Renormalized Graph Representations for Node Classification: Supplementary Material

Appendix Index

Here's a brief summary of the contents of the various sections of the appendix:

- **Appendix A: Datasets.** We explain in more detail the datasets used in the experiments.
- **Appendix B: Additional experimental results.** We present a comparison between the models analyzed in the paper and those with a single encoder that only see the original graph. We show the test accuracy results as a function of epochs for all datasets.
- **Appendix C: Execution times.** We present a study of the execution times for the algorithm that calculates the characteristic scale and for the algorithm that aligns a graph to a characteristic scale through rewiring.
- **Appendix D: Density of the renormalized graphs.** We show the final number of edges after rewiring through the procedure presented in the main paper.

A Datasets

We conducted node classification experiments on the following datasets which are 3 citation networks and 2 co-purchasing networks:

The *Cora* dataset (McCallum et al., 2000) consists of publications categorized into one of 7 classes. It contains a total of 2,708 scientific papers connected by 5,429 citation edges. Each paper is represented by a 0/1 word vector indicating the presence or absence of a term from a dictionary of 1,433 distinct words.

Citeseer (Giles et al., 1998) is another citation network that represents publications similarly to Cora, but it uses a dictionary of 3,703 terms. It includes 3,327 publications connected by 4,732 citation edges classified in 6 classes.

The *PubMed* dataset (Namata et al., 2012) is the third citation network containing 19,717 scientific publications from the PubMed database, connected by 44,338 citation edges. Each publication is classified into one of 3 classes and is represented by a Term Frequency-Inverse Document Frequency (TF-IDF) word vector based on a dictionary of 500 words.

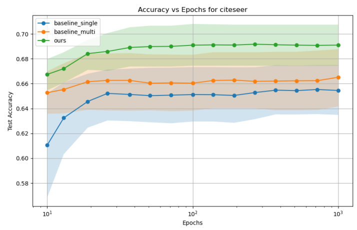
The *Amazon Photo* dataset (McAuley et al., 2015; Shchur et al., 2018) is a subset of the Amazon co-purchasing graph and contains 7,650 products with 119,043 edges. Each product in the dataset is described by a bag-of-words encoding of its reviews with a dictionary of 745 words. The graph is obtained through the "Also Bought" feature on Amazon, where an edge from item A to item B implies that item A is frequently co-purchased with item B.

The *Amazon Computers* dataset (Shchur et al., 2018) is analogous to Amazon Photo but contains 13,752 products connected by 491,722 links and represented through a dictionary of 767 words. Each product is associated to one of 10 classes.

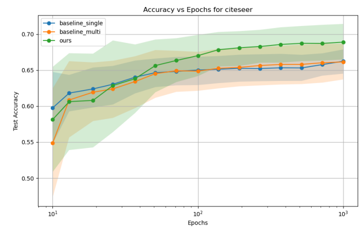
B Additional experimental results

B.1 Multi-encoder vs Single-encoder

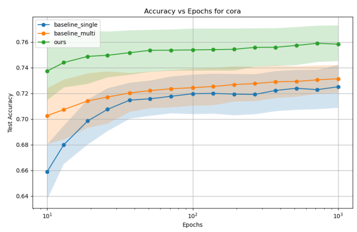
We present the comparison charts between models with multiple encoders that also observe the graphs aligned with characteristic scales ("ours"), models with multiple encoders that observe only the original graphs ("baseline_multi"), and models with a single encoder that see only the original graph ("baseline_single"). For further information about the models and the scales, please refer to the main paper.



