SIEG: Structural Information Enhanced Graph representation

Anonymous Author(s)

ABSTRACT

The expressive power of GNNs has been playing an important part in limiting the improvement of the link prediction task in the past few years. In order to make the GNNs more powerful than 1^{st} -WL algorithm, SEAL [22, 23] adds the structure information to GNN-based model with a labeling trick, which maintains two essential properties of target-nodes-distinguishing and permutation equivariance. However, since SEAL simply adds the structure features as initial node feature, the structure information will decay after several rounds of message passing. Graphormer [21] creatively applies the transformer architecture[16] and adds the graph structure information directly to its attention term. However, it is faced with a computationally expensive problem caused by the message passing between any two nodes. Therefore, we propose Structural Information Enhanced Graph representation (SIEG) in this paper to improve the expressive power of GNN by adding the structure information to GNN more properly. The global information of each node is captured by the aggregation process of a GNN model. We add the structure information to the transformer attention, but keep the transformer attention only for the two centered nodes to capture only the most important local structure information. The computationally expensive problem is avoided at the same time. Experiments shows that, SIEG outperforms the state-of-the-art work SEAL by 2.0% ROC-AUC relatively on ogbl-vessel dataset[12, 14].

KEYWORDS

graph representation learning; graph convolutional network; graph structure; graph transformer

ACM Reference Format:

Anonymous Author(s). 2022. SIEG: Structural Information Enhanced Graph representation. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Link prediction is one of the most important applications of knowledge graph, and is capable of predicting links between any two nodes in the graph. For example, it can predict users ' interest of certain products in recommendation system, and can link the accounts belonging to the same person in order to evaluate the financial credit of a natural person. There are three popular classes of link prediction methods, which are heuristic methods, shallow knowledge graph embedding methods (shallow-KGE) [18] and Graph neural networks (GNNs) based methods.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Heuristic methods, such as common neighbors, Adamic-Adar [1], Jaccard neighbors and Katz index [6] describe the graph structure features between the targeted pair of nodes. However, heuristic methods can only treat the pair of nodes with certain patterns. Shallow-KGEs embed the entities and relations directly into vectors without any aggregation operation. They include translational distance models such as TransE [2], TransH [19] etc. and semantic matching models such as RESCAL [11], ComplEx [15] etc. Shallow-KGEs are good at dealing with graphs with rich relation types. However there exist several disadvantages. First, Shallow-KGEs are lack in sharing parameters mechanism between nodes, which leads to inefficiency. Second, they do not leverage node features. Last but not least, they are inherently transductive, which means they can only infer the nodes presented in training set.

GNN-based methods are becoming increasingly popular in recent years. The basic GNN formular can be derived from non-Euclidean graph convolution [8], and it describes how the target node aggregates the node and edge features from neighbor nodes. In order to solve the link prediction problem, an encoder-decoder framework is usually applied, in which GNN models are chosen as encoder to extract the features of a subgraph and then aggregate the features to the two centered nodes, while the Shallow-KGEs are usually chosen as decoder to combine the features of two target nodes and the target edge to match the link prediction labels.

The expressive power of GNN-based methods has been continuously improved in recent years. The early GNN-based methods such as GCN [8][3][4] and GAT [17] are proved at most as powerful as 1^{st}-WL algorithm. GIN [20] has increased its expressive power to match the power of 1^{st}-WL algorithm through keeping the aggregate and update function injective.

