

# GC-LSTM: Graph Convolution Embedded LSTM for Dynamic Network Link Prediction

Jinyin Chen · Xueke Wang · Xuanheng Xu

Received: date / Accepted: date

**Abstract** Dynamic network link prediction is becoming a hot topic in network science, due to its wide applications in biology, sociology, economy and industry. However, it is a challenge since network structure evolves with time, making long-term prediction of adding/deleting links especially difficult. Inspired by the great success of deep learning frameworks, especially the convolution neural network (CNN) and long short-term memory (LSTM) network, we propose a novel end-to-end model with a Graph Convolution Network(GCN) embedded LSTM, named GC-LSTM, for dynamic network link prediction. Thereinto, LSTM is adopted as the main framework to learn the temporal features of all snapshots of a dynamic network. While for each snapshot, GCN is applied to capture the local structural properties of nodes as well as the relationship between them. One benefit is that our GC-LSTM can predict both added and removed links, making it more practical in reality, while most existing dynamic link prediction methods can only handle removed links. Extensive experiments demonstrated that GC-LSTM achieves outstanding performance and outperforms existing state-of-the-art methods.

**Keywords** Dynamic network link prediction · Graph convolution network · Long short-term memory network · Network embedding · Deep learning

---

J. Chen

Institute of Cyberspace Security and the College of Information Engineering, Zhejiang University of Technology, Hangzhou, China.

Tel.: +(86) 13666611145

E-mail: chenjinyin@zjut.edu.cn

X. Wang

the College of Information Engineering, Zhejiang University of Technology, Hangzhou, China.

Tel.: +(86) 17660478061

E-mail: 3049128970@qq.com

X. Xu

the College of Information Engineering, Zhejiang University of Technology, Hangzhou, China.

Tel.: +(86) 15757171889

E-mail: 1225910594@qq.com

## 1 Introduction

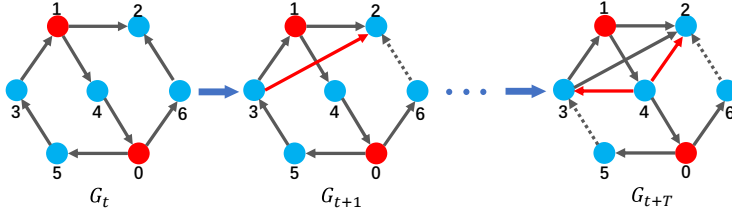
Network analysis has emerged as a powerful tool in various fields [1, 2, 3, 4], such as social networks [5, 6] and economy [7]. Most existing algorithms focus on the analysis of static networks [8, 9], but ignore the dynamics of real world systems. Dynamic networks [10] can make up for the weakness and attract researchers' attention. In particular, dynamic network link prediction (DNLP) has become a hot topic in recent studies. DNLP aims to predict future linkage status of the given network based on historical information. For instance, we can predict people's future relationships in social networks [5, 11] according to their historical behaviors, friends and even their personal attributes. Such examples could also be found in scientific collaboration networks [12], social security networks [13], disease transmission networks [14], etc.

Given a dynamic network, it is often modeled as a chronological sequence of graphs with fixed node and changed links. This characterizes the dynamics of networks while static network models could not. Fig. 1 shows an example of a dynamic network. The relationships between entities could last during the whole observed period (the black line), or occur at a certain moment (the red line), or disappear at some time (the dotted black line). Even if the different entities (the red nodes) have similar network structures at the beginning, they may evolve in different ways. DNLP tries to learn the evolution patterns of each node from historical information and then predicts the structure of next intermediate graph.

Static network link prediction approaches, including similarity indexes and network embedding

techniques, only make prediction based on a specific snapshot. Though sometimes it is effective, they could not capture the evolution patterns due to the ignorance of temporal information. Although graph convolution network (GCN) [15], graph attention network (GAT) [16] and other GNNs have shown their abilities of handling high dimensional graph-structured data, they still suffer from the same problem and are limited to static network analysis. A bunch of dynamic network analysis methods, especially DNLP algorithms, have been proposed for better learning the temporal features. Some are developed by extending time series processing to similarity indexes and network embedding approaches. And many are based on recurrent neural networks (RNNs) like long short-term memory network (LSTM). In dynamic networks, the emerging/disappeared links are more worthy of attention than those existing during the whole period.

Motivated by the excellent performance of LSTM [17] in handling time series data and GCN's feature learning capacity, we propose an unified model capable of handling the spatio-temporal data. Similar idea was once proposed by Manessi et al. [18]. They also jointly made use of GCN and LSTM in one model in a simple way of stacking GCN layer and LSTM cell sequentially, while we embed GCN into LSTM cells to better integrate structural information. However, their proposed method can not realize dynamic link prediction. In particular, we propose a novel end-to-end dynamic network link prediction deep model, called GC-LSTM, which could handle links that are going to appear or disappear. The main idea of our model is to make full use of GCN to learn network structure in the hidden state and cell



**Fig. 1** An example of dynamic networks.

state, and learn the temporal feature through LSTM model. GC-LSTM can effectively handle high-dimensional, time-dependent, and sparse structural sequence data. We conduct extensive experiments on real-world data sets. The results show that our model is significantly better than the current state-of-the-art methods. The main contributions of this paper are as follows.

1. For dynamic network link prediction, we propose a novel end-to-end deep learning model, named GC-LSTM, which extracts the structural feature of each snapshot network through graph convolution, and learns temporal structure through LSTM. The model can effectively learn spatio-temporal features to make precise predictions on dynamic networks.
2. Most existing methods can only predict the added links in the network, while our method can predict all links that are going to appear, disappear, or constant to obtain accurate predictions of the whole dynamic network evolution.
3. We conduct extensive experiments to verify the effectiveness of GC-LSTM on dynamic networks, and compare with different baseline methods. It is show that our model significantly outperforms the current state-of-the-art methods on various metrics.

The rest of the paper is organized as follows. In Section 2, we briefly review recent literatures related to dynamic network link prediction. Then, we give a detailed interpretation of GC-LSTM in Section 3. In Section 4, we show all the experimental results, as well as detailed analysis. Finally, in Section 5, we summarize the main content of the paper and conclude with future works.

## 2 Related Work

Dynamic networks are often modeled as a sequence of graphs to simplify the problem. Apparently, different snapshots play different roles in the prediction. Similarity based dynamic network link prediction is a typical one. In recent years, more and more literatures are focused on the link prediction of dynamic networks. They can be classified into similarity-based methods [19, 20, 21, 22, 23, 24], machine/deep learning-based methods [25, 26, 27, 28, 29], and so on [30, 31, 32].

**Similarity based methods.** The larger similarity, the more likely two nodes will be connected. A number of dynamic network [33] link prediction technologies [13] utilize the topology information of the network to define the similarity between pairwise nodes, named structural similarity indexes, including local indexes and global indexes. The local similarity

indexes only need the neighborhood information of nodes, e.g., Common Neighbors (CN) [34]. It simply uses the number of common neighbors of two nodes as the index but **it is** quite effective. Other common local similarity indexes **include** Jaccard (JA), Adamic-Adar (AA), Resource allocation (RA), Hub promoted index, Hub-contracted index, and Salton index etc [13]. While the global similarity index [13] makes full use of the global topology information of the network, such as Katz, Random Walk, SimRank, Leicht Holme Newman Index (LHNI) and Matrix Forest Index (MFI) etc.

Further, Chen et al. [35] proposed a supervised network link prediction method for dynamic network **by using an ensemble result** of classifiers trained for each property. However, the optimization method is computationally expensive and is limited by the **existing** similarity index. Additionally, Wang et al. [36] assumed that the ability of each node to attract links depends not only on its structural importance, but also on its current popularity. **Under** this assumption, they proposed an approach based on SPM and took the node popularity into consideration. Similarity-based methods often make use of the common characteristics of network topology and have good generalization. However, they are limited by the time-varying features.

**Machine/deep learning based methods.** Besides the similarity based prediction methods, machine/deep learning is also applied to calculate the optimal similarity [37,35,38,39] for precise network link prediction. Thereinto, Tomas et al. [37] proposed a semi-supervised learning algorithm based on spectral convolution, which can encode the graph structure data and features effectively and has a

good performance in the field of link prediction. Additionally, Ahmed et al. [40] believed that the most recent snapshots are of great importance and thus added a damping coefficient to each snapshot in the order of time. And they further proposed Time Series Random Walk (TS-RW) to capture the global information [41]. Other works may learn network features in deep learning frameworks. So, Chen et al. [38] proposed a batched **training method, named FastGCN, which defines** convolution as a form of integral transformation to solve a common memory bottleneck in GCN. They reconstructed the sampling program of loss and gradient, which is proven by the integration transformation of the embedded function. Numerous experiments have confirmed that FastGCN's prediction speed is one order of magnitude faster than GNN, and it maintains a high prediction performance.

