USING TOPOLOGY TO MEASURE DYNAMICS OF BIOLOGICAL AGGREGATIONS

August 7, 2019

Lori Ziegelmeier, Macalester College

ICERM Workshop: Applied Mathematical Modeling with Topological Techniques
INTRODUCTION
In many natural systems, particles, organisms, or agents interact locally according to rules that produce aggregate behavior.
Alignment Order Parameter: \( \varphi(t) = \frac{1}{Nv_0} \left| \sum_{i=1}^{N} \vec{v}_i(t) \right| \)
Alignment Order Parameter: \( \phi(t) = \frac{1}{Nv_0} \left| \sum_{i=1}^{N} \vec{v}_i(t) \right| \)
Alignment Order Parameter: $\varphi(t) = \frac{1}{Nv_0} \left| \sum_{i=1}^{N} \vec{v}_i(t) \right|$
1. Envision data as a point cloud
   - e.g. position-velocity for one snapshot in time

2. Create connections between proximate points
   - build simplicial complex

3. Determine topological structure of complex
   - compute homology (measure # holes)

4. Vary proximity parameter to assess different scales
   - calculate persistent homology
COMPUTE PERSISTENT HOMOLOGY
COMPUTE PERSISTENT HOMOLOGY
1. Envision data as a point cloud
   - e.g. position-velocity for one snapshot in time

2. Create connections between proximate points
   - build simplicial complex

3. Determine topological structure of complex
   - compute homology (measure # holes)

4. Vary proximity parameter to assess different scales
   - calculate persistent homology
TOPOLOGICAL DATA ANALYSIS

1. Envision data as a point cloud
   - *e.g.* position-velocity for one snapshot in time

2. Create connections between proximate points
   - build simplicial complex

3. Determine topological structure of complex
   - compute homology (measure # holes)

4. Vary proximity parameter to assess different scales
   - calculate persistent homology

5. **Measure topology as time evolves.**
   - Crocker plots
Compute the $k$th Betti number $b_k(\varepsilon, t)$,
Compute the $k$th Betti number $b_k(\varepsilon, t)$.

**CROCKER plot**

Contour Realization Of Computed $K$-dimensional hole Evolution in the Rips complex (CROCKER)

*(Topaz, Z., Halverson 2015) Topological Data Analysis of Biological Aggregation Models*
CROCKER AS EXPLORATORY TOOL
D’ORSOGNA MODEL

○ Dynamical system describing motion of interacting point particles in an unbounded plane in continuous time.

○ Model written as:

\[
\begin{align*}
\dot{x}_i &= \vec{v}_i & i = 1, \ldots, N \\
\dot{m}\vec{v}_i &= (\alpha - \beta|\vec{v}_i|^2)\vec{v}_i - \nabla_i U_i \\
U_i &= \sum_{j \neq i} C_r e^{-|\vec{x}_i - \vec{x}_j|/\ell_r} - C_a e^{-|\vec{x}_i - \vec{x}_j|/\ell_a}
\end{align*}
\]

○ After nondimensionalization, we have 4 parameters:

\[
\alpha, \beta, C = \frac{C_r}{C_a}, \ell = \frac{\ell_r}{\ell_a}
\]

SNAPSHOTS IN TIME

\[ \alpha = 1.5, \beta = 0.5, C = 2, \ell = 0.25 \text{ with } N = 500 \text{ particles} \]

- Clockwise motion
- Counter Clockwise motion
ORDER PARAMETERS TO SUMMARIZE COLLECTIVE BEHAVIOR

Polarization: \( P(t) = \left| \frac{\sum_{i=1}^{N} v_i(t)}{\sum_{i=1}^{N} |v_i(t)|} \right| \)
Polarization: $P(t) = \left| \frac{\sum_{i=1}^{N} v_i(t)}{\sum_{i=1}^{N} |v_i(t)|} \right|$
ORDER PARAMETERS TO SUMMARIZE COLLECTIVE BEHAVIOR

Polarization: \( P(t) = \left| \frac{\sum_{i=1}^{N} v_i(t)}{\sum_{i=1}^{N} |v_i(t)|} \right| \)

Angular Momentum:
\( M_{\text{ang}}(t) = \left| \frac{\sum_{i=1}^{N} r_i(t) \times v_i(t)}{\sum_{i=1}^{N} |r_i(t)||v_i(t)|} \right| \)
ORDER PARAMETERS TO SUMMARIZE COLLECTIVE BEHAVIOR

Polarization: $P(t) = \left| \frac{\sum_{i=1}^{N} v_i(t)}{\sum_{i=1}^{N} |v_i(t)|} \right|

Angular Momentum: 
$M_{\text{ang}}(t) = \left| \frac{\sum_{i=1}^{N} r_i(t) \times v_i(t)}{\sum_{i=1}^{N} |r_i(t)||v_i(t)|} \right|

Are the particles rotating (in the same direction)?

