

A Hypothesis Testing Framework for Comparing Warm and Cold Dark Matter Assumptions

Mike Wu, Jessi Cisewski

December 20, 2015

Abstract

Much of the definitions of the observable Universe and dark matter are heavily based on assumptions. Provided this uncertainty, one approach to gaining understanding of the true structure of the Universe is through cosmological simulations. Recent simulations from the EAGLE project have separated the Universe under warm and cold dark matter assumptions. Scientifically, it is believed that warm and cold dark matter have contrasting physical properties. This paper attempts to present a hypothesis testing framework to quantify the differences and similarities between warm and cold dark matter to a statistical precision. One approach to comparing and contrasting cosmological simulations is through structure and geometry. This paper uses the resulting persistent homologies from topological data analysis to create statistics for hypothesis testing. Voronoi tessellations were used as inexpensive proxies to mimic the topology of true simulations of the observable Universe to a finite precision. Through the Voronoi simulations, it is shown that the euler characteristic test, the global kernel density test, the parallel silhouette test, and the global contour test are best suited frameworks to distinguish fine differences in structure while remaining powerful. When applied to the EAGLE simulation, globally the tests failed to reject the null hypothesis, but with finer resolution through cubic splitting, it is suggested that the observable Universe under assumptions of warm and cold significantly differ by geometry and not structure.

1 Introduction

Many questions about dark matter and the observable Universe have binary answers: yes or no. For example, are there differences in structure between warm and cold dark matter? If lambda-CDM theory prevails, warm dark matter (WDM) should move faster, resulting in fragmentation and voids. Contrastingly, cold dark matter (CDM) is thought to be slower and more likely to collapse smaller objects under their self-gravity and merge in a continuous hierarchy to form conglomerate objects. Therefore, one should expect to see fewer voids and perhaps more filament loops.

Recent advancement in topological data analysis provide algorithms to summarize the presence of loops and voids from large three-dimensional data sets. More specifically, persistent homology offers an efficient approach for tracking

which topological objects “persist” over time. Provided these graphical summaries, hypothesis tests can potentially address binary questions with statistical confidence. One such test, the Euler characteristic test, is very popular and has shown success as a test statistic in several research domains. This paper explores the Euler test and several alternative testing frameworks in analyzing topological structure in the universe from EAGLE simulations.

Topological data analysis and persistence diagrams have been used for similar purposes of comparing complex structures but in different academic and industry fields. Bendich et. al. [1] demonstrated improved statistical analyses of multiple sets of brain artery trees using persistence diagrams to quantify branching and looping of vessels. Duong et. al. [5] developed the global kernel test to automate detection of structural changes in cellular morphology. Given the similarity of the data and the statistical task, the extension of similar models to astronomy is a rational one.

Conventional topological data analysis depended on distance functions to quantify differences between two persistence diagrams. Both Wasserstein [6] and Bottleneck distances [2] were proven to be stable, with many applications into shape segmentation and labeling. These offered the first approaches to comparing points on 3D shapes reliably, creating global signatures to capture the overall topology of an object. In practice, distance functions are computationally expensive and biased towards smaller data sets. In response, landscapes and silhouettes [3] were developed to further abstract a persistence diagram into functional summaries to more quickly compare signatures. Additional methods like permutation tests [7] or kernel density tests [4, 5] sprouted as further alternatives. With the plethora of summary statistics, a question this paper hopes to address is which of them is best at differentiating topological signatures.

2 Persistent Homology

Persistent homology computes topological features of a space given some parameters for spatial resolution. This algorithm is biases persistent features as those more likely to represent the true features of the underlying space, while others are deemed as artifacts of sampling or noise. Persistent homology returns diagrams of the births and deaths of different dimensional features.

