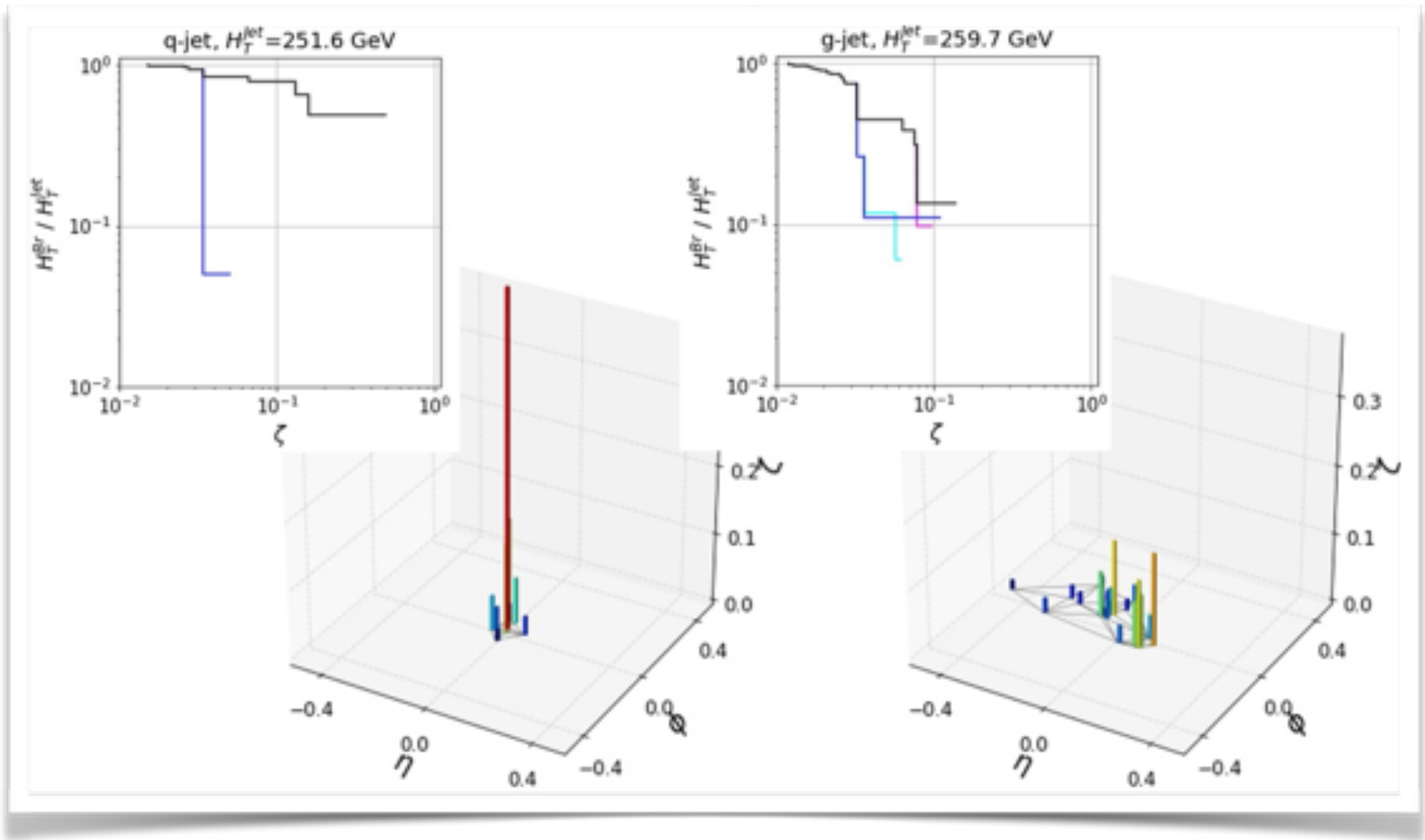
The 2020 International Workshop on the High Energy CEPC, Oct. 2020



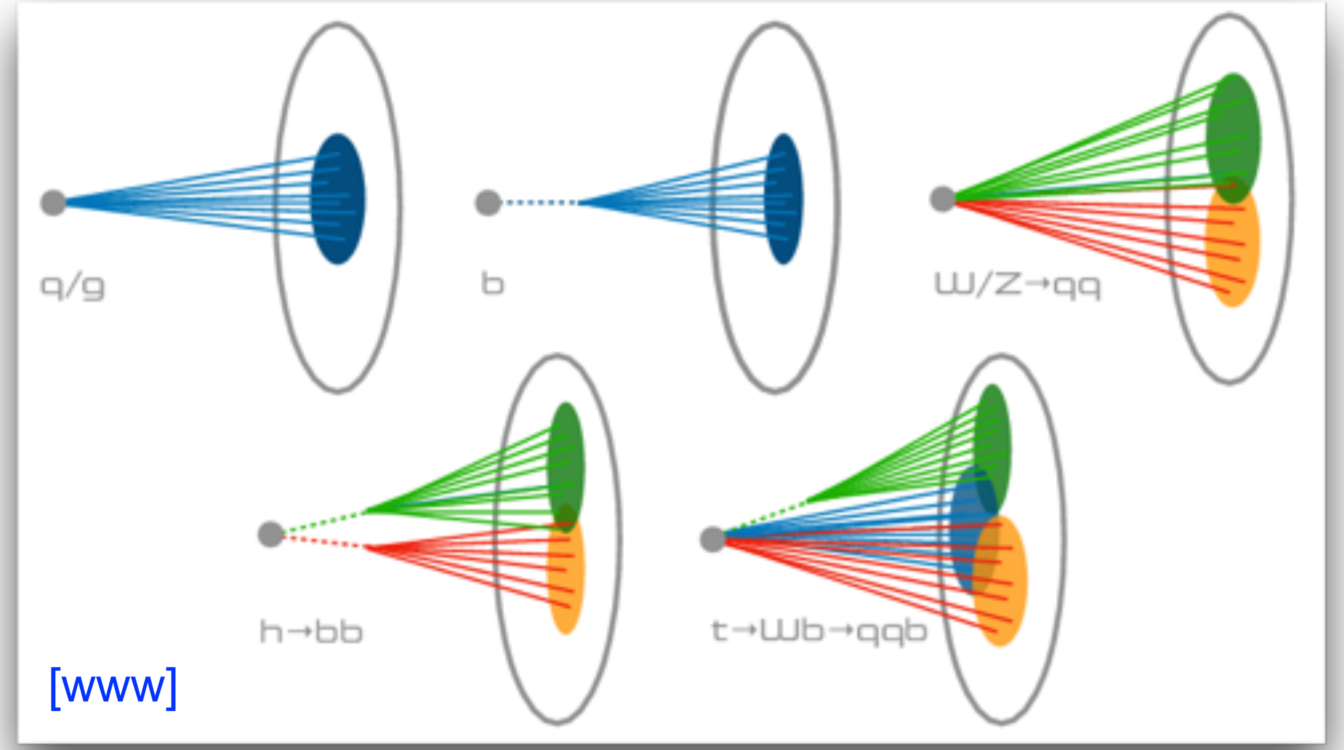
JET TOPOLOGY

Tao Liu

The Hong Kong University of Science and Technology

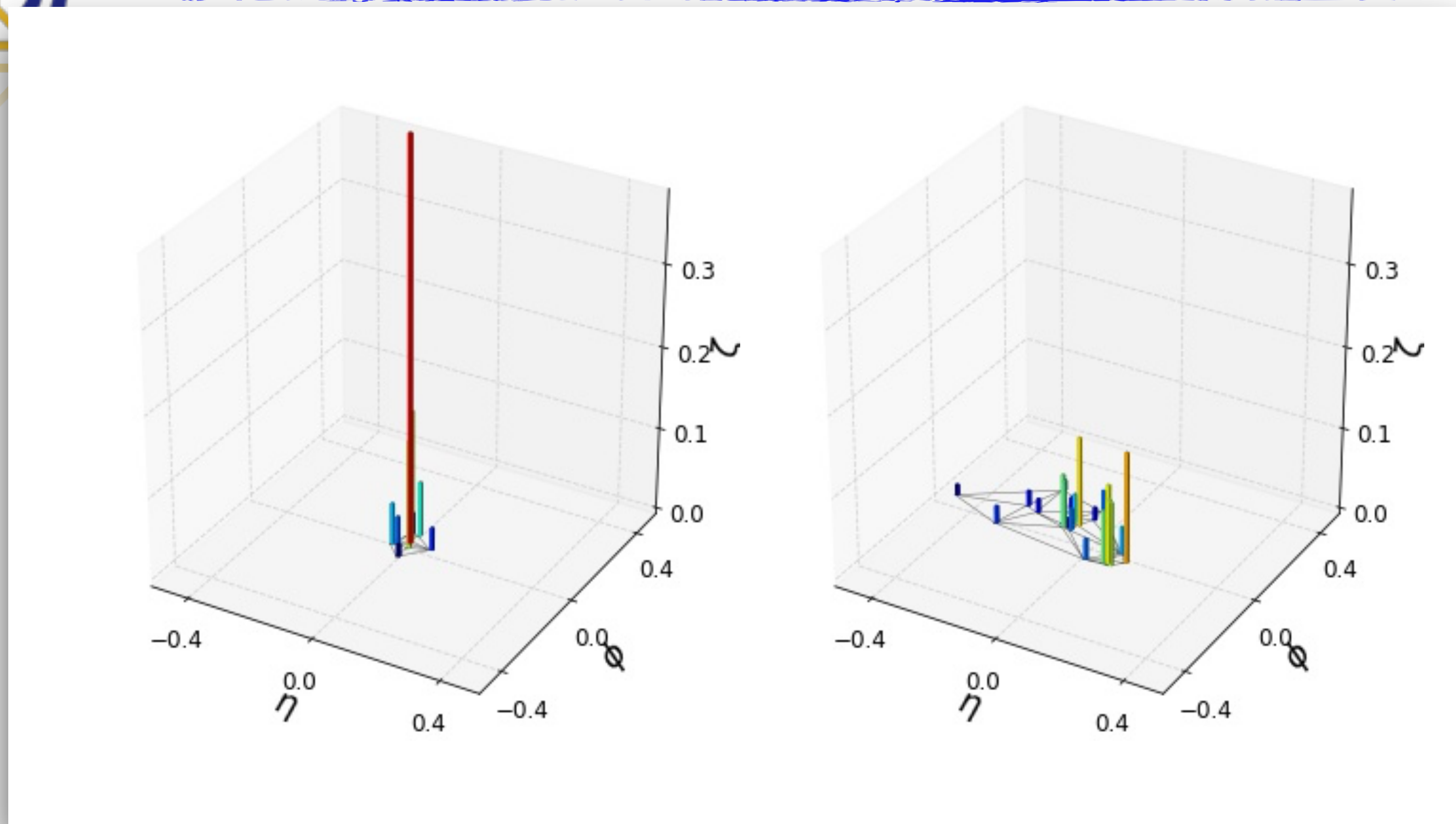
Based on arXiv:2006.12446, and ongoing project
in collaboration with Lingfeng Li and Si-jun Xu

What Are Jets?



