NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Topological Data Analysis in Urban Planning

Joenathan Halim, Jr (U2340760K)

Supervisor: Dr. Fedor Duzhin

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.

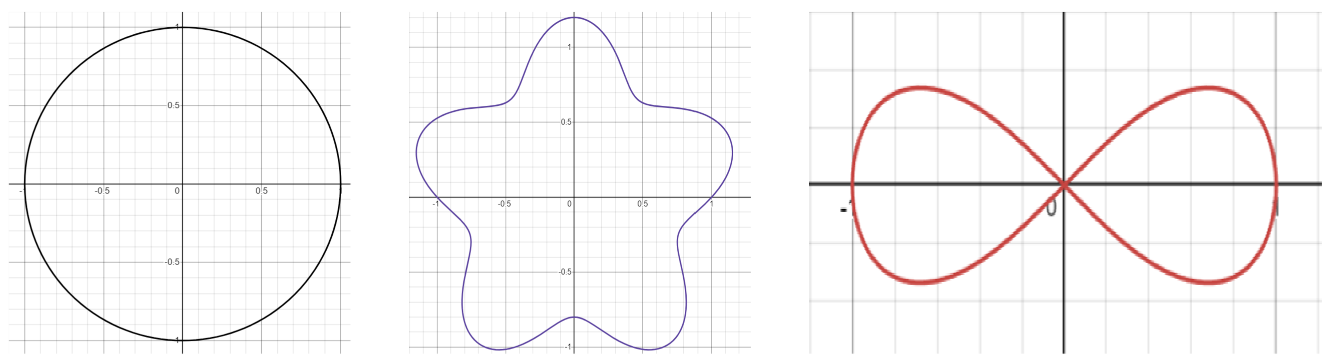

Figure 1: Graph of a circle, a random closed curve, and a wedge sum of two circles, i.e., a figure eight

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, representing 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.

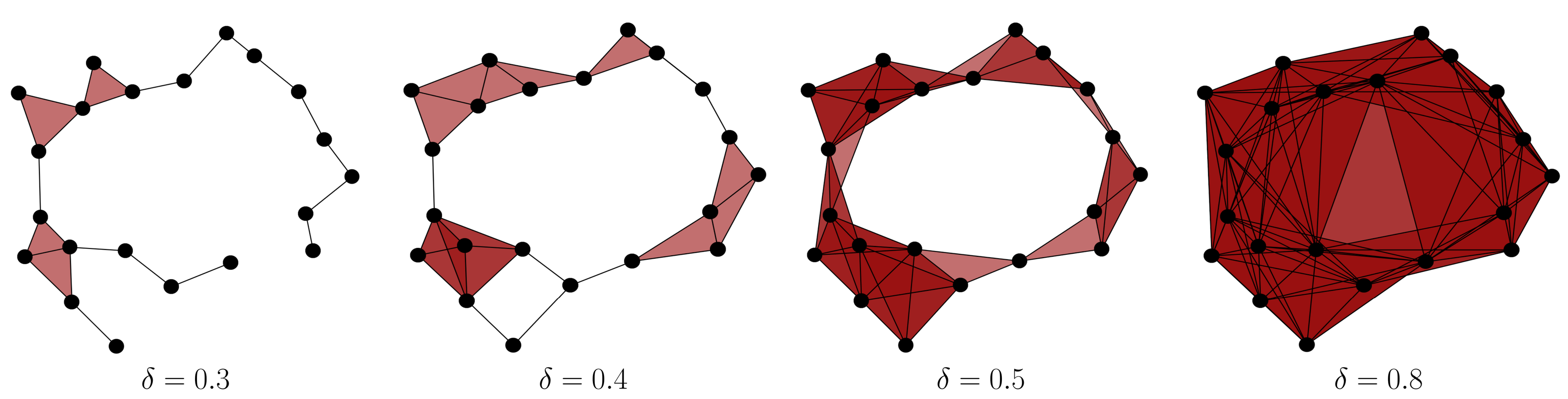

Figure 2: Vietoris-Rips complexes constructed from a point cloud with varying values of the threshold δ.

