

The Persistent Homology of Archaeological Similarity Networks in the American Southwest.

Andrew Belz

Last Updated: July 1, 2026

Abstract

Archaeological similarity networks (ASNs) provide a straightforward framework for the quantitative comparison of archaeological sites and are widely used by computational archaeologists to study the transmission of culture. The graph, a data structure that encodes relationships between entities, underpins the vast majority of the analytical methodology employed by researchers in the field to study ASNs. While this structure provides a relatively robust framework for analysis, it comes with some notable limitations. First, a threshold for the minimum similarity score must be chosen to construct a graph representation, which can potentially be a source of error. Second, graphs lack the ability to encode geometric structure beyond pairwise relationships. Persistent homology, a framework originating from the growing field of Topological Data Analysis (TDA), is introduced to address these limitations and provide a robust alternative to traditional network analysis in the study of ASNs. Common TDA techniques are applied to extract meaningful insights about the distribution of ceramic material in the American Southwest. The persistent homology of site similarities originating from the Chaco region (1050-1100 AD) reveals that these sites form dense, well-mixed clusters of similarity that cannot be explained by geography alone. Ultimately, the application of TDA provides a more comprehensive topological understanding of cultural transmission, offering archaeologists a powerful new tool to analyze complex, multidimensional relationships between sites.

Keywords: Persistent homology, topological data analysis, archaeological similarity networks, Chaco, American Southwest.

1 Introduction

A central problem in archaeology is determining the existence and extent of cultural relationships between distinct sites. What inferences can be made about the scale and structure of cultural transmission between ancient peoples? How can the properties of cultural transmission be measured for an entire region? How can the presence of cultural barriers between peoples be identified? These are some of the questions that computational archaeologists seek to answer through the analysis of Archaeological Similarity Networks, which model the cultural relationships between archaeological sites based on the similarity of the material found there.

1.1 Aims

The aim of my work is to critically evaluate, through the lens of mathematics and archaeology, the work that is being done to study Archaeological Similarity Networks, and summarize how they are being used to solve problems in the field. I also aim to address existing limitations. Finally, I apply persistent homology, originating from the growing field of Topological Data Analysis, to archaeological data, to demonstrate its effectiveness for identifying boundaries of cultural similarity between sites with various different properties.

1.2 Research Question

What are the quantitative methods being employed to study Archaeological Similarity Networks? What are the strengths and limitations of these methods, and how can they be improved?

1.3 Approach

My approach is to critique the tools being used to analyze Archaeological Similarity Networks, then comparatively expand upon them by introducing a novel framework (to archaeology) for network analysis: persistent homology. I apply the method to an existing data set representing a

common application of network analysis to show that persistent homology may provide archaeologically meaningful results in an area of active study.

1.4 Summary of Contents

Following this introduction, Section 2 reviews the modern paradigm of Archaeological Similarity Networks and identifies the limitations of simple graph models. Section 3 outlines the methodology, detailing the mathematical intuition behind persistent homology and the construction of clique complexes to estimate the network's topology. Section 4 presents the results of applying this higher-dimensional analysis to regional site similarities, demonstrating distinct differences between measures of cultural similarity and geographic proximity. Section 5 discusses the implications of these findings for the field and suggests avenues for future research. Supplemental mathematical and archaeological details and references are provided in Section 6.

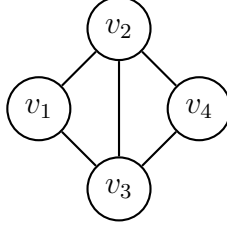


Figure 1: A simple graph with 4 vertices (v_1, v_2, v_3, v_4) and 5 edges ($\{v_1, v_2\}, \{v_2, v_3\}, \{v_2, v_4\}, \{v_3, v_4\}, \{v_1, v_3\}$).

2 Literature Review

2.1 Archaeological Similarity Networks (ASNs)

ASNs are commonly used in archaeological data analysis to represent pairwise relationships between archaeological sites and identify cultural similarity patterns and motifs (Collar, Coward, Brughmans, & Mills, 2015). Similarity in the material recovered at two distinct sites is frequently used to quantify both temporal relationships (chronology) and cultural material exchange. Much of the methodology used to study ASNs originates from Social Network Analysis (SNA) (Borck, Mills, Peeples, & Clark, 2015), a subfield of formal network theory that seeks to develop methodologies to study social networks.

2.1.1 Graphs as Network Representations

The fundamental mathematical abstraction underlying ASNs is the graph, a set of vertices and a set of pairs of vertices called edges (Fig. 1). While there are many variations in the specific mathematical properties that computational archaeologists assign to graph representations of ASNs, I address only the most basic construction: the simple, undirected graph. Typically, each archaeological site is represented by a distinct vertex, and edges are assigned a weight corresponding to the strength of pairwise similarity between sites.

2.1.2 Brainerd-Robinson Similarity

While there are many ways to weight the strength of a cultural relationship between archaeological sites (Prignano, Morer, & Díaz-Guilera, 2017), one of the most widely adopted measures is the Brainerd-Robinson similarity coefficient (Brainerd, 1951; Robinson, 1951), which I refer to as BR similarity. The BR similarity of archaeological sites x and y , denoted $BR(x, y)$, measures the difference in material proportions of all types identified at each site. Let x_i and y_i denote the percentage of material type i found at sites x and y , respectively. The BR similarity is defined as

$$BR(x, y) = 200 - \sum_{i=1}^n |x_i - y_i|.$$

Sites with very similar distributions of the same types of archaeological material, relative to site size, have high similarity scores close to 200. Sites with little in common have scores close to zero.

2.2 Limitations of ASNs

In this section, I reiterate two of the concerns of Prignano et al. regarding common approaches to ASNs and introduce one of my own.

The first is that in general, network analysis can often be sensitive to data loss. In every practical case, only a subset of archaeological sites can be represented, and at each site, only a subset of the original material can be represented. There are few other fields that suffer from data loss to the same extent as archaeology, and many traditional network analysis methods rely on the assumption of a robust representative sample.

