Soundscape Connectomes: Unsupervised Graph-Based Approach for Soundscape Mapping

Abstract

We introduce *soundscape connectomes*, which are graph representations of acoustic relationships of a landscape where nodes are geographical sites and edges reflect relations derived from each site's biophony. Soundscape connectomes are built from passive acoustic monitoring (PAM) recordings and are enabled by the unique acoustic signatures of habitats. However, in ecoacoustic analysis, ground-truth graphs or labels are often not available. We propose an unsupervised pipeline that decomposes recordings into sonotypes, builds per-site acoustic structures, infers graphs with several methods, and compares them using a smoothness-based, unsupervised criterion that scores reconstruction of held-out nodes. We apply the proposed method to a large-scale real-life data set acquired in the Colombian Andes with over 19.598 recordings on 17 sites and 292 sonotypes. Results show the stability of the generated connectomes for a variety of graph inference methods. These results provide a practical way to select a graph model without prior information and position soundscape connectomes as a complement to remote-sensing analyses for monitoring and conservation.

1 Introduction

Network maps or connectomes provide compact summaries of complex systems by linking units through data-driven relationships. In neuroscience, they have reshaped how function and disease are studied, from the "human connectome" formulation to large-scale initiatives that integrate distributed sensors into network representations [1; 2]. We adopt a similar organizing idea for ecoacoustics and define *soundscape connectomes*: graphs whose nodes are study sites and whose edges encode acoustically driven relations derived from each site's biophony. In this ecological analogue, passive acoustic monitoring (PAM) acts as the distributed sensor network, providing scalable, non-invasive, long-term measurements of biophony across space and time [3; 4; 5], enabling label-free comparison among sites. As a management tool, these networks may complement remote sensing by revealing on-the-ground dynamics of vocal communities (often at understory and sub-canopy levels) that are informative for monitoring, prioritization, and conservation decision-making [6; 5].

Advancements in ecoacoustic research, which includes both supervised and unsupervised detection and classification of vocal taxa as well as the analysis of acoustic indices [7; 8; 9; 10], have led to studies demonstrating that land-cover and habitat types can be identified directly from sound. Furthermore, it has been shown that habitats possess unique acoustic signatures, which positions acoustic data as a valuable complement to traditional site characterization methods [11; 12]. Following this idea, in [13], the authors decomposed recordings into sonotypes, time–frequency entities, and for each analyzed site counted their occurrences to construct a site-by-sonotype matrix for unsupervised cross-site comparison. Since sonotypes can later be linked to species, the representation remains biologically interpretable [14]. Using these site–level features as node attributes, relationships among sites were then inferred using a sparse Gaussian graphical model (Graphical Lasso), and edges were interpreted with sonotype information and ecological context.

In the context of graphs, several methods for graph inference can be applied to site-by-sonotype features. These methods include sparse statistical models (such as precision-matrix estimation), smoothness-based graph learning derived from graph signal processing (GSP), and simple neighborhood graphs (like k-NN), along with neural variants that can also learn the graph topol-

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

ogy [15; 16; 2; 17]. However, in ecoacoustics, when the goal is to compare sites and their connectivity rather than to identify specific species, there is typically no available ground-truth information; in this context, land-cover types are, at best, rough proxies. Selecting from different plausible graph inference methods thus becomes a label-free problem of network structure inference (unsupervised graph inference).

The objective of this study is to compare different graph inference approaches and to articulate a practical, unsupervised selection criterion to obtain a soundscape connectome. Under a smoothness–based view, we treat site features as graph signals and prefer graphs in which a held–out node is accurately reconstructed from its neighbors, summarizing reconstruction quality with the normalized mean squared error (NMSE). We implement this idea with a node-removal ablation test and report error rates (mean and standard deviation). To render edges ecologically interpretable, we analyze which sonotypes most significantly influence the similarity or dissimilarity between sites, examining their time-frequency information and temporal acoustic pattern. Ultimately, we aim to provide a usable criterion for selecting the most suitable graph-inference methods to derive soundscape connectomes and, via acoustic biodiversity, to assess ecosystem health and relationships among sites.

2 Methodology

Acoustic Dataset and node features: To perform the soundscape connectome analysis, we used a database derived from PAM conducted in a rural area of Puerto Wilches, Santander, Colombia $(7°21'52.5"N, 73°51'33.0"W)$. Recordings were collected in March 2021 (dry season) with a Song Meter Mini device, programmed to capture 1 min every 10 min at 48 kHz, yielding 19.598 clips over 10 days. 17 recorders were spaced by at least 300 m within a $\sim$9 km^2 area to reduce shared sources. The landscape is dominated by oil palm of varying ages with patches of secondary vegetation, forest, grassland, and aquatic vegetation (for map and detailed description see A.1).

