as a function of x/Λ_0 , where Λ_0 is the average inter-event time of the same group. As shown in Figure 3.10, the distributions of inter-event time of all groups in both the real-world network and the network predicted by SD model follow a similar heavy-tail distribution. Networks predicted by our SD model reproduce approximately the burstiness of inter-event times as observed in real-world networks.

3.8. CONCLUSION

In this work, we propose two network-based models to solve the short-term temporal network prediction problem. The design of these models is motivated by the time-decaying memory observed 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. Interestingly, we find that SD and SCD models perform better in temporal networks with a stronger memory.

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 network, the contribution of the neighboring links that form a triangle with the target link tends to be significant and possibly dominant.

We further apply these short-term network prediction models either recursively or

repeatedly to make the long-term network prediction., that is the prediction of the temporal network in the long-term future based on the network topology observed in the past and given the number of contacts at each prediction step. The accuracy of long-term prediction accuracy is evaluated from the perspective of the network predicted per snapshot and the predicted aggregated network. The repeated method performs, in general, better for all prediction models. This is likely because the iterative method uses both the observed network and the predicted network which is not precise enough to predict the network further in time. In general, SD model performs the best among all models in all data sets. It predicts better in networks with a stronger memory. The prediction accuracy decays as the prediction step is further ahead in time and this decay speed is positively correlated with the decay speed of network memory. Finally, networks predicted by various models respectively have a heterogeneous distribution of inter-event time similar to real-world networks, and also the burstiness of inter-event times of a link.

Our work is a starting point to explore network-based temporal network prediction methods. Our findings may shed light on the modeling of the formation of temporal networks which is crucial in understanding and controlling the dynamics of and on temporal networks. Our finding that activities of neighboring links that form a triangle with a target link have prediction power on the connection of the target link may suggest that higher-order events [48], [49] like triangles in each network snapshot may contribute to the prediction of (pairwise and higher-order) temporal networks. It is also interesting to evaluate the prediction accuracy of network-based prediction methods in comparison with state-of-the-art machine learning methods that target at high accuracy.

3.9. APPENDIX

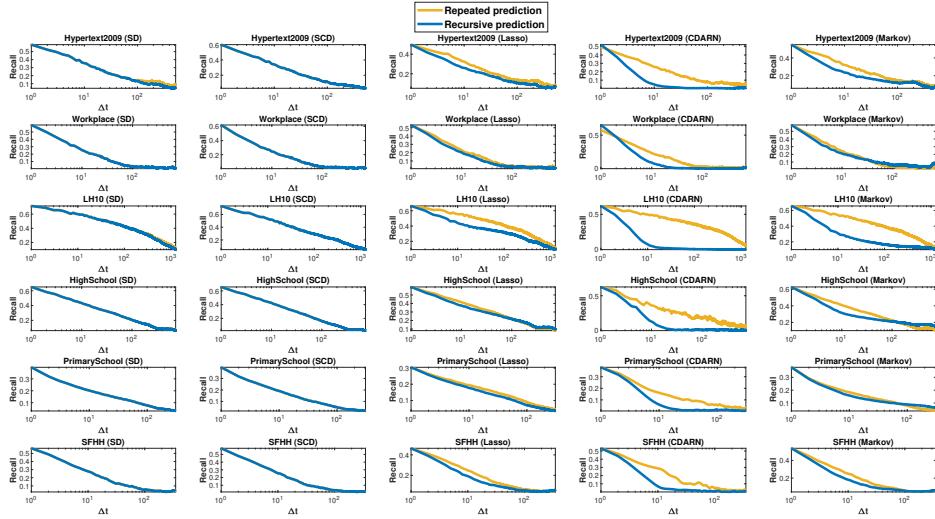


Figure 3.11: The prediction accuracy, Recall, of SD, SCD, Lasso Regression, CDARN, and Markov model, respectively applied recursively (blue curve) or repeatedly (yellow curve) in each real-world network at each prediction step that is Δt step ahead of the training/observed network.

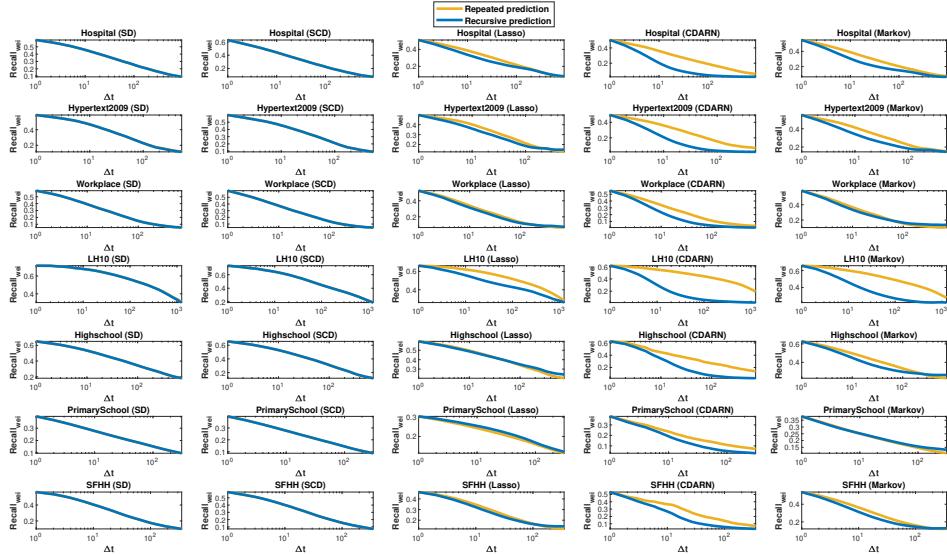


Figure 3.12: The accuracy $Recall_{wei}$ in predicting the aggregated network within $[t+1, t+\Delta t]$, for SD, SCD, Lasso Regression, CDARN, and Markov, respectively applied recursively or repeatedly, as a function of Δt .

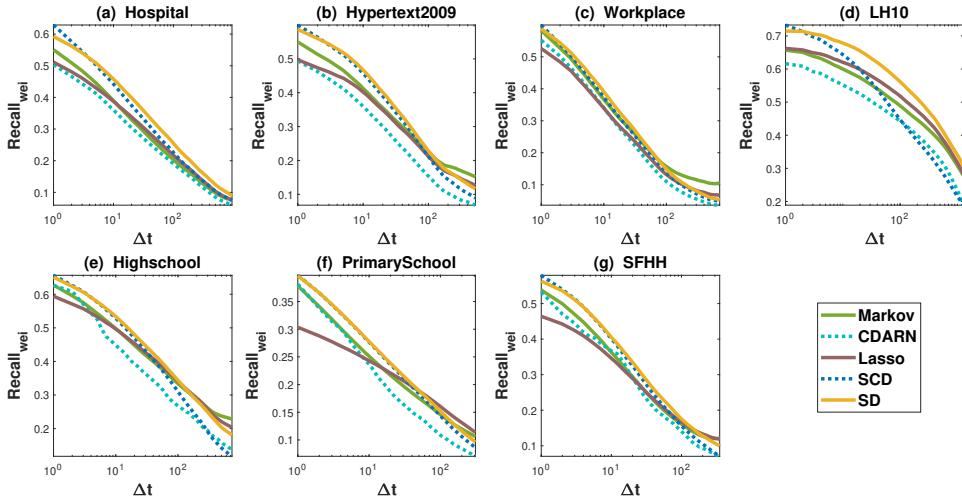


Figure 3.13: The accuracy $Recall_{wei}$ in predicting the aggregated network within $[t+1, t+\Delta t]$, of SD, SCD, Lasso Regression, CDARN, and Markov model applied repeatedly, respectively in seven temporal networks.

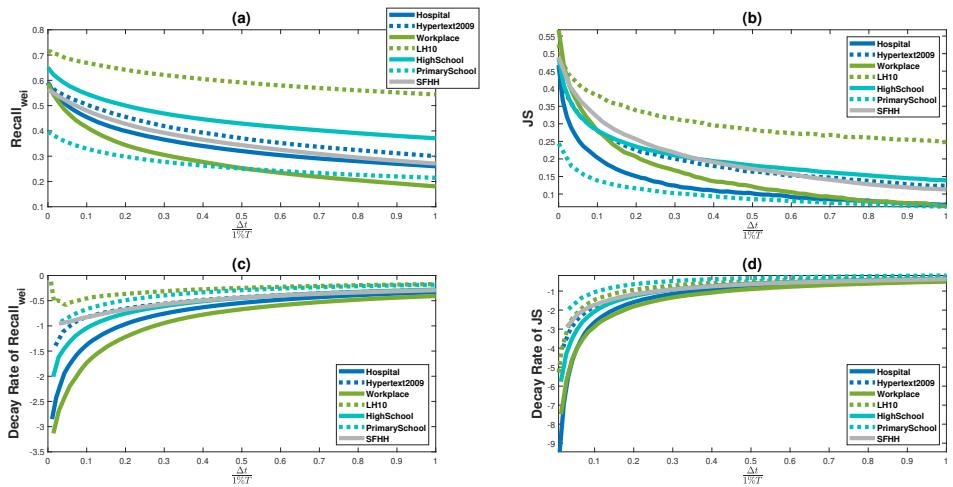


Figure 3.14: (a) The accuracy, $Recall_{wei}$, of SD model applied repeatedly in predicting the aggregated network over $[t+1, t+\Delta t]$, (b) the Jaccard similarity (JS), (c) the decay rate of $Recall_{wei}$ and (d) the decay rate of JS as a function of the normalized prediction time gap $\frac{\Delta t}{1\%T}$.

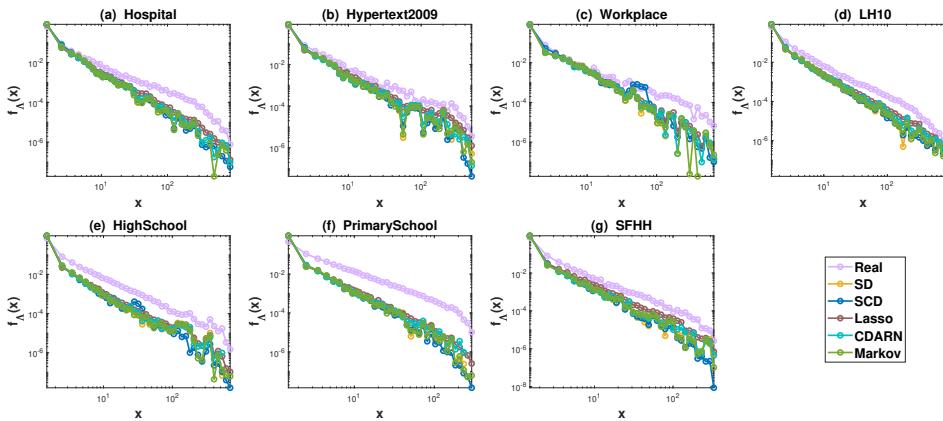


Figure 3.15: The probability density function of inter-event time (Λ) in each real network and network predicted by various models applied recursively.

