2025 IEEE 10th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) | 979-8-3315-2669-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/CAMSAP66162.2025.11423962

On the Impact of Downstream Tasks on Sampling and Reconstructing Noisy Graph Signals

Baskaran Sripathmanathan, Xiaowen Dong, Michael Bronstein
University of Oxford

***Abstract*—We investigate graph signal reconstruction and sample selection for classification tasks. We present general theoretical characterisations of classification error applicable to multiple commonly used reconstruction methods, and compare that to the classical reconstruction error. We demonstrate the applicability of our results by using them to derive new optimal sampling methods for linearized graph convolutional networks, and show improvement over other graph signal processing based methods.**

***Index Terms*—Graph signal processing, sampling, reconstruction, least squares, classification.**

I. Introduction

Signal processing and machine learning on graphs have become increasingly common, with many applications such as analysing and predicting brain fMRI signals [1], urban air pollution [2], political preferences [3] and protein function [4]. However, data defined on a network domain, often called 'graph signals', is frequently noisy and sometimes even incomplete, making such tasks challenging.

The graph signal processing (GSP) community has addressed the issue of noisy and incomplete data by generalising the tools of sampling and reconstruction from classical signal processing, via extending the classical shift operator to a graph shift operator [5] such as the Laplacian. The majority of studies on graph-based sampling focus on designing efficient sampling schemes for criteria which approximate [6]–[12] or bound [13], [14] reconstruction loss, grounded in the optimal design of experiments [15], with some work simultaneously learning the underlying graph topology [16]–[18]. Theoretical results have also been derived on the performance of sample choice and signal reconstruction under reconstruction loss [11], [19]–[23]. However, these results do not immediately carry over to other tasks such as classification, as they rely on the specific structure of reconstruction loss – e.g., the proof of near-optimality of greedy sampling on graphs [20] relies on the loss's (approximate) supermodularity. Thus, there is a significant lack of investigation in the GSP literature of the impact of incomplete data on downstream learning tasks.

The graph machine learning (GML) community, on the other hand, does address the issue of incomplete data, often in the context of downstream tasks such as node classification. They take two main approaches – label prediction and feature imputation. Label prediction sidesteps the issue of missing data by directly predicting the missing node class labels, such as by label propagation [24] or by using graph neural networks (GNNs) on the partial data with modified message passing [25], [26]. Feature imputation involves reconstructing the graph signals, potentially alongside missing edges, usually as a pre-processing step to inference. Such methods include matrix completion methods [27] and autoencoder-based methods [28]. Another example is feature propagation [29], which is a graph Laplacian-based feature imputation method that applies a GSP-motivated approach to downstream tasks such as classification with GNNs. Finally, some approaches combine label prediction and feature imputation in an end-to-end training framework [30], [31]. While work in GML engages with missing data in settings with varied downstream tasks, it has done so through an empirical lens – evaluating the performance of the proposed methods on specific datasets – and thus theoretical results remain scarce.

The separation of the two lines of work above points to an important aspect of dealing with incomplete data: sampling and reconstruction should be designed to serve the downstream task at hand and, depending on the task, the "reconstruction error" we care about might be different. For example, with recommender systems with incomplete participant data, our goal is to make effective recommendations (which can be formulated as a classification task) rather than reconstructing the attributes of the participants.

In this paper, we fill this gap in the literature by providing a theoretical characterisation of the performance of binary classification with a linearized graph convolutional network (GCN) under missing data and noisy observation (Theorem 1). We provide a bound showing the relationship between classification and reconstruction error (Corollary 1.1), and compare the change in classification versus reconstruction error as a function of the sample size both theoretically and empirically. We show that optimal sampling derived in the reconstruction setting can underperform random sampling when applied to classification tasks. We then apply our theoretical results to derive a novel sample selection scheme for classification under a linearized GCN, and demonstrate that it improves upon using sample selection schemes which are optimal in the reconstruction setting.

II. Background

A. Graphs and graph signals

A graph $\mathcal{G}$ consists of a set $\mathcal{V}$ of N vertices, a set of edges $\mathcal{E}$ between these vertices and a corresponding set of edge weights. We assume that $\mathcal{G}$ is connected and undirected. A graph signal $\boldsymbol{x}$ is a real-valued function $\boldsymbol{x} : \mathcal{V} \to \mathbb{R}$, equivalently written as a vector in $\mathbb{R}^N$ by having each component correspond to a vertex. We define the adjacency matrix $\boldsymbol{A}_{ij} = \mathbb{1}\{(i,j) \in \mathcal{E}\}$, the degree matrix $\boldsymbol{D}_{ij} = \deg(i)\mathbb{1}\{i = j\}$, the normalized adjacency matrix $\tilde{\boldsymbol{A}} = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}$ and the normalized augmented adjacency matrix $\tilde{\boldsymbol{A}}_\gamma = (\boldsymbol{D}+\gamma\boldsymbol{I})^{-\frac{1}{2}}(\boldsymbol{A}+\gamma\boldsymbol{I})(\boldsymbol{D}+\gamma\boldsymbol{I})^{-\frac{1}{2}}$ where $\boldsymbol{I}$ is the identity matrix. This augmentation can be understood as adding self-loops to $\mathcal{G}$, and note adding no loops yields $\tilde{\boldsymbol{A}}_0 = \tilde{\boldsymbol{A}}$. Throughout this paper, we consider the symmetrically normalized Laplacian $\boldsymbol{L} = \boldsymbol{I} - \tilde{\boldsymbol{A}}$.

