Topological Data Analysis of Biological Aggregation Dynamics

Chad Topaz (Williams College)

NSF DMS-1412674/1743963

This talk, in a nutshell

Part I: Applying a topological lens to biological aggregation model data

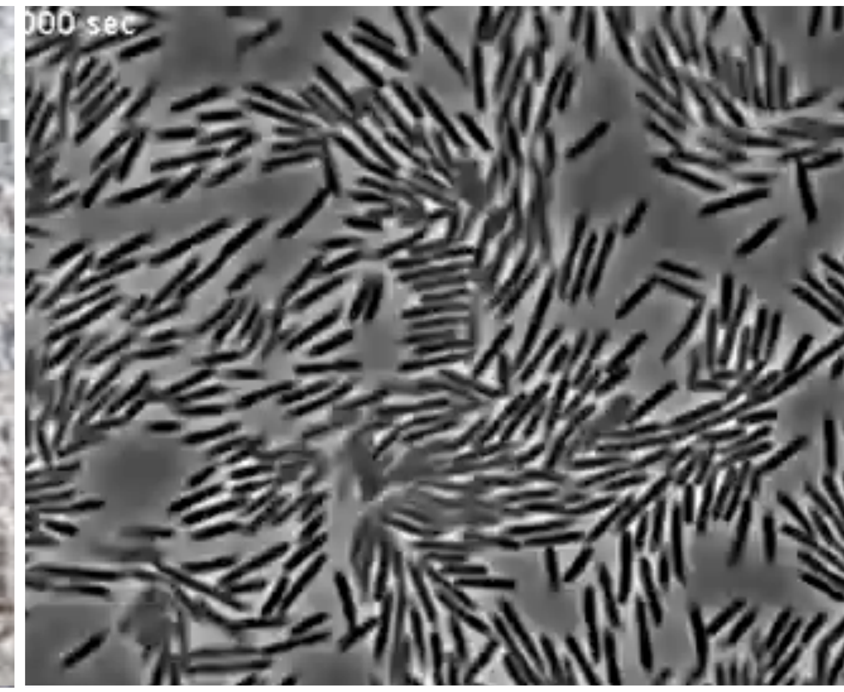
Message: TDA can be a useful tool for exploratory data analysis.

Part II: Moving towards topological reductions of a complex system

Message: When dynamics are neither highly ordered nor totally random, a topological description might be appropriate, but the approach is analytically challenging.

Biological aggregations abound in nature.



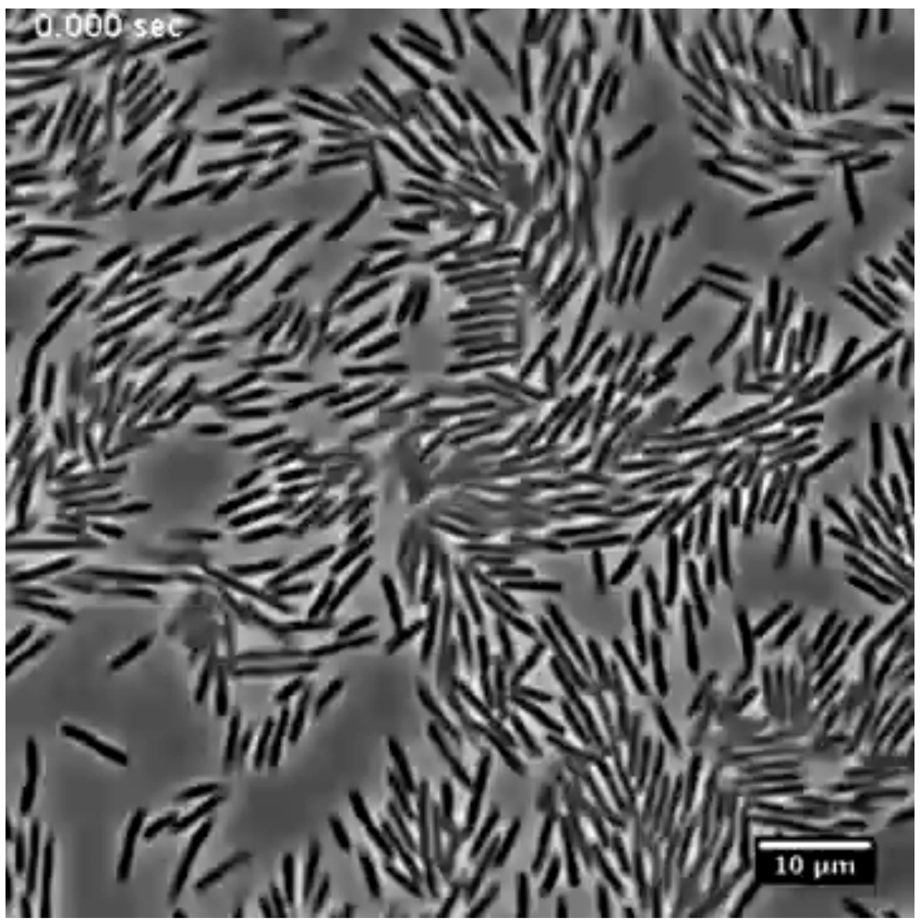



Chad's parsing of biological aggregation research:

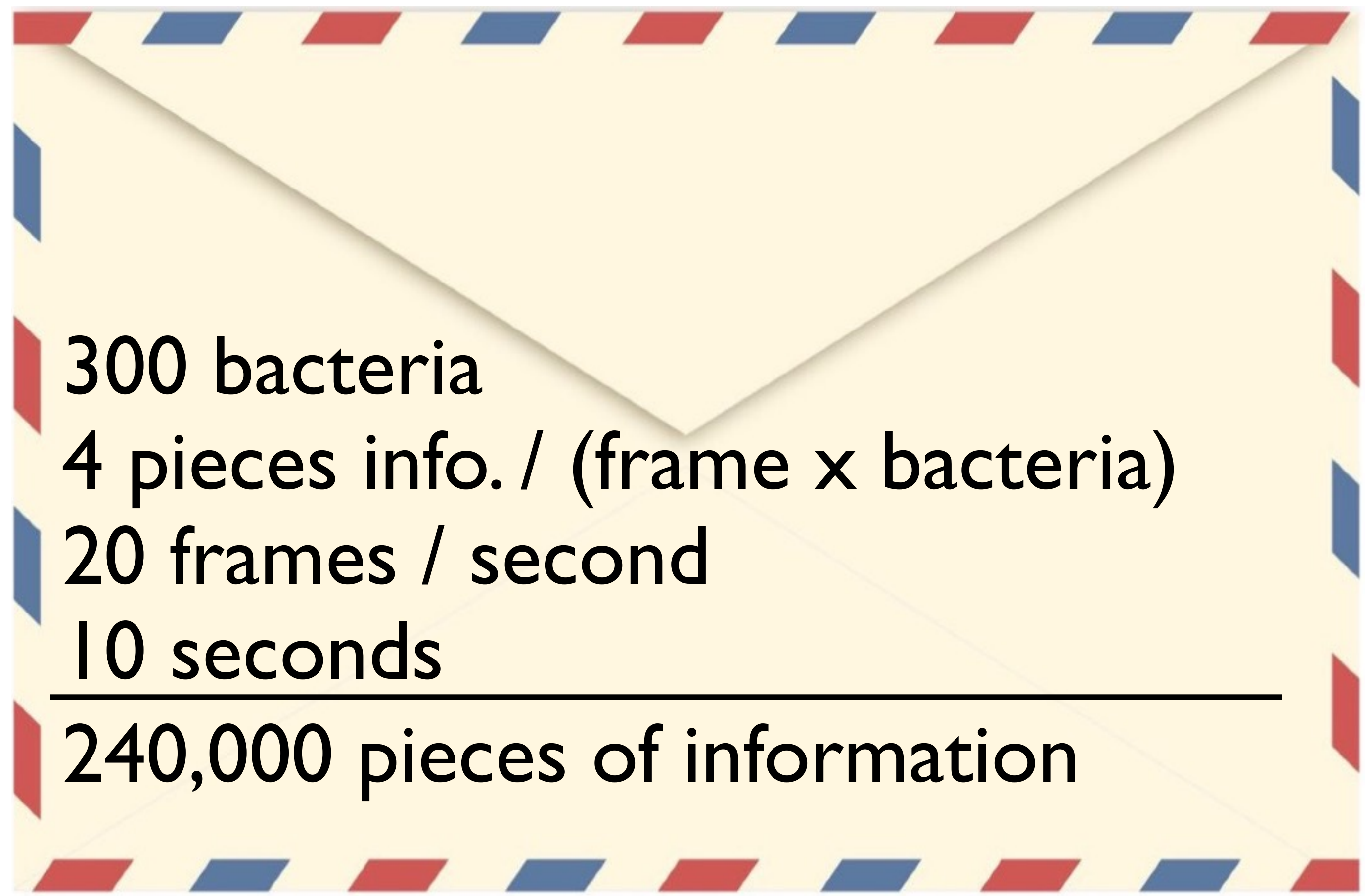
1. Determine individual-level behaviors
2. Assess macroscopic group properties
3. Elucidate the connection between these

Quantifying group dynamics is a task suited for data science.

https://youtu.be/q27Jn3h4kpE



M. Copeland, University of Wisconsin



Vicsek's seminal model describes aligning particles.

Novel type of phase transition in a system of self-driven particles

T Vicsek, A Czirók, E Ben-Jacob, I Cohen, O Shochet - Physical review letters, 1995 - APS

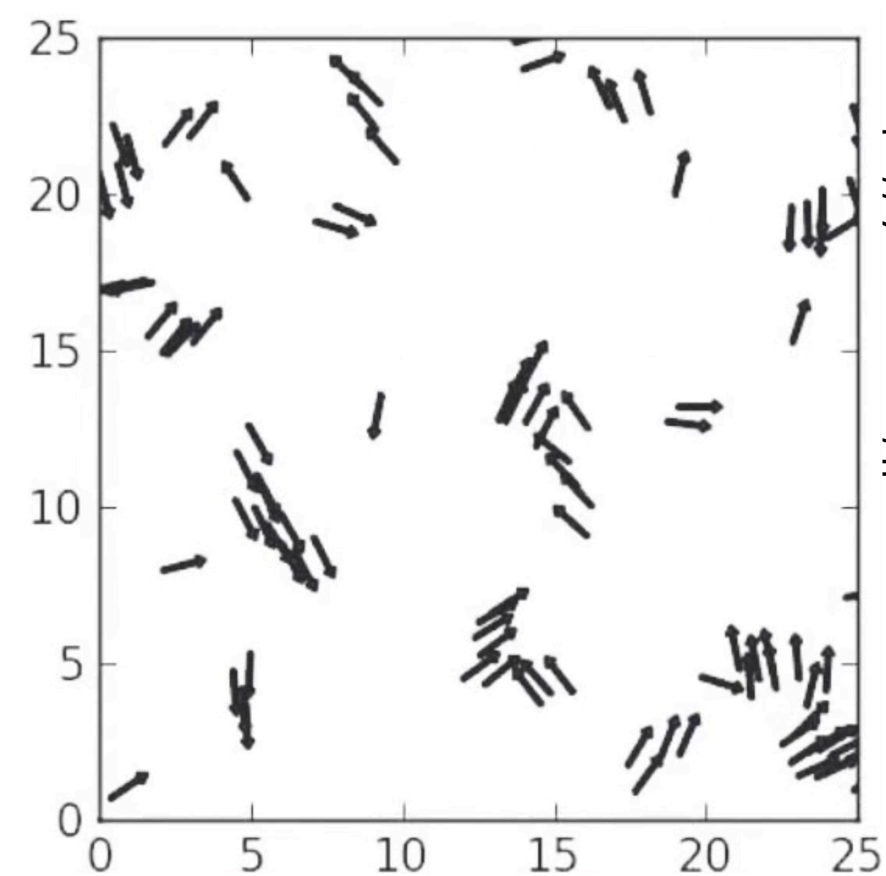
☆ ❞ Cited by 4884 Related articles All 28 versions

social alignment | uniform noise

$$\theta_i \leftarrow \langle \theta_j \rangle_{|\mathbf{x}_i - \mathbf{x}_j| \le R} + U(-\eta/2, \eta/2)$$