REFERENCES

- [1] P. Holme and J. Saramäki, “Temporal networks,” *Physics Reports*, vol. 519, no. 3, pp. 97–125, 2012, Temporal Networks, ISSN: 0370-1573.
- [2] N. Masuda and R. Lambiotte, *A Guide to Temporal Networks* (complexity science). World scientific (Europe), 2016, vol. 4.
- [3] P. Holme, “Modern temporal network theory: A colloquium,” *The European Physical Journal B*, vol. 88, no. 9, p. 234, 2015, ISSN: 1434-6036.
- [4] S. Kumar, X. Zhang, and J. Leskovec, “Predicting dynamic embedding trajectory in temporal interaction networks,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD ’19, Anchorage, AK, USA: Association for Computing Machinery, 2019, pp. 1269–1278, ISBN: 9781450362016.
- [5] Y. Dhote, N. Mishra, and S. Sharma, “Survey and analysis of temporal link prediction in online social networks,” in *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2013, pp. 1178–1183. DOI: [10.1109/ICACCI.2013.6637344](https://doi.org/10.1109/ICACCI.2013.6637344).
- [6] L. Lü 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, ISSN: 0378-4371.
- [7] A. Kumar, S. S. Singh, K. Singh, and B. Biswas, “Link prediction techniques, applications, and performance: A survey,” *Physica A: Statistical Mechanics and its Applications*, vol. 553, p. 124 289, 2020, ISSN: 0378-4371.
- [8] P. Cui, X. Wang, J. Pei, and W. Zhu, “A survey on network embedding,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, pp. 833–852, 2017.
- [9] L. Zou, C. Wang, A. Zeng, Y. Fan, and Z. Di, “Link prediction in growing networks with aging,” *Social Networks*, vol. 65, pp. 1–7, 2021, ISSN: 0378-8733.
- [10] X. Zhan, Z. Li, N. Masuda, P. Holme, and H. Wang, “Susceptible-infected-spreading-based network embedding in static and temporal networks,” *EPJ Data Science*, vol. 9, 2020.
- [11] S. M. Kazemi, R. Goel, K. Jain, et al., “Representation learning for dynamic graphs: A survey,” *J. Mach. Learn. Res.*, vol. 21, no. 1, Jan. 2020, ISSN: 1532-4435.
- [12] L. Zhou, Y. Yang, X. Ren, F. Wu, and Y. Zhuang, “Dynamic network embedding by modeling triadic closure process,” in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, ser. AAAI’18/IAAI’18/EAAI’18, New Orleans, Louisiana, USA: AAAI Press, 2018, ISBN: 978-1-57735-800-8.
- [13] Y. Wang, Y.-Y. Chang, Y. Liu, J. Leskovec, and P. Li, “Inductive representation learning in temporal networks via causal anonymous walks,” *ArXiv*, vol. abs/2101.05974, 2021.

- [14] M. Rahman, T. K. Saha, M. A. Hasan, K. S. Xu, and C. K. Reddy, *Dylink2vec: Effective feature representation for link prediction in dynamic networks*, 2018. arXiv: [1804.05755](#).
- [15] da Xu, chuanwei ruan, evren korpeoglu, sushant kumar, and kannan achan, “Inductive representation learning on temporal graphs,” in *International Conference on Learning Representations*, 2020.
- [16] H. Zhang, “A deep learning approach to dynamic interbank network link prediction,” *International Journal of Financial Studies*, vol. 10, no. 3, 2022, ISSN: 2227-7072.
- [17] A. Pareja, G. Domeniconi, J. Chen, *et al.*, “Evolvegcn: Evolving graph convolutional networks for dynamic graphs,” *ArXiv*, vol. abs/1902.10191, 2019.
- [18] L. Wu, P. Cui, J. Pei, L. Zhao, and X. Guo, “Graph neural networks: Foundation, frontiers and applications,” in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, ser. KDD ’23, , Long Beach, CA, USA, Association for Computing Machinery, 2023, pp. 5831–5832, ISBN: 9798400701030.
- [19] Y. Ma and J. Tang, “Graph embedding,” in *Deep Learning on Graphs*. Cambridge University Press, 2021, pp. 75–106.
- [20] 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 Orleans, LA, USA: Association for Computing Machinery, 2003, pp. 556–559, ISBN: 1581137230.
- [21] N. M. Ahmed, L. Chen, Y. Wang, B. Li, Y. Li, and W. Liu, “Sampling-based algorithm for link prediction in temporal networks,” *Information Sciences*, vol. 374, pp. 1–14, 2016, ISSN: 0020-0255.
- [22] H. Xu and L. Zhang, “Application of link prediction in temporal networks,” in *Proceedings of the 2012 2nd International Conference on Computer and Information Application (ICCIA 2012)*, Atlantis Press, 2014, pp. 241–244, ISBN: 978-94-91216-41-1.
- [23] X. Li, W. Liang, X. Zhang, X. Liu, and W. Wu, “A universal method based on structure subgraph feature for link prediction over dynamic networks,” in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, 2019, pp. 1210–1220.
- [24] H. Jo, J. Perotti, K. Kaski, and J. Kertész, “Correlated bursts and the role of memory range,” *Physical Review E*, vol. 92, no. 2, Aug. 2015, Publisher Copyright: © 2015 American Physical Society., ISSN: 2470-0045.
- [25] L. Zou, A. Wang, and H. Wang, “Memory based temporal network prediction,” in *Complex Networks and Their Applications XI*, H. Cherifi, R. N. Mantegna, L. M. Rocha, C. Cherifi, and S. Micciche, Eds., Cham: Springer International Publishing, 2023, pp. 661–673, ISBN: 978-3-031-21131-7.
- [26] P. Vanhemps, A. Barrat, C. Cattuto, *et al.*, “Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors,” *PLoS ONE*, vol. 8, no. 9, e73970, 2013.

- [27] M. G'enois and A. Barrat, "Can co-location be used as a proxy for face-to-face contacts?" *EPJ Data Science*, vol. 7, no. 1, p. 11, May 2018, ISSN: 2193-1127.
- [28] J. Stehlé, N. Voirin, A. Barrat, *et al.*, "High-resolution measurements of face-to-face contact patterns in a primary school," *PLoS ONE*, vol. 6, 2011.
- [29] R. Mastrandrea, J. Fournet, and A. Barrat, "Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys," *PLoS ONE*, vol. 10, no. 9, e0136497, Sep. 2015.
- [30] R. A. Rossi and N. Ahmed, "The network data repository with interactive graph analytics and visualization," in *AAAI Conference on Artificial Intelligence*, 2015. [Online]. Available: <https://api.semanticscholar.org/CorpusID:10141934>.
- [31] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck, "What's in a crowd? analysis of face-to-face behavioral networks," *Journal of Theoretical Biology*, vol. 271, no. 1, pp. 166–180, 2011, ISSN: 0022-5193.
- [32] L. Zou, X. Zhan, J. Sun, A. Hanjalic, and H. Wang, "Temporal network prediction and interpretation," English, *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 3, pp. 1215–1224, 2022, ISSN: 2327-4697.
- [33] W. Yu, W. Cheng, C. C. Aggarwal, H. Chen, and W. Wang, "Link prediction with spatial and temporal consistency in dynamic networks," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, 2017, pp. 3343–3349.
- [34] A. Paranjape, A. R. Benson, and J. Leskovec, "Motifs in temporal networks," in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, ser. WSDM '17, Cambridge, United Kingdom: Association for Computing Machinery, 2017, pp. 601–610, ISBN: 9781450346757.
- [35] J. Saramäki and E. M. Egido, "From seconds to months: An overview of multi-scale dynamics of mobile telephone calls," *The European Physical Journal B*, vol. 88, 2015.
- [36] C. Scholz, M. Atzmueller, A. Barrat, C. Cattuto, and G. Stumme, "New insights and methods for predicting face-to-face contacts," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 7, no. 1, pp. 563–572, Aug. 2021.
- [37] S. Tsugawa and H. Ohsaki, "Effectiveness of link prediction for face-to-face behavioral networks," *PLoS ONE*, vol. 8, 2013.
- [38] L. Zou, X.-X. Zhan, J. Sun, A. Hanjalic, and H. Wang, "Temporal network prediction and interpretation," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 3, pp. 1215–1224, 2022.
- [39] O. E. Williams, P. Mazzarisi, F. Lillo, and V. Latora, "Non-markovian temporal networks with auto- and cross-correlated link dynamics," *Phys. Rev. E*, vol. 105, p. 034301, 3 Mar. 2022.
- [40] A. Frigessi and B. Heidergott, "Markov chains," in *International Encyclopedia of Statistical Science*, M. Lovric, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 772–775, ISBN: 978-3-642-04898-2.

- [41] D. Tang, W. Du, L. Shekhtman, *et al.*, “Predictability of real temporal networks,” *National Science Review*, vol. 7, no. 5, pp. 929–937, Feb. 2020, ISSN: 2095-5138.
- [42] J. Davis and M. Goadrich, “The relationship between precision-recall and roc curves,” in *Proceedings of the 23rd International Conference on Machine Learning*, ser. ICML ’06, Pittsburgh, Pennsylvania, USA: Association for Computing Machinery, 2006, pp. 233–240, ISBN: 1595933832.
- [43] I. Scholtes, N. Wider, R. Pfitzner, A. Garas, C. J. Tessone, and F. Schweitzer, “Causality-driven slow-down and speed-up of diffusion in non-markovian temporal networks,” *Nature Communications*, vol. 5, 2013.
- [44] A. Vazquez, B. Rácz, A. Lukács, and A.-L. Barabási, “Impact of non-poissonian activity patterns on spreading processes,” *Phys. Rev. Lett.*, vol. 98, p. 158 702, 15 Apr. 2007.
- [45] M. E. J. Newman, “The structure and function of complex networks,” *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [46] D. Horváth and J. Kertész, “Spreading dynamics on networks: The role of burstiness, topology and non-stationarity,” English, *New Journal of Physics*, vol. 16, pp. 1–19, 2014, ISSN: 1367-2630.
- [47] K.-I. Goh and A.-L. Barabási, “Burstiness and memory in complex systems,” *Euro-physics Letters*, vol. 81, no. 4, p. 48 002, Jan. 2008.
- [48] A. Ceria and H. Wang, “Temporal-topological properties of higher-order evolving networks,” English, *Scientific Reports*, vol. 13, no. 1, 2023, ISSN: 2045-2322.
- [49] A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, and J. Kleinberg, “Simplicial closure and higher-order link prediction,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 48, E11221–E11230, 2018.