As shown by [14], the analysis of this dataset revealed that the identified sonotypes follow patterns that are similar to the acoustic indices usually linked to biophony. In this study area, these findings indicate that the acoustic patterns derived from sonotypes are predominantly influenced by strong biophonic activity. Following [13], the recordings are preprocessed to retain primarily biophonic activity, and are then decomposed into sonotypes using an unsupervised pipeline that segments acoustic activity and clusters time–frequency patterns; sonotypes behave as biophony-oriented descriptors, and the approach requires no parameterization [14]. For each site, we then build its acoustic structure by counting the occurrences of every sonotype, producing an $n \times m$ acoustic–structure matrix X. Here, n is the number of recording sites (graph nodes) and m is the number of detected sonotypes. The entry $X_{i,j}$ is the count of occurrences of sonotype j at site i. Each row $x_i \in \mathbb{R}^m$ serves as the feature vector for node i in subsequent graph–inference methods. Details of the acoustic structure generation are provided in A.2.

Graph inference models: We compare three representative graph–inference methods over the site–by–sonotype features. (i) Sparse statistical (Graphical Lasso): This method estimates a sparse precision matrix Θ (the inverse covariance). An edge appears when $\Theta_{ij} \neq 0$, meaning sites i and j are conditionally linked after accounting for all others [15]. Is the baseline introduced in [13]. (ii) Laplacian learning (GSP smoothness): This method learns nonnegative edge weights W (and Laplacian L) so that site signals vary little across connected nodes [16]. This encodes the idea that acoustically similar sites should be adjacent. We test two simple similarities to seed the learning: Euclidean distance and correlation. (iii) Distance baseline using k–nearest neighbors (kNN): This approach builds a local graph directly from pairwise distances d_{ij} (e.g., Euclidean); each site connects to its k closest neighbors, resulting in an intuitive bounded-degree topology. For comparability across methods, we rescale edge weights to $[0, 1]$.

Edge interpretation: After inferring the graphs, we explain the connections (edges) by comparing sonotype distributions between linked sites using Total Variation Distance (TVD) (more details in A.3). We focus on sonotypes with low TVD and low variance as potential drivers of similarity and summarize them with a time-frequency scatter plot (showing hours versus peak frequency) along with acoustic temporal patterns. This approach reveals which acoustic entities support or oppose the connections we observe. In this context, we refer to the entities that are common across the inferred graph methods.

Unsupervised model–selection criterion: Since no expert–defined ground truth graph exists (land–cover types are only coarse surrogates), standard graph–quality metrics are ill–suited here. Link prediction scores (AUC/AP) require labeled positive/negative edges; community measures (e.g., modularity, NMI) assume known partitions; and global statistics such as average path length or diameter presuppose a single connected component and minimum degree conditions we do not impose (isolated nodes can be ecologically meaningful). Consequently, an unsupervised model selection criterion is needed to compare candidate graphs in this setting. We therefore adopt a smoothness–based view: treat each site's feature vector as a graph signal, and assume that sites joined by an edge should carry similar acoustic information. Intuitively, a preferable graph is one that makes the signal vary smoothly along its edges. To evaluate candidate graphs without labels, we repeatedly hide a random fraction of nodes (20–80%), reconstruct their features from neighbors while keeping observed nodes close to their original values. Reconstruction balances two terms: (i) a data–fidelity term for observed nodes and (ii) a smoothness term that penalizes large differences across connected sites. Each holdout level is run 100 times with independent splits; performance is summarized with Frobenius-normalized NMSE (mean $\pm$ SD).

3 Results

From sonotypes to node features: Passive acoustic recordings were decomposed into sonotypes and per–site acoustic structures were built by counting sonotype occurrences (here, n=17 sites, m=292 sonotypes). This representation is label–free at construction time yet biologically interpretable, since sonotypes can be linked post hoc to specific species during interpretation or downstream analyses. The resulting site–by–sonotype matrix serves as the node–feature matrix for graph inference; the full step-by-step workflow from raw audio to acoustic structures is summarized in Fig. 1 (more details in A.2).

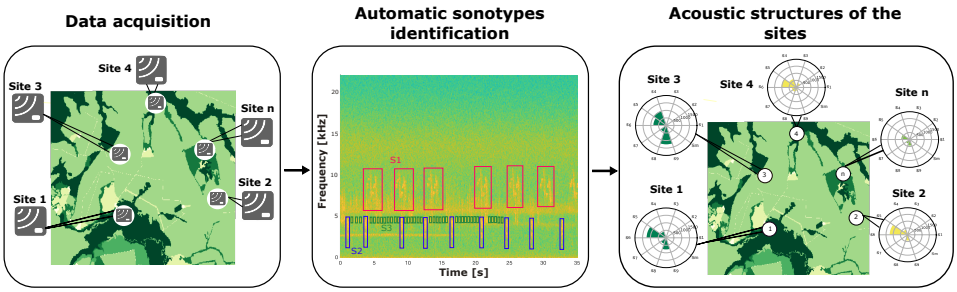

Figure 1: From recordings to node features. Left: data acquisition with passive recorders at 17 sites. Middle: unsupervised identification of sonotypes using [14]. Right: per–site acoustic structures (site–by–sonotype counts; here n=17, m=292) used as node features for graph inference.

