IL NUOVO CIMENTO **46 C** (2023) 56
DOI 10.1393/ncc/i2023-23056-3

COMMUNICATIONS: SIF Congress 2022

Using Graph Neural Networks to improve jet flavour-tagging and its modeling for the ATLAS experiment at the LHC

M. TANASINI on behalf of the ATLAS COLLABORATION

Dipartimento di Fisica, Universitá di Genova - Genova, Italy

received 30 January 2023

Summary. — Graph Neural Networks (GNNs) are machine-learning algorithms particularly suitable for modeling data with complex topological correlations. In this communication, their application to jet flavour-tagging for the ATLAS experiment at the Large Hadron Collider is presented. A new GNN algorithm to identify jets containing heavy-flavour hadrons by representing them as graphs of tracks and silicon hits is illustrated. The performance of the modern flavour-tagging algorithms poses challenges when applied on simulated events containing multiple jets, as they reduce the statistical precision of the simulated samples. To overcome this, a GNN-based technique that increases the statistical power of the samples by weighting events based on their likelihood of containing flavour-tagged jets is described.

1. – Introduction

The studies presented in this communication revolve around the application of Graph Neural Networks (GNNs) to jet flavour-tagging to increase the physics reach of the ATLAS experiment [1] at the LHC [2]. GN1, the ATLAS new GNN-based jet flavour-tagging algorithm, introduced in [3], is described. Then, an innovative technique using GNNs to parametrise jet flavour-tagging efficiencies, described in details in [4], is presented. It allows to increase the statistical power of simulated samples, which are used to model the backgrounds for many analyses conducted in ATLAS.

2. – Jet flavour-tagging

Flavour-tagging is a method used in particle physics to identify the flavour of hadrons within a jet. Jets containing b-hadrons are labeled as b-jets, jets containing c-hadrons but no b-hadrons are labeled as c-jets, jets containing τ leptons but no b or c-hadrons are labeled as τ-jets, and the remaining jets are labeled as light-flavour jets.

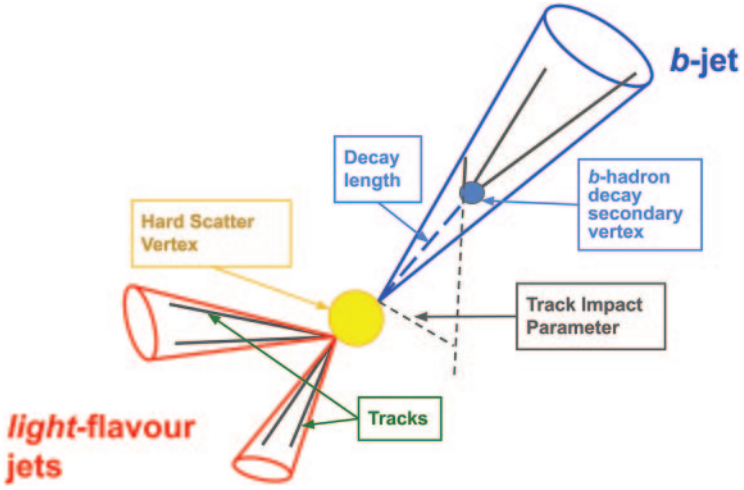


Fig. 1. – Visualisation of the features that distinguish b-jets (in blue) from light-flavour jets (in red). The displaced decays of b-hadrons lead to the presence of secondary vertices in the b-jets. Tracks associated to the charged particles produced in the decay can have large impact parameters.

The long lifetime of b-hadrons, which is on the order of 1.5 ps, allows them to fly significant distances before decaying. Hence, b-jets can be identified through the observation of their tracks' large impact parameters ([1]) and the displaced secondary vertices resulting from their decays (fig. 1). In addition, b-hadrons have relatively high masses and decay multiplicities.

The flavour-tagging algorithms used in ATLAS are trained on simulated data. They aim to strike a balance between two objectives: a high tagging efficiency, meaning a high proportion of correctly tagged jets, and a low mistagging rate, meaning a low proportion of jets that are incorrectly tagged. The performance of the algorithms is contingent on the chosen Working Point (WP), which is the desired efficiency, and the specific physics scenario being analyzed.

