Disentangling Jet Categories at Colliders

Machine Learning for Jet Physics Workshop

Eric M. Metodiev

Center for Theoretical Physics

Massachusetts Institute of Technology

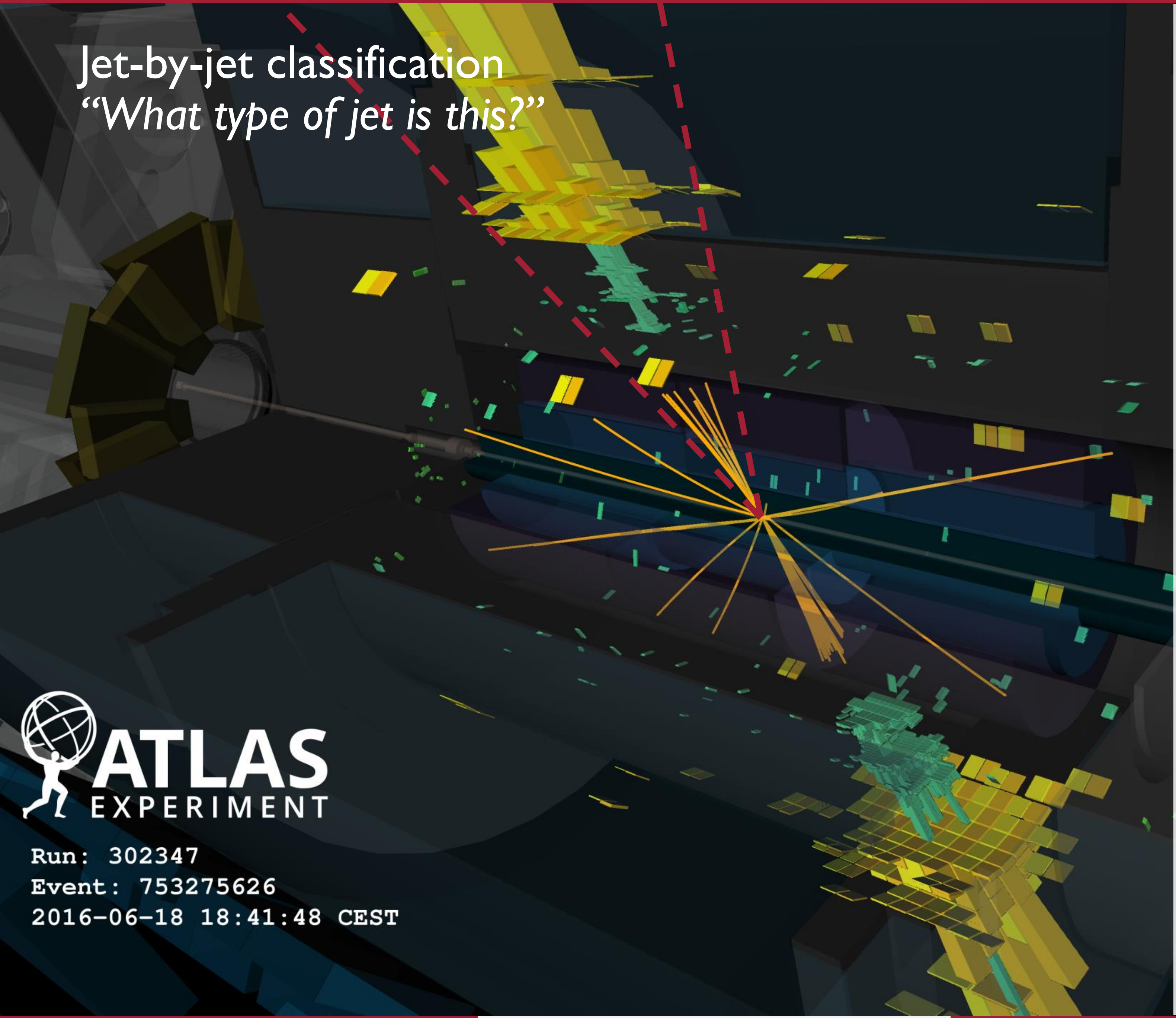
Joint work with Patrick Komiske and Jesse Thaler

November 16, 2018

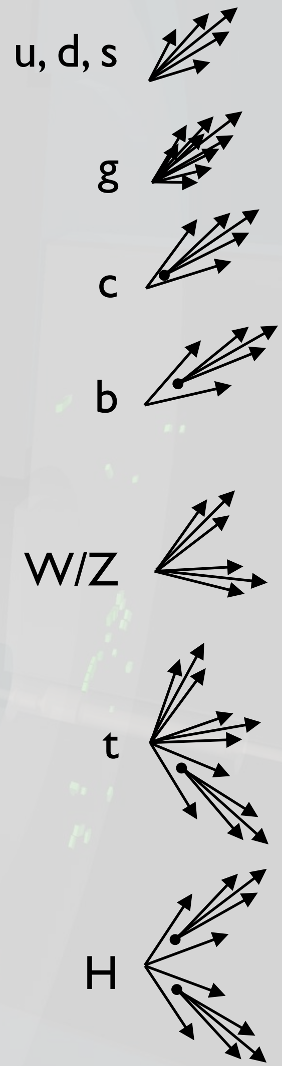
[1802.00008] [1809.01140]

Jet-by-jet classification

"What type of jet is this?"



Menu



Or

New Physics

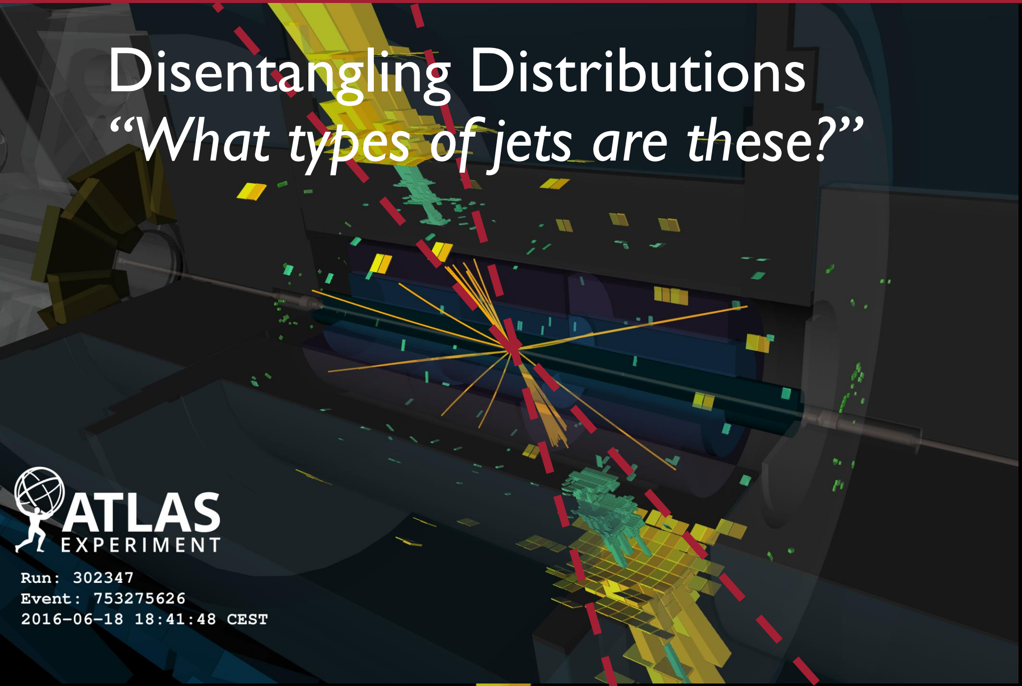
(who ordered that?)

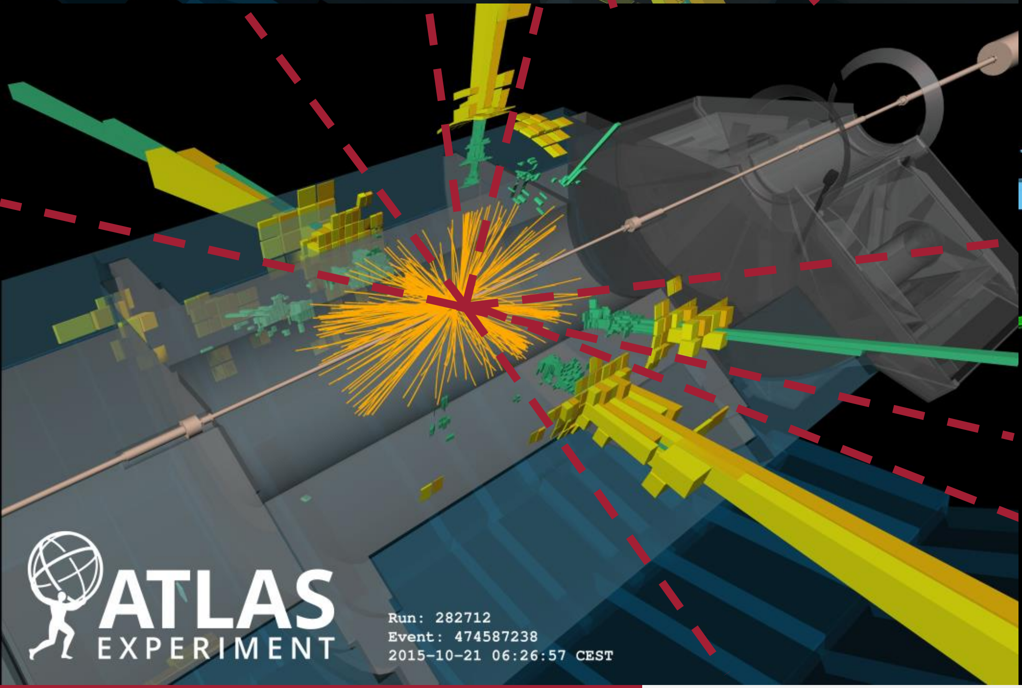
Disentangling Distributions
"What types of jets are these?"

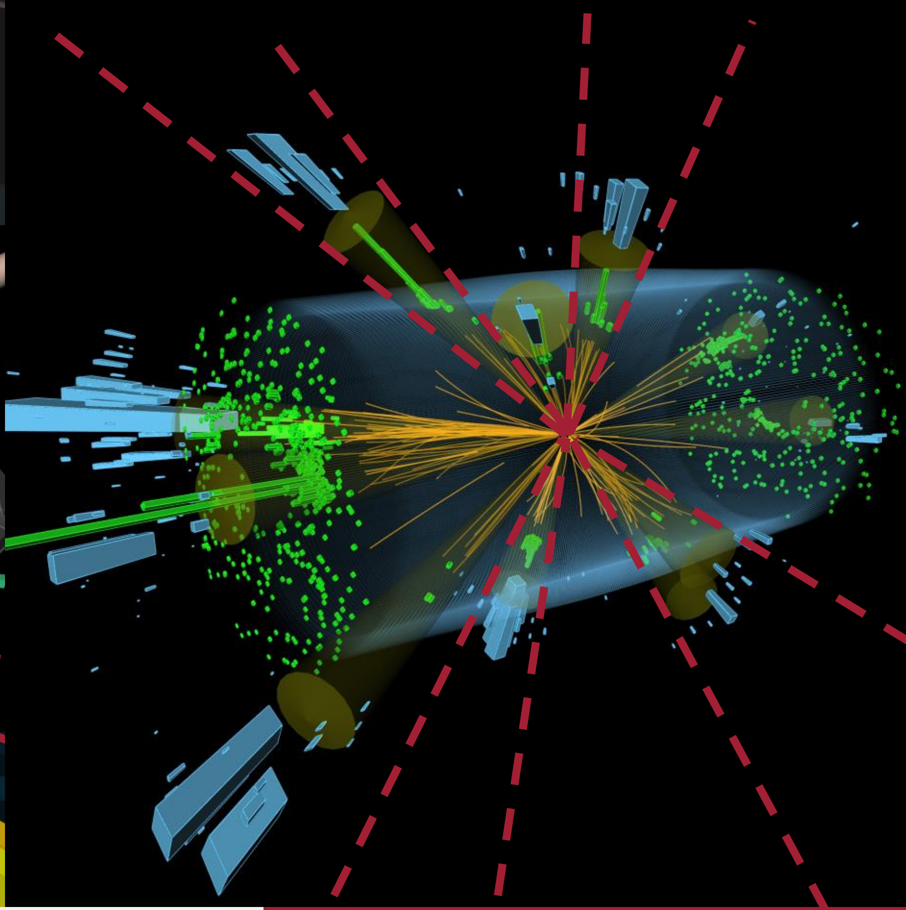
§ Menu §

Unsupervised Learning?

