Learning Topological Representation for Networks via Hierarchical Sampling

Guoji Fu, Chengbin Hou, and Xin Yao

Shenzhen Key Lab of Computational Intelligence
University Key Laboratory of Evolving Intelligent Systems of Guangdong Province
Department of Computer Science and Engineering
Southern University of Science and Technology, Shenzhen, 518055, P. R. China
Email: fugj2017@student.sustc.edu.cn, chengbin.hou10@foxmail.com, xiny@sustc.edu.cn

***Abstract*—The topological information is essential for studying the relationship between nodes in a network. Recently, Network Representation Learning (NRL), which projects a network into a low-dimensional space, has been shown their advantages in analyzing large-scale networks. However, most existing NRL methods are designed to preserve the local topology of a network, but they fail to capture the global topology. To tackle this issue, we propose a new NRL famework, named HSRL, to help existing NRL methods capture both the local and global topological information of networks. Specifically, HSRL recursively compresses an input network into a series of smaller networks using a community-awareness compressing strategy. Then, an existing NRL method is used to learn node embeddings for each compressed network. Finally, the node embeddings of the input network are obtained by concatenating the node embeddings from all compressed networks. Empirical studies for link prediction on five real-world datasets demonstrate the advantages of HSRL over state-of-the art methods.**

***Index Terms*—Networks analysis, network topology, representation learning**

I. Introduction

The science of networks has been widely used to understand the behaviour of complex systems. Those systems are typically described as networks, such as social networks in social media [1], bibliographic networks in academic field [2], protein-protein interaction networks in biology [3]. Studying the relationship between entities in a complex system is an essential topic, which benefits a various of applications [4]. Just take a few example, predicting potential new friendship between users in social networks [5], searching similar authors in bibliographic networks [2], recommending new movies to users in movie user-movie interest networks [6]. The topologies of these networks provides insight information of the behavior of nodes. We can find out strongly connected neighborhoods of a node by exploring the local topology in a network. Meanwhile, the global topology is another significant aspect for studying the relationship between communities. Such hierarchical topological information is helpful to learn the relationship between nodes in a networks.

Many networks are large-scale in real-world scenarios, such as a Facebook social network contains billion of users [7]. As a result, most traditional network analytic methods suffer from high computation and space cost [4]. To tackle this issue, **N**etwork **R**epresentation **L**earning (NRL) has been a popular technique to analyze large-scale networks recently. In particular, NRL aims to map a network into a low-dimensional vector space, while preserving as much of original network topological information as possible. Nodes in a network are represented as low-dimensional vectors which are used as input features for downstream network analysis algorithms.

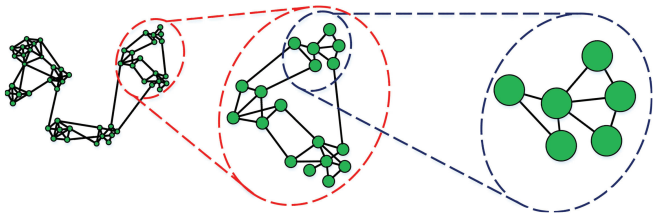


Fig. 1. An example of hierarchical view of network topology.

Traditional NRL methods such as LLE [8] and ISOMap [9] work well on small networks, while they are infeasible to large-scale networks ref to the high computational cost. Recently, some online learning methods, e.g., DeepWalk [10], node2vec [11], and LINE [12], have been proposed to learn large-scale network representation, which have been demonstrated their efficiency and effectiveness for the large-scale network analysis.

However, the above NRL methods only consider the local topology of networks and fail to capture the global topological information. DeepWalk and node2vec firstly employ short random walks to explore the local neighborhoods of nodes, and obtain node embeddings by the Skip-Gram model [13]. LINE preserves the first-order and second-order proximities, so that it can only measure the relationship between nodes at most two-hop away. These methods are efficient to capture the relationship between close nodes, however, fail to consider the case for nodes which are far away from each other. Recently, HARP [14] has been proposed to overcome this issue. It recursively compresses a network into a series of small networks using two node collapsing schemes and learns node embeddings for each compressed network based on an existing NRL method. Unfortunately, the compressed networks

may not reveal the global topology of a input network since HARP heuristically merges two closed nodes into a new node. Furthermore, when learning node embeddings on the original network, using node embeddings obtained on compressed networks as initialization solution may mislead the optimization process to local minimum.

This paper presents a new NRL framework, called **H**ierarchical **S**ampling **R**epresentation **L**earning (HSRL), to learn node embeddings for a network with preserving both their local and global topological information. Specifically, HSRL uses a community-awareness network compressing strategy, called hierarchical sampling, to recursively compress a input network into a series of smaller networks, and then engage an existing NRL method to learn node embeddings for each compressed network. Finally, the node embeddings of the original network can be abtained by concatenating all node embeddings learned on compressed networks. Besides, we mathmetically show that HSRL is able to capture the local and global topological relationship between nodes. Novel contributions of this paper include the following:

- We propose a new NRL framework, HSRL, to learn node embeddings for a network, which is able to learn both local and global topological information of a network via a community-awareness network compressing strategy.
- We mathematically show that the node embeddings obtained by HSRL explicitly embed the local and global topological information of the input network.
- We demonstrate that HSRL statistically significantly outperforms DeepWalk, node2vec, LINE, and HARP on link prediction tasks on five real-world datasets.

