Unsupervised learning universal critical behavior via the intrinsic dimension

T. Mendes-Santos*,[1] X. Turkeshi*,[1,2,3] M. Dalmonte,[1,2] and Alex Rodriguez[1]

[1] *The Abdus Salam International Centre for Theoretical Physics, strada Costiera 11, 34151 Trieste, Italy*
[2] *SISSA,via Bonomea, 265, 34136 Trieste, Italy*
[3] *INFN,via Bonomea, 265, 34136 Trieste, Italy*

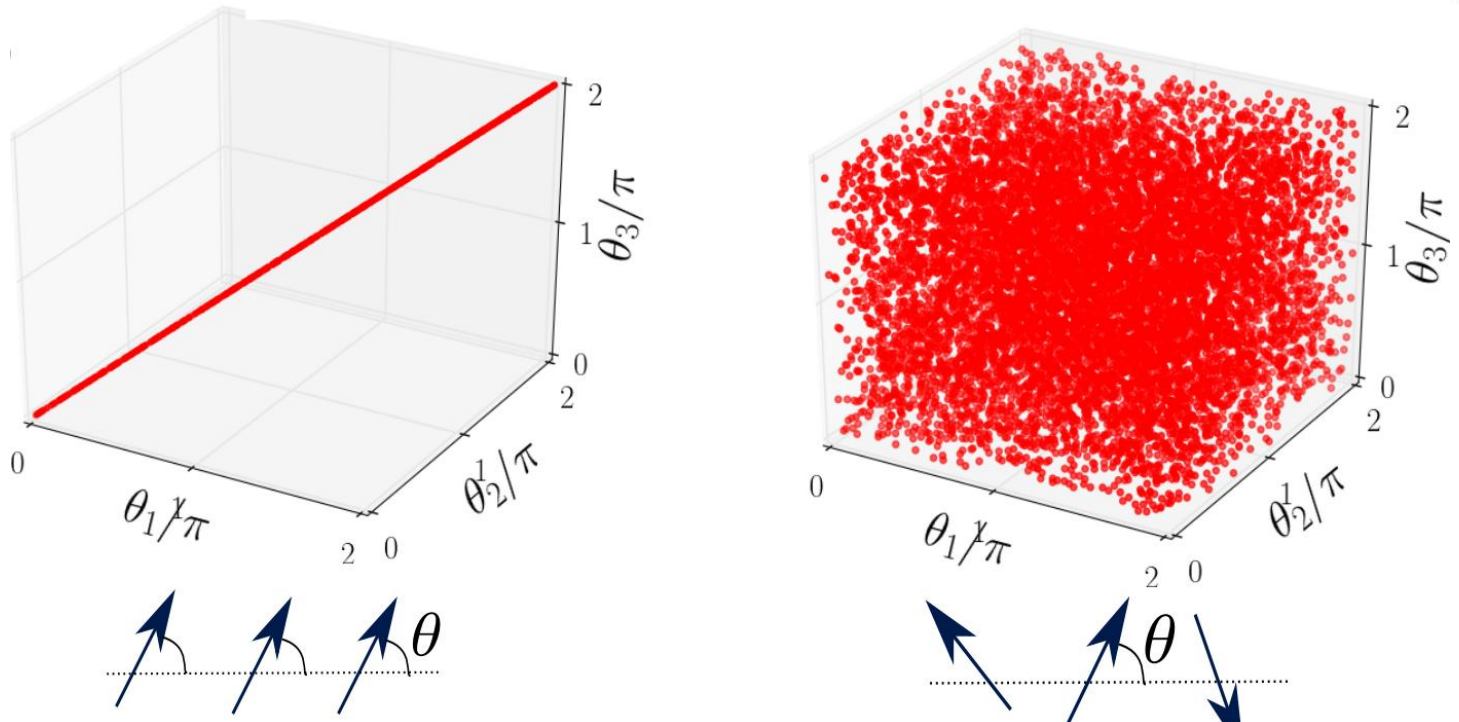
A novel way to identify phase transition through machine learning

