LoCoMapper: Local Structural Uncertainty Analysis of Mapper Graphs Under Cover Perturbations

Xinyuan Yan*
SCI Institute
University of Utah

Jixian Li†
TACC
University of Texas at Austin

Bei Wang‡
SCI Institute
University of Utah

Abstract

Mapper graphs provide compact topological summaries of high-dimensional data, but small changes in cover parameters can substantially alter their topology, making it difficult to distinguish persistent structures from those that arise only under particular parameter configurations. Despite growing interest in uncertainty visualization, methods for representing uncertainty in Mapper graphs remain largely unexplored. We present *LoCoMapper*, a framework that takes a first step toward uncertainty-aware visualization of Mapper graphs by characterizing their local structural sensitivity to perturbations of the cover parameters. Given a reference Mapper graph, *LoCoMapper* aligns local regions of the reference graph with corresponding regions in an ensemble of perturbed graphs and summarizes their variation using two node-level measures: *local structural consistency*, which quantifies the preservation of local topology across reliable alignments, and *ensemble support*, which captures the amount of reliable evidence supporting that estimate. Jointly visualizing these measures on the reference graph enables users to distinguish robust, sensitive, and weakly supported regions. We demonstrate the effectiveness of *LoCoMapper* on a controlled two-dimensional dataset, a synthetic three-dimensional ant dataset, and word embeddings from a fine-tuned BERT model.

Index Terms: Mapper graphs, topological data analysis, uncertainty visualization, structural uncertainty, parameter sensitivity.

1 Introduction

High-dimensional data are inherently difficult to explore directly. Mapper addresses this challenge by constructing a graph that captures the relationships among clusters of similar data points [25]. The resulting graph provides a structured representation of the data: connected components may indicate distinct groups, branches may reveal diverging subgroups or trajectories, and loops may expose alternative pathways between related regions. By combining structural simplification with interpretability, Mapper has become a valuable exploratory tool across a wide range of applications, including cancer research [16, 17, 19], neuroimaging [13], biological morphology [16, 19], epidemic modeling [11, 12], and neural network analysis [4, 6, 21, 22, 23, 29, 30, 28]. Together, these studies demonstrate Mapper's effectiveness in linking complex, high-dimensional measurements to interpretable structural hypotheses.

Mapper graph construction depends on several user-defined parameters, including the filter function, cover resolution and overlap, and clustering algorithm. Choosing these parameters is challenging because even small changes can substantially alter the resulting topology. In particular, cover resolution and overlap affect cluster granularity and connectivity, potentially causing branches or loops to appear, disappear, merge, or split, or changing the number of connected components. Because no universally optimal parameter setting generally exists—except under restrictive assumptions on the underlying data distribution—analysts may struggle to determine whether an observed structure is robust or sensitive to the parameter selection. Although prior work has investigated Mapper stability and parameter selection [5, 7, 8, 20], understanding how individual graph structures respond to local parameter perturbations remains a significant challenge.

We introduce *LoCoMapper*, a framework for modeling and visualizing local structural uncertainty in a reference Mapper graph induced by perturbations to the cover parameters. Given a reference graph, *LoCoMapper* uses TopoAlign [29] to compare each local region with corresponding regions across an ensemble of Mapper graphs generated from nearby cover parameter settings. Because structural comparisons are meaningful only when corresponding regions share sufficient underlying data points, *LoCoMapper* first distinguishes reliable from unreliable correspondences. It then characterizes each reference region using two complementary node-level measures. *Local structural consistency* quantifies how consistently local structure is preserved across reliable correspondences, while *ensemble support* measures the amount of reliable evidence underlying that estimate (see Section 3.3). Together, these measures distinguish consistently preserved regions from those sensitive to parameter perturbations while indicating the strength of the supporting evidence. We visualize both measures directly on the reference graph, enabling analysts to identify robust and sensitive branches, loops, and connected regions. Parameter-sensitive structures are not necessarily spurious; they may reveal meaningful features specific to particular parameter settings and should be interpreted accordingly. Analysts can further examine the underlying data points of Mapper nodes to better understand these regions. Our approach is conceptually related to recent uncertainty visualization techniques for ensembles of topological descriptors [2, 27].

