



# STWalk: Learning Trajectory Representations in Temporal Graphs

Supriya Pandhre

Indian Institute of Technology Hyderabad  
Hyderabad, India  
cs15mtech11016@iith.ac.in

Manish Gupta

Microsoft  
Hyderabad, India  
gmanish@microsoft.com

Himangi Mittal

Jaypee Institute of Information Technology  
Noida, India  
himangimittal@gmail.com

Vineeth N Balasubramanian

Indian Institute of Technology Hyderabad  
Hyderabad, India  
vineethnb@iith.ac.in

## ABSTRACT

Analyzing the temporal behavior of nodes in time-varying graphs is useful for many applications such as targeted advertising, community evolution and outlier detection. In this paper, we present a novel approach, STWalk, for learning trajectory representations of nodes in temporal graphs. The proposed framework makes use of structural properties of graphs at current and previous time-steps to learn effective node trajectory representations. STWalk performs random walks on a graph at a given time step (called *space-walk*) as well as on graphs from past time-steps (called *time-walk*) to capture the spatio-temporal behavior of nodes. We propose two variants of STWalk to learn trajectory representations. In one algorithm, we perform space-walk and time-walk as part of a single step. In the other variant, we perform space-walk and time-walk separately and combine the learned representations to get the final trajectory embedding. Extensive experiments on three real-world temporal graph datasets validate the effectiveness of the learned representations when compared to three baseline methods. We also show the goodness of the learned trajectory embeddings for change point detection, as well as demonstrate that arithmetic operations on these trajectory representations yield interesting and interpretable results.

## CCS CONCEPTS

• **Information systems** → *Data mining*; • **Human-centered computing** → *Social network analysis*;

## KEYWORDS

Representation Learning, Deep Learning, Temporal Graph Analysis

## ACM Reference Format:

Supriya Pandhre, Himangi Mittal, Manish Gupta, and Vineeth N Balasubramanian. 2018. STWalk: Learning Trajectory Representations in Temporal

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

CoDS-COMAD '18, January 11–13, 2018, Goa, India

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-6341-9/18/01...\$15.00

<https://doi.org/10.1145/3152494.3152512>

Graphs. In *CoDS-COMAD '18: The ACM India Joint International Conference on Data Science & Management of Data, January 11–13, 2018, Goa, India*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3152494.3152512>

## 1 INTRODUCTION

The rising use of social networks and various other sensor networks in real-world applications ranging from politics to healthcare has made computational analysis of graphs a very important area of research today. The use of machine learning for various graph analysis tasks such as community detection [16], link analysis [9], node classification [2] and anomaly detection [4] has exponentially increased over the last few years, considering the direct impact of the success of these methods on business outcomes. A wide variety of methods have been proposed in the aforementioned areas over the last few years. However, a large part of the efforts so far has focused on static graphs. Platforms such as Facebook, Twitter and LinkedIn correspond to graphs that change on a daily basis; however, the algorithms that analyze such graphs are primarily static, i.e., they consider a graph at a given snapshot of time for analysis. In this work, we seek to focus on temporal graphs, and on learning representations of node trajectories in such graphs.

Study of user-user interactions over time plays an important role in many applications, including user classification for targeted advertising or link prediction to suggest new connections on social networking sites. One of the crucial part of such applications is analyzing change in users' behavior over time. Aggarwal and Subbian [1] present a survey of methods that have been proposed so far to study temporal graphs, and categorize such methods into two kinds: *maintenance methods* that focus on adapting older models to a newer version of the same graph, or *analytical evolution analysis*, where the objective is to analyze the network as it changes itself. We propose to address the latter in this work. Recent efforts for analysis of temporal graphs have largely extended earlier methods proposed on static graphs, such as spectral methods or those that study statistical properties. In this work, we build on the recent emphasis on learning appropriate representations of data to temporal graphs. In particular, we propose SpaceTimeWalk (STWalk), which learns representations of node trajectories in temporal graphs. To the best of our knowledge, this is the first such effort.

STWalk is an unsupervised trajectory learning algorithm that embeds the spatial and temporal characteristics of nodes over a given time window. The algorithm carries out a random walk in

a current graph, resulting in a *space-walk*, as well as in graphs at different time steps, called *time-walk*. The traversed paths are then considered as sentences, with nodes as words from a vocabulary. The SkipGram [17] network is then employed to learn the latent representations of node trajectories such that it maximizes the probability of co-occurrence of two nodes within the specified window size. We perform extensive experimentation on three real-world time-varying networks. Performance on these datasets show the effectiveness of our framework in dealing with different kinds of temporal networks. We also study the usefulness of these representations through visualization, as well as complementary tasks such as change point detection.

The remainder of this paper is organized as follows. Section 2 reviews related work in relevant areas, highlighting the need for the proposed work. Section 3 presents the mathematical formulation of the problem, and the proposed methods are presented in Sections 3.1 and 3.2. The experimental results that study the performance of the proposed methods are described in Section 4, and Section 5 concludes the work with pointers to future directions.

## 2 RELATED WORK

Our work is most closely related to two sub-areas of graph analysis that have garnered increased interest in the last few years: network representation learning and temporal graph analysis. We review methods in both of these sub-areas below.

### 2.1 Network Representation Learning

In recent years, there have been a few approaches proposed for learning representations of networks, in particular, learning graph node embeddings. Many of the recent efforts have been inspired by the recent wave of deep learning methods that have demonstrated impressive success in learning representations of other kinds of data such as images, speech and text. DeepWalk [19], one of the earliest efforts in this regard, combined a random walk-based method with a SkipGram network (traditionally used in Natural Language Processing) to learn node representations. Node2vec [11] extended DeepWalk by employing biased random walks to learn node embeddings. In particular, Node2Vec precomputes a matrix, which is similar to a transition probability matrix, that assigns a probability of transition in a random walk, and is based on two hyper-parameters that control whether the walk will stay within a close neighborhood of the node or it will move farther away from that node.