Data-driven categories?

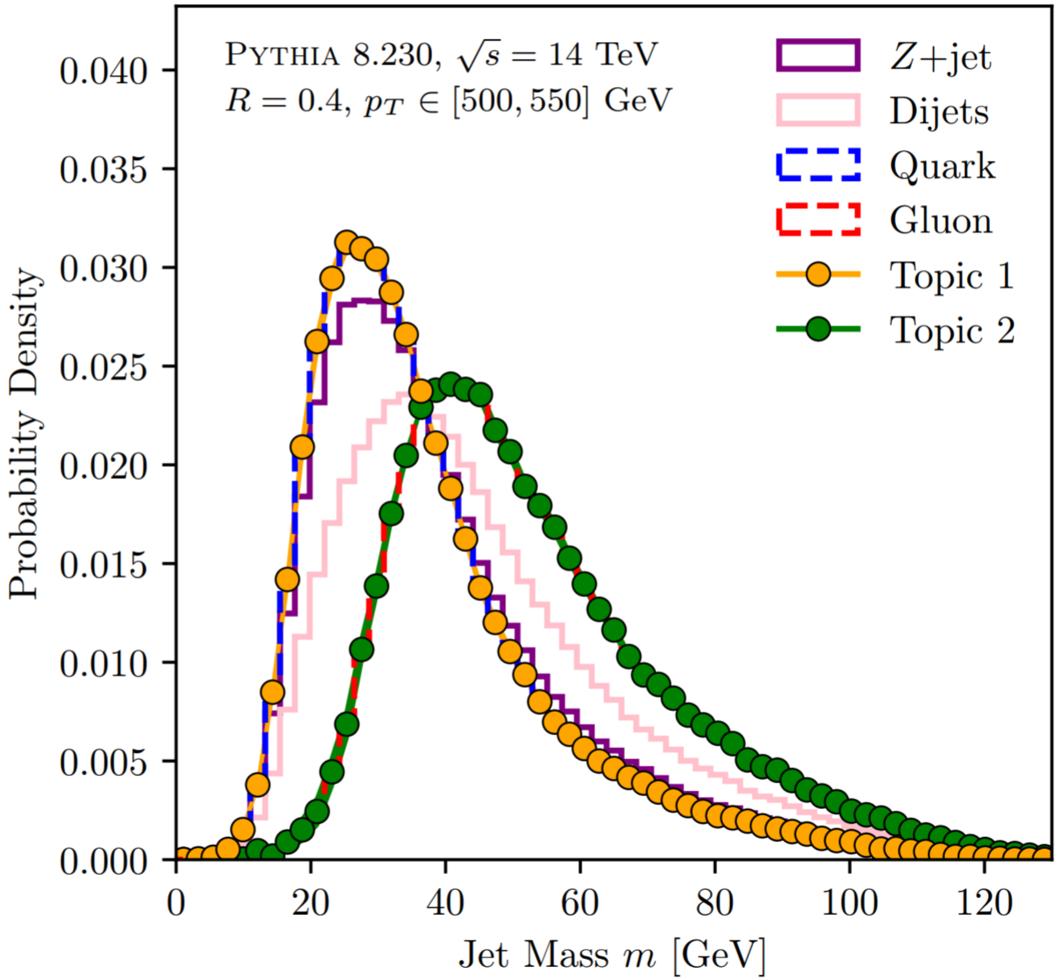






Disentangling Distributions

This talk: Towards experimentally measuring separate quark and gluon distributions



Why?

- Better understand QCD jets
- Data-driven quark/gluon taggers
- Well-defined jet categories & labels
- Parton shower tuning?
- Better extraction of α_s?

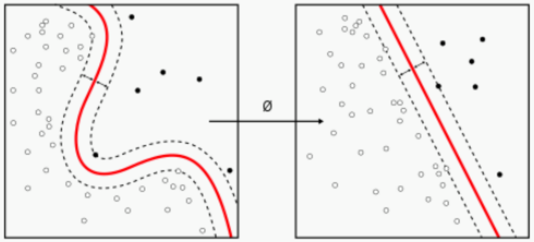
Think distribution-level, not per-jet level

Don't need a perfect tagger!

Don't need MC fractions or templates!

WIKIPEDIA

Machine learning and data mining



Problems	[show]
Supervised learning (classification • regression)	[show]
Clustering	[show]
Dimensionality reduction	[show]
Structured prediction	[show]
Anomaly detection	[show]
Artificial neural networks	[show]
Reinforcement learning	[show]
Theory	[show]
Machine-learning venues	[show]
Glossary of artificial intelligence	[show]
Related articles	[show]

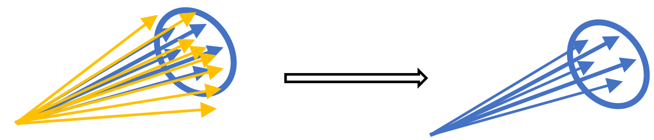
Machine learning portal

V • T • E

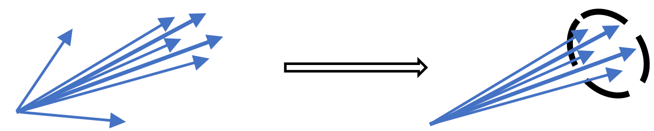
Jet-Level Techniques

Classification:
Jet Tagging

Regression:
Pileup Removal

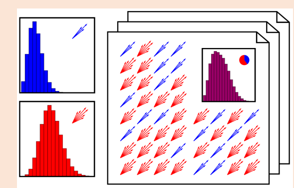


Clustering:
Jet Finding



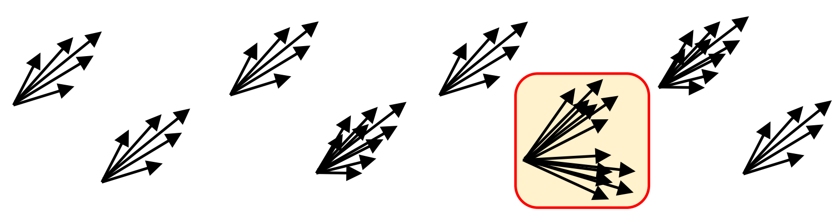
Dataset-Level Techniques

Topic Modeling:
This talk.



Our Goal: Find "jet types" that best explain the data

Anomaly Detection:
ML New Physics Searches

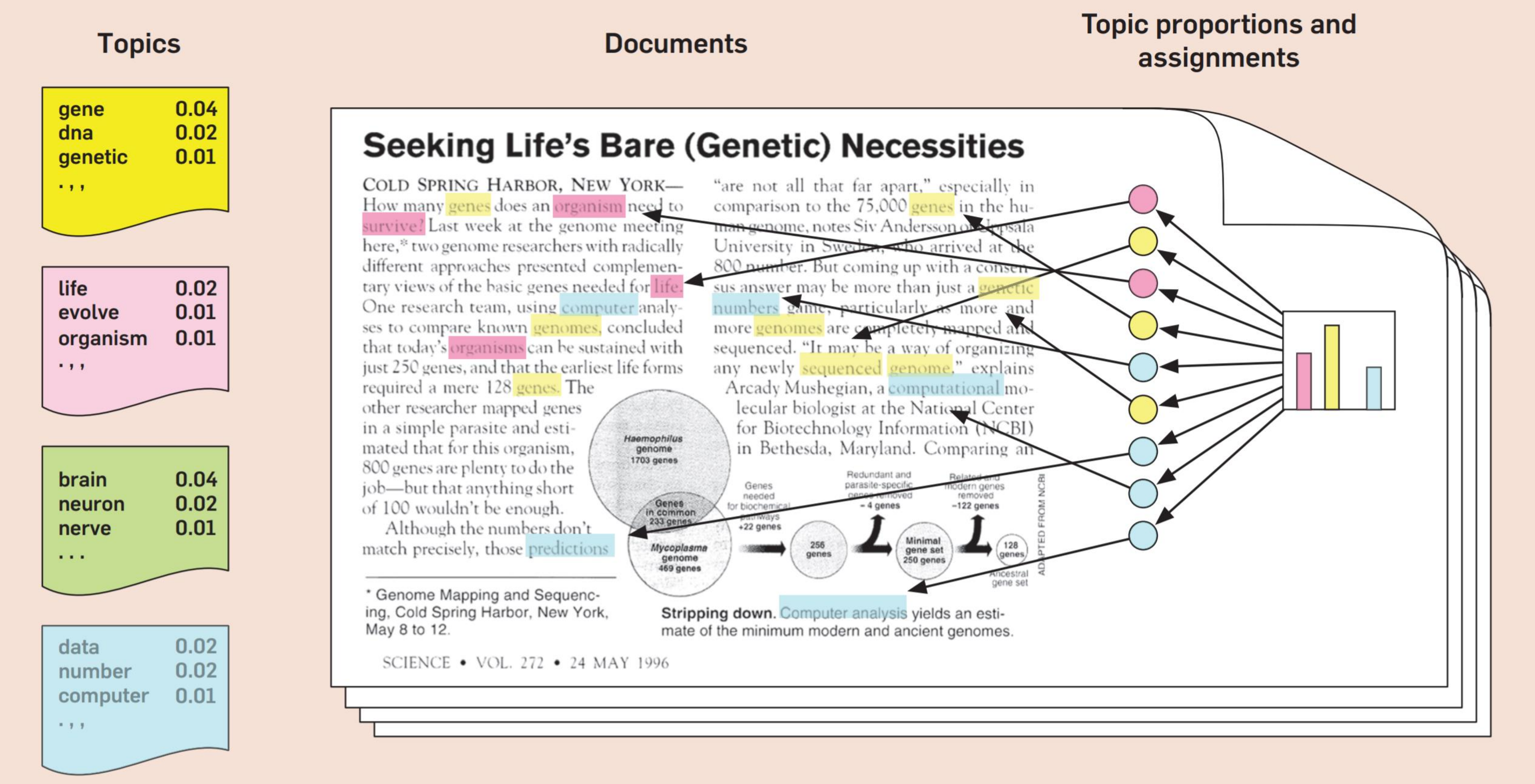


Topic modeling

Treat text documents as statistical mixtures of "topics" – distributions over words.

Can you extract the underlying "topics" given **only** the documents? Yes*

* Terms and conditions apply



Topic modeling

Treat text documents as statistical mixtures of "topics" – distributions over words.

Can you extract the underlying "topics" given **only** the documents?

Yes, as long as the topics are "mutually irreducible" (M.I.): [1710.01167] [1204.1956]

Each topic must have an "anchor" word that doesn't appear in any other topics.