Furthermore, Michael et al. [39] proposed a new link prediction **method**, named relational-GCN (R-GCN). Regarding the standard link prediction, a good competition effect is obtained and the factorization model is enriched by R-GCN. Following, Yang et al. [42] **proposed a novel link prediction model, named NetworkGAN, which tackles the** challenging temporal link prediction task [43,44] efficiently. **This method models the spatial and temporal features simultaneously** in the dynamic networks via deep learning techniques. Similarly, Rahman et al. [21] proposed DyLink2Vec, which learns metric embedding for node pairs to predict link states. Zhou et al. [25] proposed DynamicTriad, which models the network evolution as triadic closure process. Rather than turning a dynamic network into a discrete sequence. Nguyen [45] et al. proposed

temporal random walk on the original network, which is considered as continuous-time network embedding. Further, Lei et al. [46] proposed a nonlinear model GCN+GAN to solve link prediction tasks in weighted dynamic networks.

Most [existing](#) dynamic link prediction methods focus on a pair of nodes rather than the whole snapshot when predicting and thus may not be able to integrate global information [automatically](#). By comparison, our GC-LSTM method can extract both structural feature of each snapshot network and the temporal feature to make precise prediction on dynamic networks.

### 3 Methodology

In this section, we will introduce the GC-LSTM model for dynamic network link prediction. It's capable of learning both structural and temporal feature of dynamic networks for [the future links addition and deletion](#).

#### 3.1 Problem Definition

We will give necessary definitions, [which will be used](#) throughout this paper. Formally, we define dynamic networks as follows.

##### **Definition 1 (Dynamic Networks)**

Dynamic networks can be represented as a sequence of discrete snapshots,  $\{G_1, \dots, G_T\}$ , where  $G_t = (V, E_t, A_t)$  ( $t \in [1, T]$ ) represents a network at time  $t$ . Let  $V$  be the set of all nodes and  $E_t$  be the temporal links within the fixed timespan  $[t - \tau, t]$ .  $A_t$  denotes the adjacency matrix of  $G_t$ , where the element  $A_t(i, j) = 1$  if there is a link between  $i$  and  $j$ , and  $A_t(i, j) = 0$  otherwise.

In a static network, the network link prediction problem generally aims to predict the unobserved or future links by the current network. Static link prediction methods mainly focus on the structural feature of the network. Different from static network link prediction, dynamic network link prediction also needs to learn the temporal feature of the network according to the dynamic evolution processes of previous snapshots. Our goal is to extract the structural feature of each snapshot network through graph convolution and learn temporal structure [via LSTM](#).

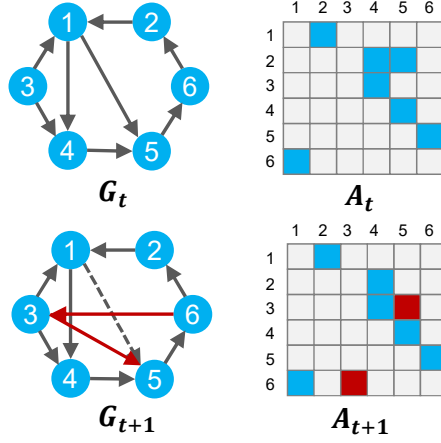
##### **Definition 2 (Network Link Prediction in Dynamic Networks)**

Given a sequence of graphs with length  $T$ ,  $\{G_{t-T}, \dots, G_{t-1}\}$ . Dynamic network link prediction is considered as a structural sequence modeling problem. It aims to learn the evolution information of the previous  $T$  snapshots to predict the probability of all links at time  $t$ . The aim is to realize the next line edge's prediction and keep the entity unchanged. defined as

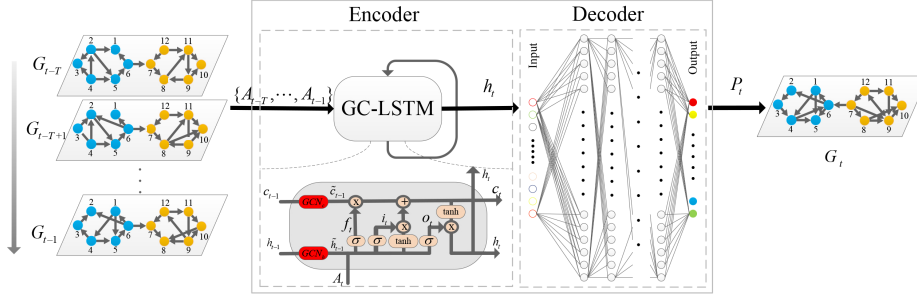
$$\hat{A}_t = \operatorname{argmax} P(A_t | A_{t-T}, \dots, A_{t-1}) \quad (1)$$

where  $\{A_{t-T}, \dots, A_{t-1}\}$  represents the adjacency matrices of previous  $T$  snapshots,  $A_t$  and  $\hat{A}_t$  are the real and predicted adjacency matrices of the snapshot at time  $t$ , respectively.

Since the adjacency matrix  $A_t$  is a precise reflection of how a network changes over time, which is adopted as the input of our prediction model, as shown in Fig. 2. We utilize a sequence of data  $\{A_{t-T}, \dots, A_{t-1}\}$  with length  $T$  to predict the adjacency matrix  $A_t$ , where  $P(A_t | A_{t-T}, \dots, A_{t-1})$  models the probability of links to appear conditioned on the past  $T$  adjacency matrices.



**Fig. 2** An illustration of dynamic network evolution and its adjacency matrix. The network changes from time  $t$  to  $t+1$ .  $E(6;3)$  and  $E(3;5)$  emerge while  $E(2;5)$  vanishes, the adjacency matrix also changes from  $A_t$  to  $A_{t+1}$ , with those elements equal to 1 represented by filled squares.



**Fig. 3** The overall framework of encoder and decoder model for end-to-end dynamic network link prediction, in which encoder model is GC-LSTM, while decoder model is a fully connected layer network. Given a sequence of networks with length  $T$ ,  $\{G_{t-T}, \dots, G_{t-1}\}$ , each network is transformed into an adjacency matrix  $A_t$  as the input. The GC-LSTM is two GCNs embedded into the LSTM to learn spatio-temporal feature from the extracted data.  $GCN_h$  and  $GCN_c$  denote the graph convolution on long-term information  $c$  and  $h$ , where  $c_0$  and  $h_0$  are zero matrixes. The decoder projects the received feature  $h_t$  maps back to the original space to get  $G_t$ . Here,  $\sigma$  in GC-LSTM is an activation function and sigmoid is adopted in this paper.

### 3.2 Overall Framework

The proposed GC-LSTM consists of an encoder and a decoder shown in Fig. 3. Encoder model is GCN embedded LSTM, using GCN to learn network structure of the cell state  $c$  and the hidden state  $h$  of each moment snapshot, while using LSTM to learn the temporal information of

the state of each link. Decoder is a fully connected layer network to convert the extracted features mapping back to the original space. GC-LSTM will output the predicted network and implement network link prediction in a unified fashion.

Before the detailed description of the proposed model, we introduce some terms and notations that will be

**Table 1** Terms and notations used in GC-LSTM.

Symbol	Definition
$N$	number of nodes at each snapshot
$A_t$	the adjacency matrix at time $t$ as the input data
$P_t$	the output probability of each link at time $t$
$d$	number of hidden layer unit in GC-LSTM
$W_{f,i,o,c}, b_{f,i,o,c}$	weight and bias of three gates in GC-LSTM
$K$	the order of graph convolution
$\sum_{k=0}^K \theta_k$	weight of graph convolution
$\tilde{L}_t$	the rescaled graph Laplacian matrix at time $t$
$D_t$	the degree matrix of $A_t$ at time $t$
$I_N$	the identity matrix
$\lambda_{max}$	the largest eigenvalue of graph Laplacian matrix
$h_t, c_t$	hidden and cell state of GC-LSTM
$\tilde{h}_t, \tilde{c}_t$	hidden and cell state extracted by GCN
$W_d^{(l)}, b_d^{(l)}$	weight and bias of $l^{th}$ layer of decoder

used in this paper, listed in TABLE 1. Notice that, in LSTM, subscript  $f$  represents the forget gate,  $i$  and  $c$  represent the input gates and  $o$  represents the output gate.

### 3.3 GC-LSTM Model

We use the proposed GC-LSTM model as an encoder to extract the corresponding temporal and structural information from the structure sequence data  $\{A_{t-T}, \dots, A_{t-1}\}$ , and the hidden layer vector  $h_t$  at the last moment will be used as the output of GC-LSTM.