Our contributions are:

- An ensemble-based framework for analyzing the sensitivity of Mapper graphs by comparing a reference graph with an ensemble of graphs generated using nearby cover parameter settings.
- Two complementary node-level measures: *local structural consistency*, which quantifies local structural variation using Portrait Divergence across reliable alignments, and *ensemble support*, which quantifies the fraction of ensemble graphs that provide reliable alignments.
- A visualization that jointly encodes *local structural consistency* and *ensemble support* on the reference graph, enabling analysts to identify robust, sensitive, and weakly supported regions while preserving their chosen Mapper representation.

We demonstrate the utility of our framework on three datasets of increasing complexity: a controlled two-dimensional dataset, a synthetic three-dimensional ant dataset, and BERT word embeddings.

2 Mapper Graph

We briefly review the construction of Mapper graphs; see [25] for a detailed treatment in the simplicial-complex setting.

*e-mail: xinyuan.yan@utah.edu
†e-mail: jixianli@tacc.utexas.edu
‡e-mail: beiwang@sci.utah.edu

As illustrated in Figure 1(left), given a point cloud $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, the Mapper algorithm constructs a graph from a filter function $f : X \to \mathbb{R}$. The image $f(X)$ is covered by a collection of overlapping intervals, called the *cover*, $\mathcal{U} = \{U_1, \ldots, U_I\}$, parameterized by the number of intervals I and the amount of overlap p. For each interval U_i, the data elements whose filter values lie in U_i are clustered. Each resulting cluster becomes a node in the Mapper graph, and two nodes are connected by an edge whenever their corresponding clusters share one or more data elements. The resulting graph $M = (V, E)$ provides a compact topological summary of the data, capturing structures such as connected components, branches, and loops. For each node $v \in V$, we denote by $X(v) \subseteq X$ the subset of data elements represented by v [25].

The structure of a Mapper graph depends on the choice of filter function, cover, metric, clustering algorithm, and clustering hyperparameters. The filter determines which aspect of the data is summarized, while the metric and clustering configuration determine how data elements are grouped within each cover interval. In the commonly used uniform cover, the filter range is partitioned into equal-length intervals with a fixed overlap, and the cover is parameterized by its resolution, specified by the number or width of intervals, together with the overlap [7]. Alternative constructions employ balanced or adaptive covers whose interval sizes vary according to the data or other selection critera [1, 9, 26]. In this work, we consider uniform covers and study the effects of perturbing the cover resolution and overlap while keeping the filter function, metric, clustering algorithm, and clustering parameters fixed.

3 *LoCoMapper* Framework

We propose *LoCoMapper*, a framework for modeling and visualizing the local structural uncertainty of Mapper graphs under small cover perturbations. Given a *reference Mapper graph* constructed using a user-specified cover, *LoCoMapper* quantifies how consistently its local structures are preserved as the cover parameters (e.g., the number of intervals and overlap percentage) are perturbed.

Overall workflow. *LoCoMapper* first generates an ensemble of *perturbed Mapper graphs* by uniformly perturbing the cover parameters of the reference graph (Section 3.1). For each reference–perturbation pair, aligned local subgraphs are identified using TopoAlign [29] (Section 3.2). We then evaluate the reliability of each alignment based on the overlap of their underlying data points and quantify the structural discrepancy between the subgraphs. These comparisons are aggregated into two complementary node-level measures: *local structural consistency*, which quantifies the expected change in local graph structure under cover perturbations, and *ensemble support*, which measures the amount of reliable evidence supporting that estimate (Section 3.3). Finally, both measures are visualized on the reference graph, enabling analysts to identify stable, sensitive, and weakly supported regions (Section 3.4).

3.1 Reference Mapper and Parameter Perturbation

LoCoMapper generates an ensemble of perturbed Mapper graphs of a reference Mapper graph by sampling neighboring cover parameter settings while keeping all other parameters fixed. Specifically, we define a local neighborhood around the reference cover parameters: the number of intervals I and the overlap p. The neighborhood is controlled by three user-specified parameters: an interval perturbation radius Δ_I, an overlap perturbation radius Δ_p, and a sampling density M. We enumerate all interval values in $\mathcal{I} = \{I - \Delta_I, \ldots, I + \Delta_I\}$, and, for each interval value, uniformly sample M overlap values from $\mathcal{P} = [p - \Delta_p,\ p + \Delta_p]$. This procedure produces $N = M(2\Delta_I + 1)$ perturbed cover parameter settings. We then construct one perturbed Mapper graph for each setting, denoted by $G_p^{(j)}$ for $1 \le j \le N$.

