Letter

Data-driven quark- and gluon-jet modification in heavy-ion collisions

Jasmine Brewer,* Jesse Thaler,† and Andrew P. Turner‡

Center for Theoretical Physics, Department of Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

(Received 2 September 2020; accepted 2 February 2021; published 26 February 2021)

Whether quark- and gluon-initiated jets are modified differently by the quark-gluon plasma produced in heavy-ion collisions is a long-standing question that has thus far eluded a definitive experimental answer. A crucial complication for quark-gluon discrimination in both proton-proton and heavy-ion collisions is that all measurements necessarily average over the (unknown) quark-gluon composition of a jet sample. In the heavy-ion context, the simultaneous modification of both the fractions and substructure of quark and gluon jets by the quark-gluon plasma further obscures the interpretation. Here, we demonstrate a fully data-driven method for separating quark and gluon contributions to jet observables using a statistical technique called topic modeling. Assuming that jet distributions are a mixture of underlying "quark-like" and "gluon-like" distributions, we show how to extract quark- and gluon-jet fractions and constituent multiplicity distributions as a function of the jet transverse momentum. This proof-of-concept study is based on proton-proton and heavy-ion collision events from the Monte Carlo event generator JEWEL with statistics accessible in Run 4 of the Large Hadron Collider. These results suggest the potential for an experimental determination of quark- and gluon-jet modifications.

DOI: 10.1103/PhysRevC.103.L021901

High-energy collisions between large nuclei at the Relativistic Heavy Ion Collider (RHIC) and the Large Hadron Collider (LHC) are a critical laboratory for studying the deconfined phase of QCD matter, the quark-gluon plasma, created in these collisions. Collimated sprays of high-momentum hadrons, called jets, are produced copiously in these collisions and provide an important probe of the quark-gluon plasma they pass through.

A long-standing question is how the quark-gluon plasma resolves the color charge of high-energy QCD partons [1–6]. Since jets can originate from either a quark or gluon, and subsequently carry information about their respective total color charge, it is crucial to understand differences in the energy loss and modification of these two categories of jets. Unfortunately, accessing independent information about quark and gluon jets experimentally is very challenging because all jet measurements involve a mixture of contributions from both.

In this paper, we demonstrate a data-driven method to estimate both the quark- and gluon-jet fractions and their separate substructure modification in heavy-ion collisions. Our method is based on a statistical technique called topic modeling, which was pioneered for applications to quark- and gluon-jet separation in proton-proton collisions in Refs. [7,8] and has been applied experimentally in Ref. [9]. We present a proof of concept that an extension of that technique can be used to extract differences in the modification of quark and gluon jets in heavy-ion collisions with the statistics anticipated in Run 4 of the LHC. In order to address the much larger statistical uncertainties present in the heavy-ion context, we develop a method that is substantially more robust to statistical fluctuations than the one used in Refs. [7,8]. This is a critical step toward a model-independent determination of quark- and gluon-jet modification in heavy-ion collisions, which would have dramatic consequences for understanding the microscopic structure of the quark-gluon plasma.

A similar type of analysis was recently performed in Ref. [10], which used a measurement of the jet charge and templates for the jet charge distributions of quark and gluon jets to extract the gluon fraction in proton-proton and heavy-ion collisions. In that study, the same Monte Carlo (MC) distributions were used as templates in both proton-proton and heavy-ion collisions, which makes the implicit assumption that the jet charge distributions of quark and gluon jets are unmodified by the quark-gluon plasma. Here, we present a method that does not require templates and does not assume that substructure observables are unmodified by the plasma, allowing for simultaneous estimates of the modification of quark- and gluon-jet fractions and of their distributions.

