RESPONSE TO REVIEWERS

Mengzhe Jia, Hanxi Zhu, Zhengpin Li, Ruyu Wang, Jian Wang

School of Data Science, Fudan University

1. RESPONSE FOR REVIEWER 4140.

We thank the reviewer for the positive assessment of the novelty of our work and for the constructive comments. Below, we clarify the main concerns and explain how we will further improve the manuscript for clarity.

1. Technical Claims: Clarify imprecise statements (e.g., "learning GFT bases"). Detail the actual implementation/complexity of $\sigma(\mathbf{R}\mathbf{R}^\top) \odot \mathbf{A}$.

The phrase "learning GFT bases" may be imprecise and we thank the reviewer for pointing this out. More precisely, GRASP does not explicitly learn eigenvectors of the graph Laplacian. Instead, it learns two graph structures (homophilic and heterophilic subgraphs) via structure learning, each of which induces its own Laplacian and corresponding spectral space. These learned Laplacians implicitly define distinct graph Fourier domains associated with smooth (low-frequency) and non-smooth (high-frequency) signal variations. Based on these induced spectral spaces, GRASP applies simple and interpretable low-pass and high-pass filters on the respective subgraphs, which effectively prevent frequency mixing and avoid suboptimal embeddings commonly observed when filtering is performed on a single, fixed graph spectrum. We will revise the manuscript to replace the term "learning GFT bases" with a more precise description emphasizing learning graph structures that induce different spectral properties, rather than explicitly computing or parameterizing Fourier bases.

Regarding the complexity of $\sigma(\mathbf{R}\mathbf{R}^\top)\odot\mathbf{A}$, the discussion in the manuscript refers to space complexity, which is $O(cN+M)$ with $c \ll N$. The corresponding time complexity is $O(cN^2 + M)$, dominated by forming $\mathbf{R}\mathbf{R}^\top$, while $\mathbf{A}$ remains sparse. We will explicitly distinguish space and time complexity to avoid potential misunderstanding.

2. Experiments: Strengthen validation by: (a) using cleaned/newer heterophily benchmarks (e.g., Roman-Empire); (b) adding comparisons to node-wise adaptive spectral methods (e.g., NODE-MoE, DSF); (c) providing diagnostics on the learned subgraphs (e.g., edge homophily, spectral concentration).

To clarify the reviewer's concern regarding experiments, we report additional comparisons in Table 1 with recent node-wise adaptive spectral methods, including DSF and NODE-MoE, and evaluate GRASP on newer heterophilic benchmarks such as Roman-Empire. Since NODE-MoE is not publicly released, we report the results as stated in the original paper. Across both homophilic and heterophilic datasets, GRASP consistently achieves stronger performance compared to these baselines. Notably, on Roman-Empire, GRASP substantially outperforms DSF (85.59% vs.75.18%), indicating strong robustness under severe heterophily.

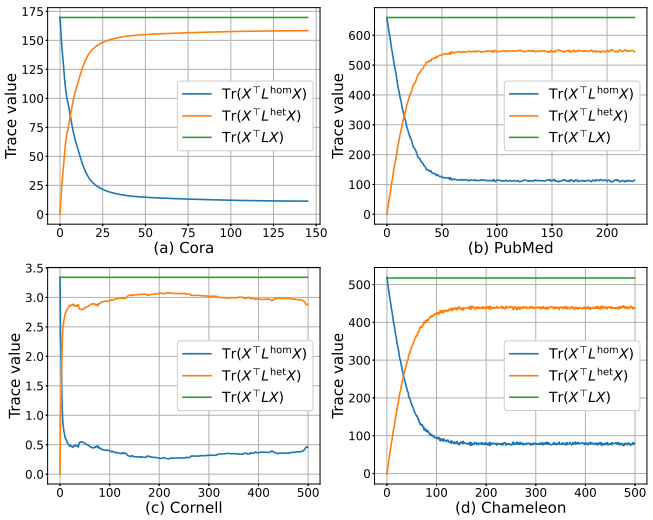

Fig. 1: The behavior of embedding smoothness and roughness regularizers during training on Cora, PubMed, Cornell and Chameleon. The x-axis stands for the training epochs.

We further provide diagnostic analyses of the learned subgraphs using spectral smoothness and roughness measured by Trace($\mathbf{X}^\top\mathbf{L}\mathbf{X}$) in Fig. 1. On Chameleon, the smoothness loss on the learned homophilic subgraph decreases from 520 to 110 over training, while the roughness loss on the heterophilic subgraph increases from near 0 to 410. This clear divergence empirically confirms that GRASP effectively separates low- and high-frequency components through graph splitting.

3. Theory: Substantiate the "tighter bound" claim with a clearer comparison to prior work or an empirical proxy.

The reviewer's concern regarding the "tighter bound" mainly relates to the clarity of comparison. In [1], prediction inconsistency is upper-bounded by a term proportional to $\|h_u - h_v\| + \rho|e_u - e_v|$, where e_u denotes local homophily. The

Corresponding author: jian_wang@fudan.edu.cn

Table 1: Node classification accuracy. We highlight the best results in bold and the second-best results by underlining.