High

Medium

Low
ORDER PARAMETERS TO SUMMARIZE COLLECTIVE BEHAVIOR

Polarization: $P(t) = \left| \frac{\sum_{i=1}^{N} v_i(t)}{\sum_{i=1}^{N} |v_i(t)|} \right|

Angular Momentum:
$M_{ang}(t) = \left| \frac{\sum_{i=1}^{N} r_i(t) \times v_i(t)}{\sum_{i=1}^{N} |r_i(t)||v_i(t)|} \right|

Absolute Angular Momentum:
$M_{abs}(t) = \left| \frac{\sum_{i=1}^{N} |r_i(t) \times v_i(t)|}{\sum_{i=1}^{N} |r_i(t)||v_i(t)|} \right|
Polarization: $P(t) = \left| \frac{\sum_{i=1}^{N} v_i(t)}{\sum_{i=1}^{N} |v_i(t)|} \right|$ 

Angular Momentum: 
$M_{ang}(t) = \left| \frac{\sum_{i=1}^{N} r_i(t) \times v_i(t)}{\sum_{i=1}^{N} |r_i(t)||v_i(t)|} \right|$ 

Absolute Angular Momentum: 
$M_{abs}(t) = \left| \frac{\sum_{i=1}^{N} |r_i(t) \times v_i(t)|}{\sum_{i=1}^{N} |r_i(t)||v_i(t)|} \right|$ 

Are the particles rotating?

High absolute angular momentum

Low absolute angular momentum
ORDER PARAMETERS TO SUMMARIZE COLLECTIVE BEHAVIOR

Polarization: $P(t) = \left| \frac{\sum_{i=1}^{N} v_i(t)}{\sum_{i=1}^{N} |v_i(t)|} \right|$.

Angular Momentum:
$M_{ang}(t) = \left| \frac{\sum_{i=1}^{N} r_i(t) \times v_i(t)}{\sum_{i=1}^{N} |r_i(t)||v_i(t)|} \right|$.

Absolute Angular Momentum:
$M_{abs}(t) = \left| \frac{\sum_{i=1}^{N} |r_i(t) \times v_i(t)|}{\sum_{i=1}^{N} |r_i(t)||v_i(t)|} \right|$.

Average Nearest Neighbor Distance:
$D_{nn}(t) = \frac{1}{N} \sum_{i=1}^{N} \min_{1 \leq j \leq N} |x_i(t) - x_j(t)|$. 
ORDER PARAMETERS TO SUMMARIZE COLLECTIVE BEHAVIOR

Polarization: \( P(t) = \left| \frac{\sum_{i=1}^{N} v_i(t)}{\sum_{i=1}^{N} |v_i(t)|} \right| \)

Angular Momentum:
\( M_{ang}(t) = \left| \frac{\sum_{i=1}^{N} \mathbf{r}_i(t) \times v_i(t)}{\sum_{i=1}^{N} |\mathbf{r}_i(t)||v_i(t)|} \right| \)

Absolute Angular Momentum:
\( M_{abs}(t) = \left| \frac{\sum_{i=1}^{N} |\mathbf{r}_i(t) \times v_i(t)|}{\sum_{i=1}^{N} |\mathbf{r}_i(t)||v_i(t)|} \right| \)

Average Nearest Neighbor Distance:
\( D_{nn}(t) = \frac{1}{N} \sum_{i=1}^{N} \min_{1 \leq j \leq N} |\mathbf{x}_i(t) - \mathbf{x}_j(t)| \)

How close are the particles?

