GNN for b-tagging in LHC

Jorge Luis David Mesa*

High-energy physics experiments, such as those conducted at the Large Hadron Collider (LHC), necessitate precision in the identification and classification of particle jets, pivotal for probing fundamental physics. This research introduces the application of the Set2Graph (S2G) model, a Graph Neural Network (GNN) framework, for enhancing b-jet tagging efficiency in the CMS and ATLAS experiment. Employing a synthetic dataset derived from proton-proton collisions, the study simulates the ATLAS detector's characteristics to test the S2G model's efficacy in identifying secondary vertex features critical for b-jet discrimination. Methodologically, the model integrates particle interaction data and vertex characteristics into a graph-based structure, capturing complex relational information that traditional methods may overlook. Results indicate a notable improvement in b-jet tagging precision, with the model achieving high F1-scores, particularly for light jets, and demonstrating robustness in adjusting the Rand Index (RI) and Adjusted Rand Index (ARI) across varied jet types. Overall, the S2G model significantly advances the particle tagging process, suggesting potential for broader applications in more complex decay scenarios within high-energy physics. This study confirms the model's enhanced performance in vertex reconstruction.

I. INTRODUCTION

The high interaction rates, radiation doses, particle multiplicities, and energies, along with the requirements for precision measurements, have set new standards for the design of particle detectors at the Large Hadron Collider (LHC)[1]. Two general-purpose detectors, ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid), have been constructed to probe proton-proton (p-p) and nucleus-nucleus (A-A) collisions [2, 3].

The ATLAS[4] experiment employs a sophisticated detector system designed for high-precision particle detection. Central to this system is a superconducting solenoid, producing a magnetic field of about 2 T, essential for the deflection of charged particles for momentum measurements. The core components include a high-granularity silicon tracker, a liquid-argon-based electromagnetic calorimeter, and a steel-scintillator tile hadron calorimeter. These elements provide detailed tracking, momentum, and energy measurements of particles resulting from high-energy proton-proton collisions, covering a pseudorapidity range of $|\eta| < 4.9$ [5].

In parallel, the CMS experiment plays a pivotal role in high-energy physics research, probing phenomena from the understanding of the Higgs boson to the search for new particles beyond the Standard Model. At the core of its scientific endeavor lies an intricate detector system that employs a superconducting solenoid to generate a 3.8 T magnetic field, facilitating the precise measurement and deflection of particle trajectories. This system integrates critical detection components such as the silicon tracker, electromagnetic calorimeter, and hadron calorimeter, enabling accurate tracking of particles up to a pseudorapidity of $\eta < 2.5$ [6].

The precision inherent in the CMS detector's design is indispensable for the identification and analysis of particles produced in high-energy collisions, particularly in the context of b-tagging. B-tagging, the process of identifying jets coming from b-quarks, is crucial for the CMS experiment, given the unique properties of b-quarks, such as their comparatively long lifetimes and their significant travel distances before decay. This technique significantly enhances the experiment's capacity to discriminate b-jet signals from background noise, thereby improving measurement accuracy and discovery potential. The evolution of b-tagging techniques, notably through the adoption of machine learning algorithms [7], has substantially enhanced the efficiency and precision of b-quark identification within the CMS framework, marking a significant advancement in the experiment's capabilities.

Similarly, flavour tagging, the identification of jets originating from b- and c-quarks, is a critical component of the physics programme of the ATLAS experiment at the LHC. It is of particular importance for the study of the Standard Model (SM) Higgs boson and the top quark, which preferentially decay to b-quarks, and additionally for several Beyond Standard Model (BSM) resonances that readily decay to heavy flavour quarks [8, 9]. This aspect of ATLAS's research underscores the broader significance of b-tagging within high-energy physics, enhancing our understanding of particle interactions and the fundamental constituents of matter. Both CMS and ATLAS utilize advanced b-tagging methodologies to advance our knowledge of fundamental particles, illustrating the synergy in technological advancements and scientific discoveries at the LHC.

A. Algorithms for b-jet identification

1. *Reconstructed objects used in b-jet identification*

In b-jet identification, jets are clustered using the particle-flow algorithm [10, 11] that integrates data from

* jorge.david1@udea.edu.co; Physics Department, Universidad de Antioquia

all subdetectors, forming a consistent set of particles for each event. These particles are then clustered into jets using the anti-kT algorithm [12] with a specific distance parameter, and raw jet energies are subsequently adjusted for uniform response and precise calibration [13].

The b-jet tagging process utilizes kinematic data from charged particles within the jet, which are reconstructed using an iterative procedure based on a Kalman filter-based method [14]. Further accuracy is achieved by integrating tracks from muon chambers with those from the silicon tracker, enhancing muon track reconstruction [15].

Primary vertices are identified by clustering tracks based on proximity to the beam line, with the most likely vertex candidate determined by an adaptive vertex fit [16]. For b-jet tagging, tracks must meet stringent criteria including a minimum transverse momentum and a specific number of detector hits, ensuring high purity and accurate track association to jets. These criteria help reduce the presence of poorly reconstructed tracks and improve the robustness of the b-jet tagging algorithms used in high-energy physics experiments [17].

2. Identification using track impact parameters

In b-jet identification, the impact parameter (IP) is a critical measure used to differentiate between decay products of b hadrons and prompt tracks, calculated with precision due to the high-resolution capabilities of the pixel detector. The IP's sign correlates with the scalar product of the vector from the primary vertex to the track's closest approach and the jet direction, typically positive for tracks from particle decay along the jet axis.

The significance of the impact parameter (SIP), defined as the ratio of the IP to its estimated uncertainty, is a key discriminator used to enhance track analysis. This measure effectively identifies tracks emanating from decay processes, with its resolution influenced by the track's transverse momentum (pT) and pseudorapidity (η) [17].