In the task of dynamic network link prediction, the linkage status (exist or non-exist) of each node with others at multiple times can be regarded as a time series, equivalent to the row vector in the adjacency matrix. Generally, LSTM is applied to capture the temporal characteristics of time series data. So in GC-LSTM, we utilize the LSTM to solve long-term dependency problems and effectively learn temporal features of the dynamic graphs. The link state of each node in the dynamic network may use LSTM to implement temporal

prediction, such as predicting links at the next moment. On the other hand, it is necessary to utilize the previous link state information of the nodes, and consider the impact of the link state of the neighbors, as well as the network structure characteristics. GCN has been proved efficient in network embedding for learning structural feature. We propose GC-LSTM model, where the Graph Convolution (GC) models are adopted to extract the structural characteristics of the snapshots at each moment, and LSTM is capable of learning temporal feature of dynamic network.

GCN is originally designed for undirected networks which means the Laplacian matrix of the symmetric network. To embed GCN into the proposed GC-LSTM, we first need to redefine the Laplacian matrix. Following Ma et al's idea, the directed Laplacian is defined as

$$L^{sym} = I - \frac{1}{2}(\Phi^{1/2}P\Phi^{-1/2} + \Phi^{-1/2}P^T\Phi^{1/2}) \quad (2)$$

where  $P = D_{out}^{-1}A$  refers to the transition probability matrix. And  $\Phi = \text{diag}(\phi_{norm}(v))$  where  $\phi_{norm}$  is the Perron vector of  $P$  and  $v$  refers to the node in  $G$ . After obtaining the directed



Laplacian, the graph convolution on directed networks is then defined as

$$g_\theta = \sum_{k=0}^K \theta_k T_k(\tilde{L}) \quad (3)$$

where  $\theta$  denotes the parameters of  $g_\theta$  and  $T_k$  is Chebyshev polynomial which is defined as  $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$  with  $T_0(x) = 1$  and  $T_1(x) = x$ . The filter  $g_\theta$  then becomes applicable to the analysis of directed networks. In the rest of the paper,  $K^{th}$  order of graph convolution is denoted as  $GCN^K(A, X)$  for simplicity.

The GC-LSTM model mainly relies on two state values, the hidden state  $h$  which is used to extract the input information at the last time, and the cell state  $c$  which is used to save the long-term information. The essential of GC-LSTM is that it has a cell state  $c$  during the forward process, resulting in information transferred over the cell state  $c$  for a long time. In the dynamic network link prediction task, we need to consider the influence of the hidden state of the neighbors on the hidden state of the node, and the influence of the cell state of the neighbors. Since the cell state  $c$  and the hidden state  $h$  respectively reflect different information. Therefore, we propose to use two GCN models to perform convolution operations on the cell layer state and the hidden layer state at the last time.

In GC-LSTM model, the first step is to decide what information will be thrown away from the previous cell state. This decision is performed by a forget gate  $f_t \in [0, 1]^d$ , where 0 means all information will be forgotten, 1 means all information will be reserved, defined as

$$f_t = \sigma(A_t W_f + GCN_f^K(\tilde{A}_{t-1}, h_{t-1}) + b_f) \quad (4)$$

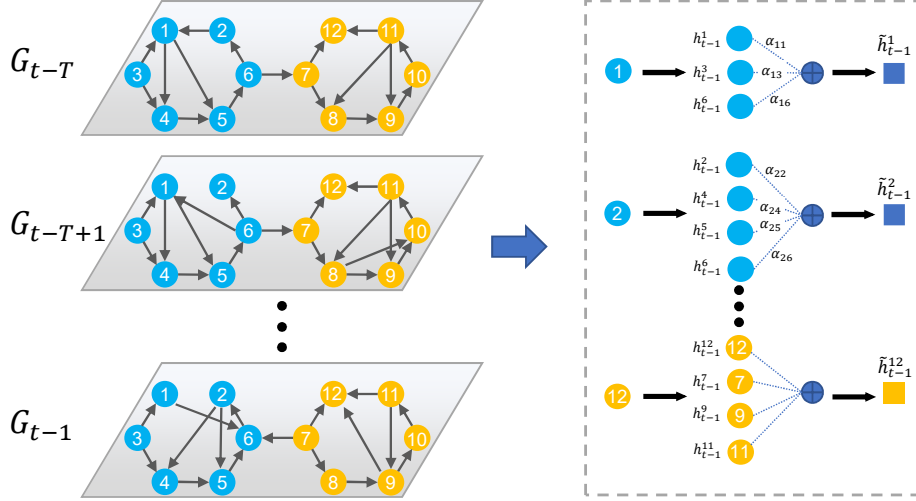
where  $A_t \in R^{N \times N}$  denotes the input of the GC-LSTM at time  $t$ ,  $h_{t-1} \in R^{N \times d}$  denotes the hidden state of the GC-LSTM at time  $t-1$ .  $W_f \in R^{N \times d}$  and  $b_f \in R^d$  are the weight and bias matrices of forget gate, respectively.  $\sum_{k=0}^K \theta_{hkf} T_k(\tilde{L}_{t-1}) h_{t-1}$  represents the graph convolution on the hidden state  $h_{t-1}$  at time  $t-1$ .  $\tilde{L}_{t-1} = \frac{2}{\lambda_{max}} L_t - I_N$  is a rescaled graph Laplacian, where  $L_t = I_N - D_t^{-\frac{1}{2}} A_t D_t^{-\frac{1}{2}}$  is the normalized graph Laplacian.  $A_t$  is the adjacency matrix.  $D_t$  is the degree matrix of  $A_t$ ,  $I_N$  is the identity matrix,  $\lambda_{max}$  denotes the largest eigenvalue of  $L_t$ .  $\sum_{k=0}^K \theta_{hkf}$  are the parameter vector, and  $K$  is the order of the graph convolution.  $\sigma(\cdot)$  denotes the sigmoid function,  $N$  denotes the input dimension,  $d$  denotes the hidden layer dimension of GC-LSTM. When  $K=1$ , the GCN model can use the information of the 1<sup>st</sup> order neighbors, so the operation of the 1<sup>st</sup> GCN on the hidden state  $h$  is as shown in the Fig. 4. In addition, the operation of the GCN on the hidden state  $c$  is the same as  $h$ , only that the input of GCN has changed from  $h$  to  $c$ .

The next step is to update the cell state. First, a  $\tanh$  layer generates a new candidate vector of the cell layers,  $\bar{c}_t \in [-1, 1]^d$ . Then,  $\bar{c}_t$  is a sigmoid layer which determines how many new candidate vector will be added to the cell state, as the input gate  $i_t \in [0, 1]^d$ . Finally, the cell state can be updated by the forget gate and the input gate, as follows.

$$\begin{aligned} \bar{c}_t &= \tanh(A_t W_c + GCN_c^K(\tilde{A}_{t-1}, h_{t-1}) + b_c), \\ i_t &= \sigma(A_t W_i + GCN_c^K(\tilde{A}_{t-1}, h_{t-1}) + b_i), \\ c_t &= f_t \odot GCN_c^K c_{t-1} + i_t \cdot \bar{c}_t. \end{aligned} \quad (5)$$

where  $W_{i,c} \in R^{N \times d}$  and  $b_{i,c} \in R^d$  are the weight and bias matrices of the input gate, respectively.  $\theta_{hkc}$ ,  $\theta_{hki}$  and  $\theta_{ck}$





**Fig. 4** The framework of 1<sup>st</sup> order GCN hidden state  $h$ .

are the parameter vectors of the graph convolution on  $h$  and  $c$ .

Therefore, the updated cell layer information can not only save long-term information, but also selectively filter out some useless information. Finally, we need to decide what information will be output, it is implemented by the output gate:

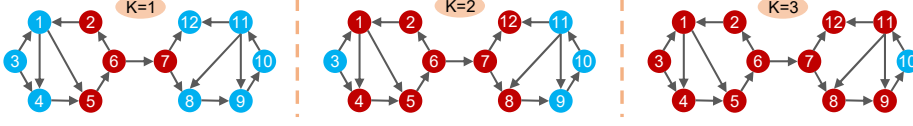
$$\begin{aligned} o_t &= \sigma(A_t W_o + GCN_o^K(\tilde{A}_{t-1}, h_{t-1}) + b_o), \\ h_t &= o_t \odot \tanh(c_t). \end{aligned} \quad (6)$$

where  $W_o \in R^{N \times d}$  and  $b_o \in R^d$  are the weight and bias matrices of the output gate, respectively.  $\theta_{hko}$  is the parameter vector of the graph convolution on  $h$ .

Above all, the three gates constitute the entire forward process of the GC-LSTM. In this paper, the adjacency matrices  $\{A_{t-T}, \dots, A_{t-1}\}$  is used as the input data, so the hidden layer vector  $h_t \in R^{N \times d}$  obtained by GC-LSTM already contains both temporal and structural information of snapshot at the previous  $T$  moments. It should be always noticed that the numbers of hidden units  $d$  in GC-LSTM varies when the number of nodes  $N$  is different for each dynamic

network. The larger  $N$  is the more units we need in the model. Generally, the number of hidden units  $d$  is selected from  $\{64, 128, 256, 512\}$ .