The second pertains to the matter of thresholding. All network analysis methods require the choice of which edges to include or exclude, and the choice is often made by including only edges that meet a strength threshold for evidence of existence.

To address these problems, computational archaeologists frequently choose data preservation and inclusion in their models at all costs, opting to apply increasingly complex statistical frame-

works to measure the sensitivity and robustness of their methods (Gjesfjeld, 2015) and reaching for generative null models. Exponential random graph models (Brughmans, Keay, & Earl, 2014) have quickly become a go-to framework for testing hypotheses of network generation.

Lastly, network models account only for pairwise relationships. In many real-world networks, such as trade, transportation, and biological networks, interactions occur at a higher order than simple pairwise representations permit. In these fields, the simple graph has been generalized to various different forms that allow more complex polyadic relationships to be represented (Battiston et al., 2020). This includes the simplicial complex, a generalization of the clique complex that we address in Section 3.

While great strides have been made in recent years to resolve some of these issues and reflect true methodological improvements, no simple solution exists. I claim that the persistence of these problems and the increasing complexity in methodology required to fully resolve them indicate the presence of asymptotic limitations inherent to the modern paradigm, and that emerging interdisciplinary fields and approaches such as TDA provide promising novel perspectives and modeling frameworks with which to approach them.

2.3 Persistent Homology as the Solution

TDA as a whole, emerging as an applied science originating from the field of algebraic topology, seeks to abstract away typical notions of geometry (Carlsson, 2009). Instead, TDA relies on estimating algebraic invariants to extract information from the data. It concerns itself more with the general class of shape to which the data belongs (its invariant topological characteristics), in contrast to the precise, locally-reliant, and combinatorial nature underpinning much of network analysis.

Persistent homology provides a way to examine the topological signature of a space by identifying topological invariants called homology groups. In this analytical process, loops, gaps, and voids are identified that characterize the abstract space formed by relationships between sites (Edelsbrunner, Letscher, & Zomorodian, 2000), which I refer to as the dissimilarity landscape. In

the dissimilarity landscape, sites that are similar are considered close together, and sites that are dissimilar are considered far apart.

Indeed, persistent homology is reasonably robust to missing data. By seeking to identify global features that inherently characterize missingness, the effects of missing data are often inherently absorbed by pre-existing voids, or local regions of connectivity, ultimately manifesting as topological noise (Cohen-Steiner, Edelsbrunner, & Harer, 2007).

Second, persistent homology examines the characteristics of the dissimilarity landscape across all scales of similarity at once by isolating features that persist across a wide range of similarity thresholds. In this way, the choice of similarity threshold is fixed by rigorous mathematical justification.

Lastly, persistent homology naturally accounts for higher-order interactions that may be present in the dissimilarity landscape by considering completely connected groups of sites to be whole functional units (Battiston et al., 2020).

I claim that persistent homology provides a robust framework for the identification of both geographic and cultural barriers of various natures between sites in a cohesive, intuitive way, grounded in rigorous mathematics.

3 Methods

The method of persistent homology is applied to data for the types and quantities of ceramic material found across 223 sites originating from pre-colonial Chaco Canyon and the surrounding region (c. 1050-1100 A.D.). The data is obtained from Peeples and Brughmans (2023), a part of a computational supplement to their book *Network Science in Archaeology* (Brughmans & Peeples, 2023).

First, the groundwork of persistent homology must be established. To gain a better intuition for the process of persistent homology, the reader is encouraged to explore interactive open-source applications that illustrate the process. The web tool provided by Majhi (2021) provides a great visual example of the filtration of a Vietoris-Rips complex in the real plane, which is the mathematical architecture that I apply here. I have also made available an interactive three-dimensional visualization developed with Python in a Jupyter notebook that can be found in the Github repository associated with this project (see Supplemental Materials 6.1).

3.1 The Dissimilarity Metric

I adapt the Brainerd-Robinson similarity coefficient to obtain a notion of cultural distance rather than similarity. That is, sites that are highly similar to each other should be relatively close together, expressed by smaller numbers, and sites that are highly dissimilar should be far apart, expressed by larger numbers.

Define the cultural distance between sites x and y to be $d(x, y) = 200 - BR(x, y)$, which I refer to as the BR dissimilarity distance, measuring cultural dissimilarity. In order to rigorously justify the approach of persistent homology, a metric must be used to define dissimilarity, which is the mathematical formalization of the notion of relative distances between points. For proof that d satisfies the axioms to be a metric, see Supplemental Materials 6.2.1.

I compute $d(x, y)$ for every pair of sites x, y for which I have data. Additionally, I scale the data to the range $[0, 1]$, so that these results can later be compared to those induced by different

metrics with different units of measurement.

3.2 Persistent Homology

Persistent homology provides a way to identify holes, voids, or loops in the surface created by placing highly similar sites close together and highly dissimilar sites far apart in a high-dimensional space I call the dissimilarity landscape.

A scale parameter ε , called a filtration parameter, begins at zero and increases. At each step in the filtration, a graph is formed by connecting pairs of sites with a BR dissimilarity of at most ε . Local densely-connected features are considered to form single functional units represented by high-dimensional solidly-filled objects called simplices, allowing true voids to appear, revealing holes in the dissimilarity landscape. The resulting complex of high-dimensional objects is called a clique complex. These holes in the clique complex, referred to as H_k features, are organized by their dimension k . Each H_k is called the k -th homology group, and reveals distinct characteristics of the dissimilarity surface.

For a more detailed description of persistent homology, see Supplemental Materials 6.3.

3.3 Persistence Diagrams

H_k features, also referred to generally as homological features, generally exist for only a short duration during the filtration. When $\varepsilon = 0$, every vertex begins as a distinct H_0 feature, a single isolated point. As ε increases, features form at various times and are said to be *born*, and eventually features *die* as enough sites connect to form fully connected simplices, or merge to create new higher-dimensional features. The *lifespan* of a feature is defined as the time between its death and birth.