- Jets, both QCD and boosted-heavy ones, are produced via perturbative interactions including parton shower + non-perturbative confinement processes.
- Although standard clustering provides a way to systematically organize jet constituents, strong motivation exists to look into its structure.
- Jet profile inherits from the kinematics of the shower-produced partons. It is in essence a manifestation of the nature (flavor, QCD charge, momentum, etc.) of its ancestral particle. Looking into its structure can greatly assist the tagging of jet flavor and deepen our understanding on jet dynamics.

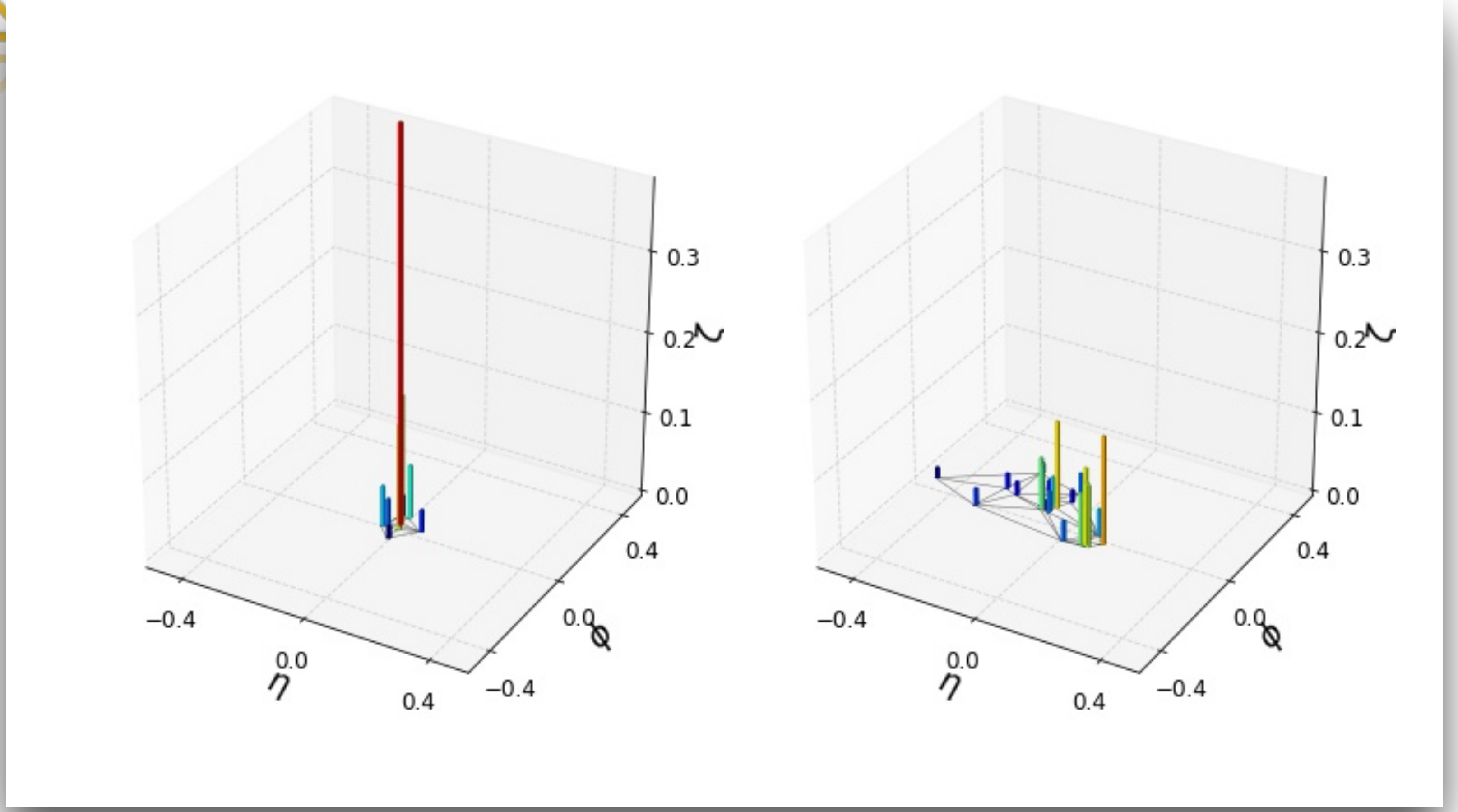
Light-quark vs. Gluon Jets



Parashorea chinensis
望天树

Banyan tree
榕树

Light-quark vs. Gluon Jets

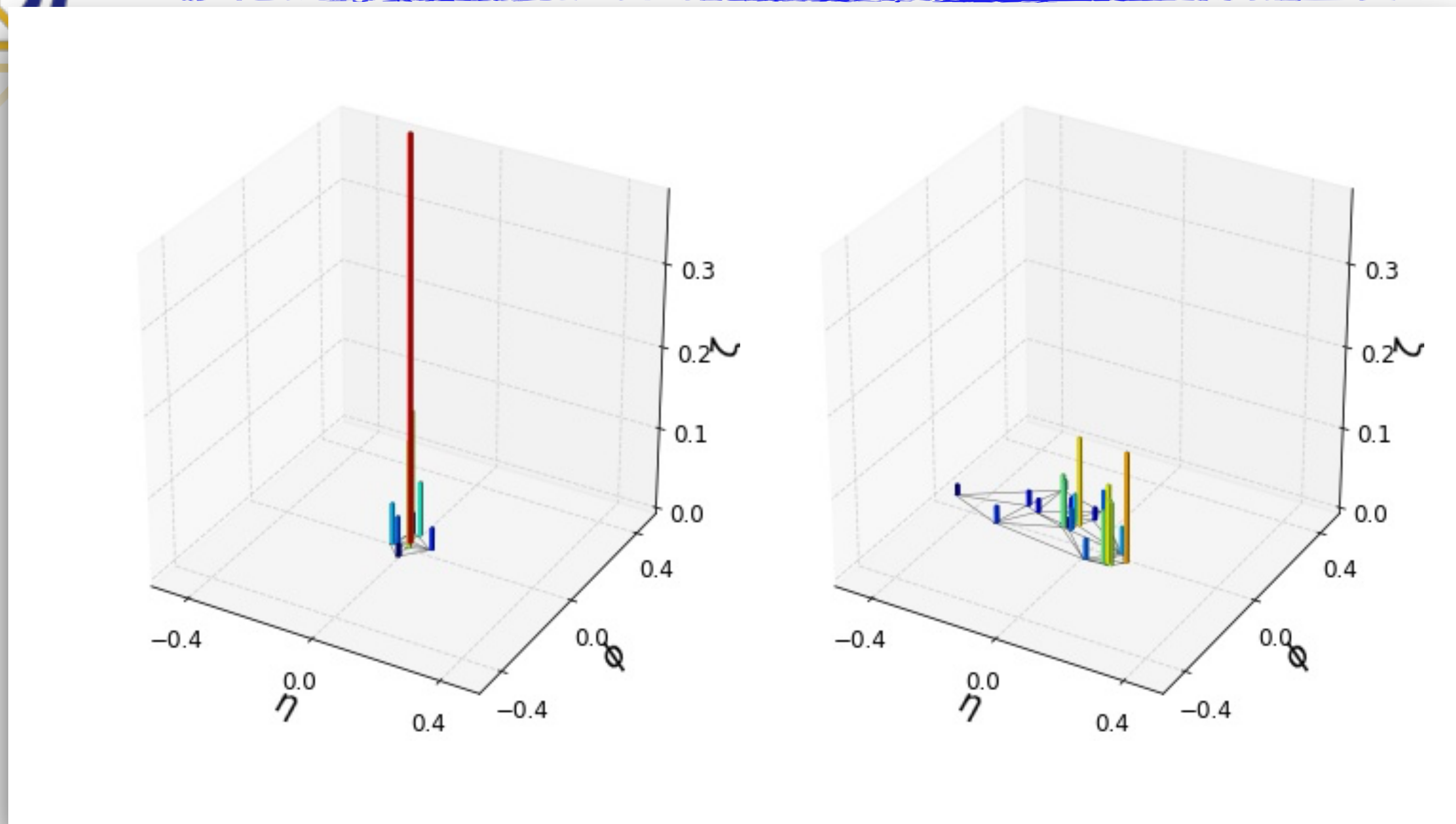


Parashorea chinensis
望天树

Banyan tree
榕树

- What make these trees look different? - Structure of their branches (the terminology ``branch'' is abused in this context, where it refers to twig, branch or trunk)
  - ``geometric'' features: shape, size, position, etc.
  - ``topological'' features: multiplicity, connectivity