$$\mathbf{v}_i \leftarrow v_0(\cos\theta_i, \sin\theta_i)$$

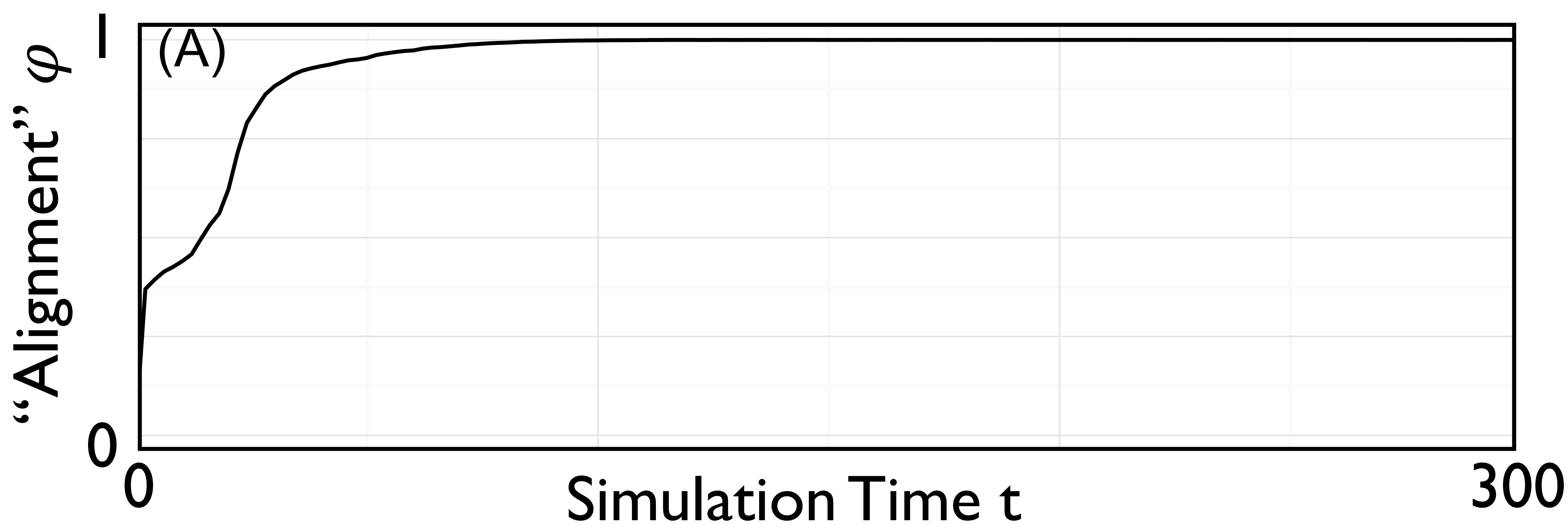
$$\mathbf{x}_i \leftarrow \mathbf{x}_i + \mathbf{v}_i \Delta t$$



http://youtu.be/iphRZV3oCaI

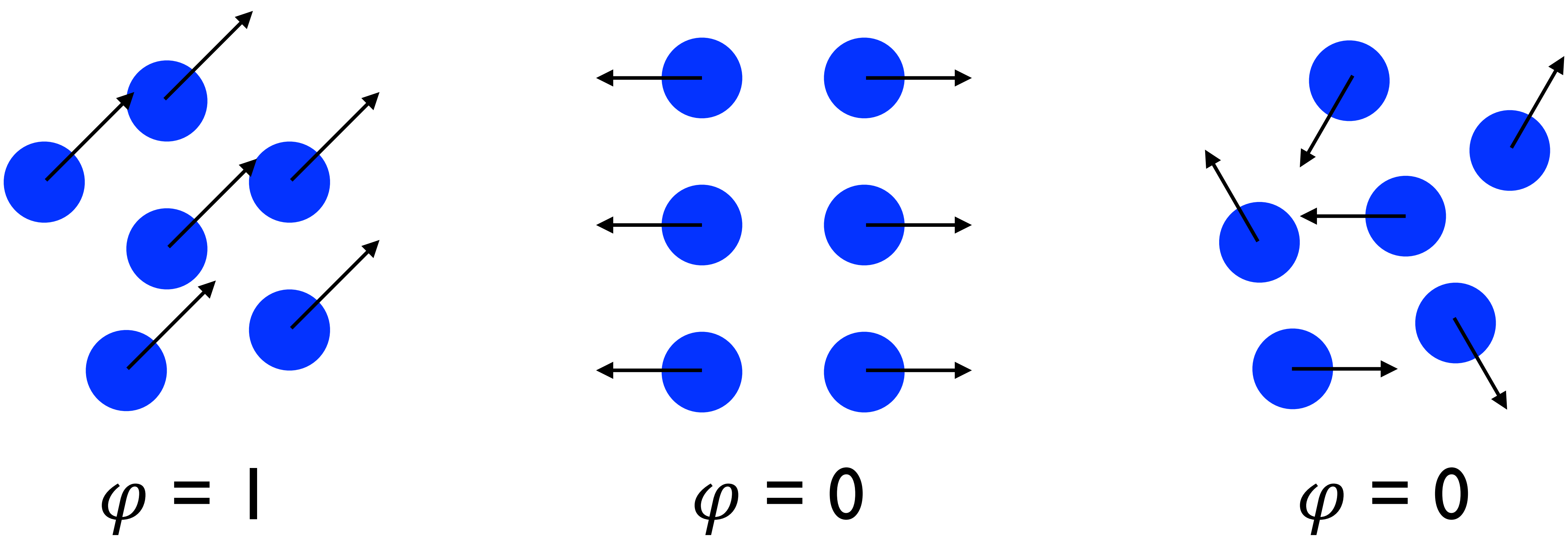
Dynamics are often assessed via order parameter time series.

Alignment order parameter: $\phi(t) = \frac{1}{Nv_0} \left| \sum_{i=1}^{N} \mathbf{v}_i(t) \right|$



are often assessed via
rameter time series.

ler parameter: $\phi(t) = \frac{1}{Nv_0}\left|\sum_{i=1}^{N} \mathbf{v}_i(t)\right|$



How about using topology as our "order parameter"?

1. Computational Homology
   T. Kaczynski, K. Mischaikow, and M. Mrozek. (2004)

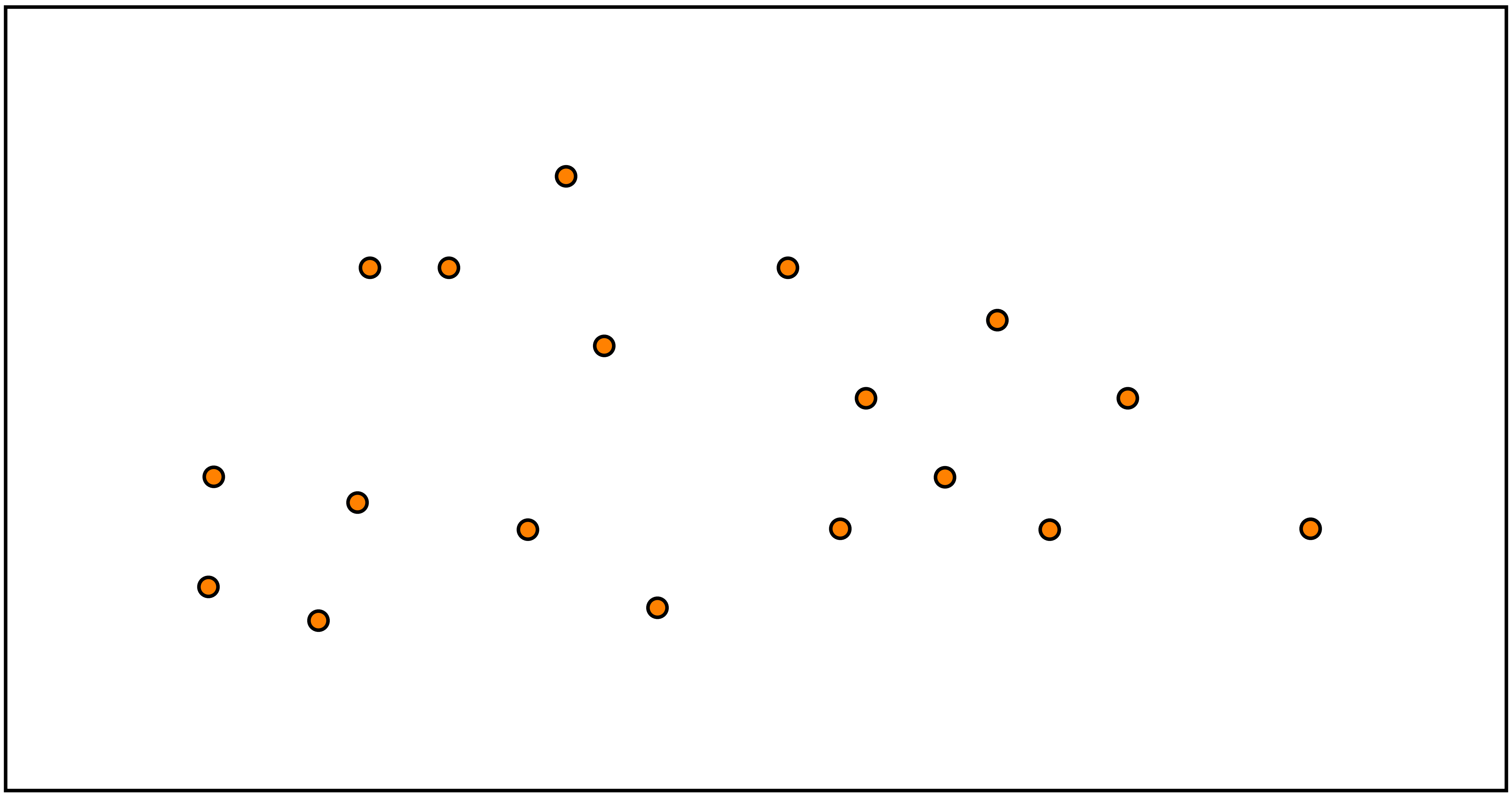
2. Computing persistent homology
   A. Zomorodian, G. Carlsson. Disc. & Comp. Geom. (2005)

3. Barcodes: The persistent topology of data
   R. Ghrist. Bull. Am. Math. Soc. (2008)

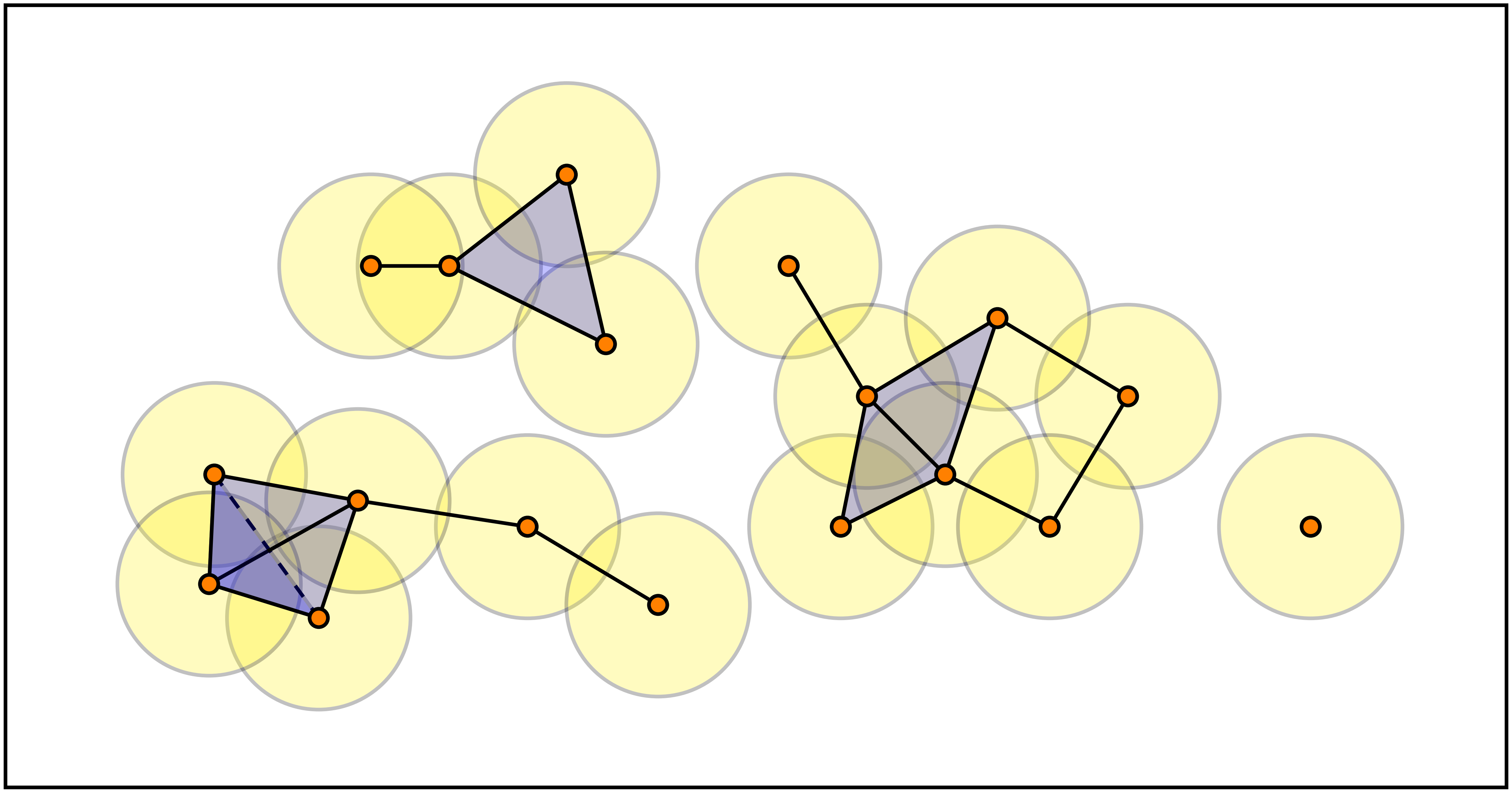
4. Persistent homology: A Survey
   H. Edelsbrunner, J. Harer. Contemp. Math. (2008)