Short-lived H_k features are considered to be topological noise, often products of sampling error or sensitive scale dependence. Longer-lived H_k features are considered to be significant topological features of the data, indicating the presence of real structural gaps in the dissimilarity landscape. The main visualization tool used in persistent homology is called the *persistence diagram* (Fig. 5).

Each H_k feature is plotted in the plane, where the x -axis corresponds to time of birth and the y -axis corresponds to time of death. Features that lie further from the line $y = x$ (representing a lifespan of zero) have longer lifespans and thus are considered more significant. Furthermore, the location of a feature in this diagram reveals exactly when in the filtration the feature lived, providing insight into the scale of the feature.

3.4 Interpretation

The homology groups of the clique complexes across the filtration provide a view from which to study the cultural boundaries that exist between archaeological sites in the similarity landscape, in various dimensions.

Importantly, the presence or absence of homological features lacks a universal archaeological interpretation, and must be interpreted in the context of the data. However, with well-founded complementary assumptions, homology can be a powerful inferential tool in the study of ASNs.

For example, in this analysis I take the following interpretative stance. Assuming that human settlements are fixed over the 50-year span represented in the data and naturally form close cultural relationships with their nearest geographic neighbors, and that homological features represent barriers to cultural transmission, then geographic regions corresponding to separations between homological features should correspond to geographic barriers to interaction or signatures of cultural connectivity beyond simple planar geography.

H_0 features correspond directly to connected components¹, and tracking H_0 features across the filtration is equivalent to the method of single-linkage hierarchical clustering, a widely used community detection method in many fields, including archaeology (Peeples & Bischoff, 2023).

Significant H_0 features may indicate the presence of some transverse barrier to cultural transmission, such as a river, or a large intractable geographic feature like a mountain range or ocean. This interpretation would be especially well-founded if the location of H_0 separations corre-

¹A connected component is a part of a graph in which there is some sequence of edges connecting every pair of vertices. Figure 2 exhibits two connected components: the single isolated point with no connections, and the rest of the graph.

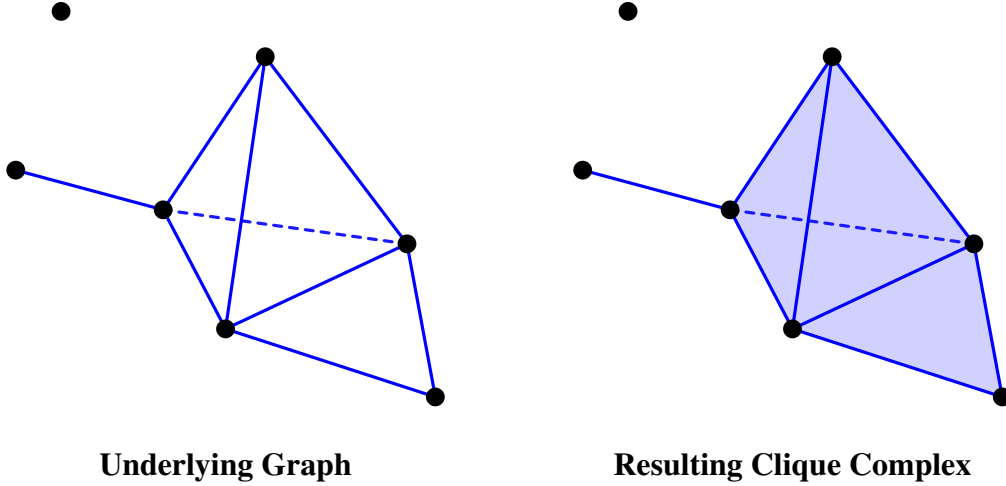


Figure 2: A clique complex, composed of an isolated point, one edge, one triangle, and one tetrahedron. There are two H_0 features (corresponding to two connected components), zero H_1 features, and zero H_2 features, as each triangle and tetrahedron is considered to be fully filled in.

sponded to identifiable geographic features.

H_1 features in the graph correspond to loops of edges forming a two-dimensional void. Similar to H_0 features, an H_1 feature may correspond to a geographic feature that did not provide such a sharp barrier to exchange, but forced the circumnavigation of this barrier. An example of such a feature may be a lake or mountain that could reasonably be bypassed.

An H_2 feature is a three-dimensional cavity entirely enclosed by completely connected triads of sites and is more difficult to interpret in this context. However, one situation where an H_2 feature may arise is in the presence of central highway systems or long-distance trade connections, where dense local clusters form with high rates of triadic closure, with selective long-distance connections made between critical hubs.

Persistent homology provides general insight into the shape of the space induced by comparing sites to one another, but the interpretation wholly depends on the assumptions and constraints that one wishes to adopt.

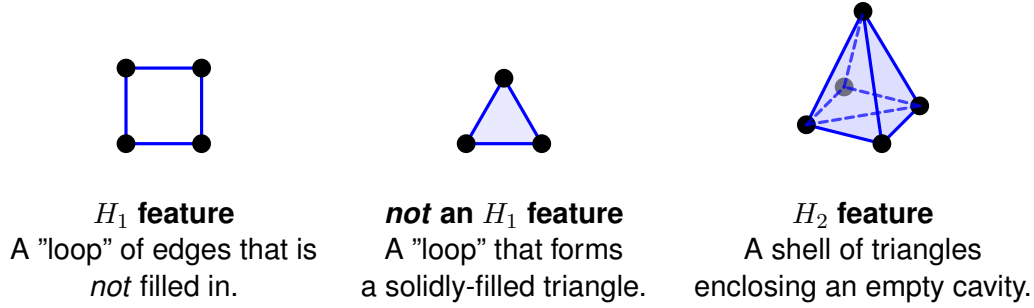


Figure 3: An H_1 feature example (left), an H_1 feature non-example (middle), and an H_2 feature (right).

