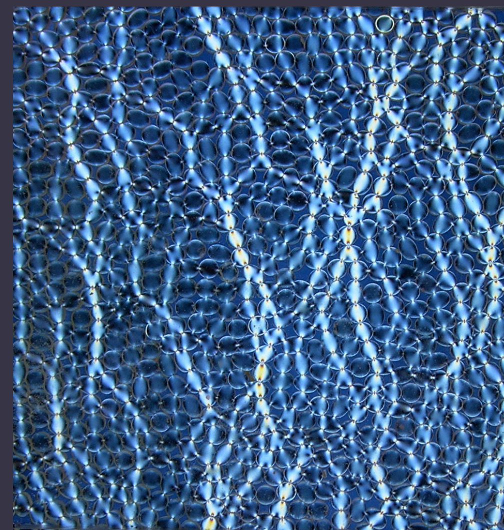
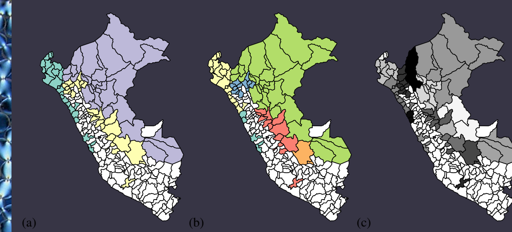
Topological Data Analysis of Spatial Systems

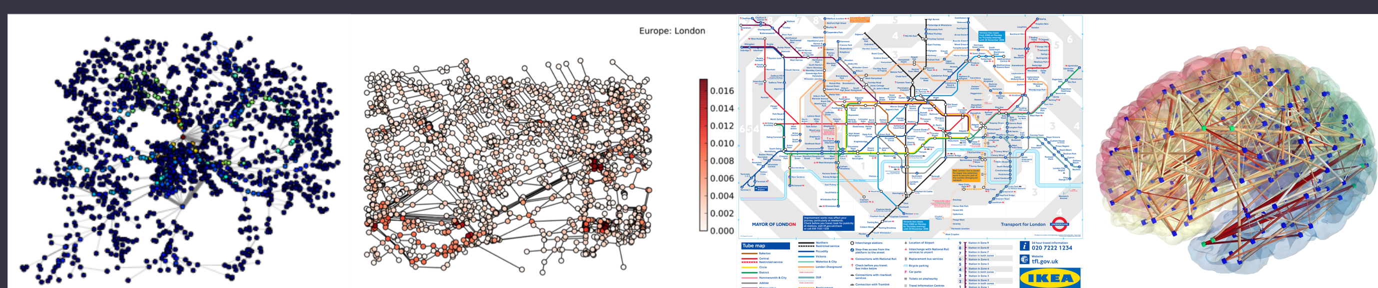
Mason A. Porter
Department of Mathematics, UCLA
Department of Sociology, UCLA
Santa Fe Institute

Spatial Systems

- Space has a major influence on the structure and dynamics of complex systems
- Useful reference: Marc Barthelemy, *Spatial Networks*, 2022

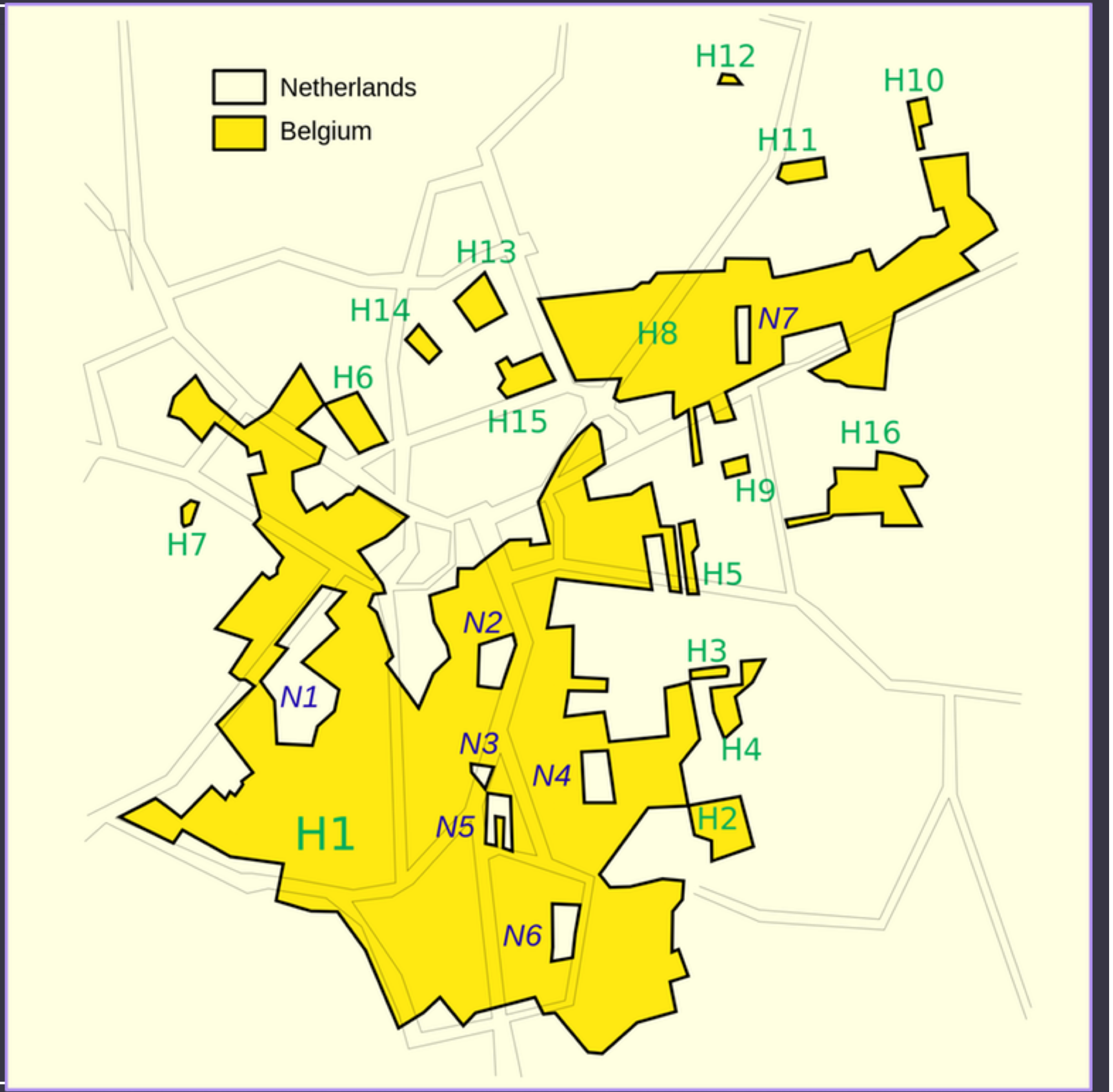






Topological Dreams
(or Nightmares)

The border between Belgium and the Netherlands at Baarle-Nassau/Baarle-Hertog



Quantifying "Political Islands"

How do we detect red voters in a sea of blue?
(Or light blue voters in a sea of dark blue?)