2.2 Persistent Homology

To calculate and keep track of these holes as δ varies, we use *Persistent Homology*, which puts all these information 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.

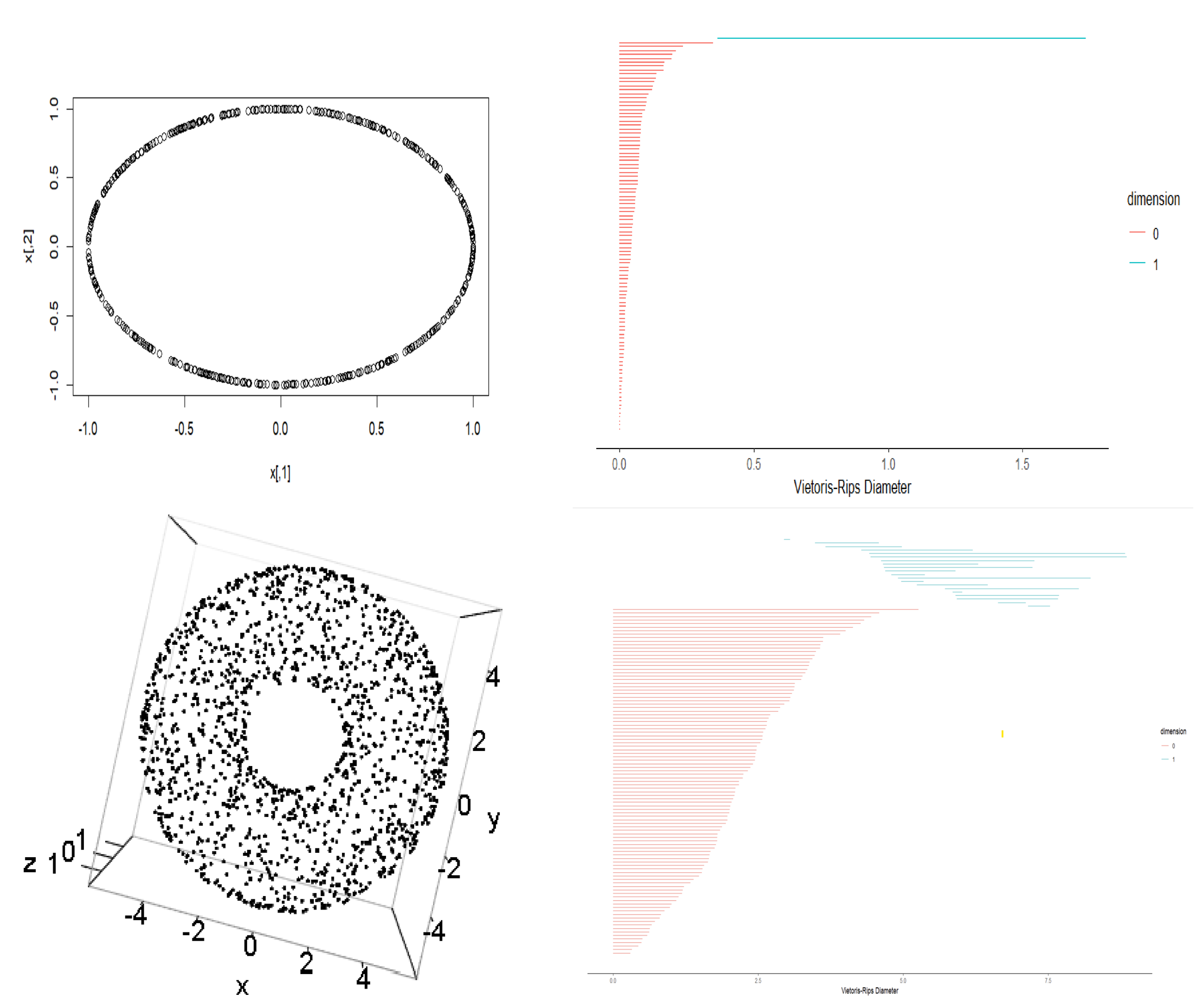

Figure 3: Datasets scattered forming a circle and a torus, with their respective barcodes.