3.5 Comparison to Geographic Distance

The topological signature induced by BR dissimilarity, characterized by its persistence diagram, can be contrasted with that induced by geographic straight-line distances. The same process can be applied, instead using straight-line (Euclidean) distances to connect pairs of sites over the filtration instead of BR dissimilarity. However, there are strengths and weaknesses associated with the use of Euclidean distance as a proxy for geographic proximity. Straight-line paths may not accurately reflect true geographic routes most commonly taken between sites, and only serve as a reasonable approximation if certain conditions are met. For a more in-depth treatment of the nuances of this matter, see Supplemental Materials 6.4.

If cultural dissimilarity between sites is largely determined by local physical proximity, then the topological signatures should be expected to be the same, or very close to the same. Similar features should persist in both filtrations. Contradictory findings would provide circumstantial evidence to suggest that cultural dissimilarity is not driven by geographic proximity alone.

4 Results

Topological analysis of the Chaco Canyon and surrounding region reveals distinct cultural patterns present in the data.

Distinct clusters emerge from hierarchical clustering across the dissimilarity range, manifesting as persistent H_0 features (Fig. 4). Notably, two primary clusters emerge around $d = 0.2$. Geospatial analysis reveals that the largest cluster (pink) diffusely covers the greater Chaco Canyon and San Juan Basin area, while the second largest cluster (yellow) is concentrated in the greater Mesa Verde area. The San Juan River appears to divide these two clusters, which aligns with recent work suggesting that the river functioned as a cultural barrier during this general time period.

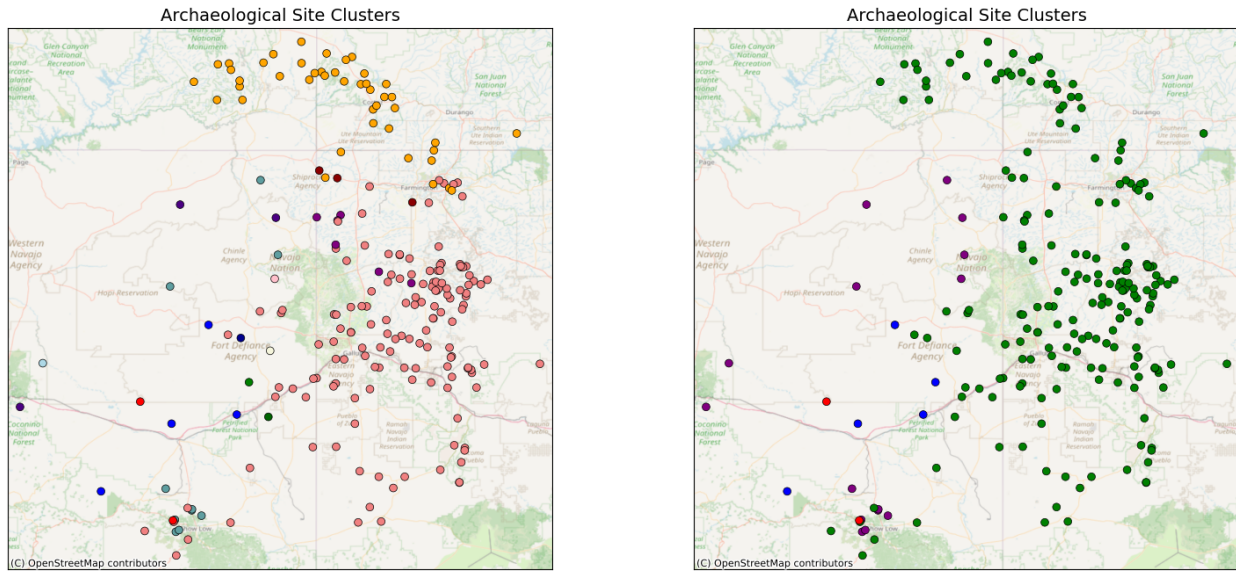


Figure 4: Geographic distribution of sites colored by hierarchical dissimilarity clustering (left: $d = 0.2$; right: $d = 0.3$).

The persistence diagram induced by cultural dissimilarity reveals the scale at which these distinct H_0 clusters emerge (Fig. 5). A giant component emerges after approximately $d = 0.25$, spanning the majority of the region, and only failing to capture peripheral sites on the southwest side of the Chaco corridor (Fig. 4). Interestingly, dissimilarity boundaries are evident at sites peripheral to these two main clusters. The same phenomenon was identified in a social network analysis of data originating from the same region and similar time period by Mills et al. (2018),