High Avg. NND

Low Avg. NND
D’ORSOGNA SIMULATION ANALYSIS

(A) Order Parameters

- **Order Parameter**
  - P
  - M
  - $M_{\text{abs}}$

- $b_0 \geq 5$
- $b_0 = 1$

(B) Proximity Parameter

- **Proximity Parameter** $\varepsilon$

- Level:
  - $\varepsilon = 1$
  - $\varepsilon = 2$
  - $\varepsilon = 3$
  - $\varepsilon = 4$
  - $\varepsilon = 5$

- $b_0 \geq 5$
- $b_0 = 1$

(C) Proximity Parameter

- **Proximity Parameter** $\varepsilon$

- Level:
  - $\varepsilon = 1$
  - $\varepsilon = 2$
  - $\varepsilon = 3$
  - $\varepsilon = 4$
  - $\varepsilon = 5$

- $b_1 \geq 5$
- $b_1 = 1$
- $b_1 = 2$
D’ORSOGNA SIMULATION ANALYSIS

(A) Order Parameters
- \(P\)
- \(M\)
- \(M_{\text{abs}}\)
- \(b_0 \geq 5\)
- \(b_0 = 1\)
- \(b_1 = 2\)

(B) Proximity Parameter \(\varepsilon\)
- \(b_0 \geq 5\)
- \(b_0 = 1\)

(C) Proximity Parameter \(\varepsilon\)
- \(b_1 = 0\)
- \(b_1 = 1\)
- \(b_1 \geq 5\)
- \(b_1 = 2\)
PARAMETER IDENTIFICATION
QUESTION OF INTEREST

Forward Problem:

- Parameters (individual interactions)
- Agent-Based Model
- Collective Behavior
QUESTION OF INTEREST

Forward Problem:

Parameters (individual interactions) → Agent-Based Model → Collective Behavior

Inverse Problem:

Collective behavior → Machine Learning (trained on ABM) → Parameters (individual interactions)

Research Question

What is an informative way to summarize collective behavior for machine learning techniques?
Different patterns emerge by altering parameters $C$ and $\ell$
Simulate Swarm Behavior

Design

Order Parameters

Problem dependent

Order Parameter based

Description

Construct

Topology

Problem independent

Topology based

Description

Construct Feature Vectors

Machine Learning

Predict Parameters & Patterns

25 parameter sets x 100 realizations
= 2500 simulations
PHENOTYPES AND FEATURE VECTORS

Single Mill:

Double Ring:

Escape:
Support Vector Machine

- Supervised machine learning algorithm
- 5-fold cross validation, with each training simulation labeled with \((C, \ell)\), 20% of data used for each test
- Accuracy = \(\frac{\text{out-of-sample simulations with } (C, \ell) \text{ correct}}{\text{out-of-sample simulations}}\)
## Classification Results

<table>
<thead>
<tr>
<th>Summary</th>
<th>Feature</th>
<th>Dimension</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order Parameters</td>
<td>$P(t)$</td>
<td>87</td>
<td>57.7%</td>
</tr>
<tr>
<td></td>
<td>$M_{ang}(t)$</td>
<td>87</td>
<td>34.4%</td>
</tr>
<tr>
<td></td>
<td>$M_{abs}(t)$</td>
<td>87</td>
<td>68.0%</td>
</tr>
<tr>
<td></td>
<td>$D_{NN}(t)$</td>
<td>87</td>
<td>91.1%</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>4*87</td>
<td>89.2%</td>
</tr>
<tr>
<td>TDA (time-delayed position)</td>
<td>$b_0$</td>
<td>200*86</td>
<td>99.6%</td>
</tr>
<tr>
<td></td>
<td>$b_1$</td>
<td>200*86</td>
<td>99.3%</td>
</tr>
<tr>
<td></td>
<td>$b_0$ &amp; $b_1$</td>
<td>2<em>200</em>86</td>
<td>99.1%</td>
</tr>
</tbody>
</table>
## CLASSIFICATION RESULTS

