Using Topological Data Analysis to Investigate Galactic Extinction Rates

Dylan Wolf

May 7, 2025

Abstract

This study explores the application of multidimensional persistent homology, a technique from topological data analysis (TDA), to investigate spatial patterns of galactic extinction rates. We probe this 4-D point-cloud geometry with multidimensional persistent homology, using 10,000 Sloan Digital Sky Survey data points (RA, Dec, redshift, r-band extinction). One- and two-parameter filtrations generate Betti heat maps which reveal tight clusters (β_0) and ring-like loops (β_1) that echo dust structures reported in large-scale astronomical surveys. The visual correspondence, however, is suggestive, not conclusive: confirming a physical link will require higher-resolution data, full rank-invariant software such as RIVET, and side-by-side comparison with 3-D dust reconstructions. Our primary contribution is methodological, a transparent, end-to-end multidimensional persistence homology workflow for extinction data that marks out the upgrades needed before firm astrophysical claims can be made.

1 Introduction

Interstellar dust is a commonly studied phenomenon as it plays a crucial role in astronomical observations. It attenuates and scatters light, altering the perceived brightness and color of celestial objects. This effect, known as extinction, depends on the column density of dust along a given line of sight and can obscure key astrophysical features. Accurately mapping extinction rates is essential for correcting observations and improving our understanding of galactic structures.

Topological Data Analysis (TDA) provides a way to investigate the geometric properties of data by examining how points in a dataset connect at multiple scales. It has been successfully employed to study cosmic matter distributions, revealing large-scale structures such as galactic filaments and clusters, by capturing both local and global geometric features. However, in astronomical observations, dust clouds between Earth and distant objects introduce an extinction component as the light we receive is attenuated creating uncertainty in the observed data. This raises the question of whether TDA can be leveraged to analyze the distribution of extinction rates.

To address this question, the project proposes constructing a higher-dimensional representation of extinction measurements so that TDA algorithms can detect topological features within the dust distribution. Identifying clusters, loops, or higher-dimensional analogues in the data could reveal how dust structures are arranged and whether they correlate with known galactic features, such as spiral arms or star-forming regions. Moreover, analyzing persistence across distance scales may provide insights into how extinction patterns evolve or cluster in different parts of the galaxy.

Accordingly, **this project seeks to determine whether TDA can be used to analyze the extinction rates of galactic structures**. If successful, this approach could expand the usage of TDA from purely studying large-scale cosmic matter to also capturing the geometric relationships in dust-based extinction data, potentially offering a fuller picture of galactic morphology and its underlying processes.

2 Brief Introduction to Topological Data Analysis

As discussed in [1], topology provides a framework for analyzing the geometric properties of datasets, offering a unique method for extracting structural information. A key component of this approach is the study of point cloud data, which consists of data points embedded in a coordinate system. The relationships between these points can be understood by examining the geometric structures formed under various parameter values.

To analyze these relationships, **simplicial complexes** are constructed. (For this and below bold terms, see Appendix A for rigorous definitions and Appendix B for illustrated examples). A commonly used construction in TDA is the **Vietoris-Rips complex** which involves increasing a radius around each data point. When two points fall within the radius of one another, they are connected, and as the radius grows, higher-order simplices emerge, forming geometric structures that encode topological information.

Each resulting geometric shape can be characterized by **Betti numbers**, denoted β_k, which describe topological features such as the number of connected components, loops, and voids. Specifically, β_0 counts the number of connected components, β_1 counts the number of one-dimensional holes (loops), and β_2 represents two-dimensional voids (cavities), and so on. These features capture essential shape properties in a dataset, helping to reveal structural patterns.

As the radius continues to increase, the topology of the dataset changes dynamically. The continuous maps between these evolving geometric structures illustrate the concept of **functoriality**, the idea that topological invariants should be preserved not just for individual objects but also for the relationships, or maps, between them. This principle plays a crucial role in data clustering, where topological structures are identified through **persistent homology**, which tracks how topological features appear and disappear across different scales. Through this methodology, a deeper understanding of the dataset's intrinsic shape can be obtained, even in high-dimensional or noisy settings.

In [2], the authors address the challenge of handling multiple parameters by introducing the notion of a **multifiltration**, which is a family of spaces parameterized by multiple

geometric dimensions. They show that although no complete discrete invariant exists for multidimensional persistence, a **rank invariant** can act as a robust discrete measure of Betti numbers across such multifiltrations.

Multidimensional persistence is thus an extension of persistent homology in which the birth and death of features are tracked across a multi-dimensional partially ordered set, rather than a single parameter. This adds considerable complexity compared to the usual one-parameter barcodes in persistent homology. The main idea is that instead of one filtration parameter, there are several independent parameters driving how to build and examine complexes. This leads to a partial order of complexes rather than a single chain.

In single-parameter persistence, a filtration can be constructed by gradually increasing one parameter, such as a distance threshold. This produces a sequence of topological spaces:

$$X_0 \subseteq X_1 \subseteq X_2 \subseteq \cdots \subseteq X_n,$$

and homology classes are tracked as they appear (birth) and disappear (death).

In multi-parameter persistence, two or more parameters may be used, such as a distance threshold and a density threshold. Instead of a single chain of spaces, a grid or partially ordered set (poset) of spaces, a multifiltration, is produced. Each node in this grid corresponds to a different combination of parameter values.

For two parameters (u, v) a topological space can be built. As (u, v) changes, these spaces are related by inclusion

$$X(u_1, v_1) \subseteq X(u_2, v_2) \quad \text{if} \quad u_1 \leq u_2 \,\text{and}\, v_1 \leq v_2.$$

Instead of a linear sequence, a 2-dimensional grid is built, where each node (u, v) points to others that are larger in both parameters.

Then the **homology groups** $H_k(X(u, v))$ are considered at each point (u, v). Because there is a partial order, homology maps forward in all directions where the parameters increase. The birth and death of features are no longer a single event in one dimension; they must be tracked across this a 2-dimensional poset, making it more complex than one-parameter barcodes.

In single-parameter persistence, barcodes or persistence diagrams summarize the lifetimes of homology classes; however, on two or more parameters there is no simple barcode. Instead, [2] introduces **rank invariants**, where a rank invariance captures how many homology classes persist throughout a geometric shape in the parameter space.

These invariants give partial information about when and where in parameter space homology classes appear and how they persist.

2.1 Applying TDA to the Cosmic Web

The ability of TDA to identify connected components, loops, and voids has made it an essential tool in astrophysics, where understanding the large-scale structure of the universe requires detecting complex geometric patterns in massive datasets. Just as TDA has been used to extract meaningful features from general point cloud data, it has also been applied to

studying the cosmic web, the vast, interconnected network of galaxies, filaments, and voids that define the large-scale matter distribution of the universe.

By leveraging persistent homology, researchers have quantified cosmic structures by analyzing how topological features such as connected components, loops, and voids evolve across different scales. As noted in [11], this approach has enabled the identification of filamentary structures and underdense regions, allowing astrophysicists to compare observations with cosmological simulations. Specifically, Betti numbers have been used to interpret these structures whereby connected components correspond to galaxy clusters, loops trace closed networks of filaments, and large low-density regions align with cosmic voids.

