SENDNet: Spectral Energy Distributions of Vertex Signals for Graph Readouts

Anonymous submission

Abstract

A long-lasting difficulty within the task of graph classification remains the graph readout step, where it is required to summarize the global information of an attributed graph with a vector of fixed size. The usual work-around for graph neural networks is to learn node embeddings of large size with hop aggregations which are ultimately pooled globally. This may be the wrong approach for a graph whose topological information commands more attention than its nodal information. Here, the graph vertex signal (the collection of node embeddings) ought to be described mainly by its behavior across the network, and less-so by the prominence of certain node-wise features. In this context, we propose to focus on the shape of the spectral energy distribution of the graph signal, i.e. the plot of graph signal energy versus vertex frequency, after having leveraged traditional message-passing layers to construct shallow node embeddings. We show that this information is aptly captured by penalized B-spline interpolation over the spectral domain with careful choice of knots.

Introduction

A popular approach to graph readouts is the graph pooling framework (Li et al. 2025), which performs recursive subsampling on the graph (often interlaced with the message propagation/filter layers) followed by the application of a global, permutation-invariant function ϕ to each graph subsample (Xu et al. 2019). A common issue lies in the summative or extremizing nature of ϕ required for permutation-invariance (). If on one channel the signal retains large support over many vertices instead of being meaningfully localized to a few vertices, then global pooling incurs an unavoidable loss of local information for that channel. Such a treatment is maybe not ideal for graphs whose vertex annotations initially hold little information. In these cases, an optimally discriminative model ought not to look at the global prominence of certain node features (as is done with global pooling), but should instead look at how they are arranged throughout the network. To this end, we rely on the framework of vertex frequency analysis as a supplement to the message-passing framework that can hopefully supersede global pooling techniques. After learning node embeddings of reasonably small size, we compare attributed graphs by looking at the spline-smoothed spectral energy distribution of their vertex signals on a per-channel basis. We name our technique SENDNet, for "Spectral ENergy Distribution Network."

Methodology

Spectral Energy Distribution of Graphs

To an attributed graph with n nodes we associate some vertex signal $f \in \mathbb{R}^{n\times C}, f = [f_1, ..., f_C]$ where C is the number of channels, along with its normalized Laplacian $\mathcal{L} \in \mathbb{R}^{n\times n}$ with diagonalization yielding the spectrum $\sigma(\mathcal{L})$–each $\lambda \in \sigma(\mathcal{L})$ has an associated eigenbasis X_λ. Consider an arbitrary channel-wise signal $f_i, 1 \leq i \leq C$; for each eigenvalue λ the energy of f_i within the associated eigenspace of λ is $E_{\lambda,i} = \sum_{\chi_\lambda \in X_\lambda} \langle \chi_\lambda, f_i \rangle^2$. The tuple-set $\{(\lambda, E_{\lambda,i})\}_{\lambda\in\sigma(\mathcal{L})}$ thus details the energy distribution (SED) of f_i.

To compare two SEDs, we turn to smooth interpolation by a finite-dimensional function space: we choose a B-spline basis of degree 3. The heuristic is that two similar graphs should have similar spectra (Wills and Meyer 2020), and their vertex signal at channel $1 \leq i \leq C$ should have similar energy distributions; for example, maybe both signals are well-represented by frequency components corresponding to a particular band (interval) of [0,2]. Hence, a B-spline with support overlapping that region should be scaled by a large coefficient in both cases. We use the Graph Isomorphism Network (GIN) (Xu et al. 2019) message-passing model to learn shallow node embeddings (at most 5 channels). Learned filter kernels (one per channel) with selective support in [0,2] modulate the Fourier coefficients before the spectral energy is computed and regression is done. The spline coefficients for each channel are then concatenated and fed to a fully-connected neural network classification layer.