4

INTERVENTION IN INFORMATION TRANSPORT ON TEMPORAL NETWORKS

Communications networks such as vehicle and social contact networks are temporal networks, where autos/individuals are connected only when they are close to each other. These networks facilitate the propagation of information. Traffic between any two nodes demanded at any time is routed along the fastest time-respecting path. In this work, we investigate the intervention in information transport on temporal networks. The objective is to understand the removal of which types of links deteriorate the efficiency/speed of information transport the most. Identifying such critical links will enable better intervention to facilitate/prohibit the spread of (mis)information. To identify critical links, we propose link-removal strategies based on transport efficiency between two end nodes of each link, properties of links in the aggregated network and in routing paths respectively. Each strategy ranks links according to its corresponding property and removes links with the highest measures. Strategies are evaluated via the relative change in transport efficiency after link removal in real-world networks. We find that the path-based strategy performs the best: links appear more often and occur early in the fastest time-respecting paths tend to be critical. Via comprehensive analysis, we explain this strategy's out-performance and its dependency on network properties.

4.1. 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 [1]. The interactions are, in many cases, not continuously active. For example, two autos or individuals in a communications network (such as an opportunistic mobile network, vehicle network, and social contact network) are connected (or have a contact) at a discrete time step if they are close to each other at that time step. Temporal networks [2], [3] could represent these systems more realistically with time-varying network topology. These communications networks facilitate the propagation of information where a piece of information is transmitted from one node (auto or individual) to another node through their contacts [4], [5]. In this work, we are going to explore the vulnerability of information transport on a temporal network to link removal. The objective is to understand the removal of which types of links deteriorate the efficiency of information diffusion the most. Identifying such critical links via their properties e.g., in the network will also allow us to better protect the network. Here, links are node pairs that have at least one contact in a temporal network. When a link is removed, all contacts along the link, i.e., between its two end nodes are removed.

Progress has been made in understanding the vulnerability of a temporal network, i.e., how the efficiency of information transport is influenced when the underlying network is subject to node removal. Scellato et al. [6], [7] and Lu et al. [8] investigated networks with which properties, are more robust against random node removal. The vulnerability of temporal networks subject to intentional attacks has also been investigated [9]–[12]. Intentional attacks remove nodes based on nodal properties like nodal degree. Trajanovski et al. [9] explored the vulnerability of both real-world networks and networks generated by synthetic models, such as Erdős-Rényi and Markov temporal network models. The Erdős-Rényi temporal model, for example, generates a sequence of Erdős-Rényi static random graphs as a temporal network. They found in real-world networks, where nodes are heterogeneous in their properties (like node degree), informa-

tion transport is affected more significantly by intentional attacks than by random node removal. In synthetic networks, where nodes are homogeneous in their network properties, random node removal, and intentional attacks affect similarly the efficiency of information transport on temporal networks. The change of network properties, such as the average degree [10] and the network reachability (the average number of nodes can be reached through time-respecting paths starting from any node at any time) [11], after node removal have also been studied.

Existing works have only considered basic attack methods that remove nodes based on simple nodal centrality metrics. However, how to identify the set of nodes whose removal decreases transport efficiency the most remains an open question. Beyond, the vulnerability of information transport on temporal networks to the removal of links, the fundamental building block of temporal networks, has not been unexplored either. Hence, in this work, we aim to understand the removal of which types of links deteriorate the efficiency of information diffusion the most.

Information transport demand (i, j, t_0) from any source node i to any target node j starting at any time t_0 is assumed to be routed along the fastest time-respecting path $p(i, j, t_0)$ on the underlying temporal network G . Time-respecting paths are sequences of links that can be traversed in a temporal network under the constraint that the next link to be traversed is activated (connected) at some point after the current one [13]. The duration of a time-respecting path is the length of time it takes to traverse that path. The fastest time-respecting path $p(i, j, t_0)$ for a traffic demand (i, j, t_0) is the time-respecting path with the minimum duration. Its duration $\tau(i, j, t_0)$ is called the temporal distance from the source node i to the target node j starting at t_0 . The transport efficiency $\sigma(i, j, t_0)$ of traffic demand (i, j, t_0) is defined as the reciprocal temporal distance, i.e., $\sigma(i, j, t_0) = \frac{1}{\tau(i, j, t_0)}$. It measures how fast information can be transported from source node i to target node j starting at t_0 . The efficiency of information transport on a temporal network G is defined as $\sigma(G) = \frac{1}{N(N-1)T} \sum_{i,j,t_0} \frac{1}{\tau(i, j, t_0)}$, the average transport efficiency over all possible traffic demands, where N is the number of nodes and T is the duration of the observation window $[1, T]$ of the temporal network [6]. Note that if the information cannot reach node j from node i starting at t_0 , $\tau(i, j, t_0)$ is set as infinity. Similarly, the transport efficiency between a node pair is the average efficiency for all traffic demands between the node pair $\sigma(i, j) = \frac{1}{2T} \sum_{t_0} \left(\frac{1}{\tau(i, j, t_0)} + \frac{1}{\tau(j, i, t_0)} \right)$.

In order to identify the links whose removal decreases the transport efficiency of the network the most, we propose link-removal strategies based on transport efficiency between two end nodes of each link, and based on the properties of each link in the time-aggregated network and in the fastest time-respecting paths. Each strategy ranks links according to a property of links and then removes a given number of links with the highest rankings. The performance of each strategy is evaluated via the relative change in transport efficiency of the network after link removal. Our work reveals that our proposed strategies are more effective in identifying critical links for information transport than random link removal in seven physical contact networks and six virtual networks. Moreover, one strategy based on time-respecting paths tends to be the most effective. We observe that removing links, that appear frequently and are active earlier in the fastest time-respecting paths, deteriorates the efficiency of information transport the most.

The rest of the paper is organized as follows. We will introduce the representation

of temporal networks in section 4.2. Real-world temporal networks to be used to evaluate link removal strategies are introduced in section 4.3. Their basic network properties are analyzed, which also inspires the design of link removal methods. Key link removal strategies are proposed in section 4.4 to motivate us to design link-removal strategies. Finally, our proposed link-removal strategies will be evaluated and interpreted in section 4.5.

4.2. TEMPORAL NETWORK

A temporal network measured at discrete times can be represented as a sequence of network snapshots $G = \{G_1, G_2, \dots, G_T\}$, where T is the duration of the observation window $[1, T]$ and $G_t = (V; E_t)$ is the snapshot at time step t with V and E_t being the set of nodes and contacts, respectively. If two nodes, 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 V . The corresponding time aggregated network G^w contains the same set of nodes V and the set of links $E = \cup_{t=1}^T E_t$. 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. The weight of each link in the aggregated network is the total number of contacts occurring along the link within $[1, T]$. The total number of nodes and links are $N = |V|$ and $M = |E|$, respectively. In this paper, links refer to links in the aggregated network.

Table 4.1: The number of nodes ($N = |V|$), the number of links (M) in the giant component of the aggregated network, the length of the observation time window (T), the time resolution (δ sec), 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
Eu1	101	745	10354	1	Virtual	email
Rad	167	3250	57791	1	Virtual	email
DNC	1833	4366	18190	1	Virtual	email
SMS	457	628	21898	1	Virtual	phone
Cal	347	477	2671	1	Virtual	phone
CollegeMsg	1089	5908	22101	1	Virtual	phone

4.3. EMPIRICAL DATA SETS

To evaluate the strategies that identify the critical links for the transport of information on temporal networks, we consider 7 empirical physical contact networks measured at various contexts (Hospital [14], Workplace [15], PrimarySchool [16], HighSchool [17], LH10 [15], SFHH [18] and Hypertext2009 [19]) and 6 virtual contact networks (SMS [20],

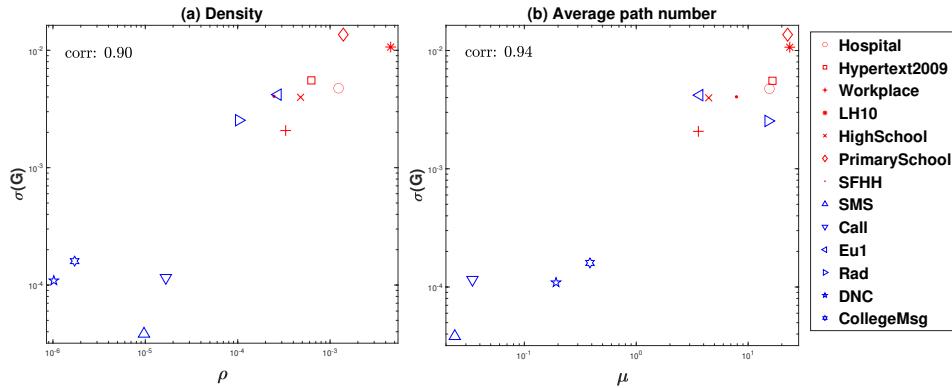


Figure 4.1: The relation between the efficiency $\sigma(G)$ of information transport and (a) the density ρ of the temporal network and (b) the average number of paths μ between a node pair in the aggregated network, respectively. The results of physical and virtual networks are marked in red and blue, respectively. The Spearman correlation coefficient between the efficiency $\sigma(G)$ of information transport and the density ρ of the temporal network (the average number of paths μ between a node pair in the aggregated network) is also shown in the upper-left corner of the left (right) figure.

Call [20], Eu1 [21], Rad [22], DNC [23] and CollegeMsg [24]). All these networks can be collected on the SociaPatterns website¹. To make the simulation more computationally feasible, we have selected a sample of the CollegeMsg network. Specifically, we focus on the contact sequence from the first 30 days of the network, while the original network spans a duration of 193 days. The basic properties of these data sets are given in Table 4.1. The time steps at which there is no contact in the whole network have been deleted in order to consider the steps that are relevant for the transport of information and to avoid the periods that have no contact due to technical errors in measurements. Also, we consider only nodes that belong to the largest connected component in the aggregated network and contacts among these nodes.

