GloDyNE: Global Topology Preserving Dynamic Network Embedding

Chengbin Hou, Han Zhang, Shan He, and Ke Tang

Abstract—Learning low-dimensional topological representation of a network in dynamic environments is attracting much attention due to the time-evolving nature of many real-world networks. The main and common objective of Dynamic Network Embedding (DNE) is to efficiently update node embeddings while preserving network topology at each time step. The idea of most existing DNE methods is to capture the topological changes at or around the most affected nodes (instead of all nodes) and accordingly update node embeddings. Unfortunately, this kind of approximation, although can improve efficiency, cannot effectively preserve the global topology of a dynamic network at each time step, due to not considering the inactive sub-networks that receive accumulated topological changes propagated via the high-order proximity. To tackle this challenge, we propose a novel node selecting strategy to diversely select the representative nodes over a network, which is coordinated with a new incremental learning paradigm of Skip-Gram based embedding approach. The extensive experiments show GloDyNE, with a small fraction of nodes being selected, can already achieve the superior or comparable performance w.r.t. the state-of-the-art DNE methods in three typical downstream tasks. Particularly, GloDyNE significantly outperforms other methods in the graph reconstruction task, which demonstrates its ability of global topology preservation.

Index Terms—Dynamic Networks, Network Embedding, Global Topology, Feature Extraction or Construction, Data Mining

1 Introduction

The interactions or connectivities between entities of a real-world complex system can be naturally represented as a network (or graph), e.g., social networks, biological networks, and sensor networks. Learning topological representation of a network, especially low-dimensional node embeddings which encode network topology therein so as to facilitate downstream tasks, has received a great success in the past few years [1], [2], [3].

Most previous Network Embedding methods such as [4], [5], [6], [7], [8] are designed for *static networks*. However, many real-world networks are dynamic by nature, i.e., edges might be added/removed between seen and/or unseen nodes as time goes on. For instance, in a wireless sensor network, devices will regularly connect to or accidentally disconnect from routers; in a social network, new friendships will establish between new users and/or existing users. Due to the *time-evolving nature* of many real-world networks, Dynamic Network Embedding (DNE) is now attracting much attention [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19]. The main and common objective of DNE is to efficiently update node embeddings while preserving network topology at each time step. Most existing DNE methods try to compromise between effectiveness (evaluated by downstream tasks) and efficiency (while obtaining node embeddings). The idea is to capture the topological changes at or around the most affected nodes (instead of all nodes), and promptly update node embeddings based on an efficient incremental learning paradigm.

Unfortunately, this kind of approximation, although can improve the efficiency, cannot effectively preserve the global topology of a dynamic network at each time step. Specifically, any changes, i.e., edges added/removed between nodes, would affect all nodes in a connected network via the *high-order proximity* as illustrated in Figure 1 a). On the other hand, as observed from Figure 1 b-d), the real-world dynamic networks usually have some *inactive sub-networks* where no change occurs lasting for several time steps. Putting both together, the existing DNE methods, which focus on the most affected nodes (belonging to the active sub-networks) but do not consider the inactive sub-networks, would overlook the accumulated topological changes propagating to the inactive sub-networks via the high-order proximity.

To tackle this challenge, the proposed DNE method–Global topology preserving Dynamic Network Embedding (GloDyNE) first partitions a current network into smaller sub-networks where one representative node in each sub-network is selected, so as to ensure the *diversity* of selected nodes. The representative node for each sub-network is sampled via a probability distribution over all nodes within each sub-network, such that a higher probability is assigned to a node with the larger accumulated topological changes. After that, GloDyNE captures the latest topologies around the selected nodes by truncated random walks [4], and then promptly updates node embeddings based on the Skip-Gram Negative Sampling (SGNS) model [21] and an incremental learning paradigm.

The contributions of this work are as follows: 1) We demonstrate the existence of inactive sub-networks in real-world dynamic networks. Together with the propagation of topological changes via the high-order proximity, we find the issue of global topology preservation for many

- *C. Hou and K. Tang are with the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, 518055, China.*
  E-mail: chengbin.hou10@foxmail.com and tangk3@sustech.edu.cn
- *H. Zhang and S. He are with the School of Computer Science, University of Birmingham, Birmingham, B15 2TT, United Kingdom.*
  E-mail: hxz325@cs.bham.ac.uk and s.he@cs.bham.ac.uk

Manuscript received March 8, 2020; revised TBD.

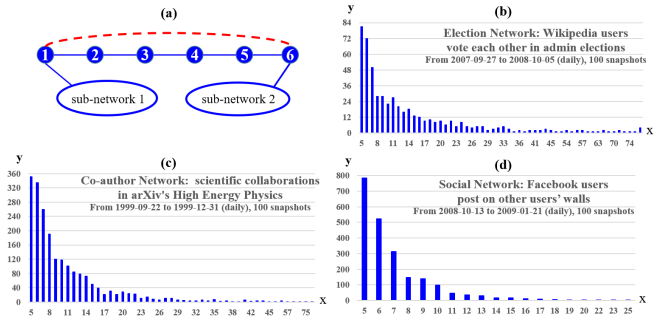

Fig. 1. a) A change (new edge in red) affects all nodes in the connected network via high-order proximity. The proximity of nodes 1-6 becomes 1^{st} order from 5^{th} order, nodes 2-6 becomes 2^{nd} order from 4^{th} order, etc. Besides, the proximity of any node in sub-network 1 to any node in sub-network 2 is reduced by 5 orders. b-d) The real-world dynamic networks have some inactive sub-networks (e.g., defined as no change occurs lasting for at least 5 time steps). The x-axis indicates the number of consecutive time steps that no change occurs in a sub-network. The y-axis gives the counts of each case in x-axis. The sub-networks, in average 50 nodes per sub-network, are obtained by applying METIS algorithm [20] on the largest snapshot of a dynamic network. The details of the three dynamic networks are described in Section 5.

existing DNE methods. 2) To better preserve the global topology, *unlike all previous DNE methods*, we propose to also consider the accumulated topological changes in inactive sub-networks. A novel node selecting strategy is thus proposed to diversely select the representative nodes over a network. 3) We further develop a new DNE method or framework, namely GloDyNE, which extends the random walk and Skip-Gram based network embedding approach to an incremental learning paradigm with a free hyper-parameter for controlling the number of selected nodes at each time step. 4) The extensive empirical studies show the superiority of GloDyNE compared with the state-of-the-art DNE methods in terms of both effectiveness and efficiency, as well as verify the usefulness of some special designs of GloDyNE, such as the node selecting strategy and the free hyper-parameter.

The remainder of the paper is organized as follows. We first review the related works in Section 2, and then formally give the definition of DNE problem in Section 3. In Section 4, we present GloDyNE step by step, as well as its pseudocode and time complexity. The empirical studies are reported and discussed in Section 5. In particular, Section 5.2 aims to compare GloDyNE with other DNE methods, whereas Section 5.3 tries to investigate GloDyNE itself. And finally, we conclude this work in Section 6.

2 Related Works

2.1 Static Network Embedding

To learn low-dimensional topological representation of a network in dynamic environments, one naive solution is to treat the snapshot of a dynamic network at each time step as a *static network*, so that a static Network Embedding method such as [4], [5], [7], [8] can be directly applied to learn node embeddings for each snapshot.

As reported in recent DNE works [9], [12], [13], [14], this naive solution obtains superior results compared with some DNE methods. One possible reason is that this solution does not suffer the aforementioned issue of global topology preservation as introduced in Section 1. However, it is *time-consuming* [12], [14], and thus may not satisfy the requirement of promptly updating embeddings for some DNE downstream tasks [16], [22].

2.2 Dynamic Network Embedding

To compromise between effectiveness and efficiency, most existing DNE methods try to capture the topological changes at or around the most affected nodes (instead of all nodes or edges), and promptly update node embeddings based on an incremental learning paradigm. BCGD [9] aims to minimize the loss of reconstructing the network proximity matrix using the node embedding matrix with a temporal regularization term, and it is optimized by the Block-Coordinate Gradient Descent algorithm. Particularly, this work further offers an efficient solution–BCGD-incremental that only updates the most affected nodes' embeddings based on their previous embeddings. DynAE [11] and NetWalk [22] both utilize an auto-encoder with some regularization terms for modeling. They continuously train the model inherited from the last time step, so that the model converges in a few iterations thanks to the knowledge transfer from previous models. To efficiently cope with dynamic changes at each time step, some DNE methods [14], [17] propose an incremental version of the Skip-Gram model [21] to update embeddings based on the most affected nodes. Likewise, DHEP [12] extends HOPE [8] to an incremental version by modifying the most affected eigenvectors using the matrix perturbation theory.

Apart from above DNE methods with the trade-off between effectiveness (evaluated by downstream tasks) and efficiency (while obtaining node embeddings), some DNE methods [15], [18], [19] aim to further improve the effectiveness *without considering the efficiency*. For example, tNodeEmbed [18] runs a static Network Embedding method to obtain node embeddings at each available time step, and then employs Recurrent Neural Networks among them (for better exploiting the temporal dependence and hence may further improve the effectiveness) to obtain the final node embeddings at each time step.