Tang et al. (LINE) [24] computed node embeddings using a function of the probability of observing first-order and second-order nodes together in a random walk within a certain window size. This method then uses Asynchronous Stochastic Gradient Descent to solve the optimization problem. GraRep [5] employed a Singular Value Decomposition-based method for dimensionality reduction and learning graph representations. Cao et al. [6] proposed a stacked denoising autoencoder for learning the vertex representations in graphs, where they used random surfer model to learn the structural properties of graph calculate the positive pointwise mutual information and then employ denoising autoencoder to reduce the dimensions of learned node features. There have also been efforts, such as [7], which have used deep neural network-based models for learning embeddings in heterogeneous networks. The main goal of

the paper [7] was to consider information from multiple sources in dynamic network and learn unified representation. A more detailed review on network embedding techniques can be found in [10]. However, all the aforementioned methods work only with static graphs. In this work, we focus on learning representations of node trajectories in dynamic graphs. Dynamic graphs are evolving by nature, and it is important to capture the addition and deletion of nodes and edges, while capturing the behavior of a node over time in a learnt representation.

### 2.2 Temporal Graph Analysis

Related to the other focal subarea for this work, time-varying graphs have been studied for applications such as node classification [3, 13, 15], link prediction [20], community evolution [25] and outlier detection [12] in recent years. In one of the earlier works in this area, Tang et al. [25] proposed a spectral clustering framework for studying the evolution of communities in time varying multi-mode networks. Aggarwal et al. [3] proposed a node classification model that used network structure and node attributes from time-varying graphs. The model was specifically designed for textual node attributes. In [20], a non-parametric model was proposed for link prediction in a series of network snapshots. More recently, Li et al. [15] presented a framework that used node attributes for learning node embeddings. The model first learns separate embeddings from network structure and node attributes respectively. The node embeddings are then combined using matrix perturbation theory. Another recent work, GraphSAGE [13], proposed an inductive algorithm that generates node embeddings using sampling and aggregating local neighborhood information and learning a function to generate node embeddings of unseen data. For more such methods, Aggarwal et al. [1] provide a comprehensive survey on methods for temporal graph analysis.

While there have been a few efforts on analysis of time-varying graphs, all the aforementioned algorithms enforce the data to have node attributes. Many real-world graph data, however, may not provide node features either because they are not available publicly for use or they are missing (for instance, it is optional to fill age or location while creating a user profile on Twitter). In this work, we seek to propose a new framework, SpaceTimeWalk (*STWalk*), that seeks to utilize only the structural properties of time-varying graphs to learn compact low-dimensional embeddings capturing spatio-temporal properties of node trajectories.

We now present the proposed *STWalk* algorithm.

## 3 PROPOSED METHOD: STWALK

Learning the change in behavior of nodes in a temporal graph requires us to consider the information from the graph in the current time step, as well as graphs from previous time steps. Using the naïve way of examining graphs at each time-step separately will fail to capture the correlation information that exists between two graphs at consecutive time steps. Hence, we propose *STWalk* a random walk based framework that learns rich trajectory representations by capturing changes in node behavior across a given time window. We hence define the following to formulate our solution.

- Given a graph  $G = (V, E)$  at  $T$  different time steps  $\{G_1, G_2, G_3, \dots, G_T\}$  having  $N$  number of nodes at each time step and varying

number of edges  $E_t, t \in \{1, \dots, T\}$ , where  $E_t$  represents the edges in  $G_t$ .

- Let  $W$  denote the adjacency tensor of size  $N \times N \times T$ , representing the adjacency matrix of the graph at different time stamps.

Our goal is to learn representations  $\Phi$  that maps a given node to a  $d$ -dimensional representation. Let  $N_t(u) \subset V$  denote the set of neighbors of node  $u$  in graph  $G_t$ . For a node  $u$  in graph  $G_t$ , the representation  $\Phi_t(u)$  will be learned in such a way that, it maximizes the probability of observing the neighbors of  $u$  in graph  $G_t$ , viz.  $N_t(u)$ , and also the representation of  $u$  at previous time stamps  $\Phi_{t-\tau}(u)$  where  $\tau \in \{1, \dots, t-1\}$ . Equation (1) below formalizes this construction. Our objective hence is to obtain a  $d$ -dimensional representation for a node  $u$  at time  $t$  that maximizes the following log probability:

$$\max_{\phi_t} \sum_{u \in V} \log \Pr(N_t(u), \phi_{t-\tau}(u) | \phi_t(u)) \quad (1)$$

where  $\tau \in \{1, \dots, t-1\}$

We assume that the neighborhood nodes in  $N_t(u)$  and past representations  $\phi_{t-\tau}(u)$  are independent of each other, conditioned on the current representation of  $u$ ,  $\phi_t(u)$ . Hence, we write:

$$\begin{aligned} & \log(\Pr(N_t(u), \phi_{t-\tau}(u) | \phi_t(u))) \\ &= \log\left(\prod_{n_i \in N_t(u)} \Pr(n_i | \phi_t(u))\right) + \log(\Pr(\phi_{t-\tau}(u) | \phi_t(u))) \end{aligned} \quad (2)$$

Learning representations using random walk has proved to measure better graph proximity, and thereby improving the performance [10] [14]. Hence, we use random walk to learn the conditional probability of observing a node  $n_i$  given the learned representation  $\phi_t(u)$  defined as follows:

$$\Pr(n_i | \phi_t(u)) = \frac{\exp(\phi_t(n_i)\phi_t(u))}{\sum_{v \in N_t(u)} \exp(\phi_t(v)\phi_t(u))} \quad (3)$$

Therefore, to learn the trajectory representation, we maximize:

$$\max_{\phi_t} \sum_{u \in V} \log\left(\prod_{n_i \in N_t(u)} \frac{\exp(\phi_t(n_i)\phi_t(u))}{\sum_{v \in N_t(u)} \exp(\phi_t(v)\phi_t(u))}\right) + \log(\Pr(\phi_{t-\tau}(u) | \phi_t(u))) \quad (4)$$

While the above problem appears to be learning only the representation of a node at time  $t$ , it actually represents a representation of a node  $t$  considering its history over the graphs at the previous time steps, and hence, the term *node trajectory representation* in this work.

To solve the maximization problem defined in Equation 4, we propose two approaches. The first approach solves equation 4 by considering it as one single expression and learning a trajectory representation by maximizing the co-occurrence of neighborhood nodes at time  $t$  and in previous  $t - \tau$  time-steps together. In the second approach, we learn the trajectory representation by solving two sub-problems separately: (i) maximizing the co-occurrence probability of a node and its neighbors at time  $t$ , and (ii) maximizing the co-occurrence probability of a node and its neighbors at

preceding  $t - \tau$  time-steps. The representations learned using these steps are then combined to form the final trajectory representation. Sections 3.1 and 3.2 describe each of the proposed approaches.

