Graph Neural Networks have been widely used in graph machine learning. The applications can be drug discovery, social analysis, and recommendation system. In this work, we focus on the learning problem of Spectral graph neural networks, a popular class of GNN models apply graph convolution with polynomial spectral filters.

We begin by briefly reviewing graph Fourier transform. An attributed simple graph can be represented by an adjacency matrix A, a degree matrix D, and a node feature matrix X. The graph Laplacian is defined by the matrix formula, and can be decomposed into the eigenvalues and eigenvectors which are respectively known as frequency and frequency component. This allows us to transform node information between graph and frequency domains, i.e., to attain graph Fourier transform.

Spectral GNNs generally define a filter function on the graph spectrum. Such filter selectively shrinks or amplifies the Fourier coefficients of node features, and map them into a new desired space.

Here, we take one-dimension X as an example. As we can see the information update comes from the changing of the frequency component.

Most existing works employ polynomial spectral filter. They either design or learn the polynomial coefficients to approximate different types of filters. Despite their success, high degree polynomials are necessary as required by their expressive power.

However, empirical evidences show that as the polynomial order increases, the filter weights tend to approach zero. This essentially reveals the local modeling nature of existing spectral GNNs. Secondly, the inherent filtering process also suggests that the same set of coefficients are used for transforming different nodes.

All of this implies that existing spectral GNNs employ a homogeneous spectral filtering. Such filtering scheme implicitly assumes similar distributions between different local graph regions. This hypothesis however may not be accurate due to the intrinsic complexity in forming real-world networks.

To make further investigation, we define two essential properties on the local graph level and observe their distribution. The first property is Local Label Homophily, which is derived by constraining the computation of the well known Label Homophily within k-hop neighborhood. The second property is called Local Graph Frequency. Likewise, it limits the calculation of graph frequency to the k-hop subgraph.

The figures display the distributions of these two properties on multiple real-world networks. Most of them exhibit diverse linking pattern with skewed and even multi-modal distributions. These phenomena imply that the local structural patterns may not be uniformly distributed between different graph parts, but showing evident heterogeneity.

To this end, we propose a diverse spectral filtering. Specifically, the original transforming coefficient is augmented from a scalar to a vector, which specifies a different transforming coefficient for a different node, and we use local graph frequency to compute it for incorporating local information.

Nevertheless, the computation of local graph frequency is mostly time consuming as it requires both Laplacian decomposition and subgraph extraction.

To avoid this, we propose a substitution strategy with the original graph frequency. Using proposition 1, our diverse spectral filtering can be formulated by personalizing each node a different set of filter weights.

However, the challenge of parameterizing a vast number of filter weights, which is directly proportional to the graph size, still persists. This would increase computational complexity, and cause severe overfitting to local noises. The figure shows a clear accuracy drop while learning with full-parameterization. In this work, we argue that a reasonable design should be built upon a shared global model while locally adapted to each node with awareness of its location in the graph.

Such a proposition is well evidenced by the key observations that nearby nodes tend to display similar local structures because of their overlapped neighborhoods. For distant nodes, they may still posses similar local contexts due to some global properties. See the nodes in green circles. Albeit being far apart, they all reside at a “line” type of subgraph. Therefore, our goal is to take advantage of such rationale as a guide to acquire node-specific filter weights, and introduce two strategies.

Our first strategy involves decomposing the original filter weights into global and local parameters. This allows for the capture of invariant graph properties and diverse node contexts. The global ones are learned as arbitrary parameters. For the learning of local filter weights, we propose incorporating node positional information as our second strategy. To achieve this, a novel objective is formulated to encode node positions into a latent space. The first two terms enforce proximity between nearby nodes while preserving their original positional information. The last penalty term ensures orthogonal feature channels for attaining a valid coordinate system.

We solve this minimization problem with an iterative gradient method, and then predict the node-specific filter weights with the encoded positional features.

Despite reasonableness, our current design still experiences computational overhead while promoting orthogonality. To tackle this, we introduce another variant with regularization in training, and remove the quadratic computational complexity.

Happily, this also works in practice, and the new variant is able to reduce more than 75% running time.

To evaluate model performance, we conducted extensive experiments on 11 node classification benchmarks. The results show that our DSF framework constantly boost classification accuracy. We observed particularly notable improvements in cases where the graphs are heterophilic, containing diverse linking patterns. Our DSF framework also improve model interpretability.

The existing spectral GNNs, e.g., BernNet, only learns a single filter that is inadequate to model the varying local patterns. This would result in limited interpretability on the micro graph mining.

In comparison, our diverse filters display similar overall shapes but different local details. It shows the capability of our DSF in capturing both conformal and disparate regional information on the graph.

For citation network citeseer, most nodes display similar local structures with uniform distribution. Therefore, the learned filters produce almost identical curves with minor variations. It further exemplifies the universal modeling capability of our DSF framework.

Here, we also present t-SNE visualizations of the filter weights specific to each node. The color likeness reflects node similarity. Overall, the different regional patterns can be differentiated, and distant nodes with comparable local structures still acquire similar filter weights. This exactly matches our previous analysis and observations, demonstrating the remarkable interpretability of our DSF framework.

In summary, our work reveals that many existing spectral GNN models are restricted in homogeneous spectral filtering, and identify the need to break this ceiling to deal with complex graphs with regional heterogeneity. To this end, we propose a novel diverse spectral filtering framework, which not only achieves substantial accuracy gains but also improves model interpretability.