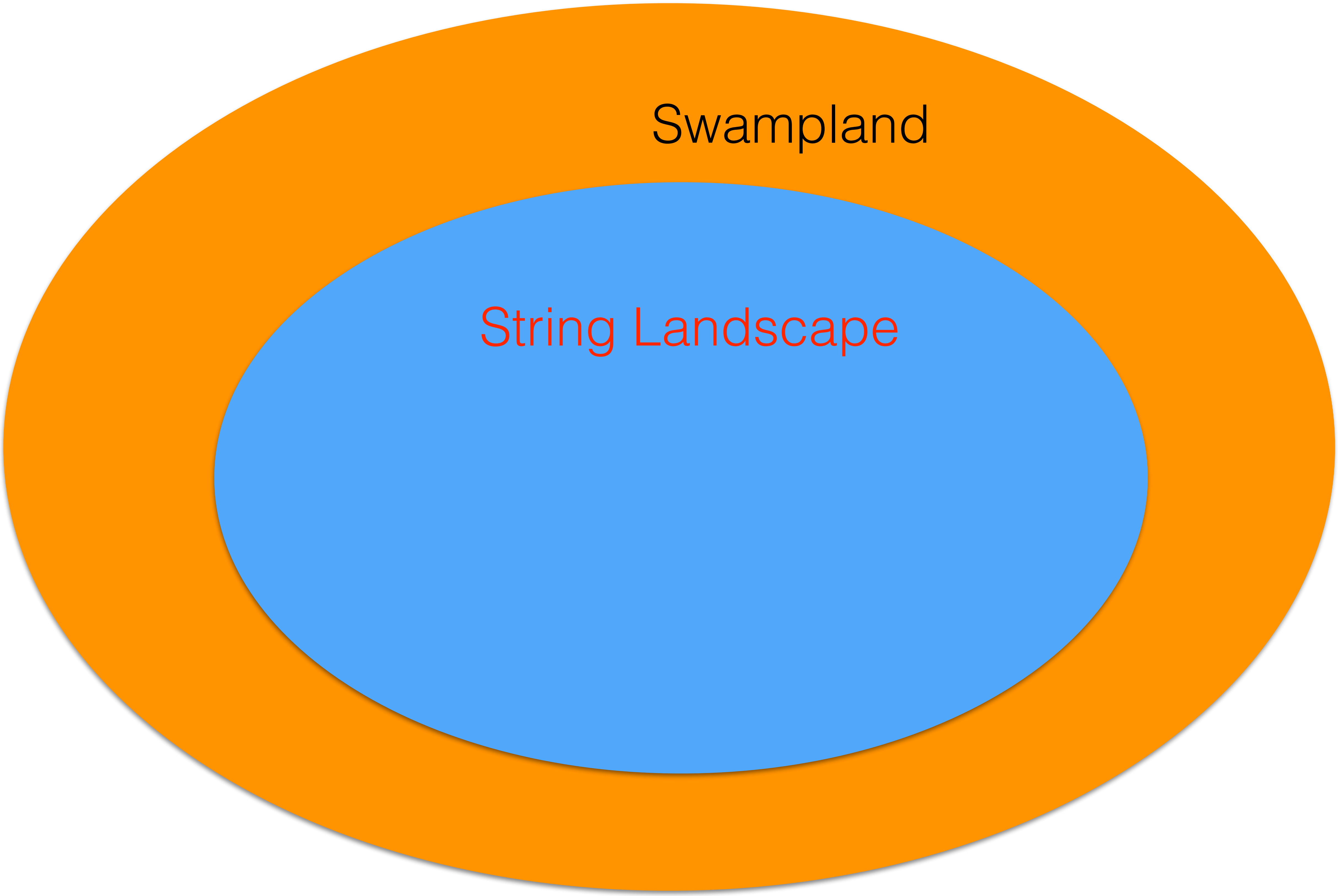
Topological Data Analysis for Cosmology & the String Landscape

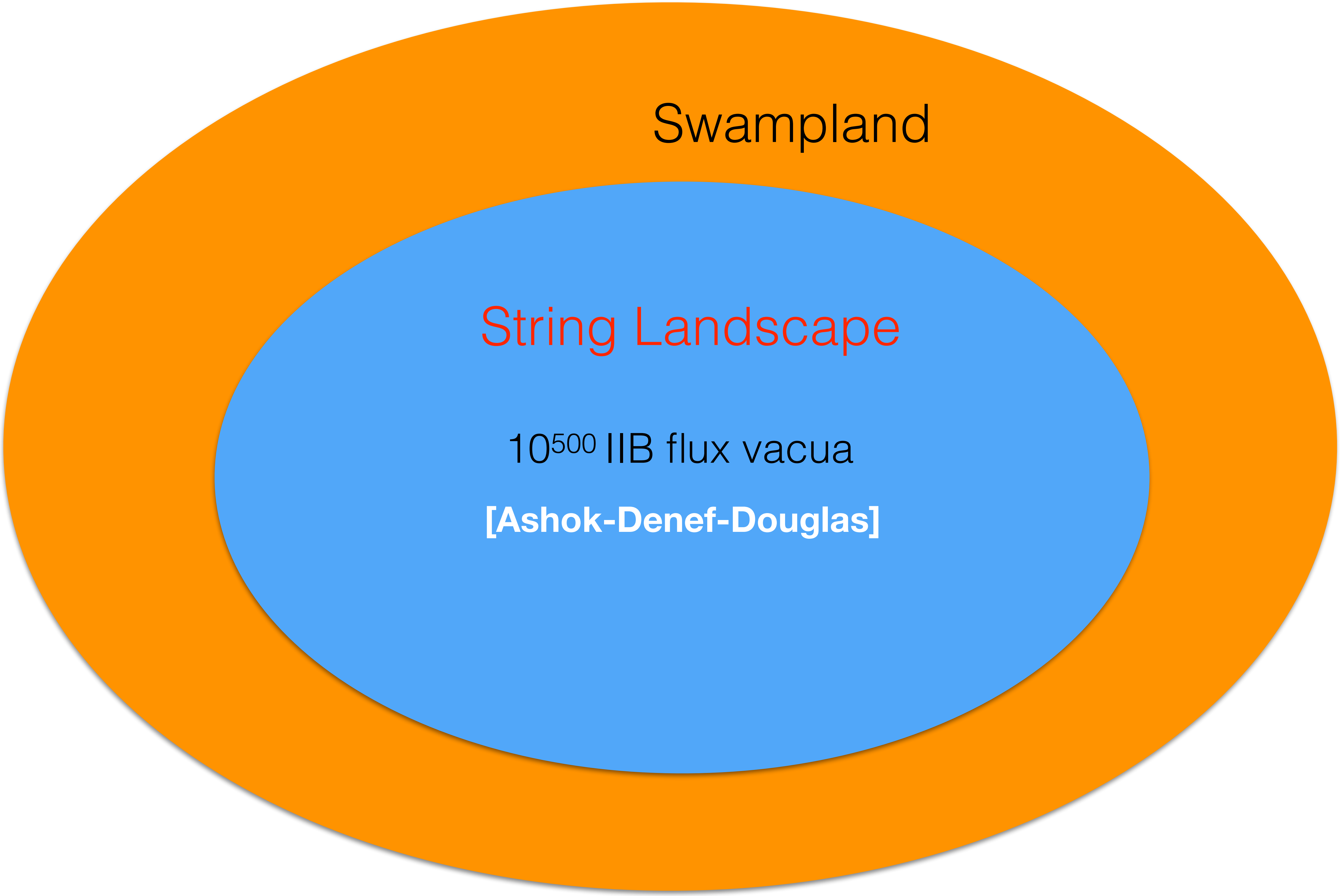
Gary Shiu
University of Wisconsin-Madison

String Data and the Swampland

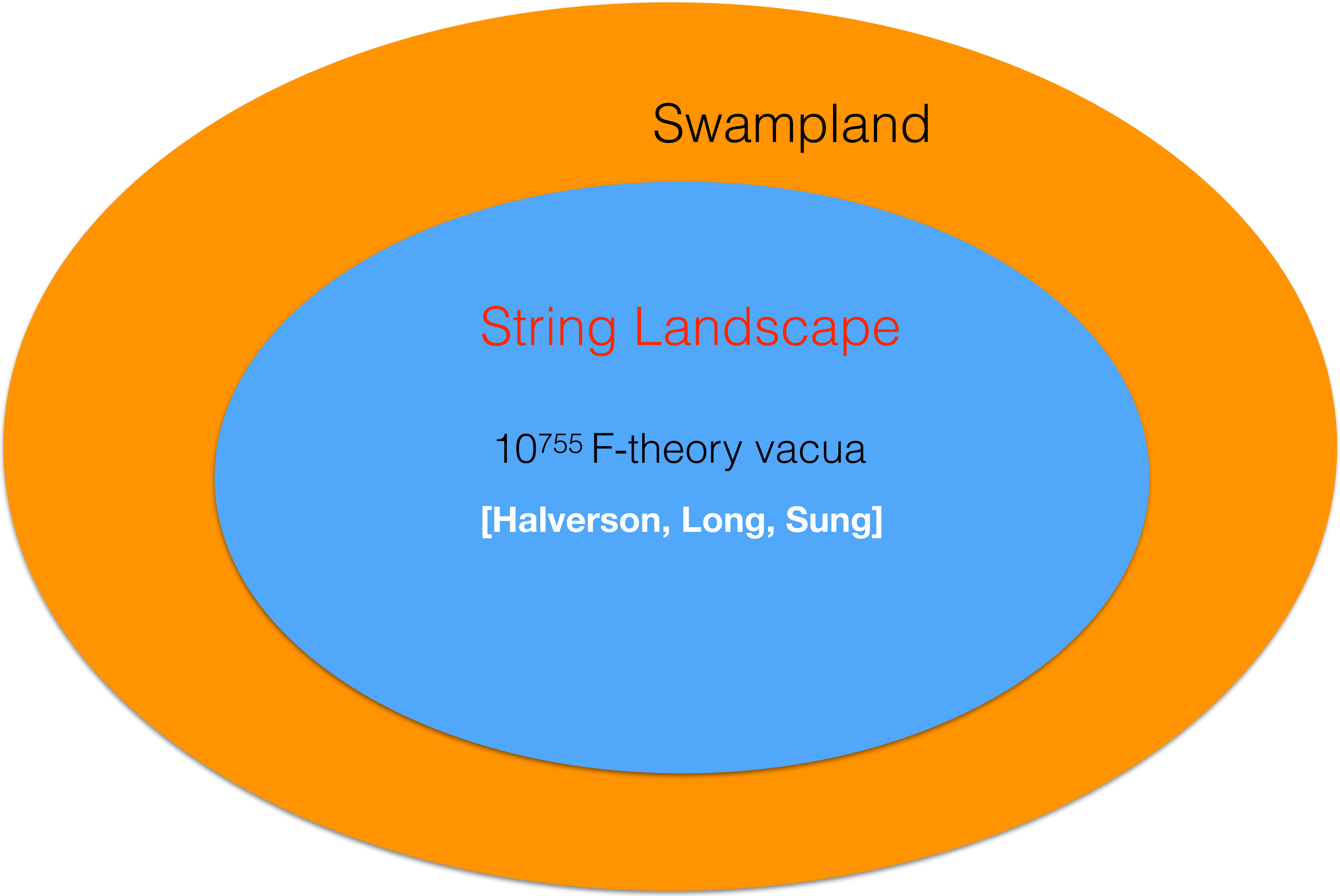
String Data and the Swampland



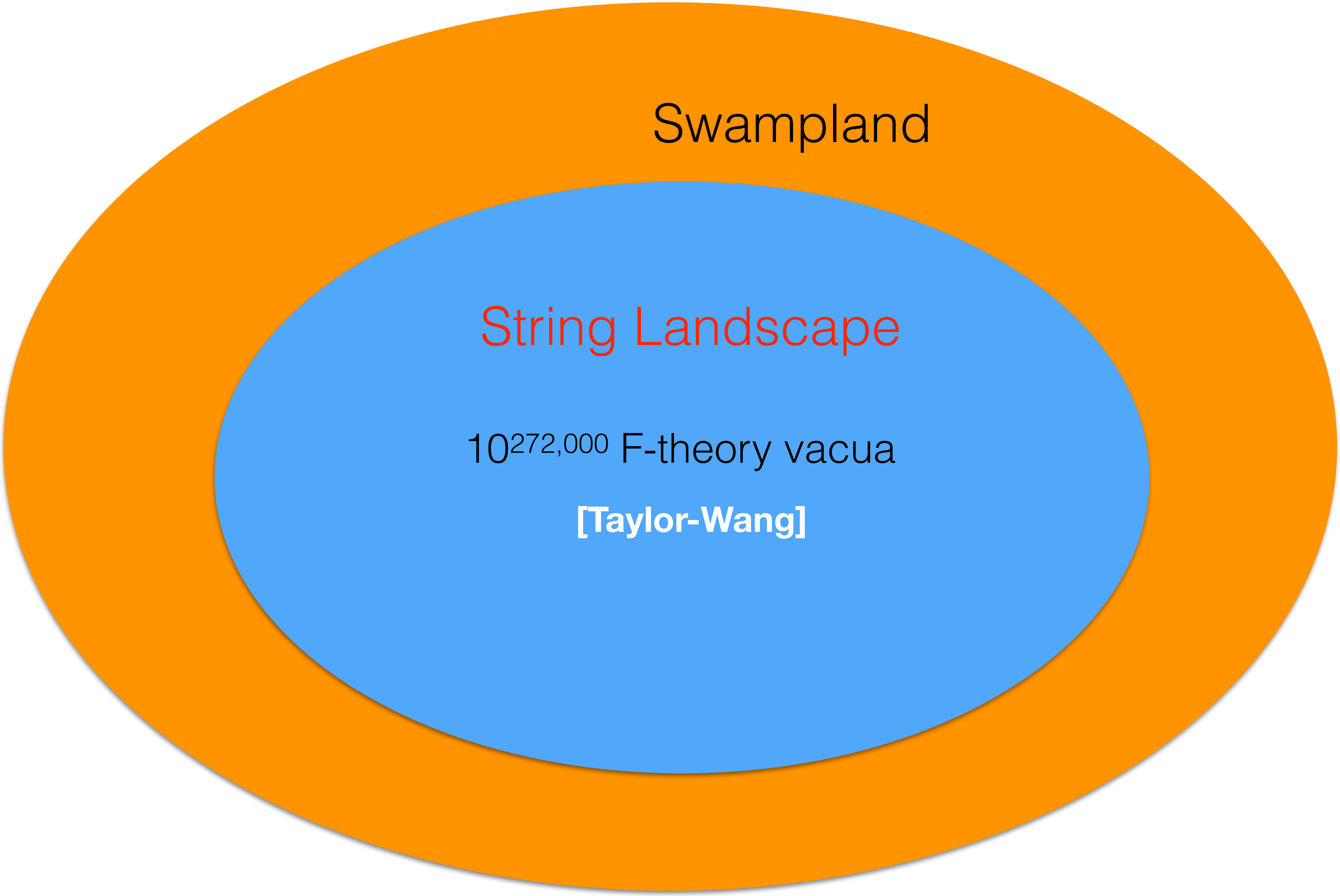
String Data and the Swampland



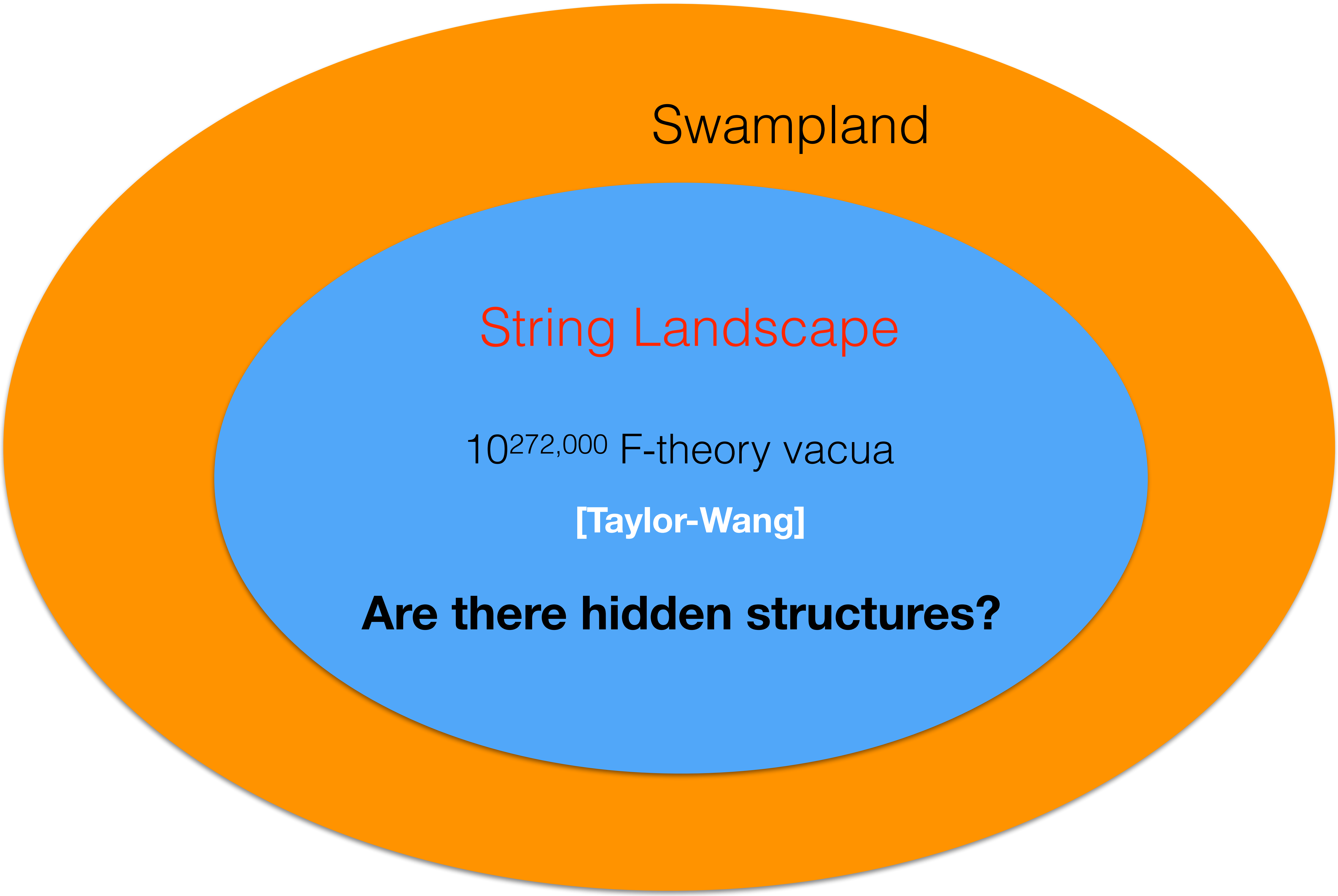
String Data and the Swampland



String Data and the Swampland

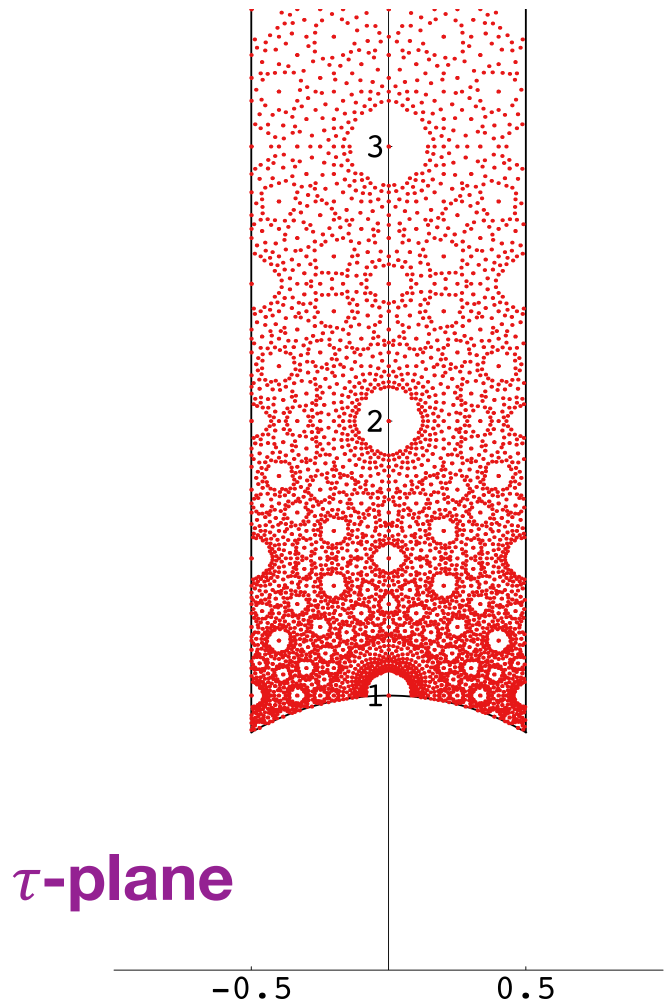


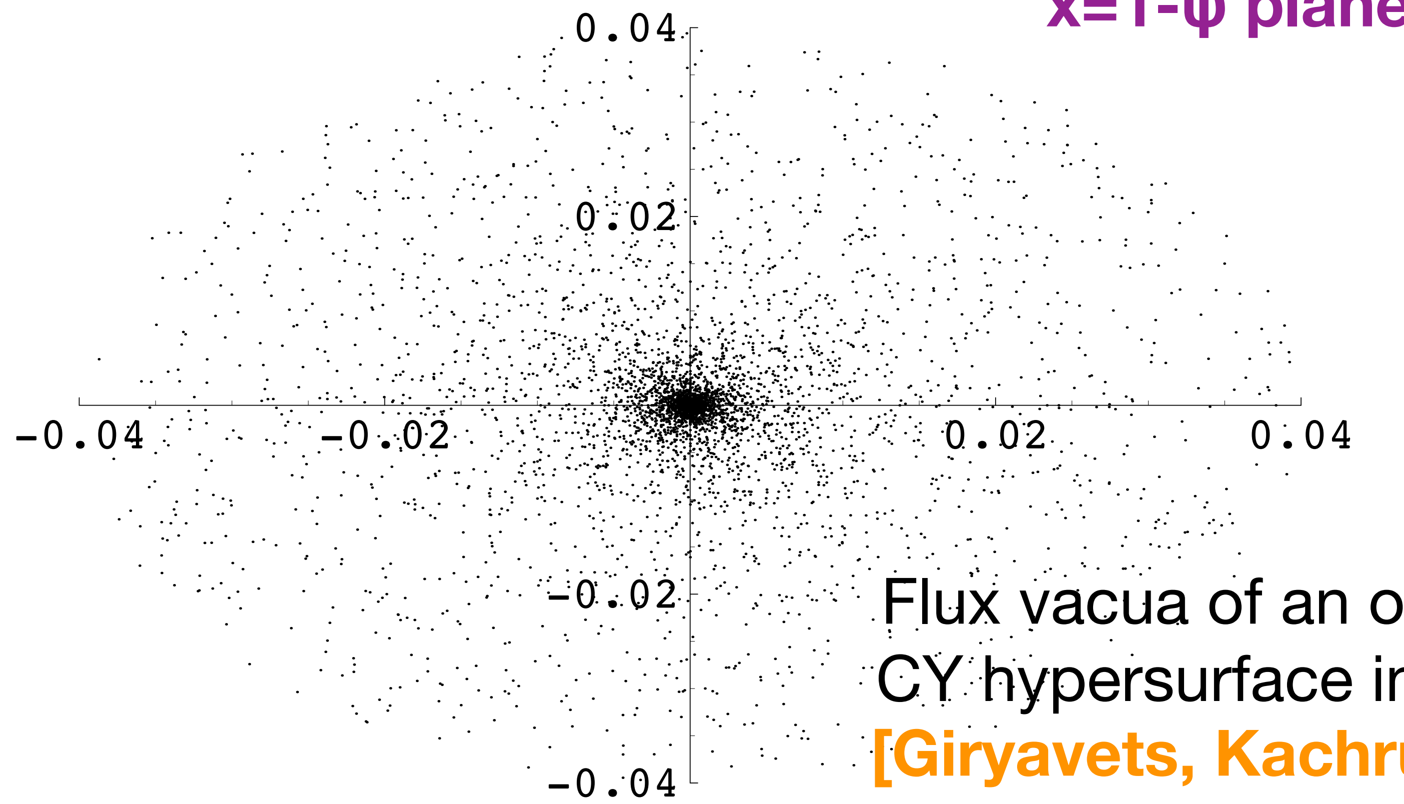
String Data and the Swampland



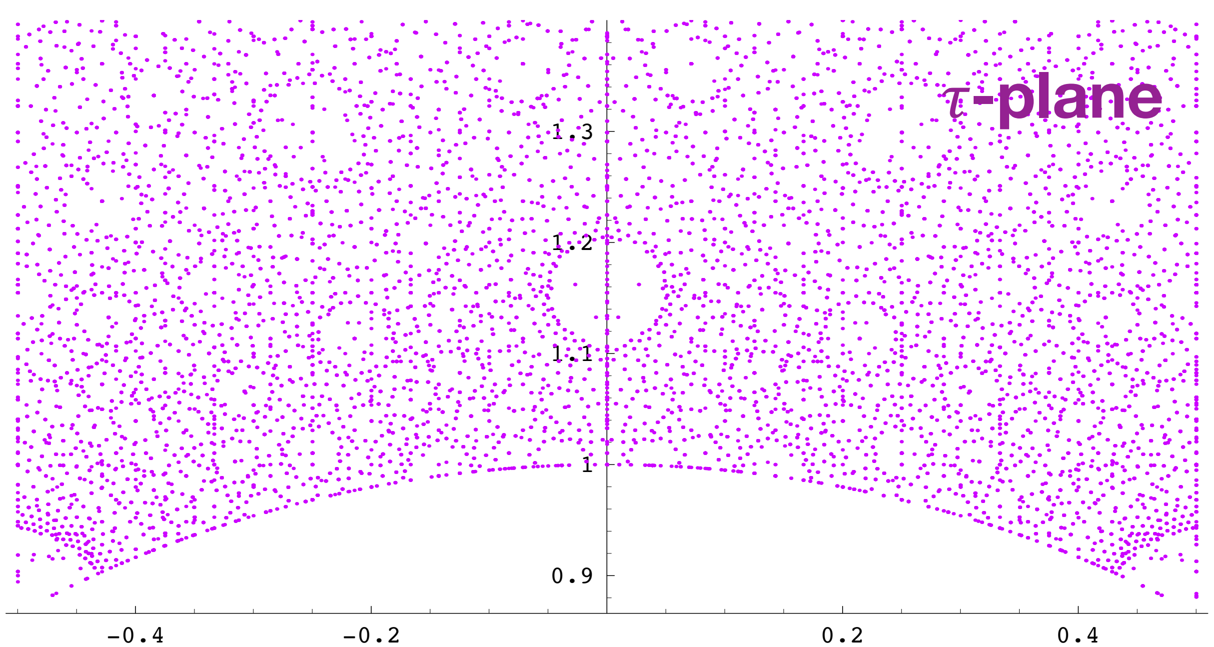
Distribution of String Vacua

Flux vacua on rigid CY
[Denef-Douglas]