<table>
<thead>
<tr>
<th>Summary</th>
<th>Feature</th>
<th>Dimension</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Order</strong></td>
<td>( P(t) )</td>
<td>87</td>
<td>57.7%</td>
</tr>
<tr>
<td>Parameters</td>
<td>( M_{\text{ang}}(t) )</td>
<td>87</td>
<td>34.4%</td>
</tr>
<tr>
<td></td>
<td>( M_{\text{abs}}(t) )</td>
<td>87</td>
<td>68.0%</td>
</tr>
<tr>
<td></td>
<td>( D_{\text{NN}}(t) )</td>
<td>87</td>
<td>91.1%</td>
</tr>
<tr>
<td></td>
<td><em>All</em></td>
<td>4*87</td>
<td>89.2%</td>
</tr>
<tr>
<td><strong>TDA</strong></td>
<td>( b_0 )</td>
<td>200*86</td>
<td>99.6%</td>
</tr>
<tr>
<td>(time-delayed position)</td>
<td>( b_1 )</td>
<td>200*86</td>
<td>99.3%</td>
</tr>
<tr>
<td></td>
<td><em>( b_0 \ &amp; \ b_1 )</em></td>
<td>2<em>200</em>86</td>
<td>99.1%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Summary</th>
<th>Feature</th>
<th>Dimension</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Order</strong></td>
<td>( P(t) ) (PCA)</td>
<td>3</td>
<td>46.7%</td>
</tr>
<tr>
<td>Parameters</td>
<td>( M_{\text{ang}}(t) ) (PCA)</td>
<td>3</td>
<td>30.0%</td>
</tr>
<tr>
<td></td>
<td>( M_{\text{abs}}(t) ) (PCA)</td>
<td>3</td>
<td>58.8%</td>
</tr>
<tr>
<td></td>
<td>( D_{\text{NN}}(t) ) (PCA)</td>
<td>3</td>
<td>81.5%</td>
</tr>
<tr>
<td></td>
<td><em>All (PCA)</em></td>
<td>3</td>
<td>68.6%</td>
</tr>
<tr>
<td><strong>TDA</strong></td>
<td>( b_0 ) (PCA)</td>
<td>87</td>
<td>99.7%</td>
</tr>
<tr>
<td>(time-delayed position)</td>
<td>( b_1 ) (PCA)</td>
<td>87</td>
<td>99.9%</td>
</tr>
<tr>
<td></td>
<td><em>( b_0 \ &amp; \ b_1 )</em> (PCA)</td>
<td>87</td>
<td>99.7%</td>
</tr>
<tr>
<td></td>
<td>( b_0 ) (PCA)</td>
<td>3</td>
<td>89.7%</td>
</tr>
<tr>
<td></td>
<td>( b_1 ) (PCA)</td>
<td>3</td>
<td>82.8%</td>
</tr>
<tr>
<td></td>
<td><em>( b_0 \ &amp; \ b_1 )</em> (PCA)</td>
<td>3</td>
<td>89.6%</td>
</tr>
</tbody>
</table>
(Bhaskar, Manhart, Milzman, Nardini, Storey, Topaz, Z. forthcoming) Analyzing Collective Motion with Machine Learning and Topology
ASSESSING MODEL VALIDITY
Social Aggregation in Pea Aphids: Experiment and Random Walk Modeling

Christa Nilsen¹, John Paige¹, Olivia Warner¹, Benjamin Mayhew¹, Ryan Sutley¹, Matthew Lam², Andrew J. Bernoff², Chad M. Topaz¹*
MOVEMENT OF PEA APHIDS
PEA APHID MODELS

- Classifies each aphid according to two motion states, moving or stationary.
- Transitions between these states are probabilistic, depending on distance $d$ to each aphid’s nearest neighbor.
Classifies each aphid according to two motion states, moving or stationary.

Transitions between these states are probabilistic, depending on distance $d$ to each aphid's nearest neighbor.

Model written as:

1. Probability of transition state

$$P_{MS}(d) = P_{MS}^\infty + (P_{MS}^0 - P_{MS}^\infty) e^{-d/d_{MS}}$$

$$P_{SM}(d) = P_{SM}^0 e^{-d/d_{SM}} + P_{SM}^\infty \frac{d}{d + \Delta_{SM}}$$

2. Step length $\ell$

$$\ell(d) = \ell^\infty + (\ell^0 - \ell^\infty) e^{-d/d_\ell}$$

3. Turning angle $\theta$ drawn from a wrapped Cauchy distribution centered at zero, with parameter $\rho(d)$ controlling the spread of the distribution.
PEA APHID MODELS

- Classifies each aphid according to two motion states, moving or stationary.
- Transitions between these states are probabilistic, depending on distance \(d\) to each aphid’s nearest neighbor.
- Two models: Interactive and Noninteractive (Control)
ASSESSING MODEL VALIDITY

Goal:
Use order parameters and crocker plots to compare each of the interactive and control models to experimental data