In [10], the topological evolution of the matter distribution in the Universe was studied by analyzing how changes in Betti numbers across a filtration can distinguish matter distributions arising from different dark energy models. Additionally, in [7] a multiscale topological measurement of the cosmic matter distribution was done by exploring the analysis of Betti numbers and topological persistence of different cosmological models.

In [8], the authors combined the Diffuse Infrared Background Experiment (DIRBE)'s well-calibrated large-beam data with Infrared Astronomical Satellite (IRAS)'s higher-resolution but more limited calibration to create high-resolution, uniformly calibrated dust emission maps. They correlated the infrared emission with known optical extinction, finding a consistent relationship between the far-infrared brightness and the column density of dust. Astronomers can determine how much dust dims or reddens light from stars or galaxies, allowing more accurate measurement of intrinsic brightness or spectral shapes. Furthermore, these comprehensive dust emission maps are not only crucial for estimating the reddening of starlight but also for identifying and subtracting Galactic foreground contamination in cosmic microwave background (CMB) observations.

Prior observations have shown that dust can be organized into numerous discrete clumps that quickly merge into filamentary and, at times, loop-like structures. In [6], isolated gas clouds and cavities are reported; in [5], partial and full dust loops are identified; and in [12], large-scale filaments aligned with the Galactic magnetic field are mapped.

Incorporating the local dust extinction rates that are derived from these maps as another dimension in TDA could enable the study of how spatial structures and dust affect observed features. Introducing an extinction rate parameter into a three-dimensional spatial domain will lead to a two-parameter setup in TDA.

3 Method

My approach comprises five main phases: Data Acquisition, Data Preprocessing, Simplicial Complex Construction, Persistent Homology Analysis, and Interpretation and Validation. The following subsections detail each phase and the corresponding code used to execute these steps.

3.1 Data Acquisition

The primary data source for this study is [9], which provides both spectroscopic and photometric observations for millions of astronomical objects, including galaxies and quasars. In addition, Schlegel, Finkbeiner, and Davis (SFD) dust maps [8] are used to quantify the amount of dust extinction, expressed in terms of $E(B-V)$, where E is a function of B and V representing the color excess caused by interstellar dust, measuring the the amount of reddening due to interstellar dust, which scatters and absorbs shorter (bluer) wavelengths more than longer (redder) wavelengths. B and V are the magnitudes of an astronomical object in the blue (B) and visual (V) filters. These SFD maps combine high-resolution infrared observations from the Infrared Astronomical Satellite (IRAS) with the well-calibrated Diffuse Infrared Background Experiment (DIRBE) data, producing uniformly calibrated maps of dust emission and, by extension, estimated extinction.

3.2 Data Preprocessing

Once acquired, the data undergo several preprocessing steps.

A subset of the sky is selected in the Sloan Digital Sky Survey (SDSS) by defining right ascension (RA) and declination (Dec) boundaries to focus on a region. Each coordinate is then associated with an average or integrated extinction value from the SFD maps.

For each coordinate in the chosen region, a feature vector is constructed. The vector includes (RA, Dec, distance) and the corresponding dust extinction value, $E(B-V)$. This transforms the dataset into a point cloud, where each data point represents a location in space endowed with a measured extinction rate.

So, the data cloud lives in a $4D$ space with each point being defined by

$$(RA,\ Dec,\ \text{distance},\ E(B-V)).$$

3.3 Simplicial Complex Construction

The point cloud dataset is then used to build a simplicial complex. The Vietoris-Rips Complex construction will be used to connect all pairs of points within distance ϵ of each other, forming a k-simplex if every pair among the $k+1$ points is within that distance.

Because an appropriate choice of ϵ, the distance threshold, is crucial, a filtration is employed to examine how topological features emerge and vanish as ϵ varies from small to large. The outcome is a sequence of simplicial complexes, each reflecting different scales of connectivity in the data.

This method was used for the 3-dimensional spatial set. It was originally intended to be used for both the 3-dimensional spatial dataset and the full 4-dimensional spatial-plus-extinction dataset, but due to complications an alternative method was used for the 4-dimensional data set. This is further explained in Section 3.6.3.

3.4 Persistent Homology Analysis

Next, persistent homology is performed on the constructed complexes to identify stable topological features.

Since extinction data inherently involve both spatial and intensity dimensions, a two-parameter persistence strategy will be applied. Here, each simplicial complex depends on two parameters, the spatial threshold ϵ_s and extinction threshold ϵ_e, forming the multifiltration as described in [2]. Topological features are then tracked across this two-dimensional parameter space, using a heatmap of Betti numbers as an approximation of rank invariants in place of simpler one-parameter barcodes or persistence diagrams.

The output will be a heatmap displaying the differing number of cluster (β_0), loops (β_1), and voids (β_2) that result when changing the spatial and extinction threshold, ϵ_s and ϵ_e. These outputs highlight which topological features persist across scales, allowing us to infer which structures are robust in the dust distribution and which may be artifacts of noise or parameter choice.

3.5 Interpretation and Validation

Finally, the results of the persistent homology analysis are interpreted in an astronomical context.

High-persistence features in β_0, β_1, and β_2 may correspond to known molecular clouds, filamentary dust, or voids where dust is absent.

The identified topological patterns are compared against known spiral arms, star-forming regions, or gas density maps to assess whether extinction structures align with larger-scale galactic morphology.

The TDA results are compared with standard dust models, like the SFD, to determine whether the method reveals structural insights.

Recognizing that astronomical measurements come with inherent noise and calibration errors, a range of parameter values is tested to evaluate the stability of the topological findings.

By coupling point cloud representations of extinction data with persistent homology, extended to handle multiple parameters, this methodology aims to uncover intricate topological structures in dust distributions. If successful, the approach has the potential to shed new light on how interstellar dust shapes our view of the universe, ultimately aiding in refining observations and improving our understanding of galactic morphology.

3.6 Implementation Details and Code Overview

The above phases are implemented and automated through a series of Python scripts. Each script corresponds to specific analysis or data-processing tasks, as outlined below.

These files use information from the returned .CSV file from the SDSS after running the SQL query in the `SDSS_search.sql` file.

The Python project for this project can be found here[1].

3.6.1 variation_heat_map.py

`variation_heat_map.py` performs an exploratory analysis of spatial variations in galactic extinction using SDSS data. After loading and cleaning the dataset including right ascension, declination, redshift, and r-band extinction (the column `extinction_r` represents the amount of extinction in magnitudes in the red optical band due to interstellar dust), it generates four complementary visualizations. First, a log-normalized scatter plot displays the spatial distribution of extinction values across the RA–Dec plane, enhancing contrast in low-intensity regions. Second, a hexbin plot reveals both spatial density and mean extinction trends. Third, a histogram characterizes the overall distribution of `extinction_r` values. Finally, the script bins `extinction_r` and computes the median redshift in each bin.