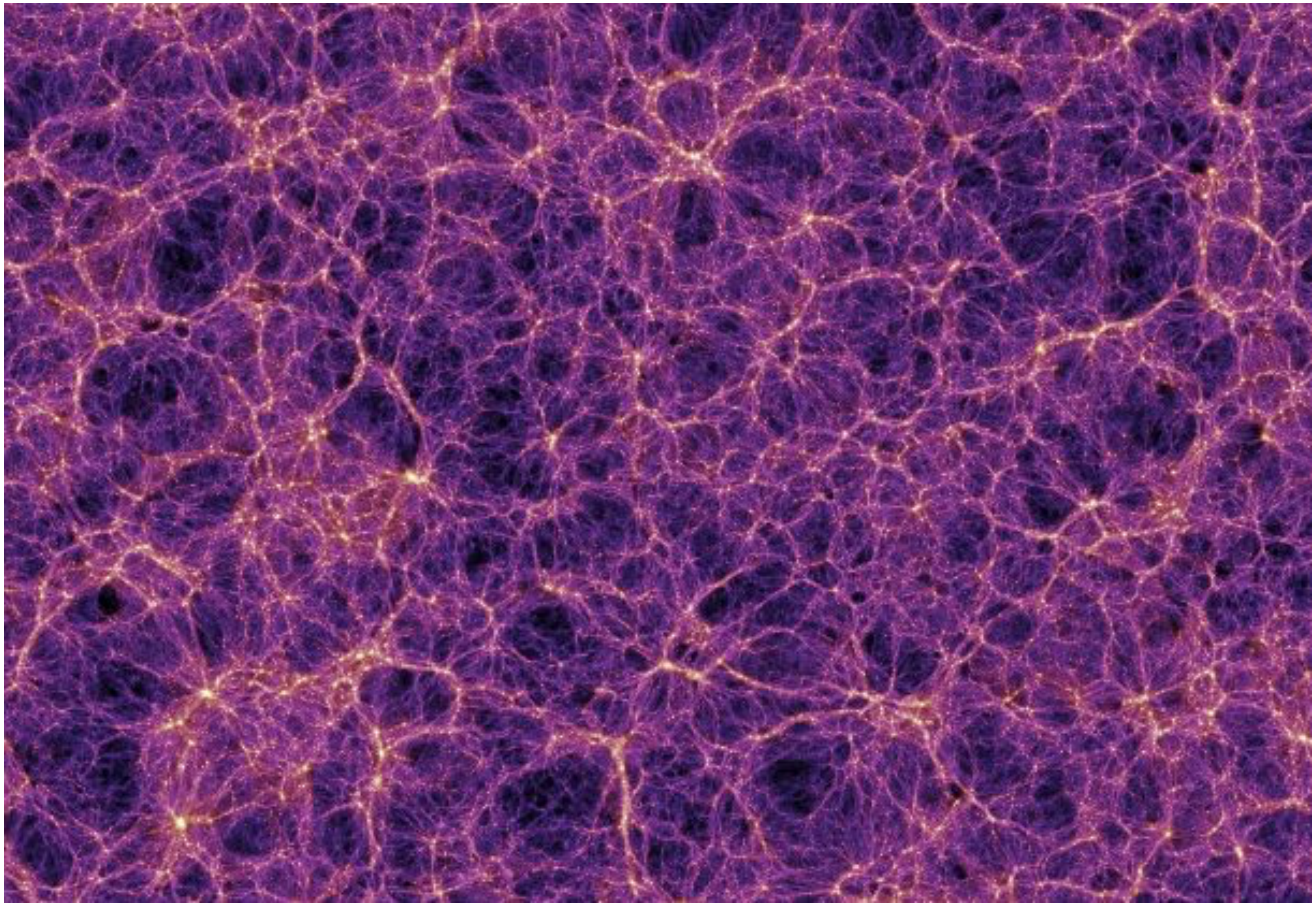
Flux vacua of an orientifold of CY hypersurface in $WP^4_{1,1,1,1,4}$
[Giryavets, Kachru, Tripathy]



Toroidal Flux vacua with W=0
[DeWolfe, Giryavets, Kachru, Taylor]

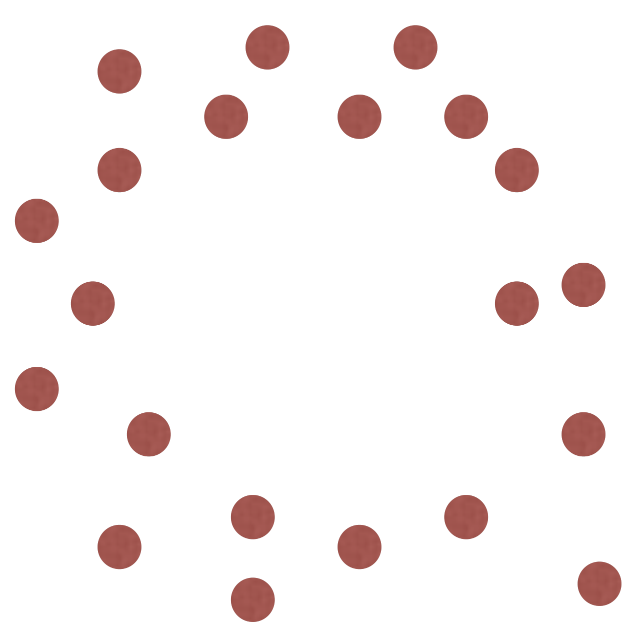
Distribution of Large Scale Structure

Similar **clustering** and **void** features also appear in LSS:



Topological Data Analysis

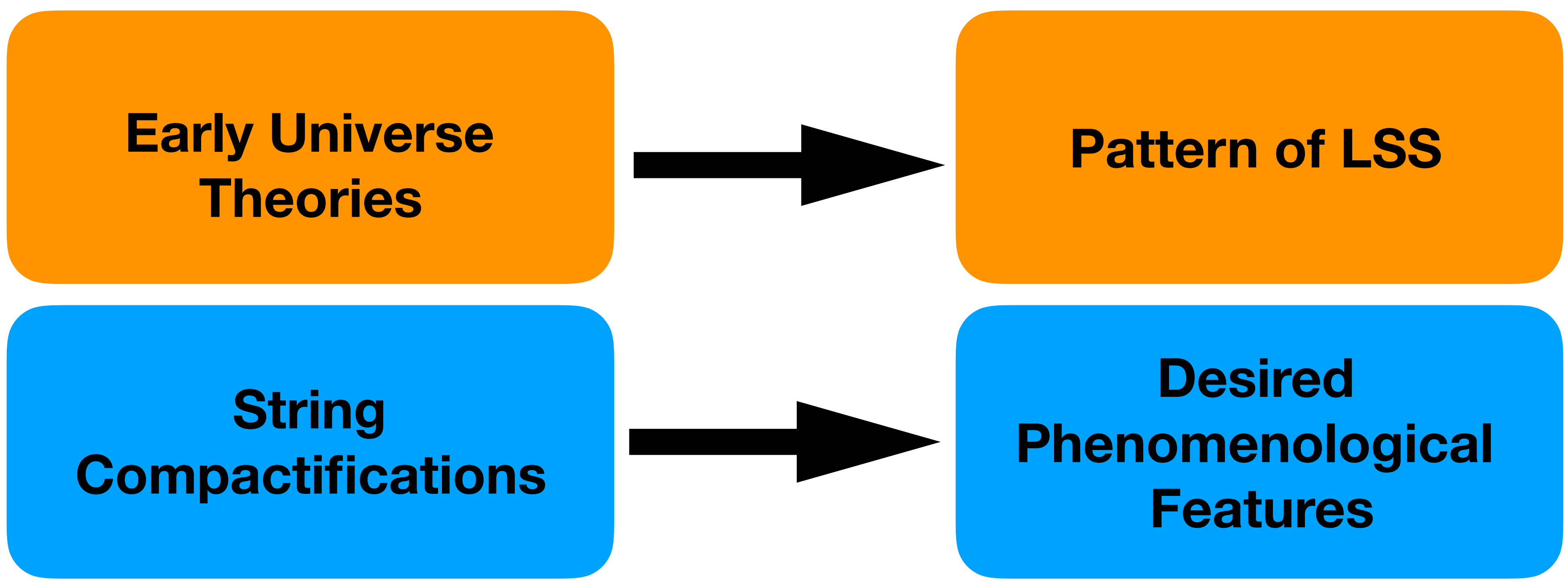
- When the space of data is huge, we cannot simply “visualize” the structure of data. We need a systematic diagnostic tool.
- Topological data analysis (TDA) is a systematic tool in applied topology to diagnose the “shape” of data.
- To turn a discrete set of data points (point cloud) into a topological space, we need a notion of ***persistence***.



Vary simplicial complexes formed by the point cloud with continuous parameters (filtration parameters)

Topological Data Analysis

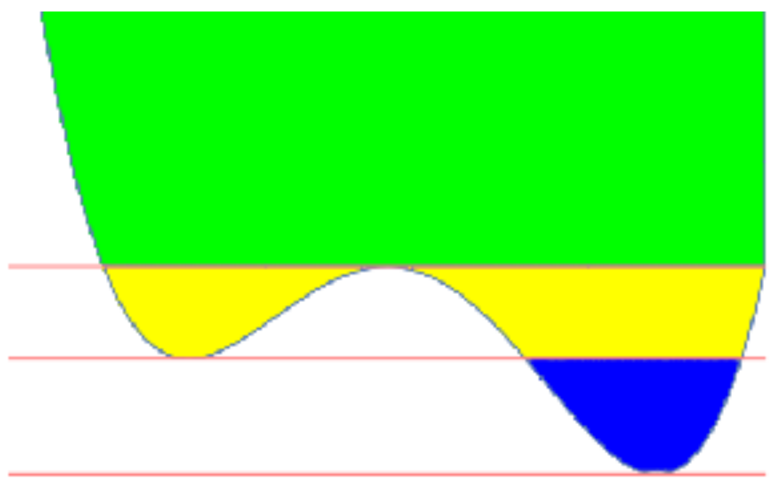
- TDA is widely used in other fields, e.g., imaging, neuroscience, and drug design. It is well suited for machine learning.
- From the persistent homology of the point cloud, we can test e.g., the effectiveness of drugs. Similarly, we can test:



- A **selector algorithm** is often used due to the huge volume of data. We can test these algorithms on cosmological datasets and the string vacua.

TDA and Energy Landscapes

- **Morse theory** connects topology of **sublevel sets** with critical points of a potential:

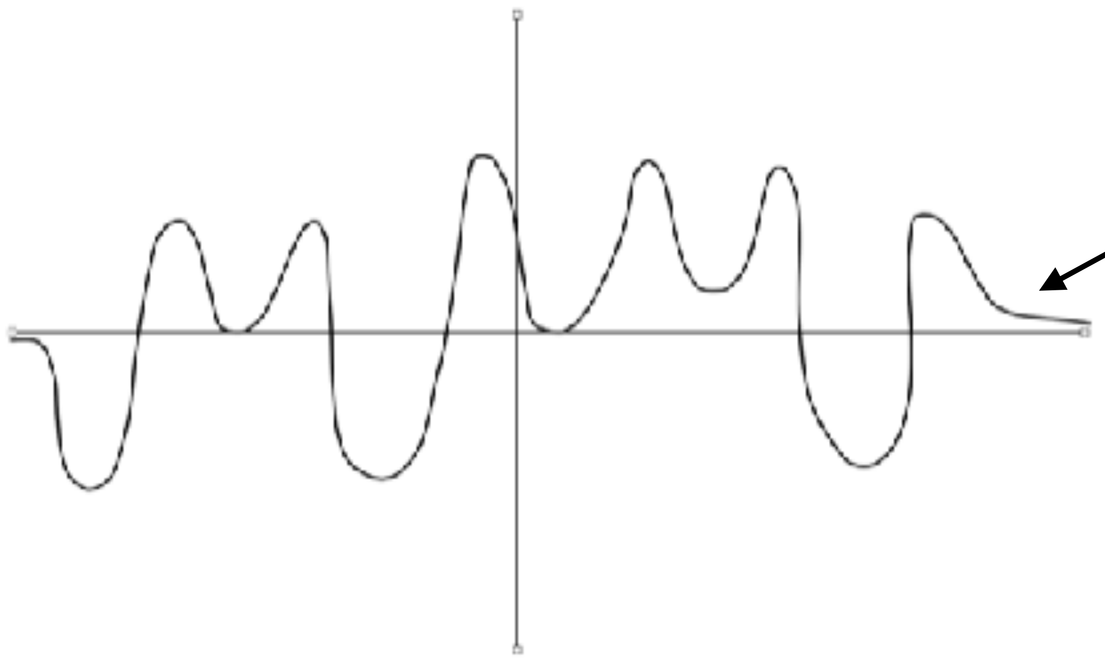


$$\textbf{Sub}_\nu = f^{-1}(-\infty, \nu]$$

- TDA can be used to find critical points of energy landscapes.
- Moreover, topology change of sublevel set of energy is often associated with a phase transition.
- Persistent homology can recover known associated topology changes. **[Donato, Gori, Pettini, Petri, De Nigris, Franzosi, Vaccarino]**

TDA and the String Landscape

- **TDA as a tool for the landscape:**



Proposed universal behavior for **V > 0** and at **large distances** in field space **[Ooguri, Palti, GS, Vafa]:**

$$V(\phi) \sim e^{-a\phi}$$

- While skepticisms have been raised regarding the existence of ***explicit, controlled dS vacua*** **[Obied, Ooguri, Spodyneiko, Vafa];[Ooguri, Palti, GS, Vafa] (see talks of Andriot and van der Schaar)**, there seem to be many AdS and Minkowski vacua.
- TDA can be used to find local energy minima in **protein conformation space** **[Haspel, Luo, González].**
- Morse theory tells us that the appearance of clusters in sublevel sets is related to local minima of the energy landscape.

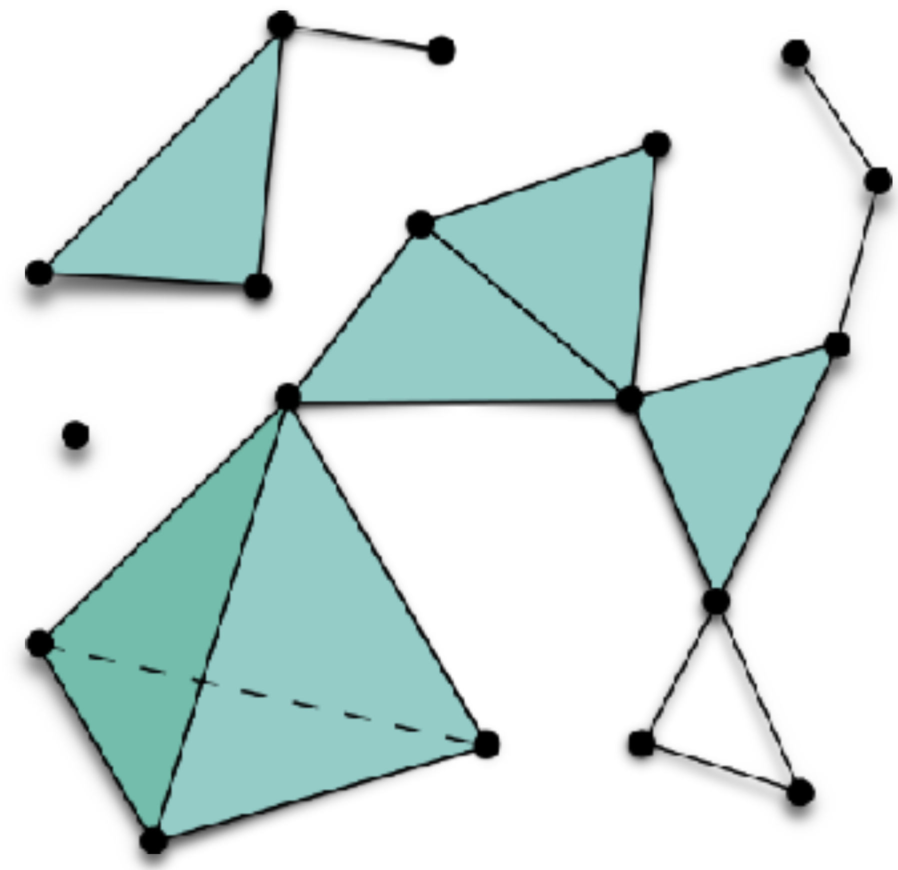
Plan of this talk

- Introduce the basic concepts of **topological data analysis**: persistent homology, barcodes, and persistence diagrams.
- Applying TDA to constrain **primordial non-Gaussianities**.
- Back to **String Data**. Computing the persistent homology of string vacua to analyze their structure.
- This talk is based on several projects done in collaboration with

- "Persistent Homology and Non-Gaussianity", A. Cole, GS, JCAP **1803**, no. 03, 025 (2018) [arXiv: 1712.08159 [astro-ph.CO]].
- "Topological Data Analysis for the String Landscape", A. Cole, GS, arXiv: 1812.06960 [hep-th].
- ...

Simplicial Complexes

- In $\mathbb{R}^3$, simplices are vertices, edges, triangles, and tetrahedra
- Simplicial complexes are collections of simplices that are:
  - Closed under intersection of simplices
  - Closed under taking faces of simplices
- Combinatorial representations — easy calculations for computers



Source: Wikipedia, "Simplicial Complex"

Simplicial Homology

- Given a simplicial complex, define a boundary operator ∂_p that maps p-simplices to (p-1)-simplices
    - We want to count independent p-cycles (i.e. p-loops) that are not boundaries of higher-dimensional objects
- Group theoretic: $Z_p = \ker \partial_p$, $B_p = \mathrm{im}\ \partial_{p+1}$,

$$H_p \equiv Z_p / B_p$$

- Betti numbers: $\beta_p \equiv \mathrm{rank} H_p$
    - 0-th Betti number is number of connected components
    - p-th Betti number is number of independent p-loops
- In practice, homology calculation is a matrix reduction

$\beta_0 = 1$

$\beta_1 = 1$

vs.