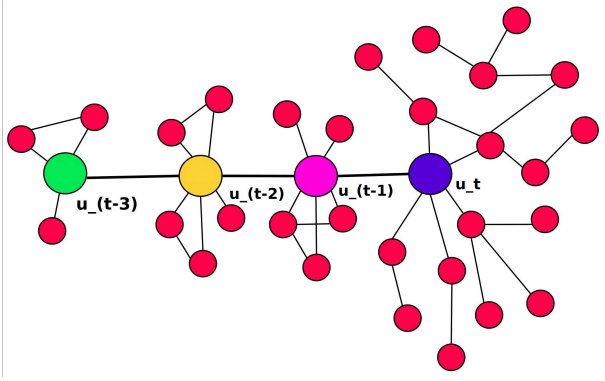
### 3.1 STWalk1

We describe the first approach, which we call STWalk1, in this section. For learning the trajectory representations of nodes in temporal graphs, here, we take into account the neighbors of a node at time steps  $t, t-1, \dots, t-\tau$  together and maximize the probability of observing these nodes together.

In STWalk1, to consider the neighbors of a node  $u$  from the current graph and graphs at previous time steps, we construct a Space-Time Graph. The space-time graph contains all nodes from graph at time  $t$  and it has a special temporal edge that links the node to itself in the graphs at previous time steps. Figure 1 illustrates this with an example of a space-time graph considered while learning the trajectory representation for node at time  $t$ ,  $u_t$ . In particular, the figure shows the temporal edge (shown as a bold line) between  $u_t$  and past self-nodes  $u_{t-1}, u_{t-2}$  and  $u_{t-3}$ . The trajectory representation thus learned for  $u_t$  will hence be influenced by all the nodes in the Space-Time graph, as shown in Figure 1, viz. nodes from the present graph as well as its preceding self nodes and their first-level neighbors. Given a node and the window size, over which we want to learn the trajectory, Algorithm 1 describes the steps to create a Space-Time Graph starting from the given node.

Now, to learn the representations of nodes in the Space-Time graph, we perform random walks on it. If two nodes share many edges or neighbors, they will be visited more often in many random walks, indicating that these two nodes share a similar graph structure. Hence, the representation learned for these two nodes should be close to each other in the embedding space. This is analogous to the concept from Natural Language Processing (NLP) that if two words co-occur in many sentences, this indicates that they represent a similar context and hence their word vectors are nearby each other in word embedding space. To learn such an embedding that captures the relationship of a word with other co-occurring words in a window, Mikolov et al. [17] presented the SkipGram network. We use a similar SkipGram algorithm (given in Algorithm 2) to learn the node embeddings from random walks of a graph, similar to DeepWalk [19]. For more details of the SkipGram network, we request the interested reader to refer to [17].

For each node in the Space-Time graph, we perform  $\rho$  number of random walks each of length  $\mathcal{L}$ . The problem of learning the trajectory representation in such a way that it maximizes probability of co-occurrence of present and past nodes is reduced to maximizing the co-occurrence of nodes in a random walk within a vocabulary window  $\mathcal{W}_v$ . Hence, we use the SkipGram algorithm 2 to learn the trajectory representations for nodes. (In particular, we use the SkipGram network, as proposed in [17] to implement the SkipGram algorithm.) For each node in a graph  $G_t$ , two representations will be learned. One representation corresponds to the node embedding when it occurs in the context of other nodes, and the other representation which is learned for the node itself. The latter one is used as the node trajectory representation. The overall algorithm STWalk1 is illustrated in Algorithm 3.



**Figure 1: Example graph of node  $u_t$  generated by Algorithm createSpaceTimeEmb 1. The node  $u_t$  represents node in current graph and nodes  $u_{t-1}$ ,  $u_{t-2}$  and  $u_{t-3}$  indicate its past self-nodes. To learn trajectory representation for  $u_t$ , STWalk1 algorithm considers the influence of nodes in current time step as well as effect from the past.**

---

**Algorithm 1: createSpaceTimeGraph**


---

**Input:** Set of graphs  $\mathcal{G}:\{G_t, G_{t-1}, \dots, G_{t-\tau}\}$ , Time window size  $\tau$ , starting node  $startNode$  at time  $t$

**Output:**  $G_{spacetime}$

```

1  $G_{spacetime} = G_t$ 
2 for each time step  $i \in \{0, \dots, \tau\}$  do
3    $pastNode = startNode$  at  $G_{t-i}$ 
4   Create  $pastNodeSubGraph$  = the subgraph of  $pastNode$ 
   and its direct neighbors from  $G_{t-i}$ 
5   Merge  $G_{spacetime}$  and  $pastNodeSubGraph$  to obtain
   updated  $G_{spacetime}$ 
6    $startNode = pastNode$ 
7 end
```

---



---

**Algorithm 2: SkipGram**


---

**Input:** Input representation:  $\Phi$ , List of nodes in a random walk:  $walk$ , windowSize:  $\mathcal{W}_v$

**Output:** Updated representation:  $\Phi$

```

1 for  $n \in walk$  do
2   for  $n_i \in walk[n - \mathcal{W}_v, n + \mathcal{W}_v]$  do
3      $J(\Phi[n]) = -\log \Pr(\Phi[n_i]|\Phi[n])$ 
4      $\Phi[n] = \Phi[n] - \frac{\partial J(\Phi[n])}{\partial \Phi[n]}$ 
5   end
6 end
```

---

### 3.2 STWalk2

In this section, we detail the second approach for learning the trajectory representation of nodes in a temporal graph. We solve the problem of learning representations of node trajectories, such that these representations are influenced by present and past neighbors, by dividing it into two sub-problems. We learn the spatial representation from the current graph,  $G_t$ , and the temporal representation

---

**Algorithm 3: STWalk1**


---

**Input:** Given node  $n \in [1, N]$ ,

time step  $t \in [1, T]$ ,

length of random walk:  $\mathcal{L}$ ,

temporal window size:  $\tau$ ,

vocabulary window size:  $\mathcal{W}_v$ ,

Set of graphs  $\mathcal{G}:\{G_t, G_{t-1}, \dots, G_{t-\tau}\}$ , size of embedding:  $d$ ,  
number of restarts (starts at same node):  $\rho$

**Output:** Updated  $n^{th}$  row of matrix of  $\Phi_t$  of size:  $N \times d$

```

1 Initialize  $\Phi_t[n] \in \mathbb{R}^d$ 
2 for  $i=0$  to  $\rho$  do
3    $new\_graph = createSpaceTimeGraph(\mathcal{G}, \tau, n)$ 
4    $spaceTimeWalk = randomWalk(new\_graph, n, \mathcal{L})$ 
5    $\Phi_t[n] = SkipGram(\Phi_t, spaceTimeWalk, \mathcal{W}_v)$ 
6 end
```

---

from past graphs  $G_{t-1}, \dots, G_{t-\tau}$ , separately. The final trajectory representation is obtained by combining the spatial and temporal representations of a given node.