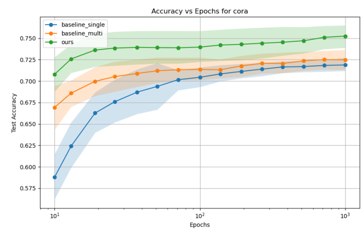
(a) Citeseer with GAT graph encoder



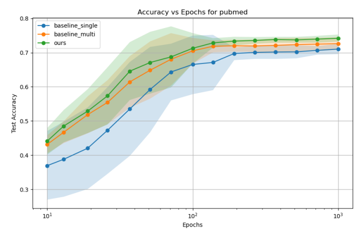
(b) Citeseer with GCN graph encoder



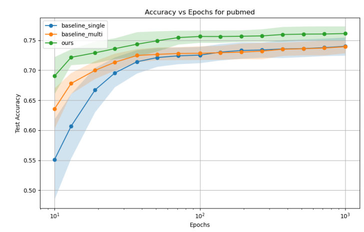
(c) Cora with GAT graph encoder



(d) Cora with GCN graph encoder

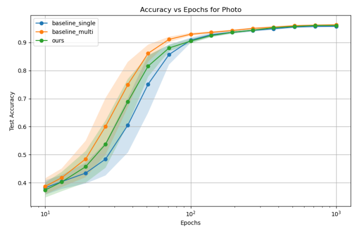


(e) Pubmed with GAT graph encoder

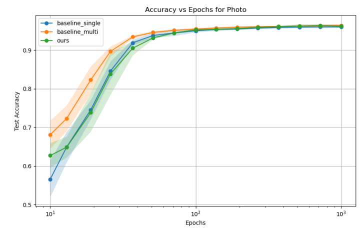


(f) Pubmed with GCN graph encoder

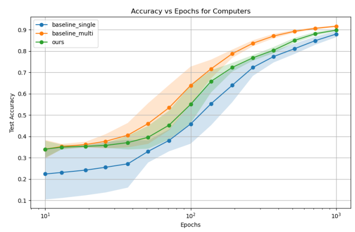
Figure 1: Test accuracy of the multiscale model, its baseline and a baseline singlescale model over epochs for the citation datasets.



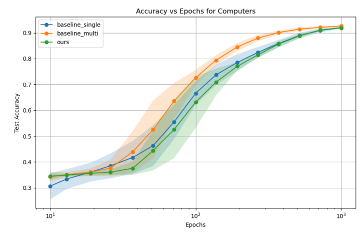
(a) Photo with GAT graph encoder



(b) Photo with GCN graph encoder



(c) Computers with GAT graph encoder

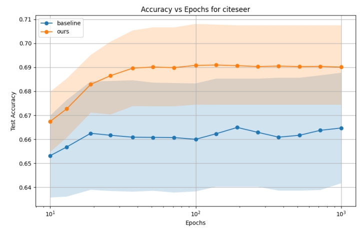


(d) Computers with GCN graph encoder

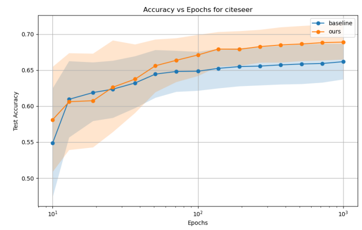
Figure 2: Test accuracy of the multiscale model, its baseline and a baseline singlescale model over epochs for the co-purchased products datasets.

B.2 Test Accuracy vs Epochs

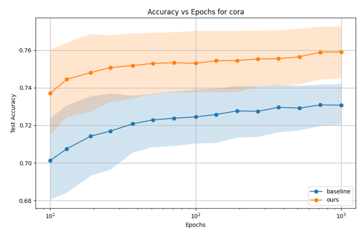
In the following section, we present graphs related to experiments similar to those shown in the main paper but conducted on all datasets. Specifically, the experiments focus on node classification tasks using a single split and 10 runs. We show the average test accuracy and its standard deviation for models that utilize not only the original graph but also the one aligned with the characteristic scale, referred to as "ours," compared to the architecturally equivalent model that only sees the original graph. For further information, please refer to the main paper.



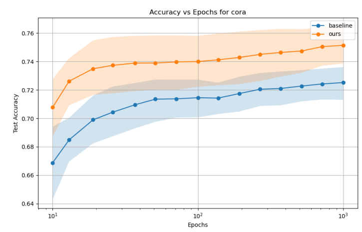
(a) Citeseer with GAT graph encoders



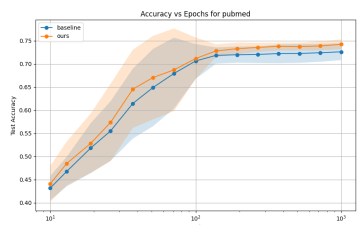
(b) Citeseer with GCN graph encoders



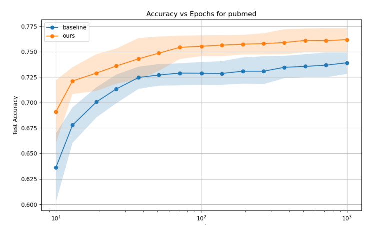
(c) Cora with GAT graph encoders



(d) Cora with GCN graph encoders

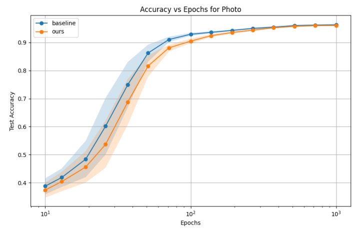


(e) Pubmed with GAT graph encoders

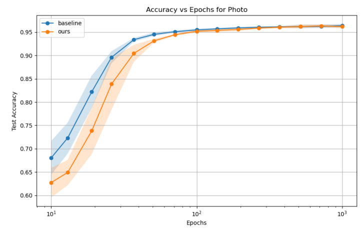


(f) Pubmed with GCN graph encoders

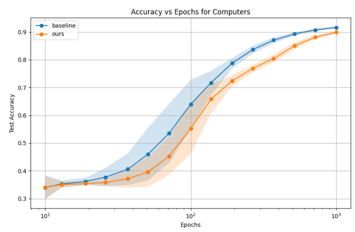
Figure 3: Test accuracy of the multiscale model and its baseline over epochs for the citation datasets.



