计算机研究与发展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202220776
61(2): 518−528, 2024

基于时序图卷积的动态网络链路预测

刘琳岚[1] 冯振兴[1] 舒 坚[2]
[1](南昌航空大学信息工程学院 南昌 330063)
[2](南昌航空大学软件学院 南昌 330063)
(765693987@qq.com)

Dynamic Network Link Prediction Based on Sequential Graph Convolution

Liu Linlan[1], Feng Zhenxing[1], and Shu Jian[2]
[1](*School of Information Engineering*, *Nanchang Hangkong University*, *Nanchang* 330063)
[2](*School of Software*, *Nanchang Hangkong University*, *Nanchang* 330063)

Abstract Dynamic network link prediction has become a hot topic in network science field because of its wide application prospect. However, the complexity of spatial correlation and temporal dependence in the evolution process of dynamic network links leads to the great challenges of dynamic network link prediction task. In this paper, a dynamic network link prediction model based on sequential graph convolution (DNLP-SGC) is proposed. On the one hand, because network snapshot sequence cannot effectively reflect the continuity of dynamic network evolution, the edge trigger mechanism is employed to modify the original network weight matrix, so as to make up loss timing information in discrete snapshot of dynamic network. On the other hand, from the view of network evolution and considering the feature similarity and historical interaction information between nodes, a temporal graph convolution method is proposed to extract node features in dynamic network, and the method integrates the spatial-temporal dependence of nodes effectively. Furthermore, the causal convolutional network is used to capture the potential global temporal features in the dynamic network evolution process to achieve dynamic network link prediction. Experimental results on two real dynamic network datasets show that DNLP-SGC outperforms the baseline model on three common indexes, such as *precision*, *recall* and *AUC*.

Key words dynamic network; link prediction; sequential graph convolution; global temporal feature; causal convolution

摘 要 动态网络链路预测广泛的应用前景,使得其逐渐成为网络科学研究的热点.动态网络链路演化过程中具有复杂的空间相关性和时间依赖性,导致其链路预测任务极具挑战.提出一个基于时序图卷积的动态网络链路预测模型(dynamic network link prediction based on sequential graph convolution, DNLP-SGC).针对网络快照序列不能有效反映动态网络连续性的问题,采用边缘触发机制对原始网络权重矩阵进行修正,弥补了离散快照表示动态网络存在时序信息丢失的不足.从网络演化过程出发,综合考虑节点间的特征相似性以及历史交互信息,采用时序图卷积提取动态网络中节点的特征,该方法融合了节点时空依赖关系.进一步,采用因果卷积网络捕获网络演化过程中潜在的全局时序特征,实现动态网络链路预测.在2个真实的网络数据集上的实验结果表明,DNLP-SGC在*precision*, *recall*, *AUC*指标上均优于对比的基线模型.

关键词 动态网络;链路预测;时序图卷积;全局时序特征;因果卷积

中图法分类号 TP393

收稿日期:2022-09-01;**修回日期**:2023-03-13

基金项目:国家自然科学基金项目(62062050, 61962037)
This work was supported by the National Natural Science Foundation of China (62062050, 61962037).

通信作者:舒坚(shujian@nchu.edu.cn)

网络可用于建模自然界和社会中许多复杂系统，其中网络中的每个顶点表示一个实体，每条边代表一对顶点之间的联系[1]. 网络分析的目的是通过挖掘网络的拓扑结构来提取有用的图模式，从而揭示底层系统的结构和功能[2]. 真实网络中节点间的交互联系随时间推移而持续变化，网络结构也因此不断演化更新. 相比于静态网络的一成不变，动态网络中节点和连边状态会随着时间不断演化，例如移动社交网络、邮件网络等，网络中的节点会随着时间的推移增加和移除，节点间的链接也随之产生和消失. 静态网络链路预测任务是发现隐藏的连接，与之不同的动态网络链路预测的目的是预测下一时刻网络中会出现的链接.几乎真实世界中的所有现象都能够用动态网络表征，理解这些网络如何演化，有助于我们研究和分析这些复杂系统，因此能否准确地预测下一时刻网络的链路状态对我们研究复杂网络具有重要意义. 本文研究移动社会网络，其链路有连接和断开2种状态.

现有的动态网络链路预测工作主要存在3方面问题：1)大量基于离散快照序列建模动态网络的模型，不能充分反映动态网络的连续性，会造成时序信息的丢失；2)现有研究通常将网络演化过程中的拓扑特征和时序特征当作独立的影响因素，忽略了两者在动态网络演化中的相互作用；3)现有研究大多注重于短期内网络链路演化受到的影响，忽略了网络长期演化规律. 为解决这3个问题，本文提出了一种基于时序图卷积的动态网络链路预测(dynamic network link prediction based on sequential graph convolution, SGC-DNLP)模型. 对于问题1，本文通过引入边缘触发机制弥补离散快照表示动态网络造成的时序信息丢失. 对于问题2，受到文献[3]的启发，本文提出时序图卷积提取网络节点演化特征，从动态网络演化机制出发，考虑到网络演化过程中受到的不同影响：首先节点会影响相同时段内其相邻节点；其次，由于时间序列中的相关性，每个节点会影响自身下个时刻的状态；最后，由于节点的时空相关性，每个节点可以影响下一个时刻其相邻节点. 对于问题3，本文通过结合因果卷积网络提取网络全局时序特征. 本文的主要贡献有3点：

1)基于边缘触发原理构建了动态网络信号矩阵，并基于该矩阵修正每个离散快照的权重，将原本独立的静态快照联系起来，一定程度上避免了每片静态快照在时序信息上的丢失，弥补了离散快照序列建模动态网络的不足.

2)提出基于时序图卷积的动态网络节点特征提取方法. 该方法从动态网络演化过程出发，针对拓扑特征和时序特征在动态网络演化中起到的共同作用，结合节点在网络演化过程中受到的主要影响因素，有效提取动态网络节点的短期时空依赖特征.

3)提出基于因果卷积的动态网络全局时序特征提取方法. 在获得短期时空依赖特征的基础上，该方法采用因果卷积网络捕获长期的网络演化规律，并结合长短期演化特征，提高链路预测的性能.

1 研究现状

近些年越来越多的工作针对动态网络的链路预测，这些工作主要分为基于相似性指标的链路预测方法、基于图嵌入和机器学习的链路预测方法、基于图神经网络的链路预测方法.

基于相似性指标的链路预测方法一般是将节点的属性、特征的相似度作为产生连接的依据. 文献[4]通过混合CN(common neighbors)，Adamic-Adar，Jaccard等指标，基于对过去节点相似性的变化以及外部因素的建模进行链路预测. 文献[5]提出一种基于内容相似性的链接预测的新指标LDAcosin，用于确定动态网络节点间未来可能发生的新交互. 文献[6]提出的ASSPL通过考虑节点属性相似度和最短路径长度来预测新节点的未来链路. 基于相似性的链路预测方法比较适用于网络拓扑结构变化不大的情况，且较为依赖人工设定的网络特征. 随着网络规模的增大和结构的频繁演化，基于相似性指标的链路预测方法略显不足.