The STWalk2 algorithm first performs a random walk in the current time step graph  $G_t$ , called as *SpaceWalk*, and learns the spatial representation for each node using the SkipGram algorithm.  $\Phi_{u,:}^{space}$  corresponds to the learned spatial representation of node  $u$ . To learn the temporal representation of a node, STWalk2 constructs  $G_{new}$ , which contains only one node from current graph and neighbors from past graphs. For example, in order to construct such a  $G_{new}$  in Figure 1, we remove all neighbors from current time-step graph of node  $u_t$ . The remaining graph will have  $u_t$  linked only to past subgraphs of  $u_{t-1}$ ,  $u_{t-2}$  and  $u_{t-3}$ . This will be used as  $G_{new}$  to learn the temporal representation (in a manner similar to learning the spatial representation above).

The random walks on  $G_{new}$  capture the temporal relationship between the current node and nodes from its past graphs.  $\Phi_{u,:}^{time}$  corresponds to the learned temporal representation of node  $u$ . Finally, the two learned embeddings are combined to get trajectory representation of nodes. Algorithm 4 summarizes this approach. Line 15 of Algorithm 4 states that the final trajectory representation is a function of spatial and temporal representations. In this work, we found vector addition to work very well in practice in the experiments.

Thus, in STWalk2, we ensure that the trajectory representation of the current node  $u_t$  has influence from past nodes by learning the temporal representation explicitly. Learning the temporal representation separately forces the random walk to explore the past graphs of the same node and capture the temporal information between the current node and its past. Therefore, STWalk2 is able to learn richer trajectory embeddings than STWalk1 because in STWalk1 we consider the present and past neighbors together, giving the random walks an option of skipping few nodes from past graphs. Experimental results in Table 2 validate that the trajectory embeddings learned by STWalk2 perform better than embeddings learned by STWalk1. We now present details of our experiments.

**Algorithm 4:** STWalk2

---

**Input:** Given node  $n \in [1, N]$ ,  
time step  $t \in [1, T]$  Set of graphs  $\mathcal{G}:\{G_t, G_{t-1}, \dots, G_{t-\tau}\}$ ,  
length of spatial random walk :  $\mathcal{L}_\alpha$ ,  
length of temporal random walk :  $\mathcal{L}_\beta$ ,  
temporal window size:  $\tau$ ,  
vocabulary window size:  $\mathcal{W}_v$ ,  
size of embedding:  $d$ ,  
number of restarts (starts from same node):  $\rho$

**Output:** Updated  $n^{th}$  row of matrix of  $\Phi_t$  of size:  $N \times d$

```

1 Initialize  $\Phi_t[n] \in \mathbb{R}^d$ 
2 Initialize  $\Phi^{space}$ , trajectory representation matrix of nodes
  only at  $G_t$ 
3 Initialize  $\Phi^{time}$ , trajectory representation matrix of past nodes
  at  $G_{t-1}, \dots, G_{t-\tau}$ 
4 for  $i=0$  to  $\rho$  do
5   //for space walk
6   spaceWalk = randomWalk( $G_t, n, \mathcal{L}_\alpha$ )
7    $\Phi^{space}[n] = \text{SkipGram}(\Phi^{space}, \text{spaceWalk}, \mathcal{W}_v)$ 
8   //for time walk
9   Construct  $G_{new}$  as described in Sec 3.2
10  timeWalk = randomWalk( $G_{new}, n, \mathcal{L}_\beta$ )
11   $\Phi^{time}[n] = \text{SkipGram}(\Phi^{time}, \text{timeWalk}, \mathcal{W}_v)$ 
12  //final trajectory embedding
13   $\Phi_t[n] = f(\Phi^{space}[n], \Phi^{time}[n])$ 
14 end

```

---

Datasets	No. of nodes	No. of edges	Duration	No. of Trajectories
DBLP	118,030	288,634	1969-2016	15,400
EPINION	21,575	2,590,798	Mar 2002-Apr 2011	21,575
CIAO	2249	27,209	Sep 2001-Mar 2011	1800

**Table 1: Statistics of datasets: number of nodes, number of edges, the duration of data considered for analysis and number of node trajectories whose class labels are known. This is further described in Sections 4.1 and 4.3.**

## 4 EXPERIMENTS

We discuss details of our datasets in Section 4.1. We evaluated the proposed methods primarily on the task of trajectory classification. We also extended this to study the goodness of the learned trajectory representations through visualization as well as a second task, change point detection. Their performance is compared with three baseline methods, which are described in Section 4.2. Section 4.3 and 4.4 provide the details of the experimental setup and results for the trajectory classification and change point detection tasks respectively.

### 4.1 Datasets

We use three real-world datasets with temporal graphs for the experiments in this work. Each of these datasets is described below.

**DBLP Dataset:** The DBLP<sup>1</sup> dataset contains bibliographic information about a large collection of computer science publications. This is a dataset commonly used in graph analysis. We have considered a subset of publications which fall under four areas of research: Database, Artificial Intelligence, Data Mining and Information Retrieval. We construct graphs from the this subset of DBLP dataset by considering authors as nodes and connecting two nodes if the corresponding authors have published a paper together. We constructed 45 such annual graphs starting from year 1969 till 2016, excluding years 1970, 1972 and 1974 (whose data was not available on the DBLP website).