It is worth noticing that the GCN operation is approximated by a  $K^{th}$  order Chebyshev polynomial[37] and the hyper-parameter  $K$  affects the performance of GC-LSTM.  $K$  is associated with the richness of structural information. That means GC-LSTM can utilize the information with the  $K$ -hop neighborhood of the central node. As shown in Fig. 5, the information of  $v_2$ ,  $v_5$  and  $v_7$  together with  $v_6$  itself is considered with  $K=1$ . And additional information,  $v_1$ ,  $v_4$ ,  $v_8$  and  $v_{12}$ , is provided when  $K=2$ . The larger  $K$  is, the larger the neighborhood will be considered. When  $K$  gets larger, the performance of GC-LSTM will be improved and it will surely increase the computational cost. To balance the performance and efficiency, we set  $K$  to 3 in this paper.



**Fig. 5** An illustration of  $K^{th}$  order ( $K=1, 2, 3$ ) GCN, which uses  $K^{th}$  order Chebyshev polynomial to approximate the graph convolution.

### 3.4 Decoder Model

In order to output the final prediction network, we use a fully connected layer network as a decoder to turn the hidden vector  $h_t$  at last moment in the sequence data into the final probability matrix.

$$\begin{aligned} Y_d^{(1)} &= ReLU(W_d^{(1)}h_t + b_d^{(1)}), \\ Y_d^{(k)} &= ReLU(W_d^{(k)}Y_d^{(k-1)} + b_d^{(k)}), \\ P_t &= o_t \odot \tanh(c_t). \end{aligned} \quad (7)$$

where  $W_d \in R^{d \times N}$  and  $b_d \in R^N$  are the weight and bias matrices of the decoder, respectively.  $L$  denotes the number of hidden layers in the fully connected layer, and the number of units in each hidden layer may vary according to the dimension of data for better performance.  $P_t \in R^{N \times N}$  represents the final probability matrix of links, satisfying  $P_t(i, j) \in [0, 1]$  represents the probability of the link from  $i$  to  $j$ . The larger  $P_t(i, j)$  is, the higher probability that the two nodes  $i$  and  $j$  are connected.

### 3.5 Loss Function and Model Training

The purpose of training the entire GC-LSTM model is to improve the accuracy of the dynamic network link prediction, so the training is designed to make the output probability matrix more similar to the adjacency matrix at time  $t$ . In the regression prediction problem, the  $L_2$  distance is usually used to reflect the degree of similarity between the predicted value and the

true value.

$$L_2 = \|P_t - A_t\|_F^2 = \sum_{i=0}^N \sum_{j=0}^N (P_t(i, j) - A_t(i, j))^2 \quad (8)$$

where  $P_t$  represents the output probability matrix at time  $t$ ,  $A_t$  represents the real adjacency matrix at time  $t$ ,  $N$  represents the number of nodes in the snapshot at time  $t$ . However, if we only use the  $L_2$  distance as the loss function, it will lead the predicted results biased towards zero [due to the sparsity of the network](#). To avoid such overfitting to certain extent, we also employ a regularization term  $L_{reg}$ .

The total loss in the training process thus is defined as

$$L = L_2 + \beta L_{reg} \quad (9)$$

where  $\beta$  is a parameter to trade-off the importance of  $L_2$  and  $L_{reg}$ , and the best  $\beta$  will be found in the model training process. The regularization loss  $L_{reg}$  is calculated by calculating the sum of squares of [the ownership weight](#) in GC-LSTM model. In order to minimize Eq. (9), we adopt the Adam as the optimizer of our model.

## 4 Experiments

The proposed GC-LSTM model is compared [with the widely used baseline methods](#) on several real benchmark dynamic networks to testify its performance. All experiments are implemented on 12G

NVIDIA GPU, 32G DDR running memory server, which are coded by Python.

#### 4.1 Setup

**Parameter Settings:** The GC-LSTM model proposed in this paper includes an encoder module composed of GCN and LSTM network. Each data set is divided according to fixed time, and 320 network sequences are taken. Among which the first 240 networks are the training set and the last 80 moments are the test set. The number of LSTM networks is 1. The number of GCN networks is 4 (corresponding to four gate units in LSTM respectively, independent with each other). Hyperparameter  $K$  of Chebyshev polynomial in each GCN is 3. LSTM hidden layer dimensions is 256 for the first five data sets and 512 for the last one. GCN network hidden layer dimension is 256 for the first five data sets and 512 for the last one. The iterations are 200. Eleven continuous network sequences in the training set were used as inputs to the model.

**Baseline Methods:** To validate the effectiveness of GC-LSTM, we compare it with node2vec [47], TNE [48], ctRBM [49], GTRBM [50], and DDNE [51], listed as follows.

- **node2vec** [47]: node2vec develops a  $2^{nd}$  order biased random walk procedure to explore neighborhood of a node, which can strike a balance between local properties and global properties of a network. In [45], node2vec is adopted as a baseline for dynamic link prediction.
- **TNE** [48]: TNE models network evolution as a Markov process and then uses the matrix factorization to learn the embedding vectors for all nodes.

- **ctRBM** [49]: ctRBM is a generative model based on temporal RBM. It first generates a vector for each node based on temporal connections and predict future linkages by integrating neighbor information.
- **GTRBM** [50]: GTRBM takes the advantages of both tRBM and GBDT to effectively learn the hidden dynamic patterns.
- **DDNE** [51]: Similar to autoencoder, DDNE uses a GRU as an encoder to read historical information and decodes the concatenated embeddings of previous snapshot into future network structure.

As the same setting in [45] for node2vec, we set the dimension of the embedded vector to 80, and find the optimal values of the hyper-parameters  $p$  and  $q$  from the scope  $\{0.5, 1, 1.5, 2\}$ . Weighted- $L_2$  [47] is used to obtain an embedding vector  $e^{u,v}$  for the link from node  $u$  to  $v$ . Each element in the vector is defined as

$$e_i^{u,v} = |u_i - v_i|^2 \quad (10)$$

where  $u_i$  and  $v_i$  are  $i^{th}$  elements of embedding vectors of nodes  $u$  and  $v$ , respectively. We compute the vector for a link by combining the learned representations of the corresponding nodes, and put the obtained representation of links into a logistic regression classifier to classify links. For TNE, we also set the dimension to 80 and regard the inner products of embedding matrix and its transpos. When we calculate the embedded vector matrix  $Z$  for each node, then  $Z \cdot Z^T$  is used to calculate the probability of all links and the threshold  $\tau=0.9$  is used to judge whether the link is existing or not. For ctRBM and GTRBM, we set the number of visible units equal to the

number of nodes in the corresponding network, and the dimension of hidden layers as 128 for the first four smaller datasets and 256 for the rest. For DDNE, we set the dimension as 128 for the first four smaller datasets and 512 for the rest. For our proposed model, GC-LSTM, we set the hidden unit  $d = 256$ , and the order  $K=3$ . In addition, we also analyze the parameter sensitivity of GC-LSTM model.

#### 4.2 Evaluation Metrics

Usually, the metrics used to evaluate static network link prediction are also applied to dynamic ones, such as Area Under the ROC Curve (AUC). Based on AUC, the Area Under the Precision-Recall Curve (PRAUC) [52] is designed to evaluate the sparsity of networks, and focus on the added links. Since there are also links removed in dynamic networks, Junuthula et al. [53] restricted the measurements to only part of node pairs and proposed the Geometric Mean of AUC and PRAUC (GMAUC) for the added and removed links. Li et al. [49] used SumD to calculate the difference between the predicted network and the true one, and evaluate network link prediction methods in a more strict way. But the absolute difference could be misleading. For example, suppose that the SumD of the two dynamic network link prediction models are both equal to 5. If one model mispredict 5 links in a total of 10 links, while the other mispredict 5 links in a total of 100 links. It is reasonable to think the latter performs better than the former, while the SumD metric could not tell this.

In our experiment, at first, we choose AUC, GMAUC, and Error Rate [48] as metrics to evaluate

all dynamic network link prediction results.

- **AUC**: If among  $n$  independent comparisons, there are  $n'$  times that the existing link gets higher probability score than the non-existing link and  $n''$  times the get the same score, then the AUC is defined as

$$\text{AUC} = \frac{n' + 0.5n''}{n}. \quad (11)$$

Before calculating the AUC index, we will select all existing links and randomly sample the same number of links that do not exist to reduce the impact of sparsity.