Flavour-tagging is an essential tool in many areas of particle physics, such as the study of the properties of the Higgs boson, the search for new physics beyond the Standard Model, and the study of the properties of the quarks and gluons themselves [5-7].

3. – Graph Neural Networks

Graph Neural Networks (GNNs) are a class of neural networks that are designed to operate on graph-structured data. A graph is a collection of nodes (or vertices) and edges that connect them. Graphs can be used to represent a wide variety of data types, such as social networks and molecular structures. In particle physics, their applications are being extensively explored, addressing jet flavour-tagging, tracking, particle reconstruction and event classification among many others [8].

([1]) The impact parameter of a track is defined as the minimum distance between the track and the primary interaction point.

Jet Input	**Description**
p_{T}	Jet transverse momentum
η	Signed jet pseudorapidity
Track Input	**Description**
q/p	Track charge divided by momentum (measure of curvature)
$\mathrm{d}\eta$	Pseudorapidity of the track, relative to the jet η
$\mathrm{d}\phi$	Azimuthal angle of the track, relative to the jet ϕ
d_0	Closest distance from the track to the PV in the longitudinal plane
$z_0 \sin\theta$	Closest distance from the track to the PV in the transverse plane
$\sigma(q/p)$	Uncertainty on q/p
$\sigma(\theta)$	Uncertainty on track polar angle θ
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
$s(d_0)$	Lifetime signed transverse IP significance
$s(z_0)$	Lifetime signed longitudinal IP significance
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nIBLHits	Number of IBL hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits
nPixHoles	Number of pixel holes
nSCTHoles	Number of SCT holes
leptonID	Indicates if track was used in the reconstruction of an electron or muon (only for GN1 Lep)

Fig. 2. – Input features to the GN1 model [3]. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. Shared hits are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [9], while split hits are hits used on multiple tracks which have been identified as merged (^{3}). A hole is a missing hit, where one is expected, on a layer between two other hits on a track. The track leptonID is an additional input to the GN1 Lep model.

GNNs are typically composed of multiple layers of neural networks, each of which operates on the nodes of the graph. In each layer, the network performs a computation on the node and its neighboring nodes, using the edges of the graph to define the connections between the nodes. The computation typically involves a combination of linear and non-linear operations, such as matrix multiplications and activation functions.

One of the key features of GNNs is that they can learn to propagate information through the graph, allowing them to make predictions or perform other tasks on the entire graph, rather than just on individual nodes. This makes them well suited for tasks such as node classification, link prediction, and graph classification.

4. – GN1

The GN1 [3] algorithm is the state-of-the-art GNN-based flavour-tagging algorithm used by the ATLAS experiment. It utilizes fully connected graph representations of jets, with the jet's associated tracks serving as the nodes. It is trained to operate on vector representation of tracks to perform jet flavour-tagging. At the same time, it performs track classification and two-track vertexing, which are defined as auxiliary tasks that guide the GNN toward an understanding of the underlying physics described in sect. **2**. These three tasks exemplify the capability of GNNs of performing graph classification, link prediction and node classification, respectively.

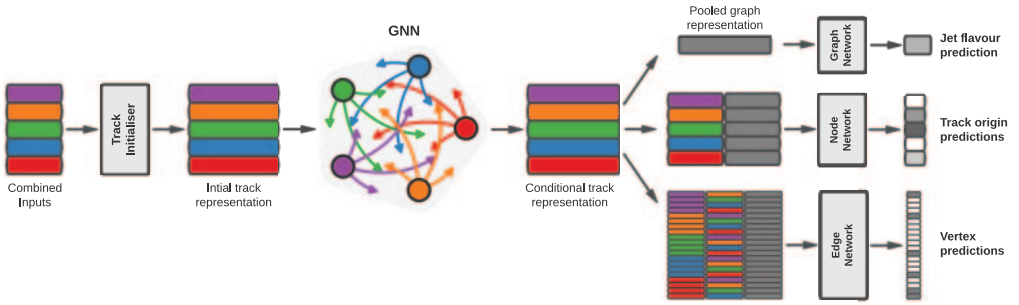


Fig. 3. – The network architecture of GN1 [3]. Inputs are fed into a per-track initialisation network, which outputs an initial latent representation of each track. These representations are then used to populate the node features of a fully connected graph network. After the graph network, the resulting node representations are used to predict the jet flavour, the track origins, and the probability of each track pair to belong to a common vertex.