Homology is calculated through representing the space as a simplicial complex. These complexes are constructed through a composition of points (0 dim), line segments (1 dim), triangles (2 dim), and their n-dimensional counterparts to represent a set of probable shapes. The filtration of a simplicial complex represents a distance function on the underlying space and induces simplicial homology groups for each dimension p . The resulting persistent homology is encoded in the form of parametrized Betti numbers into a persistent diagram. For data regarding the observable Universe, we are limited to three dimensions. Figure 1 shows an example of a large scale simulation of the Universe and its persistent homology. Generally the more interesting objects have longer lives, and are on the exterior of the persistence diagram. From figure 1, there are a fair amount of long-lasting first and second homologies, indicating the presence of loops and voids in structure, some of which of visually visible. When comparing two simulations, a measure of similarity is calculated from the number and lifespan of common loops and voids.

The births and deaths of persistent objects can be efficiently calculated using kernel density estimation (KDE). Provided a smoothing parameter h and coordinates in a grid (x, y) , KDE is defined as:

$$\frac{1}{n(\sqrt{2\pi}h)^d} \sum_{i=1}^n \exp\left(-\frac{\|y - x_i\|_2^2}{2h^2}\right)$$

KDE produces a smoothed version of the topology. Geometrically, the birth times of persistent object are extracted using KDE as the maximas of the topology: Imagine a two dimensional plane existing above the three dimensional topology. As the plane slides down to intersect with the topological structure, it passes through the peaks of the topology. Each such "pass" is considered a birth time of a homology. As the plane continues to slide, it is possible for two points of local maxima to converge, assuming the topology is continuous function. Once the plane slides down to intersect the points of convergence, a corresponding death time is defined for one of the two original maximas. The other maxima "absorbs" the second to become a higher order homology. Other algorithms exist for calculating persistent births and deaths such as RIPS or distance-to-a-measure, some of which are explored in this paper.

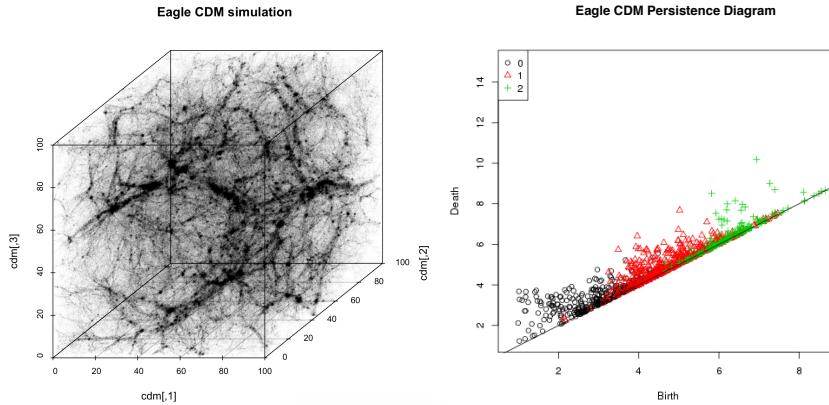


Figure 1: (left) Simulation of the observable Universe under assumptions of code dark matter. Notice the strands of filament. (right) The corresponding persistence diagram.

3 Methodology

For each testing framework, nine sets of 15 voronoi tessellations were generated, one for each filament percentage (0.1 to 0.9), and a persistence diagram was created for each tessellation. Two baseline sets were considered, 0.1 and 0.9. Each of 9 sets of persistence diagrams were compared to each baseline set to produce a probability of rejecting the null hypothesis that the two sets are sampled from the same higher dimensional distribution.

3.1 Euler Characteristic Test

The Euler characteristic is a topologically invariant number meant to describe shape. Provided some convex polyhedra where V is the number of vertices, E is the number of edges, and F is the number of faces, Euler's characteristic is the following:

$$\chi = V - E + F$$

Generalizing to the larger space of simplicial complexes, Euler's characteristic can be defined as an alternating sum where N is the number of dimensions:

$$\chi = k_0 - k_1 + k_2 - k_3 + \dots = \sum_{i=0}^N (-1)^i k_i$$

For any topological space, each k_i represents the i -th Betti number, or the rank of the i -th singular homology group. Therefore, when analyzing persistence diagrams of voronoi tessellations, there exist only 3 dimensions of data, and therefore only homologies of dimension 0, 1, 2. Given the Betti numbers b_0 , b_1 , and b_2 , the equation simplifies to:

$$\chi = b_0 - b_1 + b_2$$

For each persistence diagram, the Euler characteristic is calculated at each value along the x-axis by some resolution. The resulting Euler function is integrated with trapezoidal summation to produce an Euler statistic. Given two sets of Euler statistics, a t-test is used to extract a p-value between them.

