Homology-Separating Triangulated Euler Characteristic Curve

Nicholas O. Malott, Robert R. Lewis, and Philip A. Wilsey
Dept. of EECS, University of Cincinnati, Cincinnati, OH 45221, USA
Email: malottno@mail.uc.edu, lewis3rt@mail.uc.edu, wilseypa@gmail.com

***Abstract*—Topological Data Analysis (TDA)** **utilizes concepts from topology to analyze data. In general, TDA considers objects similar based on a topological invariant. Topological invariants are properties of the topological space that are homeomorphic; resilient to deformation in the space. The *Euler-Poincaré Characteristic* is a classic topological invariant that represents the alternating sum of the vertices, edges, faces, and higher-order cells of a closed surface. Tracking the Euler characteristic over a topological filtration produces an *Euler Characteristic Curve (ECC)*. This study introduces a computational technique to determine the ECC of $\mathbb{R}^2$ or $\mathbb{R}^3$ data; the technique generalizes to higher dimensions. This technique separates landscapes of lower-order homologies utilizing triangulations of the space.**

***Index Terms*—Euler Characteristic Curve, Topological Data Analysis, Data Mining**

I. Introduction

Topological Data Analysis (TDA) treats a point cloud as a sampling of a *topological space* and characterizes the data based on its topological invariants [1]. *Topological invariants* are measurable properties of topological spaces that are unique up to homeomorphism. Some well-known topological invariants include the cardinality, connectedness, compactness, countability conditions, and the *Euler Characteristic* [2]. The *Euler-Poincaré Characteristic* (EC), generalizes Euler's equation to higher dimensions and is defined as:

$$\chi(K) = V - E + F - C_3 + C_4 ... = \sum_{k=0}^{\infty} (-1)^k C_k. \quad (1)$$

For consistency with the higher-order cells characterizations, substitutions are made for the traditional variables: vertices ($C_0 = V$), edges ($C_1 = E$), and faces ($C_2 = F$).

One method to examine the EC is to perform a filtration, such as the sequence of topological spaces created as connecting proximity is increased. Recording the EC from each topological space creates a trend of the EC against proximity filtration, commonly referred to as the *Euler Characteristic Curve (ECC)* [3]. The ECC captures global topological structure in a manner directly related to *Persistent Homology (PH)*. PH is a well-known and invaluable tool for TDA, however, its exponential space and time complexity inhibits its application to high-dimensional data and big data applications [4]–[6].

In this study ECC is characterized alongside PH [5], [7] to identify applications where the ECC can perform fast computations to complement results from PH. Unlike PH, ECC can capture topological invariant information using favorable memory and modest run-time complexities. A synthetic analysis of $\mathbb{R}^2$ and $\mathbb{R}^3$ shapes is provided in this paper to demonstrate the results of the approach. A brief correlation analysis against $\mathbb{R}^3$ MRA Brain Artery trees is also included.

The remainder of this paper is organized as follows. Section II contains background on TDA and topological invariants. Section III identifies related studies and characterizations of ECCs. Section IV introduces the triangulated approach for ECC and dimensional homology separation. Section V presents experimental results. Finally, Section VI remarks on the experimental study and future work with ECC.

Support for this work was provided in part by the National Science Foundation under grant IIS–1909096.

II. Background

Formally, ECC and PH both examine the formation and collapse of topological features (connected components, loops, voids, and their higher dimensional analogues) over a filtration. One popular filtration is to utilize the proximity between points, increasingly adding edges, faces, and higher order components to characterize the *persistence* of topological features. Formally, a filtration can be expressed as a filtered complex, K_F that represents a nested sequence of graphs:

$$\emptyset = K_0 \subseteq K_1 \subseteq ... \subseteq K_F. \quad (2)$$

At any of these filtration spaces the EC (Equation 1) can be shown to be the alternating sum of the existent topological features, denoted as the Betti (β) count at that filtration [8]:

$$\chi(K_j) = \beta_0 - \beta_1 + \beta_2 - \beta_3 + ... = \sum_{k=0}^{\infty} (-1)^k \beta_k. \quad (3)$$

So $\chi(K_j)$ represents the EC of an arbitrary topological space.