Figure 4.1 illustrates the relation between a network property of a temporal network and the efficiency of transport on the temporal network. We consider two properties of a temporal network: density which is defined as the average number of contacts per link per time step and the average number of paths between a node pair in the aggregated network. Figure 4.1 shows that the transport efficiency of a network is strongly and positively correlated with the density of the temporal network and the average number of paths of the network, respectively. This implies that the information tends to be transported faster in temporal networks that have a higher density, as well as in networks that have a larger number of paths between each node pair in their aggregated networks. Hence, the total number of contacts of a link and the number of paths that traverse a link will be used to design link-removal strategies in the following section.

¹ SociaPatterns website: <http://www.sociopatterns.org/datasets/>

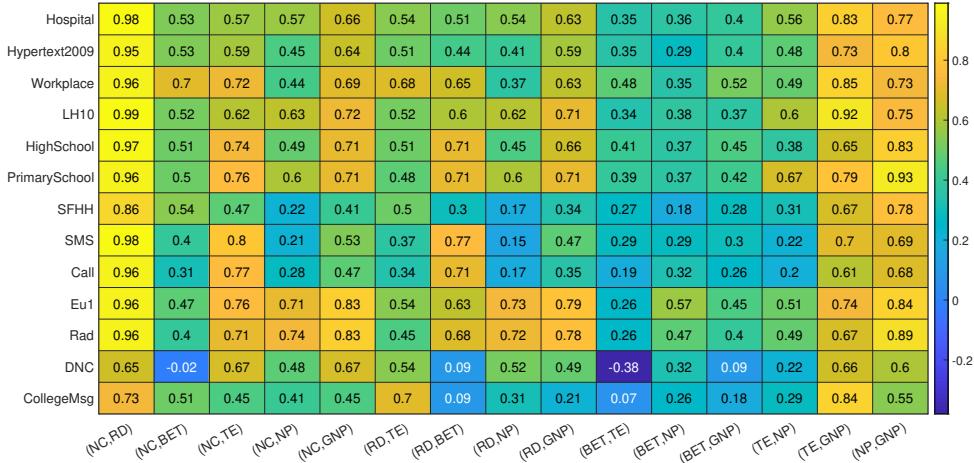


Figure 4.2: Spearman correlation of rankings of links based on two link-removal strategies respectively in 13 temporal networks.

4.4. LINK-REMOVAL STRATEGIES

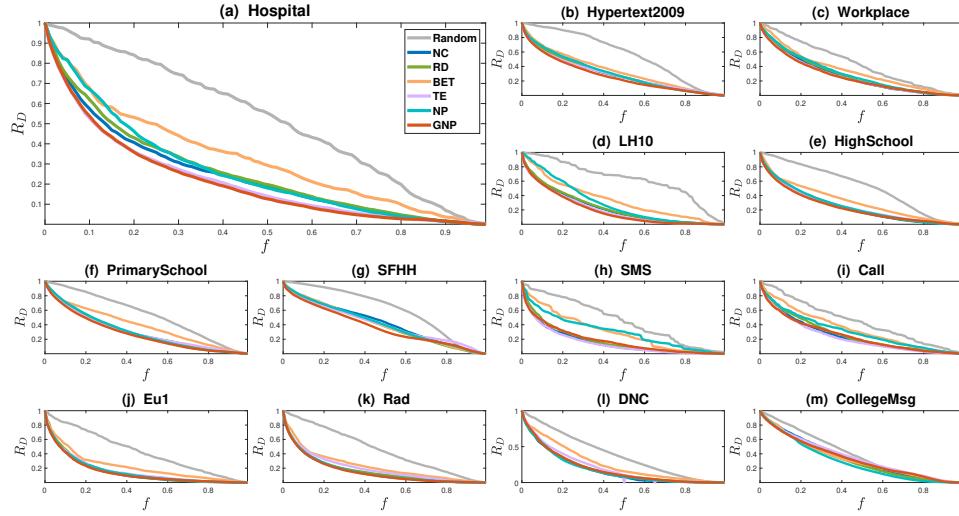
To understand the removal of which kind of links deteriorates the efficiency of information diffusion the most, in this section, we propose systematic link-removal strategies. These strategies are based on the transport efficiency of the two end nodes of each link and the properties of each link in the aggregated network and in the fastest time-respecting paths. Each strategy ranks links according to a property of links and then we remove a given number of links with the highest rankings.

4.4.1. TRANSPORT EFFICIENCY OF THE TWO END NODES OF A LINK

- **The transport efficiency (TE)** of the two end nodes of a link (i, j) has been defined as the average reciprocal temporal distance of traffic demands between node i and node j starting at all possible times, that is $\sigma(i, j) = \frac{1}{2T} \sum_{t_0} (\frac{1}{\tau(i, j, t_0)} + \frac{1}{\tau(j, i, t_0)})$. It measures how fast information can be transported between these two nodes. The transport efficiency of a temporal network is the average transport efficiency over all node pairs. Removing links with a larger contribution to the transport efficiency of a temporal network may reduce effectively the efficiency.

4.4.2. AGGREGATED NETWORK-BASED PROPERTIES

- **The number of contacts (NC)** of a link is the number of contacts occurring along the link within $[1, T]$. It is also the weight of the link in the aggregated network. In Figure 4.1 (a), we observed that temporal networks with more contacts per link per time step are more efficient for information transport. This motivates us to remove links with a large number of contacts to reduce the transport efficiency of a network.
- **Resistance Distance (RD)** of a link (i, j) is defined as the effective resistance $r(i, j)$



4

Figure 4.3: The relative change of transport efficiency $R_D(f)$ as a function of the fraction f of links removed according to random removal and 6 link-removal strategies (NC, RD, BET, TE, NP, and GNP) respectively in each temporal network. The results of random removal are averaged over 1000 simulations for each temporal network.

between node i and node j in the aggregated network G^w times the weight $w(i, j)$ of the link in G^w . The effective resistance $r(i, j)$ in G^w is the effect resistance between i and j in the electrical network that is constructed from G^w by replacing each link with a resistor whose resistance is $\frac{1}{w(i, j)}$ [25], [26]. Mathematically, the effective resistance $r(i, j)$ of link (i, j) can be calculated as:

$$r(i, j) = (\mathbf{e}_i - \mathbf{e}_j)^T L^+ (\mathbf{e}_i - \mathbf{e}_j). \quad (4.1)$$

where \mathbf{e}_i is the column vector with all elements being 0 except that the i -th element is 1 and L^+ is the pseudoinverse of the Laplacian matrix L of the aggregated network G^w . The Laplacian matrix L of G^w is an $N \times N$ matrix defined by

$$L(i, j) = \begin{cases} \sum_{k=1}^N w(i, k), & \text{if } i = j; \\ -w(i, j), & \text{otherwise.} \end{cases} \quad (4.2)$$

The effective resistance $r(i, j)$ between i and j tends to be higher if there are fewer paths, and these paths tend to be longer between node i and node j in the aggregated network G^w . If node i and node j have many contacts and also few alternate paths in the aggregated network, their resistance distance tends to be high. In this case, the direct link between i and j is likely the only path between this node pair in the aggregated network. Removing such links with high resistance distances might significantly slow down the transport of information between these two nodes. Hence, we consider resistance distance as a link-removing strategy.

Table 4.2: The average effectiveness S of each link-removal strategy in each temporal network. For each temporal network, the most effective link-removal strategy is highlighted in blue color.

Data	Random	NC	RD	BET	TE	NP	GNP
Hospital	52.63	24.74	26.05	33.10	21.84	26.58	21.27
Hypertext2009	56.88	29.90	30.29	34.47	29.78	31.06	27.28
Workplace	43.78	27.18	28.11	33.95	26.08	29.42	25.89
LH10	62.75	23.06	23.69	33.99	20.53	29.36	19.76
HighSchool	52.43	24.51	25.13	31.98	25.35	25.76	23.76
PrimarySchool	53.87	30.42	30.15	38.90	30.21	30.54	27.96
SFHII	60.93	42.22	39.99	40.90	41.51	41.45	37.33
SMS	46.17	16.28	18.21	30.26	15.39	30.69	18.58
Call	44.55	24.79	26.98	34.14	23.01	30.88	25.25
Eu1	42.99	14.77	15.68	22.44	16.57	15.90	14.72
Rad	45.72	16.52	17.61	23.12	20.13	16.40	15.84
DNC	38.92	18.44	19.34	26.63	21.13	18.31	19.44
CollegeMsg	38.31	33.71	31.49	32.98	35.43	28.70	33.43

4

- **Betweenness (BET)** of a link counts the number of shortest paths between all node pairs that traverse the link in the aggregated network [27]. To compute the shortest path of each node pair, the distance of each link is defined as $\frac{1}{w(i,j)}$, i.e., inversely proportional to its weight of the link in the aggregated network. The Betweenness strategy has been found effective in prohibiting the spreading of information starting from one seed node to minimize the average prevalence [28]. This motivates us to explore its performance in deteriorating the efficiency of information transport.

4.4.3. PATH-BASED PROPERTIES

We design link-removal strategies based on the properties of links in the fastest time-respecting paths, since links that do not contribute to the propagation of information, thus not appear in any path are likely less relevant to identify the critical link for information transport. Path-based or spreading trajectory-based properties are a special type of centrality metrics and have been seldom explored. Mainly the betweenness of a link in static networks have been studied. Recently, we have shown the effectiveness of using the frequency a link appears in the spreading trajectories/trees of an epidemic starting from each seed node respectively to identify the links to block in order to suppress the spreading [28], [29]. Hence, we design two path-based properties as a start and explore their performance in identify the critical links for information transport.

