Temporal Link Prediction Using Cluster and Temporal Information Based Motif Feature

Yi Li, Yanlong Wen*, Peng Nie, Xiaojie Yuan
College of Computer and Control Engineering
Nankai University, China
Email: {liyi,wenyanlong,niepeng,yuanxiaojie}@dbis.nankai.edu.cn

***Abstract*—Predicting future link states of a temporal network from past evolution is an important task from application perspective. Different from static networks analysis, temporal link prediction considers the trends of the links as time goes by, which is proved can achieve higher performance than that only exploring the interactions at a specific moment. The key challenge of this task is to obtain effective features of the evolving network. We propose a novel supervised method CTMF based on motif feature, which can characterize nearly complete local topological characteristic of node-pairs in the temporal network. We also attach cluster information as well as temporal transition factor to capture the influence of global structure and network transition on each motif over time. An unsupervised learning strategy is used on a concatenation of the cluster and temporal information based motif features (CTMF) from the continuous sets of snapshots to capture a better feature representation. We validate the effectiveness of the CTMF by utilizing it for link prediction on a number of temporal datasets, which shows our approach outperforms state-of-art methods using other features.**

I. Introduction

Link prediction aims to estimate the likelihood of the future linkage (interactions, contacts or edges) among node-pairs over the underlying network graph [1][2][3], which has significant applications in many fields. Link prediction in co-authorship networks like DBLP can suggest new possible collaborations for researchers. In friendship networks like Facebook and LinkedIn, link prediction tools can be used to recommend friends. In e-commerce sites, consumers can be recommended items likely to be purchased [4].

Majority of the literatures in link prediction consider various features drawn from a static snapshot of the network at a certain time. However real-world networks are often intrinsically dynamic, resulting from addition and deletion of nodes, occurrence of new links, and disappearance of existing interactions or contacts. The static approaches do not confer to the intrinsic dynamic nature of the real-world network, thus undermines prediction accuracy. Recent works show that models which consider temporality better capture link formation processes in the network [5].

Many existing temporal link prediction models compress a selection of simple information extracted from the time-varying graphs in theoretically or empirically guided ways, such as various topological similarities (common neighbors, Adamic-Adar, Jaccard coefficient). For example, [6] define a time series forecasting model to predict future similarity scores between nodes based on their past similarity score values. However these similarity based features sometimes can not capture high-quality underlying characteristics of networks, as it contains not ample but segmentary features of the evolving network.

Motifs [7] [8] are informative subgraph patterns recurring frequently in a network. It has been shown that motif frequencies often carry significant information about the local network structure in a variety of domains, which are widely adopted in various research fields such as community detection [9] and graph summerizing [10]. Given an undirected simple input graph $G = (V, E)$, a motif of size k nodes is defined as any subgraph $G_k \in G$ which consists of a subset of k nodes of the graph G. Figure 1 shows motif elements of size 4.

In this paper we propose a novel motif-based feature of both connected and disconnected node-pairs in temporal network, which is captured from local subgraph structures containing the node-pairs within specified sizes. Differ from aforementioned similarity measures, information about the nature of relationships is maintained as motif features in this way. Furthermore, instead of restricting only to the effect of local neighborhood, we integrate cluster structural effects and time-vary information of the motif features as to characterize the semi-global features as well as temporal transition factor. Then we utilize an unsupervised feature compression on the concatenation of the cluster and temporal information based motif features (CTMF) from the continuous sets of snapshots to capture a better feature representation with low dimensionality. After obtaining an appropriate temporal feature representation of the dynamic network, we use traditional supervised classification methods to predict link states in the future snapshot of the

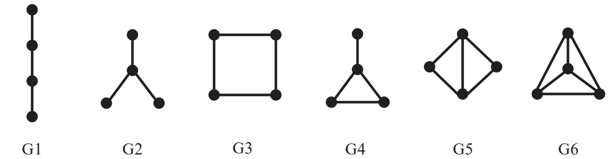


Fig. 1: Motif elements of size 4.

*Corresponding author.

978-1-5090-6014-6/18/$31.00 ©2018 IEEE

dynamic network.

The main contributions of the paper are :

- We propose a novel cluster and temporal information based motif feature (CTMF) of the node-pairs, which captures almost complete local topological characteristic as well as semi-global and time-varying features of the temporal network.
- We utilize an unsupervised learning strategy on time-combined CTMF features for an effective temporal feature representation, which contributes to the following supervised link prediction.
- We conduct experiments over three real-world networks. The experimental results show that the proposed approach achieve significant improvement in all metrics on other competed methods.

The remainder of the paper is organized as follows. Section II discuss related work on temporal link prediction. We explain the details of our methods in Section III. In Section IV we present datasets and experimental setting. We also report and discuss the results of our experiments. Finally, the conclusion is drawn and possible directions for future work are outlined.

II. Related Work

Among several techniques to address the link prediction problem, the most widespread ones explore the topological/structural patterns from the social network of interest, which can be categorized basically as unsupervised and supervised techniques [11]. The unsupervised techniques examine topological similarity measures and non-connected node-pairs are ranked according to a selected measure such as the number of common neighbors [12], Adamic/Adar [13], Katz coefficient [14] and so on. Then top-ranked node-pairs are estimated to be linked in the future. In supervised techniques [15], link prediction is treated as a binary classification problem. The pairs of nodes are determined as the instances of the classification task and each node pairs can be represented as a vector of features. These features consist of different similarity measures. The supervised link prediction can be performed by using different supervised algorithms such as Decision Trees, Random Forests, Neural Networks, SVM, etc [16].

