

A MODEL DETAILS

Here we provide some more details on the models that we use. In all cases we use the Adam optimizer and tune the learning rate. We follow the models and hyperparameters provided in OGB (Hu et al., 2020) and wikiCS (Mernyei & Cangea, 2020) and manually tune some hyperparameters on the validation data for the potential of better performance.

For our MLPs, every linear layer is followed by batch normalization, ReLU activation, and 0.5 dropout. The other parameters depend on the dataset as follows.

- Products and Arxiv: 3 layers and 256 hidden channels with learning rate equal to 0.01.
- Cora, Citseer, and Pubmed (Getoor et al., 2001; Getoor, 2005; Namata et al., 2012) and Email (Leskovec et al., 2007; Yin et al., 2017): 3 layers and 64 hidden channels with learning rate = 0.01.
- wikiCS: 3 layers and 256 hidden channels with learning rate equal to 0.005.
- US County (Jia & Benson, 2020) and Rice31 (Traud et al., 2012): 5 layers and 256 hidden channels with learning rate equal to 0.005.

SOTA models for most datasets are taken from existing benchmarks. We determined SOTA for Email, US County, and Rice31 by evaluating several models discussed in the paper. The best performing baselines were as follows. For Email, GCNII with 5 layers, 256 hidden channels, learning rate equal to 0.01. For US County, GCNII with 8 layers, 256 hidden channels, learning rate equal to 0.03. For Rice31, we reused our GCN architecture and trained it over spectral embedding, which substantially outperformed the other GNN variants.

All models were implemented with PyTorch (Paszke et al., 2019) and PyTorch Geometric (Fey & Lenssen, 2019).

B PERFORMANCE RESULTS WITH ONLY THE CORRECTION STEP

Table 5 shows results with and without smoothing in the final predictions, i.e., just the “C step” vs. C&S. Including final prediction smoothing provides a substantial performance boost in many cases.

C ANALYSIS OF PERFORMANCE GAINS FROM SPECTRAL EMBEDDINGS

Table 6 shows the effect of including spectral embeddings as node features on the accuracy of the MLP-based and GCN models. In the case of the Arxiv dataset, including the spectral embedding improves the MLP base prediction performance substantially and the C&S performance modestly, but hardly changes the performance of the GCN. For Pubmed, including the spectral embeddings barely changes the performance of any model.

D ADDITIONAL VISUALIZATION

Full visualizations of C&S and GCN-SE performance for the US County dataset are in Figures 4 to 6. Similar visualizations for the Rice31 are in Figures 7 to 9, which are generated by projecting the 128-dimensional spectral embedding used in the main text down to two dimensions with UMAP (McInnes et al., 2018).

Table 5: Performance of our C&S framework with and without the final prediction smoothing. In cases where final prediction smoothing is used, only ground truth training are used.