5. Topology and Data
   G. Carlsson. Bull. Am. Math. Soc. (2009)

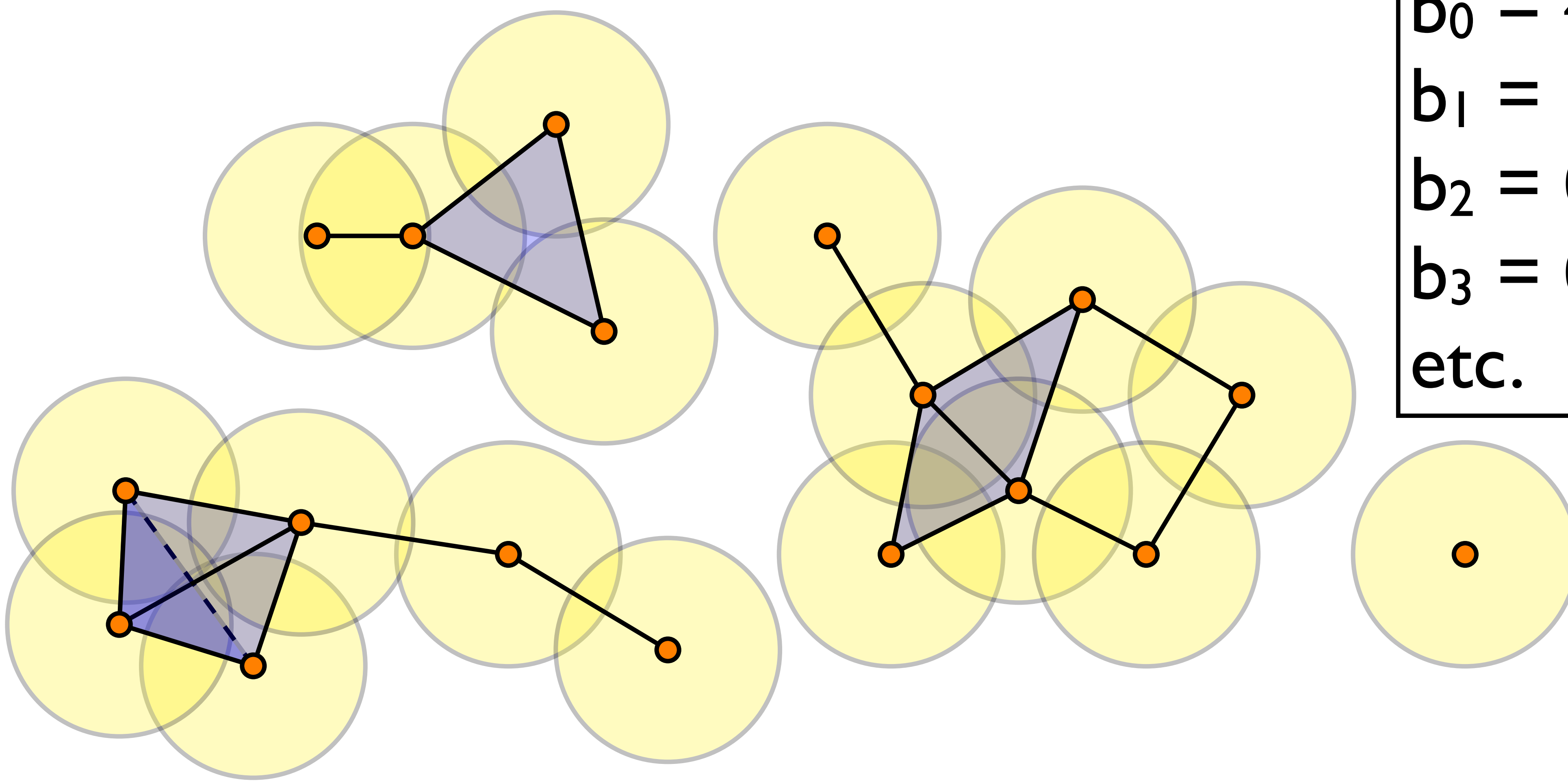
Step 1: Envision data as point cloud



Step 2:
Build simplicial complex



Step 3:
Calculate Betti numbers



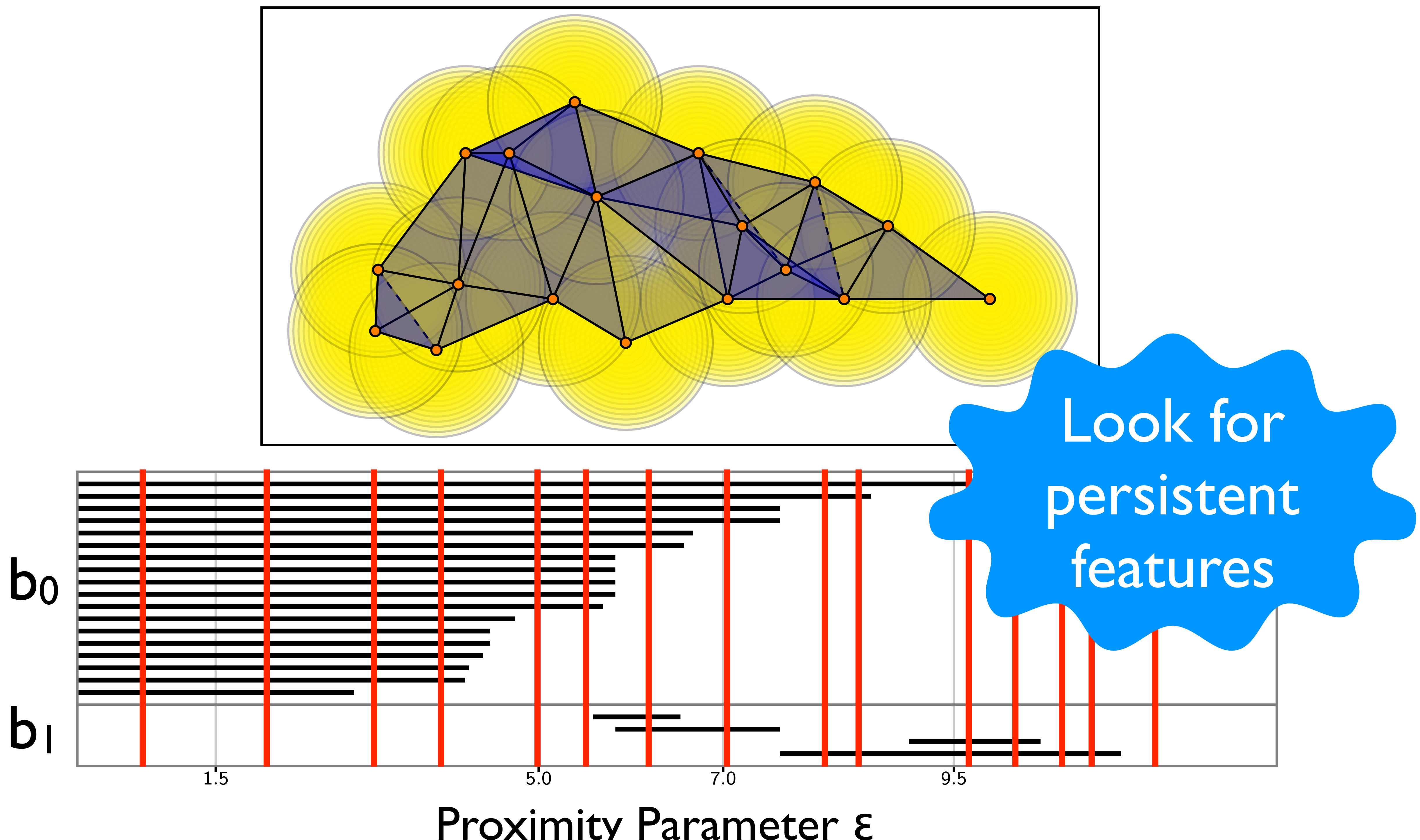
$b_0 = 4$
$b_1 = 1$
$b_2 = 0$
$b_3 = 0$
etc.

Chad's Self-Help Homology Tutorial For The Simple(x)-Minded

A full-color Extravaganza

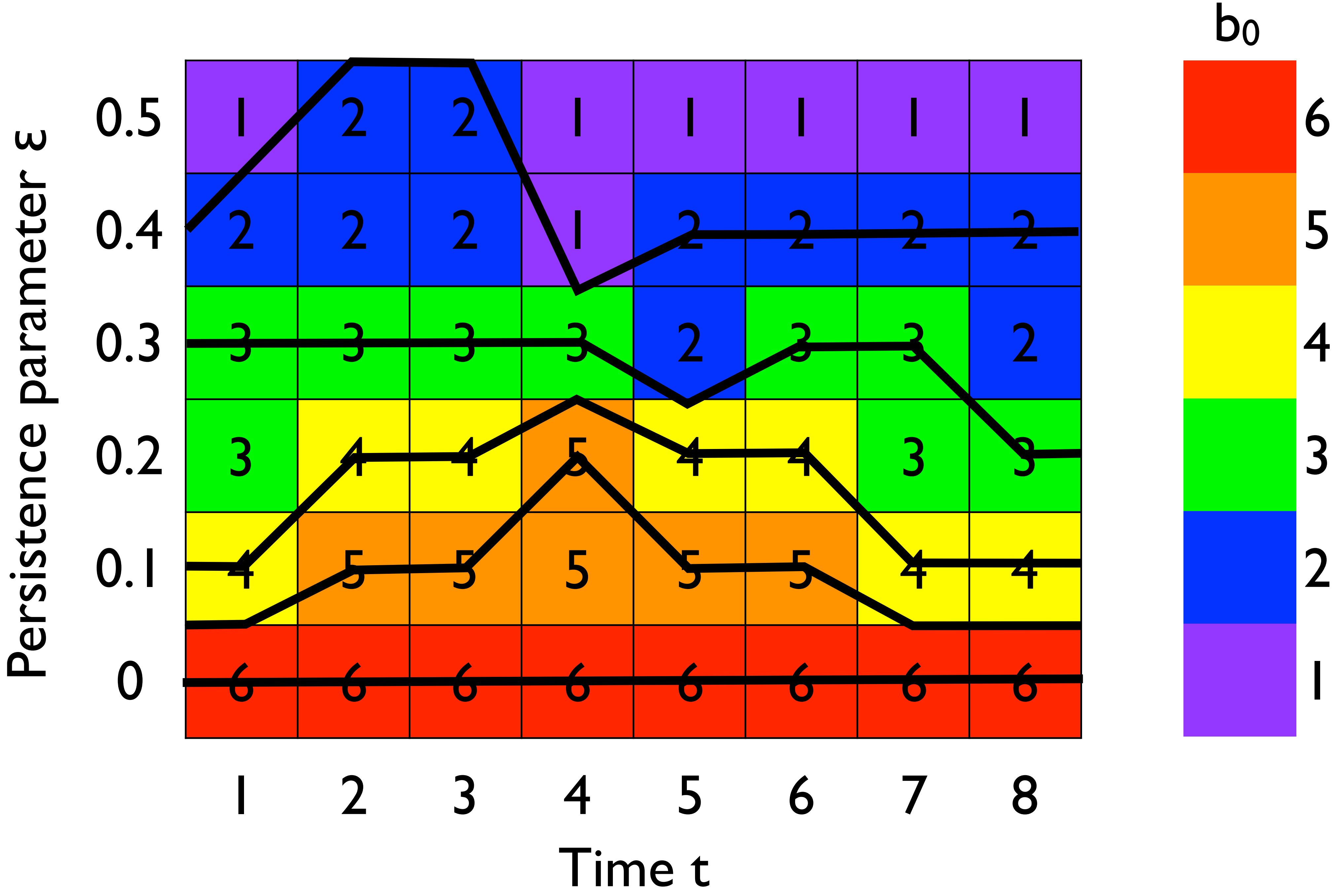
With very sincere thanks and apologies to Lori Ziegelmeier and Tom Halverson, who actually know topology and tried to explain it to me.