A filtration of the point cloud creates an ordered set of topological spaces related to the connectivity distances, $\epsilon = (\epsilon_0, \epsilon_1, \cdots, \epsilon_\infty)$ s.t. $\forall i < j, \epsilon_i < \epsilon_j$, used to create the space. The EC can be evaluated for each member of the filtration to compute: (i) the alternating sum of the Betti numbers, or (ii) the alternating sum of representational components.

A *simplicial complex* contains dimensional representations of the point cloud called simplices. Filtration of the simplices based on their weights defines a nested sequence of graphs. The filtration can be defined through various functions comparing paired samples of the data; however, we will focus on using the proximity between points to compute the weights,

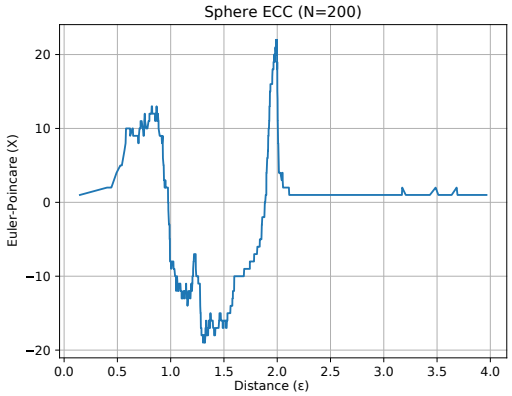

Fig. 1: The ECC for a sphere in $\mathbb{R}^3$.

such as that used for fast construction of the Vietoris-Rips complex [9]. With this filtration, we can rewrite the ECC as:

$$ECC(K) = [(\epsilon_i, \sum_{d=0}^{\infty}(-1)^d |K_{\epsilon_i,d}|), \quad \text{for } \epsilon_i \in \epsilon, \qquad (4)$$

where $|K_{\epsilon_i,d}|$ is the count of d-simplices existent in the filtered complex, K_{ϵ_i}. Figure 1 plots the ECC of the surface of a sphere; the x-axis represents the filtration values (ϵ) and y-axis measures the EC at that ϵ distance.

III. Related Work

The Euler-Poincaré Characteristic has been studied through many domain perspectives. A brief history of the EC and an overview of several famous applications are examined in Richeson [2]. Various scientific applications including neuroscience, thermodynamics, cosmology, random fields, and material science are detailed by Smith and Zavala [10].

The remainder of this section discusses work related directly to TDA with the EC; these approaches typically mine through filtrations of some domain data. Richardson and Werman [3] demonstrate the use on images and 3D mesh surfaces. Bobrowski and Skraba [11] examine the EC for detecting homological percolation. Beltramo *et al* [12] examine Euler Characteristic Surfaces, multi-parameter filtrations used for image analysis.

The Euler Characteristic Transform (ECT) examines the ECC in multiple directions. Turner *et al* [13] prove the collection of directional curves can summarize $\mathbb{R}^3$ shapes utilizing H_0 and H_1 persistence diagrams. Ghrist *et al* [15] reformulate the ECT in terms of Euler calculus on constructible functions. Curry *et al* [16] prove that any shape can be uniquely determined with the ECT using a finite number of directions. Multi-parameter ECTs have also been of recent interest in the TDA community.

Crawford *et al* [8] introduce the Smooth Euler Characteristic Curve (SECT), a smoothing function on the ECT for integrating shape information into machine learning models. The SECT is utilized to predict clinical outcomes in glioblastoma by quantifying MRI images of tumors.

IV. Overview of the Approach

In this section the direct relation between the ECC and PH is characterized. In addition, an alternative method is proposed to compute and detect higher-order homologies within a point cloud using properties of triangulation and the ECC. While few scenarios are known where higher-order homologies are salient, the potential to characterize and identify cases are limited by the computational complexity of PH [17]. The homology-separated ECC, a fast metric to compute, can alleviate these limitations.

While triangulation of a point cloud has been previously studied for computing ECCs over a filtration of data [3], [8], [11], [12], [18], a careful examination of the change of homology alongside a sequential insertion process can reveal neighborhood topological information not captured by the EC alone. This approach is demonstrated for $\mathbb{R}^2$ in Section IV-B and for $\mathbb{R}^3$ in Section IV-C.

A. ECC and PH

A connection of ECC with PH is necessary to qualify ECC as a compatible estimator. Section II provides a brief introduction to the use of the ECC with a topological filtration. The extension of this definition for the proximity-based triangulated ECC is covered here as preliminary material.