9 experiments x 100 realizations
Order Parameter based Description

9 experiments x 100 realizations
Topology based Description

Construct Summaries

Compare Distances
Two Sample t-Test to Assess Model Validity
ORDER PARAMETERS OF EXPERIMENT AND SIMULATIONS

**Polarization**

- **Experiment**
- **Interactive**
- **Control**

**Angular Momentum**

- **Experiment**
- **Interactive**
- **Control**

**Absolute Angular Momentum**

- **Experiment**
- **Interactive**
- **Control**
ORDER PARAMETERS OF EXPERIMENT AND SIMULATIONS

Average Nearest Neighbor Distance

Percent of Aphids Moving
$B_{0}(POS)$ CROCKER PLOTS OF EXPERIMENT AND SIMULATIONS
EXPERIMENT VS MODEL COMPARISON

Histogram of crocker difference between control/experiment and interactive/experiment
Summaries of statistical tests comparing models of aphid motion using order parameters.

<table>
<thead>
<tr>
<th>Exp</th>
<th>$P$</th>
<th>$D$</th>
<th>$R_{95%}$</th>
<th>$M_{ang}$</th>
<th>$D$</th>
<th>$R_{95%}$</th>
<th>$M_{abs}$</th>
<th>$D$</th>
<th>$R_{95%}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.17</td>
<td>0.13</td>
<td>1.50</td>
<td>0.16</td>
<td>-2.00</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.25</td>
<td>0.10</td>
<td>-0.97</td>
<td>0.11</td>
<td>-1.66</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.09</td>
<td>0.10</td>
<td>0.40</td>
<td>0.11</td>
<td>-1.71</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.70</td>
<td>0.08</td>
<td>-0.78</td>
<td>0.09</td>
<td>-1.24</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-2.22</td>
<td>0.09</td>
<td>-0.63</td>
<td>0.09</td>
<td>-1.10</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.04</td>
<td>0.09</td>
<td>-0.49</td>
<td>0.10</td>
<td>-1.68</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-0.97</td>
<td>0.11</td>
<td>-1.28</td>
<td>0.13</td>
<td>-1.99</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-3.30</td>
<td>0.25</td>
<td>-2.11</td>
<td>0.22</td>
<td>-2.87</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>-4.69</td>
<td>0.25</td>
<td>-3.14</td>
<td>0.22</td>
<td>-2.81</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### SUMMARY OF STATISTICAL TESTS COMPARING MODELS

Summaries of statistical tests comparing models of aphid motion using order parameters.

<table>
<thead>
<tr>
<th>Exp</th>
<th>$P$</th>
<th>$D$</th>
<th>$R_{95%}$</th>
<th>$M_{ang}$</th>
<th>$D$</th>
<th>$R_{95%}$</th>
<th>$M_{abs}$</th>
<th>$D$</th>
<th>$R_{95%}$</th>
<th>$d_a$</th>
<th>$D$</th>
<th>$R_{95%}$</th>
<th>$Mov%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.17</td>
<td>0.13</td>
<td>0.16</td>
<td>-2.00</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.26</td>
<td>0.05</td>
<td>20.2</td>
<td>0.40</td>
</tr>
<tr>
<td>2</td>
<td>-0.25</td>
<td>0.10</td>
<td>-0.97</td>
<td>-1.66</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.25</td>
<td>0.02</td>
<td>0.78</td>
<td>0.24</td>
</tr>
<tr>
<td>3</td>
<td>1.09</td>
<td>0.10</td>
<td>0.40</td>
<td>-1.71</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.54</td>
<td>0.04</td>
<td>25.1</td>
<td>0.26</td>
</tr>
<tr>
<td>4</td>
<td>1.70</td>
<td>0.08</td>
<td>-0.78</td>
<td>-1.24</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.16</td>
<td>0.02</td>
<td>1.68</td>
<td>0.21</td>
</tr>
<tr>
<td>5</td>
<td>-2.22</td>
<td>0.09</td>
<td>-0.63</td>
<td>-1.10</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.31</td>
<td>0.03</td>
<td>7.68</td>
<td>0.24</td>
</tr>
<tr>
<td>6</td>
<td>1.04</td>
<td>0.09</td>
<td>-0.49</td>
<td>-1.68</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.53</td>
<td>0.04</td>
<td>24.0</td>
<td>0.26</td>
</tr>
<tr>
<td>7</td>
<td>-0.97</td>
<td>0.11</td>
<td>-1.28</td>
<td>-1.99</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.57</td>
<td>0.04</td>
<td>23.9</td>
<td>0.29</td>
</tr>
<tr>
<td>8</td>
<td>-3.30</td>
<td>0.25</td>
<td>-2.11</td>
<td>-2.87</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.57</td>
<td>0.25</td>
<td>13.2</td>
<td>0.63</td>
</tr>
<tr>
<td>9</td>
<td>-4.69</td>
<td>0.25</td>
<td>-3.14</td>
<td>-2.81</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.48</td>
<td>0.15</td>
<td>11.7</td>
<td>0.34</td>
</tr>
</tbody>
</table>
### SUMMARY OF STATISTICAL TESTS COMPARING MODELS