II. Related Work

Most early methods in NRL field represent an input network in the form of a matrix, e.g., adjacency matrices [8], [15], Laplacian matrices [16], node transition probability matrices [17], and then factorize that matrix to obtain node embeddings. They are effective for small network, but can not scale to large-scale networks due to high computation cost.

To analyze large-scale networks, DeepWalk [10] employs truncated random walks to obtain node sequences, and then learn node embeddings by feeding node sequences into Skip-Gram model [15]. To generalize DeepWalk, node2vec [11] provides a trade-off between breadth-first search (BFS) and deepth-first search (DFS) when generating truncated random walks for a network. LINE [12] intends to preserve first-order and second-order proximities of nodes by minimizing the Kullback-Leibler divergence of two joint probability distributions for each pair nodes. These methods are scalable to large-scale networks, but fail to capture the global topological information of networks. Because random walks are only effective to explore local neighborhoods for a node, and both first-order and second-order proximities defined by LINE just measure the relationship between between nodes at most two-hop away.

To investigate global topologies of a network, HARP [14] recursively uses two collapsing schemes, edge collapsing and star collapsing, to compress a input network into a series of small networks. Starting from the smallest compressed network, it then recursively conducts a NRL method to learn node embeddings based on the node embeddings obtained from its previous level (if any) as the initlization. However, HARP has two weaknesses: 1) nodes that are connected but belongs to different communities may be merged, which leads to that the compressed networks cannot well reveal the global topology of an input network. Furthermore, taking the node embeddings learned on such compressed networks as initialization would mislead NRL methods to a bad local minima. HARP may work well on node classification tasks since close nodes tend to have the same labels, but may ineffective for the link prediction tasks. Because predicting link between two nodes needs to consider both the local and global topological information of a network, such as neighborhoods they are sharing with and communities they are both involved in. This paper proposes HSRL to tackle the above issues of existing NRL methods.

III. Preliminary and Problem definition

This section gives the notations and definitions throughout this paper.

We firstly introduce the definition of a network and related notations.

Definition 1. ***(Network)*** *[4] A network (a.k.a. graph) is defined as $G = (V, E)$, where V is a set of nodes and E is a set of edges between nodes. The edge $e \in E$ between nodes u and v is represented as $e = (u, v)$ with a weight $w_{u,v} \geq 0$. Particularly, we have $(u, v) = (v, u)$ and $w_{u,v} = w_{v,u}$ if G is undirected; $(u, v) \neq (v, u)$ and $w_{u,v} \neq w_{v,u}$, otherwise.*

In most networks, some nodes are densely connected to form a community/cluster, while nodes in different communities are sparsely connected. Detecting communities in a network is beneficial to analyze the relationship between nodes. We employ modularity as defined below to evaluate the quality of community detection.

Definition 2. ***(Modularity)*** *[18], [19] Modularity is a measure of structure of networks, which measures the density of edges between the nodes within communities as compared to the edges between nodes in different communities. It is defined as below.*

$$Q = \frac{1}{2m} \sum_{i,j} \left(w_{i,j} - \frac{k_i k_j}{2m}\right) \delta(c_i, c_j) \tag{1}$$

where $w_{i,j}$ is the weight of edge $e_{i,j}$ between nodes v_i and v_j, $k_i = \sum_i w_{i,j}$, $m = \frac{1}{2} \sum_{i,j} w_{i,j}$, c_i is the community which node v_i belongs to, and

$$\delta(u, v) = \begin{cases} 1, & \textit{if } u = v, \\ 0, & \textit{otherwise}. \end{cases}$$

Networks with high modularity have dense connections between nodes within communities but sparse connections between nodes in different communities.

We give the definition of hierarchical sampling which is used to recursively compress a network into a series of smaller networks as follows.

Definition 3. *(**Hierarchical Sampling**) Given a network $G = (V, E)$, hierarchical sampling compresses the original network level by level and obtains a series of compressed networks $G^0, G^1, ..., G^K$, which reveals the global topological information of original network at different levels, respectively.*

These compressed networks reveal the hierarchical topologies of the input network. Therefore, the node embeddings obtained in compressed networks embed the hierarchical topological information of the original network.

To learn node embeddings of a network, NRL maps the original network into a low-dimensional space and represents each node as a low-dimensional vector as formulated below.

Definition 4. *(**Network Representation Learning**) [4] Given a network $G = (V, E)$, network representation learning aims to learn a mapping function $f : v \rightarrow z \in \mathbb{R}^d$ where $d \ll |V|$, and preserving as much of the original topological information in the embedding space $\mathbb{R}^d$.*

Finally, we present the formulation of the hierarchical network representation learning problem as following:

Definition 5. *(**Hierarchical Network Representation Learning**) Given a series of compressed networks $G^0, G^1, ..., G^K$ of original network $G = (V, E)$ and a network representation learning mapping function f, hierarchical network representation learning learns the node embeddings for each compressed network by $Z^k \leftarrow f(G^k), 0 \leq k \leq K$, and finally obtains the node embeddings Z of original network G by concatenating $Z^0, Z^1, ..., Z^K$.*

IV. HSRL

In this section, we present Hierarchical Sampling Representation Learning framework which consists of two parts: 1) Hierarchical Sampling that aims to discover the hierarchical topological information of a network via a community-awareness compressing strategy; and 2) Representation Learning that aims to learn low-dimensional node embeddings while preserving the hierarchical topological information.