基于图嵌入和机器学习的链路预测方法通常通过图嵌入技术将高维的网络数据映射到低维空间，进一步通过机器学习模型处理例如链路预测等下游任务. 文献[7]将词嵌入引入到网络分析中并提出Deepwalk模型以将网络中的节点通过低维向量表示. 文献[8]通过受限的随机游走过程进一步提出node2vec模型，学得的节点嵌入能够更好体现网络特征和节点邻居特征. 在静态图嵌入的基础上，文献[9]提出CTDNE模型，它是一个将时序信息结合到网络嵌入算法中的通用框架，通过从连续的时间动态网络中学习动态网络嵌入. 文献[10]提出DNESA方法，通过在注意力机制下对开放三元组发展为封闭三元组的过程建模，捕获动态网络演化特征，学习节点在不同步长下的嵌入向量.

随着图神经网络的不断发展，越来越多的研究

基于图神经网络预测网络未来的连接. 文献 [11] 提出 EvolveGCN 模型, 通过使用循环神经网络(recurrent neural network, RNN)演化图卷积网络(graph convolution network, GCN)参数, 以捕获图序列的动态, 解决动态网络链路预测等下游任务. 文献 [12] 通过引入生成对抗网络, 结合 GCN 和长短期记忆网络(long short term memory network, LSTM)捕获动态网络演化过程中的时空特征, 实现动态网络的链路预测. 文献 [13] 通过堆叠 LSTM 模块实现一个端到端的自编码器和动态网络的链路预测任务; 刘林峰等人[14]基于自动编码器和门控循环单元实现动态网络链路预测. 文献 [15] 提出一种新的端到端模型, 通过结合 GCN 和 LSTM 用于动态网络链路预测.

文献 [9−15] 方法主要是在现有静态网络链路预测的基础上通过加入时序信息实现动态网络的链路预测, 或者通过堆叠图神经网络和处理时序的模块实现动态网络链路预测. 同时大部分的研究采用离散快照表示动态网络, 但该方法不可避免地造成了节点间时序信息的丢失. 与这些工作相比, 本文提出的 DNLP-SGC 模型首先通过引入边缘触发机制, 弥补离散快照表示动态网络造成的节点间时序信息的丢失; 从网络演化角度出发构建时序图卷积网络学习网络节点特征, 采用因果卷积网络捕获动态网络潜在的全局时序特征实现动态网络链路预测.

2 相关定义

在本节中, 主要对动态网络中的相关定义进行介绍.

定义 1. 动态网络. 网络中节点间的连接会随时间出现和消失, 动态网络 G 是一组连续的网络快照 $\{G_1, G_2,\cdots, G_T\}$, 如图 1 所示. 其中 $G_k=(V, E_k, \boldsymbol{W}_k)$表示动态网络的第 k 片时间快照, V 为网络节点集合, $E_k\subseteq\mathbb{R}^{|V|\times|V|}$为在定长时间段 $[t_{k-1}, t_k]$ 上的时序连边集合, $\boldsymbol{W}_k$ 是 G_k 的权重矩阵, 当网络中连边无权重时 $\boldsymbol{W}_k=\boldsymbol{A}_k$, $\boldsymbol{A}_k$ 为邻接矩阵.

定义 2. 动态网络链路预测. 根据前 N 片连续的网络快照 $S=\{G_{t-N},G_{t-N-1},\cdots,G_{t-1}\}$得到下一个时间切片内的网络链路状态, $G_t=F(\{G_{t-N},G_{t-N-1},\cdots,G_{t-1}\})$, $F(\cdot)$为预测模型.

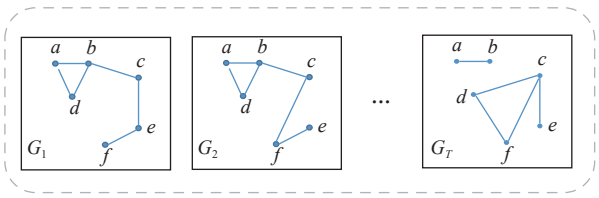


Fig. 1 Dynamic network discrete snapshot representation

图 1 动态网络离散化快照表示

定义 3. 网络样本熵. 它也指网络快照序列的变化度. 样本熵用于度量时间序列的波动性, 本文基于样本熵[16]改进得到网络样本熵 *NetSampleEn*, 网络样本熵具体计算过程如算法 1 所示.

算法 1. 网络样本熵计算算法.

输入: 网络邻接矩阵序列$\{\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_T\}$;

输出: 网络样本熵 *NetSampleEn*.

① 重构 m 维向量$(\boldsymbol{X}(1),\boldsymbol{X}(2),\cdots,\boldsymbol{X}(m))$,

$\boldsymbol{X}(i)=(\boldsymbol{A}_i,\boldsymbol{A}_{i+1},\cdots,\boldsymbol{A}_{i+m-1})$;

② for i in *range* $(1,T-m)$ do

③ $edge_i=sum\ (\|\boldsymbol{A}_i\|_1,\|\boldsymbol{A}_{i+1}\|_1,\cdots,\|\boldsymbol{A}_{i+m-1}\|_1)/m$,

$S_i^m=0$; /*记录相似序列个数*/

④ for j in *range* $(1,T-m)$ do

⑤ $d[\boldsymbol{X}(i),\boldsymbol{X}(j)]=\max\left(\left\|\boldsymbol{A}_i-\boldsymbol{A}_j\right\|_1,\cdots,\left\|\boldsymbol{A}_{i+m-1}-\boldsymbol{A}_{j+m-1}\right\|_1\right)$;

⑥ if $d[X(i),X(j)]\leqslant edge_i\times\delta$ do

S_i^m++;

⑦ end if

⑧ end for

⑨ $S^m=(T-m+1)^{-1}\sum_{i=1}^{T-m+1}S_i^m$;

⑩ end for

⑪ 令 k=m+1, 重复行①~⑩, 得S^k;

⑫ $NetSampleEn=-\ln\left(\frac{S^k}{S^m}\right)$;

⑬ return *NetSampleEn*.

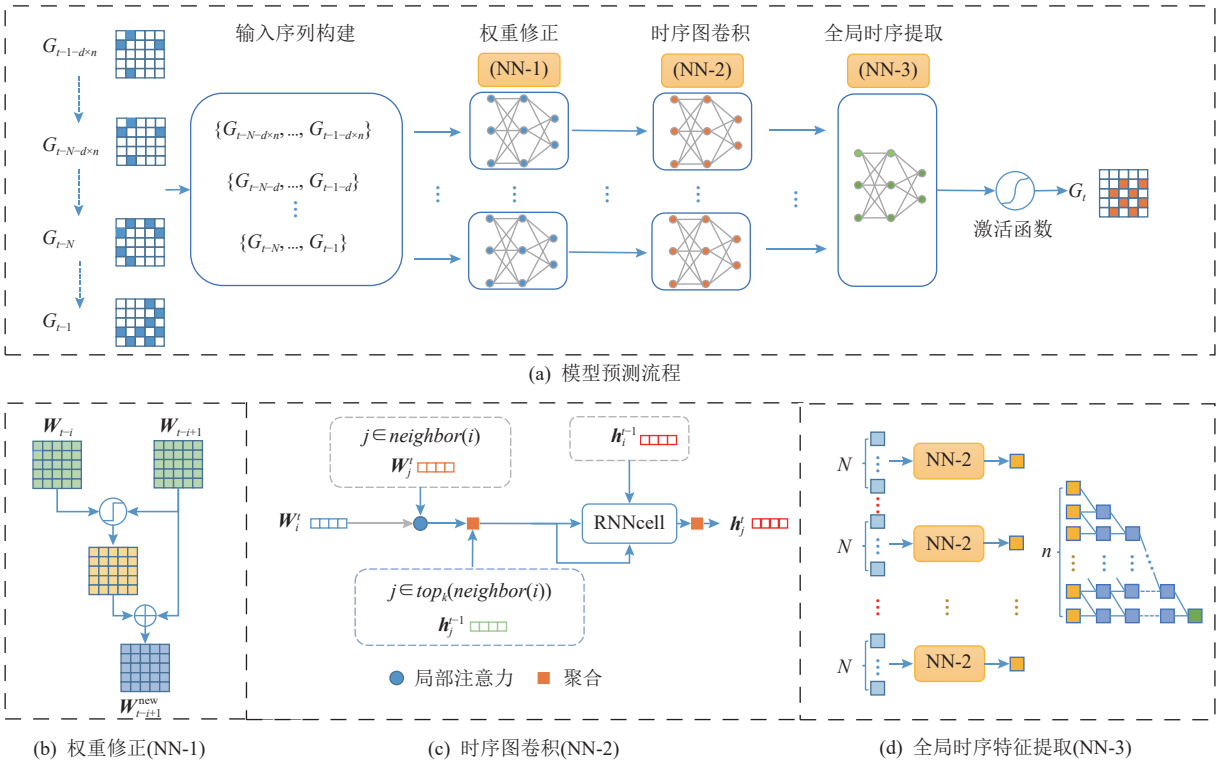
在网络样本熵计算过程中主要包含 2 个超参数 m与δ, 分别用于控制重构维度大小和调整 2 个向量间相似阈值. 基于算法 1 得到的网络样本熵随网络变化程度增加而增大.