Our method is based on a statistical technique called DEMIX [11] that separates a pair of mixed probability distributions into two common underlying base distributions. This method was demonstrated in Refs. [7,8] as a way to obtain excellent proxies for quark and gluon jets in proton-proton

*jtbrewer@mit.edu
†jthaler@mit.edu
‡apturner@mit.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

collisions. Consider two probability distributions $p_1(x)$, $p_2(x)$ for a jet observable x that are a distinct mixture of the same two underlying base probability distributions $b_1(x), b_2(x)$. Namely, we can express the mixture distributions as $p_j(x) = f_j\, b_1(x) + (1 - f_j)\, b_2(x)$, for distinct fractions f_j satisfying $0 \leqslant f_j \leqslant 1$. This expression of the p_j is always ambiguous, however, since there are infinitely many ways to mix the base distributions $b_i(x)$ among themselves to obtain new distributions $\tilde{b}_i(x)$ from which the mixture distributions can be expressed as $p_j(x) = \tilde{f}_j\, \tilde{b}_1(x) + (1 - \tilde{f}_j)\, \tilde{b}_2(x)$ with new fractions $\tilde{f}_j$. The idea behind DEMIX is to resolve this ambiguity by further requiring that the base distributions are *mutually irreducible* [12]. Qualitatively, this means that neither base distribution contains any component of the other. Mathematically, two distributions $b_1(x)$ and $b_2(x)$ are said to be mutually irreducible if

$$\inf_x \frac{b_1(x)}{b_2(x)} = \inf_x \frac{b_2(x)}{b_1(x)} = 0\,. \tag{1}$$

This motivates the definition of the *reducibility factor*

$$\kappa(q, r) = \inf_x \frac{q(x)}{r(x)} \tag{2}$$

for any two distributions $q(x), r(x)$, which allows us to express the condition of mutually irreducibility as $\kappa(b_1, b_2) = \kappa(b_2, b_1) = 0$. For any given $p_1(x), p_2(x)$, there is always a unique pair of mutually irreducible base distributions $t_1(x), t_2(x)$ from which $p_1(x), p_2(x)$ can be built, i.e., $\kappa(t_1, t_2) = \kappa(t_2, t_1) = 0$ and $p_j(x) = f_j\, t_1(x) + (1 - f_j)\, t_2(x)$ for some valid fractions f_j. We refer to these mutually irreducible base distributions $t_1(x), t_2(x)$ as *topics*, for their relation to the broader field of topic modeling established in Ref. [7]. See Refs. [13–15] for other uses of topic modeling techniques in collider physics.

The algorithm to extract the mutually irreducible base distributions (topics) is straightforward. Topics are computed from the mixture distributions via

$$\begin{aligned} t_1(x) &= \frac{p_1(x) - \kappa(p_1, p_2)\, p_2(x)}{1 - \kappa(p_1, p_2)}\,, \\ t_2(x) &= \frac{p_2(x) - \kappa(p_2, p_1)\, p_1(x)}{1 - \kappa(p_2, p_1)}\,. \end{aligned} \tag{3}$$

This justifies the name "reducibility factor" for κ, as we see that $\kappa(p_i, p_j)$ is the maximum fraction of $p_j(x)$ that can be subtracted from $p_i(x)$ such that the resulting function remains positive for every x, and can thus be normalized to yield a proper probability distribution. These κ are directly related to the mixture fractions

$$\kappa(p_1, p_2) = \frac{1 - f_1}{1 - f_2}\,, \quad \kappa(p_2, p_1) = \frac{f_2}{f_1}\,, \tag{4}$$

for which $p_j(x) = f_j\, t_1(x) + (1 - f_j)\, t_2(x)$.

The notion of "quark- and gluon-initiated jets" is not well defined at the hadron level. Even at the level of a MC generator, where parton information from the hard process is available, there is still an ambiguity about how to associate final-state jets with their initiating parton. Therefore, the quark and gluon topics we discuss in this paper do not correspond directly to any parton-level intuition about quark- and gluon-initiated jets [16]. Instead, these topics correspond to the *operational definition* of jet categories introduced in Ref. [8], which defines the quark and gluon categories as the mutually irreducible (i.e., maximally separable) distributions underlying a pair of jet samples. To minimize potential confusions, we will often use the language of quark-like and gluon-like (or "quark" and "gluon" in quotes) to refer to this operational definition.