$\beta_0 = 1$

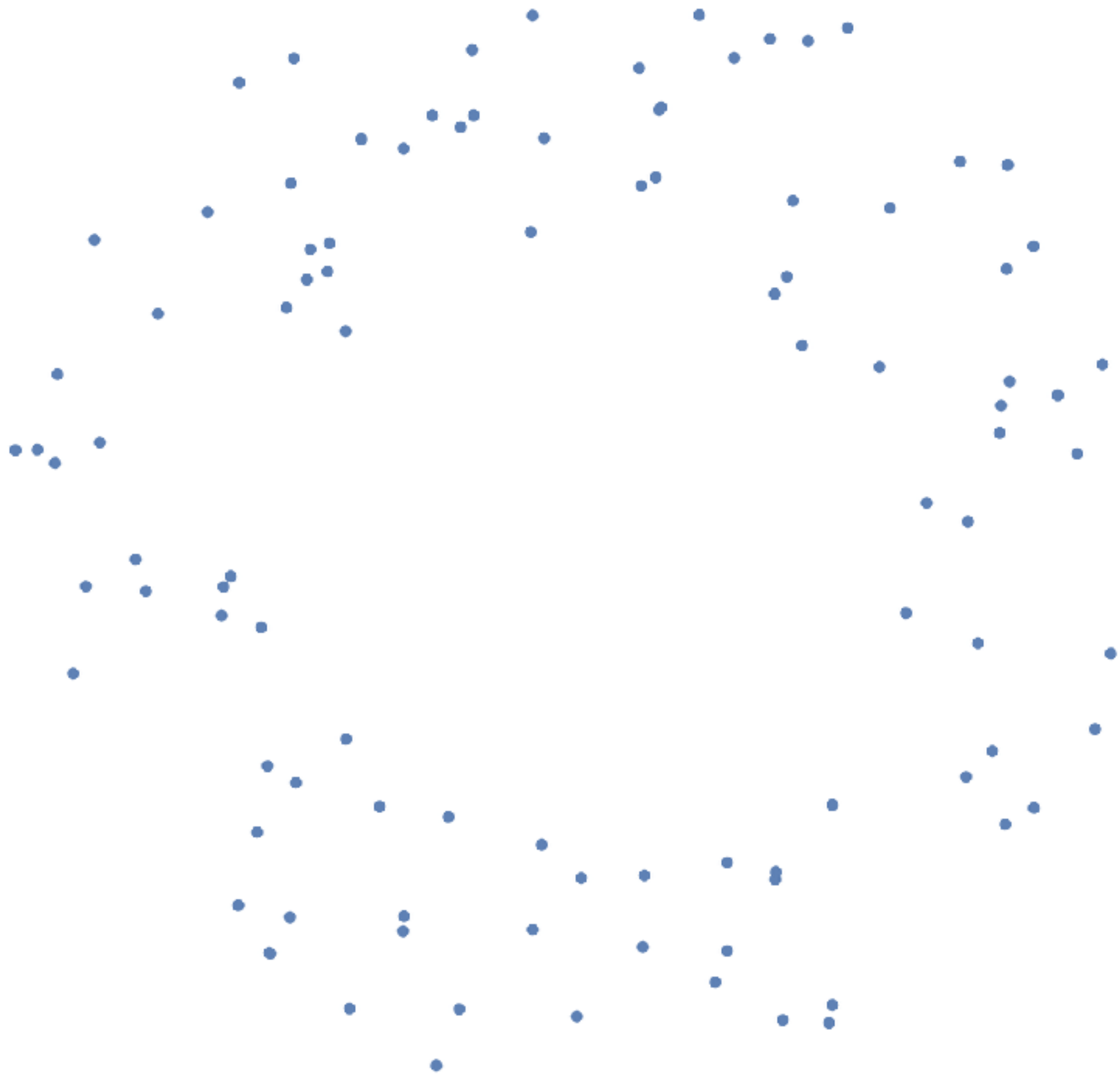
$\beta_1 = 0$

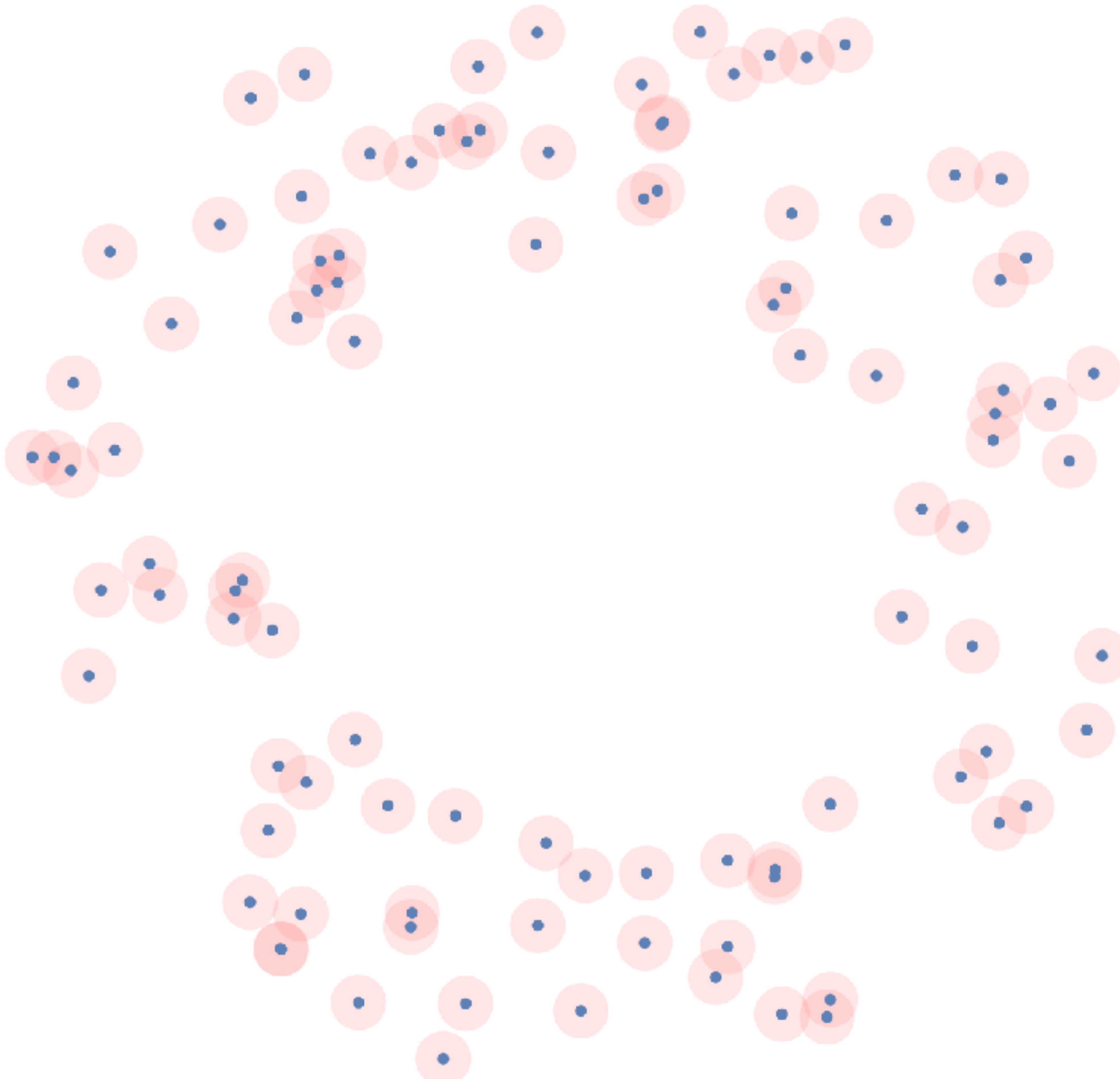
Persistence

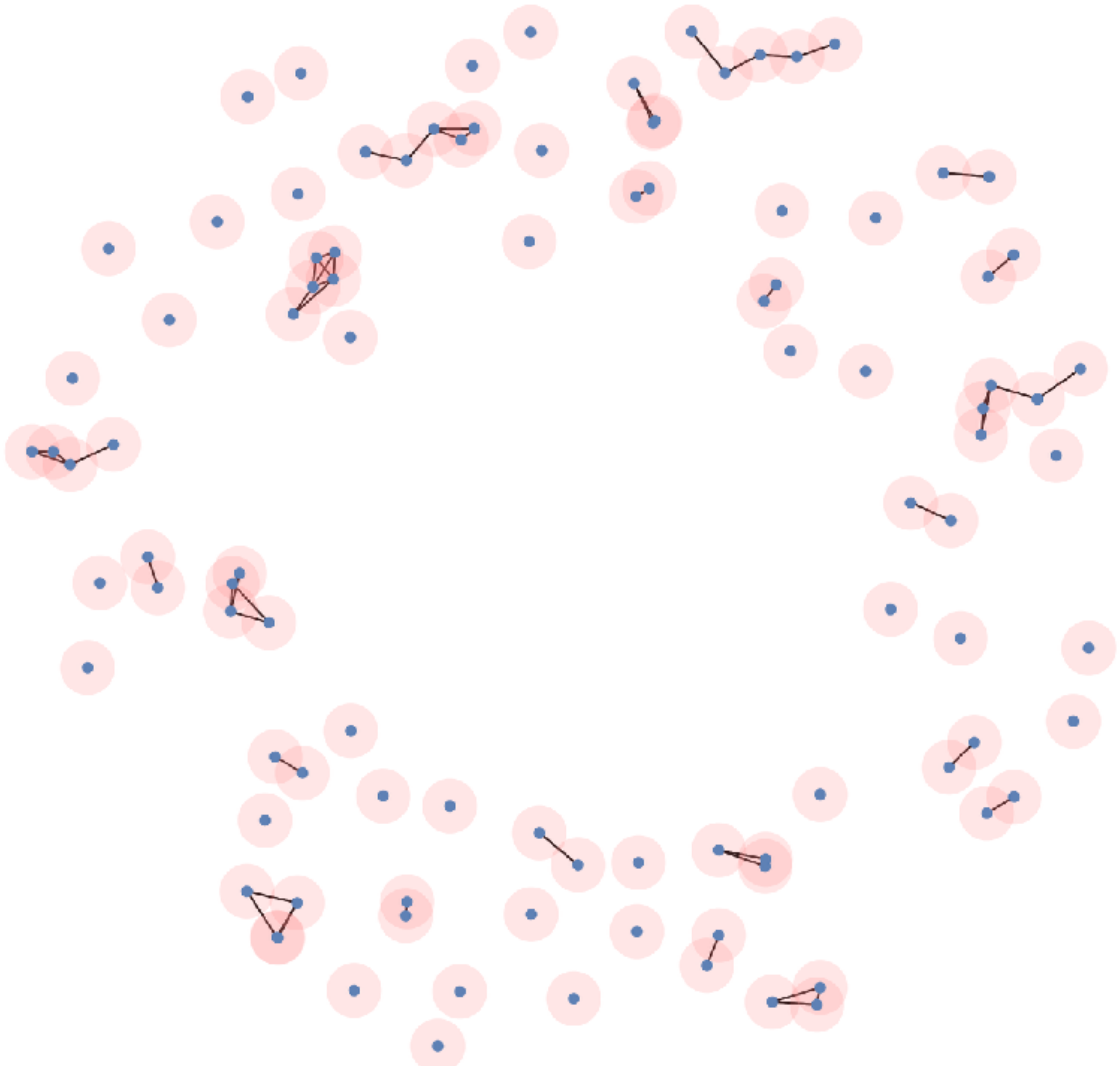
- How to choose simplicial representation of our data?
- *Persistent* homology: vary simplicial representation Σ_ν of data with some *filtration parameter* ν such that

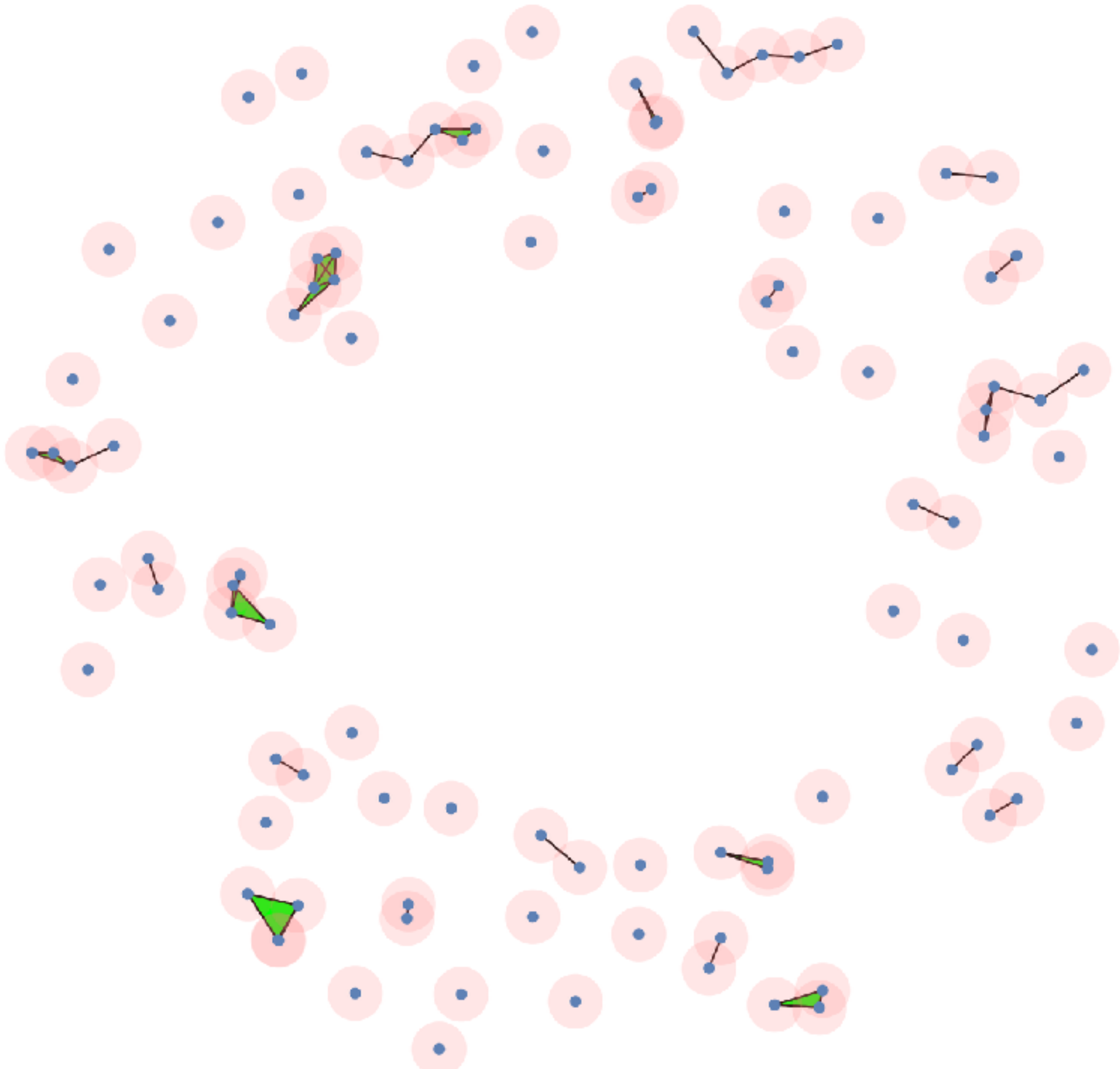
$$\nu_1 \leq \nu_2 \implies \Sigma_{\nu_1} \subseteq \Sigma_{\nu_2}$$

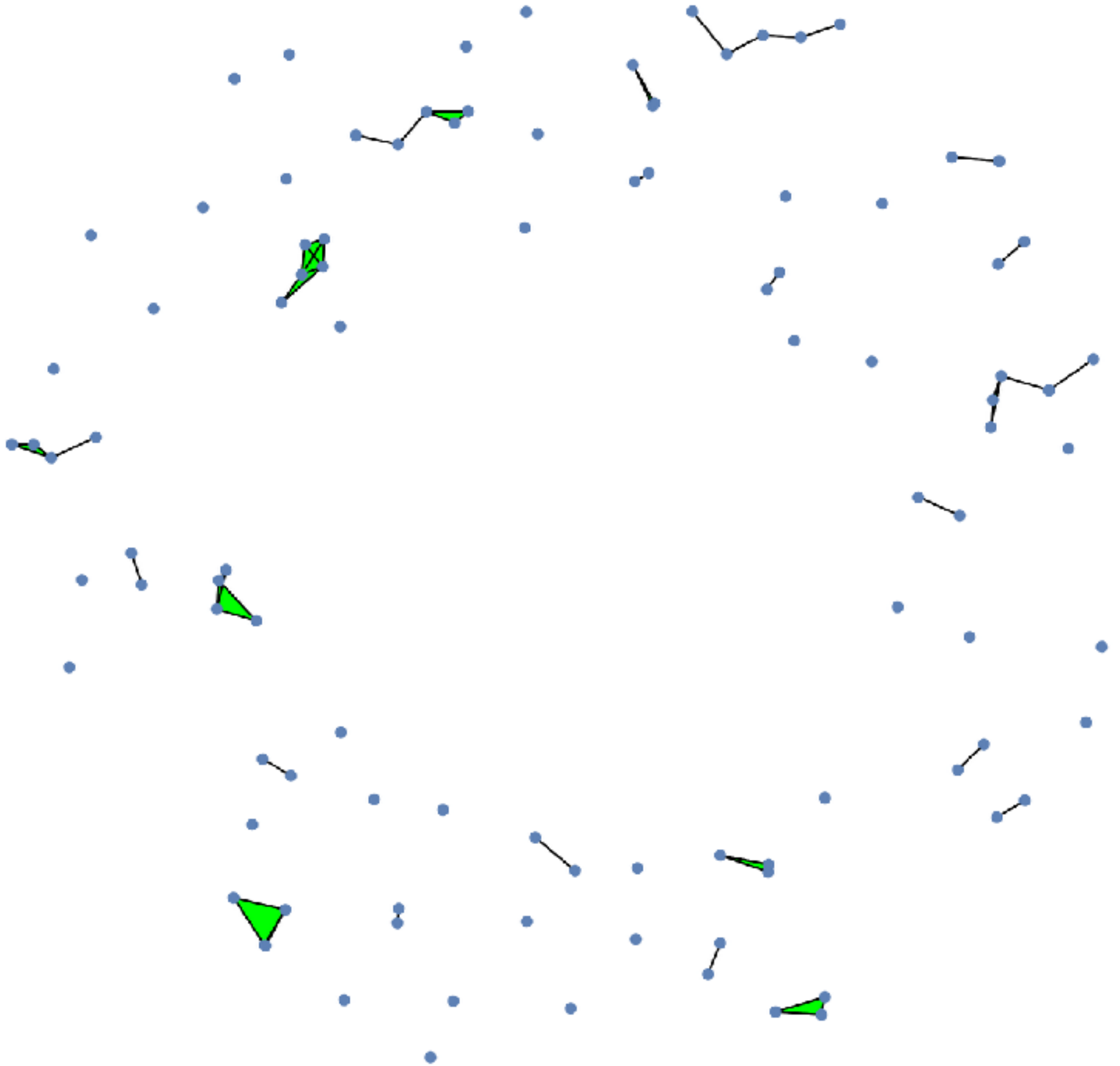
  - Track each distinct feature's lifetime (birth and death)
- Intuition: "real" topological features *persist*, short-lived features are noise
- Procedure is stable against perturbations to data **[Cohen-Steiner 2005]**

