Supplementary: Adaptive and Interpretable Graph Convolution Networks Using Generalized Pagerank

Kishan Wimalawarne KISHANWN@GMAIL.COM
University of Tokyo, Hongo, Japan

Taiji Suzuki TAIJI@MIST.I.U-TOKYO.AC.JP
University of Tokyo, Hongo, Japan
RIKEN, Nihombashi, Tokyo, Japan

Editors: Emtiyaz Khan and Mehmet Gönen

1. Summary of Hyperparamter Selection

In this section we discuss provide the details of hyperparameters selected for the proposed method though the validation process.

For both experiments, we tuned the number of coefficients of the GPR as a hyperparamter selection from the set of $\{2, 3, 4, 8, 16, 32\}$. For fully-supervised node-classification, we used the same parameter ranges as in GCNII (Chen et al., 2020); 64 hidden units, learning rate 0.01, the number of layers from $(2, 4, 8, 16, 32, 64)$, $\lambda \in (0.5, 1.0, 1.5)$, $\alpha \in (0.1, 0.2, \ldots, 0.9)$, dropout $\in (0.1, 0.2, \ldots, 0.9)$, and weight decay $\in (0.001, 5\mathrm{e}{-3}, \ldots, 1\mathrm{e}{-6})$.

For semi-supervised node-classification, we fixed the learning rate with 0.01 and $\alpha = 0.1$ as as given in (Chen et al., 2020). We set the weight decay rate $WD_2 = 0.0001$ and applied hyperparameter tuning for weight decays for WD_1 and WD_3 from the set $(1.0, 0.1, 0.01, \ldots, 0.0001)$. Further we performed hyperparameter tuning for $\lambda \in (0.1, 0.2, \ldots, 0.9)$, dropout $\in (0.1, 0.2, \ldots, 0.9)$, and number of coefficients of the GPR K from the set $(2, 3, 4, 8, 16)$.

Details of the parameters selected using hyperparamter tuning for semi-supervised node classification and fully-supervised node-classification by the AdaGPR are listed in Tables 2 and 1, respectively.

2. Further Analysis of Trained Models

Table 3 shows coefficients of a semi-supervised learning model for Cora with 8 layers and 4 Pagerank coefficients. Though there are no sparseness among coefficients, notice that

Dataset	GPR Coeffs.	LR	WD_1	WD_2	WD_3	λ	α	Dropout
Cora	4	0.01	1.0	0.0001	0.1	0.1	0.3	0.6
Citeseer	16	0.01	1.0	0.0001	0.1	0.5	0.1	0.1
Pubmed	3	0.01	0.0001	0.0001	0.1	0.1	0.1	0.5

Table 1: Hyperparameters for semi-supervised node-classification

Dataset	GPR Coeffs.	layers	LR	Weight Decay	λ	α	Dropout
Cora	3	64	0.01	0.0001	0.5	0.1	0.5
Citeseer	2	64	0.01	0.0001	0.5	0.4	0.7
Pubmed	3	4	0.01	0.0001	0.5	0.5	0.2
Chameleon	3	2	0.01	0.001	1.5	0.6	0.6
Cornell	2	4	0.01	0.0001	1.0	0.9	0.4
Texas	4	4	0.01	5e-4	1.0	0.5	0.5
Wisconsin	3	16	0.01	5e-5	1.5	0.6	0.3

Table 2: Hyperparameters for fully-supervised node-classification

Layers	GPR Coeff.			
	0	1	2	3
1	0.2664	0.2606	0.2449	0.2279
2	0.2755	0.2601	0.2435	0.2207
3	0.2626	0.2733	0.2438	0.2201
4	0.2863	0.2574	0.2467	0.2093
5	0.2412	0.2861	0.2537	0.2188
6	0.2588	0.2726	0.2574	0.2111
7	0.2664	0.2854	0.2463	0.2017
8	0.1407	0.2933	0.2919	0.2740

Table 3: GPR coefficients of Cora

there is a gradual change of coefficients from shallow layers to deep layers. As the layers increase from the first to the seventh layers the largest coefficient shifts between the first two coefficients, while the forth coefficient gradually decreases. Recall that the coefficient at 0 represent the identity matrix with no graph convolution, hence, indicates that each layer need not have graph convolution.

Table 4 shows the GPR coefficients for semi-supervised node classification for Citeseer dataset using 16 payers and 16 GPR coefficients. Notice that coefficients in shallow layers, layer 1 to layer 8, roughly equal to 1/16. By analyzing the learning parameters for coefficients, we found that this is due to small values of the learning parameters in shallow layers. This may have caused by the application of softmax-like (sparsemax) activation to a set of values that are close to zeros. As the layers increases beyond 8, coefficients start to deviate and it becomes clear that each layer applies a convolution with a different generalized Pagerank. Furthermore, it is worth noticing that with the increase in layers the value of the first coefficient becomes prominent and the coefficients for the higher order terms gradually decreases.

