

# TOPOLOGICAL DATA ANALYSIS APPLIED TO ANT NETWORKS

Adam Banatwala and Esther Rønn  
Providence College



## Ant Colonies

Our project utilized Topological Data Analysis (TDA) to analyze topological features in ant colonies in a combination of mathematics and biology. We wanted to explore the mechanics behind ant group movement. By looking at the colony as a network analyzed via TDA we were able to compare the movement to various hypotheses. The experimental data for the ant colonies was provided by Prof. James Waters and Prof. Xiaohui Guo [1].

## Topology

### VR Complexes

For our analysis using TDA, we elected to focus on the Vietoris-Rips Complex. Under this construction, if data points are within a specified proximity parameter of each other, they are considered a singular connected component. (This is a simplified explanation, please inquire for further detail. Truly, VR Complexes construct higher dimensional spaces, called simplices, out of these pairwise-connected points, but for our purposes this can be mostly abstracted.)

Let  $S =$  some finite set of points in  $\mathbb{R}^n$ , and let proximity parameter  $r \in \mathbb{R}, r > 0$

The Vietoris-Rips complex is defined as:

$$VR(S, r) = (S, \Sigma_r = \{\sigma \subseteq S \mid \text{distance}(x, y) \leq r, \forall x, y \in \sigma\})$$

$VR(S, r)$  builds a higher dimensional space from the points in  $S$ , consisting of edges, triangles, pyramids, and higher-dimensional analogs. The generalized theoretical construction of the complexes is a wide topic, but over the course of our project we primarily utilized Betti-0 numbers, which can be generally defined as,

$$\text{Betti-}p = |Z_p / B_p|$$

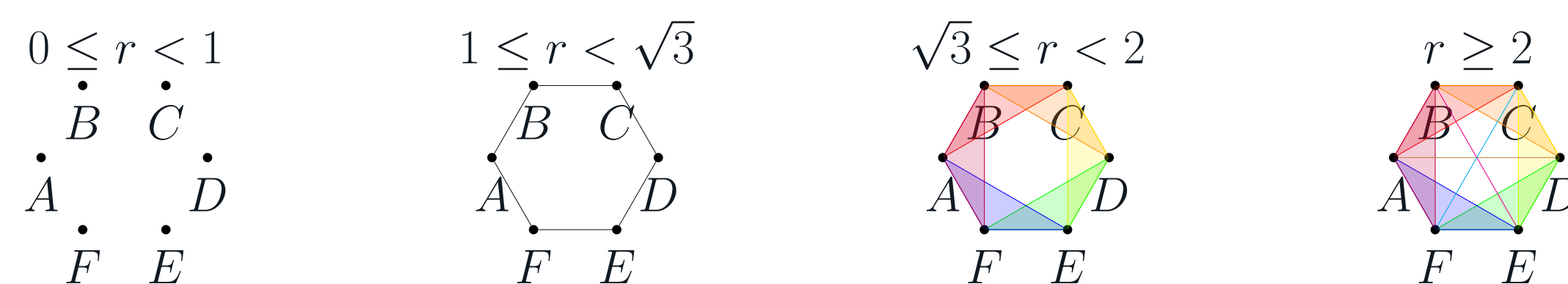
where  $Z_p = \{a \in c_p \mid \delta(a) = 0\}$  and  $B_p = \{b \in c_p \mid \exists c \in c_{p+1}, \delta(b) = c\}$

Functionally,  $c_p$  is a group of  $p$ -simplexes of dimension  $p$ , and  $\delta$  is a defined boundary operation. Hence,  $B_p$  is the image of  $c_{p+1}$  and  $Z_p$  is the kernel of  $\delta$ , both under  $\delta$ .

For our purposes, Betti-0 numbers can be thought of as the number of connected components of the network. Examining this statistic over the proximity parameter allows us to examine the persistent homology of the network in each frame, giving us more insight into the ants' movement. [3]

### Example

The following illustrates the construction and baseline analysis of a Vietoris-Rips Complex, for a simple choice of  $S$  as the six vertices of a hexagon.



### Frobenius Norms

Our analysis also utilized Frobenius norms, which are matrix norms that we used to compare the topological features of the different data sets.

The Frobenius norm of a matrix  $A$  can be defined as,

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n (a_{ij})^2}$$

such that  $a_{ij}$  iterates through every element of the matrix.