3.2 Silhouette Test

Provided a persistence diagram, a silhouette is created as a weighted sum (smoother) form of landscape functions. Like the Euler Characteristic function, a silhouette summarizes a persistence diagram through some continuous function. However, unlike Euler, the silhouette must be calculated per dimension. Landscape functions were also tested but were found to be more expensive and less accurate than their silhouette counterparts. In order to compare two sets of persistence diagrams, each diagram was summarized by its silhouette and the two sets of silhouettes were abstracted into a single p-value through a non-parametric t-test such as Kolmogorov-Smirnov. Often the different dimensions were very unstable and often resulted in disagreeing probability values.

As an attempt to stabilize the silhouette across dimensions, for each persistence diagram, a silhouette was generated for each of the three dimensions. Given two sets of persistence diagrams, the result would be two sets of 3 dimension arrays. These two sets were compared directly through a three-dimensional hotelling t-squared test. Results from the joined silhouette test were far more stable and shared features of each of the individual three.

3.3 Global Contour Test

Often in topological analysis, a viable strategy for comparing two persistence diagrams is a distance function between homologies. However, because Bottleneck and Wasserstein distances are exceedingly expensive, they are not a realistic

option. An alternative is to create a contour around the physical locations of the homologies of a single dimension in a persistence diagrams. One can then take the absolute difference between the contours of two persistence diagrams and the mean squared error as the test statistic Q . Given $c(x)$ as a contour function for x , Q is representable as:

$$Q = \frac{1}{n} \sum_i^n (c(x_1) - c(x_2))^2$$

In order to obtain a probability value between two sets of persistence diagrams, a permutation test is used. Given two independent samples $X_1, \dots, X_n \sim P$ and $Y_1, \dots, Y_m \sim Q$, the null hypothesis tests $H_0 : P = Q$. A permutation test calculates a test statistic T as a function of set $D = X_1, \dots, X_n, Y_1, \dots, Y_m$. Provided a set of labels $G = 0, \dots, 0, 1, \dots, 1$ where each X_i has a label 0 and each Y_i has a label 1. After randomly permuting the group labels, recompute the test statistic T under H_0 . After N iterations, establish a set T_1, \dots, T_N . The p-value represents the fraction of times T_j is larger than T for any j . As N approaches inf, the resulting p-value approaches the true value.

$$p = \frac{1}{N} \sum_{j=1}^n I(T_j \geq T)$$

A specific permutation test called the kernel test is used. Given some kernel, like the Gaussian kernel $K_h(x, y) = \exp(-\frac{\|x-y\|^2}{h^2})$, the test statistic T can be written as the following, where $K_h(x, y)$ can be thought of as a similarity measure between x and y . Note that h is a tuning parameter which default to a value such that $T = \sup_{h \geq 0} T_h$.

$$T = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K_h(X_i, X_j) - \frac{2}{mn} \sum_{i=1}^n \sum_{j=1}^m K_h(X_i, Y_j) + \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^n K_h(Y_i, Y_j)$$

3.4 Global Kernel Density Test

TBD.

3.5 Local Kernel Density Test

TBD.

4 Simulation Studies

4.1 Voronoi Simulations

To measure the effectiveness of different hypothesis testing frameworks, voronoi simulations are a cheap option for generating a training set. True simulations like EAGLE are computationally expensive and often unlabeled, better serving as the testing set.

A Voronoi simulation is generated from a Voronoi tessellation, where the edges of each cell represent filaments and the faces of the cell represent voids.

Give a plane with fixed size, a Voronoi tessellation partitions it into cellular regions with nuclei. These nuclei represent seeds, and every cell territory is defined as the set of points equal or closer to that cell's seed than any other. To produce a simulation in polynomial time, each point in the plane is compared to the k closest nuclei using the nearest neighbor algorithm where k is chosen to be a small integer. Gaussian noise is added to perturb the plane and inject randomness. Because the nuclei number and the plane size are variable, the points in the simulation representing filaments, clusters, walls, and clutter are customizable as well. By choosing a reasonable percentage of filaments and related structures, Voronoi simulations approach true simulations of the observable universe. By varying these percentages and repeating simulations, one can quickly generate a large, labeled data set for hypothesis testing.