arXiv number: 2006.12953

General Idea

If sampling a 3-sites XY model

$$Z = \sum_{\vec{X}} e^{-\beta E(\vec{X})}$$



Zero temperature

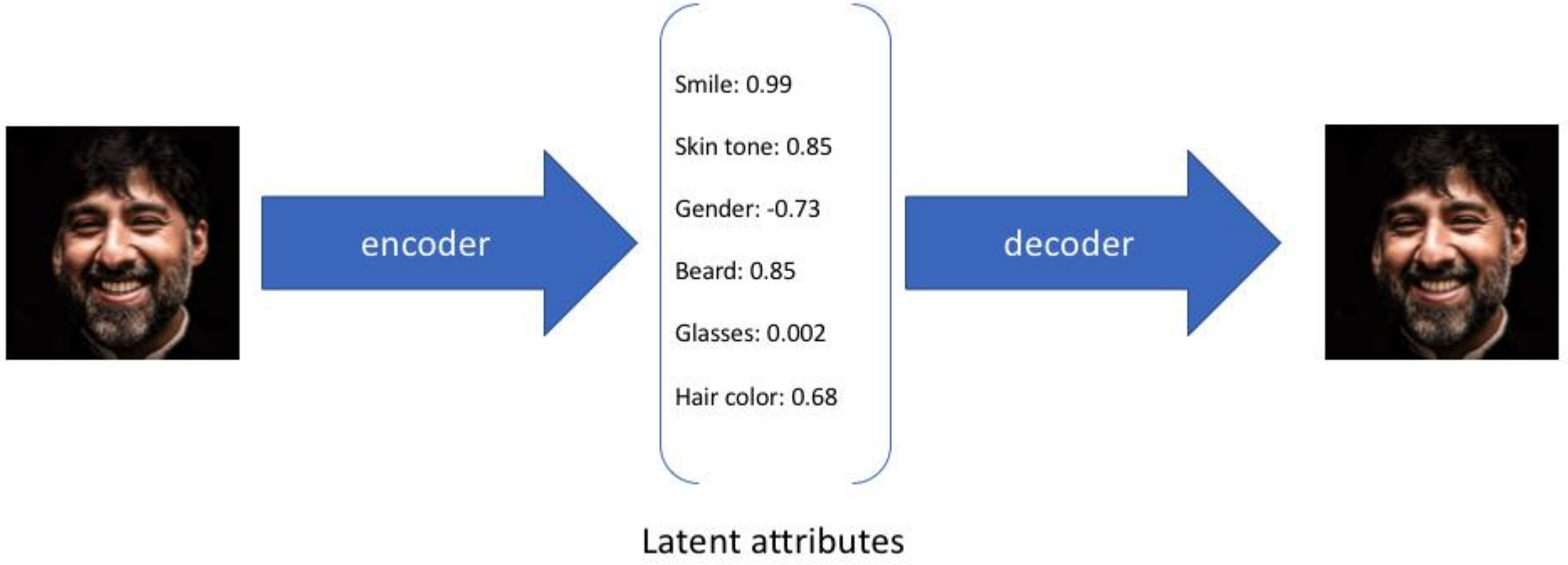
High temperature

$$E(\{\vec{\theta}\}) = -J \sum_{\langle i,j \rangle} \cos(\theta_i - \theta_j)$$

$$\vec{\theta} = (\theta_1, \theta_2, \theta_3)$$

Dimension of data varies as temperature increases

General Idea

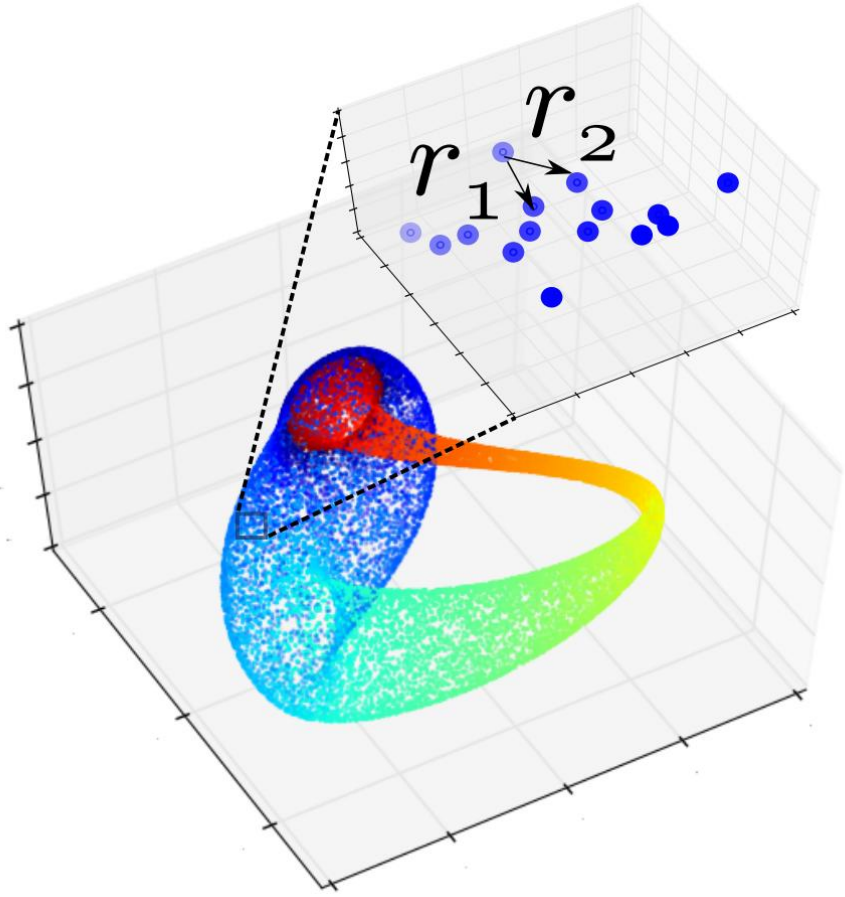


In data science, data usually has a much lower dimensionality than its original space