The resulting outputs are seen in Figures 7, 8, 9, and 10, respectively.

3.6.2 tda_spatial_gudhi

`tda_spatial_gudhi.py` performs a preliminary topological data analysis (TDA) on the spatial distribution of galaxies, using only the right ascension, declination, and redshift dimensions. This analysis intentionally omits `extinction_r`, which is reserved for inclusion in the subsequent multiparameter persistent homology analysis with RIVET. First, the script produces a 3D scatter plot of the spatial distribution of galaxies is generated to provide geometric context as seen in Figure 11. The script constructs a Rips complex from a spatial point cloud sampled from the SDSS dataset, computes persistent homology up to dimension 2, and visualizes both the persistence diagram and barcode which can be seen in Figures 12 and 13.

3.6.3 tda_bifilter_points_generator.py

Originally, `tda_bifilter_points_generator.py` prepared the dataset for multiparameter persistent homology analysis using the multidimensional persistent homology software, RIVET[2]. The script loads and cleans spatial and extinction data from the SDSS dataset, then extracts both spatial coordinates (right ascension, declination, and redshift) and r-band extinction values. However, due to technical limitations encountered while employing RIVET a customized Python-based approach using the Gudhi library was developed to approximate bifiltration visualizations.

A dataset comprising 10,000 randomly selected points from the Sloan Digital Sky Survey (SDSS) was first prepared by associating each point with spatial coordinates (right ascension, declination, redshift) and corresponding extinction measurements.

To construct a bifiltration, each point was assigned two filtration parameters: the extinction value and the spatial distance to its nearest neighbor. These values were written to an

[1] TDA_galactic_extinction_r Github repo

[2] RIVET

input file in the required bifiltration format to be used in `bifiltration_analysis.py`.

In this following script we step through a two-dimensional grid of parameter pairs, one axis for the extinction threshold, the other for the spatial-distance threshold. For each grid-point we build the corresponding Vietoris–Rips complex and record the usual Betti numbers $\beta_0, \beta_1, \beta_2$. Treating the pair $(\text{distance}, \text{extinction})$ as a single location in the grid turns these Betti numbers into functions of two variables; plotting their values across the grid yields the heat-maps shown in Section C.3, which visualise how connected components, loops, and potential cavities evolve jointly with both thresholds.

3.6.4 `bifiltration_analysis.py`

This file was then used to define a two-dimensional parameter grid, spanning extinction thresholds and nearest-neighbor distance thresholds, effectively forming a discretized multi-filtration. At each combination of parameters, the subset of points satisfying both thresholds was extracted, and a Vietoris–Rips simplicial complex was constructed. Persistent homology was computed up to dimension two, with Betti numbers β_0, β_1, and β_2 recorded for each parameter combination. Subsequently, heatmaps were generated for each Betti number, visually capturing the evolution of topological features across the multifiltration parameter space. While this method is less rigorous than established software like RIVET, it provides an effective visualization for identifying persistent structural features in the galactic extinction dataset.

4 Results

The results of the multidimensional persistent homology analysis are shown below, presenting heatmaps for Betti numbers $\beta_0, \beta_1, \beta_2$ as functions of two parameters: a spatial nearest neighbor threshold and an extinction threshold. These heatmaps offer insight into how extinction structures within the galactic dataset vary spatially.

For β_0, representing the number of connected components, the heatmap shows a rapid decrease as the nearest neighbor threshold increases. Initially, at low spatial thresholds, numerous distinct components are observed, reflecting isolated clusters of points characterized by similar extinction rates. As the spatial scale broadens, there is a sharp decline in the number of connected components, indicating that these extinction-rich regions merge quickly into fewer larger clusters. This result strongly suggests a high degree of spatial clustering in extinction values, pointing to regions in the dataset where points with comparable extinction measurements are closely grouped.

Examining the β_1 heatmap, which quantifies the loops or cycles within the dataset, distinct patterns emerge at intermediate scales of spatial proximity and extinction thresholds. The appearance of loops predominantly occurs within specific ranges, approximately 0.3 to 0.5 for the spatial threshold and around 0.1 to 0.18 for the extinction threshold. These loops likely correspond to ring-like or cyclical patterns within the spatial distribution of extinction. The presence of such features implies that the cosmic dust distributions do not merely cluster

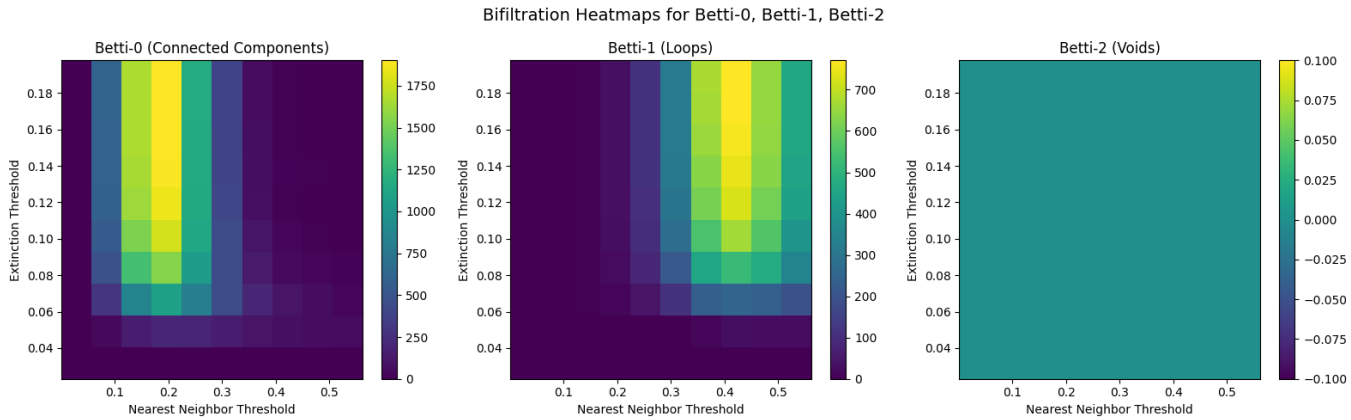

Figure 1: Bifiltration Betti Heatmap

but also exhibit more intricate geometric arrangements, possibly resembling structures such as filaments or ring-shaped plumes commonly observed in astrophysical contexts.

Finally, the heatmap for β_2, which would indicate the presence of enclosed voids or cavities, shows a uniform pattern across all parameter ranges examined, revealing the absence of significant void structures in the dataset. This suggests that within the investigated parameter space, extinction measurements form continuous distributions without isolated regions of substantially lower extinction fully enclosed by higher-extinction surroundings. This result was expected given that the data reside in a two-dimensional space (for reasons described in Sections 3.6.3 and 3.6.4) and was included primarily as a sanity check.

The methods above were again applied to two other sections of the sky and resulted in similar results as seen in Figures 14 and 15.

Collectively, these results highlight significant spatial dependence and complexity in galactic extinction distributions. The patterns observed, particularly the pronounced clustering, β_0, and loop structures, β_1, indicate that extinction rates are spatially structured in a manner consistent with astrophysical processes that give rise to complex spatial arrangements.