B-jet tagging leverages SIP through algorithms like Track Counting (TC), which ranks tracks by decreasing SIP values. The TC variants, Track Counting High Efficiency (TCHE) and Track Counting High Purity (TCHP), specifically use the SIP of the second and third-ranked tracks for further discrimination.

Extensions of these methods, such as the Jet Probability (JP) and Jet B Probability (JBP) algorithms, assess the collective likelihood of tracks originating from the primary vertex. They use data-driven probability functions for IP significance to refine the identification process, focusing on tracks with the highest SIP to enhance the detection of b-hadron decays. These approaches represent advanced techniques in particle physics experiments to improve b-jet tagging accuracy by carefully analyzing track properties and their impact parameters [17].

3. Identification using secondary vertices

In b-jet identification, secondary vertices are crucial for distinguishing b from non-b jets. Discriminative variables include the flight distance and direction from the primary to secondary vertices, and properties like the multiplicity and mass of secondary tracks. To enhance b purity, secondary vertex candidates must meet stringent criteria: they should share less than 65 % of tracks with the primary vertex, their radial distance must exceed 3σ in significance, and those with a radial distance over 2.5 cm or mass indicators of K0 or above 6.5 GeV/c^2 are excluded to avoid contamination from interactions with detector material or decays of long-lived mesons. Additionally, the flight direction must align within a cone of $\Delta R < 0.5$ around the jet direction.

The Simple Secondary Vertex (SSV) algorithms discriminate based on the flight distance significance, with versions tailored for different levels of purity based on the number of associated tracks. The Combined Secondary Vertex (CSV) algorithm extends this by incorporating track-based lifetime information, providing robust discrimination even when secondary vertices are not clearly defined, thereby significantly enhancing the accuracy of b-jet identification [17].

The algorithms employed for b-jet identification represent a critical area of focus within the CMS experiment. These algorithms leverage various aspects of particle interactions, including the decay products of B hadrons and the properties of secondary vertices, to effectively distinguish b-jets. Of particular interest in this paper is the identification strategy based on secondary vertices. This approach utilizes variables associated with secondary vertices—such as flight distance and mass—coupled with sophisticated selection criteria, to enhance discrimination between b-jets and other background signals. By employing algorithms that focus on secondary vertex characteristics, the precision and efficiency of b-jet identification are significantly improved.

This paper titled "GNN for b-tagging in LHC" delves into the application of Graph Neural Networks (GNNs), specifically through the Set2Graph (S2G)[18] model, in the identification of b-jets using secondary vertices. The S2G model, known for its capacity to approximate any equivariant function from a set to a graph, presents a novel approach to analyzing particle interactions within high-energy collisions. By exploiting the relational data among particles as represented in graph structures, GNNs, and particularly the S2G model, offer significant advantages in the b-tagging process, enhancing both the efficiency and precision of b-quark identification.

In addressing the challenges posed by the CMS experiment's vast data volumes and the complexity of particle detection, deep learning, and specifically GNNs, emerge as potent solutions. These technologies excel in feature extraction, pattern recognition, and the management of high-dimensional data, making them particularly suited

for analyzing the intricate data produced by the CMS detector. By integrating the S2G model into the b-tagging process, this paper aims to underscore the transformative potential of GNNs in particle physics research, setting a new benchmark for the application of deep learning techniques in the field.

II. METODOLOGY

A. DATA

The algorithm under study was evaluated using a synthetic dataset, derived from $pp \to t\bar{t}$ events at a collision energy of $\sqrt{s} = 14$ TeV. Event generation was facilitated through PYTHIA8 [19], followed by a rudimentary detector simulation via DELPHES [20], designed to mimic the ATLAS detector's characteristics [4]. The simulation detailed the representation of charged particle tracks through six perigee parameters as shown in TABLE I. Gaussian smearing introduced noise to these parameters, emulating realistic detector responses that vary with each track's transverse momentum (pT) and pseudorapidity (η) in a manner analogous to ATLAS's documented measurements [5]. It's noteworthy that the smearing process was applied independently to each track parameter without correlating effects, maintaining a diagonal covariance matrix configuration.

Jet reconstruction was achieved using calorimeter energy readings and the anti-kT algorithm [12] with an R parameter of 0.4. Additionally, a cone association method was employed, linking charged tracks to jets within a ΔR radius less than 0.4 around the jet axis. Jet flavors were categorized (into bottom, charm, or light) through the association of weakly decaying bottom and charm hadrons within a ΔR cone of 0.3.

A basic criterion for jet selection was established, necessitating that jets possess a pT greater than 20 GeV and an $|\eta| < 2.5$. The input variables for vertex finding algorithms encompassed the tracks associated with each jet, jet pT, η, ϕ, and jet mass as shown in TABLE I.

The dataset was methodically assembled to preserve a uniform representation of jet flavors within specified bins characterized by transverse momentum (pT), pseudorapidity (η), and the number of tracks (ntracks). This was achieved by sampling an equivalent number of jets for each flavor in each bin, with the sample size in each bin determined by the least prevalent flavor, typically c jets. This strategy ensured a balanced distribution of jet flavors across the different parameters, which is detailed in the figure provided. The figure depicts three key aspects: the frequency distribution of jet pT for light jets (top left), the pseudorapidity distribution for c jets (top middle), the proportion of each jet flavor (top right), and the secondary vertex counts for light, c, and b jets (bottom panels). The data was sectioned into sets for training (500k jets), validation, and testing (100k jets each).

B. SET2GRAPH NEURAL NETWORK

The S2G network is a neural architecture designed to transform sets into graph structures by sequentially applying three distinct modules: ϕ, β, and ψ. Each module is engineered to manipulate and refine the data progressively, facilitating the transformation from sets to a graph representation as detailed in Figure 2.