B. Sampling of graph signals

For any matrix $\boldsymbol{X}$, we define $[\boldsymbol{X}]_{\mathcal{A},\mathcal{B}}$ to be the submatrix of $\boldsymbol{X}$ with row indices in $\mathcal{A}$ and column indices in $\mathcal{B}$. Similarly, we write $[\boldsymbol{x}]_\mathcal{A}$ for the subvector of $\boldsymbol{x}$ with indices in $\mathcal{A}$. We let $\mathcal{N} = \{1,\dots,N\}, \mathcal{D} = \{1,\dots,d\}$ and $\mathcal{K} = \{1,\dots,k\}$, and for any index set $\mathcal{S}$, we write $\mathcal{S}^C = \{i \in \mathcal{N} \mid i \notin \mathcal{S}\}$. We write $\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ to be the eigendecomposition of $\boldsymbol{L}$, so the columns of $\boldsymbol{U}$ correspond to the eigenvectors of $\boldsymbol{L}$, ordered by eigenvalue in increasing order

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on April 08,2026 at 07:06:07 UTC from IEEE Xplore. Restrictions apply.

(all of which are nonnegative). We define $\mathbf{\Pi}_{bl(\mathcal{K})} = [\boldsymbol{U}]_{\mathcal{N},\mathcal{K}}[\boldsymbol{U}]^T_{\mathcal{N},\mathcal{K}}$ to be the projection matrix which makes a signal k-bandlimited, i.e., it acts as a ideal bandpass filter, keeping only the components corresponding to the k lowest eigenvalues of $\boldsymbol{L}$.

C. Reconstruction Methods

A *reconstruction method* takes (potentially) noisy observations of a graph signal on a node sample set $\mathcal{S}$. We consider linear methods such as Least Squares (LS) or Feature Propagation (FP) [29]. This means that for a fixed sample set $\mathcal{S}$, we can write the reconstruction operator as a matrix $\boldsymbol{R}_{\mathcal{S}} : \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}^N$:

$$\text{LS:} \quad \boldsymbol{R}_{\mathcal{S}} = [\boldsymbol{U}]_{\mathcal{N},\mathcal{K}}[\boldsymbol{U}]^{\dagger}_{\mathcal{S},\mathcal{K}} \tag{1}$$

$$\text{FP:} \quad \boldsymbol{R}_{\mathcal{S}} = [\boldsymbol{I}]_{\mathcal{N},\mathcal{S}} + [\boldsymbol{I}]_{\mathcal{N},\mathcal{S}^C} [\boldsymbol{L}]^{-1}_{\mathcal{S}^C} [\boldsymbol{L}]_{\mathcal{S}^C,\mathcal{S}}\,. \tag{2}$$

D. Problem Setting

We assume the following setup for all results:

- A feature matrix $\boldsymbol{X} \in \mathbb{R}^{N\times d}$, where the columns are jointly Gaussian graph signals that each follow $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$. We assume a smooth signal model (e.g. $\boldsymbol{\Sigma} = \mathbf{\Pi}_{bl(\mathcal{K})}$ [32] or $\boldsymbol{L}^{\dagger}$ [33]);
- Binary labels $\boldsymbol{l} \in \{-1, 1\}^N$ generated by a known function $f : \mathbb{R}^{N\times d} \to \mathbb{R}^N$ (i.e., $\boldsymbol{l} = \text{sign}(f(\boldsymbol{X}))$ with sign elementwise);
- Observations of $\boldsymbol{X}$ at a sample set $\mathcal{S}$ corrupted by white noise of variance η^2 (i.e., $[\boldsymbol{X} + \eta \cdot \boldsymbol{\epsilon}]_{\mathcal{S},\mathcal{D}}$);
- A linear reconstruction $\hat{\boldsymbol{X}}$ of the feature matrix, such as LS or FP, giving $\hat{\boldsymbol{X}} = \boldsymbol{R}_{\mathcal{S}}\,[\boldsymbol{X} + \eta \cdot \boldsymbol{\epsilon}]_{\mathcal{S},\mathcal{D}}$.

We consider the following losses:

$$\text{Classification Loss} = \sum_{i\in\mathcal{V}} \mathbb{P}\left(\text{sign}\,(f(\boldsymbol{X})_i) \neq \text{sign}(f(\hat{\boldsymbol{X}})_i)\right) \tag{3}$$

$$\text{Reconstruction Loss} = \mathbb{E}\left[||\boldsymbol{X} - \hat{\boldsymbol{X}}||^2_F\right]. \tag{4}$$

The classification loss defined above is the sum of per-node misclassification probabilities. We refer to $f(\boldsymbol{X}) \in \mathbb{R}^N$ and $f(\hat{\boldsymbol{X}}) \in \mathbb{R}^N$ as vectorial 'clean output' and 'reconstructed output', respectively, and $f(\boldsymbol{X}) - f(\hat{\boldsymbol{X}})$ as the 'output error'. We let $f(\boldsymbol{X})_i$ denote the component of $f(\boldsymbol{X})$ at node i.