Graph inference models: Figure 2 presents the soundscape connectomes inferred using four different approaches: Graphical Lasso (the baseline from 13), and the Kalofolias Laplacian learning with both Euclidean and correlation variants, as well as the k–NN method with k=3. The four graphs are broadly similar and identify a consistent set of strong connections. Most of these connections link sites with the same land cover type (for example, sites with oil palm are connected), while a smaller number connect sites with different land covers. The obtained graphs link places that sound alike, not just places that look alike on a land-cover map. A notable example is the connection between sites 2 and 8, which remain strongly linked despite their differing land covers and geographic distances. This connection is explained by shared sonotypes and synchronized acoustic temporal activity (see Appendix Fig. A4). Conversely, sites with sparse or unique sonotype profiles, such as sites 6 and 9, rarely establish connections with other sites. For visualization purposes, nodes are positioned based on their geographic coordinates and colored according to land cover types (orange for oil palm, light green for secondary vegetation, and dark green for forest). It is important to note that these labels were not used to construct the graphs; they serve solely for qualitative interpretation of the connectome. The width of the edges reflects the weight assigned by each method.

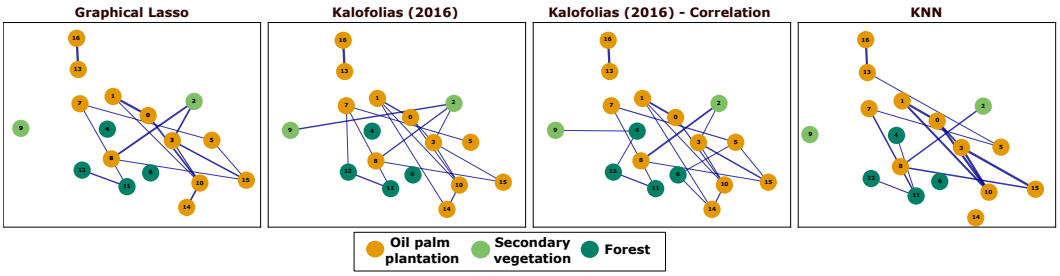

Figure 2: Soundscape connectomes inferred with four methods. The nodes represent 17 recording sites, each designated by geographic coordinates. The color of each node indicates the type of land cover: orange for oil palm plantations, light green for secondary vegetation, and dark green for forests. The edges illustrate the inferred similarities among these sites.

Unsupervised selection criterion: Patterns obtained in Figure 2 are encouraging but raise a practical question: *if several methods yield similar graphs, which should be used for mapping and downstream analysis?* Table 1 reports an unsupervised model selection criterion summary, mean $\pm$ SD of Frobenius-normalized NMSE over 20–80% node removal, which keeps methods comparable despite differences in scale or sparsity. As expected, error increases with the held-out fraction; the full ablation (NMSE as a function of the removal level) is shown in Appendix Fig. A5. Across the 20–80% range, the Kalofolias–correlation variant attains the lowest average NMSE with the tightest dispersion, with GLasso and kNN close behind. Although the gaps are modest, they are consistent, indicating that a simple reconstruction view can help choose among graph-inference methods even when no ground-truth topology exists.

Table 1: Unsupervised graph selection criterion summary. Reported is the mean $\pm$ SD of Frobenius–normalized NMSE (lower is better), averaged over holdout levels from 20% to 80% with repeated random removals per level. All methods use identical preprocessing; NMSE vs holdout curves are in Appendix Fig. A5.

Method	**Avg NMSE $\downarrow$ (20–80%)**
Glasso	0.506 ± 0.056
Laplacian learning (Kalofolias-Euclidean)	0.527 ± 0.053
Laplacian learning (Kalofolias-Correlation)	$\mathbf{0.494 \pm 0.045}$
kNN (k=3)	0.526 ± 0.051