3 DNLP-SGC 模型

本节介绍模型 DNLP-SGC, 如图 2(a)所示, 该模型包括 4 部分: 1)输入序列构建; 2)权重修正; 3)时序图卷积; 4)全局时序特征提取. 模型首先基于原始数据样本按照指定规则构建输入序列, 接着基于边缘触发机制将每个离散快照序列进行权重修正, 进一步采用时序图卷积学习节点短期时空特征, 最后利用因果卷积提取全局时序特征实现动态网络链路预测.

3.1 输入序列构建

动态网络通常以一组连续的网络快照表示,



(a) 模型预测流程

(b) 权重修正(NN-1) (c) 时序图卷积(NN-2) (d) 全局时序特征提取(NN-3)

Fig. 2 Block diagram of dynamic network link prediction based on sequence graph convolution

图 2 基于时序图卷积的动态网络链路预测框架图

其中每个快照被当作静态网络进行处理. 在本文中, 模型的输入由 n 个间隔为 d 的快照序列构成: $\{G_{t-N},\cdots,G_{t-1}\},\{G_{t-N-d},\cdots,G_{t-1-d}\},\cdots,\{G_{t-N-d\times k},\cdots,G_{t-1-d\times k}\}$. 其中 $k\in\{0,1,\cdots,n-1\}$, n 为网络快照序列个数, N 为每个序列内的快照个数. d 为每个序列间隔的快照数, 其物理意义为快照序列间的间隔时长, 例如, 当间隔时长为 1 天时, d=24×60×60s/*slice_time*, *slice_time* 为快照的切片时长. 为了防止快照序列间出现重合, 在具体构建过程中有 $d>N$. 如图 3 所示, 选取了 $n=3$ 个连续动态网络快照序列, 序列内快照个数 $N=2$, 序列间快照间隔数 $d=4$.

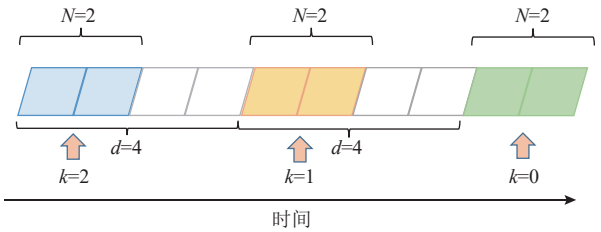


Fig. 3 Input sequence construction

图 3 输入序列构建

3.2 权重修正

由于离散快照的表示方式不能充分表达节点间连接的时序性, 进而造成了部分时序信息的丢失. 在动态有权网络中, 每个快照中连边的权重值, 仅表示某个特征在当前快照内的状态值, 例如权重值 $w_{k,i,j}$ 若反映第 k 个时间片内节点 i,j 间链接时长, 其主要体现该节点对的链接在这一时间片内的连接时长状态, 并不能反映其于与上一时间片内的变换状态. 针对该问题, 本研究借鉴数字电路技术中的边缘触发思想, 设计一个能够反映快照间连边状态变化的信号矩阵, 其构造方式如式(1)所示.

$$E_{k,i,j}=\begin{cases}1, & w_{k,i,j}>w_{k-1,i,j},\\ -1, & w_{k,i,j}<w_{k-1,i,j},\\ 0, & \text{其他}.\end{cases} \tag{1}$$

当连边强度增加时, 信号矩阵 $\boldsymbol{E}$ 会给予一个正信号, 反之将给予一个负信号, 当连边强度未发生改变时不会产生信号, 其具体生成方式如图 4 所示. 图 4 中 $\boldsymbol{W}$ 表示权重矩阵, 其中每个元素的值为归一化后的权重. 同时为了使构建的信号矩阵 $\boldsymbol{E}$ 数目与快照数目一致, 会初始化一个网络快照 $\boldsymbol{W}_{\text{init}}=\boldsymbol{W}_0$, 此时基于快照 $\boldsymbol{W}_{\text{init}}$ 和 $\boldsymbol{W}_0$ 构建的信号矩阵 $\boldsymbol{E}_0$ 为一个内部元素全为 0 的矩阵, 在信号传递上表现为无信号产生. 最终基于上述设计得到的信号矩阵 $\boldsymbol{E}$ 和初始的权重矩阵 $\boldsymbol{W}$, 模型构建一个全新的输入权重矩阵, 如图 2(b)所示, 其构建方式如式(2)所示.

$$\boldsymbol{W}_{\text{new}}=\boldsymbol{W}+\theta\times\boldsymbol{E}, \tag{2}$$

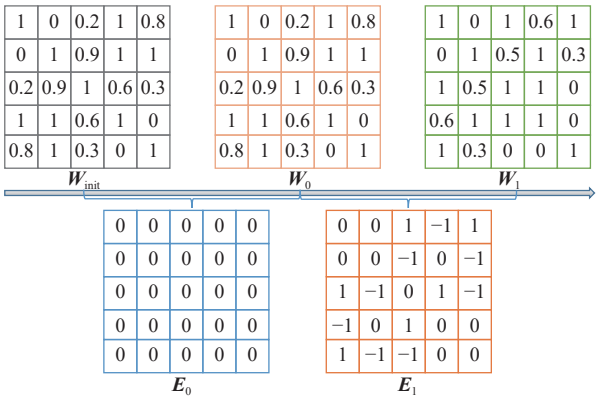


Fig. 4 Signal matrix construction process

图 4 信号矩阵构建过程

其中θ用于控制信号传递的大小, 是一个可训练参数, $\boldsymbol{W}_{new}$为更新后的权重矩阵.