III. Theoretical Results and Implications

A. General Results

We begin with a core lemma:

Lemma 1. Consider X, Y real scalar random variables with $\binom{X}{Y} \sim \mathcal{N}\left(\boldsymbol{0}, \left(\begin{smallmatrix}\sigma^2 & c\\ c & \nu^2\end{smallmatrix}\right)\right)$ with $\sigma, \nu, c \in \mathbb{R}$. Let $\rho = \text{Corr}(X, Y) = \frac{c}{\sigma\nu}$. Then

$$\mathbb{P}\left(\text{sign}(X) \neq \text{sign}(Y)\right) = \frac{\arccos\,\rho}{\pi}. \tag{5}$$

Proof. Let $X' = \frac{X}{\sigma}, Y' = \frac{Y}{\nu}$, then $\binom{X'}{Y'} \sim \mathcal{N}\left(\boldsymbol{0}, \left(\begin{smallmatrix}1 & \rho\\ \rho & 1\end{smallmatrix}\right)\right)$. Note that $\text{sign}(X) \neq \text{sign}(Y) \iff XY < 0 \iff X'Y' < 0$; therefore $\mathbb{P}\left(\text{sign}(X) \neq \text{sign}(Y)\right) = \mathbb{P}\left(X'Y' < 0\right)$. Therefore we only need to prove the lemma for X' and Y', i.e., when $\sigma = \nu = 1$ and $c = \rho$.

Let $r = \frac{\rho}{\sqrt{1-\rho^2}}$ and $Z' = rX' - \sqrt{1+r^2}Y'$. Then $\binom{X'}{Z'} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and $\mathbb{P}\left(X'Y' < 0\right)$ $= \mathbb{P}\left(rX'^2 < X'Z'\right)$ $= \mathbb{P}(\{rX' < Z' \text{ and } X' > 0\} \cup \{rX' > Z' \text{ and } X' < 0\})$. We depict this set as the shaded area in Figure 1.

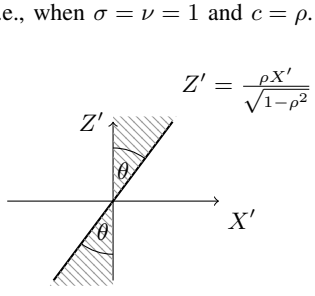


Fig. 1: $\{rX'^2 < X'Z'\}$

Note that $\theta = \arccos(\rho)$. The distribution of $\binom{X'}{Z'}$ is rotationally invariant, so for any positive integer m, if $\frac{m\pi}{\theta} \in \mathbb{Z}$, we can copy and rotate the shaded area $\frac{m\pi}{\theta}$ times to cover the plane m times (one covering of the plane corresponds to $\mathbb{P}\left(X' \in \mathbb{R}, Z' \in \mathbb{R}\right) = 1$). Thus the shaded area is $\mathbb{P}\left(X'Y' < 0\right) = \frac{\theta}{\pi}$. Otherwise let $\theta_m = \frac{m\pi}{\lceil \frac{m\pi}{\theta} \rceil}$ and our argument applies if the shaded area were shrunk to θ_m. Then θ_m tends to θ from below as $m \to \infty$ and by countable additivity of measures, $\mathbb{P}\left(X'Y' < 0\right) = \lim_{m\to\infty} \frac{\theta_m}{\pi} = \frac{\theta}{\pi}$. □

We now characterise our losses for a class of linear f which we interpret as a linearized GCN (see Sec IV-A for a specific example):

Theorem 1. If the function f generating the labels from the features can be written as

$$f(\boldsymbol{X}) = \boldsymbol{G}\boldsymbol{X}\boldsymbol{w}$$

where $\boldsymbol{G} \in \mathbb{R}^{N\times N}$ and $\boldsymbol{w} \in \mathbb{R}^d \backslash \{\boldsymbol{0}\}$, then for a sample set $\mathcal{S}$,

$$\text{Classification Loss} = \sum_{i\in\mathcal{V}} \frac{1}{\pi} \arccos\left(\rho_i(\boldsymbol{G})\right) \tag{6}$$

and if $d = 1$,

$$\text{Reconstruction Loss} = \sum_{i\in\mathcal{V}} \left(\sigma_i(\boldsymbol{I})\right)^2 + \left(\nu_i(\boldsymbol{I})\right)^2 - 2c_i(\boldsymbol{I}) \tag{7}$$

where for any matrix $\boldsymbol{M}$, we let

$$\boldsymbol{C} = \text{Cov}(\boldsymbol{X}\boldsymbol{w}) \cdot (||\boldsymbol{w}||^2_2)^{-1}, \tag{8}$$

$$c_i(\boldsymbol{M}) = (\boldsymbol{M}\boldsymbol{R}_{\mathcal{S}}[\boldsymbol{C}]_{\mathcal{S},\mathcal{N}}\boldsymbol{M}^T)_{ii}, \tag{9}$$

$$\left(\sigma_i(\boldsymbol{M})\right)^2 = (\boldsymbol{M}\boldsymbol{C}\boldsymbol{M}^T)_{ii}, \tag{10}$$

$$\left(\nu_i(\boldsymbol{M})\right)^2 = (\boldsymbol{M}\boldsymbol{R}_{\mathcal{S}}\left[\boldsymbol{C} + \eta^2\boldsymbol{I}_N\right]_{\mathcal{S}}\boldsymbol{R}^T_{\mathcal{S}}\boldsymbol{M}^T)_{ii}, \tag{11}$$

$$\rho_i\left(\boldsymbol{M}\right) = \frac{c_i(\boldsymbol{M})}{\sigma_i(\boldsymbol{M})\nu_i(\boldsymbol{M})}. \tag{12}$$