The proposed method–GloDyNE considers both effectiveness and efficiency. Nevertheless, unlike the DNE methods that focus on the most affected nodes or sub-networks, GloDyNE additionally considers the accumulated topological changes in inactive sub-networks (i.e., no change occurs lasting for several time steps) for the better global topology preservation of a dynamic network at each time step.

3 Notation and Problem Definition

Definition 1. A Static Network. Let $G = (\mathcal{V}, \mathcal{E})$ be a static network where $\mathcal{V}$ denotes a set of $|\mathcal{V}|$ nodes or vertices, and $\mathcal{E}$ denotes a set of $|\mathcal{E}|$ edges or links. The adjacency matrix of G is denoted as $W \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$ where w_{ij} is the weight of edge e_{ij} between a pair of nodes (v_i, v_j), and if $w_{ij} = 0$, there is no edge between the two nodes. For the unweighted and undirected network, $w_{ij} \in \{0, 1\}$ and $w_{ij} = w_{ji}$.

Definition 2. A Dynamic Network. A dynamic network $\mathcal{G}$ is represented by a sequence of snapshots G^t taken at each time step t, i.e., $\mathcal{G} = (G^0, G^1, ..., G^t, G^{t+1}, ...)$. Each snapshot G^t can be treated as a static network.

Definition 3. Static Network Embedding. The static network embedding aims to find a mapping function $Z = f(G)$, where $Z \in \mathbb{R}^{|\mathcal{V}|\times d}$, $d \ll |\mathcal{V}|$, and each row vector $Z_i \in \mathbb{R}^d$ is the node embedding for v_i, such that the pairwise similarity of node embeddings in Z best preserves the pairwise topological similarity of the nodes in G.

Definition 4. Dynamic Network Embedding. The DNE problem, under an incremental learning paradigm, can be defined as $Z^t = f^t(G^t, G^{t-1}, f^{t-1}, Z^{t-1})$ where $Z^t \in \mathbb{R}^{|\mathcal{V}^t|\times d}$ is the latest node embeddings, f^{t-1} and Z^{t-1} are the model and embeddings from the last time step respectively. The main objective of DNE in this work is to efficiently update node embeddings at each current time step t, such that the pairwise similarity of node embeddings in Z^t best preserves the pairwise topological similarity of the nodes in G^t.

Definition 5. Sub-networks of A Snapshot. Let G_k^t denote a sub-network of a snapshot G^t. All sub-networks of a snapshot G^t, after network partition [23], should be non-overlapping, i.e, $\mathcal{V}_m^t \cap \mathcal{V}_n^t = \emptyset, \forall \ \ m \neq n$. And their node sets should satisfy $\mathcal{V}^t = \bigcup_k \mathcal{V}_k^t$.

4 The Proposed Method

The proposed DNE method–GloDyNE consists of four important components which are introduced step by step in Section 4.1. Intuitively, Step 1 and 2 ensure the selected nodes *diversely distributed over a network*, and meanwhile, bias to the nodes with larger accumulated topological changes for each sub-network. Step 3 encodes the latest topologies around the selected nodes into random walks, which are then decoded by a sliding window and the SGNS model for incrementally training node embeddings as described in Step 4. Note that, these four steps would be *repeatedly executed at each time step*. The implementation details (via pseudocode) and complexity analysis are presented in Section 4.2 and 4.3 respectively.

4.1 Method Description

4.1.1 Step 1. Partition A Network

In order to realize inactivate sub-networks of a snapshot G^t, it is needed to divide G^t into sub-networks $G_1^t, G_2^t, ..., G_K^t$ where K is the number of sub-networks of a snapshot. The sub-networks are desirable to be non-overlapped and to cover all nodes in the original snapshot as defined in Definition 5, so that the later Step 2 can select unique nodes from each sub-network and the later Step 3 is easier to explore the whole snapshot G^t based on the selected nodes from each sub-network. A network partition algorithm [23] is therefore used to achieve the desirable goals. The most common objective function is to minimize the edge cut and can be formulated as

$$\min \sum_{1\leq m,n\leq K} \{w_{i,j}^t | v_i^t \in \mathcal{V}_m^t, v_j^t \in \mathcal{V}_n^t, (v_i^t, v_j^t) \in \mathcal{E}^t\} \quad (1)$$

where the subscripts i, j indicate node ID and m, n indicate sub-network ID. Note that Eq. (1) should subject to two constraints $\mathcal{V}_m^t \cap \mathcal{V}_n^t = \emptyset, \forall \ \ m \neq n$, and $\mathcal{V}^t = \bigcup_k \mathcal{V}_k^t$ for the reasons as discussed above.

Moreover, an additional constraint of balanced sub-networks is introduced to let the number of nodes be similar among all sub-networks, so as to facilitate the later steps to fairly explore all sub-networks and hence better preserve the global topology. The third constraint of balanced sub-networks can be defined as

$$\forall k \in \{1, ..., K\}, |\mathcal{V}_k^t| \leq (1+\epsilon)\frac{|\mathcal{V}^t|}{K} \quad (2)$$

where $|\mathcal{V}_k^t|$ is the number of nodes in G_k^t and ϵ is the tolerance parameter. Note that, if ϵ is 0, network partitions are perfectly balanced. In practice, ϵ is set to a small number to allow a slight violation. However, such a (K, ϵ) balanced network partition is a NP hard problem [23]. In order to address this problem, METIS algorithm [20] is employed. There are roughly three steps. Firstly, the coarsening phase, the original network is recursively transformed into a series of smaller and smaller abstract networks, via collapsing nodes with common neighbors into one collapsed node until the abstract network is small enough. Secondly, the partition phase, a K-way partition algorithm is applied on the smallest abstract network to get the initial partition of K sub-networks. Thirdly, the uncoarsening phase, it recursively expands the smallest abstract network back to the original network, and meanwhile recursively swaps the collapsed nodes (or the original nodes lastly) at the boarder of sub-networks between two neighboring sub-networks, so as to minimize the edge cut as describe in Eq. (1).

4.1.2 Step 2. Select Representative Nodes

In order to ensure the selected nodes diversely distributed over a snapshot G^t, one natural idea is to select one representative node from each sub-network. As a result, the total number of selected nodes is K. To increase the total number of selected nodes, one can simply increase the number of sub-networks by network partition. We let $K = \alpha|\mathcal{V}^t|$, so that α can freely control the total number of selected nodes for the trade-off between effectiveness and efficiency.

The problem now becomes as how to select one representative node from a sub-network. According to the recent DNE works such as [14], [17], [22], the nodes affected greatly by edge steams are selected for updating their embeddings, since their topologies are altered greatly. Similarly, in this work, the representative node to be selected is biased to the node with larger topological changes. Motivating by the concept of inertia[1] from Physics, an efficient scoring function is designed to evaluate the accumulated topological changes of a node v_i^t in a current snapshot G^t as follows

$$\begin{aligned} S(v_i^t) &= \frac{|\Delta\mathcal{E}_i^t| + \mathcal{R}_i^{t-1}}{Deg(v_i^{t-1})} \\ &= \frac{|\,\mathcal{N}(v_i^t) \cup \mathcal{N}(v_i^{t-1}) - \mathcal{N}(v_i^t) \cap \mathcal{N}(v_i^{t-1})\,| + \mathcal{R}_i^{t-1}}{Deg(v_i^{t-1})} \end{aligned} \quad (3)$$

where the reservoir $\mathcal{R}_i^{t-1}$ stores the accumulated changes[2] of v_i up to $t-1$. For simplicity, we treat G^t as an undirected

1. Here the node degree is regarded as the inertia of this node.

2. The accumulated changes in reservoir are used to fix the case when a node has small changes at each time step for a long time, which will greatly affect network topology but may be ignored if not recorded.

and unweighted network[3], so that the current changes of v_i at t, denoted as $|\Delta\mathcal{E}_i^t|$, can be easily obtained by the set operations on neighbors of v_i as shown in Eq. (3), which is equivalent to count the number of the edges with node v_i from current edge steams $\Delta\mathcal{E}^t$. The representative node of a sub-network G_k^t is then selected based on the probability distribution over its node set $\mathcal{V}_k^t$, i.e.,

$$P(v_i^t) = \frac{e^{S(v_i^t)}}{\sum_{v_j^t \in \mathcal{V}_k^t} e^{S(v_j^t)}} \quad \forall\, v_i^t \in \mathcal{V}_k^t \tag{4}$$

where e is Euler's number and $S(v_i^t)$ is the score of the accumulated topological changes of node v_i^t given by Eq. (3). Note that, if $S(v_i^t) = 0$, $e^0 = 1$ and $P(v_i^t) \neq 0$, so that even for an inactivate sub-network with no change at all nodes, the probability distribution over this sub-network is still a valid uniform distribution. Intuitively, within a sub-network, the higher score of a node given by Eq. (3) is, the higher probability of this node will be selected as the representative node for this sub-network. Because one representative node from each sub-network is selected, all the selected nodes are therefore diversely distributed over the whole snapshot, and meanwhile, biased to the larger accumulated topological changes for each sub-network.