Recall that the EC represents the alternating sum of dimensional components in a space (Equation 1). Computing the ECC from a triangulation is one method to associate the input point cloud to a simplicial complex. Using the maximal proximity distance ϵ between any two pairs of points in the simplex yields a proximity filtration. The EC is recorded at insertion of each triangle, in sequence by weight from smallest to largest, to obtain the ECC (Equation 4); these curves capture a proximity-series representation of the EC of a discrete point cloud. Many applications have been built on the ECC due to its significance as a topological invariant. While result analysis is beyond the scope of this study, there is considerable interest in understanding how the EC can be paired with recent advances in data mining and machine learning applications [10], [18].

Figure 1 depicts the ECC computed on points sampled from a sphere in $\mathbb{R}^3$. The curve is generated using the triangulation and proximity-based filtration described above. The positive and negative influence of the dimensional components interact across the filtration — leading to a characteristic curve that suffers from significant noise due to contributions overlapping ϵ-ranges of multiple dimensions. As discussed below, separating the contributions by dimension permits a sharper characterization of the topological structure of the point cloud.

Recall the relationship between the dimensional components and Bettis of $\chi(K_F)$. The dimensional components of K are compressed in a lossy manner to an integer; the dimensional homologies of K_F are also compressed into that same topological invariant. One might wonder: if the Bettis of $\chi(K_f)$ has topological significance, can we infer topological changes from the dimensional components through $\chi(K)$? Through strategic comparison of the triangulated shape this decomposition becomes possible.

	Relation to $K_{\epsilon_{-1}}$	$VCEC$	$\Delta(\chi)$	Adj. $\Delta(\chi)$
1	Independent	[3,3,1]	1	0
2	Shares 1 Vertex	[2,3,1]	0	0
3	Shares 2 Vertices	[1,3,1]	-1	*Case*(-1 or 0)
4	Shares 3 Vertices	[0,3,1]	-2	*Case*(-2, -1, or 0)
5	Shares 1 Edge	[1,2,1]	0	0
6	Shares 1 Edge, 1 Vertex	[0,2,1]	-1	*Case*(-2, -1, or 0)
7	Shares 2 Edges	[0,1,1]	0	0
8	Shares 3 Edges	[0,0,1]	+1	+1

TABLE I: VCEC changes for data in $\mathbb{R}^2$; raw EC changes are denoted $\Delta(\chi)$ and adjusted EC are denoted Adj. $\Delta(\chi)$

Separation of the dimensional curves can result in individual landscapes of the individual homology groups for further analysis. This approach utilizes characteristics of generalized d-tetrahedra to infer the change of homology of the space. The result is separation of high-order features. Dimensional topological landscapes provide a unique characteristic for applications of topological data analysis.

B. The Euler Characteristic Curve in $\mathbb{R}^2$

This work develops an alternate method to calculate ECC that exploits the structure of triangles in higher dimensions to separate dimensional ECCs. The approach utilizes an insertion of higher-order triangles examined over the filtration of the distance between points. This construction of ECC is presented in $\mathbb{R}^2$ to demonstrate the unique formation of H_1 loops in the space; insertion of triangles is utilized for separation. An extension to $\mathbb{R}^3$ in Section IV-C reveals the computational shortcut used in the approach.

Consider all possible combinations of adding triangles to a point cloud; insertion occurs sequentially by maximal edge weight of the triangle. Triangles always have 3 vertices, 3 edges, and 1 face. Triangles can share several, or none, of these components as they are enumerated. The sequence of insertions can be computed as the *Vector of Changes in the Euler Characteristic* (VCEC), a vector recording each subsequent change in the EC. Formally, VCEC becomes [18]:

$$\begin{aligned} VCEC_0 &= ECC_0 \\ VCEC_i &= ECC_i - ECC_{i-1}, \text{ for } 0 < i, \end{aligned} \tag{5}$$

and $VCEC_i$ becomes an enumeration of vector changes to the EC for each member of the filtration.

Instead of tracking the change in ECC at each triangle insertion, this approach tracks the component changes (vertices, edges, faces, cells) at each insertion. There are a finite number of possibilities for the component changes with only a subset impacting the topology in a filtration. The deterministic changes to the EC in $\mathbb{R}^2$ are described in Table I.