Most of the previous works on link prediction lay emphasis on using various static measures whereas there are few studies analyzing the role of temporality in the link prediction task. The dynamicity of social networks and time effect on link prediction task specify that time information should be used for achieving better results.

An unsupervised temporal measure called Time-score is defined in [17], which is an extension of Common neighbour (CN) measure and proved can predict future links more accurately than CN measures, since traditional neighborhood based methods cannot differentiate two pairs of nodes that share same number of common neighbors, but having different likelihoods of link formation. A temporal random-walk based method called T_Flow is proposed in [18], which computes information flow between nodes by considering link activeness which varies over time. However, the limitation of random-walk based measures is that they suffer from high execution time. [19] propose a supervised link prediction framework (HPLP) which combines various topological information such as node degree and common link predictors into a bagged random forests classification framework, which achieves beter performance compared to unsupervised models.

However, features that above method utilizes are segmentary information extracted from neighbor structure of node-pairs which neglect the effect of underlying characteristics on network evolution. Also previous methods only consider new link prediction for disconnected node-pairs, without recurring links of connected node-pairs. We propose a novel temporal feature based on motif elements to resolve abovemetioned limitations in temporal link prediction.

Motifs have been used successfully in static network analysis for significant applications. For a given network, [20] count the frequencies of various motifs in the network for designing a fingerprint of a biological graph. [9] use motif as a building block to conduct community detection process. While the above methods address motif based analysis of static networks, in our work, we utilize motif as features to capture underlying characteristics of the temporal network.

III. Methods

In this section, we will first introduce the problem setting, and then discuss the cluster and temporal information based motif features (CTMF) extraction and compression. Last we show the supervised classification process using temporal features to predict future link states.

Consider an undirected network $G(V, E)$, where V is the node set and E is the edge set. A dynamic network consists of a continuous set of snapshots $\mathbb{G} = \{G_1, G_2, ..., G_t\}$. t is the number of time stamps in observatin interval and $G_i(V_i, E_i)$ is a snapshot of $\mathbb{G}$ corresponding to time stamp $i : 1 \leqslant i \leqslant t$. Given a continuous set of snapshots $\mathbb{G} = \{G_1, G_2, ..., G_t\}$ of a evolving network, the link prediction problem in temporal network can be specified as the task of predicting whether a link will form between a node-pair (u, v) in G_{t+1}.

A. Initial Motif Feature Extraction

The fundamental challenge for dynamic link prediction is choosing an effective feature set for this task. Earlier works choose features by considering various similarity based features of node-pairs in different snapshots. In order to capture as much neighboring structure characteristic of the node-pairs as we can, we extract motifs of connected and disconnected node-pairs in each snapshot as features for temporal link prediction. The initial extracted motif features are various collections of motif elements of specific sizes involving target node-pairs. The position of the nodes are under consideration since it can be a remarkable topological characteristic to distinguish different motif features with same motif element structure.

For a connected node-pair, the motif features are extracted from its local subgraphs including the link between them,

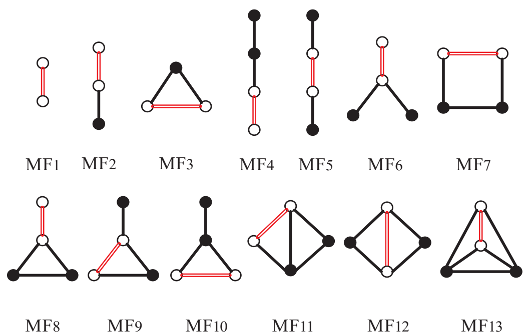


Fig. 2: Motif features of node-pairs within size 4.

considering the position of the link simultaneously. Figure 2 shows 13 distinguish motif features within size 4, including 9 kinds of motif elements. As we can see, the number of motif features is corresponding to the sum of the counts of isomerism edges in each motif element. As for disconnected node-pairs, our purpose is to obtain their neighboring structure characteristic which will reflect the topological feature around them. So we establish a virtual link between the separate node-pair in order to obtain motif features containing these two nodes. We use the same enumeration method as connected node-pairs. After the enumeration of motif features of certain sizes, the virtual link is deleted and the real local topological structures of the separate node-pair remain. As shown in figure 3(a), the virtual edges between node-pairs are represented by dashed line, and the real structures of motif features are shown by full line.

Consider a simple snapshot of a dynamic network shown in figure 3(a). We illustrate the process of the initial motif features extraction of connected node-pairs $(2, 3)$ and disconnected node-pairs $(2, 4)$ within size 4. Using the network motif discovery method introduced in [21], we can obtain fundamental motif elements of each node-pair within 4 nodes. Then we conduct a position estimate process to distinguish disparate motif features (MF_i) with same motif element structure, like MF_4 and MF_5 in figure 2. The total initial motif features of connected and disconnected node-pairs are displayed in the first and second table in figure 3(a).

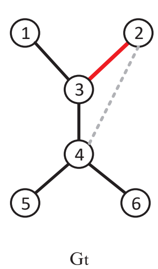
Notice that some motif features MF_i consist of several $mf_{i,j}$ that contain different node sets (like $mf_{2,1}^{(2,3)}$ and $mf_{2,2}^{(2,3)}$ in figure 3(a)). These $mf_{i,j}$ substantially carry various information about the node-pair from each other , which indicates that we should not simply count the frequency of them as the feature value of MF_i. Since motif feature concentrates on capturing local topology characteristics of the node-pairs, we evaluate the semi-global effect and temporal transition information by defining cluster concentration and temporal transition reward of each $mf_{i,j}$, which constitute our cluster and temporal information based motif feature (CTMF).