Splines, Penalty, and Choice of Knots

The basis expansion is done with K B-splines $\{B_1, ..., B_K\}$ with penalty $\mu||D_2\alpha||_2$ on the spline coefficients α, where D_2 is a second-order finite difference operator. In practice, the penalty parameter $\mu > 0$ is kept small. Hence, the choice of knots $\{t_i\}_{i=1}^{K+4} \subset [0, 2]$ should be done in a way so as to minimize the likelihood that the Gramian matrix $B^T_{\sigma(\mathcal{L})}B_{\sigma(\mathcal{L})}$ of the B-spline evaluation matrix $B_{\sigma(\mathcal{L})}(\lambda, k) := B_k(\lambda), 1 \leq k \leq K, \lambda \in \sigma(\mathcal{L})$ is singu-

lar (this occurs when a spline is left with empty support). To this end, we contend ourselves with approximating Laplacian spectra of reasonably large size N by N i.i.d. copies of some $x \sim P'$, where P' is the marginal distribution of a single eigenvalue when sampling spectra from the true model. Per the coupon collector problem, the arrangement of M bins (here, interpreted as the knot intervals $[t_i, t_{i+1}]$) that minimizes the waiting time for any m bins to be hit by i.i.d. samples of P' corresponds to the M-quantiles of P' (Brown and Ross 2016).

Locality of Spline Regression

For the B-spline coefficients to be comparable, we require that regression be done on a local scale: the coefficients of splines supported on one region of [0,2] should be minimally influenced by the activity occuring in its complement. That this requirement may be satisfied in the modest K (and small μ) regime can be seen through the decay rate of the inverse of $A = B_{\sigma(\mathcal{L})}^T B_{\sigma(\mathcal{L})} + \mu D_2^T D_2$ (recall that the coefficients are given by $\alpha = A^{-1} B_{\sigma(\mathcal{L})}^T y$, where y is the vector of y-values to be interpolated). Namely, if $r = \nu_{\max}(A)/\nu_{\min}(A)$ is the condition number of A, then by a result of Demko et al. on inverses of banded matrices (Demko, Moss, and Smith 1984), we have that the entry at row i, column j of A satisfies

$$|A^{-1}(i,j)| \leq \frac{(1+r^{1/2})^2}{2\nu_{\max}(A)} \left(1 - \frac{2}{\sqrt{r}-1}\right)^{\frac{1}{2}|i-j|}.$$

Thus, for modest K such that a sample spectrum $\sigma(\mathcal{L})$ is likely to hit all spline supports $[t_k, t_{k+4}], 1 \leq k \leq K$, we can expect the condition number $r \geq \sigma_{\max}(A)/\sigma_{\min}(B_{\sigma(\mathcal{L})}^T B_{\sigma(\mathcal{L})})$ to behave well-enough to yield considerable exponential decay in the kth row of A^{-1} when moving away from the kth index. Consequently, the kth coefficient should be mostly determined by some sub-vector of $B^T y$ centered about index k corresponding to the activation of the kth and neighboring splines.

Experiments and Results

At this stage, we only assess the performance of SENDNet as an alternative readout layer for classification.

References

Brown, M.; and Ross, S. M. 2016. Optimality Results for Coupon Collection. *Journal of Applied Probability*, 53(3): 930–937.

Demko, S.; Moss, W. F.; and Smith, P. W. 1984. Decay rates for inverses of band matrices. *Mathematics of Computation*, 43(168): 491–499.

Li, Z. P.; Wang, S. G.; Zhang, Q. H.; et al. 2025. Graph Pooling for Graph-level Representation Learning: A Survey. *Artificial Intelligence Review*, 58: 45.

Wills, P.; and Meyer, F. G. 2020. Metrics for Graph Comparison: A Practitioner's Guide. *PLoS ONE*, 15(2): e0228728.

Xu, K.; Hu, W.; Leskovec, J.; and Jegelka, S. 2019. How Powerful are Graph Neural Networks? In *International Conference on Learning Representations*.