

Article

A Feasible Temporal Links Prediction Framework Combining with Improved Gravity Model

Xinyu Huang , Dongming Chen * and Tao RenSoftware College, Northeastern University, Shenyang 110169, China; neuhyx@163.com (X.H.);
rent@swc.neu.edu.cn (T.R.)

* Correspondence: chendm@mail.neu.edu.cn

Received: 19 December 2019; Accepted: 2 January 2020; Published: 5 January 2020



Abstract: Social network analysis is a multidisciplinary study covering informatics, mathematics, sociology, management, psychology, etc. Link prediction, as one of the fundamental studies with a variety of applications, has attracted increasing focus from scientific society. Traditional research based on graph theory has made numerous achievements, whereas suffering from incapability of dealing with dynamic behaviors and low predicting accuracy. Aiming at addressing the problem, this paper employs a diagonally symmetrical supra-adjacency matrix to represent the dynamic social networks, and proposes a temporal links prediction framework combining with an improved gravity model. Extensive experiments on several real-world datasets verified the superiority on competitors, which benefits recommending friends in social networks. It is of remarkable significance in revealing the evolutions in temporal networks and promoting considerable commercial interest for social applications.

Keywords: social network; temporal links prediction; gravity model; multilayer network

1. Introduction

The analysis of social networks has drawn increasing attention in the field of sociology. It analyzes and explores the potential relations between social objects [1]. The rapid development of social media has brought us plentiful data sources, along with enormous challenges such as data incompleteness and dynamic changes [2]. On the one hand, researchers are facing data incompleteness problem since only part of social information can be collected from social platforms. On the other hand, the dynamic changes may lead the nodes and links to appear and disappear in the future, which makes the underlying graph longitudinal [3].

Link prediction [4], as fundamental research in social network analysis, is proposed to detect unobserved links from existing parts of the network [5] or forecast future links from current network structures [6]. The former research, also known as missing links prediction, has been fruitful during the last decade, whereas the prediction of future links is more challenging to estimate the upcoming connections with limited social information. This study is of great importance, not only in revealing the evolution of social networks, but also benefiting network management, such as promoting useful links or prohibiting harmful interactions. For instance, a recommendation system [7], as a typical application of temporal links prediction, is designed for individuals to make friends or purchase goods via efficient predicted results, which brings considerable benefits for corporations.

Numerous attempts have been made to address the problem of temporal links prediction, but it is a really difficult task. Firstly, the observed social information is quite limited, which leads to the smoothness assumption [8] being frequently adopted in studies, thereby the methods may be incapable when the network changes seriously. Secondly, longitudinal bias is inevitable, as the dynamic changes must be shifted towards future [3]. Finally, the different observation of extensive network changes

over time may also lead to various social structures, which may yield extremely different predicting results. To solve the problem of temporal links prediction, this paper proposes a dynamic similarity framework with an improved gravity model to estimate the future links in temporal networks.

The rest of this paper is organized as follows. Section 2 introduces the related works on link prediction. Section 3 presents the mathematical model, the improved gravity model, and the framework for predicting temporal links. Section 4 presents the experiments and analysis, including comparison experiments on real-world dataset separated by different levels, which verified the feasibility and veracity of the method. Section 5 summarizes the whole paper and provides concluding remarks.

2. Related works

Link prediction was first proposed on SIGKDD in 2005 [4]. Afterwards, Liben-Nowell and Kleinber reviewed the link prediction in social networks, which attracted more and more scholars devoting themselves to this field [9]. In 2011, Lü summarized the existing methods and classified them into three categories: structure similarity index, maximum likelihood approximation, and probabilistic model [10]. In 2017, Pech et al. [11] introduced robust principal component analysis (robust PCA) method into link prediction and estimated the missing links of the adjacency matrix. Experimental results show that, when the target network is connected and sufficiently dense, the proposed method achieves much higher accuracy compared of some the state-of-the-art algorithms. A brief summary of the classic predictors is shown in Table 1.

The prediction of future links [12,13], i.e., temporal links prediction, aims to predict the links in a network that would appear in the next state of period. As for temporal networks, a series of mathematic models are proposed, such as temporal graphs [14], evolving graphs [15], time-varying graphs, dynamic networks [16], etc. Three representative models are available to depict dynamic behaviors: Snapshots, Contact sequences and Interval graphs.

Snapshots, as a favorable model to exhibit dynamic behaviors, have been widely used in various application scenarios. By employing such representation, unsupervised learning methods are feasible to estimate the links at time t with the observed network structures at time $[1, t - 1]$ [17–19]. Besides, statistical methods, such as Exponential Smoothing (EPS) [20] and Autoregressive Integrated Moving Average (ARIMA) [21], are also employed to predict temporal links with snapshots representation. However, Snapshots suffer from coarse-grained depiction of continuous changes, which probability result in poor predictive performance and misleading results [22]. Distinguished by the duration of interaction being negligible or not, Contact sequences and Interval graphs are proposed to illustrate the network dynamics. A temporal network can be represented by a series of triplet (i, j, t) where i and j are entities and t is the time of interaction. If the duration of interaction is not negligible, the framework can be represented by $(i, j) = (t_1, t'_1), \dots, (t_n, t'_n)$, namely Interval graphs. Although they encode more interactions information of dynamic behaviors, the above-mentioned two models are scarcely used in studying temporal link prediction for their complicated expressions.

By combining with temporal correlations and evolutions of link occurrences, Özcan and Öğüdücü [23] proposed Multivariate Time Series Link Prediction method. Experiments on real-world bibliographic datasets showed that the proposed method, which can incorporate covariance structures, achieved better results for temporal links prediction than classic competitors. Considering both temporal dynamics and multi-relational properties in bibliographic networks, Sett et al. [3] proposed a robust and efficient set of features named time-aware multi-relational link prediction (TMLP) features to predict future links using supervised learning framework in dynamic multi-relational network. They analyzed unsupervised performance of individual features, and then applied a supervised learning method that combines multiple features towards link prediction. To overcome the inherent problem of longitudinal bias, random forests supervised learning framework is utilized in the experiments. Experiments on bibliographic datasets showed the effectiveness. However, the above-mentioned methods rely on the plentiful data features and the performance on general

temporal network is uncertain. Overall, the study on temporal links prediction is blossoming and still requires great endeavors to achieve better performance.