B. Cluster Concentration and Temporal Transition Factor of Motif Feature

1) Cluster Concentration: Initial motif feature can provide almost complete local topological information of the node-pairs at each snapshot. Since many literature prove that global measures can provide higher accuracy than local measures in static link prediction [1] [3], we define a cluster concentration measure of the initial motif feature to combine local feature with global cluster information, by means of awarding nodes that share more clusters with target node-pairs. We firstly apply a partitioning scheme to divide each snapshot of the temporal network into communities. After initial motif features extraction of node-pairs in every snapshot, we compare the cluster set that every node in the motif belongs to with the target node-pairs. The cluster concentration of a motif $mf_{i,j}$ of a node-pair (a, b) at time stamp t is defined as :

$$CC_t(mf_{i,j}^{(a,b)}) = \sum_{v \in mf_{i,j}^{(a,b)}} \frac{|C(v) \cap C(a) \cap C(b)|}{\max\limits_{v \in V_t} |C(v)| \left| mf_{i,j}^{(a,b)} \right|} \quad (1)$$

$C(i)$ represents the cluster set that node i belongs to. Since social networks exhibit inherent community property, edges are more likely to be formed within communities rather than between communities. If a motif consists of nodes within a community, we can identify this motif as an intensive motif whose structure may carry more valuable information compared with those containing nodes sharing disjoint communities.



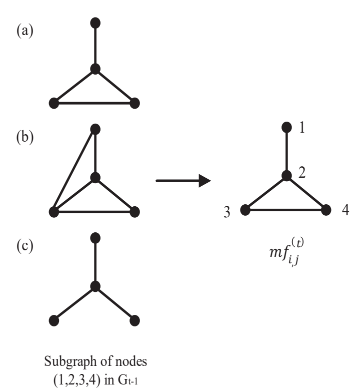
Initial motif features of connected node-pair (2,3)

Nodes	Motifs	Features	
23		mf1,1	MF1
132		mf2,1	MF2
423		mf2,2	
2345		mf4,1	MF4
2346		mf4,2	
2314		mf6,1	MF6

Initial motif features of disconnected node-pair (2,4)

Nodes	Motifs	Features	
24		mf1,1	MF1
245		mf2,1	MF2
246		mf2,2	
243		mf3,1	MF3
2456		mf6,1	MF6
5423		mf9,1	MF9
6423		mf9,2	
1324		mf10,1	MF10

(a) Initial motif feature extraction of two node-pairs in a snapshot Gt.



(b) Examples of temporal transition computation.

Fig. 3: Examples of initial motif feature extraction process and temporal transition reward computation.

2) Temporal Transition Factor: The above local and cluster features are both static characteristics extracted from each snapshot of the temporal network. The final temporal motif feature is a concatenation of all MF at each time stamp, without considering every single $mf_{i,j}$ transition between snapshots. As time goes on, if the structure of a certain $mf_{i,j}$ of a node-pair remains or changes, the underlying impact of this motif on temporal network will reveal. To this end, we take the temporal transition of each $mf_{i,j}$ into account. Firstly we define three kinds of temporal events of edges: persistent, emerging and vanishing, defined below.

- Persistent: A persistent event arise when the interaction between two nodes is not dropped when network evolves.
- Emerging: Emerging events represent the creation of a new link between two nodes in contiguous snapshots.
- Vanishing: Vanishing events are opposite to emerging ones. This kind of event represents the removal of an existing link between two nodes from a snapshot to its subsequent.

We define non-negative reward p, e and negative reward v to persistant, emerging and vanishing edges respectively. We compare the edge set of $mf_{i,j}$ at a certain snapshot with the subgraph containing same node set at previous snapshot and attach corresponding reward to each type of edges. The total temporal transition reward of a $mf_{i,j}$ at time stamp t is defined as:

$$TT_t(mf_{i,j}^{(a,b)}) = \frac{p\left|E_t^p(mf_{i,j})\right| + e\left|E_t^e(mf_{i,j})\right| + v\left|E_t^v(mf_{i,j})\right|}{\left|E_t^p(mf_{i,j}) \cup E_t^e(mf_{i,j}) \cup E_t^v(mf_{i,j})\right|} \tag{2}$$

Where $E_t^p(mf_{i,j})$, $E_t^e(mf_{i,j})$ and $E_t^v(mf_{i,j})$ represents persistant, emerging and vanishing edge set of $mf_{i,j}(a,b)$ at snapshot t respectively. Figure 3(b) illustrates three examples of temporal transition reward computation of a $mf_{i,j}$ at snapshot t. $mf_{i,j}^{(t)}$ is a motif feature extracted at stamp t, and three subgraphs of G_{t-1} containing same node set as $mf_{i,j}^{(t)}$ on the left side denotes three different evolving cases. By Eq. (2), we can obtain respective $TT_t(mf_{i,j})$ in three cases: p for case(a), $(4p+v)/5$ for case(b) and$(3p+e)/4$ for case(c). In our experiments, the number of parameters to define was reduced to two by making the values of p and v proportional to e (in this case, e was set to 1 for the sake of simplicity).