➡ learning the actual dimension of data is important for pattern recognization

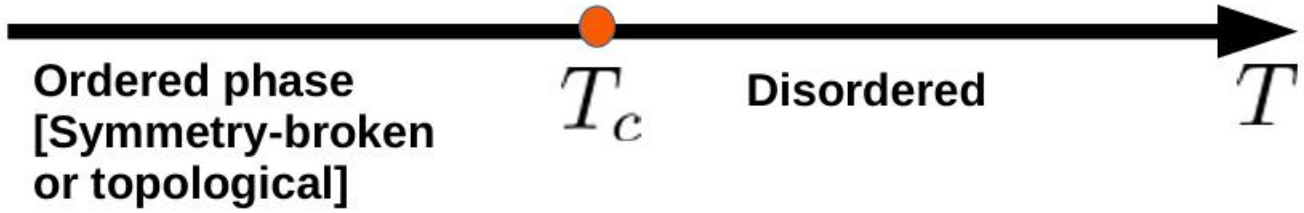
General Idea

Intrinsic Dimension:
Data set lies in a manifold whose I_d is lower than the number of coordinates



$I_d = 2$

General Idea



Phase transition in configuration space?

ID of data emerge at the vicinity of phase transition

(Does the ice also melt in data structure?)

Extract information of system from raw data - Universal property?

Methods – Model and Data

- Three types of model are considered in the article:

- Second order PT: 2D Ising, q=3 Pott model

$$E(\vec{s}) = -\sum_{\langle i,j \rangle} s_i s_j$$

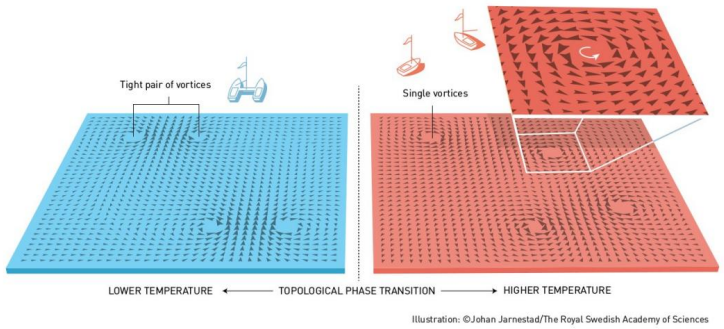
- First order PT: q=8 Pott model

$$E(\vec{\sigma}) = -\sum_{\langle i,j \rangle} \delta_{\sigma_i,\sigma_j}$$

2D-Ising configurations

- Topological PT: 2D XY model

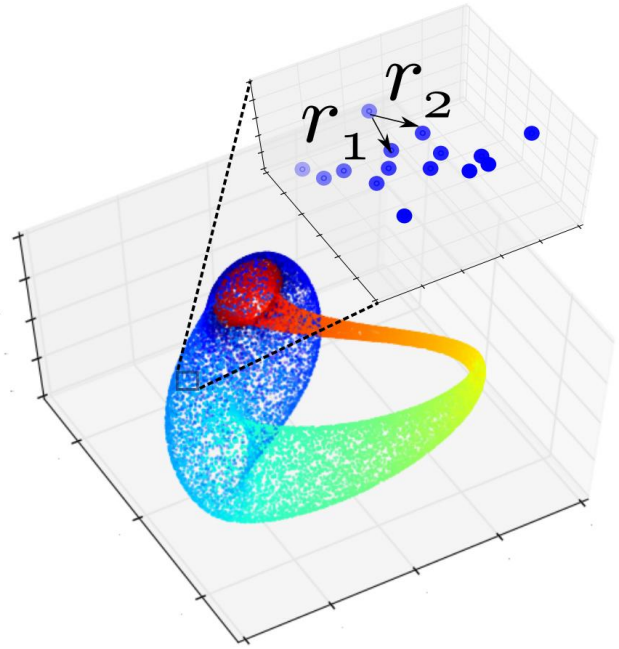
$$E(\vec{\theta}) = -\sum_{\langle i,j \rangle} \vec{S}_i \cdot \vec{S}_j$$



- Using MCMC simulations for thermal equilibrium configurations

Methods - I_d Estimation

- Two-NN method
  - Estimates I_d using first and second nearest-neighbor distances.
  - Computes ratio $\mu = r_2/r_1$, fits distribution to derive I_d.
  - Handles nonlinear, nonuniform manifolds.



