A. Proof: Correlation Between High-frequency Area S_{high} and Homophily Level h

We aim to prove the negative correlation between the high-frequency area of the graph spectrum, denoted as S_{high}, and the homophily level h. To begin with, according to the graph Laplacian energy, we have:

$$x^T \mathbf{L} x = \sum_{(u,v)\in E} (x_u - x_v)^2 = \mathcal{E}_{\text{diff}},$$

where $\mathcal{E}_{\text{diff}}$ represents the sum of signal differences across edges. The expected value of the sum is given by:

$$\mathbb{E}[\mathcal{E}_{\text{diff}}] = \mathbb{E}\left[\sum_{(u,v)\in E} (x_u - x_v)^2\right].$$

Assuming that the difference $(x_u - x_v)^2$ between connected nodes u and v in graph G is independent and identically distributed (i.i.d) and that its expectation is influenced by the homophily level h, we have:

$$\mathbb{E}[(x_u - x_v)^2 \mid \mathbf{y}_u = \mathbf{y}_v] \ll \mathbb{E}[(x_u - x_v)^2 \mid \mathbf{y}_u \neq \mathbf{y}_v],$$

where $\mathbf{y}_u$ and $\mathbf{y}_v$ denote the labels of nodes u and v, respectively. Therefore, the overall expectation of the signal difference is approximately:

$$\mathbb{E}[(x_u - x_v)^2] \approx (1-h) \cdot \mathbb{E}[(x_u - x_v)^2 \mid \mathbf{y}_u \neq \mathbf{y}_v].$$

$$\mathbb{E}[\mathcal{E}_{\text{diff}}] \approx |\mathcal{E}| \cdot (1-h) \cdot \mathbb{E}[(x_u - x_v)^2 \mid \mathbf{y}_u \neq \mathbf{y}_v],$$

where $|\mathcal{E}|$ is the number of edges in the graph. Assuming that the signal x has length F and its elements x_f are i.i.d. following a normal distribution $\mathcal{P}(0, \sigma^2)$, we have:

$$\mathbb{E}[x^T x] = \mathbb{E}\left[\sum_{f=1}^{F} x_f^2\right] = \sum_{f=1}^{F} \mathbb{E}[x_f^2] = F\sigma^2.$$

Consequently, the expected value of S_{high} is expressed as:

$$\mathbb{E}[S_{\text{high}}] = \mathbb{E}\left[\frac{\mathcal{E}_{\text{diff}}}{F\sigma^2}\right] \approx \frac{|\mathcal{E}| \cdot (1-h) \cdot \mathbb{E}[(x_u - x_v)^2 | \mathbf{y}_u \neq \mathbf{y}_v]}{F\sigma^2},$$

$$\mathbb{E}[S_{\text{high}}] \propto (1-h),$$

which shows the negative correlation between the high-frequency area S_{high} and the homophily level h.

B. Fisher Information Matrix

Parameter sensitivity is quantified using the Fisher Information Matrix (FIM) [1, 17, 52], which provides a measure of the amount of information that the observed data carries about the GNN parameters. Let $\mathcal{G}$ represent the graph data, and let θ represent the parameters of the distribution $p(\mathcal{G}|\theta)$ underlying the GNN. The FIM is defined as:

$$\mathbf{I}(\theta) = \mathbb{E}_{\mathcal{G}\sim p(\mathcal{G}|\theta)}\left[\left(\frac{\partial \log p(\mathcal{G}|\theta)}{\partial \theta}\right)\left(\frac{\partial \log p(\mathcal{G}|\theta)}{\partial \theta}\right)^{\top}\right], \tag{13}$$

where the expectation is taken over the distribution of the observed graph data. The FIM essentially captures the curvature of the likelihood function in the parameter space, indicating how sensitive the likelihood is to small changes in the parameters. This sensitivity is crucial in various applications, including parameter estimation, uncertainty quantification, and model selection.

Nevertheless, the FIM for GNNs is computationally infeasible due to the high dimensionality of the parameter space. Therefore, an approximation is leveraged in our method. Specifically, the FIM can be approximated using an empirical distribution derived from the observed data. This approximation, particularly its diagonal form, simplifies the computation significantly:

$$\mathbf{I}_{\text{diag}}(\theta) \approx \mathbb{E}_{\mathcal{G}\sim p(\mathcal{G}|\theta)}\left[\left(\frac{\partial \log p(\mathcal{G}|\theta)}{\partial \theta}\right)^2\right]. \tag{14}$$

To further motivate, we consider the effect of small perturbations in the model parameters on the output of the GNN. This relationship can be rigorously quantified using the Kullback-Leibler (KL) divergence, which measures the difference between the original distribution and a perturbed distribution. Specifically, if the parameters are perturbed by a small amount δ, the second-order Taylor expansion of the KL divergence leads to the following approximation:

$$\mathbb{E}_{\mathcal{G}}\left[D_{\text{KL}}\left(p(\mathcal{G}|\theta) \| p(\mathcal{G}|\theta+\delta)\right)\right] \approx \frac{1}{2}\delta^{\top}\mathbf{I}(\theta)\delta, \tag{15}$$

where δ is a small perturbation in the parameter space. This result shows that the FIM not only measures the sensitivity of the likelihood function to parameter changes but also quantifies the expected change in model output as a result of these parameter perturbations. In other words, the FIM provides a fundamental connection between the parameter space and the output space of the GNN.