Recently, making GNNs more powerful than 1^{st}-WL algorithm is becoming an active area of research. One direction is to create a GNN by following the k^{th}-WL algorithm, since it is known that the k^{th}-WL algorithm is more powerful than 1^{st}-WL. Morris et al. proposed the k-GNN [9], which is a differentiable and continuous analog of the k^{th}-WL algorithm, and they also proved that k-GNN is as expressive as the k^{th}-WL algorithm. In another related work, Murphy et al. developed the relational pooling (RP) method [10], which takes an average of the results from all possible permutations of a graph in order to improve the expressive power of GNN. In addition, it keeps the permutation invariant. However, these works are computationally intractable theoretically. Therefore, the subsequent works focus on various approximations in order to make the models tractable in practice. Another direction tries to improve the expressive power of GNN by bringing the additional structure information to the GNN-based models. Zhang et al. pointed out that GNN computes the embedding of each target node independently without being aware of another target node while treating link prediction tasks, so that GNN cannot recognize the common neighbor and the shortest distance type of structure. They further summarized the label techniques and proposed SEAL [22, 23] to add the structure information to GNN-based models with so called

labeling trick, which records the distance of each node to the two target nodes. The success of the labelling trick is due to it maintains two important properties of target-nodes-distinguishing and permutation equivariance. Although it is proved that SEAL indeed can recognize some structures that 1^{st}-WL algorithm cannot. However, since SEAL simply adds the structure features as initial node feature, the structure information will decay after several rounds of message passing between layers.

Graphormer [21] creatively adds the graph structure information to its attention term, although Graphormer is better known for the first work to apply the famous transformer framework to graph representation. The distance of shortest paths and the relation type of one of the shortest paths have been added to the attention term. However, since the attention of any two nodes needs to be computed, Graphormer is computationally expensive. Therefore, it has only been applied to small dataset. Moreover, the way of Graphormer to add structure information is so rough that it only assigns a trainable embedding to each structure value, and take a lookup operation while iterating.

We propose Structural Information Enhanced Graph representation (SIEG) in this paper to improve the expressive power of GNN by adding the structure information to GNN more properly. The global information of each node is captured by the aggregation process of a GNN model. We add the structure information to the transformer attention term, but keep the transformer attention only for the two centered nodes to capture the most important local structure information. A reasonable assumption is implied that the structure patterns between two centered nodes are more important than the patterns between other nodes in the link prediction task. We also get an additional benefit that the computationally expensive problem is avoided. Eventually, more structure information is added to our model than Graphormer such as the number of path for each hop and the edge type information of each path. Finally, the labeling trick is also applied to initialize the node embeddings.

2 METHODOLOGY

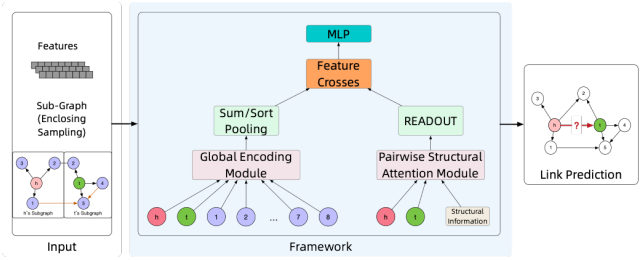

Figure 1: The model architecture of SIEG. The left tower is a GCN model to obtain the global information. The right tower is a transformer only calculating the attention between the two centered nodes.

As shown in figure1, our proposed SIEG model has a double tower structure. GCN is used as the first tower to capture the global information, and Graphormer[21] is used as the second tower to capture the important local information.

As the first tower, the classic GCN is applied to aggregate the information of neighbor nodes to the centered nodes.

$$H^{(t+1)} = \sigma(\tilde{\mathbf{A}}H^{(t)}W^{(t)}) \tag{1}$$

$\tilde{\mathbf{A}} = \hat{D}^{-1/2}\hat{\mathbf{A}}\hat{D}^{-1/2}$ is the normalized adjacency matrix, and $\hat{\mathbf{A}} = \mathbf{A} + I$ where $\mathbf{A}$ is adjacency matrix. $H^{(t)}$ is the node embedding of layer t. W is a full connected layer.

As the second tower, we apply the Transformer attention from Graphormer to illustrate the similarity of two centered nodes.

$$Q = HW_Q, K = HW_K, V = HW_V, \tag{2}$$

$$A = \frac{QK^\top}{\sqrt{d_K}}, Attn(H) = softmax(A)V \tag{3}$$

We propose an unified form S_{ij} to introduce the structure information and add it to the attention, therefore to enhance the model.

$$A_{ij} = \frac{(h_iW_Q)(h_jW_K)^\top}{\sqrt{d}} + S_{ij}, \tag{4}$$

The single-head self-attention is considered in the first term, therefore $d_K = d_V = d$.

