Graph-Based Approaches to Learning Epileptogenic Zone Localization Using Stereo-EEG Recordings

Daniel Wendelken
Dept. of Computer Science
University of Cincinnati
Cincinnati, USA
wendeldr@mail.uc.edu

Brian Ervin
Dept. of Neurology
Cincinnati Children's Hospital Medical Center
Cincinnati, USA
Brian.Ervin@cchmc.org

Ravindra Arya
Dept. of Neurology
Cincinnati Children's Hospital Medical Center
Cincinnati, USA
Ravindra.Arya@cchmc.org

Ali A. Minai
Dept. of Electrical and Computer Engineering
University of Cincinnati
Cincinnati, USA
minaiaa@ucmail.uc.edu

***Abstract*—The epileptogenic zone (EZ) is the brain region that generates seizures in an individual, and is the target of epilepsy surgery. Localizing the EZ from stereo-EEG (sEEG) recordings supports surgical planning, but manual interpretation is time-consuming and focuses on seizure recordings. Graphical learning models of resting-state functional connectivity among the recorded brain regions are an attractive alternative, but depend crucially on the network topology chosen for the model.**

We present a controlled study of graph-based models to explore how graph topology affects EZ localization from resting-state sEEG in 40 patients. Using the same simple learnable model and leave-one-patient-out evaluation, we compare dense graphs, anatomy- and geometry-informed priors, budgeted sparsification methods, and learned sparsification, including the proposed Region-Bridge-c topology. To compare graph constructions fairly, we control the number of incoming edges per node and vary graph sparsity.

At $\approx 30\%$ edge retention, Region-Bridge-c achieves the highest observed mean PR-AUC (0.371 ± 0.015; ROC-AUC 0.743 ± 0.010) while using $\approx 69\%$ fewer edges than Dense (PR-AUC 0.349 ± 0.014). Spatial-k is competitive, whereas random pruning requires near-dense retention. Learned sparsification benefits from anatomical node metadata but, on average, does not surpass the best fixed prior. Across all topologies, the best choice varies by patient. These results suggest that graph construction should be evaluated explicitly rather than treated as fixed preprocessing.

I. Introduction

Epilepsy affects about 50 million people worldwide [1]. Approximately 30% of patients have drug-resistant epilepsy [2], and for surgically eligible cases, resective or ablative intervention can substantially improve seizure outcomes [3]. Stereo-electroencephalography (sEEG) is a key technique for localizing the epileptogenic zone (EZ) [4]. sEEG involves implanting multiple needle-like multi-contact depth electrodes into targeted brain regions and recording from them. Clinical interpretation of sEEG recordings is labor-intensive and focuses on seizure (ictal) recordings. Limited seizure capture and time-consuming review have driven growing interest in analyzing resting-state between-seizure (interictal) signals.

One practical approach follows from the view that epilepsy is primarily a network disorder [5]. Abnormal functional couplings underlying seizures can persist interictally and be captured by connectivity measures between implanted contacts. Accordingly, functional connectivity graphs provide a natural representation for learning-based EZ localization. Graph neural networks (GNNs) can learn non-linear representations over such graphs for node-level classification [6], [7].

Dense functional connectivity produces fully connected graphs that mix informative pathways with weak or irrelevant interactions; dense aggregation can amplify noise and cause oversmoothing [8]. Sparsifying via thresholding, k-NN, or anatomical masks reduces computation but introduces priors that substantially alter graph statistics and downstream results [9], motivating controlled comparisons under matched per-node edge budgets.

Using resting-state sEEG from 40 patients, we benchmark dense functional graphs, spatial and anatomy-informed priors, and a differentiable learned-topology module that can be projected to a hard per-node top-k graph at inference, sweeping per-node incoming-edge budgets to characterize when sparsification improves over dense connectivity.

Our primary contribution is a controlled evaluation framework for matched-budget topology comparisons; within this framework, Region-Bridge-c, a simple anatomy+geometry prior, achieves the highest observed mean performance while retaining $\approx 30\%$ of incoming edges. **Contributions:**

- We benchmark graph topologies for resting-state sEEG EZ node classification under matched per-node incoming-edge budgets across 40 leave-one-patient-out (LOPO) folds and five random seeds.
- We propose and evaluate Region-Bridge-c, a sparse and interpretable structural prior that improves on average over dense graphs at substantially lower edge retention.
- We test an entmax-based learned sparsifier, projected to a hard per-node top-k graph at inference, and analyze its sensitivity to realized retention and patient-level heterogeneity.

Code and data cannot be shared as both contain patient-specific information subject to confidentiality.

II. Related Work

A. Interictal Network Features for Localization

Pairwise connectivity features, such as phase-locking value and coherence, can be used to build functional networks

that characterize interactions relevant to seizure initiation and propagation [10]. Because pairwise analyses scale rapidly with implant size and the space of interaction measures is large, automated learning-based approaches are increasingly used to summarize and interpret these networks. Yet network statistics and their localizing relevance are highly sensitive to how connectivity is estimated and how graphs are constructed; for example, weighted versus binary representations and sparsification settings substantially change graph properties, complicating cross-patient comparisons [11], [9].

B. Learning Over Connectivity Graphs

Given this sensitivity, prior learning pipelines can be viewed as jointly choosing a graph construction procedure and then a predictive model on top of the resulting network. Classical approaches often rely on handcrafted network summaries with conventional machine learning [12], [13], [14]. More recent work explores learned representations over connectivity graphs, including GNN-based models in related settings such as resting-state fMRI graphs and dynamic functional connectivity [15]. Recent work has applied GNNs to sEEG contact graphs for EZ/seizure onset zone (SOZ) localization using fixed graph constructions, including spatial-distance graphs over interictal recordings for surgical planning [16] and correlation-thresholded graphs over ictal recordings for SOZ localization [17]; in both cases the topology is not treated as an experimental variable. Across these directions, reported performance and interpretability remain tightly coupled to the underlying graph construction and sparsification strategy, motivating controlled evaluations of topology choices and methods that can adapt or learn sparse graph structure [18], [11].