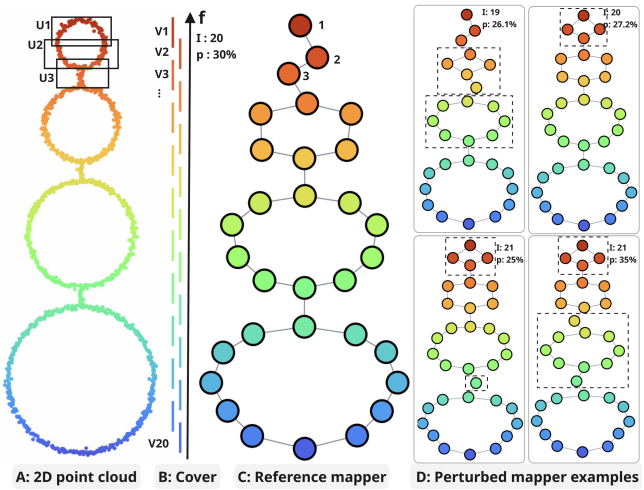


Figure 1: (A) A synthetic two-dimensional point cloud containing 4,000 points. (B) A uniform cover constructed from the height filter function f, using $I = 20$ intervals and $p = 30\%$ overlap. (C) The reference Mapper graph constructed from the cover in (B) using DBSCAN ($\mathrm{minPts} = 5$, $\varepsilon = 0.15$). (D) Perturbed Mapper graphs generated by varying the cover parameters within $I \in [19, 21]$ and $p \in [25\%, 35\%]$. Dotted boxes highlight representative local structural changes induced by the perturbations.

Figure 1 illustrates our running example. We construct a reference Mapper graph from a synthetic two-dimensional point cloud using height as the filter function, DBSCAN clustering, and a uniform cover with $I = 20$ intervals and $p = 30\%$ overlap. We then generate 30 perturbed graphs using $\Delta_I = 1$, $\Delta_p = 5$, and $M = 10$. Even these small perturbations produce noticeable local structural changes, including new loops and variations in loop size (Figure 1D), motivating uncertainty analysis of Mapper graphs.

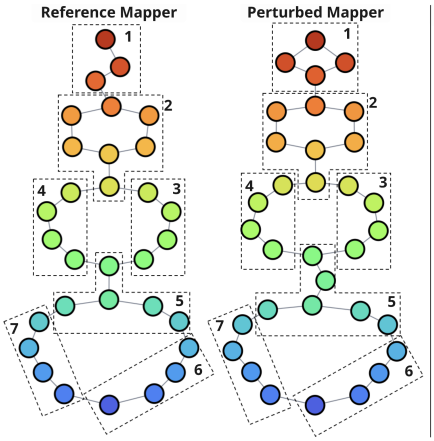
3.2 Local Structure Correspondence

To characterize local structural changes between the reference Mapper graph and the perturbed Mapper graphs, we first establish correspondences between their local regions. We employ TopoAlign [29], which automatically identifies aligned subgraphs between two Mapper graphs constructed from the same dataset by leveraging the overlap of their underlying data points. At a high level, TopoAlign constructs a joint graph by connecting nodes across the two Mapper graphs according to their shared data points and then partitions the joint graph using spectral clustering [18] (see [29] for details). Each resulting cluster contains one subgraph from the reference Mapper graph and one from the perturbed Mapper graph, forming a candidate local alignment. Applying TopoAlign to each reference–perturbation pair yields a collection of locally aligned subgraph pairs for subsequent structural comparison. Figure 2 (left) shows an example where TopoAlign identifies seven local alignments between a reference and a perturbed graph.

3.3 Node-Level Structural Uncertainty and Support

To characterize local structural variation in a reference Mapper graph, we first identify reliable local alignments from Section 3.2 based on the overlap of their underlying data points. We then aggregate their structural variation into two complementary node-level measures: *local structural consistency* and *ensemble support*.