A quick example:

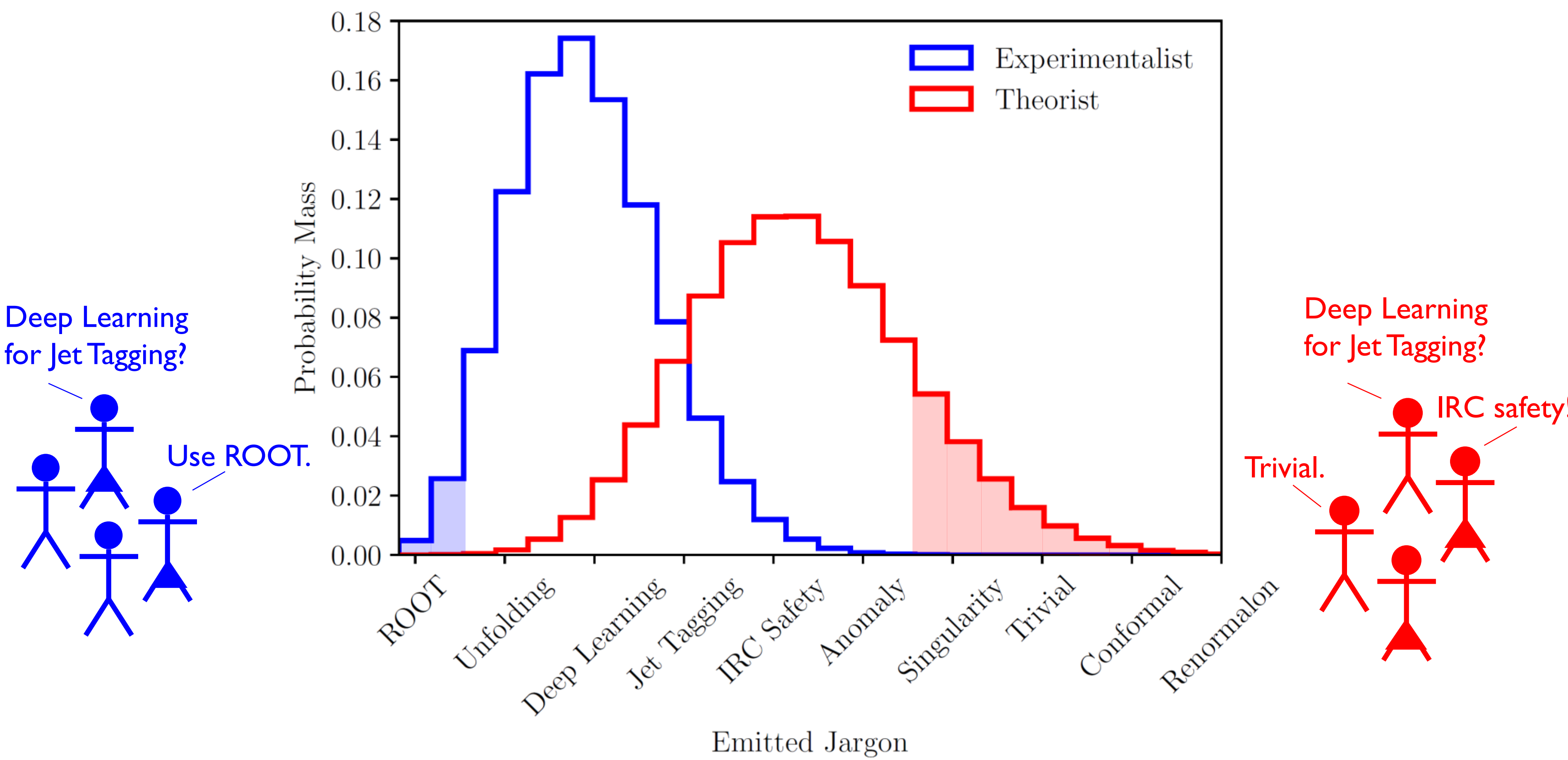
The term "energy conservation" appears in Physics papers and in Climate Science papers.

However, only Physics papers contain "Noether's Theorem" and only Climate Science papers contain "Kyoto Protocol". These are the anchor words.

Hence Physics and Climate Science are mutually irreducible topics.

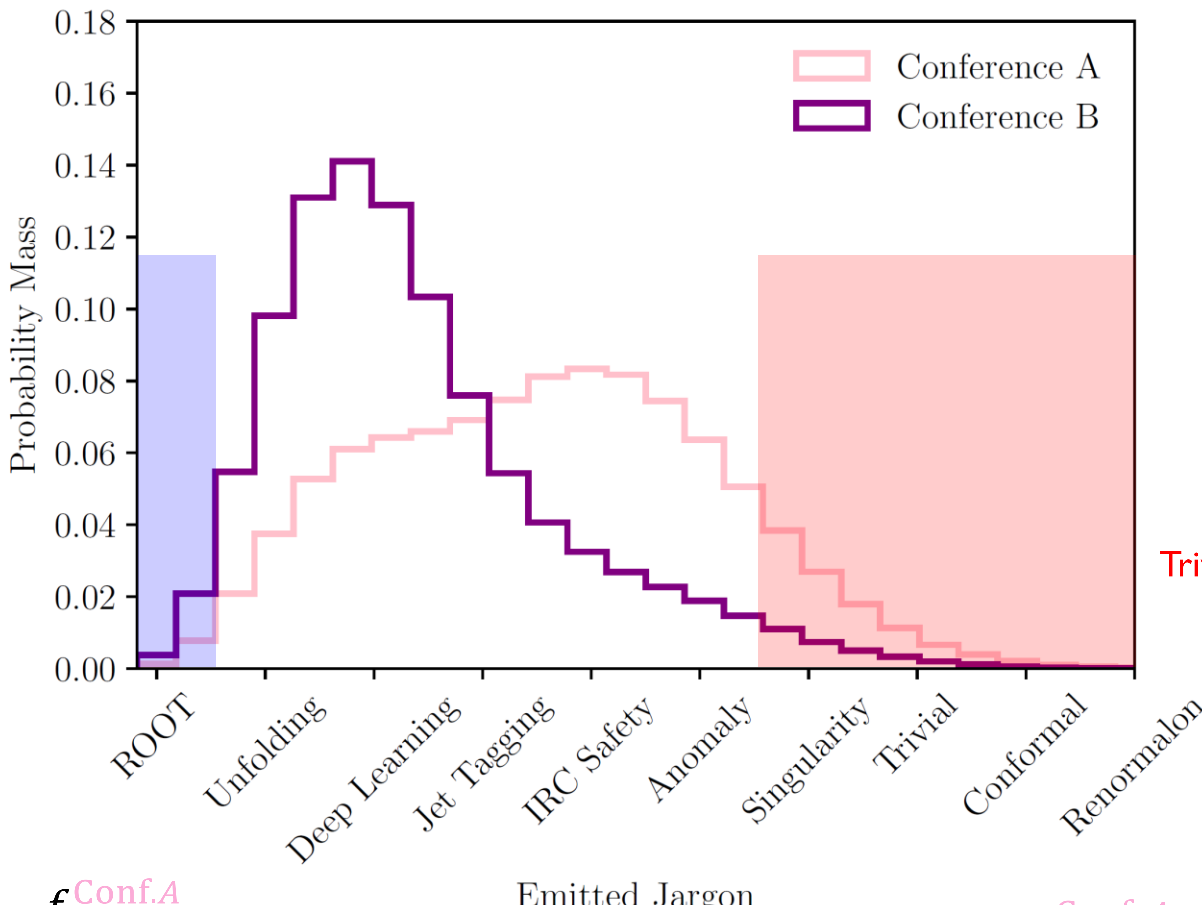
An Example

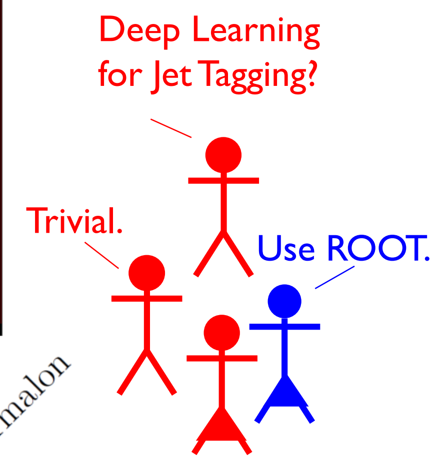
Let's model physicists as random jargon emitters.



An Example

Listen to the jargon emitted from two different conferences.



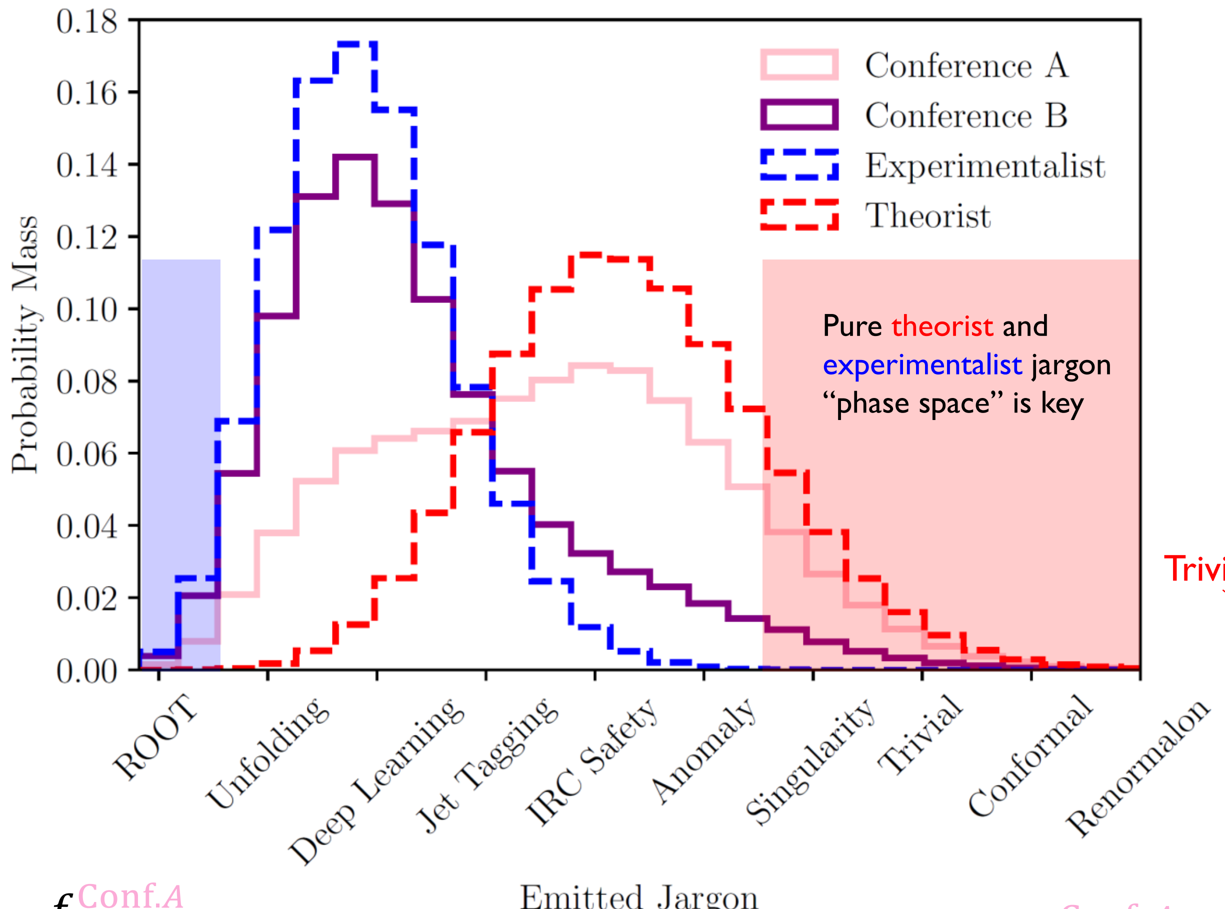


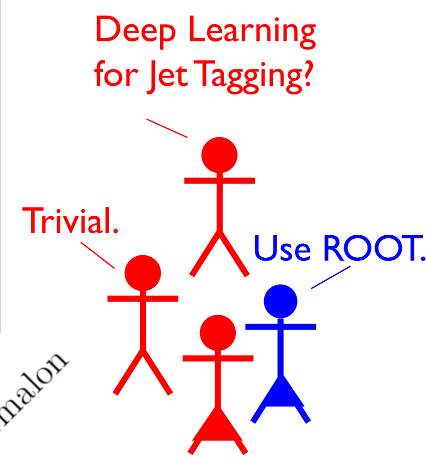
$$\frac{N^{\text{Conf. }A}_{\text{"ROOT"}}}{N^{\text{Conf. }B}_{\text{"ROOT"}}} = \frac{f^{\text{Conf.}A}_{\text{Expt.}}}{f^{\text{Conf.}B}_{\text{Expt.}}}$$

$$\frac{N^{\text{Conf. }A}_{\text{"Trivial"}}}{N^{\text{Conf. }B}_{\text{"Trivial"}}} = \frac{1 - f^{\text{Conf.}A}_{\text{Expt.}}}{1 - f^{\text{Conf.}B}_{\text{Expt.}}}$$

An Example

Disentangle theorist and experimentalist vocabularies from the jargon at conferences.