Since the base distributions extracted from a jet observable x using DEMIX are mutually irreducible, they can only agree with the MC quark- and gluon-jet distributions of x if those are also mutually irreducible. It was argued in Ref. [7] that quark-gluon mutual irreducibility is approximately satisfied for the constituent multiplicity (number of constituents) of groomed and ungroomed jets and n_{SD} [17] in proton-proton collisions, though not for other common jet observables like jet mass. This stems from counting observables having exact quark-gluon mutual irreducibility in the high-energy limit [17]. Reference [8] further showed that constituent multiplicity is a nearly optimal classifier to separate operationally defined quark and gluon jets (see Ref. [18] for further developments). We thus focus on constituent multiplicity for extracting quark and gluon fractions in our case study.

Finally, it is important to note that this approach relies on factorization, namely that hadronic observables can be described by contributions from underlying partonic processes. In proton-proton collisions, this is comparatively well established, but in heavy-ion collisions the modification of jets by the medium could alter this picture. Given three independent jet samples, this assumption can be tested by quantifying the extent to which all three can be described by mixtures of only two common underlying distributions. We discuss the methods to carry out this quantification further in the Supplemental Material [19]. It may prove challenging, however, to find three independent jet samples with sufficient statistics to be distinguished in this way experimentally. In principle, jets produced in dijet events and those produced in association with a high-energy photon or Z boson have different quark-gluon fractions. However, the latter two have quark-gluon fractions that are not distinguishable within expected uncertainties in near-term measurements. While jets produced at different rapidities also have different quark-gluon fractions, in heavy-ion collisions they additionally traverse different amounts of medium and therefore are presumably not described by identical underlying distributions. In this study we therefore restrict our attention to two-category classification of jets from dijet production and those produced in association with a high-energy photon (γ + jet).

For two analytically known mixtures of two base distributions, it is essentially trivial to compute $\kappa(p_1, p_2)$ using Eq. 2 and then to extract the mutually irreducible underlying distributions using Eq. 3. As an example, two Gaussian distributions with different mean and the same standard deviation are mutually irreducible, so any two convex mixtures built from them can be demixed exactly. When dealing with finite-sampled distributions, however, one encounters substantial technical difficulties using Eq. 2 directly. A histogram of samples from a probability distribution $p(x)$ has a