Alignment reliability assessment. Because TopoAlign partitions a joint graph using global spectral clustering, some resulting alignments may correspond to subgraphs with limited overlap in their underlying data points. Following TopoAlign, we quantify the content similarity of each alignment using the Jaccard similarity, $J(A, B) = \frac{|A \cap B|}{|A \cup B|}$, where A and B denote the sets of data points



Local Structure Alignments

ID	Content similarity	Structural similarity
1	0.863	0.728
2	0.906	1.0
3	0.672	1.0
4	0.668	1.0
5	0.883	0.943
6	0.954	1.0
7	0.935	1.0

Figure 2: Left: Seven local alignments identified by TopoAlign between the reference and a perturbed graph. Right: Content similarity (Jaccard similarity) and structural similarity ($1 - \mathrm{PD}$; PD denotes Portrait Divergence) for each alignment. The two alignments with the lowest Jaccard similarity, shown in red, are discarded as unreliable.

contained in the aligned reference and perturbed subgraphs, respectively. To remove unreliable correspondences while maintaining sufficient coverage of the reference graph, we discard the 15% of alignments with the lowest Jaccard similarity.

Local structural consistency. For each retained alignment, we compute the Portrait Divergence [3] between the aligned reference and perturbed subgraphs. Portrait Divergence compares the distributions of shortest-path structures encoded by network portraits, without requiring node correspondence [3]. It captures structural differences across multiple distance scales while accommodating graphs of different sizes. The resulting divergence value is assigned to every node in the corresponding reference subgraph. For a reference node v, let $\mathcal{G}(v)$ denote the set of perturbed Mapper graphs that provide a retained alignment containing v, and let $k(v) = |\mathcal{G}(v)|$. We define the *local structural uncertainty* of v as

$$U(v) = \frac{1}{k(v)} \sum_{i \in \mathcal{G}(v)} \mathrm{PD}(R_i(v), P_i(v)) \,,$$

where $R_i(v)$ and $P_i(v)$ denote the aligned reference and perturbed subgraphs containing v, respectively. A larger $U(v)$ indicates greater structural variation under nearby cover perturbations. We define the complementary measure, *local structural consistency*, as

$$L(v) = 1 - U(v).$$

Ensemble support. Since unreliable alignments are discarded, the structural estimate for a node may be supported by fewer than N perturbed Mapper graphs. We therefore define its *ensemble support* as $S(v) = \frac{k(v)}{N}$, which measures the proportion of perturbed Mapper graphs that provide a reliable alignment for v. Thus, $U(v)$ quantifies the structural variation of v under cover perturbations, whereas $S(v)$ quantifies the amount of reliable evidence supporting that estimate. Estimates with low *ensemble support* should be interpreted cautiously, as they are based on relatively few reliable observations.

3.4 Uncertainty Visual Encoding

We encode the two node-level measures using independent visual variables applied to the border of each Mapper node. Because node color and size are often reserved for intrinsic data attributes, we treat uncertainty as an auxiliary visual layer on the node boundary, preserving the original node encoding while conveying the reliability of the local graph structure. Following prior work on contour-based uncertainty visualization [15], we explore several border-based visual variables, including wave amplitude, wave frequency, dash frequency, border thickness, and opacity (Figure 3). We avoid color to prevent interference with the existing node color encoding.

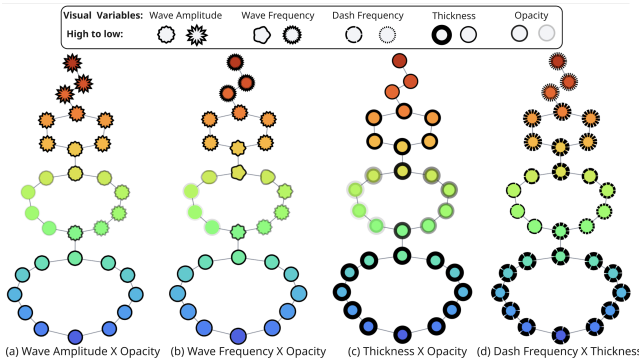


Figure 3: Top: Legend illustrating the five visual variables used to encode *local structural consistency* and *ensemble support*, ordered from high to low. Bottom: The four combinations of high/low *local structural consistency* and high/low *ensemble support*.