4 Conclusions

Soundscape connectomes provide a compact complement to structural landscape analyses by linking sites that sound similar and revealing recurrent acoustic relationships that land cover maps often overlook. Using methods such as GLasso, Laplacian learning (Euclidean/correlation), and k-Nearest Neighbors (k-NN), we observed a consistent core of connections, even though there were method-specific differences. This highlights the need for an unsupervised model selection criterion; our smoothness-based reconstruction approach is a preliminary step in that direction. The findings suggest that graph-based representations can improve mapping and monitoring workflows, allowing for the prioritization of sites based on information that remote sensing might overlook. However, there are important limitations to consider: this is a static analysis, we aggregate all recordings and ignore temporal dynamics (diurnal/seasonal), and it relies on edge visualizations that employ thresholds. Future work should aim to formalize the unsupervised selection criterion, evaluate graph-inference methods on unthresholded, weighted graphs, scale the analysis to incorporate temporal structure (diurnal and seasonal) to reveal ecosystem dynamics, and strengthen ecological explanations for observed edges, deriving them directly from graph structure and complementing sonotype-based interpretations. The long-term goal is to deliver decision-ready soundscape connectome maps to support monitoring and conservation.

References

[1] Olaf Sporns, Giulio Tononi, and Rolf Kötter. The human connectome: A structural description of the human brain. *PLOS Computational Biology*, 1(4):e42, 2005. doi: 10.1371/journal.pcbi.0010042.

[2] Ivan Brugere, Brian Gallagher, and Tanya Berger-Wolf. Network structure inference, a survey: Motivations, methods, and applications. *ACM Computing Surveys*, 51, 10 2018. doi: 10.1145/3154524.

[3] Bryan C. Pijanowski, Almo Farina, Stuart H. Gage, Sarah L. Dumyahn, and Bernie L. Krause. What is soundscape ecology ? an introduction and overview of an emerging new science. *Landscape Ecol*, 26:1213–1232, 2011. doi: 10.1007/s10980-011-9600-8.

[4] Rory Gibb, Ella Browning, Paul Glover-Kapfer, and Kate E. Jones. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, 2019(September 2018):169–185, 2018. ISSN 2041210X. doi: 10.1111/2041-210X.13101.

[5] Dan Stowell and Jérôme Sueur. Ecoacoustics: acoustic sensing for biodiversity monitoring at scale. *Remote Sensing in Ecology and Conservation*, 6:217–219, 2020. doi: 10.1002/rse2.174.

[6] Jérôme Sueur and Almo Farina. Ecoacoustics: the ecological investigation and interpretation of environmental sound. *Biosemiotics*, 8:493–502, 2015. doi: 10.1007/s12304-015-9248-x.

[7] Carol Bedoya, Claudia Isaza, Juan M. Daza, and José D. López. Automatic recognition of anuran species based on syllable identification. *Ecological Informatics*, 24:200–209, 2014. doi: 10.1016/j.ecoinf.2014.08.009.

[8] Jack LeBien, Ming Zhong, Marconi Campos-Cerqueira, Julian P. Velev, Rahul Dodhia, Juan Lavista Ferres, and T. Mitchell Aide. A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. *Ecological Informatics*, 59:101113, 2020. doi: 10.1016/j.ecoinf.2020.101113.

[9] David Rendon, Susana Rodriguez-Buritica, Camilo Sánchez-Giraldo, Juan Daza, and Claudia Isaza. Automatic acoustic heterogeneity identification in transformed landscapes from colombian tropical dry forests. *Ecological Indicators*, 140:109017, 07 2022. doi: 10.1016/j.ecolind.2022.109017.

[10] Luc Barbaro, Anne Sourdril, Jeremy Froidevaux, Maxime Cauchoix, François Calatayud, Marc Deconchat, and Amandine Gasc. Linking acoustic diversity to compositional and configurational heterogeneity in mosaic landscapes. *Landscape Ecology*, 37:1–19, 04 2022. doi: 10.1007/s10980-021-01391-8.

[11] Andrés Castro-Ospina, Miguel Solarte-Sanchez, Laura Vega-Escobar, Claudia Isaza, and Juan Martinez-Vargas. Graph-based audio classification using pre-trained models and graph neural networks. *Sensors*, 24:2106, 03 2024. doi: 10.3390/s24072106.

[12] Dimitrios Bormpoudakis, Jérôme Sueur, and John Pantis. Spatial heterogeneity of ambient sound at the habitat type level: Ecological implications and applications. *Landscape Ecology*, 28:495–506, 03 2013. doi: 10.1007/s10980-013-9849-1.

[13] Maria J. Guerrero, Camilo Sánchez-Giraldo, Cesar Uribe, Victor Martinez Arias, and Claudia Isaza. Graphical representation of landscape heterogeneity identification through unsupervised acoustic analysis. *Methods in Ecology and Evolution*, 16:1255–1272, 05 2025. doi: 10.1111/2041-210X.70041.

[14] Maria. J. Guerrero, Carol. L. Bedoya, José. D. López, Juan. M. Daza, and Claudia. Isaza. Acoustic animal identification using unsupervised learning. *Methods in Ecology and Evolution*, 14:1500–1514, 2023. doi: 10.1111/2041-210X.14103.