4.1.3 Step 3. Capture Topological Changes

Given the selected representative nodes from Step 2, this step will explain how to capture the topological changes based on the selected nodes. As the topological changes at the selected nodes can propagate to other nodes via the high-order proximity, the truncated random walk sampling [4] (instead of edge sampling [5]) strategy is employed to capture the topological changes around (instead of at) the selected nodes. Concretely, for each selected node, r truncated random walks with length l are conducted starting from the selected node. For a random walk, the next node v_j^t is sampled based on the probability distribution over its previous node's neighbors $\mathcal{N}(v_i^t)$, i.e.,

$$P(v_j^t|v_i^t) = \begin{cases} \frac{w_{ij}^t}{\sum_{v_{j'} \in \mathcal{N}(v_i^t)} w_{ij'}^t} & \text{if } v_j \in \mathcal{N}(v_i^t) \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

4.1.4 Step 4. Update Node Embeddings

After Step 3, the latest topological information around the selected nodes is encoded in random walks. Step 4 aims to utilize the random walks to update node embeddings. Following [4] and [7], a sliding window with length $s + 1 + s$ is used to slide along each walk (i.e., node sequence), and the positive node-pair samples in a set $\mathcal{D}^t$ is built via $(v_{center+i}^t, v_{center}^t)$ where $i \in [-s, +s], i \neq 0$. As a result, the node-pair samples can encode $1^{st} \sim s^{th}$-order proximity of a given center node with another node. Note that, several network embedding works have shown the advantage of using the high-order proximity [6], [8], [24].

3. If one wants to consider edge's weight in Eq. (3), let $|\Delta\mathcal{E}_i^t| = \sum_{v_j^t \in \mathcal{N}(v_i^t)} |w_{i,j}^t - w_{i,j}^{t-1}| + \sum_{v_j^{t-1} \in [\mathcal{N}(v_i^{t-1}) - \mathcal{N}(v_i^t)]} |w_{i,j}^{t-1}|$ where the first term gives the total weight changes of i' neighbors presented at both t and $t-1$, and presented at t but not presented at $t-1$; while the second term gives the total weight changes of i' neighbors presented at $t-1$ but not presented at t. The operator $|\cdot|$ on a set gives its cardinal number, and on a scalar gives its absolute value.

Assuming the observations of node pairs in $\mathcal{D}^t$ are mutually independent [7], the objective function to maximize the node co-occurrence log probability over all node pairs in $\mathcal{D}^t$ can be written as

$$\max \sum_{(v_i^t, v_c^t) \in \mathcal{D}^t} \log P(v_i^t|v_c^t) \tag{6}$$

where v_c^t is the center node, and v_i^t is another node with $1^{st} \sim s^{th}$ order proximity to v_t^c. Unlike [4] which defines $P(v_i^t|v_c^t)$ as a softmax, we follow [25] to treat it as a binary classification problem, so as to further reduce the complexity. Concretely, it aims to distinguish a positive sample $(v_i^t, v_j^t) \in \mathcal{D}^t$ from q negative samples $(v_i^t, v_{j'}^t)$s. The probability of observing a positive sample (v_i^t, v_j^t) can be defined as

$$P(B = 1|v_i^t, v_j^t) = \sigma(Z_i^t \cdot Z_j^t) = \frac{1}{1 + e^{Z_i^t \cdot Z_j^t}} \tag{7}$$

where Z_i^t is the node embedding vector parameterized by the mapping function $f^t(v_i^t)$, the operator $\cdot$ represents the dot product between two vectors, and $P(B = 1|v_i^t, v_j^t)$ gives the probability of a positive prediction given a positive sample (v_i^t, v_j^t). Likewise, the probability of observing a negative sample $(v_i^t, v_{j'}^t)$ can be defined as

$$P(B = 0|v_i^t, v_{j'}^t) = 1 - \sigma(Z_i^t \cdot Z_{j'}^t) = \frac{1}{1 + e^{-Z_i^t \cdot Z_{j'}^t}} = \sigma(-Z_i^t \cdot Z_{j'}^t) \tag{8}$$

where $P(B = 0|v_i^t, v_{j'}^t)$ gives the probability of a negative prediction given a negative sample $(v_i^t, v_{j'}^t)$. The above Skip-Gram Negative Sampling (SGNS) model [25] then try to maximize $P(B = 1|v_i^t, v_j^t)$ for each positive sample in $\mathcal{D}^t$ and $P(B = 0|v_i^t, v_{j'}^t)$ for the q negative samples corresponding to each positive sample, i.e.,

$$\max \;\; log\sigma(Z_i^t \cdot Z_j^t) + \sum\nolimits_q \mathbf{E}_{v_{j'}^t \sim P_{\mathcal{D}^t}} [log\sigma(-Z_i^t \cdot Z_{j'}^t)] \tag{9}$$

where q negative samples are drawn from a unigram distribution $P_{\mathcal{D}^t}$ [25]. The overall objective of SGNS is to sum over all positive samples and their corresponding negative samples, i.e.,

$$\max \sum_{(v_i^t, v_j^t) \in \mathcal{D}^t} \#(v_i^t, v_j^t) \;\; Eq.(9) \tag{10}$$

where $\#(v_i^t, v_j^t)$ denotes the number of times a positive sample occurs in $\mathcal{D}^t$. Intuitively, the more frequently a pair of nodes co-occurs, the closer their embeddings should be.

Finally, we extend the SGNS model as described above to an *incremental learning paradigm*. The overall framework of GloDyNE can be formalized as

$$Z^t = \begin{cases} f^0(G^0, f_{rand}^0, Z_{rand}^0) & t = 0 \\ f^t(G^t, G^{t-1}, f^{t-1}, Z^{t-1}) & t \geq 1 \end{cases} \tag{11}$$

where f^{t-1} is the trained SGNS model from last time step, Z^t is the current embedding matrix directly taken from newly trained f^t via an index operator, and G^{t-1} and G^t are the two consecutive snapshots for generating the edge steams $\Delta\mathcal{E}^t$ if not directly given. The implementation details are presented in Section 4.2 and 4.3.

Algorithm 1 GloDyNE

Input: snapshots of a dynamic network $G^0...G^{t-1}, G^t...$; coefficient to determine the number of selected nodes α; walks per node r; walk length l; sliding window size s; negative samples per positive sample q; embedding dimensionality d
Output: embedding matrix $Z^t \in \mathbb{R}^{|\mathcal{V}^t| \times d}$ at each time step

1: **for** $t = 0$ **do**
2: conduct random walks with length l starting from each node in $\mathcal{V}^0_{all}$ for r times by Eq. (5)
3: build positive node-pair samples $\mathcal{D}^0$ based on each sliding window with size s along each walk
4: initialize SGNS model f^0_{rand}, and train it using $\mathcal{D}^0$ with q negative samples per positive sample by Eq. (9)
5: return f^0 and Z^0
6: **for** $t \geq 1$ **do**
7: calculate $K = \alpha|\mathcal{V}^t|$
8: partition G^t into K sub-networks $G^t_1, G^t_2, ..., G^t_K$ by METIS based on Eq. (1) and Eq. (2)
9: read edge streams $\Delta\mathcal{E}^t$ (or obtain it by differences between G^{t-1} and G^t if not given)
10: update reservoir dictionary via $\mathcal{R}^t_{v_i} = |\Delta\mathcal{E}^t_i| + \mathcal{R}^{t-1}_i$ for accumulating new changes of v^t_i by Eq. (3)
11: **for** $k \in (1, ..., K)$ **do**
12: calculate a probability distribution over all nodes in a sub-network G^t_k by Eq. (4)
13: select one representative node based on the probability distribution, and add it to $\mathcal{V}^t_{sel}$
14: remove selected nodes $\mathcal{V}^t_{sel}$ from the reservoir $\mathcal{R}^t$ if exists
15: conduct random walks with length l starting from each node in $\mathcal{V}^t_{sel}$ for r times by Eq. (5)
16: build node-pair positive samples $\mathcal{D}^t$ based on each sliding window with size s along each walk
17: initialize SGNS model $f^t = f^{t-1}$, and train it using $\mathcal{D}^t$ with q negative samples per positive sample by Eq. (9)
18: return f^t and Z^t

4.2 Algorithm

The pseudocode of GloDyNE is summarized in Algorithm 1, and the open source code is provided at https://github.com/houchengbin/GloDyNE.

4.3 Complexity Analysis

According to Eq. (11) and Algorithm 1, GloDyNE consists of two stages. During the offline stage, i.e., $t = 0$, Step 3 (specifically $\mathcal{V}^t_{sel} = \mathcal{V}^0_{all}$) and Step 4 are employed to obtain the initial SGNS model and node embeddings. Lines 2-5 are indeed a static network embedding method–a modified version of DeepWalk [4], which trains a SGNS model instead of Skip-Gram Hierarchical Softmax (SGHS) model. As such, the time complexity of lines 2-5 is further reduced to $O(rlw(1+q)d|\mathcal{V}^0_{all}|)$ [7] where $(1+q)$ is due to one positive sample corresponding to q negative samples.