C. Graph Structure Learning

Graph neural networks typically treat the graph topology as fixed and perform message aggregation over the resulting edge set [6], [7]. In epilepsy network analysis, graph construction is not merely a preprocessing step, and choices such as connectivity estimator, thresholding strategy, and use of spatial or anatomical constraints can meaningfully affect inferred network organization and subsequent prediction performance [19].

Graph structure learning (GSL) instead seeks to learn edge weights or discrete connectivity patterns jointly with the predictive model [20]. Since discrete edge selection is not differentiable, many approaches learn continuous edge weights and then impose sparsity through a projection step. In this work, we use the α-entmax transformation [21], a sparse alternative to softmax, to produce sparse, nonnegative edge weights and apply a hard top-k projection at inference to satisfy an explicit per-node budget for comparison.

III. Methods

A. Problem Formulation

We treat the identification of the EZ as a binary classification problem over a graph of sEEG contacts. For patient p, we define the node set $V_p = \{1, \ldots, n_p\}$, where n_p is the total count of contacts/channels. The patient is represented by the graph $G_p = (V_p, E_p^{\text{full}})$, containing all pairwise connections $E_p^{\text{full}} = \{(u, v) : u, v \in V_p,\ u \neq v\}$. Each edge can be assigned F scalar features representing various statistical relations between the nodes it connects.

Let $y_v \in \{0, 1\}$ be the label for node v that indicates EZ membership. We define $\mathbf{X}_p \in \mathbb{R}^{n_p \times n_p \times F}$ as the tensor of edge features, where the entry $\mathbf{x}_{uv}$ corresponds to the directed edge from u to v. Different graph topologies are realized by applying a mask or weight matrix to the edge set E_p^{full}.

B. Data, Labels, and Electrode Localization

Data acquisition, electrode localization, and EZ labeling follow our prior study [22]. For the present work, we analyzed a cohort of 40 patients, including 21 males and 19 females, with favorable post-surgical outcomes (International League Against Epilepsy [ILAE] scores 1–3) and at least one EZ-labeled contact. Patient age was 13.3 ± 5.6 years (range: 2.0–21.2). Each patient contributed approximately 5 minutes of awake resting-state sEEG (mean 5.04 ± 0.02 min) sampled at 2,048 Hz. Patients had 132.1 ± 31.4 contacts (range: 75–197), of which 15.9 ± 11.2 (range: 3–59) were clinically labeled as EZ. Across the cohort, this yielded 5,282 total contacts with 635 EZ contacts (12.0%) and 4,647 non-EZ contacts (88.0%), resulting in a class imbalance ratio of 7.3:1. Contacts were labeled as EZ vs. non-EZ based on standard clinical practice using visual analysis of ictal sEEG. Electrode localization was performed by co-registering post-implant CT to pre-implant MRI, transforming to the Montreal Neurological Institute (MNI) standard space, and assigning each contact to anatomical parcels and derived neuroanatomical groups using FASCILE software [23]. First, a fine-grained atlas label (called **MICCAI**) [24] serves as categorical node metadata for embedding lookup. Second, we use a coarse region label for topology priors. Starting from nine neuroanatomic groups defined in [22], we add an `unknown` category and split each group by hemisphere. This produces a 20-class label space that we refer to as **Region20**. Region20 is used for topology construction in Intra-Region and Region-Bridge-c, and in one learned variant it is also used as node metadata for edge scoring.

C. Graph Preprocessing and Construction

a) Graph Representation: For each patient p, we constructed a directed graph from precomputed functional connectivity tensors stored in patient feature files. Connectivity was computed after artifact rejection, notch filtering at 60 Hz and harmonics, common average re-referencing, and segmentation into non-overlapping 0.5 s windows ($W \approx 600$ per patient). For each window, we used PySPI v1.1.1 to compute a comprehensive suite of pairwise connectivity measures [25], yielding a windowed edge tensor of shape (n_p, n_p, W, Q), where $Q = 185$ is the number of potential connectivity features per electrode pair. We then averaged across windows to obtain a graph of shape (n_p, n_p, Q). Each feature was

independently normalized within each patient using a robust z-score (median-centered, median absolute deviation [MAD]-scaled), with the median and MAD computed separately for each patient-feature pair; thus, no normalization statistics were shared across patients or LOPO folds.

b) Feature Selection: To reduce redundancy across connectivity measures, we performed unsupervised edge-feature selection prior to training. Using the normalized graph features, we removed features with near-zero variance and features with substantial NaN counts. For the remaining features, we computed absolute Spearman correlations and performed complete-linkage hierarchical clustering using the distance $\delta = 1 - |\rho|$, where ρ denotes the Spearman correlation. We cut the dendrogram at $\delta \leq 0.25$ to form clusters of highly correlated features, and selected one representative per cluster. This representative was the feature with the largest sum of intra-cluster correlations, producing the final $F = 63$ features used in all experiments. This selection used only edge-feature correlations (no EZ labels) and was performed once before LOPO splitting. A leave-one-out check confirmed high stability (22/40 folds identical; Jaccard mean 0.98, min 0.88). The retained features span phase, coherence, correlation, distance, covariance/precision, and directed/information-theoretic measures. We converted each (n_p, n_p, F) tensor to an edge list by extracting all non-diagonal entries.

D. Graph Topologies

We evaluated several topologies, illustrated schematically in Fig. 1.

*1) **Dense**:* The dense graph configuration served as a baseline, retaining all pairwise connections except self-loops, yielding a complete directed graph.

2) Structural Priors: These topologies use fixed anatomical or physical constraints:

- **Intra-Lead:** Preserves connections between contacts on the same electrode shaft, removing inter-electrode edges.
- **Lead-Chain:** Restricts connectivity to adjacent contacts on the same electrode shaft.
- **Intra-Region:** Retains edges only between nodes sharing the same **Region20** label.