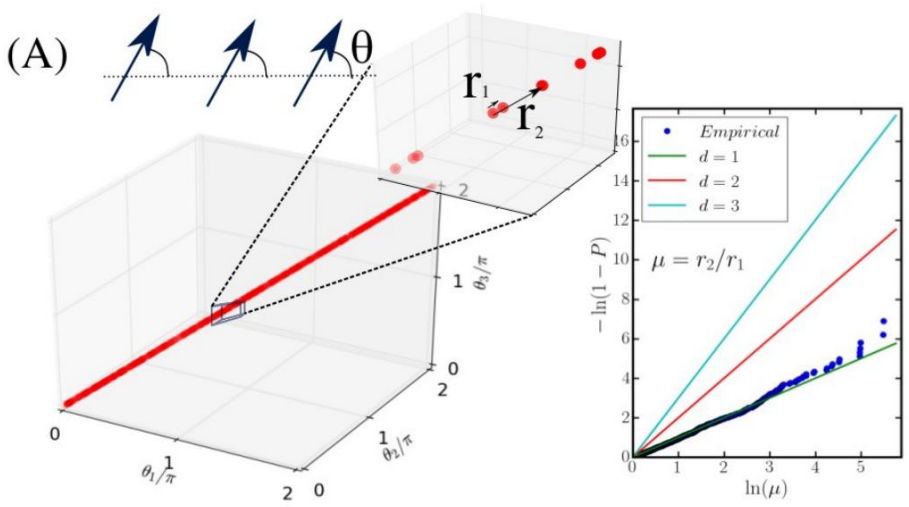
Probability distribution function

$$f(\mu) = I_d \mu^{-1-I_d}$$

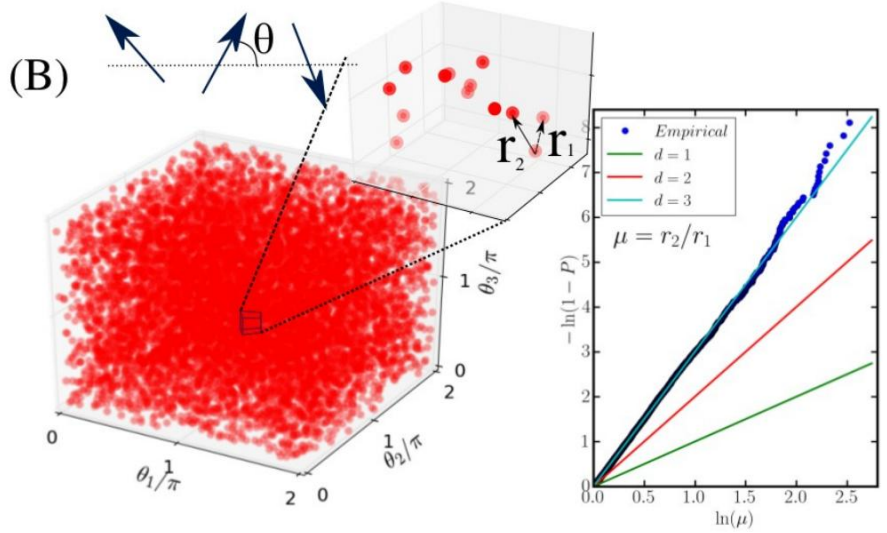
Elena Facco et. al. Scientific Reports (2017)

Assumption: data set is locally uniform in density

Methods - Intrinsic Dimension Estimation



$I_d \approx 1$



$I_d \approx N_s$

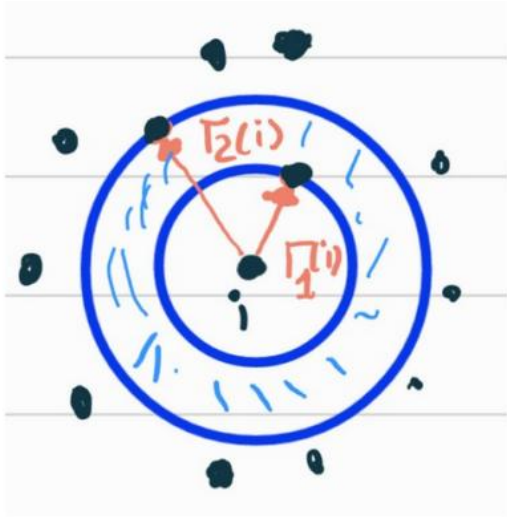
I_d (TWO-NN)

- Local feature of configuration space
- Depends on the typical value of distances [scale-dependent quantity]

N_r: Number of points in the data sets

Consider a 2-dimensional space with density λ first. The number of points $N(A)$ falling into region A is distributed as a Poisson variable with parameter $\lambda\mu(A)$, with $\mu(A)$ being the measure of A:

$$P(A \text{ contains exactly } n \text{ points}) \doteq P(n, A) = \frac{(\lambda\mu(A))^n}{n!} e^{-\lambda\mu(A)}$$



The probability of having no points in A is given by:

$$P(0, A) = e^{-\lambda\mu(A)}$$

The probability of first distance r_1 to fall in an infinitesimally small annulus:

$$\begin{aligned} P(d_1 \in C_{r_1, r_1+dr_1}) &= P(N(B_{o,r_1}) = 0, N(C_{r_1,r_1+dr_1}) \geq 1) \\ &= P(N(B_{o,r_1}) = 0) P(N(C_{r_1,r_1+dr_1}) \geq 1) \\ &= P(N(B_{o,r_1}) = 0)(1 - P(N(C_{r_1,r_1+dr_1}) = 0) \\ &= e^{-\lambda r_1^2 \pi}(1 - e^{-\lambda\pi r_1 dr_1}). \end{aligned}$$