- **The number of paths (NP)** of a link counts the number of fastest time-respecting paths between all node pairs starting from all possible times that traverse the link. Removing links that participate more in routing paths affects the information transport of more traffic demands and thus may reduce effectively the efficiency of information transport $\sigma(G)$ on the network.
- **Generalized Number of Paths (GNP).** The number of paths only counts how many

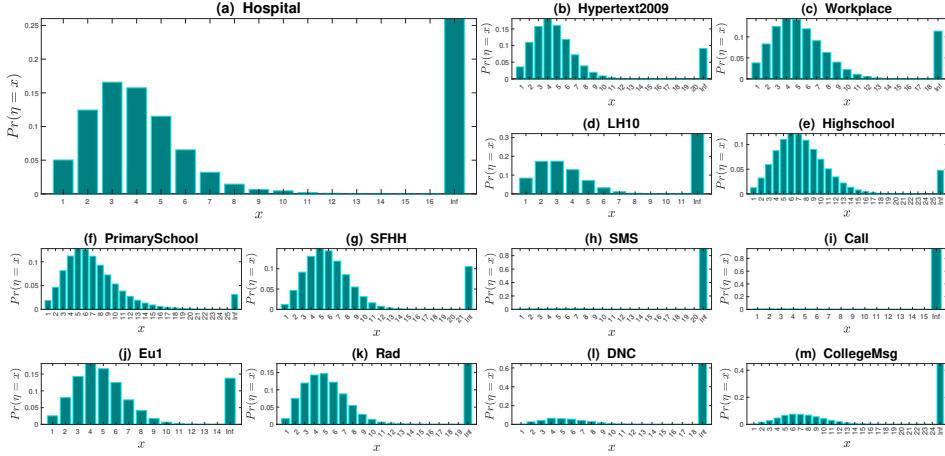


Figure 4.4: The distribution of the hopcount η of the fastest time-respecting path to transport a random traffic demand from a random source node to a random target node starting at a random time on each temporal network.

4

times a link appears in the fastest time-respecting paths, and ignores the temporal information of the link, i.e., when the link is active (appear) in each of the fastest time-respecting paths. Hence, we propose to consider the generalized number of paths of a link. The generalized number of paths for any link l is defined as the sum of the reciprocal relative active time of the link in each fastest time-respecting path. The relative active time $t_l(p(i, j, t_0))$ of the link l in the fastest time-respecting path $p(i, j, t_0)$ is the time when the link is active in the path minus the starting time t_0 of the path (traffic demand). If the path does not traverse link l , $t_l(p(i, j, t_0))$ is infinity. Thus, the generalized number of paths of a link is $\sigma(G) = \sum_{i,j,t_0} \frac{1}{t_l(p(i, j, t_0))}$. Removing links that participate early in many routing paths may increase the temporal distance much for many traffic demands and thus may reduce effectively the efficiency of information transport $\sigma(G)$ on the network.

In order to investigate the relationship between any two link-removal strategies, we analyze the Spearman correlation between their rankings of links. This correlation provides insight into how similar or dissimilar the rankings assigned by different strategies are. In Figure 4.2, we observe a positive correlation between the rankings assigned by most pairs of link-removal strategies. In particular, the link ranking assigned by the NC strategy exhibits a strong correlation with those of other strategies. On the other hand, the link rankings of the BET strategy are less correlated with other strategies. This correlation analysis will be used to explain the different performance of link-removal strategies in deteriorating the efficiency of information transport in the next section.

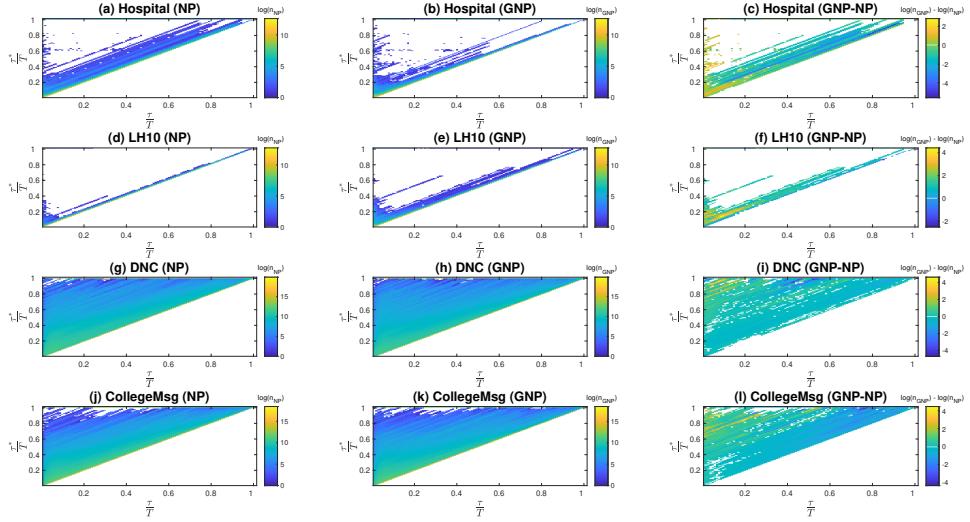


Figure 4.5: Normalized temporal distance of traffic demands before (τ/T) and after (τ^*/T) link removal in Hospital, LH10, DNC, and CollegeMsg. (a) and (b) illustrates the number of traffic demands in logarithm that have a temporal distance τ in the original network Hospital and temporal distance τ^* after link removal according to NP and GNP strategy respectively. (c) shows the difference between (a) and (b). Same for other rows for other networks. A fraction $f = 10\%$ links are removed.

4.5. RESULTS

In this section, we will evaluate and interpret the performance of our proposed link-removal strategies.

4.5.1. EVALUATION

Given a temporal network G , each link-removal strategy ranks the links in the aggregated network G^w according to a property of each link. A fraction f , where $0 \leq f \leq 1$, of links with the highest ranks are selected/removed. After removing all the contacts associated with the selected links, the resultant temporal network is denoted as G_f . The vulnerability of information transport on the temporal network G to this removal of links, or equivalently the effectiveness of the strategy in deteriorating information transport, is evaluated by the relative change of transport efficiency $R_D(f) = \frac{\sigma(G_f)}{\sigma(G)}$ and $0 \leq R_D(f) \leq 1$. A more effective link-removal strategy will lead to a smaller $R_D(f)$.

Figure 4.3 shows the relative change of transport efficiency $R_D(f)$ as a function of the fraction f of removed links in 13 real-world temporal networks. We find all our proposed strategies can reduce transport efficiency more significantly than random link removal in all networks. Moreover, for most link-removal strategies, removing just e.g., 10% of the links can result in a drop of over 50% in transport efficiency. This implies that the removal of a small percentage of links can lead to a large impact on the transport efficiency of

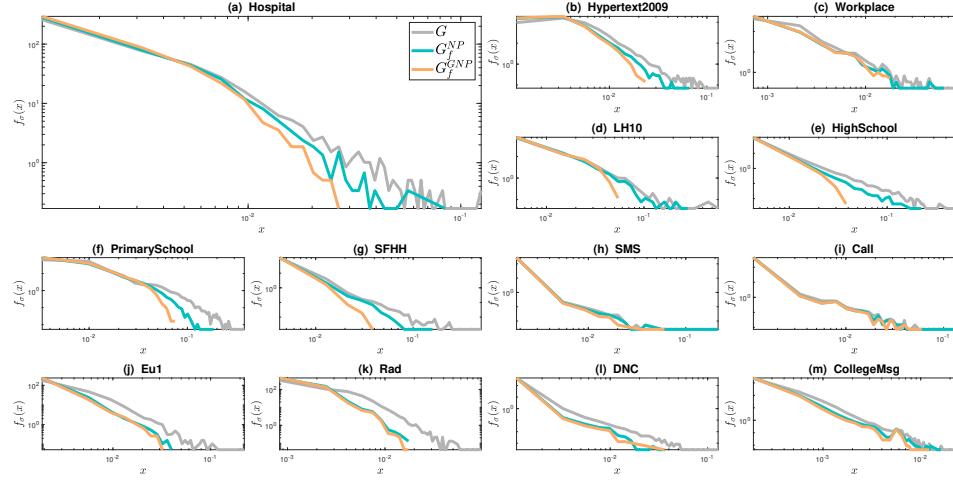


Figure 4.6: The probability density function of transport efficiency σ of traffic demands from any source node to any target node starting at any time. After removing all the contacts associated with the selected $10\%M$ links based on NP (GNP) strategy, the resultant temporal network is denoted as G_f^{NP} (G_f^{GNP}). The probability density function in the networks G , G_f^{NP} , and G_f^{GNP} are denoted by grey, blue, and orange, respectively.

temporal networks. This motivates us to find such critical links.

We also observe that link-removal strategies that exhibit a larger (smaller) discrepancy in their $R_D(f)$ curve in Figure 4.3 also display a lower (larger) degree of correlation in their rankings of links in Figure 4.2. For instance, the $R_D(f)$ curve corresponding to strategy BET is significantly higher than the other strategies at each f , whereas the correlation between BET and any other strategy in ranking links is relatively low.

We further calculate the area under each $R_D(f)$ curve, denoted as S , for each link-removal strategy in each temporal network. This variable implies the average effect of a link removal method over all possible fraction f of links removed. The result is presented in Table 4.2. A smaller S of a link-removal strategy suggests that the strategy better identifies critical links for information transport. GNP strategy performs the best in most networks.

Figure 4.1 shows that our networks are divided into two categories. The first category of networks (virtual network SMS, Call, DNC, and CollegeMsg) have evidently lower transport efficiency, lower network density, and a smaller number of paths between a node pair in aggregated networks than the second category (all the other networks: all physical contact networks and two virtual networks). We find that our GNP strategy works the best in networks of the second category. In the four networks of the first category, the performance of GNP is close to the optimal, whereas TE is the most effective in SMS and Call, and NP performs the best in DNC and CollegeMsg. In practice, usually, a small percentage of links are removed or attacked. The same has been observed when link removal strategies are evaluated by the area under the $R_D(f)$ curve where $f \in [0, 0.5]$,

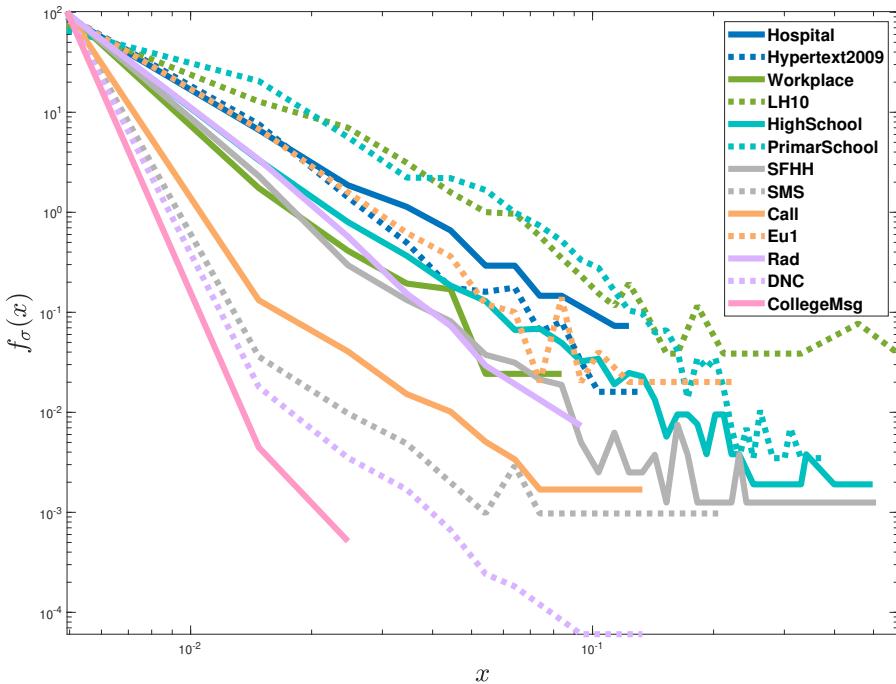


Figure 4.7: The probability density function of transport efficiency σ of traffic demands from any source node to any target node starting at any time before link removal in each temporal network.

i.e., the average effectiveness of a strategy when a percentage of links, between 0 and 50% are removed.