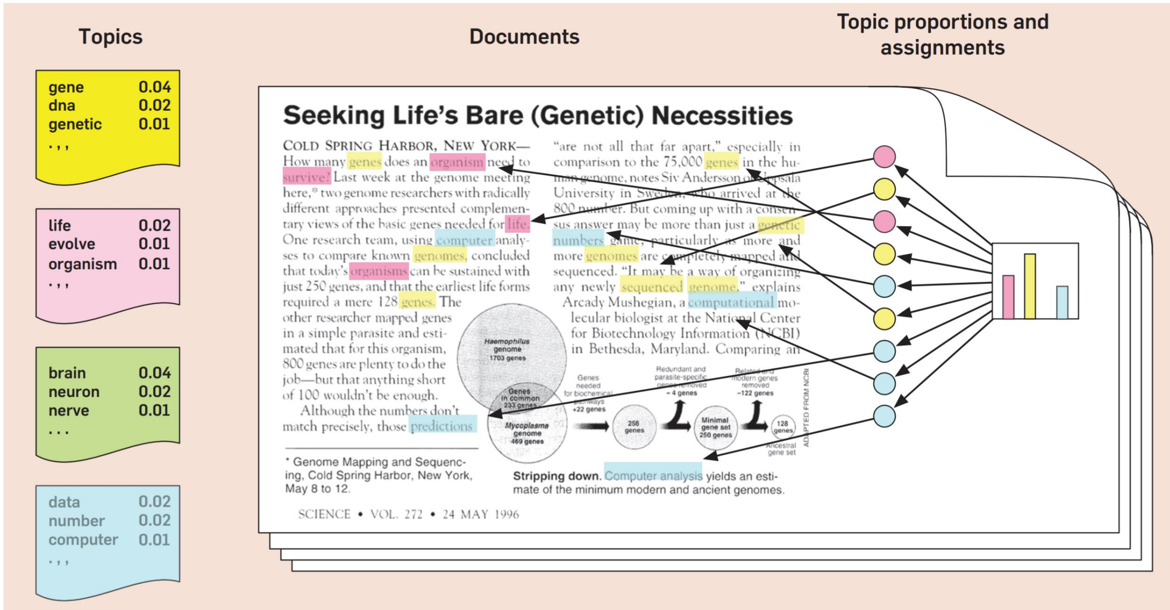


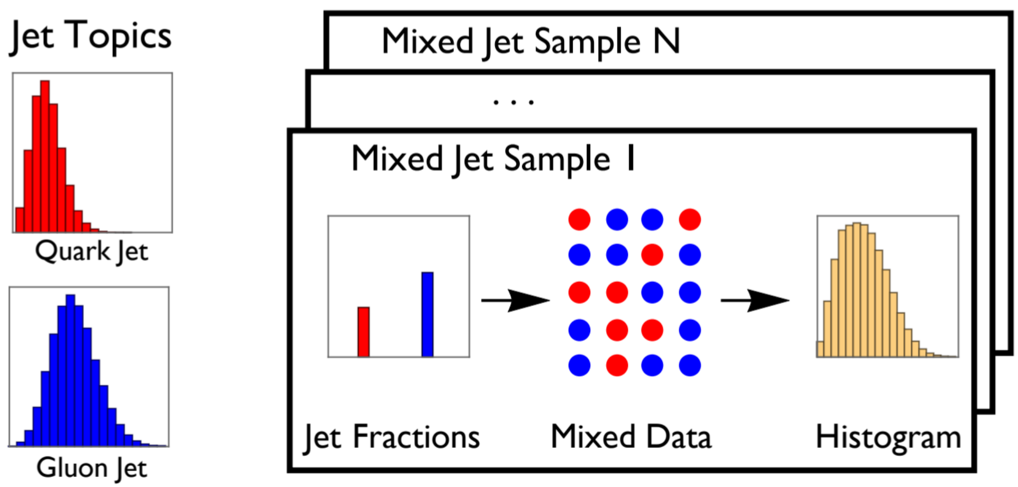
$$\frac{N^{\text{Conf. }A}_{\text{"ROOT"}}}{N^{\text{Conf. }B}_{\text{"ROOT"}}} = \frac{f^{\text{Conf.}A}_{\text{Expt.}}}{f^{\text{Conf.}B}_{\text{Expt.}}}$$

$$\frac{N^{\text{Conf. }A}_{\text{"Trivial"}}}{N^{\text{Conf. }B}_{\text{"Trivial"}}} = \frac{1 - f^{\text{Conf.}A}_{\text{Expt.}}}{1 - f^{\text{Conf.}B}_{\text{Expt.}}}$$

Collider data as mixtures of jet types

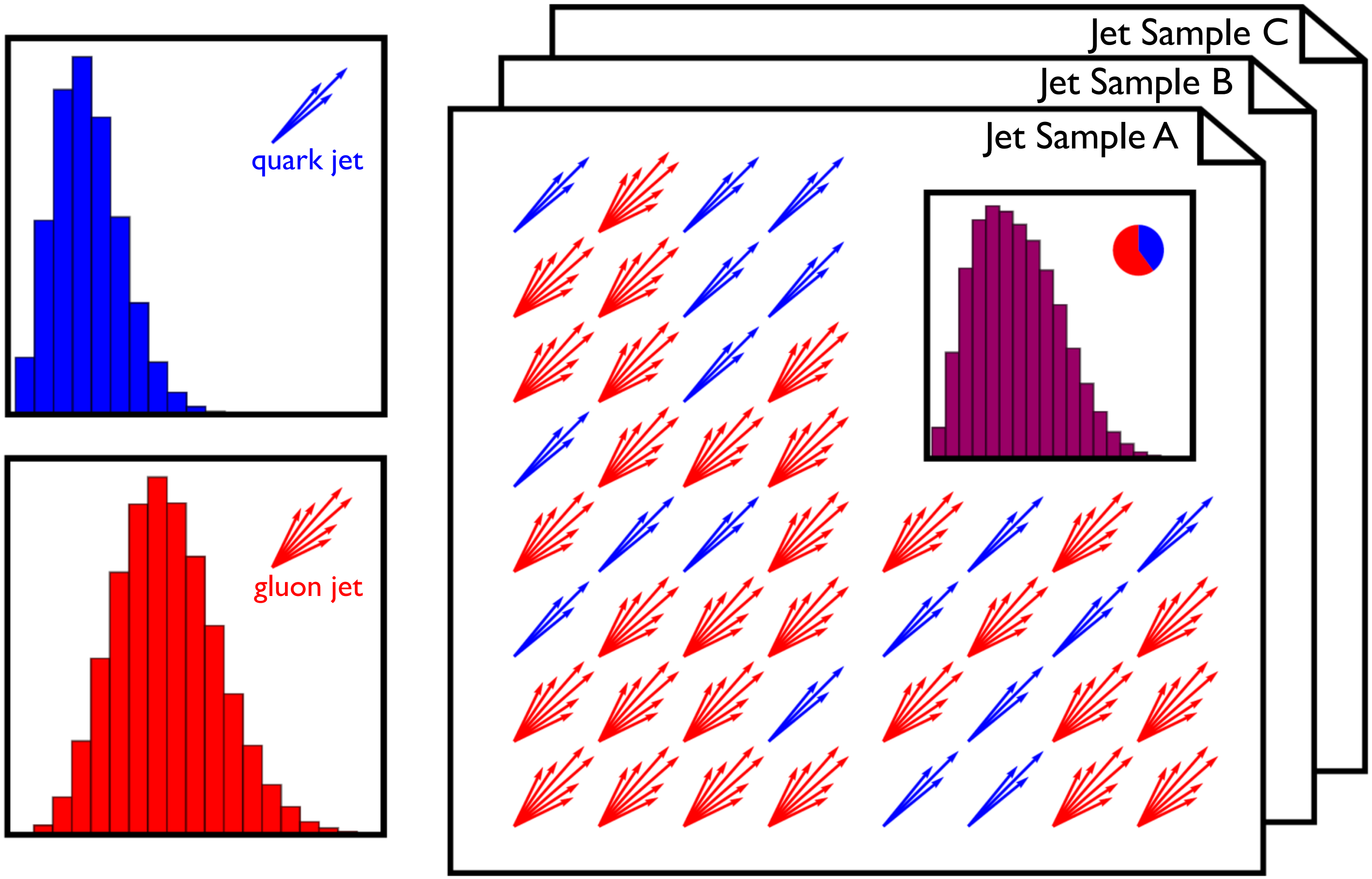
A mathematical correspondence between topic models and jet distributions.





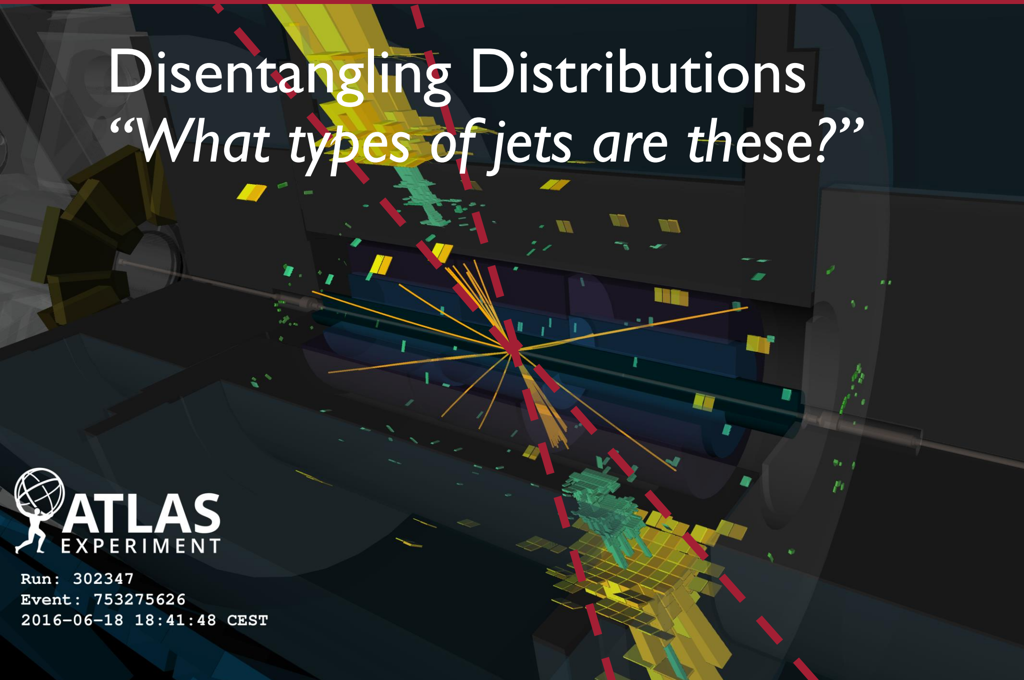
Collider data as mixtures of jet types

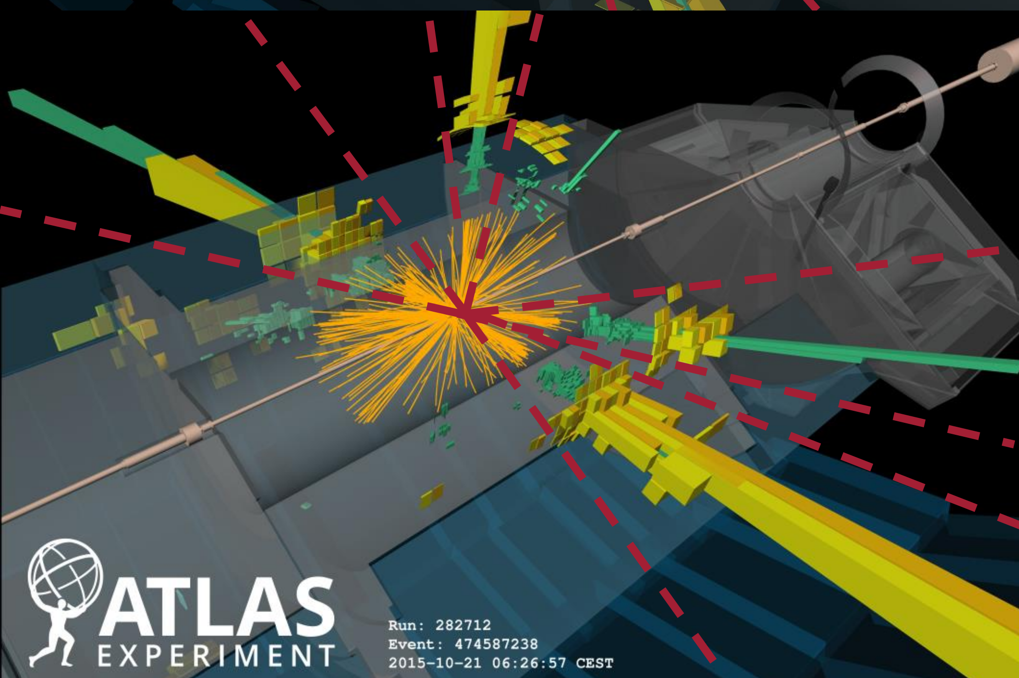
This is an unfamiliar way to think about machine learning and jet physics.



We are going to use observables and model outputs *not* as classifiers, but as feature spaces to extract mixture fractions.

Disentangling Distributions
"What types of jets are these?"