Light-quark vs. Gluon Jets



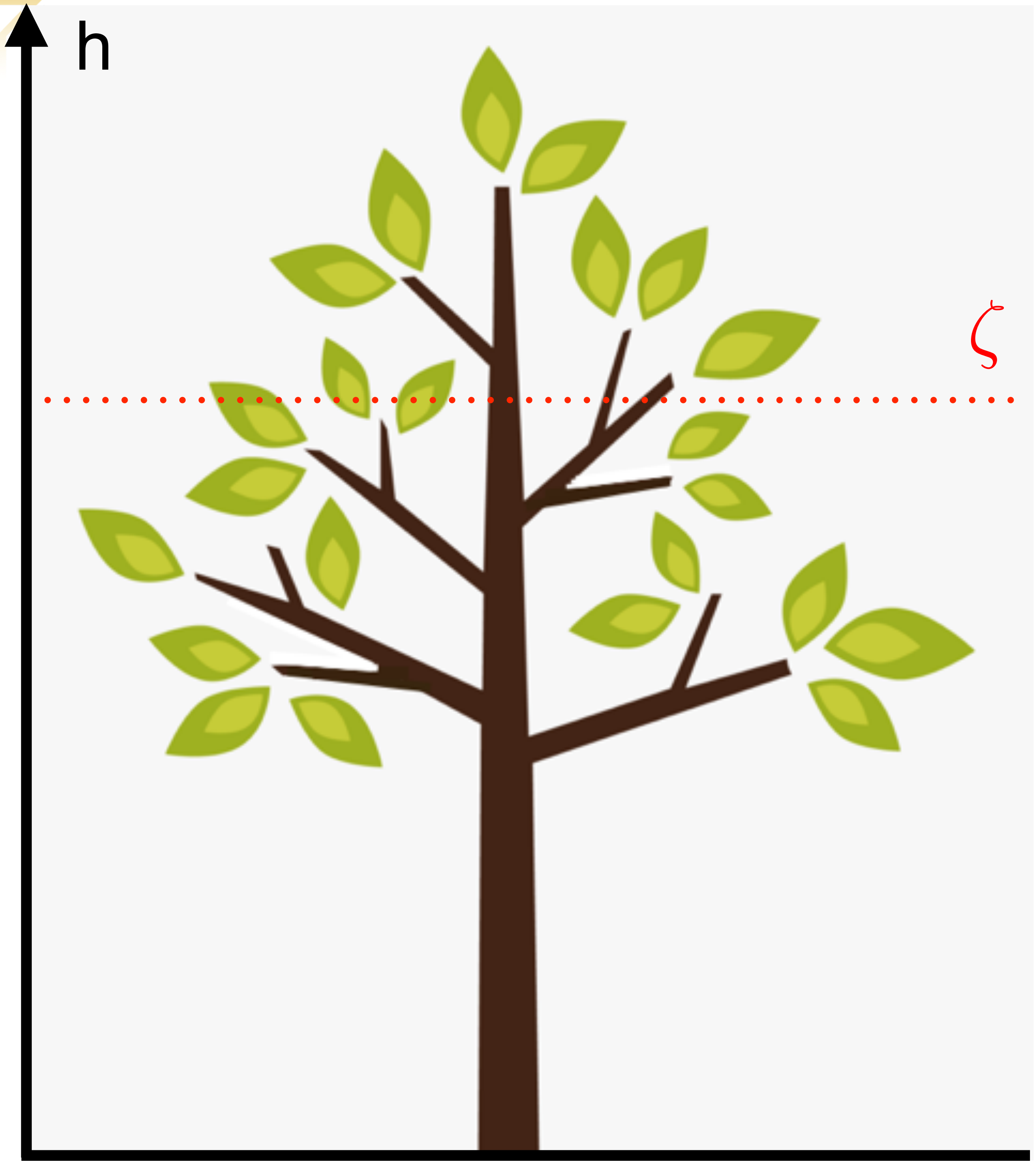
- What make these trees look different? - Structure of their branches (the terminology ``branch'' is abused in this context, where it refers to twig, branch or trunk)
  - ``geometric'' features: shape, size, position, etc.
  - ``topological'' features: multiplicity, connectivity

The ``topology'' of a tree is manifested as its dependence on some characteristic parameter!

Parashorea chinensis
望天树

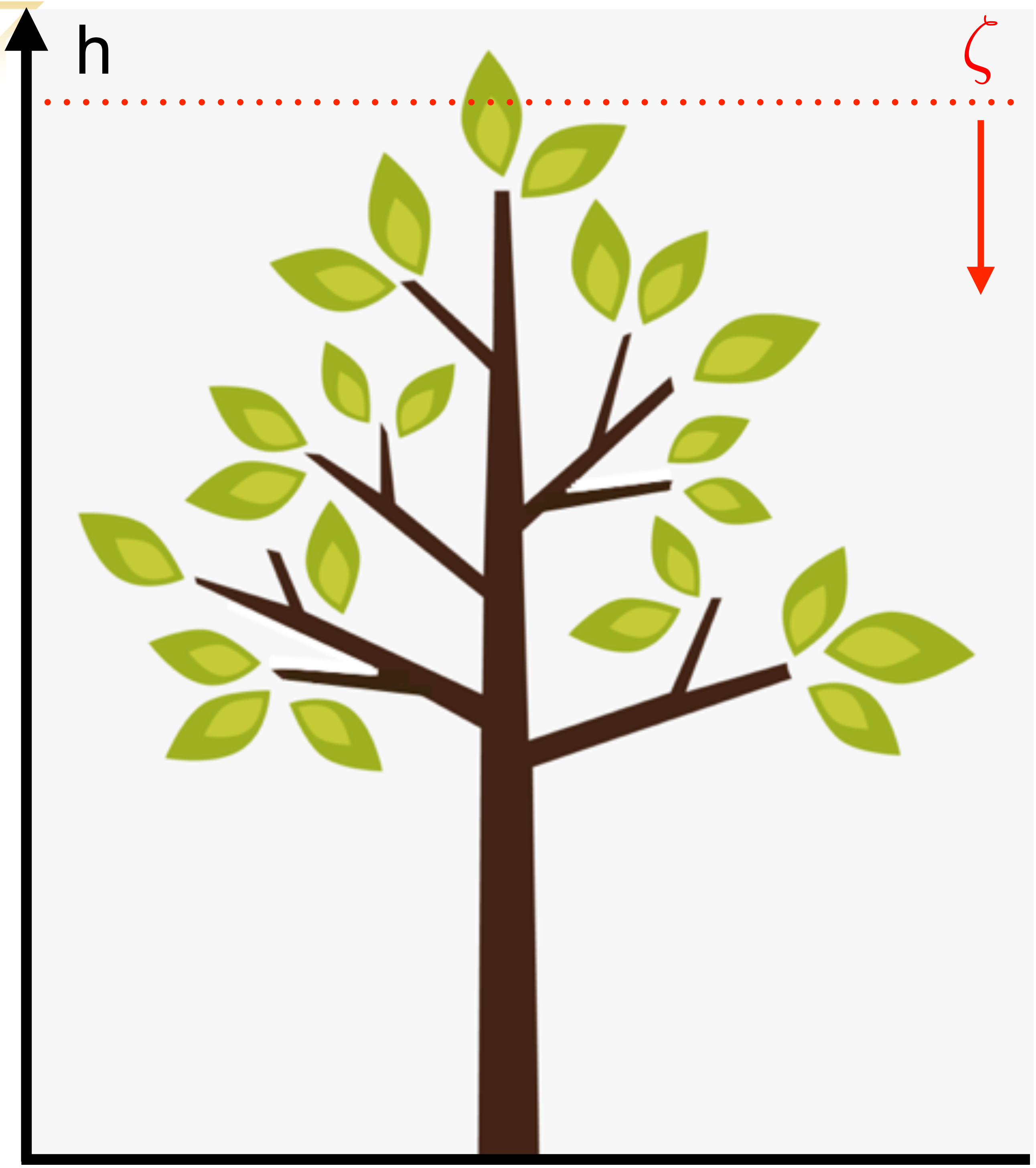
Banyan tree
榕树

Topology of A Tree



Above the threshold => totally three connected branches

Topology of A Tree



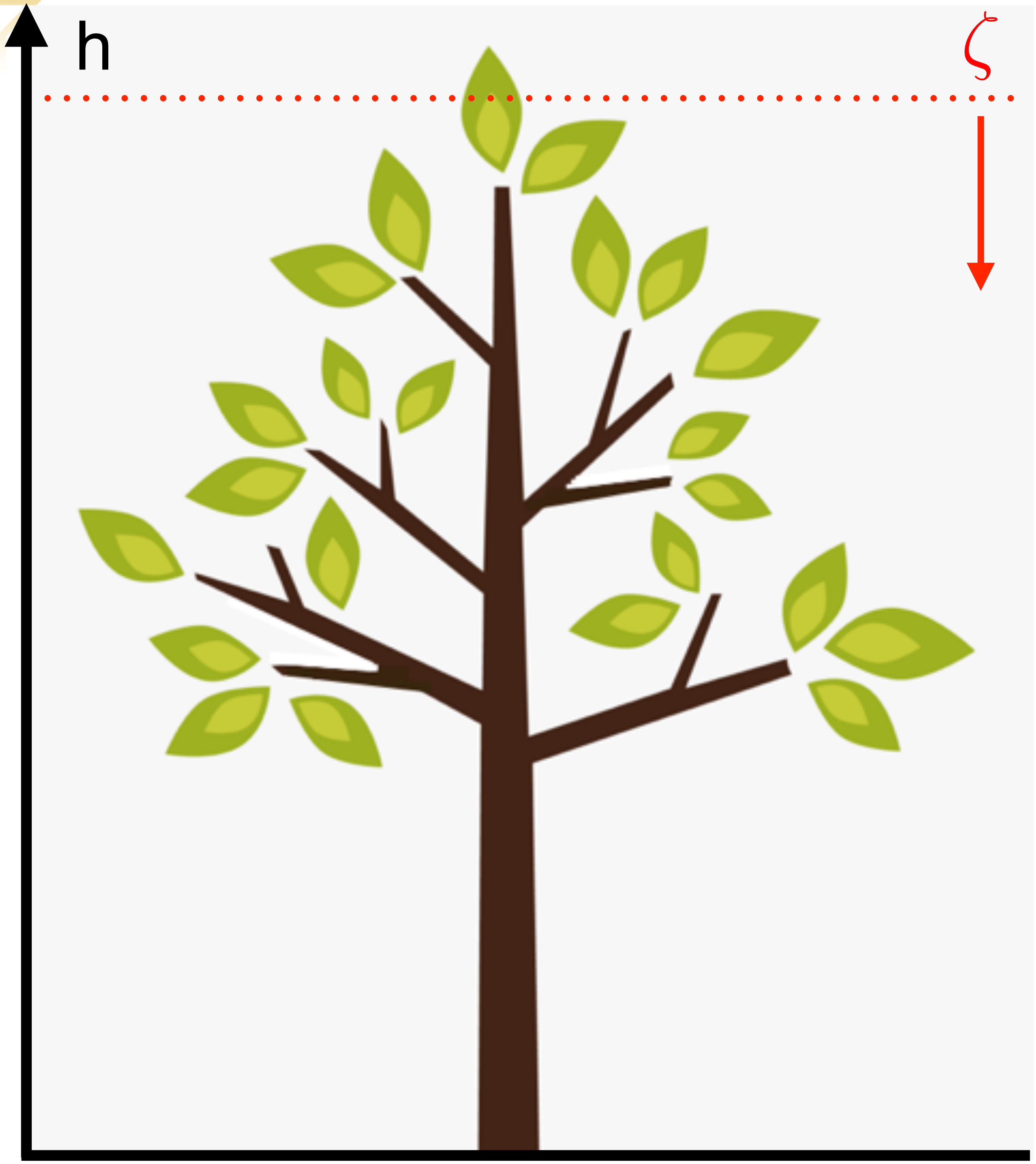
As the threshold decreases from above the top, one can record the evolution of each branch:

- birth (the height of its tip)
- growth (the increase of its weight above the threshold)
- death (the height of the place where it merges with a bigger branch)