Set-to-Set Component ϕ: As a core component of set-to-graph functions, ϕ serves as a set-to-set equivariant model. It accepts an input matrix of dimensions $n_{\text{tracks}} \times d_{\text{in}}$ and transforms it into a hidden representation of each track with dimensions $n_{\text{tracks}} \times d_{\text{hidden}}$. This transformation is achieved through a series of deep set layers [21] incorporating self-attention mechanisms and linear transformations, allowing ϕ to approximate arbitrary continuous set-to-set functions universally. The attention mechanism [22, 23] within this module is defined as:

$$\text{Attention}(X) = \text{softmax}\left(\frac{\tanh f_1(X) \cdot f_2(X)^T}{\sqrt{d_{\text{small}}}}\right) \cdot X$$

where X is the input, and f_1, f_2 are the key and query Multilayer Perceptrons (MLPs).

Broadcasting Layer β: β acts as a broadcasting layer mapping the outputs from ϕ to k-edges. It constructs a representation for each directed edge by concatenating the individual track representations and their summation, resulting in an output dimension of $(n_{\text{track}}(n_{\text{track}} - 1)) \times 3d_{\text{hidden}}$. This module enables the transition from sets to complex relational structures by linearly combining the transformed sets into higher-dimensional tensors.

Edge Classifier ψ: Concluding the transformation process, ψ maps k-tensors to output tensors, serving as an edge classifier. Utilizing a MLP, ψ processes the edge representations to generate an edge score for each pair. These scores are symmetrized during inference to reflect the unordered nature of track pairs:

$$s_{ij} = \sigma\left(\frac{1}{2}(\psi(\text{track}_i, \text{track}_j) + \psi(\text{track}_j, \text{track}_i))\right)$$

ψ illustrates the application of minimal yet effective neural operations on complex structured data, emphasizing the model's capability to adaptively process and classify relational data.

In essence, the S2G model effectively encapsulates the functionality needed to transform sets through set-to-set, broadcasting, and edge classification stages into a comprehensive graph structure, aligning with the architectural and functional descriptions provided in Figure 2.

S2G Training

The training of the S2G model was undertaken with a batch size of 2048, using an Adam optimizer as described in [24], with a learning rate set at 10^{-3}. The entire training process was conducted on a MacBook Air equipped

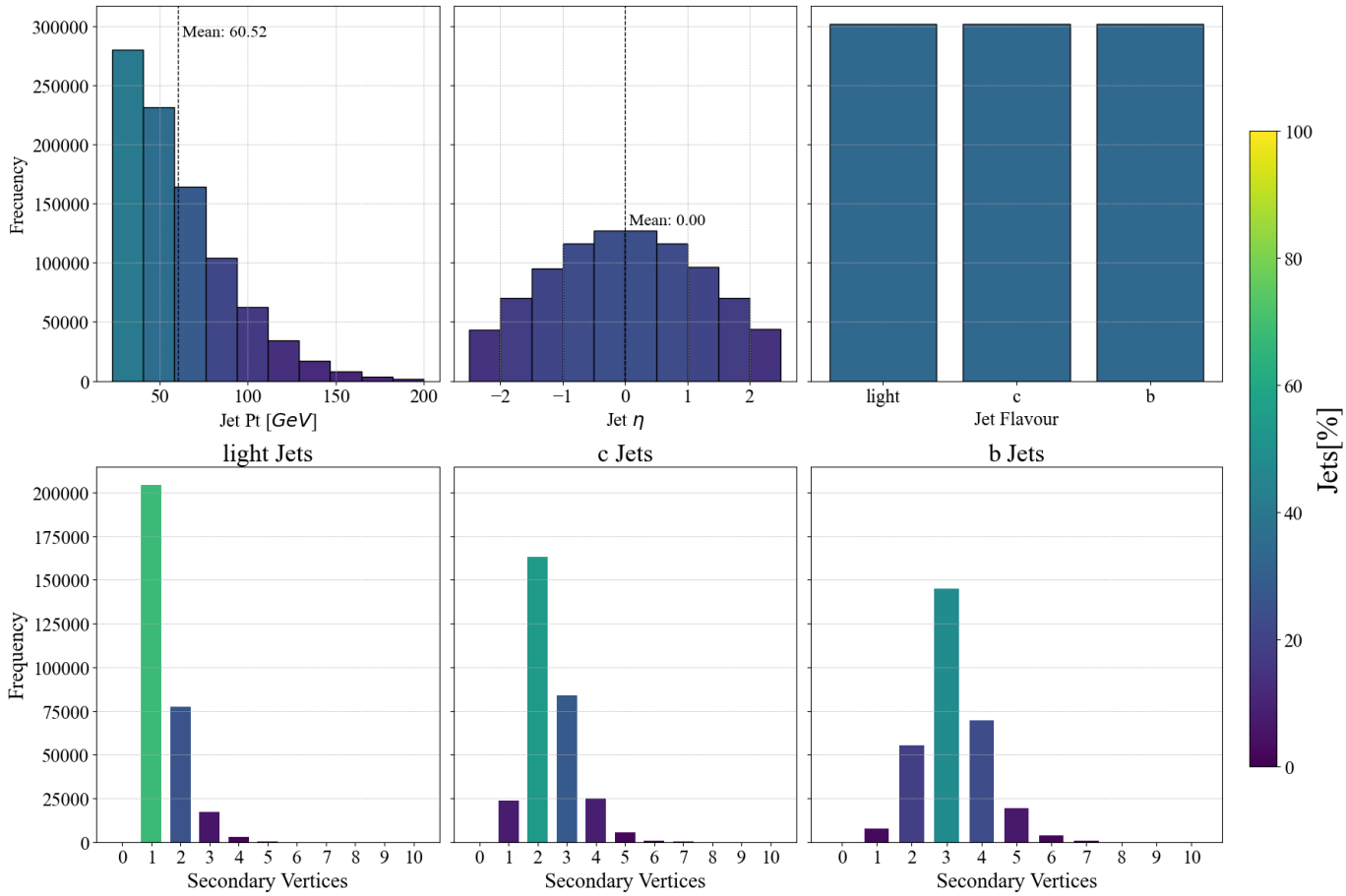

FIG. 1. Distribution of jet characteristics from the simulated dataset. The upper left and upper middle histograms indicating the frequency of the transverse momentum (pT) and pseudorapidity (η) for all jets. The upper right bar graph enumerates the relative composition of jet flavors in the dataset. The lower histograms detail the number of secondary vertices associated with each jet flavor. Collectively, these visualizations depict the dataset's underlying structure, which informs the jet flavor tagging and vertex reconstruction in simulation.

with an M1 chip and 16 GB of RAM. Over the course of three separate training sessions, the model underwent a total of 151 epochs, with the first session completing 30 epochs, the second 86, and the third 45 epochs.