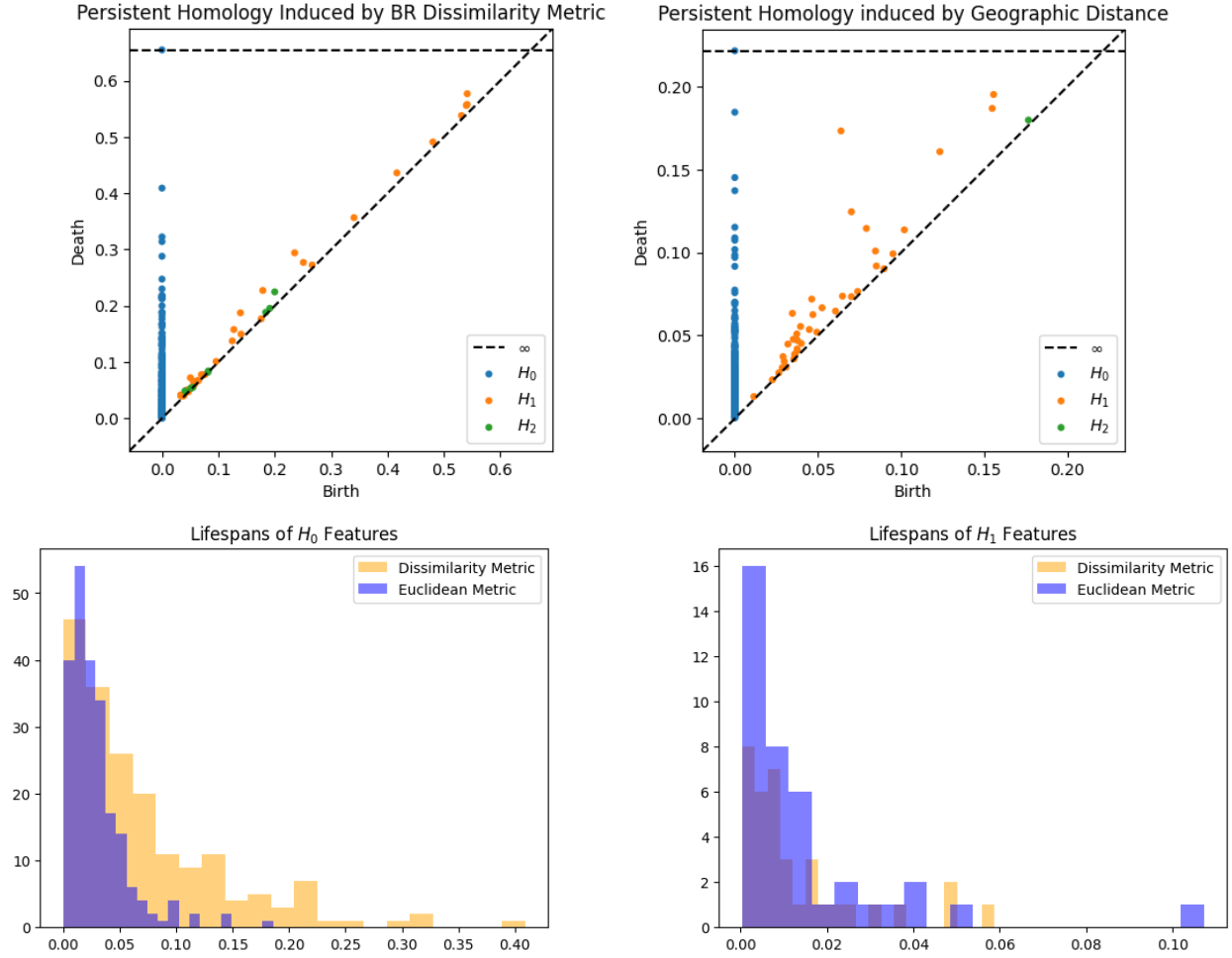


Figure 5: Persistent homology induced by cultural dissimilarity measured by Brainerd-Robinson dissimilarity (top left) and geographic distance measured by Euclidean distance. Tail distributions of feature lifespans for H_0 and H_1 (bottom) reveal significant topological features. H_0 features induced by dissimilarity are much longer lived overall relative to Euclidean distance. A single prominent H_1 feature is induced by Euclidean distance that has no evidence for existence in the dissimilarity landscape, indicating a barrier overcome by cultural similarity. See Supplemental Materials 6.4 for more on Euclidean distance as a measure of geographic distance.

who suggested that this was due to the presence of sites sharing strong similarity to both the northern and southern clusters. I hypothesize that these smaller peripheral clusters serve as topological bottlenecks in the dissimilarity landscape.

Moreover, the small threshold window ($0.25 < d < 0.3$) where the Mesa Verde cluster merges with the Chaco cluster lends support to the theory that the San Juan River region functioned as a barrier to cultural mixing.

The lifespans of the H_1 features detected by persistent homology are very short, remaining close to zero (represented by the line $y = x$). This implies that clusters of similarity form and merge densely across the filtration, indicating that the majority of any true barriers detected are axially-restrictive for sites in this study.

The persistent homology induced on the sites by cultural dissimilarity stands in stark contrast to that induced by geographic distance. Overall, H_0 features are much longer lived in the dissimilarity landscape compared to the geographic landscape. While dissimilarity induces small, insignificant H_1 features, geographically there exists at least one topologically significant H_1 motif.

Interestingly, no significant H_2 features appear in either persistence diagram (Fig. 5). This is a hallmark of networks constrained to locality by planar geometry and a lack of higher-order organization. This would suggest that cultural connections between sites were constrained by the geometry of the topography, and could not make the long-distance connections that would be required to see H_k features with $k \geq 2$.

Overall, the results of persistent homology strongly suggest there is evidence of cultural mixing that may not easily be explained by geography alone without significant cultural conductance or other forces driving cultural transmission in the region.

One possible explanation for the results is a topological analog to the small-world phenomenon explained mathematically by Watts and Strogatz (1998). In general, sites in dense geographic proximity are most likely to have high levels of cultural transmission, with several key sites forming long distance cultural connections spanning disparate clusters.

Sites in the Mesa Verde region tend to be extremely similar to each other compared to sites

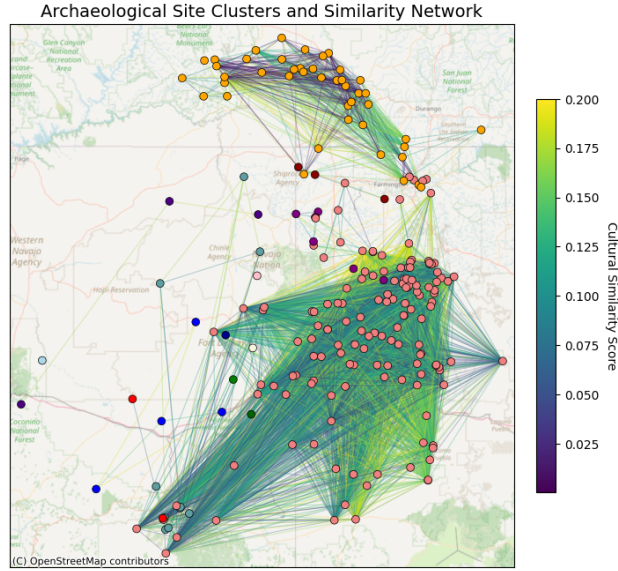


Figure 6: Geographic network with edges colored by dissimilarity. Clusters are colored based on hierarchical clustering at threshold $d = 0.2$.

found in the Chaco region. Most notably, sites in the Shiprock area tend to be highly similar to sites in both the Mesa Verde and Chaco clusters while maintaining central geographic proximity to both, perhaps suggesting that they may have served as a cultural bridge. Interestingly, there are sites to the west and southwest (teal and navy blue) that are also relatively central to both clusters, but are more similar to sites in the deep southwest than any sites that are geographically closer.

