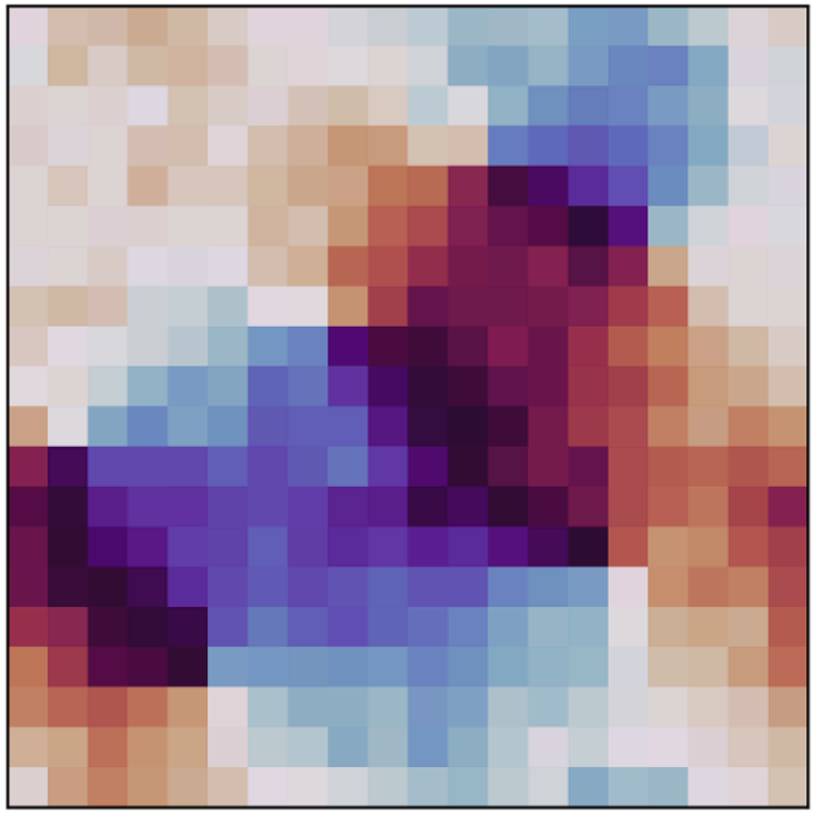
Interpretable features from topology: application to phase transitions

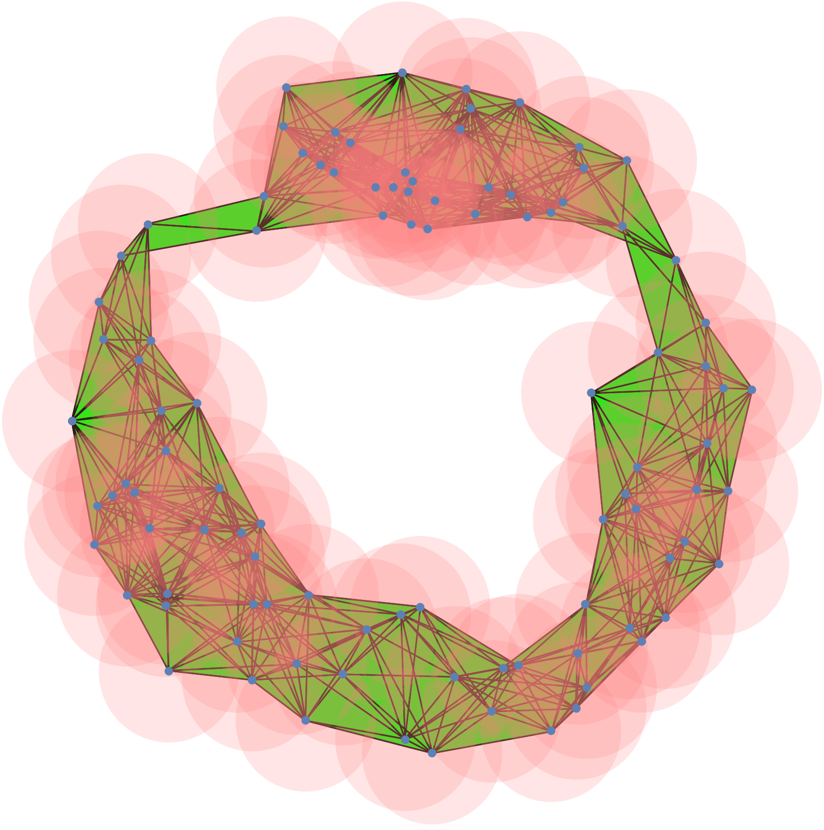
Alex Cole
Univ. of Amsterdam, GRAPPA and ITFA