Table 1. A brief summary of the existing link prediction methods.

Indicator	Topology	Definition ¹	Complexity ²
CN [24]	Local	$S_{xy} = \Gamma(x) \cap \Gamma(y) $	$O(n^2)$
Salton [25]	Local	$S_{xy} = \frac{ \Gamma(x) \cap \Gamma(y) }{\sqrt{k_x k_y}}$	$O(n^2)$
Jaccard [26]	Local	$S_{xy} = \frac{ \Gamma(x) \cap \Gamma(y) }{ \Gamma(x) \cup \Gamma(y) }$	$O(2n^2)$
Sorenson [27]	Local	$S_{xy} = \frac{2 \times \Gamma(x) \cap \Gamma(y) }{k_x + k_y}$	$O(n^2)$
HPI [28]	Local	$S_{xy} = \frac{ \Gamma(x) \cap \Gamma(y) }{\min(k_x, k_y)}$	$O(n^2)$
HDI [29]	Local	$S_{xy} = \frac{ \Gamma(x) \cap \Gamma(y) }{\max(k_x, k_y)}$	$O(n^2)$
LHN-I [29]	Local	$S_{xy} = \frac{ \Gamma(x) \cap \Gamma(y) }{k_x k_y}$	$O(n^2)$
AA [30]	Local	$S_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log k_z}$	$O(2n^2)$
RA [31]	Local	$S_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k_z}$	$O(2n^2)$
PA [32]	Local	$S_{xy} = k_x k_y$	$O(2n)$
LP [33]	Semi-Local	$S = A^2 + \alpha \cdot A^3$	$O(n^3)$
Katz [34]	Global	$S = (I - \alpha \cdot A)^{-1} - I$	$O(n^3)$
LHN-II [35]	Global	$S = 2m\lambda_1 D^{-1} (I - \frac{\phi}{\lambda_1})^{-1} D^{-1}$	$O(n^3)$
LRW [36]	Semi-Local	$S_{xy}^{LRW}(t) = q_x \cdot \pi_{xy}(t) + q_y \cdot \pi_{yx}(t)$	$O(nk^t)$
SRW [36]	Semi-Local	$S_{xy}^{SRW}(t) = \sum_{l=1}^t S_{xy}^{LRW}(l) = q_x \sum_{l=1}^t \pi_{xy}(l) + q_y \sum_{l=1}^t \pi_{yx}(l)$	$O(nk^t)$
RWR [37]	Semi-Local	$S_{xy}^{LRW}(t) = q_{xy} + q_{yx}$	$O(n^3)$
ACT [38]	Semi-Local	$S_{xy}^{ACT} = \frac{1}{l_{xx}^+ \cdot l_{yy}^+ - 2l_{xy}^+}$	$O(n^3)$
SimR [39]	Global	$S_{xy}^{SimR} = C \frac{\sum_{v \in \Gamma(x)} \sum_{v' \in \Gamma(y)} S_{zz'}^{SimR}}{k_x k_y}$	$O(n^3)$
Cos+ [40]	Semi-Local	$S_{xy}^{cos+} = \cos(x, y)^+ = \frac{l_{xy}^+}{\sqrt{l_{xx}^+ \cdot l_{yy}^+}}$	$O(n^3)$
TS [41]	Global	$S^{Tr} = (I - \varepsilon S)^{-1} S$	$O(n^3)$
LowRank [11]	Global	$S = \min_{X^*, E} \ X^*\ _* + \lambda \ E\ _1$	$O(nk^3)$
MFI [42]	Global	$S = (I + \alpha \cdot L)^{-1}, \alpha > 0$	$O(n^3)$

¹ $\Gamma(x)$ and $\Gamma(y)$ denote the neighbors of node x and y , respectively; k_x and k_y are the degrees of node x and y , respectively; m is the number of edges; l_{xy}^+ denotes the element at row x , column y of the pseudo-inverse matrix L^+ ; q_{xy} represents the probability of random walk from node x to node y ; ε represents an adjustable parameter; $\pi_{xy}(l)$ is the random walk probability from node x to node y at time l ; ² n denotes the number of nodes; k is the average degree of nodes; and t is the step of random walk steps.

3. Modeling and Methods

3.1. Model

The problem of link prediction is described as estimating the likelihood of all the possible links with a given network model $G = (V, E)$, where $V = \{v_1, v_2, \dots, v_n\}$ is the node set and $E = \{(v_i, v_j)\}$, $(v_i, v_j \in V)$ is the edge set. Suppose the likelihood (or similarity) of the link between every two nodes in G can be calculated by a certain algorithm, and then sorted by ascending order; the future links are obtained by top- k links, where k is the target links amount.

In this paper, we employ a generative multilayer network model [43], which can transform the temporal network into a considerable collection of snapshots. An illustration of the multilayer network model with data derived from a large European research institution [44] is shown in Figure 1 and the corresponding supra-adjacency matrix is shown in Figure 2.

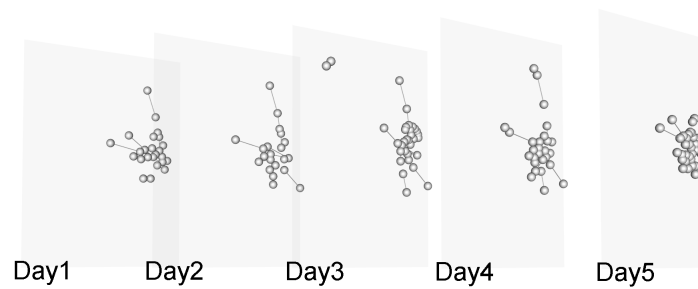


Figure 1. Snapshot of temporal network of the first five days.