3) Budget Controlled Topologies: We employ budget-controlled topologies to sparsify the graph by explicitly limiting the number of incoming edges per node. For a patient p with n_p contacts, each destination node v retains $k = \lceil r \cdot (n_p - 1) \rceil$ incoming neighbors, where $r \in (0, 1]$ is the target retention ratio (fraction of possible incoming edges retained per node).

- **Random-k:** For each destination node, we sample k incoming edges uniformly at random to serve as an uninformed baseline.
- **Spatial-k:** This topology connects each node to its k nearest spatial neighbors, determined by Euclidean distance in 3D MNI space.
- **Region-Bridge-c:** For each node v, we retain all incoming edges from nodes in the same coarse anatomical region. We then add $b_v = \lceil c\,(n_p - 1) \rceil$ additional incoming edges from nodes in different regions, chosen as the nearest contacts in MNI space. This hybrid topology preserves dense intra-region coupling while allowing a limited number of cross-region "bridges." We set c so that the realized retention ratio is close to the target retention ratio r; see Sec. III-D.
- **Learned-edges-k:** This approach utilizes a scoring network to predict edge importance from $\mathbf{x}_{uv}$ and applies a per-node hard top-k projection during inference; see Sec. III-F.

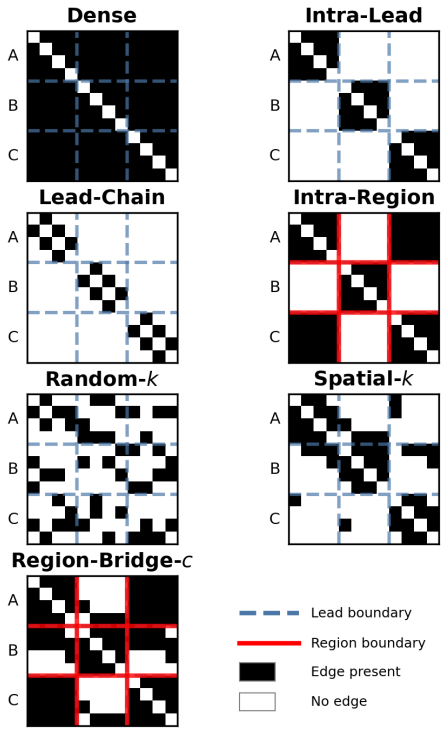

Fig. 1. Adjacency-matrix schematics of evaluated topologies on a simulated 12-node sEEG graph (3 leads, 2 regions). Black entries indicate retained edges.

E. Backbone Model

For each patient, the input is a directed, fully connected graph without self-loops (nodes: sEEG contacts; edges: feature vectors). The nodes and edges of the graph are processed for embedding via two distinct streams:

Node Stream: Each node is assigned an embedding vector based on its atlas label. Each atlas label is represented by a learned vector (one vector per label, shared across contacts).

Edge Stream: The normalized F-dimensional feature vector for each edge is passed through a multilayer perceptron (MLP) network to generate an edge embedding with a lower dimension.

1) Node Label Embedding: Each node v is represented by a learnable atlas embedding $\mathbf{e}_v \in \mathbb{R}^{d_{\text{emb}}}$ (with $d_{\text{emb}} = 16$), obtained by an embedding lookup indexed by the node's standardized atlas label.

a) Atlas Vocabulary Standardization:: MICCAI labels were standardized by merging hemispheric variants to reduce label fragmentation, then frequency filtered. Labels in at least four patients receive unique IDs, and the rest map to `unknown`, producing 39 atlas IDs in this cohort that index a learnable embedding table trained end-to-end with the GNN. In the learned-topology variant, Region20 labels map without additional standardization to a separate embedding table (Sec. III-B).

2) Edge Feature Embedding: For each directed edge (u, v) with edge feature vector $\mathbf{x}_{uv} \in \mathbb{R}^F$, we compute a node-level edge aggregation by first using an edge-level MLP $\phi_{\text{edge}} : \mathbb{R}^F \to \mathbb{R}^H$:

$$\mathbf{m}_{uv} \;=\; \phi_{\text{edge}}(\mathbf{x}_{uv}) \in \mathbb{R}^H \tag{1}$$

with hidden dimension $H = 32$ and learnable weights. Incoming projected edges for each node v are aggregated via a weighted mean using nonnegative edge weights w_{uv} defined by the topology. The weighted mean ensures aggregation is invariant to the number of contacts per patient. The resulting *edge feature vector* h_v is:

$$\mathbf{h}_v \;=\; \frac{\sum_{u \neq v} w_{uv}\, \mathbf{m}_{uv}}{\sum_{u \neq v} w_{uv}} \tag{2}$$

For all topologies, other than the learned sparsifier, $w_{uv} = 1$ for retained edges and 0 otherwise. For learned topologies, w_{uv} is defined as described in Sec. III-F.

The final embedding for each node v is obtained by concatenating its aggregated edge feature vector with the atlas label embedding, followed by layer normalization:

$$\mathbf{z}_v \;=\; \mathrm{LN}\big([\mathbf{h}_v; \mathbf{e}_v]\big) \tag{3}$$

A *classifier MLP* ϕ_{cls} with learnable weights uses this to generate 2-class logits for the node:

$$\boldsymbol{\ell}_v \;=\; \phi_{\text{cls}}(\mathbf{z}_v) \in \mathbb{R}^2 \tag{4}$$

All topology variants use the same two-stream backbone network; only edge masking and weights change. For fixed topology cases, there are three sets of learnable parameters: 1) The weights of the atlas label embeddings; 2) The weights of the edge embedding MLP; and 3) The weights of the classifier MLP generating logits. All parameters are trained end-to-end using the node classification loss (Eq. (12)).