The mean runtime per epoch was recorded at 16 minutes and 53 seconds, accumulating to a comprehensive training duration of 1 day, 18 hours, and 47 minutes. The criterion for cessation of training in each session was a lack of improvement in the validation loss for a continuous span of 20 epochs, a strategy implemented to mitigate overfitting and ensure that the model's generalization capabilities were maintained.

C. Jet Classification Model Architecture

The Jet Classification model, depicted in Figure 3, integrates four core components, synergistically operating to classify jets effectively. This architecture employs the S2G network, as outlined earlier, for vertex finding, a pivotal preliminary step in the process. The components are:

1. Deep Set Network: As described in Appendix B, the Deep Set Network forms the foundation for hidden representation. Within this classification framework, it operates with dimensions (126, 126, 126, 126), creating a hidden representation for each track in the input, facilitating information exchange between tracks.

2. Vertex Finding Module: Utilizing the S2G network, this module ascertains vertex assignment, essential for edge feature determination between any two tracks within the input set. The resulting hidden representation from the vertex finding module, in conjunction with the Deep Set Network, provides enriched edge features for the fully connected graph of tracks.

3. Graph Network: This network comprises a succession of Graph Network (GN) blocks, which include both an edge update and a node update Multilayer Perceptron (MLP). The edge update MLP E_l has dimensions $126 \cdot 3 + 1, 100, 20$, and the node update MLP U_l is dimensioned at $126 + 20, 100, 126$. These GN blocks iterate to evolve the node hidden representation h_i^l and the global

TABLE I. variables.

	Variable	Definition
Track features	d_0	Transverse Impact Parameter: Shortest distance from the trajectory to the primary interaction point in the plane perpendicular to the beam axis.
	$z0$	Longitudinal Impact Parameter: Distance along the beam axis from the primary interaction point to the point where trk_d0 is measured.
	ϕ	Azimuthal Angle: Angle in the transverse plane from a reference direction to the projection of the trajectory.
	ctgθ	Cotangent of the Polar Angle: Cotangent of the angle between the trajectory and the beam axis.
	pt	Transverse Momentum: Component of momentum perpendicular to the beam axis.
	q	Charge: Electric charge of the particle, influencing its trajectory in a magnetic field.
Jet features	pt	Jet Transverse Momentum: The component of a jet's total momentum that is perpendicular to the beam axis.
	η	Jet Pseudorapidity: A measure of the angle of the jet relative to the beam axis, where a value of 0 means the jet is perpendicular to the beam and values approaching infinity indicate the jet is aligned with the beam.
	ϕ	Jet Azimuthal Angle: The angle around the beam axis from a fixed reference direction, measuring the direction of the jet in the transverse plane.
	m	Jet Mass: The invariant mass calculated from the energy and momentum of all the particles that make up the jet, representing the "weight" of the jet.

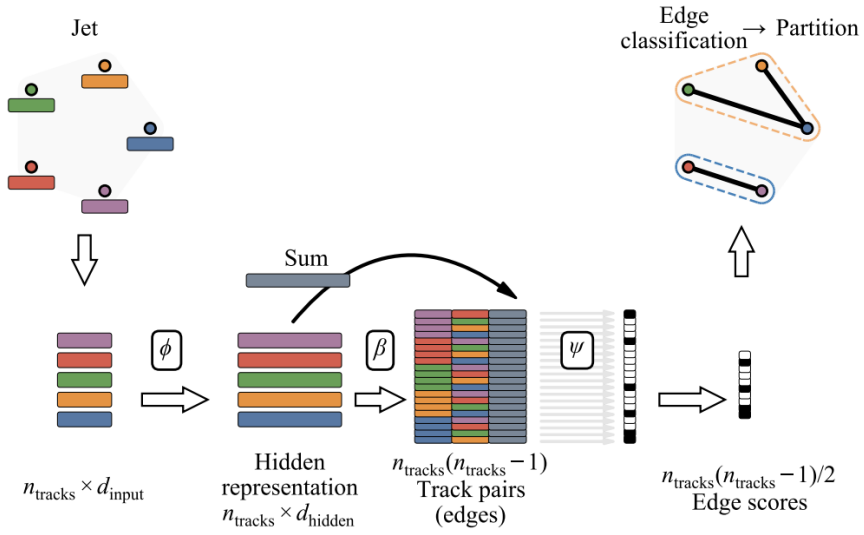

FIG. 2. S2G model, jet tracks are effectively partitioned using a neural network approach. Initially, a set-to-set component, denoted as ϕ, processes each track to generate a hidden representation of dimension d_{hidden}. Subsequently, a broadcasting layer, β, constructs a representation for each directed edge, which consists of an ordered pair of tracks within the jet. This is achieved by merging the representations of the individual tracks along with the aggregate of all track representations. Finally, an edge classifier, ψ, is applied to these directed edges to complete the partitioning process. The figure was taken from [25]

graph representation g^l, with the updates governed by:

$$g^l = \sum_i h_i^l$$

$$m_i^{l+1} = \sum_{j \in N(i)} E_l(h_i^l, h_j^l, e_{ij}^l)$$

$$h_i^{l+1} = U_l(h_i^l, m_i^{l+1})$$

where $N(i)$ denotes the node neighborhood, and e_{ij} the edge prediction between nodes i and j.