- **GMAUC**: It is a metric proposed to evaluate dynamic network link prediction, which combines both PRAUC and AUC by taking geometric mean of the two quantities, and the GMAUC is defined as

$$\text{GMAUC} = \left( \frac{\text{PRAUC}_{new} - \frac{L_A}{L_A + L_R}}{1 - \frac{L_A}{L_A + L_R}} \cdot 2(\text{AUC}_{prev} - 0.5) \right)^{1/2} \quad (12)$$

where  $L_A$  and  $L_R$  are the numbers of added and removed links, respectively,  $\text{PRAUC}_{new}$  is the PRAUC value calculated among the new links and  $\text{AUC}_{prev}$  is the AUC for the originally existing links.

- **Error Rate (ER)**: It is defined as the ratio of the number of mis-predicted links  $N_{false}$  to the total number of truly existent links  $N_{true}$ , which is defined as

$$\text{ER} = \frac{N_{false}}{N_{true}} \quad (13)$$

Unlike the SumD which only calculates the absolute link difference in two networks, ER considers relative link difference for precise evaluation. The smaller ER corresponds to the better prediction effect.

In particular, we further divide ER into two parts as an effective supplement, Error Rate+ (ER+) and Error Rate- (ER-). ER+ indicates the ratio of the number of mis-predicted links in truly existent links to the total number of truly existent links. ER- indicates the ratio of the number of mis-predicted links in truly non-existent links to the total number of truly existent links.

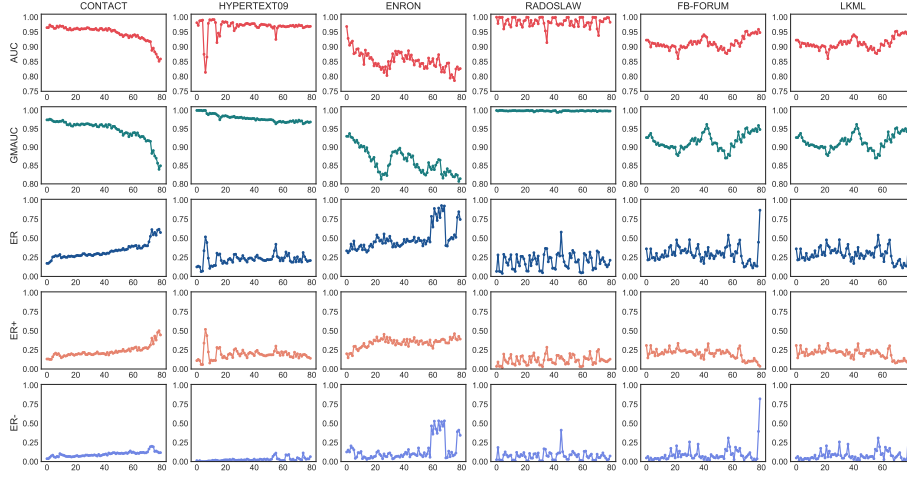
#### 4.3 Datasets

We carry out the experiments on the real-world dynamic networks. Most networks are human relation networks [The nodes](#) representing humans and links capturing their connections. The contacts can be face-to-face, phone, emailing etc. The data sets are described as follows.

- **HYPERTEXT09** [54] and **CONTACT** [55]: They are human contact dynamic networks of face-to-face proximity. The networks are collected through wireless devices carried by people who attended corresponding activity. If a person contacts with another one within a certain timestamp, a corresponding link will occur. Both of them are divided into snapshots at the time interval of 10 minutes.
- **ENRON** and **RADOSLAW** [56]: ENRON and RADOSLAW are email networks. Each node represents an employee of a medium-sized company, and a link occurs each time such as an email sent from one employee to another. ENRON records email data for nearly three years and RADOSLAW lasts nearly nine months. The time intervals for ENRON and RADOSLAW are 3 days and 8 hours, respectively.
- **FB-FORUM** [57]: FB-FORUM is attained from a Facebook-like online forum of students at University of California, Irvine, in 2004. It is an online social network where nodes are users and links represent interactions (e.g., messages) between users. And the records span more than 5 months. We split FB-FORUM at the time interval of 8 hours.
- **LKML**: It is collected from Linux kernel mailing list. In this network, the nodes represent users which are identified by their email address. And each link donates a reply from one user to another. In this paper, we only focus on the 2210 users that were recorded from 2007-01-01 to 2007-04-01, and then construct a dynamic network based on the links between these users that appear from 2007-04-01 to 2013-12-01. The time interval for LKML is 7 days.

TABLE 2 provides a summary of all dynamic networks and their statistics for evaluation.  $V$  represents the number of nodes.  $|E_T|$  shows the number of dynamic links in the network,  $\bar{d}$  represents the average dynamic node degree,  $d_{max}$  represents the maximum dynamic node degree. Timespan (days) is the span from the first network to the last network in sequential networks.

Before training, we generate snapshots for each dataset at fixed intervals, and then sort them in an ascending order of time to generate dynamic network structure sequence data. Considering that the connection between people may be temporary, we remove the links that do not reappear within 8 intervals. In order to get enough samples, we split each dataset into 331 snapshots according to an ascending order of time, then use continuous 11 snapshots as one



**Fig. 6** The performance of GC-LSTM on AUC, GMAUC, ER, ER+ and ER-, obtained by our model, as functions of time  $t$  for the six datasets.

**Table 2** The basic statistics of the dynamic networks.

Dataset	$ V $	$ E_T $	$d$	$ d_{max} $	Timespan (days)
HYPERTEXT09	113	20.8K	368.5	1,483	2.46
ENRON	151	50.5K	669.8	5,177	1,137.55
RADOSLAW	167	82.9K	993.1	9,053	271.19
CONTACT	274	28.2K	206.2	2,092	3.97
FB-FORUM	899	50.5K	669.8	5,177	164.49
LKML	2,210	422.4K	34.6	47,995	2,436.30

sample, where the first 10 snapshots are adopted as input and the last snapshot network is used as output, equal to the sampling period  $T=10$ . Therefore, we can get a total of 320 samples. Then, we group the first 240 samples as a training set and the remaining 80 samples as a test set.

#### 4.4 Dynamic Prediction Results

When training the model, we take 10 consecutive snapshots  $\{G_{t-10}, \dots, G_{t-1}\}$  as a sample, and feed the model with it to predict the next snapshot  $G_t$ . There are two typical strategies to process networks:

1) using only (t-1)-th snapshot  $G_{t-1}$  to predict the next snapshot  $G_t$  [58]; 2) integrating the previous 10 snapshots  $\{G_{t-10}, \dots, G_{t-1}\}$  into one sample to predict the next snapshot  $G_t$  [51]. In this paper, we choose the first strategy for node2vec, while choose the second strategy for the others.

Since the evolution pattern of dynamic network may change over time, we select the average of the three evaluation metrics of the first 20 test samples and all 80 samples to reflect the short-term and long-term prediction performance of the model. All results are presented in TABLE 3, we compared several baseline methods, including node2vec, TNE, ctRBM,

GTRBM, and DDNE. Link prediction performance indicators include AUC, GMAUC, ER, ER+ and ER-. We can find that GC-LSTM outperforms all baseline methods on ER, ER+, and ER-, while node2vec still performs the worst. However, under the data sets FB-FORUM and LKML, the indicators AUC and GMAUC are not optimal under GC-LSTM link prediction method, while TNE has better effect. Next, we analyze each indicator in detail. Where we can see that the GC-LSTM model outperforms all baseline methods for both the short-term and long-term prediction capabilities, while the performance of node2vec is worst in most cases. That is to say, network embedding methods, designed for static networks, are not suitable for dynamic network link prediction, while other dynamic prediction models perform much better due to their ability of capturing dynamic characteristics. Our GC-LSTM model still performs better than other dynamic prediction models on all datasets in most cases, especially on the GMAUC which reflects the predictive performance of both added and removed links in dynamic networks.