Step 4:
Find persistent homology



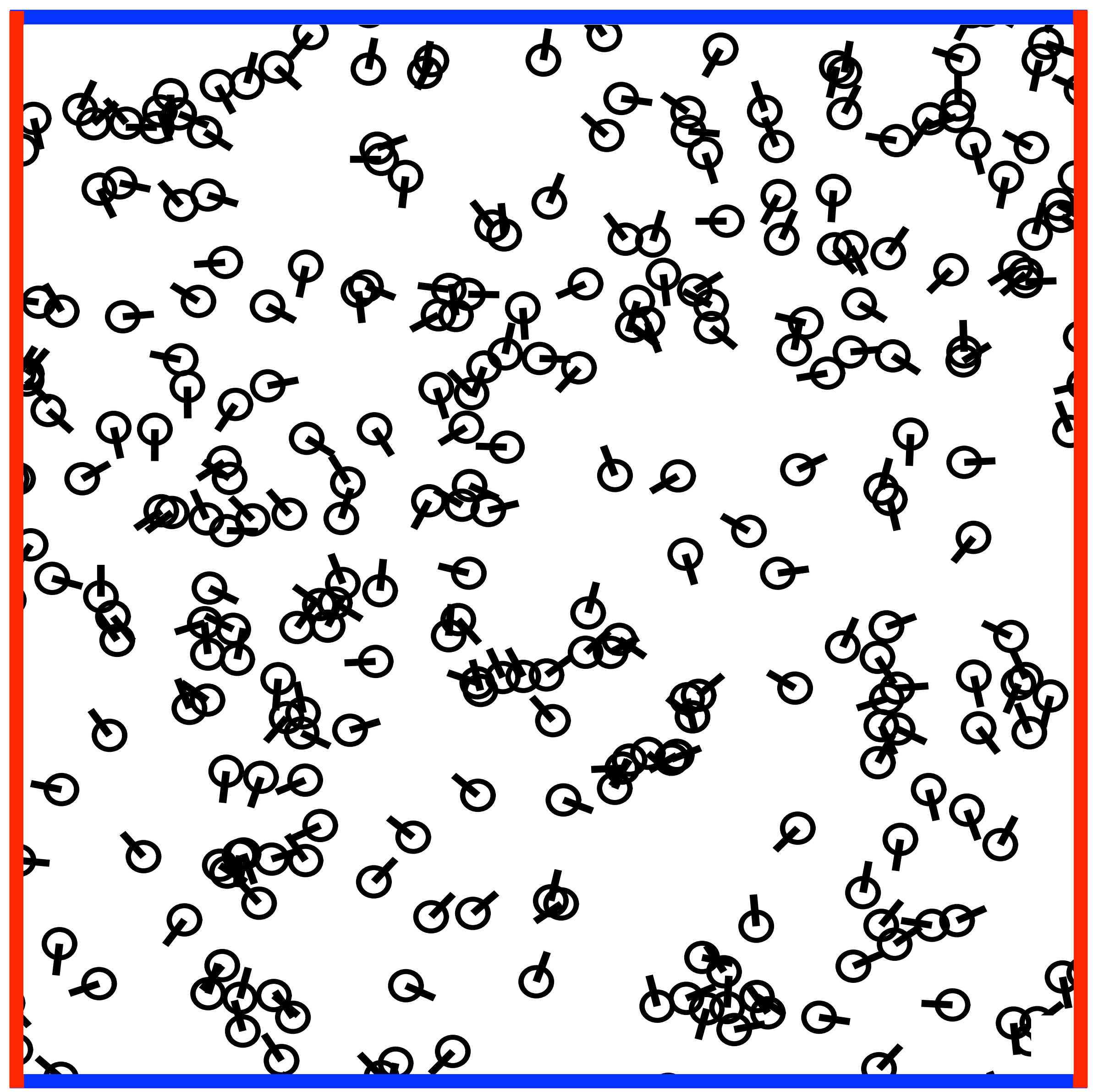
Step 4:
Find persistent homology

Step 5:
Evolve in time



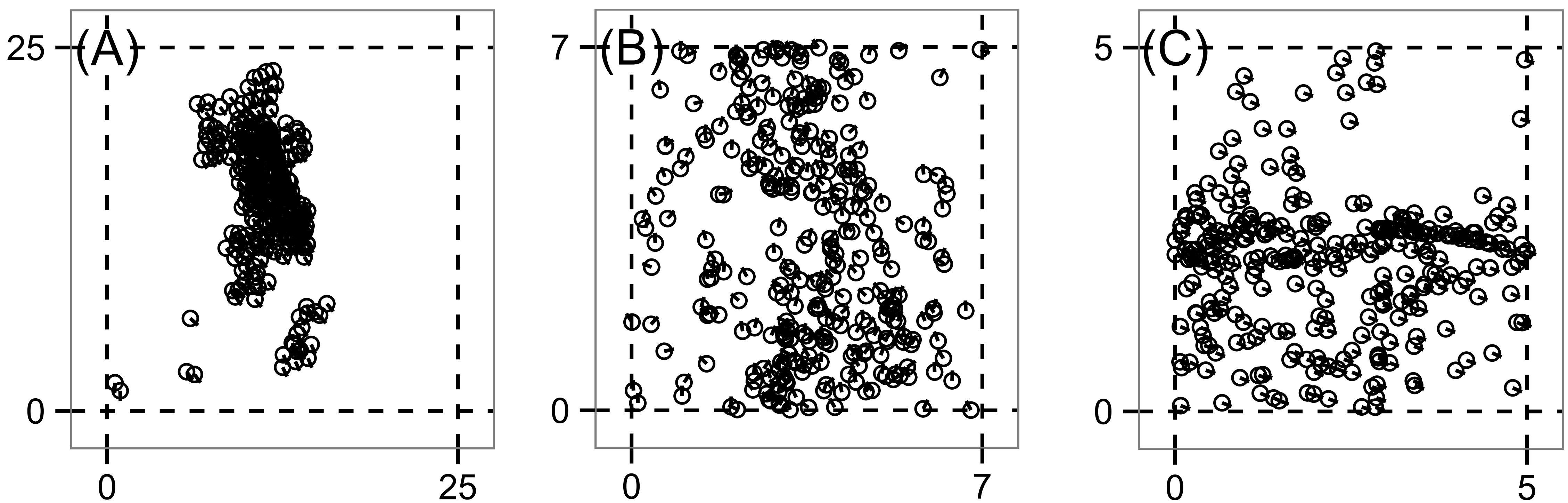
Step 5:
Evolve in time (CROCKER)

Initial condition for Vicsek model covers a three-torus.



b = (1,3,3,1,0,…)

The Vicsek model has several prototypical behaviors.



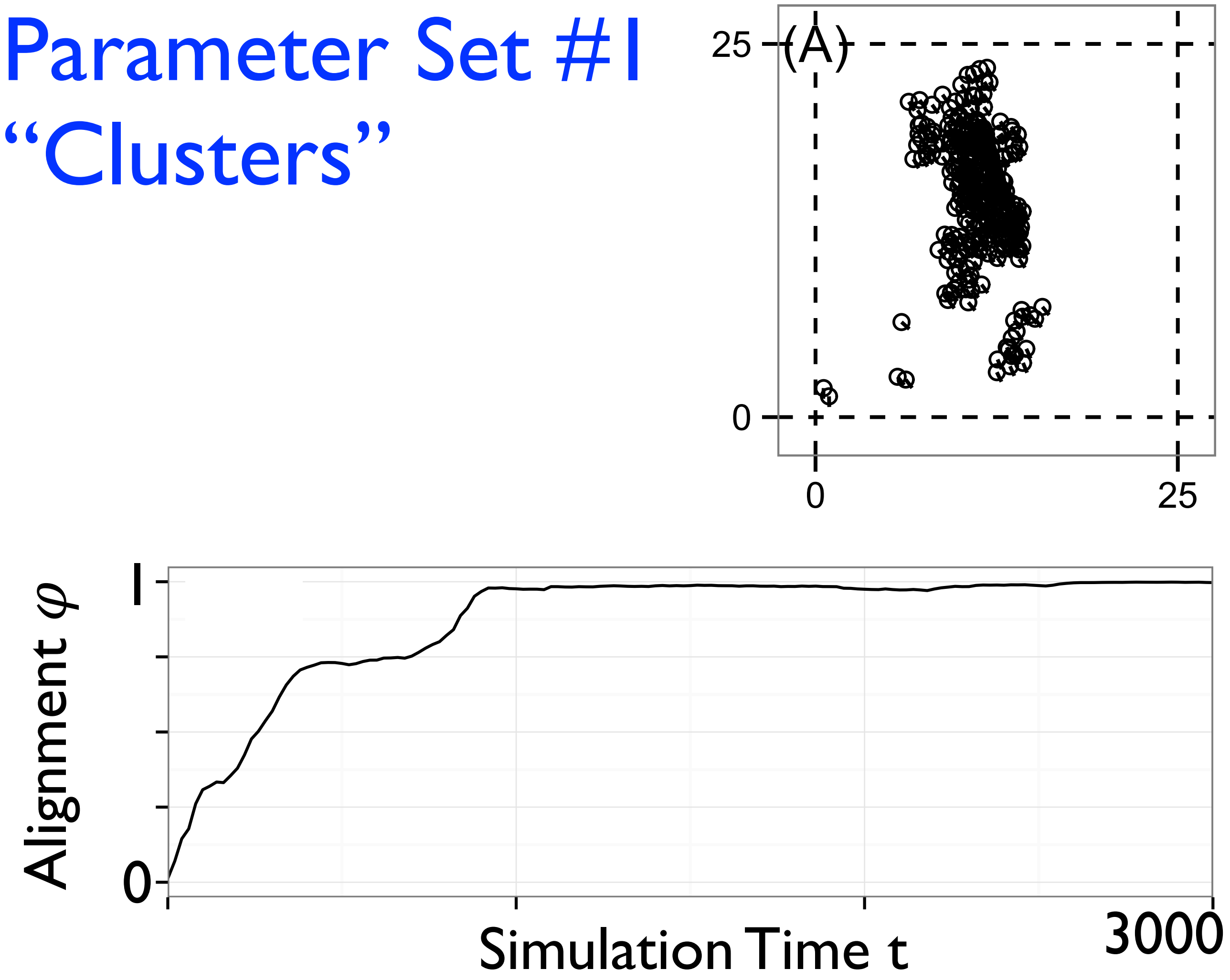
Clusters?

Loose alignment?

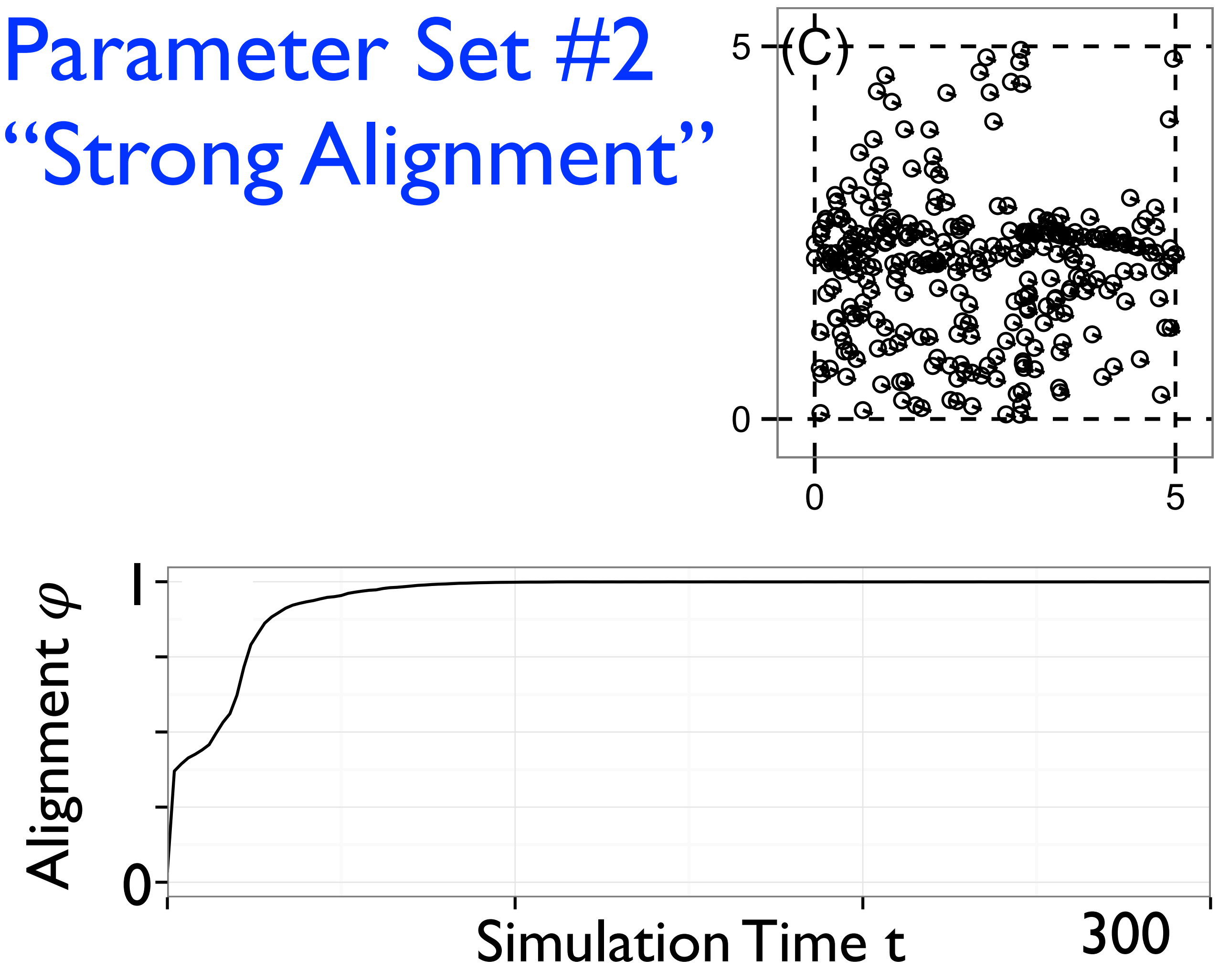
Strong alignment?