A potentially rewarding direction for future research would be to rigorously quantify the spatial anisotropy of site dissimilarity. Exploratory post-hoc analysis of the geometry of dissimilarity indicates that pairs of sites that are collinear in the southeast-northwest direction exhibit significantly higher correlation between cultural dissimilarity and geographic distance compared to sites that are collinear in the southwest-northeast direction (Fig. 6), as measured by Pearson correlation (r) between dissimilarity and geographic distance metrics. A p-value of 0.017 was obtained by a matrix permutation test, confirming statistical significance.

This implies that geographic distance is a stronger predictor of cultural dissimilarity for pairs of sites related in the northwest-southeast direction ($r = 0.69$), and there is a significantly weaker relationship in these two metrics for pairs of sites related in the northeast-southwest direction ($r = 0.57$).

5 Discussion

Persistent homology reveals distinct, interpretable characteristics of the spatial distribution as well as the dissimilarity landscape of sites in the pre-colonial Chaco World. Although it provides a new mathematical perspective, persistent homology recovers many of the same findings identified by researchers employing more common methods of network analysis (Mills et al., 2018), providing some validation for its application.

Persistent homology has the potential to serve as a robust alternative to traditional network analysis in identifying gaps, loops, or voids in the data. This analysis was limited to a basic application of the method, and many rich variations and supporting mathematical concepts exist for further testing. For example, one extension of persistent homology that may be viable in archaeological data analysis is multiparameter persistent homology (Schenck, 2022), where the homology can be computed across filtrations of both space and dissimilarity simultaneously. Additionally, further analysis could account for the dynamics of time, analyzing the cultural properties of the dissimilarity landscape as a time-dynamic process.

Archaeological network analysis derives many of its methods from Social Network Analysis and other interdisciplinary sources, but the questions and problems of this field remain unique. While the incorporation of interdisciplinary methodology is inevitably a healthy development for the field, I believe that conversely, archaeological contexts provide a unique exploratory playground for the development of new quantitative methodologies that may see broader use across the social sciences and beyond. I encourage mathematicians and computational scientists to fully embrace the unique challenges that archaeological data presents and discover utility in its solutions.

6 Supplemental Materials

6.1 Source Code

The code used in this analysis is available on Github, located at

<https://github.com/andrew-belz/persistent-homology-ASNs/>

.

6.2 Metric Space Topology

Persistent homology can be applied to the set of archaeological sites S because equipping the set with the BR-dissimilarity metric induces a valid *metric space*. A metric space is simply any set of points S together with a *metric*. Similarity may be defined in many different ways for many different purposes (Prignano et al., 2017), and so too can a metric. I include this section to lay the necessary groundwork for the development of new dissimilarity metrics, to which persistent homology can be applied.

In this application, I embed S in the vector space $\mathbb{R}^n$, where each archaeological site x has a vector representation with components x_i corresponding to the percentage of material type i recovered at site x . This naturally leads to a data matrix representation of S , with rows corresponding to sites and columns corresponding to material types.

6.2.1 Axioms of a Metric Space

There are three primary axioms that a function $d : S \times S \rightarrow [0, \infty)$ must satisfy to be considered a metric. For any $x, y, z \in S$, the following must be satisfied:

1. $d(x, y) \geq 0$ (Non-negativity), with $d(x, y) = 0$ if and only if $x = y$.
2. $d(x, y) = d(y, x)$ (Symmetry).
3. $d(x, y) \leq d(x, z) + d(z, y)$ (Triangle Inequality).

While the notion of a metric space is actually less strict than that of a vector space, a vector space equipped with a norm always induces a metric space. This is because norms measure vector length, and every norm induces a metric by considering the distance between two vectors to be the length of their difference vector, computed component-wise.

A very common metric used is the standard Euclidean metric that measures the length of a straight line drawn between two points in Euclidean space. The Euclidean metric belongs to a larger class of metrics that are derived from an important class of vector norms called the L_p norms, denoted $\|\cdot\|_p$. Note that the Euclidean metric is derived from the L_2 norm. That is, the Euclidean metric is defined as

$$d(x, y) = \|x - y\|_2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

Another commonly used L_p norm is the L_1 norm, also called the taxicab norm, or Manhattan norm in some contexts. The L_1 metric is defined as

$$d(x, y) = \|x - y\|_1 = \sum_{i=1}^n |x_i - y_i|.$$

Intuitively, this norm measures the distance traveled if each axis must be traveled separately. This is analogous to the distance a taxicab must travel on city streets situated in a north-south and east-west grid orientation.

In fact, BR dissimilarity is simply the L_1 metric between site material distributions. Here, the L_1 norm of a site x measures the total percentage of material quantity recovered at the site, which is always 100. The metric induced by the L_1 norm measures the difference in material quantities between two different sites, compared across each type i . The resulting difference is summed to determine a measure of total differences between the two sites.