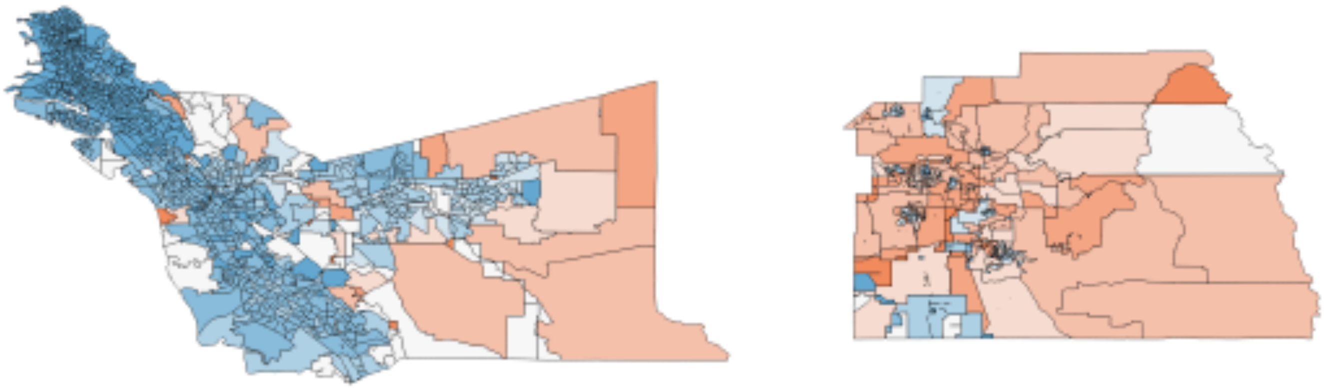


Fig. 1: The counties of (left) Alameda and (right) Tulare. Red precincts voted predominantly for Donald Trump, and blue ones voted predominantly for Hillary Clinton. Darker shading in a precinct indicates a stronger majority for the winning candidate, so Trump won dark-red precincts by a large margin and Clinton won dark-blue precincts by a large margin. We use the color white for precincts with an equal number of votes for the two candidates.

Spiders Spinning Under the Influence

- The Marshall Space Flight Center studied the webs of spiders that were exposed to various chemicals. (There is a NASA Tech Brief from 1995.)
  - Earlier work, starting in 1948 by Swiss pharmacologist Peter N. Witt
- They concluded that more toxic chemicals resulted in more deformed spiderwebs

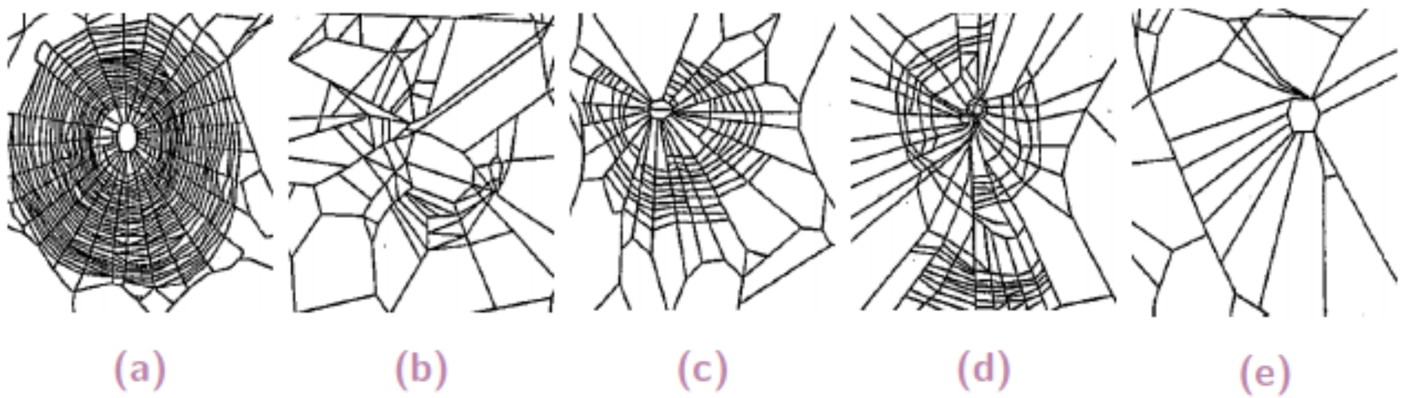


Figure 15: Drug-free, caffeine, marijuana, speed, and chloral hydrate influenced spiderwebs

Resource Coverage: Pubs in the United Kingdom

P. Corcoran & C. B. Jones [2023], "Topological Data Analysis for Geographical Information Science Using Persistent Homology", *International Journal of Geographical Information Science*, Vol. 37, No. 3: 712–745

(a) (b)

Figure 1. The set of Cardiff and Manchester city pub locations are displayed using red dots in (a) and (b), respectively. In both cases, the city street network in question is also represented in the background to provide context.

A Couple of References

- Gunnar Carlson [2020]: "Topological Methods for Data Modelling", *Nature Reviews Physics*, Vol. 2: 697–707

- Nina Otter, MAP, Ulrike Tillmann, Peter Grindrod, and Heather A. Harrington [2017]: "A Roadmap for the Computation of Persistent Homology", *European Physical Journal — Data Science*, Vol. 6: 17
    - Includes a tutorial for installing and using software for PH. If you are trying this stuff for the first time, this is an article to help guide you through things.

Some Lectures Notes and Books

- Vidit Nanda [2021]: "Computational Applied Topology", Mathematical Institute, University of Oxford
  - Lecture notes: http://people.maths.ox.ac.uk/nanda/cat/TDANotes.pdf
  - Individual topics and videos: http://people.maths.ox.ac.uk/nanda/cat/

- Book: Herbert Edelsbrunner and John Harer [2010]: *Computational Topology: A Introduction*
  - One can purchase it, but some versions are available online (e.g., at https://www.maths.ed.ac.uk/~v1ranick/papers/edelcomp.pdf)

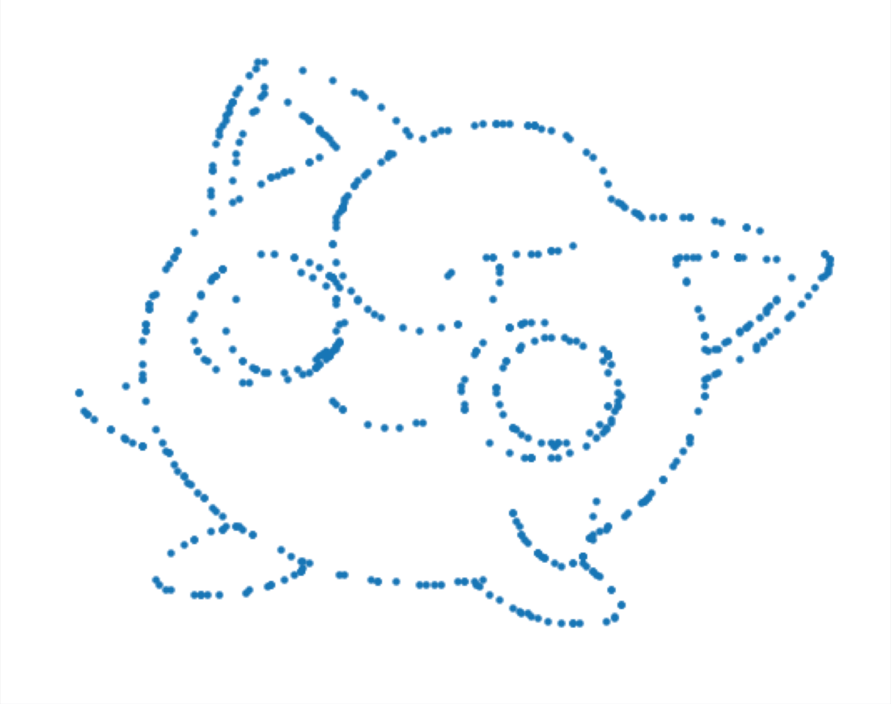
- Book: Tamal Krishna Dey and Yusu Wang [2022]: *Computational Topology for Data Analysis*
  - One can purchase it, but some versions are available online (e.g., at https://www.cs.purdue.edu/homes/tamaldey/book/CTDAbook/CTDAbook.html)

Donuts and Coffee Cups

https://youtu.be/9NlqYr6-TpA

(Henry Segerman and Keenan Crane)

If we squint at a point cloud, what does it look like?