⇩

- Evolution of the number of the connected branches as the threshold decreases

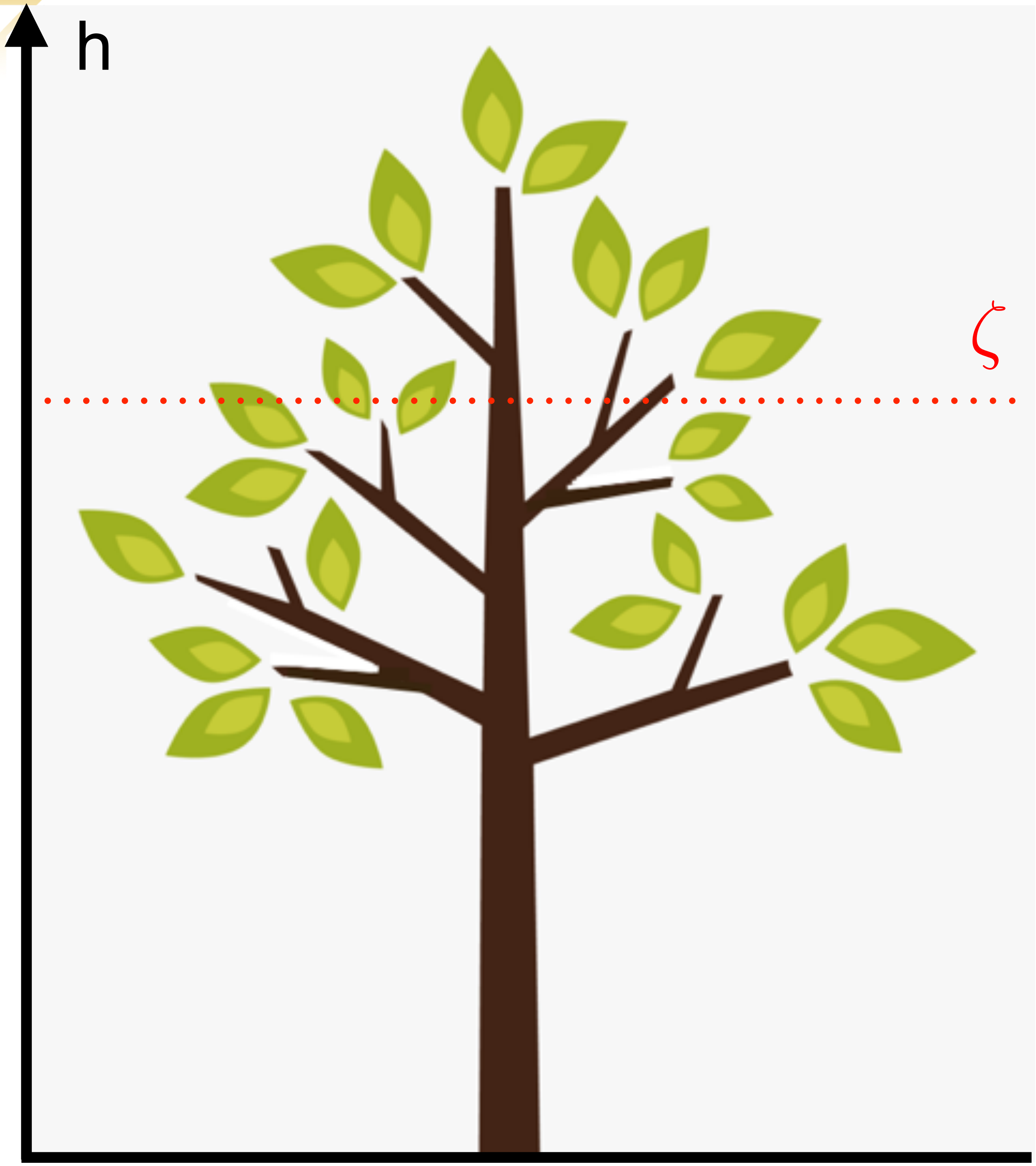
Topology of A Tree



Some other structure exists, e.g., a cavity

- Evolution of each hole and its number as the threshold decreases

Persistent Homology



A powerful tool in computational topology

([T. Martinetz, 1993; C. Correa et.al., 2011; S. Liu, 2016])

- Graph (G): the whole tree
- Super level set (G(ζ)): the subgraph above the threshold
- Persistence: lifetime of each feature

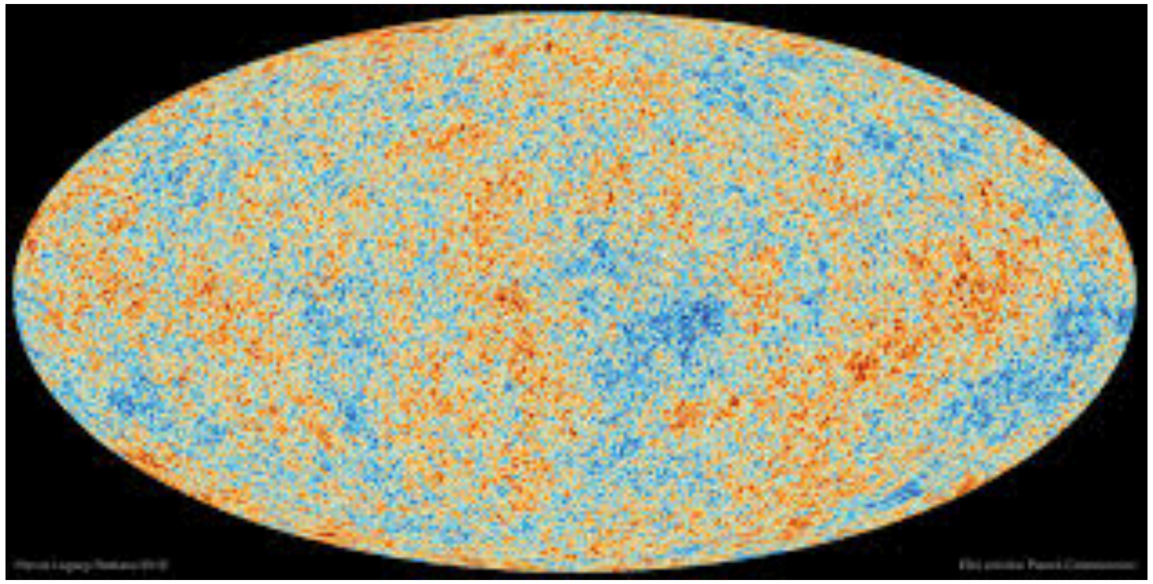
⇩

- Evolution of the number of the connected branches as the threshold decreases (Betti number β_0)
- Evolution of the number of the holes as the threshold decreases (Betti number β_1)
- Betti number is the rank of the i-th Homology group in algebraic topology. The latter characterizes the topology of a space based on the relationship between the cycles and its boundaries.

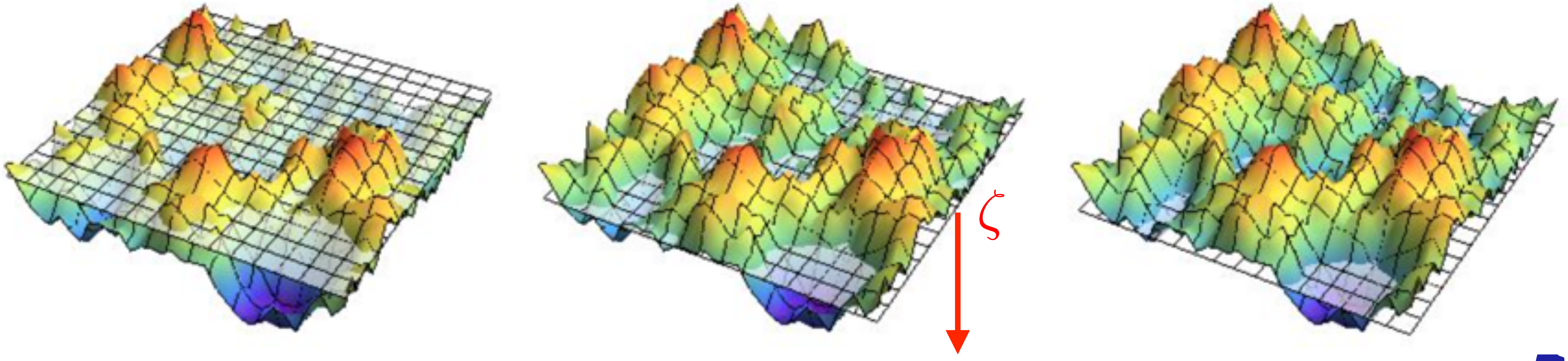
Persistent Homology

A powerful tool in computational topology
([T. Martinetz, 1993; C. Correa et.al., 2011; S. Liu, 2016])

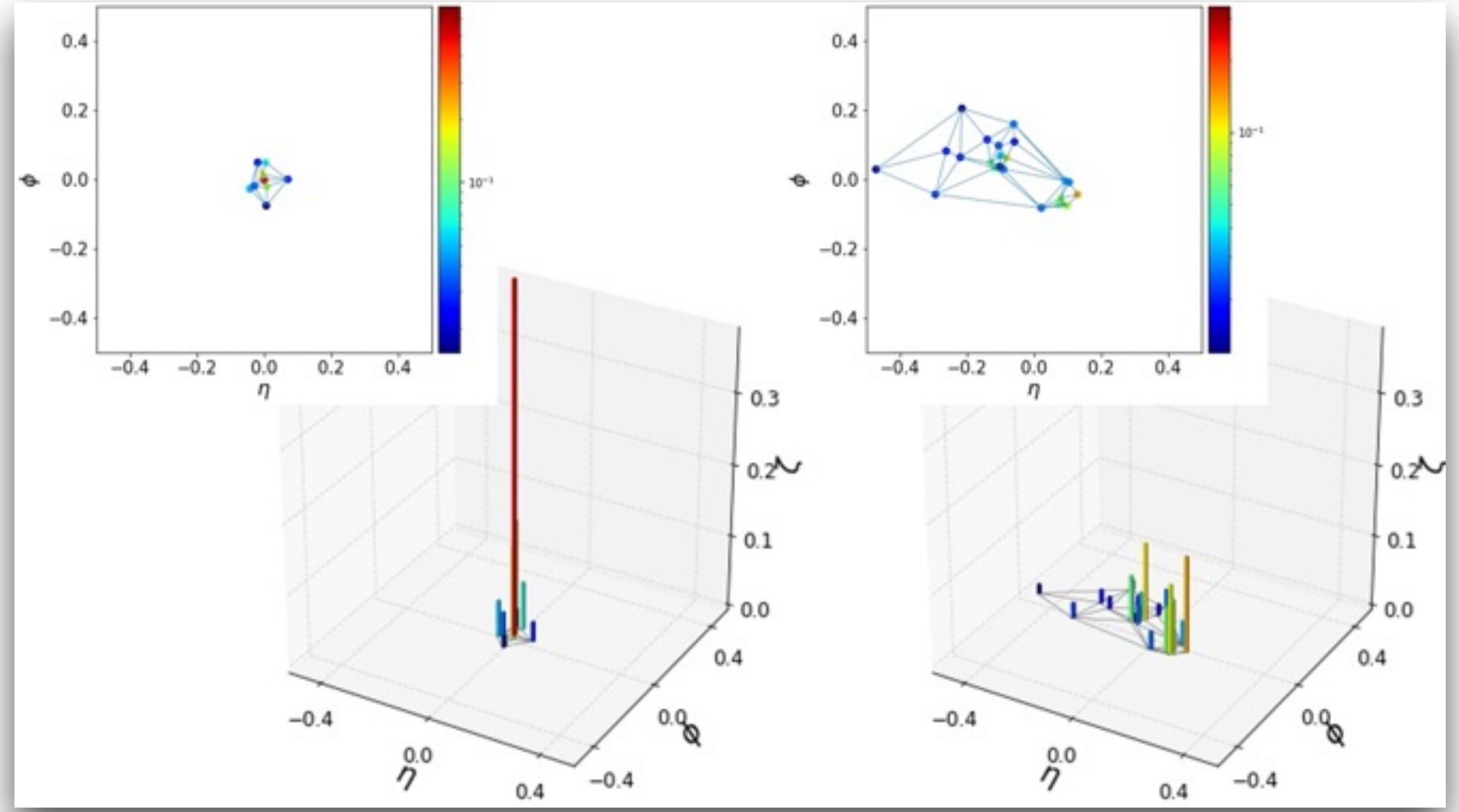
Which has been applied in biology (protein), chemistry (molecule), astrophysics (cosmic web), cosmology (CMB), etc., to study the topology of various dynamical systems.