A. Hierarchical Sampling

Here we present the hierarchical sampling which is intended to compress a network into a series of compressed networks according to different compressing levels. Each compressed network reveals one of hierarchical levels of global topology of the original network.

A community is one of the significant patterns of networks. Nodes in the same community are densely connected and nodes in different communities are sparsely connected. The relationship between nodes inside a community presents the local topological information of a network, while the relationship between communities reveals its global topology. It is worthy to notice that in most large-scale networks, there are several natural organization levels - communities divide themselves into sub-communities - and thus communities with different hierarchical levels reveal the hierarchical topological information of original networks [20], [21]. Consequently, we compress a network into a new network based on communities by taking each community as a new node in the compressed network. Based on different hierarchical levels of communities, we can obtain a series of compressed networks which reveal the hierarchical global topological information of the input network.

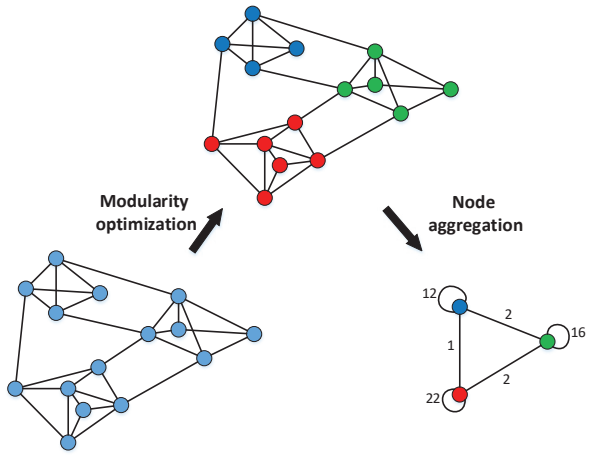


Fig. 2. An exampling of compressing a network

The quality of the partitions obtained by community dectection algorithms can be measured by the modularity of the partition [21], [22]. As a result, we can detect communities through optimizing the modularity of a network. As shown in Fig.2, inspired by the Louvain method [21], hierarchical sampling compresses a network into a new network by implementing two phases, modularity optimization and node aggregation.

Modularity optimization. The first phase initializes each node in a network as a community, and merges two connected nodes into one community if it can improve the modularity of the network. The implementation of community amalgamation will be repeated until a local maxima of the modularity is attained.

Node aggregation. The second phase builds a new network whose nodes are the communities found in the previous phase. The weights of edges between new nodes are the sum of the weights of edges between nodes in the corresponding two communities.

As shown in Algorithm 1, by iteratively repeating the above two phases, hierarchical sampling obtains a series of compressed networks which reveal hierarchical global topology of the original network.

B. Representation Learning

This section introduces representation learning on compressed networks obtained by previous section, and concatenating the learned embeddings into node embeddings of the original network. We further provide a mathematical proof to demonstrate that HSRL embeds both local and global

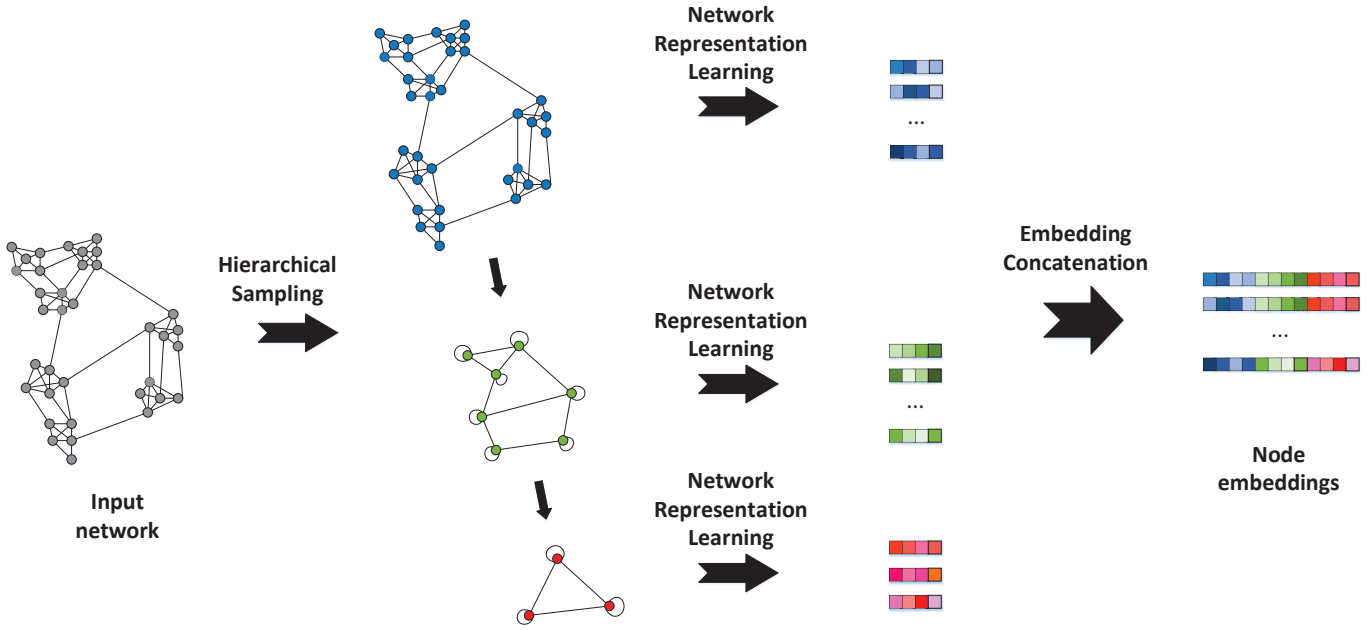


Fig. 3. The framework of HSRL.