4.1.1 Parameter Specifications

The Voronoi simulations used in this paper have a 1.25×10^5 box volume, 0.1 resolution, 1×10^4 points, 64 cells, 0.02 cluster, 0.00 clutter, and variable percentage of filament [0.1, 0.9] and walls [0.08, 0.88]. Persistent homology are calculated with a distance-to-metric (dtm) shell with a 0.01 tuning parameter. Persistence diagrams are preprocessed to remove the known 0-dimensional artifact and optionally are normalized by the maximum birth or death value after removal.

4.2 Voronoi Results

The hypothesis tests were run on 20 iterations of sets of 15 voronoi simulations. Results are shown for each of the hypothesis testing frameworks comparing a set of Voronoi simulations for each percent filament from 0.1 to 0.9 with a independent baseline simulation of 0.1 percent filament. Trials were also completed against a baseline simulation of 0.9 percent filament and identical results were found. See figure 1 and 2 for plots of p-values and IQR variances.

With a baseline set of 0.1 percent filament, one should expect a p-value around 0.5, which suggests that most of tests are "powerful". The best three testing frameworks are the (1) Euler test, (2) the parallel silhouette test, and (3) the contour test on 2-dimensional homologies. These three frameworks are able to uncover a significant difference between 0.1 and 0.5 percent filaments sets whereas other tests were less confident. However, the global KDE test may be useful since it does not immediately produce a p-value close to 0 given smaller differences in percent filament, making it better suited to answer questions of "how different" instead of "if different". The 0th and 2nd dimensional silhouette tests show no discernible pattern between scaled percent filaments.

4.3 EAGLE Simulations

Training the hypothesis testing frameworks on the Voronoi simulations provided some ranks of effectiveness, susceptibility to noise, stability, etc. With that information, testing was conducted on the Virgo Consortium's EAGLE project, which is a suite of hydrodynamical simulations that represent the appropriate formation of galaxies and volumes. A large improvement of this simulation over others is the treatment of feedback from massive stars and AGN in which

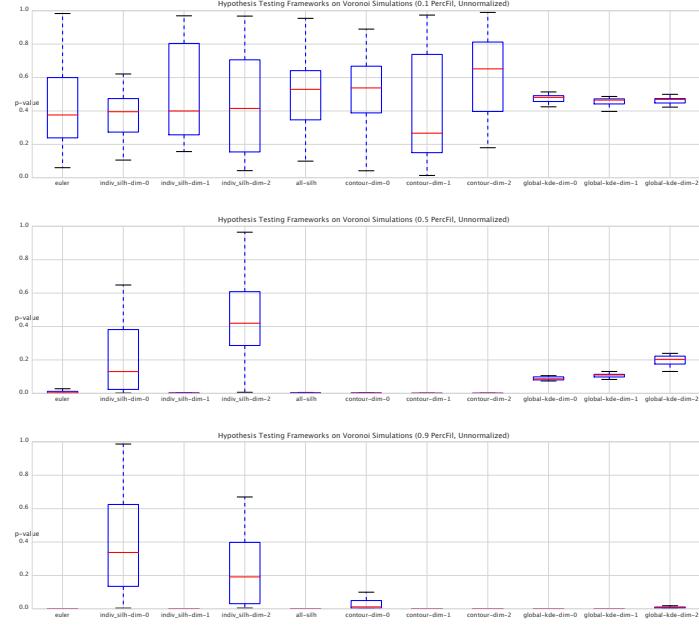


Figure 2: Box plots for each testing framework when comparing sets of percent filament 0.1, 0.5, and 0.9 with the 0.1 percent filament baseline.