§ Menu §

Take your favorite jet algorithm

Anti-kT R=0.4

Consider *multiple* jet samples

Sample A: Z + jet
Sample B: dijets

Select a substructure feature space

Constituent Multiplicity
Jet Mass
Soft Drop Multiplicity
Model Output

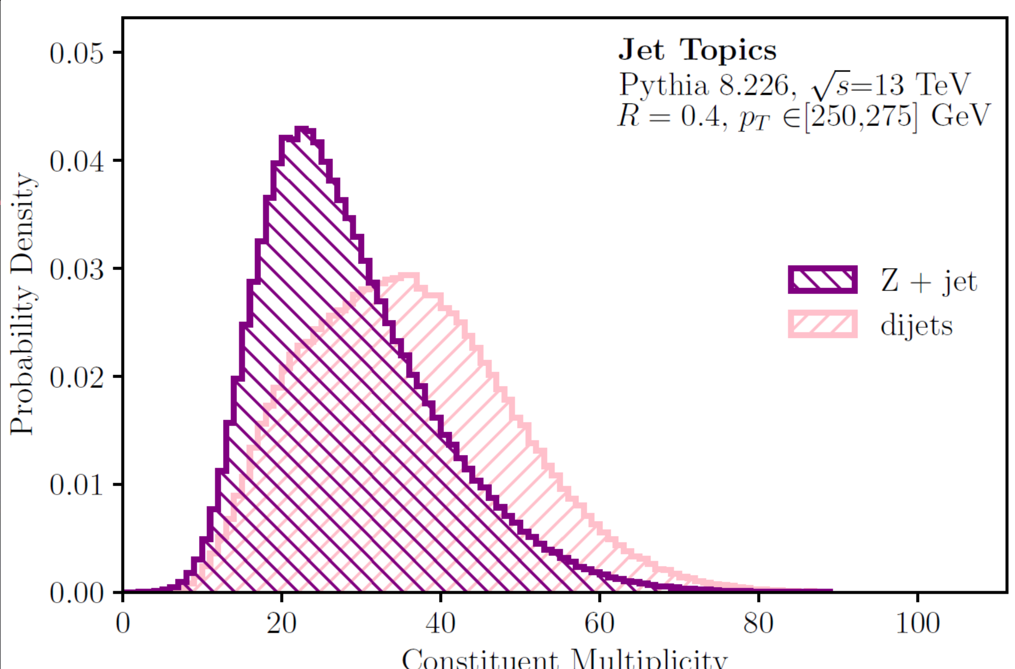
Goal: Find the underlying categories which explain the variation in substructure among the samples.

Disentangling Distributions

"What types of jets are these?"

$$p_{\text{sample }A}(\boldsymbol{x}) = f_A^q\, p_{\text{quark}}(\boldsymbol{x}) + \left(1 - f_A^q\right) p_{\text{gluon}}(\boldsymbol{x})$$

$$p_{\text{sample }B}(\boldsymbol{x}) = f_B^q\, p_{\text{quark}}(\boldsymbol{x}) + \left(1 - f_B^q\right) p_{\text{gluon}}(\boldsymbol{x})$$



Menu

Take your favorite jet algorithm

Anti-kT R=0.4

Consider *multiple* jet samples

Sample A: Z + jet
Sample B: dijets

Select a substructure feature space

Constituent Multiplicity
Jet Mass
Soft Drop Multiplicity
Model Output

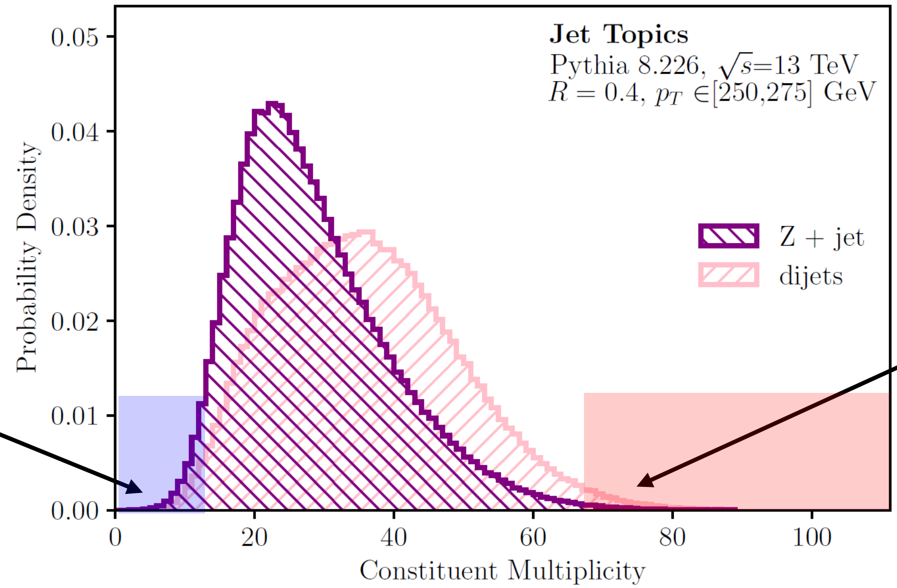
Goal: Find the underlying categories which explain the variation in substructure among the samples.

Demixing the mixtures

$$p_A(\boldsymbol{x}) = f_A^q \, p_{\text{quark}}(\boldsymbol{x}) + \left(1 - f_A^q\right) p_{\text{gluon}}(\boldsymbol{x})$$

$$p_B(\boldsymbol{x}) = f_B^q \, p_{\text{quark}}(\boldsymbol{x}) + \left(1 - f_B^q\right) p_{\text{gluon}}(\boldsymbol{x})$$

$$\kappa_{\text{BA}} \equiv \min_{\boldsymbol{x}} \frac{p_B(\boldsymbol{x})}{p_A(\boldsymbol{x})} = \frac{f_B^q}{f_A^q}$$



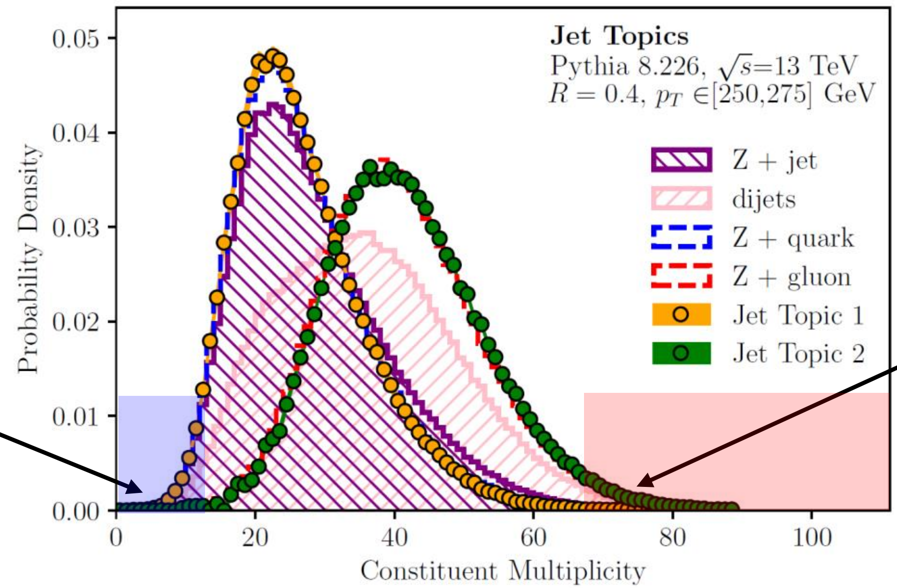
$$\kappa_{\text{AB}} \equiv \min_{\boldsymbol{x}} \frac{p_A(\boldsymbol{x})}{p_B(\boldsymbol{x})} = \frac{1-f_A^q}{1-f_B^q}$$

Demixing the mixtures

$$p_A(\boldsymbol{x}) = f_A^q\, p_{\text{quark}}(\boldsymbol{x}) + \left(1 - f_A^q\right) p_{\text{gluon}}(\boldsymbol{x})$$

$$p_B(\boldsymbol{x}) = f_B^q\, p_{\text{quark}}(\boldsymbol{x}) + \left(1 - f_B^q\right) p_{\text{gluon}}(\boldsymbol{x})$$

$$\kappa_{\text{BA}} \equiv \min_{\boldsymbol{x}} \frac{p_B(\boldsymbol{x})}{p_A(\boldsymbol{x})} = \frac{f_B^q}{f_A^q}$$



$$\kappa_{\text{AB}} \equiv \min_{\boldsymbol{x}} \frac{p_A(\boldsymbol{x})}{p_B(\boldsymbol{x})} = \frac{1-f_A^q}{1-f_B^q}$$

With **reducibility factors** κ_{AB} and κ_{BA}, solve for the quark and gluon fractions and distributions:

$$f_A^q = \frac{1-\kappa_{\text{AB}}}{1-\kappa_{\text{AB}}\kappa_{\text{BA}}} \qquad f_{\text{B}}^q = \frac{\kappa_{\text{BA}}(1-\kappa_{\text{AB}})}{1-\kappa_{\text{AB}}\kappa_{\text{BA}}}$$

$$p_{\text{quark}}(\boldsymbol{x}) = \frac{p_A(\boldsymbol{x}) - \kappa_{\text{AB}}\, p_B(\boldsymbol{x})}{1-\kappa_{\text{AB}}} \qquad p_{\text{gluon}}(\boldsymbol{x}) = \frac{p_B(\boldsymbol{x}) - \kappa_{\text{BA}}\, p_A(\boldsymbol{x})}{1-\kappa_{\text{BA}}}$$

Exploring substructure feature spaces

Why restrict ourselves to multiplicity? It works, but we can explore this choice.

We can also use a *trained classifier* (with CWoLa) as an observable in its own right.

Observables

- Multiplicity n_{const}
  Number of particles in the jet
- Soft Drop Multiplicity n_{SD}
  Probes number of perturbative emissions
- Image Activity N_{95}
  Number of pixels with 95% of jet p_T
- N-subjettiness $\tau_2^{(\beta=1)}$
  Probes how multi-pronged the jet is
- Jet Mass m
  Mass of the total jet four-vector
- Width w
  Probes the girth of the jet

Models

- PFN-ID
  Full particle-level information
- PFN
  Full four-momentum information
- EFN
  Full IRC-safe information

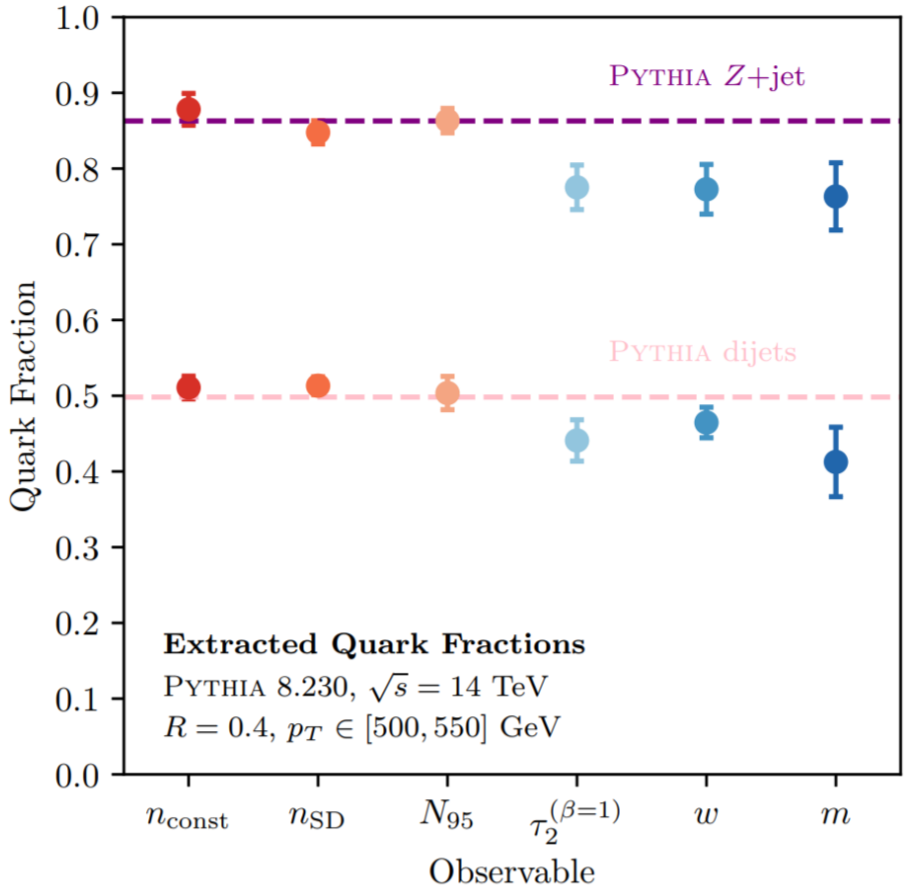
See Patrick's talk!