Tables 5 and 6 further show learned Pagerank coefficients of trained models for Chameleon and Texas under fully-supervised node-classification. The trained model for Chameleon has graph convolution only at the second layer with the normalized adjacency matrix and the first layer act as a residual layer. The learned model for Cornell shows that only the first two layers apply graph convolutions with gradual adaptations of the GPR from shallow layers to deeper layers. An interesting observation is with the trained model for Texas, where it

Layers	GPR Coeff.															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062
2	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062
3	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062
4	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062
5	0.064	0.064	0.063	0.063	0.063	0.063	0.063	0.063	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.061
6	0.064	0.064	0.064	0.064	0.063	0.063	0.063	0.062	0.062	0.062	0.062	0.062	0.061	0.062	0.061	0.061
7	0.065	0.065	0.064	0.064	0.064	0.063	0.063	0.063	0.062	0.062	0.062	0.061	0.061	0.061	0.060	0.060
8	0.066	0.066	0.066	0.065	0.064	0.064	0.063	0.063	0.062	0.062	0.061	0.060	0.060	0.060	0.060	0.059
9	0.068	0.069	0.067	0.066	0.065	0.064	0.063	0.063	0.062	0.061	0.061	0.060	0.059	0.059	0.058	0.056
10	0.072	0.070	0.069	0.067	0.065	0.065	0.064	0.063	0.062	0.061	0.060	0.058	0.058	0.057	0.056	0.055
11	0.077	0.075	0.072	0.069	0.070	0.067	0.064	0.062	0.060	0.059	0.057	0.056	0.055	0.054	0.052	0.052
12	0.090	0.084	0.078	0.074	0.071	0.067	0.064	0.061	0.059	0.056	0.054	0.052	0.050	0.048	0.046	0.046
13	0.135	0.102	0.091	0.080	0.074	0.068	0.062	0.057	0.055	0.049	0.045	0.042	0.040	0.036	0.033	0.030
14	0.280	0.132	0.107	0.085	0.072	0.061	0.052	0.044	0.038	0.032	0.027	0.022	0.018	0.014	0.011	0.007
15	0.557	0.147	0.109	0.069	0.050	0.032	0.021	0.010	0.001	0.001	0.002	0.001	0.000	0.000	0.001	0.000
16	0.879	0.080	0.039	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 4: GPR coefficients of a trained AdaGPR model for Semi-supervised learning with Citeseer.

Layers	GPR Coeff.		
	0	1	2
1	1	0	0
2	0	1	0

Table 5: GPR coefficients of Chameleon

Layers	GPR Coeff.			
	0	1	2	3
1	1	0	0	0
2	1	0	0	0
3	1	0	0	0
4	1	0	0	0

Table 6: GPR coefficients of Texas

has no graph convolution in all four layers. By looking at these sparse GPR coefficients one may draw a conclusion that many of the above models (e.g. Texas) do not need any graph convolution at all. We have found that graph convolutions with higher orders are important during the learning process though the final trained model may have zeros or small values. In Figures 1a,1b,1c,1d, we show the change of values in GPR coefficients at each iteration with fully-supervised node classification for Texas with a AdaGPR model that consists of 4 convolution layers and 4 GPR coefficients.

3. Ablation Studies

We conducted ablations studies to understand the oversmoothing effect under layer-wise adaptive learning of AdaGPR. We compared AdaGPR with vanilla GCN and GPR convolution without adapted layer-wise coefficients. It is difficult to design a general GPR convolution with appropriate user specified coefficients. For simplicity, we considered the spacial case where all GPR coefficients are equal and assigned values of 1/(number of GPR coefficients).