Algorithm 1 Hierarchical Sampling

Input: network $G = (V, E)$, the largest # compressed levels K

Output: a series of compressed networks G^0, G^1, ..., G^K

$G^0 \leftarrow G$
for $k \leq K$ **do**
  $C^k \leftarrow ModularityOptimization(G^k)$
  $G^{k+1} \leftarrow NodeAggregation(C^k)$
  $k \leftarrow k + 1$
end for
return G^0, G^1, ..., G^K

topological relationship of nodes in the original network into the learned embeddings.

As shown in Fig.3, we conduct representation learning on each compressed network. It is worth noticing that any NRL method can be used for this purpose. The embeddings of nodes in each compressed network are used to generate the final node embeddings of the original network. Particularly, the embedding Z_i of node v_i in the original network G is the concatenation of the embeddings of hierarchical communities it involved in, as shown below.

$$Z_i = [Z^0_{c^0_i}, Z^1_{c^1_i}, ..., Z^K_{c^K_i}], \tag{2}$$

where c^k_i is the k-th hierarchical community v_i belongs to.

The node embeddings learned by the above representation learning process hold the following two Lemmas.

Lemma 1. *Nodes within the same hierarchical communities will get similar embeddings. The more the same hierarchical communities in which nodes involved, more similar embeddings they have.*

From Eq.2, it is easy to find that the above lemma holds. Lemma 1 shows that HSRL preserves the relationship between densely connected nodes in the original network. Therefore, HSRL is capable to preserve the local topological information of a network.

Lemma 2. *The cosine similarity between embedding Z_i and embedding Z_j is proportional to the sum of similarities of the embeddings between their hierarchical communities.*

$$sim(Z_i, Z_j) \propto \sum_{k=0}^{K} sim(Z^k_{C^k_i}, Z^k_{C^k_j}). \tag{3}$$

Proof.

$$\begin{aligned} sim(Z_i, Z_j) &= \frac{Z_i \cdot Z_j}{|Z_i||Z_j|} \\ &\propto [Z^0_{C^0_i}, Z^1_{C^1_i}, ..., Z^K_{C^K_i}] \cdot [Z^0_{C^0_j}, Z^1_{C^1_j}, ..., Z^K_{C^K_j}] \\ &= \sum_{k=0}^{K} Z^k_{C^k_i} \cdot Z^k_{C^k_j} \\ &\propto \sum_{k=0}^{K} sim(Z^k_{C^k_i}, Z^k_{C^k_j}). \end{aligned}$$

□

From Lemma 2, we know that two nodes will obtain similar embeddings if they are involved in similar hierarchical communities no matter the distance between them in the original network. The relationship between communities in different hierarchies is embedded in the embeddings of their

involved nodes. Hence, HSRL can preserve the hierarchical global topological information of a network.

Finally, HSRL is presented in Algorithm 2.

Algorithm 2 HSRL

Input: network $G = (V, E)$, compressing levels K, NRL mapping function f
Output: node embeddings Z

$G^0, G^1, ..., G^K \leftarrow HierarchicalSampling(G, K)$
for $k \leq K$ **do**
  $Z^k \leftarrow f(G^k)$
  $k \leftarrow k + 1$
end for
$Z \leftarrow Concatenating(Z^0, Z^1, ..., Z^K)$
return Z

V. Experiments

In this section, five real-world datasets are used to evaluate the representation learning performance of HSRL on link prediction task. The source code is available at *https://github.com/fuguoji/HSRL.*

A. Datasets

We evaluate our method on various real-world datasets, including Movielens[1], MIT [23], DBLP [2], Douban [24], and Yelp[2]. These datasets are commonly used in NRL field. The detailed statistics of datasets are shown in Table I and the brief descriptions of each dataset are presented as below.

- **Movielens**: Movielens is a user-movie interest network which contains three types of nodes: *users*, *movies*, and *terms*.
- **DBLP**: DBLP is a bibliographic network in computer science collected from four research areas: database, data mining, machine learning, and information retrieval. Nodes in the network including *authors*, *papers*, *venues*, and *terms*.
- **MIT**: MIT is a Facebook friendship network at one hundred American colleges and universities at a single point in time. It contains a single type of nodes, *users*.
- **Douban**: Douban is a user-movie interest network collected from a user review website Douban in China. The network contains four types of nodes including *users*, *movies*, *actors*, and *directors*.
- **Yelp**: Yelp is a user-business network collected from a website Yelp in America. It contains four types of nodes including *users*, *businesses*, *locations*, and *business categories*.

B. Baselines

We compare our method with four start-of-the-art algorithms, which are introduced as below.

