On the Topic of Jets

BOOST 2018

Eric M. Metodiev
Center for Theoretical Physics
Massachusetts Institute of Technology

Joint work with Patrick T. Komiske and Jesse Thaler

July 19, 2018
**Quark and Gluon Jets**

**Quarks** are color triplets and **Gluons** are color octets. We observe color-singlet hadrons.

No unambiguous hadron-level definition of jet flavor.

We often rely on unphysical notions such as parton shower event records to define jet flavor in practice.

Can **quark** and **gluon** be made well-defined nonetheless? Similar to defining jets themselves.

Ubiquitous concepts. From BOOST 2018 so far:
What are “Quark” and “Gluon” Jets?

What is a Quark Jet?

*From lunch/dinner discussions*

- **Ill-Defined**
  - What people sometimes think we mean
  - A quark parton
  - A Born-level quark parton
  - The initiating quark parton in a final state shower
  - An eikonal line with baryon number $1/3$ and carrying triplet color charge
  - A quark operator appearing in a hard matrix element in the context of a factorization theorem
  - A parton-level jet object that has been quark-tagged using a soft-safe flavored jet algorithm (automatically collinear safe if you sum constituent flavors)
  - A phase space region (as defined by an unambiguous hadronic fiducial cross section measurement) that yields an enriched sample of quarks (as interpreted by some suitable, though fundamentally ambiguous, criterion)

- **Well-Defined**
  - What we mean

[Les Houches 2015 Report]
[P. Gras, et al., 1704.03878]
Systematics of quark/gluon tagging

Philippe Gras, Stefan Höche, Deepak Kar, Andrew Larkoski, Leif Lonnblad, Simon Plätzer, Andrzej Siódmok, Peter Skands, Gregory Soyez, and Jesse Thaler

2 What is a quark/gluon jet?

The definition we adopt for this study is inspired by the idea that one should think about quark/gluon tagging in the context of a specific measurement, regardless of whether the observable in question has a rigorous factorization theorem.

- A phase space region (as defined by an unambiguous hadronic fiducial cross section measurement) that yields an enriched sample of quarks (as interpreted by some suitable, though fundamentally ambiguous, criterion). Here, the goal is to tag a phase space region as being quark-like, rather than try to determine a truth definition of a quark. This definition has the advantage of being explicitly tied to hadronic final states and to the discriminant variables of interest. The main challenge with this definition is how to determine the criterion that corresponds to successful quark enrichment. For that, we have to rely to some degree on the other less well-defined notions of what a quark jet is.
Systematics of quark/gluon tagging

Philippe Gras, Stefan Höche, Deepak Kar, Andrew Larkoski, Leif Lonnblad, Simon Plätzer, Andrzej Siödmok, Peter Skands, Gregory Soyez, and Jesse Thaler

2 What is a quark/gluon jet?

The definition we adopt for this study is inspired by the idea that one should think about quark/gluon tagging in the context of a specific measurement, regardless of whether the observable in question has a rigorous factorization theorem.

- A phase space region (as defined by an unambiguous hadronic fiducial cross section measurement) that yields an enriched sample of quarks (as interpreted by some suitable, though fundamentally ambiguous, criterion). Here, the goal is to tag a phase space region as being quark-like, rather than try to determine a truth definition of a quark. This definition has the advantage of being explicitly tied to hadronic final states and to the discriminant variables of interest. The main challenge with this definition is how to determine the criterion that corresponds to successful quark enrichment. For that, we have to rely to some degree on the other less well-defined notions of what a quark jet is.

To define the truth-level jet flavor, we use a simple definition: a quark jet is a jet produced by a parton-shower event generator in $e^+e^- \rightarrow (\gamma/Z)^* \rightarrow u\bar{u}$ hard scattering, while a gluon jet is a jet produced in $e^+e^- \rightarrow h^* \rightarrow gg$. 

[P. Gras, et al., 1704.03878]
Our Plan: An operational definition of quark and gluon jets

That definition:

[A quark jet is defined by:]
A phase space region (as defined by an unambiguous hadronic fiducial cross section measurement) that yields an enriched sample of quarks (as interpreted by some suitable, though fundamentally ambiguous, criterion)

This talk: Translating those 30 words to these 2 equations:

\[ p_{\text{quark}}(x) \equiv \frac{p_A(x) - \kappa_{AB} p_B(x)}{1 - \kappa_{AB}} \]
\[ p_{\text{gluon}}(x) \equiv \frac{p_B(x) - \kappa_{BA} p_A(x)}{1 - \kappa_{BA}} \]
A picture of quark and gluon jets

1. Take your favorite jet algorithm

2. Consider two jet samples A and B of QCD jets

3. Choose a jet substructure observable $x$

4. “Assume” that “quark” and “gluon” jets exist

5. “Assume” “quark/gluon” jet mutual irreducibility

The samples A and B are statistical mixtures of quark and gluon:

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

$$p_{\text{sample } B}(x) = f_B^q p_{\text{quark}}(x) + f_B^g p_{\text{gluon}}(x), \quad f_B^g = 1 - f_A^q$$

Similar picture to template- and fraction-based methods.
A/B Likelihood Ratio

\[ p_{\text{sample } A}(x) = f_A^q \, p_{\text{quark}}(x) + (1 - f_A^q) \, p_{\text{gluon}}(x) \]

\[ p_{\text{sample } B}(x) = f_B^q \, p_{\text{quark}}(x) + (1 - f_B^q) \, p_{\text{gluon}}(x) \]

\[ L_{A/B}(x) \equiv \frac{p_A(x)}{p_B(x)} = \frac{f_A^q \, L_{\text{quark}}(x) + (1 - f_A^q)}{f_B^q \, L_{\text{quark}}(x) + (1 - f_B^q)} \]

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.

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.
Systematics of quark/gluon tagging

Philippe Gras, Stefan Höche, Deepak Kar, Andrew Larkoski, Leif Lönnblad, Simon Plätzer, Andrzej Siódmok, Peter Skands, Gregory Soyez, and Jesse Thaler

To better understand this last definition, consider a quark/gluon discriminant $\lambda$.

For example, the user could choose that small $\lambda$ jets should be tagged as “quark-like” while large $\lambda$ jets should be tagged as “gluon-like”. Alternatively, the user might combine $\lambda$ with other discriminant variables as part of a more sophisticated classification scheme.

These concepts are not new in physics, and have been around for a while.

Quark/gluon mutual irreducibility: There are some substructure phase space regions where quark and gluon jets are pure.

$$\min_x \frac{p_B(x)}{p_A(x)} = \frac{f_B^q}{f_A^q}$$  $$\min_x \frac{p_A(x)}{p_B(x)} = \frac{1 - f_A^q}{1 - f_B^q}$$

[P. Gras, et al., 1704.03878]
Demixing the mixtures

\[ p_A(x) = f_A^q \ p_{\text{quark}}(x) + (1 - f_A^q) \ p_{\text{gluon}}(x) \]
\[ p_B(x) = f_B^q \ p_{\text{quark}}(x) + (1 - f_B^q) \ p_{\text{gluon}}(x) \]

\[ \kappa_{BA} \equiv \min_x \frac{p_B(x)}{p_A(x)} = \frac{f_B^q}{f_A^q} \]
\[ \kappa_{AB} \equiv \min_x \frac{p_A(x)}{p_B(x)} = \frac{1 - f_A^q}{1 - f_B^q} \]

Solve for the quark and gluon distributions and fractions:

\[ f_A^q = \frac{1 - \kappa_{AB}}{1 - \kappa_{AB} \kappa_{BA}} \]
\[ f_B^q = \frac{\kappa_{BA}(1 - \kappa_{AB})}{1 - \kappa_{AB} \kappa_{BA}} \]

\[ p_{\text{quark}}(x) = \frac{p_A(x) - \kappa_{AB} \ p_B(x)}{1 - \kappa_{AB}} \]
\[ p_{\text{gluon}}(x) = \frac{p_B(x) - \kappa_{BA} \ p_A(x)}{1 - \kappa_{BA}} \]
Demixing the mixtures

Defined from data

\[ p_A(x) = f_A^q p_{\text{quark}}(x) + (1 - f_A^q) p_{\text{gluon}}(x) \]
\[ p_B(x) = f_B^q p_{\text{quark}}(x) + (1 - f_B^q) p_{\text{gluon}}(x) \]

Ambiguous?

\[ \kappa_{BA} \equiv \min_x \frac{p_B(x)}{p_A(x)} \]
\[ = \frac{f_B^q}{f_A^q} \]

\[ \kappa_{AB} \equiv \min_x \frac{p_A(x)}{p_B(x)} \]
\[ = \frac{1 - f_A^q}{1 - f_B^q} \]

Solve for the quark and gluon distributions and fractions:

\[ f_A^q = \frac{1 - \kappa_{AB}}{1 - \kappa_{AB} \kappa_{BA}} \]
\[ f_B^q = \frac{\kappa_{BA}(1 - \kappa_{AB})}{1 - \kappa_{AB} \kappa_{BA}} \]
\[ p_{\text{quark}}(x) = \frac{p_A(x) - \kappa_{AB} p_B(x)}{1 - \kappa_{AB}} \]
\[ p_{\text{gluon}}(x) = \frac{p_B(x) - \kappa_{BA} p_A(x)}{1 - \kappa_{BA}} \]
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 $x$, quark and gluon jet distributions are defined to be:

$$p_{\text{quark}}(x) \equiv \frac{p_A(x) - \kappa_{AB} p_B(x)}{1 - \kappa_{AB}}$$

$$p_{\text{gluon}}(x) \equiv \frac{p_B(x) - \kappa_{BA} p_A(x)}{1 - \kappa_{BA}}$$

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

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

Defined in the context of a specific pair of samples A and B, regardless of whether the observable in question has a rigorous factorization theorem.