The broad hierarchy we recover show a rapid fall-off in β_0 signaling the merger of small extinction clumps, followed by a narrow ridge of elevated β_1 that marks the scale at which filament segments wrap into loops. This closely parallels the morphology inferred from the dust maps seen in [4], [5], and [12]. Consequently, even though our analysis relies on a different dataset and a purely topological pipeline, the bifiltration appears to recover the same clump-to-filament and occasional loop organization that observers have already reported for the interstellar medium.

5 Discussion and Future Work

The analysis presented demonstrates that multidimensional persistent homology identifies complex spatial structures in galactic extinction data, particularly emphasizing connected

clusters and loop-like formations. While the cluster and loop patterns we recover resemble filamentary and ring-shaped dust lanes reported in infrared surveys, we treat the match as qualitative only; confirming a physical correspondence will require higher-resolution data, full two-parameter rank invariant analysis, and direct comparison with 3-D dust reconstructions.

Future work will also extend the same physical comparison workflow described above to multiple, widely separated sky fields, expanding the survey footprint while experimenting with alternative distance metrics. By juxtaposing the topological summaries from each field with published 3-D dust reconstructions, we can test whether the cluster-and-loop motif is a local peculiarity or a galaxy-wide signature. Additionally, future analyses could involve more extensive variation of parameter radii, examining how larger radius settings influence the emergence of large-scale or complex topological features. Furthermore, exploring finer radius adjustments within narrower parameter ranges could yield clearer interpretations of the patterns observed in the Betti heatmaps, providing deeper insights into the underlying spatial structure of extinction distributions.

Potentially, future work will include expanding the dimensionality of the analysis, integrating additional observational parameters, and applying this methodology to broader astronomical datasets to further validate and refine the approach.

6 Conclusion

This research shows that a multidimensional persistence homology workflow can be applied to extinction data, and the resulting visualisations are compatible with known dust structures; however, the analysis is not yet precise enough to draw astrophysical conclusions. The main contribution is therefore methodological as it maps out where standard TDA tools fit into the astronomer's toolkit and where more specialised software or richer data are needed. With those enhancements in place, multidimensional persistence homology could become a quantitative complement to existing dust-mapping techniques.

7 Acknowledgements

I would like to express my sincere gratitude to Dr. Peter Muller for his steady encouragement and technical advice at every stage of this project; his feedback and commitment often turned obstacles into opportunities for clearer understanding and interesting solutions. I am also deeply indebted to Dr. Klaus Volpert for guiding me through the thickets of topology, suggesting further readings that broadened my perspective, and generously allowing me to bounce half-formed ideas off him whenever inspiration (or confusion) struck. Their combined support made this work possible.

As this project marks the end of my graduate studies, I would also like to thank my friends and family for their constant love, support, and for occasionally dragging me into much-needed diversions — especially Natalie, whose belief in me and unwavering encouragement have meant more than words can express.

References

[1] G. Carlsson, *Topology and data*, Bulletin of the American Mathematical Society, 46 (2009), pp. 255–308.

[2] G. Carlsson and A. Zomorodian, *The theory of multidimensional persistence*, Discrete and Computational Geometry, 42 (2007), pp. 71–93.

[3] R. Ghrist, *Barcodes: The persistent topology of data*, Bulletin of the American Mathematical Society, 45 (2007), pp. 61–75.

[4] M. Juvela, I. Ristorcelli, L. Montier, D. Marshall, V.-M. Pelkonen, J. Malinen, N. Ysard, L. Toth, J. Harju, J. Bernard, N. Schneider, E. Verebélyi, L. Anderson, P. Andre, M. Giard, O. Krause, K. Lehtinen, J. Macias-Perez, P. Martin, and A. Zavagno, *Galactic cold cores*, http://dx.doi.org/10.1051/0004-6361/201014619, 527 (2010).

[5] V. Könyves, D. Ward-Thompson, Y. Shimajiri, P. Palmeirim, and P. André, *A low-mass hub-filament with double centre revealed in ngc2071-north*, Monthly Notices of the Royal Astronomical Society, 520 (2023).

[6] I. Makarenko, A. Shukurov, R. Henderson, L. F. S. Rodrigues, P. Bushby, and A. Fletcher, *Topological signatures of interstellar magnetic fields – i. betti numbers and persistence diagrams*, Monthly Notices of the Royal Astronomical Society, 475 (2018), pp. 1843–1858.

[7] P. Pranav, H. Edelsbrunner, R. Van de Weygaert, G. Vegter, M. Kerber, B. J. Jones, and M. Wintraecken, *The topology of the cosmic web in terms of persistent betti numbers*, Monthly Notices of the Royal Astronomical Society, 465 (2017), pp. 4281–4310.

[8] D. J. Schlegel, D. P. Finkbeiner, and M. Davis, *Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds*, The Astrophysical Journal, 500 (1998), p. 525.

[9] Sloan Digital Sky Survey, *Sloan digital sky survey website*. https://www.sdss.org/, 2025. Accessed: April 3, 2025.

[10] R. van de Weygaert and W. Schaap, *The cosmic web: Geometric analysis*, Data analysis in cosmology, (2009), pp. 291–413.

[11] X. Xu, J. Cisewski-Kehe, S. B. Green, and D. Nagai, *Finding cosmic voids and filament loops using topological data analysis*, Astronomy and Computing, 27 (2019), pp. 34–52.

[12] S. Zaroubi, V. Jelić, A. de Bruyn, F. Boulanger, A. Bracco, R. Kooistra, M. Alves, M. Brentjens, K. Ferrière, T. Ghosh, L. Koopmans, F. Levrier, M.-A. Miville-Deschenes, L. Montier, V. Pandey, and J. Soler, *Galactic interstellar filaments as probed by lofar and planck*, Monthly Notices of the Royal Astronomical Society: Letters, 454 (2015).

A Overview of Topological Data Analysis

The following definitions and information were taken from [1].

- **Homotopic** Two continuous maps $f, g : X \to Y$ are homotopic if there exists a continuous function $H : X \times [0,1] \to Y$ such that $H(x,0) = f(x)$ and $H(x,1) = g(x)$.
- $\mathbf{H_k(X, A)}$ Given a topological space X, an abelian group A, and an integer $k \geq 0$, we associate a group $H_k(X, A)$.
- **Functoriality** For any A and k as above, and any continuous map $f : X \to Y$, there is an induced homomorphism $H_k(f, A) : H_k(X, A) \to H_k(Y, A)$. This satisfies $H_k(f \circ g, A) = H_k(f, A) \circ H_k(g, A)$ and $H_k(\mathrm{Id}_X, A) = \mathrm{Id}_{H_k(X,A)}$. These conditions define functoriality.
- **Homotopy Invariance** If f and g are homotopic, then $H_k(f, A) = H_k(g, A)$. As a result, if X and Y are homotopy equivalent, then $H_k(X, A)$ and $H_k(Y, A)$ are isomorphic.
- **Normalization** $H_0(*, A) \cong A$, where $*$ represents a single-point space.
- **Betti Numbers** For any field F, $H_k(X, F)$ forms a vector space over F. If this space is finite-dimensional, its dimension is denoted $\beta_k(X, F)$, called the k-th Betti number with coefficients in F. Intuitively, $\beta_k(X, F)$ counts independent k-dimensional surfaces in X. If two spaces are homotopy equivalent, they have the same Betti numbers.
- **Abstract Simplicial Complex**: A pair (V, Σ), where V is a finite set and Σ is a collection of non-empty subsets of V satisfying the condition that if $\sigma \in \Sigma$ and $\tau \subseteq \sigma$, then $\tau \in \Sigma$.