During the online stage, i.e., $t \geq 1$, steps 1-4 are employed to incrementally update the SGNS model and node embeddings. For lines 7-8 corresponding to Step 1, the time complexity is $O(|\mathcal{V}^t| + |\mathcal{E}^t| + KlogK)$ [20] where $K = \alpha|\mathcal{V}^t|, \alpha \in (0, 1)$. For lines 9-14 corresponding to Step 2, the time complexity of lines 9-10 is $O(|\Delta\mathcal{E}^t|)$; the time complexity of lines 11-13 using alias sampling method [7] requires $O(|\mathcal{V}^t|)$; the time complexity of line 14 is $O(\alpha|\mathcal{V}^t|)$ due to $|\mathcal{V}^t_{sel}| = K = \alpha|\mathcal{V}^t|$. For lines 14-18 corresponding to Step 3 and Step 4, similarly to lines 2-5 above, the complexity of lines 14-18 is $O(rlw(1+q)d|\mathcal{V}^t_{sel}|)$ where $|\mathcal{V}^t_{sel}| = \alpha|\mathcal{V}^t|$. Because most real-world networks are sparse, edges in a snapshot $|\mathcal{E}^t| = b_1|\mathcal{V}^t|$ such that the average degree b_1 is a very small number compared with $|\mathcal{V}^t|$. Besides, since the edge steams between two consecutive snapshots are often much less than the edges in the snapshot, edge steams $|\Delta\mathcal{E}^t| = b_2|\mathcal{V}^t|$ such that $b_2 < b_1 \ll |\mathcal{V}^t|$. Regarding those real-world assumptions, the overall complexity of online stage at each time step can be approximated as $O(|\mathcal{V}^t| + \alpha|\mathcal{V}^t|log\ \alpha|\mathcal{V}^t| + rlw(1+q)d\alpha|\mathcal{V}^t|)$ where $\alpha \in (0, 1)$ is used to control the number of selected nodes, $|\mathcal{V}^t|$ denotes the number of nodes at t, and others are negligible constants compared to $|\mathcal{V}^t|$. Consequently, GloDyNE is scalable w.r.t. $|\mathcal{V}^t|$, as there is no quadratic or higher term.

5 Empirical Studies

5.1 Experimental Settings

5.1.1 Datasets

In this work, six real-world datasets are employed to empirically evaluate the effectiveness and efficiency of the proposed method. *To construct the dynamic networks*, except AS733 (which is given as the snapshot representation), all other ones (which are the edge steams $\{(v_i, v_j, timestamp), ...\}$) are constructed by continuously adding edge steams to the existing network based on the ascending order of timestamps. The snapshots of a dynamic network are then taken at the specified timestamps successively, and the gap between the specified timestamps are identical. For each of six dynamic networks, the largest connected component is finally used in experiments, and the details are as follows.

AS733 contains 733 daily instances of the Autonomous System of routers exchanging traffic flows with neighbors. Since AS733 is directly given as the snapshot representation, we directly take out the recent 21 snapshots (13/Dec./1991–02/Jan./2000) to form its dynamic network. The initial snapshot has 1476 nodes and 3123 edges, and the final snapshot has 3570 nodes and 7033 edges. The original dataset comes from https://snap.stanford.edu/data/as-733.html.

Elec is the network of English Wikipedia users vote for and against each other in admin elections. The gap between the specified timestamps for taking snapshots is set to one calendar day. We take out the recent 21 snapshots (16/Dec./2007–05/Jan./2008) to form its dynamic network.

The initial snapshot has 6968 nodes and 98947 edges, and the final snapshot has 7058 nodes and 100521 edges. The original dataset comes from http://konect.uni-koblenz.de/networks/elec.

HepPh is a co-author network extracted from the papers of High Energy Physics Phenomenology in the arXiv. The gap between the specified timestamps for taking snapshots is set to one calendar day. We take out the recent 21 snapshots (11/Dec./1991–31/Dec./1991) to form its dynamic network. The initial snapshot has 16729 nodes and 1170677 edges, and the final snapshot has 16910 nodes and 1194275 edges. The original dataset comes from http://konect.uni-koblenz.de/networks/elec.

FBW is a social network of Facebook Wall posts where nodes are the users and edges are built based on the interactions in their wall posts. The gap between the specified timestamps for taking snapshots is set to one calendar day. We take out the 21 snapshots (01/Jan./2009–21/Jan./2009) to form its dynamic network. The initial snapshot has 41603 nodes and 169427 edges, and the final snapshot has 43889 nodes and 181974 edges. The original dataset comes from http://konect.uni-koblenz.de/networks/facebook-wosn-wall.

Cora is a citation network where each node represent a paper, a edge between two nodes representation a citation. Each paper is assigned with a label (from xx different labels) based on its field of the publication. Following [18], the gap between the specified timestamps for taking snapshots is set to one year. The 11 snapshots (1989–1999) are taken out to form its dynamic network. The initial snapshot has 348 nodes and 481 edges, and the final snapshot has 12022 nodes and 45421 edges. The original dataset comes from https://people.cs.umass.edu/~mccallum/data.html.

DBLP is a co-author network in computer science field. Each author is associated with a label (from 15 different labels). The label of an author is defined by the fields in which the author has the most publications. Following [18], the gap between the specified timestamps for taking snapshots is set to one year. The 12 snapshots (1984–1995) are taken out to form its dynamic network. The initial snapshot has 391 nodes and 848 edges, and the final snapshot has 24157 nodes and 53431 edges. The original dataset comes from https://dblp.org/xml/release/.

5.1.2 Methods

The proposed DNE method–GloDyNE is compared with five state-of-the-art DNE methods for demonstrating the effectiveness and efficiency. All the compared methods can be regarded as the unsupervised approach, because no explicit node label is used to learn node embedding, and the learned node embeddings are not dedicated to a specific downstream task. The details of each method are as follows.

BCGD^g (2016) [9]: The general objective of BCGD is to minimize the quadratic loss of reconstructing the network proximity matrix using the node embedding matrix with a temporal regularization term. BCGD^g (or BCGD-global) employs all historical snapshots to jointly and cyclically update embeddings for all time steps.

BCGD^l (2016) [9]: Unlike BCGD^g but following the same general objective of BCGD as above, BCGD^l (or BCGD-local) iteratively employs the previous snapshot and initializes current embeddings with the previous embeddings to update embeddings for current time step.

DynAE (2017) [11]: This work proposes a strategy to modify the structure of a deep auto-encoder model based on the size of a current snapshot. At each time step, the auto-encoder model is initialized by its previous model. DynAE continuously trains the adaptive auto-encoder model based on the existing edges in a current snapshot.

DynTriad (2018) [15]: DynTriad models the triadic closure process, social homophily, and temporal smoothness in its objective function to learn node embeddings at each time step. It optimizes the objective function according to the existing edges of each snapshot respectively.

tNodeEmbed (2019) [18]: tNodeEmbed runs a static Network Embedding method to obtain node embeddings at each time step, and then exploits the temporal dependence among all currently available static node embeddings using Recurrent Neural Networks to obtain the final node embeddings for current time step.

The original open source codes with the default settings of BCGD[4], DynAE[5], DynTriad[6], and tNodeEmbed[7] are adopted in the experiments. Note that, BCGD^g and BCGD^l are two proposed algorithms in BCGD correspondingly to the type of algorithm 2 and 4. Moreover, we adopt the link prediction architecture of tNodeEmbed to obtain node embeddings, so that all methods only use network linkage information as the supervised signal to learn node embeddings. Furthermore, for fair comparison, the dimensionality of node embeddings is set to 128 for all methods.

Regarding our method–GloDyNE, following [4] and [7], the hyper-parameters of walks per node, walk length, window size, and negative samples are set to 10, 80, 10, and 5 respectively. The hyper-parameter α to control the number of selected nodes for freely trade-off between effectiveness and efficiency, is set to 0.1 unless otherwise specified.

5.2 Comparative Studies of Different Methods

In this section, three typical types of downstream tasks are employed to evaluate the quality of obtained node embeddings by the six methods on the six datasets. In particular, the *graph reconstruction* task is used to demonstrate the ability of global topology preservation, while the *link prediction* task and *node classification* task are used to show the benefit of global topology preservation. For fairness, we first take out the node embeddings obtained by each method respectively, and then feed them to exactly the same downstream tasks with the same training and testing sets. The above process is repeated for 10 runs, and we report their average results in Section 5.2.1, 5.2.2, and 5.2.3. Moreover, the average results of the *wall-clock time* to obtain node embeddings by each method, are reported in Section 5.2.4 for comparing the efficiency of the implementation of the six methods.

All experiments in Section 5.2 are conducted in the following hardware specification. For all methods, we enable

4. https://github.com/linhongseba/Temporal-Network-Embedding
5. https://github.com/palash1992/DynamicGEM
6. https://github.com/luckiezhou/DynamicTriad
7. https://github.com/urielsinger/tNodeEmbed

32 Intel-Xeon-E5-2.2GHz CPUs and 512G memory. In addition, for DynAE, DynTriad, and tNodeEmbed that can use GPU for acceleration, we also enable 1 Nvidia-Tesla-P100 GPU with 16G memory. The N/A values for tNodeEmbed on AS733 are due to that tNodeEmbed cannot handle node deletions. The N/A values for DynAE on HepPh, DBLP, and FBW are because of running out of GPU memory.