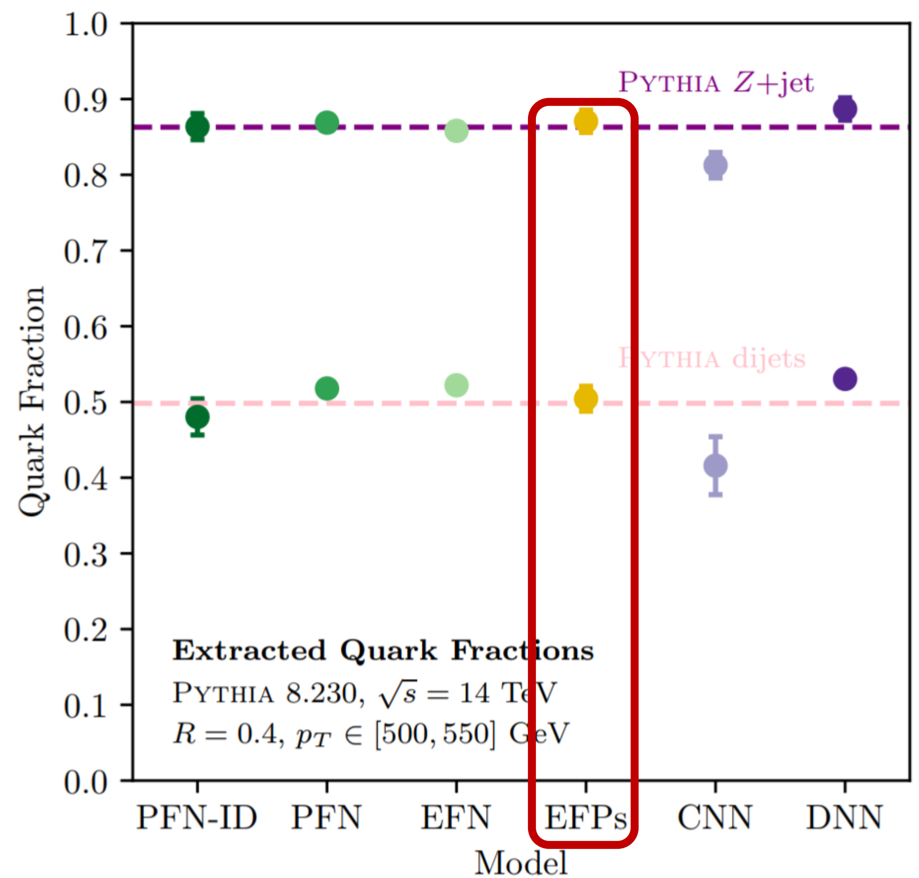
- EFPs
  Full IRC-safe information, linearly
- CNN
  Trained on two-channel jet images
- DNN
  Trained on an N-subjettiness basis

[P.T. Komiske, EMM, J. Thaler, 1810.05165]

Extracting quark and gluon fractions

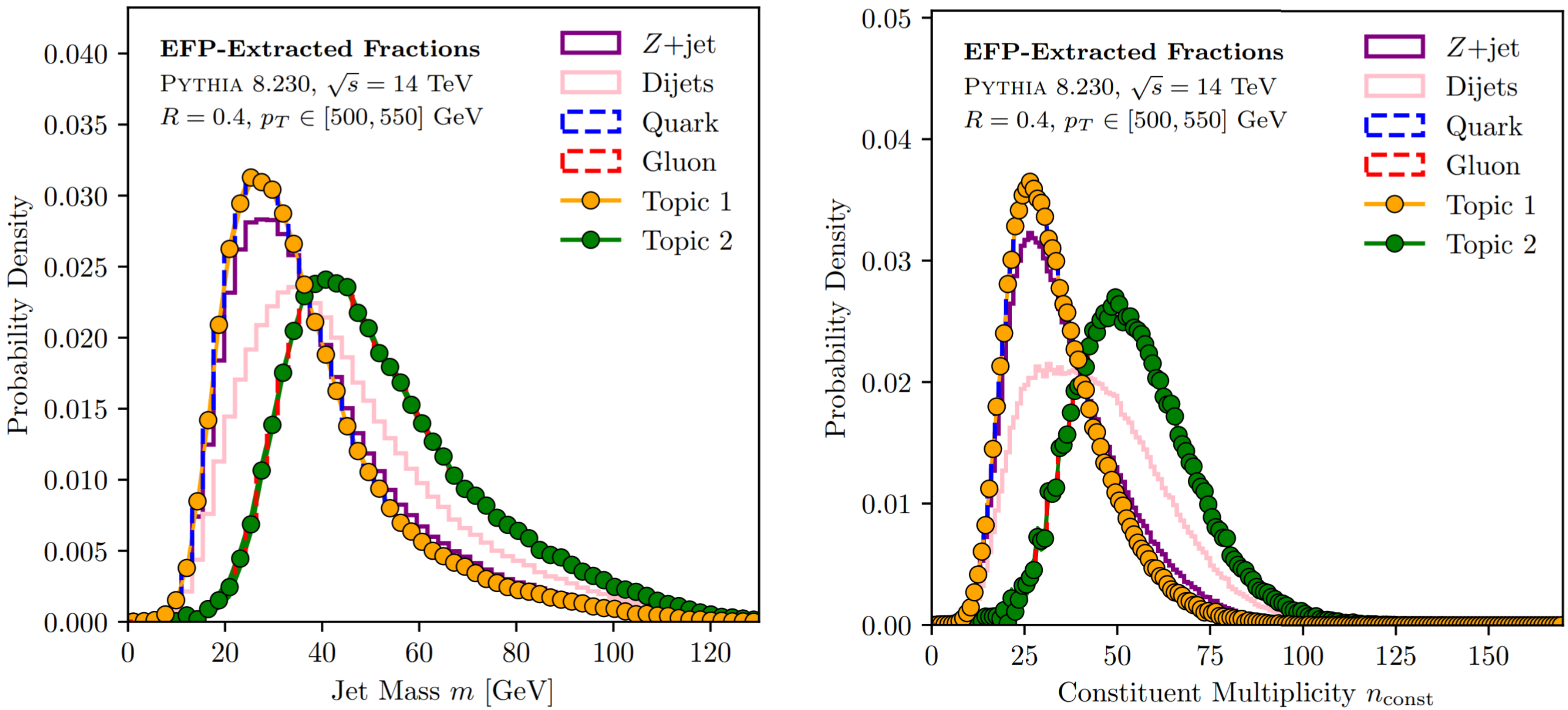
With the topics procedure, the quark and gluon fractions of the samples can be obtained.





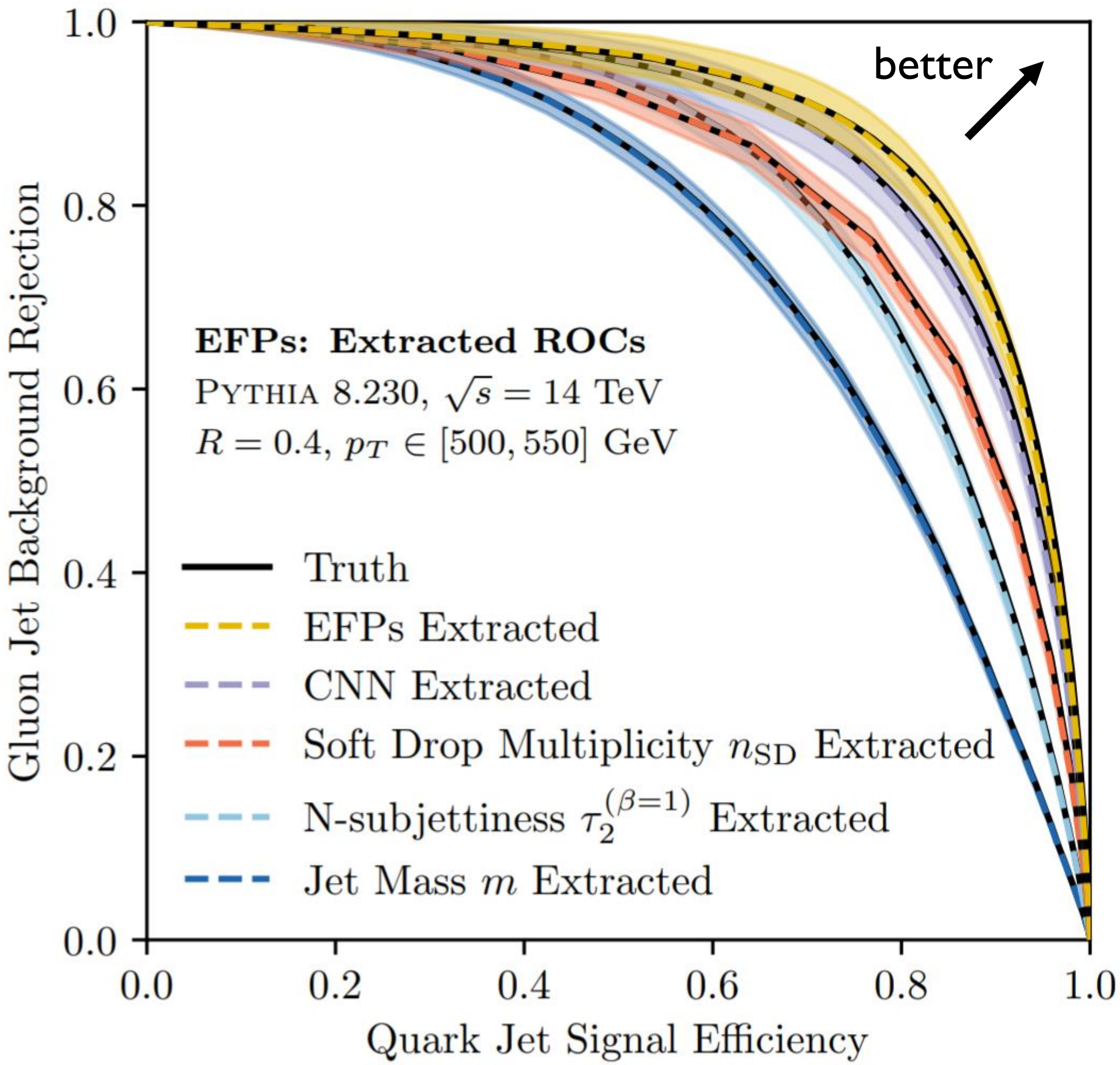
Extracting quark and gluon distributions

The extracted quark and gluon fractions can be used to obtain any quark/gluon distributions.



(Self-)calibrating quark and gluon classifiers

The extracted quark and gluon fractions can calibrate any data-driven quark/gluon classifiers.



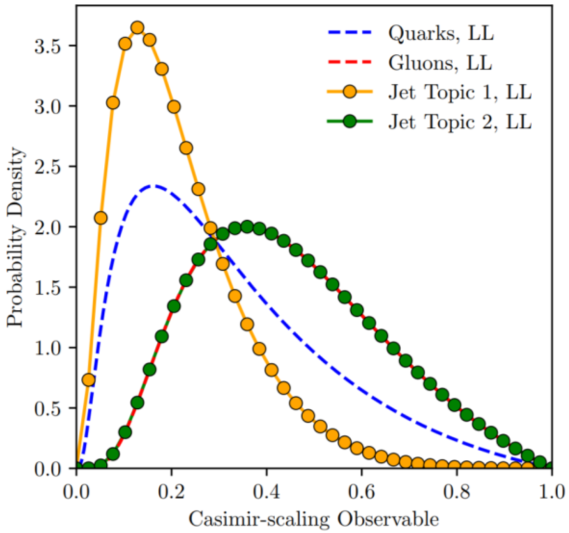
Jet topics from perturbative QCD

Topic modeling for jets can be understood and calculated from perturbative QCD.