3) CTMF: Each eigenvalue of cluster and temporal information based motif feature (CTMF) of a node pair (a,b) at snapshot t is defined as:

$$CTMF_{t,i}^{(a,b)} = \sum_{mf_{i,j}^{(a,b)} \in MF_i^{(a,b)}} (CC_t(mf_{i,j}^{(a,b)})+1) \cdot TT_t(mf_{i,j}^{(a,b)}) \tag{3}$$

And we normalize $CTMF_t^{(a,b)}$ vector of each node-pair by the largest value of all $CTMF_{t,i}^{(a,b)}$. Here $CTMF_t^{(a,b)}$ for linked and separate node-pairs are represented by two feature vectors: $CTMF_t^+$ for connected ones, and $CTMF_t^-$ for disconnected ones. The eventual $CTMF_t^{(a,b)}$ for each node-pair contains both of them. The eigenvalue of $CTMF_t^+$ for disconnected node-pairs is all 0, so do $CTMF_t^-$ for connected ones. Then we aggregate $CTMF_t^{(a,b)}$ in all snapshots by concatenating to get final temporal feature representation $CTMF_{[1,t-1]}^{(a,b)}$ of the evolving network.

$$CTMF_{[1,t-1]}^{(a,b)} = CTMF_1^{(a,b)}||CTMF_2^{(a,b)}||...||CTMF_{t-1}^{(a,b)}$$

We also attach link states of node-pairs $LS_{[1,t-1]}^{(a,b)}$ in past snapshots to the feature vector. $LS_{[1,t-1]}^{(a,b)} = E_1(a,b)||E_2(a,b)||...||E_{t-1}(a,b)$. $E_t(a,b) = 1$ if there is a link between node-pair (a,b) in G_t, and 0 otherwise. Finally, the feature vector of node-pair (a,b), $e_{[1,t-1]}^{a,b}$, is the concatenation of CTMF feature set ($CTMF_{[1,t-1]}^{(a,b)}$) and LS feature set ($LS_{[1,t-1]}^{(a,b)}$).

$$e_{[1,t-1]}^{(a,b)} = CTMF_{[1,t-1]}^{(a,b)}||LS_{[1,t-1]}^{(a,b)}$$

Feature representation of all node-pairs in temporal network:

$$\begin{aligned} E_{train} &= \{e_{[1,t-1]}^{a,b}\}_{a,b \in V} \\ E_{test} &= \{e_{[2,t]}^{a,b}\}_{a,b \in V} \end{aligned} \tag{4}$$

Here, E_{train} and E_{test} corresponds to the training and test dataset. E_{train} and E_{test} are matrices of dimensions (m,k). m denotes all node-pairs in the temporal network. The size of the feature vector is $k = \left|e_{[1,t-1]}^{a,b}\right| = 2N(t-1)+t-1$, where N is the total number of distinct motif features that we consider. Here $N = 70$ since we extract motif features within 5 nodes.

C. Unsupervised Feature Compression

As a combination of features extracted from each timestamp of the evolving graph, the final temporal feature representation could contain entire characteristic from time span perspective. But simple as the feature aggregation method it use, it may not capture the most accurate temporal feature and also incorporate superfluous attributes. We address this issue by conducting an unsupervised learning process on the time-combined features. Unsupervised learning methods are often used to produce pre-processors and feature extractors for image and audio analysis systems [22]. Here we utilize it on temporal feature of the evolving graph to obtain a preferable feature representation.

The unsupervised learning method aims to find a linear transformation φ which reduces feature vectors e of k-dimension to feature vectors y of h-dimension (where $h < k$) with the minimum mean squared error to reconstruct e from y. The compressed feature vector y preserves essential information contained in e such that it can reconstruct e almost inerrably, thus we consider y as proper temporal features for supervised link prediction. The compression and reconstruction process can be accomplished by the encoder and decoder module, which are designed as Eq.5 and Eq.6. Encoder can be understood as a nonlinear transformation of linear combination of the input features, and decoder performs the reverse operation to encoder.

$$y = f(\varphi^{(c)} e + b^{(c)}) \tag{5}$$

$$\widehat{e} = f(\varphi^{(r)} y + b^{(r)}) \tag{6}$$

y is the compressed feature vector using the encoder, and $\widehat{e}$ is the reconstructed feature vector produced by the decoder. $\varphi^{(c)}$ is a (h, k) dimensional matrix representing the weight matrix for encoding and $\varphi^{(r)}$ represents the weight matrix for decoding of dimensions (k, h). $b^{(c)}$ and $b^{(r)}$ represent biases. $f(\cdot)$ is designed as the sigmoid function. The mean squared error of reconstruction $J(\varphi, b)$ is :

$$J(\varphi, b) = \frac{1}{2m} \sum_{e \in e_{[1,t]}} \frac{1}{2} \left\| \widehat{e} - e \right\|^2 \tag{7}$$

Here, $\|\cdot\|$ represents the norm squared value. The desired parameters are obtained by minimizing the reconstruction error. In the training process of parameters, we firstly initialize the parameters with random values, and utilize gradient descent based method to achieve a local minimum within limited number of iterations. Then with Eq.5, the compressed temporal feature representation y of node-pairs are obtained using these trained parameters. Choosing LBFGS algorithm for training process, two steps are conducted in each iteration: calculating descend value of the loss function and updating parameter. Time complexity is $O(kh)$ for a node-pair, where k and h is the length of initial and compressed feature, and the time complexity in one iteration is $O(mkh)$.

D. Supervised Temporal Link Prediction Model

Supervised temporal link prediction is treated as a binary classification problem, where linked node-pairs correspond to positive class and separated node-pairs correspond to negative class. For training the classifier, the continuous snapshots $[1, t-1]$ are regarded as training set and snapshots $[2, t]$ as testing set. After obtaining the compressed feature representation of the temporal network, several classification algorithms are used to predict link state at G_{t+1}. The pseudo-code of our CTMF temporal link prediction model is given in Algorithm 1.