| Method | Arxiv | Products | Cora | Citeseer | Pubmed |
|-------------------------------|-------|----------|-----------|----------|--------|
| Linear + C (Autoscale) | 66.89 | 74.63 | 79.56 | 72.56 | 88.56 |
| Linear + C&S (Autoscale) | 71.11 | 80.24 | 88.62 | 76.31 | 89.99 |
| Linear-SE + C (Autoscale) | 71.52 | 70.93 | 79.08 | 70.77 | 88.84 |
| Linear-SE + C&S (Autoscale) | 72.07 | 80.25 | 88.73 | 76.75 | 89.93 |
| MLP-SE + C (Autoscale) | 71.97 | 69.85 | 74.11 | 71.78 | 87.35 |
| MLP-SE + C&S (Autoscale) | 72.62 | 78.60 | 87.39 | 76.31 | 89.33 |
| Linear + C (Fdiff-scale) | 65.62 | 80.97 | 76.48 | 70.48 | 87.52 |
| Linear + C&S (Fdiff-scale) | 70.60 | 82.54 | 89.05 | 76.22 | 89.74 |
| Linear-SE + C (Fdiff-scale) | 70.26 | 73.89 | 79.32 | 70.53 | 84.47 |
| Linear-SE + C&S (Fdiff-scale) | 71.57 | 83.01 | 88.66 | 77.06 | 89.51 |
| MLP-SE + C (Fdiff-scale) | 71.55 | 72.72 | 74.36 | 71.45 | 86.97 |
| MLP-SE + C&S (Fdiff-scale) | 72.43 | 84.18 | 87.39 | 76.42 | 89.23 |
| Method | Email | Rice31 | US County | wikiCS | |
| Linear + C (Autoscale) | — | 43.97 | 82.60 | 77.49 | |
| Linear + C&S (Autoscale) | — | 75.99 | 85.25 | 79.57 | |
| Linear-SE + C (Autoscale) | 73.39 | 86.19 | 84.08 | 74.06 | |
| Linear-SE + C&S (Autoscale) | 72.50 | 86.42 | 86.15 | 79.53 | |
| MLP-SE + C (Autoscale) | 71.64 | 84.61 | 88.83 | 78.72 | |
| MLP-SE + C&S (Autoscale) | 74.55 | 85.50 | 89.64 | 78.10 | |
| Linear + C (Fdiff-scale) | — | 72.44 | 87.16 | 75.98 | |
| Linear + C&S (Fdiff-scale) | — | 73.66 | 87.38 | 79.54 | |
| Linear-SE + C (Fdiff-scale) | 71.31 | 85.22 | 88.27 | 73.86 | |
| Linear-SE + C&S (Fdiff-scale) | 72.53 | 87.55 | 88.11 | 79.25 | |
| MLP-SE + C (Fdiff-scale) | 72.59 | 85.42 | 89.62 | 78.40 | |
| MLP-SE + C&S (Fdiff-scale) | 75.74 | 85.74 | 89.85 | 78.24 | |

Table 6: Comparison of models with and without spectral embeddings, using only ground truth training labels for final prediction smoothing within C&S.

| Method | Arxiv | Products | Cora | Citeseer | Pubmed |
|----------------------------|-------|----------|-----------|----------|--------|
| GCN | 71.74 | 75.64 | 85.77 | 73.68 | 88.13 |
| GCN-SE | 71.76 | 76.12 | 85.83 | 73.60 | 88.32 |
| MLP | 59.67 | 59.23 | 74.21 | 69.34 | 86.73 |
| MLP-SE | 71.51 | 63.41 | 74.06 | 68.10 | 86.85 |
| MLP + C&S (Autoscale) | 71.76 | 79.42 | 87.56 | 76.42 | 89.29 |
| MLP-SE + C&S (Autoscale) | 72.62 | 78.60 | 87.39 | 76.31 | 89.33 |
| MLP + C&S (FDiff-scale) | 71.57 | 83.8 | 87.61 | 76.44 | 89.28 |
| MLP-SE + C&S (FDiff-scale) | 72.43 | 84.18 | 87.39 | 76.42 | 89.23 |
| Method | Email | Rice31 | US County | wikiCS | |
| GCN | — | 15.45 | 84.13 | 78.61 | |
| GCN-SE | 74.51 | 38.54 | 89.72 | 78.15 | |
| MLP | — | 15.73 | 87.77 | 71.42 | |
| MLP-SE | 69.13 | 17.16 | 87.70 | 73.07 | |
| MLP + C&S (Autoscale) | — | 85.05 | 89.67 | 78.92 | |
| MLP-SE + C&S (Autoscale) | 74.55 | 85.50 | 89.64 | 78.10 | |
| MLP + C&S (FDiff-scale) | — | 86.40 | 89.64 | 78.10 | |
| MLP-SE + C&S (FDiff-scale) | 75.74 | 85.74 | 89.85 | 78.24 | |