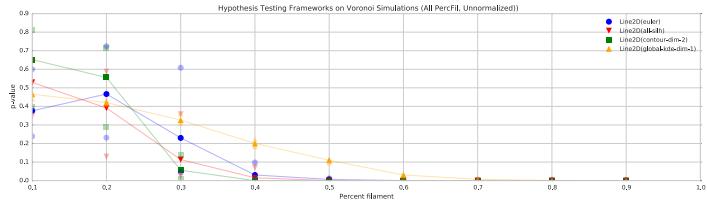


Figure 3: Line plots of four hypothesis testing frameworks. X-axis represents the percent filament of the set being compared to the baseline set; y-axis shows the p-values. The lines plot the median p-value of the 20 iterations and the fainter dots at each percent filament represent the 25th and 75th percentiles.

thermal energy is injected into the gas without the need to turn off cooling or hydrodynamical forces, allowing winds to develop without predetermined speed or mass loading factors. From the EAGLE project, two data sets were available modeling the state of the Universe under assumptions of warm dark matter and cold dark matter. For the WDM simulation, disingenuous objects were filtered out by sub-halo mass and sub-halo sphere sizes. The same procedure of normalizing, running persistence diagrams, and hypothesis tests was performed to evaluate similarity. Additionally, local kernel density tests were used to visually quantify that similarity.

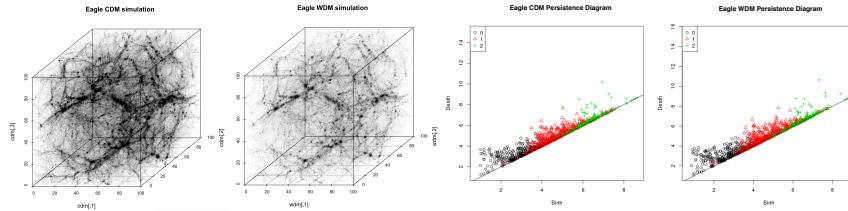


Figure 4: (left) Visualization of the complete CDM and WDM simulations. (right) Their corresponding persistence diagrams (unnormalized).

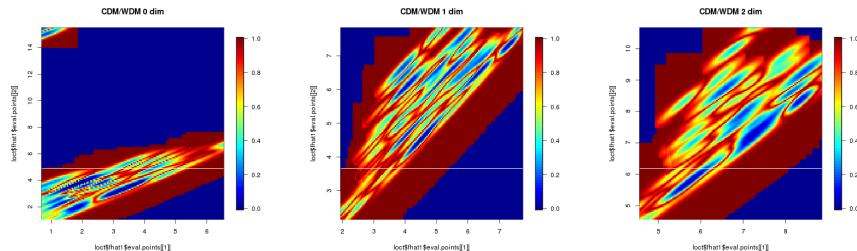


Figure 5: The differences in contour between the WDM and CDM data sets. Ignore the artifact in the upper left corner of the 0th homologies figure.

Because the WDM and CDM simulation are each only a single data set, most of the aforementioned hypothesis tests do not apply. Only the global KDE test is relevant, which produced a 0.694 p-value, failing to reject the null hypothesis that the two simulations differ. The persistence diagrams were generated a volume of 1×10^6 , resolution of 2 and a dtm distance function with tuning parameter 0.001.

4.3.1 Cubic Slicing

Since a closer look at the persistence diagrams and the local KDE tests suggest differences, a better examination of the WDM and CDM simulations is worthwhile. In order to use the rest of the hypothesis tests, a set of simulations are required. Assuming that each simulation is large and repetitive, one can split the cubic simulation into smaller cubes evenly. For example, a “double split” would produce a set of size 8 while a “triple split” would produce a set of size 27. Figure 5 shows the results of each hypothesis test across three split sizes.

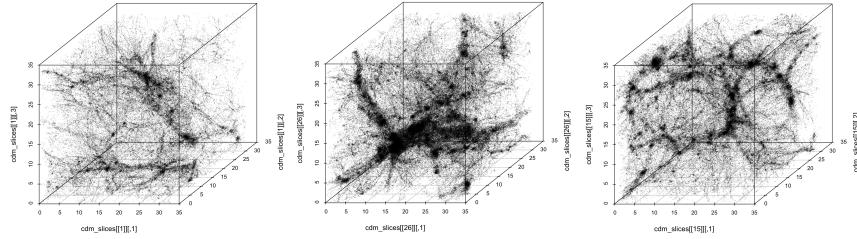


Figure 6: Examples of three triple-split samples from CDM simulation.