5.2.1 Graph Reconstruction (GR)

In order to demonstrate the ability of the global topology preservation of each method, one possible way is to use the obtained node embeddings to reconstruct the original network. For this purpose, precision at k or $P@k$ is used as the metric to evaluate how well the top-k similar nodes of each node in the embedding space can match the ground-truth neighbors of each node in the original network [6], [12], [24]. Concretely, $P@k(v_i) = |\mathcal{Q}(v_i)_{@k} \cap \mathcal{N}(v_i)| \,/\, min(k, |\mathcal{N}(v_i)|)$, where $\mathcal{Q}(v_i)_{@k}$ gives a set of the top-k similar nodes of a queried node v_i based on the cosine similarity between node embeddings, and $\mathcal{N}(v_i)$ denotes a set of the ground-truth neighbors of v_i. To show the ability of global topology preservation, we further calculate the mean of $P@k$ over all nodes in a current snapshot, i.e., Mean$P@k$ $=$ $[\ \sum_{v_i^t \in \mathcal{V}^t} P@k(v_i^t)\ ] \,/\, |\mathcal{V}^t|$ where $\mathcal{V}^t$ is a set of all nodes in a current snapshot G^t, and $|\mathcal{V}^t|$ counts the number of nodes in $\mathcal{V}^t$. Note that, each result shown in Table 1 is calculated by the mean of Mean$P@k$ over all time steps and over 10 runs, and finally, Mean$P@5$, Mean$P@10$, Mean$P@20$, and Mean$P@40$ are employed.

In general, GloDyNE significantly outperforms all other methods on all datasets, except that it obtains the second best result on Elec dataset under Mean$P@5$ metric[8]. Specifically, Mean$P@5$ measures how well the top-5 similar nodes of each node in the embedding space can match the ground-truth neighbors of each node in the original network. For Elec dataset, DynTriad outperforms GloDyNE by 2.81% under Mean$P@5$, however, GloDyNE outperforms DynTriad by 3.84%, 5.28%, 4.85% under Mean$P@10$, Mean$P@20$, Mean$P@40$ respectively. For other five datasets, it is easy to verify that GloDyNE consistently achieves the best results (often with a large margin over 10%) compared with all other methods under all four metrics.

The main reason of such superiority of GloDyNE in the GR task is due to that GloDyNE is designed to better preserve the global topology of a dynamic network at each time step, while the GR task is also used for demonstrating the ability of global topology preservation.

5.2.2 Link Prediction (LP)

The (dynamic) LP task aims to predict future edges at time step $t+1$ using the obtained node embeddings at t. The testing edges include both added and removed edges from t to $t+1$, plus other edges randomly sampled from the snapshot at $t+1$ for balancing existent edges (or positive samples) and non-existent edges (or negative samples). The LP task is then evaluated by Area under the ROC Curve (AUC) score based on the cosine similarity between node embeddings [9], [26], [27]. Each result shown in Table 2 is calculated by the mean of AUC scores over all time steps and over 10 runs.

8. The tendency of recall at k or $R@k$ is exactly the same as $P@k$, since the only difference between them is in the denominator. For $R@k(v_i)$, the denominator is $|\mathcal{N}(v_i)|$.

TABLE 1
Mean$P@k$ scores of graph reconstruction tasks. Each result (in %) is calculated by the mean of Mean$P@k$ over all time steps and over 10 runs. The N/A values for tNodeEmbed and DynAE are due to inability of handling node deletions and running out of memory respectively.

	AS733	Elec	Cora	HepPh	DBLP	FBW
			Mean$P@5$			
BCGD^g	1.62	8.32	11.78	28.06	5.92	0.14
BCGD^l	38.63	17.35	8.12	58.95	2.62	4.80
DynAE	0.55	3.36	10.17	N/A	N/A	N/A
DynTriad	58.99	**60.44**	51.77	62.06	69.33	54.15
tNodeEmbed	N/A	5.62	58.86	50.12	64.76	25.28
GloDyNE	**65.54**	57.63	**76.60**	**65.11**	**77.00**	**80.03**
			Mean$P@10$			
BCGD^g	2.51	9.32	21.06	31.50	16.95	0.12
BCGD^l	49.55	17.62	9.69	61.44	5.20	4.40
DynAE	0.59	3.62	10.39	N/A	N/A	N/A
DynTriad	65.80	64.55	55.09	66.43	74.79	54.39
tNodeEmbed	N/A	5.81	69.44	50.12	80.09	26.40
GloDyNE	**76.56**	**68.39**	**88.03**	**69.32**	**93.02**	**87.06**
			Mean$P@20$			
BCGD^g	43.70	9.23	29.55	32.85	30.56	0.11
BCGD^l	66.52	18.42	15.55	59.76	13.36	3.86
DynAE	0.63	3.64	11.81	N/A	N/A	N/A
DynTriad	72.38	67.92	60.19	66.95	79.26	56.60
tNodeEmbed	N/A	5.90	79.36	47.20	88.39	28.85
GloDyNE	**84.64**	**73.20**	**96.27**	**69.92**	**98.85**	**91.24**
			Mean$P@40$			
BCGD^g	89.29	8.64	40.15	32.39	43.73	0.12
BCGD^l	80.51	26.39	24.87	57.45	29.39	4.11
DynAE	0.78	3.60	14.17	N/A	N/A	N/A
DynTriad	78.74	71.74	65.92	66.75	83.26	60.68
tNodeEmbed	N/A	6.36	84.72	43.22	91.40	32.61
GloDyNE	**90.51**	**76.59**	**98.76**	**70.40**	**99.86**	**94.84**

TABLE 2
AUC scores of (dynamic) link prediction tasks. Each result (in %) is calculated by the mean of AUC over all time steps and over 10 runs.

	AS733	Elec	Cora	HepPh	DBLP	FBW
BCGD^g	69.96	82.65	68.37	79.74	66.71	82.26
BCGD^l	62.24	87.69	80.94	**89.23**	**88.25**	83.41
DynAE	59.70	60.56	59.09	N/A	N/A	N/A
DynTriad	64.69	**94.87**	62.74	87.60	59.99	77.58
tNodeEmbed	N/A	80.88	51.62	85.04	56.82	73.37
GloDyNE	**81.94**	87.20	**93.59**	88.84	77.05	**88.00**

GloDyNE obtains either the best result (on 3 datasets) or the acceptable result compared to other methods. Specifically, GloDyNE outperforms the second best method on AS733, Cora, and FBW by 11.98%, 12.56%, and 4.59% respectively. For other three datasets, GloDyNE obtains the third best result on Elec (only 0.49% gap from the second best result), the second best result on HepPh (only 0.39% gap from the best result), and the second best result on DBLP.

Overall, GloDyNE is also a good method for the (dynamic) LP task on most datasets, thanks to the high-order proximities being used for better preserving the global

topology [6]. In fact, the high-order proximity between nodes is an important temporal feature for predicting future edges. For example, the triadic closure process which tries to predict the third edge among three nodes if there have already been two edges among them, as modelled in DynTriad [15], can be easily realized by considering the second-order proximity via setting $l \geq 3$ and $s \geq 3$ (see Section 4.1.3 and 4.1.4). In the experiments, we set $l = 80$ and $s = 10$. As a result, much higher order proximities (up to 10^{th} order according to s) are considered for better preserving the global topology, which therefore provides more advanced temporal features (analogous to triadic closure process) to improve the performance of GloDyNE in LP tasks on most datasets. However, this kind of temporal features might not be always very useful, e.g., both GloDyNE and DynTriad receive degraded performance on DBLP dataset.

5.2.3 Node Classification (NC)

The NC task aims to infer the most likely label for the nodes without labels. Specifically, 50%, 70%, and 90% nodes are randomly picked respectively to train a one-vs-rest logistic regression classifier based on their embeddings and labels. The left nodes respectively are treated as the testing set. At each time step, the latest node embeddings are employed as the input features to logistic regression classifier. The prediction of the trained classifier over the testing set are evaluated by Micro-F1 and Macro-F1 [4], [7], [18] respectively. Each result shown in Table 3 is calculated by the mean of a F1 metric over all time steps and over 10 runs.

TABLE 3
Micro-F1 and Macro-F1 scores of node classification tasks. Each result (in %) is calculated by the mean of a F1 metric over all time steps and over 10 runs. Three different proportions of training set, 0.5, 0.7, and 0.9, are evaluated respectively.

	Cora			DBLP		
	0.5	0.7	0.9	0.5	0.7	0.9
Micro-F1						
BCGD^g	32.12	32.82	31.99	49.63	49.41	50.35
BCGD^l	36.76	37.15	37.28	50.91	49.96	51.03
DynAE	33.74	34.83	34.64	N/A	N/A	N/A
DynTriad	35.91	35.85	36.17	50.84	50.91	51.17
tNodeEmbed	66.28	65.53	65.19	58.82	58.01	58.45
GloDyNE	**73.88**	**73.87**	**73.89**	**59.49**	**59.27**	**59.93**
Macro-F1						
BCGD^g	7.95	8.48	8.02	10.29	10.15	10.19
BCGD^l	12.20	12.38	12.59	11.36	11.24	11.27
DynAE	7.04	7.49	7.61	N/A	N/A	N/A
DynTriad	15.61	15.92	16.23	14.39	14.87	14.46
tNodeEmbed	52.00	51.40	51.81	23.91	23.69	23.96
GloDyNE	**60.76**	**62.20**	**61.09**	**26.60**	**27.39**	**26.57**

It is obvious that GloDyNE obtains the best result for all cases on Cora and DBLP datasets, which shows the benefit of global topology preservation in NC tasks. Comparing Cora and DBLP, GloDyNE achieves better performance on Cora than DBLP. The reason is that Cora is a citation network where the label/field of nodes/papers contains less noise (the field of a journal or conference often remains the same), while DBLP is a co-author network where the label/field of nodes/authors contains more noise (the field of an author varies over time or an author with few papers is not accurate). Note that, the approach to construct the dynamic network of Cora and DBLP, and to generate node labels are described in Section 5.1.1.