[15] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics (Oxford, England)*, 9:432–41, 08 2008. doi: 10.1093/biostatistics/kxm045.

[16] Vassilis Kalofolias. How to learn a graph from smooth signals. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pages 920–929, Cadiz, Spain, 09–11 May 2016. PMLR.

[17] Ziyu Jia, Youfang Lin, Jing Wang, Ronghao Zhou, Xiaojun Ning, Yuanlai He, and Yaoshuai Zhao. Graphsleepnet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)*, pages 1324–1330. International Joint Conferences on Artificial Intelligence Organization, 7 2020. doi: 10.24963/ijcai.2020/184.

A Appendix

A.1 Dataset and study area

The study area is a rural landscape in Puerto Wilches (Santander, Colombia) dominated by oil–palm stands of varying ages (75%), interspersed with secondary vegetation (7.6%), forest patches (6.1%), grasslands (5.5%), and aquatic vegetation (3.2%). Human activity near recorders is low: roads are private and infrequently used, and dwellings are sparse and distant from sensors, reducing anthropogenic noise. The map below complements the main text by showing (a) land–cover context and (b) exact recorder positions with IDs used throughout the analysis.

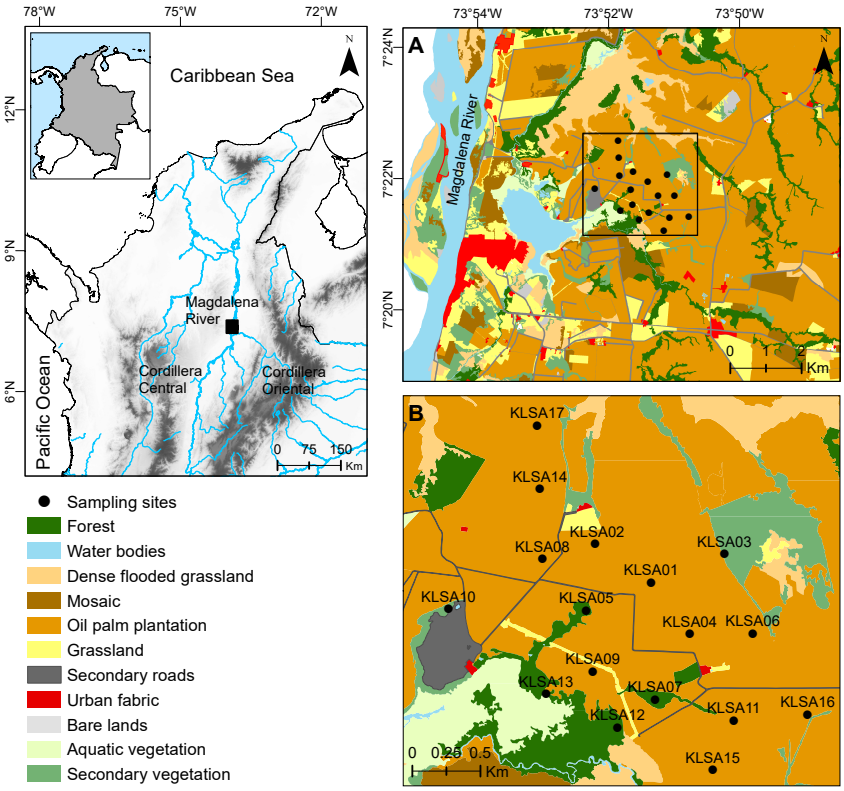

Fig. A 1: Study site map. (a) Land–cover context. (b) Geographic positions and IDs of the 17 recorders.

A.1.1 Codes and Dataset repository

Codes and dataset is available in: `https://anonymous.4open.science/r/Graph-inference-and-model-selection-1FE8`

A.2 Acoustic structures of the sites

First, sonotypes are obtained with the unsupervised method provided by Guerrero et al. [14], which automatically segments acoustic activity in the recordings and clusters events by similarities in time–frequency descriptors. The resulting clusters exhibit distinct acoustic patterns that can be linked to species calls. In our study, we do not assign species labels; instead, we treat sonotypes as soundscape descriptors and use their occurrence frequencies to construct each site's acoustic structure. The procedure is fully unsupervised and requires no manual parameterization.

After extracting sonotypes, we compile for each site the counts per sonotype, forming an $n \times m$ matrix (n sites, m sonotypes). In the figure 2, each bar ($S1 \ldots Sm$) denotes a unique sonotype and bar length is proportional to its number of detections, providing a rapid view of the site's dominant

acoustic patterns. These structures enable quantitative comparisons within and across sites, revealing sonotype richness, acoustic diversity, and differences in biophony composition without training data or prior knowledge of how many species are present.