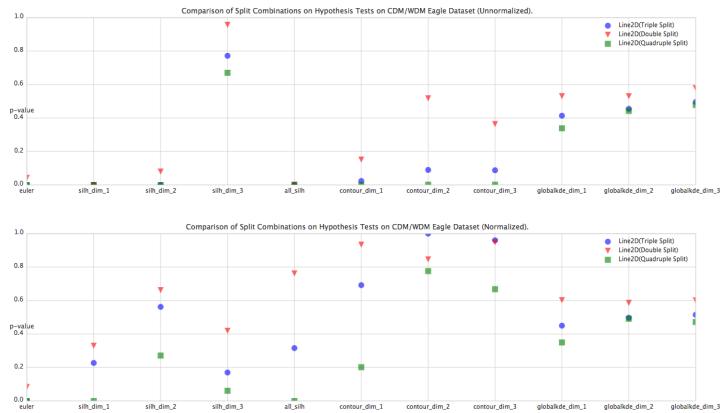


Figure 7: (top) Results of unnormalized hypothesis tests on a split simulation set of size 8, 27, and 64 respectively. (bottom) An equivalent set of tests with normalized persistence diagrams.

The results suggest that the greater the number of splits, the lower the p-value regardless of hypothesis test. Additionally, normalization dramatically reduces the effectiveness of the tests. With normalization, most of the hypothesis testing frameworks across all three split types drastically change. Again, global KDE tests are relatively invariant; the 3rd dimension silhouette test produces a far lower p-value; all other tests produce significantly higher p-values. Since normalizing standardizes object sizes, this could suggest that WDM and CDM simulations are relatively agreeable in structure but differ in geometry.

5 Discussion

A possible improvement is to explore the global KDE test on the CDM and WDM EAGLE simulations with a variety of parameters. With a smaller resolution, it may be possible to catch more minute differences. The resolution of 2 may have been too coarse for the global KDE test to extract the structural differences. This would agree with the higher level of differences found after cubic splitting since a smaller resolution was used. As a measure of the effect of resolution, the CDM and WDM simulations were triple split, and each of the resulting 27 divisions were paired together for a global KDE test. The process was repeated for resolutions 1, 0.5, 0.25. Figure 10 shows a plot of the corresponding p-values.

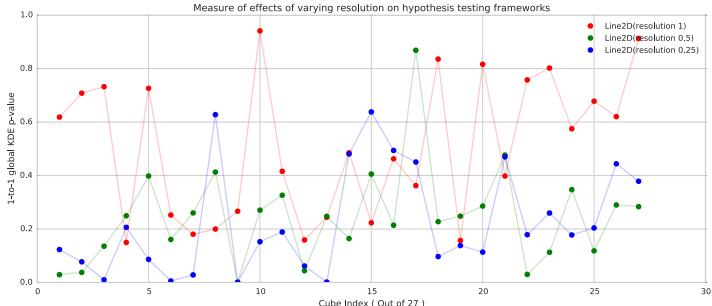


Figure 8: Plot of individual global kernel density test results on CDM and WDM simulations with varying resolution.

From figure 10, two possible conclusions are (1) the 27 divisions are not very similar seeing that different indexes produce a large range of p-values, and (2) the finer the resolution, the lower the p-value. This suggests that a finer resolution is able to preserve smaller topological structure into the persistence diagrams. Tuning the resolution will be an important task.

6 Appendix

6.1 Normalization

A possible preprocessing step to hypothesis testing is normalizing the persistence diagrams. Depending on the Voronoi simulation parameters, the ranges of the

x, y, z axis are different, produced different axes on the persistence diagrams. Normalizing these axes may increase comparability between simulation sets. For each persistence diagram, all of the homology coordinates for (birth,death) were divided by $\max(\text{births}, \text{deaths})$.