Algorithm 1 CTMF Temporal Link Prediction Model

Input: $\mathbb{G}$, stamps t, compression size h, TT factors p, e, v
Output: $\widehat{G}_{t+1}$: predicted link state of G_{t+1}

for all $t' \in t$ **do**
  Enumerate all motif elements of each node-pair
  Get $mf_{i,j}^{(a,b)}$ by node-pair position estimate process
  Get $CC_{t'}(mf_{i,j}^{(a,b)})$ and $TT_{t'}(mf_{i,j}^{(a,b)})$ using Eq.1 2
  Get $CTMF_{t'}^{(a,b)}$ using Eq.3
end for
Get E_{train} and E_{test} using Eq.4
Learn optimal parameters (φ, b) with h
Get $y_{[1,t-1]}$ and $y_{[2,t]}$ using Eq.5
Train classifier C using $y_{[1,t-1]}$ and G_t
Predict $\widehat{G}_{t+1}$ using C and $y_{[2,t]}$
return $\widehat{G}_{t+1}$

IV. Experiment

A. Dataset and Experimental Setup

In order to evaluate the performance of the proposed method, data from three real-world dynamic networks were used in our experiments : UCI Msg, DBLP and Facebook. Table I lists the three real-world datasets used in our experiments.

TABLE I: Datasets

	Vertices	Avg(Edge)	Avg(Density)	$\lvert T \rvert$
UCI Msg	709	1782	7.1×10^{-3}	8
DBLP	1620	3016	2.3×10^{-3}	10
Facebook	2142	9631	4.2×10^{-3}	9

- UCI Msg is a network created from an online community [23]. This network dataset covers the period from April to November 2004. The nodes represent students at the University of California, Irvine and edges represent online messages sent between pairs of students.
- DBLP dataset [24] is a bibliographic database which consists of research publications of 28 conferences in the fields of Data Mining, Databases and Machine Learning held during the years 2003 to 2012. Each vertex in this dataset is an author and the edges represent co-authorship.
- Facebook dataset [25] contains wall-to-wall post relationship among users between August 2015 and April 2016. The nodes of the network are from the New Orleans Region. For each snapshot, there is a link between two nodes if there is a wall publication between them.

We generate overlapping communities of each snapshot using the Louvain method for greedy modularity optimization (Mod)[26]. The iteration time in feature compression strategy is limited to 50 or reaching a local convergence. For supervised temporal link prediction we use AdaBoost and Support Vector Machine (SVM) as classifiers. We conduct experiments with compression size h from 50 to 800. The disparity of performance are tiny and less than 3.6% in PRAUC. Moreover $h = 200$ and 400 obtain most of the best results.

Competing Methods. We compare our CTMP model against supervised HPLP method [19] and unsupervised time series methods TS [6]. HPLP is a supervised link prediction framework which combines simple topological information such as node degree and common link predictors into a bagged random forests classification framework. TS is an unsupervised method where several local topological similarity metrics and historical link states of node-pairs are adopted in ARIMA model to predict future similarity scores for unsupervised link prediction. We choose three similarity metrics with best performance: Common Neighbors(TS-CN), Adamic-Adar(TS-AA) and Jaccard Coefficient(TS-JA). Additionally, we also compare our method with unsupervised feature compression (called $CTMF$) to the method without unsupervised feature compression (called $CTMF_k$).

Evaluation Metrics. Area Under Receiver Operating Characteristic (AUC) curve is used commonly for evaluating the performance of a classifier. However, most of real-world networks are imbalanced, with very few linked node-pairs considering all node-pairs. Area under Precision-Recall curve (PRAUC) [27] is considered a more relevant measure by the authors of [28][29] for imbalanced classification problems, since true negative metric with little valuable information is not taken into account [30]. We also use an information retrieval metric, Normalized Discounted Cumulative Gain (NDCG), which is widely used in recommendation systems. NDCG evaluates the relevance of the predicted links in recommened list of length p , the value of which varies from 0.0 to 1.0. Here we set $p = 50$.

B. Performance

1) Effect of Temporal Transition Factors: To analyse the effect of temporal transition factors on our CTMF method, we set emerging reward $e = 1.0$ and vary persistant reward p and vanishing reward v proportionally to e: $p = \{0.0, 0.25, 0.5, 0.75, 1.0, 2.0\}$ and $v = \{-2.0, -1.0, -0.75, -0.5, -0.25\}$. Table II shows the PRAUC value obtained by each combination of the factor values in all three datasets using SVM as classifier with $h = 200$. The best four observed results in each dataset are presented in bold.

We can find that the optimal combinations of factor values in three datasets are fairly close to each other, as the best observed results gather to similar area in the tables of three datasets. Also, we can discovery that persistent and vanishing rewards can balance each other, since greater the value difference between them are, the lower PRAUC value they get. The best observed results were achieved when $p \leqslant e$ in most cases. As for vanishing reward v, smaller absolute factor values always perform much better than larger ones. We assume that vanishing events of edges may not contain as much valuable information as persisting and emerging ones, which makes vanishing factor less important to link prediction task. The CTMF method in following experiments follow the best combination of parameters of each dataset.