4.5.2. EXPLANATION

Here, we perform further analysis to explain: (1) why the transport efficiency $\sigma(G)$ in networks of the first category is lower than that of the second category of networks; (2) why GNP works better than NP in most networks; and (3) why TE works the best in SMS and Call in identifying critical links for information transport.

TRANSPORT EFFICIENCY IN RELATION TO NETWORK PROPERTIES

As traffic demand from any source node to any target node at any time is routed along the fastest time-respecting path, we will begin with examining the properties of these paths, specifically the distribution of their hopcounts. The hopcount η of a path is the number of links contained in the path.

The distribution of the hopcount η of the fastest time-respecting path to transport a random traffic demand from a random source node to a random target node starting at a random time on each temporal network is given in Figure 4.4. This distribution is derived from the hopcounts of all fastest time-respecting paths to transport all traffic

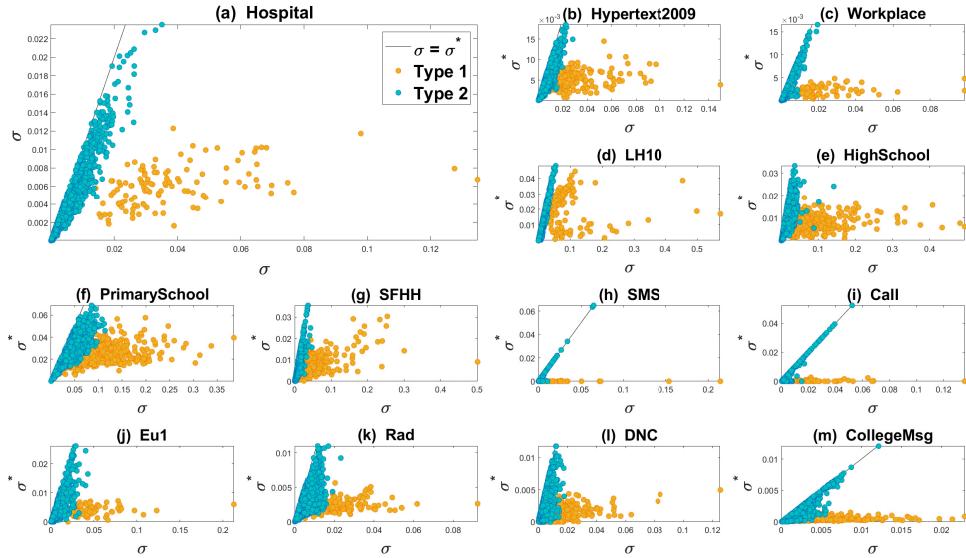


Figure 4.8: The transport efficiency of each traffic demand before (σ) and after (σ^*) 10% M links are removed based on the GPN strategy in 13 temporal networks. Each point in each subgraph corresponds to a traffic demand. Traffic demands between two end nodes of a removed link are marked in orange, otherwise in blue. The black dashed line represents $\sigma = \sigma^*$.

demands, between all possible node pairs starting at each possible time step. If there is not any feasible fastest time-respecting path for a given traffic demand, the corresponding hopcount is defined as infinity. The networks (virtual network SMS, Call, DNC, and CollegeMsg) in the first category have an evidently higher probability for the hopcount to be infinity than those in the second category, in line with their lower network density compared with networks of the second category. In other words, a higher percentage of traffic demands have no time-respecting paths in the first category of networks than that in the second category. This explains why the transport efficiency $\sigma(G)$ in networks of the first category is relatively lower.

WHY GNP WORKS BETTER THAN NP

Our GNP strategy performs better than GN in all networks except DNC and CollegeMsg. Here, we explain why GNP works better than NP.

The removal of any link l in the fastest time-respective path $p(i, j, t_0)$ that routes the traffic demand (i, j, t_0) will lead to a larger temporal distance $\tau(i, j, t_0)$ of the traffic demand. If the relative time $t_l(p(i, j, t_0))$ that link l is active in the fastest time-respective path $p(i, j, t_0)$ is small, the temporal distance $\tau(i, j, t_0)$ tends to small, because $\tau(i, j, t_0) > t_l(p(i, j, t_0))$. The removal of such a link l with a small $t_l(p(i, j, t_0))$, tends to lead to a far smaller transport efficiency $\frac{1}{\tau^*(i, j, t_0)}$ than the efficiency $\frac{1}{\tau(i, j, t_0)}$ before link removal. Hence, removing links that appear in many fastest paths and appear relatively early in each path tends to effectively reduce the transport efficiency of the given net-

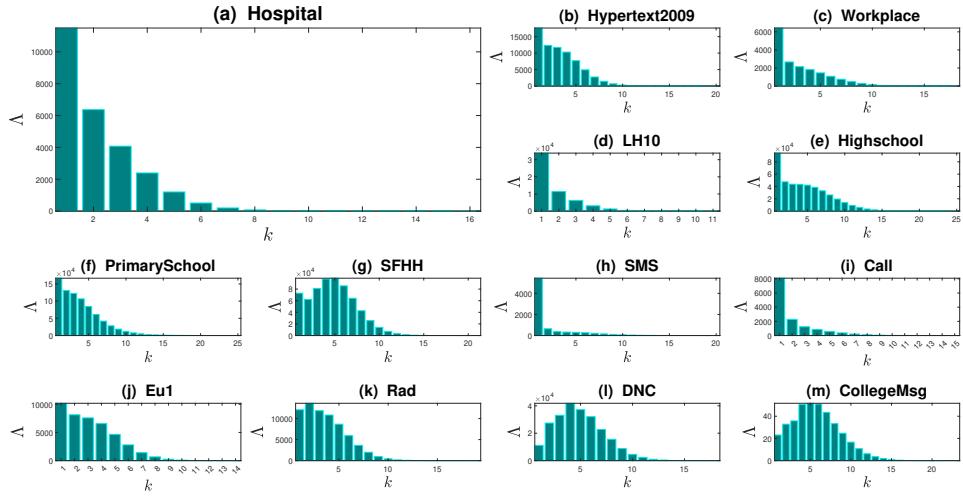


Figure 4.9: The total transport efficiency $\Lambda(k)$ of traffic demands that have hopcount k , as a function of k in each temporal network.

work. This explanation is further supported by the following analysis.

We compare temporal distance of traffic demands before and after link removal, using Hospital, LH10, DNC, and CollegeMsg as examples. The objective is to understand the temporal distance of which traffic demands are more likely to be influenced by link removal. Figure 4.5 (a) and (b) illustrates the number of traffic demands in logarithm that have a temporal distance τ in the original network Hospital and temporal distance τ^* after link removal according to NP and GNP strategy respectively. We find that strategy GNP tends to affect more traffic demands that have a low temporal distance before link removal, than NP. This comparison is clearer in Figure 4.5 (c) which shows the difference between Figure 4.5 (a) and (b). We observe that GNP (NP) affects more traffic demands with a small (medium and large) temporal distance τ in the original network to have a large temporal distance τ^* after link removal. This is reflected in the positive values on the left top corner in Figure 4.5 (c), as well as (f), but less evidently in (i) and (l). As a consequence, GNP effectively reduces the transport efficiency of traffic demands with a large transport efficiency in the original network. This is supported by the probability density function of the transport efficiency σ of traffic demands from any source node to any target node starting at any time in Figure 4.6 before and after link removal (a). After removing links according to GNP(NP), the probability of having a large transport efficiency is evidently (slightly) lower than that in the original network. The same has been observed most networks (not evident in DNC and CollegeMsg).

From Figure 4.7, we can see that the percentage of traffic demands with a large efficiency is the smallest in DNC and CollegeMsg among all networks. In other words, transport efficiency of traffic demands varies less in DNC and CollegeMsg. The trend that, if the relative time $t_l(p(i, j, t_0))$ that link l is active in the fastest time-respective path $p(i, j, t_0)$ is small, the temporal distance $\tau(i, j, t_0)$ tends to small, is thus not evident

in DNC and CollegeMsg. This might limit the effectiveness of GNP. On the other hand, DNC and CollegeMsg have the lowest density among all networks. Removing links that appear in many fastest paths could be possibly the most effective in reducing network efficiency. These two perspectives might explain why NP performs the best, slightly better than GNP in DNC and CollegeMsg.

WHY TE WORKS THE BEST IN SMS AND CALL

The removal of a link could decrease the transport efficiency of traffic demands between its two end nodes and also other node pairs if this link appears in the fastest time-respective paths of these node pairs. SMS and Call differ from all the other networks in the sense that the probability that there is not any feasible time-respective path for a traffic demand is the highest in these two networks, as shown in Figure 4.4. Hence, once a link is removed from SMS or Call, the chance is high that there is no time-respective path for any traffic demand between its end nodes starting at any time. This is supported by Figure 4.8, where we examine the transport efficiency of each node pair before and after removing 10% M links based on GPN strategy. The node pairs are grouped into Type 1 (including the pair of end nodes of each removed link) and Type 2 node pairs (all the other node pairs). In SMS and Call, the efficiency of Type 1 node pairs is almost reduced to 0 after link removal. This confirms that after link removal there is hardly any feasible time-respective path between the end nodes of each removed link. In contrast, the transport efficiency of Type 2 node pairs hardly decreases in SMS and Call. This is because the removed links seldom appear in the fastest time-respective path of a traffic demand between a type 2 node pair, supported by the low probability for the hopcount of a random traffic demand to be larger than one but not infinity (see Figure 4.4). Even if the removed link appears in the path between a type 2 node pair, the influence of the link removal on the efficiency of a type 2 node pair is small. This is because the efficiency of any node pair depends largely on the efficiency of the traffic demands between the node pair that have hopcount 1, as shown in add Figure 4.9 (h) and (i).