In addition, we also compare the difference of links between the predicted network and the ground truth network to calculate ER. The significant difference in the ER indicates that this metric is a good complement to fully measure the performance of dynamic network link prediction. As shown in TABLE 3, from the results of ER+ and ER-, we further find that most of baselines may predict more invalid links than the actually existed links, resulting in a relatively large ER. And node2vec always predict a large number of invalid links, which is 10 times or even 70 times the number of links actually

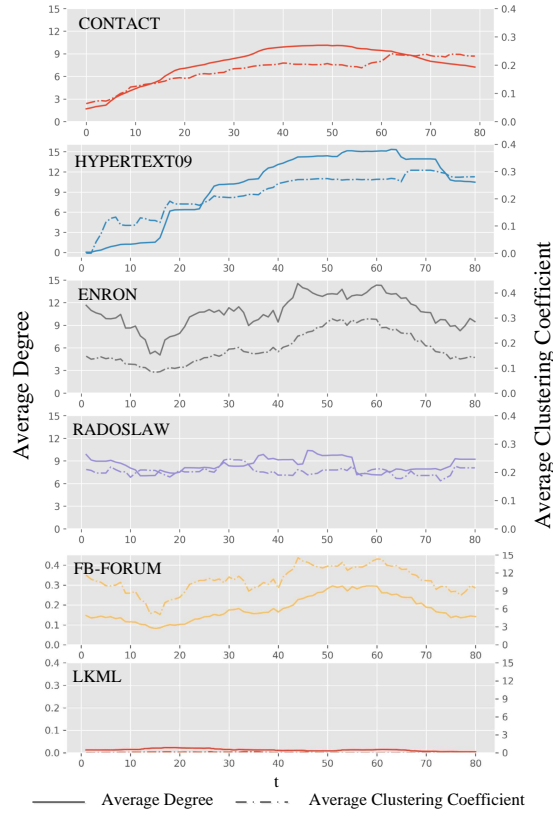
exist. In particular, TNE also has a poor performance on the ER, because the matrix decomposition method used by TNE cannot balance the positive and negative samples on the sparse adjacency matrix.

Actually, the results prove once again that our GC-LSTM model has better performance in dynamic network link prediction. Because the baselines directly predict the next snapshot network based on the first 10 snapshot networks, while our method requires pre-training models to learn the evolution pattern of dynamic networks more effectively from the first 240 samples. And GC-LSTM model not only uses LSTM to learn the temporal characteristics of the sequence network, but also uses GCN to learn the network characteristics of each snapshot. Therefore, in the test process, it can predict future links more accurately with lower ER.

Interestingly, in TABLE 3, the ER- is much larger than ER+ in most cases, especially for node2vec, indicating that most of the methods are more likely to predict non-existent links than existent ones. This phenomenon may be due to the fact that most of the dynamic networks in the experiments are relatively sparse. For our GC-LSTM method, although ER+ is slightly larger than ER-, both of its ER+ and ER- are smaller than that of baselines in most cases. That is, GC-LSTM method not only has relatively high prediction accuracy on the non-existing links, but also has better prediction performance on the existing ones, i.e., it can better predict the dynamic links of the network. In general, the link prediction method proposed in this paper has better prediction effect.

Next, for test samples  $G_t$  ( $t$  changes from 1 to 80), we plot the AUC, GMAUC, ER, ER+, and ER- metric





**Fig. 7** The trends over time of network structure attributes including average degree and average cluster coefficient.

curves over time obtained by our GC-LSTM model on all networks to reflect the dynamic network link prediction performance, as shown in Fig. 6. As time evolves, AUC and GMUUC curves are gradually decreasing, while the ER, ER+, and ER- curves are gradually increasing. This phenomenon indicates that long-term link prediction for dynamic network is relatively difficult due to the uncertainty of dynamic network evolution over a long period of time. Interestingly, since the network structure of RADOSLAW and LKML evolves periodically, the performance of GC-LSTM on RADOSLAW and LKML are relatively stable. In order

to further explain this phenomenon, we studied the trend of the two most common structural properties (i.e., the average degree and the average clustering coefficient) of the dynamic networks with increasing time  $t$ , and the results are shown in Fig. 7. We can see that the average degree and average clustering coefficients of the four datasets, CONTACT, HYPERTEXT09, ENRON and FB-FORUM, have changed significantly over time, while those of RADOSLAW and LKML are relatively stable. These results explain why we can get better long-term predictive performance.

**Table 3** Dynamic network link prediction performances on AUC, GMAUC, ER, ER+ and ER- for the first 20 samples and all the 80 samples.

Metric	Method	CONTACT		HYPERTEXT09		ENRON		RADOSLAW		FB-FORUM		LKML	
		20	80	20	80	20	80	20	80	20	80	20	80
AUC	node2vec	0.5212	0.5126	0.6348	0.6591	0.7659	0.6806	0.6103	0.7676	0.5142	0.5095	0.6348	0.5892
	TNE	0.9443	0.9297	0.9076	0.8517	0.8096	0.8314	0.8841	0.8801	<b>0.9810</b>	<b>0.9749</b>	<b>0.9861</b>	<b>0.9867</b>
	ctRBM	0.9385	0.9109	0.9029	0.8847	0.8468	0.8295	0.8834	0.8590	0.8728	0.8349	0.8091	0.7729
	GTRBM	0.9451	0.9327	0.9204	0.8936	0.8527	0.8491	0.9237	0.9104	0.9023	0.8749	0.8547	0.8329
	DDNE	0.9347	0.9433	0.9282	0.9105	0.7985	0.7638	0.9027	0.8974	0.9238	0.8729	0.9328	0.9115
	GC-LSTM	<b>0.9649</b>	<b>0.9453</b>	<b>0.9573</b>	<b>0.9675</b>	<b>0.8425</b>	<b>0.8135</b>	<b>0.9838</b>	<b>0.9833</b>	0.9021	0.9098	0.9082	0.9000
GMAUC	node2vec	0.1805	0.1398	0.4891	0.5163	0.4069	0.5417	0.7241	0.7203	0.2744	0.2886	0.2309	0.2193
	TNE	0.9083	0.8958	0.8856	0.8392	0.8233	0.7974	0.8282	0.8251	<b>0.9689</b>	<b>0.9629</b>	<b>0.9839</b>	<b>0.9778</b>
	ctRBM	0.9126	0.8893	0.9139	0.8779	0.7207	0.6921	0.8004	0.7998	0.8926	0.8632	0.7723	0.7206
	GTRBM	0.9240	0.9136	0.9407	0.9201	0.9148	0.8675	0.9157	0.8849	0.9329	0.9117	0.6529	0.6038
	DDNE	0.8925	0.8684	0.9484	0.9127	0.8724	0.8476	0.8938	0.8724	0.9126	0.9023	0.7894	0.7809
	GC-LSTM	<b>0.9680</b>	<b>0.9456</b>	<b>0.9921</b>	<b>0.9788</b>	<b>0.8571</b>	<b>0.8134</b>	<b>0.9980</b>	<b>0.9983</b>	0.9062	0.9123	0.9184	0.9049
ER	node2vec	44.7753	25.2278	24.3980	12.8260	23.9053	24.8060	20.7240	21.2489	40.5109	48.5376	53.2895	61.0274
	TNE	13.1410	7.1556	1.6453	1.9675	23.1276	19.9167	16.7078	16.7175	19.1058	24.4350	18.5702	18.2091
	ctRBM	0.9126	0.8893	0.6663	0.6970	2.4890	2.7328	1.8920	2.0937	3.4509	3.6782	2.9903	3.3089
	GTRBM	0.9240	0.9136	0.5745	0.5992	1.5947	1.8836	1.9079	2.0031	2.2347	2.4396	2.5351	2.7942
	DDNE	0.8925	0.8684	0.5820	0.6084	1.7664	1.9014	1.6316	1.5941	1.9014	1.8266	2.0134	2.2258
	GC-LSTM	<b>0.2324</b>	<b>0.3213</b>	<b>0.1988</b>	<b>0.2194</b>	<b>0.3763</b>	<b>0.3928</b>	<b>0.1783</b>	<b>0.1891</b>	<b>0.2702</b>	<b>0.3012</b>	<b>0.7369</b>	<b>0.7798</b>
ER+	node2vec	5.0931	1.3049	3.0721	2.0933	4.9831	7.1045	2.8204	3.1249	7.8192	9.1033	6.3719	13.0938
	TNE	3.2049	3.2307	0.4201	0.2049	3.2924	3.9837	3.2145	3.6537	7.0203	7.9924	4.3001	7.9342
	ctRBM	0.2909	0.3334	0.1896	0.1263	0.6037	0.9026	0.4722	0.6777	0.9764	1.1120	0.9024	1.0017
	GTRBM	0.3029	0.2198	0.1308	<b>0.1038</b>	0.4908	<b>0.3296</b>	0.3790	0.3802	0.9022	0.9982	0.8517	1.3826
	DDNE	0.2132	0.2205	<b>0.1253</b>	0.1104	0.3923	0.4828	0.3001	0.3841	0.6437	0.8029	0.8429	0.9997
	GC-LSTM	<b>0.1703</b>	<b>0.2304</b>	0.1902	0.1923	<b>0.3070</b>	0.3763	<b>0.1008</b>	<b>0.1167</b>	<b>0.2186</b>	<b>0.1997</b>	<b>0.6198</b>	<b>0.6459</b>
ER-	node2vec	39.6822	23.9229	21.3259	10.7327	18.9222	17.7015	17.9036	18.124	32.6917	39.4343	46.9176	47.9336
	TNE	9.9361	3.9249	1.2252	1.7626	19.8352	15.933	13.4933	13.0638	12.0855	16.4426	14.2701	10.2749
	ctRBM	0.6217	0.5559	0.4767	0.5707	1.8853	1.8302	1.4198	1.4160	2.4745	2.5662	2.0879	2.3072
	GTRBM	0.6211	0.6938	0.4437	0.4954	1.1039	1.5540	1.5289	1.6229	1.3325	1.4414	1.6834	1.4116
	DDNE	0.6793	0.6479	0.4567	0.4980	1.3741	1.4186	1.3315	1.2100	1.2577	1.0237	1.1705	1.2261
	GC-LSTM	<b>0.0621</b>	<b>0.0908</b>	<b>0.0084</b>	<b>0.0270</b>	<b>0.0857</b>	<b>0.1418</b>	<b>0.0704</b>	<b>0.0724</b>	<b>0.0515</b>	<b>0.1014</b>	<b>0.1171</b>	<b>0.1339</b>