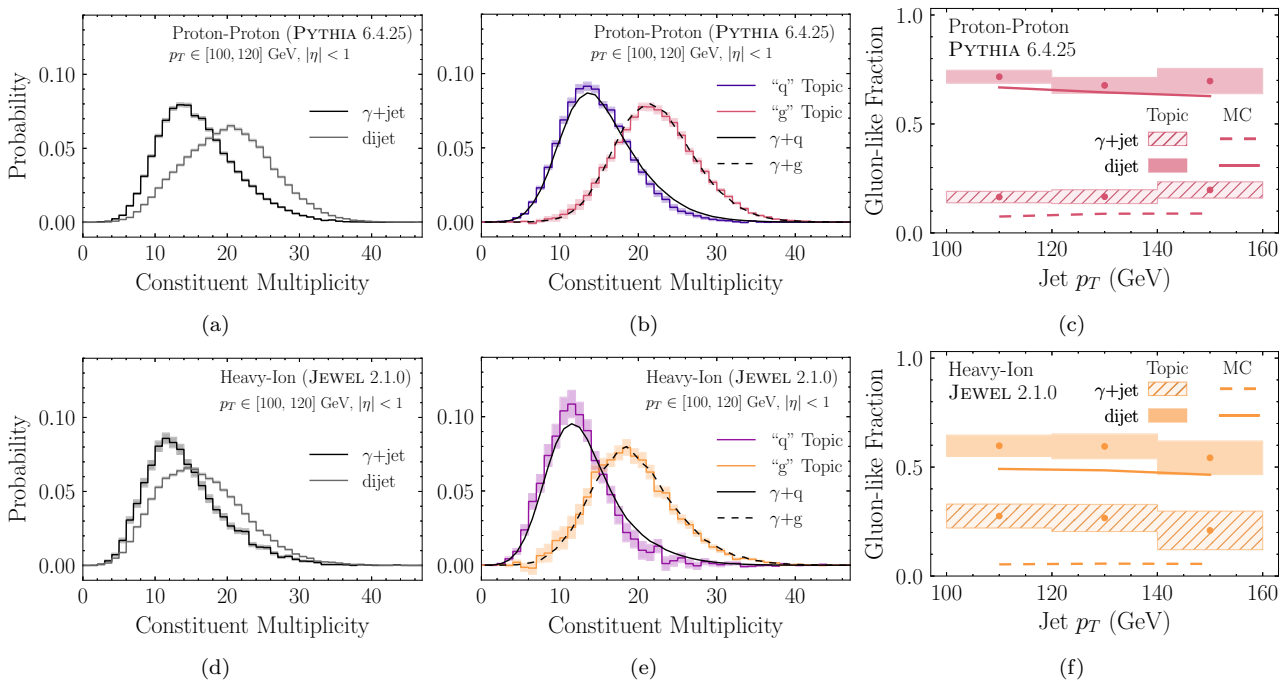

FIG. 1. Extracting quark-like and gluon-like jet topics from (top row) proton-proton collisions and (bottom row) heavy-ion collisions, as generated by PYTHIA and JEWEL, respectively. (a) Distributions of jet constituent multiplicity for the γ + jet and dijet samples in proton-proton collisions. (b) The two underlying topics extracted from these distributions using the DEMIX method (colorful bands), compared to the MC-level definition of quark- and gluon-initiated jets (black). (c) Fractions of the gluon-like topic in the γ + jet and dijet samples. The corresponding results for heavy-ion collisions are shown in (d)–(f). Possible reasons for the higher gluon-like topic fractions compared to the MC label fractions are provided in the text.

finite, discretized range of histogram bins $\{x_\ell\}$ at which $p(x)$ is estimated from the finite-statistics sampled distribution, $\hat{p}(x_\ell)$. We need a method of defining the reducibility factors $\hat{\kappa}(\hat{p}_1, \hat{p}_2)$ for a pair of sampled histograms. Naively, the infimum of Eq. 2 becomes a minimum of the ratio of the histograms over $\{x_\ell\}$; simply taking the minimum, however, is very sensitive to statistical fluctuations. A more robust approach, introduced in Ref. [8], is to define $\hat{\kappa}$ to be the ratio of histograms in the bin for which the ratio plus its uncertainty is minimized. This method turns out to be insufficient to deal with the much more limited statistics we aim to utilize in this work, particularly because $\hat{\kappa}$ is typically extracted at the low-statistics end points of the distributions. For this reason, we must improve upon the method of Refs. [7,8] to make it significantly more robust to statistical fluctuations.

The novel method we present in this work uses fitting to leverage information about the interior of the distribution, where the statistics are better, to put additional constraints on the tails. We note that the dijet and γ + jet histograms shown in Figs. 1(a) and 1(d) are exceptionally well described by a simultaneous fit to two distinct sums of a pair of skew-normal distributions $\mathrm{SN}(x; \mu, \sigma, s)$. That is, they are well described by the form

$$f_N(x; \alpha_i, \theta) = \sum_{k=1}^{N} \alpha_{i,k}\, \mathrm{SN}(x; \mu_k, \sigma_k, s_k) \tag{5}$$

with $N = 2$. Here $\theta = (\mu_1, \sigma_1, s_1, \ldots, \mu_N, \sigma_N, s_N)$, and $\alpha_i = (\alpha_{i,1} \ldots, \alpha_{i,N})$ contains $N - 1$ independent fractions, with the Nth fraction constrained by $\sum_{k=1}^{N} \alpha_{i,k} = 1$, for jet samples $i = 1, 2$ [dijet and γ + jet in Figs. 1(a) and 1(d)]. For further generality of the functional form, we consider $N = 4$ and simultaneously fit the two input distributions to $f_4(x; \alpha_1, \theta)$ and $f_4(x; \alpha_2, \theta)$, respectively, with 18 fit parameters α_1, α_2, and θ. To estimate the uncertainty on such fits, we use the Markov chain Monte Carlo (MCMC) ensemble sampler emcee [20] to do posterior estimation using the likelihood function [21]

$$\ln \frac{C}{p} = \sum_{i,\ell} n_i \left[f(x_{i,\ell}; \alpha_i, \theta) - y_{i,\ell} + y_{i,\ell} \ln \frac{y_{i,\ell}}{f(x_{i,\ell}; \alpha_i, \theta)} \right]. \tag{6}$$