F. Learned Topology Module

For the learned topology, we compute a scalar score for each directed edge (u, v) from its edge feature vector $\mathbf{x}_{uv}$ using an MLP ϕ_{score} with trainable weights:

$$s_{uv} \;=\; \phi_{\text{score}}(\mathbf{x}_{uv}) \in \mathbb{R} \tag{5}$$

As the graph starts fully connected, for a destination node v, its in-degree is $d_v = n_p - 1$. We collect the incoming scores into a vector

$$\mathbf{q}_v \;=\; [\, s_{u_1 v}, \ldots, s_{u_{d_v} v} \,] \in \mathbb{R}^{d_v}. \tag{6}$$

where $u_1, \ldots, u_{d_v}$ are the source nodes $u \neq v$ (in any fixed order). We convert scores to sparse incoming weights per destination node using the α-entmax transform with $\alpha = 1.5$ [21]:

$$\boldsymbol{w}_v^{\text{soft}} \;=\; \mathrm{entmax}_{1.5}(\mathbf{q}_v), \qquad \sum_{u \neq v} w_{uv}^{\text{soft}} = 1, \;\; w_{uv}^{\text{soft}} \geq 0. \tag{7}$$

Here w_{uv}^{soft} denotes the entry of $\boldsymbol{w}_v^{\text{soft}}$ for source node u.

During inference, we enforce the exact per-node budget by keeping the top-k_v scores for each destination node v, where $k_v = \lceil r \cdot d_v \rceil$ and r is the target retention ratio. Let $\mathcal{K}_v$ denote the top-k_v source nodes for v according to the scores in $\mathbf{q}_v$. We first mask the soft weights,

$$\tilde{w}_{uv} \;=\; \begin{cases} w_{uv}^{\text{soft}} & \text{if } u \in \mathcal{K}_v \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

and then renormalize,

$$w_{uv}^{\text{hard}} \;=\; \frac{\tilde{w}_{uv}}{\sum_{u' \neq v} \tilde{w}_{u'v}}. \tag{9}$$

We report the realized edge retention ratio as the fraction of edges with $w_{uv}^{\text{hard}} > 0$ relative to $n_p(n_p - 1)$.

a) Auxiliary Budget Compliance Loss: To encourage budget compliance in the learned model, we regularize the effective in-degree induced by the soft incoming weights,

$$d_{\text{eff}}(v) \;=\; \frac{1}{\sum_{u \neq v} \big(w_{uv}^{\text{soft}}\big)^2}, \qquad k_v = \lceil r \cdot d_v \rceil \tag{10}$$

via a log-space matching loss

$$\mathcal{L}_{\text{aux}} = \frac{1}{N_{\text{train}}} \sum_{p \in \mathcal{P}_{\text{train}}} \sum_{v \in V_p} \big(\log d_{\text{eff}}(v) - \log k_v\big)^2 \tag{11}$$

During training we use $w_{uv} = w_{uv}^{\text{soft}}$; during evaluation we use $w_{uv} = w_{uv}^{\text{hard}}$.

b) Learned Topology Variants: We evaluate an edge-only scorer (**Learned-edges-k**) and node-aware variants that include source/destination embeddings. We additionally test atlas-augmented variants that concatenate atlas embeddings to the scorer input (**Learned-MICCAI-k**, **-detached-k**, **-separate-k**) and a coarse Region20 variant (**Learned-Region20-k**). These variants performed similarly; only the best (Learned-Region20-k) is reported individually. Unless stated otherwise, learned topology results use the hard top-k projection at inference time.

G. Experimental Protocol

a) LOPO Cross Validation: We evaluate model performance using leave-one-patient-out cross-validation. In each fold, one patient is held out for testing, and the remaining patients are split at the patient level into train and validation sets with an 80%:20% split and a seeded shuffle. Each fold is repeated five times with random seeds. For topologies that depend on randomness, such as Random-k, we fix a separate data seed so that the structural masks are identical across training seeds.

b) Optimization and Objectives: We train with AdamW for up to 250 epochs and apply early stopping with a patience of 30 based on the validation loss. The main objective is the focal loss [26] over node labels, which was chosen to address class imbalance. For each node v, let $\boldsymbol{\ell}_v \in \mathbb{R}^2$ denote the logits and $\boldsymbol{\pi}_v = \mathrm{softmax}(\boldsymbol{\ell}_v)$ the class probabilities. Let $y_v \in \{0, 1\}$ be the ground-truth label and π_{v,y_v} the predicted probability of the true class.

Let $\mathcal{P}_{\mathrm{train}}$ denote the set of training patients in a fold, and N_{train} the total number of training nodes in the set. The focal loss is

$$\mathcal{L}_{\mathrm{focal}} \;=\; -\frac{1}{N_{\mathrm{train}}} \sum_{p \in \mathcal{P}_{\mathrm{train}}} \sum_{v \in V_p} \left(1 - \pi_{v,y_v}\right)^{\gamma} \log \pi_{v,y_v}, \quad (12)$$

with $\gamma = 1.0$ for the reported topology comparisons. For the learned topologies, we minimize

$$\mathcal{L} \;=\; \mathcal{L}_{\mathrm{focal}} + \lambda_{\mathrm{aux}} \mathcal{L}_{\mathrm{aux}}, \qquad \lambda_{\mathrm{aux}} = 0.02. \quad (13)$$

c) Regularization and Training Stabilization: We apply dropout on edge features (0.2 on $\mathbf{x}_{uv}$ before ϕ_{edge}/ϕ_{score}), classifier input (0.1 on $\mathbf{z}_v$ before ϕ_{cls}), and atlas embeddings (0.5 on $\mathbf{e}_v$ before concatenation). For learned topology, we stabilize training by freezing the scorer network for the first 10 epochs and then unfreezing it.

d) Hyperparameter Selection: Backbone hyperparameters were selected in a separate grid search using a fixed baseline topology (Spatial-k), with validation macro precision-recall AUC (PR-AUC) as the selection metric (MCC as a tiebreaker), and then fixed for all comparisons and sweeps. Fixing hyperparameters from a single anchor topology (Spatial-k) is deliberate: per-topology tuning would confound topology comparisons, and Spatial-k is not top-performing, so this choice does not favor the observed ranking. Unless stated otherwise, we use $H = 32$, $d_{\mathrm{emb}} = 16$, $\gamma = 1.0$, learning rate 0.02 for fixed-topology models and 0.001 for learned-topology models. Dropout probabilities are given in the Regularization and Training Stabilization paragraph above. For budgeted top-k comparisons, we use target retention $r = 0.3$, chosen during the initial Spatial-k grid search as a moderate sparsity anchor for matched-budget comparisons; the full retention sweep is reported in Fig. 2. All MLPs (ϕ_{edge}, ϕ_{cls}, ϕ_{score}) are 2-layer with ReLU activations.