Proof. As $\boldsymbol{X}$ is independent to the noise and $\mathbb{E}\left[\boldsymbol{X}\right] = 0$, $\text{Cov}(f(\hat{\boldsymbol{X}})_i, f(\boldsymbol{X})_i) = \mathbb{E}\left[f(\hat{\boldsymbol{X}})f(\boldsymbol{X})^T\right]_{ii} = \boldsymbol{G}\boldsymbol{R}_{\mathcal{S}}[\boldsymbol{I}]_{\mathcal{S},\mathcal{N}}\mathbb{E}\left[(\boldsymbol{X} + \eta\cdot\boldsymbol{\epsilon})\boldsymbol{w}\boldsymbol{w}^T\boldsymbol{X}^T\right]\boldsymbol{G}^T$ $= \boldsymbol{G}\boldsymbol{R}_{\mathcal{S}}\left[\text{Cov}(\boldsymbol{X}\boldsymbol{w})\right]_{\mathcal{S},\mathcal{N}}\boldsymbol{G}^T = ||\boldsymbol{w}||^2_2\, c_i(\boldsymbol{G})$. Similarly, $\text{Var}(f(\boldsymbol{X})_i) = ||\boldsymbol{w}||^2_2\left(\sigma_i(\boldsymbol{G})\right)^2$ and $\text{Var}(f(\hat{\boldsymbol{X}})_i) = ||\boldsymbol{w}||^2_2\left(\nu_i(\boldsymbol{G})\right)^2$ giving $\rho_i(\boldsymbol{G}) = \text{Corr}(f(\hat{\boldsymbol{X}})_i, f(\boldsymbol{X})_i)$. For classification loss, as $(f(\hat{\boldsymbol{X}})_i$ and $f(\boldsymbol{X})_i)$ are linear transformations of $\boldsymbol{X}$, they are jointly Gaussian and we can apply Lemma 1 to get that the misclassification probability at node i is $\frac{\arccos(\rho_i(\boldsymbol{G}))}{\pi}$; sum across all nodes to get the total loss. For the Reconstruction loss, $\mathbb{E}\left[||\boldsymbol{X} - \hat{\boldsymbol{X}}||^2_F\right] = \mathbb{E}\left[\text{tr}(\boldsymbol{X}\boldsymbol{X}^T)\right] + \mathbb{E}\left[\text{tr}(\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T)\right] - 2\mathbb{E}\left[\text{tr}(\hat{\boldsymbol{X}}\boldsymbol{X}^T)\right] = \sum_{i\in\mathcal{V}} d\text{Var}(\boldsymbol{X}_i) + d\text{Var}(\hat{\boldsymbol{X}}_i) - 2d\text{Cov}(\hat{\boldsymbol{X}}_i, \boldsymbol{X}_i)$. As $\mathbb{E}\left[\boldsymbol{X}\right] = \mathbb{E}\left[\hat{\boldsymbol{X}}\right] = 0$, if $d = 1$ this expands to (7). □

If we let $\boldsymbol{G}$ be a polynomial p of $\tilde{\boldsymbol{A}}_\gamma$, then $\boldsymbol{l} = \text{sign}(p(\tilde{\boldsymbol{A}}_\gamma)\boldsymbol{X}\boldsymbol{w})$, which corresponds to a multi-layer SGC [34] or a single layer polynomial-based GNN[1] like Chebnet [35], FavardGNN [36], JacobiNet [37] or BernNet [38]. Then the classification loss in Theorem 1 is the difference in output when applying the GNN to the full data $\boldsymbol{X}$ versus to the reconstructed data $\hat{\boldsymbol{X}}$. Note that in the case where the columns of $\boldsymbol{X}$ are independent, $\boldsymbol{C} = \boldsymbol{\Sigma}$, and therefore for SGC the classification loss does not depend on the weights, only on the number of layers.

To interpret Theorem 1, we consider any $\boldsymbol{G}$ and LS reconstruction of a noise-free k-bandlimited signal; specifically $\boldsymbol{\Sigma} = \mathbf{\Pi}_{bl(\mathcal{K})}$. In this case, $\sqrt{c_i(\boldsymbol{G})} = \nu_i(\boldsymbol{G})$ and the classification loss is

[1]We assume that all feature dimensions share the same polynomial filter.

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on April 08,2026 at 07:06:07 UTC from IEEE Xplore. Restrictions apply.