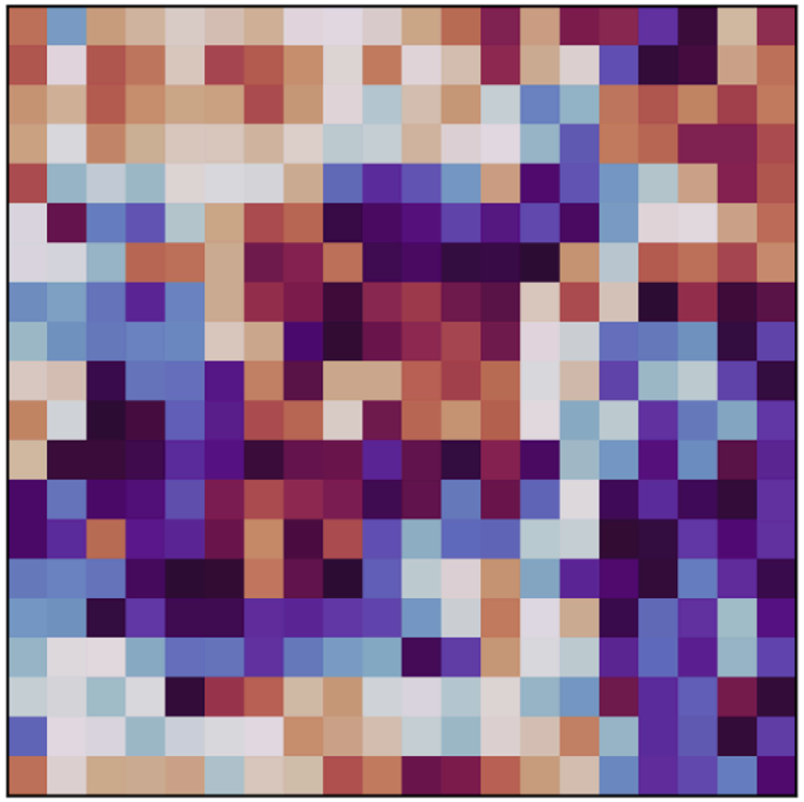
string_data 2020, CERN

Based on 2009.14231 w/ G. Loges, G. Shiu.

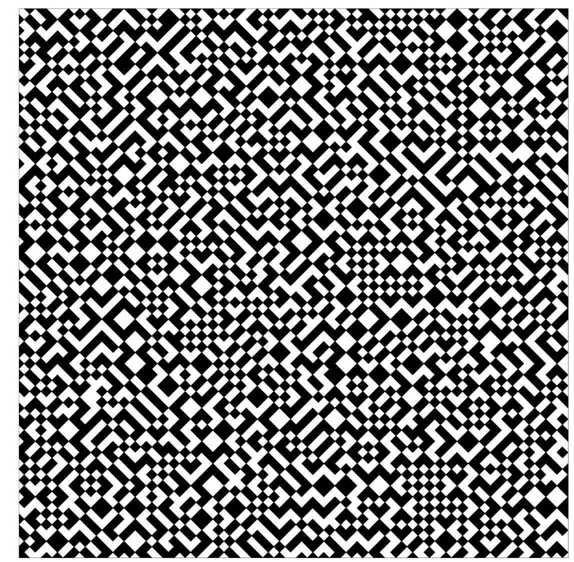
Related work:
1712.08159 (w/Shiu, JCAP)
1812.06960 (w/Shiu, JHEP)
2009.04819 (w/Biagetti, Shiu)







(T, P, λ_i) $\xrightarrow{\text{draw sample}}$



Q: what is the system's phase structure?

Q: given a sample, can we determine the phase?

Q: what is the interpretation of differences between phases?