For Region-Bridge-c, intra-region edges alone yield mean retention 0.182. To match the comparison target $r = 0.3$, we choose $c = 0.12$, so each node adds $b_v = \lceil c\,(n_p - 1) \rceil$ cross-region bridges, giving a realized retention of 0.306.

e) Evaluation Metrics: For each topology, metrics are computed per LOPO fold, averaged across the 40 folds within each seed, and then summarized as mean and standard deviation across the 5 seeds. To compute thresholded metrics, we choose a decision threshold on pooled validation nodes using Youden's J statistic [27]. This threshold is then applied to the held-out test patient to compute sensitivity, specificity, and Matthews correlation coefficient (MCC).

f) Edge Retention Reporting and Sweep: We study sensitivity to sparsity by sweeping 20 target retention ratios $r \in \{0.05, 0.10, \ldots, 1.00\}$. For each r, we vary the per-node incoming-edge budget $k = \lceil r(n_p - 1) \rceil$ and evaluate the resulting topology. The sweep uses the same LOPO cross-validation protocol, training procedure, reporting, and five random seeds as the main experiments. We report performance as a function of the mean realized edge retention on the held-out test patients. Note that because Region-Bridge-c is swept via c using the same target ratio grid, its realized retentions may not precisely match those of the other topologies at the same grid point. This is presented in Fig. 2 where the y-axis is mean PR-AUC and shaded bands denote ± 1 standard deviation across seeds.

g) Ablations: We run diagnostic ablations using the Dense topology, following the same LOPO protocol and metric computation as in the main experiments (Table II). Node-only ablations remove all edges, eliminating edge aggregation while retaining atlas node embeddings. Edge-only ablations set node embeddings to zero and retain the original edge features. Shuffle ablations permute node metadata (atlas indices) or edge attribute vectors within each patient.

IV. Results

We report PR-AUC as the primary metric due to strong class imbalance: 12% EZ nodes and a random PR-AUC baseline of ≈ 0.12. Table I compares topologies under LOPO evaluation with a fixed backbone. Budget-controlled methods are compared at a matched realized edge retention of $\approx 30\%$ (target $r = 0.3$) incoming edges.

a) Fixed Structural Priors: Among the fixed topologies, performance varies with the restrictiveness of the prior (Table I). Intra-Region achieves the strongest PR-AUC among structural priors (0.340) and produces higher sensitivity than specificity under the Youden-selected threshold (sens. 0.694; spec. 0.625). The more restrictive lead-based masks underperform vs. Dense. Intra-Lead achieves a PR-AUC of 0.320, and Lead-Chain achieves a PR-AUC of 0.306. By design, the structural masks induce different graph densities (retention: Intra-Region 0.182; Intra-Lead 0.082; Lead-Chain 0.015).

b) Topology Comparisons at Matched Retention ($\approx 30\%$): Region-Bridge-c achieves the highest observed mean PR-AUC (0.371 vs. 0.349 for Dense) at 0.306 retention, a 69% edge reduction. It also has the highest receiver operating characteristic AUC (ROC-AUC) 0.743 and MCC 0.221. Spatial-k attains 0.363 PR-AUC at 0.303 retention; Random-k attains 0.321.

c) Learned Topology: The edge-only learned sparsifier, Learned-edges-k, attains a PR-AUC of 0.339 and ROC-AUC of 0.709 at 0.303 retention. The best learned variant overall, shown in Table I, is Learned-Region20-k (PR-AUC 0.354; ROC-AUC 0.722 at 0.304 retention). The additional atlas-augmented variants such as Learned-MICCAI-* were evaluated under matched conditions but were omitted from the table for brevity. All performed similarly, with the group-average PR-AUC at 0.345 ± 0.027 and ROC-AUC at 0.710 ± 0.015.

TABLE I
PERFORMANCE ACROSS TOPOLOGIES AT FIXED RETENTION OF 0.3 (MEAN ± STD). BEST VALUES ARE BOLD.

Topology	PR-AUC	ROC-AUC	MCC	Sens.	Spec.	Edge Retention
Dense	0.349(14)	0.728(17)	0.212(11)	0.680(26)	0.650(31)	1.000
Intra-Lead	0.320(14)	0.691(6)	0.171(5)	0.602(18)	**0.653(17)**	0.082
Lead-Chain	0.306(22)	0.673(15)	0.160(23)	0.591(20)	0.636(39)	0.015
Intra-Region	0.340(9)	0.733(13)	0.210(16)	**0.694(22)**	0.625(20)	0.182
Region-Bridge-c	**0.371(15)**	**0.743(10)**	**0.221(11)**	0.693(18)	0.641(28)	0.306
Random-k	0.321(13)	0.698(2)	0.172(8)	0.624(40)	0.642(21)	0.303
Spatial-k	0.363(19)	0.716(8)	0.193(16)	0.651(29)	0.647(26)	0.303
Learned-edges-k	0.339(9)	0.709(14)	0.186(16)	0.621(45)	0.647(20)	0.303
Learned-Region20-k	0.354(28)	0.722(21)	0.198(8)	0.670(34)	0.632(33)	0.304

TABLE II
ABLATIONS USING THE DENSE TOPOLOGY (MEAN ± STD).