In $\mathbb{R}^2$ only homology groups H_0 and H_1 are present. There are two methods to separate these features within ECC. First, the H_0 connected components can be tracked by a UnionFind structure to identify when connected components form within the point cloud. Alternatively, the H_1 components can be computed using the VCEC dimensional changes. Either may be removed from the original ECC to compute the other:

$$\begin{aligned} ECC - ECC_0 &= ECC_1 \\ ECC - ECC_1 &= ECC_0 \end{aligned} \tag{6}$$

Counting H_0 connected components is straightforward with the UnionFind structure. Identifying H_1 loops requires more analysis. Surprisingly, formation of a loop can only happen in a general case: two vertices are shared with some connected component without an adjoining edge. This can happen with 3 vertices as well, and potentially form up to 3 H_1 loops with the insertion of one triangle. However, if any of the original components are independent from one another, no loop is formed from the adjoining edge addition.

Closure of a loop follows a similar result. The death of a loop in the filtration happens only when a triangle is inserted with only a face; all vertices and edges are constituent of a previous filtration. Change in the EC's dimensional components indicates the death of an H_1 loop (Table I, Row 8).

Counting the changes noted above is relatively simple when computing the ECC. However, some cases expose a deterministic change in the homology of the space. If the inserted triangle is independent (Table I, Row 1) and shares no components with the existing space the EC is expected to increase by 1, indicating an unconnected component (H_0) topological feature has been inserted into the space.

Adjusting the ECC to account for formation of lower order topological features is captured in Table I. Note that only 5 $\mathbb{R}^2$ cases induce a change in the EC; of these, 4 are recorded by the Adjusted characteristic. This adjustment is made by tracking the connected sets of vertices in the space, implemented through a UnionFind algorithm and a mapped lookup table for adjustments from $\Delta(C_k)$. Between the UnionFind algorithm and tracking the change in vertices, edges, and faces, all formations of H_1 loops can be tracked through the filtration, free of noise from lower dimensional components.

These observations do not hold for data in $\mathbb{R}^3$. Higher dimensional topological features begin to create noise in the EC that is not easily manipulated to characterize H_1 features in the filtration. Higher-dimensional triangulations of the space must be explored to generalize the approach to $\mathbb{R}^n$.

C. The Euler Characteristic Curve in $\mathbb{R}^3$

In $\mathbb{R}^3$ there are many combinations of tetrahedrons that can occur; however, only those that change the EC are of interest to this approach. Table II lists the Euler tetrahedron insertions that induce change in the ECC; additional combinations are omitted for brevity. The formation of H_1 features in the space must also be considered; fortunately this is achieved with the same adjusted values from Table I.

The goal of the EC in $\mathbb{R}^2$ was to separate H_0 components and H_1 loops. In $\mathbb{R}^3$, an additional focus is directed to the H_2 voids of the space. H_2 voids are encompassed by connected faces forming a surface, such as the interior of a sphere. The goal of the adjusted ECC is to only increase when the birth of a void occurs and only to decrease at the filling of the void.

	Relation to $K_{\epsilon_{-1}}$	$VCEC$	$\Delta(\chi)$	Adj. $\Delta(\chi)$
1	Independent	[4,6,4,1]	+1	0
2	Shares 2 Vertices	[2,6,4,1]	-1	*Case*(-1 or 0)
3	Shares 3 Vertices	[1,6,4,1]	-2	*Case*(-2, -1, or 0)
4	Shares 4 Vertices	[0,6,4,1]	-3	*Case*(-3, -2, -1, or 0)
5	Shares 1 Edge, 1 Vertex	[1,5,4,1]	-1	*Case*(-1 or 0)
6	Shares 1 Edge, 2 Vertices	[0,5,4,1]	-2	*Case*(-2, -1, or 0)
7	Shares 2 Edges	[0,4,4,1]	-1	*Case*(-1 or 0)
8	Shares 2 Faces, 1 Edge	[0,0,2,1]	+1	*Case*(+1 or 0)
9	Shares 4 Faces	[0,0,0,1]	-1	-1

TABLE II: VCEC changes for data in $\mathbb{R}^3$; raw EC changes are denoted $\Delta(\chi)$ and adjusted EC are denoted Adj. $\Delta(\chi)$

The simplest representation of a void is that of a tetrahedron with no internal cell, only faces encapsulating the void. The closing of the tetrahedron (*birth*) occurs when a face is inserted adding no new vertices or edges. When inserting tetrahedrons into the filtration many instances of this birth can occur. *Death* of the void occurs when a tetrahedral cell is inserted, with no faces, edges, or vertices. This mirrors the observations of $\mathbb{R}^2$.