Figure 2a,2b shows ablation plots of Cora and Citeseer for semi-supervised node classification. We can see that adaptive learning with AdaGPR improves accuracy with the

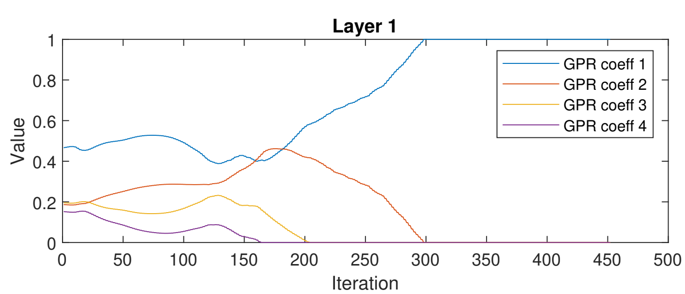

Figure 1a: Coefficient evolution of Layer 1 for Texas

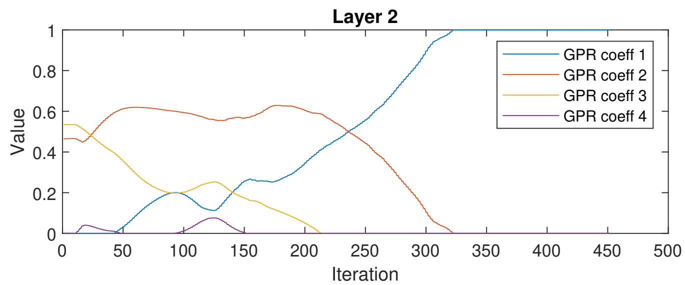

Figure 1b: Coefficient evolution of Layer 2 for Texas

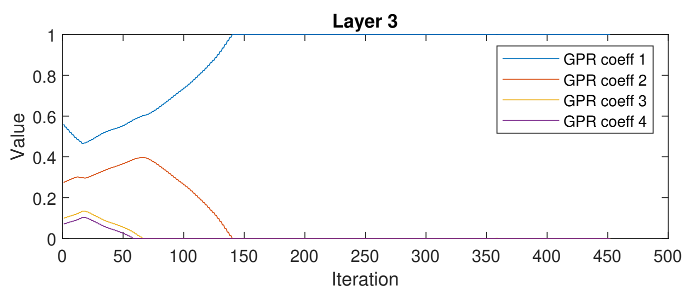

Figure 1c: Coefficient evolution of Layer 3 for Texas

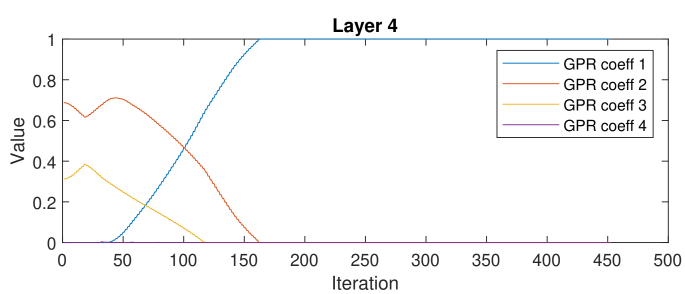

Figure 1d: Coefficient evolution of Layer 4 for Texas

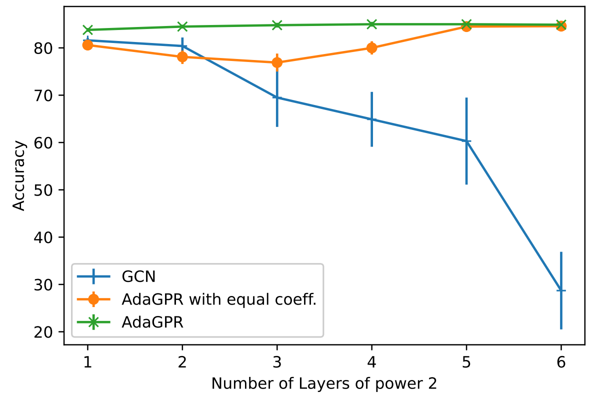

Figure 2a: Ablation study of Cora

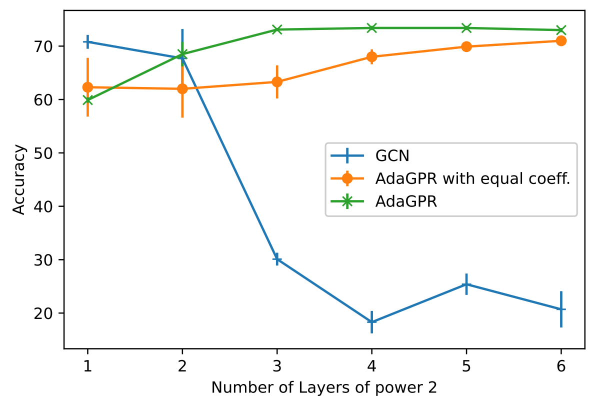

Figure 2b: Ablation study of Citeseer

increase of layers. Adaptive layer-wise learning of GPR coefficients consistently improve accuracy with the increasing number of layers compared to having constant GPR coefficients.

References

Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. Simple and deep graph convolutional networks. In *ICML*. PMLR, 2020.