Jet mass (like many shape observables) exhibits **Casimir scaling** at Leading Logarithmic accuracy:

$$\Sigma_g(m) = \Sigma_q(m)^{\frac{C_A}{C_F}}$$

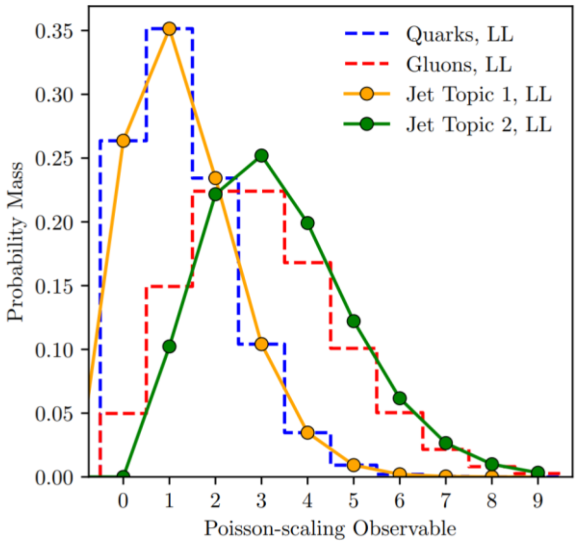
$$\boxed{\kappa_{qg}^{\text{Cas.}} = \frac{C_F}{C_A} = \frac{4}{9} \qquad \kappa_{gq}^{\text{Cas.}} = 0}$$



Soft Drop Multiplicity (like many count observables) exhibits **Poisson scaling** at Leading Logarithmic accuracy:

$$p_q(n) = \text{Pois}(n; C_F\lambda),$$
$$p_g(n) = \text{Pois}(n; C_A\lambda).$$

$$\boxed{\kappa_{qg}^{\text{Pois.}} = 0 \qquad \kappa_{gq}^{\text{Pois.}} = e^{\lambda(C_F - C_A)}}$$



See back-up slides for more.

An operational definition of quark and gluon jets

Quark and Gluon Jet Definition (Operational): Given two samples A and B of QCD jets at a fixed p_T obtained by a suitable jet-finding procedure, taking A to be "quark-enriched" compared to B, and a jet substructure feature space $\boldsymbol{x}$, quark and gluon jet distributions are defined to be:

$$p_{\text{quark}}(\boldsymbol{x}) \equiv \frac{p_A(\boldsymbol{x}) - \kappa_{\text{AB}}\, p_B(\boldsymbol{x})}{1 - \kappa_{\text{AB}}} \qquad p_{\text{gluon}}(\boldsymbol{x}) \equiv \frac{p_B(\boldsymbol{x}) - \kappa_{\text{BA}}\, p_A(\boldsymbol{x})}{1 - \kappa_{\text{BA}}}$$

Well-defined and operational statement in terms of hadronic cross sections.

Not a per-jet flavor label, but rather an aggregate distribution label.

Jets themselves are operationally defined.

Sample dependence?

How does "sample dependence" manifest in this language?

Pairs of samples define quark and gluon.

Different pairs of samples may yield different flavor definitions.

Comparing definitions from different pairs of samples (dijets, Z+jet, gamma+jet, …) in data could probe how universal quark and gluon are. Can grooming improve this?

There are ways to quantify how "explainable" a new sample C is by quark and gluon:

$$\max(f^q + f^g) \quad \text{s.t.} \;\; p_C(\boldsymbol{x}) = f^q\, p_q(\boldsymbol{x}) + f^g\, p_g(\boldsymbol{x}) + (1 - f^q - f^g)\, p_{\text{other}}(\boldsymbol{x})$$

Thus topic modeling techniques could be an interesting avenue to explore issues of sample dependence directly in data.

Summary

Jet categories can be extracted (or defined!) using topic modeling ideas.

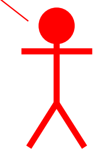
In our two-category case, this allows quark and gluon jet distributions to be measured separately without fractions or templates:

$$p_{\text{quark}}(\boldsymbol{x}) \equiv \frac{p_A(\boldsymbol{x}) - \kappa_{\text{AB}}\, p_B(\boldsymbol{x})}{1-\kappa_{\text{AB}}} \qquad p_{\text{gluon}}(\boldsymbol{x}) \equiv \frac{p_B(\boldsymbol{x}) - \kappa_{\text{BA}}\, p_A(\boldsymbol{x})}{1-\kappa_{\text{BA}}}$$

$$\kappa_{\text{AB}} \equiv \min_{\boldsymbol{x}} \frac{p_A(\boldsymbol{x})}{p_B(\boldsymbol{x})} \qquad \kappa_{\text{BA}} \equiv \min_{\boldsymbol{x}} \frac{p_B(\boldsymbol{x})}{p_A(\boldsymbol{x})}$$

These methods are theoretically tractable and operate directly in terms of hadronic cross sections.

Trivial.

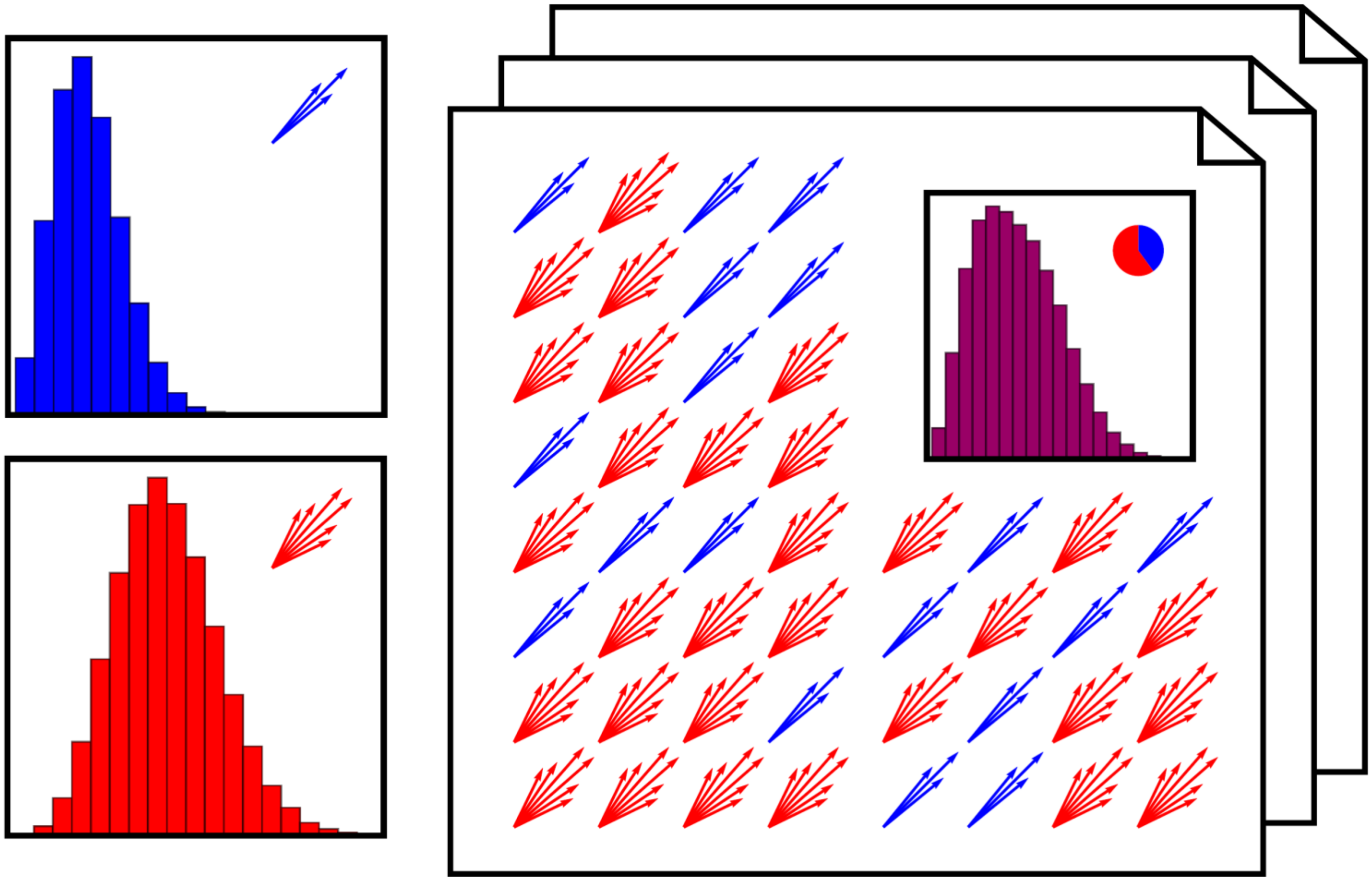


Can we do this for more categories?

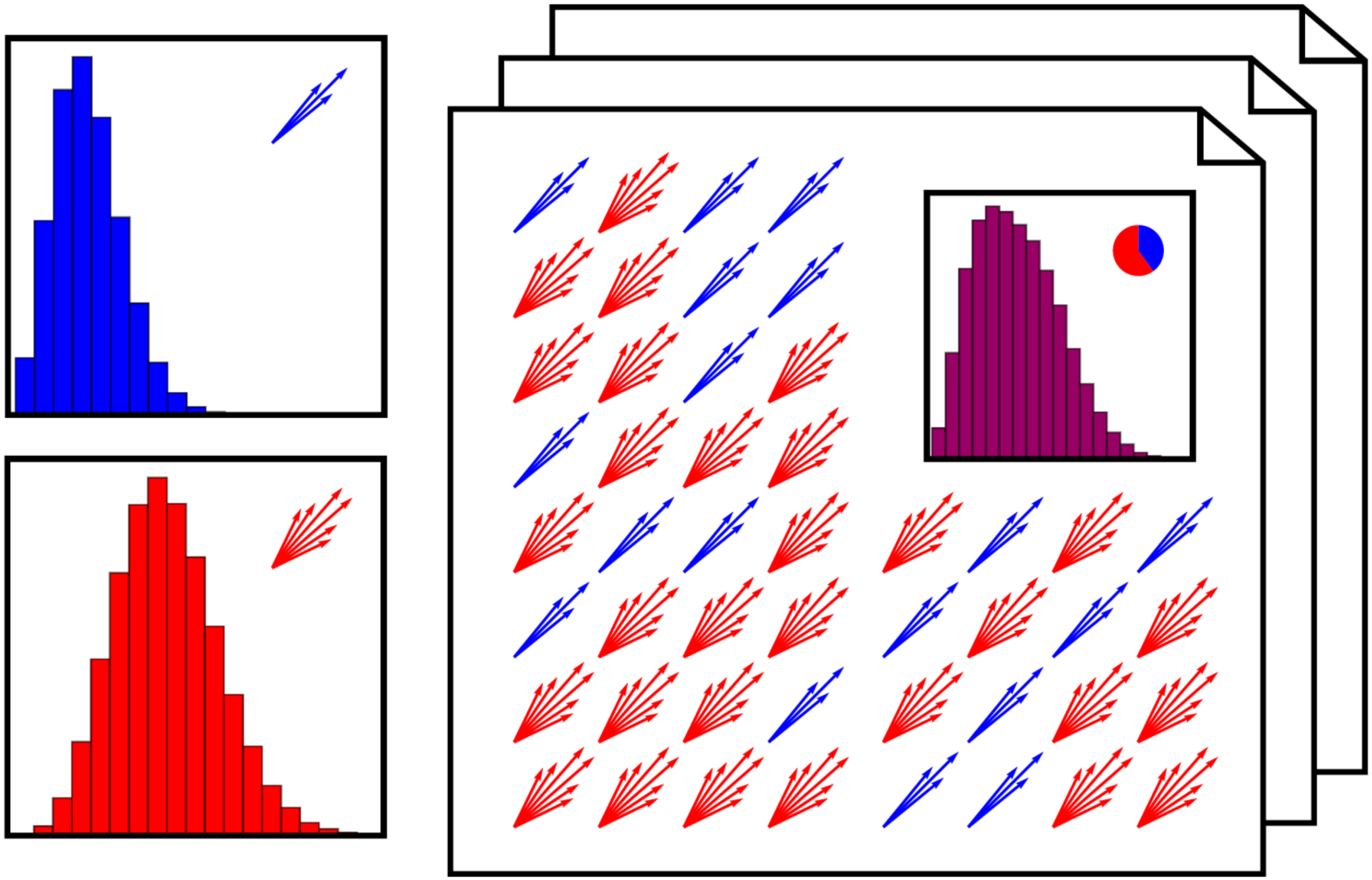
- Need to specify (using experts or ML) a pure phase space region for each category.
- Need different mixtures of the different categories.
- Something new to think about!

The End

Thank you!