Condition	PR-AUC	ROC-AUC
Edge-only (no node embedding)	0.351(18)	0.722(15)
Edge-only (shuffled edge attributes)	0.162(8)	0.506(19)
Node-only (no edges)	0.191(9)	0.527(14)
Node-only (shuffled node metadata)	0.136(2)	0.489(6)

d) Sensitivity to Retention: Structured priors peak at moderate retention: Region-Bridge-c 0.384 (at 0.485 retention) and Spatial-k 0.363 (0.303). Learned variants peak similarly: Learned-edges-k 0.357 (0.354) and Learned-Region20-k 0.354 (0.304). In contrast, Random-k requires near-dense retention to reach its best performance (0.358 at 0.904).

e) Patient-Level Variability: The per-patient (seed averaged) change in PR-AUC relative to Dense ranges from -0.172 to $+0.489$ for Region-Bridge-c (mean $+0.021$). Region-Bridge-c has higher PR-AUC than Dense in 24 of 40 patients (60%). A Wilcoxon signed-rank test found no significant shift between Region-Bridge-c and Dense ($W = 344$, $p = 0.383$); $r_{\mathrm{rb}} = 0.161$ (95% bootstrap CI $[-0.200, 0.502]$) and mean ΔPR-AUC $+0.021$ (95% bootstrap CI $[-0.011, 0.058]$) indicate at most a weak tendency favoring Region-Bridge-c. Counting the topology with the highest seed-averaged PR-AUC per patient: Learned-Region20-k (7/40), Intra-Region (7/40), Learned-edges-k (5/40), Dense (5/40), Lead-Chain (4/40), Region-Bridge-c (4/40), Spatial-k (3/40), Intra-Lead (3/40), and Random-k (2/40).

f) Ablations: Table II reports ablation results. With Dense, the edge-only ablation attains a PR-AUC of 0.351 and a ROC-AUC of 0.722, while the node-only ablation attains a PR-AUC of 0.191 and a ROC-AUC of 0.527. Shuffling edge attributes (PR-AUC 0.162; ROC-AUC 0.506) or node metadata (PR-AUC 0.136; ROC-AUC 0.489) substantially degrades performance, indicating both contribute useful signal.

V. DISCUSSION

a) Topology choice meaningfully affects performance: Under a fixed mean edge-aggregation backbone, topology changes the mean PR-AUC by up to 0.065 across all evaluated topologies (range 0.306–0.371; Table I). Among budget-matched topologies at $\approx 30\%$ realized retention, the spread is 0.050 (0.321–0.371). Region-Bridge-c achieves the highest observed mean performance, compared with Dense (0.371 vs. 0.349 PR-AUC) while using $\approx 69\%$ fewer edges. Spatial-k performs well, substantially outperforming Random-k in PR-AUC at the same retention (0.363 vs. 0.321), indicating that which edges are retained matters beyond sparsity level alone. The most restrictive structural masks (Intra-Lead and Lead-Chain) retain far fewer edges (0.082 and 0.015) and exhibit reduced discrimination (0.320 and 0.306 PR-AUC), suggesting that very local interactions and node metadata carry signal. Still, broader connectivity patterns are needed to reach the best overall performance. Together, these results argue that graph construction is a fundamental design choice and should be reported alongside realized retention (and, when possible, sensitivity to retention) rather than treated as an unexamined preprocessing detail.

b) Why might Region-Bridge-c work well?: Region-Bridge-c combines two priors: dense within-Region20 connectivity and a small set of distance-based cross-region connections. This biases mean edge-aggregation toward anatomically coherent neighborhoods while preserving controlled inter-regional information flow; it also reduces reliance on many weak, long-range edges present in the dense graph. The sensitivity and specificity pattern in Table I is consistent with this interpretation: Intra-Region attains high sensitivity but lower specificity, while adding bridges improves specificity with similar sensitivity and yields higher overall discrimination. This aligns with evidence that interictal abnormalities remain locally structured while seizure propagation follows fewer cross-regional pathways [5], [10].

c) Sparsification is beneficial, but not uniformly: The retention sweep (Fig. 2) shows that sparsification can help, but the effect depends on how edges are chosen. Spatial-k and Region-Bridge-c peak at moderate retention, whereas Random-k approaches its best performance near the dense graph. Thus, different topologies have different optimal budgets, and reporting only a single sparsity level can be misleading.

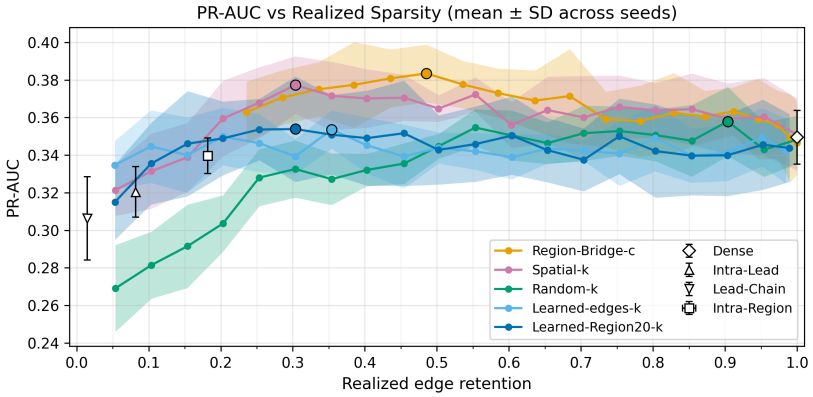

Fig. 2. PR-AUC versus realized edge retention (fraction of incoming edges retained on the held-out test patients). Lines show the mean across seeds, and error (bars and shaded bands) denote ± 1 standard deviation. Markers show the maximum observed mean PR-AUC across the sweep grid for each topology.