More motivation

- Already in introductory physics courses students learn that solving problems is much easier once a “good” coordinate system is chosen. Alternatively, we can focus on features that reflect relevant symmetries, etc.

- How can we automatically extract relevant features from data?

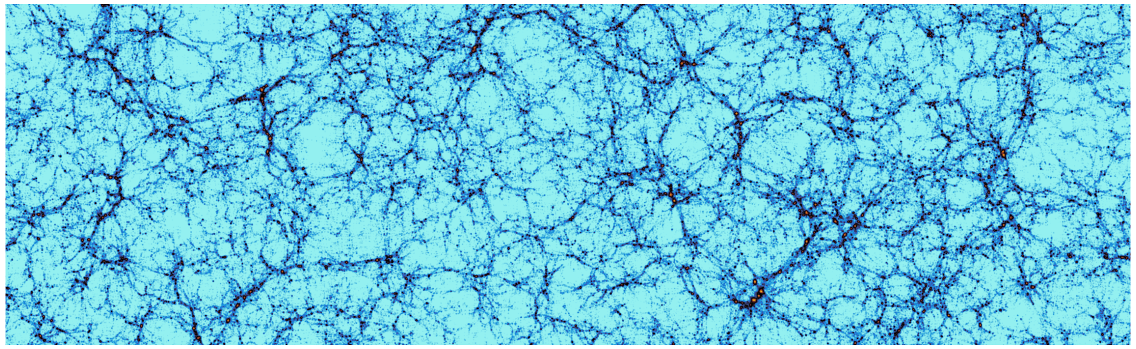
Topological Data Analysis

- Recognize the **shape** of data in a **systematic** way
  - Generalize to **high-dimensional** data sets in **abstract** spaces
- Main tools: **persistent homology**, Mapper

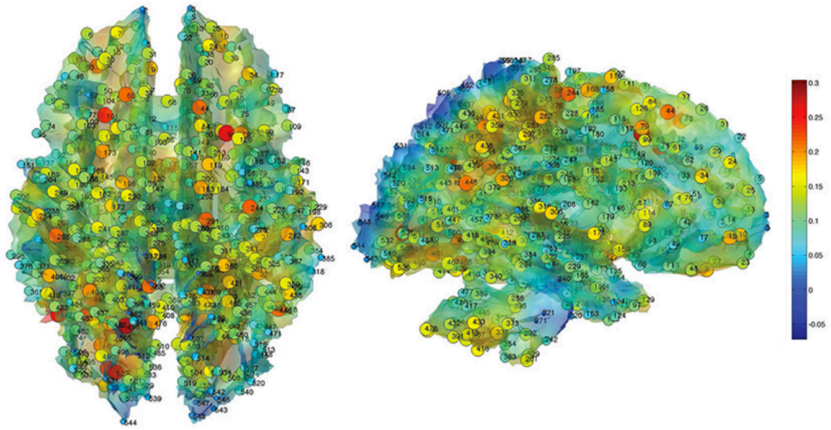
[AC, Shiu; Biagetti,AC,Shiu]

- Used in **cosmology**, neuroscience, **string theory**, drug design, sensor networks, image analysis, virology, computer vision, materials science, sports analytics, computational biology, QCD, dynamical systems, …

[AC, Shiu]