$\sum_{i\in\mathcal{V}} \frac{1}{\pi}\arccos\left(\sqrt{\frac{c_i}{\sigma_i}}\right)$ where $0 \leq \frac{c_i}{\sigma_i} \leq 1$. Thus, we expect the behaviour of classification loss as it changes with sample size to look similar to $x \to \arccos(\sqrt{x})$.

B. Relationship between Reconstruction and Classification Loss

We now relate the reconstruction and classification losses via the error in reconstructing the output of f. All statements in the following Corollary hold for any integer value of $d \geq 1$.

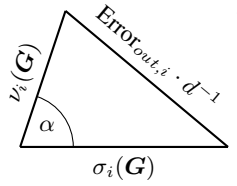
Corollary 1.1. Assume the setting and definitions of Theorem 1. Define the normalized output error at node i as

$$\text{Error}_{out,i} = \left(||\boldsymbol{w}||_2^2\right)^{-1} \mathbb{E}\left[||(f(\boldsymbol{X}))_i - (f(\hat{\boldsymbol{X}}))_i||_2^2\right]. \tag{13}$$

Let $||\boldsymbol{G}||$ be the spectral norm of $\boldsymbol{G}$, then

$$\sum_{i\in\mathcal{V}} \text{Error}_{out,i} \leq ||\boldsymbol{G}||^2 \cdot \text{Reconstruction Loss} \tag{14}$$

and for every node i, the following is always a valid triangle.



α = Node i misclassification probability

Fig. 2: Relationship between misclassification probability (angle) and normalized output error at node i

Proof. We first prove (14). By linearity of f, $(f(\boldsymbol{X}))_i - (f(\hat{\boldsymbol{X}}))_i = (\boldsymbol{G}(\boldsymbol{X} - \hat{\boldsymbol{X}})\boldsymbol{w})_i$. Then $\sum_{i\in\mathcal{V}} ||f(\boldsymbol{X})_i - f(\hat{\boldsymbol{X}})_i||_2^2 = ||\boldsymbol{G}(\boldsymbol{X} - \hat{\boldsymbol{X}})\boldsymbol{w}||_2^2 \leq ||\boldsymbol{G}||^2 \cdot ||(\boldsymbol{X} - \hat{\boldsymbol{X}})\boldsymbol{w}||_2^2 \leq ||\boldsymbol{G}||^2 \cdot (\boldsymbol{X} - \hat{\boldsymbol{X}})||_F^2 \cdot ||\boldsymbol{w}||_2^2$ where the last inequality is because $||\cdot||_2 = ||\cdot||_F$ for vectors and from sub-multiplicativity of $||\cdot||_F$. Taking expectations on both sides gives $||\boldsymbol{w}||_2^2 \cdot \sum_{i\in\mathcal{V}} \text{Error}_{out,i} \leq ||\boldsymbol{G}||^2 \cdot \text{Reconstruction Loss} \cdot ||\boldsymbol{w}||_2^2$.

Secondly, we prove Fig. 2 forms a valid triangle. Through the same expansion and independence arguments as in the proof of Theorem 1, $\text{Error}_{out,i} = d \cdot \left((\sigma_i(\boldsymbol{G}))^2 + (\nu_i(\boldsymbol{G}))^2 - 2\rho_i(\boldsymbol{G})\right)$. As $\rho_i(\boldsymbol{G})$ is the probability of misclassification of node i (see the proof of Theorem 1), the sides and angle of Fig. 2 obey the law of cosines and thus form a valid triangle. □

For SGCs, $||\boldsymbol{G}|| \leq 1$ (as $||\tilde{\boldsymbol{A}}|| \leq 1$ by Gershgorin's Disc Theorem [39], so $\forall\gamma \geq 0 : ||\tilde{\boldsymbol{A}}_\gamma|| \leq 1$ by [34, Theorem 1] and therefore $||\tilde{\boldsymbol{A}}_\gamma^r|| \leq 1$ for all $r \geq 1$). Therefore, by (14), the sum of the per-node normalized output errors (which are directly related to the per-node misclassification probabilities, as shown in Fig. 2) is upper bounded by the reconstruction loss.

While inequality (14) suggests optimising reconstruction loss to minimise $\sum_{i\in\mathcal{V}} \text{Error}_{out,i}$, the non-linear relationship between $\text{Error}_{out,i}$ and the classification loss at node i make this sum a poor proxy for total classification error. We discuss experimental results around optimising reconstruction loss as a proxy for classification loss in Section IV-C.

C. Sampling for Classification

We use Theorem 1 to design a sampling scheme specifically for classification, minimising mean classification loss rather than the common mean reconstruction loss objective [6]–[8], [32]. We propose greedily choosing $\mathcal{S}$ to optimize (6) from Theorem 1.

IV. Empirical Results

A. Experimental Setup

1) Synthetic Graph Generation: We generate 32 graphs with $N = 500$ nodes from each of the following models:

- Barabási-Albert (BA) with a preferential attachment to 3 vertices at each step of construction;
- Stochastic Blockmodel (SBM) with intra- and inter- cluster edge probabilities of 0.7 and 0.1, respectively.

2) Synthetic Signal Generation: We draw features $\boldsymbol{X}$ as described in Section II-D with $\boldsymbol{\Sigma} = \boldsymbol{\Pi}_{bl(\mathcal{K})}$, i.i.d. columns, bandwidth $k = \frac{N}{10}$ and noise variance $\eta^2 = 10^{-3}$ so the SNR is $20dB$, like in [13].