Traditional order parameter time series that look similar…

Parameter Set #1
"Clusters"

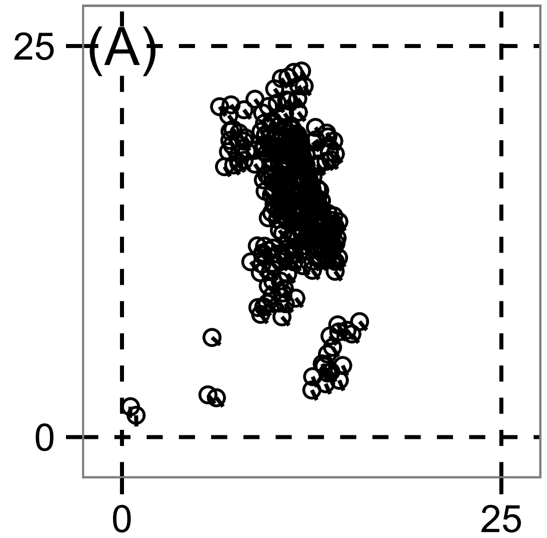


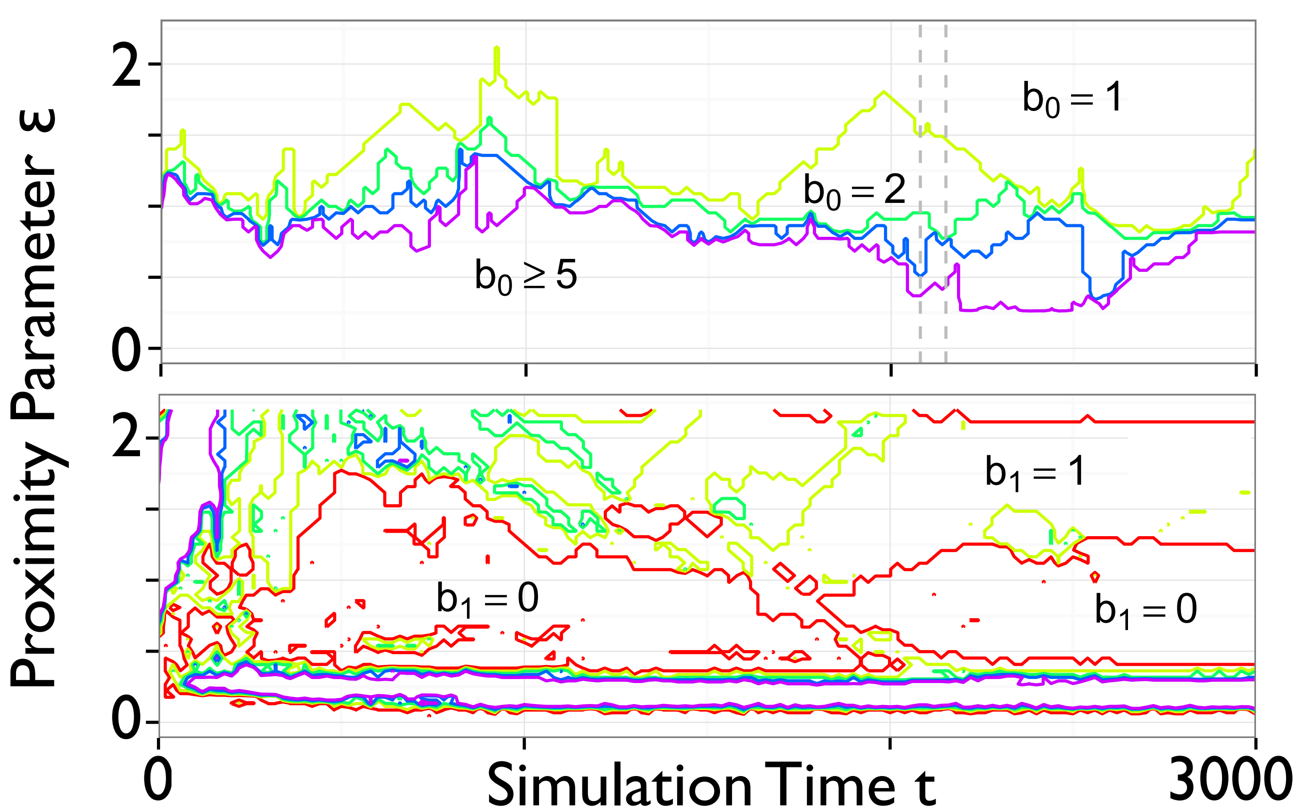
Parameter Set #2
"Strong Alignment"



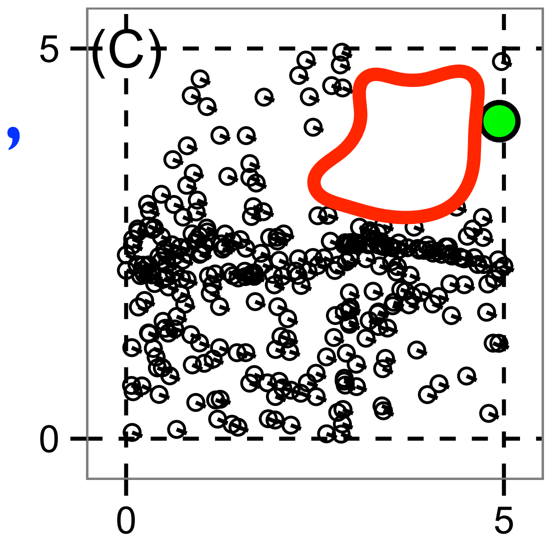
…can have drastically different topological signatures.

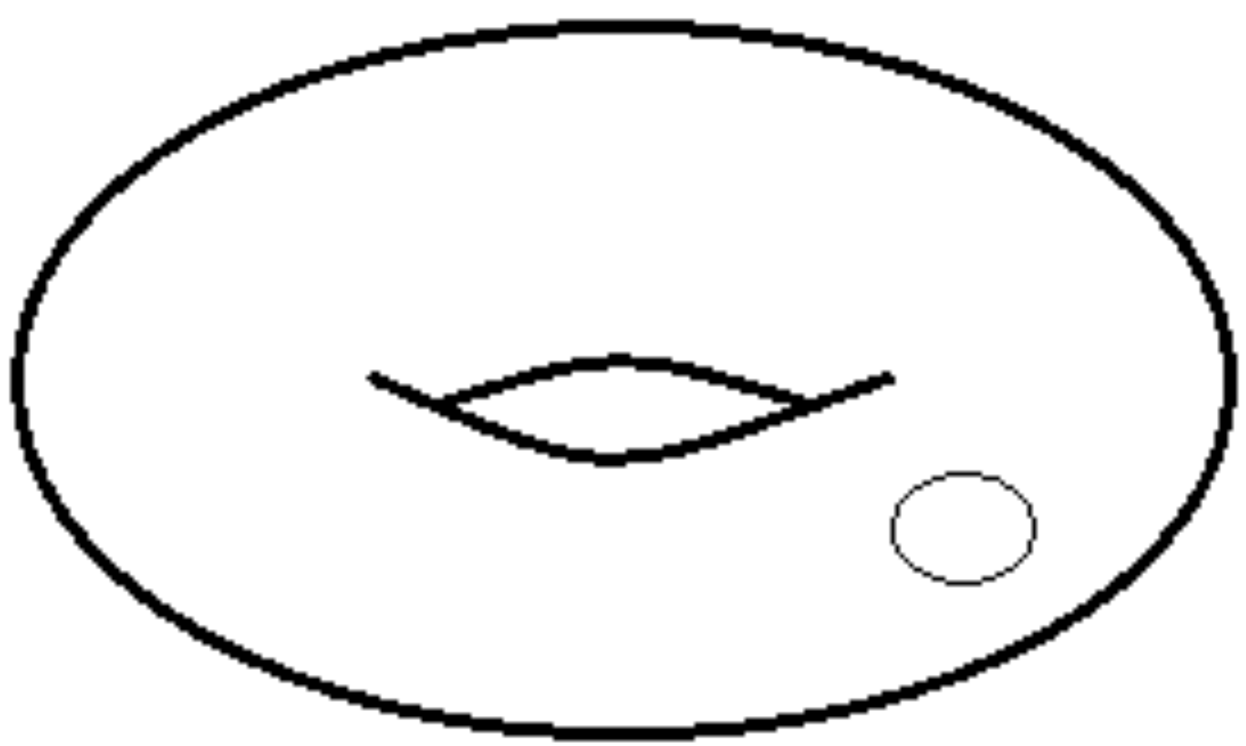
Parameter Set #1
"Clusters"





Parameter Set #2
"Strong Alignment"