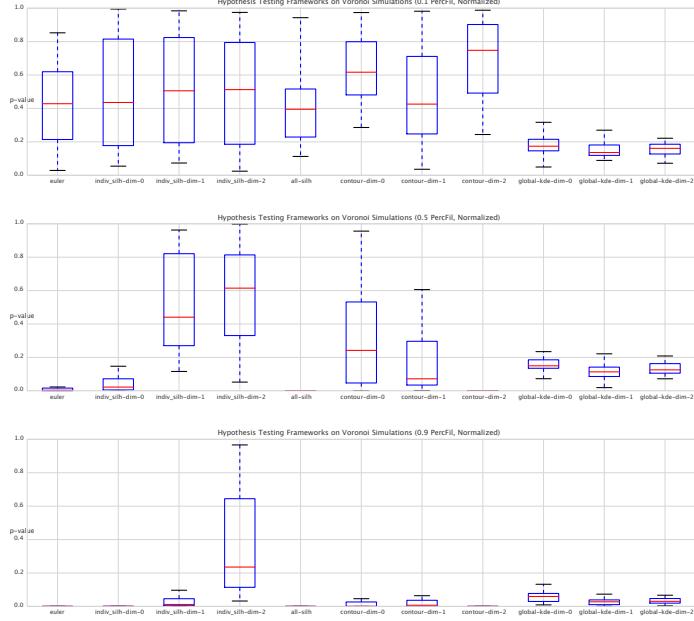


Figure 9: Boxplots of hypothesis tests on normalized persistence diagrams for 0.1, 0.5, and 0.9 percent filaments.

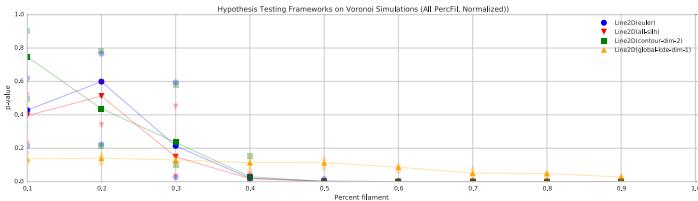


Figure 10: Another set of four hypothesis testing results. In this case, all persistence diagrams were normalized prior to processing.

Normalization has a drastic impact on hypothesis test results. Primarily, it seems that the global KDE test has lost effectiveness entirely, remaining roughly the same across the three percent filaments. Remarkably, the Euler test, the parallel silhouette test, and the 2nd homology contour test remain superior. Additional patterns emerge: when comparing the effectiveness of the tests over the range of percent filaments, one notices that the lower dimensional homologies are more effective for silhouette tests while the higher dimensional homologies are more effective for contour tests. Geometrically, normalizing the persistence

diagrams is equivalent to disregarding size in favor of structure. For example, a circle of radius 5 is suddenly equal to a circle of radius 500. The results possibly indicate that the global KDE test is primed to detect geometric differences (size) while the top three tests are better at detecting structural differences.

References

- [1] Paul Bendich, JS Marron, Ezra Miller, Alex Pieloch, and Sean Skwerer. Persistent homology analysis of brain artery trees. *arXiv preprint arXiv:1411.6652*, 2014.
- [2] Mathieu Carrière, Steve Y Oudot, and Maks Ovsjanikov. Stable topological signatures for points on 3d shapes. In *Computer Graphics Forum*, volume 34, pages 1–12. Wiley Online Library, 2015.
- [3] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman. Stochastic convergence of persistence landscapes and silhouettes. In *Proceedings of the thirtieth annual symposium on Computational geometry*, page 474. ACM, 2014.
- [4] Tarn Duong. Local significant differences from nonparametric two-sample tests. *Journal of Nonparametric Statistics*, 25(3):635–645, 2013.
- [5] Tarn Duong, Bruno Goud, and Kristine Schauer. Closed-form density-based framework for automatic detection of cellular morphology changes. *Proceedings of the National Academy of Sciences*, 109(22):8382–8387, 2012.
- [6] Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A stable multi-scale kernel for topological machine learning. *arXiv preprint arXiv:1412.6821*, 2014.
- [7] Andrew Robinson and Katharine Turner. Hypothesis testing for topological data analysis. *arXiv preprint arXiv:1310.7467*, 2013.