CMB temperature map
[A. Cole, G. Shiu, 2017; P. Pranav et. al., 2018]



Delaunay Triangulation of Jets

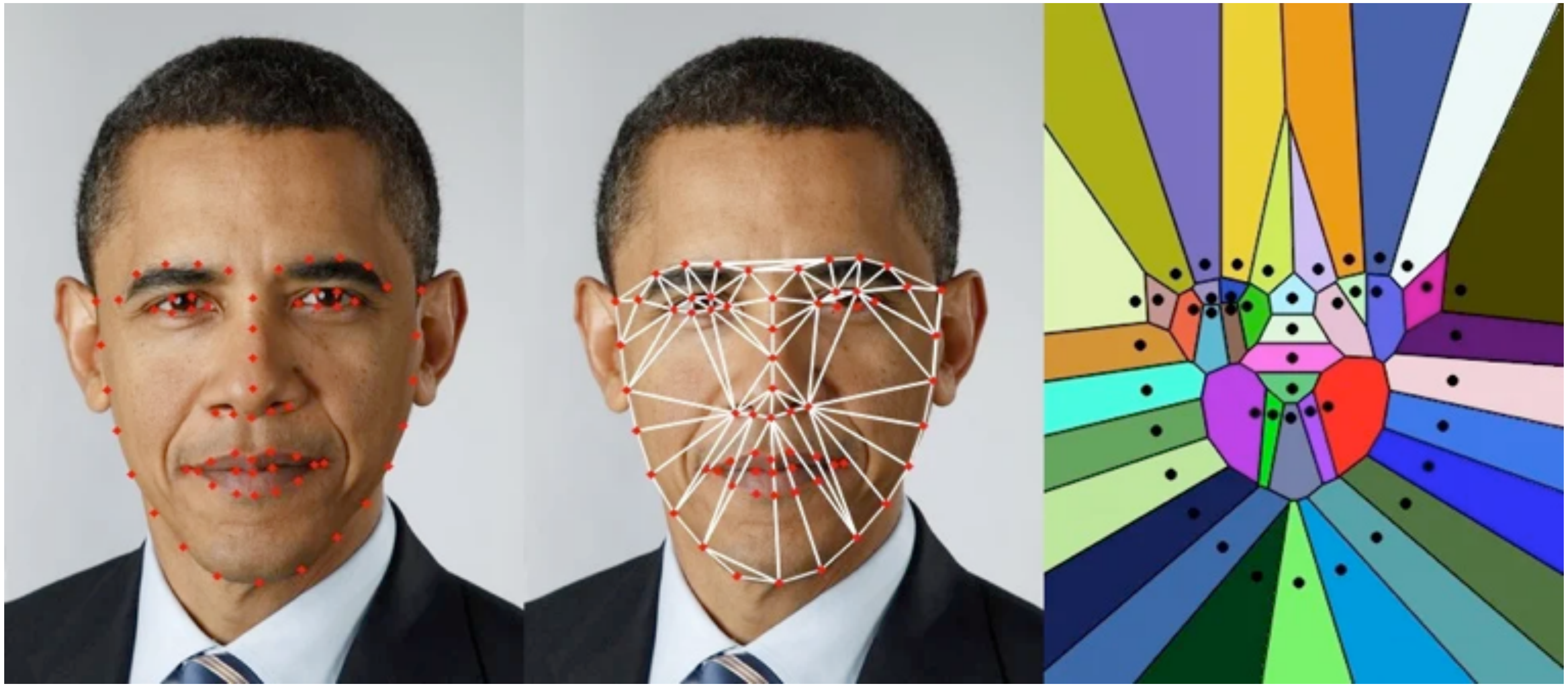


Each jet is a discrete system of its constituents.
How to define its topology?

— Treat its constituents as the sampling of a continuous profile

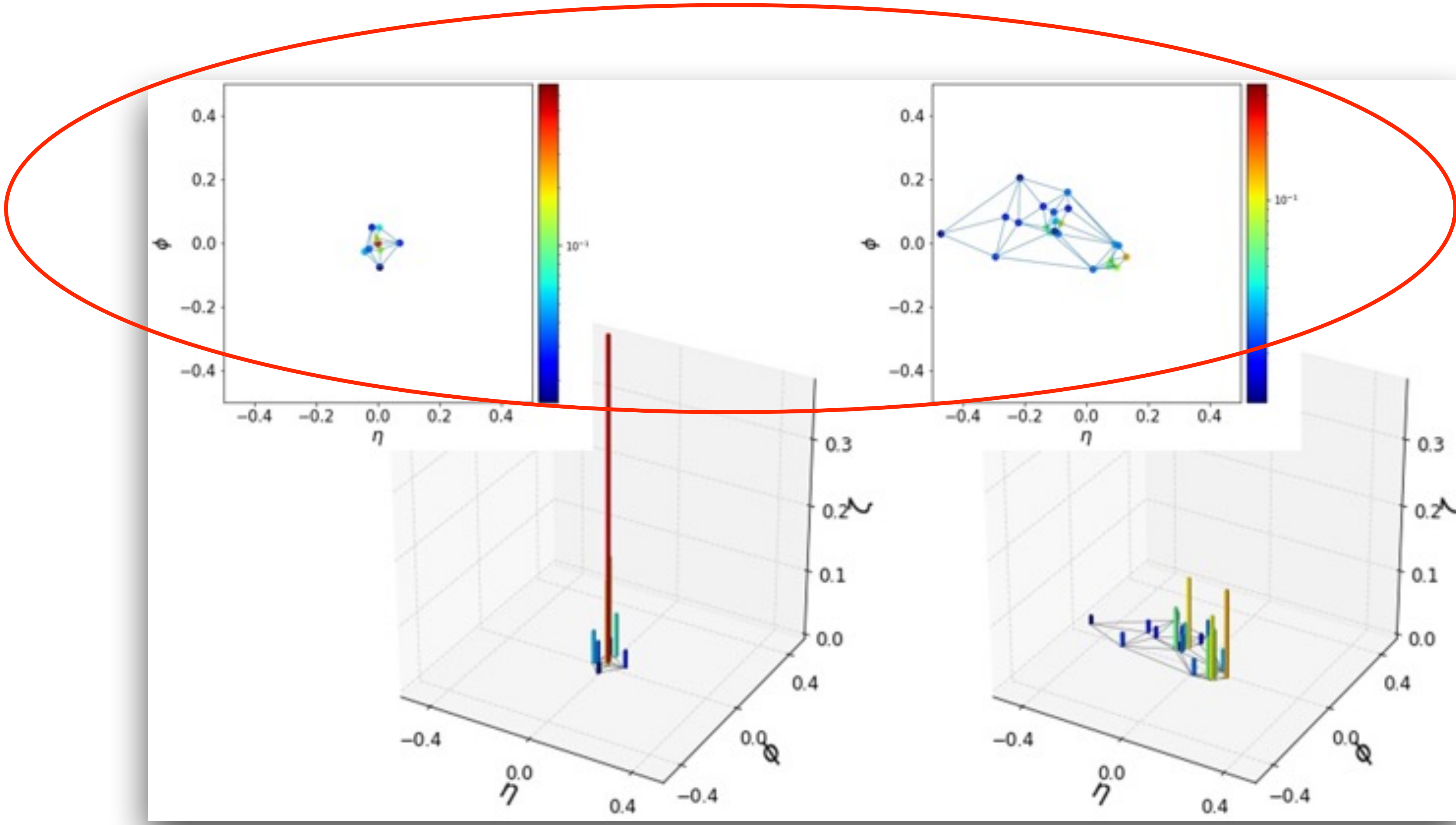
Delaunay Triangulation

- To reconstruct the continuous pT profile of each jet, we take a method of Delaunay triangulation (DT).
- DT is the dual of Voronoi Diagram (VD) (vertex <=> cell center)
- The cell of Voronoi diagram (VD) in jet physics has a natural explanation as the passive catchment area for each jet constituent, for the kt jet clustering [M. Cacciari, G. P. Salam, and G. Soyez, 2008]



Obama with detected landmarks DT VD [www.youtube.com]