4˙1. *Inputs and architecture.* – Figure 2 shows the variables that are used for representing the tracks as input for the GN1 and GN1 Lep ([2]) models.

The architecture of the GN1 algorithm is schematized in fig. 3. It is composed of a Feed Forward Neural Network (FFNN) that transforms the input vector representation of each track into an higher-dimensional one; a GNN that processes the transformed representations to create new ones which embed information about the overall structure of the graph; and three FFNNs that use different combinations of the resulting track representations to predict the flavour of the jet, the origin of the tracks, and the presence of two-track secondary vertices.

4˙2. *Training and performance.* – GN1 is trained to minimize a total loss function L_{total} through a supervised gradient-descent procedure:

$$L_{\text{total}} = L_{\text{jet}} + \alpha L_{\text{vertex}} + \beta L_{\text{track}}. \tag{1}$$

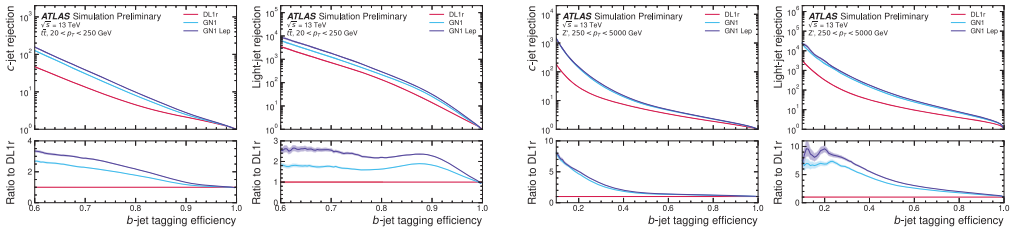
The terms on the right side of eq. (1) are the losses assigned to the FFNNs responsible for the graph and node classification, and the link prediction tasks. These losses are designed to target truth-labels derived from the simulation. L_{jet} targets the jet-flavour labels (b, c, or light). L_{track} targets the track-labels as defined in fig. 4 after analysing their formation process. L_{vertex} targets track-pair labels, where a value of 1 is assigned if neither track is labelled as Fake or Pileup. The link prediction network assigns binary labels for each track-pair, with a value of 1 if the two tracks are predicted to originate from the same point in space. Therefore, the network is trained to predict links only for tracks which are not fake or from Pileup and that come from the same point. The values of $\alpha = 1.5$ and $\beta = 0.5$ are chosen to ensure that L_{jet} converges to a larger value than L_{vertex} and L_{track}, to reflect the priority of the jet-classification task as the final goal of GN1.

([2]) GN1 Lep includes an additional input, leptonID, which indicates whether a track in the jet was used in the reconstruction of an electron, a muon, or neither. This input is useful as heavy-flavour hadrons can decay semi-leptonically. The presence of a reconstructed lepton in a jet can provide valuable information about the jet's flavour.

([3]) In high-p_{T} dense jet cores, tracks are tightly packed and the distance between them is similar

Truth Origin	Description
Pileup	From a pp collision other than the primary interaction
Fake	Created from the hits of multiple particles
Primary	Does not originate from any secondary decay
fromB	From the decay of a b-hadron
fromBC	From a c-hadron decay, which itself is from the decay of a b-hadron
fromC	From the decay of a c-hadron
OtherSecondary	From other secondary interactions and decays

Fig. 4. – Truth origins which are used to categorise the physics process that led to the production of a track [3]. Tracks are matched to charged particles using the truth-matching probability [9]. A truth-matching probability of less than 0.5 indicates that reconstructed track parameters are likely to be mismeasured and may not correspond to the trajectory of a single charged particle. The "OtherSecondary" origin includes tracks from photon conversions, K^0_S and Λ^0 decays, and hadronic interactions.