#### Summaries of statistical tests comparing models of aphid motion using order parameters.

<table>
<thead>
<tr>
<th>Exp</th>
<th>$P$</th>
<th>$D$</th>
<th>$R_{95%}$</th>
<th>$M_{ang}$</th>
<th>$D$</th>
<th>$R_{95%}$</th>
<th>$M_{abs}$</th>
<th>$D$</th>
<th>$R_{95%}$</th>
<th>$d_{a}$</th>
<th>$D$</th>
<th>$R_{95%}$</th>
<th>$M_{mov%}$</th>
<th>$D$</th>
<th>$R_{95%}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.17</td>
<td>0.13</td>
<td>1.50</td>
<td>0.16</td>
<td>-2.00</td>
<td>0.15</td>
<td>0.26</td>
<td>0.05</td>
<td>20.2</td>
<td>0.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.25</td>
<td>0.10</td>
<td>-0.97</td>
<td>0.11</td>
<td>-1.66</td>
<td>0.09</td>
<td>0.25</td>
<td>0.02</td>
<td>0.78</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.09</td>
<td>0.10</td>
<td>0.40</td>
<td>0.11</td>
<td>-1.71</td>
<td>0.10</td>
<td>0.54</td>
<td>0.04</td>
<td>25.1</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.70</td>
<td>0.08</td>
<td>-0.78</td>
<td>0.09</td>
<td>-1.24</td>
<td>0.07</td>
<td>0.16</td>
<td>0.02</td>
<td>1.68</td>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-2.22</td>
<td>0.09</td>
<td>-0.63</td>
<td>0.09</td>
<td>-1.10</td>
<td>0.07</td>
<td>0.31</td>
<td>0.03</td>
<td>7.68</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.04</td>
<td>0.09</td>
<td>-0.49</td>
<td>0.10</td>
<td>-1.68</td>
<td>0.09</td>
<td>0.53</td>
<td>0.04</td>
<td>24.0</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-0.97</td>
<td>0.11</td>
<td>-1.28</td>
<td>0.13</td>
<td>-1.99</td>
<td>0.11</td>
<td>0.57</td>
<td>0.04</td>
<td>23.9</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-3.30</td>
<td>0.25</td>
<td>-2.11</td>
<td>0.22</td>
<td>-2.87</td>
<td>0.26</td>
<td>0.57</td>
<td>0.25</td>
<td>13.2</td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>-4.69</td>
<td>0.25</td>
<td>-3.14</td>
<td>0.22</td>
<td>-2.81</td>
<td>0.20</td>
<td>0.48</td>
<td>0.15</td>
<td>11.7</td>
<td>0.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Summaries of statistical tests comparing models of aphid motion using topology.