**Epinion Dataset:** Epinion<sup>2</sup> is a popular product review site where users write critical reviews about products from various categories. The Epinion dataset<sup>3</sup> contains review details such as user ids, product ids, category ids, time-stamp when the ratings were created, along with few other fields. This dataset has been used in earlier work such as [22] and [23]. We create a dynamic network from this dataset by considering users as nodes and inducing an edge between two users if they have rated products from the same category. We created 110 such graphs from monthly data starting from March 2002 until April 2011.

**Ciao Dataset:** Ciao<sup>4</sup> is another popular product review site. This dataset<sup>5</sup> contains fields similar to those in Epinion: user ids, product ids, category ids, and time-stamps when the ratings were created. The temporal graph for the Ciao dataset also contains user-ids as nodes and edges between nodes when the corresponding users review products from the same category. We created 115 such graphs from monthly data starting from September 2001 until March 2011.

More details of each of these datasets are shown in Table 1.

### 4.2 Baseline Methods

We compare the performance of our proposed methods against the following baseline algorithms.

**Node PageRank:** This method defines a simple approach to use graph structure to study changes in behavior of nodes over time, viz. node trajectories. In this method, we compute the PageRank value of nodes at each time step of a graph within a window. The vector containing these PageRank values at different time steps is used as a trajectory representation for the tasks evaluated in this work.

**Avg DeepWalk:** DeepWalk [19] uses a truncated random walk to represent each node in a static graph. To compare performance of STWalk against DeepWalk, we run the DeepWalk algorithm at each time step. The average of these node embeddings is considered as

<sup>1</sup><https://www.informatik.uni-trier.de/~ley/db/>

<sup>2</sup>Website: <http://www.epinions.com>.

<sup>3</sup><http://www.cse.msu.edu/~tangjili/trust.html>

<sup>4</sup><http://www.ciao.co.uk>

<sup>5</sup><http://www.cse.msu.edu/~tangjili/trust.html>

		Methods				
		Node PageRank	Avg DeepWalk	LSTM-AE	STWalk1	STWalk2
Emb Dim	Win Size	Acc (%)	Acc (%)	Acc (%)	Acc (%)	Acc (%)
DBLP						
64	5	70.42	72.61	73.89	77.66	<b>78.74</b>
64	10	61.94	70.21	73.31	70.43	<b>77.53</b>
128	10	74.53	71.53	74.32	<b>83.72</b>	78.25
EPINION						
64	5	50.03	51.44	50.08	51.82	<b>53.10</b>
64	10	53.24	52.21	50.50	55.51	<b>56.70</b>
128	10	57.36	53.79	52.70	59.03	<b>62.03</b>
CIAO						
64	5	51.94	52.02	54.17	56.31	<b>57.27</b>
64	10	25.56	24.65	43.95	58.17	<b>62.20</b>
128	10	32.03	53.15	53.80	57.15	<b>60.29</b>

**Table 2: Trajectory classification results on the DBLP, EPINION and CIAO datasets. STWalk1 and STWalk2 are the proposed algorithms, while Node PageRank, Avg DeepWalk and LSTM-AE are the baseline methods. Column ‘Emb Dim’ indicates the dimension of the trajectory representation, while ‘Win Size’ indicates the size of time window considered for learning trajectory representations.**

trajectory representation of the node across time.

**LSTM-AE:** Inspired by the work in [21] to learn representations of video sequences using an unsupervised approach, we use a similar approach for learning node trajectories in a temporal graph. DeepWalk [19] is used to learn  $d$ -dimensional node representations at each time-step. A Long Short-term Memory (LSTM) AutoEncoder (AE) is used to learn trajectory representations (similar to [21], which used this approach for videos though). The encoder-LSTM takes the sequence of  $d$ -dimensional node embeddings and encodes the sequence, and the decoder-LSTM reconstructs the original sequence from the embedding obtained using the encoder. Once this LSTM-AE network is completely trained, the output of the encoder-LSTM is used as the trajectory representation. This baseline algorithm was implemented using Keras [8], and trained using the Adam optimizer with the mean square error loss function. The low-dimensional embedding obtained through encoder was chosen to be of  $d/2$  dimensions after empirical studies.

### 4.3 STWalk for Trajectory Classification

We studied the effectiveness of the representations of trajectories learned using the proposed methods on the task of trajectory classification (where the node’s trajectory in a pre-specified time window is considered). We now describe how each of the aforementioned datasets were studied in this work for trajectory classification, along with the corresponding results. We note that in case of each of the datasets, the total number of node trajectories are divided into a 70% training set and the remaining 30% as a test set. An ensemble voting-based classifier (comprised of random forests and Support Vector Machines) was trained on the training set using stratified  $k$ -fold cross-validation with  $k = 10$  to finetune the hyperparameters of the classifier (this classifier was chosen after empirical studies,

and is also known popularly to be a very robust classifier). The classifier with the best hyperparameters is then trained on the complete training dataset to obtain the final model, and used to study performance on test data. We use the classification accuracy as the performance metric for this task. We repeat the process five times and report the average performance in our results (Table 2). This process is maintained for each of the considered datasets.

**DBLP Dataset:** To evaluate our model on the temporal graphs in the DBLP dataset, we considered non-overlapping windows of 5 years, leading to 9 such windows over the 45 graphs in this dataset. (For example, while learning the trajectory embedding of nodes at time  $t_5$ , we consider the graph at  $t_5$  as well as the previous 4 graphs:  $t_4, t_3, t_2$  and  $t_1$ .) A similar process is used for window of 10 years too.

The representations of author trajectories learned using the proposed methods and baseline methods are then used for the binary classification task, where class labels are “Expert Researcher” and “Interdisciplinary Researcher”. These labels are assigned to each node trajectory in the following manner. Based on the conference where each author has published his/her work, we assign one of the labels from the following to each node of a graph at each time step. Label *Database*: if published in ICDE, VLDB, SIGMOD, PODS or EDBT; label *Artificial Intelligence*: if published in IJCAI, AAAI, ICML, or ECML; label *Data Mining*: if published in KDD, PAKDD, ICDM, PKDD or SDM; label *Information Retrieval*: if published in SIGIR, WWW, ECIR or WSDM. Subsequently, the trajectory labels are determined based on the node labels assigned at each time-step. If the node label is constant throughout the considered time window, the trajectory label corresponding to that node is “Expert Researcher”. If the node label changes within the time window, we assign the label “Interdisciplinary Researcher” to the author trajectory.

