Topological Data Analysis in Urban Planning

Joenathan Halim, Jr (U2340760K)

Supervisor: Dr. Fedor Duzhin

Abstract

This poster briefly explain the application of algebraic topology in urban planning to study behavioral patterns of a city over time. The goal of this poster is to extract these patterns using topological data analysis implemented in R.

1 Basics of Algebraic Topology

In algebraic topology, one makes a rigorous treatment of intuitive concepts such as the number of holes in a geometric structure by assigning algebraic structures to topological spaces - see Figure 1.

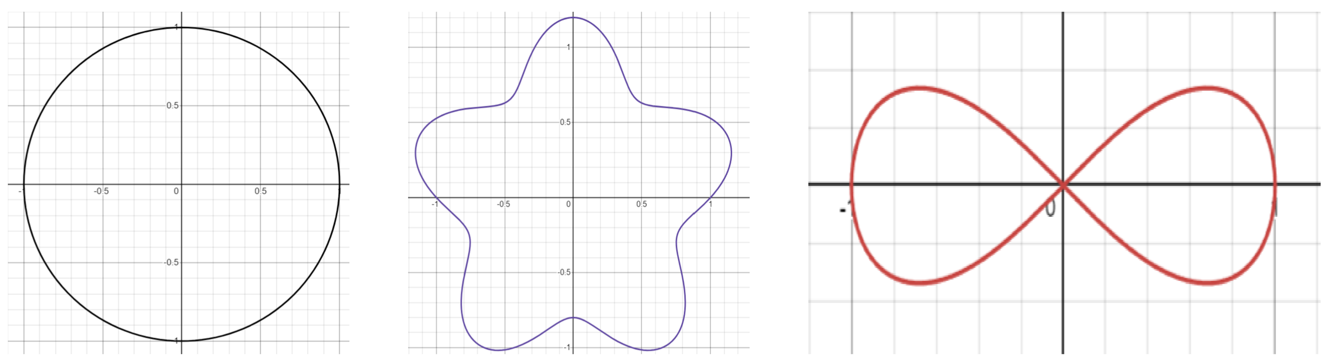

Figure 1: Graph of a circle, a random closed curve, and a wedge sum of two circles, i.e., a figure eight

A method to calculate holes in different dimensions within a topological space is called *homology*. As seen in Figure 1, a circle and a random closed curve has a one dimensional hole, while a figure eight has two dimensional holes.

2 Topological Data Analysis

2.1 From a dataset to a topological space

The input of topological data analysis is a finite point cloud with known distances between every two points. This can represent various data types, including images, sound waves, text, etc. Then, a *Vietoris-Rips* complex is constructed on the point cloud to build a topological space by connecting points with simplices (line, triangle, tetrahedron, etc.) whenever pairwise distances are less than a threshold $\delta > 0$ [1]. By adjusting the values of δ, various topological spaces are formed, as depicted in Figure 2.

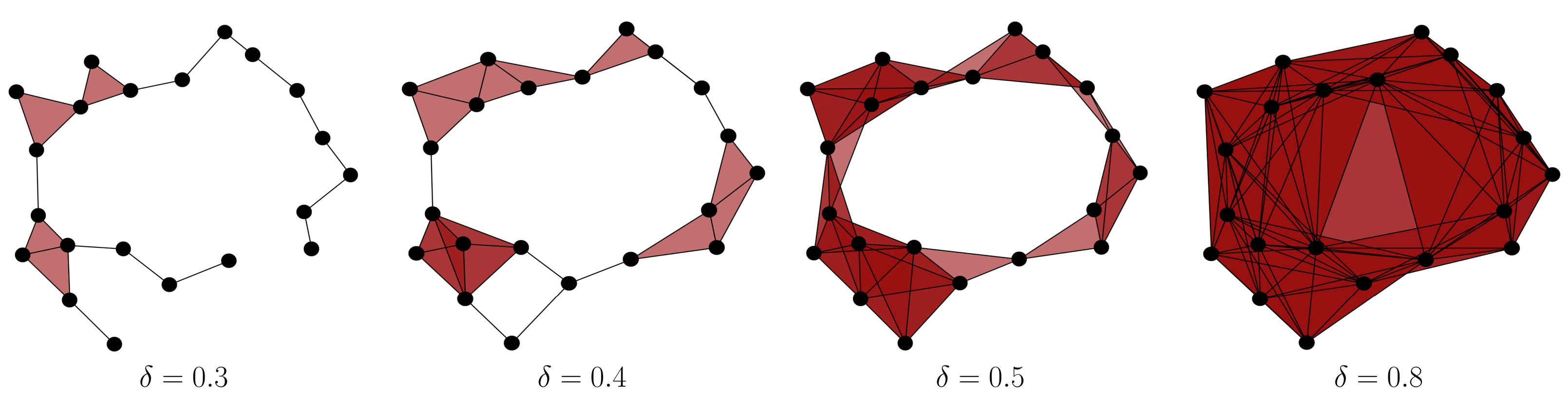

Figure 2: Vietoris-Rips complexes constructed from a point cloud with varying values of the threshold δ.