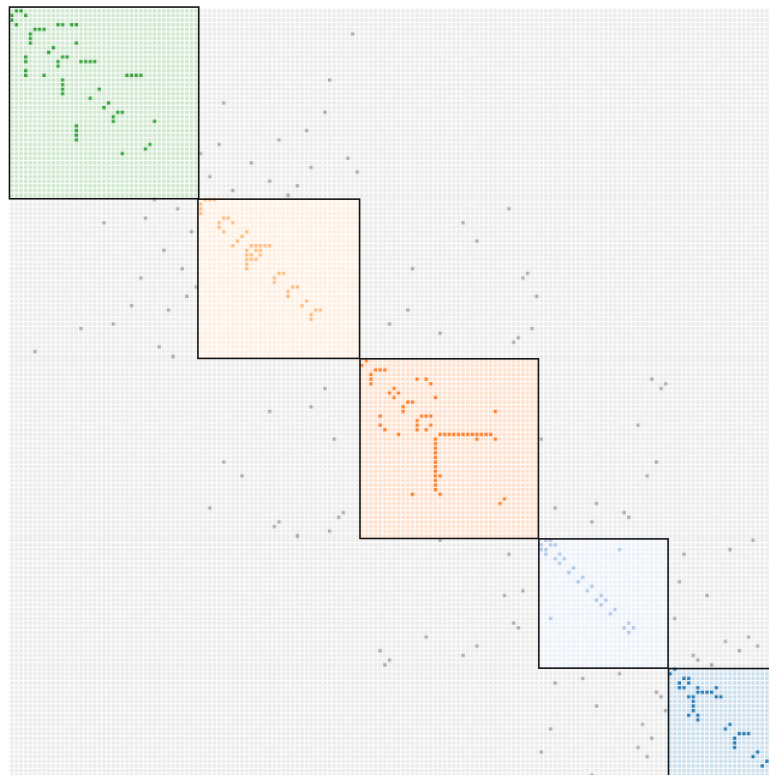


Figure 2. Supra-adjacency matrix representation. The diagonal blocks with different colors represent the network structures at different snapshots.

Suppose given a temporal network G^T with T separated slices, where each slice is modeled as a mono-layer network (i.e., single-layer network), T is the total number of layers, and $t = 1, 2, \dots, T$; the model is denoted by

$$G^T = \{\sum V^t, \sum E^t\}, \quad (1)$$

where $\sum V^t$ and $\sum E^t$ are the union of the nodes and edges at each slice. To simplify the research problem, we utilize undirected graph to present the network structures in each snapshot. Thus, the temporal network can be represented by a diagonally symmetrical supra-adjacency matrix (denoted by $\tilde{M}$), described as

$$\tilde{M} = \begin{bmatrix} A_1 & I_{1,2} & & \\ I_{2,1} & A_2 & \ddots & \\ & \ddots & \ddots & I_{T-1,T} \\ & & I_{T,T-1} & A_T \end{bmatrix} \in \mathcal{R}^{N \times N}, \quad (2)$$

where $A_1, A_2, \dots, A_T$ are the adjacency matrix of time 1, 2, $\dots$, T , respectively, representing the links (i.e., intra-layer edges). N is the total numbers of the nodes, which can be calculated by

$N = \sum_{1 \leq l \leq T} |V^l|$. I denotes the relationship of the node located in the continuous snapshots (i.e., the inter-layer edges). By utilizing the former temporal information from time 1 to $T - 1$, our goal is to predict the links at the last time T . The problem is described as

$$P_{(u,v)} \otimes T = \sum_{1 \leq t \leq T-1} \tau(p_{(u,v)} \otimes t) \cdot E_{(u,v)} \otimes t, \quad (3)$$

where $P_{(u,v)} \otimes T$ is the predicted likelihood of link connected node u and v at time T , $p_{(u,v)} \otimes t$ is the likelihood indicator of node u and v at time t , τ depicts the varying function of determining the influence at different time t and $E_{(u,v)} \otimes t$ is the existence function of nodes u and v at time t , described as

$$E_{(u,v)} \otimes t = \begin{cases} \sum I_u I_v, & \text{if node } u \text{ links } v \text{ at time } t \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

where I_u and I_v represent the interlayer edges of node u and v , respectively. The prediction of future links (at time T) relies on the former structures (i.e., from time 1 to time $T - 1$). The links with the maximum score are the prediction results of future links. The prediction result is evaluated by precision, recall, F_1 -value, accuracy, etc. Considering the significance of prediction, AUC [45] is employed for evaluation in this paper.

3.2. Definitions

Gravity is the force between objects, which relates to the qualities of the objects and the distance between the two objects. The gravity we focus on in this paper is utilized to describe the strength of the interactions between the two nodes. The gravity in networks [46] is defined as

$$G_{i,j} = \frac{k_i \cdot k_j}{d_{ij}^2}, \quad (5)$$

where $G_{i,j}$ represents the gravity between node i and j , d_{ij} depicts the shortest path length between node i and node j , and k_i is the degree of node i . Inspired by this model, we simplify the time-consuming process, namely the calculation of shortest path, by merely considering the neighbors within two steps (i.e., neighbors and second-order neighbors), denoted by GR. Thus, it can be reduced to the accumulation of gravity between common neighbors, and the likelihood of existing links between node i and j is given by

$$S_{i,j}^{GR} = \sum_{z \in \Gamma(i) \cap \Gamma(j)} \frac{G_{i,z} \cdot G_{z,j}}{\delta^2}, \quad (6)$$

where $\Gamma(i)$ and $\Gamma(j)$ are the neighbors of node i and node j , respectively. δ is the steps between node i and j if there are common neighbors between i and j , and the number of nodes otherwise. Thus, the predicted likelihood of node i and j at time t (marked as $S_{i,j}^{GR} \otimes T$) is given by

$$S_{i,j}^{GR} \otimes T = \sum_{t=0}^{t-1} G_{i,j}^{GR|t} \times e^{-\alpha(t-p)}, \quad (7)$$

where $G_{i,j}^{GR|t}$ is the gravity of node i and j at time t . $e^{-\alpha(t-p)}$ is the above-mentioned τ function to enforce temporal effect on the similarity evaluation [47], namely dynamic similarity process (for short DS). p is the existence duration of i and j . α is the attenuation constant in the range of $[0, 1]$. The values change with x under different α , as shown in Figure 3.