Consider the birth and death instances of both $\mathbb{R}^2$ (edge/triangle) and $\mathbb{R}^3$ (triangle/tetrahedron). We can guarantee detection of these cases due to the strategic insertion of $(d+1)$-tetrahedra into the space. In any dimension d, the face of the representational $(d+1)$-tetrahedron creates the birth and the $(d+1)$-tetrahedron insertion with no lower-order components represents a death.

The same strategy of mapping dimensional VCEC adjustments is taken for the tetrahedral representations. We can generalize our mapping approach to arbitrary dimensions by examining the VCEC and UnionFind count. In addition, the EC components and VCEC maps can be compared across different cardinalities. H_1 loops, for example, can be identified in the $\mathbb{R}^3$ filtration by comparing the leading components when maps are stored as $[\Delta(C_0), \Delta(C_1), \Delta(C_2)]$. Selection of certain dimensions of topological features to analyze, such as only enabling H_2 adjusted ECC mappings, may provide flexibility in representation for different applications.

D. An Algorithm for computing ECC in $\mathbb{R}^n$

Extrapolating the ECC constructions in $\mathbb{R}^2$ and $\mathbb{R}^3$, we can formalize ECC separation with Algorithm 1. The algorithm inputs data in $\mathbb{R}^n$ and a parameter for the maximum dimension of homology (H_d) to compute ($d < n$)[1]. The algorithm initializes a UnionFind data structure for tracking the connected components of the space (Line 3) and creates sets for tracking each dimensional component C_k for easy duplicate removal when counting (Line 4).

The Delaunay triangulation (Line 5) is the primary bottleneck for the algorithm.[2] The ECC utilizes generalized $\mathbb{R}^{d+1}$ triangles for representing the H_d features of the space; therefore we set the triangulation to create (H_d+1)-dimensional representations. The triangulation are ordered by filtration weight (Lines 6 - 7). Cases where the weight of two triangles are equal are recorded for the same weight value in the ECC.

Algorithm 1 Separated Euler Characteristic Curve

Input: PC, a point cloud of dimension $\mathbb{R}^n$;
H_d, the dimension of homology to characterize
Output: ECC, a series of triangulated ECC values

1: **function** $ECC(\epsilon_\infty)$
2: $dmap \leftarrow Adj.\Delta(\chi)$
3: $uf \leftarrow UnionFind()$
4: $C_k \leftarrow \{\}$ for d
5: $tri_d \leftarrow Delaunay(PC, H_d+1)$
6: **for** $u \in tri_d$ **do**
7: $tri_d[u][\epsilon] \leftarrow max(dist(i,j) \forall i,j \in u)$
8: **for** $t \in ordered(tri_d)$ **do**
9: $uf.insert((i,j) \forall i,j \in t))$
10: $C_k.insert(t_k \forall k <= d)$
11: $VCEC_i = C_{k_i} - C_{k_{i-1}}$
12: $ECC.insert(\epsilon, dmap[VCEC_i] + \Delta uf)$

For each triangle, the constituent pairs of points are inserted into the UnionFind algorithm to identify any new connected components that form (Line 9). The change in connected components adjusts the ECC to ignore cases identified in Sections IV-B and IV-C. Each d-dimensional subset of the triangle is inserted into dimensional sets (Line 10). For example, each constituent vertex is inserted into the first set. If vertices already exist there is no change; if new vertices are added they are counted; likewise for edges, faces, and so on.

Finally, the ECC is computed at the filtration (Line 12). The filtration weight is recorded with the ECC value. Depending on the dimensions of homology to filter, the mappings can have variable cardinality and only compare the leading order change in the ECC components.

E. ECC Homology Separation

The ECC sequence considers topological features in all dimensions; if instead the ECC is computed from a strategic insertion of higher-dimensional components, such as tetrahedrons, lower-dimensional features can be anticipated and removed from the results. In essence, by computing a series of triangulations of the point cloud from 2 to d, the H_d homologies can be separated. This is the critical contribution of this work; separation of ECC results by homological dimension can provide insights into high-dimensional structures currently unobtainable with other topological methods such as PH.