d) Learned sparsification: edge-only is insufficient under hard budgets: Under hard top-k budgets, learned sparsifiers do not exceed the highest performing structural prior, Region-Bridge-c, at matched retention (Table I). Incorporating node metadata improves over edge-only scoring (Learned-Region20-k 0.354 vs. Learned-edges-k 0.339 PR-AUC), suggesting alignment with anatomical structure aids generalization. Functional connectivity can be noisy and patient-specific; a globally trained scorer may overfit without contextual cues, especially in small-data settings (40 patients, LOPO). Dense-topology ablations support this: edge attributes carry signal, but node metadata provides complementary information (Table II). In the retention sweep, learned topologies remain below the highest performing fixed priors (Fig. 2), suggesting that hybrid approaches in which anatomy-informed priors guide learned sparsification may be more effective.

e) Patient Heterogeneity and Implications: Patient-level results are heterogeneous across all tested topologies, indicating that topology choice can materially affect outcomes for a given implantation. Using mean PR-AUC per patient, no single topology dominates (Sec. IV-E). Across patients, the median gap between the best and worst topology is 0.16 PR-AUC. This variability likely reflects differences in implant coverage, contact counts, and the scale at which resting-state connectivity aligns with a given patient's epileptogenic network organization. This suggests that a single global retention ratio and topology may not be optimal; patient-adaptive sparsification is promising but would require care to avoid tuning on held-out patients.

f) Limitations: First, labels are based on clinical EZ annotations with inherent inter-rater variability; restricting to favorable ILAE outcomes reduces but does not eliminate this. The 40 patient cohort supports only within-cohort LOPO generalization; external cross-dataset and multi-center validity remain untested. Second, window-averaged connectivity may obscure transient dynamics relevant to EZ characterization. Third, topology priors rely on coarse region definitions and MNI-space distances; localization errors can affect both structural masks and learned metadata. Fourth, topology rankings may depend on matched retention, a single Spatial-k reference setting, and the fixed mean-aggregation backbone (Sec. III-E), which we deliberately keep simple to isolate topology effects; as this model overfits without early stopping, limited capacity is unlikely the main bottleneck, but stronger architectures could still alter the rankings. We report a sweep but do not provide a principled rule for choosing r. Finally, enforcing a hard top-k for fair comparison may limit expressivity relative to fully weighted or adaptive schemes. This is a controlled topology comparison, not a performance maximization study or a benchmark against clinical interpretation.

VI. Conclusions and Future Work

We presented a controlled study of graph-topology choices for resting-state sEEG EZ localization, using a mean edge-aggregation backbone and LOPO evaluation across 40 patients. By matching per-node incoming-edge budgets and using a shared backbone, we isolate the effect of topology from confounds caused by edge density. Our results show that topology is a key design decision. Under a matched budget of $\approx 30\%$, Region-Bridge-c achieves the highest observed mean performance (PR-AUC 0.371 ± 0.015; ROC-AUC 0.743 ± 0.010), peaking at PR-AUC 0.384 at 48.5% retention in a sweep. Spatial-k is competitive at the same budget, while random pruning requires $\approx 90\%$ retention to approach strong priors. In contrast, our learned sparsifiers are competitive with structural baselines at matched retention, but they do not exceed Region-Bridge-c on average. The best sparsification varies by patient; the learned sparsifier performs best for some, reinforcing the need for patient-adaptive topology choices over a single global prior. Finally, dense topology ablations show both edge attributes and node metadata are important, as shuffling either source degrades performance (Table II).

Several directions can extend this work. First, learned sparsifiers could incorporate stronger inductive biases that reflect successful priors, such as region-aware regularization, spatial smoothness penalties, or explicit constraints that favor sparse cross-region bridges. A practical alternative is to start from a structural prior and then fine-tune edge weights or

selections with a learned module. Second, patient-adaptive sparsity, choosing an r per patient or learning a patient-specific budget, may better accommodate the heterogeneity in implants and network organization. Third, moving beyond window-averaged static graphs to dynamic time-resolved connectivity may capture transient coupling patterns that are obscured by aggregation. Finally, validating topology choices across multi-center cohorts, electrode layouts, and labeling practices will be necessary to assess robustness and generalization.

ACKNOWLEDGMENTS

We want to acknowledge support with data acquisition and clinical phenotyping from Jason Buroker, Craig Scholle, Hansel M. Greiner, Jeffrey R. Tenney, Jesse Skoch, Francesco T. Mangano, and Katherine D. Holland. All conceptual work, methods, experimental design, data analysis, results interpretation, and the main text were produced by the authors. AI tools, specifically ChatGPT 5.2, were used for assistance in LaTeX figures, equations, table formatting, grammar review, and minor edits.

REFERENCES

[1] (2024) WHO epilepsy. World Health Organization. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/epilepsy

[2] P. Kwan, A. Arzimanoglou, A. T. Berg, M. J. Brodie, W. Allen Hauser, G. Mathern, S. L. Moshé, E. Perucca, S. Wiebe, and J. French, "Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies," *Epilepsia*, vol. 51, no. 6, pp. 1069–1077, Jun. 2010.

[3] S. Wiebe, W. T. Blume, J. P. Girvin, M. Eliasziw, and Effectiveness and Efficiency of Surgery for Temporal Lobe Epilepsy Study Group, "A randomized, controlled trial of surgery for temporal-lobe epilepsy," *The New England Journal of Medicine*, vol. 345, no. 5, pp. 311–318, Aug. 2001.

[4] J. A. Gonzalez-Martinez, "The Stereo-Electroencephalography: The Epileptogenic Zone," *Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society*, vol. 33, no. 6, pp. 522–529, Dec. 2016.

[5] M. A. Kramer and S. S. Cash, "Epilepsy as a disorder of cortical network organization," *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, vol. 18, no. 4, pp. 360–372, Aug. 2012.