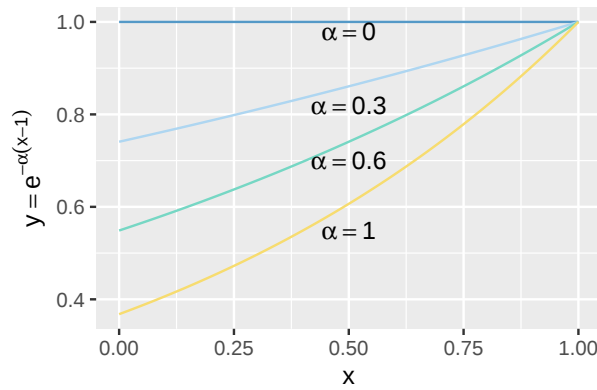


Figure 3. Illustration of τ function with different parameters. α varies in $[0, 1]$, which depicts the attenuation level. The larger is α , the less effect there is on the current x .

Suppose k is the number of links to be predicted in the future. The algorithm of predicting future links is described as the following steps.

Step 1: Obtain all the node pairs at $T - 1$ layer as the target-predicting links.

Step 2: Collect all the existing links from the start time to time $T - 1$, marked as training set.

Step 3: Calculate the likelihood of node pairs in the training set in terms of Equation (7).

Step 4: Sort the possible links in descending order and obtain the top- k results as the predicted result.

The pseudo-code of the above process is shown as Algorithm 1.

Algorithm 1: Temporal links prediction framework.

Input: temporal network G^T , target links amount k

Output: future links L

```

1  $T = \text{len}(G^T.\text{layers}());$ 
2  $g^{T-1} = G^T.\text{layers}(T - 1);$  // Obtain the network structures at time  $T - 1$ .
3  $\text{targets} = \text{array}();$ 
4  $\text{candidates} = \{\};$ 
5 for  $i = 0; i < \text{len}(g^{T-1}.\text{nodes}); i++$  do
6   for  $j = i + 1; j < \text{len}(g^{T-1}.\text{nodes}); j++$  do
7     if  $(i, j) \in g^{T-1}.\text{edges}()$  then
8        $\text{targets} += (i, j);$  // Step 1: Obtain target-predicting links.
9 for  $t = 0; t \leq T - 1; t++$  do
10    $g^t = G^T.\text{layers}(t);$ 
11   for  $k = 0; k < \text{len}(\text{targets}); k++$  do
12      $(u, v) = \text{targets}[k];$  // Step 2: Obtain testing set.
13     if  $(u, v) \in g^t.\text{edges}()$  then
14        $S_{u,v} += S_{u,v}^{GR|t} * e^{-\alpha(t-p)};$  // Step 3: Calculate the likelihood of node pairs.
15        $\text{candidates}[(u, v)] = S_{u,v};$ 
16  $L = \text{sort}(\text{candidates.values}(), \text{descending}, k);$  // Step 4: Obtain the predictive results.
17 return  $L;$ 

```

3.3. Complexity Analysis

Suppose m and n are the number of edges and nodes, respectively, the average degree of nodes is d , and the total layers of temporal networks is T . The complexity of calculating common neighbors is $O(d^2)$. Thus, the complexity of the proposed indicator $S_{i,j}^{GR}$ is lower than $O(d^2 + d)$, which can

be simplified as $O(d^2)$. Traversing every two nodes at time $T - 1$ needs complexity of $O(n^2)$, thus the total complexity is $O(n^2 d^2)$. Actually, d^2 is much smaller than n , thus the proposed method can be simplified as $O(n^2)$. It is very close to competitive indicators, e.g. CN, AA, and RA are all with complexity of $O(n^2)$ [48]. The process of temporal links prediction our method $S_{i,j}^{GR} \otimes T$ is $O(n^2 T)$, which is the same as the representative Linear Regression methods, EPS [20], and so on.

4. Experiments and Discussion

The experimental environment was a Intel(R) Core (TM) i5-7200U CPU @ 2.50 GHz (4 CPUs), 2.7 GHz, the memory was 8 GB DDR3, the operating system was Windows10 64 bit, the programming language was Python 3.7.1, and the relevant libs were NetworkX 2.2 and Multinetx. The goal of the experiments was to validate the performance of the proposed method and compare with competitive indicators.

4.1. Experimental Datasets

To verify the proposed indicator, seven real-world datasets were employed in the experiments, as shown in Table 2. The real-world email dataset from a large European research institution [44] was employed to check the proposed dynamic similarity framework and the data statistics are shown in Table 3.

Table 2. Statistics of seven real-world datasets.

Dataset Name	$ V $	$ E $	$\langle k \rangle$	$ C $	$\langle c \rangle$	$ D $	r
Zachary karate Club [49]	34	78	4.59	0.57	2.41	2.22	−0.48
Dolphins social network [50]	62	159	5.13	0.26	3.06	3.36	−0.04
Terriers of 9/11 [51]	69	159	4.61	0.47	1.76	3.22	−0.04
NEUSNCP dataset ¹	89	365	4.10	0.54	3.15	1.92	−0.40
Books about US politics [52]	105	411	8.40	0.49	5.26	3.08	−0.13
American college football network [53]	115	613	10.66	0.40	10.23	2.51	0.16
Scientist collaboration network [52]	1589	2742	4.60	0.64	0.08	5.99	−0.09

Note: $|V|$ denotes the number of nodes; $|E|$ denotes the number of edges; $\langle k \rangle$ is the average degree; $|C|$ is the average clustering index; $\langle c \rangle$ is the average connectivity; $|D|$ is the average shortest path; and r represents assortativity coefficient. ¹ We developed an experimental social platform and invited hundreds of users to register. Data availability: <https://www.neusncp.com/api/about>.

Table 3. Data statistics of Email-Eu-core temporal network.