A topological space $|(V, \Sigma)|$ is associated with a simplicial complex and can be defined using a bijection $\varphi : V \to \{1, 2, \ldots, N\}$. This space is given as a subspace of $\mathbb{R}^N$, formed by the union of $c(\sigma)$, where $c(\sigma)$ is the convex hull of the set $\{e_{\varphi(s)}\}_{s \in \sigma}$ for $\sigma \in \Sigma$, with e_i denoting the i-th standard basis vector.

The use of Vietoris-Rips complexes provide methods for building simplexes from a given set of points.

To construct these simplicial complexes, it is necessary to define coverings of the space. When the space in question is a metric space, one covering is given by the family $B_\epsilon(X) = \{B_\epsilon(x)\}_{x \in X}$, for some $\epsilon > 0$. More generally, for any subset $V \subseteq X$ for which $X = B_\epsilon(v)$, for each $v \in V$ one can construct the nerve of the covering $\{B_\epsilon(v)\}_{v \in V}$.

Suppose (X, d) is a metric space and let $\epsilon > 0$ be given. The **Vietoris–Rips complex** $V_R(X, \epsilon)$ is defined as the simplicial complex whose vertices are the points of X. A finite subset $\{x_0, x_1, \ldots, x_k\} \subset X$ spans a k-simplex in $V_R(X, \epsilon)$ precisely when $d(x_i, x_j) \leq \epsilon$ for all indices i, j satisfying $0 \leq i, j \leq k$. See Section B.1 for an illustrated example.

Since different values of ϵ may produce different topological structures, a method is needed to track the evolution of these structures as ϵ varies. This leads to the concept of persistent homology.

Persistent homology postulates that given a sample set of data homological information can be obtained by avoiding a selection of a fixed value of the threshold ϵ, and instead use all possible different values of ϵ at once.

Information from persistence homology can be represented as a **barcode**, intervals on a number line that represent the data, or a **persistence diagram**, points on a two-dimensional plane where each point represents the birth and death time of a topological feature.

B Illustrated Examples

B.1 Data Point Cloud and Vietoris-Rips Complex

The following illustrations show a Vietoris-Rips Complex being applied to a point set data cloud.

In Figure 2, there is a set of data points.

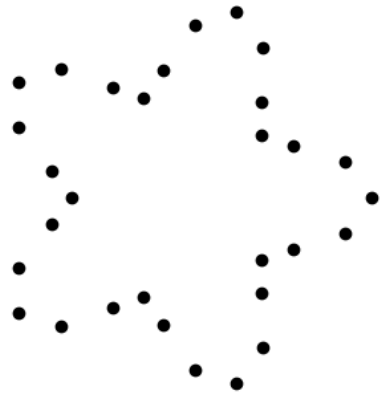

Figure 2: Point Set Data Cloud

Below in Figure 3, a circle of radius ϵ is drawn around each data point.

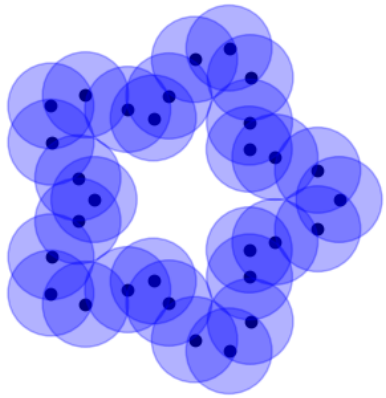

Figure 3: Point Set Data Cloud with a Given Radius ϵ

Whenever two ϵ-balls intersect, the corresponding points are connected by an edge; higher-dimensional simplices form analogously. The resulting simplicial complex is shown in Figure 4.

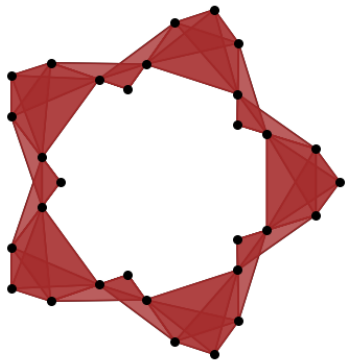

Figure 4: Corresponding Vietoris–Rips Simplicial Complex $V_R \leq (X, \epsilon)$

B.2 Betti Number

A 2-Dimensional Object with Three Holes $\beta_1 = 3$

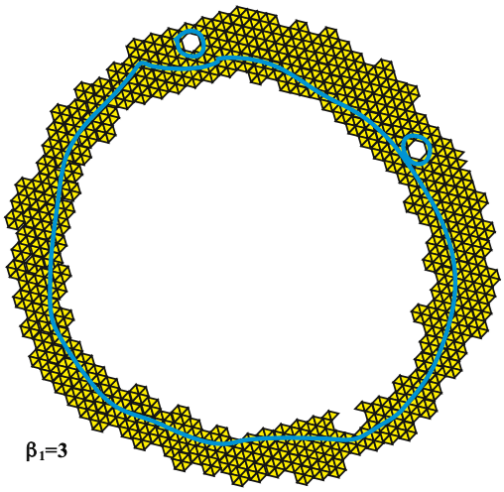

Figure 5: This illustration was taken from [1]

B.3 Barcode

Below is an example of a barcode. As the radius ϵ increases, topological features, such as connected components and loops, are created and destroyed, and their lifespans are recorded in the barcode.