Canceling lower order features in each of the ECs being computed create an inverted relationship:

$$ECC^d(K) \subset ECC^{d-1}(K) \subset ... \subset ECC^0(K). \quad (7)$$

Therefore, the ECC can be separated by dimension:

$$ECC^d(K) = ECC(K) - \sum_{j=0}^{d-1} (-1)^j (ECC^j(K)). \quad (8)$$

[1]The algorithm is defined for $\mathbb{R}^n$ but, since we only have ECC adjustments in $\mathbb{R}^2$ and $\mathbb{R}^3$, the application is limited to data in these dimensions.

[2]The Delaunay triangulation is used here for expediency and can be replaced with other methods to form a triangulation of the data.

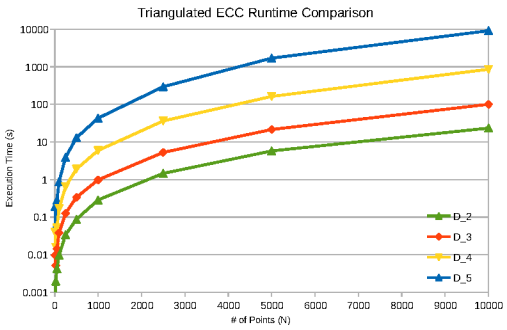

Fig. 2: Scalability of the ECC against data set size.

This briefly introduces homology separation methods for the Triangulated ECC. Additional computational techniques and salient topological data sets are necessary to begin further exploration of this relationship in higher dimensions.

V. Experimental Results

Synthetic data was generated to demonstrate the scalability of the approach in Section V-A and the separability of dimensional components in Section V-B. Real-world ECC analysis of MRA Brain Artery trees are compared to previously published PH correlation results in Section V-C. All experiments in this section were performed on an AMD Ryzen 7 3800X 8-Core Processor @ 3900MHz with 64GB of RAM.

A. Synthetic Scalability

Scalability is one of the primary benefits of this work. The ECC complexity is dominated by the triangulation of the input data; the remaining computation is a counting problem that can be easily computed with linear programming [18].

Figure 2 charts the scalability of the ECC for increasing number of points. Each curve is measuring a different dimension of synthetically generated d-spheres in $\mathbb{R}^{d+1}$. As the number of points increase the runtime grows; however the rate of growth compared to a more complex approaches, such as PH, is modest. Where the ECC approach excels is in memory complexity; PH is typically bounded by memory [4], [6].

B. Synthetic Data Analysis

Figure 3 presents the ECC plot (left), and a PH comparison (right) for a sphere (top row) and a torus (bottom row). These shapes are fundamental in Topology and are expected to emit structural differences when compared.

The ECC plot (left) is the raw ECC computed over the triangulated filtration separated using the method described in Section IV. Here the ECC, ECC_0, and ECC_1 are computed to compute ECC_2. This reduces the cost of determining the ECC at the highest dimension.

The right plots of Figure 3 overlay the $ECC_1 + ECC_2$ curves with the H_1 and H_2 barcode results from PH. The barcode results do not fully depict the filtration captured through the ECC separation approach; instead they are approximated from a sparse distance matrix recording edges of each tetrahedron inserted. The edges are all assigned the weight of the tetrahedron itself, resulting in all edges and the tetrahedron appearing in the filtration simultaneously for PH.

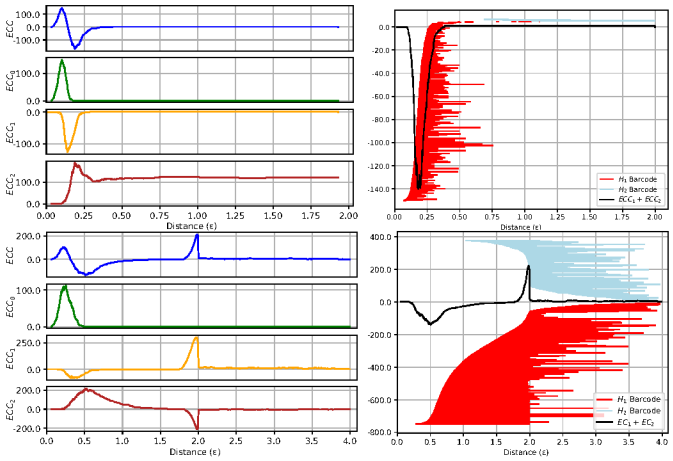

Fig. 3: Analysis of $\mathbb{R}^3$ sphere (top row) and torus (bottom row) data: ECC curves (left column), and the $ECC_1 + ECC_2$ curve overlayed with H_1 and H_2 PH results (right column).