Dataset Name	$ V $	$ E $	Days
Email-Eu-core temporal network	986	332,334	803
Email-Eu-core-temporal-Dept1	309	61,046	803
Email-Eu-core-temporal-Dept2	162	46,772	803
Email-Eu-core-temporal-Dept3	89	12,216	803
Email-Eu-core-temporal-Dept4	142	48,141	803

The email dataset is convictive without providing any feature information. The e-mails only represent communication between institution members (the core), and the dataset does not contain incoming messages from or outgoing messages to the rest of the world. To illustrate the connections of each period, the dataset was separated by different intervals (daily, weekly, and monthly), as shown in Figure 4.

As shown in Figure 4, the more layers there are in the multilayer network model, the fewer connections there are in each slice. The fitting results of connections are marked with red lines in the above three panels. Here, we can see the number of edges is varying over time among all three separations. When the dataset is daily separated, the fitting result is not obvious and indicates poor predictive performance. When the dataset is weekly separated, the improved fitting result provides

a better predictive training set. When the dataset is monthly separated, an obvious fitting seems to reveal a better predictive result.

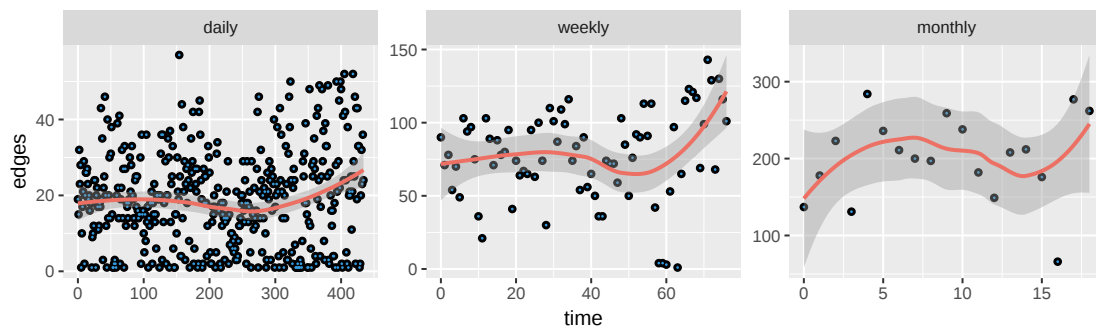


Figure 4. Email-Eu-core-temporal-Dept3 network for link prediction in the dynamic network (daily, weekly, and monthly separated).

4.2. Performance Comparison

First, we compared the proposed GR indicator with several classic methods. The results were evaluated by AUC via the average result of 100 independent experiments, as shown in Table 4.

Table 4. AUC of the GR and other indicators comparison.

Dataset Name	GR	AA	RA	JC	PA	CN
Zachary karate Club	0.8790	0.8784	0.8784	0.6281	0.8773	0.8433
Dolphins social network	0.7442	0.7428	0.7425	0.7431	0.6621	0.7379
Terriers of 9/11	0.9374	0.9339	0.9371	0.9151	0.7144	0.9103
NEUSNCP dataset	0.9110	0.9105	0.9096	0.8855	0.6725	0.9012
Books about US politics	0.8310	0.8299	0.8299	0.8397	0.2573	0.8304
American college football network	0.8775	0.8750	0.8769	0.7494	0.8381	0.8657
Scientist collaboration network	0.9431	0.9431	0.9431	0.9430	0.6725	0.9429

Table 4 shows that the AUC obtained by the proposed GR indicator generally outperforms the competitive methods, as marked by boldface. Although the performance of GR on Books about US politics dataset is inferior to that of JC indicator, it is still competitive. Generally, the experiments verified the performance of the proposed indicator in predicting unknown links.

Secondly, we compared the proposed method with the existing linear regression method, ARIMA, EPS methods, and the AUC results of six indicators are shown in Figure 5.

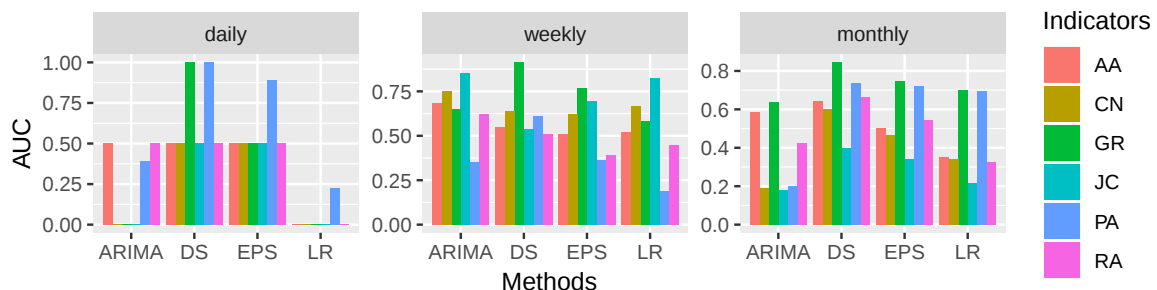


Figure 5. Comparison of temporal links prediction on Email-Eu-core-Dept3 network.

Obviously, the result of temporal links prediction obtained from our proposed method (DS combined with GR indicator) outperforms the competitive methods. When the datasets were daily separated, the proposed method obtained maximum AUC, achieving 0.913 and 0.8444 when weekly and monthly separated, respectively, which are greater than the result of the other methods.

4.3. Parameters Analysis

In the proposed framework, parameters are crucial to the prediction results. Thus, in this subsection, the parameter α is analyzed and the results are plotted in Figure 6.