3) Choice of f: For each graph, we construct f as a linearized GCN with randomly initialized weights and added self-loops ($\gamma = 1$). We choose our architecture to mimic the strong baselines in [40], but keep to a single convolutional layer to minimise oversmoothing. Specifically, we set

$$f(\boldsymbol{X}) = \tilde{\boldsymbol{A}}_1 \boldsymbol{X}\,(\boldsymbol{W}_1\boldsymbol{W}_2) \tag{15}$$

where $\boldsymbol{W}_1 \in \mathbb{R}^{64\times 32}$, $\boldsymbol{W}_2 \in \mathbb{R}^{32\times 1}$ are both populated via Glorot initialization [41].

4) Real-world Datasets: We consider an FMRI dataset with nodes corresponding to brain regions and 292 graph signals corresponding to blood oxygen levels [32], [42], subsampling a connected graph with $N = 367$ nodes using neighbourhood sampling [43]. We do not bandlimit the given signals. We construct a classification task by subtracting the mean from all signals and taking the sign as node class labels, representing high or low blood oxygen levels.

5) Sample Set Selection: We consider three sampling methods:

- Random sampling;
- Greedy sampling optimizing classification loss (6), which is described in Sec III-C;
- Greedy sampling optimizing reconstruction loss (7). For LS, this is the same as optimizing for the frequently used A-optimal objective for reconstruction [6]–[8], [32].

To sample on real world data, we need to compute losses for our sampling to optimize. This requires making assumptions about the noise level and distribution of the features. For the FMRI dataset, we assume the features are i.i.d. $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Pi}_{bl(\mathcal{K})})$ with $k = \lfloor\frac{N}{10}\rfloor$ and that the SNR is $20dB$ (i.e., $\eta^2 = 10^{-3}$).

B. Experimental Validation

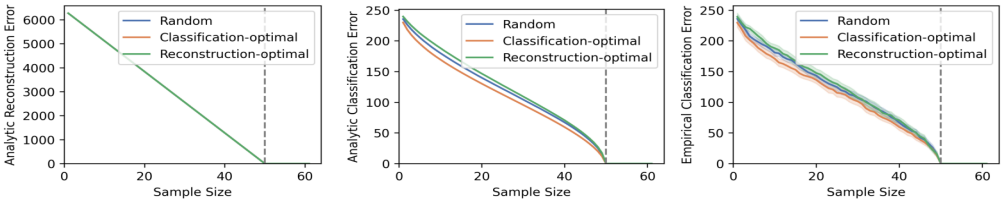
We experimentally validate our derivation of expected classification loss for different graph models, noise levels, reconstruction methods and versions of f. Subfigures (b) and (c) of Figs. 3-6 display the empirical loss calculations via directly reconstructing signals, and the analytic loss calculations using Theorem 1 respectively. We see that the empirical results largely agree with our theoretical derivations.

We also see that under noiseless LS reconstruction (Figs. 3 and 6), our classification loss looks like $x \to \arccos(\sqrt{x})$, which follows the theoretical interpretation in III-A.

C. Optimal Sampling

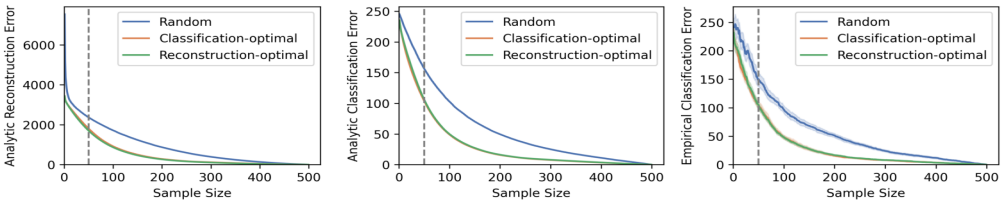
In the noiseless setting, under LS reconstruction, our method outperforms both random and A-optimal sampling (Fig. 3b) by a small but consistent margin, with no difference in reconstruction error (Fig. 3a). This margin is even larger in the SBM case (Fig. 6b). Surprisingly, A-optimal sampling underperforms random sampling; optimizing for reconstruction loss is worse than random sampling under classification loss. This emphasises how sampling schemes need to be designed in light of the downstream task.

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on April 08,2026 at 07:06:07 UTC from IEEE Xplore. Restrictions apply.



