Appendix of PathNet

1 Proof of Theorem 1.

Proof. A tag-limited transition matrix $\mathbf{A}^{(l)}$ is derived by setting the same elements as the original adjacency matrix $\mathbf{A}$ if the end node can be tagged as l, while other elements are tagged as zero:

$$\mathbf{A}_{i,j}^{(l)} = \begin{cases} \mathbf{A}_{i,j}, & \mathbf{Y}_j = l, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

During the path sampling process, we sample the neighbors with the probability matrix of maximal entropy calculated as the Eq. 2. Thus, the probability matrix $\mathbf{P}^{\{l_1,l_2,\ldots,l_k\}}$ of sampling a attributed path is as follows:

$$\mathbf{P}^{\{l_1,l_2,\ldots,l_k\}} = \prod_{i=1}^{k}(\mathbf{P}_{\mathbf{u}}^{(l)}) = \prod_{i=1}^{k}(\frac{\mathbf{D}_{\mathbf{u}}^{-1}\mathbf{A}^{(1)}\mathbf{D}_{\mathbf{u}}}{\lambda}). \tag{2}$$

The element in the i-th row and j-th column of $\mathbf{P}^{\{l_1,l_2,\ldots,l_k\}}$ is the probability of walking from i to j through an attributed path. Moreover, $P_{S,\{l_1,l_2,\ldots,l_k\}}$, the probability of sampling an attributed path starting from node S, is derived as follows:

$$P_{S,\{l_1,l_2,\ldots,l_k\}} = \sum_{i=1}^{|\mathcal{V}|}\mathbf{P}_{S,i}^{\{l_1,l_2,\ldots,l_k\}}. \tag{3}$$

Here, $N_{S,\{l_1,l_2,\ldots,l_k\}}$ denotes the number of attributed paths in m times of sampling from graph. Since we use the same sampling strategy for every epoch, the probability of attributed paths being sampled is $P_{S,\{l_1,l_2,\ldots,l_k\}}$, while the probability of not being sampled is $1 - P_{S,\{l_1,l_2,\ldots,l_k\}}$. Whether an attributed path would be sampled in two samples is independent and identically distributed. Therefore, $N_{S,\{l_1,l_2,\ldots,l_k\}}$ follows a binomial distribution:

$$N_{S,\{l_1,l_2,\ldots,l_k\}} \sim \mathrm{B}(m, P_{S,\{l_1,l_2,\ldots,l_k\}}). \tag{4}$$

According to Hoeffding's inequality[Hoeffding, 1963], we have

$$\Pr(|\frac{N_{S,\{l_1,l_2,\ldots,l_k\}}}{M} - P_{S,\{l_1,l_2,\ldots,l_k\}}| > \epsilon) \leq 2\exp(-2\epsilon^2 m). \tag{5}$$

Thus, as the number of sampled paths m increases, the probability on the left of inequality decreases exponentially. □

2 Computational Complexity

Here we discuss the computational complexity of PathNet. In general, our method is rather efficient. Concretely, the preprocessing includes the path sampling and shortest distance calculation, with the precise time complexity of $O(d^k n)$, where n denotes number of nodes, d is the average degree of the graph ($d = 3.89$ in Cora) and k is the length of sampled paths ($k = 4$ in our experiments). Moreover, the overall time complexity of PathNet is $O(d^k n + mtnhh')$, where m denotes the number of sampled path ($m = 40$ in our experiments), t is the iteration (epoch) value, h and $h^{'}$ represent the dimensions of input and output.

3 Experiment Settings and Hyperparameters

As important parts of preprocess, the path sampling and shortest distance between nodes are finished before model training with the time complexity shown in supplemental material. The final hyper-parameter settings for PathNet are as follows: 1000 epochs, learning rate as 0.005 and weight decay as 0.0005. For other baselines, we choose the hyper-parameter settings which can make them perform best. For simplicity, we set the number of paths as 40 for both heterophily and homophily graphs. Moreover, we set the length of paths as 3 for Cora, Citeseer and NBA while 4 for other datasets. We use 10 random splits with 48%, 32%, and 20% for training, validation and testing. For Cora, Pubmed, Citeseer and Cornell, we use the the same splits provided by [Zhu *et al.*, 2020]. For NBA, BGP, and Electronics, we randomly generate the 10 splits.

References

[Hoeffding, 1963] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[Zhu *et al.*, 2020] Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. In *NIPS*, volume 33, pages 7793–7804, 2020.