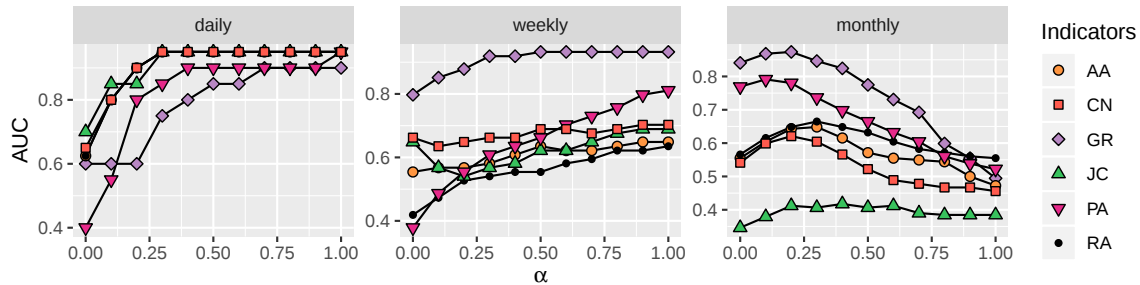


Figure 6. Analyzing parameter α in the temporal network with different separations. In general, the obtained AUC is larger with α increasing when the networks are daily and weekly separated.

As shown in Figure 6, in a temporal network with a large number of slices, a larger α is preferred to be selected. On the contrary, when the temporal network is separated with fewer slices, a smaller α contributes to obtaining better performance. Finally, the robustness of the proposed method was verified by conducting experiments on different scales of temporal networks, as shown in Figure 7.

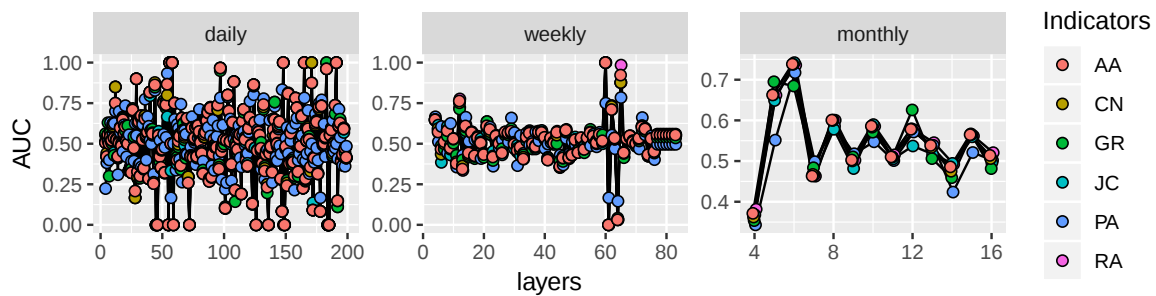


Figure 7. Robustness verification by varying layers (or time slices).

As shown in Figure 7, with the increasing of the number of slices (i.e., layers), the AUC result is changing periodically, and the varying range is stable generally. The results of the six indicators are in the same tendency, which verifies the robustness of the proposed method. To analyze the slicing effect on network features and performance, we conducted experiments with DS method and GR indicator. The results are shown in Figures 8 and 9.

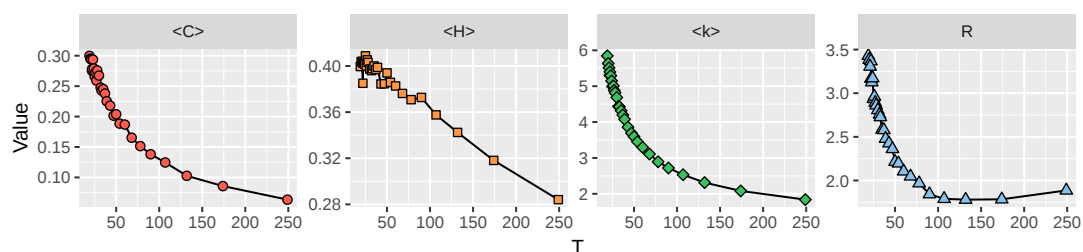


Figure 8. Illustration of network features with varying numbers of slicing. T is the number of slices ranging from 17 to 249. $\langle C \rangle$, $\langle H \rangle$, and $\langle k \rangle$ are the average clustering coefficient [54], average heterogeneity [55], and average degree of the nodes in all slices, respectively. R is the ratio of intralayer edges comparing interlayer edges. In general, R declines when the temporal network is separated into more slices.

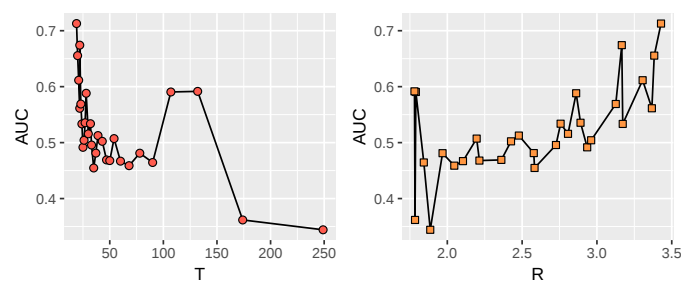


Figure 9. Relationship of performance with T (i.e., number of slices) and R (i.e., intralayer edges comparing interlayer edges). AUC declines with T increasing (or R declining) in general.

As shown in Figure 9 (left), the AUC result is declining as the number of slices increases. This results from the large numbers of intralayer edges of each time slice, which providing more edges to calculate the existence likelihood. Thus, we can utilize the network information more comprehensively. The performance is better with the ratio (i.e., intralayer edges comparing interlayer edges) increasing, as shown in Figure 9 (right).

5. Conclusions

Aiming at solving the problem of temporal links prediction in social networks, this paper proposes a novel dynamic similarity framework combining with an improved gravity model. Experiments on real-world datasets with different separations were conducted, and the experimental results show that the proposed method outperforms competitors. Afterwards, the determination of parameter α was analyzed by conducting a series of experiments, and we give the recommended selection for different temporal structures. Finally, the robustness of the proposed framework was also verified by comparing the obtained AUC results with varying time slices. Overall, the proposed framework is capable of predicting temporal links with reasonable results.