## Process

We began our project by analyzing the experimental data. The experimental data is from an ant colony of 61 ants, and has been divided into two data sets. Each is a 162 second recording (with 30 frames per second), with the first consisting of unalarmed ants as a baseline and the second being the same ants in an alarmed state. The alarm division is from a biological perspective based on the ants' speeds [1]. The persistent homology at each frame is described by barcodes. Below is the barcode of the first frame of the unalarmed data. Each bar represents one connected component in the ant network. As the proximity parameter increases, more ants are connected; thus the number of connected components decreases.

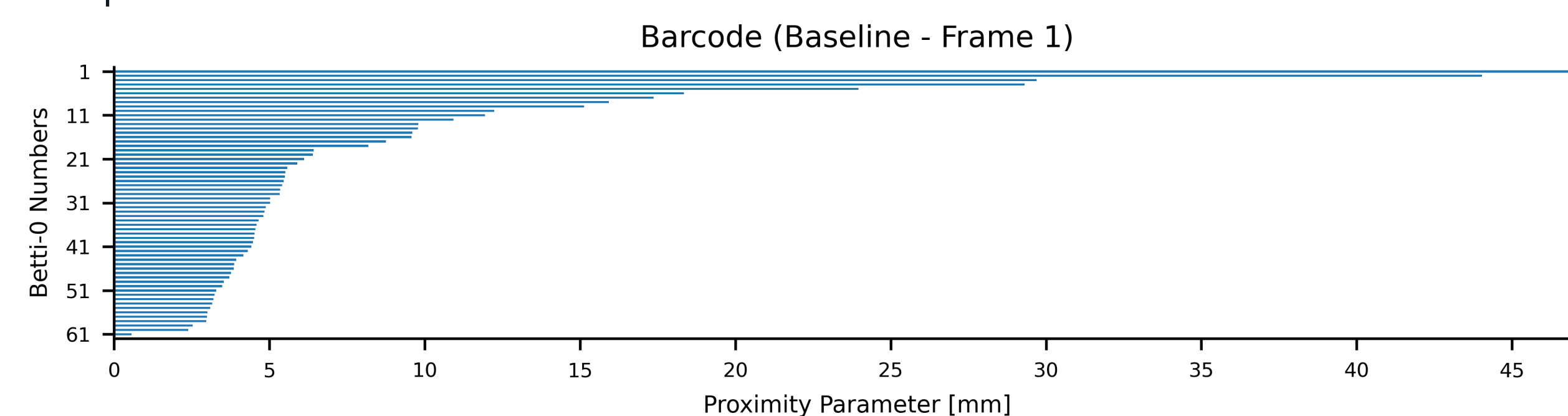


Fig. 1: An example of a barcode plot from the data

How persistent homology varies over time can be visualized using a Crocker plot [2]. Each x-coordinate represents a frame. The y-values represent the barcode of that corresponding frame. Barcode values associated with different Betti numbers are represented by different colors, eg. purple corresponds to Betti-0=61 and red corresponds to Betti-0=1. The derived data from both the alarmed and unalarmed data set can be seen below.