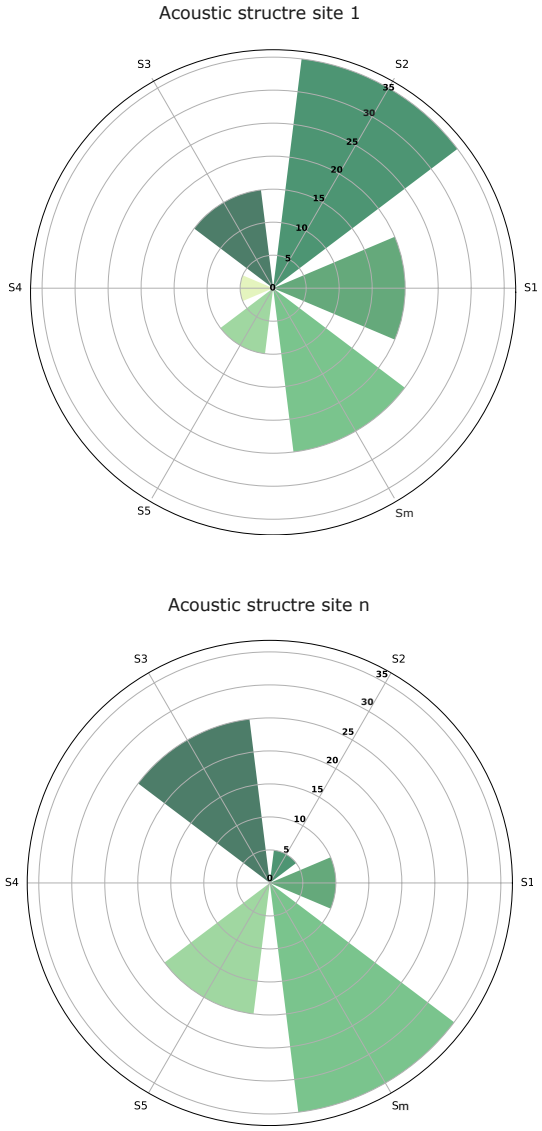

Fig. A 2: (a) Site 1; (b) Site n. Each radial spoke ($S1 \ldots Sm$) is a distinct sonotype (colored consistently), and its radius encodes the number of detections. The plot serves as an acoustic fingerprint, highlighting sonotype diversity and the site's biophony. Even when two sites share many sonotypes, their occurrence profiles can differ, enabling direct comparative analysis.

A.2.1 Santander soundscape fingerprints (study area)

The acoustic dataset was processed as described in Section 2 yielding a 17×292 site-by-sonotype count matrix spanning 10 recording days. Each site is thus summarized by its acoustic structure. For illustration, Fig. 3 shows a simplified rendering in which only the first ten sonotypes (the same set for all sites) are displayed to improve readability. The sonotype–occurrence matrix was then normalized and used as the input to the graph inference models.

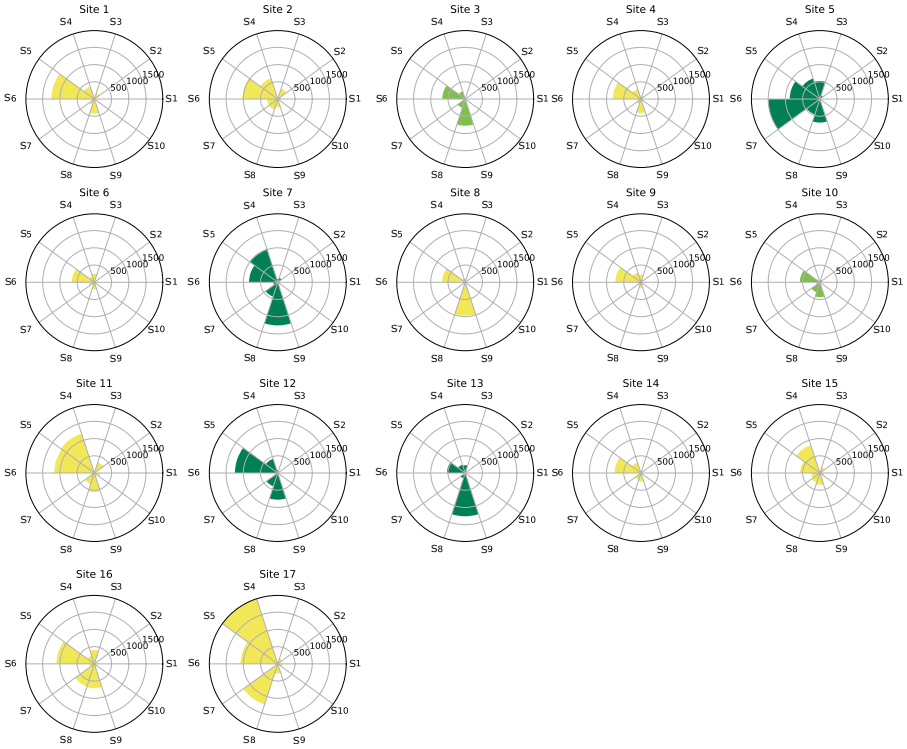

Fig. A 3: For each site sampled in Puerto Wilches (Santander, Colombia), we built an acoustic structure using a shared sonotype set. For visualization only, colors indicate land cover as documented by experts: yellow = oil palm plantations, light green = secondary vegetation, dark green = forest. Crucially, land-cover labels were not used to estimate acoustic heterogeneity or to infer connections among sites.