(a) $t\bar{t}$ (b) Z'

Fig. 5. – The c-jet and light-jet rejections as a function of the b-jet tagging efficiency for jets in a simulated $t\bar{t}$ sample, with transverse momentum $20 < p_{\mathrm{T}} < 250\,\mathrm{GeV}$ (a); and for jets in a simulated Z' sample, with transverse momentum $250 < p_{\mathrm{T}} < 5000\,\mathrm{GeV}$ (b) [3].

The performance of a jet flavour-tagging algorithm can be measured by its ability to accurately reject jets with different flavours at a specified jet flavour-tagging WP. This can be visualized using Receiver Operating Characteristics (ROC) curves, which plot the proportion of correctly tagged jets against the rate of mistagged jets for any WP. The larger the area under the ROC curve, the better the performance. Figure 5 illustrates the b-tagging ROC curves of DL1r ([4]) [10], GN1, and GN1Lep. At a b-jet tagging efficiency of 70%, GN1 outperforms DL1r light-flavour (c)-jet rejection by a factor of $\sim 1.8 (\sim 2.1)$ for simulated jets coming from $t\bar{t}$ decays with transverse momentum p_{T} between 20 and 250 GeV. For simulated jets coming from Z' decays with transverse momentum p_{T} between 250 and 5000 GeV the light-flavour (c)-jet rejection improves by a factor $\sim 6 (\sim 2.8)$ for a comparative 30% b-jet efficiency.

Figure 6 exemplifies the pivotal role that the node classification and link prediction tasks play in boosting the performance of the GN1 algoritm.

to the size of the sensor, which leads to merged clusters and tracks that share hits.
([4]) DL1r is the ATLAS FFNN-based jet flavour-tagging algorithm that preceded GN1.

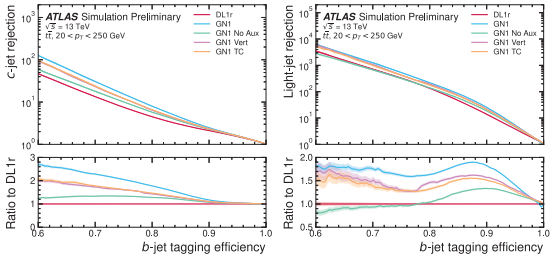


Fig. 6. – The c-jet (left) and light-jet (right) rejections as a function of the b-jet tagging efficiency for $t\bar{t}$ jets with transverse momentum $20 < p_{\mathrm{T}} < 250\,\mathrm{GeV}$, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed [3]. Binomial error bands are denoted by the shaded regions.

5. – Truth flavour-tagging

The simplest way of using jet flavour-tagging algorithms to identify jets both in simulated and real data samples can be defined as "pass-or-fail", meaning that only the jets selected with a given WP are considered as flavour-tagged. Unfortunately, this approach can largely reduce the statistical power of the simulated samples, which are used for modeling the background processes in many physics analyses. The problem is exacerbated when analysing collision events with multiple flavour-tagged jets, such as in the measurement of the branching ratio of the Higgs boson decaying to pairs of $b\bar{b}$ and $c\bar{c}$ quarks [5,11]. For instance, for a 70% b-tagging WP, only $70\% \times 70\% = 49\%$ of the simulated events containing two "true" b-jets ([5]) pass the pass-or-fail flavour-tagging selection. Therefore, the statistical power of the sample for these events is halved. Figure 5(a) shows that the associated light-flavour jet rejection rate of the DL1r algorithm is of $\sim 10^3$, when applied on $t\bar{t}$ jets with transverse momentum p_{T} between 20 and 250 GeV. Thus, only $10^{-3}\% \times 10^{-3}\% = 10^{-6}\%$ of the simulated $t\bar{t}$ events with two light-flavour jets can be used to model the scenario in which both are mistagged as b-jets.