3.3 时序图卷积

本文基于文献 [3] 所提到的动态网络演化可能受到的影响因素, 设计了一个时序图卷积模块, 用于学习动态网络中的节点短期时空特征, 权重修正后的快照序列会作为该模块的输入. 其时空信息聚合过程如图 2(c)所示. 时序图卷积通过融合时空特征提取过程, 能够学习节点和其邻居的时空依赖关系, 网络中节点 i 的更新过程如式(3)所示.

$$\boldsymbol{h}_i^t = F_{\text{update}}(\boldsymbol{w}_i^t, \boldsymbol{w}_{j\in neighbor(i)}^t, \boldsymbol{h}_i^{t-1}, \boldsymbol{h}_{j\in neighbor(i)}^{t-1}), \tag{3}$$

其中$\boldsymbol{w}_i{}^t$为节点 i 在 t 时间片的输入信息, $\boldsymbol{h}_i{}^t$为节点 i 在 t 时间片的状态信息, 时序图卷积模块对动态网络中的节点状态更新过程分别从空间依赖$(\boldsymbol{w}_i^t, \boldsymbol{w}_{j\in neighbor(i)}^t)$和时间依赖$(\boldsymbol{h}_i^{t-1}, \boldsymbol{h}_{j\in neighbor(i)}^{t-1})$ 2 部分出发. 首先, 在每个快照内节点基于注意力机制聚合其邻居信息, 具体细节如式(4)(5)所示.

$$\boldsymbol{att}_i^t = \boldsymbol{w}_i^t + \sum_{k\in neighbor(i)} \alpha_{i,j} \cdot \boldsymbol{w}_j^t, \tag{4}$$

$$\alpha_{i,j} = \frac{\exp\left(LeakyReLU(\boldsymbol{w}_i^t, \boldsymbol{w}_j^t)\right)}{\sum_{k\in neighbor(i)} \exp\left(LeakyReLU(\boldsymbol{w}_i^t, \boldsymbol{w}_k^t)\right)}, \tag{5}$$

其中$\alpha_{i,j}$为节点 i 和邻居 j 的注意力系数, $\boldsymbol{att}_i{}^t$为时刻 t 节点 i 聚合空间上邻居后的节点表示. 经过式(4)(5)的处理后, 每个节点能够聚合快照内其邻居的信息, 实现网络局部空间依赖关系的提取.

动态网络演化过程中, 节点下个快照内的状态除了受到当前拓扑结构的影响, 还应受到其上一个时刻的状态$\boldsymbol{h}_i^{t-1}$以及其邻居上一个时刻的状态$\boldsymbol{h}_{j\in neighbor(i)}^{t-1}$的影响. 考虑到模型的计算开销以及其对重要信息的提取, 在聚合节点邻居的节点历史信息的过程中, 可考虑节点 i 邻居的注意力系数α中排名前 top_k 个邻居的历史信息, 具体细节如式(6)~(8)所示.

$$\boldsymbol{h}_i^{t-1} = \boldsymbol{h}_i^t + \boldsymbol{m}_i^t, \tag{6}$$

$$\boldsymbol{m}_i^t = \gamma \boldsymbol{att}_i^t + \eta \sum_{j\in top_k} \boldsymbol{h}_j^{t-1}, \tag{7}$$

$$\boldsymbol{h}_i^t = RNNCell(\boldsymbol{m}_i^t, \boldsymbol{h}_i^{t-1}), \tag{8}$$

其中$\boldsymbol{h}_i^t$为节点 i 在时刻 t 的状态, $\boldsymbol{m}_i^t$为节点 i 更新后的输入, $\boldsymbol{h}_i^{t-1}$为节点 i 上一个时刻状态信息, γ,η为可学习的权重参数, $RNNCell(\cdot)$表示循环神经网络中的最小单元. 式(6)~(8)节点下一时刻的状态不仅取决于其自身上一个时刻的状态, 还与其邻居的状态相关. 由单个节点对应到整个网络权重矩阵的输入如式(9)所示.

$$\boldsymbol{H}^t = RNNCell(\boldsymbol{M}^t, \boldsymbol{H}^{t-1}) + \boldsymbol{M}^t, \tag{9}$$

其中$\boldsymbol{H}^t$为时序图卷积的最终输出. 模型通过时序图卷积单元处理时间间隔为 d 的 n 个网络快照序列, 最终得到了网络链接状态预测序列$T = \left(\boldsymbol{H}^{t-d\times(n-1)}, \boldsymbol{H}^{t-d\times(n-2)}, \cdots, \boldsymbol{H}^t\right)$.

3.4 基于全局时序特征的链路预测

3.3 节中时序图卷积用于处理每个快照序列, 由于每个序列长度有限, 导致时序图卷积感受野受限只能提取短期特征. 本文利用因果卷积[17]提取网络全局时序特征. 动态网络中存在一些全局的时序特征, 例如在校园社交网络中, 学生节点会在中午时段产生大量链接, 工作日相较于周末更容易产生连接等. 这种全局时序特征构成了动态网络演化规律的一部分, 本文通过在时序图卷积后堆叠因果卷积, 提取网络的全局时序性特征, 其结构如图 2(c)所示. 因果卷积的具体细节如式(10)所示.

$$Hid(n) = (\boldsymbol{T} \times f)(n) = \sum_{i=0}^{m-1} f(i) \boldsymbol{T}^{n-p\times i}, \tag{10}$$

其中 m 为卷积核大小, p 为扩展系数, $\boldsymbol{T}$ 为上一节时序图卷积处理得到的网络拓扑状态预测序列, $Hid(n)$为因果卷积过程中第一层中单个卷积核计算函数. 在深度学习模型中, 较深的神经网络模型会遇到梯度爆炸和网络退化的问题, 通过引入残差连接能够一定程度上解决这 2 个问题[18-19]. 引入残差连接后的因果卷积公式如式(11)所示.

$$Hid(n) = (\boldsymbol{T} \times f)(n) = \left(\sum_{i=0}^{m-1} f(i) \boldsymbol{T}^{n-p\times i}\right) + \boldsymbol{T}^n. \tag{11}$$

最终, 序列$\left[\boldsymbol{H}^{t-d\times(n-1)}, \boldsymbol{H}^{t-d\times(n-2)}, \cdots, \boldsymbol{H}^t\right]$在经过带残差的因果卷积网络处理后, 在其最后一层得到网络的最终输出$\boldsymbol{H}_{new}^t$, 通过激活函数映射到 {0, 1}得到动

态网络下一个时刻的链路预测结果 $\boldsymbol{A}_t$，如式(12)所示.

$$\boldsymbol{A}_t^{\text{pred}} = sigmoid(\boldsymbol{H}_{\text{new}}^t). \tag{12}$$

链路预测过程如图 5 所示.

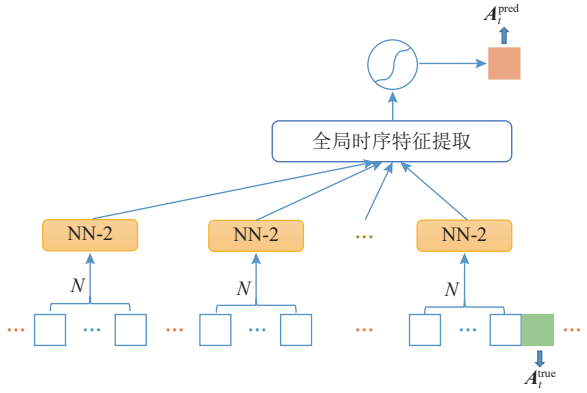


Fig. 5 Link prediction process

图 5 链路预测过程

3.5 损失函数

整个模型的训练目标是使预测结果 $\boldsymbol{A}_t^{\text{pred}}$ 接近于真实的网络邻接矩阵 $\boldsymbol{A}_t^{\text{true}}$，由于在大部分的动态网络真实出现的节点间，连边与其所有可能出现的情况相比，真实连边的占比较小，这使得模型在训练过程中会注重预测无连边的状态，即邻接矩阵中的 0 值，同时随着模型参数的增加不可避免地在训练过程中会遇到过拟合问题. 为了解决此问题，本文在逻辑回归的基础上做修改，增强目标函数对动态网络中不平衡数据的优化，对非零元素添加惩罚值，增强反向传播过程中真实存在链接的影响. 同时通过添加正则化损失函数计算模型中所有权重的 F 范数的平方和，防止网络过拟合现象.