If we squint at a point cloud, what does it look like?

Traditional Persistent Homology: Thickening a Point Cloud and Tracking Changes in Topology

[Figure from: Michelle Feng, Abigail Hickock, Yacoub H. Kureh, MAP, & Chad M. Topaz, "Connecting the Dots: Discovering the "Shape" of Data", *Frontiers for Young Minds*, 2021]

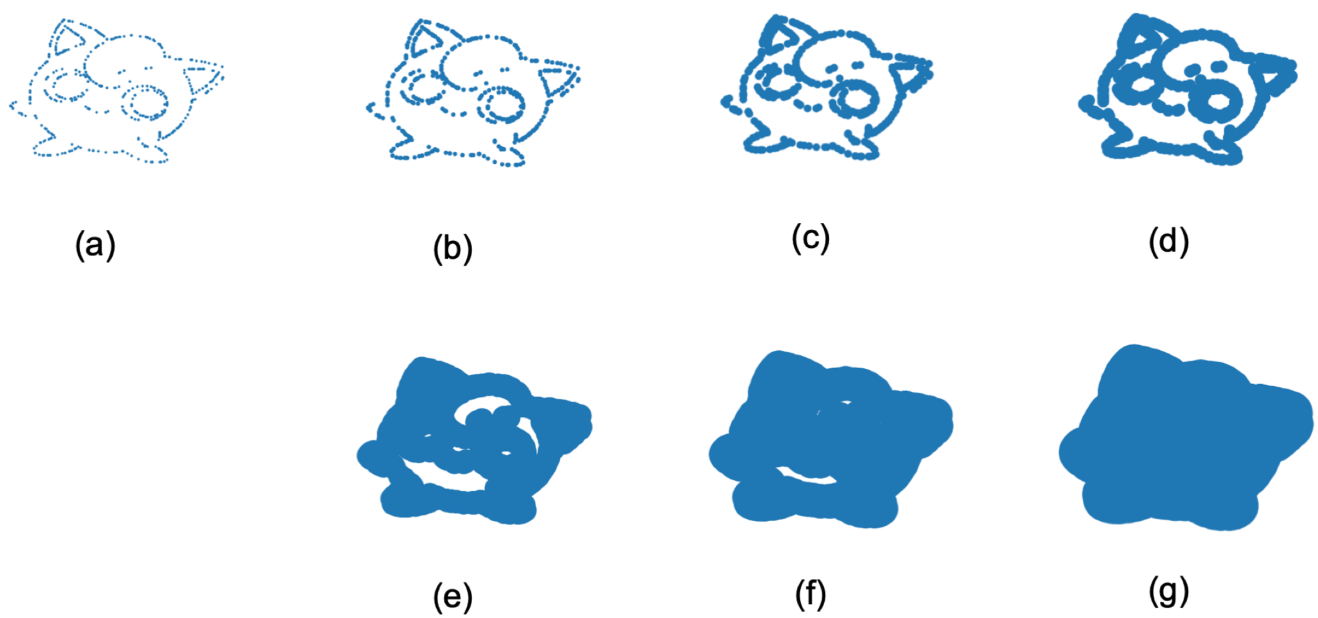


Fig 3. Increasing the size of dots in data. At first, Jigglypuff becomes easier to see as the dots get larger. Jigglypuff then gets harder to see.

Counting the Numbers of Components (H_0) and 1D Holes (H_1)

	Picture comes from	Number of pieces	Number of holes
	Figure 3A	224	0
	Figure 3B	101	0
	Figure 3C	17	2
	Figure 3D	1	6
	Figure 3E	1	6
	Figure 3F	1	3
	Figure 3G	1	0

Table 1

What is the point?

- Algorithmic methods to study (potentially high-dimensional) data in a quantitative manner
  - Data from point clouds, networks, images, time series, etc.
- Examine the "shape" of data
- Persistent homology (PH)
  - Mathematical formalism to study topological invariants
  - Fast algorithms
  - Persistent structures: a way to cope with noise in data

Algebraic Topology

The key subject with underlying mathematical ideas for topological data analysis (TDA)

Key Idea

- We want to describe the properties of an object that stay the same if we stretch it or shrink it or bend it, but without us gluing things together or tearing the object
- Seek *topological invariants*
  - Number of components
  - Number of holes
  - Number of "holes" in different dimensions (loops, cavities, etc.)
  - Remark: Today we are considering *homological invariants*, but there are also other types of topological invariants

Simplicial Complexes

- A *k-simplex* is a k-dimensional polytope that is the convex hull of its k+1 vertices.
- An *m-face* of a k-simplex is a subset of size m+1 and is an m-simplex itself
- A *simplicial complex S* is a set of simplices that satisfies the following conditions:
  - Every face of a simplex of *S* is also in *S*
  - The non-empty intersection of any two simplices $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in S$ is a face of both $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$