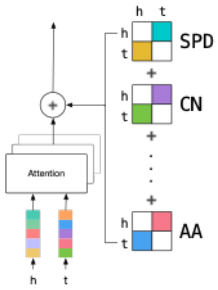

Figure 2: The structure attention between two centered nodes, which is extensible. SPD means shortest path distance, CN represents common neighbors, AA represents Adamic-Adar index.

$$S_{ij} = b_{ij}^{dis} + b_{ij}^{num} + f_{ij}^{AA} + f_{ij}^{Jac} \tag{5}$$

As shown in figure2, any structure information between two centered nodes could be added to S_{ij}. For example, b_{ij}^{dis} represents the distance of the shortest path which is inherited from Graphormer [21]. b_{ij}^{num} represents the number of the shortest path between two centered nodes, f_{ij}^{AA} describes the common neighbor status with the Adamic-Adar form. f_{ij}^{Jac} also describes the common neighbor status but with the Jaccard form.

SIEG's attentions have several advantages compared with Graphormer's, which also introduces the structure information with spatial encoding. First, only the attention between two centered nodes and self-attention are remained greatly reduces the computation calculation. second, more fine-grained structure information is added to the model.

Finally, the labeling trick of SEAL[23] is also applied to initialize the node embeddings.

3 EXPERIMENTS

3.1 Dateset and Task

In order to prove the effectiveness of the proposed method, we conduct link prediction tasks on the ogbl-vessel dataset. The dataset is an undirected, unweighted spatial graph of the mouse brain [12, 14], where edges and nodes represent vessels and the corresponding bifurcation points. The node features are 3-dimensional, representing the spatial (x, y, z) coordinates of the nodes in Allen Brain atlas reference space. So, our link prediction task is to predict if a vessel exists or not. The dataset is randomly splitted to training set, validation set and test set with a 80/10/10 split ratio, and can be easily get from Open Graph Benchmark (OGB) [5].

3.2 Experimental Setup

We use BCEWithLogitsLoss and Adam [7] as optimization objective and algorithm for model training, where the learning rate and batch size are 0.0002 and 256, respectively. We train model for 10 epochs, and chose the parameters perform best on validation set for test. The code development is based on PyTorch and PyTorch Geometric, and program runs on Tesla P100 (16GB).

3.3 Experimental Results

We repeat experiment ten times with the same settings, and report the average ROC-AUC and (sample) standard deviation in table 1. Meanwhile, we also list state of the art methods and results for comparison. It can be seen that the proposed method significantly outperforms all of them, specially with an approximate 2% relative improvement over SEAL (no-xfeat)[22]. We analyse the possible reasons for the improvement as follows.

1) Global information interaction. GNN-based methods, such as SEAL [22, 23], only pass messages in the local domain in single-layer aggregation, so the receptive field is small and restricted. The proposed method make it easier to learn global feature through Graphormer, where nodes receive all vessel information in a focused (attentional) way, facilitating the model to observe an overview of the mouse brain.

2) Structural Information Enhancement. In the vessel dataset, node features are the simple spatial coordinates of the nodes, which may not be very important for vessel connections, and the comparison of SEAL (no-xfeat) [23] and SEAL [22] proves this point. Instead, it may be the topology structure rather than the spatial coordinates that really matter. Therefore, we compute structural feature explicitly, especially pairwise feature such as Jaccard index and Adamic-Adar[1], then embed them into the bias of Self-attention. Generally, Transformer is better at attribute features extraction, so our operation make up for the lack of structure feature extraction ability of model, and help to learn vessel structure in the mouse brain.