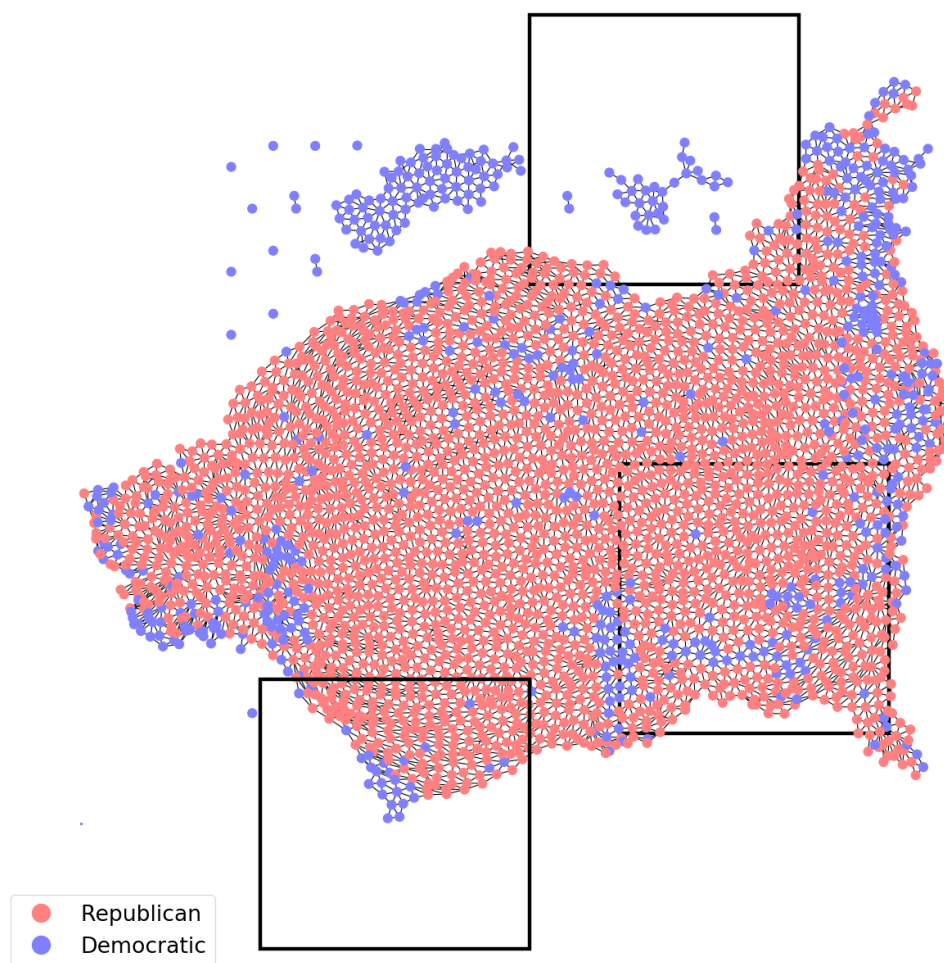


Figure 4: US County ground truth class labels.