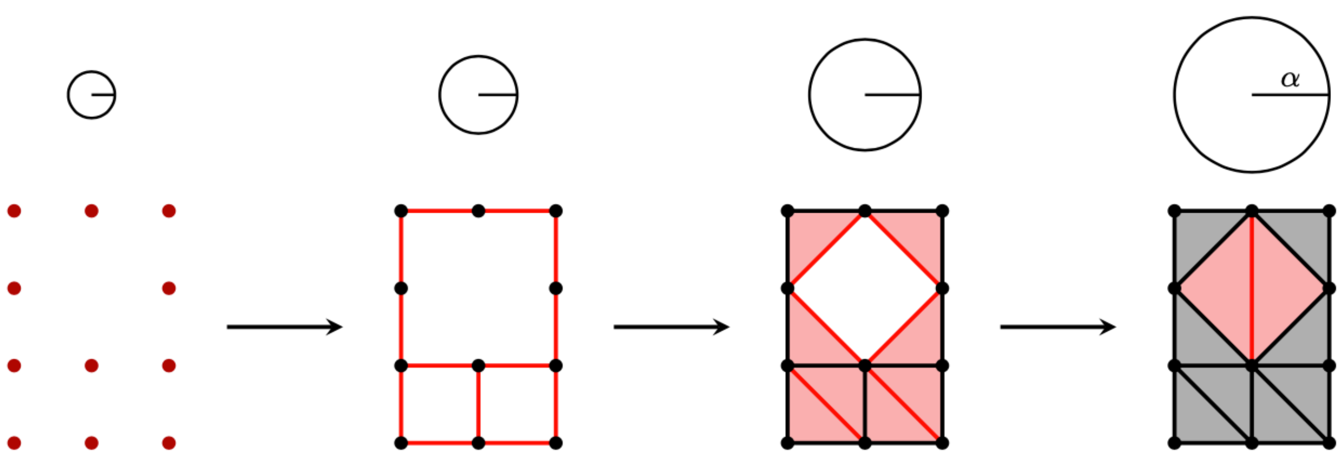
[Biagetti, AC, Shiu]

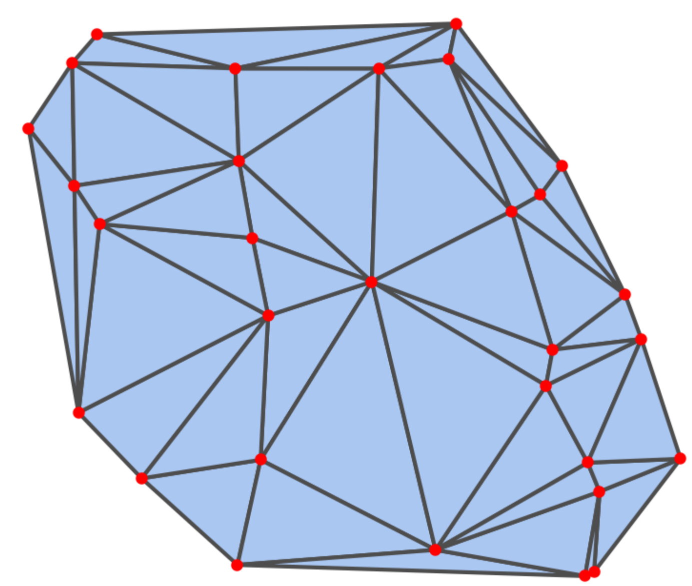


arXiv:1409.0177

α-filtrations (discrete spins)

- For discrete spins, put a point at every lattice site agreeing with the *majority*.
- For this point cloud, perform an α-filtration, probing multiscale topology of majority spins.