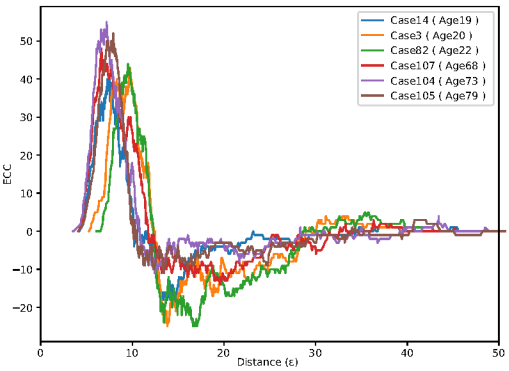

Fig. 4: The ECC for several $\mathbb{R}^3$ Brain Artery data sets. Legend shows patient age labels. Statistics of these trends were utilized for evaluation against the entire set of patients with correlation results in Table III.

C. Real-world Data Analysis

This section examines the use of ECC to analyze MRA Brain Artery Trees in $\mathbb{R}^3$. This data set contains brain artery trees for $n = 98$ labeled patients; patient labels include age, gender, and handedness. The data for each patient contains roughly $100k$ points. Previous studies utilizing PH [19], [20] have identified correlations between the H_1 persistence intervals (loops) and the ages of the patients; however, the results with PH had to use additional subsampling of the test data. In this study the MRA Brain Artery Trees are reduced to 3000 points through k-means clustering (the previous PH studies used up to 2000 points). During experimentation it was observed larger samplings of the MRA scans generally increased the ECC correlation measures.

Figure 4 depicts the ECC results for a sampling of several of the oldest and youngest patients in the data set. The younger

	Measure	$Corr(Age)$	$Corr(Gender)$	$Corr(Hand)$
ECC	Max	0.553	0.071	0.029
	Min	0.444	**0.418**	0.165
	Avg	**0.587**	0.363	0.071
	MaxT	0.456	0.096	**0.222**
	MinT	0.336	0.089	0.045
	Area	0.259	0.263	0.152
ECC_0	Max	0.497	0.079	0.008
	Min	0	0	0
	Avg	0.481	0.134	0.057
	MaxT	0.506	0.152	**0.171**
	MinT	**0.568**	**0.242**	0.096
	Area	0.178	0.154	0.107
ECC_1	Max	0.340	0.132	0.037
	Min	0.434	0.305	0.109
	Avg	0.528	**0.348**	0.051
	MaxT	0.262	0.311	0.051
	MinT	0.169	0.128	**0.269**
	Area	**0.589**	0.311	0.071
ECC_2	Max	0.459	0.334	0.091
	Min	0.197	0.019	**0.177**
	Avg	**0.541**	**0.465**	0.087
	MaxT	0.197	0.227	0.157
	MinT	0.223	0.153	0.087
	Area	0.533	0.271	0.033

TABLE III: Pearson Correlation Coefficients for the full set of MRA results. **Bold** values represent best curve correlation, and underlined values denote the best overall correlation. MaxT, MinT are the filtration where the max, min values occurred.

cases $(3, 14, 82)$ have a lower overall peak compared to the older cases $(104, 107, 105)$. The latter also have a greater minimum value; these characteristics may be associated to the H_1 loops identified in previous studies that correlate with age. However, Figure 4 only examines a sampling of the patients. Evaluation of several ECC measures against the full set of patients are provided in Table III.

While the ECC measures presented in Table III are certainly naive, they demonstrate potential measures related to the results from Malott [20]. Further exploration of similarity metrics, time-series analysis, and machine learning methods are of interest for application of the ECC.

In some cases the correlation of the measurement with the patient's label is close to that of the previous study (Table I of [20]). The best correlation with the Age labels was found with the area under each patient's ECC_1, corresponding to the H_1 loops of the data. This result is notably similar to the previous study; the area under each ECC curve represents the total length of persistence intervals in the filtration. The other two labels, gender and handedness, do not have published baselines for correlation. Handedness, overall, does not perform notably. In the case of gender, the average value of the ECC_2 curve provides the best correlation. The presence of H_2 voids in the MRA scans have not been previously studied.