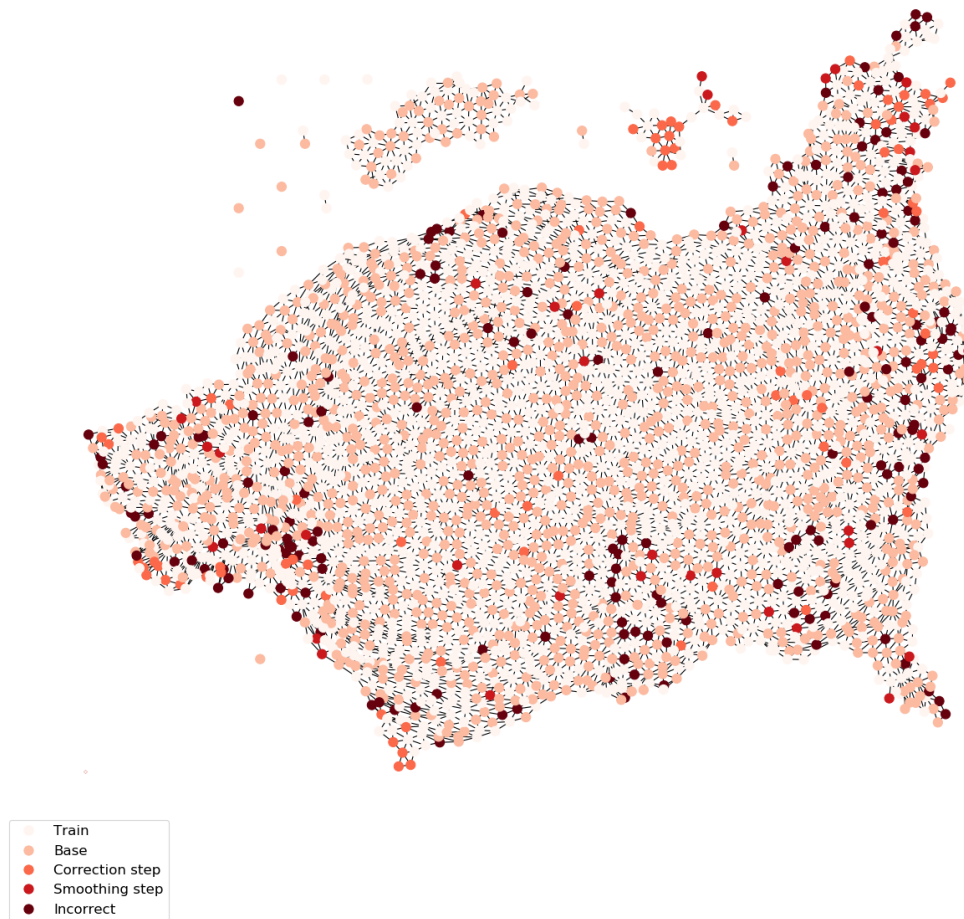


Figure 5: Linear-SE + C&S prediction performance on US County.

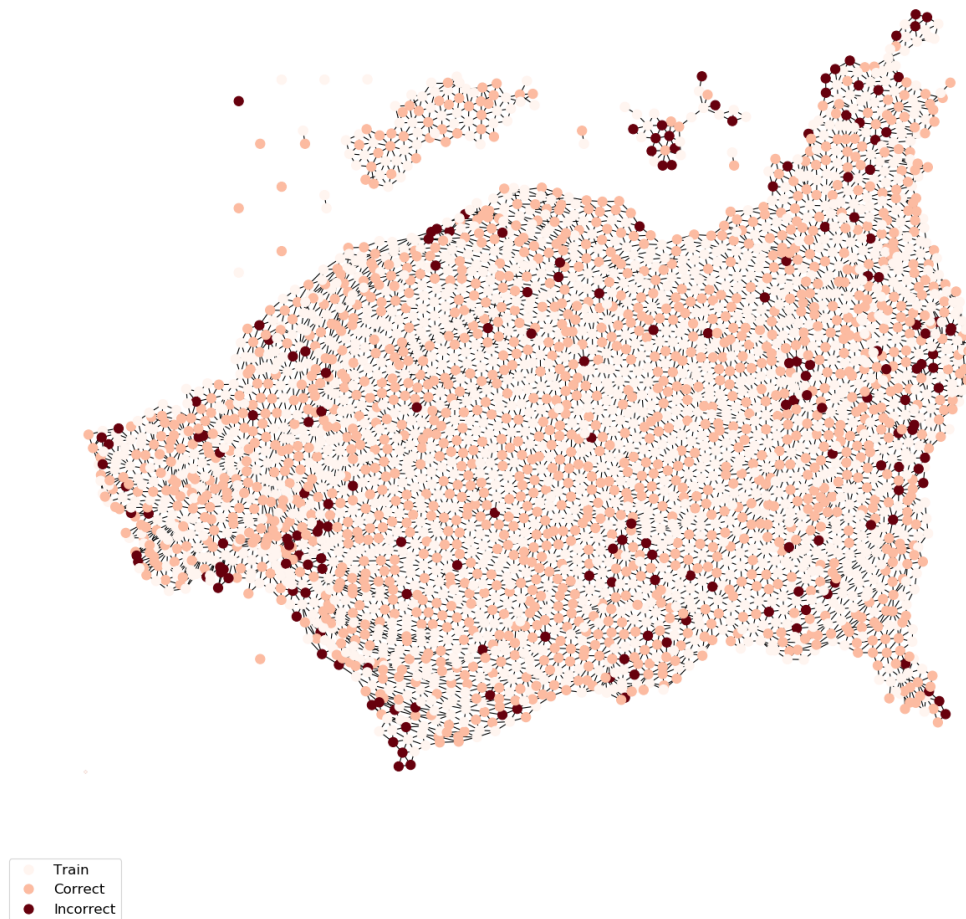


Figure 6: GCN-SE prediction performance on US County.