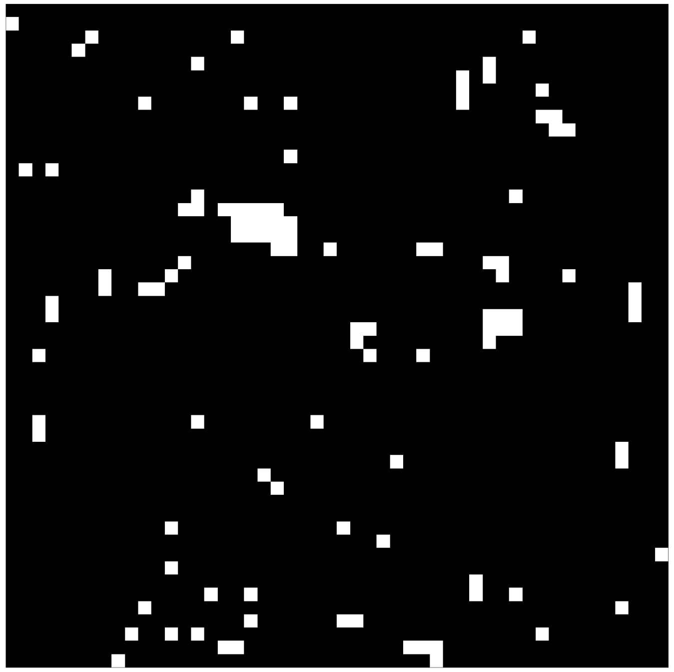




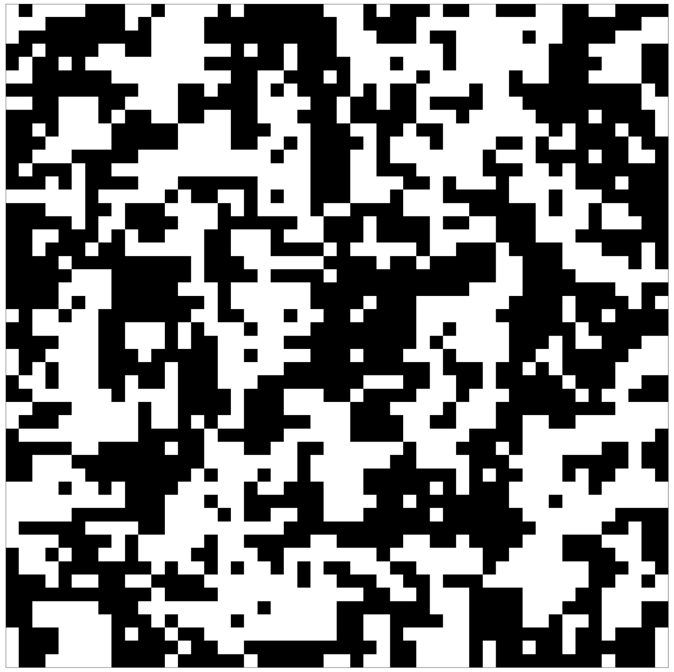
Example: Ising model

- $$H_{\mathrm{Is}} = -\sum_{\langle i,j \rangle} s_i s_j, \quad s_i \in \{-1,1\}$$

- Spontaneous magnetization below $T_c \approx 2.27$, breaking $\mathbb{Z}_2$ symmetry.