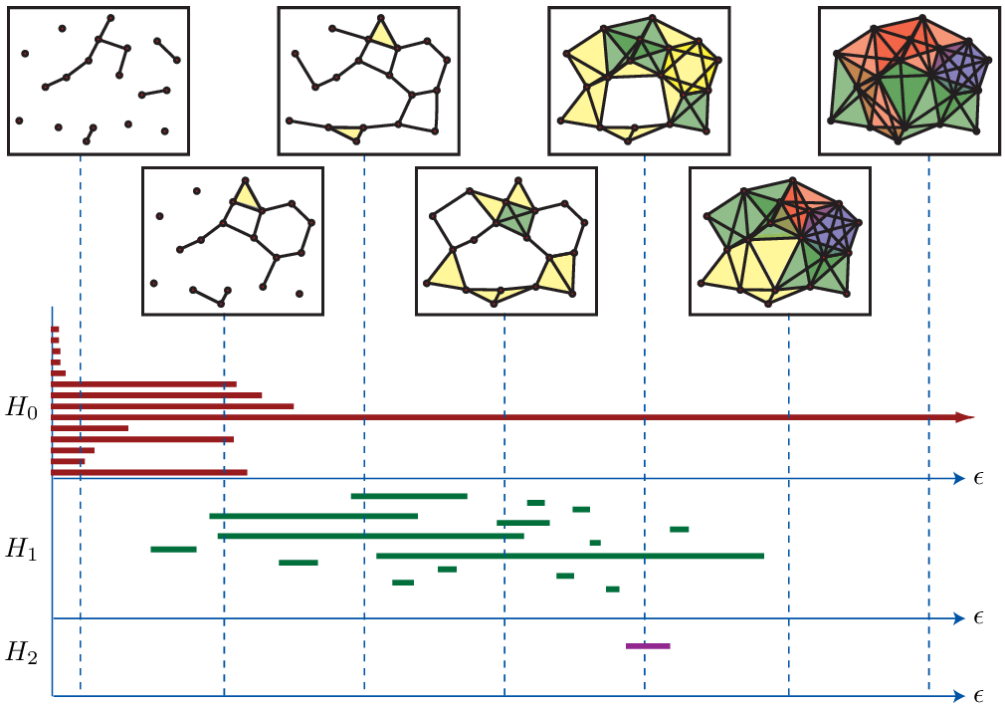

Figure 6: This illustration was taken from [3]

C Results

C.1 Initial Review

The below heat map in Figure 7 visualizes the spatial variation of galactic extinction rates across the selected region of the sky. Brighter colors indicate higher extinction values, corresponding to regions with more interstellar dust.

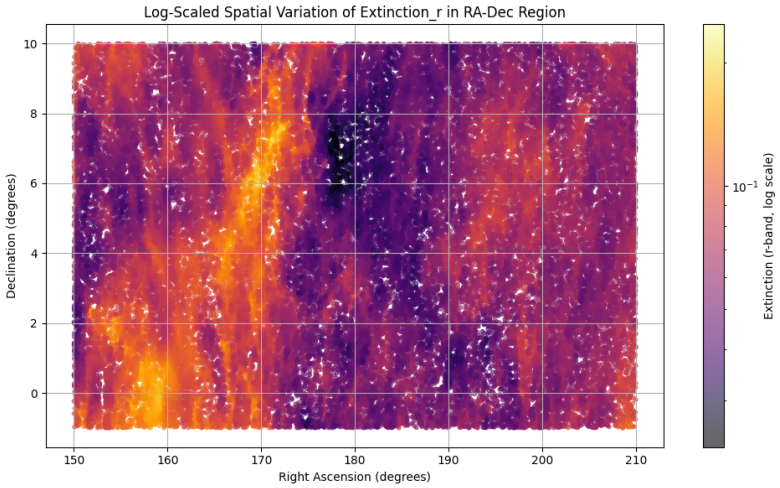

Figure 7: Extinction Rate Heat Map

The Figure 8 hexbin plot shows the spatial density of data points along with average extinction values per hexagonal bin. It highlights both clustering trends and extinction intensity variations across the region.

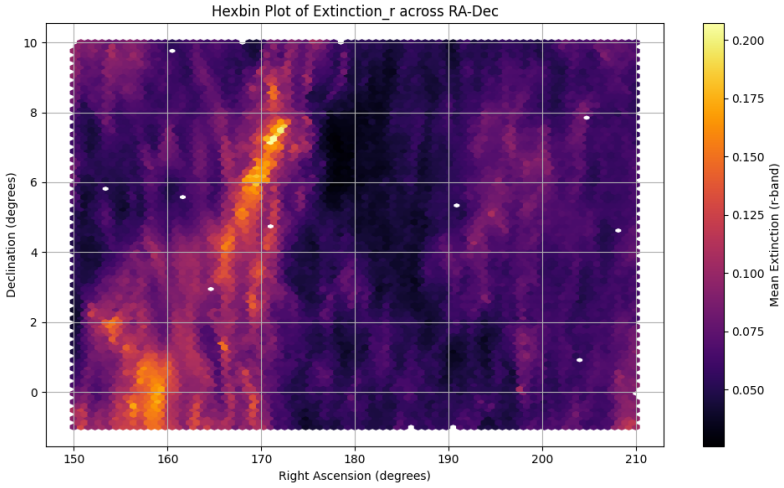

Figure 8: Extinction Rate Hexbin Plot

The below histogram presents the distribution of extinction values throughout the dataset, revealing the frequency of different extinction levels among the observed data points.

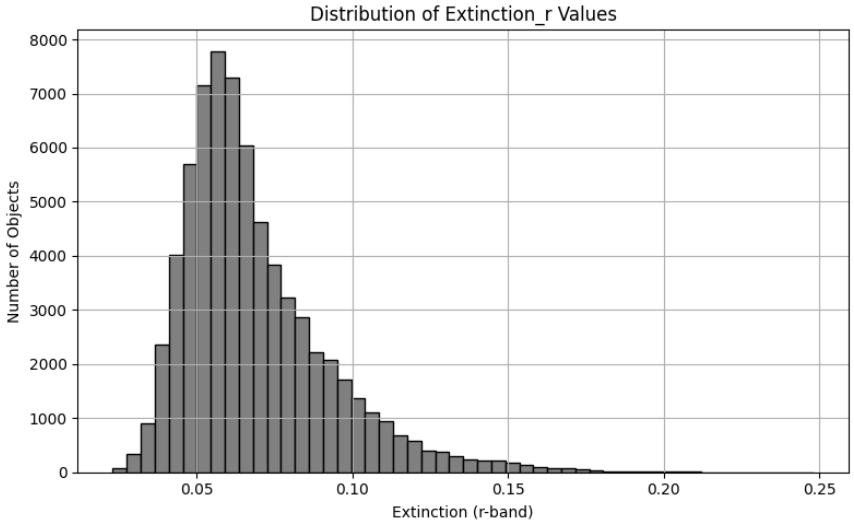

Figure 9: Extinction Rate Histogram

The below plot shows the median redshift associated with different bins of extinction values, offering insights into how extinction correlates with distance via redshift in the observed region.

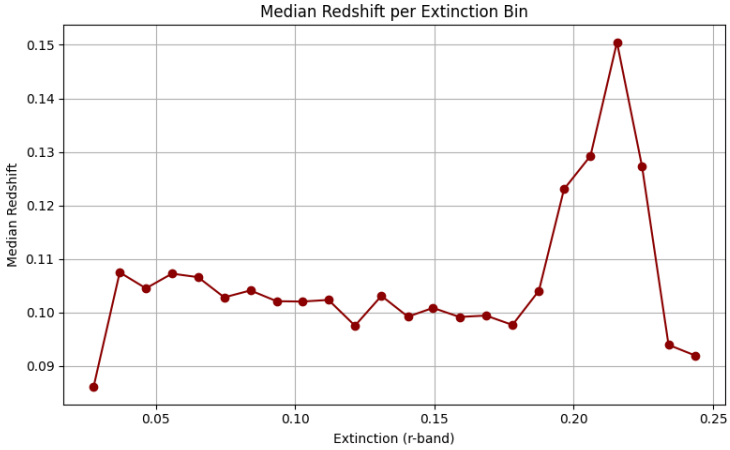

Figure 10: Median Redshift per Extinction Bin

C.2 Spatial TDA Results

This 3D scatter plot illustrates the spatial distribution of galaxies based on their right ascension, declination, and redshift coordinates, providing a geometric view of the sample before topological analysis.