Extra Slides

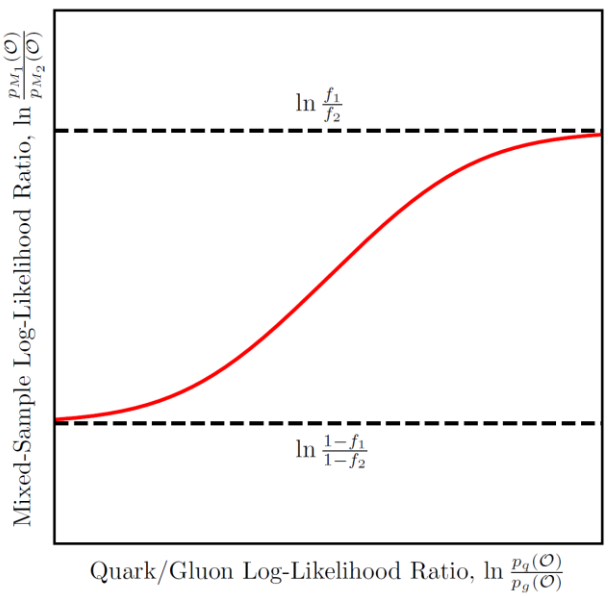


A/B Likelihood Ratio

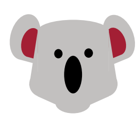
$$p_{\text{sample } A}(\boldsymbol{x}) = f_A^q \, p_{\text{quark}}(\boldsymbol{x}) + \left(1 - f_A^q\right) p_{\text{gluon}}(\boldsymbol{x})$$

$$p_{\text{sample } B}(\boldsymbol{x}) = f_B^q \, p_{\text{quark}}(\boldsymbol{x}) + \left(1 - f_B^q\right) p_{\text{gluon}}(\boldsymbol{x})$$

$$\boxed{L_{\frac{\mathrm{A}}{\mathrm{B}}}(\boldsymbol{x}) \equiv \frac{p_A(\boldsymbol{x})}{p_B(\boldsymbol{x})} = \frac{f_A^q \, L_{\frac{\text{quark}}{\text{gluon}}}(\boldsymbol{x}) + \left(1 - f_A^q\right)}{f_B^q \, L_{\frac{\text{quark}}{\text{gluon}}}(\boldsymbol{x}) + \left(1 - f_B^q\right)}}$$



The A/B and quark/gluon likelihood ratios are monotonic!

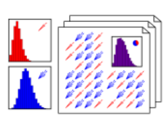


Classification without labels (CWoLa)

- Optimal A/B classifier is the optimal quark/gluon classifier.
- Use machine learning to approximate A/B likelihood ratio.

[EMM, B. Nachman, J. Thaler, 1708.02949]

The A/B likelihood ratio is bounded between $\frac{f_A^q}{f_B^q}$ and $\frac{1-f_A^q}{1-f_B^q}$!



Jet Topics

- "Mutually irreducibility" means the bounds saturate
- Obtain the maxima and minima of the A/B likelihood ratio.
- Solve for the quark/gluon fractions and distributions.

[EMM, J. Thaler, 1802.00008]

Jet topics from QCD: Casimir scaling

Jet mass (and many substructure observables) exhibits Casimir scaling at Leading Logarithmic accuracy:

$$\Sigma_g(m) = \Sigma_q(m)^{\frac{C_A}{C_F}}$$

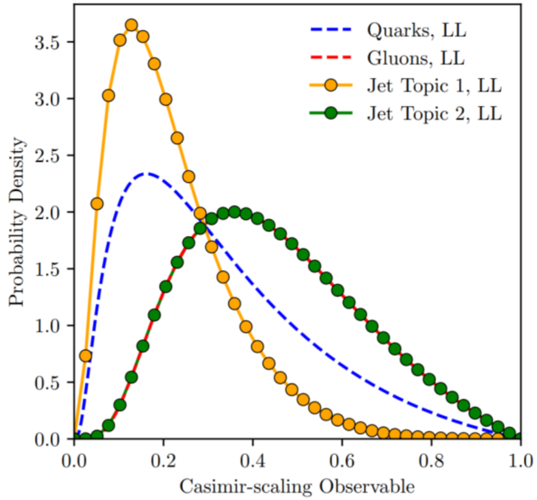
$C_F = \frac{4}{3}$ for quarks

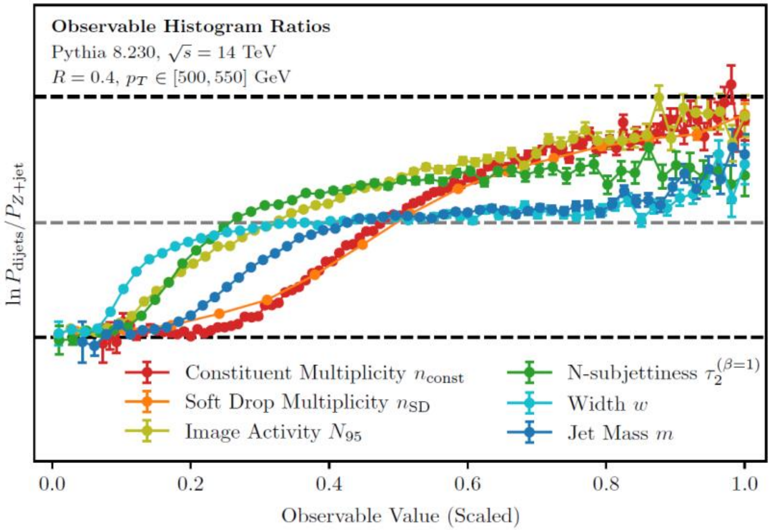
$C_A = 3$ for gluons

The quark/gluon reducibility factors at LL for any Casimir scaling observable are:

$$\kappa_{gq} = \min_m \frac{p_g(m)}{p_q(m)} = \min_m \frac{\Sigma_g{}'(m)}{\Sigma_q{}'(m)} = \frac{C_A}{C_F} \min_m \Sigma_q'(m)^{\frac{C_A}{C_F}-1} = 0$$

$$\kappa_{qg} = \min_m \frac{p_q(m)}{p_g(m)} = \min_m \frac{\Sigma_q{}'(m)}{\Sigma_g{}'(m)} = \frac{C_F}{C_A} \min_m \Sigma_q'(m)^{1-\frac{C_A}{C_F}} = \frac{C_F}{C_A} = \frac{4}{9}$$





Jet topics from QCD: Poisson scaling

Soft Drop Multiplicity (and other count observables) exhibits Poisson scaling at Leading Logarithmic accuracy:

$C_F = \frac{4}{3}$ for quarks

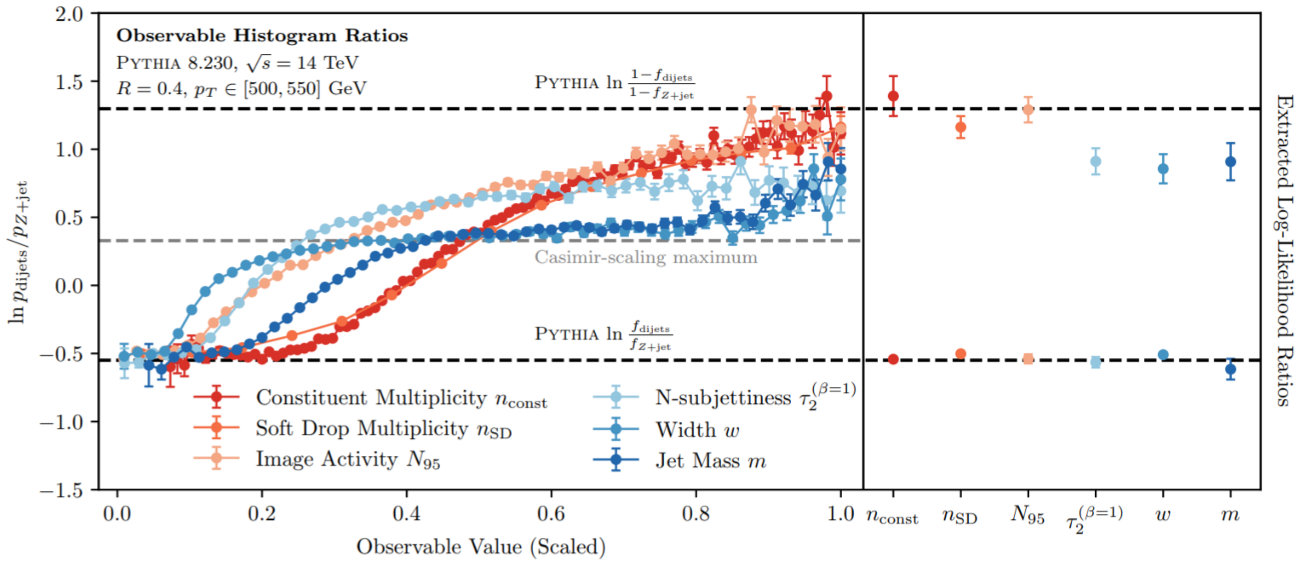
$C_A = 3$ for gluons

$$p_q(n) = \mathrm{Pois}(n; C_F\lambda), \qquad p_g(n) = \mathrm{Pois}(n; C_A\lambda).$$

The quark/gluon reducibility factors at LL for any Poisson scaling observable are:

$$\boxed{\begin{aligned} \kappa_{gq} &= \min_n \frac{p_g(n)}{p_q(n)} = \min_n \frac{(C_A\lambda)^n \ e^{-C_A\lambda}}{(C_F\lambda)^n \ e^{-C_F\lambda}} = e^{\lambda(C_F - C_A)} \min_n \left(\frac{C_A}{C_F}\right)^n = e^{\lambda(C_F - C_A)} \\ \kappa_{qg} &= \min_n \frac{p_q(n)}{p_g(n)} = \min_n \frac{(C_F\lambda)^n \ e^{-C_F\lambda}}{(C_A\lambda)^n \ e^{-C_A\lambda}} = e^{\lambda(C_A - C_F)} \min_n \left(\frac{C_F}{C_A}\right)^n = 0 \end{aligned}}$$

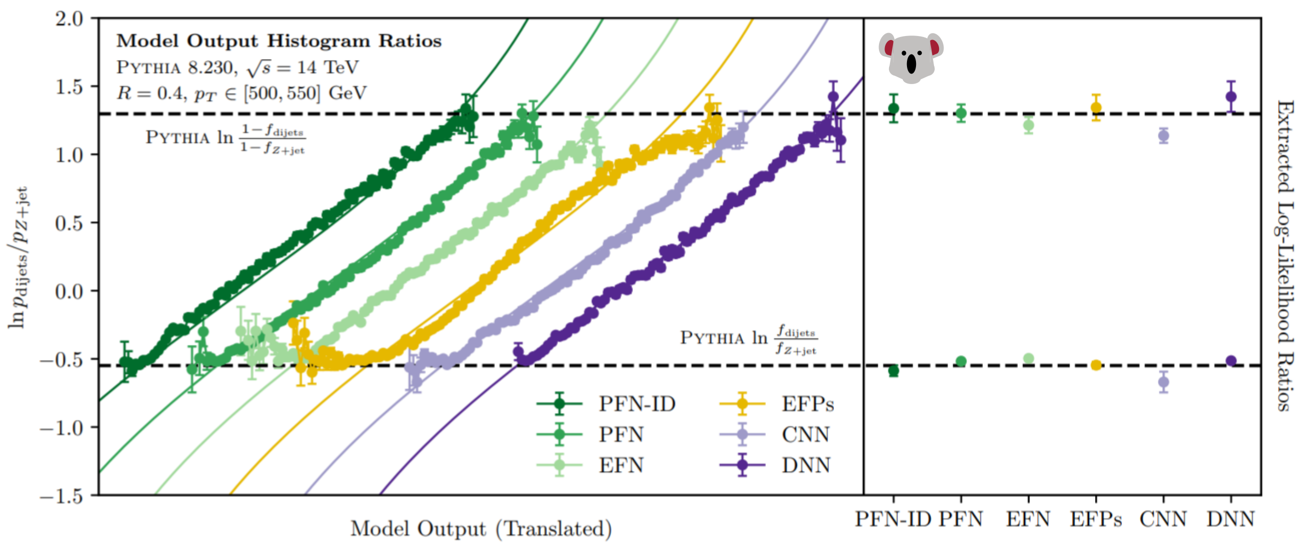
Exploring substructure feature spaces



Casimir scaling of mass and width is observed (gray).

Count observables come closer to saturating the bounds (black).

Lower bound easier to extract than upper. (i.e. Gluons are easy!)



Models CWoLa-trained. Fully data-driven.

Well-behaved likelihoods close to S/(S+B) expectation.

All different models manifest the same bounds.

Parton-labeled sample dependence in Pythia