Our design is guided by two objectives: (1) structurally unstable regions should be immediately visually salient, and (2) uncertainty estimates supported by fewer observations should receive less visual emphasis. Figure 3 compares several candidate visual encodings. Among those considered, wave amplitude most effectively conveys *local structural consistency*, with increasingly irregular boundaries indicating greater structural variation. For *ensemble support*, opacity naturally conveys the strength of supporting evidence by reducing the visual prominence of nodes with lower confidence. Based on these observations, we encode *local structural consistency* using wave amplitude and *ensemble support* using opacity throughout the remainder of the paper.

Uncertainty interpretation. As shown in Figure 3, we inspect the local alignments associated with selected nodes. The local region around the top three nodes forms a loop in 13 of the 30 perturbed Mapper graphs, indicating an alternative local structure and consequently higher *local structural consistency*. In contrast, the bottom loop is consistently preserved across the ensemble, resulting in low *local structural consistency*. Although the second loop itself is largely preserved, the two nodes connecting it to neighboring loops frequently shift into adjacent structures (e.g., alignments 2 and 5 in Figure 2). As a result, these nodes participate in fewer reliable alignments, leading to lower *ensemble support*.

4 Use Cases

We demonstrate the effectiveness of *LoCoMapper* on two datasets. The first is a 3D mesh dataset from [10] (Section 4.1), and the second consists of contexualized word embeddings generated by a fine-tuned BERT model (Section 4.2).

4.1 3D Ant Dataset

Data processing. We apply *LoCoMapper* to the 3D ant dataset [10], which contains 6,370 points. As shown in Figure 4, we construct a reference Mapper graph using the distance to the abdomen tip as the filter function and a uniform cover with 15 intervals and 30% overlap. Under this parameter setting, the two midlegs are merged into a single branch. We then generate an ensemble of 30 perturbed Mapper graphs by varying the cover parameters within ± 1 interval and $\pm 5\%$ overlap.

Figure 5(a) shows the reference Mapper graph with uncertainty encoded on node borders: lower wave amplitudes indicate higher *local structural consistency*, while lower opacity indicates lower *ensemble support*. The head and abdomen exhibit low uncertainty, indicating that their local structures are consistently preserved under cover perturbations. In contrast, the leg regions are more uncertain. Among them, the forelegs and midlegs have higher *en-*

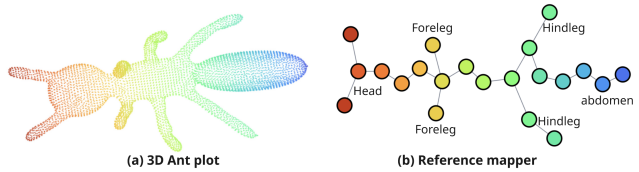


Figure 4: The 3D ant dataset (left) and its reference Mapper graph (right). The Mapper graph is constructed using the distance to the abdomen tip as the filter function, DBSCAN ($\text{minPts} = 5$, $\varepsilon = 0.05$), and a uniform cover with 15 intervals and 30% overlap. Data points and Mapper nodes are colored by their filter values.

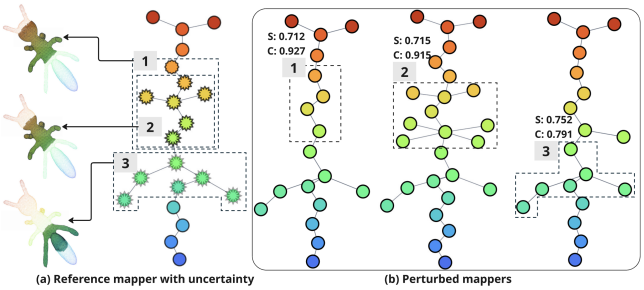


Figure 5: The reference mapper of the 3D ant dataset with uncertainty encoding. Wave amplitude represents structural uncertainty and opacity represents support. Three local alignments (1–3) are shown, with the aligned subgraphs outlined by dotted boundaries, the corresponding 3D points highlighted, and their structural similarity (S) and content similarity (C).

semble support than the hindlegs, suggesting that their uncertainty estimates are supported by more reliable local alignments.