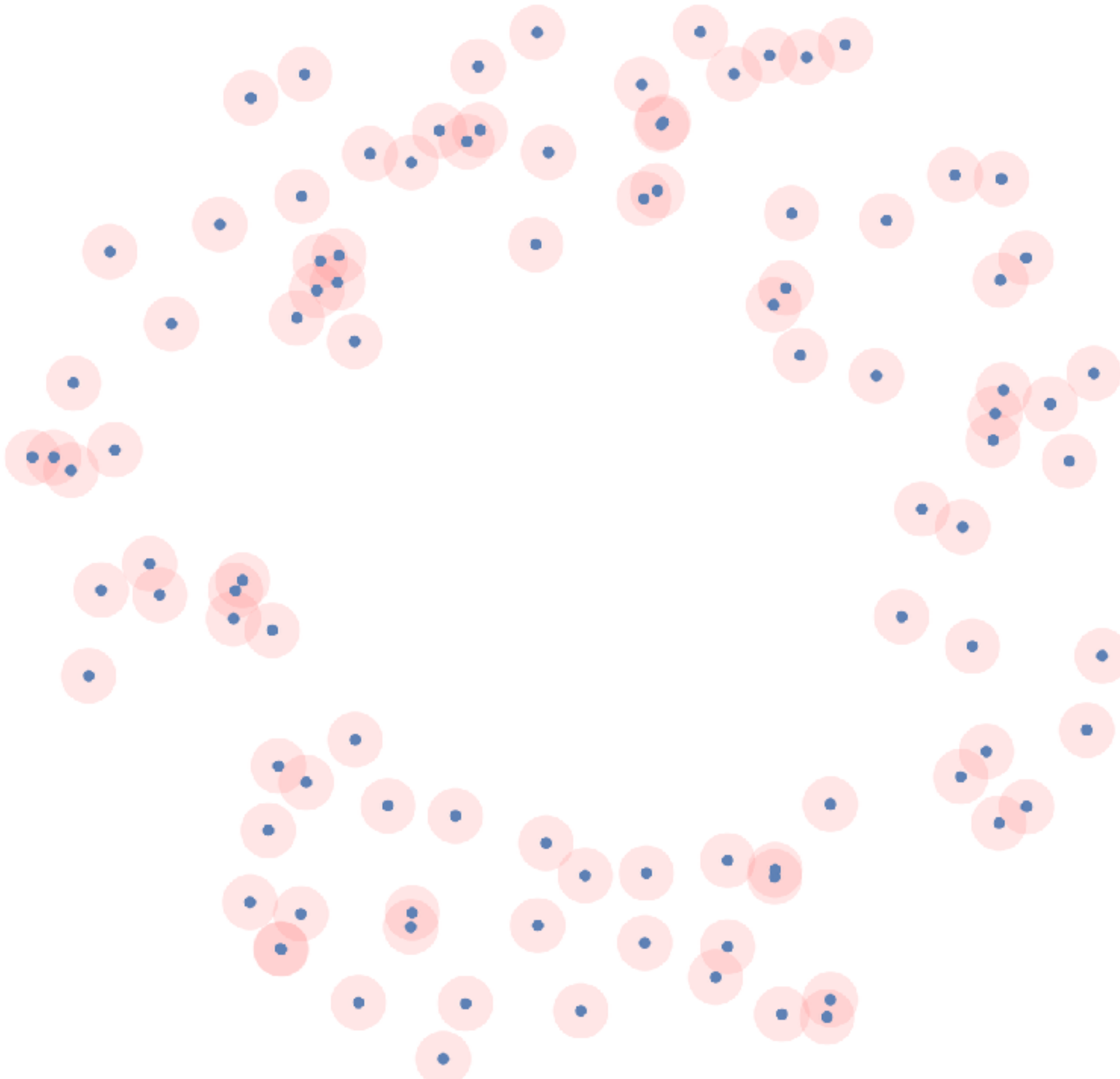


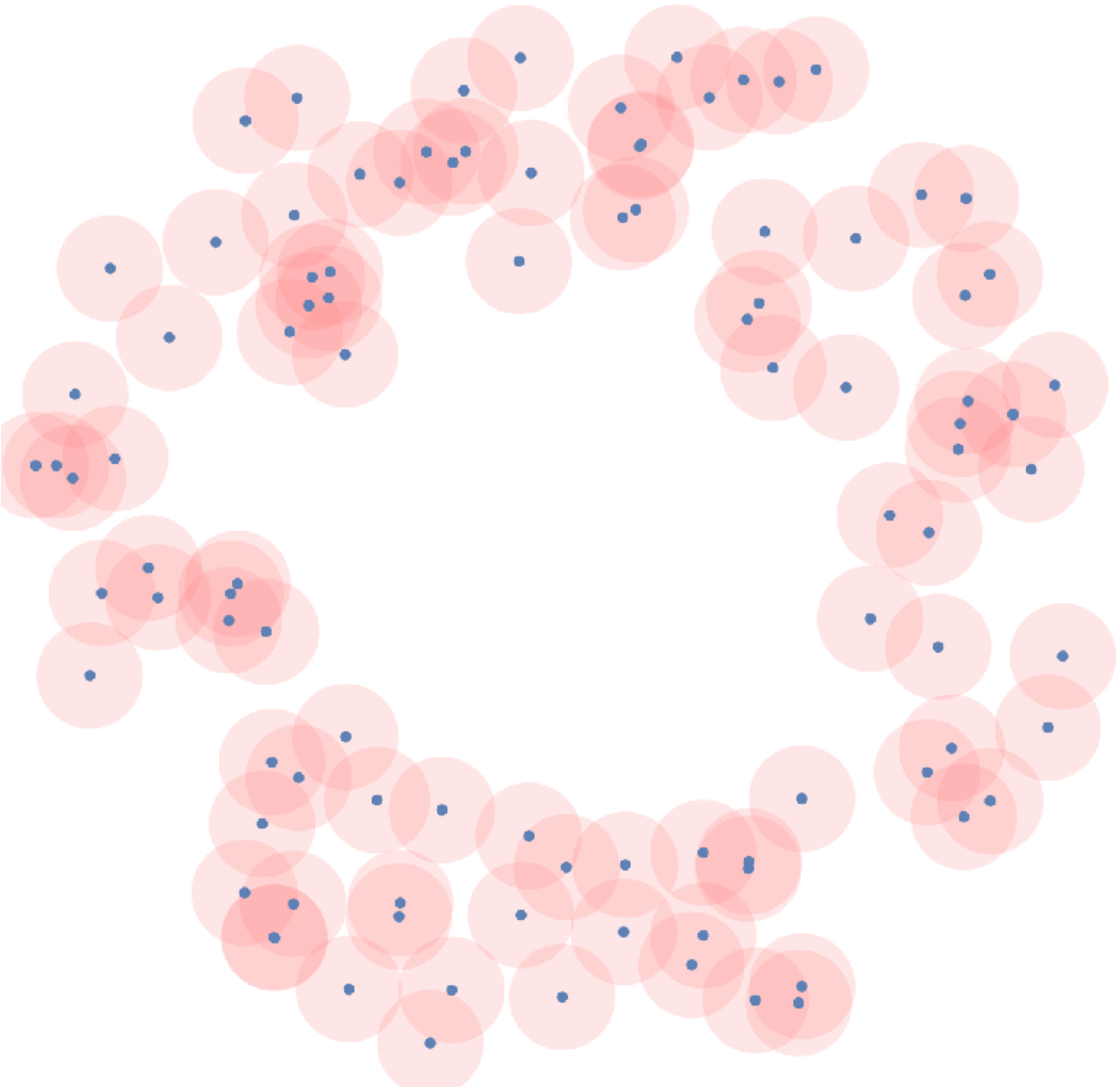


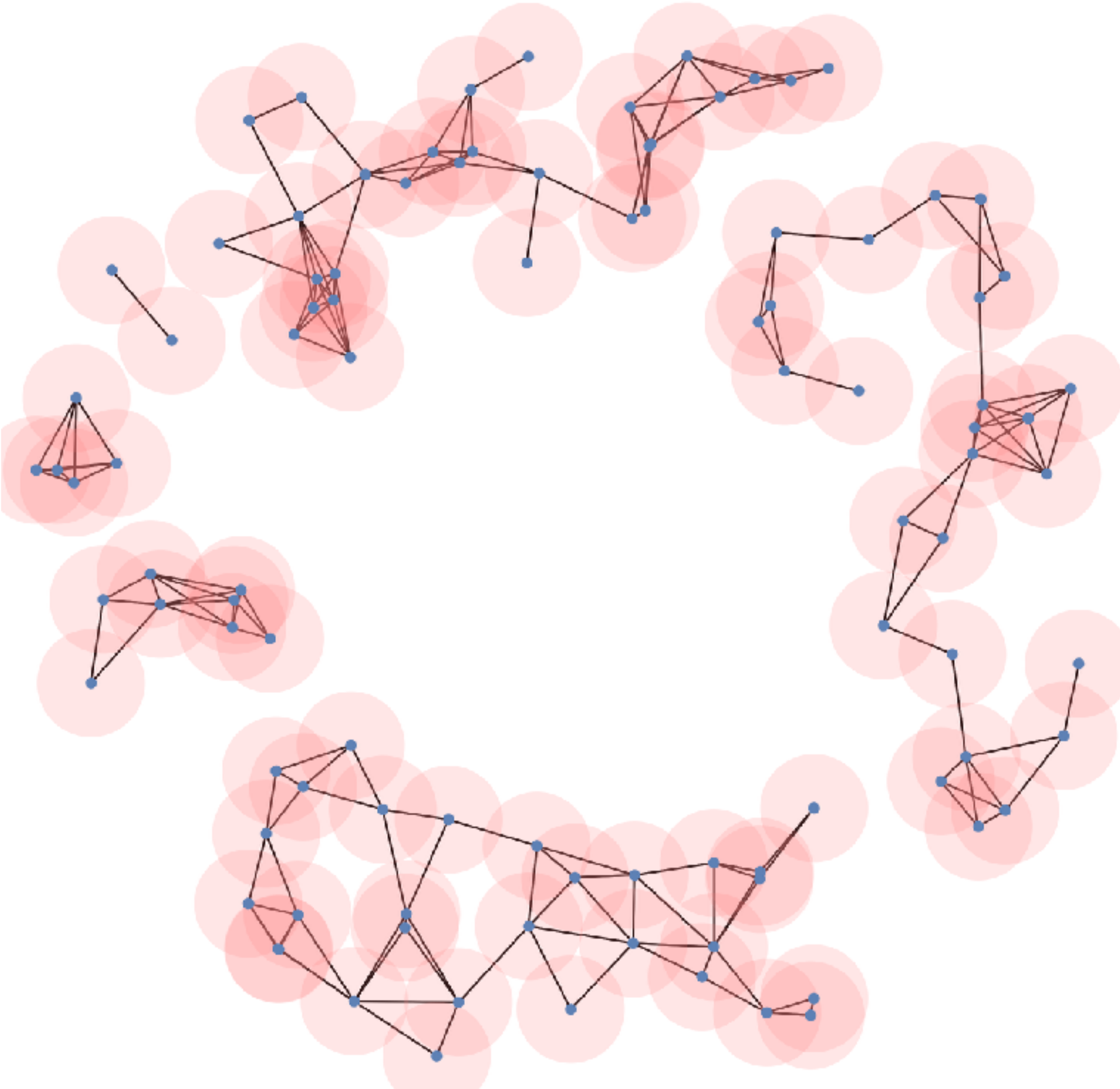


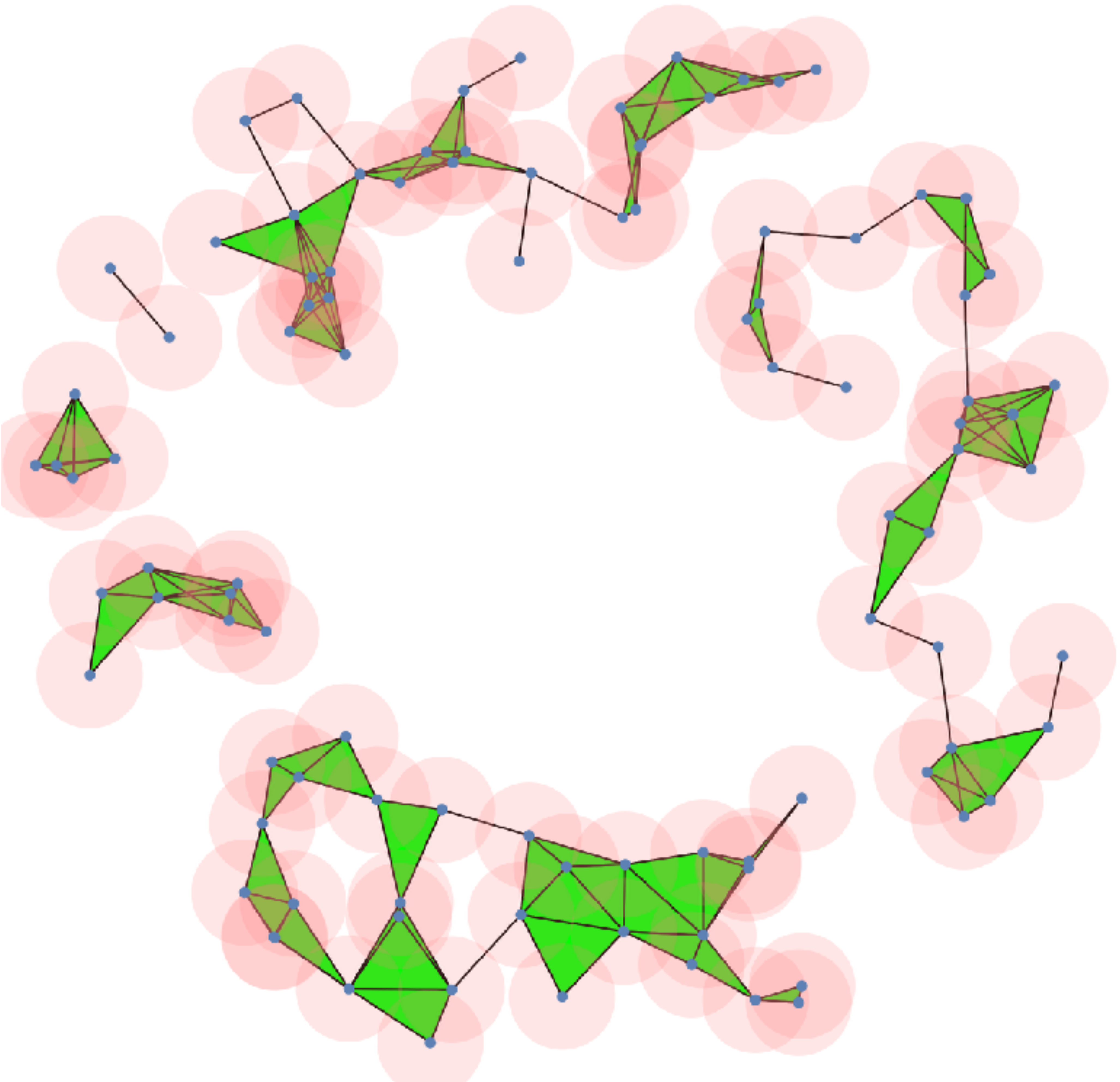


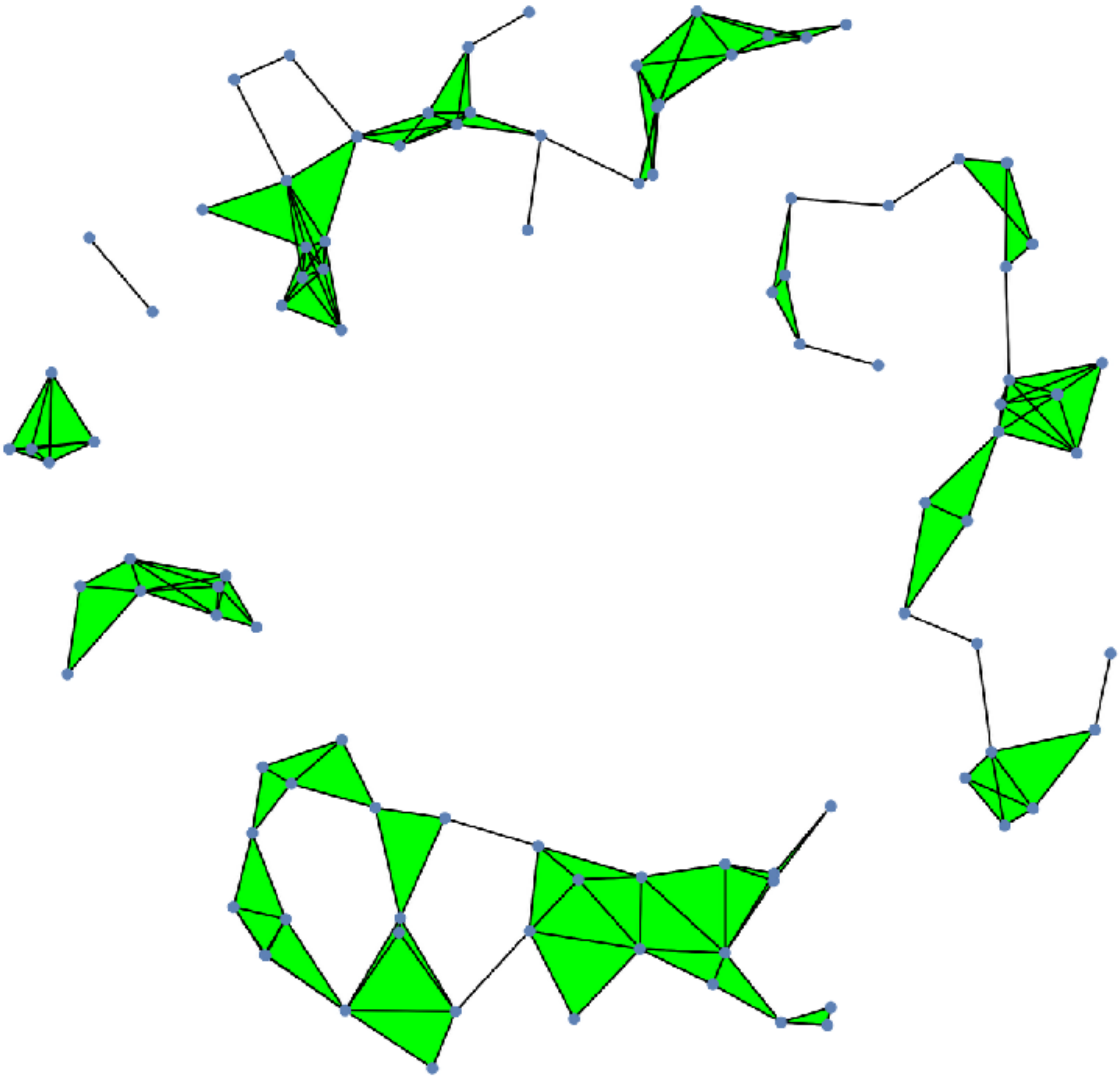


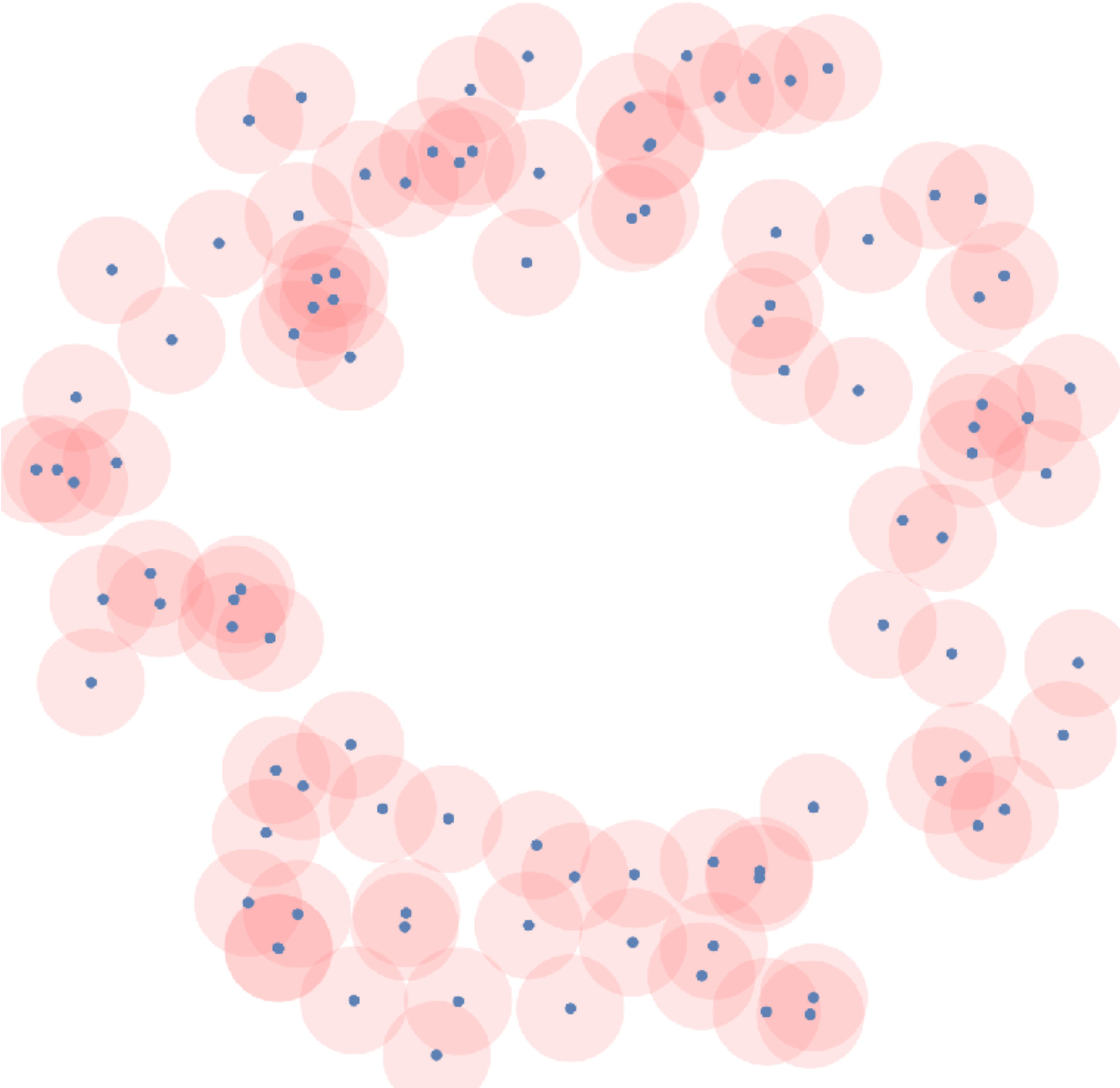


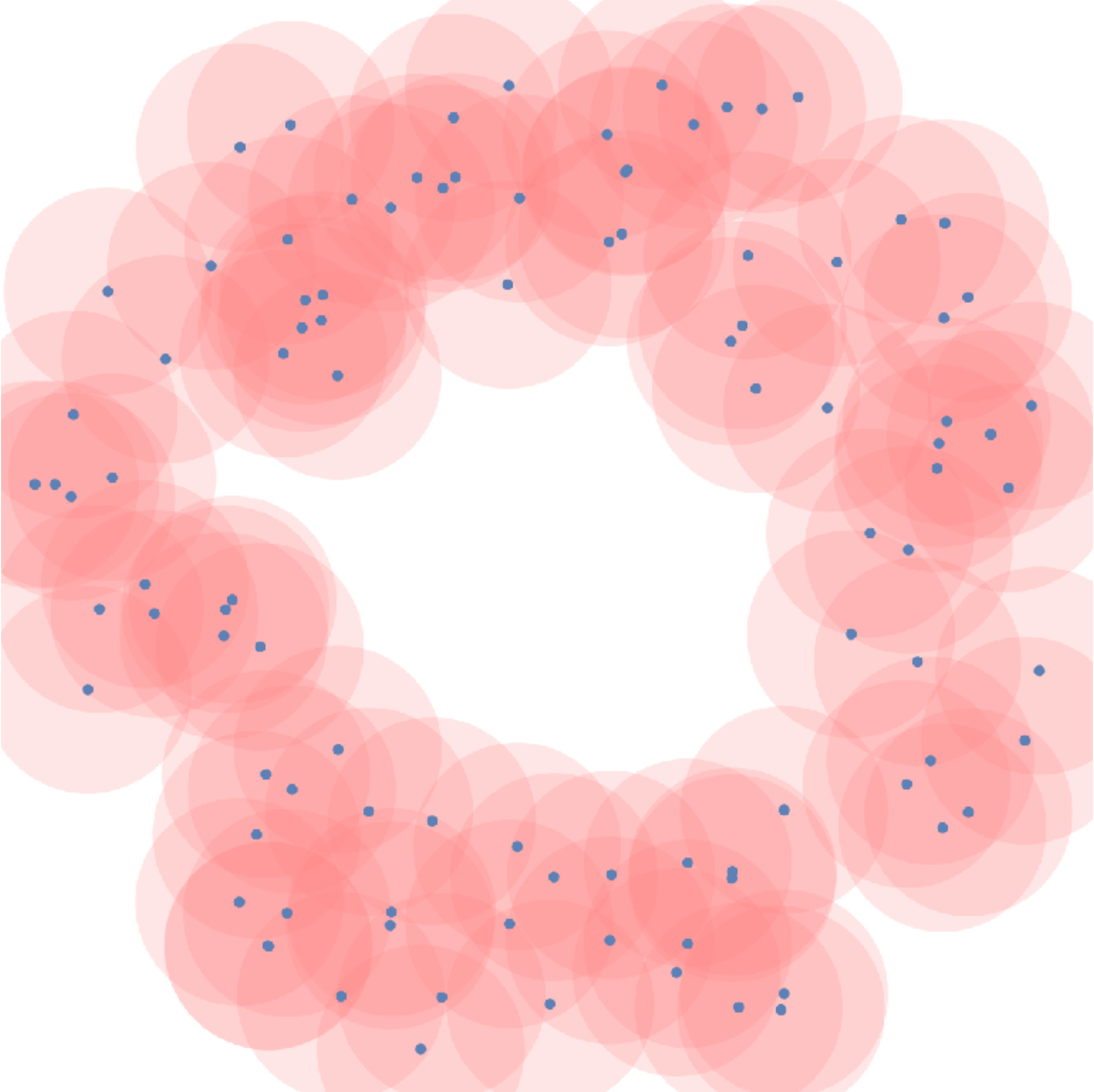


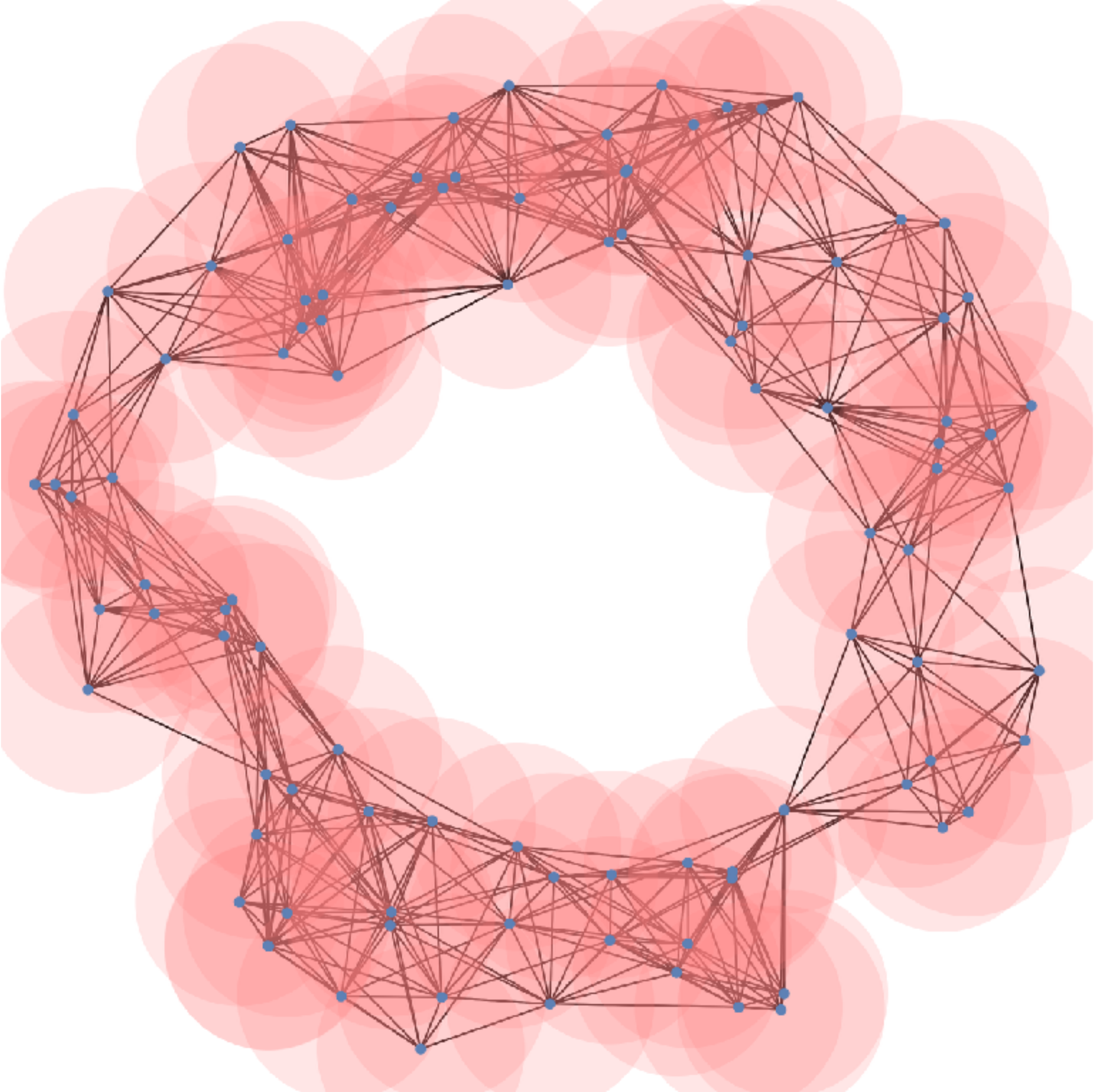


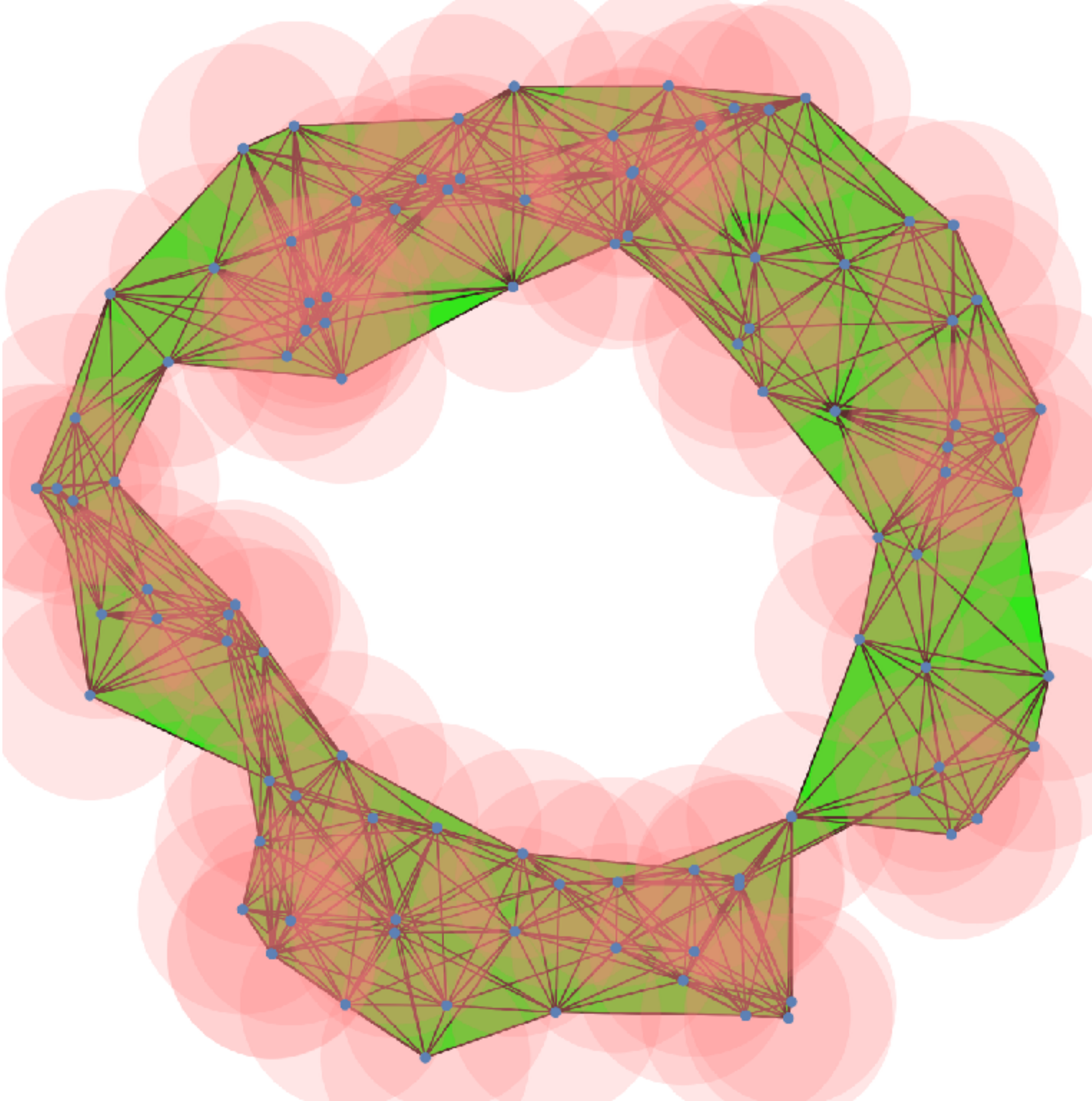


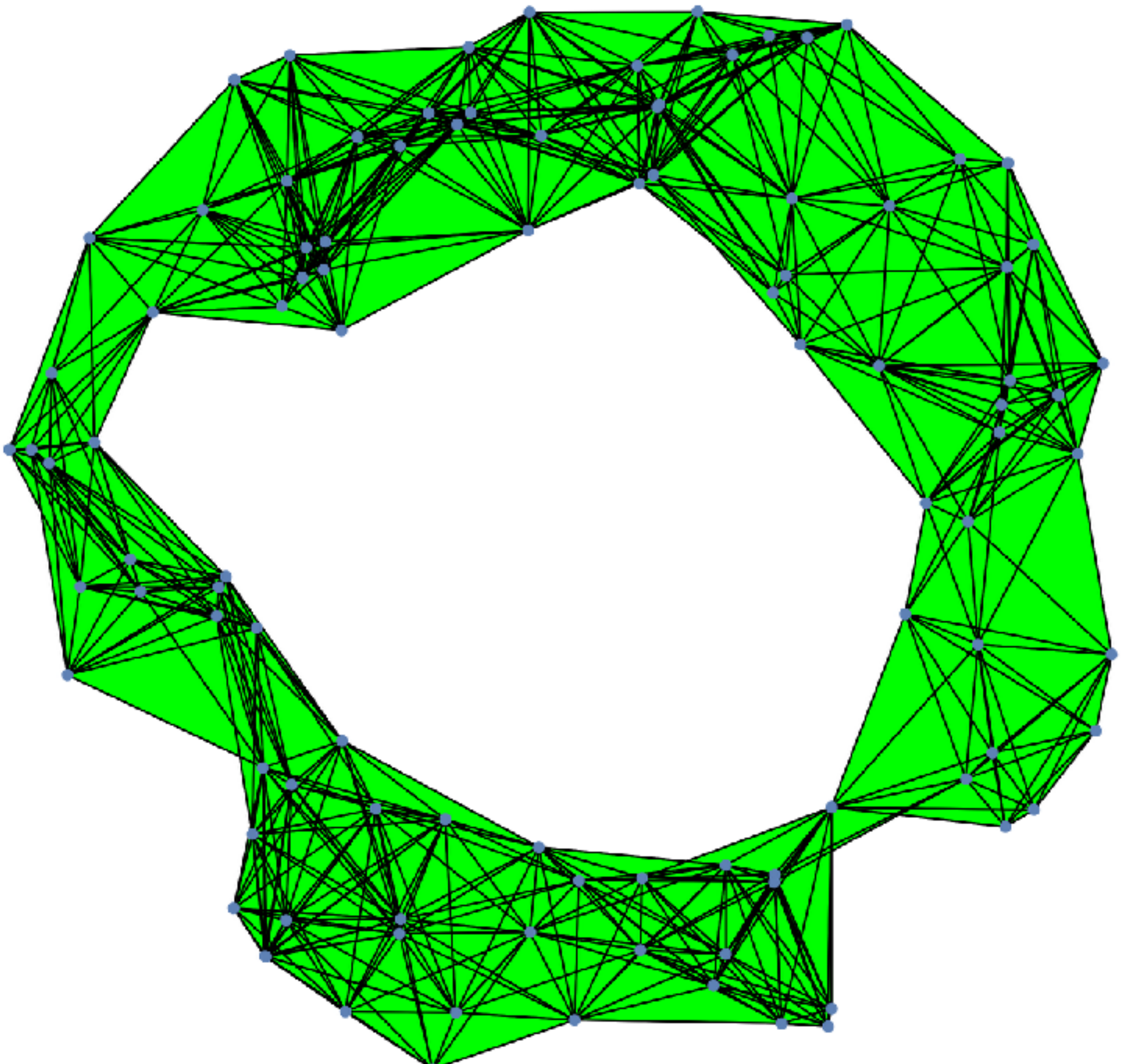






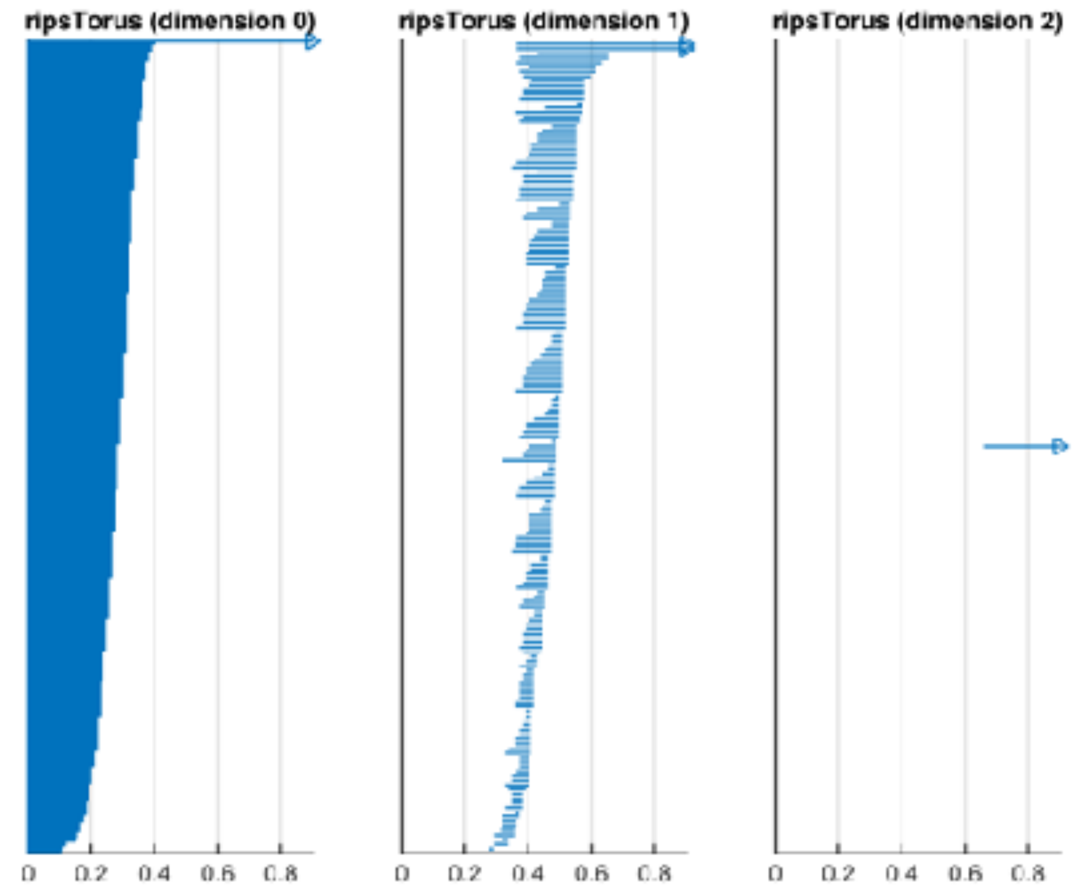




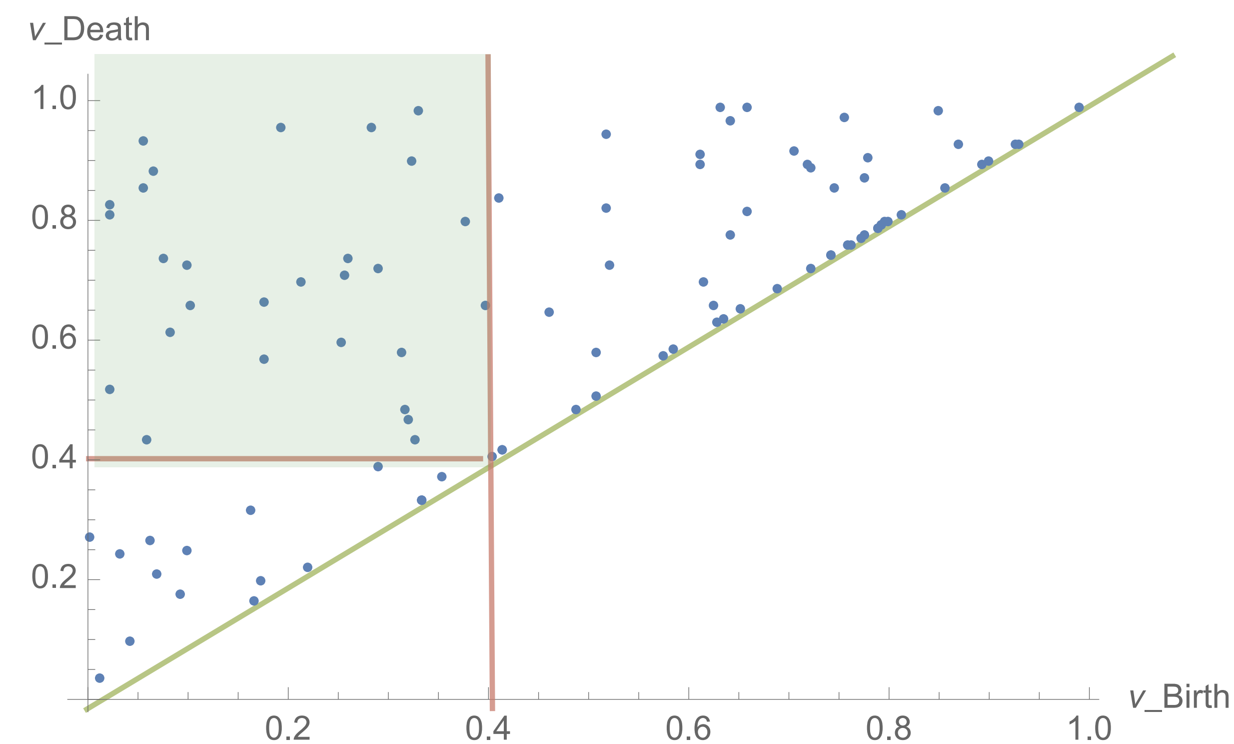


Visualizing Persistent Homology

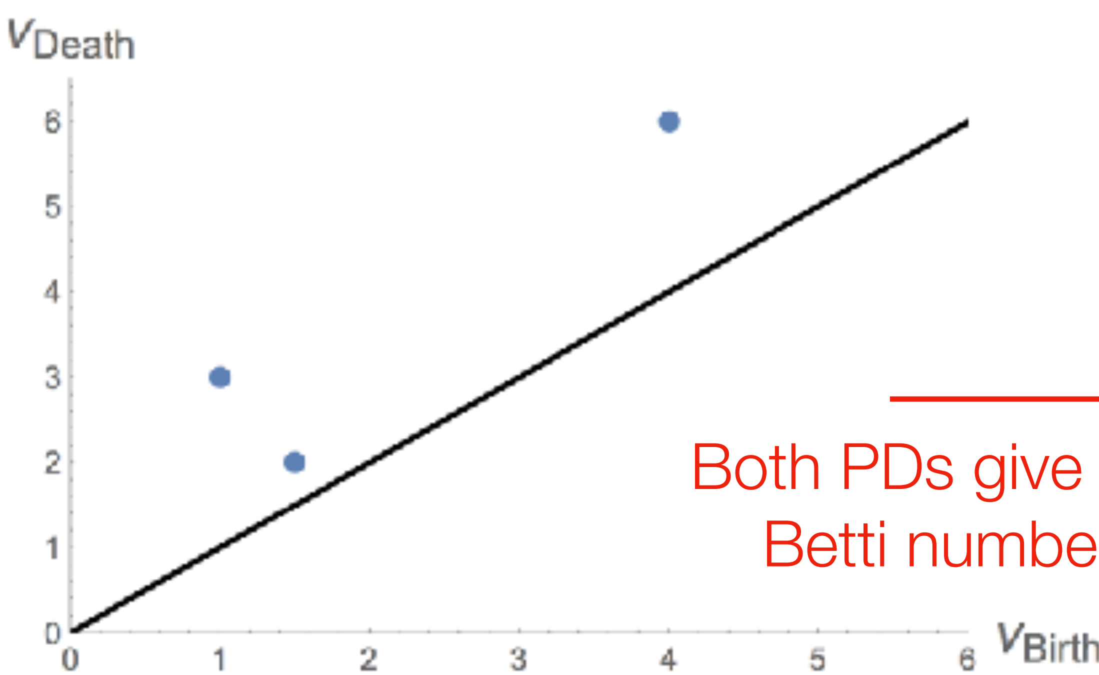
- **Barcodes**:
    - Each horizontal line represents an independent cycle contributing to a particular Betti number (i.e. a connected component, loop, void…)
    - Lines start at birth and end at death
    - To calculate Betti number, make vertical slice and count intersections

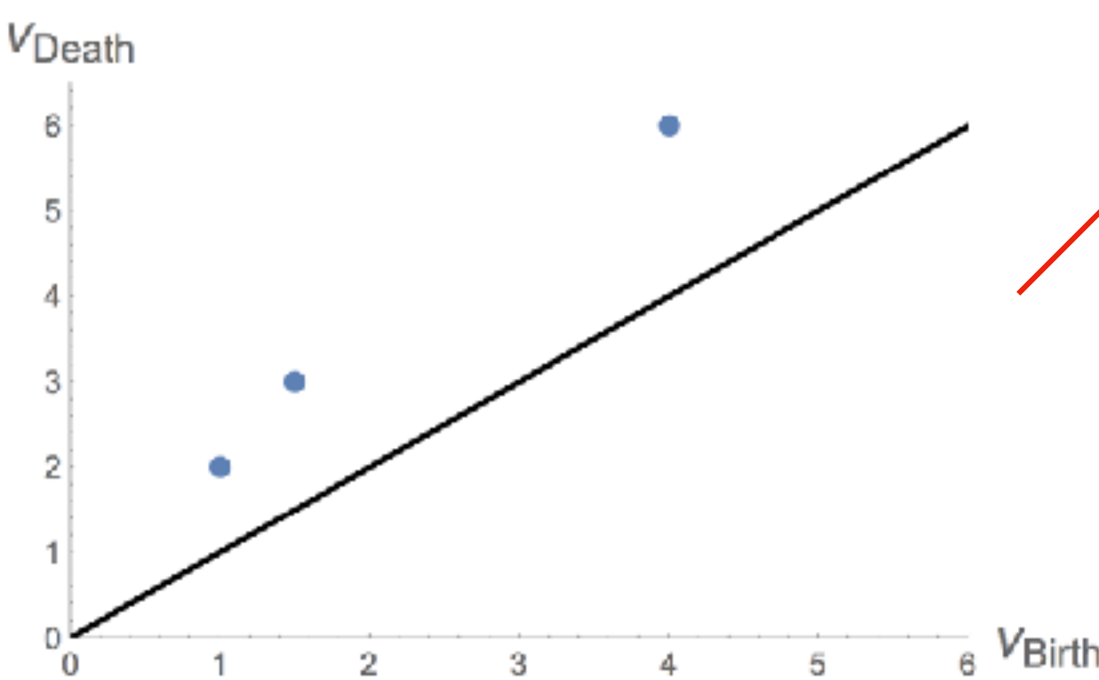


- **Persistence diagrams**:
    - Scatter plot, each point representing an independent cycle
    - Calculate Betti number by counting “living” cycles

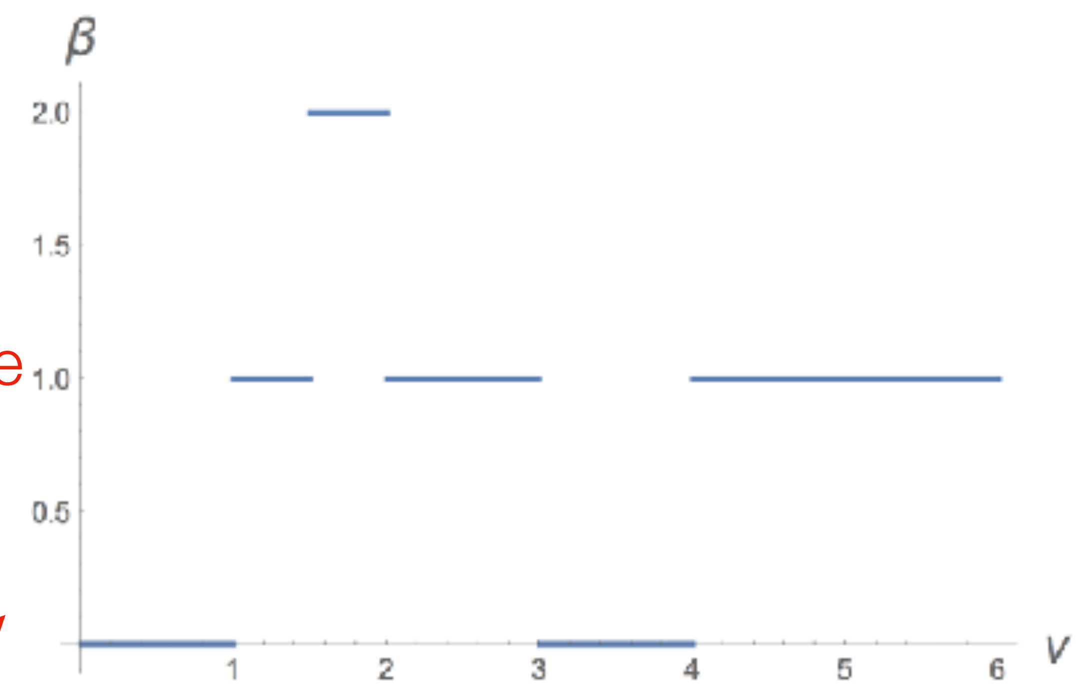


Persistence diagrams contain more information than Betti number curves!





Both PDs give the same Betti number curve



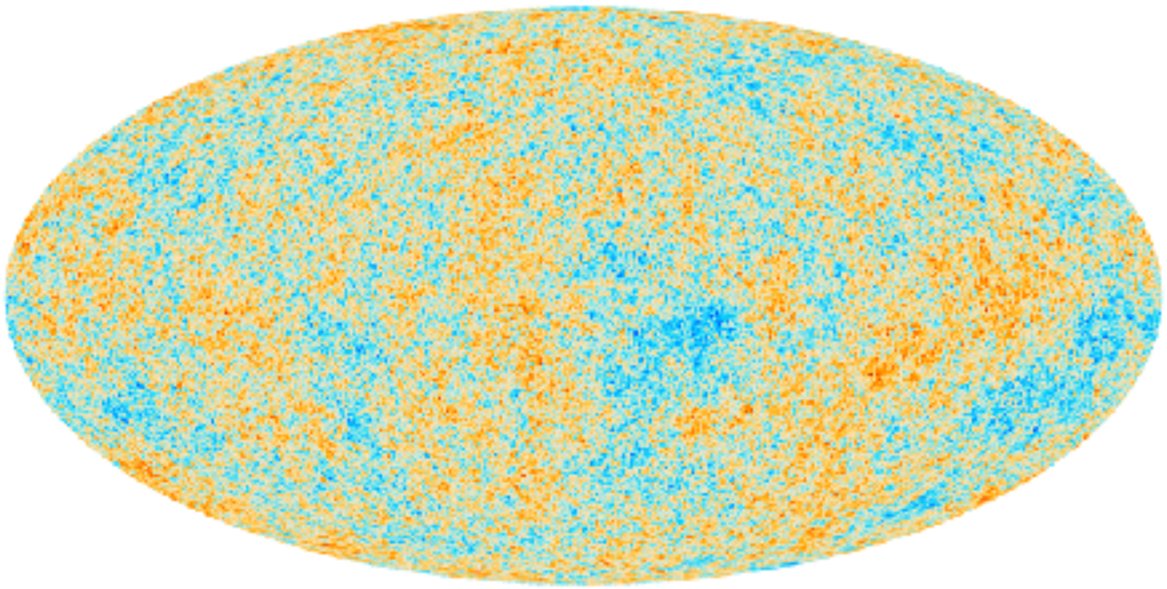
We can exploit this to improve CMB data analysis

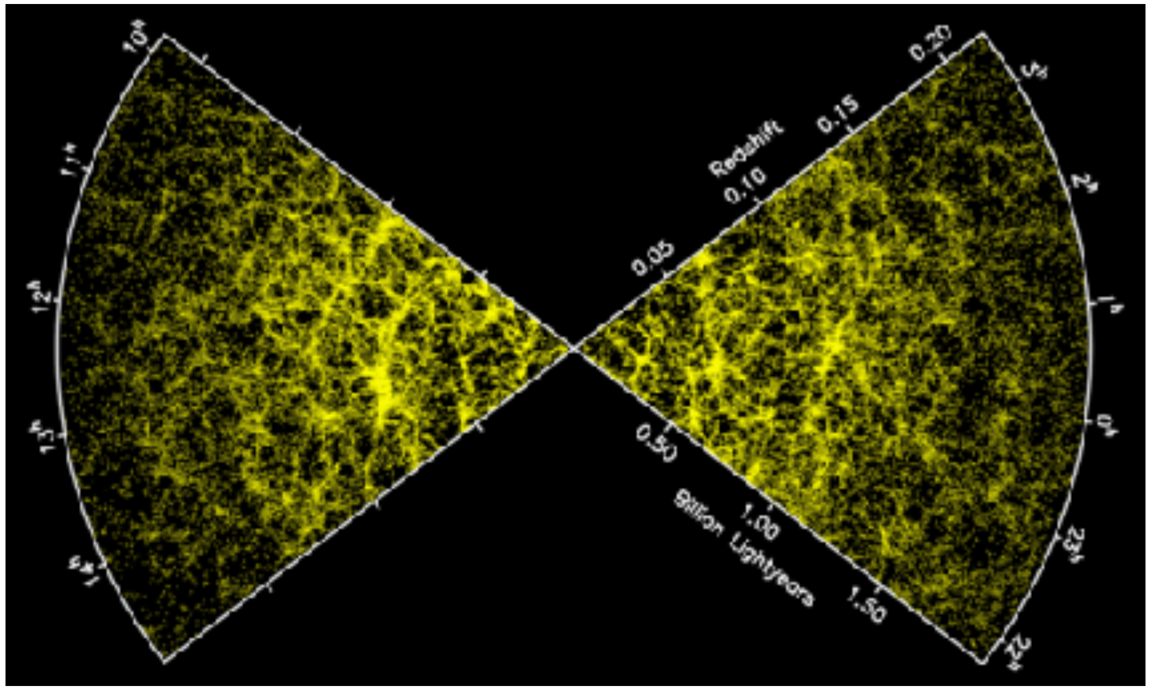
Applying TDA to Cosmology

Inflation

[Starobinsky];[Guth];[Linde];[Albrecht, Steinhardt];…

- Period of **accelerated expansion** in early universe
  - Solves flatness, horizon, and monopole problems
  - Predicts **nearly scale-invariant**, **Gaussian** curvature fluctuations
    - Source anisotropies in CMB, inhomogeneities in LSS
- A myriad of models. Taxonomy done mostly through their observables (n_s, r)





Anisotropies

- The lowest order correlation we can extract from the anisotropies is the **power spectrum**

$$\left\langle 0 \left| \hat{\mathcal{R}}_{\mathbf{k_1}} \hat{\mathcal{R}}_{\mathbf{k_2}} \right| 0 \right\rangle = (2\pi)^3 P_{\mathcal{R}}(k_1) \delta(\mathbf{k}_1 + \mathbf{k}_2) \qquad \Delta^2_{\mathcal{R}} = \left(\frac{k^3}{2\pi^2}\right) P^2_{\mathcal{R}} \propto k^{n_s - 1}$$