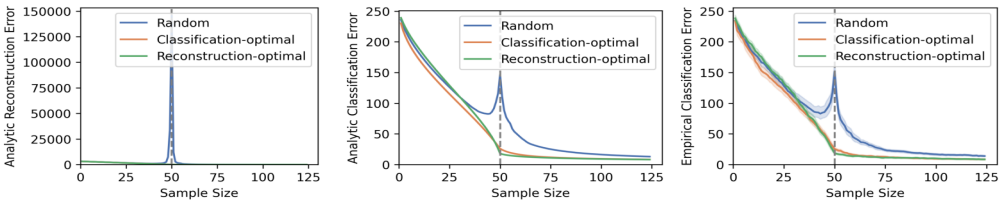
(a) Reconstruction Loss (b) Class. Loss (Analytic) (c) Class. Loss (Empirical)

Fig. 3: LS Reconstruction, BA graph model, noiseless



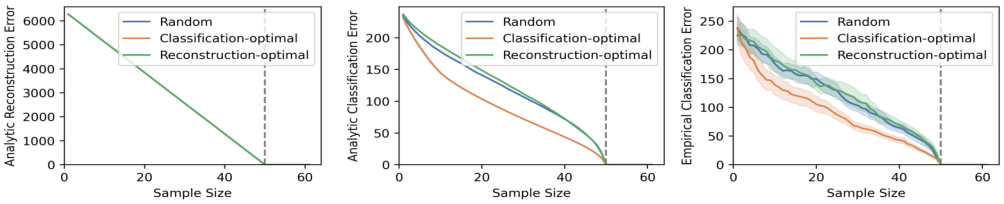
(a) Reconstruction Loss (b) Class. Loss (Analytic) (c) Class. Loss (Empirical)

Fig. 4: FP Reconstruction, BA graph model, noiseless



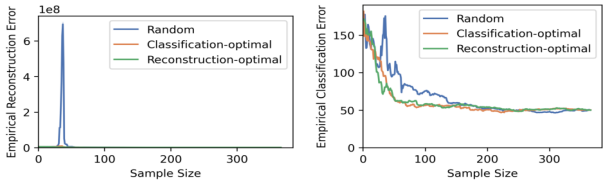
(a) Reconstruction Loss (b) Class. Loss (Analytic) (c) Class. Loss (Empirical)

Fig. 5: LS Reconstruction, BA graph model, Noisy ($20dB$ signal)



(a) Reconstruction Loss (b) Class. Loss (Analytic) (c) Class. Loss (Empirical)

Fig. 6: LS Reconstruction, SBM graph model, noiseless



(a) Rec. Loss (Empirical) (b) Class. Loss (Empirical)

Fig. 7: FMRI dataset, LS reconstruction, sampled assuming $k = 36$ and SNR $= 20dB$

In the FP reconstruction case (Fig. 4), we see that both forms of optimal sampling are approximately equal, with both beating random sampling significantly. We do see that reconstruction-optimal sampling marginally beats classification sampling under reconstruction loss, and vice-versa for classification loss.

In the noisy case (Fig. 5), we see classification-optimal sampling has lower classification loss except for sample sizes near the bandwidth (marked with the vertical dashed line); near the bandwidth, the impact of noise is the highest for LS [32] and thus reconstruction-optimal sampling, which optimizes for noise sensitivity more aggressively, has both lower reconstruction and classification loss. This is an artefact of the greedy construction of our sampling sets. We refer to [32] for analysis of the spike in reconstruction error in Fig. 5a.

In our real-world dataset (Fig. 7) we find, for classification loss, both reconstruction and classification-optimal sampling outperforming random sampling, with our novel classification-optimal sampling scheme marginally outperforming at sample sizes over 90.

V. Conclusion

In this paper we have studied how classification differs from reconstruction in the presence of noise and missing data. We have given theoretical and empirical characterisation of the relationship between classification and reconstruction error, where classification is done by a linearized GCN. We have shown that reconstruction-optimal sampling can underperform random sampling, which emphasises the point that sampling schemes need to be designed with the specific task in mind. We have then presented a novel optimal sampling scheme specifically for classification with a GCN, and showed its performance across multiple graph models, reconstruction methods and noise levels.

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on April 08,2026 at 07:06:07 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] S. Itani and D. Thanou, "A graph signal processing framework for the classification of temporal brain data," in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1180–1184.

[2] R. K. Jain, J. M. Moura, and C. E. Kontokosta, "Big data+ big cities: Graph signals of urban air pollution [exploratory sp]," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 130–136, 2014.

[3] B. Renoust, G. Cheung, and S. Satoh, "Estimating political leanings from mass media via graph-signal restoration with negative edges," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2017, pp. 1009–1014.

[4] D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork *et al.*, "String v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," *Nucleic acids research*, vol. 47, no. D1, pp. D607–D613, 2019.

[5] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[6] F. Wang, Y. Wang, and G. Cheung, "A-optimal sampling and robust reconstruction for graph signals via truncated neumann series," *IEEE Signal Processing Letters*, vol. 25, no. 5, pp. 680–684, 2018.

[7] F. Wang, G. Cheung, and Y. Wang, "Low-complexity graph sampling with noise and signal reconstruction via neumann series," *IEEE Transactions on Signal Processing*, vol. 67, no. 21, pp. 5511–5526, 2019.

[8] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4845–4860, 2016.

[9] N. Tremblay, P.-O. Amblard, and S. Barthelmé, "Graph sampling with determinantal processes," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 1674–1678.

[10] A. Jayawant and A. Ortega, "Practical graph signal sampling with log-linear size scaling," *Signal Processing*, vol. 194, p. 108436, 2022.

[11] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *Applied and Computational Harmonic Analysis*, vol. 44, no. 2, pp. 446–475, 2018.

[12] D. Pakiyarajah, E. Pavez, and A. Ortega, "Graph-based signal sampling with adaptive subspace reconstruction for spatially-irregular sensor data," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.