| Exp | $b_{0\,pos}(\hat{q})$ | $b_{0\,pos}(\hat{q})$ | $b_{0\,pos}(\hat{q})$ | $b_{0\,pos}(\hat{q})$ | $b_{0\,pos}(\hat{q})$ | $b_{0\,pos}(\hat{q})$ | $b_{0\,pos}(\hat{q})$ | $b_{0\,pos}(\hat{q})$ | $b_{0\,pos}(\hat{q})$ | $b_{0\,pos}(\hat{q})$ | $b_{0\,pos}(\hat{q})$ | $b_{0\,pos}(\hat{q})$ | $b_{0\,pos}(\hat{q})$ | $b_{0\,pos}(\hat{q})$ |
|-----|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1   | 1147                  | 38.73                 | 135.1                 | 5.149                 | 831.9                 | 42.81                 | 124.2                 | 5.640                 |                      |                      |                      |                      |                      |                      |                      |
| 2   | 1703                  | 23.98                 | 297.0                 | 6.500                 | 858.9                 | 16.88                 | 261.1                 | 7.258                 |                      |                      |                      |                      |                      |                      |                      |
| 3   | 1893                  | 37.36                 | 290.4                 | 6.836                 | 1347                  | 35.60                 | 271.3                 | 7.218                 |                      |                      |                      |                      |                      |                      |                      |
| 4   | 1825                  | 18.36                 | 359.0                 | 5.741                 | 552.3                 | 19.04                 | 332.5                 | 6.301                 |                      |                      |                      |                      |                      |                      |                      |
| 5   | 1370                  | 15.81                 | 249.2                 | 6.039                 | 270.8                 | 15.25                 | 225.1                 | 6.401                 |                      |                      |                      |                      |                      |                      |                      |
| 6   | 2086                  | 40.53                 | 322.9                 | 7.061                 | 1484                  | 37.35                 | 307.7                 | 7.116                 |                      |                      |                      |                      |                      |                      |                      |
| 7   | 1747                  | 34.32                 | 264.8                 | 6.235                 | 1289                  | 33.69                 | 247.5                 | 6.69                  |                      |                      |                      |                      |                      |                      |                      |
| 8   | 298.5                 | 20.81                 | 10.28                 | 1.920                 | 147.9                 | 19.81                 | 3.573                 | 1.641                 |                      |                      |                      |                      |                      |                      |                      |
| 9   | 467.5                 | 19.66                 | 22.34                 | 2.148                 | 245.3                 | 21.10                 | 10.93                 | 2.067                 |                      |                      |                      |                      |                      |                      |                      |

(Ulmer, Z., Topaz 2018) Assessing Biological Models Using Topological Data Analysis
PERSISTENT CROCKER PLOTS
BENEFITS AND DRAWBACKS OF CROCKERS

Benefits:

- Can be used to summarize time-varying metric spaces
- Displays topological information at all times simultaneously
- Can be vectorized and applied to statistical and machine learning tasks

Drawbacks:

- Crockers are not stable. Perturbing the dynamic metric space only slightly could produce changes of unbounded size.
BENEFITS AND DRAWBACKS OF CROCKERS

Benefits:

- Can be used to summarize time-varying metric spaces
- Displays topological information at all times simultaneously
- Can be vectorized and applied to statistical and machine learning tasks

Drawbacks:

- Crockers are not stable. Perturbing the dynamic metric space only slightly could produce changes of unbounded size.
Stability

Ideal if tools for data analysis are stable with respect to small perturbations of the inputs.
Stability

Ideal if tools for data analysis are stable with respect to small perturbations of the inputs.
The bottleneck distance between two PDs $B$ and $B'$ is given by

$$d_\infty(B, B') = \inf_{\gamma: B \to B'} \sup_{u \in B} \|u - \gamma(u)\|_\infty,$$

ranging over all bijections between $B$ and $B'$. 
DISTANCE BETWEEN METRIC SPACES

**Definition**

If $X$ and $Y$ are two subsets of a metric space $Z$, then the *Hausdorff distance* between $X$ and $Y$ is

$$d^Z_{H}(X, Y) = \max\{\sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y)\}.$$  

The *Gromov–Hausdorff* distance between metric spaces $X$ and $Y$ is

$$d_{GH}(X, Y) = \inf_{Z, f, g} d^Z_{H}(f(X), g(Y)),$$

where the infimum is taken over all metric spaces $Z$ and isometric embeddings $f: X \to Z$ and $g: Y \to Z$. 
Stability of PDs

(Chazal, de Silva, Oudot 2013) PDs are stable (Lipschitz) with respect to the bottleneck metric

\[ d_\infty(PD(VR(X)), PD(VR(Y))) \leq 2d_{GH}(X, Y) \]

for metric spaces \( X \) and \( Y \).
RANK INVARIANT

**Definition**

For a persistence module $V$ and $\varepsilon < \varepsilon'$, the *rank* of the map $V(\varepsilon) \to V(\varepsilon')$, is the number of intervals in the persistence barcode that contain the interval $[\varepsilon, \varepsilon']$.

The collection of all natural numbers $\text{rank}(V(\varepsilon) \to V(\varepsilon'))$ for all $\varepsilon < \varepsilon'$ is called the *rank invariant*.