- **DeepWalk**: DeepWalk is a random walk based NRL method. It conducts random walks on each node to sample node sequences from a network, and uses the Skip-Gram model to learn node embeddings by treating node sequences as sentences and nodes as words.
- **node2vec**: node2vec is a biased random walk based method that provides a trade-off between DFS and BFS when employing random walks on nodes. Then the Skip-Gram model is used to learn node embeddings based on the sampling node sequences.
- **LINE**: LINE defines the first-order and second-order proximities to measure the similarity between nodes, and learns node embeddings by preserving the aforementioned proximities of nodes in the embedding space.
- **HARP**: HARP recursively uses two collapsing schemes, edge collapsing and star collapsing, to compress a input network into a series of small networks. Starting from the smallest compressed network, it recursively conducts a NRL method to learns node embeddings in each network using node embeddings of the previous level network as initilization.

[1]https://movielens.org/
[2]https://www.yelp.com

TABLE I
Statistics of five datasets.

Datasets	# Nodes	# Edges	Network Types
Movielens	1332	2592	User-movie
DBLP	37791	170794	Bibliography
MIT	6402	251230	Friendship
Douban	13786	214392	User-movie
Yelp	28759	247698	User-business

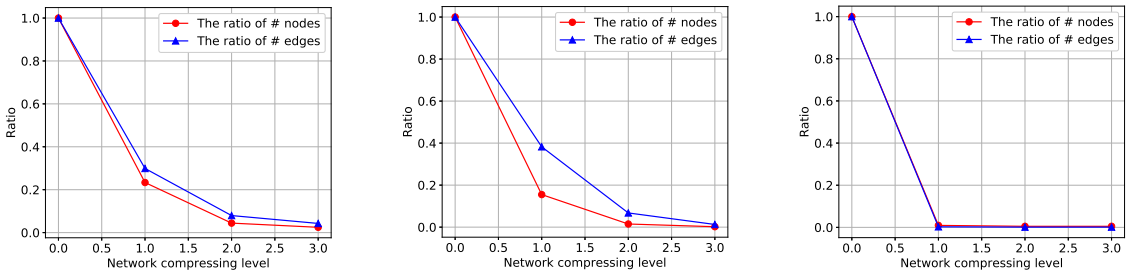
C. Parameter Settings

Here we discuss the parameter setting for our method and baselines:

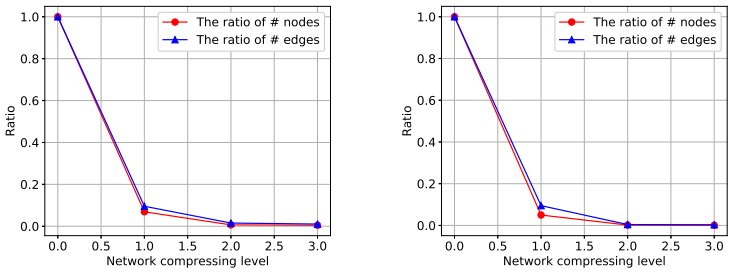
- **DeepWalk**, **HARP(DW)**, and **HSRL(DW)**. The number of random walks t, the length of each walk l, window size of Skip-Gram model w, representation size d, and learning rate η for DeepWalk, HARP(DW), and HSRL(DW) are set as $t = 10$, $l = 40$, $w = 5$, $d = 64$, $\eta = 0.025$. The largest number of compressing levels for HSRL(DW) K is set as 3.
- **node2vec**, **HARP(N2V)**, and **HSRL(N2V)**. The parameter setting for node2vec, HARP(N2V), and HSRL(N2V) is $t = 10$, $l = 40$, $w = 5$, $d = 64$, $\eta = 0.025$. The largest number of compressing levels for HSRL(N2V) K is set as 3.
- **LINE**, **HARP(LINE)**, and **HSRL(LINE)**. The number of negative sampling λ, the learning rate η, and representation size d for LINE, HARP(LINE), and HSRL(LINE) are set as $\lambda = 5$, $\eta = 0.025$, $d = 64$. The largest number of compressing levels for HSRL(LINE) K is set as 3.

D. Hierarchical Sampling for Networks

We firstly discuss the results of hierachical sampling on testing networks. Fig.4 presents the network compressing results by the hierarchical sampling on five datasets. As shown in Fig.4, the number of nodes and edges of compressed networks

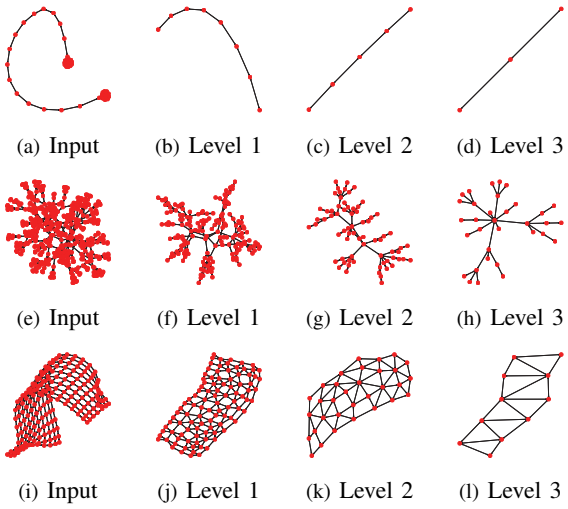


(a) Movielens (b) DBLP (c) MIT



(d) Douban (e) Yelp

Fig. 4. The compressing ratio of nodes/edges of compressed networks to the original network.