Bernadette J. Stolz et al., *Chaos*, 2017

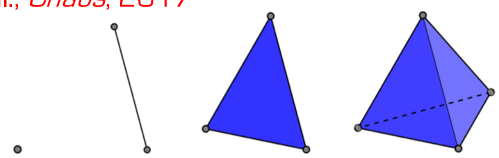
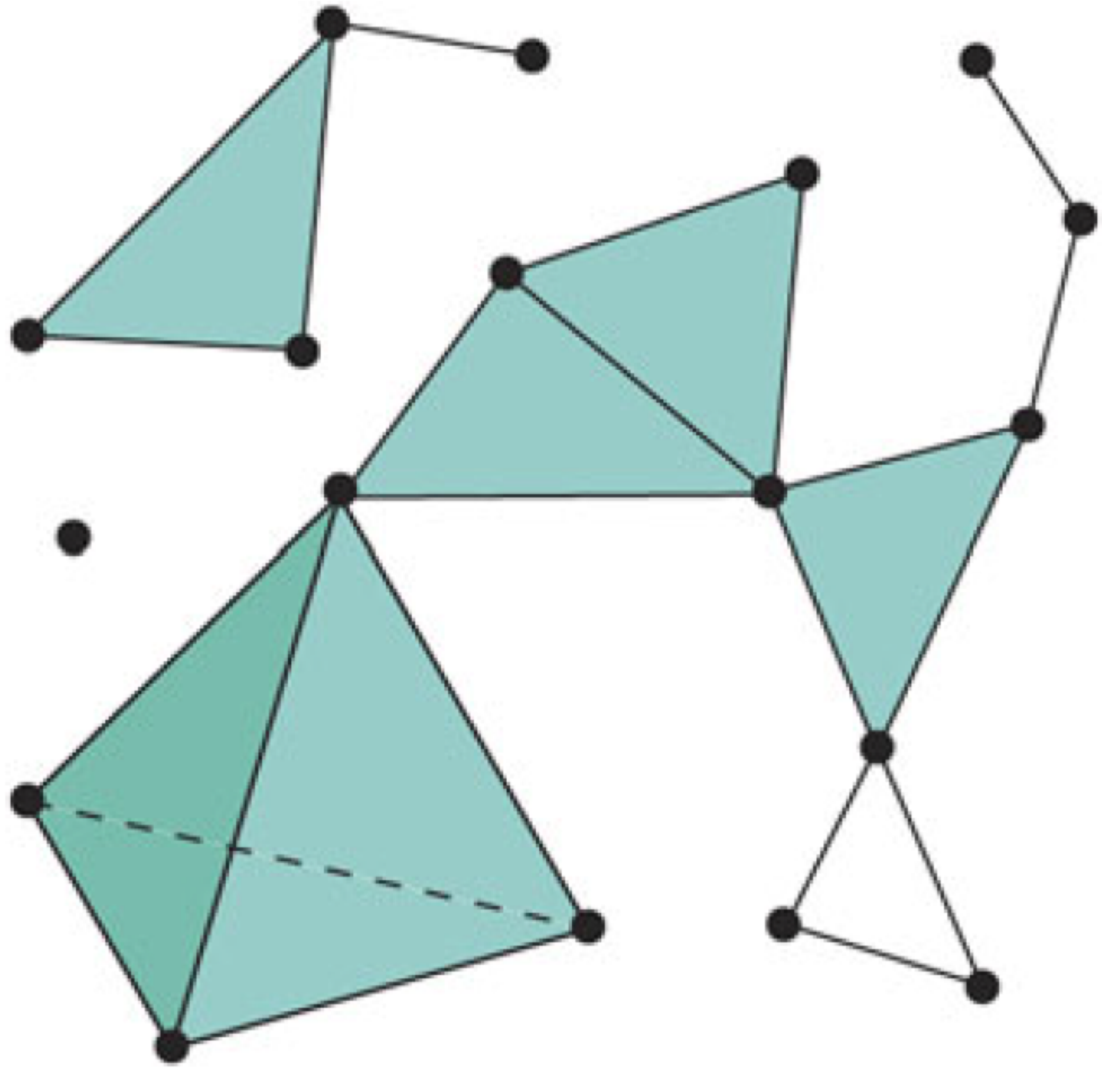


FIG. 1. From left to right, we show examples of a 0-simplex, a 1-simplex, a 2-simplex, and a 3-simplex. [We adapt these examples and the figure from [2].]

An Example of a Simplicial Complex

(picture from Wikipedia)



Persistent Homology (PH)

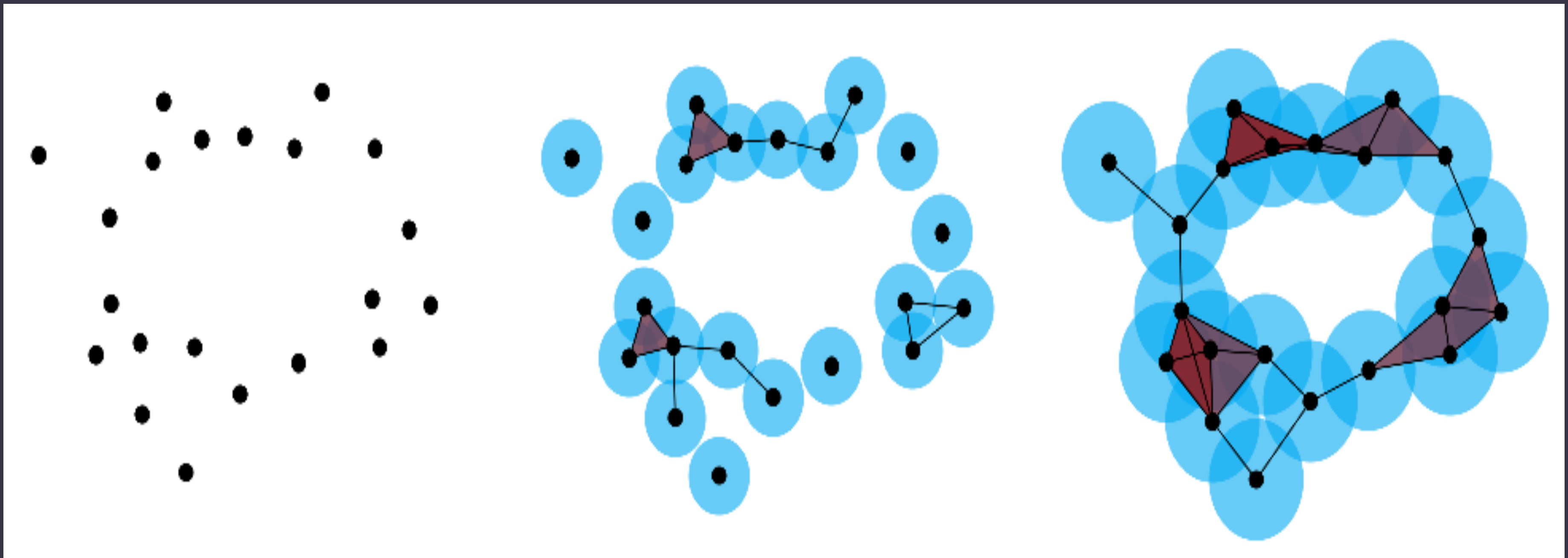
Use ideas from algebraic topology to analyze data

Filtered Simplicial Complex

- A *filtered simplicial complex* (i.e., a *filtration*) is a sequence of simplicial complexes.

$$S_0 \subseteq S_1 \subseteq S_2 \subseteq \ldots \subseteq S_n$$

- Think of each S_i as looking at a data set at a different scale.
  - Which topological structures (holes, etc.) exist across a range of values of i? In other words, which structures *persist* across scales?



A single-parameter filtration: $K_{r_0} \subseteq K_{r_1} \subseteq \cdots \subseteq K_{r_n}$

Vietoris–Rips Filtration

(this is what we did with Jigglypuff)

- 1. Fix a value of **ε**
- 2. For each point in a point cloud, center a ball of radius **ε** on it
- 3. Whenever k+1 of these **ε**-balls all overlap pairwise, create a k-simplex
- 4. The resulting collection of simplices is a simplicial complex S_ε
- 5. Increment **ε** and do steps 2–4 again (and keep doing this until you have a giant blob)