To better understand these patterns, we inspect the corresponding local alignments in the perturbed Mapper graphs (Figure 5(b)). **Alignments 1 and 2**, associated with the forelegs and midlegs, exhibit low structural similarity. In the corresponding perturbed Mapper graphs, these leg branches appear or disappear together, indicating that their local structures are sensitive to small cover perturbations. **Alignment 3** corresponds to the hindleg region and has the lowest Jaccard similarity (0.752), causing it to be discarded as unreliable. A possible explanation is that the hindleg region spans a wider range of filter values and contains more branching structures, making the underlying point assignments less consistent across perturbations and therefore more difficult to align reliably.

4.2 Word Embeddings from BERT

Following Explainable Mapper [30], we fine-tune BERT-base on the word supersense classification task from STREUSLE v4.2 [24] and extract contextualized word embeddings from the final transformer layer. Restricting the analysis to the five most frequent semantic categories yields 1,525 embeddings of dimension 768.

We construct a Mapper graph using the ℓ_2 norm as the filter function, which captures the strength of an LLM's activation for an input token [23, 14], DBSCAN clustering ($\text{minPts} = 3$, $\varepsilon = 6$), and a uniform cover with 30 intervals and 30% overlap. *LoCoMapper* then generates an ensemble of 50 perturbed Mapper graphs by varying the cover parameters (intervals ± 2 and overlap $\pm 5\%$). Figure 6 shows the reference Mapper graph with uncertainty encoded on node borders. Each node is rendered as a pie chart showing its ground-truth supersense composition; pure nodes and components indicate that the fine-tuned embedding space clearly separates the supersense categories.

LoCoMapper reveals several representative forms of structural uncertainty. **Alignments 1 and 2** exhibit *component merges*, in which disconnected components containing words with the same supersense category (e.g., occurrences of *"to"* in Alignments 1 and

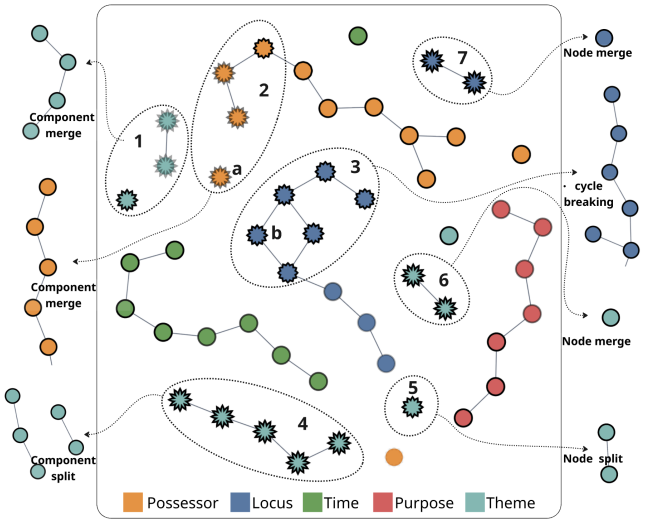


Figure 6: Reference Mapper graph of the fine-tuned BERT word embeddings with uncertainty encoded on node borders. Border wave amplitude and opacity encode *local structural consistency* and *ensemble support*, respectively. Local alignment examples (1–7) illustrate parameter-sensitive structural variations, including component merges, component splits, cycle breaking, and node merges/splits.

possessive pronouns such as *"our"* and *"my"* in Alignments 2) merge under nearby cover perturbations, suggesting that their separation is not robust. **Alignment 3** illustrates *cycle breaking*: a four-node loop disappears in 37 of the 50 perturbed Mapper graphs, indicating that the loop is sensitive to cover perturbations and is not consistently preserved across the sampled parameter neighborhood.

In contrast, **Alignment 4** demonstrates a *component split*, where a component containing different usages of *"for"* separates into two semantically coherent groups (e.g., *"wait for"* versus *"look for"* and *"ask for"*), revealing finer semantic distinctions that are obscured in the reference Mapper graph.

Finally, **Alignments 5–7** illustrate *node merges* and *node splits*, where neighboring nodes merge or split while largely preserving their associated word sets. These changes primarily occur when samples near cover boundaries are assigned to one or two overlapping intervals under different cover parameter settings. Together, these examples demonstrate that *LoCoMapper* distinguishes parameter-sensitive structures from stable semantic patterns, helping analysts identify which Mapper structures are reliable for downstream interpretation.