- For a Gaussian theory, the power spectrum dictates all higher-pt correlations.
- However, the inflationary fluctuations are not perfectly Gaussian.
- The leading **non-Gaussianity** is the **bispectrum**:

$$\langle 0 | \, \hat{\mathcal{R}}_{\mathbf{k_1}} \hat{\mathcal{R}}_{\mathbf{k_2}} \hat{\mathcal{R}}_{\mathbf{k_3}} \, | 0 \rangle = (2\pi)^3 \, \delta^3(\mathbf{k_1} + \mathbf{k_2} + \mathbf{k_3}) F(\mathbf{k_1}, \mathbf{k_2}, \mathbf{k}_3)$$

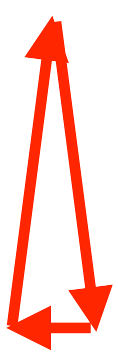
- Scaling and symmetries imply that F(**$\mathbf{k_1}$**, **$\mathbf{k_2}$, $\mathbf{k_3}$**) is fixed by an overall **size** ~ f_{NL} and its ''**shape"** F(1, k_2/k_1, k_3/k_1).
- More **powerful discriminator** of inflationary models.

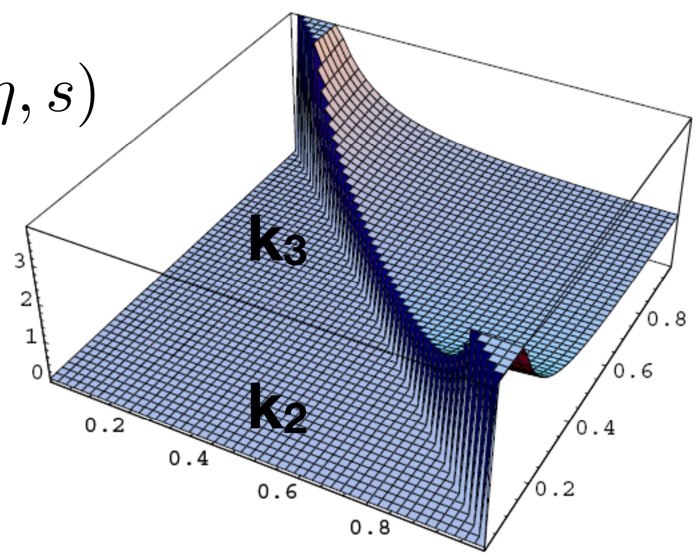
Non-Gaussianities

- The bispectrum for **single field slow-roll** inflation was computed in **[Maldacena, '02];[Acquaviva et al, '02]**; its size is $f_{NL} \sim \mathcal{O}(\varepsilon,\boldsymbol{\eta})$:
- The bispectrum for **general single field inflation** was found to be parametrized by 5 parameters **[Chen, Huang, Kachru, GS, '06]**:

Local shape

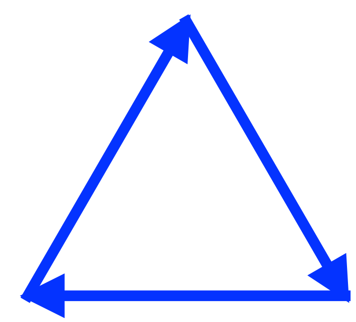
$$f_{NL}^{local} \sim \mathcal{O}(\epsilon, \eta, s)$$

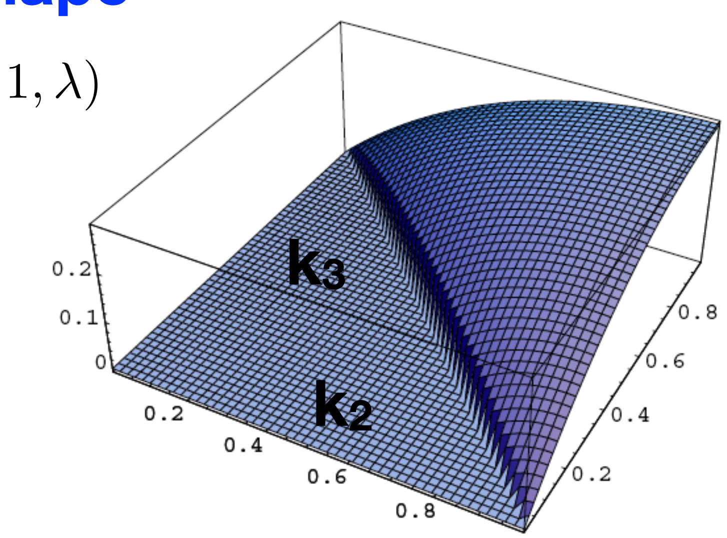




Equilateral shape

$$f_{NL}^{equil} \sim \mathcal{O}(\frac{1}{c_s^2} - 1, \lambda)$$



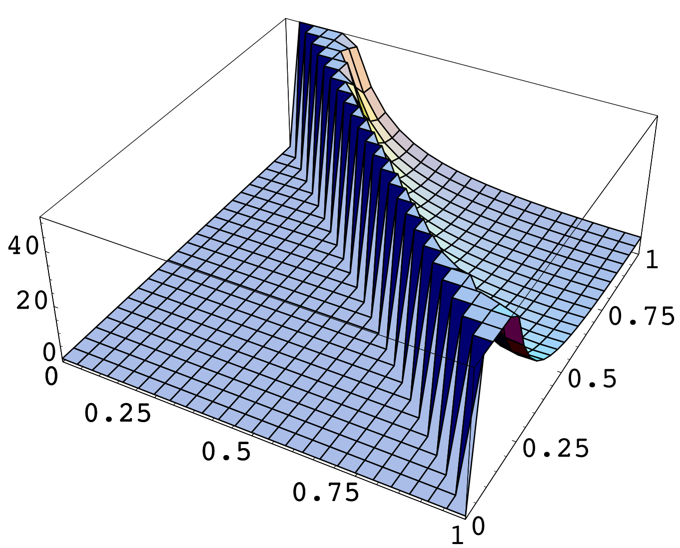


- There is also an **"orthogonal shape"** but it "looks" qualitatively like the equilateral shape (***challenge for machine learning?***).

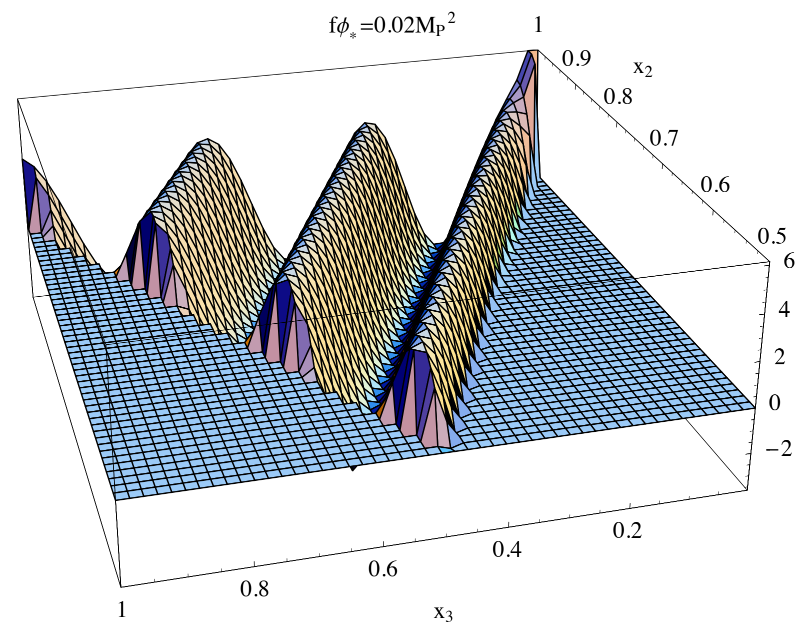
Non-Gaussianities

- More complicated models which involve non-standard initial conditions, features in potential (e.g. axion monodromy), or multiple fields or quasi-single field can give rise to more shapes:

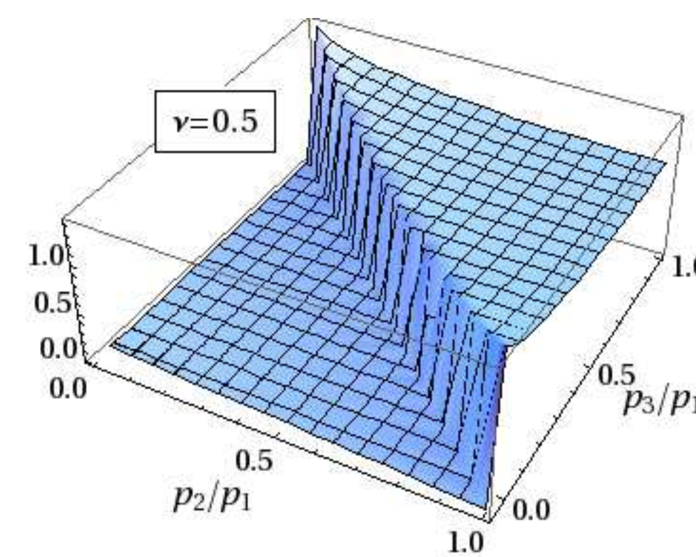
Non Bunch-Davis



Axion Monodromy



Quasi-single field



- Like scattering amplitudes in particle physics, non-Gaussianties can reveal interactions governing inflation: ***cosmological collider.***

- In collider physics: use ***different strategies*** for different particles.