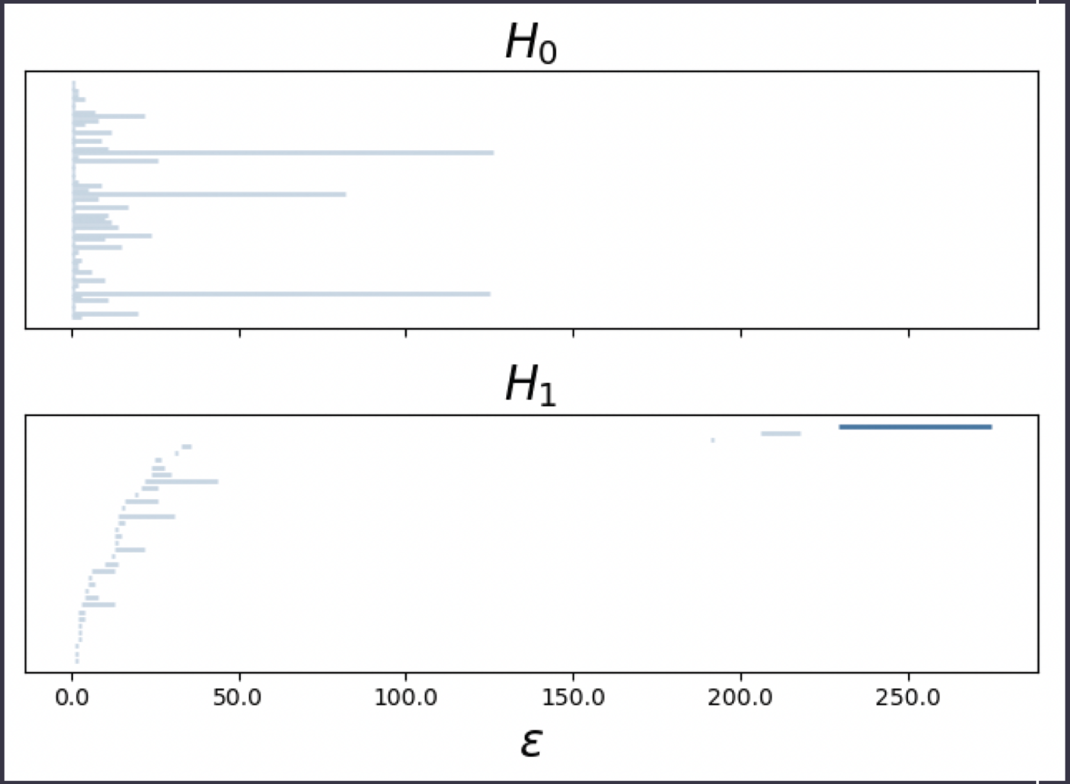
Birth and Death of Features

- A feature is *born* in S_i if this is the smallest i for which a feature exists
- A feature *dies* in S_i if this is the largest i for which a feature exists
  - Some features live forever
- There are various ways to track the birth and death of features in different dimensions
  - Barcodes
  - Persistence diagrams
  - …

Barcodes

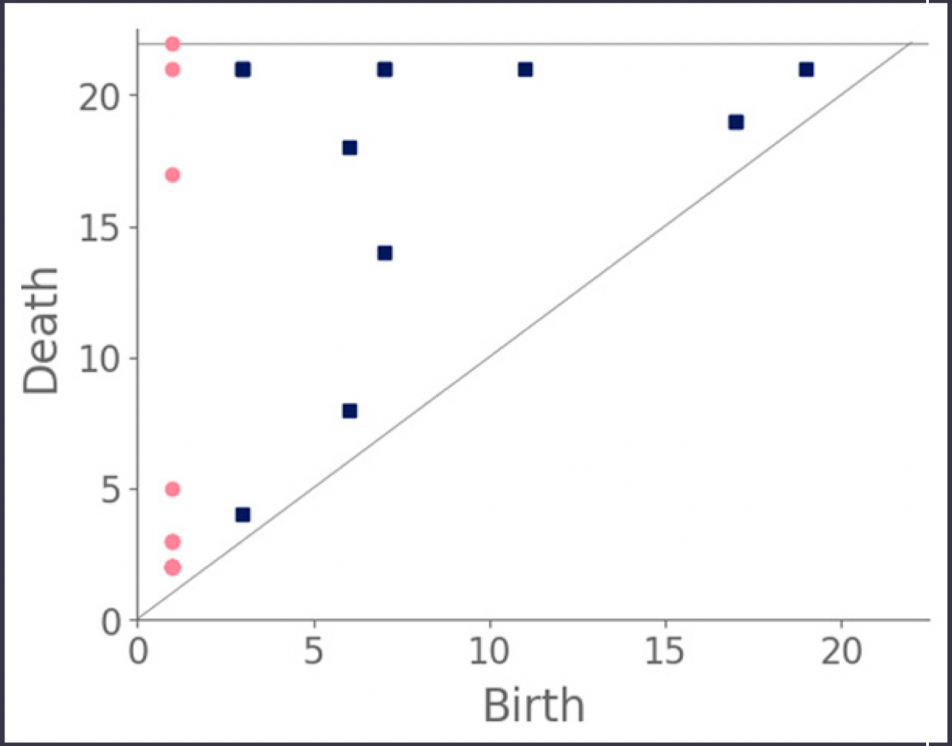
- Each interval represents a feature in dimension *n*
  - Left endpoint = “birth” of a feature
  - Right endpoint = “death” of a feature

- Visually, longer features are “more persistent”



Persistence Diagrams

- If a feature is born at *b* and dies at *d*, we place a point at (*b*,*d*)
- The height above the diagonal indicates the persistence
- Pink circles: H_0
- Blue squares: H_1



Topological Data Analysis of 2D Voting Data

Michelle Feng & MAP [2021], "Persistent Homology of Geospatial Data: A Case Study with Voting", *SIAM Review*, Vol. 63, No. 1: 67–99

Quantifying "Political Islands"

How do we detect red voters in a sea of blue?
(Or light blue voters in a sea of dark blue?)

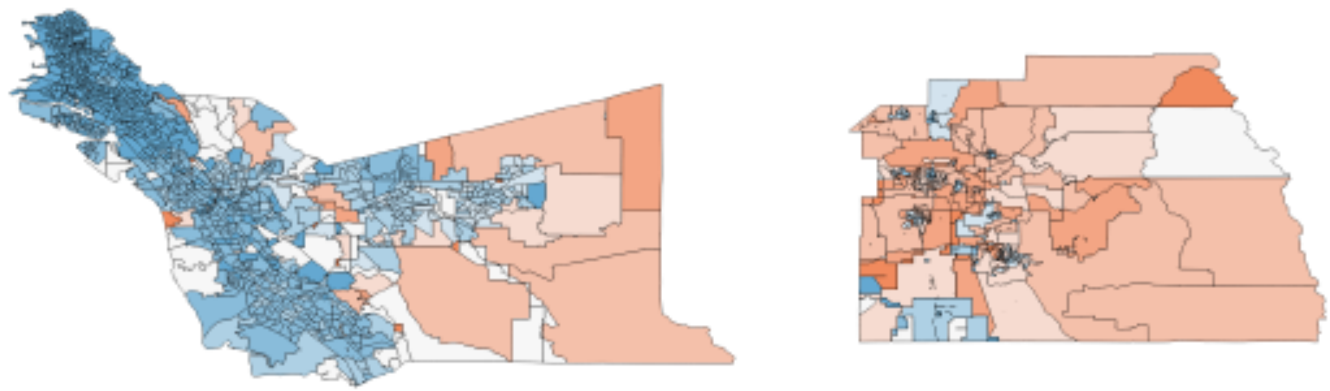
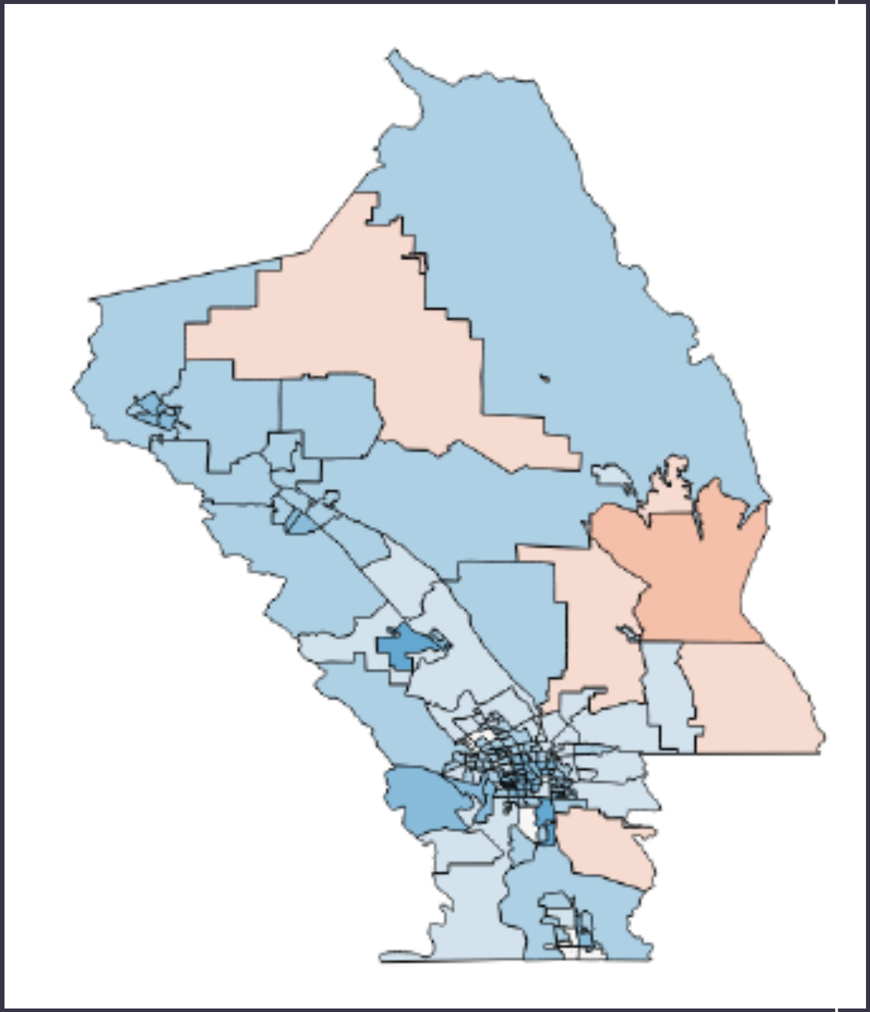