2.2 Persistent Homology

Topological data analysis is concerned with the persistent features that are created over various δ. One such feature is homology. We use *Persistent Homology* to keep track of homology by putting it into a diagram called a *barcode*. The key observation in persistent homology is that long lines represent holes in the specified dimension, while short lines are irrelevant noise, as shown in Figure 3.

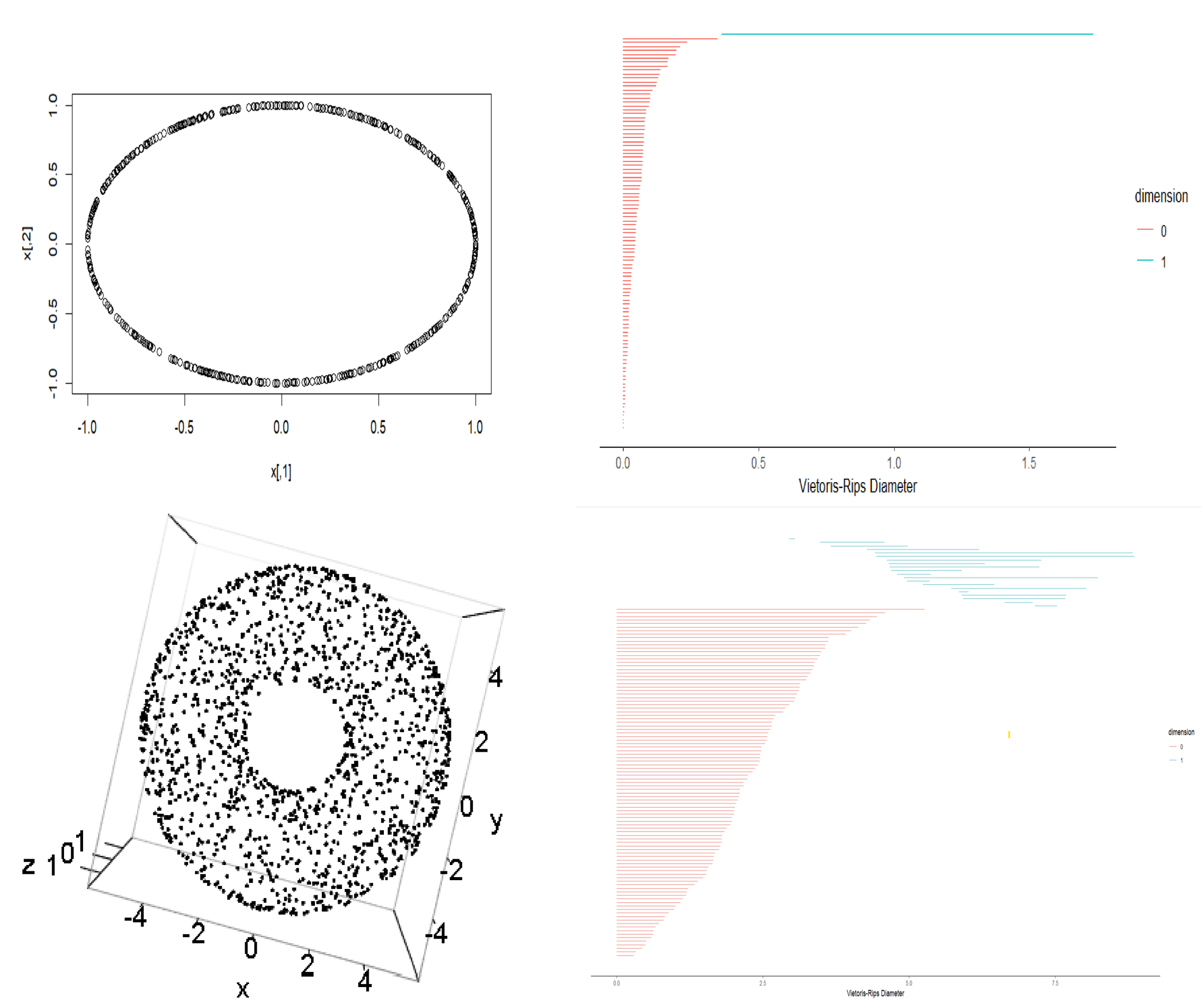

Figure 3: Datasets scattered forming a circle and a torus, with their respective barcodes.