Table 1: Validation and test results of SIEG on ogbl-vessel, ROC-AUC (%)

Method	Validation AUC (%)	Test AUC (%)
SEAL (no-xfeat)[23]	0.8073 ± 0.0001	0.8077 ± 0.0001
SEAL[22]	0.8053 ± 0.0022	0.8050 ± 0.0021
LRGA[13]	0.5418 ± 0.0439	0.5415 ± 0.0437
SIEG (Ours)	**0.8255 ± 0.0040**	**0.8249 ± 0.0041**

4 CONCLUSIONS

We propose a structural information enhanced graph representation model (SIEG) in this work, which is able to aggregate the global information with GCN, and capture the important local structure information using transformer attention only for the two centered nodes. In this way, not only the expressive power is improved but also the computationally expensive problem is avoided. The experiment on ogbl-vessel dataset shows that, SIEG achieves 2.0% ROC-AUC relative improvement than the state-of-the-art work SEAL.

REFERENCES

[1] Lada A. Adamic and Eytan Adar. 2003. Friends and neighbors on the Web. *Social Networks* 25 (2003), 211–230.

[2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In *Neural Information Processing Systems*.

[3] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. 2014. Spectral Networks and Locally Connected Networks on Graphs. In *International Conference on Learning Representations*.

[4] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In *Neural Information Processing Systems*.

[5] Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. 2020. Open graph benchmark: Datasets for machine learning on graphs. *Advances in neural information processing systems* 33 (2020), 22118–22133.

[6] Leo Katz. 1953. A new status index derived from sociometric analysis. *Psychometrika* 18 (1953), 39–43.

[7] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[8] Thomas Kipf and Max Welling. 2016. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv: Learning* (2016).

[9] Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. 2019. Weisfeiler and Leman Go Neural: Higher-Order Graph Neural Networks. In *National Conference on Artificial Intelligence*.

[10] Ryan L. Murphy, Balasubramaniam Srinivasan, Vinayak Rao, and Bruno Ribeiro. 2019. Relational Pooling for Graph Representations. In *International Conference on Machine Learning*.

[11] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A Three-Way Model for Collective Learning on Multi-Relational Data. In *International Conference on Machine Learning*.

[12] Johannes C Paetzold, Julian McGinnis, Suprosanna Shit, Ivan Ezhov, Paul Büschl, Chinmay Prabhakar, Mihail I Todorov, Anjany Sekuboyina, Georgios Kaissis, Ali Ertürk, et al. 2021. Whole brain vessel graphs: a dataset and benchmark for graph learning and neuroscience (vesselgraph). *arXiv preprint arXiv:2108.13233* (2021).

[13] Omri Puny, Heli Ben-Hamu, and Yaron Lipman. 2020. Global attention improves graph networks generalization. *arXiv preprint arXiv:2006.07846* (2020).

[14] Mihail Ivilinov Todorov, Johannes Christian Paetzold, Oliver Schoppe, Giles Tetteh, Suprosanna Shit, Velizar Efremov, Katalin Todorov-Völgyi, Marco Düring, Martin Dichgans, Marie Piraud, et al. 2020. Machine learning analysis of whole mouse brain vasculature. *Nature methods* 17, 4 (2020), 442–449.

[15] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International Conference on Machine Learning*.

[16] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[17] Petar Veli{c}ković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2017. Graph Attention Networks.

[18] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Transactions on Knowledge and Data Engineering* 29 (2017), 2724–2743.

[19] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *National Conference on Artificial Intelligence*.

[20] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2018. How Powerful are Graph Neural Networks. In *International Conference on Learning Representations*.

[21] Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. 2021. Do Transformers Really Perform Bad for Graph Representation. *arXiv: Learning* (2021).

[22] Muhan Zhang and Yixin Chen. 2018. Link prediction based on graph neural networks. *Advances in neural information processing systems* 31 (2018).

[23] Muhan Zhang, Pan Li, Yinglong Xia, Kai Wang, and Long Jin. 2021. Labeling Trick: A Theory of Using Graph Neural Networks for Multi-Node Representation Learning. In *Neural Information Processing Systems*.