5 Limitations

LoCoMapper has several limitations that motivate future work. First, the framework relies on user-specified or heuristic parameters, including the perturbation range, sampling density, and alignment filtering threshold; their effects warrant systematic sensitivity analysis and more principled selection strategies. Second, although we use Portrait Divergence to quantify local structural differences, the framework can incorporate alternative graph similarity measures or combine structural and content-based information to capture application-specific characteristics. Third, regions with low *ensemble support* require cautious interpretation, as they may reflect ambiguous local structure, limitations of the TopoAlign optimization, or insufficient alignment evidence rather than genuine structural instability. Fourth, future work should explore alternative uncertainty encodings, such as color and luminance, and evaluate the perceptual effectiveness of visual variable combinations and mapping directions. Finally, application-driven evaluations with domain experts are needed to assess the framework's utility in real-world analytical workflows.

Acknowledgments

This work was supported in part by the National Science Foundation (NSF) under Grants IIS-2205418, DMS-2134223, and DMS-2301361.

References

[1] E. Alvarado, R. Belton, E. Fischer, K.-J. Lee, S. Palande, S. Percival, and E. Purvine. G-Mapper: Learning a cover in the mapper construction. *SIAM Journal on Mathematics of Data Science*, 7(2):572–596, 2025. doi: 10.1137/24M1641312 2

[2] T. M. Athawale, D. Maljovec, L. Yan, C. R. Johnson, V. Pascucci, and B. Wang. Uncertainty visualization of 2D Morse complex ensembles using statistical summary maps. *IEEE Transactions on Visualization and Computer Graphics*, 28(4):1955–1966, 2022. doi: 10.1109/TVCG.2020.3022359 1

[3] J. P. Bagrow and E. M. Bollt. An information-theoretic, all-scales approach to comparing networks. *Applied Network Science*, 4(1):45, 2019. doi: 10.1007/s41109-019-0156-x 3

[4] R. Ballester, C. Casacuberta, and S. Escalera. Topological data analysis for neural network analysis: A comprehensive survey, 2024. doi: 10.48550/arXiv.2312.05840 1

[5] F. Belchi, J. Brodzki, M. Burfitt, and M. Niranjan. A numerical measure of the instability of Mapper-type algorithms. *Journal of Machine Learning Research*, 21(202):1–45, 2020. doi: 10.48550/arXiv.1906.01507 1

[6] C. Bodnar, C. Cangea, and P. Liò. Deep graph Mapper: Seeing graphs through the neural lens. *Frontiers in Big Data*, Volume 4 - 2021, 2021. doi: 10.3389/fdata.2021.680535 1

[7] M. Carrière, B. Michel, and S. Oudot. Statistical analysis and parameter selection for Mapper. *Journal of Machine Learning Research*, 19(12):1–39, 2018. doi: 10.48550/arXiv.1706.00204 1, 2

[8] M. Carrière and S. Oudot. Structure and stability of the one-dimensional Mapper. *Foundations of Computational Mathematics*, 18(6):1333–1396, 12 2018. doi: 10.1007/s10208-017-9370-z 1

[9] N. Chalapathi, Y. Zhou, and B. Wang. Adaptive covers for mapper graphs using information criteria. In *2021 IEEE International Conference on Big Data*, pp. 3789–3800, 2021. doi: 10.1109/BigData52589.2021.9671324 2

[10] X. Chen, A. Golovinskiy, and T. Funkhouser. A benchmark for 3D mesh segmentation. *ACM Transactions on Graphics*, 28(3):1–12, 2009. doi: 10.1145/1576246.1531379 3

[11] Y. Chen and I. Volić. Topological data analysis model for the spread of the coronavirus. *PLoS One*, 16(8):e0255584, 8 2021. doi: 10.1371/journal.pone.0255584 1

[12] P. Dlotko and S. Rudkin. Visualising the evolution of english COVID-19 cases with topological data analysis Ball Mapper. *arXiv preprint arXiv:2004.03282*, 2020. doi: 10.48550/arXiv.2004.03282 1

[13] C. Geniesse, S. Chowdhury, and M. Saggar. NeuMapper: A scalable computational framework for multiscale exploration of the brain's dynamical organization. *Network Neuroscience*, 6(2):467–498, 06 2022. doi: 10.1162/netn_a_00229 1