➡ $P(d_1 \in C_{r_1,r_1+dr_1}) \sim e^{-\lambda r_1^2\pi} 2\pi\lambda r_1 dr_1$ (small r_1)

The volume of hypersferical shell enclosed between two successive neighbors $l-1$ and l is given by

$$\Delta v_l = \omega_d (r_l^d - r_{l-1}^d)$$

So as discussed in the last page

$$P(\Delta v_l \in [v, v + dv]) = \rho e^{-\rho v} dv.$$

Set R be the ratio $\frac{\Delta v_i}{\Delta v_j}$, the probability distribution (pdf) of R

$$\begin{aligned} & P(R \in [\overline{R}, \overline{R} + d\overline{R}]) \\ &= \int_0^\infty dv_i \int_0^\infty dv_j \rho^2 e^{-\rho(v_i+v_j)} 1_{\left\{\frac{v_j}{v_i} \in [\overline{R}, \overline{R}+d\overline{R}]\right\}} \\ &= d\overline{R} \frac{1}{(1+\overline{R})^2}, \end{aligned}$$

Define $\mu = \frac{r_2}{r_1} \geq 1$, then

$$R = \mu^d - 1$$

$$
\begin{aligned}
P(r_2 \mid r_1) \doteq & P(\text{the second nearest neighbour is at a distance } r_2 \text{ given that the first is at a distance } r_1) \\
= & P(\text{the second nearest neighbour is at a distance } r_2 \mid N(B_{o,r_1}) = 0, N(C_{r_1,r_1+dr_1}) \geq 1) \\
= & P(N(C_{r_1,r_2}) = 0, N(C_{r_2,r_2+dr_2}) \geq 1 \mid N(B_{o,r_1}) = 0, N(C_{r_1,r_1+dr_1}) \geq 1) \\
= & P(N(C_{r_1,r_2}) = 0 \mid N(B_{o,r_1}) = 0, N(C_{r_1,r_1+dr_1}) \geq 1) \cdot \\
& \cdot P(N(C_{r_2,r_2+dr_2}) \geq 1 \mid N(B_{o,r_1}) = 0, N(C_{r_1,r_1+dr_1}) \geq 1). \\
\sim & e^{-\lambda\pi(r_2^2 - r_1^2)} 2\lambda\pi r_2 dr_2
\end{aligned}
$$

And the joint probability $P(r_1, r_2)$

$$P(r_1, r_2) = P(r_2 \mid r_1)P(r_1) \sim e^{-\lambda\pi r_2^2}(2\lambda\pi)^2 r_1 r_2 dr_1 dr_2.$$

After doing a similar integration we get

$$f(\mu) = d\mu^{-d-1} 1_{[1,+\infty]}(\mu),$$

The cumulative distribution (cdf) is obtained by an integration over μ

$$F(\mu) = (1 - \mu^{-d}) 1_{[1,+\infty]}(\mu)$$

Methods - Intrinsic Dimension Estimation

- I_d can be obtained through the following steps:

1. For each point i of the data set $(i = 1, 2, ..., N_r)$, compute its first- and second-nearest neighbor, $r_1(i)$, $r_2(i)$, respectively.
2. For each point i, compute the ratio $\mu_i = r_2(i)/r_1(i)$.
3. The empirical cumulate is defined as $P^{\text{emp}}(\mu) = i/N_r$, while the values of μ_i are sorted in an ascending order through a permutation, i.e., $(\mu_1, \mu_2, ...\mu_{N_r})$, where $\mu_i < \mu_j$, for $i < j$.
4. Finally, the resulting $S = \{(\ln(\mu), -\ln[1 - P^{\text{emp}}(\mu)]\}$ are fitted with a straight line passing through the origin. The slope of this line is equal to I_d (see Eq.(2)).

How to define distance between configuration points?

For Ising model & Pott model:

- Hamming distance (How many spins are different)

For BKT model:

- Euclidean distance

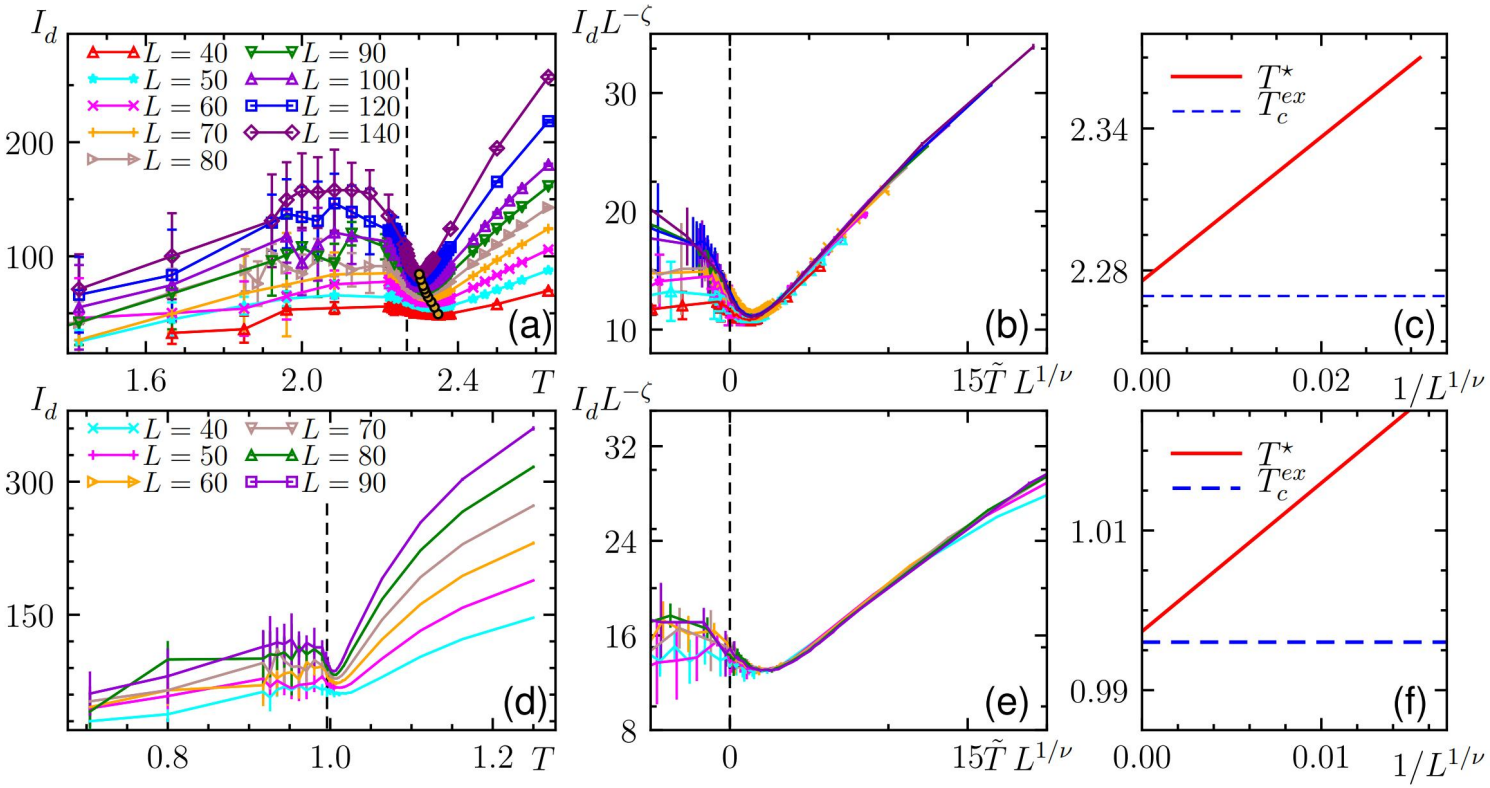
$$r(\vec{\theta^i}, \vec{\theta^j}) = \sqrt{2 \sum_{k=1}^{N_s} \left(1 - \vec{S}_k^i \cdot \vec{S}_k^j\right)}.$$

- Advantages
  - Overcomes limitations of PCA and Isomap for complex data.
  - Directly analyzes raw Monte-Carlo configuration.