STWalk has four hyper-parameters: *embedding dimension*: dimension of the trajectory representation; *window size*: number of time steps to consider while learning trajectory embedding; *walk length*: length of random walk performed on the spatio-temporal graph; and *restarts*: number of random walks starting from the same node. Table 2 shows the performance of the proposed and baseline methods in three different hyper-parameter settings. The *walk length* value was set to 30, and the *restarts* value to 40.

**Epinion Dataset:** In this dataset, when we set the hyperparameter, window size, to five, 22 windows are considered for analysis (among the 110 graphs available in this dataset), and similarly when the window size is set to 10. Category IDs are used as node labels, which are subsequently used to determine the ground truth labels for node trajectories. If a user has reviewed products from the same category during all the time steps considered within a time window, the trajectory corresponding to the user will be assigned the label “Proficient reviewer”, indicating that the user has good knowledge about the products in the category. If the user reviews products from different categories during the considered time window, the user is assigned the label “Versatile reviewer”, indicating that the user has expertise in assessing products from many categories. Thus, trajectory classification is modeled as a binary classification problem on this dataset, and we report the result of executing our algorithm on Epinion dataset in various hyper-parameter settings in Table 2. Similar to the DBLP dataset, we set the hyper-parameters *walk length* to 30 and *restarts* to 40 for this dataset.

**Ciao Dataset:** In this dataset, when we set the hyperparameter, window size, to five, 23 windows are considered for analysis (among the 115 graphs available in this dataset), and similarly when the window size is set to 10. The category ID of the product that a user reviews is considered as the node label, and ground truth trajectory labels are determined based on these node labels. Similar to the Epinion dataset, if the node labels remain constant for the entire time window, the corresponding node trajectory is assigned the label “Experienced reviewer”, and if the user reviews products from multiple categories within the same time window, the node trajectory is assigned the label “Multifaceted reviewer”. Binary classification is then used to validate the effectiveness of trajectory representations learned on this dataset. The *walk length* and *restarts* hyperparameters are chosen as on the Epinion dataset.

Table 2 shows the classification accuracies of the proposed and baseline methods on all the three datasets. The proposed algorithms, especially STWalk2 shows superior performance in most settings. In case of the CIAO dataset, the proposed algorithms show significant increase in accuracy over the baseline methods. Interestingly, the LSTM-AE method, which was introduced as a baseline method in this work, also performs better than the other two baseline methods. The results support the proposed methods in general. In particular, the improved performance of STWalk2 shows that having separate space and time walks provides better representations. We surmise the reason for this to be the improved coverage of temporal neighborhoods through this approach when compared to STWalk1 under the same random walk lengths.

#### 4.4 STWalk for Change Point Detection

In a temporal setting, analyzing change points can be critical in applications such as concept drift and anomaly detection. As the trajectory representation captures the change in behavior of nodes in temporal graph over a given time period, this makes the trajectory representations suitable features to study such change points.

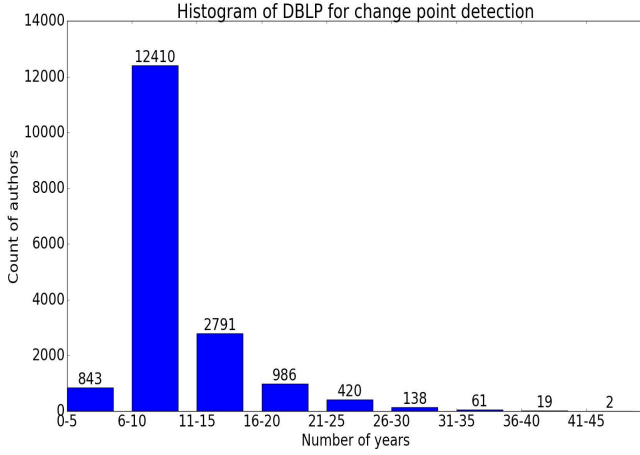
In the DBLP dataset, we use such an approach to find the average time spent by an author in his/her research area before switching to another domain, thus finding the ‘points of change in research domain’ in the author’s trajectory. Specifically, we consider the trajectory labels across multiple windows of the author (as obtained from Section 4.3). We study the sequence of trajectory labels and based on the pattern of occurrence of “Interdisciplinary” and “Expert” labels, we calculate the average duration of time spent in a single domain. Figure 2a shows the histogram of authors who spent certain number of years in one domain before working in other research area. This analysis allows us to answer queries such as the minimum (or average) number of years spent before a researcher becomes interdisciplinary. For example, Figure 2a shows evidently that the average duration of time before a researcher becomes interdisciplinary is between 6-10 years, which agrees with common understanding. We also used the histogram to analyze cases of individual authors, and found the change points to be fairly accurate. This supports the effectiveness of the trajectory representations learned by the proposed methods (STWalk2 was used in these studies, considering its superior performance in the earlier experiments).

Similarly, analyzing change points in the Epinion dataset gives insight into the behavior of reviewers. Figure 2b shows the trends for the Epinion dataset. It shows the histogram of reviewers and the number of months they review products from one single category, before switching to products from other categories. In this case, it is clear from the figure that a reviewer, on an average, writes reviews for products from one category for 1 to 2 months before switching to other categories, which matches with common understanding too. This corroborates the effectiveness of the proposed method (STWalk2 in this case again) for learning trajectory representations in temporal graphs.