[6] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," 2017, arXiv:1609.02907 [cs]. [Online]. Available: http://arxiv.org/abs/1609.02907

[7] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural Message Passing for Quantum Chemistry," Jun. 2017, arXiv:1704.01212 [cs]. [Online]. Available: http://arxiv.org/abs/1704.01212

[8] K. Oono and T. Suzuki, "Graph Neural Networks Exponentially Lose Expressive Power for Node Classification," 2020, arXiv:1905.10947 [cs]. [Online]. Available: http://arxiv.org/abs/1905.10947

[9] B. C. M. v. Wijk, C. J. Stam, and A. Daffertshofer, "Comparing Brain Networks of Different Size and Connectivity Density Using Graph Theory," *PLOS ONE*, vol. 5, no. 10, p. e13701, Oct. 2010. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0013701

[10] S. Jenssen, C. M. Roberts, E. J. Gracely, D. J. Dlugos, and M. R. Sperling, "Focal seizure propagation in the intracranial EEG," *Epilepsy Research*, vol. 93, no. 1, pp. 25–32, Jan. 2011.

[11] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, Sep. 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S105381190901074X

[12] F. Panzica, G. Varotto, F. Rotondi, R. Spreafico, and S. Franceschetti, "Identification of the Epileptogenic Zone from Stereo-EEG Signals: A Connectivity-Graph Theory Approach," *Frontiers in Neurology*, vol. 4, p. 175, Nov. 2013.

[13] I. Vlachos, B. Krishnan, D. M. Treiman, K. Tsakalis, D. Kugiumtzis, and L. D. Iasemidis, "The Concept of Effective Inflow: Application to Interictal Localization of the Epileptogenic Focus From iEEG," *IEEE transactions on bio-medical engineering*, vol. 64, no. 9, pp. 2241–2252, Sep. 2017.

[14] G. Varotto, G. Susi, L. Tassi, F. Gozzo, S. Franceschetti, and F. Panzica, "Comparison of Resampling Techniques for Imbalanced Datasets in Machine Learning: Application to Epileptogenic Zone Localization From Interictal Intracranial EEG Recordings in Patients With Focal Epilepsy," *Frontiers in Neuroinformatics*, vol. 15, p. 715421, 2021.

[15] N. Nandakumar, D. Hsu, R. Ahmed, and A. Venkataraman, "A Deep Learning Framework to Localize the Epileptogenic Zone from Dynamic Functional Connectivity Using a Combined Graph Convolutional and Transformer Network," in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2023, pp. 1–5, iSSN: 1945-8452. [Online]. Available: https://ieeexplore.ieee.org/document/10230831

[16] P. Nejedly, V. Hrtonova, M. Pail, J. Cimbalnik, P. Daniel, V. Travnicek, I. Dolezalova, F. Mivalt, V. Kremen, P. Jurak, G. A. Worrell, B. Frauscher, P. Klimes, and M. Brazdil, "Leveraging interictal multimodal features and graph neural networks for automated planning of epilepsy surgery," *Brain Communications*, vol. 7, no. 3, p. fcaf140, 2025.

[17] S. A. Amir, A. Agaronyan, W. Gaillard, C. Oluigbo, and S. M. Anwar, "Dual-task graph neural network for joint seizure onset zone localization and outcome prediction using stereo EEG," 2025, arXiv:2505.23669. [Online]. Available: http://arxiv.org/abs/2505.23669

[18] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, Mar. 2009. [Online]. Available: https://www.nature.com/articles/nrn2575

[19] M. Mazrooyisebdani, V. A. Nair, C. Garcia-Ramos, R. Mohanty, E. Meyerand, B. Hermann, V. Prabhakaran, and R. Ahmed, "Graph Theory Analysis of Functional Connectivity Combined with Machine Learning Approaches Demonstrates Widespread Network Differences and Predicts Clinical Variables in Temporal Lobe Epilepsy," *Brain Connectivity*, vol. 10, no. 1, pp. 39–50, Feb. 2020.

[20] L. Franceschi, M. Niepert, M. Pontil, and X. He, "Learning Discrete Structures for Graph Neural Networks," 2019, arXiv:1903.11960 [cs]. [Online]. Available: http://arxiv.org/abs/1903.11960

[21] B. Peters, V. Niculae, and A. F. T. Martins, "Sparse Sequence-to-Sequence Models," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, A. Korhonen, D. Traum, and L. Màrquez, Eds. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 1504–1519. [Online]. Available: https://aclanthology.org/P19-1146/

[22] D. Wendelken, B. Ervin, J. Buroker, C. Scholle, H. M. Greiner, J. R. Tenney, K. D. Holland, J. Skoch, F. T. Mangano, A. Minai, and R. Arya, "Intracranial High-Frequency Oscillations and Epileptogenic Zone: Incorporating Neuroanatomic Variation," *Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society*, Jun. 2025.

[23] B. Ervin, L. Rozhkov, J. Buroker, J. L. Leach, F. T. Mangano, H. M. Greiner, K. D. Holland, and R. Arya, "Fast Automated Stereo-EEG Electrode Contact Identification and Labeling Ensemble," *Stereotactic and Functional Neurosurgery*, vol. 99, no. 5, pp. 393–404, 2021.

[24] "Neuromorphometrics MICCAI 2012 multi-atlas labeling challenge dataset," Neuromorphometrics, Inc., 2012, 138 labeled anatomical structures from OASIS database MRI scans. [Online]. Available: https://www.neuromorphometrics.com/2012_MICCAI_Challenge_Data.html

[25] O. M. Cliff, A. G. Bryant, J. T. Lizier, N. Tsuchiya, and B. D. Fulcher, "Unifying pairwise interactions in complex dynamics," *Nature Computational Science*, vol. 3, no. 10, pp. 883–893, Oct. 2023. [Online]. Available: https://www.nature.com/articles/s43588-023-00519-x

[26] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," Aug. 2017, arXiv:1708.02002 [cs]. [Online]. Available: http://arxiv.org/abs/1708.02002

[27] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, Jan. 1950.