Here, ℓ indexes the histogram bins of jet sample i, with the ℓth bin having constituent multiplicity $x_{i,\ell}$ and probability density $y_{i,\ell}$, and the ith sample having total count n_i. This form assumes that the number of counts in each histogram bin, $n_{i,\ell} = n_i y_{i,\ell}$, is independently Poisson distributed around the value $f(x_{i,\ell}; \alpha_i, \theta)$ and estimates distributions of the parameters α_i, θ for which the observed data are most likely. Following Refs. [21,22], the likelihood function p in Eq. 6 is rescaled by a fit-independent constant C that cancels a $\ln(n_{i,\ell}!)$ that arises when taking the logarithm of the Poisson probability distribution. We take a uniform prior on the parameters θ and α_i in the range $\mu_k \in [0, 50]$, $\sigma_k \in [1, 15]$, $s_k \in [-20, 20]$, and $\alpha_{i,k} \in [0, 1]$, and we start the MCMC walkers in a Gaussian ball of standard deviation 10% around the least-squares fit parameters. We use the distribution of

fits to obtain distributions of $\hat{\kappa}(\hat{p}_i, \hat{p}_j)$ via Eq. 2. To combat finite statistics effects, we compute the infimum in Eq. 2 as a minimum of the MCMC walkers over a reduced range. We consider only the range for which at least one input histogram is nonzero. For each reducibility factor, we further identify whether the minimum will occur on the left or right side of the range, and truncate the opposite tail at the outermost bin that has at least ten data points for each input histogram. The distribution of $\hat{\kappa}(\hat{p}_i, \hat{p}_j)$ is used to compute a distribution of fractions, and its mean and standard deviation are used as the value and uncertainty of $\hat{\kappa}(\hat{p}_i, \hat{p}_j)$ used to extract the topics.

The samples for our proof-of-concept study come from the heavy-ion MC event generator JEWEL 2.1.0 [23,24], based on vacuum jet production in PYTHIA 6.4.25 [25]. We consider two mixed distributions coming from photon-jet ($\gamma +$ jet) production and dijet production. For each process, we generate proton-proton and 0% to 10% centrality heavy-ion events at 5.02 TeV and reconstruct anti-k_t jets using FASTJET 3.3.0 [26,27] with radius parameter $R = 0.4$ within the pseudorapidity range $|\eta| < 1$. We include initial-state radiation, but do not include medium recoil effects. (There is no underlying event model in JEWEL, but we verified that similar results can be obtained after aggressively grooming jets using the soft drop algorithm [28] with $z_{\text{cut}} = 0.5$ and $\beta = 1.5$ as in Ref. [29].) For $\gamma +$ jet events we consider the recoiling jet with the highest transverse momentum (p_T), and for dijet events we consider the two highest-p_T jets. In the case of heavy-ion collisions, we downsample our JEWEL events to mimic the statistics that will be available with the anticipated luminosity $\int \mathcal{L}\, dt = 13\,\text{nb}^{-1}$ after Run 4 [30] (see the Supplemental Material [19] for details). The equivalent statistics of our dijet sample are already less than those achievable in Run 4, but this substantially reduces the statistics of our $\gamma +$ jet sample. We emphasize that we are only using JEWEL for demonstration purposes, and this data-driven technique can be applied directly to experimental collider measurements for a range of jet observables beyond just multiplicity.

Starting with proton-proton collisions in the top row of Fig. 1, we show the distributions of jet constituent multiplicity for $\gamma +$ jet and dijet samples [Figs. 1(a)] and the "quark-like" and "gluon-like" topics extracted from these distributions via the data-driven method described above [Fig. 1(b)]. The corresponding heavy-ion results are shown in Figs. 1(d) and 1(e), keeping in mind that the proton-proton and heavy-ion analyses are completely independent. The extracted topics are in good agreement with the distributions of constituent multiplicity for quark- and gluon-initiated jets as defined at the MC level.