3 Urban Planning

Given a dataset of geotagged photos from Flickr comprising latitude, longitude, and epoch time, in New York City, one can construct a density function at location p to quantify activity levels.

$$f(p) = \sum_{x_i \in N(p)} e^{\frac{-d(p,x_i)^2}{\epsilon^2}} \tag{1}$$

The resulting dataset is a point cloud where the function value correspond to its y-coordinate - refer to Figure 4.

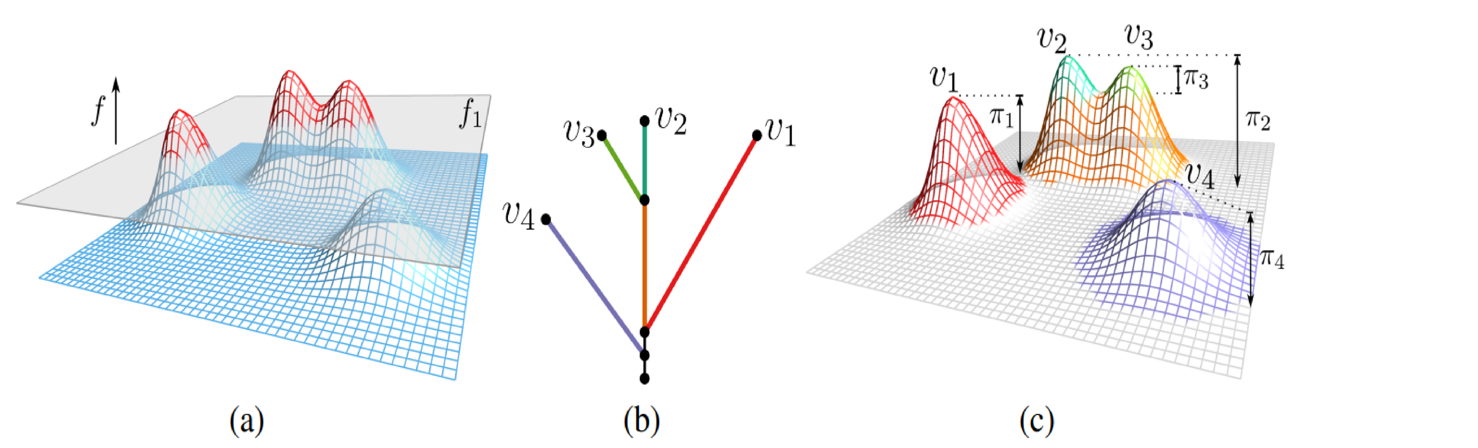

Figure 4: Topology of scalar functions. **(a)** The height function f defined on a graph. f_1 is the set of all points above the highlighted plane (sweep of the function value in decreasing order) colored red. **(b)** Join tree of f. **(c)** The label peaks denote the set of maxima. π_i denotes the persistence of maximum v_i.

To identify locations with high activity levels at different times, We observe changes in the number of connected components of Figure 4 (a) as the function value decreases from the maximum. The number of connected components emerge at a maxima, preserved at regular points, and merge at saddle points. Persistent homology track these changes, as shown in Figure 4(b) and (c).