Dataset	Homophilic			Heterophilic				
	Cora	Citeseer	PubMed	Chameleon	Cornell	Texas	Squirrel	Roman-Empire
GCN	87.78 ± 0.96	81.39 ± 1.23	88.90 ± 0.32	64.18 ± 2.62	82.46 ± 3.11	83.11 ± 3.20	44.76 ± 1.39	73.69 ± 0.74
GraphSAGE	86.58 ± 0.26	78.24 ± 0.30	86.85 ± 0.11	62.15 ± 0.42	71.41 ± 1.24	79.03 ± 1.20	41.26 ± 0.26	**85.74 ± 0.67**
GOAL	88.75 ± 0.87	77.15 ± 0.95	89.25 ± 0.55	71.65 ± 1.66	84.90 ± 0.66	92.02 ± 0.74	60.53 ± 1.60	72.30 ± 0.48
FAGCN	88.85 ± 1.36	82.37 ± 1.46	89.98 ± 0.54	65.47 ± 2.84	88.03 ± 5.60	88.85 ± 4.39	42.24 ± 1.20	65.22 ± 0.56
GPR-GNN	79.51 ± 0.36	67.63 ± 0.38	85.07 ± 0.09	67.48 ± 0.40	91.36 ± 0.70	92.92 ± 0.61	49.93 ± 0.53	64.85 ± 0.27
BernNet	88.52 ± 0.95	80.09 ± 0.79	88.48 ± 0.41	68.29 ± 1.58	92.13 ± 1.64	93.12 ± 0.65	51.35 ± 0.73	65.56 ± 1.34
UniFilter	88.28 ± 1.36	77.98 ± 5.28	91.24 ± 0.69	73.63 ± 2.03	85.74 ± 3.36	89.02 ± 3.52	**65.75 ± 0.96**	74.43 ± 0.25
DSF	89.63 ± 0.17	78.22 ± 0.29	90.51 ± 0.07	71.64 ± 0.55	84.93 ± 0.90	85.56 ± 0.93	58.44 ± 0.30	75.18 ± 0.37
NODE-MoE	89.38 ± 1.26	77.78 ± 1.36	89.58 ± 0.60	73.64 ± 1.80	closed-source	closed-source	62.31 ± 1.98	closed-source
GRASP	**90.77 ± 0.30**	**82.80 ± 0.43**	**91.40 ± 0.55**	**73.78 ± 2.27**	**93.82 ± 1.20**	**94.08 ± 2.64**	64.75 ± 1.71	85.59 ± 0.41

second term can dominate under heterophilic graphs due to unstable neighborhood compositions. In contrast, our bound (Theorem 1) eliminates the $|e_u - e_v|$ dependency by explicitly constructing homophilic and heterophilic subgraphs, yielding a tighter bound under heterophilic settings. We will make this comparison more explicit in the revised manuscript.

4. Clarity: Define ablation settings precisely (e.g., "w/o rel", "L_{F-hom}") and clarify hyperparameter details (e.g., τ, layer-wise sharing).

We will define ablation settings precisely, and further clarify statements that may be ambiguous to readers, precisely define all ablation settings, and explicitly specify hyperparameters such as τ, which are shared across layers.

We believe these clarifications fully address the reviewer's concerns and further demonstrate the soundness of our approach.

2. RESPONSE FOR REVIEWER 8E71.

We respectfully clarify several points that may stem from a misunderstanding or from focusing primarily on the introduction.

1. The paper would be strengthened by providing concrete examples of real-world networks that exhibit heterophily where connected nodes are likely to have different labels/attributes to better motivate the practical significance of the problem setting.

The heterophilic setting studied in this paper is motivated by widely used real-world benchmarks, including web and Wikipedia networks such as Chameleon, Squirrel, Cornell, and Texas, where connected nodes frequently belong to different semantic categories. These datasets are standard in heterophilic GNN research and have been extensively adopted in prior work. We will further emphasize this motivation in the revised introduction.

2. The statement that "traditional GNNs often produce unsatisfactory representations on heterophilic graphs" is currently asserted without sufficient support. Please add appropriate citations to prior work discussing this issue and/or include brief experimental evidence (even a small reference to baseline results) to substantiate the claim.

The statement that traditional GNNs struggle on heterophilic graphs is supported both empirically and by prior studies. As shown in our experimental results (Table X), classical message-passing models such as GCN and GraphSAGE exhibit clear performance degradation on heterophilic benchmarks. We will make this connection more explicit by adding citations and cross-referencing the experimental results in the introduction.

3. The manuscript would benefit from a clearer and more detailed explanation of what "low-frequency" and "high-frequency" information correspond to in heterophilic graphs. In particular, please clarify the intuition behind why high-frequency components become informative under heterophily, why separating low/high frequencies is important for node classification, and how these concepts relate to smoothing vs. preserving inter-class differences.

The discussion of low- and high-frequency information follows standard interpretations in spectral GNN literature and is already included in the manuscript. We will further streamline the explanation in the revised version for better readability.

4. There are still some typos and minor language issues in the manuscript (e.g., "ghe" → "the"). A careful proofreading is recommended to improve readability and presentation quality

We will carefully proofread the manuscript and correct the remaining typos and minor language issues.

3. REFERENCES

[1] Haitao Mao, Zhikai Chen, Wei Jin, Haoyu Han, Yao Ma, Tong Zhao, Neil Shah, and Jiliang Tang, "Demystifying structural disparity in graph neural networks: Can one size fit all?," in *Proceedings of the 37th Conference on Neural Information Processing Systems*, 2023, pp. 37013–37026.