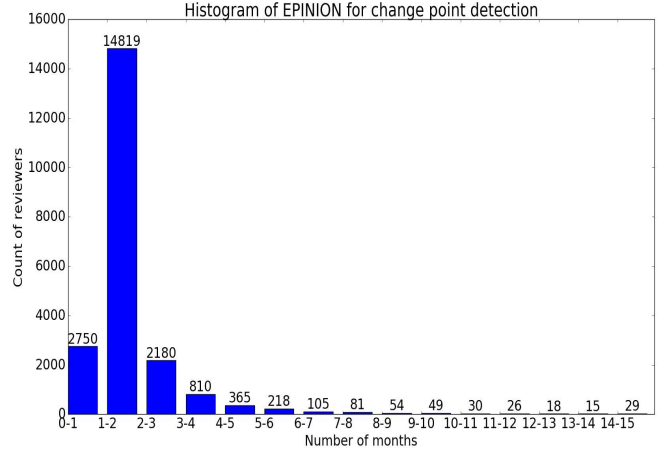
#### 4.5 Arithmetic Operations on Trajectory Representations

Mikolov et al. [18] demonstrated that arithmetic operations on learned word representations show a linear structure in embedding space, in the case of text processing. For example,  $\text{vector}(\text{“Paris”}) - \text{vector}(\text{“France”}) + \text{vector}(\text{“England”})$  gives a vector representation nearest to the word embedding for London. Here, we examine whether similar behavior is shown by the learned trajectory representations. We perform this study with the representations learned on the DBLP dataset. In particular, we consider the representations for interdisciplinary authors (shown as yellow points in Figure 3, best viewed in color) and expert authors (shown as cyan points in the figure). We then perform the following difference operation:  $\text{trajectory}(\text{“interdisciplinary author”}) - \text{trajectory}(\text{“Expert author”})$  and examine the resulting representations (plotted in red color in the figure).





a) Histogram shows the count of authors vs number of years spent by any author in one research area before switching to other domain



b) Histogram of number of users vs average duration spent on reviewing product of same category before switching to other category.

Figure 2: Results of change point detection on the DBLP and Epinion datasets

Figure 3 shows the Principal Component Analysis (PCA) plot for the result of performing arithmetic difference operation on the trajectory representations of DBLP authors, as described above. For this plot, we considered only the trajectory vectors of authors working in the *Database* (DB) domain (shown in pink color on the scatter plot). We also considered the trajectory embeddings of authors who focus only on the *Artificial Intelligence* (AI) domain (violet color on the scatter plot). Thus, the pink and violet clusters represent authors with trajectory label “Expert Researcher”. To evaluate the goodness of learned trajectory embeddings, we consider the representations for authors with labels “Interdisciplinary Researcher”. In particular, we consider trajectories of authors who have published in both *Database* and *Artificial Intelligence* (AI+DB) venues. These interdisciplinary authors are shown in green color data points on the scatter plot. Three distinct clusters, in Figure 3 indicating DB (pink), AI (violet) and AI+DB (green), validate that the proposed model STWalk is able to learn rich trajectory embeddings for “Expert” authors and “Interdisciplinary” authors.

To validate if the arithmetic operation:  $(AI+DB)-DB=AI$  holds for the learned trajectory embeddings, we perform the following steps. We take the trajectory vectors of a randomly picked “Interdisciplinary” author from the collection of vectors mentioned above. The selected “Interdisciplinary” points are shown in yellow color in Figure 3. These are part of the green cluster, which represents all “Interdisciplinary” authors. Similarly, we pick trajectory embeddings of “Expert” authors from a DB background (shown in cyan color points). We then apply vector difference operation on vectors from the above two selected sets, and the resulting vectors are plotted in red color. As seen in Figure 3, the resulting red points are near the violet points, which indicates the cluster for “Expert” authors in AI. This demonstrates that simple vector arithmetic on the learned trajectory representations show linear structure and hence, can be used for developing interesting applications.

Figure 4 shows the results of similar vector operation on Epinion dataset. We considered trajectory representation of users who

review products from category *Movies* and *Books* as “Proficient reviewer”, represented by violet and pink clusters respectively in figure 4. The trajectory representation of users reviewing products from both categories *Movies* and *Books* as “Versatile reviewer”, represented by green cluster. We randomly select trajectory embedding of “Versatile reviewer”, the yellow points and “Proficient reviewer”, the cyan points, and then perform vector subtraction. The resulting points are plotted in red color. We can conclude from 4 that these points are near violet cluster which represents *Movies* cluster, validating that the learned trajectory representations show linear structure

## 5 CONCLUSION

In this paper, we proposed STWalk a novel approach to learn trajectory representations in temporal graphs, by capturing change in nodes’ behavior over time. We presented two algorithms to learn such trajectory representations: STWalk1 learns the representations by considering the present neighbors and past neighbors together, and STWalk2 models it as two subproblems, solving each separately before combining the results to get the final trajectory embedding. We validated the effectiveness of the learned trajectory representations on three real-world datasets: DBLP, Epinion and Ciao. We compare the results against three baseline algorithms, and show that the proposed STWalk outperforms baseline methods in various settings.

Furthermore, we demonstrated that simple arithmetic operations are valid on the trajectory representations learned through the proposed methods, i.e., the vector difference:  $\text{trajectory}(\text{“interdisciplinary author”}) - \text{trajectory}(\text{“specialized author”}) = \text{trajectory}(\text{“other specialized author”})$ . We also examined the usefulness of trajectory embeddings in change point detection. Each of these experimental studies showed the effectiveness of trajectory representations learned by STWalk.



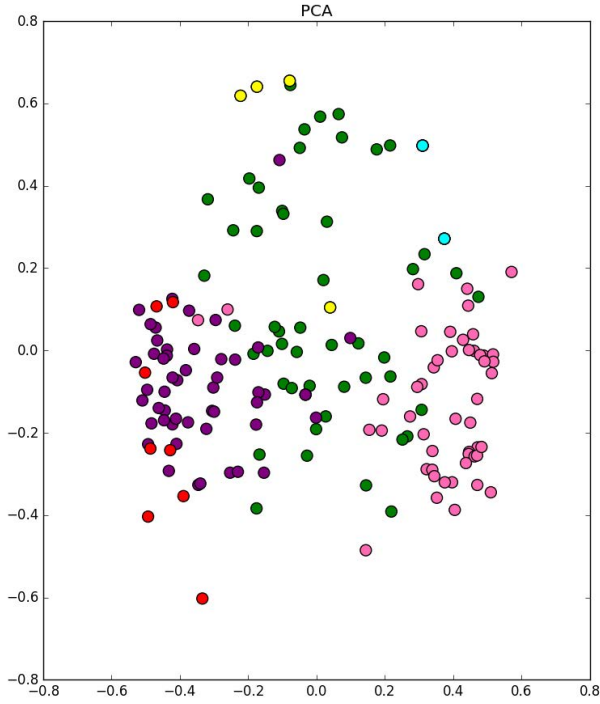


Figure 3: Analyzing the goodness of trajectory embeddings of DBLP dataset (Best viewed in color). (i) Arithmetic operations are valid on trajectory embeddings. E.g. (interdisciplinary embedding)-(specialized embedding)=(other specialized embedding), i.e., taking arithmetic difference as in (AI+DB) (yellow points)-DB (cyan points)=AI (red points). (ii) Pink color cluster represents DBLP authors working in Database (DB). Violet color represents authors working in Artificial Intelligence (AI) and the green color cluster indicates the interdisciplinary authors working in both areas AI and DB.