$$loss = l_1 + \beta l_{\text{reg}}, \tag{13}$$

$$l_1 = -\sum_{i=1}^{|V|}\sum_{j=1}^{|V|}(Q \times label_1 \times \log(A_{ij}^{\text{pred}}) + label_0 \times \text{lb}(1 - A_{ij}^{\text{pred}})), \tag{14}$$

$$l_{\text{reg}} = \|\boldsymbol{W}_{\text{model}}\|_2, \tag{15}$$

其中惩罚系数 Q 用于对 $label_1$ 即有连边的情况增大其训练权重，l_{reg} 用于避免模型陷入过拟合，$\boldsymbol{W}_{\text{model}}$ 为模型参数，β 为控制正则化损失重要程度的参数.

由于模型在时序图卷积模块采用循环神经网络单元来存储上一个状态的信息，当序列长度增加时，反向传播过程中梯度连乘项数目增多. 同理，在全局时序特征提取模块中随着序列长度 n 的增大，隐藏层数增加，也会导致梯度连乘项增加，容易引起梯度爆炸. 对于梯度爆炸问题，本文通过引入梯度裁剪技术缓解该问题[20]，其细节如式(16)所示.

$$\boldsymbol{g} = \begin{cases} \dfrac{threshold}{\|\boldsymbol{g}\|}\boldsymbol{g}, & \|\boldsymbol{g}\| \geqslant threshold, \\ \boldsymbol{g}, & \text{其他}, \end{cases} \tag{16}$$

其中 $\boldsymbol{g}$ 为模型参数梯度值，$\|\boldsymbol{g}\|$ 为二范数值，$threshold$ 为阈值.

4 实验结果与分析

4.1 实验数据集

本文采用动态网络数据集 ITC 和 MIT[21] 对模型的有效性进行验证，这 2 个数据集的主要信息如表 1 所示.

Table 1 Datasets Information

表 1 数据集信息

数据集	节点数	连边数	持续天数
ITC	50	6 800	12
MIT	97	102 500	246

1)ITC.该数据集源于剑桥大学校园学生轨迹的可视化实验，记录了 12 天中 50 名学生携带 iMote 产生的连接数据.

2)MIT.该数据集记录了 97 个携带 Nokia6600 手机用户自 2004 年 10 月至 2005 年 5 月共 246 天，利用蓝牙接口相互通信的数据.

图 6 展示了数据集在切片时长为 30 min 时的每个快照内的连边总数(MIT 数据集抽取 30 天数据)变化. 网络在经过连续的 5 天频繁连接后，后 2 天网络的连接数目急剧下降，该现象交替出现，呈现出长期时序特征，该现象在持续时间较长的 MIT 数据集中更为明显. 在相邻天数中，网络的连接数目交替增减，对应到校园网络中白天学生产生大量交互，晚上几乎不产生交互. 通过分析数据集，可以得到结论：网络随时间演化过程中，节点间链接的产生与消失存在全局时序特征，动态网络中节点间连接状态不仅与其最近的时间状态相关，还与其在前一天的同一时段状态相关.

4.2 评价指标与对比方法

由于在动态网络的演化过程中，网络的链路状态只有连接和断开 2 种，因此动态网络的链路预测可看作是一个二分类问题，即判断下一时刻节点间是否产生连接. 通常用来衡量链路预测算法的有效性指标一般有精准率、召回率和 AUC 这 3 种.

精准率($precision$)是在模型预测为正例的所有结果中模型预测正确的比重，其计算如式(17)所示.

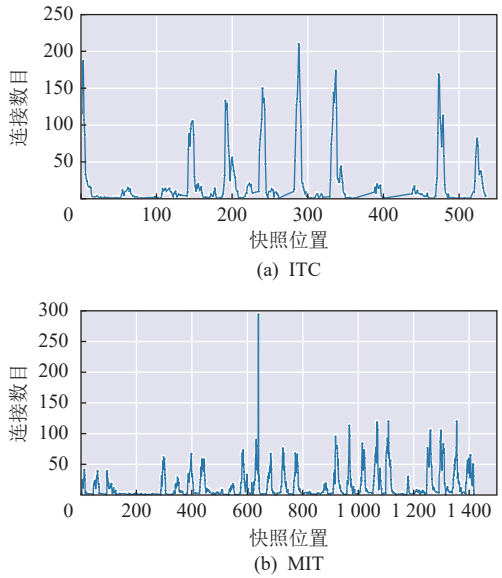


Fig. 6 Dataset nodes connection status

图 6 数据集节点连接状况

$$precision = \frac{TP}{TP+FP}, \tag{17}$$

其中 TP 表示模型将正样本预测为正例的数目, FP 表示模型将负样本预测为正例的数目.

召回率(*recall*)是在真实值为正例的所有结果中模型预测正确的比重, 其计算如式(18)所示.

$$recall = \frac{TP}{TP+FN}, \tag{18}$$

其中 FN 表示模型将正样本预测为负例的数目.

AUC 是衡量二分类模型优劣的一种评价指标, 其计算如式(19)所示.

$$AUC = \frac{n' + 0.5n''}{n}, \tag{19}$$

其中 n 为比较次数, n' 为正例中所选边分数大于负例中所选边分数的次数, n'' 为两者相等的次数.

为了验证模型的效果, 本文选取了 Self-Attention[22], TCN[17], E-LSTM-D[13], GCN-GAN[12], GC-LSTM[15] 作为基线模型, 为确保对比实验的公平性, 这些模型均以连续的 5 片网络快照作为模型的输入.

4.3 实验结果及分析

实验过程中首先需要对动态网络数据集进行时间切片, 将数据集划分为多个连续的网络快照, 在本文设计的实验中以 300 s 作为切片时长, 同时考虑到实验所采用的数据集中节点均代表人类, 因此选用符合人类行为周期的超参数作为模型输入的选取依据, 最终以 d=(24×60×60 s)/(300 s)为时间跨度, 选取 n=7 个这样连续的序列作为整个模型的输入参数, 训练过程中, 为防止梯度爆炸, *threshold* 取经验值 $10^{[20]}$.

表 2 总结了本文提出的 DNLP-SGC 和基线方法的 *AUC*, *precision*, *recall*. 从实验结果中观察到, 与所有数据集的最佳基线模型相比, DNLP-SGC 实现了 1%~5% 的平均评分增益.

Table 2 Comparison of Prediction Models

表 2 预测模型对比

评价指标	模型	ITC	MIT
AUC	Self-Attention	0.956 1	0.959 6
	TCN	0.948 3	0.950 1
	E-LSTM-D	0.944 2	0.950 9
	GCN-GAN	0.932 1	0.955 1
	GC-LSTM	0.949 8	0.960 1
	DNLP-SGC (本文)	**0.965 6**	**0.970 1**
precision	Self-Attention	0.918 8	0.919 3
	TCN	0.899 4	0.901 5
	E-LSTM-D	0.900 4	0.905 8
	GCN-GAN	0.861 4	0.871 7
	GC-LSTM	0.900 1	0.908 6
	DNLP-SGC (本文)	**0.939 4**	**0.943 8**
recall	Self-Attention	0.804 8	0.849 9
	TCN	0.897 4	0.901 4
	E-LSTM-D	0.899 5	0.910 1
	GCN-GAN	0.857 5	0.900 4
	GC-LSTM	0.900 4	0.912 3
	DNLP-SGC (本文)	**0.931 9**	**0.940 2**

注:加粗数字表示最优值.