5.2.4 Walk-Clock Time During Embedding

To conduct the downstream tasks in Section 5.2.1, 5.2.2, and 5.2.3, the common step is to first obtain node embeddings which serve as the *low dimensional hidden features* of each node in the downstream tasks. In this section, the wall-clock time (or running time) of obtaining node embeddings over all time steps are reported, and each result shown in Table 4 is given by the mean over 10 runs.

TABLE 4
Wall-clock time (in seconds) of obtaining node embeddings (not including downstream tasks) over all time steps. Each result is given by the mean over 10 runs. The total number of nodes and edges of a dynamic network over all snapshots is also attached.

	AS733	Elec	Cora	HepPh	DBLP	FBW
BCGD^g	2402	6111	3868	25972	6682	40975
BCGD^l	704	2314	1543	15228	2878	13569
DynAE	536	9473	1126	N/A	N/A	N/A
DynTriad	156	2546	233	34718	377	4458
tNodeEmbed	N/A	3734	892	9224	1377	48631
GloDyNE	**62**	**205**	**105**	**1254**	**155**	**923**
# of nodes	45k	147k	66k	353k	92k	900k
# of edges	91k	2093k	216k	24888k	202k	3690k

According to Table 4, GloDyNE is the most efficient method among all methods on all datasets. In addition, the superiority of efficiency of GloDyNE grows, as the size of a dynamic network (given by the number of nodes or edges over all snapshots) grows. There are two reasons to explain the observations: 1) GloDyNE is scalable since there is no quadratic or higher term appeared in $|\mathcal{V}|$ and $|\mathcal{E}|$ as analyzed in Section 4.3; and 2) the Step 3 and Step 4 in Section 4.1 are parallelized in the implementation of GloDyNE.

5.2.5 Effectiveness and Efficiency

To better visualize the comparison among the six methods in terms of both effectiveness and efficiency, we further make the scatter plots as shown in Figure 2 based on the quantitative results in Section 5.2.1, 5.2.2, 5.2.3, and 5.2.4. Note that, the experiments in these sections are conducted under the fair hardware environment as mentioned in Section 5.2. Besides, the N/A values shown in the tables are now located at the most bottom-left corner (i.e., the origin) such as t-NodeEmbed on AS733 in GR-Mean$P$@5 task and DynAE on HepPh in LP-AUC task.

It is worth noticing that Figure 2 only shows the two tasks, i.e., GR-Mean$P$@5 and LP-AUC for further illustration, since GloDyNE does not always obtain the best results in terms of effectiveness on all datasets in the two tasks. Concretely, we make the following observations according to Figure 2, Table 1, Table 2, and Table 4. Firstly, overall, GloDyNE is the *best choice* for the all twelve sub-figures, if one prefers efficiency (ranked at top-1 for all cases) as well as considers effectiveness (ranked at top-1 for most cases and at least top-3 for all cases). Secondly, in the GR-Mean$P$@5 task, GloDyNE is outperformed by DynTriad on Elec by

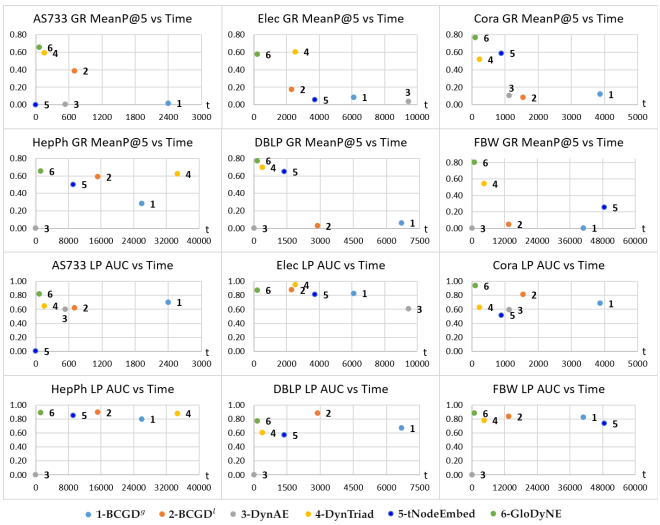

Fig. 2. The comparison among the six methods in terms of both effectiveness (y-axis for scores in decimal) and efficiency (x-axis for wall-clock time in seconds). Note that, we only show the two tasks, i.e., GR-Mean$P@5$ and LP-AUC for further illustration, since GloDyNE does not always obtain the best results in terms of effectiveness on all datasets. But for other nine tasks such as GR-Mean$P@10$ and NC-0.5-Micro-F1, GloDyNE is always the best choice in terms of both effectiveness and efficiency on all datasets, and hence, their plots are omitted.

2.81%. However, GloDyNE is $\times 12.4$ faster than DynTriad. Thirdly, in the LP-AUC task, GloDyNE is outperformed by DynTriad, BCGD^l and BCGD^l on Elec, HepPh and DBLP by 7.67%, 0.39% and 11.2% respectively. However, GloDyNE is $\times 12.4$, $\times 12.1$ and $\times 18.6$ faster than DynTriad, BCGD^l and BCGD^l on Elec, HepPh and DBLP respectively.

Last but not least, it is also worth noticing that, for other nine tasks such as GR-Mean$P@10$ and NC-0.5-Micro-F1, although their plots are omitted, one can easily image that GloDyNE is always located at the most top-left corner, i.e., it is always the best choice in terms of both effectiveness and efficiency on all datasets in all nine tasks.

5.3 Further Investigations of Proposed Method

In this section, we further investigate the proposed method–GloDyNE. Because GloDyNE is proposed to better preserve the global topology, we will focus on the ability of global topology preservation, and thus adopt the *graph reconstruction* task to quantify the effectiveness. Besides, due to the good time and space efficiency of GloDyNE and its variants, all experiments in this section are conducted with less expensive hardware specification: 16 Intel-Xeon-E5-2.2GHz CPUs and 8G memory. All experiments are repeated for 20 runs, and their average results are reported.

5.3.1 Necessity of Dynamic Network Embedding

One advantage of DNE is that, it promptly updates node embeddings at each time step, so that the latest node embeddings can better reflect the original network topology at each time step. To demonstrate this point, two variants of GloDyNE based on the SGNS model, namely SGNS-static and SGNS-retrain, are used for comparison. For SGNS-static, we only perform the $t = 0$ part of Algorithm 1, and the obtained node embeddings at $t = 0$ will be identically used in the downstream task at each time step. For SGNS-retrain, we repeatedly perform the $t = 0$ part of Algorithm 1 at each time step, and the obtain node embeddings at each time step will be used in the downstream task at each time step respectively.

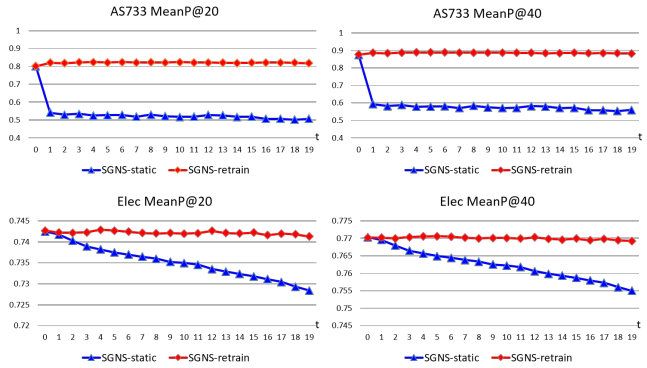

Fig. 3. SGNS-static vs SGNS-retrain in graph reconstruct tasks for showing the necessity of dynamic network embedding: y-axis indicates Mean$P@k$ scores; x-axis indicates time steps; and each point depicts the average result over 20 runs at a time step.

According to Figure 3, SGNS-retrain outperforms SGNS-static on both tested datasets. For AS733, SGNS-retrain maintains the performance at a superior level all the time under Mean$P@20$ and Mean$P@40$, whereas the performance of SGNS-static suddenly decreases at $t = 1$ and then maintains a poor level afterward. For Elec, SGNS-retrain maintains the performance at a superior level all the time, whereas the performance of SGNS-static gradually decreases. The difference of the sudden drops on AS733 and the gradual drops on Elec is due to that the network topology between two consecutive time steps on AS733 varies more severely than on Elec (see Section 5.1.1), so that the obtained node embeddings at $t = 0$ is less useful afterward. Consequently, it is needed to promptly update node embeddings at each time step (i.e., the necessity of DNE) as what SGNS-retrain–the naive DNE method does.