Recall the definition of BR similarity coefficient for sites x and y , where each vector component x_i corresponds to the percentage of material at site x of type i :

$$BR(x, y) = 200 - \sum_{i=1}^n |x_i - y_i|.$$

Mathematically, BR dissimilarity d is defined as

$$\begin{aligned} d(x, y) &= 200 - BR(x, y) \\ &= \sum_{i=1}^n |x_i - y_i| \\ &= \|x - y\|_1. \end{aligned}$$

A simple check of the axioms reveals that a metric remains a metric when a multiplicative scaling factor is applied, and in fact induces an equivalent topology. When comparing two different metric spaces induced on the same set, distances should be normalized to equivalent scales so that persistence diagrams derived from different metrics can be accurately compared. Ergo, in this analysis BR dissimilarity is scaled to the range $[0, 1]$ by dividing each observation by the maximum BR dissimilarity value, and Euclidean distances are scaled to this range in the same manner.

Thus, BR dissimilarity provides a mathematically valid notion of cultural distance between two archaeological sites, to which persistent homology can be rigorously applied.

6.3 Persistent Homology

For a motivated, thorough introduction to applied Topological Data Analysis (TDA), see Kemme and Agyingi (2025). The authors provide a comprehensive, insightful perspective on the application of persistent homology and other common tools of TDA to the biological sciences, although the mathematical machinery is universal. Crucially, they develop rigorous mathematical definitions that are necessary for the formalization of TDA. In this section, I rely on the formalization of Kemme and Agyingi and introduce the most pertinent critical ideas necessary to understand the application of persistent homology to study ASNs.

6.3.1 The Vietoris-Rips Complex

First, I extend the notion of a graph to a simplicial complex, which allows us to represent higher-order relationships among groups of vertices. A *clique* is a subset of vertices in a graph in which every pair of vertices is connected by an edge. Each clique is represented as a geometric object called a *simplex*: a clique of size two becomes an edge (1-simplex); a clique of size three becomes a filled triangle (2-simplex), and a clique of size four becomes a filled tetrahedron (3-simplex). A lone vertex is itself a 0-simplex. The collection of all simplices obtained from the cliques of a graph G is called the *clique complex* of G . Considering cliques in G to be a single, functional group represented by a simplex allows us to analyze higher-order geometric and topological structure implied by patterns of pairwise similarity.

The Vietoris-Rips complex, denoted $VR_\varepsilon(G)$, is simply the clique complex of the graph G that is obtained by connecting every pair of vertices that have a BR dissimilarity of at most ε . By initializing $\varepsilon = 0$, and letting ε increase, one obtains a sequence of nested Vietoris-Rips complexes

$$VR_{\varepsilon_1} \subseteq VR_{\varepsilon_2} \subseteq \cdots ,$$

which, all together, capture the structure of the similarity landscape of the archaeological sites spanning a range of scales, called a filtration.

Indeed, this notion captures a sense of scale, or resolution, because the filtration parameter ε controls which distances are considered "close" together.

Unlike traditional ASN representations, the Vietoris-Rips complex allows us to consider higher-order interactions between sites that simple graphs cannot encode. As a result, holes that appear significant in a graph may disappear once densely interconnected groups are represented as simplices, allowing only the consideration of holes that persist after accounting for higher-order structure.

6.3.2 Homology

The advantage of looking at the Vietoris-Rips complexes over the range of the filtration parameter is that tracking the topological properties of the complex provides some insight into the underlying shape of the similarity landscape. Homology, a concept originating from algebraic topology, provides a way to look at the holes of a space in various dimensions. I forego a formal discussion of homology in favor of developing the intuition, but encourage the reader to see Kemme and Agyingi for a rigorous formalization of homology.

Naively, persistent homology is concerned with identifying the fundamental, independent loops formed by connected chains of simplices that enclose an empty void, or gap.

These holes are organized according to dimension. An H_0 feature corresponds to a connected component, an H_1 feature corresponds to a loop or cycle of edges, and an H_2 feature corresponds to a three-dimensional cavity or void. While higher-order homology groups are mathematically valid, I restrict my discussion of homology to the first three groups, which are most interpretable because they can be intuitively visualized.

As depicted in Figure 3, an H_1 feature appears as a closed cyclic loop. An H_2 feature appears as a 3-dimensional void enclosed by triangles. The H_0 features require a slightly different interpretation: each H_0 feature corresponds to a connected component of the graph. When the filtration parameter is initialized to $\varepsilon = 0$, every vertex begins as an H_0 feature, each a single, isolated component.

6.3.3 The Persistence Diagram

An H_k feature is considered to be born at $t_{\text{birth}} = \varepsilon$ if it first appears in the complex $VR_\varepsilon(G)$ corresponding to the parameter ε . An H_k feature is considered to die at time $t_{\text{death}} = \varepsilon$ when it is no longer present in $VR_\varepsilon(G)$. For example, all individual vertices are born at $t_{\text{birth}} = 0$, each forming an H_0 feature. However, as pairs of vertices connect as ε increases, the number of connected components shrinks. Every time two components join, an H_0 feature dies. An H_1 feature dies when it becomes filled with higher-dimensional simplices or merges with another

feature to become a new higher-dimensional feature.

The lifespan of a feature is given by $t_{\text{death}} - t_{\text{birth}}$. Long-lived features are considered to be more robust than short-lived features, which are more likely to exist due to local noise or small-scale fluctuations.

One of the best ways to visualize the lifespans of these features is in the plane, with birth times t_{birth} on the x -axis and t_{death} on the y -axis. Because every feature must be born before it can die, every point will lie strictly above the line $y = x$. Features with long lifespans will be significantly higher above this line than features with very short lifespans. The resulting plot is called a *persistence diagram*.

Examining persistence diagrams can provide some remarkable insight into the topological structure of the data. For example, short-lived H_0 features that appear early on in the filtration show groups of sites with strong pairwise similarity. If simultaneously there are persistent H_1 features in this interval, this means that while dense local clusters are forming, globally these clusters are disparate enough that voids in the similarity landscape are allowed to persist for long periods of time. By looking at the persistence diagram, one can gain a sense of what the global similarity landscape looks across all sites.