The contribution of this work is likely to benefit many real-world social applications, such as recommending new friends, protecting teenagers from harmful interactions, etc. Inspired by the multiple interactions among social actors, we have established a social platform, namely NEUSNCP (<https://www.neusncp.com>), for college students to make friends and share knowledge in various manners. By applying the proposed framework into recommending friends for newcomers, we have observed an obvious increase in user interactions. As part of future works, link prediction on more complicated models, i.e., multi-relational networks and bipartite networks on social platforms can be further studied. Notably, the research of recommendation for dynamic “user-blog” networks is in development, via a combination of collaborative filtering algorithm with temporal changes computation. In brief, the application of temporal links prediction is just unfolding.

Author Contributions: X.H. designed the framework and wrote the original draft; D.C. revised the manuscript; and T.R. checked the manuscript and made some modifications. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by Liaoning Natural Science Foundation under Grant No. 20170540320, the Doctoral Scientific Research Foundation of Liaoning Province under Grant No. 20170520358, the National Natural Science Foundation of China under Grant No. 61473073, and the Fundamental Research Funds for the Central Universities under Grant Nos. N161702001 and N172410005-2.

Acknowledgments: We would like to thank the anonymous reviewers for their careful reading and useful comments that helped us to improve the final version of this paper.

Conflicts of Interest: The author declares that there are no conflicts of interest regarding the publication of this paper.

Abbreviations

The following abbreviations are used in this manuscript:

AA	Adamic-Adar
ACT	Average commute time
ARIMA	Autoregressive Integrated Moving Average
AUC	Area Under the receiver operating characteristic Curve
CN	Common Neighbors
DS	Dynamic Similarity
EPS	Exponential Smoothing
GR	Gravity
JC	Jaccard
LR	Linear Regression
LRW	Local Random Walk
MFI	Matrix-forest index
PCA	Principal Component Analysis
RA	Resource Allocation
RWR	Random Walk with Restart
SimR	SimRank
SRW	Superposed Random Walk
TMLP	Time-aware Multi-relational Link Prediction
TS	Transferring Similarity

References

1. Wasserman, S.; Faust, K. *Social Network Analysis: Methods and Applications*; Cambridge University Press: Cambridge, UK, 1994; Volumn 8.
2. Antonacci, G.; Fronzetti Colladon, A.; Stefanini, A.; Gloor, P. It is rotating leaders who build the swarm: Social network determinants of growth for healthcare virtual communities of practice. *J. Knowl. Manag.* **2017**, *21*, 1218–1239. [[CrossRef](#)]
3. Sett, N.; Basu, S.; Nandi, S.; Singh, S.R. Temporal link prediction in multi-relational network. *World Wide Web* **2018**, *21*, 395–419. [[CrossRef](#)]
4. Getoor, L.; Diehl, C.P. Link mining: A survey. *ACM SIGKDD Explor. Newslett.* **2005**, *7*, 3–12. [[CrossRef](#)]
5. Srinivas, V.; Mitra, P. Link Prediction Using Thresholding Nodes Based on Their Degree. In *Link Prediction in Social Networks*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 15–25.
6. Oyama, S.; Hayashi, K.; Kashima, H. Cross-temporal link prediction. In Proceedings of the 2011 IEEE 11th International Conference on Data Mining, Vancouver, BC, Canada, 11–14 December 2011; pp. 1188–1193.
7. Slokom, M.; Ayachi, R. A New Social Recommender System Based on Link Prediction Across Heterogeneous Networks. In Proceedings of the International Conference on Intelligent Decision Technologies, Sorrento, Italy, 17–19 June 2017; pp. 330–340.
8. Kim, W.; Kwon, K.; Kwon, S.; Lee, S. The identification power of smoothness assumptions in models with counterfactual outcomes. *Quantit. Econ.* **2018**, *9*, 617–642. [[CrossRef](#)]
9. Liben-Nowell, D.; Kleinberg, J. The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.* **2007**, *58*, 1019–1031. [[CrossRef](#)]
10. Lü, L.; Zhou, T. Link prediction in complex networks: A survey. *Phys. A Stat. Mech. Appl.* **2011**, *390*, 1150–1170. [[CrossRef](#)]
11. Pech, R.; Hao, D.; Pan, L.; Cheng, H.; Zhou, T. Link prediction via matrix completion. *EPL (Europhys. Lett.)* **2017**, *117*, 38002. [[CrossRef](#)]
12. Munasinghe, L.; Ichise, R. Time aware index for link prediction in social networks. In Proceedings of the International Conference on Data Warehousing and Knowledge Discovery, Toulouse, France, 29 August–2 September 2011; pp. 342–353.
13. Yasami, Y.; Safaei, F. A novel multilayer model for missing link prediction and future link forecasting in dynamic complex networks. *Phys. A Stat. Mech. Appl.* **2018**, *492*, 2166–2197. [[CrossRef](#)]
14. Kostakos, V. Temporal graphs. *Phys. A Stat. Mech. Appl.* **2009**, *388*, 1007–1023. [[CrossRef](#)]