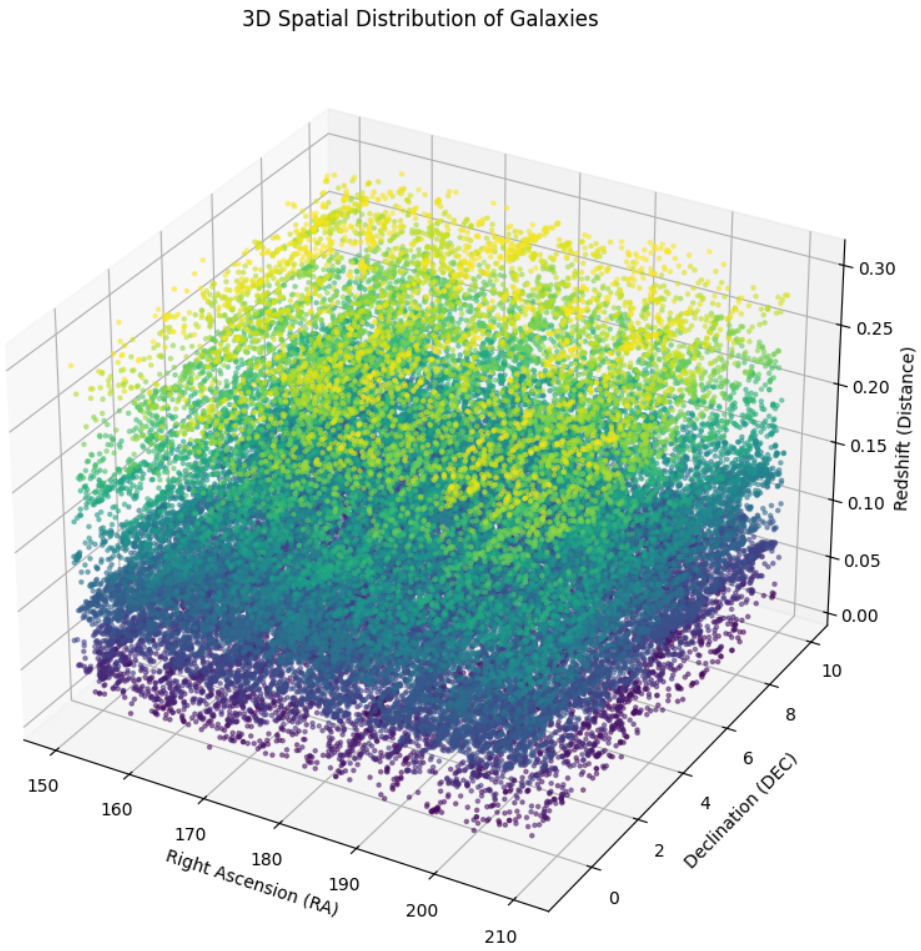

Figure 11: Spatial Distribution

The persistence diagram in Figure 12 captures the birth and death of topological features, connected components and loops, as the scale parameter varies, quantifying the dataset's spatial structure through TDA.

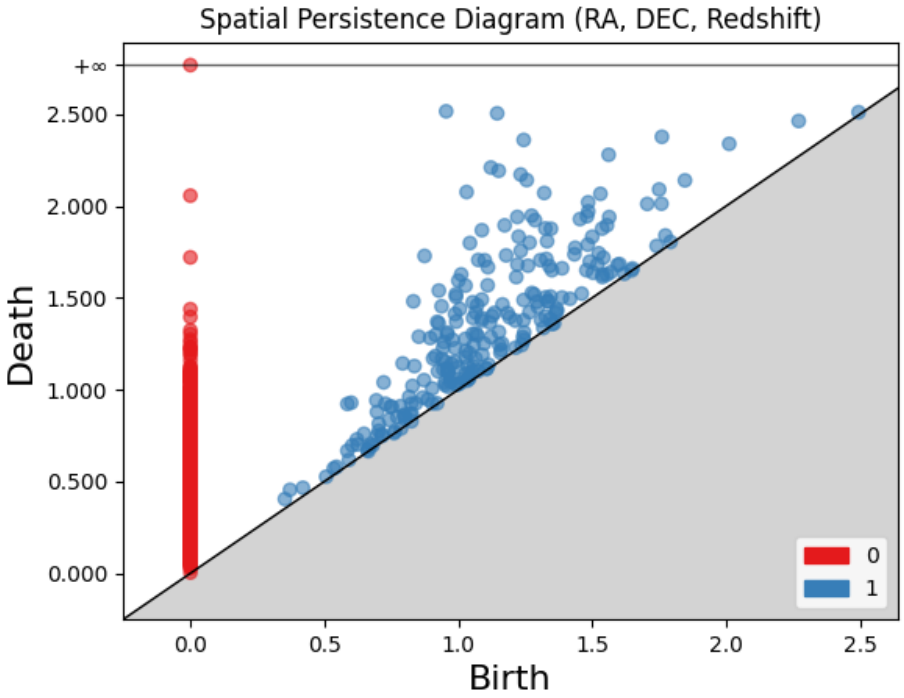

Figure 12: Spatial Persistence Diagram

The barcode in Figure 13 represents the lifespan of topological features across scales. Longer bars indicate features that persist across a wide range of scales, suggesting robust spatial structures.

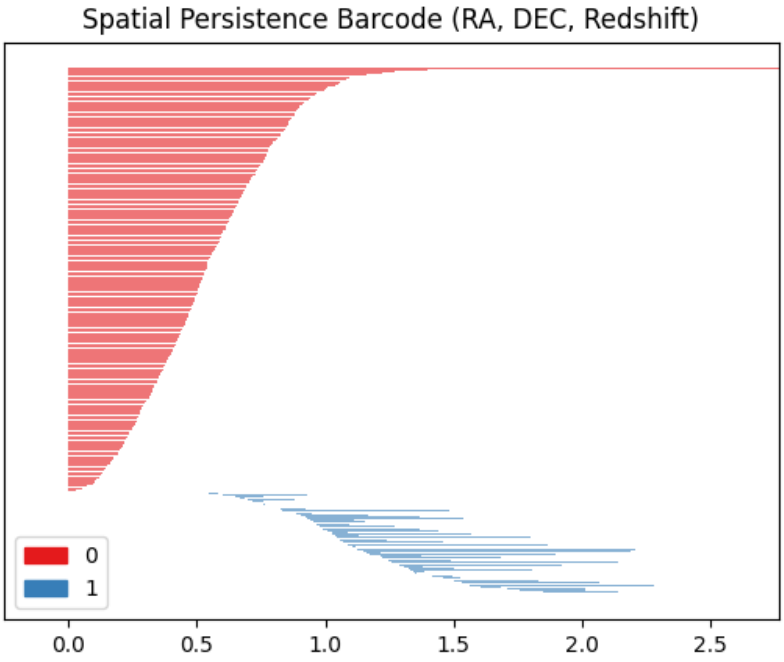

Figure 13: Spatial Persistence Barcode

C.3 Other Bifiltration Results

The bifiltration heatmap in Figure 14 visualizes the Betti numbers computed across two parameters—spatial distance and extinction threshold—for a second region of the sky, highlighting patterns of clustering and loops.