Delaunay Triangulation of Jets

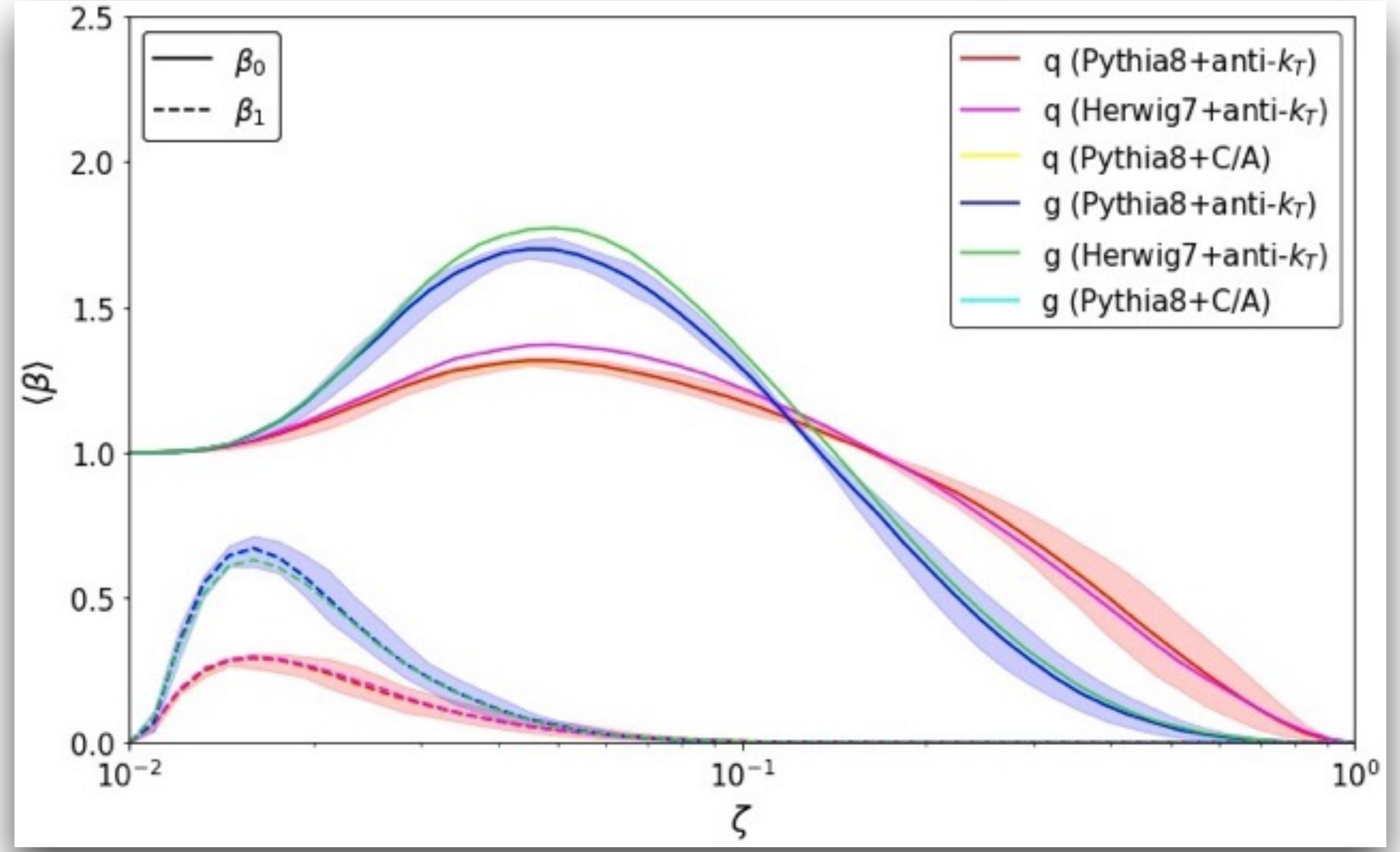


- DT: insensitive to Lorentz boost along beam direction (= translation in eta)
- Jet branches: peaks over the DT profile
  - topology(connectivity)-based (as a comparison, the clustering of subjects is ``distance''-based)

Preparation of the Samples

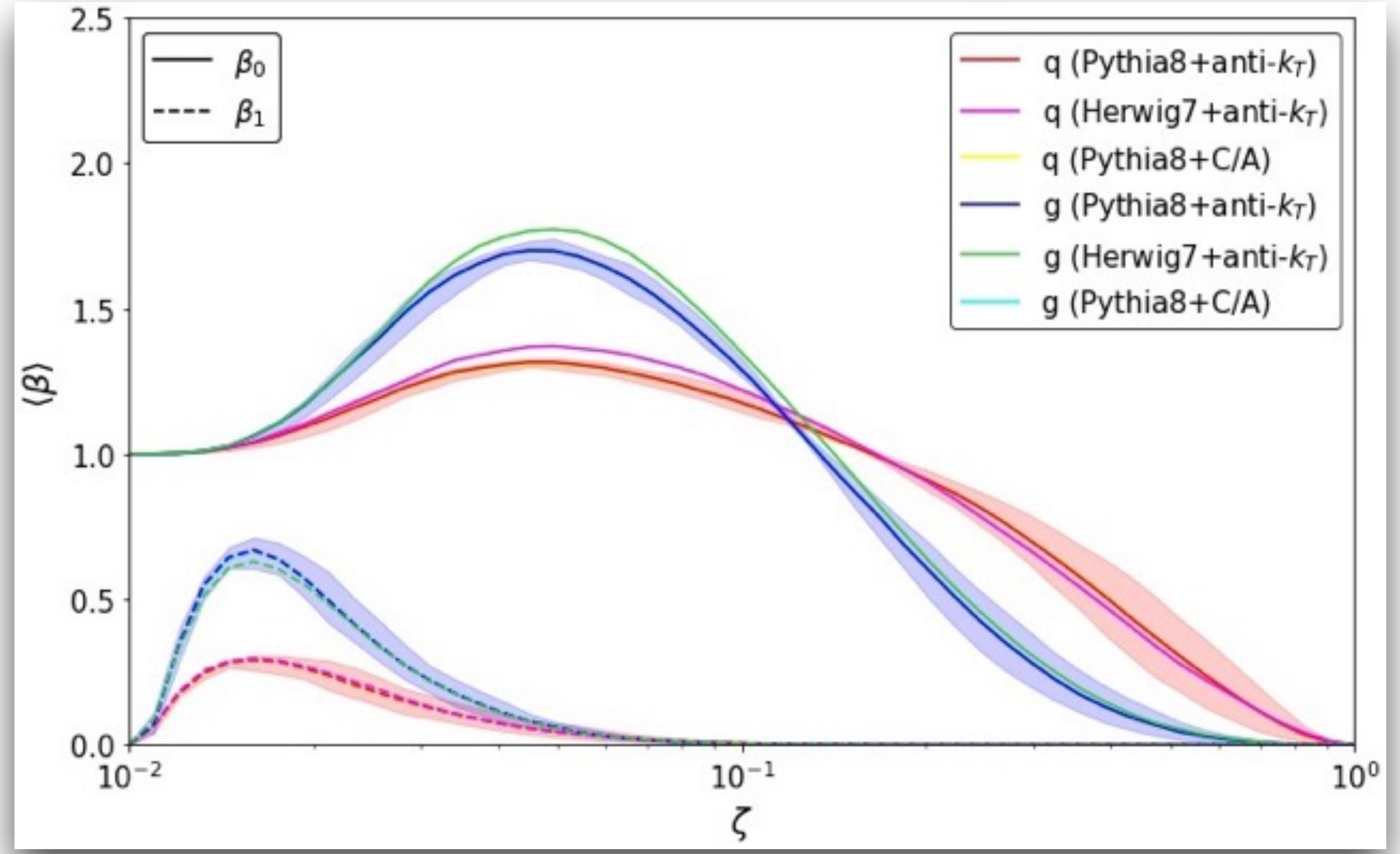
- The sample jets are generated from the pp → Z + q/g events with Z → νν.
- Showered with Pythia8 and clustered with anti-k_T algorithm ($\Delta R = 0.6$), unless otherwise specified.
- The q and g jets are then equally selected from five 50GeV-wide, non-overlapped bins, with the jet pT ranging from 100GeV to 350GeV.
- The scale threshold is defined to be pT (normalized by jet pT).
- To ensure these topological invariants to be collinear- and infrared-safe technically, we graph the jets by merging their collinear constituents with $\Delta R_{ij} \leq 0.01$ and then vetoing the soft ones with $\zeta_i \leq 0.01$.
- No dedicated detector simulation (part of the detector effects have been resolved into the said cutoffs of ΔR_{ij} and ζ_i for jet constituents)

Evolution of Betti Numbers



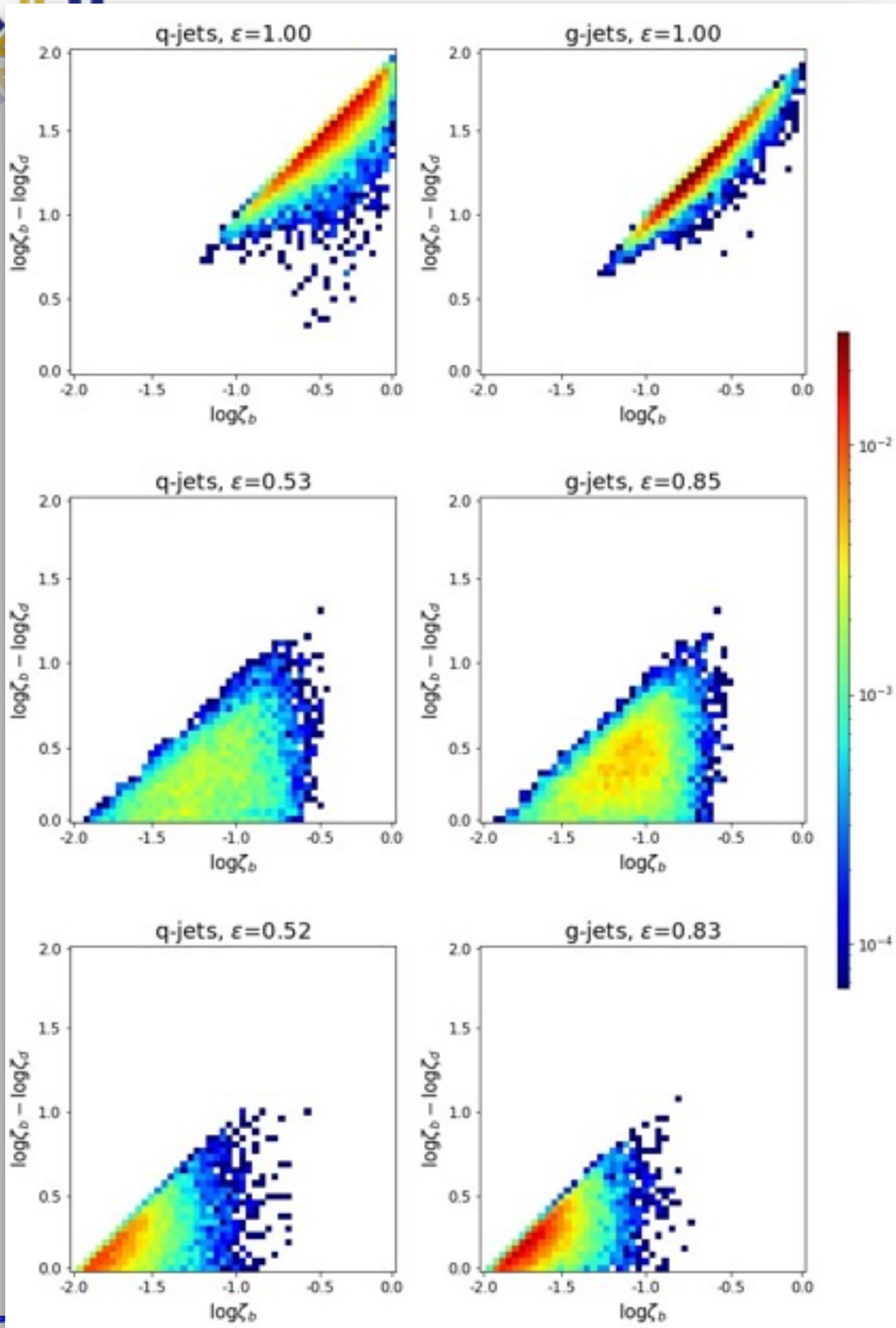
- At ζ > 1, all jets have β_0 = β_1 = 0 and hence <β_0> = <β_1> = 0.
- As ζ decreases, the q-jet < β_0> rises up earlier and then enters slow evolution.
- The g-jet <β_0> rises up later, grows a sharper and higher peak.
- As ζ keeps decreasing, soft jet constituents serve as a connector, yielding a decrease of <β_0> and an increase of < β_1>.
- The g jets show a bigger chance to form the holes, compared to the q jets.
- As ζ goes down to 10^{-2}, <β_0> evolves to one, and < β_1> to zero

Evolution of Betti Numbers



- Robust against the parton-shower models and jet clustering algorithms => 1-10% level shift to the curves.