[14] S. Ghosh, R. Datta, A. K. Das, and A. Chakrabarti. Towards explainability of SLMs by investigating token level activation. *arXiv preprint arXiv:2605.22377*, 2026. doi: 10.48550/arXiv.2605.22377 4

[15] J. Görtler, C. Schulz, D. Weiskopf, and O. Deussen. Bubble Treemaps for uncertainty visualization. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):719–728, 2018. doi: 10.1109/TVCG.2017.2743959 3

[16] S. Lockwood and B. Krishnamoorthy. Topological features in cancer gene expression data. *Pacific Symposium on Biocomputing*, pp. 108–119, 2015. doi: 10.1142/9789814644730_0012 1

[17] C. F. Loughrey, S. Maguire, P. Dłotko, L. Bai, N. Orr, and A. Jurek-Loughrey. A novel method for subgroup discovery in precision medicine based on topological data analysis. *BMC Medical Informatics and Decision Making*, 25(1):139, 3 2025. doi: 10.1186/s12911-025-02852-9 1

[18] A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, vol. 14. MIT Press, 2001. 2

[19] M. Nicolau, A. J. Levine, and G. Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences*, 108(17):7265–7270, 2011. doi: 10.1073/pnas.1102826108 1

[20] M. Pegoraro, A. Conforti, and M. Carrière. Building confidence regions for Reeb graphs using the interleaving distance. In *Statistical Science: From Theory to Applied Research III*, pp. 26–31. Springer Nature Switzerland, Cham, 2026. doi: 10.1007/978-3-032-30881-8_5 1

[21] E. Purvine, D. Brown, B. Jefferson, C. Joslyn, B. Praggastis, A. Rathore, M. Shapiro, B. Wang, and Y. Zhou. Experimental observations of the topology of convolutional neural network activations. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, 2023. doi: 10.1609/aaai.v37i8.26134 1

[22] A. Rathore, N. Chalapathi, S. Palande, and B. Wang. TopoAct: Exploring the shape of activations in deep learning. *Computer Graphics Forum*, 40(1):382–397, 2021. doi: 10.1111/cgf.14195 1

[23] A. Rathore, Y. Zhou, V. Srikumar, and B. Wang. TopoBERT: Exploring the topology of fine-tuned word representations. *Information Visualization*, 22(3):186–208, 2023. doi: 10.1177/14738716231168671 1, 4

[24] N. Schneider and N. A. Smith. A corpus and model integrating multiword expressions and supersenses. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1537–1547. Association for Computational Linguistics, Denver, Colorado, 2015. doi: 10.3115/v1/N15-1177 4

[25] G. Singh, F. Memoli, and G. Carlsson. Topological methods for the analysis of high dimensional data sets and 3D object recognition. In *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007. doi: 10.2312/SPBG/SPBG07/091-100 1, 2

[26] Y. Tao and S. Ge. A distribution-guided Mapper algorithm. *BMC Bioinformatics*, 26(1):73, 3 2025. doi: 10.1186/s12859-025-06085-5 2

[27] L. Yan, Y. Wang, E. Munch, E. Gasparovic, and B. Wang. A structural average of labeled merge trees for uncertainty visualization. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):832–842, 2020. doi: 10.1109/TVCG.2019.2934242 1

[28] X. Yan, S. Liu, K. Thopalli, and B. Wang. Visual exploration of feature relationships in sparse autoencoders with curated concepts. In *Mechanistic Interpretability Workshop at NeurIPS 2025*, 2025. doi: 10.48550/arXiv.2511.06048 1

[29] X. Yan, R. Sevastjanova, M. El-Assady, and B. Wang. TopoAlign: Topology-aware visual representation alignment. *arXiv preprint arXiv:2605.25541*, 2026. doi: 10.48550/arXiv.2605.25541 1, 2

[30] X. Yan, R. Sevastjanova, S. van der Ben, M. El-Assady, and B. Wang. Explainable Mapper: Charting LLM embedding spaces using perturbation-based explanation and verification agents. *IEEE Transactions on Visualization and Computer Graphics*, 2026. To appear. doi: 10.48550/arXiv.2507.18607 1, 4