[13] Y. Bai, F. Wang, G. Cheung, Y. Nakatsukasa, and W. Gao, "Fast graph sampling set selection using gershgorin disc alignment," *IEEE Transactions on signal processing*, vol. 68, pp. 2419–2434, 2020.

[14] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE transactions on signal processing*, vol. 63, no. 24, pp. 6510–6523, 2015.

[15] F. Pukelsheim, *Optimal design of experiments*. SIAM, 2006.

[16] X. Zhang and Q. Wang, "Graph learning from incomplete graph signals: From batch to online methods," *Signal Processing*, vol. 226, p. 109663, 2025.

[17] P. Berger, G. Hannak, and G. Matz, "Efficient graph learning from noisy and incomplete data," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6, pp. 105–119, 2020.

[18] S. N. Sridhara, E. Pavez, and A. Ortega, "Towards joint graph learning and sampling set selection from data," in *2024 58th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2024, pp. 1168–1172.

[19] I. Pesenson, "Sampling in paley-wiener spaces on combinatorial graphs," *Transactions of the American Mathematical Society*, vol. 360, no. 10, pp. 5603–5627, 2008.

[20] L. F. Chamon and A. Ribeiro, "Near-optimality of greedy set selection in the sampling of graph signals," in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2016, pp. 1265–1269.

[21] ——, "Greedy sampling of graph signals," *IEEE Transactions on Signal Processing*, vol. 66, no. 1, pp. 34–47, 2017.

[22] H. Shomorony and A. S. Avestimehr, "Sampling large data on graphs," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2014, pp. 933–936.

[23] S. Chen, R. Varma, A. Singh, and J. Kovačević, "Signal recovery on graphs: Fundamental limits of sampling strategies," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 4, pp. 539–554, 2016.

[24] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," *Advances in neural information processing systems*, vol. 16, 2003.

[25] B. Jiang and Z. Zhang, "Incomplete graph representation and learning via partial graph neural networks," *arXiv preprint arXiv:2003.10130*, 2020.

[26] X. Chen, S. Chen, J. Yao, H. Zheng, Y. Zhang, and I. W. Tsang, "Learning on attribute-missing graphs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 2, pp. 740–757, 2020.

[27] V. Kalofolias, X. Bresson, M. Bronstein, and P. Vandergheynst, "Matrix completion on graphs," *arXiv preprint arXiv:1408.1717*, 2014.

[28] I. Spinelli, S. Scardapane, and A. Uncini, "Missing data imputation with adversarially-trained graph convolutional networks," *Neural Networks*, vol. 129, pp. 249–260, 2020.

[29] E. Rossi, H. Kenlay, M. I. Gorinova, B. P. Chamberlain, X. Dong, and M. Bronstein, "On the unreasonable effectiveness of feature propagation in learning on graphs with missing node features," *arXiv preprint arXiv:2111.12128*, 2021.

[30] H. Taguchi, X. Liu, and T. Murata, "Graph convolutional networks for graphs containing missing features," *Future Generation Computer Systems*, vol. 117, pp. 155–168, 2021.

[31] J. You, X. Ma, Y. Ding, M. J. Kochenderfer, and J. Leskovec, "Handling missing data with graph representation learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 19 075–19 087, 2020.

[32] B. Sripathmanathan, X. Dong, and M. Bronstein, "On the impact of sample size in reconstructing noisy graph signals: A theoretical characterisation," *arXiv preprint arXiv:2406.16816*, 2024.

[33] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6160–6173, 2016.

[34] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger, "Simplifying graph convolutional networks," in *International conference on machine learning*. Pmlr, 2019, pp. 6861–6871.

[35] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," *Advances in neural information processing systems*, vol. 29, 2016.

[36] Y. Guo and Z. Wei, "Graph neural networks with learnable and optimal polynomial bases," in *Proceedings of the 40th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, Eds., vol. 202. PMLR, 23–29 Jul 2023, pp. 12 077–12 097. [Online]. Available: https://proceedings.mlr.press/v202/guo23i.html

[37] J. Guo, L. Du, X. Chen, X. Ma, Q. Fu, S. Han, D. Zhang, and Y. Zhang, "On manipulating signals of user-item graph: A jacobi polynomial-based graph collaborative filtering," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 602–613.

[38] M. He, Z. Wei, H. Xu *et al.*, "Bernnet: Learning arbitrary graph spectral filters via bernstein approximation," *Advances in Neural Information Processing Systems*, vol. 34, pp. 14 239–14 251, 2021.

[39] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[40] Y. Luo, L. Shi, and X.-M. Wu, "Classic GNNs are strong baselines: Reassessing GNNs for node classification," in *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2024. [Online]. Available: https://openreview.net/forum?id=xkljKdGe4E

[41] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.

[42] Y.-C. Zhi, Y. C. Ng, and X. Dong, "Gaussian processes on graphs via spectral kernel learning," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 9, pp. 304–314, 2023.

[43] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.

Authorized licensed use limited to: Bodleian Libraries of the University of Oxford. Downloaded on April 08,2026 at 07:06:07 UTC from IEEE Xplore. Restrictions apply.