Furthermore, we can use Eq. 4 to extract the topic fractions, i.e., the proportions of the topics in the original input distributions. Figures 1(c) and 1(f) show the extracted fraction of the gluon-like topic in the $\gamma +$ jet and dijet samples as a function of jet p_T. The gluon topic fractions are marginally higher than the MC-level fraction of gluon-initiated jets in proton-proton collisions, and more dramatically higher in heavy-ion collisions. In interpreting these results, however, one has to be mindful of the inherent ambiguity in using MC-level information to label jets, which we explore in the Supplemental Material [19]. In addition, limited statistics drive the extraction of $\hat{\kappa}$ into the interior of the distribution where the true minimum is not yet achieved. In the Supplemental Material [19], we repeat the heavy-ion analysis using a $\gamma +$ jet sample with a factor of about 2.8 higher luminosity. Though the results are consistent within (large) uncertainties, the method with limited statistics will tend to overestimate the gluon-like fraction.

Even accounting for these issues, though, we find a persistently larger gluon-like fraction compared to the MC labeling, at least in the context of JEWEL. One possible explanation for this effect is that a "quark-initiated" jet may become more gluon-like through gluon radiation, an effect which may be enhanced by medium-induced gluon radiation in heavy-ion collisions. For methods like this one, as well as for the method in Ref. [10], this would result in a larger fraction of jets being classified as gluon jets. It is also possible that constituent multiplicity, though apparently nearly mutually irreducible in proton-proton collisions, may be less mutually irreducible in the presence of medium effects, so alternative observables (perhaps from machine learning [8]) might be required. Understanding these issues will be important for interpreting eventual LHC Run 4 data, but the operational definition used to define the quark-like and gluon-like topics is independent of its interpretation.

As a final proof of concept, in Fig. 2 we show the modification of the jet constituent multiplicity distributions for the quark-like [Fig. 2(a)] and gluon-like [Fig. 2(b)] jet topics as a function of p_T. To our knowledge, this represents the first fully data-driven method to separate the modification of a jet observable for "quark" and "gluon" jets. Though we show here the modification of the constituent multiplicity distribution for clarity, we emphasize that once the topic fractions have been extracted, they can be used to extract separate quark and gluon distributions for any jet observable. Since both jet observable distributions and the quark and gluon fractions may change between proton-proton and heavy-ion collisions, it is substantially simpler to interpret the separate modification of quark and gluon topics compared to, e.g., the modification of the dijet distribution. Though not shown here, the method of Ref. [31] could be used to match the proton-proton and heavy-ion jet p_T quantiles and further clarify the interpretation.

In summary, we have illustrated a data-driven method to extract quark-like and gluon-like topic fractions and distributions in proton-proton and heavy-ion collisions. Our method improves significantly over previous approaches in its robustness to statistical fluctuations, making it viable for heavy-ion collision data and for lower-statistics or more differential applications in the proton-proton context. Using JEWEL samples of comparable statistics to those anticipated in Run 4 of the LHC, we have shown that these topics have a similar qualitative interpretation to the (physically ambiguous) definition of quark and gluon jets at parton level available from MC generators. We have further shown, as an example, the modification of the constituent multiplicity in heavy-ion collisions separately for quark- and gluon-jet topics. This study offers an exciting proof-of-concept demonstration of the power of topic modeling to interpret future heavy-ion collision data, though