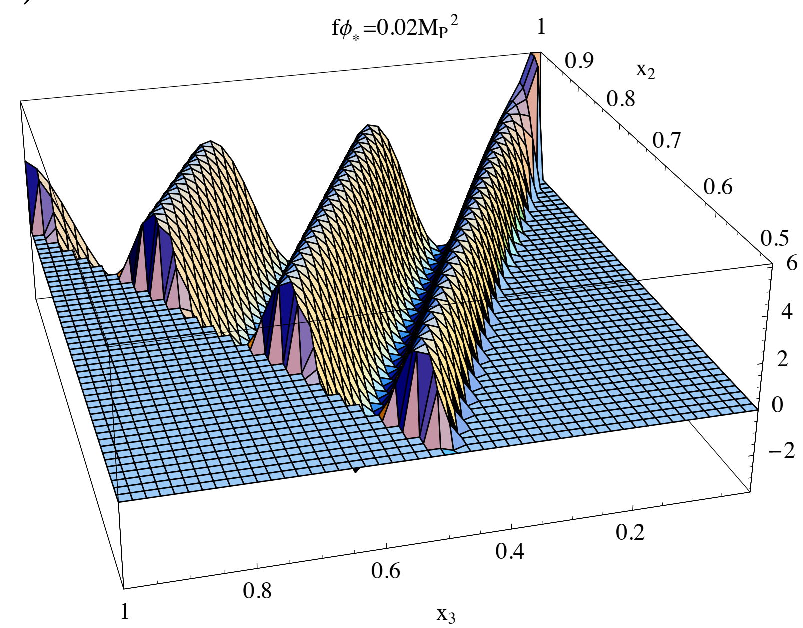
Measuring Non-Gaussianity

- **Harmonic space:** fits with *templates* of bispectrum, trispectrum, etc. One can define a "cosine" between distributions:

$$\cos(F_1, F_2) = \frac{F_1 \cdot F_2}{(F_1 \cdot F_1)^{1/2} (F_2 \cdot F_2)^{1/2}}$$

- Some shapes are harder to find, e.g.,

Resonant shape (axion monodromy)



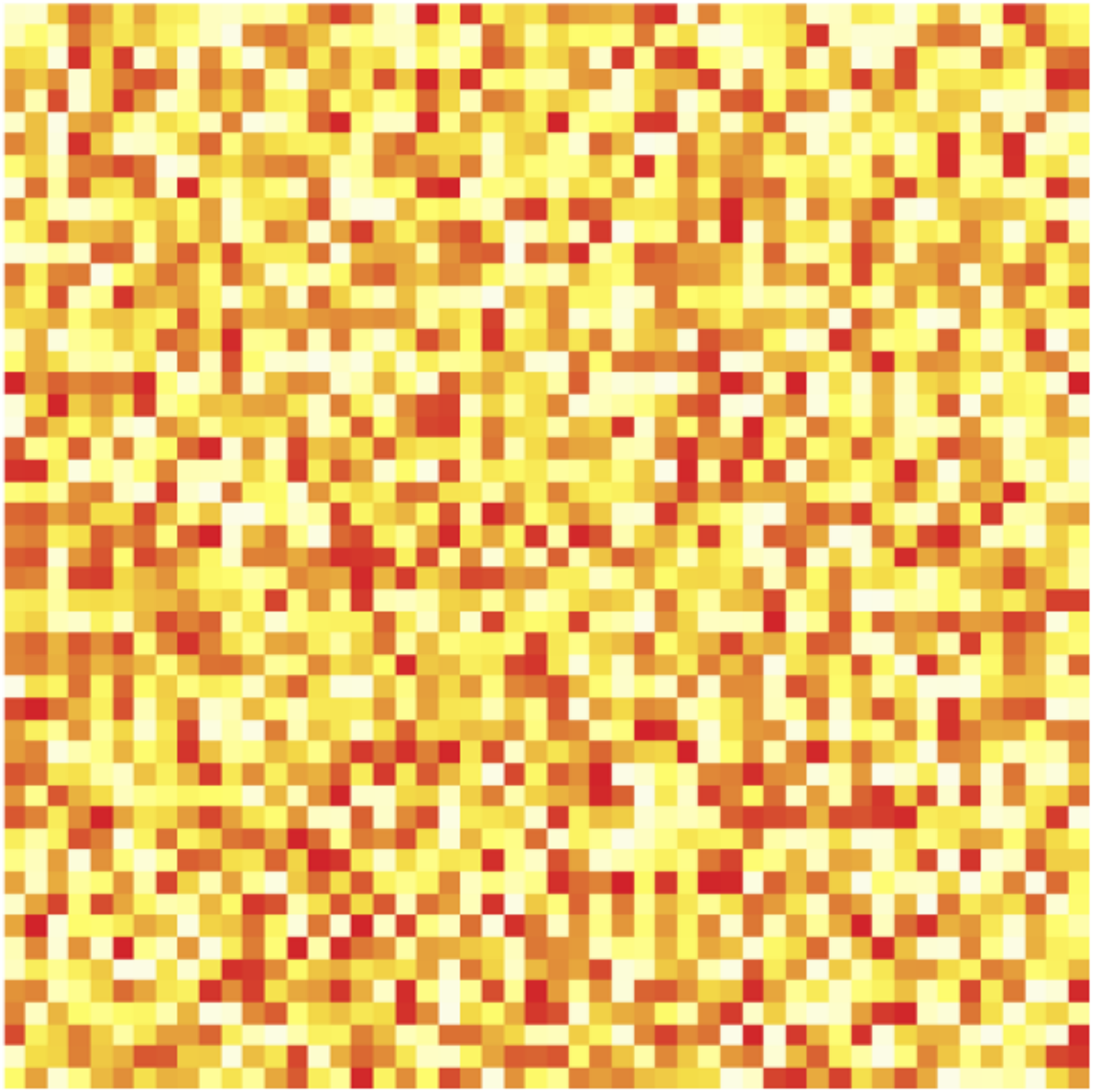
- Geometrical/topological: **Minkowski functionals** (for CMB: area fraction, length of boundaries, and genus of excursion sets)

- Current bound on non-Gaussianity (Planck '15):

$$f_{NL}^{local} = 2.5 \pm 5.7 \qquad\qquad f_{NL}^{equil} = -16 \pm 70$$

Sublevel Filtration

(Hotter points are deeper red)

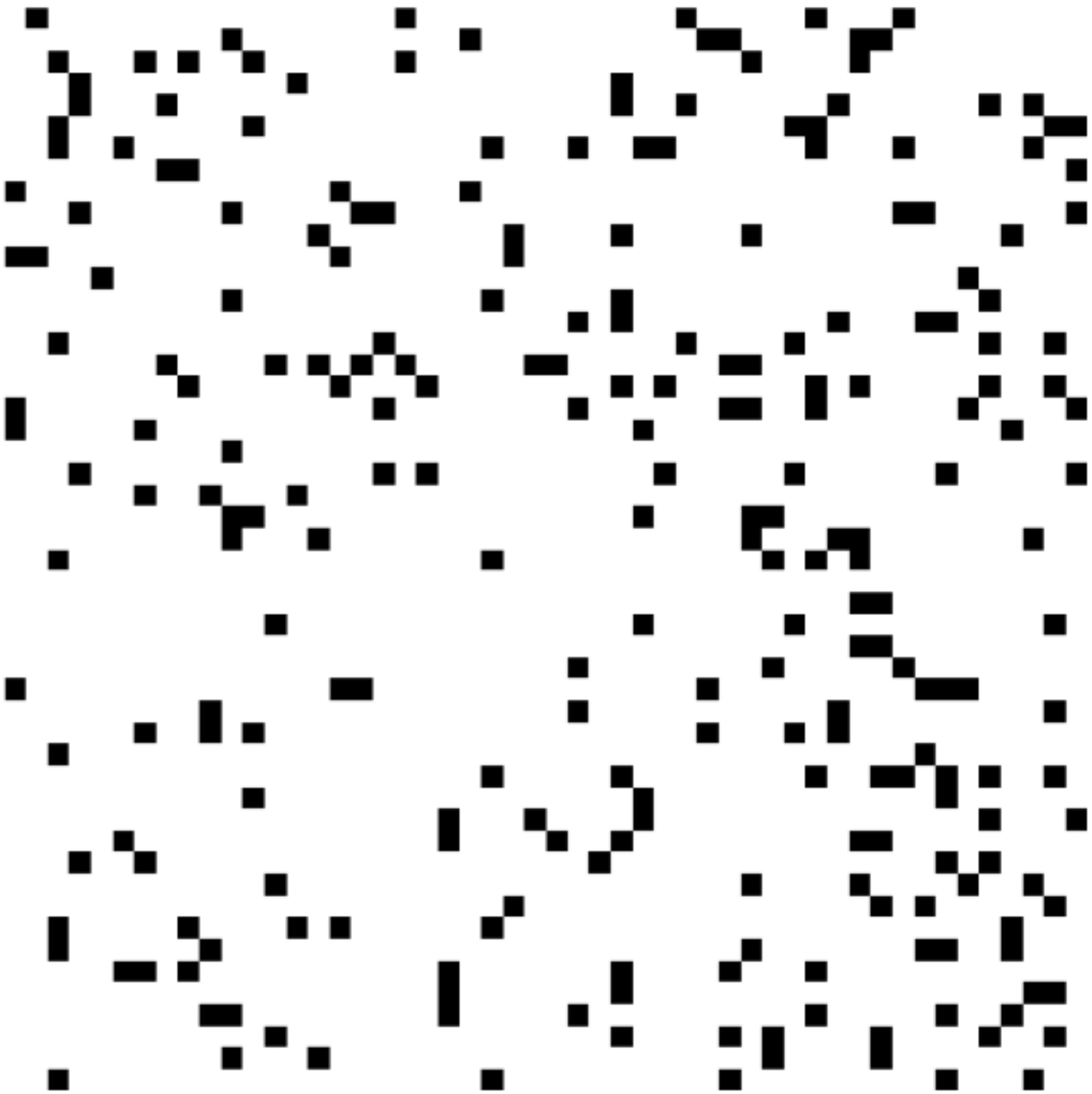


Sublevel Filtration

$\nu = -1$

Many distinct components, no loops

(Sublevel set in black)

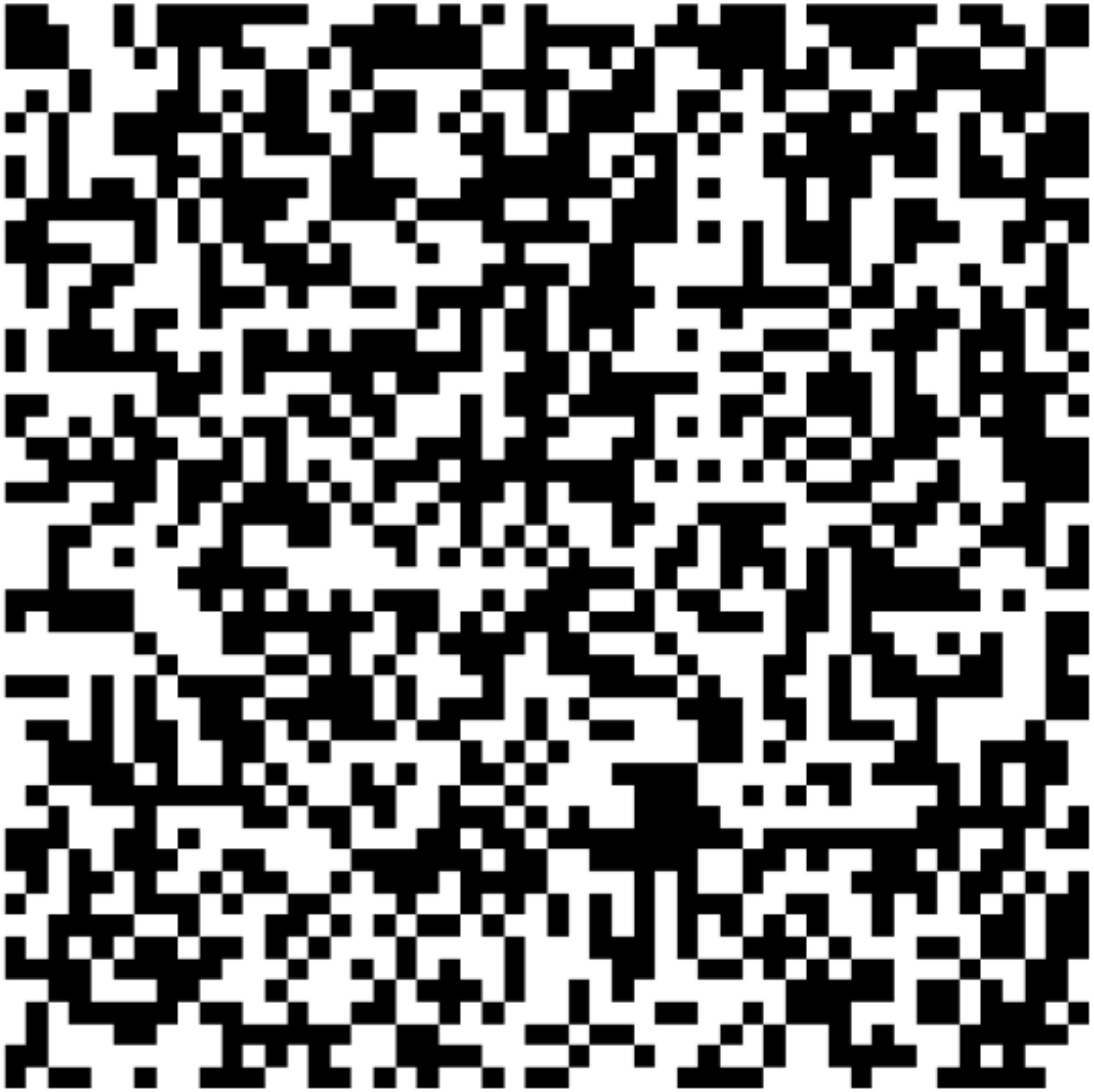


Sublevel Filtration

$\nu = 0$

Many loops, fewer distinct components

(Sublevel set in black)

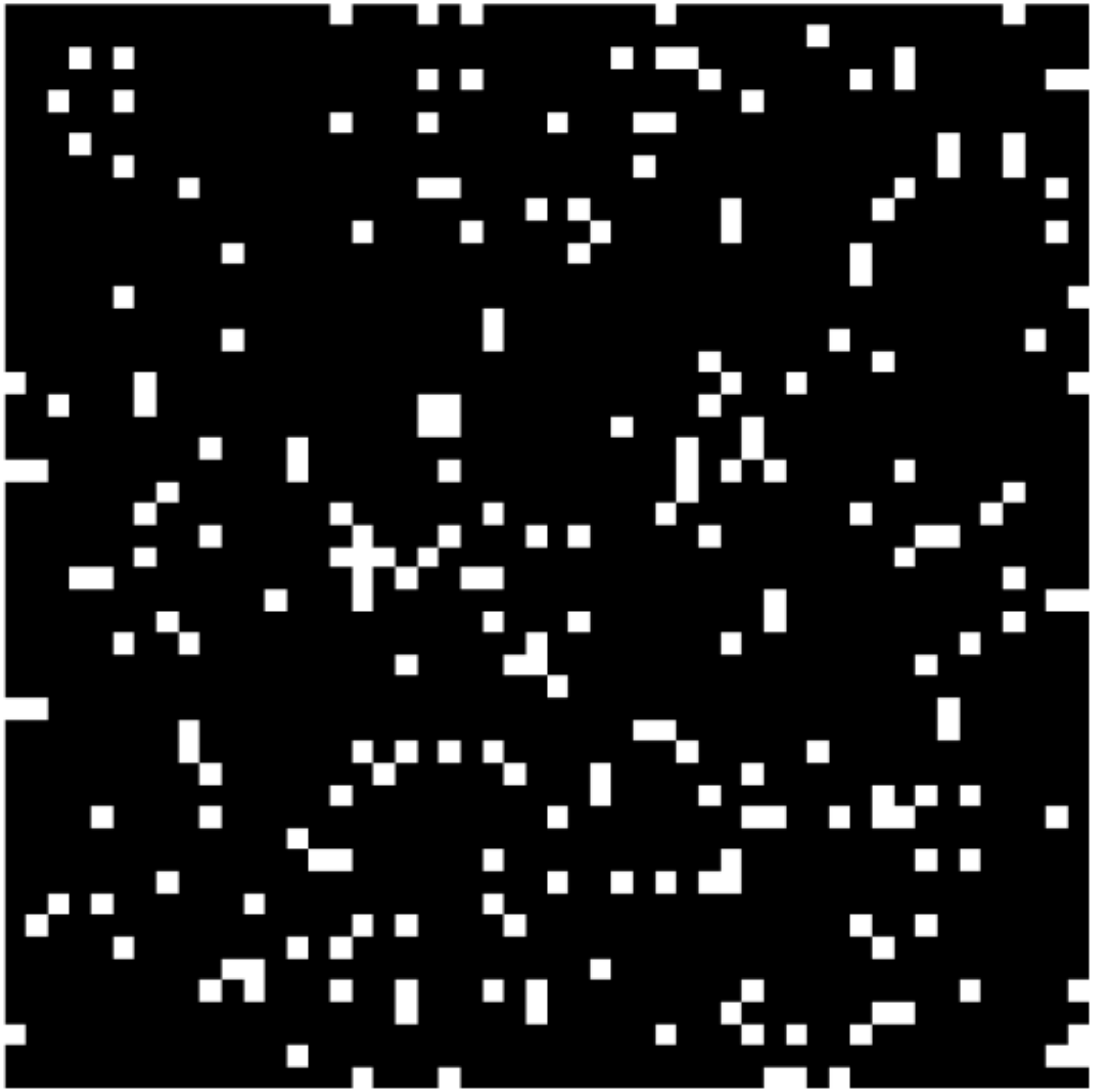


Sublevel Filtration

$\nu = 1$

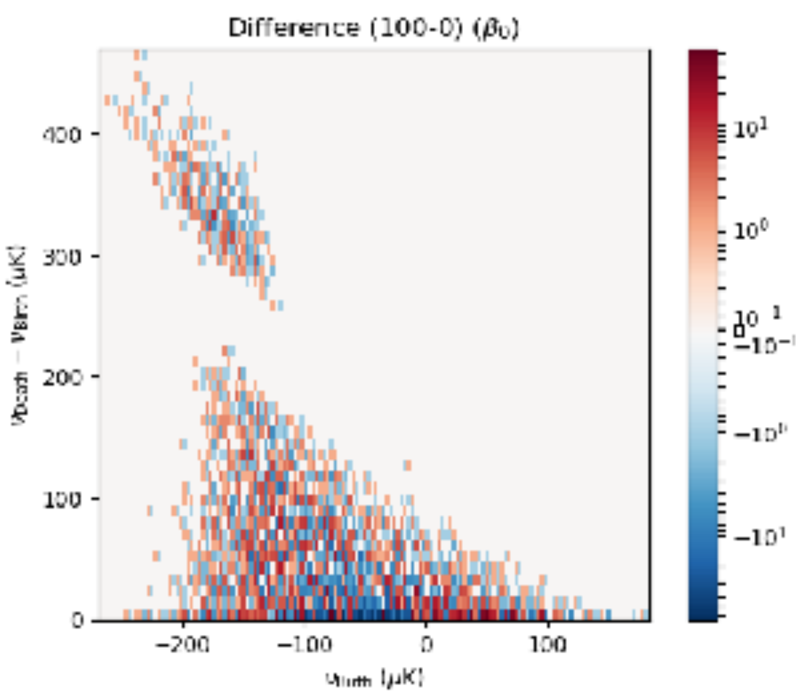
One connected component, many loops have been filled in

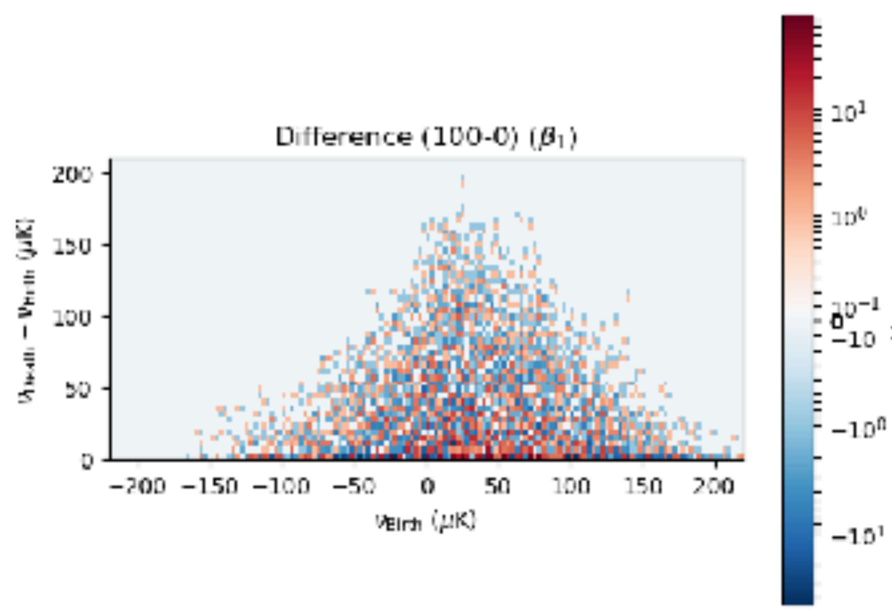
(Sublevel set in black)



Sensitivity to Non-Gaussianity

- We first carried out TDA for local NG and with low-resolution maps ($\ell_{max} \sim 1024$) as a warmup, more in our pipeline.
- Bined the PDs for different f_{NL}, & computed the likelihood function:





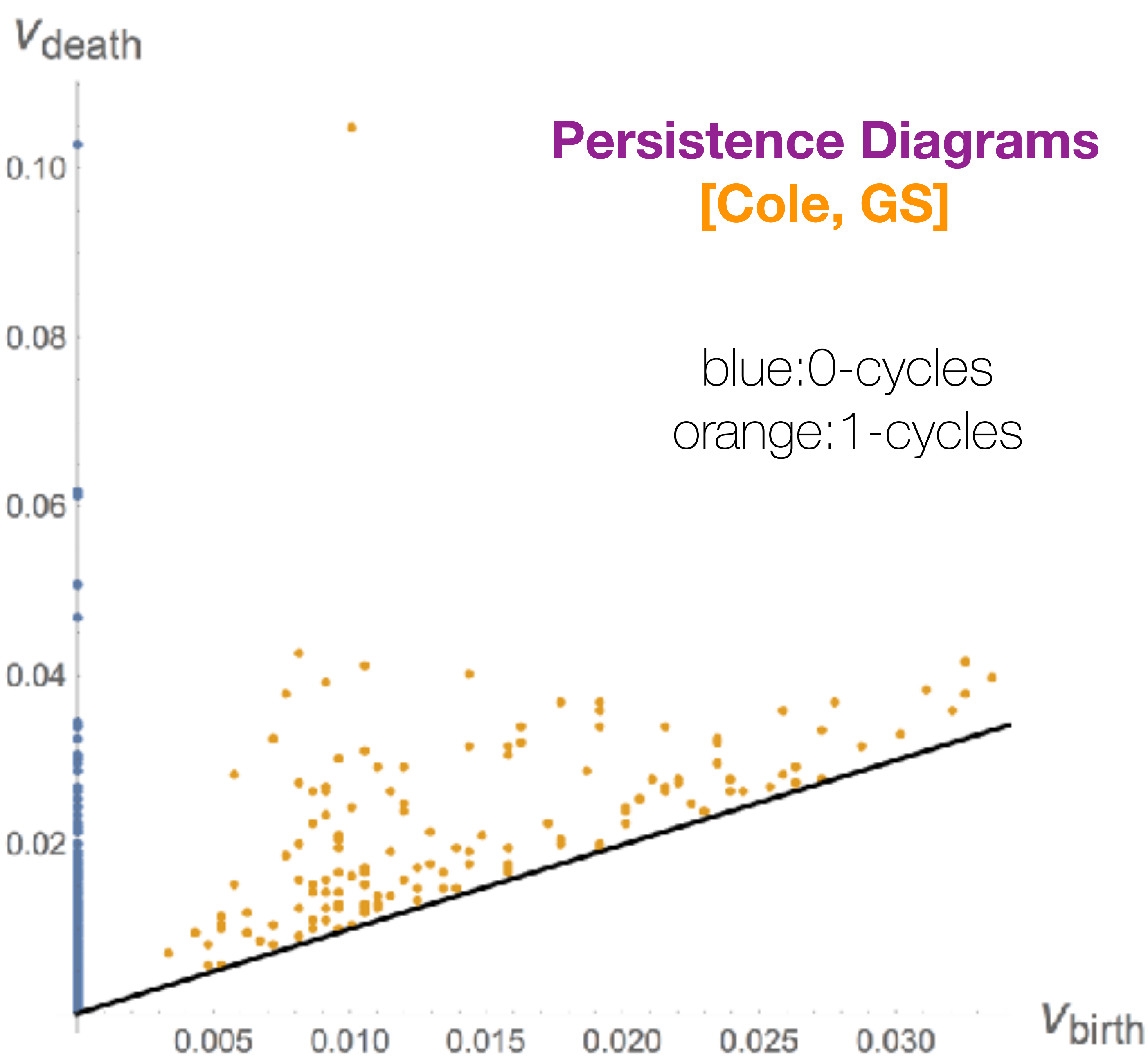
1σ error bar

Statistic	Δf_{NL}
β_0	67.4
β_1	66.1
$\beta_0 + \beta_1$	60.6
PD_0	39.1
PD_1	37.4
$PD_0 + PD_1$	35.8

- More sensitive statistic than Minkowski functionals or Betti number curves, **PDs strengthens topological analysis significantly**.
- N.B. Lower resolution maps used here compared to Planck's.
- Other subtle shapes (e.g., resonant non-Gaussianity).