## REFERENCES

- [1] Charu Aggarwal and Karthik Subbian. 2014. Evolutionary network analysis: A survey. *ACM Computing Surveys (CSUR)* 47, 1 (2014), 10.
- [2] Charu C Aggarwal. 2011. An introduction to social network data analytics. *Social network data analytics* (2011), 1–15.
- [3] Charu C Aggarwal and Nan Li. 2011. On node classification in dynamic content-based networks. In *Proceedings of the 2011 SIAM International Conference on Data Mining*. SIAM, 355–366.
- [4] Leman Akoglu, Hanghang Tong, and Danai Koutra. 2015. Graph based anomaly detection and description: a survey. *Data Mining and Knowledge Discovery* 29, 3 (2015), 626–688.
- [5] Shaosheng Cao, Wei Lu, and Qiongkai Xu. 2015. Grarep: Learning graph representations with global structural information. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*. ACM, 891–900.
- [6] Shaosheng Cao, Wei Lu, and Qiongkai Xu. 2016. Deep Neural Networks for Learning Graph Representations. In *AAAI* 1145–1152.
- [7] Shiyu Chang, Wei Han, Jiliang Tang, Guo-Jun Qi, Charu C Aggarwal, and Thomas S Huang. 2015. Heterogeneous network embedding via deep architectures. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 119–128.
- [8] François Chollet et al. 2015. Keras. <https://github.com/fchollet/keras>. (2015).

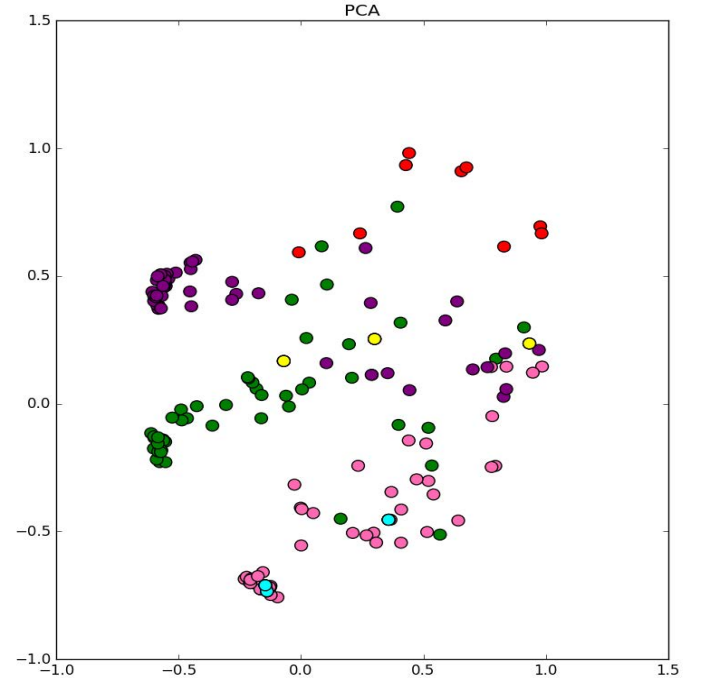


Figure 4: Analyzing the goodness of trajectory embeddings of Epinion dataset (Best viewed in color). (i) Arithmetic difference of users' trajectory who review products from category (Movies+Books) (yellow points)-Books (cyan points)=Movies category (red points). (ii) Pink color cluster represents users who review Books. Violet color represents users who review Movies and the green color cluster indicates the versatile users who review products from both categories Movies and Books.

- [9] Lise Getoor and Christopher P Diehl. 2005. Link mining: a survey. *Acm Sigkdd Explorations Newsletter* 7, 2 (2005), 3–12.
- [10] Palash Goyal and Emilio Ferrara. 2017. Graph Embedding Techniques, Applications, and Performance: A Survey. *arXiv preprint arXiv:1705.02801* (2017).
- [11] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 855–864.
- [12] Manish Gupta, Jing Gao, Charu Aggarwal, and Jiawei Han. 2014. Outlier detection for temporal data. *Synthesis Lectures on Data Mining and Knowledge Discovery* 5, 1 (2014), 1–129.
- [13] William L Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. *arXiv preprint arXiv:1706.02216* (2017).
- [14] William L Hamilton, Rex Ying, and Jure Leskovec. 2017. Representation Learning on Graphs: Methods and Applications. *arXiv preprint arXiv:1709.05584* (2017).
- [15] Jundong Li, Harsh Dani, Xia Hu, Jiliang Tang, Yi Chang, and Huan Liu. 2017. Attributed Network Embedding for Learning in a Dynamic Environment. *arXiv preprint arXiv:1706.01860* (2017).
- [16] Fragkiskos D Malliaros and Michalis Vazirgiannis. 2013. Clustering and community detection in directed networks: A survey. *Physics Reports* 533, 4 (2013), 95–142.
- [17] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).
- [18] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.
- [19] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: Online Learning of Social Representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*. ACM, New York, NY, USA, 701–710. <https://doi.org/10.1145/2623330.2623732>

- [20] Purnamrita Sarkar, Deepayan Chakrabarti, and Michael Jordan. 2012. Nonparametric link prediction in dynamic networks. *arXiv preprint arXiv:1206.6394* (2012).
- [21] Nitish Srivastava, Elman Mansimov, and Ruslan Salakhutdinov. 2015. Unsupervised Learning of Video Representations using LSTMs.. In *ICML*. 843–852.
- [22] J. Tang, H. Gao, and H. Liu. 2012. mTrust: Discerning multi-faceted trust in a connected world. In *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 93–102.
- [23] J. Tang, H. Gao, H. Liu, and A. Das Sarma. 2012. eTrust: Understanding trust evolution in an online world. (2012), 253–261 pages.
- [24] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 1067–1077.
- [25] Lei Tang, Huan Liu, Jianping Zhang, and Zohreh Nazeri. 2008. Community evolution in dynamic multi-mode networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 677–685.