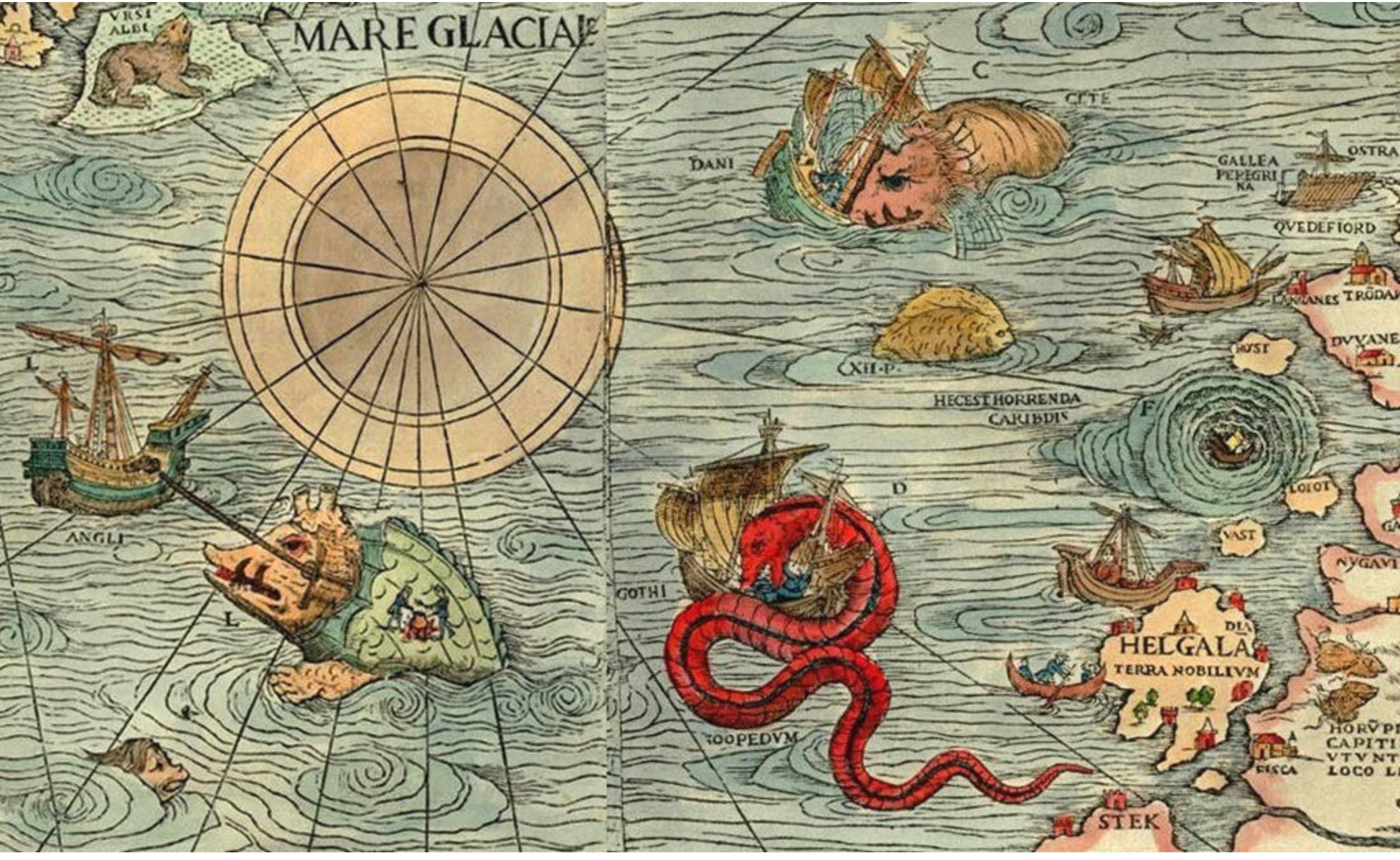
This talk, in a nutshell

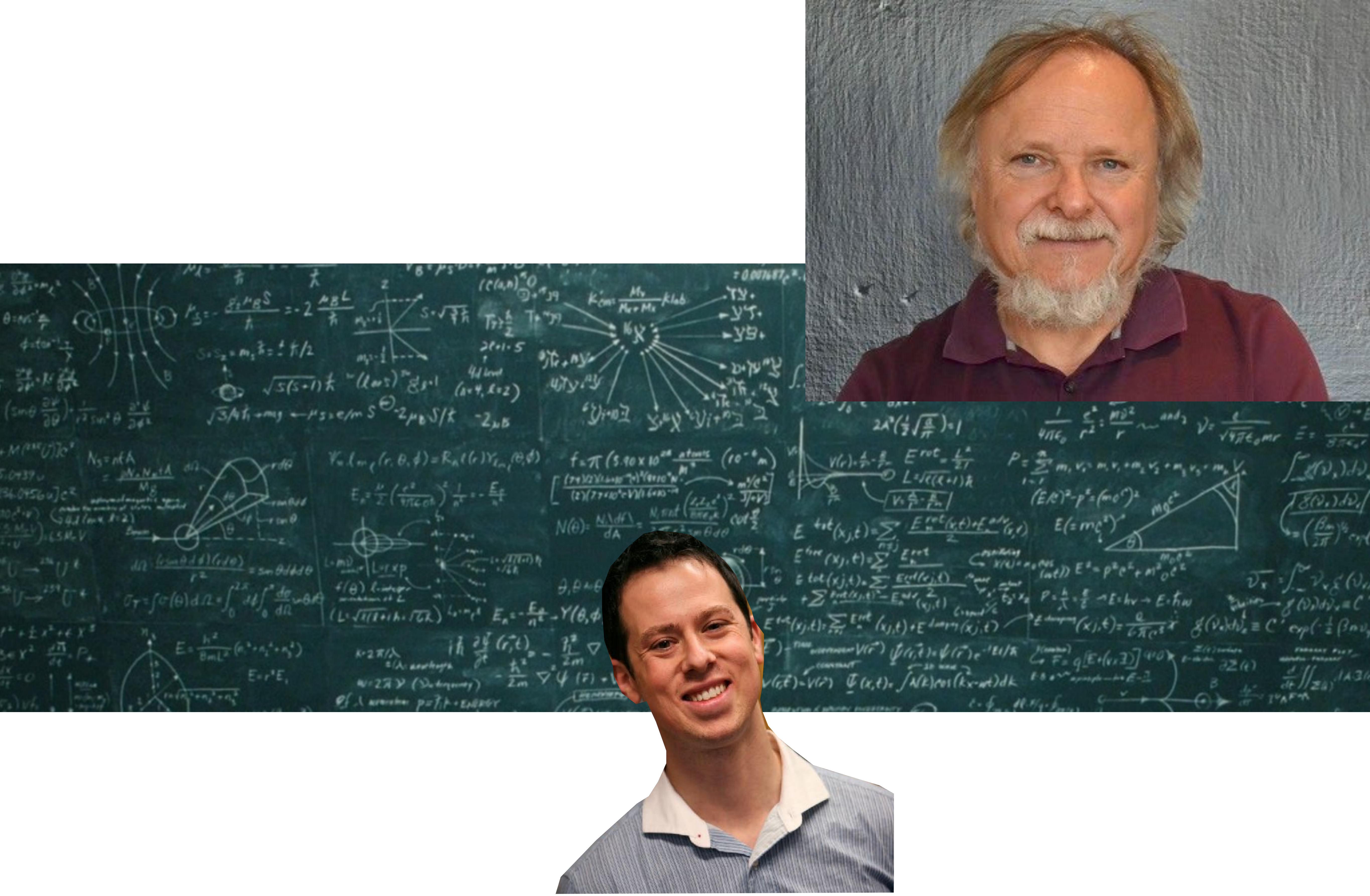
Part I: Applying a topological lens to biological aggregation model data

Message: TDA can be a useful tool for exploratory data analysis.

Part II: Moving towards topological reductions of a complex system

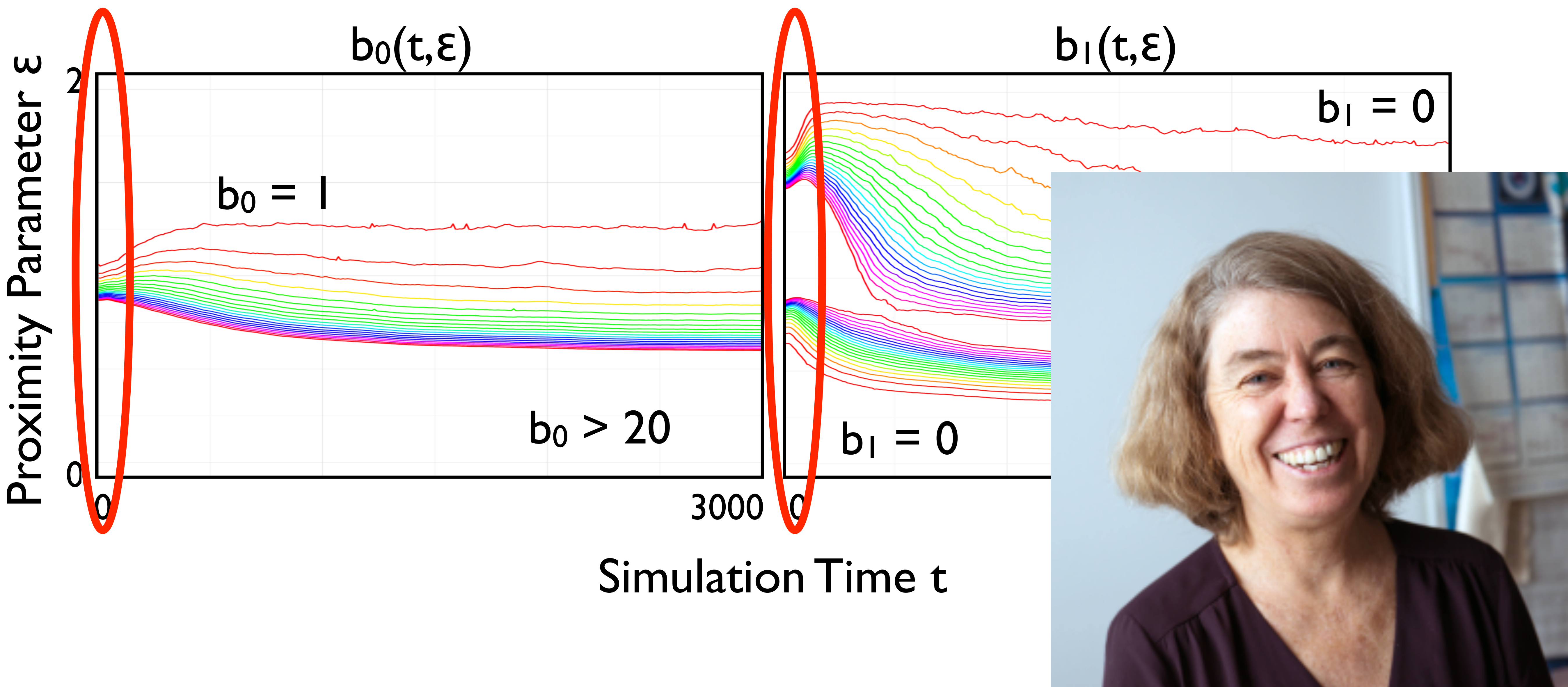
Message: When dynamics are neither highly ordered nor totally random, a topological description might be appropriate, but the approach is analytically challenging.


MARE GLACIA
VRSI ALBI
DANI
CETE
GALLEA PEREGRI NA
QVEDEFIORD
HEC EST HORRENDA CARIBDIS
ANGLI
GOTHI
HELGALA
TERRA NOBILIVM
STEK
VAST

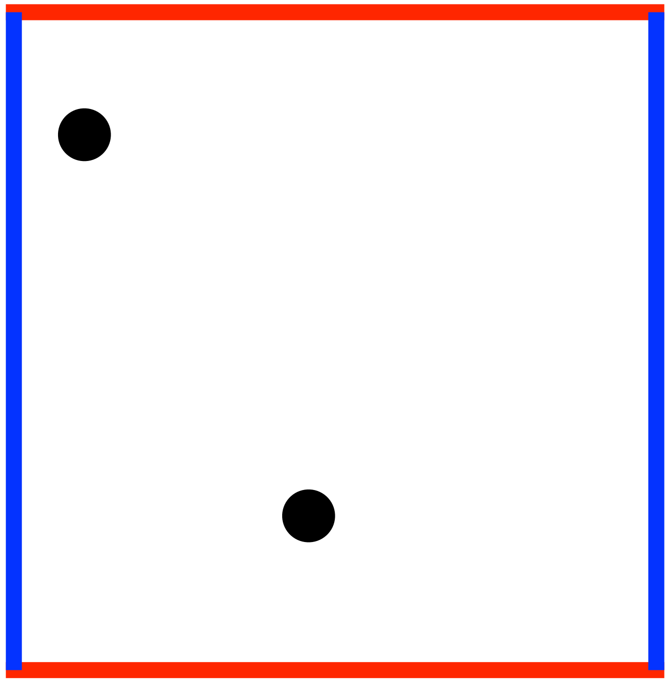


Do time series of random processes have average homology?

Vicsek model (naive) average over n = 1000 simulations



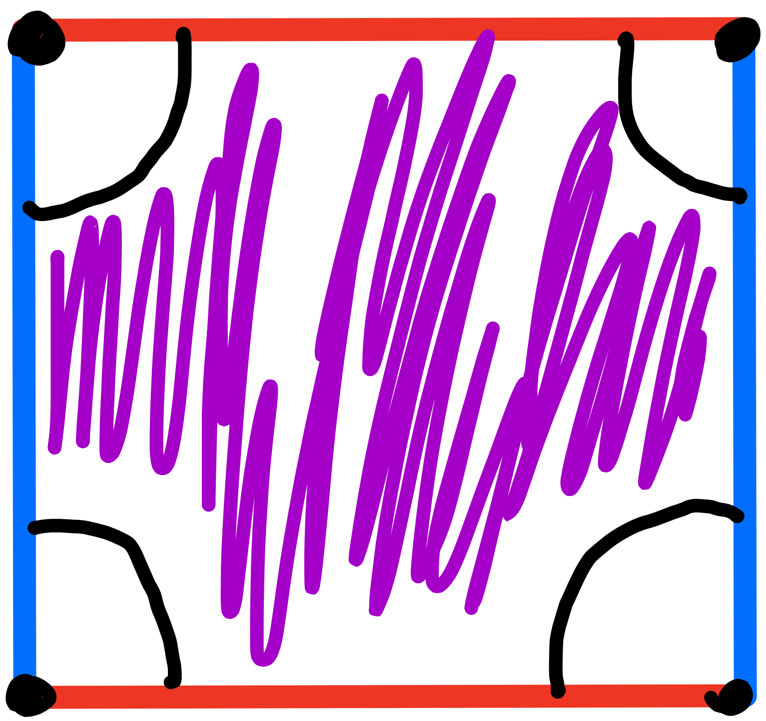
Expected value of $b_0(\epsilon)$ for an impressive ensemble of 2 points?



$$
\begin{aligned}
b_0(\epsilon) &= 2 \cdot P(\text{disconn.}) + 1 \cdot P(\text{conn.}) \\
&= 2 \cdot [1 - P(\text{conn.})] + P(\text{conn.}) \\
&= 2 - P(\text{conn.})
\end{aligned}
$$