(a) Input (b) Level 1 (c) Level 2 (d) Level 3

(e) Input (f) Level 1 (g) Level 2 (h) Level 3

(i) Input (j) Level 1 (k) Level 2 (l) Level 3

Fig. 5. Examples of hierarchical sampling.

drastically decrease as the compressing process continues and finally becomes stable when the compressing level is large than 3. Therefore, in the following link prediction tasks, we set the largest number of compressing level as 3.

As shown in Fig.5, we present three networks, Fig.5(a), Fig.5(e), and Fig.5(i), as the intuitive examples to illustrate how hierarchical sampling works.

The network in Fig.5(a) contains two dense communities which are merged into a new node in the following compressed networks respectively. Therefore in the compressed networks, the local topological information of original networks are preserved by considering densely connected nodes in the same community as a whole. Meanwhile, the compressed networks, Fig.5(b), Fig.5(c), and Fig.5(d), reveal the hierarchical topological information of the input network. For a balanced tree network and a grid network as shown in Fig.5(e) and Fig.5(i), their hierarchical topologies can be revealed by their compressed networks as well. The results of Fig.5 show that the network compressing strategy of HSRL works well on different types of networks.

E. Link Prediction

We conduct link prediction tasks to evaluate the performance of our method on five real-world datasets. Specifically, we prediction the link between nodes based on the cosine similarity of their embeddings. The evaluation metric used in this task is AUC. Higher AUC indicates the better performance of NRL methods.

We randomly split the edges of a network into 80% edges as training set and the left 20% as testing set. Each experiment is independently implemented for 20 times and the average performances on testing set are reported in Table II.

We summarize the observations from Table.II as following:

- HSRL significantly outperforms all baselines on all datasets. For the small and sparse network movielens, the imporvements of HSRL(DW), HSRL(N2V), and HSRL(LINE) are 3.6%, 2.5%, and 16.6% respectively. For the dense network MIT, HSRL(DW), HSRL(N2V), and HSRL(LINE) outperform the baselines by 2.9%, 8.5%, and 0.5%. For three large networks, DBLP, Yelp, and Douban, the improvement of HSRL is striking: the improvements of HSRL(DW), HSRL(N2V), and

TABLE II
AUC OF LINK PREDICTION.

Algorithm	Dataset				
	Movielens	DBLP	MIT	Yelp	Douban
DeepWalk	0.847	0.794	0.899	0.842	0.687
HARP(DW)	0.817	0.659	0.902	0.743	0.559
HSRL(DW)	**0.879**†	**0.847**†	**0.926**†	**0.901**†	**0.842**†
Gain of HSRL(%)	**3.6**	**6.3**	**2.9**	**6.5**	**18.4**
node2vec	0.843	0.673	0.843	0.742	0.569
HARP(N2V)	0.828	0.647	0.879	0.708	0.552
HSRL(N2V)	**0.865**†	**0.840**†	**0.921**†	**0.892**†	**0.819**†
Gain of HSRL(%)	**2.5**	**19.9**	**8.5**	**16.8**	**30.5**
LINE	0.613	0.641	0.814	0.752	0.624
HARP(LINE)	0.220	0.387	0.702	0.306	0.399
HSRL(LINE)	**0.735**†	**0.664**†	**0.819**	**0.799**†	**0.756**†
Gain of HSRL(%)	**16.6**	**3.5**	**0.5**	**5.9**	**17.5**

† denotes the performance of HSRL is significantly better than the other peers according to the Wilcoxon's rank sum test at a 0.05 significance level.

HSRL(LINE) are 6.3%, 19.9%, 3.5% for DBLP, 6.5%, 16.8%, 5.9% for Yelp, and 18.4%, 30.5%, 17.5% for Douban.

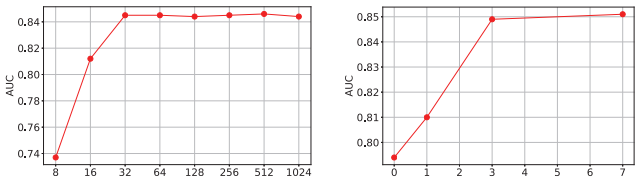
- The results of HARP on movielens, DBLP, Yelp, and Douban are worse than the original NRL methods. Moreover, the performance of HARP(LINE) is drastically worse than LINE. It only works better than DeepWalk, node2vec on MIT which is a small and dense network. The compressed networks generated by HARP on a network could not reveal its global topologies. Hence, using node embeddings of compressed networks as initialization could mislead the NRL methods to a bad local minima. Such a issue could occur especially when the input network is large-scale and the objective function of the NRL method is highly non-convex, e.g., LINE.
- The improvements of HSRL on DBLP, Yelp, and Duban are larger than that on Movielens and MIT. It demonstrates that HSRL works much better than baselines on large-scale networks.

F. Parameter Sensitivity Analysis

We conduct link prediction task on DBLP to study the paramenter sensitivity of HSRL. Without loss of generality, we used DeepWalk to learn node embeddings for each compressed network. Fig.6 shows using 80% edges as training set and the left as testing set, the link prediction performance (AUC) as a function of one of chosen parameters when fixing the others.