3 Urban Planning

Given a dataset of photos with location and time in a planar domain, one can construct a density function at location p to measure the level of activity in said location that maps the planar domain into real values.

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

The resulting dataset is a point cloud containing the location of the picture paired with the density function value representing the activity level, and is then grouped into a discrete time steps to identify regions with high activity levels, as in Figure 4.

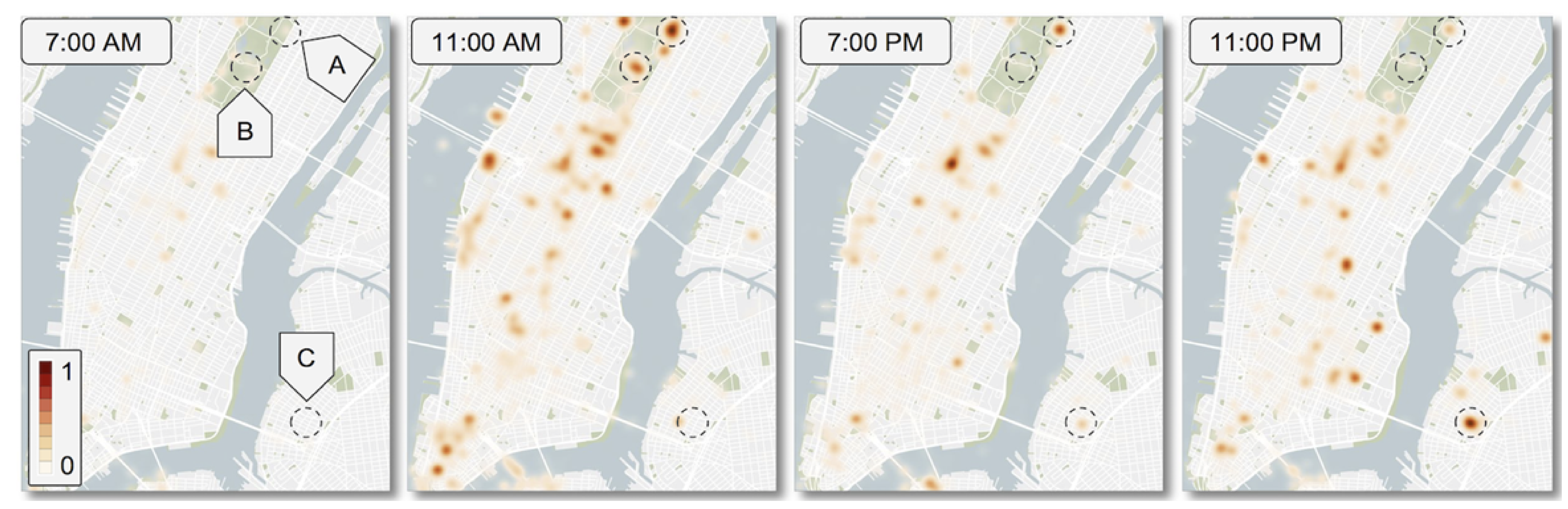

Figure 4: Density function computed on the Flickr dataset at different time steps with respect to New York City across region A (Metropolitan Museum), region B (Central Park), and region C (Williamsburg). [2]

The function is then represented as a piece-wise linear function, while the planar domain is represented by a triangular mesh with the function as the vertices of the mesh. Now, we will consider region A, B, and C from figure 4 to calculate the persistent homology to see when these locations become prominent, i.e., high in activity - see Figure 5.