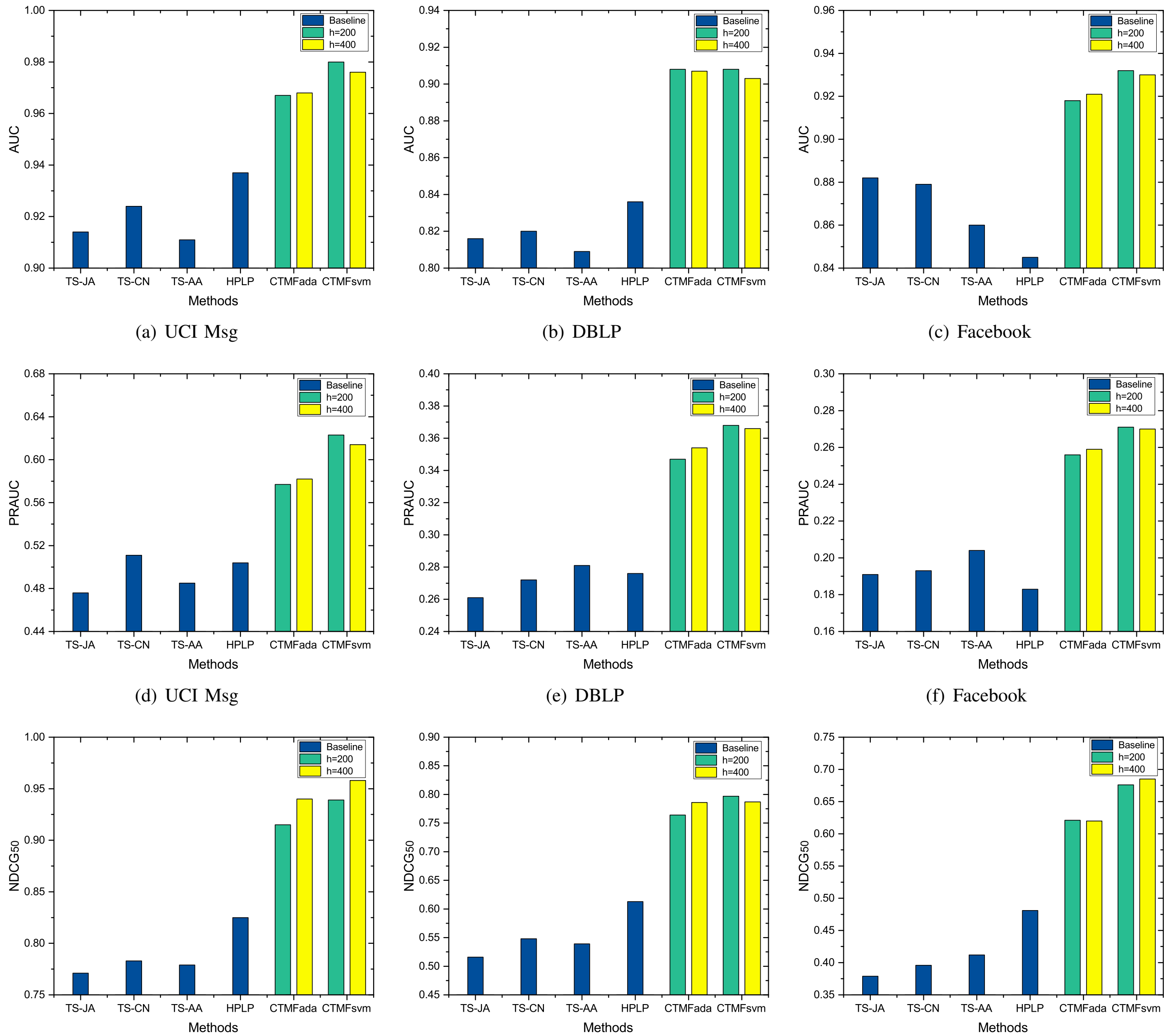
2) Performance compared with baselines: Performance comparison of CTMF with several competing methods in three datasets using AUC, PRAUC and $NDCG_{50}$ metrics is presented in Figure 4. In each chart, the first 4 bars are baseline methods: first three corresponding to time series based methods, the next representing HPLP method; and the final four bars represent $CTMF_{Ada}$ and $CTMF_{SVM}$ both with $h = 200$ and $h = 400$.

We observe that our CTMF methods using AdaBoost and SVM both outperform the competing methods with respect to all evaluating metrics in three datasets. The improvement in PRAUC metric (Figure 4(d)-4(f)) of our CTMF model on competing methods is more prominent than AUC and $NDCG_{50}$. As the most relevant metric for imbalanced classification problems, the performance improvement for PRAUC is a strong evidence of the superiority of CTMF methods over the baselines. The best PRAUC value achieved by unsupervised time series methods is 0.511, 0.281, 0.204 in UCI Msg, DBLP and Facebook dataset. While the best PRAUC values of $CTMF_{Ada}$ and $CTMF_{SVM}$ are 0.582, 0.354, 0.259 and 0.623, 0.368, 0.271, corresponding to 13.9%, 26.0%, 27.0% and 21.9%, 30.1%, 32.8% improvement respectively in three datasets. As for supervised HPLP method, the best improvement of our methods is 23.6%, 33.3% and 48.1% in three datasets from $CTMF_{SVM}$. The significant superiority of CTMF over all baseline methods proves that our motif based local topological feature combined with cluster and temporal information can better capture the underlying characteristics of the dynamic network than traditional similarity based metrics such as CN, AA and JA. Also we can notice that CTMF methods using SVM as classifier always achieve a higher PRAUC value than that using Adaboost in all datasets, and the optimal compression size for them are opposite: $h = 200$ for SVM and $h = 400$ for Adaboost. We can infer that SVM is an effcient classifier for temporal link prediction requiring low-dimensional feature representation.