Fig. 1: The counties of (left) Alameda and (right) Tulare. Red precincts voted predominantly for Donald Trump, and blue ones voted predominantly for Hillary Clinton. Darker shading in a precinct indicates a stronger majority for the winning candidate, so Trump won dark-red precincts by a large margin and Clinton won dark-blue precincts by a large margin. We use the color white for precincts with an equal number of votes for the two candidates.

TDA and Voting Data

- Precinct-level voting data
- Topological methods allow us to find and identify holes
  - They also allow us to relate the presence of holes to global structure
- Want to find "political islands"
  - Red voters in a sea of blue, etc.
  - Consider these islands as "holes" in a manifold in which all precincts vote similarly



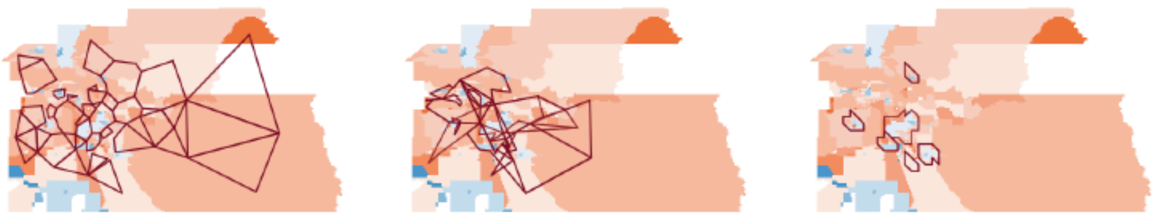
Adjacency Simplicial Complex

- Use network adjacency to define simplices
- If $n + 1$ nodes are all pairwise adjacent, define an n–simplex
- Given appropriate node data (or edge data), we construct a filtration
  - The filtration is not determined by distance
- In our data, the filtration parameter tracks the strength of preference for a specific candidate
  - For example, we can find light-blue precincts in a sea of dark blue

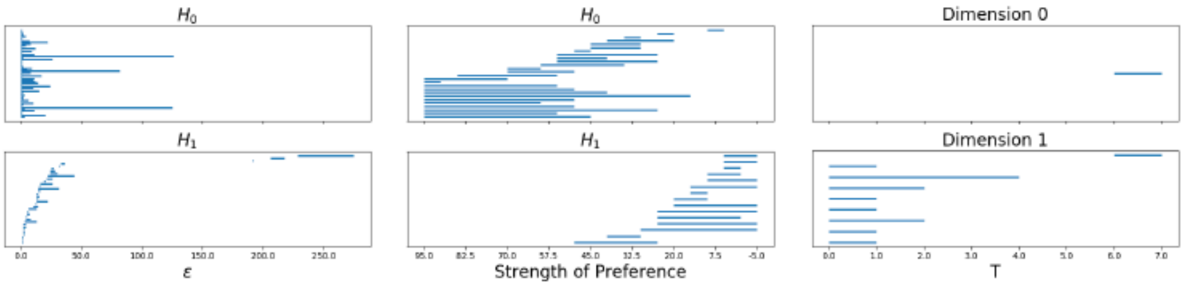
Level-Set VR Complex

- Use data in surface form
- Take map of all precincts with similar voting patterns, and consider the outer contour to be the 0 level set of some 3D object
- Evolve the surface outward with forces on a triangular grid according to the level-set PDE: $$\frac{\partial \phi}{\partial t} = v|\nabla \phi|$$
- Take the collection of filled grid cells to be 2-simplices (and take grid lines to be edges, and take points to be vertices)
- The filtration parameter is the time step of the evolution

Direct Comparison – Tulare County



(a) Alpha (b) Adjacency (c) Level-set



(a) Alpha (b) Adjacency (c) Level-set

Persistent Homology for Resource Coverage: A Case Study of Access to Polling Sites

Abigail Hickok, Benjamin Jarman, Michael Johnson, Jiajie Luo, & MAP [2023],
arXiv:2206.04834 (*SIAM Review*, in press)