4. Jet Classifier MLP: The final classification stage involves the Jet Classifier MLP, which sizes up the collective hidden representations of the tracks alongside jet features (transverse momentum p_T, pseudorapidity η, and jet mass) to categorize the jet as either a b, c, or light jet. The dimensions for this MLP are (126 + 4, 100, 50, 3).

Jet Classifier Training

The training of the Jet Classifier was conducted over

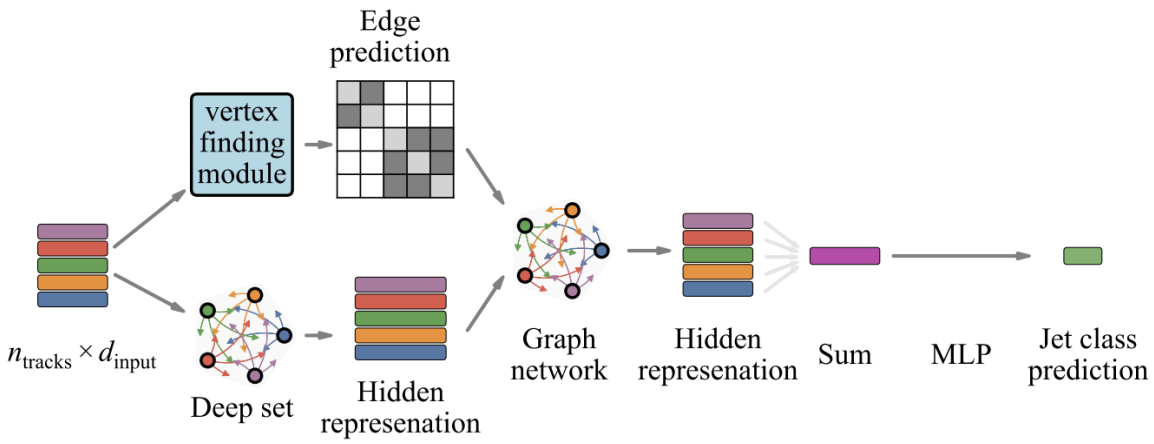

FIG. 3. The depicted Jet Classification model architecture integrates a S2G network as its vertex finding module, as detailed in the S2G model subsection. This module processes the input of track data, given as an $n_{\text{tracks}} \times d_{\text{input}}$ matrix, to perform edge prediction. The predictions then flow into the Graph network, which further refines the data to enhance hidden representations of the tracks. These representations are summarized and passed through a Multilayer Perceptron (MLP), culminating in the jet class prediction. The S2G network's weights are frozen during the classifier's training. The figure was taken from [25]

32 epochs with a total runtime of 2 hours, 21 minutes, and 40 seconds. On average, each epoch took approximately 4 minutes and 26 seconds. The training was performed on a MacBook Air with an M1 chip and 16 GB of RAM, showcasing the chip's capability for machine learning tasks.

A batch size of 1000 was employed, optimizing the network with an Adam optimizer set to a learning rate of 5×10^{-4}. Cross-entropy loss was used as the loss function to guide the learning process. Training proceeded until a plateau in performance improvement was detected, characterized by a lack of decrease in validation loss for a span of 20 epochs, at which point training was halted to prevent overfitting and ensure generalization of the model to unseen data.

D. Overall Jet Performance

To assess the performance of jet clustering algorithms comprehensively, we employ the adjusted Rand index (ARI), which offers a quantifiable metric of similarity between two set partitions, as referenced in [26]. Particularly pertinent to vertex finding, where ground truth data is precisely defined, the ARI serves as a scoring system, indicating the fidelity with which our algorithm replicates the true set partition.

The ARI is an enhancement of the Rand index (RI), a statistical measure used for gauging the agreement between two partitions. At its core, the RI is the ratio of the number of correct decisions (correct edges) made by the algorithm to the total number of decisions (edges in the set), expressed as:

$$RI = \frac{\text{number of correct edges}}{\text{number of edges in the set}}$$

Correct edges are those for which the algorithm's label matches the actual label in the ground truth, accounting for true positives and true negatives. To adjust for chance, the ARI is calculated by normalizing the RI against its expected value:

$$ARI = \frac{RI - \mathbb{E}[RI]}{1 - \mathbb{E}[RI]}$$

The expected value of the RI, $\mathbb{E}[RI]$, is contingent upon the clustering model adopted. In this context, we use the "one-sided" comparison model as detailed in [27]. This model treats the true vertex assignment as fixed and computes the expectation value based on random vertex assignments for algorithm predictions. The formula for the expected RI in this model is:

$$\mathbb{E}[RI] = \frac{B_{N-1}}{B_N} \frac{\sum_i \binom{g_i}{2}}{\binom{N}{2}} + \left(1 - \frac{B_{N-1}}{B_N}\right)\left(1 - \frac{\sum_i \binom{g_i}{2}}{\binom{N}{2}}\right)$$

Here, N equals the number of tracks (n_{tracks}), B_N denotes the Bell number representing all possible partitions of a set with N elements, and the summation is taken over the tracks in the jet with g_i denoting the number of tracks in the i-th vertex.

The ARI score interprets to 1 for an algorithm that flawlessly identifies the correct cluster assignments, akin to a perfect classification. A score of 0 would indicate a level of cluster assignment as random as guessing.

III. RESULTS

The performance of the S2G model is systematically assessed across three different jet types—b jets, c jets, and light jets—with the results summarized in Table II. The metrics employed for this evaluation include Precision (P), Recall (R), F1-Score (F1), Rand Index (RI), and Adjusted Rand Index (ARI). These metrics collectively gauge the model's accuracy, sensitivity, and the similarity between the predicted and actual clustering of vertex data.