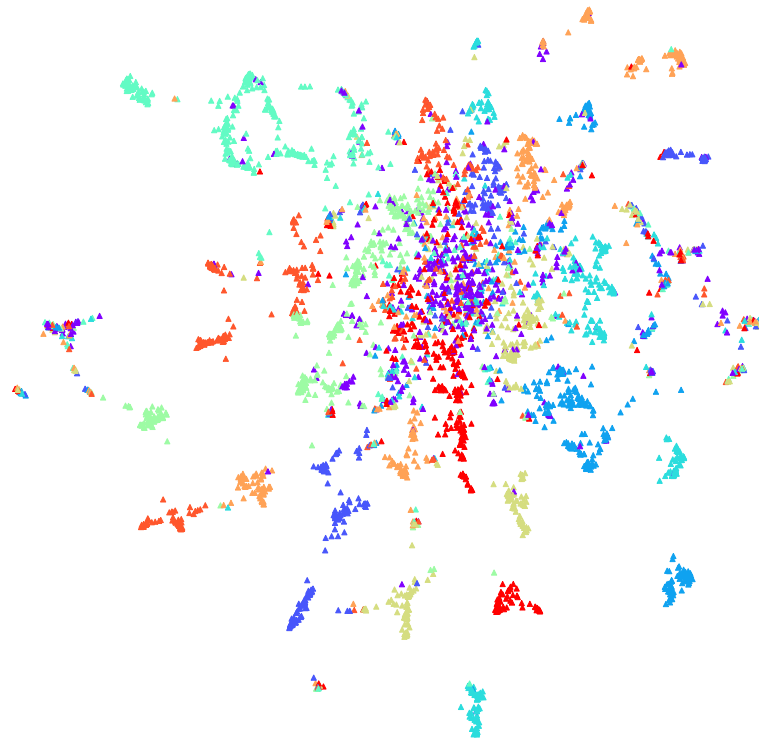


Figure 7: Rice31 ground truth class labels.

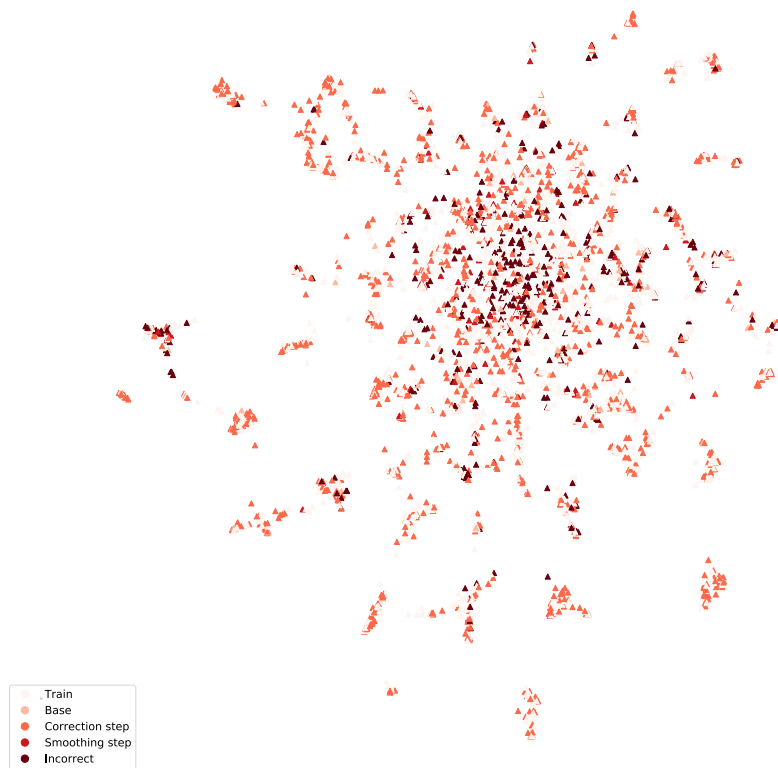


Figure 8: Linear-SE + C&S prediction performance on Rice31.



Figure 9: GCN-SE prediction performance on Rice31.