5.3.2 Incremental Learning vs Retraining

Instead of SGNS-retrain, recent DNE methods often adopt the incremental learning paradigm by continuously training the previous model on a new training set. Accordingly, another baseline–SGNS-increment, which follows Algorithm 1 but replaces all operations in lines 4-17 with $\mathcal{V}^t_{sel} = \mathcal{V}^t$. The difference between SGNS-increment and SGNS-retrain is whether they reuse the previous model as the initialization of next model.

According to Figure 4, SGNS-increment outperforms SGNS-retrain on both tested datasets. The general tendency on the two datasets under the two metrics are the same, although the performances of SGNS-increment and SGNS-retrain are both less stable on AS733 than on Elec, due to the larger variations between two consecutive snapshots on AS733 than on Elec (see Section 5.1.1). These observations show that reusing the previous model as the initialization of next model might be not only useful for a dynamic network with small dynamic changes (e.g., Elec), but also useful for a dynamic network with large dynamic changes (e.g., AS733).

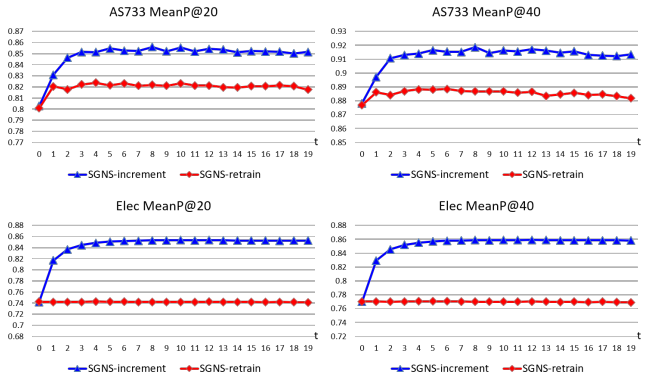

Fig. 4. SGNS-increment vs SGNS-retrain in graph reconstruct tasks for showing the advantage of reusing previous models: y-axis indicates Mean$P@k$ scores; x-axis indicates time steps; and each point depicts the average result over 20 runs at a time step.

5.3.3 Different Node Selecting Strategies

According to Figure 3 and Figure 4, the ranking of performances among the three baselines is as follows (from high to low): SGNS-increment, SGNS-retrain, and SGNS-static. Although, SGNS-increment (i.e., GloDyNE with $\alpha = 1.0$) achieves the best performance, it is not efficient enough since all nodes in a current snapshot are selected for conducting random walks and then training the SGNS model. One natural idea for further improving the efficiency is to select some representative nodes as the *approximate solution*, such that it can significantly reduce the running time but meanwhile, still retain a good performance. Consequently, in this work, we propose a node selecting strategy, denoted as $S4$, as described in Section 4.1.1 and Section 4.1.2.

In order to show the advantage of $S4$ used in GloDyNE, the following baselines with different node selecting strategies are used for comparison. For fairness, the number of selected nodes at each time step is set to $\alpha|\mathcal{V}^t| = 0.1|\mathcal{V}^t|$ for all strategies. Concretely, $S1$ selects the nodes randomly with replacement from the reservoir $\mathcal{R}^t$ which records the most affected nodes (see Section 4.1.2); $S2$ selects the nodes randomly without replacement from $\mathcal{R}^t$ and then, from all nodes in a current snapshot if $|\mathcal{R}^t| < 0.1|\mathcal{V}^t|$; $S3$ selects the nodes randomly without replacement from all nodes in a current snapshot. Intuitively, from the perspective of the *diversity of selected nodes*, $S1 < S2 < S3 < S4$ due to 1) sampling nodes from $\mathcal{R}^t$ cannot be aware of inactive sub-networks which exist in many real-world dynamic networks; 2) sampling nodes from all nodes in a current snapshot cannot guarantee the selected nodes have an enough distance from each other; 3) sampling one node from each sub-network after network partition as introduced in $S4$, however, can ensure the selected nodes have an enough distance from each other.

To compare the performance of GloDyNE with different node selecting strategies, the length of random walks l (see Section 4.1.3) should be also considered. Because as l increases, the generated random walks (or node sequences) become less distinguishable. An extreme case is that, if l goes to infinity, a random walker starting from any node in a network can well explore its global topology. As a result, we compare the four different node selecting strategies w.r.t. different ls as shown in Table 5.

TABLE 5
The performance of GloDyNE with different node selecting strategies w.r.t. different length of random walks in graph reconstruction tasks. Each result (in %) is calculated by the mean of Mean$P@k$ over all time steps and over 20 runs. The two-tailed and two-sample equal variance T-Test is applied to $S3$ and $S4$. * and ** indicate the p-value of T-Test < 0.05 and < 0.01 respectively. Note that, $S4$ is the proposed node selecting strategy of this work.

	AS733				Elec			
	$S1$	$S2$	$S3$	$S4$	$S1$	$S2$	$S3$	$S4$
l	Mean$P@20$							
3	17.71	22.78	24.74	**25.74****	4.17	5.83	6.05	**6.11***
5	45.52	46.84	47.49	**47.85****	6.12	10.57	11.01	**11.27****
8	51.47	51.72	51.81	**51.88**	8.31	13.54	14.16	**14.41****
10	52.02	52.30	52.54	**52.56**	9.68	16.54	17.31	**17.65****
15	56.79	57.23	57.58	**57.72**	17.09	26.69	27.47	**27.70****
20	63.34	63.82	64.38	**64.40**	28.71	37.35	37.84	**37.97**
30	73.04	73.31	73.66	**73.69**	49.09	52.90	53.04	**53.16***
40	77.79	78.19	**78.63**	78.45	58.93	60.80	60.89	**60.90**
50	80.49	80.68	**81.14**	81.13	64.40	65.49	65.58	**65.64**
60	82.05	82.34	**82.80**	82.70	68.08	68.81	68.84	**68.89**
70	83.24	83.61	83.93	**83.94**	70.80	71.34	71.38	**71.39**
80	84.34	84.62	**84.92**	**84.92**	72.99	73.34	73.35	**73.40***
l	Mean$P@40$							
3	21.38	27.29	29.38	**30.35****	4.90	6.49	6.73	**6.79***
5	53.06	54.28	54.67	**55.02***	5.81	11.30	11.83	**12.14****
8	58.60	58.96	59.08	**59.15**	7.42	14.30	15.07	**15.37****
10	60.01	60.38	60.51	**60.64***	9.17	18.34	19.31	**19.71****
15	66.21	66.68	**67.36**	**67.36**	19.69	31.60	32.46	**32.73***
20	73.31	73.65	**74.37**	74.31	35.03	44.23	44.64	**44.75**
30	81.62	81.99	82.50	**82.52**	56.99	60.14	60.24	**60.32***
40	85.37	85.87	**86.36**	86.31	65.81	67.06	67.12	**67.13**
50	87.48	87.79	88.19	**88.21**	70.27	70.85	70.86	**70.92***
60	88.83	89.09	**89.35**	89.29	73.11	73.39	73.38	**73.40**
70	89.75	89.98	90.07	**90.11**	75.13	75.24	75.25	**75.27**
80	90.57	90.64	90.73	**90.75**	76.71	76.70	76.69	**76.71**

There are three important observations from Table 5. Firstly, the overall ranking of the performance among four strategies under a same l is $S1 < S2 < S3 < S4$, which exactly matches the ranking of the diversity of selected nodes among them as discussed above. Secondly, as l increases, the four strategies become less distinguishable, which verifies the analysis above of four strategies w.r.t. l. Thirdly, under a smaller l, $S4$ statistically significantly[9] outperforms $S3$ on both datasets, while as l increases, $S4$ is still a better choice than $S3$ on Elec (but $S4$ and $S3$ become less distinguishable on AS733). It suggests that using $S4$ with GloDyNE on a lager dataset (Elec is larger than AS733 as shown in Table 4) might gain more benefits.

5.3.4 The Free Hyper-Parameter

The hyper-parameter α, which determines the number of selected nodes of GloDyNE at each time step, is designed for freely trade-off between effectiveness and efficiency. We vary α from 0.1 to 1.0 with step 0.1, together with four additional small values: 0.001, 0.005, 0.01, and 0.05. Each bar as shown in Figure 5 has two results: the blue one shows the effectiveness which is measured by the mean of Mean$P@k$

9. Two-tailed and two-sample equal variance T-Test is applied to $S3$ and $S4$ with the null hypothesis that there is no statistically significant difference of the mean over 20 runs between $S3$ and $S4$.

over all time steps and over 20 runs; while the red one shows the efficiency which is measured by the mean over 20 runs of the total wall-clock time during all time steps.

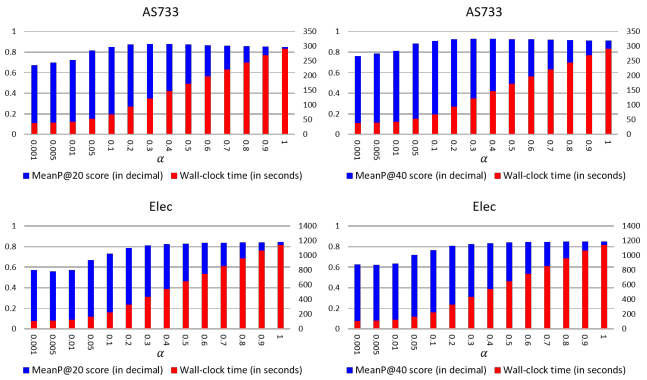

Fig. 5. The effectiveness (in blue corresponding to the left y-axis) and efficiency (in red corresponding to the right y-axis) of GloDyNE w.r.t. different α which determines the number of selected nodes.