#### 4.5 Prediction of Important Links

In summary, although some methods have superior performance in the statistical metrics AUC or GMAUC, they are not satisfactory in terms of ER and predict many wrong invalid links. In some real-world scenarios, we may focus more on whether the most important links are predicted correctly. Therefore, we further evaluate the performance of all models on some important links. Here, we use two metrics to measure the importance of each link: degree centrality (DC) and edge intermediate centrality (EBC). DC is generally used to measure the node's importance according to the number of neighbors. In this paper, we use the sum of the two nodes' DC to measure the importance of the link between them. We select the top 10% important links based on DC and EBC, and then calculate the ER for these

important links, as shown in TABLE 4. In most cases, our GC-LSTM model has the lowest ER on all data sets for short-term and long-term prediction. This validates that our GC-LSTM model still performs quite well in predicting the important links. Furthermore, comparing the results of TABLE 3 and TABLE 4, we find that the ER on the top 10% important links is much smaller than that on all links. This indicates that the performance of our model on those important links are better than those on less important ones.

#### 5 Conclusion

In this study, we propose a new deep learning model, GC-LSTM, as an encoder-decoder architecture for dynamic network link prediction. The entire GC-LSTM model consists of LSTM and GCN, where LSTM is used

**Table 4** Prediction ER of the top 10% important links in terms of DC and EBC.

Metric	Method	CONTACT		HYPERTEXT09		ENRON		RADOSLAW		FB-FORUM		LKML	
		20	80	20	80	20	80	20	80	20	80	20	80
DC	node2vec	0.6279	0.6297	0.4397	0.4837	0.4900	0.4524	0.4735	0.5203	0.3873	0.3454	0.5034	0.5289
	TNE	0.9622	0.9551	0.2869	0.3360	0.3446	0.3315	0.5068	0.4413	0.0595	0.0558	0.6390	0.6288
	ctRBM	0.2739	0.3307	0.1842	0.2105	0.4193	0.4410	0.3028	0.3097	0.1095	0.1137	0.3291	0.3341
	GTRBM	0.2209	0.2390	0.2073	0.2377	0.4098	0.4322	0.2109	0.2198	0.1127	0.1239	0.2973	0.3030
	DDNE	0.1293	0.1359	0.1096	0.1147	0.2270	0.2133	0.0803	0.1249	0.1190	0.1088	<b>0.1653</b>	<b>0.1821</b>
	GC-LSTM	<b>0.1679</b>	<b>0.1422</b>	<b>0.0438</b>	<b>0.0663</b>	<b>0.1173</b>	<b>0.1455</b>	<b>0.0289</b>	<b>0.0357</b>	<b>0.0043</b>	<b>0.0075</b>	0.2006	0.2955
EBC	node2vec	0.6747	0.6509	0.5602	0.5064	0.4607	0.5953	0.4657	0.4397	0.6517	0.6799	0.8729	0.8698
	TNE	0.9998	0.9987	0.6655	0.4705	0.9598	0.9590	1.0000	1.0000	0.9986	1.0000	0.9992	0.9992
	ctRBM	0.5396	0.5619	0.4998	0.5203	0.6512	0.7381	0.2165	0.2291	0.4432	0.4508	0.7279	0.7503
	GTRBM	0.4418	0.4573	0.4040	0.4324	0.6906	0.7420	0.2399	0.2511	0.4507	0.4529	0.6370	0.6524
	DDNE	<b>0.2713</b>	<b>0.2849</b>	0.3019	0.3157	0.4988	0.5471	0.2083	0.2508	0.2697	0.3014	0.6435	0.6614
	GC-LSTM	0.3252	0.4551	<b>0.1948</b>	<b>0.2615</b>	<b>0.3584</b>	<b>0.3985</b>	<b>0.1455</b>	<b>0.1588</b>	<b>0.1694</b>	<b>0.1591</b>	<b>0.4975</b>	<b>0.6038</b>

to learn the temporal characteristics from continuous snapshots, while GCN is used to learn the structural characteristics of the snapshot at each moment. A fully connected layer network is used as a decoder to convert the extracted spatio-temporal features back to the original space. The proposed model can capture not only the time dependence between a sequence of snapshots, but also the impact of the network structure, i.e., it can better understand the network evolution pattern. Finally, we conducted many experiments to compare our GC-LSTM model with traditional network link prediction methods on various dynamic network datasets. The results validate that our model outperforms the others in terms of AUC, GMUUC and ER. Meanwhile, we also find that GC-LSTM model behaves well in predicting those important links characterized by DC and EBC. *It is better than the others in most cases.*

Since GC-LSTM needs pre-training, its time complexity is relatively high. In the future, we will focus on accelerating the model and predicting dynamic link for large-scale dynamic

networks which may consist of hundreds of thousands nodes and links, to make it more practical in reality.

### Acknowledgements

This research was supported by the National Natural Science Foundation of China under Grant No. 62072406, the Natural Science Foundation of Zhejiang Province under Grant No. LY19F020025, the Major Special Funding for Science and Technology Innovation 2025 in Ningbo under Grant No. 2018B10063, the National Key Research and Development Program of China under Grant No. 2018AAA0100801, the Special Technology Project for Pre-research of Military Common Information System Equipment under Grant No. JZX6Y201907010467.

### Compliance with Ethical Standards

**Ethical approval:** This article does not contain any studies with human participants or animals performed by any of the authors.

**Conflict of Interest:** The authors declare that they have no conflict of interest.

## References

1. Hui Liu, Xuanhong Xu, Guanrong Chen, and Jun An Lu. Optimizing pinning control of complex dynamical networks based on spectral properties of grounded laplacian matrices. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, PP(99):1–11, 2018.
2. Jiancheng Sun, Yong Yang, Neal N Xiong, Liyun Dai, Xiangdong Peng, and Jianguo Luo. Complex network construction of multivariate time series using information geometry. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 1(99):1–16, 2017.
3. Lin Wang and Jeff Orchard. Investigating the evolution of a neuroplasticity network for learning. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 1(99):1–13, 2017.
4. Yao Ma, Ziyi Guo, Zhaocun Ren, Jiliang Tang, and Dawei Yin. Streaming graph neural networks. In *vProceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages v719–728, 2020.
5. Nur Nasuha Daud, Siti Hafizah Ab Hamid, Muntadher Saadoun, Firdaus Sahran, and Nor Badrul Anuar. Applications of link prediction in social networks: A review. *Journal of Network and Computer Applications*, 166:102716, 2020.
6. Ricky Laishram, Kishan Mehrotra, and Chilukuri K Mohan. Link prediction in social networks with edge aging. In *IEEE International Conference on Tools with Artificial Intelligence*, pages 606–613. IEEE, 2017.
7. Mansooreh Kazemilari and Maman Abdurachman Djauhari. Correlation network analysis for multi-dimensional data in stocks market. *Physica A Statistical Mechanics & Its Applications*, 429:62–75, 2015.
8. Chenbo Fu, Minghao Zhao, Lu Fan, Xinyi Chen, Jinyin Chen, Zhefu Wu, Yongxiang Xia, and Qi Xuan. Link weight prediction using supervised learning methods and its application to yelp layered network. *IEEE Transactions on Knowledge & Data Engineering*, PP(99):1–1, 2018.
9. Qi Xuan, Zhi Yuan Zhang, Chenbo Fu, Hong Xiang Hu, and Vladimir Filkov. Social synchrony on complex networks. *IEEE Transactions on Cybernetics*, 48(5):1420–1431, 2018.
10. Rakshit Trivedi, Hanjun Dai, Yichen Wang, and Le Song. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In *International Conference on Machine Learning*, pages 3462–3471. PMLR, 2017.
11. Jinyin Chen, Yangyang Wu, Xuanheng Xu, Yixian Chen, Haibin Zheng, and Qi Xuan. Fast gradient attack on network embedding. *arXiv preprint arXiv:1809.02797*, 2018.
12. Dmytro Lande, Minglei Fu, Wen Guo, Iryna Balagura, Ivan Gorbov, and Hongbo Yang. Link prediction of scientific collaboration networks based on information retrieval. *World Wide Web*, pages 1–19, 2020.
13. Seven-layer model in complex networks link prediction: A survey.
14. Hartmut HK Lentz, Andreas Koher, Philipp Hövel, Jörn Gethmann, Carola Sauter-Louis, Thomas Selhorst, and Franz J Conraths. Disease spread through animal movements: a static and temporal network analysis of pig trade in germany. *PLoS One*, 11(5):e0155196, 2016.
15. Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *ICLR*, 2017.
16. Ting Zhang, Bang Liu, Di Niu, Kunfeng Lai, and Yu Xu. Multiresolution graph attention networks for relevance matching. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 933–942, 2018.
17. Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
18. Franco Manessi, Alessandro Rozza, and Mario Manzo. Dynamic graph convolutional networks. *Pattern Recognition*, 97:107000, 2020.
19. Ting Zhang, Kun Zhang, Laishui Lv, and Xun Li. Temporal link prediction using node centrality and time series. *International Journal of Future Computer and Communication*, 9(3):62–65, 2020.
20. Xiaomin Wu, Jianshe Wu, Yafeng Li, and Qian Zhang. Link prediction of time-evolving network based on node ranking. *Knowledge-Based Systems*, 195:105740, 2020.
21. Mahmudur Rahman, Tanay Kumar Saha, Mohammad Al Hasan, Kevin S Xu, and Chandan K Reddy. Dylink2vec: Effective feature representation for link prediction in dynamic networks. *arXiv preprint arXiv:1804.05755*, 2018.