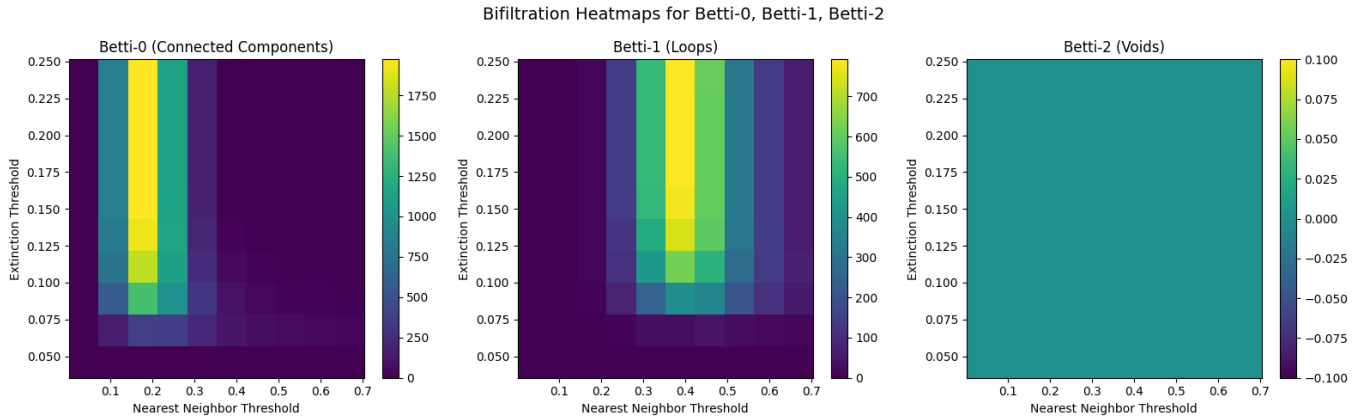

Figure 14: Bifiltration Betti Heatmap for Second SDSS Selection

Similar to the above figure, Figure 15 shows the bifiltration results for a third sky region, allowing for comparisons of topological patterns across different parts of the galaxy.

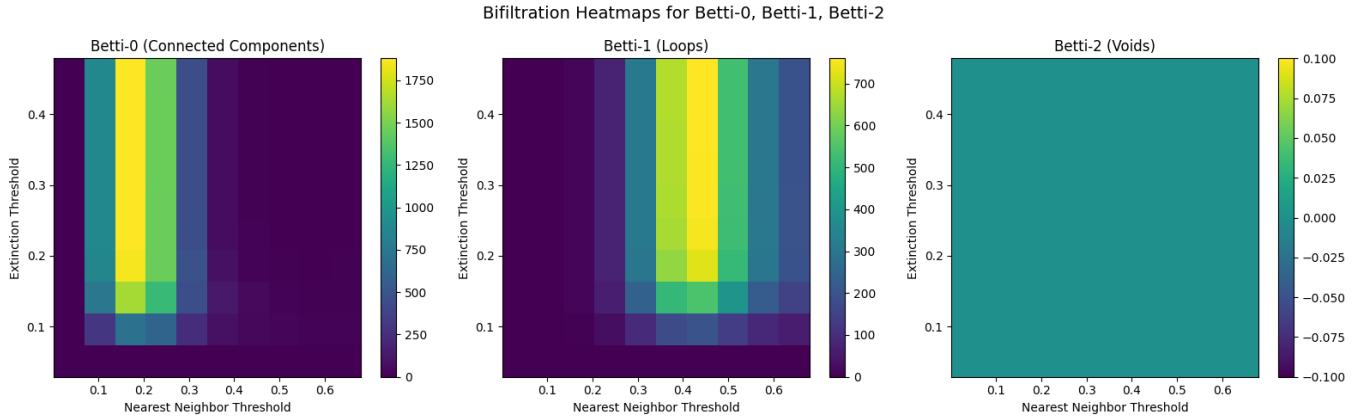

Figure 15: Bifiltration Betti Heatmap for Third SDSS Selection

Memo

To: Observational and theoretical astronomers interested in interstellar dust
From: Dylan Wolf
Subject: A topological snapshot of Galactic extinction
Date: 9 May 2025

Accurate extinction corrections underpin distance scales, stellar parameters, and extragalactic surveys. Yet existing three-dimensional dust maps disagree at fine scales and reveal little about the geometry of dusty regions. Topological Data Analysis (TDA) offers a complementary, scale-aware lens on that geometry.

TDA looks at the shape of data rather than its exact values. Imagine sprinkling dust on a sheet of paper and then slowly expanding each speck: at first you see lots of isolated specks, but as the specks get larger, some specks touch and form little islands, and eventually rings. By keeping track of when islands form and merge, and when rings appear or disappear, we get a high-level summary of the dust pattern. This provides a rough concise "shape signature" that is robust to noise.

In this study the "specks" are 10,000 galaxy data points from the Sloan Digital Sky Survey, each recorded by its right ascension, declination, redshift, and r-band extinction. In other words, $(\mathrm{RA}, \mathrm{Dec}, z, A_r) \in \mathbb{R}^4$. Each point carries its sky position and an estimate of how much dust sits in front of it. When we run the TDA workflow we find that nearby points with similar dust values tend to clump together quickly, and occasionally they wrap around to form loops. At intermediate thresholds a modest number of loops survive, hinting at ring-like or filamentary dust structures reminiscent of those seen in infrared surveys. The resemblance is encouraging but strictly qualitative: resolving a physical connection will demand higher-resolution extinction estimates, rank-invariant tools such as RIVET, and side-by-side comparison with recent 3-D dust reconstructions. This behaviour matches what astronomers already know, that dust collects in filaments and clouds, but the present analysis is only a proof of concept.

The immediate value of this work is therefore methodological. It supplies an end-to-end, open-source workflow that converts survey catalogues into interpretable topological summaries—runnable in minutes on a laptop. The next stage is to replicate the pipeline across multiple sky fields, test its sensitivity to parameter choices, and integrate richer photometric data. If those steps confirm that TDA can reliably flag filaments, voids, or shells in extinction space, the technique could become a quantitative companion to established dust-mapping efforts and help tighten the uncertainties that ripple through modern astronomy. To further turn the idea into a practical tool, future work will need better software, more data, and careful cross-checks with physical dust models.