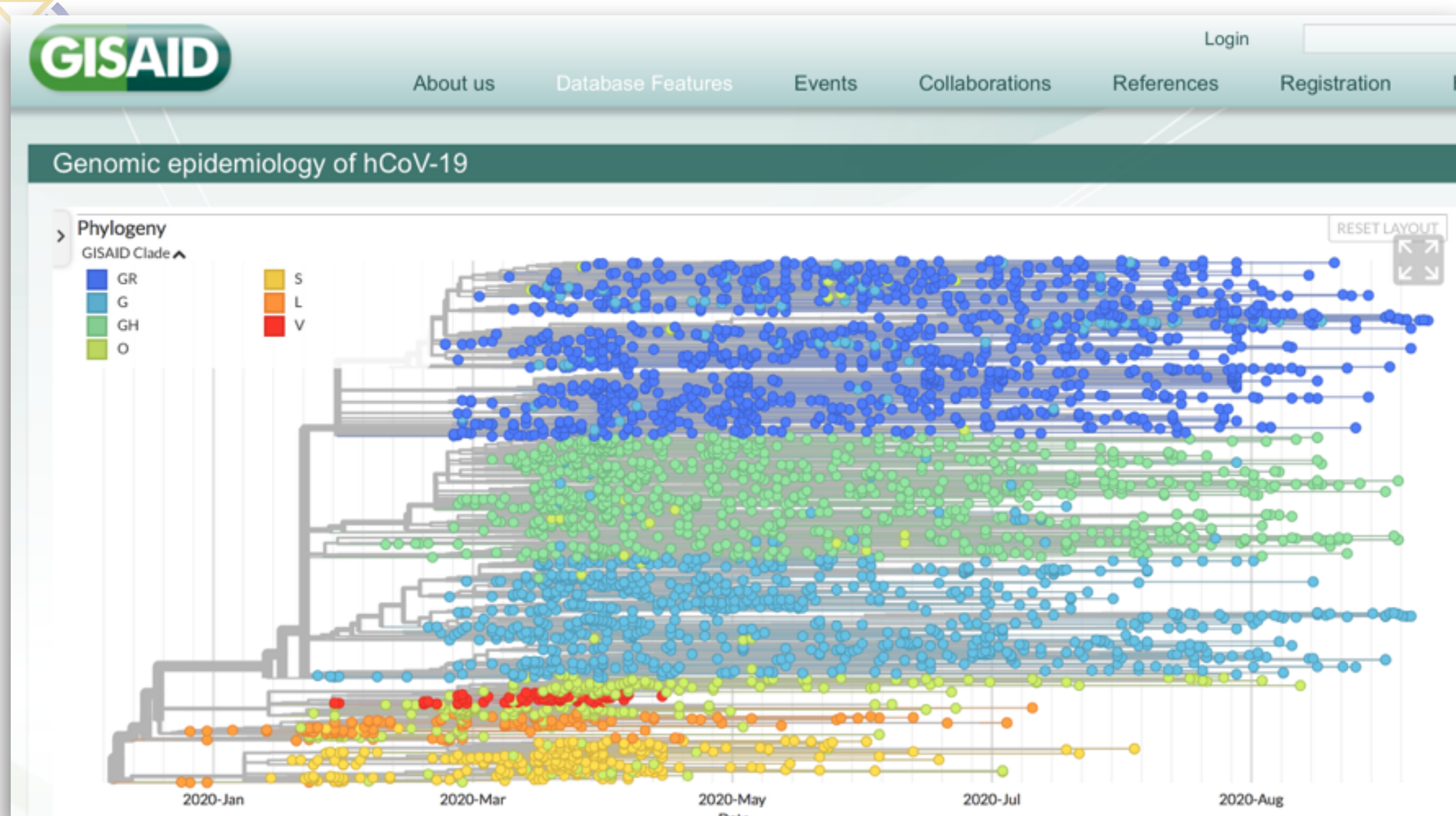
Topological Persistence



q Jets vs. g Jets

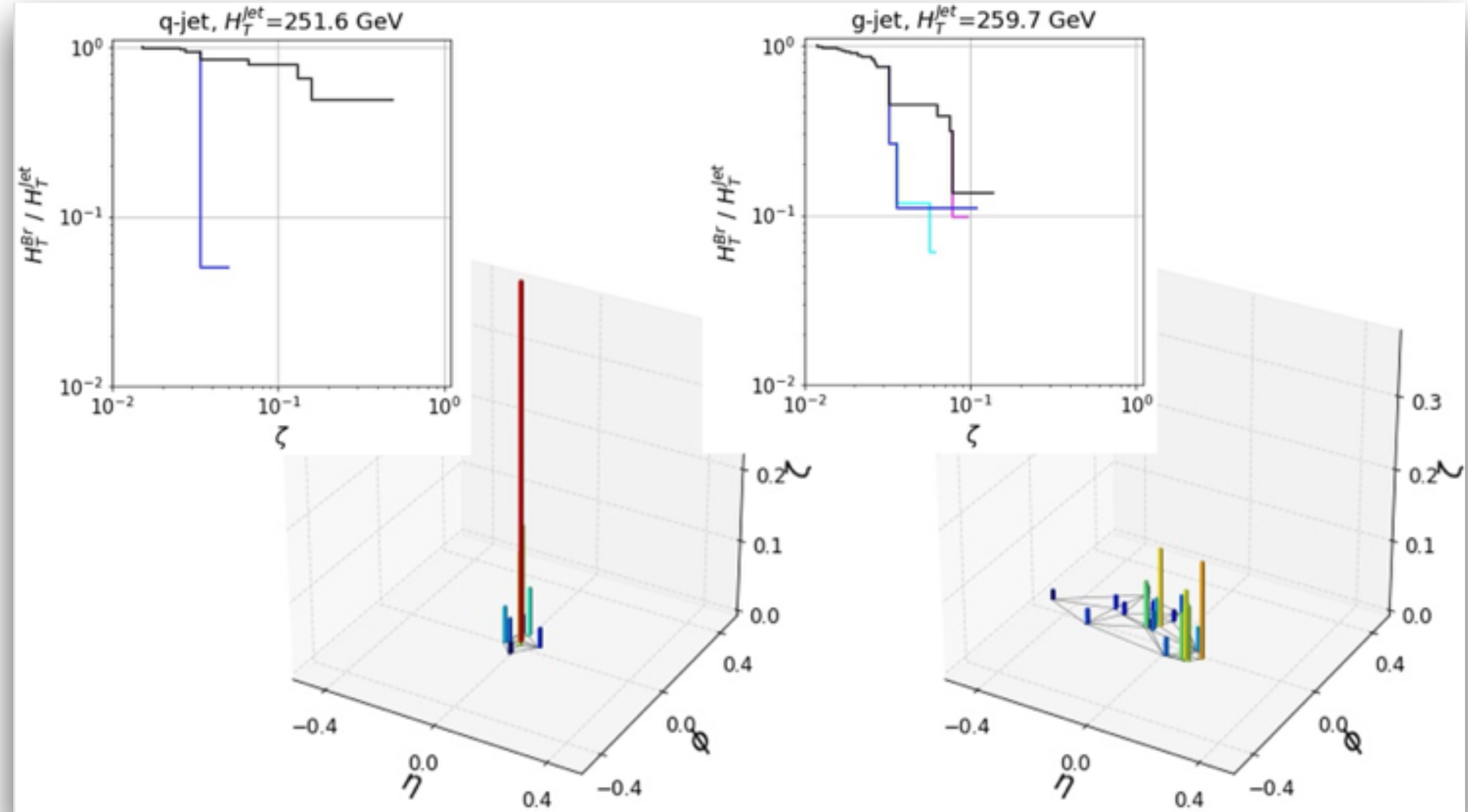
- 1st β_0 feature:
  - big ζ_b and $\log \frac{\zeta_b}{\zeta_d}$ vs. small ζ_b and $\log \frac{\zeta_b}{\zeta_d}$
- 2nd β_0 and 1st β_1 features:
  - < 55% grow vs. > 80% grow
  - small $\log \frac{\zeta_b}{\zeta_d}$ vs. big $\log \frac{\zeta_b}{\zeta_d}$

Phylogenetic Tree



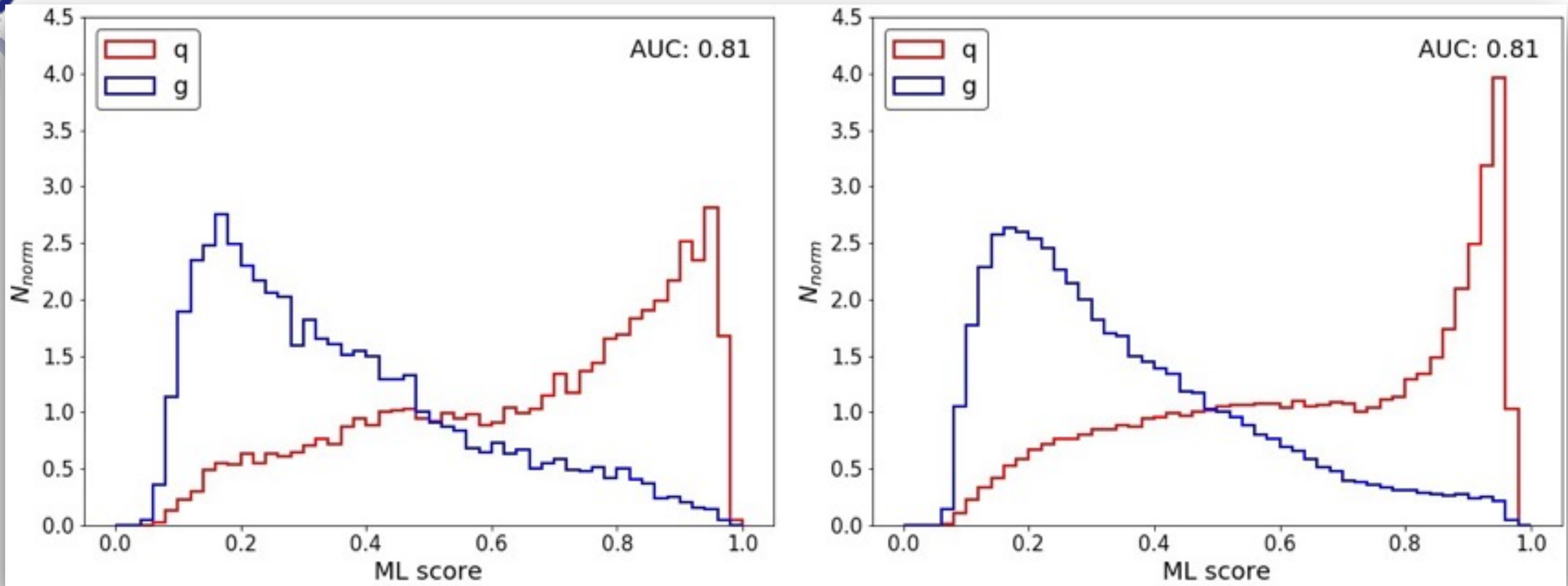
https://www.gisaid.org/epiflu-applications/phylodynamics/