A.3 Edge interpretation with sonotypes

For each connected pair of sites, we turn sonotype counts into relative frequencies and compare their sonotype profiles with the Total Variation Distance (TVD). TVD tells us how different the two distributions are: 0 means "the profiles look the same," 1 means "they are completely different." We then keep the sonotypes that show low TVD between the two sites, these are the shared, acoustic entities that most plausibly drive the similarity (they explain why the edge exists). Sonotypes with high TVD are noted as contrasts (they explain differences).

$$\mathrm{TVD}(P, Q) \;=\; \frac{1}{2} \sum_{k} \big|P(k) - Q(k)\big|$$

$P(k)$ and $Q(k)$ are the relative frequencies of sonotype k at each site; $|P(k) - Q(k)|$ is the absolute difference for that sonotype; the sum adds differences across all sonotypes; the factor $\frac{1}{2}$ scales the result to the range $[0, 1]$.

For each pair in Fig. A4 we show: (i) per–site diurnal profiles (hourly counts) to assess alignment in daily activity, and (ii) a time–frequency scatter of the selected low–TVD sonotypes, the shared drivers of similarity, where each point marks a detection at its hour and peak frequency. In Fig. 4A (Sites 2–8), the connection is strong, supported by many shared sonotypes and tightly aligned activity with peaks around dawn and dusk.

In Fig. 4B (Sites 6–9), connectivity is weak: few shared sonotypes and poor diurnal alignment. Only shared (low–TVD) sonotypes are plotted in the time–frequency panels.