Truth flavour-tagging is an alternative approach to pass-or-fail flavour-tagging. It consists in weighting the events with their likelihood of having the required number of flavour-tagged jets. Weighting a set of N events with their probabilities of passing a selection, rather than applying the selection itself, reduces the relative statistical uncertainty on the estimated number of selected events, if, as it typically happens, the variance of the weights $\sqrt{(\sum_{i\leq N} w_i^2)}$ is small compared to the statistical uncertainty $\sqrt{N_p}$ on the number N_p of events selected with the pass-or-fail technique [12]:

$$\frac{\sqrt{N_p}}{N_p} > \frac{\sqrt{N_p}}{N} > \frac{\sqrt{(\sum_{i\leq N} w_i^2)}}{N} \text{ if } \sqrt{N_p} > \sqrt{\left(\sum_{i\leq N} w_i^2\right)}. \tag{2}$$

([5]) In this context, "true" b-jets are jets that are labelled as such at the simulation level, as determined by the presence of simulated b-hadrons associated with them.

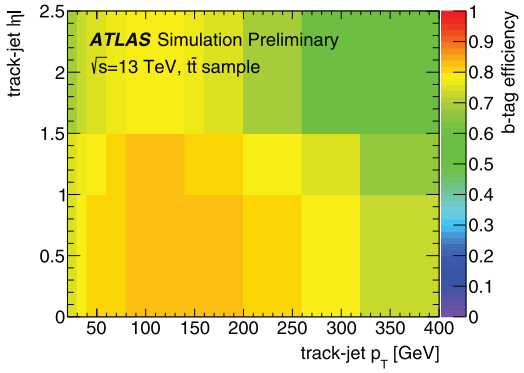


Fig. 7. – Efficiency map for b-jets derived from simulated $t\bar{t}$ events using the DL1r flavour-tagging algorithm at the 77% WP [4].

Therefore, if the flavour-tagging efficiency function

$$\epsilon_f = \frac{\text{number of jets of flavour } f \text{ selected with a given WP}}{\text{number of jets of flavour } f} \tag{3}$$

is well known, it is possible to apply the truth flavour-tagging technique to obtain a sample distributed with the same overall shapes and normalisations that would be obtained with pass-or-fail flavour-tagging, but with a reduced relative statistical uncertainty.

6. – Flavour-tagging efficiency parametrization

The function ϵ_f is not known *a priori*. It depends on a set of parameters θ whose elements are only partially known. Some refer to the jet kinematics and composition, some to the relations between jets, and some to the general features of the collisions in which the jets were produced. Figure 7 shows a 2D map displaying the dependence of ϵ_b on the transverse momentum and pseudorapidity of track-jets ([6]) in $t\bar{t}$ events for the 77% WP of the DL1r algorithm.

Truth flavour-tagging using such 2-dimensional $\epsilon_f(p_{\mathrm{T}}, \eta)$ maps was successfully implemented in [5, 11]. Nevertheless, despite their simplicity in production ([7]), maps have inherent limitations that can impede their accuracy in parametrizing ϵ_f. First of all, scaling maps to higher dimensionalities is extremely challenging, if not impossible. As the number n of dimensions increases, the statistics needed to populate bins in the map scales exponentially with n. Thus, all dependences on further parameters have to be neglected. Secondly, maps have information about features of the individual jets in the events, but not about the relations between them, or about features of the events themselves. For instance, light-flavour jets close to b-jets are more likely to be b-tagged, since tracks produced by charged particles from a b-hadron decay can be wrongly assigned to the close-by light-flavour jet [11]; and the number of tracks associated to the

([6]) Track-jets are jets clustered using the variable-radius jet algorithm [13,14] with $\rho = 30\,\mathrm{GeV}$, $R_{\mathrm{min}} = 0.02$ and $R_{\mathrm{max}} = 0.4$ on a subset of the reconstructed tracks.
([7]) Even though the optimization of the maps' binning can require a painful trial-and-error procedure.