VI. Conclusion

This study has examined the Euler-Poincaré Characteristic Curve (ECC) as a topological invariant and complementary tool for existing methods of Topological Data Analysis, most notably Persistent Homology (PH). The natural connection between the ECC and PH is beneficial for studying topological relationships within data sets. Classification, clustering, and similarity search applications built upon the ECC require significantly less memory resources than PH and can extend result interpretation to larger datasets and of higher dimensions.

The ECC method proposed has the ability to separate dimensional homological landscapes by taking advantage of neighborhood information ignored in the traditional ECC approach. Analysis of these landscapes can be complementary to the results of PH for many applications, as has been shown through study of MRA Brain Artery Trees.

References

[1] G. Carlsson, "Topology and data," *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 255–308, Apr. 2009.

[2] D. S. Richeson, *Euler's Gem: The Polyhedron Formula and the Birth of Topology*. Princeton, NJ: Princeton University Press, 2008.

[3] E. Richardson and M. Werman, "Efficient classification using the euler characteristic," *Pattern Recognition Letters*, vol. 49, pp. 99–106, Nov. 2014.

[4] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, "A roadmap for the computation of persistent homology," *EPJ Data Science*, vol. 6, no. 1, Aug. 2017.

[5] F. Chazal and B. Michel, "An introduction to topological data analysis: Fundamental and practical aspects for data scientists," *Frontiers in Artificial Intelligence*, vol. 4, Sep. 2021.

[6] N. O. Malott, S. Chen, and P. A. Wilsey, "A survey on the high-performance computation of persistent homology," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 5, pp. 4466–4484, Jan. 2022.

[7] C. S. Pun, K. Xia, and S. X. Lee, "Persistent-homology-based machine learning and its applications – a survey," pp. 1–42, Nov. 2018.

[8] L. Crawford, A. Monod, A. X. Chen, S. Mukherjee, and R. Rabadán, "Predicting clinical outcomes in glioblastoma: an application of topological and functional data analysis," *Journal of the American Statistical Association*, vol. 115, no. 531, pp. 1139–1150, Jul. 2020.

[9] A. Zomorodian, "Fast construction of the vietoris–rips complex," *Computer and Graphics*, vol. 34, pp. 263–271, Jun. 2010.

[10] A. Smith and V. M. Zavala, "The euler characteristic: A general topological descriptor for complex data," *Computers & Chemical Engineering*, vol. 154, p. 107463, Nov. 2021.

[11] O. Bobrowski and P. Skraba, "Homological percolation and the euler characteristic," *Physical Review E*, vol. 101, no. 3, pp. 032 304:1–16, Mar. 2020.

[12] G. Beltramo, R. Andreeva, Y. Giarratano, M. O. Bernabeu, R. Sarkar, and P. Skrab, "Euler characteristic surfaces," Jun. 2021.

[13] K. Turner, S. Mukherjee, and D. M. Boyer, "Persistent homology transform for modeling shapes and surfaces," *Information and Inference: A Journal of the IMA*, vol. 3, no. 4, pp. 310–344, Dec. 2014.

[14] R. Ghrist, R. Levanger, and H. Mai, "Persistent homology and euler integral transforms," Jun. 2018.

[15] J. Curry, S. Mukherjee, and K. Turner, "How many directions determine a shape and other sufficiency results for two topological transforms," Oct. 2019.

[16] R. R. Verma, N. O. Malott, and P. A. Wilsey, "Data reduction and feature isolation for computing persistent homology on high dimensional data," in *Workshop on Applications of Topological Data Analysis to Big Data*. USA: IEEE, Dec. 2021, pp. 3860–3864.

[17] F. Wang, H. Wagner, and C. Chen, "GPU computation of the euler characteristic curve for imaging data," in *The 38th International Symposium on Computational Geometry*, ser. Leibniz International Proceedings in Informatics (LIPIcs), X. Goaoc and M. Kerber, Eds. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2022, pp. 63.1–63.17.

[18] P. Bendich, J. S. Marron, E. Miller, A. Pieloch, and S. Skwerer, "Persistent homology analysis of brain artery trees," *The Annals of Applied Statistics*, vol. 10, no. 1, pp. 198–218, Mar. 2016.

[19] N. O. Malott, R. R. Verma, and P. A. Wilsey, "Parallel computation of partitioned persistent homology," in *2021 IEEE International Conference on Cluster Computing*, ser. CLUSTER 2021. Piscataway, New Jersey: IEEE, 2021, pp. 344–354.