15. Alhajj, R.; Rokne, J. *Encyclopedia of Social Network Analysis and Mining*; Springer: Berlin/Heidelberg, Germany, 2014.
16. Casteigts, A.; Flocchini, P.; Quattrociocchi, W.; Santoro, N. Time-varying graphs and dynamic networks. *Int. J. Parallel Emerg. Distrib. Syst.* **2012**, *27*, 387–408. [\[CrossRef\]](#)
17. Hua, T.D.; Nguyen-Thi, A.T.; Nguyen, T.A.H. Link prediction in weighted network based on reliable routes by machine learning approach. In Proceedings of the 2017 4th NAFOSTED Conference on Information and Computer Science, Hanoi, Vietnam, 24–25 November 2017; pp. 236–241.
18. Zhou, J.; Huang, D.; Wang, H. A dynamic logistic regression for network link prediction. *Sci. China Math.* **2017**, *60*, 165–176. [\[CrossRef\]](#)
19. Tabourier, L.; Bernardes, D.F.; Libert, A.S.; Lambiotte, R. RankMerging: A supervised learning-to-rank framework to predict links in large social networks. *Mach. Learn.* **2019**, *108*, 1729–1756. [\[CrossRef\]](#)
20. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*; OTexts: Melbourne, Australia, 2018.
21. Hyndman, R.; Koehler, A.B.; Ord, J.K.; Snyder, R.D. *Forecasting with Exponential Smoothing: The State Space Approach*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2008.
22. Divakaran, A.; Mohan, A. Temporal Link Prediction: A Survey. *New Gener. Comput.* **2019**. doi:10.1007/s00354-019-00065-z. [\[CrossRef\]](#)
23. Özcan, A.; Ögüdücü, Ş.G. Multivariate temporal link prediction in evolving social networks. In Proceedings of the 2015 IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS), Las Vegas, NV, USA, 28 June–1 July 2015; pp. 185–190.
24. Lorrain, F.; White, H.C. Structural equivalence of individuals in social networks. *J. Math. Soc.* **1971**, *1*, 49–80. [\[CrossRef\]](#)
25. Worth, D. Introduction to modern information retrieval. *Aust. Acad. Res. Libr.* **2010**, *41*, 305–306. [\[CrossRef\]](#)
26. Jaccard, P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull. Soc. Vaudoise Sci. Nat.* **1901**, *37*, 547–579.
27. Sorensen, T.A. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biol. Skar.* **1948**, *5*, 1–34.
28. Ravasz, E.; Somera, A.L.; Mongru, D.A.; Oltvai, Z.N.; Barabási, A.L. Hierarchical organization of modularity in metabolic networks. *Science* **2002**, *297*, 1551–1555. [\[CrossRef\]](#)
29. Molloy, M.; Reed, B. A critical point for random graphs with a given degree sequence. *Random Struct. Algorithms* **1995**, *6*, 161–180. [\[CrossRef\]](#)
30. Adamic, L.A.; Adar, E. Friends and neighbors on the web. *Soc. Netw.* **2003**, *25*, 211–230. [\[CrossRef\]](#)
31. Zhou, T.; Lü, L.; Zhang, Y.C. Predicting missing links via local information. *Eur. Phys. J. B* **2009**, *71*, 623–630. [\[CrossRef\]](#)
32. Barabási, A.L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509–512. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Lü, L.; Jin, C.H.; Zhou, T. Similarity index based on local paths for link prediction of complex networks. *Phys. Rev. E* **2009**, *80*, 046122. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **1953**, *18*, 39–43. [\[CrossRef\]](#)
35. Leicht, E.A.; Holme, P.; Newman, M.E. Vertex similarity in networks. *Phys. Rev. E* **2006**, *73*, 026120. [\[CrossRef\]](#)
36. Liu, W.; Lü, L. Link prediction based on local random walk. *EPL (Europhys. Lett.)* **2010**, *89*, 58007. [\[CrossRef\]](#)
37. Vragović, I.; Louis, E. Network community structure and loop coefficient method. *Phys. Rev. E* **2006**, *74*, 016105. [\[CrossRef\]](#)
38. Klein, D.J.; Randić, M. Resistance distance. *J. Math. Chem.* **1993**, *12*, 81–95. [\[CrossRef\]](#)
39. Jeh, G.; Widom, J. SimRank: A measure of structural-context similarity. In Proceedings of the eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, AB, Canada, 23–26 July 2002; pp. 538–543.
40. Fouss, F.; Pirotte, A.; Renders, J.M.; Saerens, M. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Trans. Knowl. Data Eng.* **2007**, *19*, 355–369. [\[CrossRef\]](#)

41. Sun, D.; Zhou, T.; Liu, J.G.; Liu, R.R.; Jia, C.X.; Wang, B.H. Information filtering based on transferring similarity. *Phys. Rev. E* **2009**, *80*, 017101. [[CrossRef](#)]
42. Chebotarev, P.Y.; Shamis, E. A matrix-forest theorem and measuring relations in small social group. *Avtomatika i Telemekhanika* **1997**, *58*, 125–137.
43. Boccaletti, S.; Bianconi, G.; Criado, R.; Del Genio, C.I.; Gómez-Gardenes, J.; Romance, M.; Sendina-Nadal, I.; Wang, Z.; Zanin, M. The structure and dynamics of multilayer networks. *Phys. Rep.* **2014**, *544*, 1–122. [[CrossRef](#)]
44. Paranjape, A.; Benson, A.R.; Leskovec, J. Motifs in temporal networks. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, Cambridge, UK, 6–10 February 2017; pp. 601–610.
45. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
46. Li, Z.; Ren, T.; Ma, X.; Liu, S.; Zhang, Y.; Zhou, T. Identifying influential spreaders by gravity model. *Sci. Rep.* **2019**, *9*, 8387. [[CrossRef](#)] [[PubMed](#)]
47. Chen, D.; Kong, L.; Wang, D.; Huang, X.; Fang, B. TNLCD: A Feasible Algorithm for Local Community Discovery in Temporal Networks. In *FSDM*; IOS Press: Amsterdam, The Netherlands, 2018; pp. 459–464.
48. Wang, P.; Xu, B.; Wu, Y.; Zhou, X. Link prediction in social networks: The state-of-the-art. *Sci. China Inf. Sci.* **2015**, *58*, 1–38. [[CrossRef](#)]
49. Zachary, W.W. An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* **1977**, *33*, 452–473. [[CrossRef](#)]
50. Lusseau, D.; Schneider, K.; Boisseau, O.J.; Haase, P.; Slooten, E.; Dawson, S.M. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behav. Ecol. Sociobiol.* **2003**, *54*, 396–405. [[CrossRef](#)]
51. Tsvetovat, M.; Kouznetsov, A. *Social Network Analysis for Startups: Finding Connections on the Social Web*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2011.
52. Newman, M.E. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8577–8582. [[CrossRef](#)]
53. Girvan, M.; Newman, M.E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [[CrossRef](#)]
54. Watts, D.J.; Strogatz, S.H. Collective dynamics of ‘small-world’ networks. *Nature* **1998**, *393*, 440. [[CrossRef](#)]
55. Hu, H.B.; Wang, X.F. Unified index to quantifying heterogeneity of complex networks. *Phys. A Stat. Mech. Appl.* **2008**, *387*, 3769–3780. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).