进一步分析, 动态网络演化的过程中, 其链路状态受到时间和空间的影响, 并且时间和空间之间并非独立, 将两者分开处理将忽视其相互作用所产生的影响. Self-Attention, TCN 等时间序列模型仅考虑时序变化, 只学习每条节点连边自身在时间上的演化规律, 并未考虑动态网络中节点间的相互作用, 忽视了节点间的空间依赖关系对网络演化的影响, 造成了信息的丢失, 因此动态网络链路预测效果较差. 基于堆叠 GNN 与时序模型的方法 E-LSTM-D, GCN-GAN, GC-LSTM 等, 将动态网络演化过程中节点间的时空特征分开处理, 然而该过程中时间和空间并非独立, 将两者分开处理将忽视其相互作用对网络演化所产生的影响, 进而导致预测精度受限.

大量动态网络链路预测研究方案中, 网络被划分为一组连续的网络快照, 每个快照被当作静态网络处理, 这种做法不可避免地破坏了网络演化过程中的连续性. 因此本文提出了动态网络信号矩阵, 并基于该矩阵对快照原始的权重矩阵进行修正, 从而

弥补离散快照表示网络造成的连续信息丢失的问题. 为了验证本文方法的有效性，将去除该模块后的模型与原始的模型进行对比，实验结果如图 7 所示.

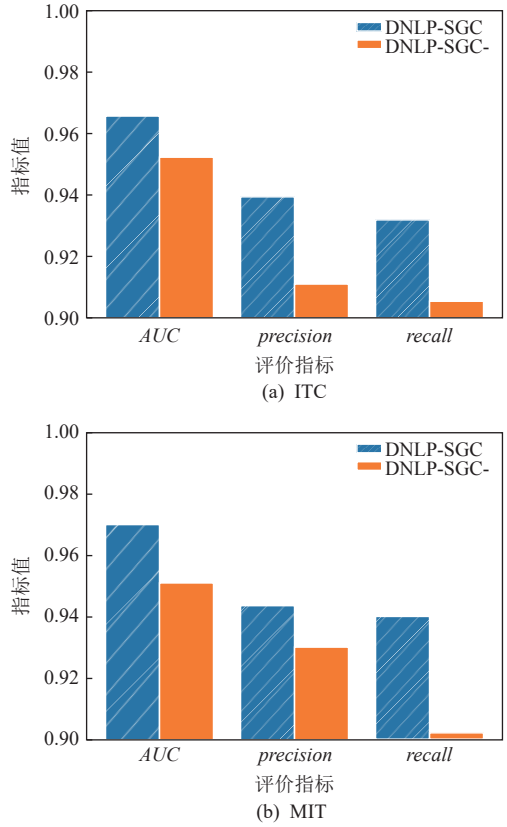


Fig. 7 Effective validation of signal matrix

图 7 信号矩阵有效性验证

图 7 中 DNLP-SGC 为原始的模型，DNLP-SGC-表示去除权重矩阵修正模块后的模型，实验结果中 DNLP-SGC 相较于 DNLP-SGC-实现了 1%~3% 的平均评分增益. 该实验同时也证明了，基于离散快照的动态网络表示方法造成了相邻快照内节点部分信息的丢失，进而降低了模型的预测效果，本文设计的基于信号矩阵修正网络权重的方案，能在一定程度上弥补由于离散化快照表示动态网络造成的节点时序信息的丢失.

为了分析本文提出的信号矩阵是如何弥补离散化快照表示动态网络造成的时序信息丢失，本文基于定义 3 提出的网络样本熵和式(2)中的参数θ，综合分析两者间的相关性，验证模型是否能够根据不同切片下的网络样本熵做出相应的修正. 基于离散化快照表示动态网络的方法首先需要按照确定的时长对原网络进行划分，在不同的切片时长下，相邻切片间的变化程度不一致，本文提出网络样本熵用于量化这种变化程度. 实验过程中向量重构大小[16]m=2，δ是用于判断向量间是否相似的阈值，$\delta\in\{0.05, 0.10, 0.20, 0.30, 0.40\}$. 如图 8 所示，2 个数据集在不同切片时长下的网络样本熵，可以看出在 240 s, 300 s, 360 s, 420 s, 480 s, 540 s, 600 s 的快照划分粒度下 2 个数据集的网络样本熵整体呈现上升趋势，反映在 240~600 s 切片时长下，随着切片时长的增大，2 个数据集相邻快照间的相似度减小，动态网络离散快照序列的变化度增大.

在不同的切片时长下，DNLP-SGC 的权重修正参数θ在训练过程中初始值均设为 0.1，通过不断地迭代训练模型最终达到收敛. 分别在切片时长为 240 s, 300 s, 360 s, 420 s, 480 s, 540 s, 600 s 时，收集 30 次模型达到收敛情况下对应的θ值，其变化规律如图 9 所示.DNLP-SGC 和去除权重矩阵修正模块后的模型 DNLP-SGC-在切片时长为 240 s, 300 s, 360 s, 420 s, 480 s, 540 s, 600 s 时的 *AUC*, *precision*, *recall* 值如图 10~12 所示.

图 9 为δ=0.2 时，2 个数据集在不同切片时长下

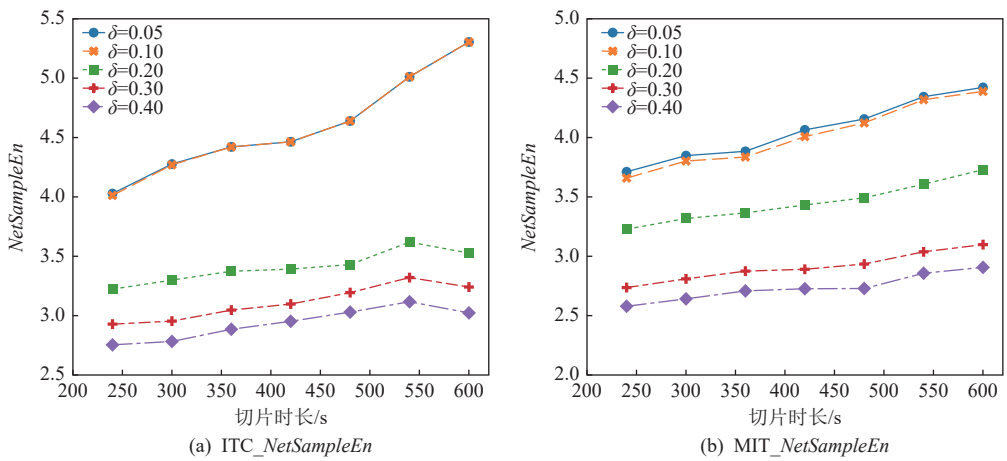
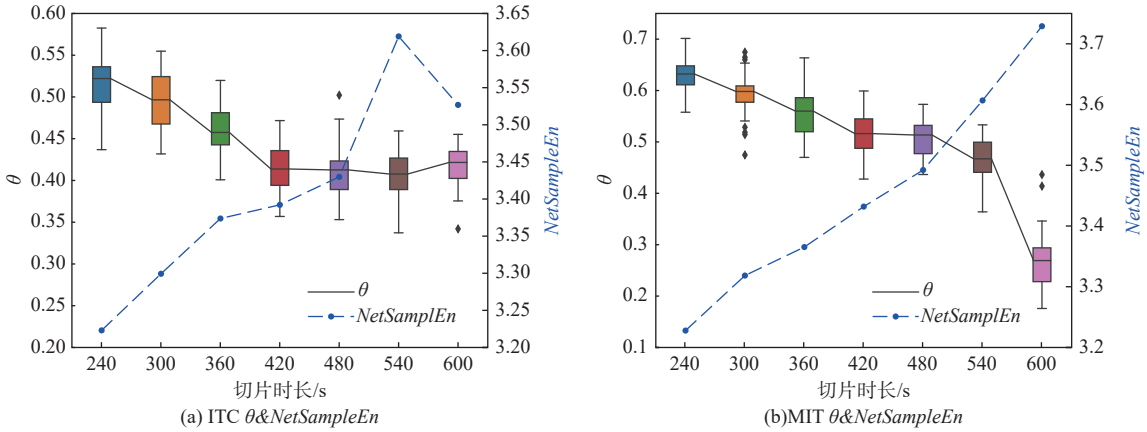