In general, removing links in SMS and Call only reduces the efficiency of type 1 node pairs to zero without affecting evidently the efficiency of type 2 node pairs. This has been observed when different percentages of links are removed according to diverse link removal methods. Hence, removing links whose end nodes have the the highest transport efficiency TE is the most effective in reducing the transport efficiency of the network in SMS and Call.

4.6. CONCLUSION

In this work, we investigate the intervention in information transport on temporal networks to link removal to understand the removal of which types of links deteriorate the transport efficiency of information diffusion the most. Link removal strategies have been proposed systematically, based on the transport efficiency of the two end nodes of each link, properties of each link in the aggregated network, and in the fastest time-respecting paths.

We evaluate the effectiveness of these link removal strategies in reducing the efficiency of information transport on a given temporal network. Seven physical contact networks and six virtual networks have been considered. Interestingly, we find that the

strategy based on the property of each link in the fastest time-respecting paths performs the best in most networks. Links that appear in more paths and occur earlier in each path better identify critical links for information transport.

We explain why and in which kind of networks this strategy performs the best via comprehensive analysis, such as the temporal network properties, and the influence of link removal on the transport efficiency of different types of traffic demands and of different kinds of node pairs.

Properties of a link in diffusion trajectories or routing paths are special types of centrality metrics, that have not been explored systematically. Within this class, mainly the betweenness of a link in a static network has been widely studied. In this work, we have designed two properties of a link in the fastest time-respecting paths, by taking into account how frequently a link appears in these routing paths, and when it is activated in these paths or equivalently when it propagates the information. This illustrates the possibilities in designing such centrality metrics. Our findings also reveal the relation between the property of a link in diffusion paths and the effect of moving the link on information transport.

Our work is a starting point to explore the intervention in dynamic processes on temporal networks to link removal. Properties of links in spreading paths can be further designed, e.g., by considering the properties of the paths that a link belongs to. Properties of links in spreading paths usually have a high computational complexity. It is interesting to explore their relation with other network properties of links, especially those with a low computational complexity. This may enable efficient identification of critical links.

REFERENCES

- [1] M. Newman, *Networks: An Introduction*. OUP Oxford, 2010, ISBN: 9780191500701.
- [2] P. Holme and J. Saramäki, “Temporal networks,” *Phys. Rep.*, vol. 519, no. 3, pp. 97–125, 2012, ISSN: 0370-1573.
- [3] N. Masuda and R. Lambiotte, *A Guide to Temporal Networks* (complexity science). World scientific (Europe), 2016, vol. 4.
- [4] M. Karsai, N. Perra, and A. Vespignani, “Time varying networks and the weakness of strong ties,” *Scientific Reports*, vol. 4, p. 4001, 2014.
- [5] J.-P. Eckmann, E. Moses, and D. Sergi, “Entropy of dialogues creates coherent structures in e-mail traffic,” *Proceedings of the National Academy of Sciences*, vol. 101, no. 40, pp. 14 333–14 337, 2004.
- [6] S. Scellato, I. Leontiadis, C. Mascolo, P. Basu, and M. Zafer, “Understanding robustness of mobile networks through temporal network measures,” in *2011 Proceedings IEEE INFOCOM*, 2011, pp. 1–5.
- [7] S. Scellato, I. Leontiadis, C. Mascolo, P. Basu, and M. Zafer, “Evaluating temporal robustness of mobile networks,” *Mobile Computing, IEEE Transactions on*, vol. 12, pp. 105–117, 2013.
- [8] X. Lu, H. Wang, and Y. Deng, “Evaluating the robustness of temporal networks considering spatiality of connections,” *Chaos, Solitons & Fractals*, vol. 78, pp. 176–184, 2015.

- [9] S. Trajanovski, S. Scellato, and I. Leontiadis, “Error and attack vulnerability of temporal networks,” *Phys. Rev. E*, vol. 85, p. 066 105, 6 Jun. 2012.
- [10] N. Wang, N. Wu, L.-I. Dong, H.-k. Yan, and D. Wu, “A study of the temporal robustness of the growing global container-shipping network,” *Scientific Reports*, vol. 6, p. 34 217, 1 2016.
- [11] H. H. Sano and L. Berton, “Topology and robustness analysis of temporal air transport network,” in *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, ser. SAC ’20, Brno, Czech Republic: Association for Computing Machinery, 2020, pp. 1881–1884, ISBN: 9781450368667.
- [12] J. W. Matthew and M. Mirco, “Spatio-temporal networks: Reachability, centrality and robustness,” *R. Soc. open sci.*, vol. 3, p. 160 196, 2016.
- [13] R. K. Pan and J. Saramäki, “Path lengths, correlations, and centrality in temporal networks,” *Phys. Rev. E*, vol. 84, p. 016 105, 1 Jul. 2011.
- [14] P. Vanhemps, A. Barrat, C. Cattuto, *et al.*, “Estimating potential infection transmission routes in hospital wards using wearable proximity sensors,” *PLOS ONE*, vol. 8, no. 9, pp. 1–9, Sep. 2013.
- [15] G. Mathieu and B. Alain, “Can co-location be used as a proxy for face-to-face contacts?” *EPJ Data Science*, vol. 7, p. 11, 2018, ISSN: 2193-1127.
- [16] J. Stehlé, N. Voirin, A. Barrat, *et al.*, “High-resolution measurements of face-to-face contact patterns in a primary school,” *PLOS ONE*, vol. 6, no. 8, e23176, Aug. 2011.
- [17] R. Mastrandrea, J. Fournet, and A. Barrat, “Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys,” *PLOS ONE*, vol. 10, no. 9, pp. 1–26, Sep. 2015.
- [18] R. A. Rossi and N. K. Ahmed, “The network data repository with interactive graph analytics and visualization,” ser. AAAI’15, Austin, Texas: AAAI Press, 2015, pp. 4292–4293, ISBN: 0262511290.
- [19] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. V. den Broeck, “What’s in a crowd? analysis of face-to-face behavioral networks,” *J. Theor. Biol.*, vol. 271, pp. 166–180, 2011, ISSN: 0022-5193.
- [20] P. Sapiezynski, A. Stopczynski, D. D. Lassen, and S. Lehmann, “Interaction data from the copenhagen networks study,” vol. 6, 2019, p. 315.
- [21] A. Paranjape, A. R. Benson, and J. Leskovec, “Motifs in temporal networks,” in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, ser. WSDM ’17, Cambridge, United Kingdom: Association for Computing Machinery, 2017, pp. 601–610, ISBN: 9781450346757.
- [22] R. Michalski, S. Palus, and P. Kazienko, “Matching organizational structure and social network extracted from email communication,” in *Lecture Notes in Business Information Processing*, vol. 87, Springer Berlin Heidelberg, 2011, pp. 197–206.
- [23] J. Kunegis, “Konekt: The koblenz network collection,” in *Proceedings of the 22nd International Conference on World Wide Web*, ser. WWW ’13 Companion, Rio de Janeiro, Brazil: Association for Computing Machinery, 2013, pp. 1343–1350.

- 4
- [24] P. Panzarasa, T. Opsahl, and K. Carley, “Patterns and dynamics of users’ behavior and interaction: Network analysis of an online community,” *JASIST*, vol. 60, pp. 911–932, May 2009.
 - [25] A. Mercier, S. Scarpin, and C. Moore, “Effective resistance against pandemics: Mobility network sparsification for high-fidelity epidemic simulations,” *PLoS Comput. Biol.*, vol. 18, e1010650, 11 2022.
 - [26] D. A. Spielman and N. Srivastava, “Graph sparsification by effective resistances,” *SIAM Journal on Computing*, vol. 40, no. 6, pp. 1913–1926, 2011.
 - [27] H. Wang, J. M. Hernandez, and P. Van Mieghem, “Betweenness centrality in a weighted network,” *Phys. Rev. E*, vol. 77, p. 046105, 4 Apr. 2008.
 - [28] X.-X. Zhan, A. Hanjalic, and H. Wang, “Suppressing information diffusion via link blocking in temporal networks,” in *Complex Networks and Their Applications VIII*, H. Cherifi, S. Gaito, J. F. Mendes, E. Moro, and L. M. Rocha, Eds., Cham: Springer International Publishing, 2020, pp. 448–458, ISBN: 978-3-030-36687-2.
 - [29] X.-X. Zhan, A. Hanjalic, and H. Wang, “Information diffusion backbones in temporal networks,” *Scientific Reports*, vol. 9, p. 6798, 2019.

5

CONCLUSION

In this thesis, we aim to contribute to the development of mitigation strategies as means to intervene in spreading processes on temporal networks. We focus on two aspects. Firstly, we design methods to predict temporal networks in the future and also provide insightful interpretation regarding which inherent properties of temporal networks or mechanism of network formation these prediction methods capture or utilize. The future temporal network predicted and the underlying mechanism of network formation detected may facilitate the development of mitigating strategies. Secondly, assuming the temporal network in the future is known or perfectly predicted, we investigate the design of advanced mitigation strategies that select links to block based not only on classic network properties of links, but also new type of link properties/centralities, such as properties of links in spreading trajectories.

The previous part of the thesis is organized into an introduction chapter and three technical chapters where we describe our proposed methodologies, findings, and empirical results. In this final section, we will provide a concluding summary of our research and reflect on its main contributions. Additionally, we will outline potential avenues for future research given the insights gained and the limitations identified during our study.

5.1. MAIN CONTRIBUTION AND REFLECTIONS

We first address the following temporal network prediction problem: predict the temporal network in the future based on the network observed in the past of a given duration. We develop two types of interpretable network prediction methods, described in chapters 2 and 3.