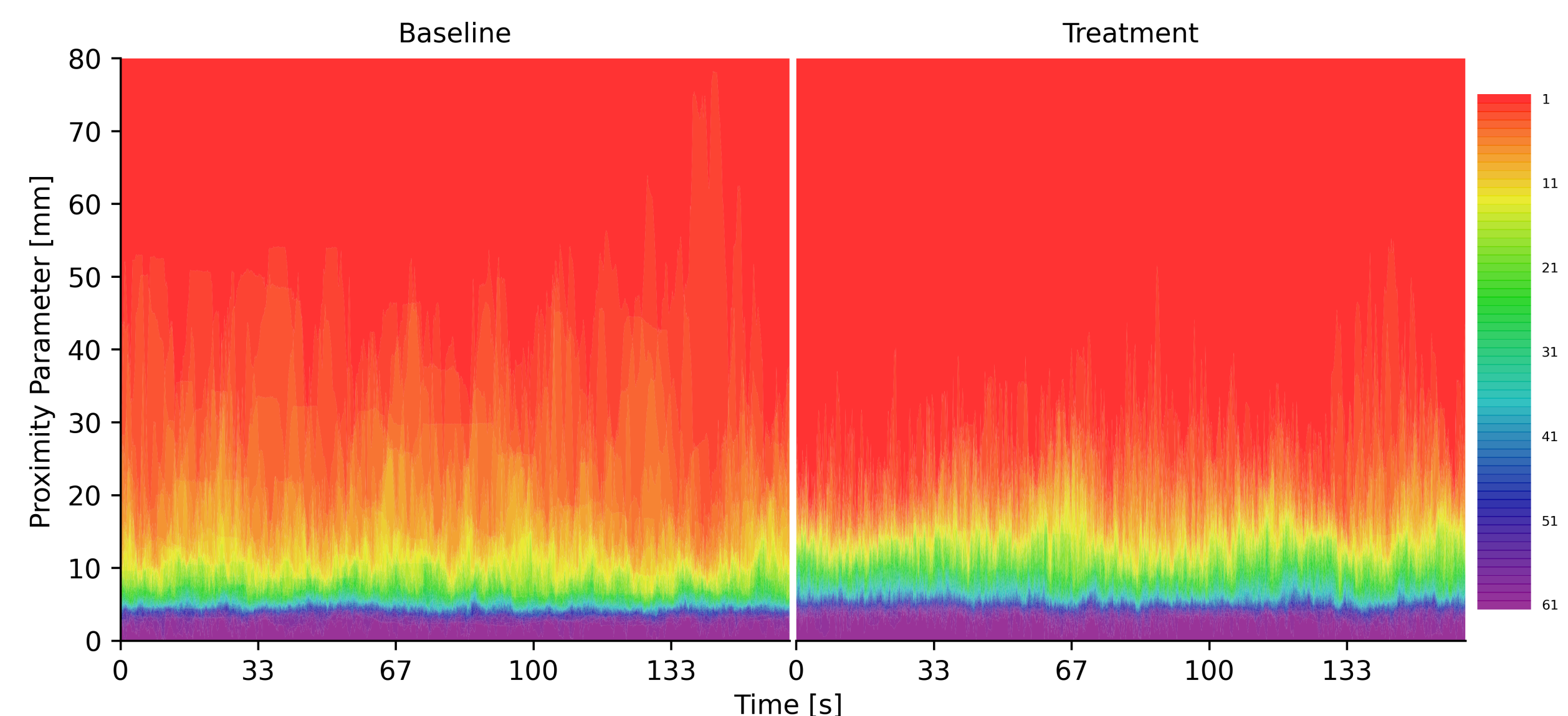


Fig. 2: Experimental Crocker Plots

### Frobenius Norms

Frobenius norm for the unalarmed dataset:

Increment 0.1: 30646.6024, increment 0.01: 96490.6943.

Frobenius norm for the alarmed dataset:

Increment 0.1: 34252.0631, increment 0.01: 107934.8501.

### Code

Our group was able to construct the necessary VR Complexes and analyze this data through coding primarily via Python and R. These scripts were developed by our group, off of foundations in R's TDA and C++'s Ripser packages. Code is available upon request.

### Hypotheses Simulations

After creating this analysis for the experimental data, our group developed several hypotheses for ant movement, and coded simulations to exhibit the movement of each. We began with three simulations: random, center-of-mass, and point-to-point. Each simulation was based fully on the experimental data, aiming to change only the movement of the ants. As such, the speeds of the ants, the size of the arena, and the timeframes measured were all kept similar to the experimental data for each simulation [1]. The point-to-point simulation was excluded from the Crocker plots because it did not produce noteworthy data.

## Comparison to Simulations

From our hypotheses simulations, we then performed the same TDA analysis that was done on the experimental data. Based on this, we compared the topological features between the experimental data and the simulations. Observe the simulation results below.