Fig. 8 *NetSampleEn* at different scales

图 8 不同尺度下的 *NetSampleEn*

的网络样本熵. 2 个数据集的 *NetSampleEn* 整体呈现上升趋势, 意味着相邻快照间的相关性减小, 网络变化度增大; 其次当模型收敛时, 参数 θ 值总能稳定在某个范围内, 随着切片时长的增加, θ 整体呈现下降趋势, 即随着网络样本熵的增大 θ 减小. 实验表明, 模型能够根据不同切片下的网络变化度做出相应的修正, 随着网络样本熵的增大, 模型对其时序信息弥补能力下降, 随着网络样本熵的增加, 各评价指标也呈现出一定的下降趋势. 同时, 如图 10~12 所示; 在不同切片时长下, DNLP-SGC 各评价指标结果均高于无权重修正的 DNLP-SGC-, 且随着网络变化度的增大, DNLP-SGC 的下降趋势也明显低于 DNLP-SGC-, 侧面反映出 DNLP-SGC 能够根据不同的网络变化度做出相应的修正.

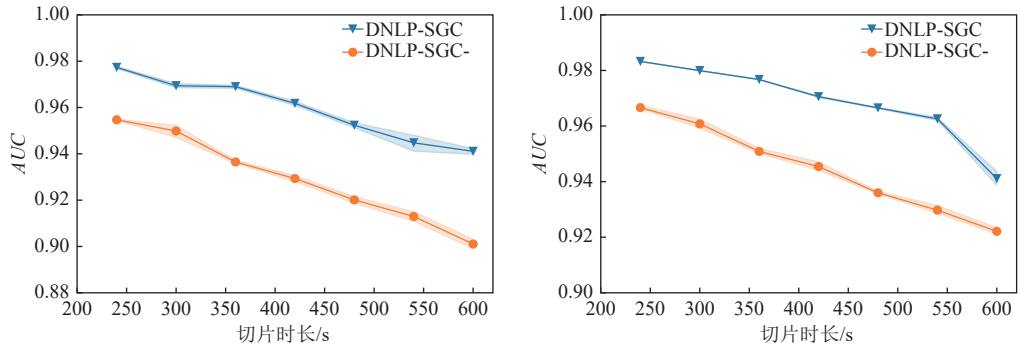
同时进一步通过实验分析了每个序列内快照个数 N 对模型预测效果的影响, 实验结果如图 13 所示.