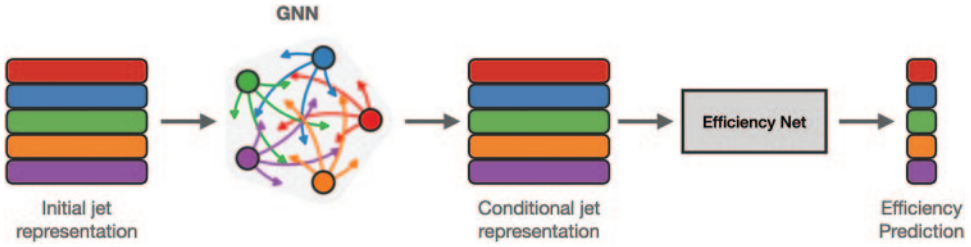


Fig. 8. – The network architecture used for truth flavour-tagging [4]. The initial jet representation is created with the jet and the event features listed in fig. 9. This representation is then updated by a fully connected graph network resulting in a jet representation which is also conditioned by its neighbours. A second fully connected network is then applied on each of these updated jet representations independently to obtain the estimates of the flavour-tagging efficiencies.

jets depends on the amount of pile-up collisions that occur in the event, influencing the flavour-tagging efficiency.

GNNs operating on representations of the events as fully connected graphs of jets are particularly suitable to address the high-dimensional and exquisitely relational nature of the problem at hand.

7. – Flavour-tagging efficiency parametrization with GNNs

Figure 8 shows the scheme of a GNN-based algortihm developed to parametrize the efficiency of flavour-tagging jets in proton-proton collision events. Each event is represented as a fully connected graph in which the jets are the nodes. Initially, jets are represented as vectors that concatenate the track-jet and event variables listed in fig. 9. As opposed to GN1, also edges have an initial representation, which is the angular distance $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ between the jets they connect. A GNN is then used to provide new vector representations of the jets embedding also information about the relations between the jets and the overall structure of the graph. The resulting jet representations are then used by a FFNN that provides a five dimensional vector for each jet, which represents its probability of being b-tagged with each of the five WPs of DL1r which are maintained by the flavour-tagging group of the ATLAS Collaboration ([8]) [4].

The training is done by targeting the DL1r bin in which the jet falls with a multi-class cross entropy loss function. The training converges to an estimate of the true b-tagging efficiency in each of the 5 WPs by virtue of the density ratio estimation method introduced in [15].

7˙1. *Performance.* – For use in physics analyses, the distributions obtained through truth flavour-tagging must be consistent with those obtained through pass-or-fail flavour-tagging within statistical uncertainty. Figure 10 shows the distribution of the large-radius

([8]) The flavour-tagging WPs are defined by selecting jets above a given threshold on the discriminant variable built from the continuous outputs of the flavour-tagging algorithms. In the approach known as Pseudo Continuous B-Tagging, such discriminant variable is binned. Jets in high bins are tagged with increasing likelihood, and are used for tighter WPs. Jets belonging to the first bin are rejected.

Track-jet variables
Track-jet p_{T}
Track-jet η
Track-jet ϕ
Track-jet flavour label
Mass of the p_{T}-leading b- or c-hadron in the track-jet
p_{T} of the p_{T}-leading b- or c-hadron in the track-jet
η of the p_{T}-leading b- or c-hadron in the track-jet
ϕ of the p_{T}-leading b- or c-hadron in the track-jet
Event variables
Average number of interactions per event, $\langle\mu\rangle$
Jet-pair variables
Angular separation between two track-jets, ΔR

Fig. 9. – The list of variables used as features of the GNN [4].