According to Figure 5, it obviously demonstrates that the hyper-parameter α can be used to freely compromise between effectiveness and efficiency. Note that, all experiments in the above sections set $\alpha = 0.1$, which indicates one can obtain better results by increasing α at the risk of consuming more running time. Besides, with this free hyper-parameter, one can compromise between effectiveness and efficiency to fulfill the real-world requirements, e.g., have to promptly update node embeddings within a specified period.

Furthermore, an interesting observation is that increasing α to a certain level obtains a very competitive performance as $\alpha = 1.0$ (GloDyNE with $\alpha = 1.0$ is equivalent to SGNS-increment), but consumes much less running time. This observation also supports that GloDyNE especially the proposed node selecting strategy that selects partial nodes, makes a good approximation to SGNS-increment that selects all nodes for further computation.

6 Conclusion

This work proposed a new DNE method–GloDyNE, which aims to efficiently update node embeddings while better preserving the global topology of a dynamic network at each time step, by extending the SGNS model to an incremental learning paradigm. In particular, unlike all previous DNE methods, a novel node selecting strategy is proposed to diversely select the representative nodes over a network, so as to additionally considers the inactive sub-networks for better global topology preservation. The extensive experiments not only confirmed the effectiveness and efficiency of GloDyNE w.r.t. other five state-of-the-art DNE methods, but also verified the usefulness of some special designs or considerations in GloDyNE.

From a high-level view, GloDyNE can also be seen as a general DNE framework based on the incremental learning paradigm of SGNS model. Under this framework, one may design a different node selecting strategy to preserve other desirable topological features into node embeddings for a specific application. On the other hand, the idea of selecting diverse nodes could be adapted to other existing DNE methods for better global topology preservation. Besides, one more future work, according to Figure 5, is to further investigate why selecting partial nodes can receive almost the same performance or even the superior performance compared to selecting all nodes.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2017YFB1003102), the Natural Science Foundation of China (Grant No. 61672478), the Program for Guangdong Introducing Innovative and Enterpreneurial Teams (Grant No. 2017ZT07X386), the Shenzhen Peacock Plan (Grant No. KQTD2016112514355531) and the National Leading Youth Talent Support Program of China.

References

[1] P. Cui, X. Wang, J. Pei, and W. Zhu, "A survey on network embedding," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 5, pp. 833–852, 2018.

[2] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation learning on graphs: Methods and applications," *IEEE Data Eng. Bull.*, vol. 40, no. 3, pp. 52–74, 2017.

[3] P. Goyal and E. Ferrara, "Graph embedding techniques, applications, and performance: A survey," *Knowledge-Based Systems*, vol. 151, pp. 78–94, 2018.

[4] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: online learning of social representations," in *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2014, pp. 701–710.

[5] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "LINE: large-scale information network embedding," in *International Conference on World Wide Web (WWW)*, 2015, pp. 1067–1077.

[6] S. Cao, W. Lu, and Q. Xu, "Grarep: Learning graph representations with global structural information," in *Proceedings of the 24th ACM international on conference on information and knowledge management*. ACM, 2015, pp. 891–900.

[7] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2016, pp. 855–864.

[8] M. Ou, P. Cui, J. Pei, Z. Zhang, and W. Zhu, "Asymmetric transitivity preserving graph embedding," in *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2016, pp. 1105–1114.

[9] L. Zhu, D. Guo, J. Yin, G. V. Steeg, and A. Galstyan, "Scalable temporal latent space inference for link prediction in dynamic social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 10, pp. 2765–2777, 2016.

[10] J. Li, H. Dani, X. Hu, J. Tang, Y. Chang, and H. Liu, "Attributed network embedding for learning in a dynamic environment," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017.

[11] P. Goyal, N. Kamra, X. He, and Y. Liu, "Dyngem: Deep embedding method for dynamic graphs," in *IJCAI International Workshop on Representation Learning for Graphs*, 2017.

[12] D. Zhu, P. Cui, Z. Zhang, J. Pei, and W. Zhu, "High-order proximity preserved embedding for dynamic networks," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 11, pp. 2134–2144, 2018.

[13] Z. Zhang, P. Cui, J. Pei, X. Wang, and W. Zhu, "TIMERS: error-bounded SVD restart on dynamic networks," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

[14] L. Du, Y. Wang, G. Song, Z. Lu, and J. Wang, "Dynamic network embedding : An extended approach for skip-gram based network embedding," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI)*, 2018.

[15] L. Zhou, Y. Yang, X. Ren, F. Wu, and Y. Zhuang, "Dynamic network embedding by modeling triadic closure process," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

[16] X. Chen, P. Cui, L. Yi, and S. Yang, "Scalable optimization for embedding highly-dynamic and recency-sensitive data," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018, pp. 130–138.

[17] S. Mahdavi, S. Khoshraftar, and A. An, "dynnode2vec: Scalable dynamic network embedding," in *IEEE International Conference on Big Data (Big Data)*, 2018, pp. 3762–3765.
[18] U. Singer, I. Guy, and K. Radinsky, "Node embedding over temporal graphs," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI)*, 2019.
[19] R. Trivedi, M. Farajtabar, P. Biswal, and H. Zha, "Dyrep: Learning representations over dynamic graphs," in *International Conference on Learning Representations*, 2019.
[20] G. Karypis and V. Kumar, "A fast and high quality multilevel scheme for partitioning irregular graphs," *SIAM Journal on scientific Computing*, vol. 20, no. 1, pp. 359–392, 1998.
[21] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems (NIPS)*, 2013, pp. 3111–3119.
[22] W. Yu, W. Cheng, C. C. Aggarwal, K. Zhang, H. Chen, and W. Wang, "Netwalk: A flexible deep embedding approach for anomaly detection in dynamic networks," in *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2018.
[23] A. Buluç, H. Meyerhenke, I. Safro, P. Sanders, and C. Schulz, "Recent advances in graph partitioning," in *Algorithm Engineering*. Springer, 2016, pp. 117–158.
[24] Z. Zhang, P. Cui, X. Wang, J. Pei, X. Yao, and W. Zhu, "Arbitrary-order proximity preserved network embedding," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018, pp. 2778–2786.
[25] O. Levy and Y. Goldberg, "Neural word embedding as implicit matrix factorization," in *Advances in Neural Information Processing Systems (NIPS)*, 2014, pp. 2177–2185.
[26] G. Fu, C. Hou, and X. Yao, "Learning topological representation for networks via hierarchical sampling," in *2019 International Joint Conference on Neural Networks (IJCNN)*, July 2019, pp. 1–8.
[27] L. Liao, X. He, H. Zhang, and T.-S. Chua, "Attributed social network embedding," *IEEE Transactions on Knowledge and Data Engineering*, 2018.

PLACE PHOTO HERE

Chengbin Hou received the B.Eng. first class degree and the M.Sc. distinction degree from University of Liverpool (July 2014) and Imperial College London (November 2015) respectively. He started his Ph.D. in September 2017. Before his PhD study, he worked as an engineer at Huawei. Currently, he is pursuing his Ph.D. degree at Southern University of Science and Technology (SUSTech) and University of Birmingham. He is interested in machine learning and data mining on networked data.

PLACE PHOTO HERE

Han Zhang received his B.Eng. degree in Software Engineering from Xidian University in 2012, then he received MSc degree in Financial Engineering from University of Birmingham in 2014. He is currently pursuing his PhD degree in Computer Science at the University of Birmingham. His current research interests include machine learning and bioinformatics.

PLACE PHOTO HERE

Shan He received the Ph.D. degree in electrical engineering and electronics from University of Liverpool, Liverpool, U.K., in 2007. He is a Senior Lecturer in Computational Biology in School of Computer Science, the University of Birmingham. He is also an affiliate of the Centre for Computational Biology. His research interests include complex network analysis, machine learning, optimization, agent-based modelling and their applications to biological and clinical problems.

Dr. He is an Associate Editor of IEEE Transactions on Nanobioscience. He also serves as the Editorial Board Member of Complex & Intelligent Systems (Springer).

PLACE PHOTO HERE

Ke Tang received the B.Eng. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2002 and the Ph.D. degree from Nanyang Technological University, Singapore, in 2007. From 2007 to 2017, he was with the School of Computer Science and Technology, University of Science and Technology of China, Hefei, China, first as an Associate Professor from 2007 to 2011 and later as a Professor from 2011 to 2017. He is currently a Professor with the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China. He has over 8000 Google Scholar citations with an H-index of 42. His current research interests include evolutionary computation, machine learning, and their applications.

Dr. Tang was a recipient of the Royal Society Newton Advanced Fellowship in 2015 and the 2018 IEEE Computational Intelligence Society Outstanding Early Career Award. He is an Associate Editor of the IEEE Trans. on Evolutionary Computation and served as a member of Editorial Boards for a few other journals.