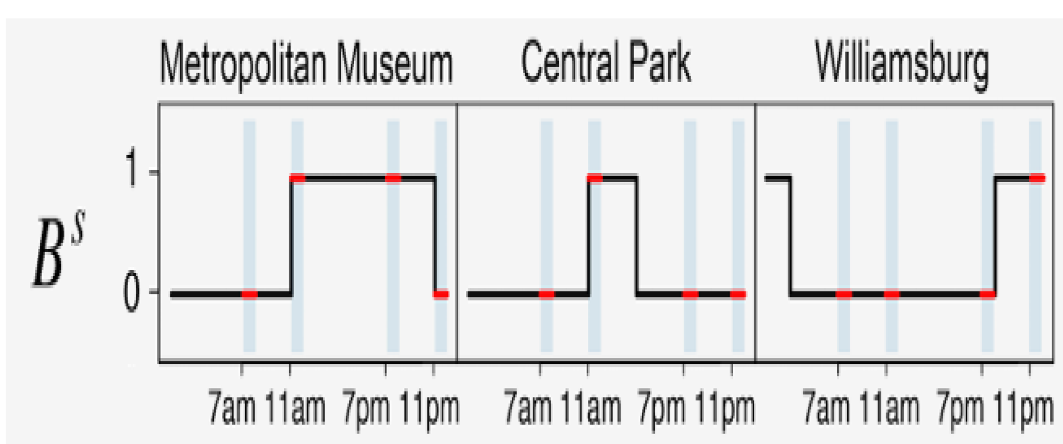

Figure 5: Persistent homology is calculated on the density function across region A (Metropolitan Museum), region B (Central Park), and region C (Williamsburg) to see how frequently the region has high activity. [2]

The aforementioned method has served as a framework in urban planning by defining the notion of beats (B^s), i.e., level of activity in an area, and urban pulses, i.e., a set of beats in a region - refer to Figure 6.

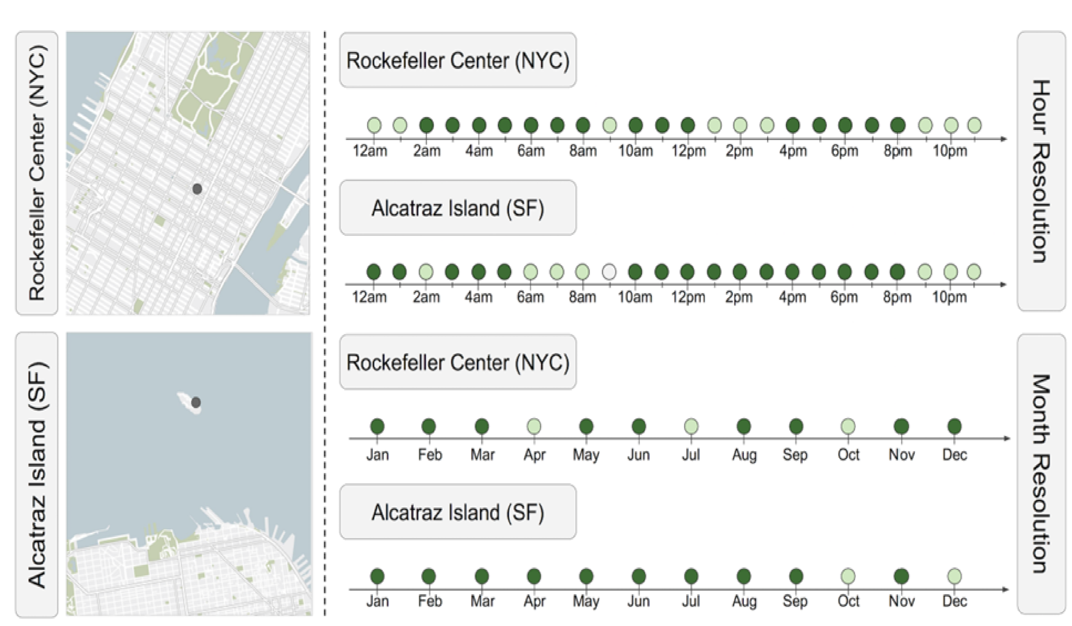

Figure 6: Comparing the pulses of Rockefeller Center (NYC) and Alcatraz Island (SF) over an hourly and monthly resolution, based on Flickr activity. Dark green beats represent a notably high activity at the location, while light green beats represent proportionally high activity in comparison with its neighbouring areas.[2]

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.