Applying TDA to String Vacua

TDA for String Vacua

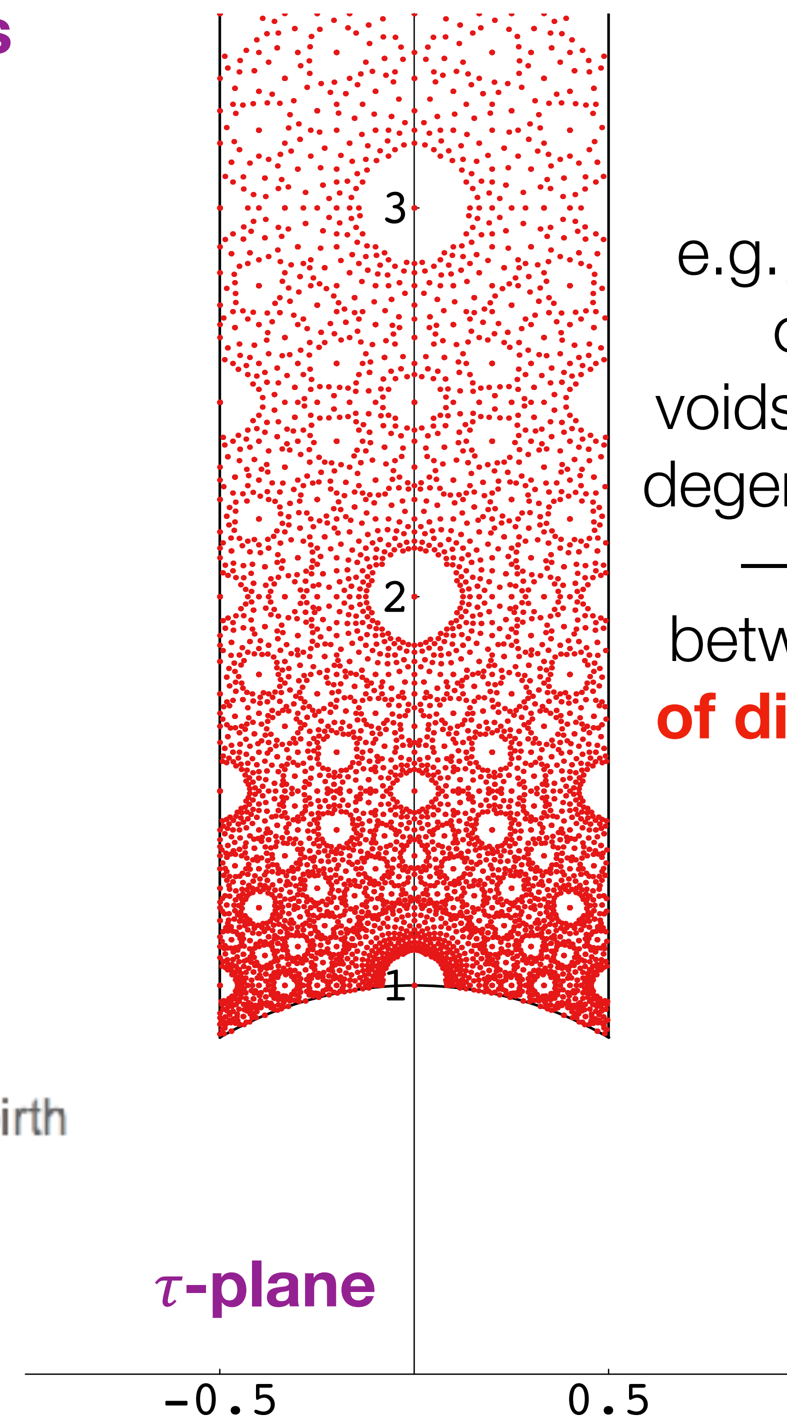


Persistence Diagrams
[Cole, GS]

blue:0-cycles
orange:1-cycles

See also **[Cirafici]** for the barcodes

"Topological Complexity"



τ-plane

e.g., for flux vacua on rigid CY, voids correspond to degeneracy of vacua — relationship between **topology of distribution** and **physics**

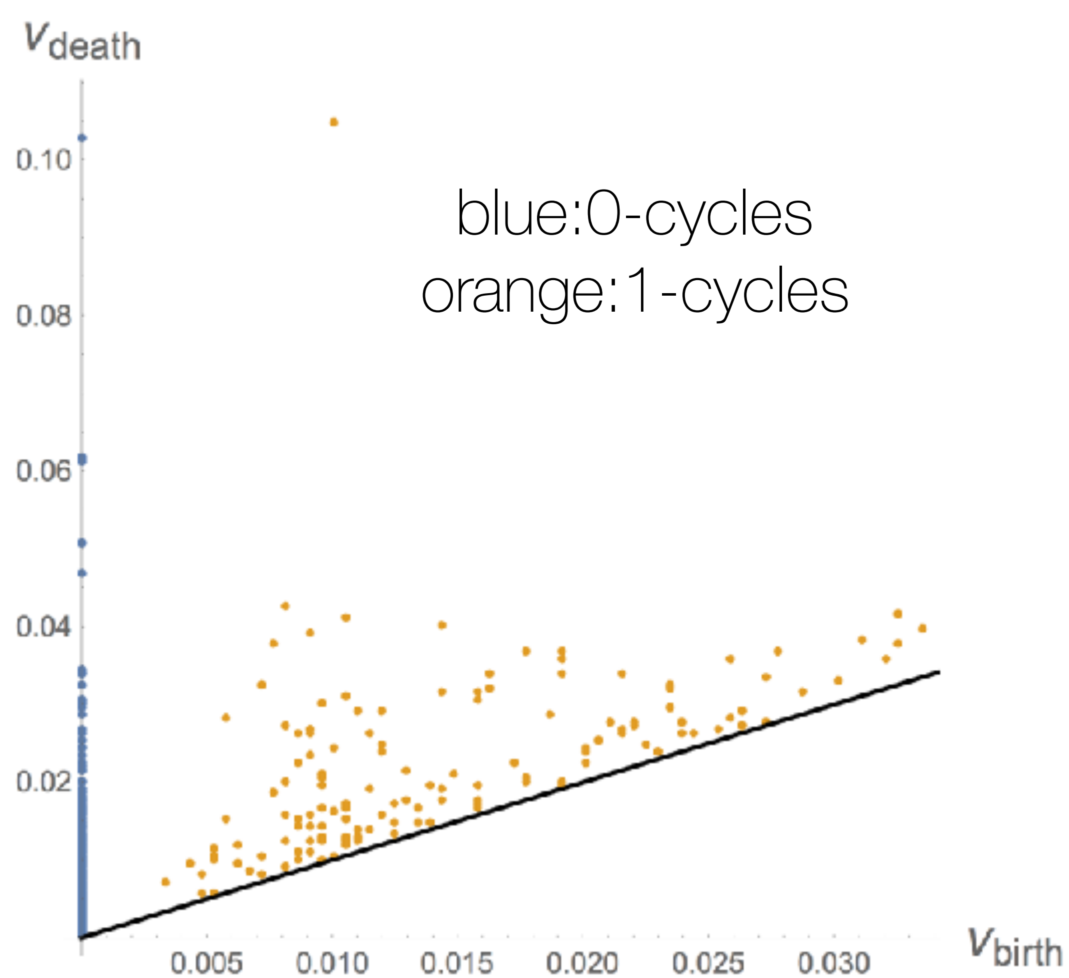
Toy Example: IIB Flux Vacua on Rigid CY

- **Superpotential** W = Aτ+B where the flux quanta:

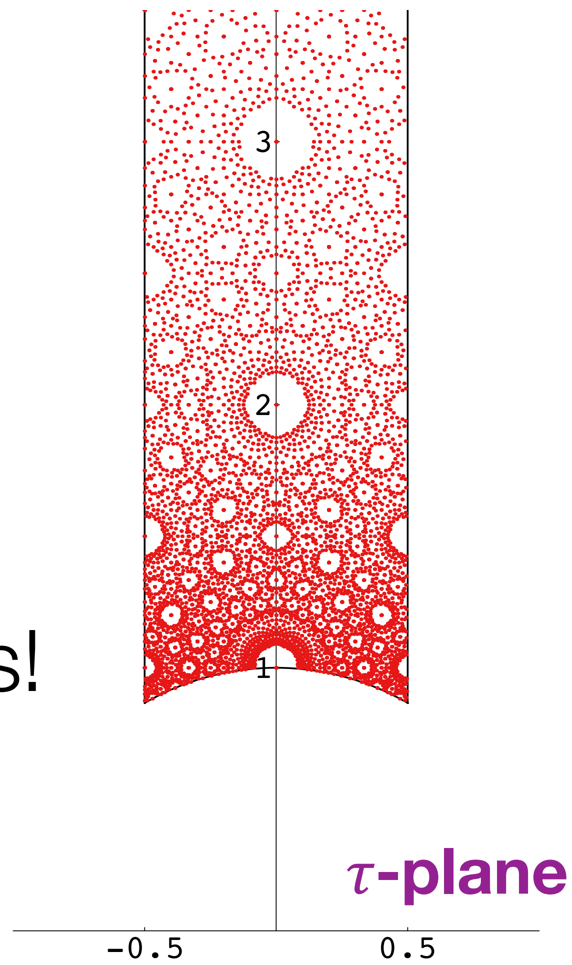
$$A = -h_1 - ih_2, \quad B = f_1 + if_2, \quad h_1, h_2, f_1, f_2 \in \mathbb{Z}$$

subject to **tadpole cancellation**: $N_{\text{flux}} = f_1 h_2 - h_1 f_2 \leq L_{\max}$

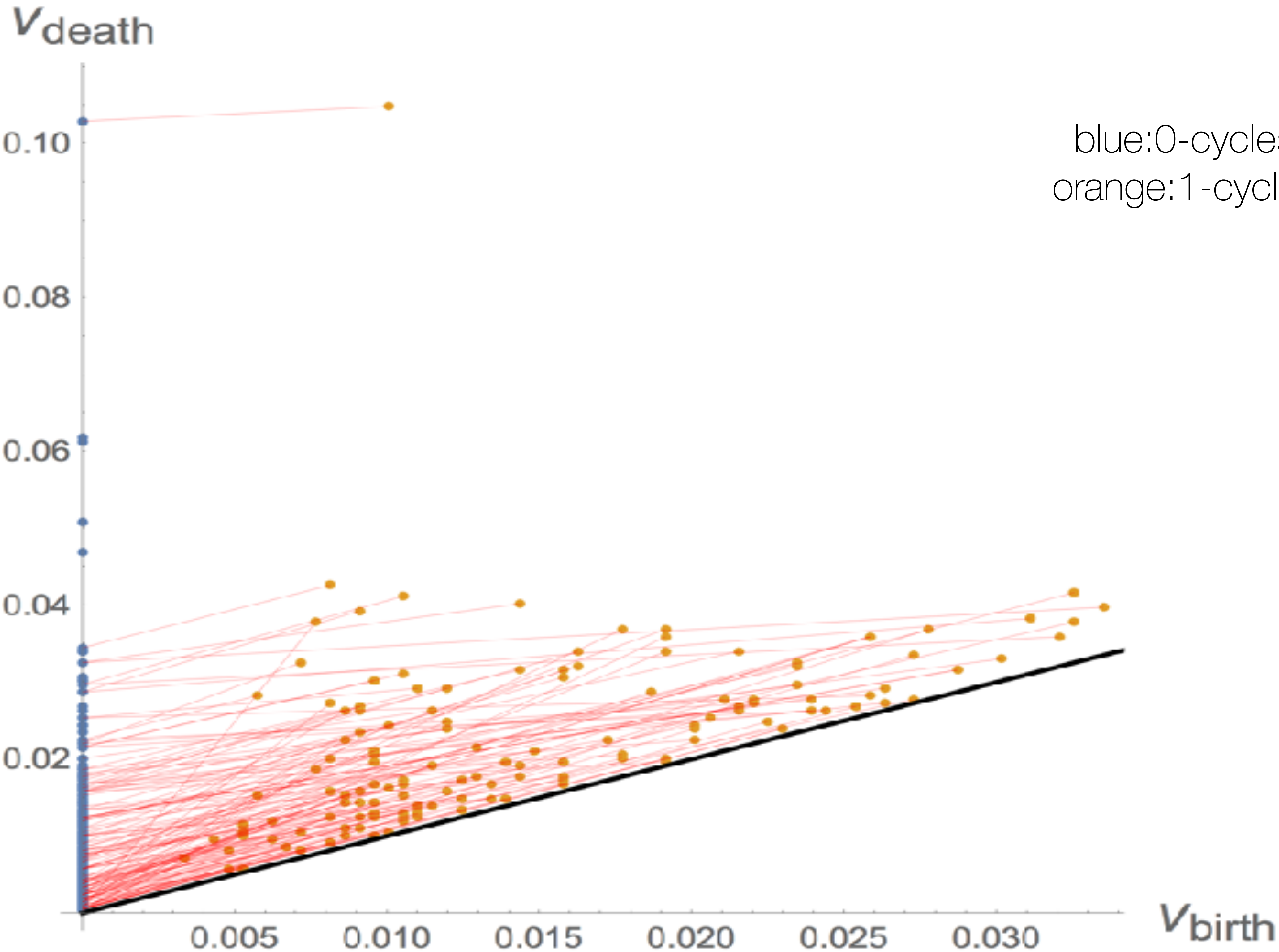
- Vacua are mapped to the **fundamental domain** using SL(2,Z).



Short-lived
but **correlated**
topological features
most apparent in
persistence diagrams!



Persistence Pairing



blue:0-cycles
orange:1-cycles

Work in progress [Cole,GS]

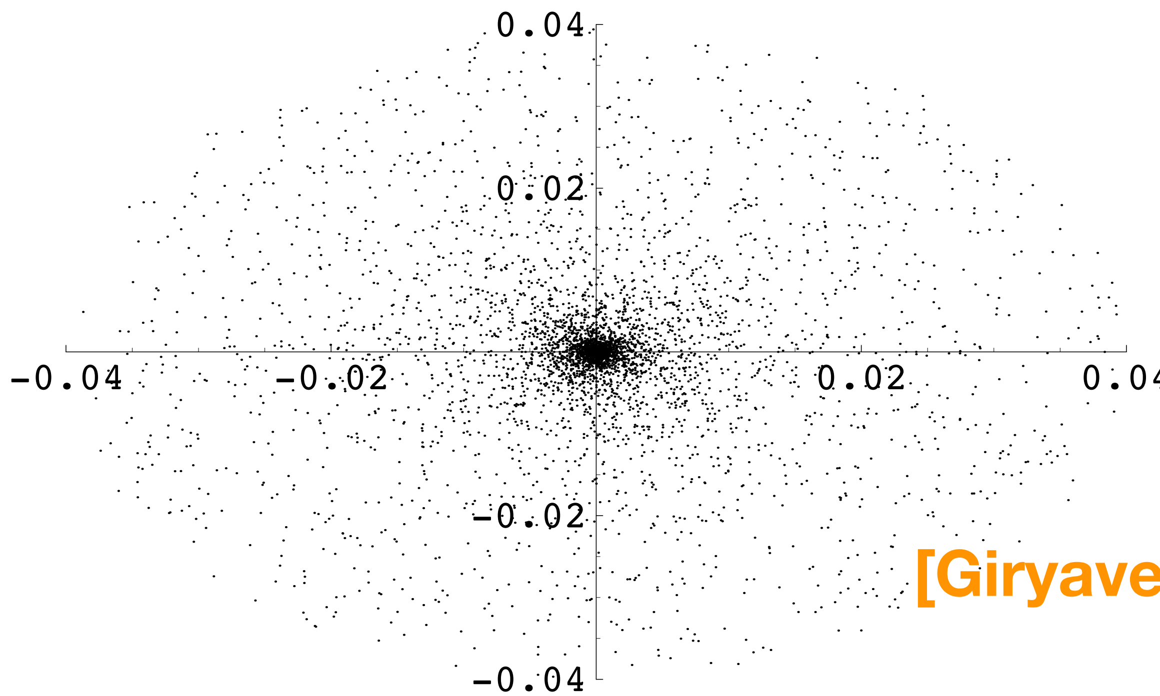
Flux Vacua on CY Hypersurface

- In general, not possible to visualize a *higher dim.* data space.
- For example, flux vacua of IIB orientifold on CY hypersurface:

$$\sum_{i=1}^{4} x_i^8 + 4x_0^2 - 8\psi\, x_0 x_1 x_2 x_3 x_4 = 0\,, \quad x_i \in \mathbf{WP}^4{}_{1,1,1,1,4}$$

has $h^{1,1}=1$, $h^{2,1}=149$ and discrete symmetry $\Gamma=Z_8^2 \times Z_2$. The only Γ-invariant moduli: complex structure modulus ψ & axio-dilaton τ.

Projecting onto the x=1-ψ plane



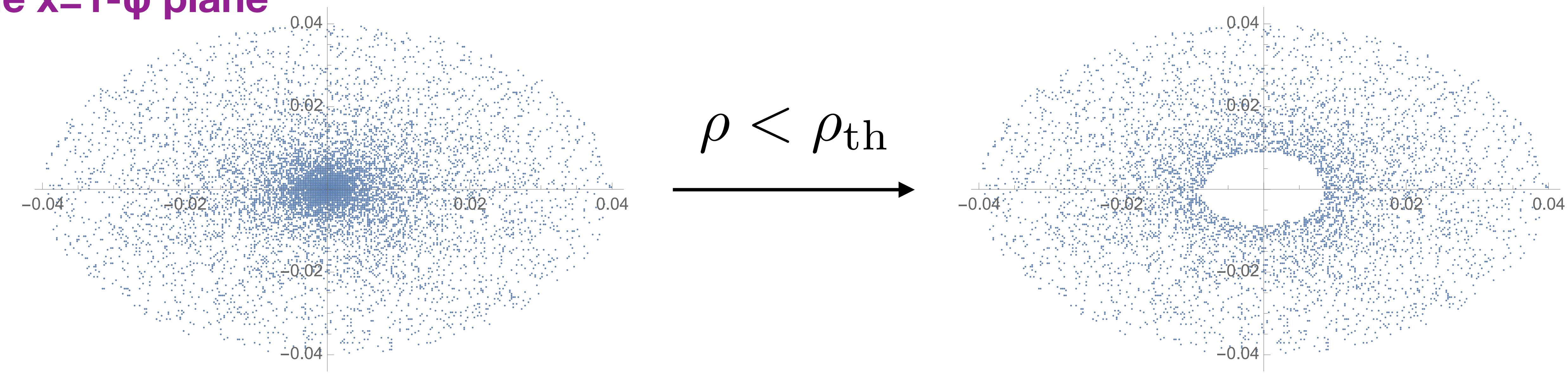
[Giryavets, Kachru, Tripathy]

- TDA can more systematically diagnose the vacuum structure.

Work in progress [Cole,GS]

Flux Vacua on CY Hypersurface

Projecting onto the x=1-ψ plane

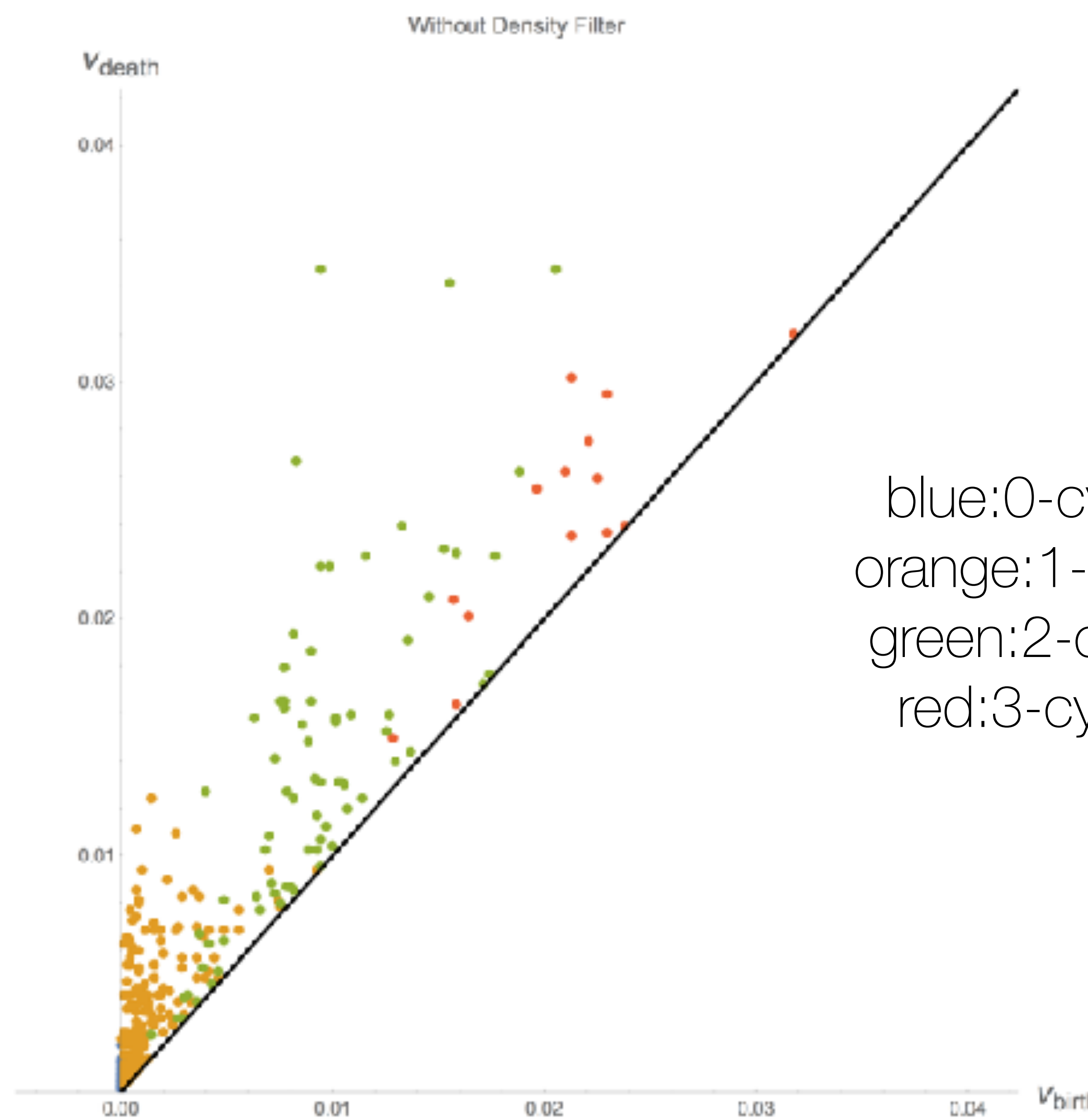


- To identify cluster, apply density cutoff (excises cluster, results in identifiable void)
- Does this cluster/void exist in the full four-dimensional space? (Might not if clustering correlates with structure in axiodilaton.) Are there significant higher dimensional features?
    - These questions can be answered with persistent homology

Work in progress [Cole,GS]

Flux Vacua on CY Hypersurface

- To identify cluster, apply density cutoff (excises cluster, results in identifiable void)
- We found a long-lived 1-cycle in the full four-dim. space and only observe short-lived higher dimension features (sampling noise)

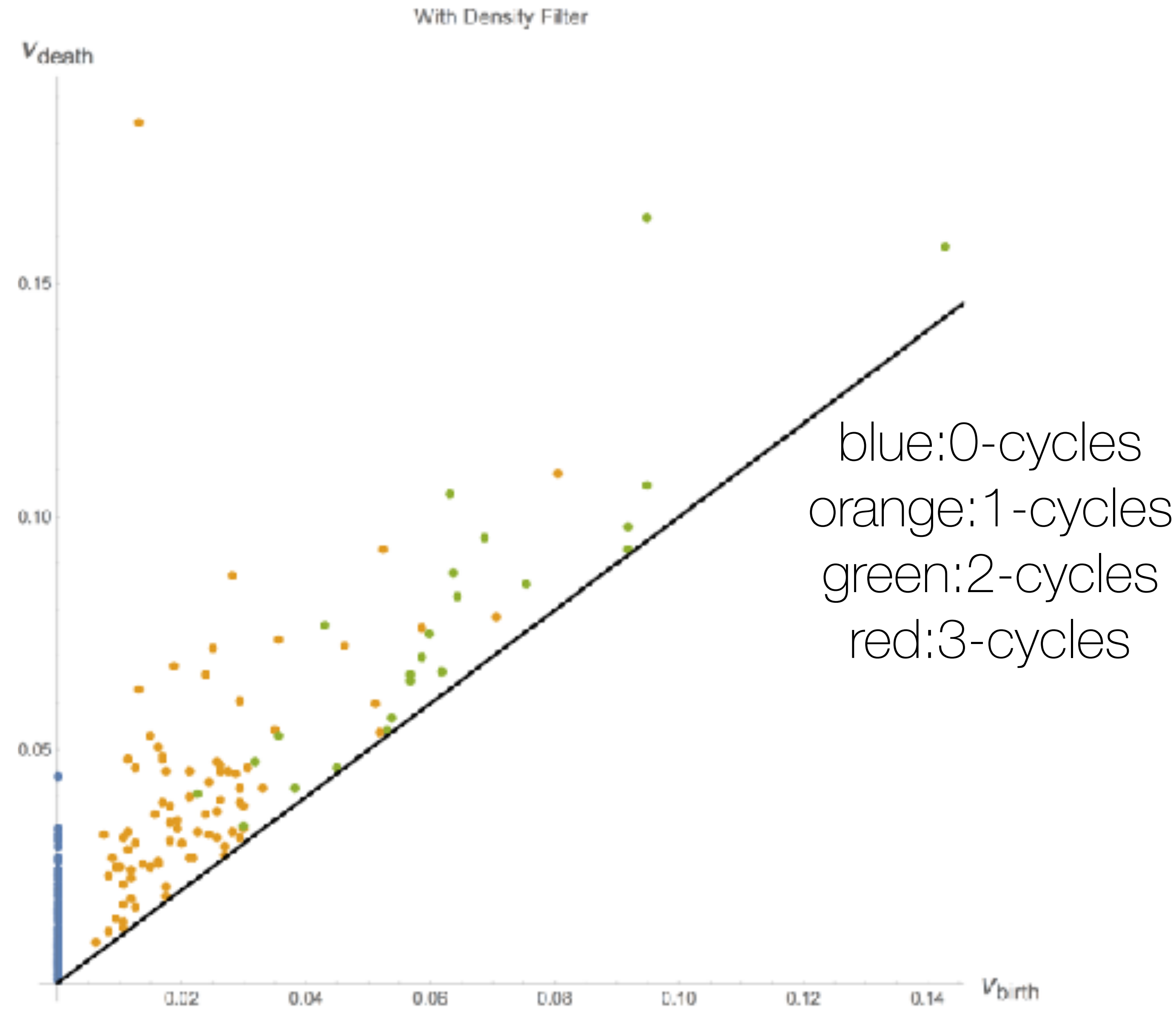


blue:0-cycles
orange:1-cycles
green:2-cycles
red:3-cycles

Work in progress [Cole,GS]

Flux Vacua on CY Hypersurface

- To identify cluster, apply density cutoff (excises cluster, results in identifiable void)
- We found a long-lived 1-cycle in the full four-dim. space and only observe short-lived higher dimension features (sampling noise)