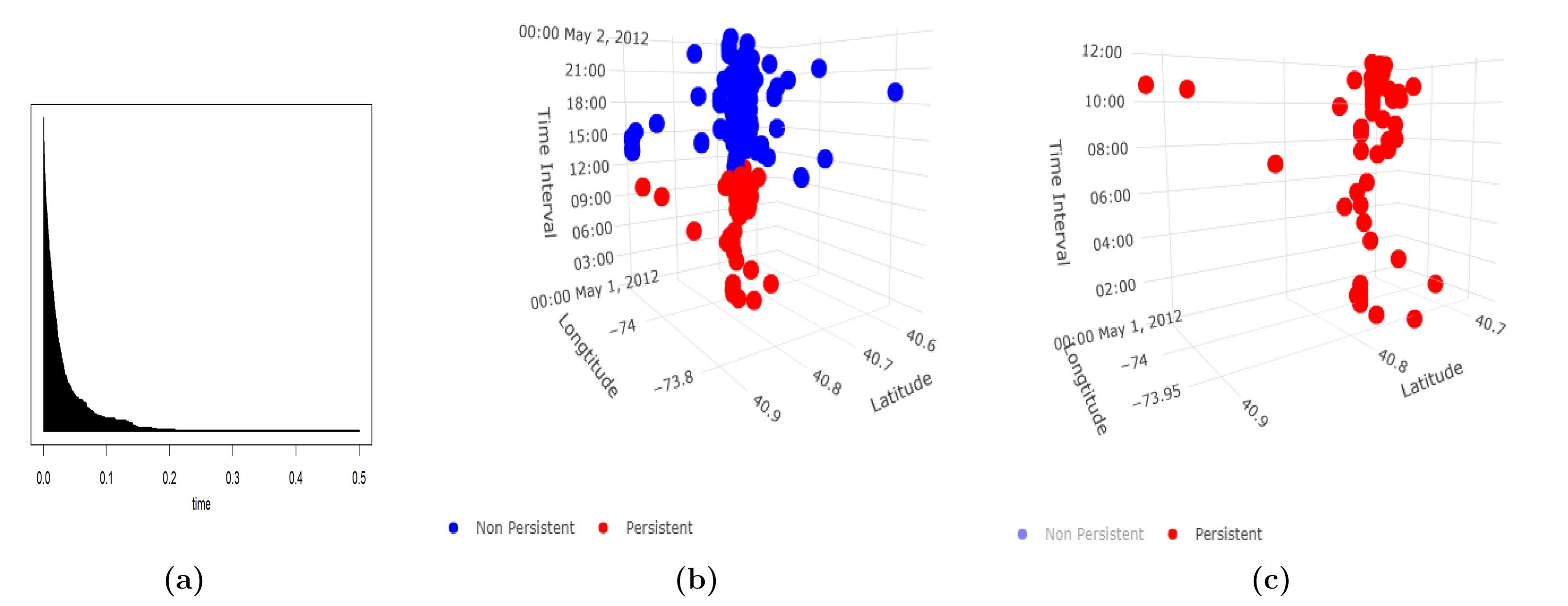

Figure 5: (a) Persistence Barcode, (b) Spatial Temporal visualization of the activity level of New York City over hourly time steps. (c) The red dots denotes the locations with high persistence value

The result of persistent homology is a barcode, shown in Figure 5 (a). Identifying the vertices that has high persistence value indicates locations with high activity over hourly time steps, as illustrated in Figure 5 (b) and (c). A software visualization of Figure 5 is depicted in Figure 6.

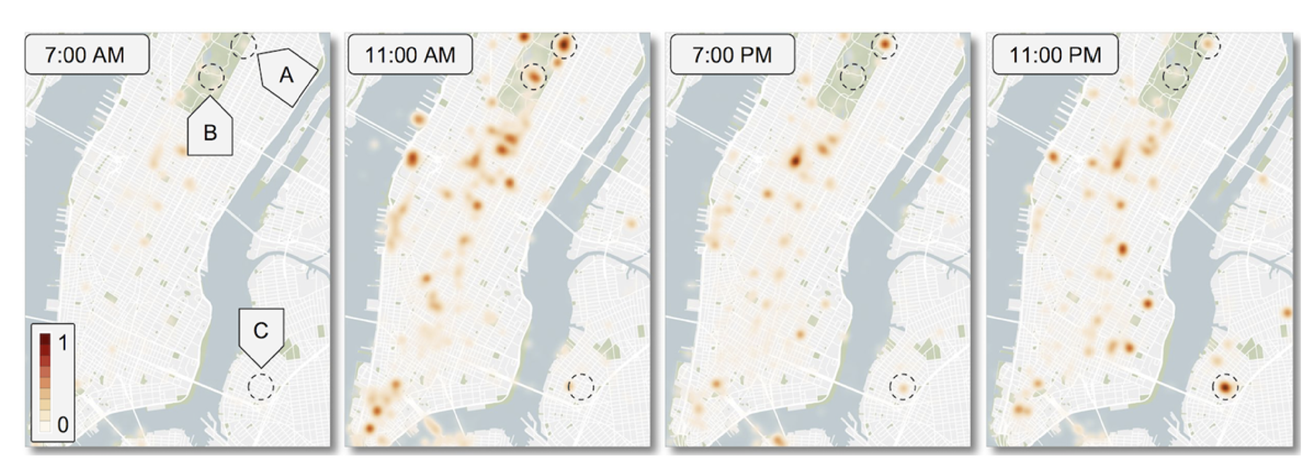

Figure 6: Density function computed on the Flickr dataset at different time steps with respect to New York City across region A (Metropolitan Museum), region B (Central Park), and region C (Williamsburg). [2]

4 Conclusion

In conclusion, the topological data analysis method implemented above has allowed us to analyze a city by identifying significant clusters of urban activity that persist over time. The visualization of these behavior across different times reveal patterns of activity that may be crucial for urban planning.

References

[1] Ran Deng and Fedor Duzhin. Topological data analysis helps to improve accuracy of deep learning models for fake news detection trained on very small training sets. *Big Data and Cognitive Computing*, 6(3):74, 2022.

[2] Fabio Miranda, Harish Doraiswamy, Marcos Lage, Kai Zhao, Bruno Gonçalves, Luc Wilson, Mondrian Hsieh, and Cláudio T Silva. Urban pulse: Capturing the rhythm of cities. *IEEE transactions on visualization and computer graphics*, 23(1):791–800, 2016.

School of Physical and Mathematical Sciences

Odyssey Research Programme