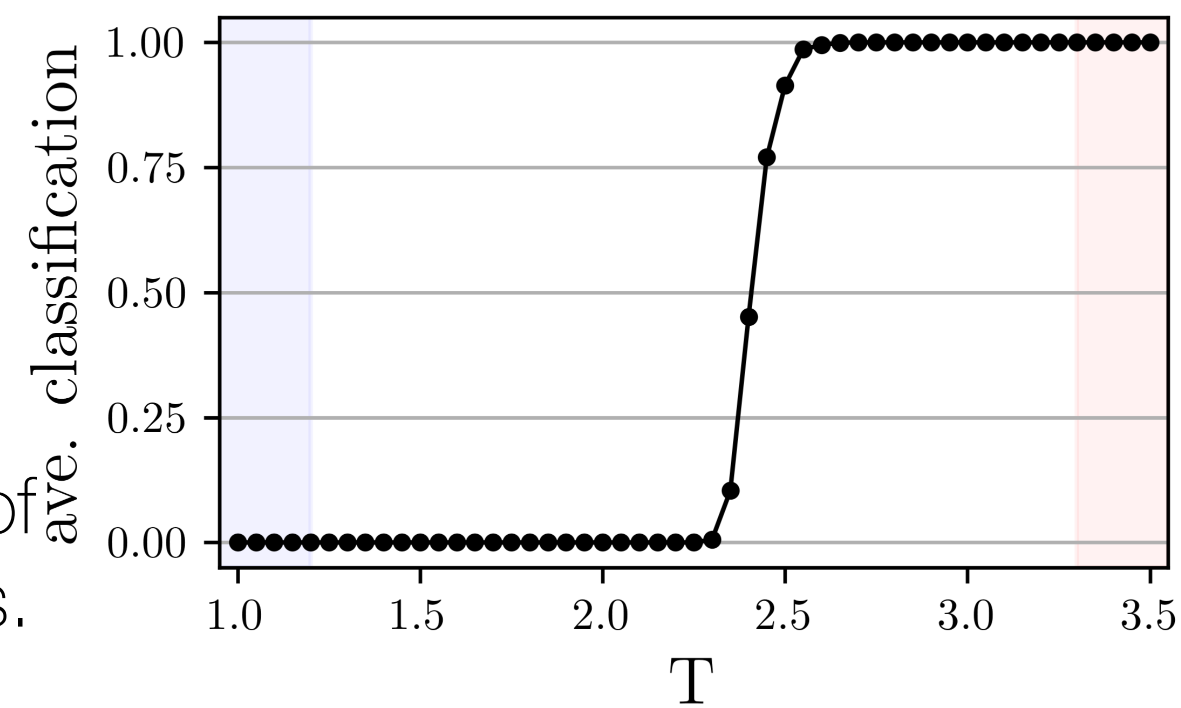
$T = 2.0$



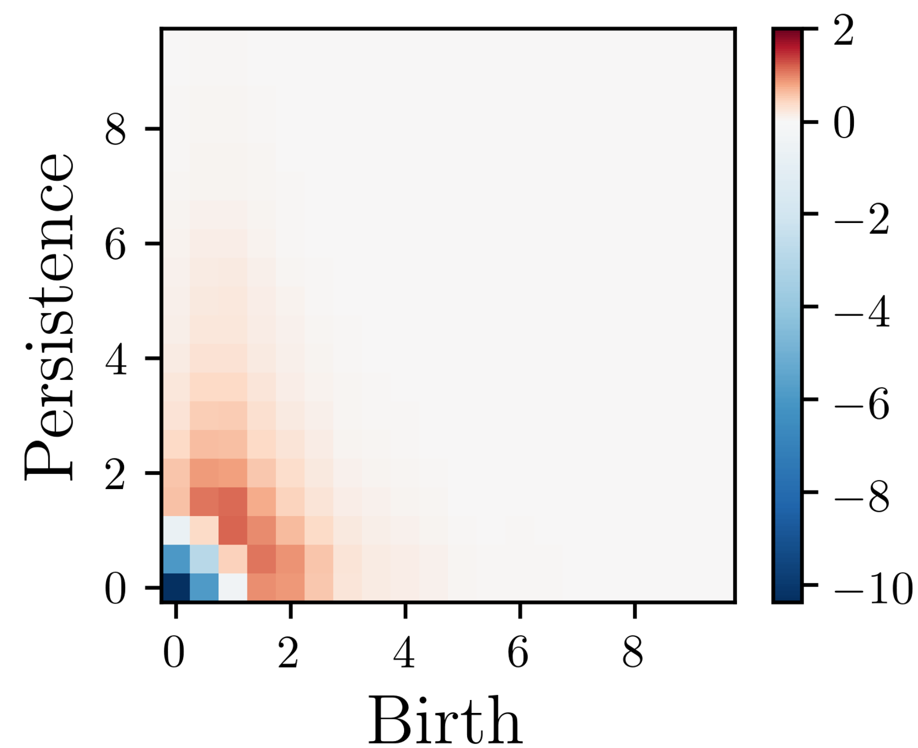
$T = 3.5$

Ising phase classification

- For each $T \in \{1.00, 1.05, \ldots 3.50\}$, generate 1000 samples (MCMC with Wolff cluster update).



- Train logistic regression using 25% of simulations at extreme temperatures.
    - The trained LR estimates $T_c \approx 2.37$ (lattice effects)



- LR coefficients identify the **magnetization** (features at lattice scale) as order parameter.

TL;DR: a “smart” dimensional reduction allows simple and interpretable classification of physical states.

See Gary’s talk tomorrow for more details and other physics applications of TDA!

Fun topics in topology and ML

- NeurIPS 2020 workshop “TDA and Beyond” — https://tda-in-ml.github.io/. *Some* personal highlights:
    - Filtered hierarchical clustering via multiparameter persistent homology (Bauer et al.)
    - Topological loss function enhances image segmentation, generation. Topological regularization of decision boundaries to improve training with noisy labels (Chen et al.)
    - Topology of learning (Carlsson et al.; Wang et al.)
    - Probabilistic aspects of Mapper and persistent homology (Carrière et al.; Hiraoka et al.)
    - Adventures in generalized graph laplacians (Ghrist et al.)

General questions

- How can theoretical priors from string theory inform ML architectures? Duality, modularity, ….

- What are "good coordinates" on the space of compactifications? Can these be identified explicitly? cf. information geometry

- What (constrained) optimization techniques are appropriate for navigating the landscape? Do they teach us something about early universe dynamics or vice versa?

  - Advances in DL-assisted combinatorial optimization

Thanks!