In summary, the Fisher Information Matrix serves as a measure of parameter sensitivity for GNNs. By approximating the FIM, we can quantify the posterior parameter sensitivity bias between a client and the globe. Correspondingly, a posterior bias-driven aggregation is proposed for reasonably measuring client contribution and enabling the global GNN to benefit more from those with lower biases.

C. Contextual Stochastic Block Model

In this research, we utilize the Contextual Stochastic Block Model (CSBM) [13] to generate synthetic graphs. These graphs feature variable edge probabilities both within and between different groups. The primary idea is that nodes of the same class share a uniform feature distribution. The resulting graph is denoted as $\mathcal{G} \sim \text{CSBM}(N, F, \sigma, \mu)$, where n is the total number of nodes, F represents the feature dimension, and σ and μ are the hyperparameters. These hyperparameters, σ and μ, control the influence of the graph structure and node features respectively. We consider two equal-sized classes, c_1 and c_0, each with $N/2$ nodes.

The CSBM produces features of a node u as follow:

$$\mathbf{x}_u = \sqrt{\frac{\mu}{N}}\mathbf{y}_u\zeta + \frac{q_u}{\sqrt{F}}, \tag{16}$$

where $\mathbf{y}_u \in \{-1, +1\}$ indicates the class label of node v_u, μ represents the mean of the Gaussian distribution, $\zeta \sim \mathcal{N}(0, I/F)$, and q_u consists of independently distributed standard normal variables. The average degree of the generated graph is denoted as d, and the adjacency matrix $\mathbf{A}$ for the CSBM graph is defined by:

$$P(\mathbf{A}_{uv} = 1) = \begin{cases} \frac{1}{N}(d + \sigma\sqrt{d}) & \text{if } \mathbf{y}_u = \mathbf{y}_v \\ \frac{1}{N}(d - \sigma\sqrt{d}) & \text{if } \mathbf{y}_u \neq \mathbf{y}_v. \end{cases} \tag{17}$$

The level of homophily h can be tuned by adjusting $\sigma = \sqrt{d(2h-1)}$, within the range $-\sqrt{d} \leq \sigma \leq \sqrt{d}$. A fully heterophilic graph is represented by $\sigma = -\sqrt{d}$, while a fully homophilic graph is described by $\sigma = \sqrt{d}$.

D. Datasets

We perform experiments on node classification tasks on both homophilic and heterophilic graph datasets to demonstrate the superiority of our proposed method FedSPA.

Cornell, Wisconsin: These datasets are subsets of the WebKB dataset [12]. The WebKB dataset was introduced in 1998, comprising web pages from the computer science departments of various universities, including Cornell University and the University of Wisconsin. These pages are categorized into five classes: student, faculty, course, project, and staff. In this dataset, each node represents a webpage, and edges denote hyperlinks between them. The dataset is commonly used for tasks such as webpage classification and link prediction, serving as a benchmark for evaluating machine learning models in graph-based learning scenarios.

Cora: The Cora dataset [54] is a widely used benchmark in machine learning and graph analysis, particularly for tasks like node classification and link prediction. It comprises 2,708 scientific publications categorized into seven classes: Case-Based Reasoning, Genetic Algorithms, Neural Networks, Probabilistic Methods, Reinforcement Learning, Rule Learning, and Theory. Each publication is represented as a node, and edges between nodes denote citation relationships, forming a citation network with 5,429 links. Node features are binary vectors indicating the presence or absence of 1,433 unique words from the publication's content. This dataset is instrumental in evaluating the performance of various graph-based algorithms and models.

Coauthor-CS, Coauthor-Physics: The Coauthor-CS and Coauthor-Physics datasets [63] are derived from the Microsoft Academic Graph and were used in the KDD Cup 2016 challenge. In these datasets, nodes represent authors, and edges denote co-authorship relationships. The Coauthor-CS dataset contains 18,333 nodes and 81,894 edges, with node features representing the keywords of papers authored by each individual and labels indicating classification into 15 fields of study. The Coauthor-Physics dataset includes 34,493 nodes and 247,962 edges, with similar node features and labels representing classification into 5 main research areas. These datasets are widely used for node classification tasks in graph neural network research as standard benchmarks to evaluate model performance.

Minesweeper: The Minesweeper dataset [5] is a synthetic graph dataset inspired by the classic Minesweeper game. In this dataset, the graph is structured as a regular 100x100 grid, where each node represents a cell connected to its eight neighboring nodes, except for edge nodes which have fewer neighbors. Approximately 20% of the nodes are randomly designated as mines. The primary task is to predict which nodes contain mines. Node features are one-hot encoded to represent the number of neighboring mines. However, for a randomly selected 50% of the nodes, these features are unknown, indicated by a separate binary feature. This dataset is commonly used to evaluate the performance of GNNs under heterophily.

ArXiv-year: The arXiv-year dataset [27] is a benchmark dataset designed for graph learning tasks, consisting of a citation network with nodes, edges, node features, and temporal labels. Each node represents a paper from the arXiv repository, and the edges denote citation relationships between papers, forming an undirected graph. The node features are typically embeddings related to the content of the paper, such as text-based representations. The node labels correspond to the publication year of each paper.