(a) ITC θ&*NetSampleEn* (b)MIT θ&*NetSampleEn*

Fig. 9 θ and *NetSampleEn*

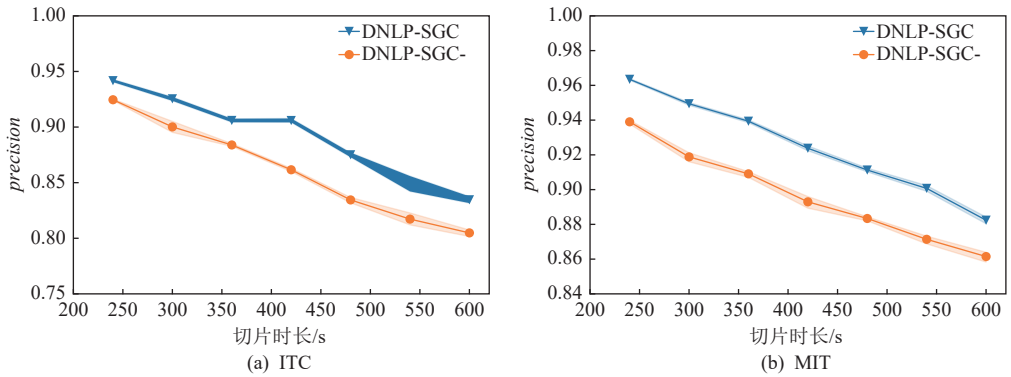
图 9 θ与网络样本熵



(a) ITC (b) MIT

Fig. 10 *AUC* at different slicing time

图 10 不同切片时长下的 *AUC*



(a) ITC (b) MIT

Fig. 11 *precision* at different slicing time

图 11 不同切片时长下的 *precision*

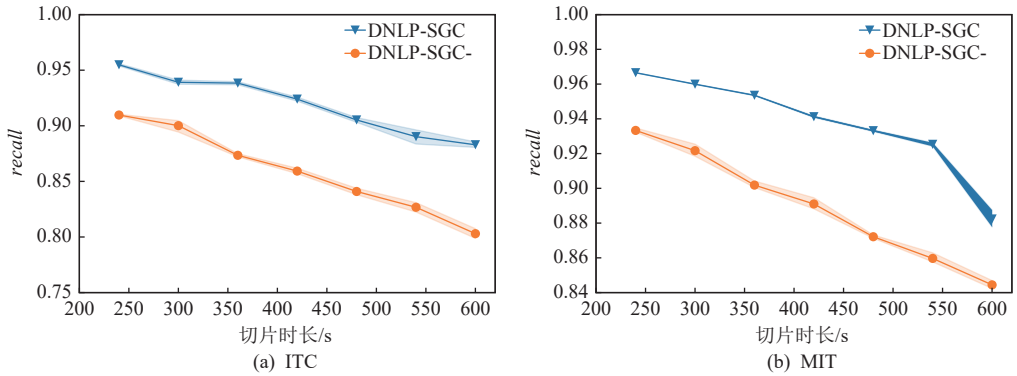


Fig. 12 *recall* at different slicing time

图 12 不同切片时长下的 *recall*

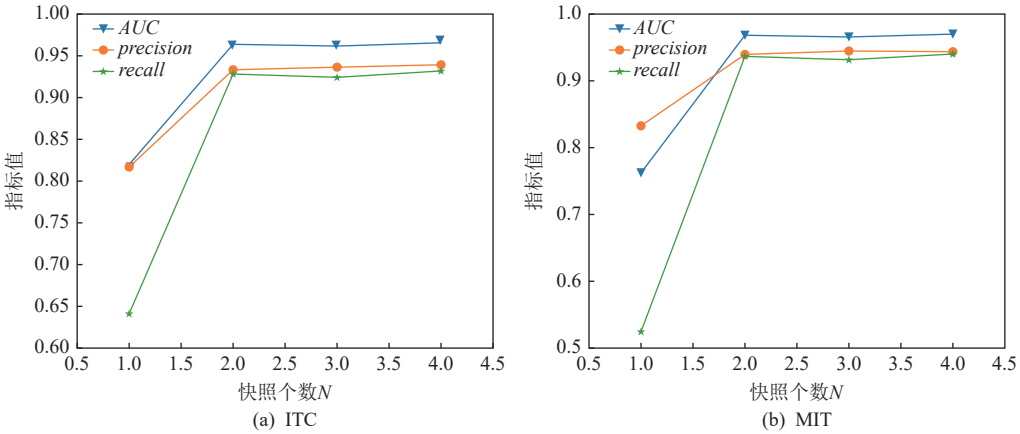


Fig. 13 The effect of sequence length on model effectiveness

图 13 序列长度对模型效果的影响

首先随着快照个数 N 的增加，在 2 个数据集上模型的预测效果在 3 个指标上都有一定的提升. N=1~2 的效果提升最明显，这是由于当 N=1 时，模型只会考虑最近的网络快照对网络下一时刻的影响，此时每个序列只是一张静态图，节点间大量的时间信息丢失，并且信号矩阵为零，网络的权重矩阵没有得到任何的修正，随着连续网络快照个数的增加，时序上的信息被模型捕获进而提高了模型预测的效果.

5 总结展望

本文提出了一种基于时序图卷积的动态网络链路预测方法，主要由权重修正单元、时序图卷积和全局时序特征提取模块 3 部分组成. 权重修正单元用于弥补离散快照表示动态网络造成的时序信息丢失，时序图卷积用于学习网络演化过程中短期的时空依赖，全局时序特征提取模块用于捕获长期的网络演化规律. 通过在 2 个真实数据集上进行对比实验，本文提出的 DNLP-SGC 模型具有更高的 *AUC*, *precision*, *recall*, 因此本文模型能够有效地捕获网络演化过程中链路的变化状态. 未来工作将专注于大型动态网络和不同类型动态网络的链路预测，进一步分析网络中节点的局部和全局信息.

作者贡献声明：刘琳岚指导和修改论文；冯振兴负责模型设计与实现、实验设计与分析以及论文撰写；舒坚指导和撰写论文.

参 考 文 献

[1] Watts D J, Dodds P S, Newman M E J. Identity and search in social networks[J]. Science, 2002, 296(5571): 1302−1305

[2] Ma Xiaoke, Tan Shiyin, Xie Xianghua, et al. Joint multi-label learning and feature extraction for temporal link prediction[J]. Pattern Recognition, 2022, 121: 108216

[3] Song Chao, Lin Youfang, Guo Shengnan, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting [C] //Proc of the 34th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020:

914−921

[4] Güne İ, Gündüz-Öğüdücü Ş, Çataltepe Z. Link prediction using time series of neighborhood-based node similarity scores[J]. Data Mining and Knowledge Discovery, 2016, 30(1): 147−180

[5] Chuan P M, Son L H, Ali M, et al. Link prediction in co-authorship networks based on hybrid content similarity metric[J]. Applied Intelligence, 2018, 48(8): 2470−2486

[6] Ran Yijun, Liu Siyuan, Yu Xiaoyao, et al. Predicting future links with new nodes in temporal academic networks[J]. Journal of Physics: Complexity, 2022, 3(1): 015006

[7] Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations [C] //Proc of the 20th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2014: 701−710

[8] Grover A, Leskovec J. node2vec: Scalable feature learning for networks [C] //Proc of the 22nd ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2016: 855−864

[9] Nguyen G H, Lee J B, Rossi R A, et al. Continuous-time dynamic network embeddings [C] // Proc of the 18th Web Conf (WWW'18). New York: ACM, 2018: 969−976

[10] Chen Zhang, Yiming Fan, Yu Xie, et al. Dynamic network embedding via structural attention[J]. Expert Systems with Applications, 2021, 176: 114895

[11] Pareja A, Domeniconi G, Chen Jie, et al. EvolveGCN: Evolving graph convolutional networks for dynamic graphs [C] //Proc of the 34th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020: 5363−5370

[12] Lei Kai, Qin Meng, Bai Bo, et al. GCN-GAN: A non-linear temporal link prediction model for weighted dynamic networks [C] // Proc of the 38th IEEE Conf on Computer Communications. Piscataway, NJ: IEEE, 2019: 388−396

[13] Chen Jinyin, Zhang Jian, Xu Xuanheng, et al. E-LSTM-D: A deep learning framework for dynamic network link prediction [J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(6): 3699−3712

[14] Liu Linfeng, Yu Zixing, Zhu He. A link prediction method based on gated recurrent units for mobile social network [J/OL]. Journal of Computer Research and Development, 2023, 60(3): 705−716 (in Chinese) (刘林峰, 于子兴, 祝贺. 基于门控循环单元的移动社会网络链路预测方法 [J]. 计算机研究与发展, 2023, 60(3): 705−716)

[15] Chen Jinyin, Wang Xueke, Xu Xuanheng. GC-LSTM: Graph convolution embedded LSTM for dynamic network link prediction[J]. Applied Intelligence, 2022, 52(7): 7513−7528

[16] Randall M J. Physiological time-series analysis using approximate entropy and sample entropy[J]. American Journal of Physiology Heart & Circulatory Physiology, 2000, 278(6): H2039

[17] Bai Shaojie, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. arXiv preprint, arXiv: 1803. 01271, 2018

[18] Veit A, Wilber M J, Belongie S. Residual networks behave like ensembles of relatively shallow networks [C] // Proc of the 30th Int Conf on Neural Information Processing Systems (NIPS'16). La Jolla, CA: NIPS, 2016: 550–558

[19] Balduzzi D, Frean M, Leary L, et al. The shattered gradients problem: If resnets are the answer, then what is the question? [C] // Proc of the 34th Int Conf on Machine Learning. New York: PMLR, 2017: 342−350

[20] Zhang Jingzhao, He Tianxing, Sra S, et al. Why gradient clipping accelerates training: A theoretical justification for adaptivity[J]. arXiv preprint, arXiv: 1905. 11881, 2019

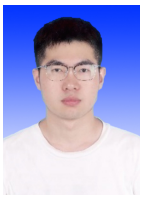
[21] KotzA D, Henderson D, Abyzov I, et al. CRAWDAD: Community resource for archiving wireless data [DB/OL]. Hanover, New Hampshire: Dartmouth College Libraries, [2022-10-10]. http://www.crawdad.org

[22] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C] // Proc of the 31st Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2017: 5998−6008



Liu Linlan, born in 1968. Bachelor, professor. Member of CCF. Her main research interests include software engineering and distributed system.

刘琳岚, 1968 年生. 学士, 教授. CCF 会员. 主要研究方向为软件工程、分布式系统.



Feng Zhenxing, born in 1998. Master candidate. Student member of CCF. His main research interests include network analysis and link prediction.

冯振兴, 1998 年生. 硕士研究生. CCF 学生会员. 主要研究方向为网络分析、链路预测.



Shu Jian, born in 1964. Master, professor. Senior member of CCF. His main research interests include complex networks, embedded system, and software engineering.

舒　坚, 1964 年生. 硕士, 教授. CCF 高级会员. 主要研究方向为复杂网络、嵌入式系统、软件工程.