Naive Approach versus PH Approach

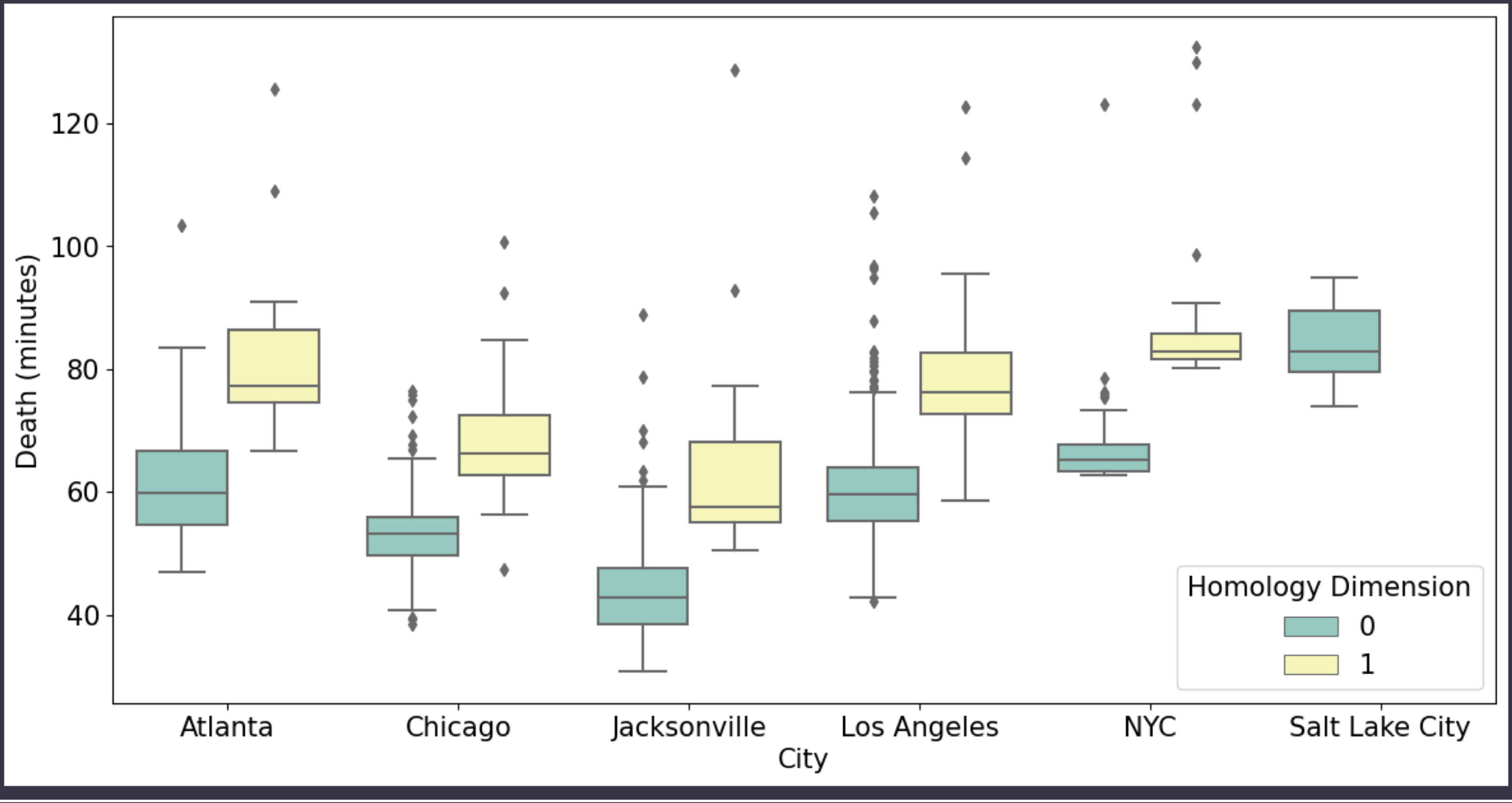
- Naive approach
  - Select some distance threshold D
  - Calculate the percent of people within distance D of a resource cite (e.g., a polling site, a DMV location, a park, etc.), or calculate which locations are within distance D of a resource cite
- Persistent homology
  - Calculate holes in coverage at all scales (not just an arbitrary threshold D)
  - Identify entire zones that are not covered (rather than pointwise locations)

Measure "Distance" in terms of Time

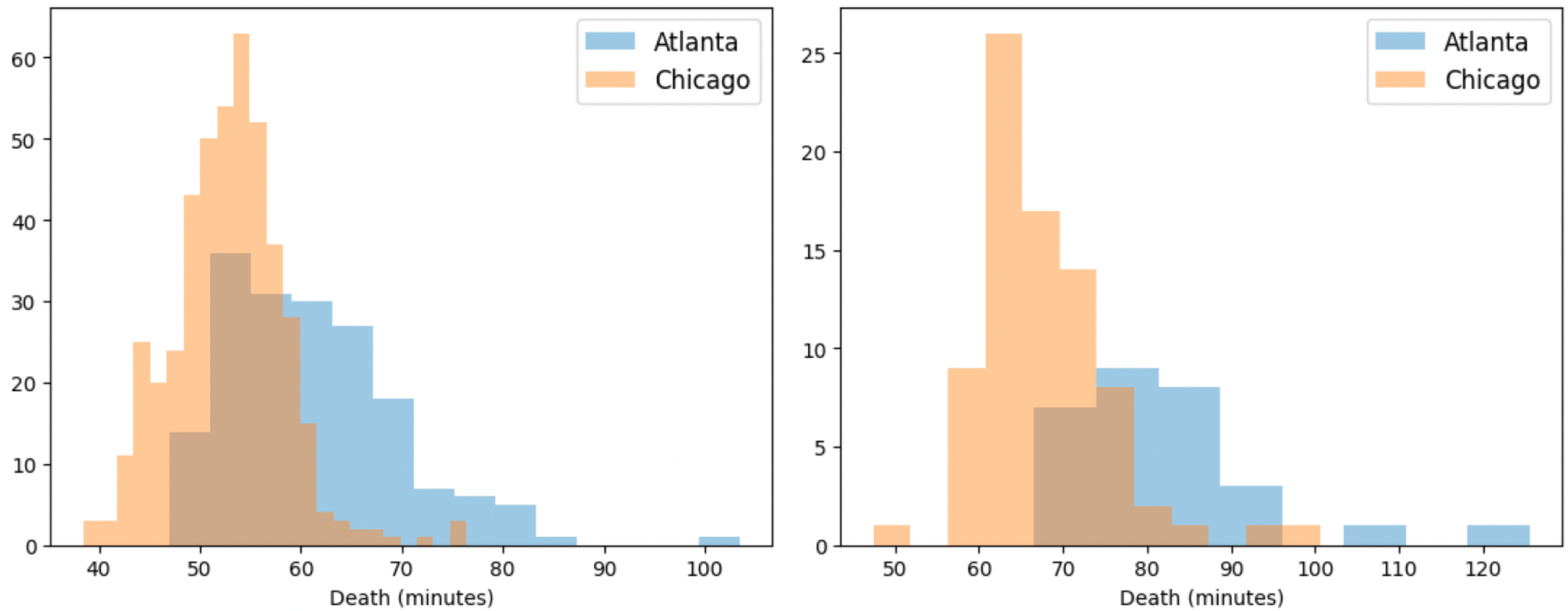
- Time is a better choice than geographical distance for this application
- Time = Travel Time + Waiting Time
  - Travel time to and from a polling site
  - Different modes of transportation, with different situations in different zip codes of a city
  - Waiting times: coverage radii around locations only start expanding from 0 after the waiting time (weighted VR filtration)
    - We used 2016 estimates at the level of Congressional districts

Death Values are the Important Feature

- A death value indicates how long it takes to vote (no matter what the birth time is)



Voting Sites in United States Cities



(a) 0D homology classes (b) 1D homology classes

- PH death values indicate when a hole in coverage closes.
  - We measure cost in time (including both travel time and waiting time)