When fixing the largest number of compressed level to 3, Fig.6(a) shows the AUC of link prediction drastically improves as the number of embedding dimension d increases, and finally becomes stable when d is larger than 32. When d is small, it is inadequate to embody rich information of networks. However, when d is large enough to embody all original network information, increasing d will not improve the performance of link prediction.

Fig.6(b) shows that the impact of the largest number of network compressing level K on the performance of link prediction by fixing the representation size d to 64. As the we increase the number of network compressing level K at the beginning, the AUC of link prediction drastically improves. It demonstrates that the hierarchical topologies help to capture the potential relationship between nodes in a network. When K is larger than 3, the performance of link prediction becomes stable. It is reasonable since the DBLP network could not be compressed further after level 3 as shown in Fig.4(b).



(a) representation size d (b) network compressing level K

Fig. 6. Parameter sensitivity on link prediction.

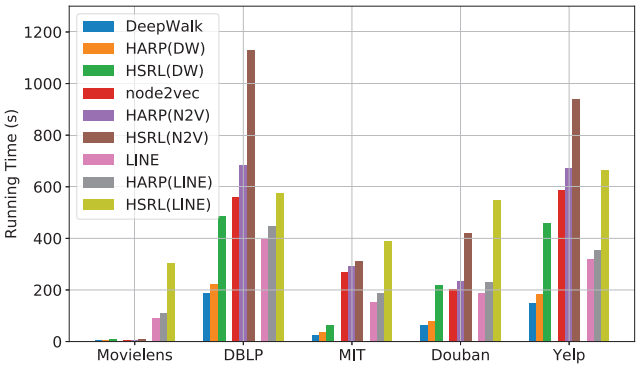


Fig. 7. Running time

G. Running Time

Fig.6 shows the actual running time of all NRL methods on five testing networks. All experiments are conducted on a single machine with 32GB memory, 16 CPU cores at 3.2 GHZ. The results show that the actual running time of HSRL is at most three times higher than others. The running time of HSRL is linear to the corresponding baselines as the scale of input networks growing. Moreover, the running time of HSRL can be reduced by parallelizing the training processes on all compressed networks.

VI. CONCLUSIONS

Most conventional NRL methods aim to preserve the local topological information of a network but overlook their global topology. Recently, HARP was proposed to preserve both local and global topological information. However, it could easily get stuck at a bad local minimum due to poor network compressing schemes. In this paper, we propose a new NRL framework, HSRL, to tackle these issues. Specifically, HSRL employs a community-awareness network compressing scheme to obtain a series of smaller networks from an input network, and conducts a NRL method to learn node embeddings for each compressed network. Finally, the node embeddings of

the original network can be obtained by concatenate all node embeddings of compressed networks. Empirical studies on link prediction on various real-world networks demonstrate HSRL significantly outperforms the state-of-the-art algorithms.

Our future work includes combining HSRL with deep learning-based methods, such as DNGR [25], SDNE [26], and GCN [27]. It is also very interesting to extend HSRL to learn node embeddings of more complex networks which may be more common in real-world applications, e.g., heterogeneous networks, attributed networks, and dynamic networks.

ACKNOWLEDGMENT

This work was supported by National Key R&D Program of China (Grant No. 2017YFC0804003), a 2018 IEEE Computational Intelligence Society Graduate Student Research Grant, the Program for Guangdong Introducing Innovative and Enterpreneurial Teams (Grant No. 2017ZT07X386), Shenzhen Peacock Plan (Grant No. KQTD2016112514355531), the Science and Technology Innovation Committee Foundation of Shenzhen (Grant No. ZDSYS201703031748284) and the Program for University Key Laboratory of Guangdong Province (Grant No. 2017KSYS008).

REFERENCES

[1] J. Scott, *Social network analysis*. Sage, 2017.

[2] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, "Pathsim: Meta path-based top-k similarity search in heterogeneous information networks," *Proceedings of the VLDB Endowment*, vol. 4, no. 11, pp. 992–1003, 2011.

[3] A. Theocharidis, S. Van Dongen, A. J. Enright, and T. C. Freeman, "Network visualization and analysis of gene expression data using biolayout express 3d," *Nature protocols*, vol. 4, no. 10, p. 1535, 2009.

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

[5] S.-H. Yang, B. Long, A. Smola, N. Sadagopan, Z. Zheng, and H. Zha, "Like like alike: joint friendship and interest propagation in social networks," in *Proceedings of the 20th international conference on World wide web*. ACM, 2011, pp. 537–546.

[6] Q. Diao, M. Qiu, C.-Y. Wu, A. J. Smola, J. Jiang, and C. Wang, "Jointly modeling aspects, ratings and sentiments for movie recommendation (jmars)," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014, pp. 193–202.

[7] N. B. Ellison, C. Steinfield, and C. Lampe, "The benefits of facebook "friends:" social capital and college students' use of online social network sites," *Journal of Computer-Mediated Communication*, vol. 12, no. 4, pp. 1143–1168, 2007.

[8] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[9] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[10] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014, pp. 701–710.

[11] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2016, pp. 855–864.

[12] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "Line: Large-scale information network embedding," in *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2015, pp. 1067–1077.

[13] 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*, 2013, pp. 3111–3119.