Additional jet processing (e.g. grooming) can be folded into definition of A and B.

Extracting topics well is fundamentally easier than tagging well.
A picture of quark and gluon jets

1. Take your favorite jet algorithm

2. Consider two jet samples $A$ and $B$ of QCD jets

3. Choose a jet substructure observable $x$

4. “Assume” that “quark” and “gluon” jets exist

5. “Assume” “quark/gluon” jet mutual irreducibility

The samples $A$ and $B$ are statistical mixtures of quark and gluon:

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

$$p_{\text{sample } B}(x) = f_B^q p_{\text{quark}}(x) + f_B^g p_{\text{gluon}}(x), \quad f_B^g = 1 - f_A^q$$

Firm foundation for data-driven methods.
Exploring substructure feature spaces

Why restrict ourselves to multiplicity? It works, but we can explore this choice. We can also use a trained model (with CWoLa) as an observable in its own right.

**Observables**

- **Multiplicity** $n_{\text{const}}$
  Number of particles in the jet

- **Soft Drop Multiplicity** $n_{\text{SD}}$
  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

- **EFPs**
  Full IRC-safe information, linearly

- **CNN**
  Trained on two-channel jet images

- **DNN**
  Trained on an N-subjettiness basis

See Patrick’s talk!
Exploring substructure feature spaces

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

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

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.

Insensitive to the model details.

[P.T. Komiske, EMM, J. Thaler, Upcoming.]
Extracting quark and gluon distributions

Preliminary graphs showing distributions for jet mass, constituent multiplicity, width, soft drop multiplicity, image activity, and N-subjectness.
(Self-)calibrating quark and gluon classifiers

The extracted quark and gluon fractions can calibrate quark/glauon classifiers and evaluate tagging performance.

Even the classifier that was used to extract the fractions in the first place!

Note: To compare classifiers, one can just use the performance on A vs B directly.
Looking ahead

How does “sample dependence” manifest in this language?

Pairs of samples define quark and gluon. Different pairs of samples yield different flavor definitions.

Comparing definitions from different 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:

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

Beyond quarks and gluons?

- Multi-sample & multi-category generalizations of these ideas exist (though become more complicated).
- These ideas may be useful for other boosted hadronic objects as well.

Max (f^q + f^g) s.t. p_C(x) = f^q p_q(x) + f^g p_g(x) + (1 - f^q - f^g) p_{other}(x)
Summary

• An operational definition of quark and gluon jets defined directly in terms of hadronic cross sections:

\[ p_{\text{quark}}(x) \equiv \frac{p_A(x) - \kappa_{AB} p_B(x)}{1 - \kappa_{AB}} \quad p_{\text{gluon}}(x) \equiv \frac{p_B(x) - \kappa_{BA} p_A(x)}{1 - \kappa_{BA}} \]

• Allows quark and gluon jet distributions to be measured separately without fraction or template inputs:

• Provide a firm foundation for data-driven techniques
  • Template methods, classification without labels, etc.

• Potential to probe questions of sample dependence in data
The End
Thank you!
Extra Slides
Jet topics from QCD: Casimir scaling

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

\[ \Sigma_g(m) = \Sigma_q(m) \frac{C_A}{C_F} \]

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

\[ \kappa_{gq} = \min_m \frac{p_g(m)}{p_q(m)} = \frac{\Sigma_g'(m)}{\Sigma_q'(m)} = \frac{C_A}{C_F} \min_m \frac{\Sigma_q'(m)}{C_F} = 0 \]

\[ \kappa_{qg} = \min_m \frac{p_q(m)}{p_g(m)} = \frac{\Sigma_q'(m)}{\Sigma_g'(m)} = \frac{C_F}{C_A} \min_m \frac{\Sigma_g'(m)}{C_A} = \frac{C_F}{C_A} = \frac{4}{9} \]
Jet topics from QCD: Poisson scaling

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

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

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

\[
\kappa_{gq} = \min_n \frac{p_g(n)}{p_q(n)} = \min_n \left( \frac{C_A \lambda}{C_F \lambda} \right)^n e^{-C_A \lambda} e^{C_F \lambda} = e^{\lambda(C_F-C_A)} \min_n \left( \frac{C_A}{C_F} \right)^n = e^{\lambda(C_F-C_A)}
\]

\[
\kappa_{qg} = \min_n \frac{p_q(n)}{p_g(n)} = \min_n \left( \frac{C_F \lambda}{C_A \lambda} \right)^n e^{-C_F \lambda} e^{-C_A \lambda} = e^{\lambda(C_A-C_F)} \min_n \left( \frac{C_F}{C_A} \right)^n = 0
\]

\[ C_F = \frac{4}{3} \text{ for quarks} \]
\[ C_A = 3 \text{ for gluons} \]
Extracting **quark** and **gluon** fractions

From the reducibility factors, the **quark** and **gluon** fractions of the samples can be obtained.
MC-labeled sample dependence in Pythia