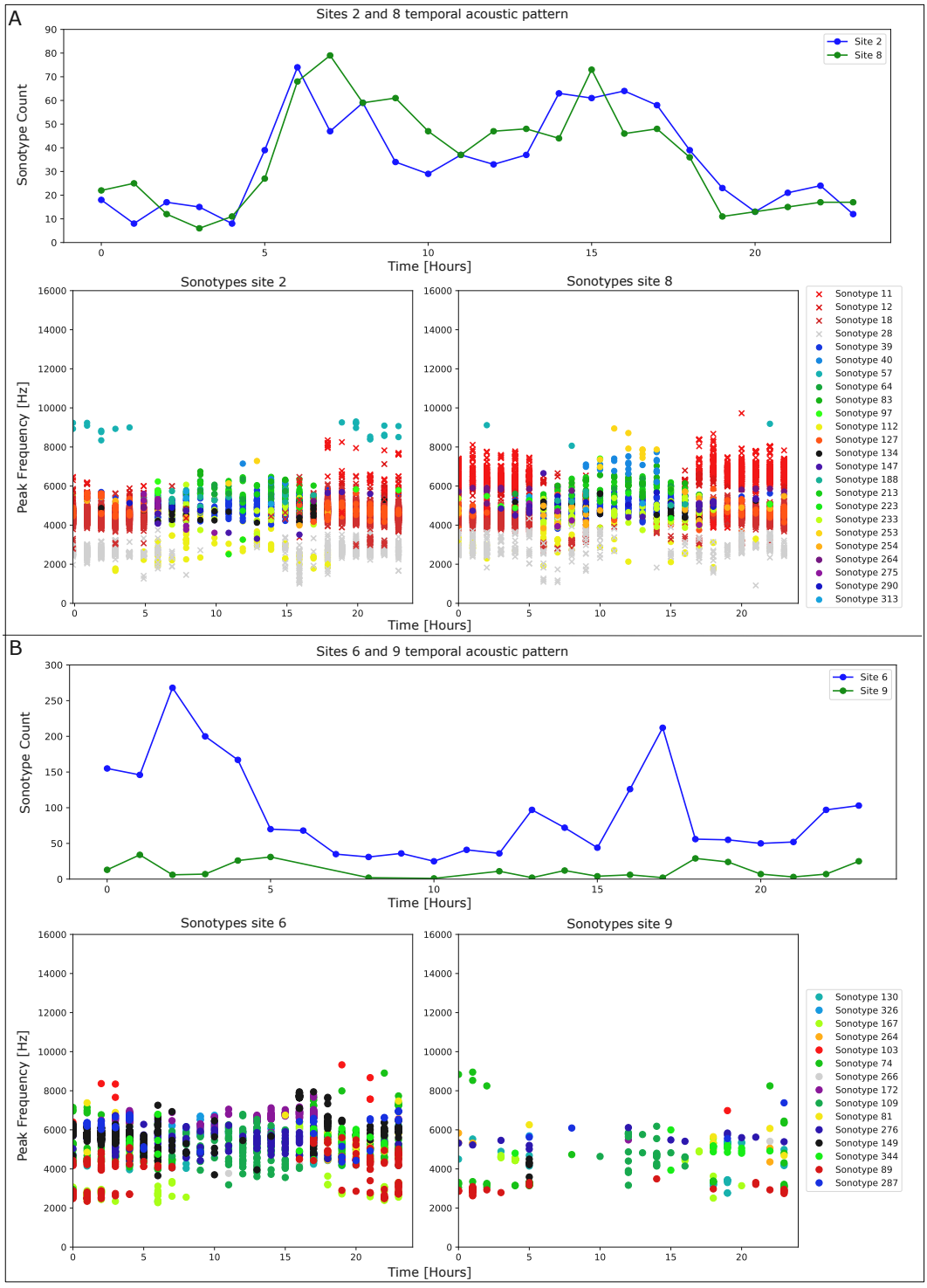

Fig. A 4: **Sonotype-based interpretation of graph structure.** **(A)** *Stable cross-cover link* (Sites 2–8): top, hourly sonotype counts show aligned acoustic time activity; bottom, time–frequency plots display shared sonotypes (colored markers) across sites. **(B)** *Low-degree/non-connecting sites* (Sites 6–9): weak alignment and sparse, site-specific sonotype sets explain the lack of edges.

A.4 Ablation curves for the selection criterion

We repeatedly hide a random fraction of nodes (20–80%), reconstruct held–out features from neighbors under a smoothness prior, and summarize performance with Frobenius–normalized NMSE. Each holdout level is run 100 times with independent splits. Curves (median; shaded SD) complement the summary table in the main text.

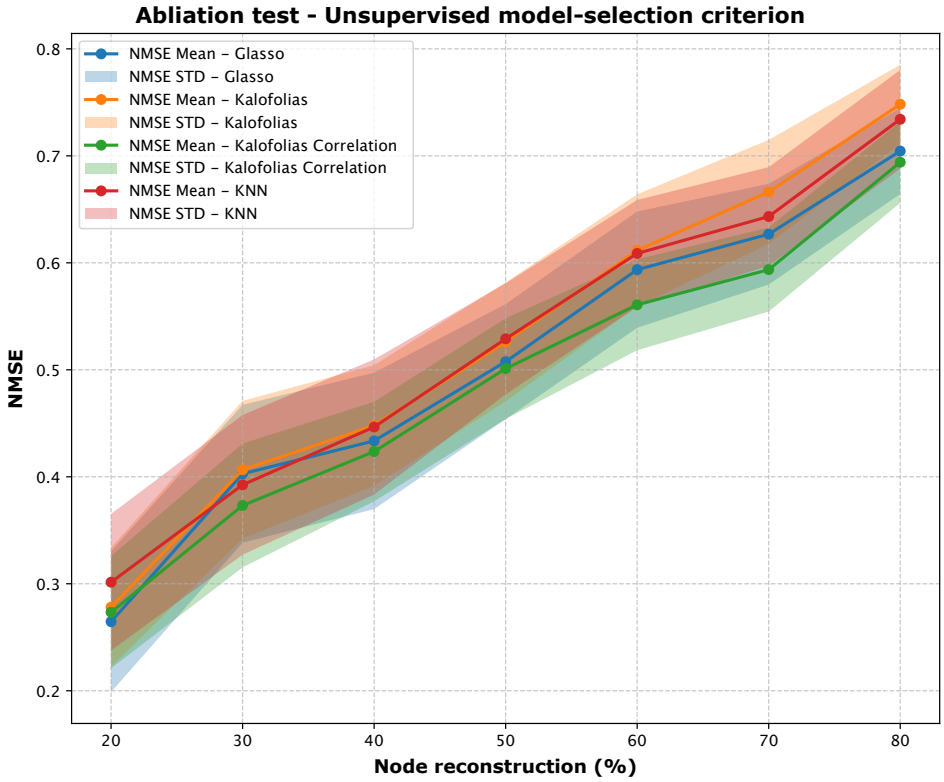

Fig. A 5: **Node-removal ablation.** Normalized MSE (median; shaded = SD) as a function of the fraction of held-out nodes (20–80%) for each graph-inference approach. Curves illustrate the smoothness-based reconstruction perspective and complement the summary in Table 1.