$P(\text{conn.}) = P(\text{conn.};\epsilon) = ???$

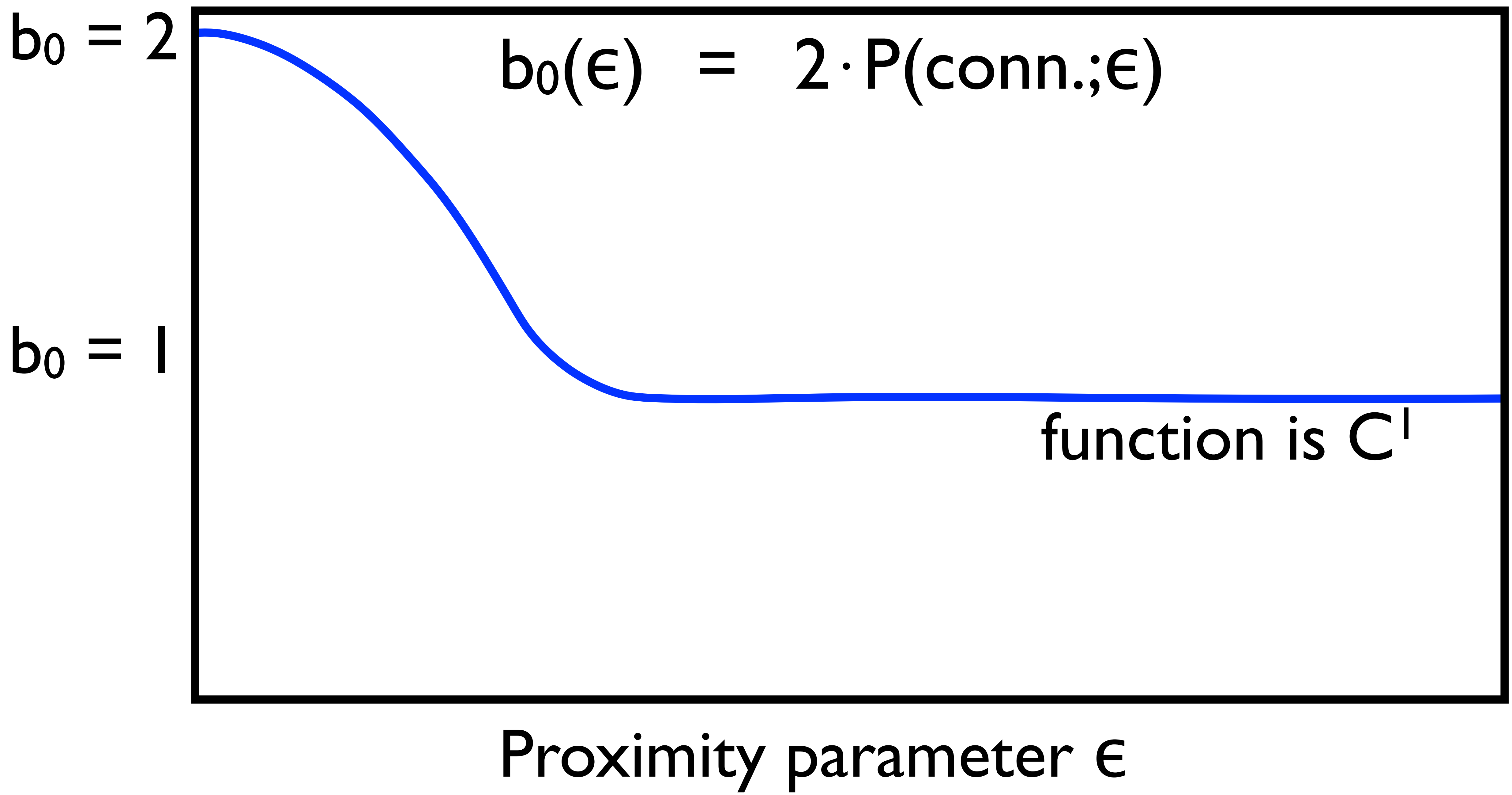
Expected value of $b_0(\epsilon)$ for an impressive ensemble of 2 points?



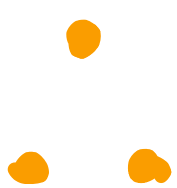
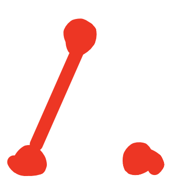
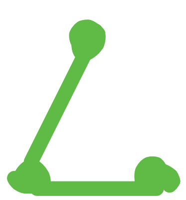
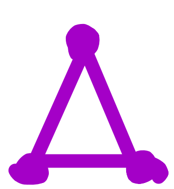
$\varepsilon \leq \frac{1}{2}$

$P = \frac{\pi \varepsilon^2}{1^2}$

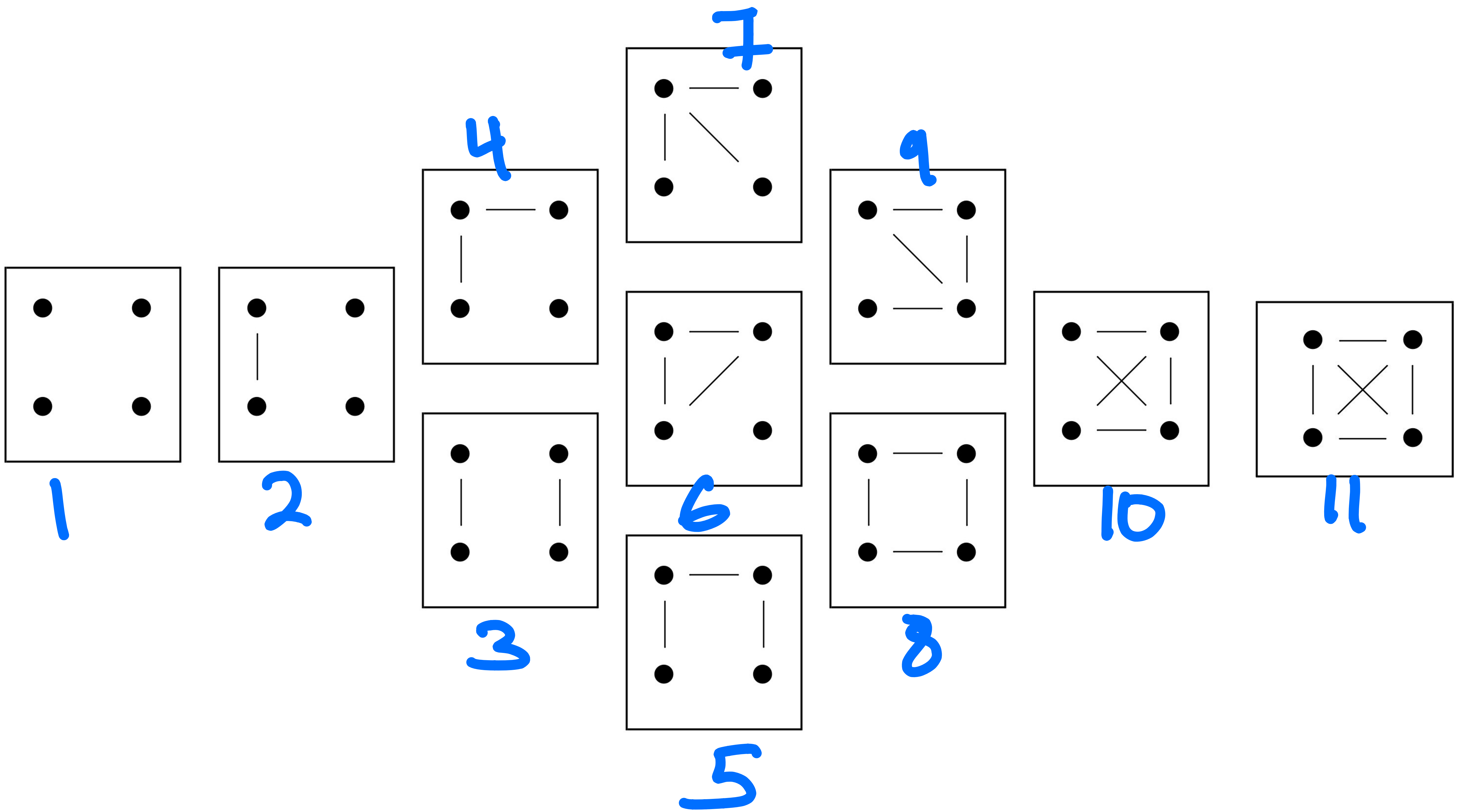
Expected value of $b_0(\epsilon)$ for an impressive ensemble of 2 points?



Expected value of $b_0(\epsilon)$ for an impressive ensemble of 3 points?

Class	# in class	b_0	b_1	Probability
	1	3	0	?
	3	2	0	?
	3	1	0	?
	1	1	0	?

Expected value of $b_0(\epsilon)$ for an impressive ensemble of 4 points?



Expected value of $b_0(\epsilon)$ for an impressive ensemble of 4 points?

Class	# in class	b_0	b_1	Probability
1	1	4	0	?
2	6	3	0	?
3	3	2	0	?
4	12	2	0	?
5	12	1	0	?
6	4	2	0	?
7	4	1	0	?
8	3	1	1	?
9	12	1	0	?
10	6	1	0	?
11	1	1	0	?

What is the homology of a random geometric graph?

- Random Geometric Graphs [M. Penrose, 2003]
- Topology of random geometric complexes: A Survey [Bobrowski and Kahle, 2014]

Types of results:

- Bounds
- Limiting results ($N \to \infty$, $\varepsilon \to 0$)
- Hard expressions

What is the homology of a random geometric graph?

Theorem 3.2.1 (Penrose, [47]). *If* $\Lambda = \lambda \in (0, \infty)$, *then:*

$$\frac{\beta_0(n)}{n} \xrightarrow{L^2} \int_{\mathbb{R}^d} \left(\sum_{k=1}^{\infty} k^{-1} p_k(\lambda f(x)) \right) f(x) dx,$$

where

$$p_k(t) = \frac{t^{k-1}}{k!} \int_{(\mathbb{R}^d)^{k-1}} h(0, y_1, \ldots, y_{k-1}) e^{-tA(0,y_1,\ldots,y_{k-1})} dy_1 \cdots dy_{k-1},$$

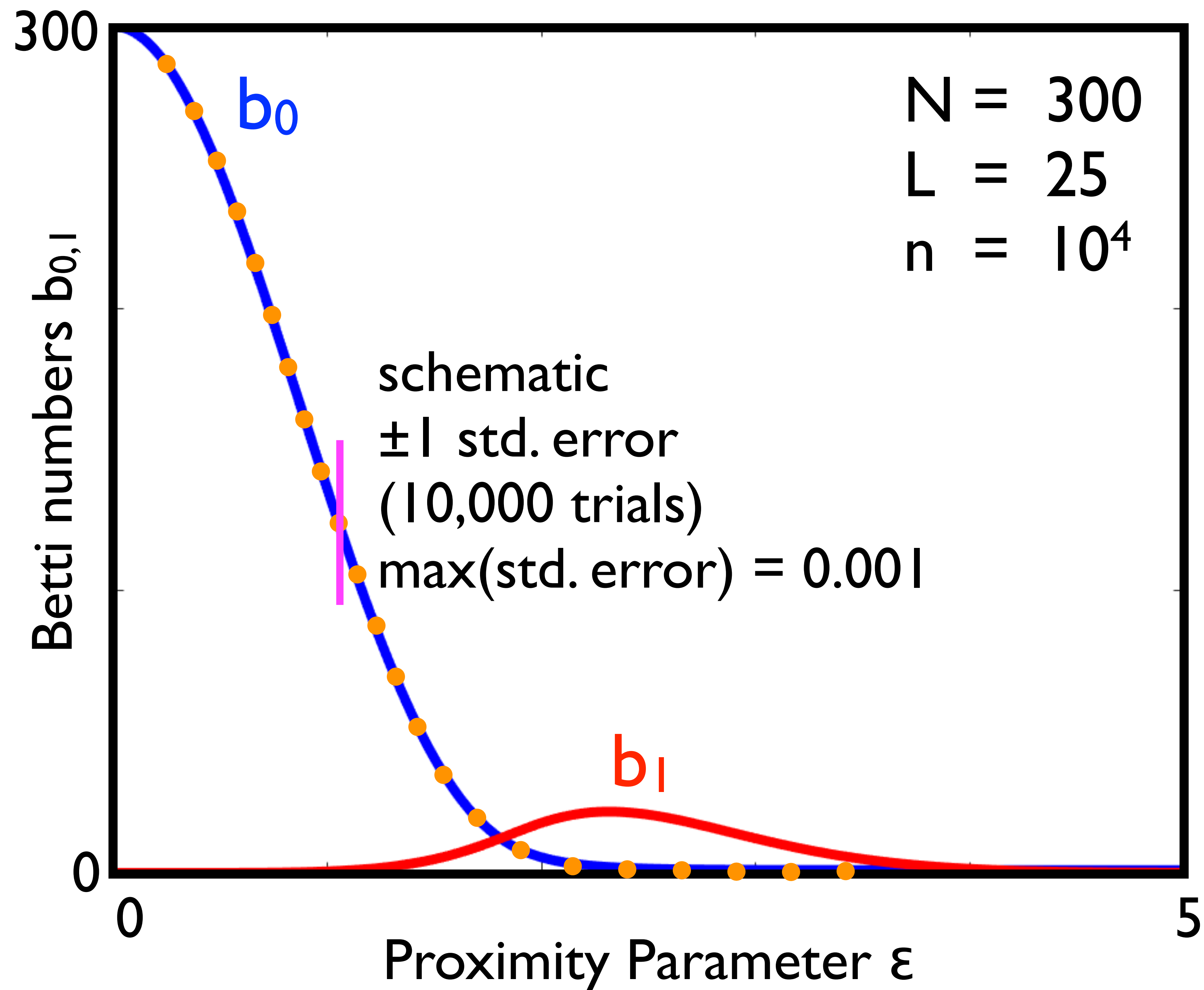
$$h(x_1, x_2, \ldots, x_k) = \begin{cases} 1 & G(\{x_1, x_2, \ldots, x_k\}, 1) \text{ is connected}, \\ 0 & otherwise, \end{cases}$$