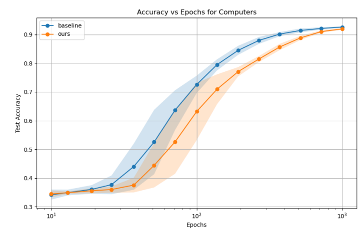
(a) Photo with GAT graph encoders



(b) Photo with GCN graph encoders



(c) Computers with GAT graph encoders



(d) Computers with GCN graph encoders

Figure 4: Test accuracy of the multiscale model and its baseline over epochs for the co-purchased products datasets.

C Execution times

We present the execution times obtained experimentally using the algorithm to find the characteristic scale and the algorithm to align a graph to a given scale through rewiring.

C.1 Specific Heat

We present the execution times obtained experimentally by calculating the derivative of the spectral entropy with respect to the logarithm of time, i.e., the specific heat, across various datasets. The specific heat was calculated using 100 values of τ within the range $[10^{-2}, 10^{4}]$ (defined by the numpy function numpy.logspace(-2, 4, 100) of the numpy library (Harris et al., 2020)) and 5 runs were performed. The mean and standard deviation across the runs are shown.

Table 1: Experimental execution times, in seconds, to obtain 100 specific heat values over a logarithmic scale interval between 10^{-2} and 10^4. The data refer to different graphs, identified based on the number of edges in the original graph.

Initial Edges	**Execution Time (s)**
7,358	400 ± 100
10,138	140 ± 40
88,648	21300 ± 500
238,086	788 ± 8
491,556	4400 ± 300

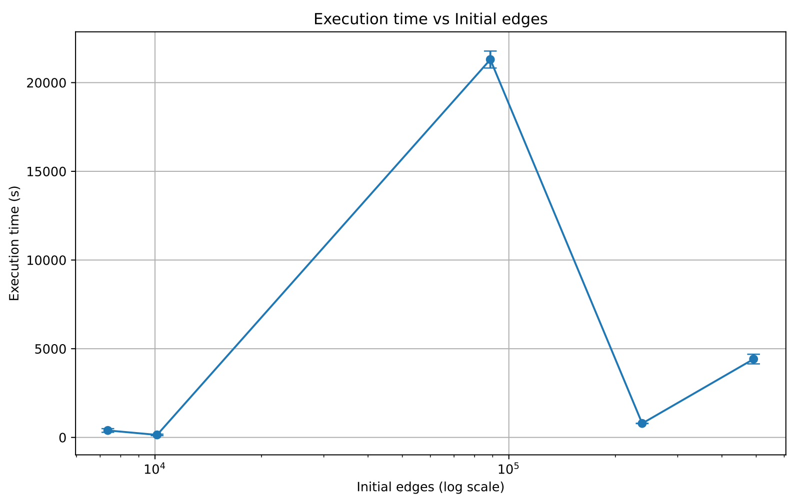


Figure 5: Experimental execution times, in seconds, to obtain 100 specific heat values over a logarithmic scale interval between 10^{-2} and 10^4 versus the number of edges in the original graph.

C.2 Rewiring

We present the execution times obtained experimentally using our implementation of the rewiring algorithm proposed in the paper to align a graph to a given scale. The results for the datasets used are reported. For each dataset, we show the values for three arbitrarily chosen τ values, as well as for the characteristic scale, which depends on the graph. The estimated characteristic scale is also provided. The statistics were obtained from 5 runs.

Table 2: Experimental execution times, in seconds, for aligning the graph to the respective τ scales using the proposed rewiring method. The data refer to different graphs, identified based on the number of edges in the original graph. The last column shows the value of the characteristic scale, $\tau^\star$.