TABLE II: PRAUCs obtained for several combinations of p and v

v / p	0.0	0.25	0.5	0.75	1.0	2.0
(a) UCI Msg						
-2.0	0.1972	0.2011	0.2159	0.2405	0.2641	0.2856
-1.0	0.3530	0.3783	0.3865	0.4036	0.4201	0.4280
-0.75	0.4613	0.4717	0.4873	0.4955	0.5102	0.4912
-0.5	0.5604	0.5748	0.5761	**0.5769**	0.5758	0.5631
-0.25	0.5628	0.5749	**0.5768**	**0.5774**	**0.5770**	0.5687
(b) DBLP						
-2.0	0.1051	0.1289	0.1443	0.1588	0.1737	0.1955
-1.0	0.2087	0.2294	0.2354	0.2458	0.2528	0.2541
-0.75	0.2761	0.2829	0.2946	0.3070	0.3101	0.3117
-0.5	0.3384	0.3461	**0.3474**	**0.3472**	0.3456	0.3361
-0.25	0.3381	0.3466	**0.3466**	**0.3467**	0.3462	0.3390
(c) Facebook						
-2.0	0.1421	0.1462	0.1608	0.1688	0.1773	0.1897
-1.0	0.1833	0.1877	0.1957	0.2026	0.2112	0.2120
-0.75	0.2161	0.2185	0.2268	0.2294	0.2349	0.2336
-0.5	0.2457	0.2554	**0.2561**	**0.2558**	0.2498	0.2373
-0.25	0.2530	0.2557	**0.2564**	**0.2563**	0.2499	0.2381

From the charts in the third row of figure 4, we can see CTMF method still perform better than other methods in $NDCG_{50}$ metric, which puts higher weight on the more relevant recommended links. CTMF methods outperform the best method of baselines by 16.1%, 30.0%, and 42.4% in $NDCG_{50}$ in three datasets. Its notable that compressing size for different classifiers of CTMF in $NDCG_{50}$ prefers the longer one, except $CTMF_{SVM}$ in DBLP dataset. It indicates that high relevant recommended links are better explained by longer feature representation, which encodes more information to represent the entire network. Finally we compare the AUC performance in the first row. The improvement in AUC of our CTMF methods on baseline methods comes between 4.6%



(a) UCI Msg (b) DBLP (c) Facebook

(d) UCI Msg (e) DBLP (f) Facebook

(g) UCI Msg (h) DBLP (i) Facebook

Fig. 4: Comparison between CTMF methods and competing methods.

to 8.6% in all three datasets, which is smaller compared with PRAUC and $NDCG_{50}$ metrics, due to its inadequacy in imbalanced classification problem.

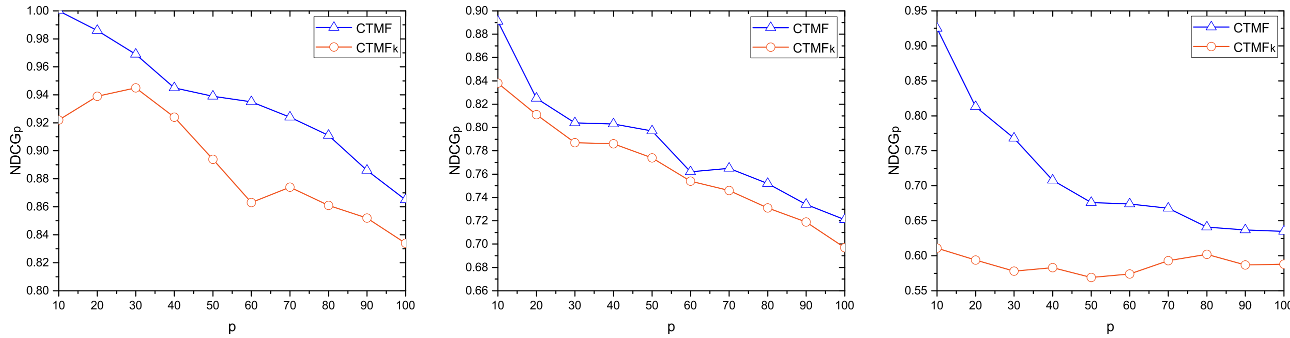
As for different CTMF methods, generally $CTMF_{SVM}$ achieves higher performance than $CTMF_{Ada}$ for all metrics in three datasets, except using AUC and $NDCG_{50}$ metrics in DBLP dataset, where two classifiers have comparable results. We assume that Adaboost is more suitable for collaboration network than communication and friendship networks due to different interaction ways of individuals, and SVM is capable for temporal link prediction tasks in most networks.

Concerning overall performance of all methods through all three datasets, we can find Facebook dataset always fall behind in all metrics, showing the challenge for temporal link prediction in Facebook dataset. Our CTMF methods achieve highest improvement over the best baseline method in Facebook among all three datasets.

3) Effect of Unsupervised Feature Compression: We also conduct another version of CTMF: $CTMF_k$, whose feature representation is the CTMF features without compression, and compare $CTMF_k$ with CTMF using $NDCG_p$ metric for various p values (using SVM as classifier with $h = 200$) in Figure 5. We can see that CTMF achieves higher $NDCG_p$ values than $CTMF_k$ in all the datasets. Comparing the $NDCG_{50}$ value of $CTMF_k$ with all baseline methods in each datasets in figure 4, we can conclude that $CTMF_k$ also outperforms all baseline methods. This indicates that CTMF features are more effective temporal features for link prediction than the similar based features.