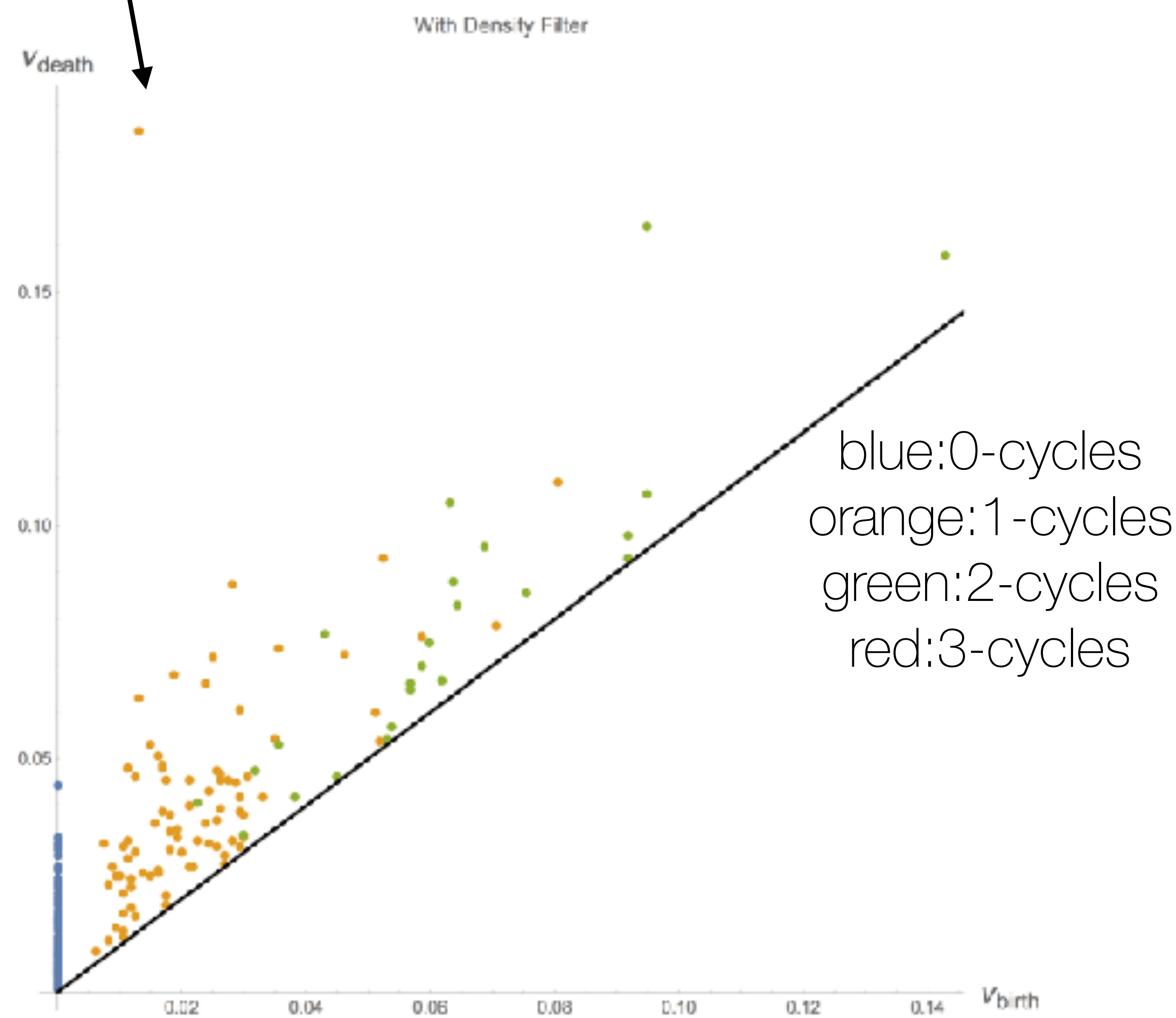


Work in progress [Cole,GS]

Flux Vacua on CY Hypersurface

- To identify cluster, apply density cutoff (excises cluster, results in identifiable void)
- We found a long-lived 1-cycle in the full four-dim. space and only observe short-lived higher dimension features (sampling noise)

long-lived 1-cycle

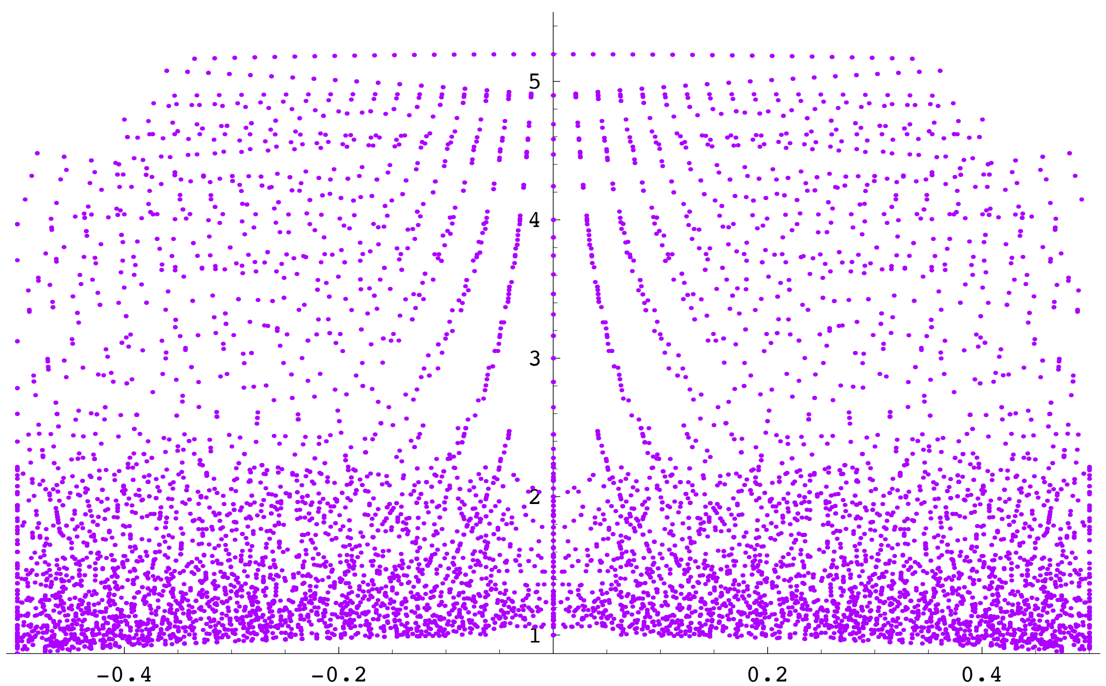


Work in progress [Cole,GS]

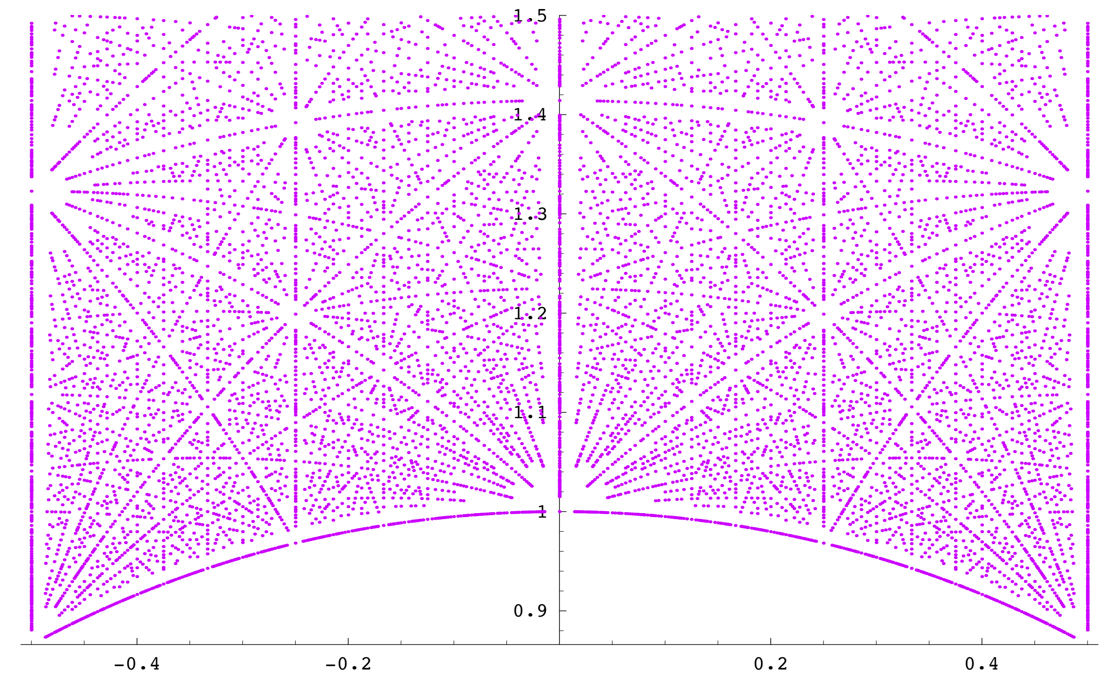
Flux Vacua on Symmetric T^6

- Factorizable $T^6 = (T^2)^3$ with equal complex structure $z_1 = z_2 = z_3 = z$
- Two complex moduli: complex structure modulus z and axio-dilaton τ.
- Number-theoretical methods were used to find distributions of vacua with W=0 and with discrete symmetries **[DeWolfe, Giryavets, Kachru,Taylor]**

Generic vacua on z-plane



W=0 vacua on z-plane



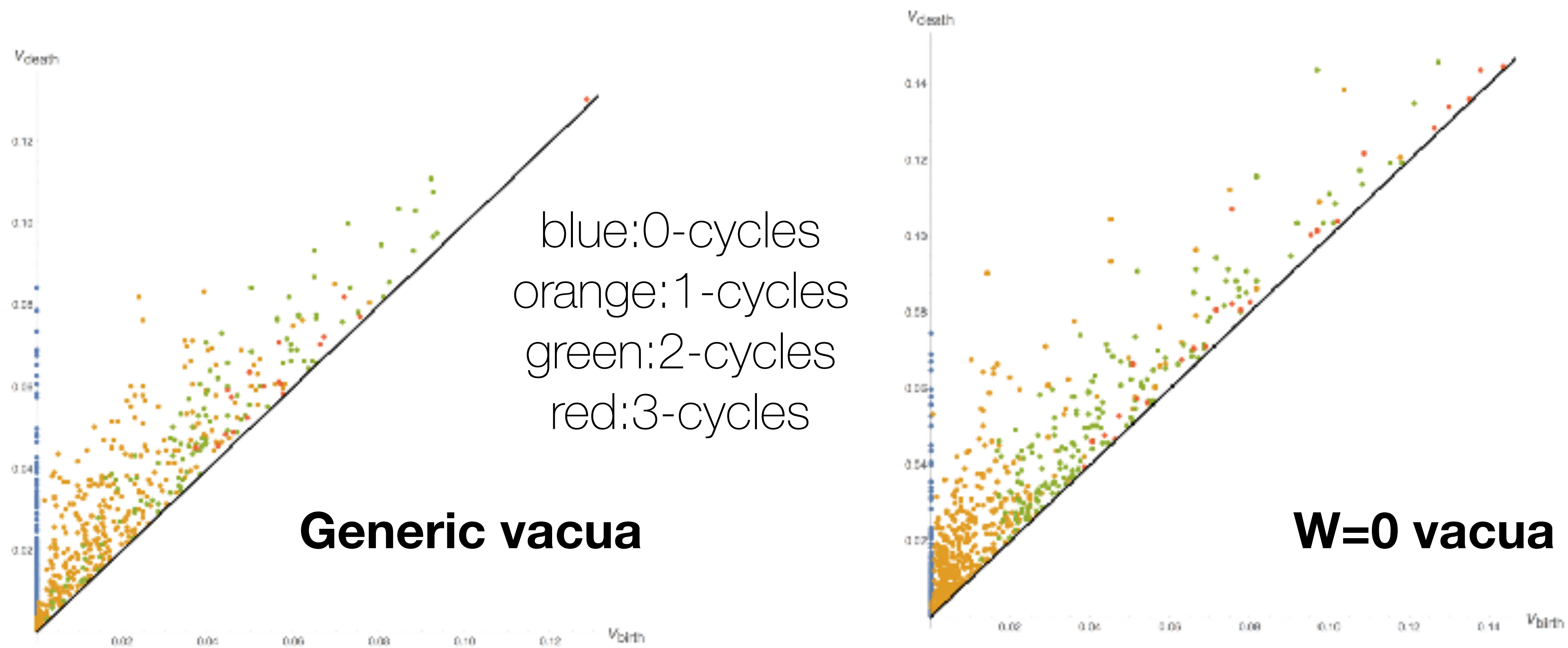
- How do “cuts” like restricting to W=0 vacua (e.g., discrete R-symmetry, motivated by **[Nelson, Seiberg]**) change the topology of distribution?

Flux Vacua on Symmetric T^6

- Reasonable expectation:

generic moduli dist topology	"topology" of cut	resulting topology
trivial	trivial	trivial
nontrivial	trivial	reduced complexity
trivial	nontrivial	nontrivial
nontrivial	nontrivial	complicated

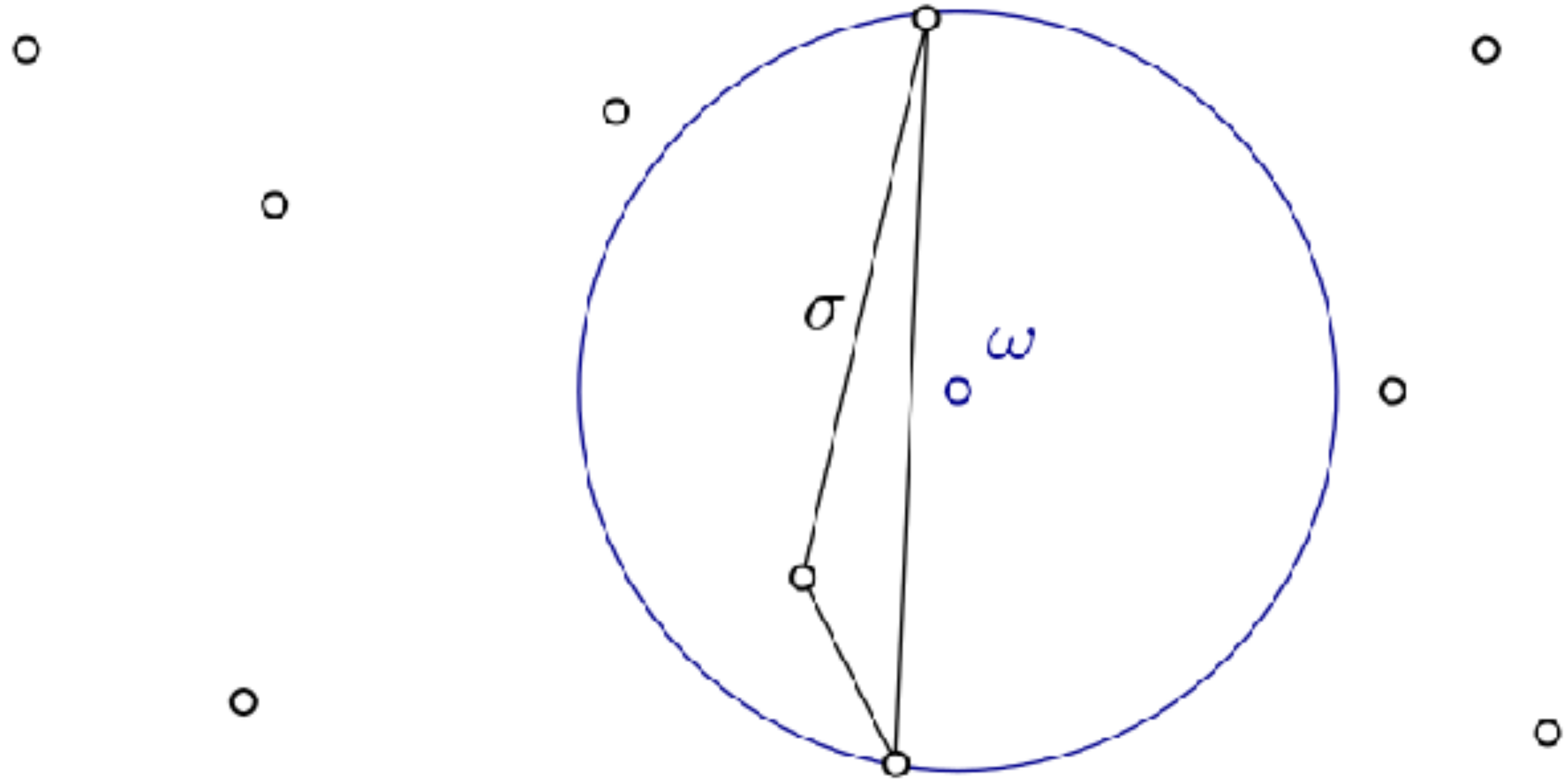
- Comparing persistent homology:

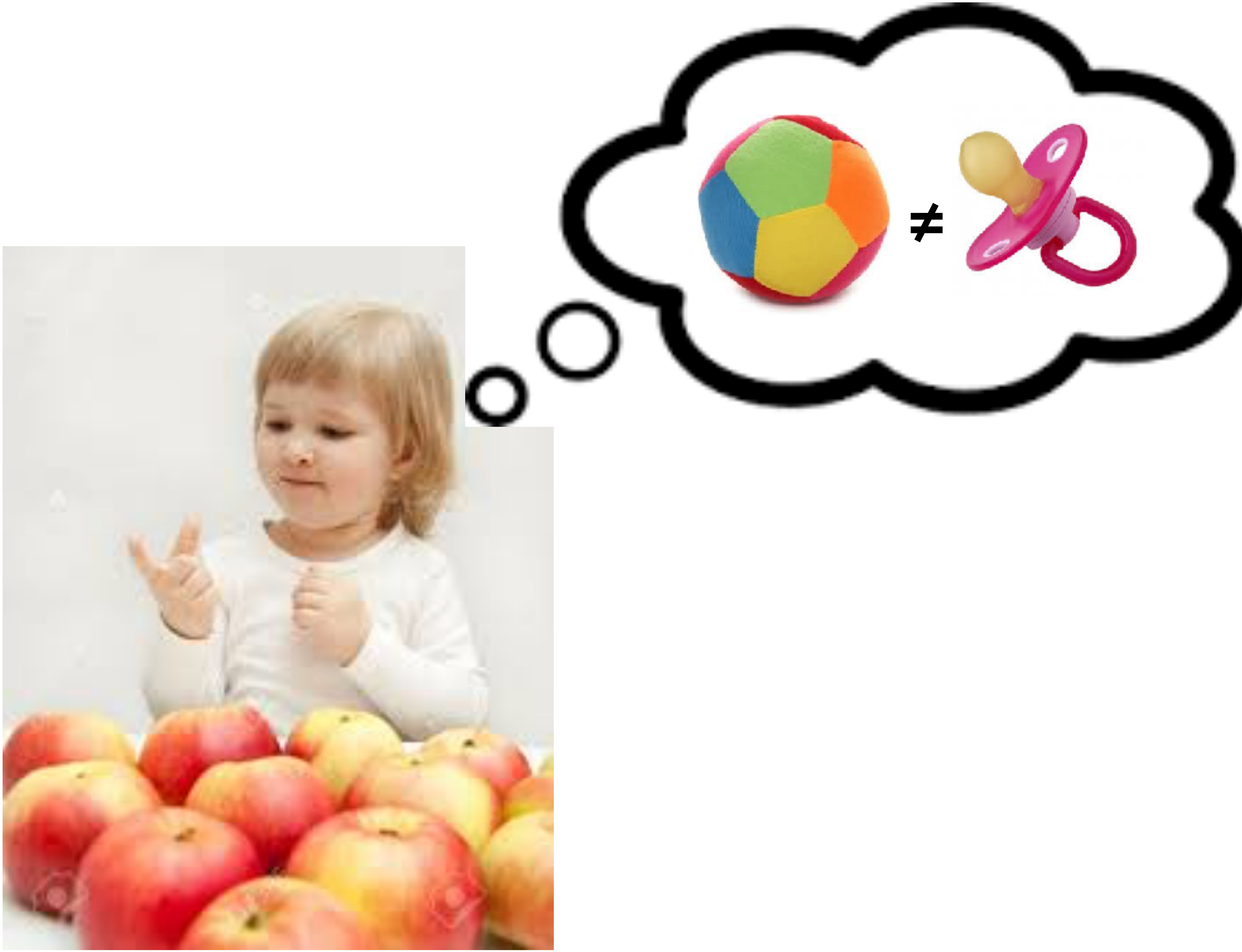


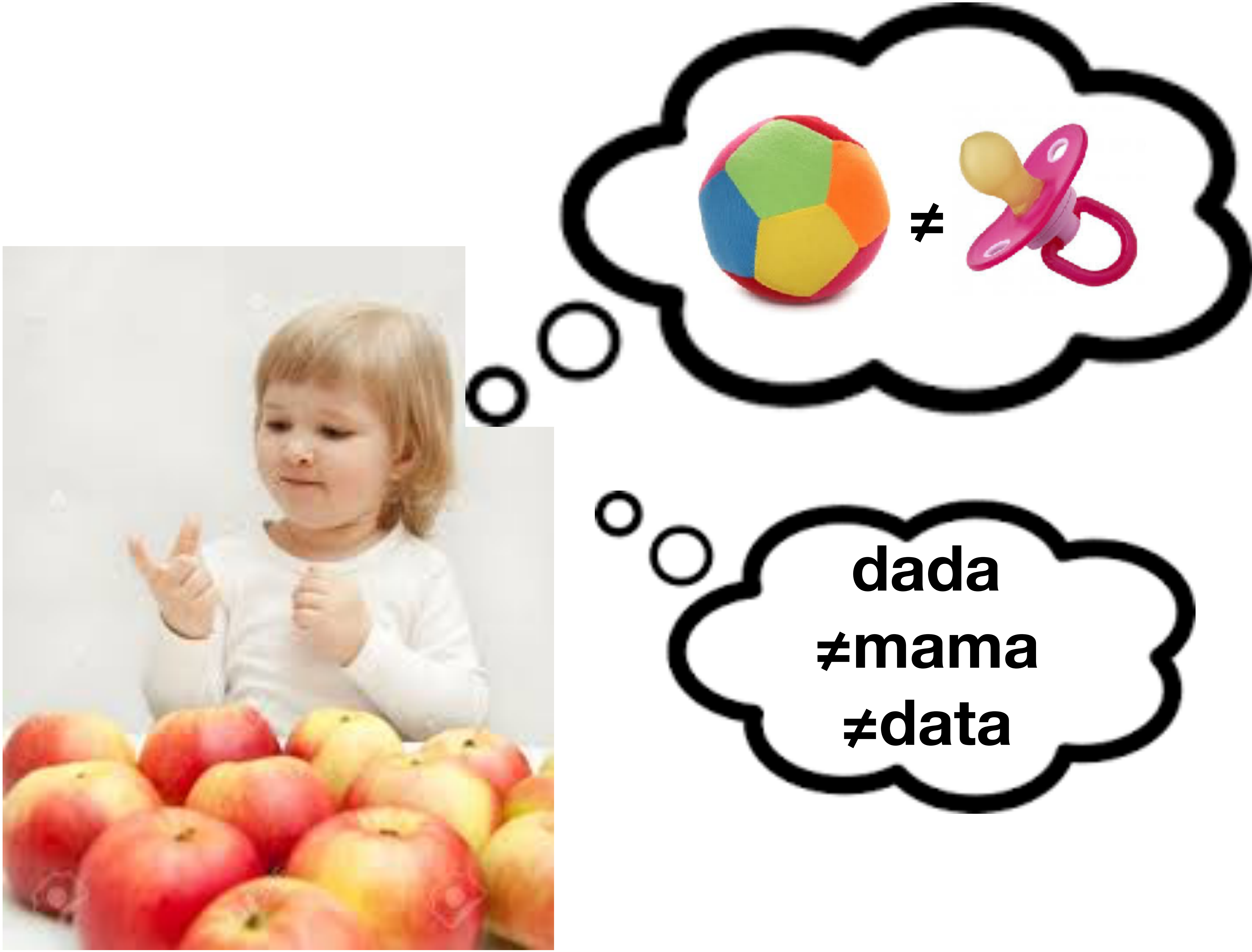
- W=0 cut adds complexity! Long-lived higher dimensional topological features differs from that for generic vacua.

Sampling in TDA

- We can't realistically include all 10^{500} vacua as vertices
- Can sample the topology via the **witness complex**:
  - From the entire point cloud Z, choose a ***landmark set L*** as the complex's vertices. Often chosen randomly or via sequential maxmin algorithm
  - Let $m_k(z)$ be the distance from some z∈Z to the (k+1)-nearest landmark point. Then, given filtration parameter ν, the simplex $[l_0 l_1 \dots l_k]$ is included in the witness complex if $\max\{d(l_0, z), d(l_1, z), \dots, d(l_k, z)\} \leq \nu + m_k(z)$






≠
dada
≠mama
≠data

Conclusions

Conclusions

- Applications of TDA to **cosmological datasets** and **string vacua**.
- Persistence diagrams strengthen constraints on local **non-Gaussianities**, and potentially other shapes & other observables.
- Techniques we developed can be applied to analyze the structure of string vacua. We performed initial study of simple flux vacua.
- Next step is to examine the **topology** of string vacua point clouds with desired features, supplementing earlier work on ***statistics***:
    - Enhanced symmetries **[DeWolfe, Giryavets, Kachru, Taylor], …**
    - Particle physics features **[Marchesano, GS, Wang];[Dienes];[Gmeiner, Blumenhagen, Honecker, Lust, Weigand], [Douglas, Taylor], …**
- Topology of Energy Landscape of String theory?
- String Landscape vs the Swampland?

Conclusions

- Applications of TDA to **cosmological datasets** and **string vacua**.
- Persistence diagrams strengthen constraints on local **non-Gaussianities**, and potentially other shapes & other observables.
- Techniques we developed can be applied to analyze the structure of string vacua. We performed initial study of simple flux vacua.
- Next step is to examine the **topology** of string vacua point clouds with desired features, supplementing earlier work on ***statistics***:
  - Enhanced symmetries **[DeWolfe, Giryavets, Kachru, Taylor], …**
  - Particle physics features **[Marchesano, GS, Wang];[Dienes];[Gmeiner, Blumenhagen, Honecker, Lust, Weigand], [Douglas, Taylor], …**
- Topology of Energy Landscape of String theory?
- String Landscape vs the Swampland?