6.4 Least-cost Path Length as Geographic Distance

In many cases, archaeologists are more concerned with understanding the paths that humans would have been most likely to take in their environment, which may be different from straight-line (Euclidean) paths. The topographic features of a region deeply influence the way that humans move through their environment, as they are assumed to be governed by the principle of *least cost*. The fundamental assumption is that humans will seek routes that minimize time and energy expenditure, avoiding geographic hurdles like steep mountainous ascents and dangerous river crossings. A reasonable measure of physical distance in most sociological contexts should account for routes of least-cost, rather than just geographic straight-line distances.

To address this, I obtained a Digital Elevation Model (DEM) from the United States Geological

Survey (2026) measuring elevation at a resolution of 10 arc-seconds (approximately 308 meters), and computed least-cost paths for every pair of sites using Tobler's hiking function (Tobler, 1993). I found that for this region, least-cost path length was nearly perfectly correlated with straight-line distances ($r > 0.99$), so I find it reasonable to use the latter as a measure of effective geographic distance when the alternative is least-cost paths computed from a DEM. Additionally, straight-line distance is a valid metric, and least-cost-path distance may not be, simplifying the mathematical justification.

A more robust least-cost path calculation should incorporate geographic features that are not effectively measured by elevation, such as bodies of water or weather-related factors. Distinct barriers emerge as H_0 features that are not accounted for by Euclidean distance (Fig. 5) and the failure to account for other topographical features could confound the results. While the results can be directly interpreted in this context as geographic barriers, further development of a better least-cost metric is required to truly study the topological properties of dissimilarity as a true diffusion process.

A future refinement of this analysis should examine these factors along with the use of straight-line distances to model human travel in this region and time period in order to more rigorously scrutinize the use of straight-line distances. Importantly, least-cost path lengths are typically not valid metrics, requiring extra care for the mathematical details.

References

- Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., ... Petri, G. (2020). Networks beyond pairwise interactions: structure and dynamics. *Physics Reports*. doi: 10.1016/j.physrep.2020.05.004
- Borck, L., Mills, B. J., Peeples, M. A., & Clark, J. J. (2015). Are social networks survival networks? An example from the late pre-Hispanic US Southwest. *Journal of Archaeological Method and Theory*. doi: 10.1007/s10816-014-9236-5
- Brainerd, G. W. (1951). The place of chronological ordering in archaeological analysis. *American Antiquity*. doi: 10.2307/276979
- Brughmans, T., Keay, S., & Earl, G. (2014). Introducing exponential random graph models for visibility networks. *Journal of Archaeological Science*. doi: 10.1016/j.jas.2014.05.027
- Brughmans, T., & Peeples, M. A. (2023). *Network science in archaeology*. Cambridge, UK: Cambridge University Press. doi: 10.1017/9781009170659
- Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society*. doi: 10.1090/S0273-0979-09-01249-X
- Cohen-Steiner, D., Edelsbrunner, H., & Harer, J. (2007). Stability of persistence diagrams. *Discrete & Computational Geometry*. doi: 10.1007/s00454-006-1276-5
- Collar, A., Coward, F., Brughmans, T., & Mills, B. J. (2015). Networks in archaeology: Phenomena, abstraction, representation. *Journal of Archaeological Method and Theory*. doi: 10.1007/s10816-014-9235-6
- Edelsbrunner, H., Letscher, D., & Zomorodian, A. (2000). Topological persistence and simplification. In *Proceedings 41st annual symposium on foundations of computer science*. doi: 10.1109/SFCS.2000.892133
- Gjesfjeld, E. (2015). Network analysis of archaeological data from hunter-gatherers: Methodological problems and potential solutions. *Journal of Archaeological Method and Theory*. doi: 10.1007/s10816-014-9232-9
- Kemme, A. J., & Agyingi, C. A. (2025). *Persistent homology: A pedagogical introduction with*

- biological applications*. (Preprint) doi: 10.48550/arXiv.2505.06583
- Majhi, S. (2021, October). *Vietoris-rips complex*. Retrieved from <https://observablehq.com/@sushovan4/vietoris-rips> (Accessed: 2026-06-17)
- Mills, B. J., Peeples, M. A., Aragon, L., Bellorado, B., Clark, J. J., Giomi, E., & Windes, T. (2018). Evaluating Chaco migration scenarios using dynamic social network analysis. *Antiquity*. doi: 10.15184/aqy.2018.86
- Peeples, M. A., & Bischoff, R. J. (2023). Archaeological networks, community detection, and critical scales of interaction in the u.s. southwest/mexican northwest. *Journal of Anthropological Archaeology*. doi: 10.1016/j.jaa.2023.101511
- Peeples, M. A., & Brughmans, T. (2023). *Online companion to network science in archaeology*. Retrieved from <https://archnetworks.net> (Accessed: 2026-06-16)
- Prignano, L., Morer, I., & Díaz-Guilera, A. (2017). Wiring the past: A network science perspective on the challenge of archeological similarity networks. *Frontiers in Digital Humanities*. doi: 10.3389/fdigh.2017.00013
- Robinson, W. S. (1951). A method for chronologically ordering archaeological deposits. *American Antiquity*. doi: 10.2307/276978
- Schenck, H. (2022). Multiparameter persistent homology. In *Algebraic foundations for applied topology and data analysis*. Springer. doi: 10.1007/978-3-031-06664-1
- Tobler, W. (1993). Three presentations on geographical analysis and modeling: Non-isotropic geographic modeling; speculations on the geometry of geography; and global spatial analysis. *NCGIA Technical Report*. Retrieved from <https://escholarship.org/uc/item/05r820mz>
- United States Geological Survey. (2026). *Tnm downloader - the national map*. Retrieved from <https://apps.nationalmap.gov/downloader/> (Accessed: 2026-06-16)
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*. doi: 10.1038/30918