and

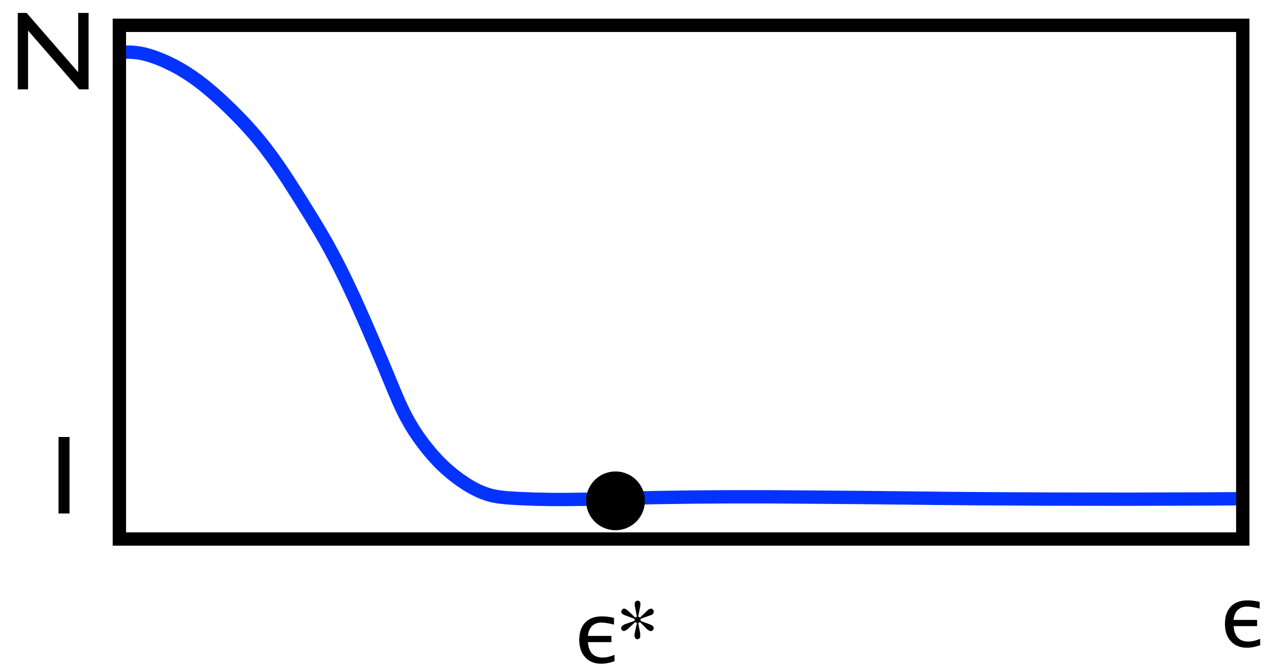
$$A(x_1, x_2, \ldots, x_k) := |\bigcup_{j=1}^{k} B_1(x_j)|.$$

The infinite sum in (3.2) comes from the fact that we need to count the number of components consisting of any possible number of vertices. The limiting expression provided by the theorem is highly intricate, and at this point impossible to evaluate analytically. Nonetheless, as we will

What is the homology of a random points on flat torus?



Try modeling the topological signature.

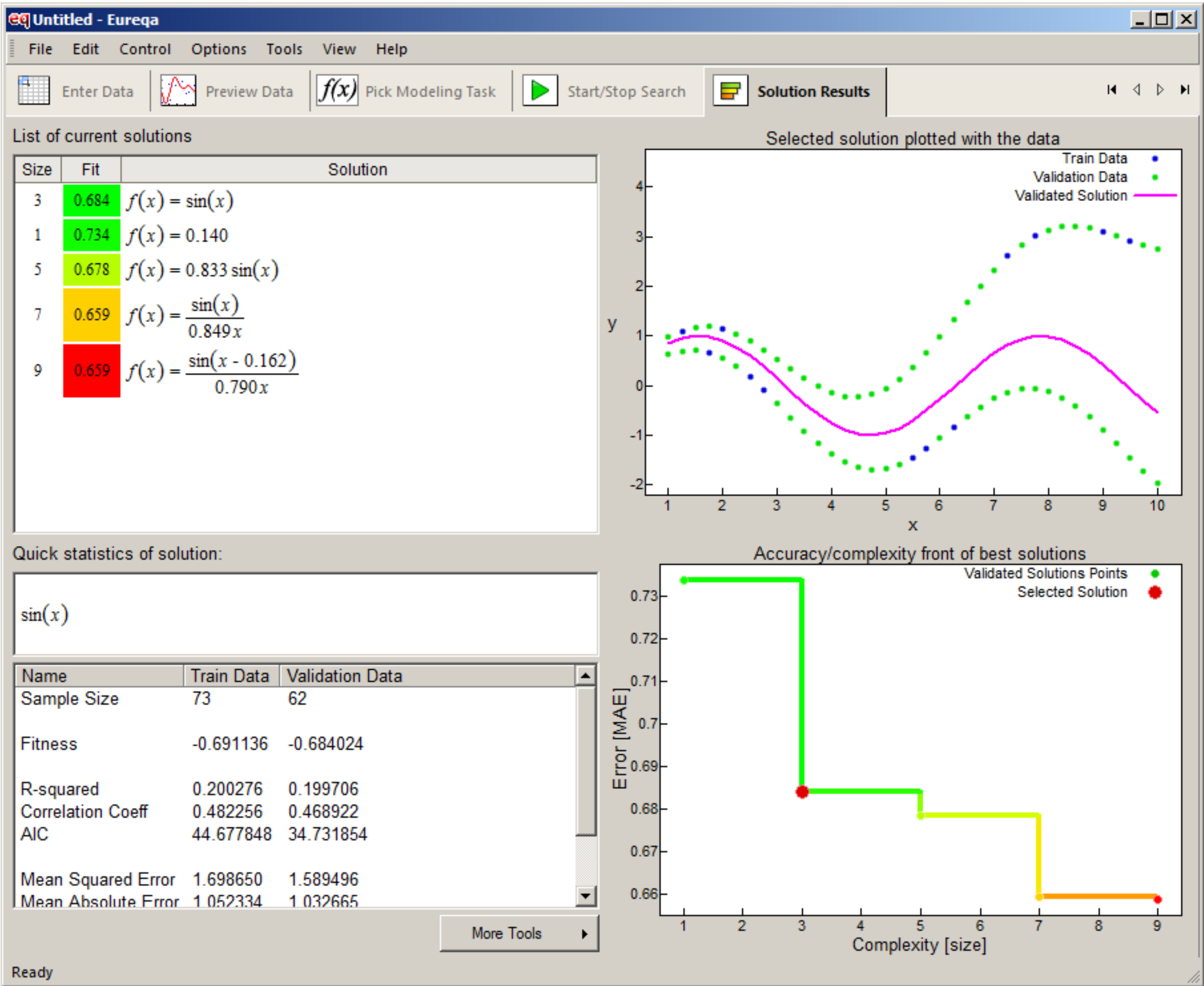


+ dimensional analysis

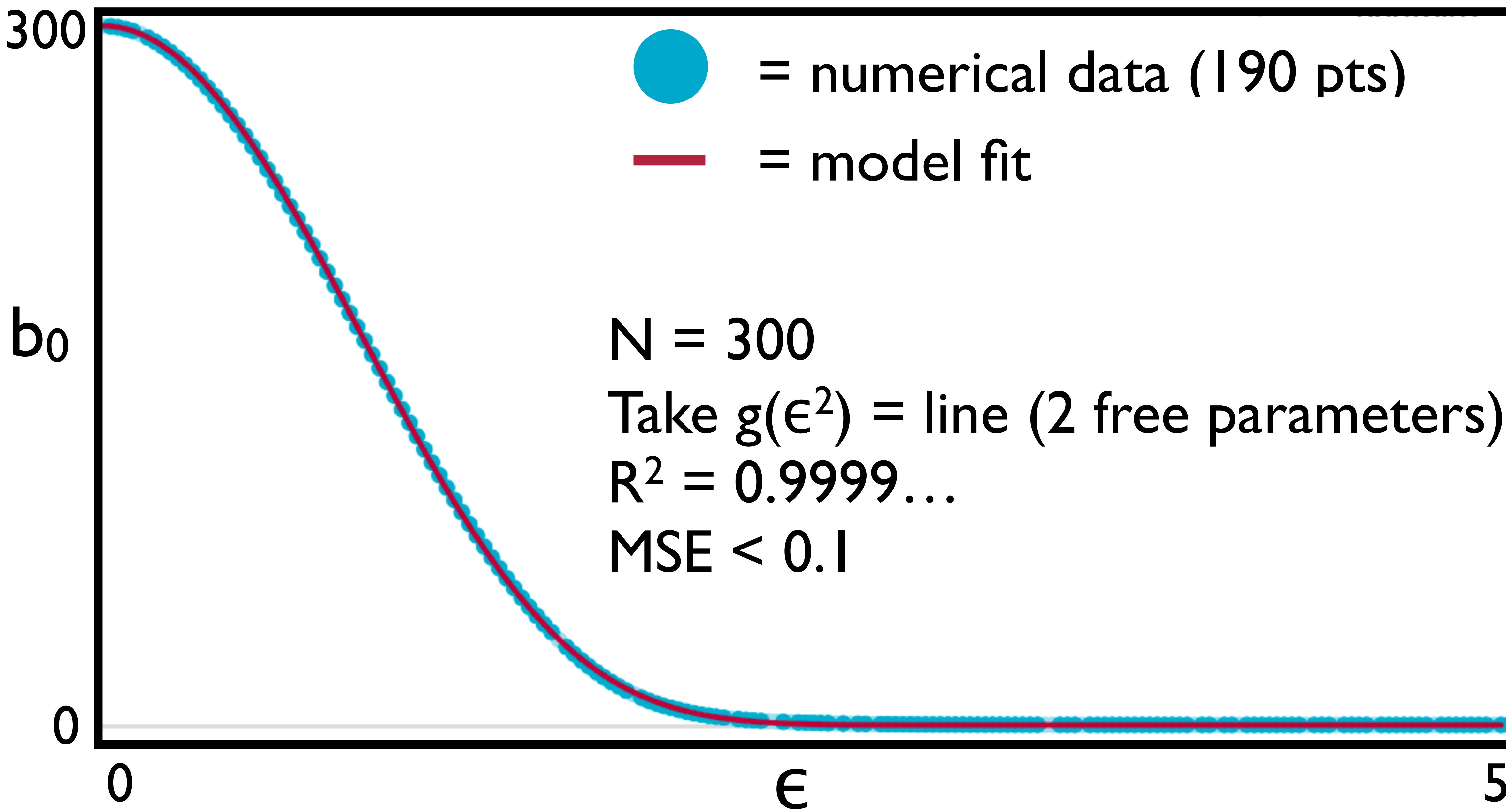
Let $b_0(\epsilon) = 1 + (N - 1) \cdot f(\epsilon^2)$, $f(0) = 1, f(\epsilon^2) = 0$ for $\epsilon \geq \epsilon_*$

$$f(\epsilon^2) = \exp\left[\frac{1}{g(0)\epsilon_*^2} - \frac{1}{\epsilon_*^2 - \epsilon^2}\frac{1}{g(\epsilon^2)}\right]$$

Try modeling the topological signature.



Try modeling the topological signature.



This talk, in a nutshell

Part I: Applying a topological lens to biological aggregation model data

Message: TDA can be a useful tool for exploratory data analysis

Part II: Moving towards topological reductions of a complex system

Message: When dynamics are neither highly ordered nor totally random, a topological description might be appropriate, but the approach is analytically challenging

Applied mathematical modeling with topological techniques Summer (???) 2019

Organizers:
Henry Adams, Colorado State University
Maria D'Orsogna, Cal State Northridge
Rachel Neville, University of Arizona
Jose Perea, Michigan State University
Chad Topaz, Williams College