![Persistence Barcode](image)

Figure: $\text{rank}(V(4) \to V(8)) = 2$

- The rank invariant is equivalent to a persistence barcode i.e. can obtain one from the other (Carlsson and Zomorodian, 2009).
Let $X$ be a time-varying metric space, with $X_t$ the space at time $t$.

An \( \alpha \)-smoothed crocker plot, for \( \alpha \geq 0 \), at time $t$ and scale $\varepsilon$ is equal to $\text{rank}(H_k(VR(X_t; \varepsilon - \alpha)) \to H_k(VR(X_t; \varepsilon + \alpha)))$.

A standard crocker plot is a 0-smoothed crocker plot.
A crocker video is a sequence of $\alpha$-smoothed crocker plots as $\alpha$ increases continuously from 0. It is an integer-valued function

$$f : \mathbb{R}^3 \rightarrow \mathbb{N}$$

$$f(t, \epsilon, \alpha) = \text{rank}(H_k(\text{VR}(X_t; \epsilon - \alpha))) \rightarrow H_k(\text{VR}(X_t; \epsilon + \alpha))).$$
A crocker video is a non-increasing function,

\[ f(t, \varepsilon, \alpha) \leq f(t, \varepsilon, \alpha') \text{ for } \alpha \geq \alpha'. \]
A crocker video is a non-increasing function,

\[ f(t, \varepsilon, \alpha) \leq f(t, \varepsilon, \alpha') \quad \text{for} \quad \alpha \geq \alpha'. \]
A crocker video is a non-increasing function,

\[ f(t, \varepsilon, \alpha) \leq f(t, \varepsilon, \alpha') \text{ for } \alpha \geq \alpha'. \]
A crocker video is a non-increasing function,

\[ f(t, \varepsilon, \alpha) \leq f(t, \varepsilon, \alpha') \text{ for } \alpha \geq \alpha'. \]
A crocker video is a non-increasing function,

\[ f(t, \varepsilon, \alpha) \leq f(t, \varepsilon, \alpha') \text{ for } \alpha \geq \alpha'. \]
A crocker video is a non-increasing function,

\[ f(t, \varepsilon, \alpha) \leq f(t, \varepsilon, \alpha') \text{ for } \alpha \geq \alpha'. \]
The crocker video is equivalent to the vineyard (Cohen-Steiner, Edelsbrunner, Morozov 2006) i.e. can obtain one from the other.

The crocker video inherits some nice interpretability properties of the crocker plot.

All times are represented in each frame $\alpha$, whereas in a vineyard one only sees information about a single time.

Crocker videos satisfy a stability property similar to the stability of vineyards.
CROCKER VIDEOS ARE STABLE

Theorem (Stability theorem for crocker videos)

If $\mathbf{X}$ and $\mathbf{Y}$ are totally-bounded time-varying metric spaces, and if $d_{\infty}^{GH}(\mathbf{X}, \mathbf{Y}) \leq \delta / 2$, then the crocker videos for $\mathbf{X}$ and $\mathbf{Y}$ are close in the sense that for all $t$, $\varepsilon$, and $\alpha$, we have

- $f_{\mathbf{X}}(t, \varepsilon, \alpha + \delta) \leq f_{\mathbf{Y}}(t, \varepsilon, \alpha)$, and
- $f_{\mathbf{Y}}(t, \varepsilon, \alpha + \delta) \leq f_{\mathbf{X}}(t, \varepsilon, \alpha)$.

Here, by an abuse of notation, we let $f_{\mathbf{X}}$ and $f_{\mathbf{Y}}$ denote $f_{\text{PH}(\text{VR}(\mathbf{X}))}$ and $f_{\text{PH}(\text{VR}(\mathbf{Y}))}$.

(Adams, Topaz, Xian, Z. forthcoming) Capturing Dynamics of Time-Varying Systems with Topology
CONCLUSION
Crocker plots (0-persistence crockers) are useful summaries of
topological properties of time-varying systems.
- Can be combined with machine learning and statistical tools
- Provide means of identifying global behaviors, parameter
  identification, and model selection

Crocker videos are stable.
- If two time-varying metric spaces are “close”, then their
  corresponding crocker videos will also be “close”
QUESTIONS?

Lori Ziegelmeier
lziegel1@macalester.edu
Department of Mathematics, Statistics, and Computer Science

NSF:CDS&E-MSS-1854703
Join WinCompTop: email WinCompTop+subscribe@googlegroups.com