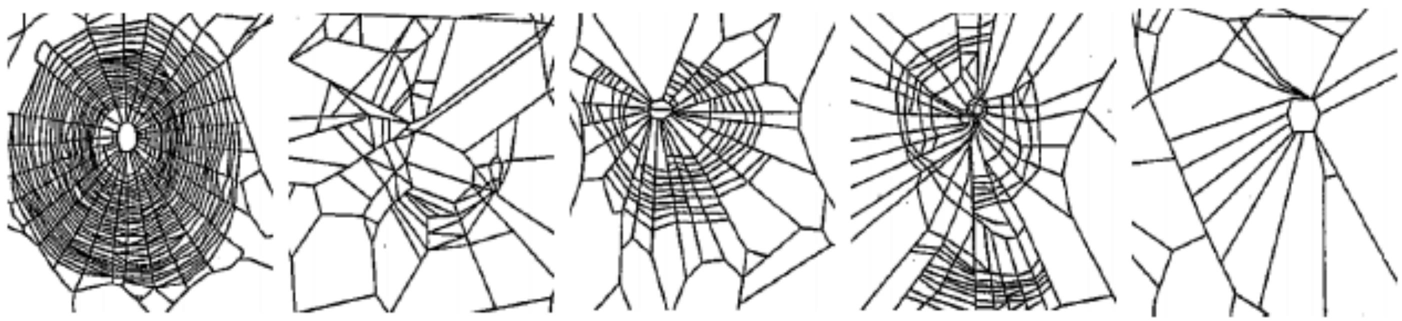
Spiders on Drugs

Michelle Feng & MAP [2020], *Physical Review Research*, Vol. 2, No. 3: 033426

[one example from this paper]

Spiders Spinning Under the Influence

- The Marshall Space Flight Center studied the webs of spiders that were exposed to various chemicals. (There is a NASA Tech Brief from 1995.)
  - Earlier work, starting in 1948 by Swiss pharmacologist Peter N. Witt
- They concluded that more toxic chemicals resulted in more deformed spiderwebs



(a) (b) (c) (d) (e)

Figure 15: Drug-free, caffeine, marijuana, speed, and chloral hydrate influenced spiderwebs

PH with Level-Set Complexes on Spiderwebs

Pink circles: H_0

Blue squares: H_1

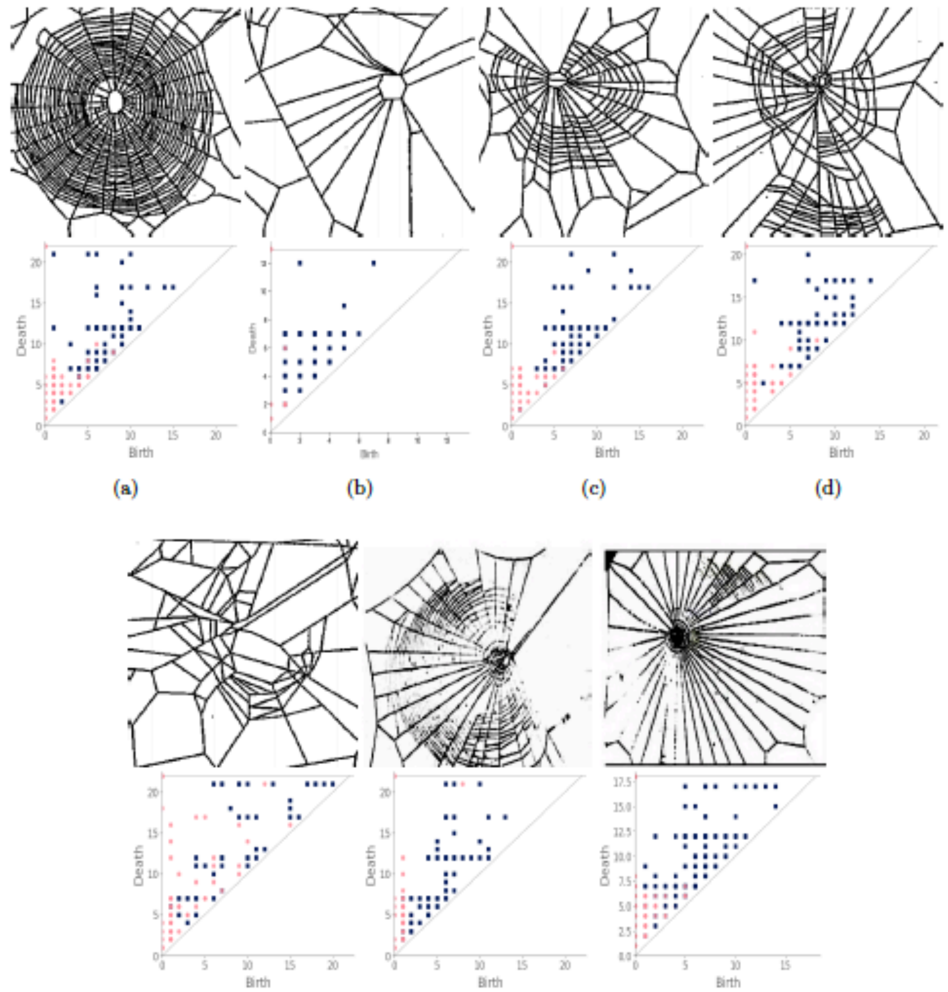


FIG. 19: Webs spun by (a) drug-free spiders compared with webs spun by spiders that were under the influence of (b) chloral hydrate (which is used in some sleeping pills), (c) marijuana, (d) speed, (e) caffeine, (f) peyote, and (g) LSD. [The images for panels (a)–(e) are from [74], and the images for panels (f) and (g) are from [73].]

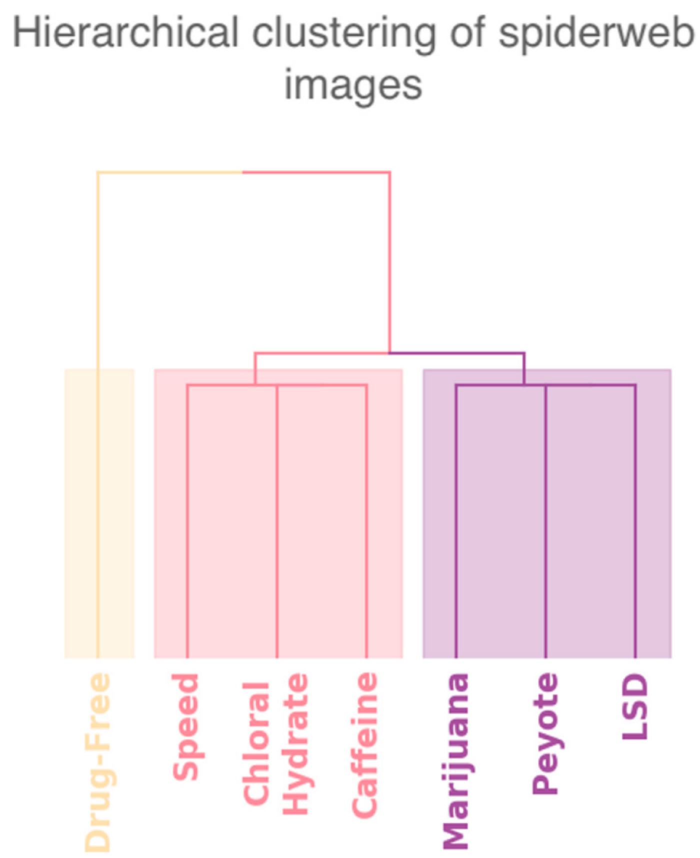


FIG. 18. Classification of webs that were spun by spiders under the influence of various psychotropic substances.

Conclusions

Conclusions

- Topological data analysis (TDA), such as by computing persistent homology (PH), can give insights into structures and dynamics in spatial complex systems.

- Persistent homology of spatial and spatiotemporal data
  - By looking at 2D data, we can do systematic comparisons between different types of constructions (topologically, fewer things can happen)
  - Incorporate information from applications of interest into PH approaches