Initial Edges	$\tau = 1.0$	$\tau = 2.0$	$\tau = 4.0$	$\tau = \tau^\star$	$\tau^\star$
7,358	5.2 ± 0.7	5.1 ± 0.6	5.2 ± 0.9	7 ± 3	0.57
10,138	10 ± 10	2.5 ± 0.6	2.8 ± 0.7	5 ± 5	0.57
88,648	240 ± 30	290 ± 20	410 ± 20	400 ± 100	1.52
238,086	140 ± 40	220 ± 20	290 ± 20	19 ± 2	0.07
491,556	750 ± 10	1400 ± 200	2600 ± 300	130 ± 90	0.07

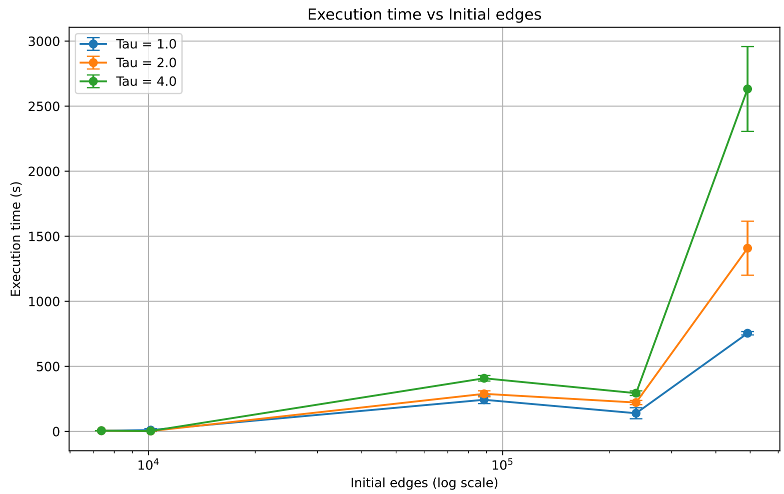


Figure 6: Experimental execution times, in seconds, for aligning the graph to the respective τ scales using the proposed rewiring method versus the number of edges in the original graph.

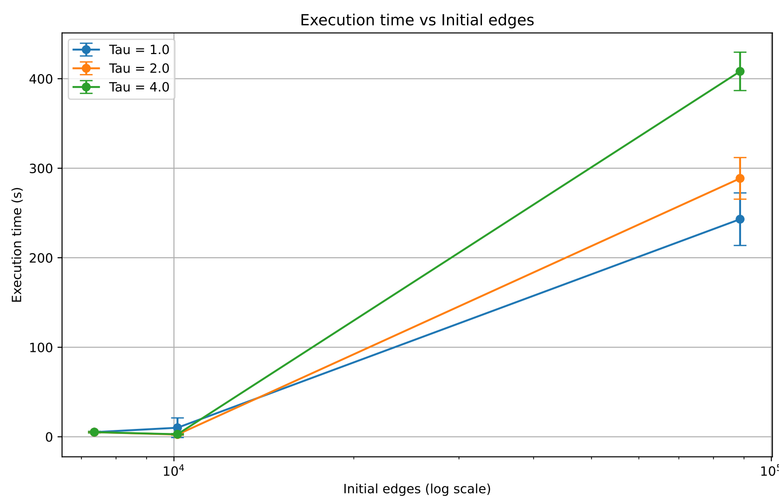


Figure 7: Experimental execution times, in seconds, for aligning the graph to the respective τ scales using the proposed rewiring method versus the number of edges in the original graph. Here, we present a zoomed-in view of the values for the initial number of edges that are less than 10^5.

D Density of the renormalized graphs

We show the original number of edges for the various datasets, along with the number of edges in the rewired graph aligned to three arbitrary τ scales and the respective characteristic scales. The estimated characteristic scale is also provided.

Table 3: Final number of edges after applying the rewiring procedure for various values of the scale τ across different datasets identified by the number of edges in the original graph. The last column shows the value of the characteristic scale, $\tau^\star$.

Initial Edges	$\tau = 1.0$	$\tau = 2.0$	$\tau = 4.0$	$\tau = \tau^\star$	$\tau^\star$
7,358	56,244	100,366	131,900	39,040	0.57
10,138	121,522	223,588	378,536	83,660	0.57
88,648	7,479,910	11,822,244	22,786,512	9,926,802	1.52
238,086	9,193,648	12,090,284	5,158,006	6,000,168	0.07
491,556	19,195,502	2,320,654	23,914	34,784,570	0.07

References

Giles, C. L., Bollacker, K. D., and Lawrence, S. (1998). Citeseer: An automatic citation indexing system. In *Proceedings of the third ACM conference on Digital libraries*, pages 89–98.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.

McAuley, J., Targett, C., Shi, Q., and Van Den Hengel, A. (2015). Image-based recommendations on styles and substitutes. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, pages 43–52.

McCallum, A. K., Nigam, K., Rennie, J., and Seymore, K. (2000). Automating the construction of internet portals with machine learning. *Information Retrieval*, 3:127–163.

Namata, G., London, B., Getoor, L., Huang, B., and Edu, U. (2012). Query-driven active surveying for collective classification. In *10th international workshop on mining and learning with graphs*, volume 8, page 1.

Shchur, O., Mumme, M., Bojchevski, A., and Günnemann, S. (2018). Pitfalls of graph neural network evaluation. *arXiv preprint arXiv:1811.05868*.