Results – 2nd Order Phase Transition

2D Ising

$q = 3$ Pott model



I_d exhibits a local minimum at T^*

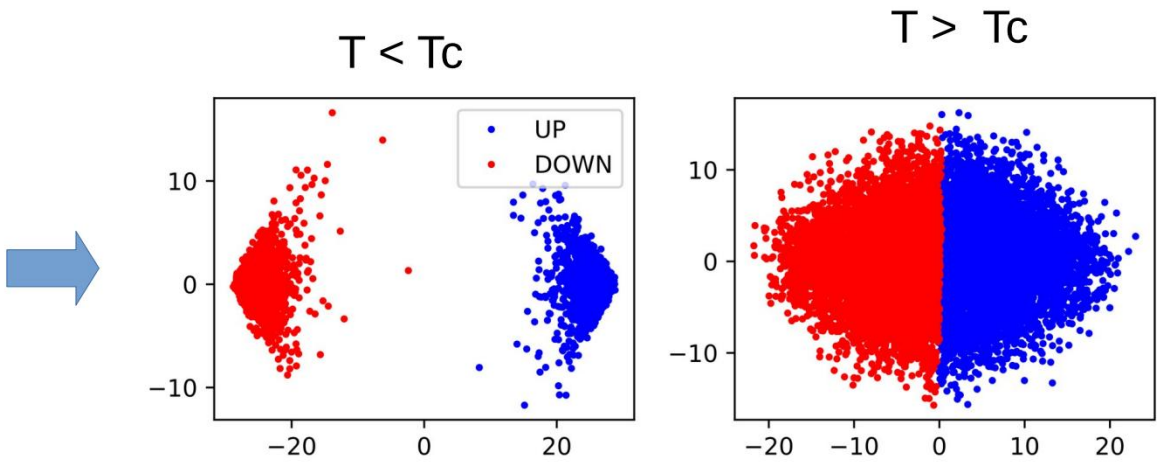
Finite size scaling

$$Q(t, L) = L^{\sigma} f(\xi / L)$$

$$T^*(L) - T_c \sim \frac{1}{L^{1/\nu}}$$

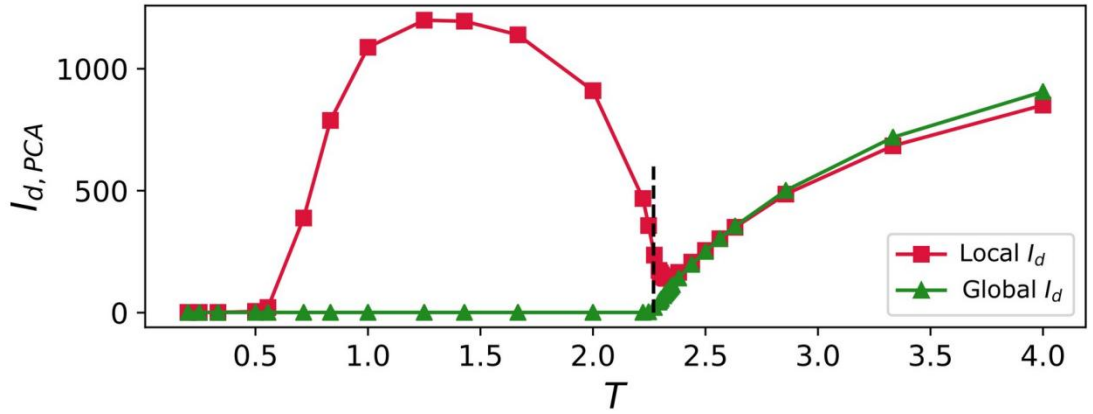
Results – Compared to PCA

Projection of the Ising data set in the two leading PC

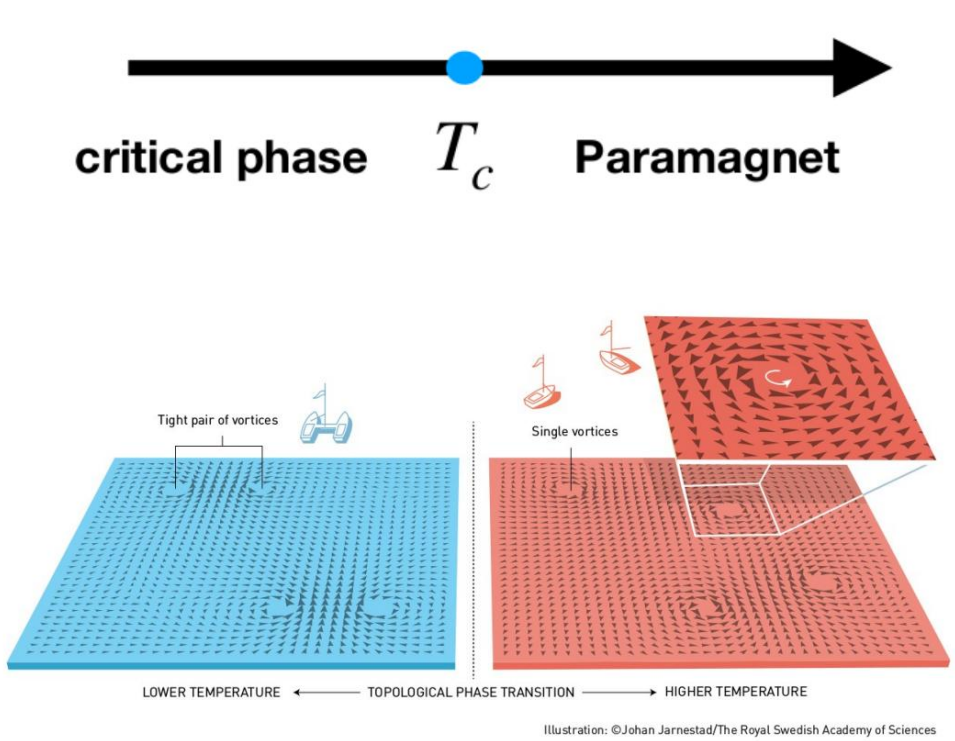