In Chapter 2, we propose a method that deploys interpretable learning algorithms, Lasso Regression and Random Forest, to predict the activity (connected or not) of each link at the next time step based on the current activities of all links. The coefficients learned from each algorithm are further used to construct the prediction backbone network, presenting the influence or contribution of all links in determining each link's activity. Via exploring the properties of the backbone network and its relation to the activity

time series of links and its relation to the aggregated network, we conclude the following based on six real-world physical and virtual contact networks: (a) A link's next step activity is mainly influenced by the current activity of the link itself and of other links that are better connected with the link; (b) Two links are better connected if they have shorter and/or more shortest paths in the aggregated network; (c) The influence between two links tends to be large if their corresponding activity time series are strongly correlated. The linear regression assumed by Lasso could be one elementary mechanism for modeling temporal networks. The influence patterns that we have discovered can be further used to adapt other dynamic processes to model temporal networks. Our findings of the backbone network and its association with other network properties may also inspire the solution of network classification and optimization problems.

To overcome the problem of a high computational cost of machine learning algorithms, we further design network-based models for temporal network prediction in Chapter 3. The design of these models is motivated by the time-decaying memory observed 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. Interestingly, we find that SD and SCD models perform better in temporal networks with a stronger memory. 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 network, the contribution of the neighboring links that form a triangle with the target link tends to be significant and possibly dominant.

We further apply these short-term network prediction models to generate long-term network predictions, that is, the prediction of the temporal network in the long-term future based on the network topology observed in the past and given the number of contacts at each prediction step. The accuracy of long-term prediction accuracy is evaluated from the perspective of the network predicted per snapshot and the predicted aggregated network. The repeated method performs, in general, better for all prediction models. This is likely because the iterative method uses both the observed network and the predicted network which is not precise enough to predict the network further in time. In general, SD model performs the best among all models in all data sets. It predicts better in networks with a stronger memory. The prediction accuracy decays as the prediction step is further ahead in time and this decay speed is positively correlated with the decay speed of network memory. Finally, networks predicted by various models respectively have a heterogeneous distribution of inter-event time similar to real-world networks, and also the burstiness of inter-event times of a link.

This work is a starting point to explore network-based temporal network prediction methods. Our findings may shed light on the modeling of the formation of temporal networks which is crucial in understanding and controlling the dynamics of and on temporal networks. Our finding that activities of neighboring links that form a triangle with a target link have prediction power on the connection of the target link may suggest that higher-order events like triangles in each network snapshot may contribute to the prediction of (pairwise and higher-order) temporal networks.

In Chapter 4, we assume the future temporal networks are given or perfectly predicted, and develop advanced strategies to mitigate an information diffusion process. Specifically, we investigate the intervention in information transport on temporal networks to link removal to understand the removal of which types of links deteriorate the transport efficiency of information diffusion the most. Link removal strategies have been proposed systematically, based on the transport efficiency of the two end nodes of each link, properties of each link in the aggregated network, and in the fastest time-respecting paths. We evaluate the effectiveness of these link removal strategies in reducing the efficiency of information transport on a given temporal network. Seven physical contact networks and six virtual networks have been considered. Interestingly, we find that the strategy based on the property of each link in the fastest time-respecting paths performs the best in most networks. Links that appear in more paths and occur earlier in each path better identify critical links for information transport. We explain why and in which kind of networks this strategy performs the best via comprehensive analysis, such as the temporal network properties, and the influence of link removal on the transport efficiency of different types of traffic demands and of different kinds of node pairs.

Properties of a link in diffusion trajectories or routing paths are special types of centrality metrics, that have not been explored systematically. Within this class, mainly the betweenness of a link in a static network has been widely studied. In this work, we have designed two properties of a link in the fastest time-respecting paths, by taking into account how frequently a link appears in these routing paths, and when it is activated in these paths or, equivalently, when it propagates the information. This illustrates the possibilities in designing such centrality metrics. Our findings also reveal the relation between the property of a link in diffusion paths and the effect of moving the link on information transport.

Our work is a starting point to explore the intervention in dynamic processes on temporal networks to link removal. Properties of links in spreading paths can be further designed, e.g., by considering the properties of the paths that a link belongs to. Properties of links in spreading paths usually have a high computational complexity. It is interesting to explore their relation with other network properties of links, especially those with a low computational complexity. This may enable efficient identification of critical links.

5.2. FUTURE WORK

Based on the results and insights in this thesis, we raise some promising future directions related to temporal network prediction and interpretation.

Temporal network modelling. In this thesis, we mainly focus on temporal network prediction, i.e., what is the future structure of a temporal network once we observe the structure of the temporal network in the past. We revealed the relation of links, i.e., to what extent the activity of a link in the future is influenced by the activity of other links in the past. This link relation may shed light on the modeling of the formation of temporal networks which is crucial in understanding and controlling the dynamics of and on temporal networks.

Predicting higher-order temporal networks. A social interaction (so-called higher-order event/interaction) among any number of people can be regarded as the activation of a hyperlink. Hence, social interactions can be represented as higher-order temporal

networks, that record the higher-order events occurring at each time stamp over time. The prediction of higher-order interactions is usually overlooked in traditional temporal network prediction methods, where a higher-order interaction is regarded as a set of pairwise interactions. Our finding that activities of neighboring links that form a triangle with a target link have prediction power on the connection of the target link may suggest that higher-order events like triangles in each network snapshot may contribute to the prediction of higher-order temporal networks.

Intervention in information transport on temporal networks via blocking higher-order structure. The higher-order interactions on temporal networks have been found to impact the dynamic process on temporal networks. The understanding of the removal of which types of higher-order interactions negatively influence the efficiency of information diffusion on temporal networks the most is important to identify critical higher-order interactions. Such understanding also can enable better protection (intervention) to facilitate (prohibit) the spread of (mis)information. Therefore, studying the Intervention in information transport on temporal networks via blocking higher-order structures is also a promising topic for future studies.

ACKNOWLEDGEMENTS

First of all, I am immensely grateful to my promoter, Alan, and co-promoter, Huijuan, for their invaluable patience, guidance, and feedback. They have provided me with tremendous opportunities for learning and improvement. I express my gratitude to Alan for fostering a welcoming, international, diverse, and inclusive group environment for us PhD students. This atmosphere allowed me to thoroughly enjoy my time in the office, engaging in conversations with colleagues from various backgrounds.

Words cannot express my gratitude to Huijuan, my daily supervisor. To be honest, there were moments during the early stages of my Ph.D. studies when I complained about your strictness and frequent pointed comments. However, whenever people ask me which work I am most proud of, I always answer it's the work finished under your strict supervision. I've come to realize that your strict approach prompts me to think deeply, make significant progress, and elevate the quality of my research. The methods of thinking and conducting research you taught me are the most valuable acquisition I have ever gained. Also, I appreciate your patience, encouragement, and support in both work and life.

Secondly, I express my gratitude to my colleagues in the MMC group. I cherished the moments of brainstorming, and dinners, as well as our experiences attending conferences abroad together with Xiu-Xiu, Alberto, Shilun, Tongjing, Lingbo, and Omar. Having group meetings and lunch with Yuhui, Varnika, and Mathieu was also a good and special time. Participating in a half-marathon and bouldering with Dimme, Andrea, and Mohammad was truly an incredible experience. I also enjoyed the moments of coffee breaks with Manel, Jorge, Roger, Antony, Bence, Zhengjun, Yuanyuan, Sandy, Saskia, Xinsheng, Maosheng, Bishwadeep, Jay, Megha, Odette, Julian, Cynthia, Elvin, Patrick, Matteo, Tianrui, Tianqi, Chenggen, and Tianqi. Additionally, I want to extend my thanks to Ruud for providing the hardware support that enabled me to complete all the simulation work on time.

Thirdly, my appreciation goes to Prof. Piet Van Mieghen and Prof. Rob Kooij for allowing me to participate in the NAS meeting and build connections with colleagues in the NAS group. Engaging in conversations with Fenghua, Elizaveta, Zhihao, Xinhua, Yingyue, Edgar, Sergey, Fredrick, and other members of the NAS group made my Ph.D. journey so memorable.

Next, I want to express my gratitude to the friends I made throughout my Ph.D. journey. Special thanks to Lili and Long for coming to my rescue whenever I found myself locked outside of my place, as well as providing all kinds of sincere advice. Thank Shilun and Fenghua for their kind support during tough times, and Shilun specifically for being my coach on my gym journey. I am thankful to Mo for being my first friend at TU Delft and introducing me to your research group. Additionally, I'm grateful to Jeroen for sharing Dutch culture and engaging in many interesting discussions. Junhan, Xiangwei, and Qisong, thank you for sharing wonderful topics. Playing badminton with Xiaowei, Yun,

Guotai, and other friends was also a source of fun for me. My time in Delft was made delightful by all these friends. Last but not least, my gratitude goes to my mother, my father, and my sister. Thanks for your unconditional love and support.

CURRICULUM VITÆ

Li Zou

EDUCATION

- 2011–2015 Chengdu University of Technology, China
B.Sc. in Mathematics and Applied Mathematics
- 2016–2019 Beijing Normal University, China
M.Sc. in Systems Theory
Supervisor: Prof. dr. Y. Fan
Thesis: Quantifying network structural dissimilarity and predicting node age
- 2019–2023 Delft University of Technology, the Netherlands
Ph.D. in Computer Science
Promotor: Prof. dr. A. Hanjalic
Promotor: Dr. ir. H. Wang
Thesis: Prediction and intervention strategy design on temporal networks

LIST OF PUBLICATIONS

6. **L. Zou**, and H. Wang, *Intervention in information transport on temporal networks*, IEEE Transactions on Network Science and Engineering (submitted).
5. **L. Zou**, C. Ceria, and H. Wang, *Short- and long-term temporal network prediction based on network memory*, [Applied Network Science 18, 639 \(2023\)](#).
4. **L. Zou**, A. Wang, and H. Wang, *Memory Based Temporal Network Prediction*, In Conference on Complex Networks and Their Applications XI, Springer, Cham, 661-673 (2023).
3. **L. Zou**, X.-X. Zhan, J. Sun, A. Hanjalic, and H. Wang, *Temporal Network Prediction and Interpretation*, IEEE Transactions on Network Science and Engineering 9, 1215-1224 (2022).
2. **L. Zou**, C. Wang, A. Zeng, Y. Fan, and Z. Di, *Link prediction in growing networks with aging*, Social Networks 65, 1-7 (2021).
1. **L. Zou**, A. Wang, A. Zeng, Y. Fan, and Z. Di, *Connecting node age with centrality measurement in growing networks*, Journal of Statistical Mechanics: Theory and Experiment 7, 073403 (2019).