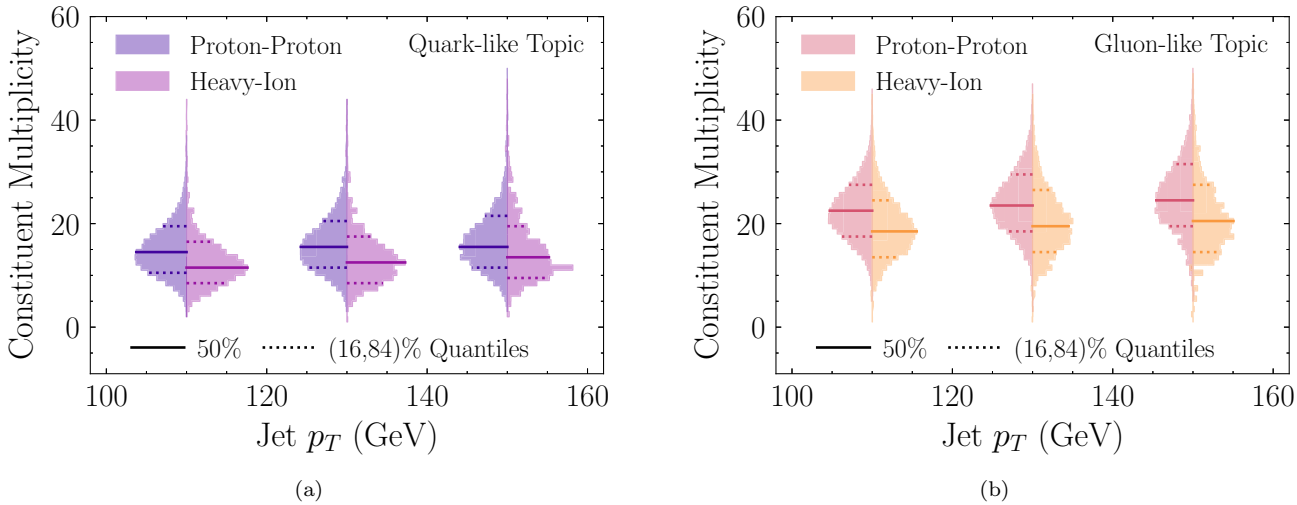

FIG. 2. Constituent multiplicity distributions for (a) the quark-like topic and (b) the gluon-like topic as a function of jet p_T. Each violin plot has results for both (left side) proton-proton and (right side) heavy-ion collisions, and the change between the two sides illustrates the modification of the constituent multiplicity distribution for the corresponding topic. Horizontal lines indicate the median (solid) and 16% and 84% quantiles (dashed) of the multiplicity distributions.

more quantitative studies will be necessary to understand the feasibility of this analysis and the best way to incorporate systematic uncertainties. We leave a detailed study of the impact of underlying event and background subtraction, medium response, and experimental inefficiencies as important future work. This method is well defined for any jet observable, and the aforementioned effects may render it important to study the performance of this analysis with additional observables beyond the constituent multiplicity.

Code to run this analysis on input histograms is publicly available [33].

The authors acknowledge valuable discussions with Liliana Apolinário, Raghav Kunnawalkam Elayavalli, Dhanush Anil Hangal, Patrick Komiske, Yen-Jie Lee, Eric Metodiev, Guilherme Milhano, and Krishna Rajagopal, in addition to technical assistance and feedback from Jacob Bandes-Storch, William Lewis, Lina Necib, and Anastasia Patterson. The authors were supported by the U.S. Department of Energy (DOE) Office of Nuclear Physics under Contract No. DE-SC0019128 and the U.S. Department of Energy (DOE) Office of High Energy Physics under Contract No. DE-SC0012567, and A.P.T. was additionally supported by funding from the Tushar Shah and Sara Zion Fellowship.

[1] S. Caron-Huot, $O(g)$ plasma effects in jet quenching, Phys. Rev. D **79**, 065039 (2009).
[2] M. Spousta and B. Cole, Interpreting single jet measurements in Pb + Pb collisions at the LHC, Eur. Phys. J. C **76**, 50 (2016).
[3] Y.-T. Chien and R. Kunnawalkam Elayavalli, Probing heavy ion collisions using quark and gluon jet substructure, arXiv:1803.03589.
[4] Y. Mehtar-Tani and S. Schlichting, Universal quark to gluon ratio in medium-induced parton cascade, J. High Energy Phys. 09 (2018) 144.
[5] J.-W. Qiu, F. Ringer, N. Sato, and P. Zurita, Factorization of Jet Cross Sections in Heavy-Ion Collisions, Phys. Rev. Lett. **122**, 252301 (2019).
[6] L. Apolinário, J. Barata, and G. Milhano, On the breaking of Casimir scaling in jet quenching, Eur. Phys. J. C **80**, 586 (2020).
[7] E. M. Metodiev and J. Thaler, Jet Topics: Disentangling Quarks and Gluons at Colliders, Phys. Rev. Lett. **120**, 241602 (2018).
[8] P. T. Komiske, E. M. Metodiev, and J. Thaler, An operational definition of quark and gluon jets, J. High Energy Phys. 11 (2018) 059.
[9] G. Aad *et al.* (ATLAS Collaboration), Properties of jet fragmentation using charged particles measured with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV, Phys. Rev. D **100**, 052011 (2019).
[10] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of quark- and gluon-like jet fractions using jet charge in PbPb and pp collisions at 5.02 TeV, J. High Energy Phys. 07 (2020) 115.
[11] J. Katz-Samuels, G. Blanchard, and C. Scott, Decontamination of mutual contamination models, J. Mach. Learn. Res. **20**, 1 (2019).
[12] G. Blanchard, M. Flaska, G. Handy, S. Pozzi, and C. Scott, Classification with asymmetric label noise: Consistency and maximal denoising, Electron. J. Stat. **10**, 2780 (2016).
[13] B. M. Dillon, D. A. Faroughy, and J. F. Kamenik, Uncovering latent jet substructure, Phys. Rev. D **100**, 056002 (2019).
[14] E. Alvarez, F. Lamagna, and M. Szewc, Topic model for four-top at the LHC, J. High Energy Phys. 01 (2020) 049.
[15] B. M. Dillon, D. A. Faroughy, J. F. Kamenik, and M. Szewc, Learning the latent structure of collider events, J. High Energ. Phys. 10 (2020) 206.
[16] P. Gras, S. Höche, D. Kar, A. Larkoski, L. Lönnblad, S. Plätzer, A. Siódmok, P. Skands, G. Soyez, and J. Thaler, Systematics of quark/gluon tagging, J. High Energy Phys. 07 (2017) 091.