Topological Phylogenetic Tree of Jets



- x-axis: threshold; y-axis: branch weight (scalar sum of pT)
- q jets: less branches and higher primary branch
- g jets: more branches and lower primary branch

A DNN-based Classifier

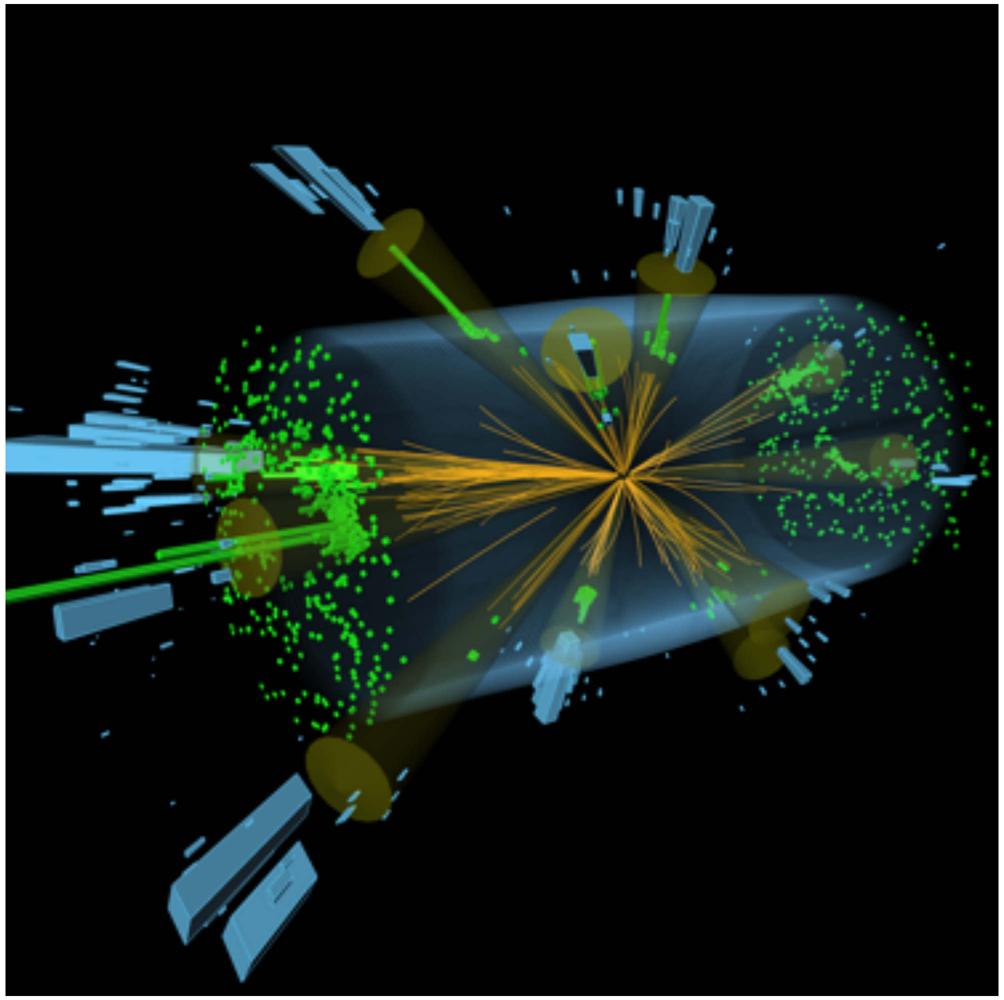


- Inputs: the first six β_0 features (birth, death, weight at death, ID of merger) + the first five β_1 features
- Left: self-generated samples (100 GeV < pT < 350 GeV)
- Right: public samples (500 GeV < pT < 550 GeV) [P. T. Komiske, E. M. Metodiev, and J. Thaler, 2018]
- Compare with literatures:
  - AUC: 0.81 - 0.88, for various samples (different pT range, detector effects, etc.)
  - AUC: 0.88 - 0.90, for public samples [P. T. Komiske et. al., 2018; H. Qu et. al., 2019]; no cutoff for jet constituents; in this limit, the topological classifier raises its AUC to > 0.85

Outlook I - Jet Physics

- Optimize this study:
  - Better reconstruction of the jet profile than DT?
  - Synergization of the missed ``geometric'' messages in this study such as spatial distribution of jet branches, into the construction of q-g jet classifier?
  - Proper simulation of detector effects: persistence may serve as a grooming tool to suppress soft contaminations
- New taggers for heavy jets (top, Higgs, W/Z, etc.)
- A new handle to probe jet dynamics

Outlook II - Event Study



- Extend the application of this tool to study the collider-event topology
  - Event branches (or ``Topological'' jets) and their evolution
  - These topological features are insusceptible to Lorentz boost along beam direction => Unlike some well-known event-level observables such as Fox-Wolfram moments, they will not suffer the boost effect caused by proton PDF at hadron colliders and hence could be well-applied in both e-e+ and pp colliders

Summary

We introduce persistent homology to characterize topological structure of jets.

These topological invariants measure multiplicity and connectivity of jet branches at a given scale threshold, while their persistence records the evolution of each topological feature as this threshold varies.

With this knowledge, we are able to reconstruct the topological phylogenetic tree for each jet.

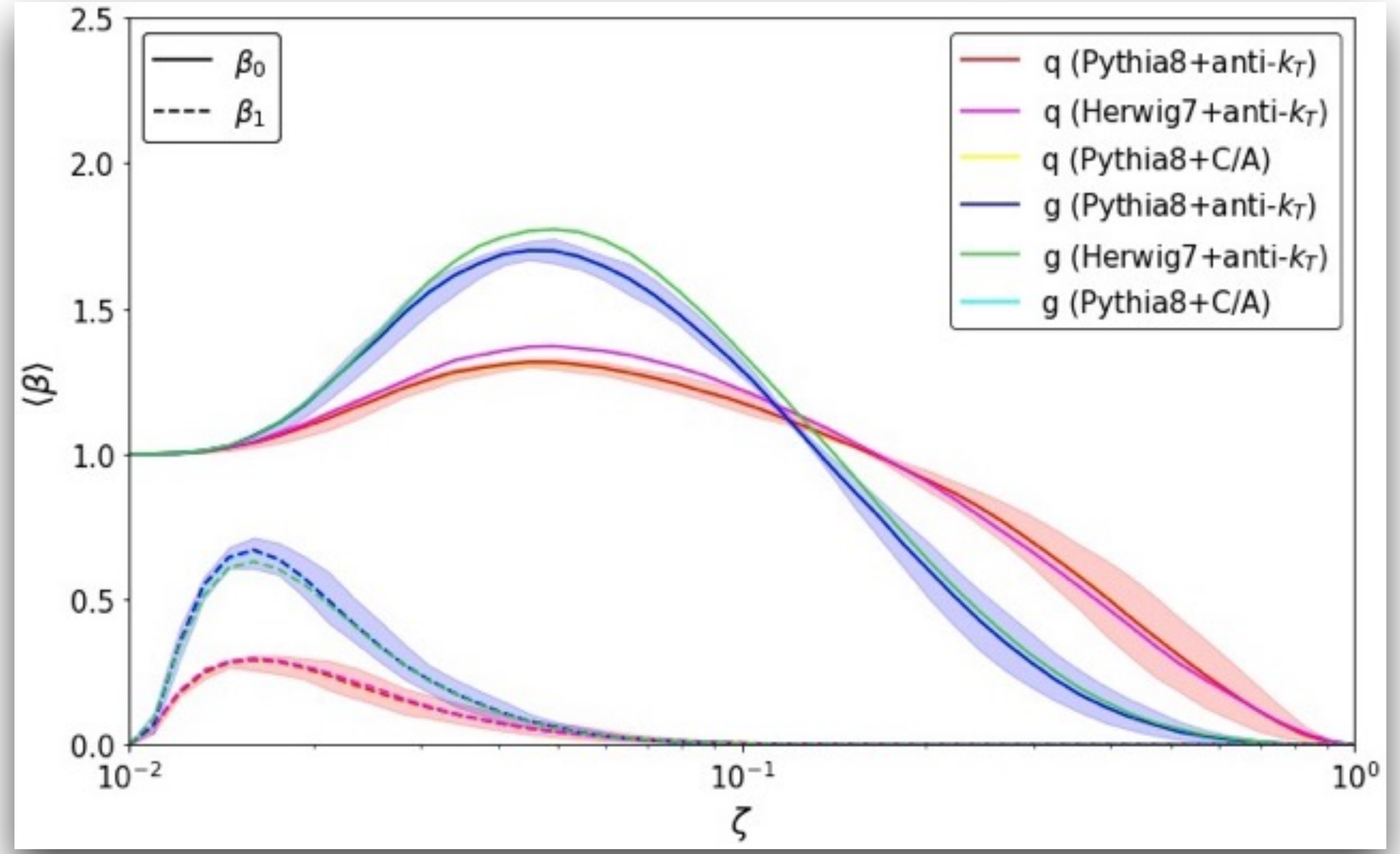
We demonstrate these points in a benchmark scenario of light-quark versus gluon jets and then develop classifiers to distinguish between them based on these topological features. We show that encouraging sensitivities can be achieved.

We hope - this topological method will open a new angle to look into jet (and event) physics.

GRF under grant No. 16305219

AoE under grant No. AoE/P-404/18-3

Backup - Evolution of Betti Numbers



- Jet pT > 100 GeV: the evolution of Betti numbers are not very sensitive to jet pT
- Jet pT < 100 GeV: the peaks will take an overall downward shift, due to the quickly weakening of parton shower.