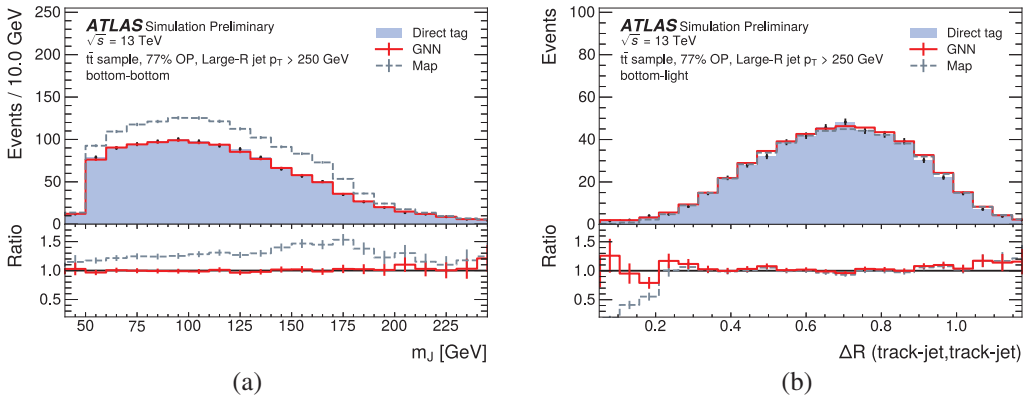


Fig. 10. – Comparison of the large-R jet mass (a) and ΔR (b) distribution obtained using direct and truth flavour-tagging with DL1r at the 77% WP on the first two p_{T}-leading track-jets associated to the large-R jet in $t\bar{t}$ events. The pass-or-fail flavour-tagging is represented by the blue-filled histogram. The black error bands represent the statistical error in direct flavour-tagging. The truth flavour-tagging obtained parametrising the efficiencies with the map method is in grey, while the GNN method is in red. The bottom pad shows the ratio with respect to the direct flavour-tagging distribution. Each plot refers to a different truth flavour combination of the track-jets associated to the large-R jet and used for flavour-tagging. Figure from [4].

jet ([9]) mass and of the ΔR between the two p_{T}-leading track-jets associated to the large-R jet, comparing the results obtained using the pass-or-fail flavour-tagging method to those obtained using maps and GNN-based truth flavour-tagging. The comparison illustrates that the shapes and normalizations of the pass-or-fail flavour-tagging distributions are better replicated by the GNN-based approach.

([9]) Large-radius jets are reconstructed from topological clusters of energy deposited in the calorimeters using the anti-k_{T} algorithm [16] with radius parameter $R = 1.0$.

8. – Conclusion

In the context of jet flavour-tagging in ATLAS, the capability of GNNs of modeling data with complex relational structures allows to surpass by far the performance of the existing algorithms. GN1, the ATLAS state-of-the-art flavour-tagging algorithm, improves jet rejection by factors of up to 6 when applied on simulated data.

The general nature of the problem that GNNs are designed to address, namely to capture the relations between nodes of a graph and with its overall structure, allows to use them for a wide variety of problems. A GNN-based method for regressing the flavour-tagging efficiency dependence on an high-dimensional set of parameters has been presented. It improves the accuracy in modeling the background for analyses of proton-proton collision events having multiple flavour-tagged jets.

REFERENCES

[1] ATLAS Collaboration, *JINST*, **3** (2008) S08003.
[2] Evans L. and Bryant P., *JINST*, **3** (2008) S08001.
[3] ATLAS Collaboration, ATL-PHYS-PUB-2022-027, https://cds.cern.ch/record/2811135.
[4] ATLAS Collaboration, ATL-PHYS-PUB-2022-041, http://cds.cern.ch/record/2825433.
[5] ATLAS Collaboration, *Phys. Lett. B*, **786** (2018) 59.
[6] ATLAS Collaboration, *JHEP*, **03** (2020) 145.
[7] ATLAS Collaboration, ATL-CONF-2022-070, https://cds.cern.ch/record/2842916.
[8] Shlomi Jonathan, Battaglia Peter and Vlimant Jean-Roch, *Mach. Learn.: Sci. Technol.*, **2** (2020) 021001.
[9] ATLAS Collaboration, *Eur. Phys. J. C*, **77** (2017) 673.
[10] ATLAS Collaboration, *JHEP*, **03** (2020) 145.
[11] ATLAS Collaboration, *Eur. Phys. J. C*, **82** (2022) 717.
[12] Di Bello F. A., Shlomi J., Badiali C. *et al.*, *Comput. Softw. Big Sci.*, **5** (2021) 14.
[13] Krohn D., Thaler J. and Wang L. T., *JHEP*, **06** (2009) 059.
[14] ATLAS Collaboration, ATL-PHYS-PUB-2017-010, https://cds.cern.ch/record/2268678.
[15] Sugiyama M., Suzuki T. and Takafumi K., *Density Ratio Estimation in Machine Learning* (Cambridge University Press) 2012.
[16] Cacciari M., Salam G. P. and Soyez G., *JHEP*, **04** (2008) 063.