[14] H. Chen, B. Perozzi, Y. Hu, and S. Skiena, "Harp: hierarchical representation learning for networks," *arXiv preprint arXiv:1706.07845*, 2017.

[15] A. Ahmed, N. Shervashidze, S. Narayanamurthy, V. Josifovski, and A. J. Smola, "Distributed large-scale natural graph factorization," in *Proceedings of the 22nd international conference on World Wide Web*. ACM, 2013, pp. 37–48.

[16] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in neural information processing systems*, 2002, pp. 585–591.

[17] 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.

[18] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[19] ——, "Analysis of weighted networks," *Physical review E*, vol. 70, no. 5, p. 056131, 2004.

[20] M. Sales-Pardo, R. Guimera, A. A. Moreira, and L. A. N. Amaral, "Extracting the hierarchical organization of complex systems," *Proceedings of the National Academy of Sciences*, vol. 104, no. 39, pp. 15 224–15 229, 2007.

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

[22] A. Clauset, M. E. Newman, and C. Moore, "Finding community structure in very large networks," *Physical review E*, vol. 70, no. 6, p. 066111, 2004.

[23] A. L. Traud, P. J. Mucha, and M. A. Porter, "Social structure of facebook networks," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 16, pp. 4165–4180, 2012.

[24] J. Zheng, J. Liu, C. Shi, F. Zhuang, J. Li, and B. Wu, "Recommendation in heterogeneous information network via dual similarity regularization," *International Journal of Data Science and Analytics*, vol. 3, no. 1, pp. 35–48, 2017.

[25] S. Cao, W. Lu, and Q. Xu, "Deep neural networks for learning graph representations." in *AAAI*. ACM, 2016, pp. 1145–1152.

[26] D. Wang, P. Cui, and W. Zhu, "Structural deep network embedding," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2016, pp. 1225–1234.

[27] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

APPENDIX

In this appendix, we present a case study of searching relevant authors for "Qiang Yang" to illustrate the performance of HSRL.

A. A Case Study: Relevant Authors Search for "Qiang Yang"

"Qiang Yang" is an active and influential researcher on machine learning and data mining. We conduct a case study of finding relevant researchers to him on the DBLP network. Table.A shows the results of his top-25 relevant researchers given by HSRL(DW), HARP(DW), and DeepWalk, respectively.

TABLE A1
TOP-25 RELEVANT AUTHORS TO "QIANG YANG"

Rank	HSRL(DW)	HARP(DW)	DeepWalk
1	Dou Shen	Dou Shen	Dou Shen
2	Gui-Rong Xue	Jian-Tao Sun	Zheng Chen
3	Wenyuan Dai	Zheng Chen	Jian-Tao Sun
4	Zheng Chen	Pu Wang	Hua-Jun Zeng
5	Weizhu Chen	Min Qin	Gui-Rong Xue
6	Jian-Tao Sun	Bin Cao	Weizhu Chen
7	Hua-Jun Zeng	Lin Deng	Wenyuan Dai
8	Benyu Zhang	Benyu Zhang	Benyu Zhang
9	Hua Li	Weizhu Chen	Jeffrey Junfeng Pan
10	Jun Yan	Lujun Fang	Sinno Jialin Pan
11	Bin Cao	Jeffrey Junfeng Pan	Bin Cao
12	Xiao Ling	Jun Yan	Yong Yu
13	Dikan Xing	Jilin Chen	Chun-Yi Shi
14	Wei-Ying Ma	Zhong Zhang	Min Qin
15	**QianSheng Cheng**	Hua-Jun Zeng	Gongyi Wu
16	**Ning Liu**	Wenyuan Dai	Hua Li
17	Min Qin	Raymond Chan	Jie Yin
18	Yong Yu	Dikan Xing	Xiao Ling
19	**Jun Zhu**	Dong Zhuang	Lujun Fang
20	**Junhui Zhao**	Tielin Chen	Dikan Xing
21	**Peng Bai**	Jie Yin	Jun Yan
22	**Hang Li**	Chenyong Hu	Wei-Ying Ma
23	**Ji-Rong Wen**	Gui-Rong Xue	Dong Zhuang
24	**Jinwen Ma**	Xiao Ling	Wensi Xi
25	**HongJiang Zhang**	Yiqiang Chen	Evan Wei Xiang

From the results, we have the following observations:

- The top-14 relevant researchers of "Qiang Yang" obtained by HSRL(DW) are also the researchers in the top-25 lists of both HARP(DW) and DeepWalk with slightly different orders. We can easily find out that most of these relevant researchers are Ph.D students of "Qiang Yang". They have strong collaboration with "Qiang Yang" in many works and directly related with him in the DBLP network. HSRL(DW), HARP(DW), and DeepWalk all perform well on finding the directly relevant researchers of "Qiang Yang".
- The last 10 relevant researchers in the top-25 list of HSRL(DW) are very different with respect to the other two methods. The researchers in bold in the Table.A1 are not in the top-25 relevant researchers. However, researchers including "Hang Li", "Ji-Rong Wen", "HongJiang Zhang", are active and influential on machine learning or data mining. These researchers are similar with "Qiang Yang". Contrary to HARP(DW) and DeepWalk which only work well on finding directly related researchers of "Qiang Yang", HSRL(DW) is able to find his relevant researchers who perform similar contributions to the same research communities.