V. Conclusion

In this paper, we present a novel supervised temporal link prediction method CTMF. Unlike existing similarity based methods, our method extracts almost complete local topological structures of the node-pairs as motif features. Cluster concentration and temporal transition information of the



(a) UCI Msg (b) DBLP (c) Facebook

Fig. 5: Performance comparison between $CTMF$ and $CTMF_k$.

motif features are attached to initial motif features to capture global and temporal characters of the temporal network. We conduct an unsupervised learning strategy for better feature representation, and a supervised classification for temporal link prediction. The experimental results on three real dynamic network datasets demonstrate significant advantages of CTMF over the existing approaches in various metrics. In future work, other baseline methods could be tested as well other metris for temporal link prediction. Also we will investigate other effective temporal features for link prediction in temporal networks.

Acknowledgment

This work is supported by National Natural Science Foundation of China under Grant No. 61772289 and 61702285.

References

[1] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *journal of the Association for Information Science and Technology*, vol. 58, no. 7, pp. 1019–1031, 2007.

[2] J. Valverde-Rebaza and A. de Andrade Lopes, "Exploiting behaviors of communities of twitter users for link prediction," *Social Network Analysis and Mining*, vol. 3, no. 4, pp. 1063–1074, 2013.

[3] L. Lü and T. Zhou, "Link prediction in complex networks: A survey," *Physica A: statistical mechanics and its applications*, vol. 390, no. 6, pp. 1150–1170, 2011.

[4] T. J. Lakshmi and S. D. Bhavani, "Temporal probabilistic measure for link prediction in collaborative networks," *Applied Intelligence*, vol. 47, no. 1, pp. 83–95, 2017.

[5] P. Wang, B. Xu, Y. Wu, and X. Zhou, "Link prediction in social networks: the state-of-the-art," *Science China Information Sciences*, vol. 58, no. 1, pp. 1–38, 2015.

[6] İ. Güneş, Ş. Gündüz-Öğüdücü, and Z. Çataltepe, "Link prediction using time series of neighborhood-based node similarity scores," *Data Mining and Knowledge Discovery*, vol. 30, no. 1, pp. 147–180, 2016.

[7] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, 2002.

[8] A. Paranjape, A. R. Benson, and J. Leskovec, "Motifs in temporal networks," pp. 601–610, 2016.

[9] A. Arenas, A. Fernandez, S. Fortunato, and S. Gomez, "Motif-based communities in complex networks," *Journal of Physics A: Mathematical and Theoretical*, vol. 41, no. 22, p. 224001, 2008.

[10] D. Koutra, U. Kang, J. Vreeken, and C. Faloutsos, "Summarizing and understanding large graphs," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 8, no. 3, pp. 183–202, 2015.

[11] A. Özcan and Ş. G. Öğüdücü, "Supervised temporal link prediction using time series of similarity measures," in *Ubiquitous and Future Networks (ICUFN), 2017 Ninth International Conference on*. IEEE, 2017, pp. 519–521.

[12] M. E. Newman, "Clustering and preferential attachment in growing networks," *Physical review E*, vol. 64, no. 2, p. 025102, 2001.

[13] L. A. Adamic and E. Adar, "Friends and neighbors on the web," *Social networks*, vol. 25, no. 3, pp. 211–230, 2003.

[14] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.

[15] M. Al Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link prediction using supervised learning," in *SDM06: workshop on link analysis, counter-terrorism and security*, 2006.

[16] H. R. Sa and R. B. Prudencio, "Supervised learning for link prediction in weighted networks," in *III International Workshop on Web and Text Intelligence*, 2010.

[17] L. Munasinghe and R. Ichise, "Time aware index for link prediction in social networks," in *International Conference on Data Warehousing and Knowledge Discovery*. Springer, 2011, pp. 342–353.

[18] L. Munasinghe, "Time-aware methods for link prediction in social networks," Ph.D. dissertation, Ph. D. thesis, The Graduate University for Advanced Studies, 2013.

[19] R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla, "New perspectives and methods in link prediction," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 243–252.

[20] N. Pržulj, D. G. Corneil, and I. Jurisica, "Modeling interactome: scale-free or geometric?" *Bioinformatics*, vol. 20, no. 18, pp. 3508–3515, 2004.

[21] T. Hocevar and J. Demsar, "A combinatorial approach to graphlet counting," *Bioinformatics*, vol. 30, no. 4, pp. 559–565, 2014.

[22] C. Poultney, S. Chopra, Y. L. Cun *et al.*, "Efficient learning of sparse representations with an energy-based model," in *Advances in neural information processing systems*, 2007, pp. 1137–1144.

[23] P. Panzarasa, T. Opsahl, and K. M. Carley, "Patterns and dynamics of users' behavior and interaction: Network analysis of an online community," *Journal of the Association for Information Science and Technology*, vol. 60, no. 5, pp. 911–932, 2009.

[24] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "Arnetminer: extraction and mining of academic social networks," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2008, pp. 990–998.

[25] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, "On the evolution of user interaction in facebook," in *Proceedings of the 2nd ACM workshop on Online social networks*. ACM, 2009, pp. 37–42.

[26] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[27] V. Raghavan, P. Bollmann, and G. S. Jung, "A critical investigation of recall and precision as measures of retrieval system performance," *ACM Transactions on Information Systems (TOIS)*, vol. 7, no. 3, pp. 205–229, 1989.

[28] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 233–240.

[29] R. Lichtnwalter and N. V. Chawla, "Link prediction: fair and effective evaluation," in *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*. IEEE Computer Society, 2012, pp. 376–383.

[30] T. J. Lakshmi and S. D. Bhavani, "Heterogeneous link prediction based on multi relational community information," in *Communication Systems and Networks (COMSNETS), 2014 Sixth International Conference on*. IEEE, 2014, pp. 1–4.