22. Behnaz Moradabadi and Mohammad Reza Meybodi. Link prediction based on temporal similarity metrics using continuous action set learning automata. *Physica a: statistical mechanics and its applications*, 460:361–373, 2016.
23. Ke-ke Shang, Wei-sheng Yan, and Michael Small. Evolving networks—using past structure to predict the future. *Physica A: Statistical Mechanics and its Applications*, 455:120–135, 2016.
24. Kuo Chi, Guisheng Yin, Yuxin Dong, and Hongbin Dong. Link prediction in dynamic networks based on the attraction force between nodes. *Knowledge-Based Systems*, 181:104792, 2019.
25. Lekui Zhou, Yang Yang, Xiang Ren, Fei Wu, and Yueting Zhuang. Dynamic network embedding by modeling triadic closure process. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
26. Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, Tao Schardl, and Charles Leiserson. Evolvegcn: Evolving graph convolutional networks for dynamic graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5363–5370, 2020.
27. Palash Goyal, Sujit Rokka Chhetri, and Arquimedes Canedo. dyngraph2vec: Capturing network dynamics using dynamic graph representation learning. *Knowledge-Based Systems*, 187:104816, 2020.
28. Jinyin Chen, Jian Zhang, Xuanheng Xu, Chenbo Fu, Dan Zhang, Qingpeng Zhang, and Qi Xuan. E-lstm-d: A deep learning framework for dynamic network link prediction. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019.
29. Carter Chiu and Justin Zhan. Deep learning for link prediction in dynamic networks using weak estimators. *IEEE Access*, 6:35937–35945, 2018.
30. Behnaz Moradabadi and Mohammad Reza Meybodi. Link prediction in fuzzy social networks using distributed learning automata. *Applied Intelligence*, 47(3):837–849, 2017.
31. Deepak Mangal, Niladri Sett, Sanasam Ranbir Singh, and Sukumar Nandi. Link prediction on evolving social network using spectral analysis. In *2013 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, pages 1–6. IEEE, 2013.
32. Seyed Mehran Kazemi, Rishab Goel, Sepehr Eghbali, Janahan Ramanan, Jaspreet Sahota, Sanjay Thakur, Stella Wu, Cathal Smyth, Pascal Poupard, and Marcus Brubaker. Time2vec: Learning a vector representation of time. *arXiv preprint arXiv:1907.05321*, 2019.
33. Rakshit Trivedi, Mehrdad Farajtabar, Prasenjeet Biswal, and Hongyuan Zha. Dyrep: Learning representations over dynamic graphs. In *International Conference on Learning Representations*, 2019.
34. Mark EJ Newman. Clustering and preferential attachment in growing networks. *Physical review E*, 64(2):025102, 2001.
35. Ke-Jia Chen, Yang Chen, Yun Li, and Jingyu Han. A supervised link prediction method for dynamic networks. *Journal of Intelligent & Fuzzy Systems*, 31(1):291–299, 2016.
36. Tong Wang, Xing Sheng He, Ming Yang Zhou, and Zhong Qian Fu. Link prediction in evolving networks based on popularity of nodes. *Scientific Reports*, 7(1), 2017.
37. Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
38. Jie Chen, Tengfei Ma, and Cao Xiao. Fastgcn: fast learning with graph convolutional networks via importance sampling. *arXiv preprint arXiv:1801.10247*, 2018.
39. Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pages 593–607. Springer, 2018.
40. Nahla Mohamed Ahmed, Ling Chen, Yulong Wang, Bin Li, Yun Li, and Wei Liu. Sampling-based algorithm for link prediction in temporal networks. *Information Sciences*, 374:1–14, 2016.
41. Nahla Mohamed Ahmed and Ling Chen. An efficient algorithm for link prediction in temporal uncertain social networks. *Information Sciences*, 331:120–136, 2016.
42. Min Yang, Junhao Liu, Lei Chen, Zhou Zhao, Xiaojun Chen, and Ying Shen. An advanced deep generative framework for temporal link prediction in dynamic networks. *IEEE transactions on cybernetics*, 50(12):4946–4957, 2019.
43. Xiu-Xiu Zhan, Ziyu Li, Naoki Masuda, Petter Holme, and Huijuan Wang. Susceptible-infected-spreading-based network embedding in static and temporal networks. *EPJ Data Science*, 9(1):30, 2020.

44. Hao Peng, Jianxin Li, Hao Yan, Qiran Gong, Senzhang Wang, Lin Liu, Lihong Wang, and Xiang Ren. Dynamic network embedding via incremental skip-gram with negative sampling. *Science China Information Sciences*, 63(10):1–19, 2020.
45. Giang Hoang Nguyen, John Boaz Lee, Ryan A Rossi, Nesreen K Ahmed, Eunye Koh, and Sungchul Kim. Continuous-time dynamic network embeddings. In *3rd International Workshop on Learning Representations for Big Networks (WWW BigNet)*, 2018.
46. Kai Lei, Meng Qin, Bo Bai, Gong Zhang, and Min Yang. Gcn-gan: A non-linear temporal link prediction model for weighted dynamic networks. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 388–396. IEEE, 2019.
47. Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864. ACM, 2016.
48. Linhong Zhu, Dong Guo, Junming Yin, Greg Ver Steeg, and Aram Galstyan. Scalable temporal latent space inference for link prediction in dynamic social networks. *IEEE Transactions on Knowledge and Data Engineering*, 28(10):2765–2777, 2016.
49. Xiaoyi Li, Nan Du, Hui Li, Kang Li, Jing Gao, and Aidong Zhang. A deep learning approach to link prediction in dynamic networks. In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pages 289–297. SIAM, 2014.
50. Taisong Li, Bing Wang, Yasong Jiang, Yan Zhang, and Yonghong Yan. Restricted boltzmann machine-based approaches for link prediction in dynamic networks. *IEEE Access*, 6:29940–29951, 2018.
51. Taisong Li, Jiawei Zhang, S Yu Philip, Yan Zhang, and Yonghong Yan. Deep dynamic network embedding for link prediction. *IEEE Access*, 6:29219–29230, 2018.
52. Yang Yang, Ryan N Lichtenwalter, and Nitesh V Chawla. Evaluating link prediction methods. *Knowledge and Information Systems*, 45(3):751–782, 2015.
53. Ruthwik R Junuthula, Kevin S Xu, and Vijay K Devabhaktuni. Evaluating link prediction accuracy in dynamic networks with added and removed edges. In *Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom), 2016 IEEE International Conferences on*, pages 377–384. IEEE, 2016.
54. Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter Van den Broeck. What’s in a crowd? analysis of face-to-face behavioral networks. *Journal of theoretical biology*, 271(1):166–180, 2011.
55. Augustin Chaintreau, Pan Hui, Jon Crowcroft, Christophe Diot, Richard Gass, and James Scott. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing*, 6(6):606–620, 2007.
56. Ryan Rossi and Nesreen Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI*, volume 15, pages 4292–4293, 2015.
57. Tore Opsahl. Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks*, 35(2):159–167, 2013.
58. Lin Yao, Luning Wang, Lv Pan, and Kai Yao. Link prediction based on common-neighbors for dynamic social network. *Procedia Computer Science*, 83:82–89, 2016.