	P	R	F1	RI	ARI
b jets	0.5796	0.6752	0.5895	0.7552	0.3742
c jets	0.6611	0.8409	0.7078	0.7191	0.3828
light jets	0.9588	0.9422	0.9427	0.9376	0.8917

TABLE II. Performance metrics of the S2G model. Precision (P), recall (R), F1 score, Rand Index (RI), and Adjusted Rand Index (ARI) are tabulated, with light jets exhibiting high scores across all metrics. B jets and C jets show lower values, especially in precision and ARI.

	Algorithm	**F1**	**RI**	**ARI**
b jets	AVR	0.56	0.61	-0.01
	Track Pair	0.62	0.74	0.32
	RNN	0.59	0.75	0.37
	Set2Graph	**0.66**	**0.78**	**0.43**
c jets	AVR	0.70	0.65	0.22
	Track Pair	0.74	0.73	0.40
	RNN	0.71	0.72	0.40
	Set2Graph	**0.75**	**0.75**	**0.45**
light jets	AVR	0.97	0.96	0.93
	Track Pair	0.96	0.96	0.93
	RNN	0.93	0.92	0.87
	Set2Graph	**0.97**	**0.96**	**0.94**

TABLE III. Performance metrics comparison across different algorithms for b jets, c jets, and light jets from [25].

Precision quantifies the proportion of true positive predictions among all positive predictions made by the model. In our analysis, light jets exhibit the highest precision of 0.9588, indicating a substantial accuracy in the model's ability to identify correct vertices for this jet type. In contrast, b jets and c jets display lower precision scores of 0.5796 and 0.6611 respectively, suggesting a propensity towards higher false positive rates.

Recall, on the other hand, measures the model's capability to identify all actual true positive instances. Here again, light jets demonstrate a high recall of 0.9422, confirming that the model effectively detects the true vertices. Conversely, c jets and b jets show lower recall values of 0.8409 and 0.6752, respectively, indicating missed true vertex instances.

The F1-Score, a harmonic mean of precision and recall, reflects a balanced measure of the model's precision and recall capabilities. Light jets score the highest F1-score at 0.9427, while c jets and b jets achieve scores of 0.7078 and 0.5895, respectively, highlighting a need for enhanced accuracy and reliability in vertex identification for these jet types.

The Rand Index and Adjusted Rand Index further corroborate these findings. Light jets report a high RI of 0.9376 and an ARI of 0.8917, affirming a close match between the predicted and actual clustering patterns. The indices for b jets and c jets, however, suggest room for refinement with RI scores of 0.7552 and 0.7191, and ARI scores of 0.3742 and 0.3828 respectively, pointing towards potential misclassifications.

For comparative analysis, Table III from [25] serves as a benchmark to evaluate the consistency of the S2G model across independent research. The S2G model shows similar performance trends in this study for b jets and c jets with comparable F1, RI, and ARI scores. Particularly, b jets in both studies show closely aligned outcomes with F1 scores around 0.66, RI scores near 0.78, and ARI values around 0.43, reinforcing the model's effectiveness in vertex finding. For c jets and light jets, our findings also mirror those of the external study, with nearly identical performance metrics, thus demonstrating consistent model performance.

In the left plot Figure 4, the Rand Index (RI) is plotted against the number of training epochs. This index is a measure of the similarity between two partitions, offering insight into how well the model is clustering or finding vertices in the data. The blue line represents the RI for the training set, while the cyan line indicates the RI for the validation set. Notably, the RI for both datasets rapidly increases in the early epochs, indicating that the model quickly establishes a reliable clustering pattern. As training progresses, the RI plateaus with minor fluctuations, signifying that the model has reached a consistent level of performance.

The right plot Figure 4 illustrates the loss function over the same training epochs, with the training loss (blue) and validation loss (cyan) depicted separately. The loss function reflects the error rate during training, indicating how well the model is optimizing over time. In this graph, the training loss shows a steady decline, indicative of the model learning effectively from the data. Meanwhile, the validation loss exhibits more fluctuations, but maintains a downward trend, suggesting that the model is generalizing to unseen data. This behavior points to a balance between learning and avoiding overfitting.

The left plot Figure 5 for jet-classification displaying accuracy over 30 epochs indicates a steady increase in training accuracy, represented by the blue line. This upward trajectory is a positive indication of the model's learning process, suggesting that it is progressively improving its predictions. The cyan line representing validation accuracy shows more variability, with a series of fluctuations that could be due to noise in the validation data or potential overfitting. Nonetheless, the general upward trend in validation accuracy suggests that the model is learning to generalize beyond the training dataset.

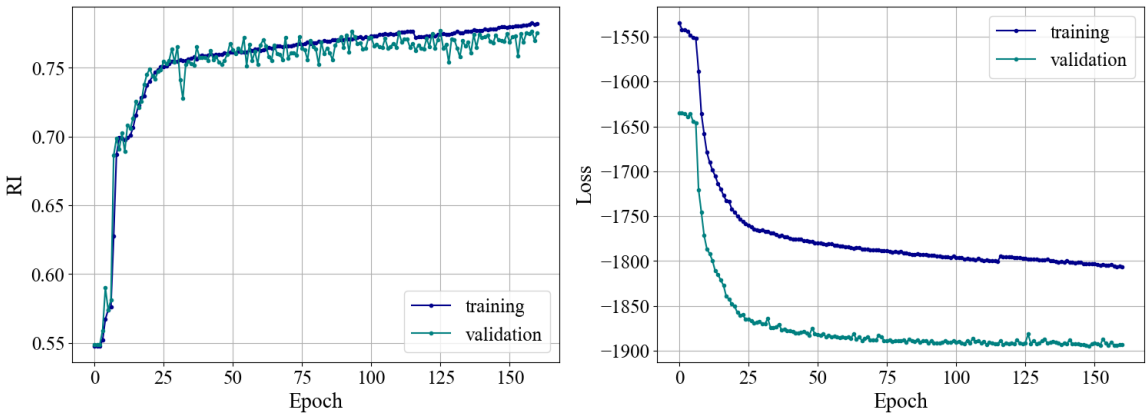

FIG. 4. Training performance of the S2G model, illustrated through two metrics over epochs. The left graph displays the Rand Index (RI) for the training (blue) and validation (cyan) datasets, showing rapid initial improvement followed by stabilization. The right graph demonstrates the decrease in loss for both training (blue) and validation (cyan), with the validation loss plateauing, indicative of the model reaching optimal generalization.