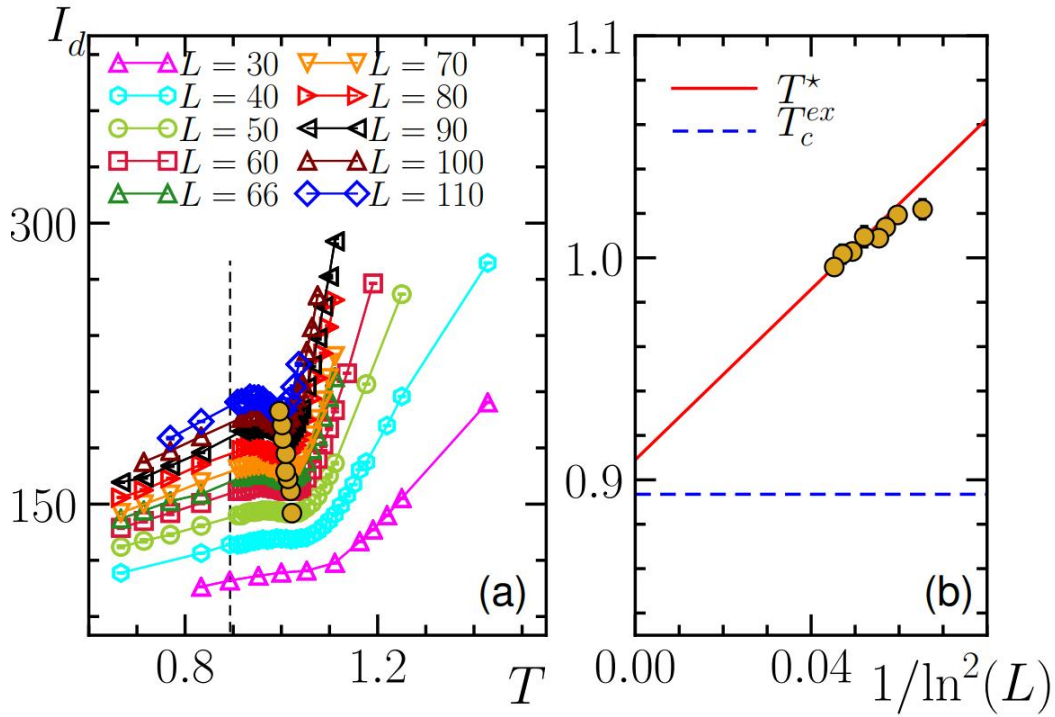
ID obtained with PCA

$$\sum_{n=1}^{I_{d,\mathrm{PCA}}} \tilde{\lambda}_n \approx f,$$



Results – BKT Transition

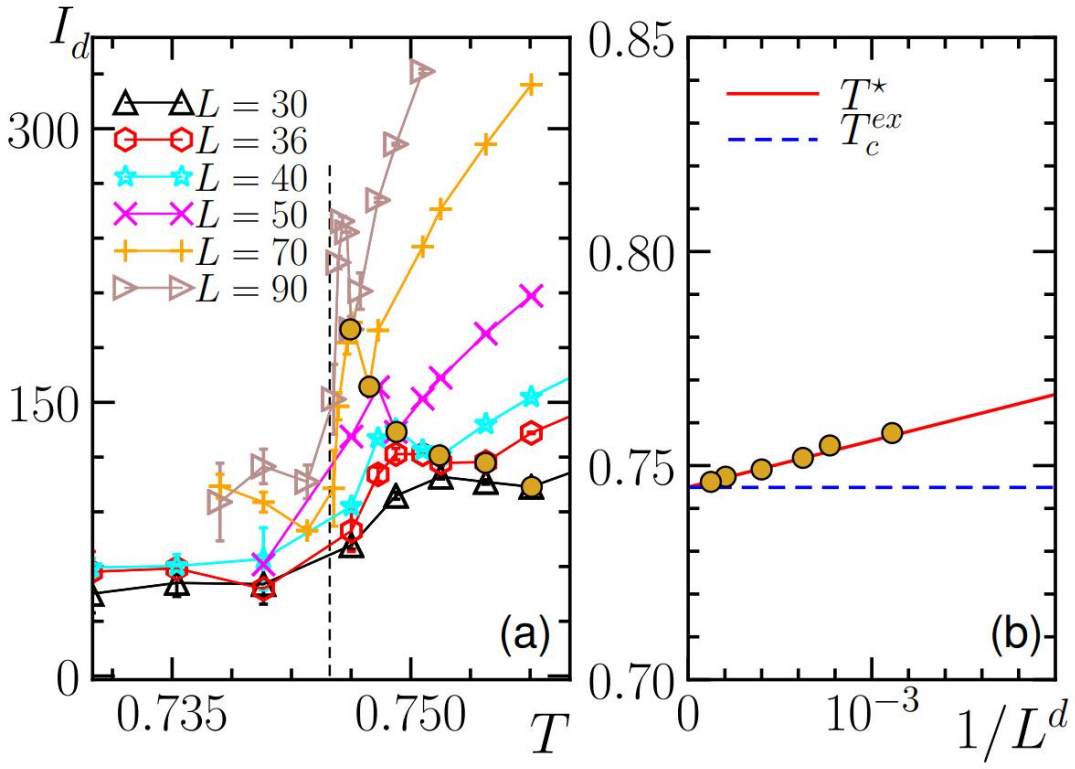




I_d exhibit a local minimum at T^*

Finite size scaling $\xi \sim \exp\left(\frac{a}{\sqrt{T - T_c}}\right)$

Results – 1st Order Phase Transition



- I_d peak at T_c from two-phase coexistence (metastability).