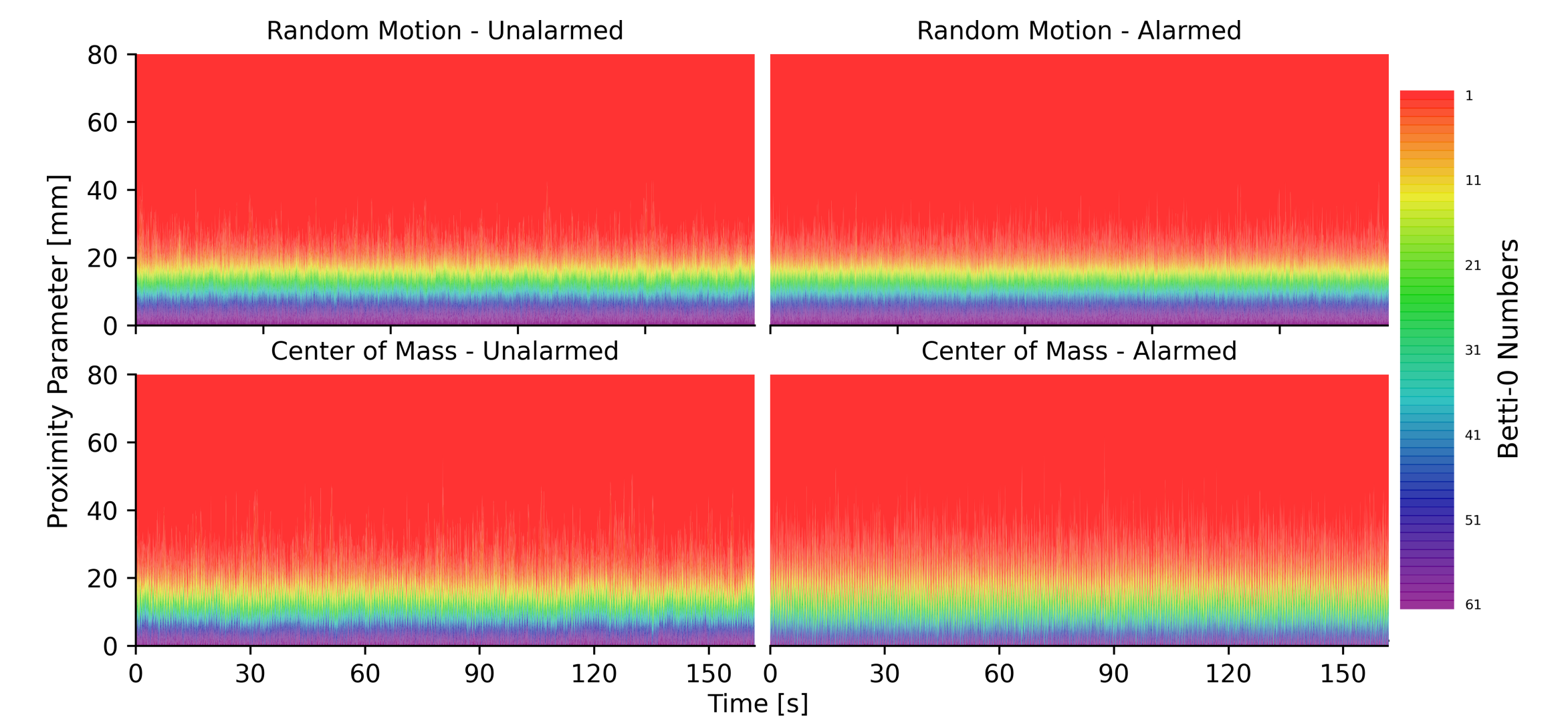


Fig. 3: CROCKER Plots of simulations with different restrictions

### Frobenius Norms

**Frobenius norm for the simulation with random movement:**

Unalarmed: Increment 0.1: 39337.9789, increment 0.01: 124069.4846

Alarmed: Increment 0.1: 39224.9056, increment 0.01: 123711.1316

**Frobenius norm for the simulation with movement towards center of mass:**

Unalarmed: Increment 0.1: 36186.8349, increment 0.01: 114095.3721

Alarmed: Increment 0.1: 33254.4579, increment 0.01: 104831.0618.

### Discussion

As expected, clearly the TDA analysis differentiates between the experimental data and our simulations. The simulations' Frobenius norms are roughly 25% larger in each case. Comparison of the Crocker plots of experimental unalarmed and alarmed state indicates that the ants stay in a few midsize clusters in the unalarmed state and tend towards one larger cluster in alarmed state. The simulated data seems more similar to the alarmed state than the unalarmed state when comparing the topological features illustrated by the Crocker plots. Nonetheless, continuing this process and further refining our simulations will grant us a deeper insight into the ants' movement, and eventually a more complex simulation may generate movement that produces far closer results.

## Acknowledgements

Many thanks to Prof. Laura Murray for guiding us through this process, many steps would have been impossible without her support. Additional thanks go to Prof. James Waters and Prof. Xiaohui Guo for providing us with data for our research.

## References

- [1] Guo Xiaohui & Lin Michael R. & Azizi Asma & Saldyt Lucas P. & Kang Yun & Pavlic Theodore P. & Fewell Jennifer H. "Decoding alarm signal propagation of seed-harvester ants using automated movement tracking and supervised machines learning". In: *Proceedings of the Royal Society* (Jan. 2022), pp. 20-24. DOI: <https://doi.org/10.6084/m9.figshare.c.5772172.v2>.
- [2] Chad M. Topaz & Lori Ziegelmeier & Tom Halverson. "Topological Data Analysis of Biological Aggregation Models". In: *PLOS ONE* 10.5 (May 2015). DOI: <https://doi.org/10.1371/journal.pone.0126383>.
- [3] Herbert Edelsbrunner & John L. Harer. *Computational Topology - An Introduction*. American Mathematical Society, 2010. ISBN: 9780821849255.