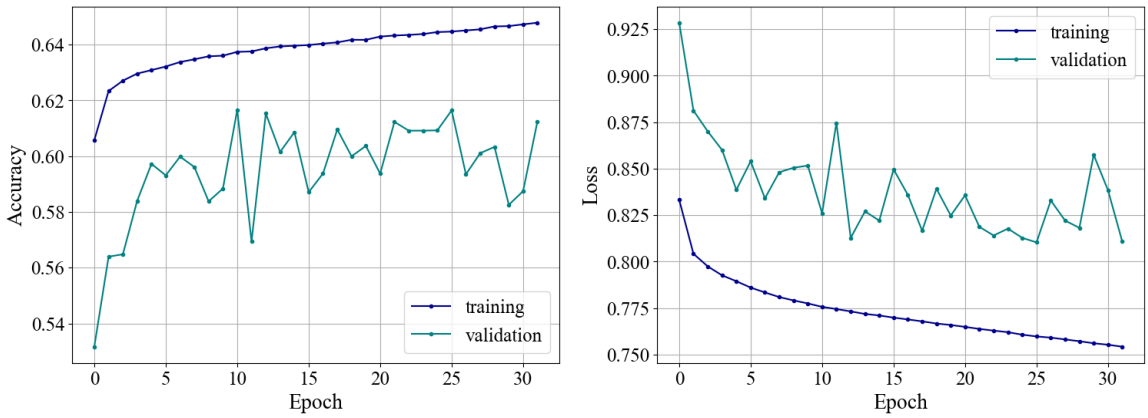

FIG. 5. Training and validation curves for the jet tagging model. The left graph tracks accuracy over 30 epochs, with training (blue) demonstrating a steady increase before plateauing, and validation (cyan) showing variability around an upward trend. The right graph plots loss, revealing a sharp decrease for training (blue) and significant oscillations in validation (cyan), suggesting variability in model performance on validation data.

The right plot Figure 5, focusing on loss over the same 30 epochs, reveals a consistent decline in training loss, also depicted by the blue line. This steady decrease suggests that the model is effectively minimizing error during training. The validation loss, however, presents a more fluctuating pattern, though it follows a downward trend overall, indicating that the model's generalization capability improves over time. The observed fluctuations might point to areas where further tuning could enhance the stability of the model.

The accuracy graph for b jets in Figure 6 shows stable training accuracy, with slight variations in the validation accuracy. Despite these fluctuations, the validation accuracy remains relatively high, suggesting that the model performs adequately with b jets. The variability could be due to differences in the dataset or model sensitivity to certain b jet features, indicating a need for further refinement to reduce this inconsistency.

In the case of c jets, the training accuracy shows a minor upward trend, while the validation accuracy exhibits considerable fluctuations. These variations could indicate overfitting or the model's struggle to maintain consistent performance across different subsets of c jet data. Nonetheless, the overall upward trend in validation accuracy suggests that the model can adapt over time, albeit with some instability.

For light jets, the accuracy graph demonstrates high and stable training accuracy, suggesting that the model has learned to classify this jet type effectively.

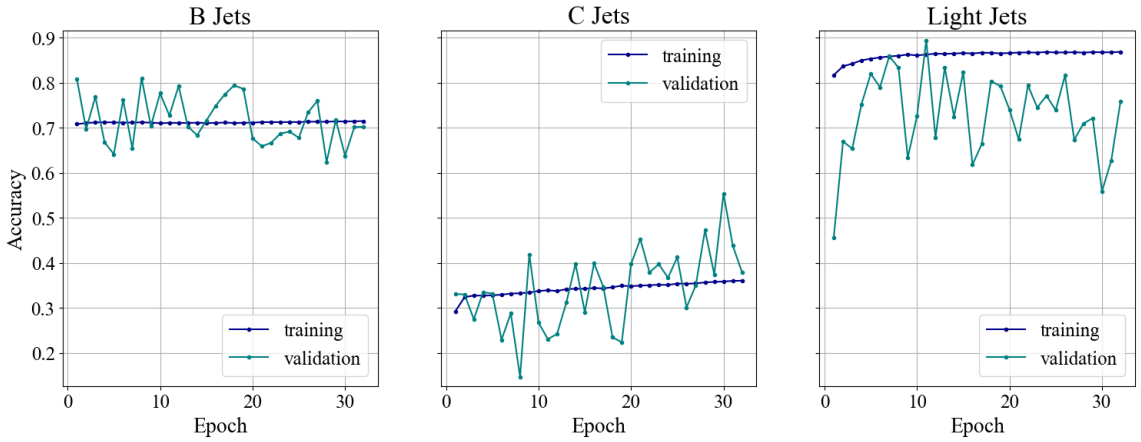

FIG. 6. Training (blue) and validation (cyan) accuracy over 30 epochs for jet tagging, separated by jet types: B Jets (left), C Jets (center), and Light Jets (right). B Jets show a relatively stable accuracy for both training and validation with minor fluctuations. C Jets exhibit greater variability, especially in validation accuracy. Light Jets demonstrate a similar trend with training accuracy outpacing validation, indicating potential overfitting or a need for improved generalization in the model for C and Light Jets.

Epoch	Mean Accuracy	b-jets	c-jets	Light jets
8	0.5961	**0.8107**	0.1465	0.8336
30	0.5825	0.6378	**0.5538**	0.5588
32	**0.6124**	0.7022	0.3788	0.7585

TABLE IV. Results of jet-classification where accuracy is maximum for b-jets, c-jets, and overall. Bold entries highlight the maximum accuracies for each category.