[17] C. Frye, A. J. Larkoski, J. Thaler, and K. Zhou, Casimir meets Poisson: Improved quark/gluon discrimination with counting observables, J. High Energy Phys. 09 (2017) 083.
[18] A. J. Larkoski and E. M. Metodiev, A theory of quark vs. gluon discrimination, J. High Energy Phys. 10 (2019) 014.
[19] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevC.103.L021901 for further information about testing for additional distributions, limited statistics effects, and MC label ambiguities, which includes Ref. [32].
[20] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, emcee: The MCMC hammer, Publ. Astron. Soc. Pac. **125**, 306 (2013).
[21] S. Baker and R. D. Cousins, Clarification of the use of chi square and likelihood functions in fits to histograms, Nucl. Instrum. Methods **221**, 437 (1984).
[22] M. Tanabashi *et al.* (Particle Data Group), Review of Particle Physics, Phys. Rev. D **98**, 030001 (2018).
[23] K. C. Zapp, JEWEL 2.0.0: directions for use, Eur. Phys. J. C **74**, 2762 (2014).
[24] R. Kunnawalkam Elayavalli and K. C. Zapp, Simulating V+jet processes in heavy ion collisions with JEWEL, Eur. Phys. J. C **76**, 695 (2016).
[25] T. Sjostrand, S. Mrenna, and P. Z. Skands, PYTHIA 6.4 physics and manual, J. High Energy Phys. 05 (2006) 026.
[26] M. Cacciari, G. P. Salam, and G. Soyez, The anti-k_t jet clustering algorithm, J. High Energy Phys. 04 (2008) 063.
[27] M. Cacciari, G. P. Salam, and G. Soyez, FastJet user manual, Eur. Phys. J. **C72**, 1896 (2012).
[28] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, Soft drop, J. High Energy Phys. 05 (2014) 146.
[29] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of the groomed jet mass in PbPb and pp collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV, J. High Energy Phys. 10 (2018) 161.
[30] Z. Citron *et al.*, Report from Working Group 5: Future physics opportunities for high-density QCD at the LHC with heavy-ion and proton beams, in *Report on the Physics at the HL-LHC, and Perspectives for the HE-LHC*, Vol. 7, edited by A. Dainese, M. Mangano, A. B. Meyer, A. Nisati, G. Salam, and M. A. Vesterinen (CERN, Geneva, 2019), pp. 1159–1410.
[31] J. Brewer, J. G. Milhano, and J. Thaler, Sorting Out Quenched Jets, Phys. Rev. Lett. **122**, 222301 (2019).
[32] G. Roland, K. Safarik, and P. Steinberg, Heavy-ion collisions at the LHC, Prog. Part. Nucl. Phys. **77**, 70 (2014).
[33] https://github.com/jasminebrewer/jet-topics-from-MCMC.