IV. CONCLUSIONS

In this study, we conducted a comprehensive comparative analysis between our results and those from a previously published paper, focusing particularly on the application of the Set2Graph (S2G) model in vertex finding and jet classification. Our findings affirm and reinforce the conclusions drawn in the previous study [25], demonstrating remarkable similarity in the performance metrics obtained, particularly in the context of jet b-tagging.

Both studies illustrate the efficacy of the S2G model, highlighting its universality and equivariance as crucial attributes that enhance vertex reconstruction. The universality of the S2G model enables effective learning from comprehensive track information within the jet, while its inherent equivariance provides a significant advantage by effectively incorporating relational information among data points, which is especially beneficial in jet b-tagging. Our results specifically underscore the model's robustness in improving jet classification accuracy through the precise identification of vertex structures.

By utilizing the vertex findings as a form of supervised input to the jet classifiers, we have demonstrated substantial improvements in classification performance. This approach, akin to a supervised attention mechanism, distinctly outperforms traditional models that rely on unsupervised attention mechanisms. The improved accuracy in vertex reconstruction directly contributes to enhanced b jet tagging, reinforcing the critical role of precise vertex identification in complex classification tasks.

Looking forward, the promising results obtained here suggest significant potential for applying the S2G model to more complex decay scenarios and in more challenging experimental environments, such as those involving boosted Higgs decays to ($b\bar{b}/c\bar{c}$) and datasets incorporating full detector simulations with pileup interactions. Future research will benefit from exploring these advanced applications, which may further substantiate the model's utility and extend its applicability in particle physics research, particularly enhancing its impact on b jet tagging within these complex environments. This progression will be crucial in pushing the boundaries of accuracy and efficiency in high-energy physics experiments.

[1] L. Evans and P. Bryant, Lhc machine, JINST **3** (08), S08001, cit. on p. 2.
[2] G. Aad *et al.*, *ATLAS detector and physics performance: Technical Design Report*, Tech. Rep. (ATLAS Collaboration, CERN, Geneva, 1999) electronic version not available.
[3] G. L. Bayatian *et al.*, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*, Tech. Rep. (CMS Collaboration, CERN, Geneva, 2006).
[4] G. A. et al., The atlas experiment at the cern large hadron collider, Journal of Instrumentation **3** (08), S08003.

[5] *The ATLAS Detector* (Springer, Cham, 2018) pp. 71–92.

[6] CMS Collaboration, The cms experiment at the cern lhc, JINST **3** (08), S08004.

[7] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban, and D. Whiteson, Jet flavor classification in high-energy physics with deep neural networks, Phys. Rev. **D94**, 112002 (2016).

[8] ATLAS Collaboration, Observation of h $\rightarrow$ bb^{-} decays and vh production with the atlas detector, Phys. Lett. B **786**, 59 (2018), cit. on p. 2, arXiv:arXiv:1808.08238 [hep-ex].

[9] ATLAS Collaboration, Observation of higgs boson production in association with a top quark pair at the lhc with the atlas detector, Phys. Lett. B **784**, 173 (2018), cit. on p. 2, arXiv:arXiv:1806.00425 [hep-ex].

[10] CMS Collaboration, Particle-flow event reconstruction in cms and performance for jets, taus, and missing et, CMS Physics Analysis Summary **CMS-PAS-PFT-09-001** (2009).

[11] CMS Collaboration, Commissioning of the particle-flow reconstruction in minimum-bias and jet events from pp collisions at 7 tev, CMS Physics Analysis Summary **CMS-PAS-PFT-10-002** (2010).

[12] M. Cacciari, G. P. Salam, and G. Soyez, The anti-kt jet clustering algorithm, JHEP **04**, 063, arXiv:0802.1189.

[13] CMS Collaboration, Determination of jet energy calibration and transverse momentum resolution in cms, JINST **06**, P11002, arXiv:1107.4277.

[14] CMS Collaboration, Cms tracking performance results from early lhc operation, Eur. Phys. J. C **70**, 1165 (2010).

[15] CMS Collaboration, Performance of cms muon reconstruction in pp collision events at sqrt(s) = 7 tev, JINST **7**, P10002, arXiv:1206.4071.

[16] W. Waltenberger, R. Fruhwirth, and P. Vanlaer, Adaptive vertex fitting, J. Phys. G **34**, N343 (2007).

[17] T. C. collaboration, Identification of b-quark jets with the cms experiment, Journal of Instrumentation **8** (04), P04013–P04013.

[18] H. Serviansky, N. Segol, J. Shlomi, K. Cranmer, E. Gross, H. Maron, and Y. Lipman, Set2graph: Learning graphs from sets (2020), arXiv:2002.08772 [cs.LG].

[19] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, An introduction to pythia 8.2, Computer Physics Communications **191**, 159 (2015).

[20] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, Delphes 3: a modular framework for fast simulation of a generic collider experiment, Journal of High Energy Physics **2014** (2014).

[21] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov, and A. J. Smola, Deep sets, in *Advances in Neural Information Processing Systems* (2017) pp. 3391–3401.

[22] M. Ilse, J. M. Tomczak, and M. Welling, Attention-based deep multiple instance learning, arXiv preprint (2018), arXiv:1802.04712.

[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, in *Advances in Neural Information Processing Systems 30* (2017) pp. 5998–6008.

[24] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv preprint (2014), arXiv:1412.6980.

[25] J. Shlomi, S. Ganguly, E. Gross, K. Cranmer, Y. Lipman, H. Serviansky, H. Maron, and N. Segol, Secondary vertex finding in jets with neural networks, The European Physical Journal C **81**, 10.1140/epjc/s10052-021-09342-y (2021).

[26] L. Hubert and P. Arabie, Comparing partitions, Journal of Classification **2**, 193 (1985).

[27] A. Gates and Y.-Y. Ahn, The impact of random models on clustering similarity, Journal of Machine Learning Research **18** (2017).