Tagging Emerging Jets using Graph Neural Networks

by

Paras Pokharel

B.Sc., Howard University, 2021

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master in Science

in the
Department of Physics
Faculty of Science

SIMON FRASER UNIVERSITY
Spring 2024

Declaration of Committee

Name:	**Paras Pokharel**
Degree:	**Master in Science**
Thesis title:	**Tagging Emerging Jets using Graph Neural Networks**
Committee:	**Chair:** Andrei Frolov Professor, Physics
	Matthias Danninger Supervisor Assistant Professor, Physics
	Bernd Stelzer Committee Member Professor, Physics
	Dugan O'Neil Examinor Professor, Physics

Abstract

Many extensions to the Standard Model predict the existence of strongly interacting dark sectors, which behave similarly to QCD but interact weakly with SM particles. Depending on the parameters of the dark sector, one potential signature could be that of an "emerging jet". In particle detectors such as the ATLAS experiment, this signature could be seen as sprays of many *displaced tracks* - particle trajectories that do not originate from the main proton-proton interaction point, and many *displaced vertices* - the common origin of tracks that are away from the primary proton-proton interaction point. Graph Neural Networks (GNN) have shown great promise in capturing complex dependencies and patterns in graph-structured data, making them well-suited for analyzing the intricate topology of emerging jets. A GNN-based flavor tagging algorithm has recently been deployed in ATLAS and significantly outperforms previous taggers. Its architecture is used to tag emerging jets with high accuracy while significantly suppressing the QCD background. The architecture of the GNN also enables the classification of displaced tracks as well as the identification of displaced vertices within the jet cone, providing valuable insight into the topology of the jet.

Keywords: Graph Neural Networks (GNNs); Emerging Jets; Jet Tagging

Dedication

To my wonderful parents, with love and gratitude.

Acknowledgements

None of the work presented in this thesis would have been possible without the help and support of the incredible community around me which I am extremely grateful for.

Firstly, I owe a deep sense of gratitude to my parents. Words can hardly capture the extent of their influence over the past few years. Their unwavering faith in me has been a constant source of strength, empowering me to face and overcome diverse challenges.

I must express my sincere appreciation to Dr. Matthias Danninger, my supervisor, for the opportunity and trust he placed in me to pursue graduate research in experimental particle physics. His consistent support, encouragement, valuable guidance, energetic approach, and steadfast assistance have been pivotal throughout my academic journey. Also, a special thanks to Jackson Burzynski, my academic role model, for patiently and kindly addressing my endless stream of questions. His thoughtful answers have played a crucial role in transforming my curiosity into a passion for learning. My gratitude also extends to members of the Department of Physics at Simon Fraser University, especially those in the experimental high-energy physics group, whose friendship and support have made this journey enjoyable. I am also grateful to my committee members Dr. Bernd Stelzer and Dr. Dugan O'Neil, for their valuable feedback, constructive criticisms, and encouragement which significantly contributed to enhancing the quality of this thesis.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The mystery of human existence lies not in just staying alive, but in finding something to live for.

– Fyodor Dostoevsky

1.1 Overview

The pursuit of understanding the physical world and uncovering its secrets is a fundamental human endeavor. Particle physics, focusing on the study of the fundamental blocks of matter known as elementary particles and their interactions, is driven by our innate curiosity about the universe's basic principles.

The Standard Model (SM) is the theoretical framework that provides the best understanding to date of how matter is formed of quarks and leptons and interacts via the strong, electromagnetic, and weak forces. Despite the SM's remarkable success, it is not a complete theory of the universe as it still fails to explain several crucial questions about the universe. Countless theories beyond the SM (BSM) have been proposed, but none have been experimentally verified to date, despite a tremendous effort made by experimentalists in the last decades.

The work presented in this thesis contributes to this ongoing effort, focusing on the investigation of the unexplored areas in the lifetime frontier at the Large Hadron Collider (LHC), which studies remnants of proton-proton collisions. Among many BSM theories, the work of this thesis is based on the extension of SM with a hypothetical collection of quantum fields and their corresponding particles called the "dark sector". While most

searches for BSM physics have predominantly concentrated on particles that decay near the primary collision point of the particle detector, the work on this thesis relates to the search for particles that traverse considerable distances before decaying, commonly referred to as long-lived particles (LLPs). One potential signature of this extended SM is the *Emerging Jets* (EJs) - whose identification via the use of GNNs is the main work presented in this thesis.

1.2 Thesis structure

The thesis starts with a brief introduction to the SM and its shortcomings. It then provides a discussion of BSM physics, introducing the idea of EJs.

The next section provides an overview of the ATLAS experiment at the LHC and briefly outlines the sub-detector components within the ATLAS detector. This is followed by a discussion of the current search efforts for emerging jet signatures using the ATLAS detector.

Transitioning to the Machine Learning domain, the introductory chapters present the concepts behind GNNs. Through these, the fundamental notions of functions on graphs, and the common architecture of GNNs are presented. Their application to high-energy experimental particle physics is also discussed. In particular, the discussion of flavor tagging - a significant challenge in the identification of a distinct quark flavor using GNNs is presented. The subsequent leap to EJs identification by re-purposing the ideas of flavor tagging using GNN is highlighted.

The performance evaluation of GNNs forms a pivotal part of this work and is discussed in the last chapter. GNN can classify the jets into distinct categories, classifying individual particles within these jets, and enabling the identification of whether these particle trajectories originate from a common decay point (vertex). The evaluation of this classification task is presented by studying several performance metrics. Furthermore, the robustness of GNNs is demonstrated through model validation with the ATLAS dataset and its application in EJs Run 3 analysis - i.e. analysis on the third data-taking period of LHC from 2022-2025.

Chapter 2

Physics Theory Overview

The human soul needs beauty more than bread.

– D.H Lawrence

This chapter presents an overview of the Standard Model (SM) of particle physics, which is a theoretical framework that describes the fundamental particles and their interactions in the universe. Limitations of the SM - especially in the context of dark matter are briefly discussed which leads to the concept of extending the SM with dark sectors. One possible phenomenological consequence of this extension is the production of long-lived particles whose decay gives rise to emerging jet(EJ) signatures.

2.1 Standard Model

The fundamental particles and interactions are described by a mathematical framework of Quantum Field Theory (QFT) and unifies quantum mechanics and special relativity. In the SM formalism, particles are described by a local field $\psi(x)$ that depends upon space and time coordinates. The dynamics of this field are governed by the SM Lagrangian from which the phenomenology is derived. This Lagrangian also encodes symmetries observed in nature. It is Lorentz invariant and also invariant under the transformation of the "SM gauge group" i.e. $SU(3)_c \times SU(2)_L \times U(1)_Y$. Likewise, this Lagrangian is built to use the fewest terms to describe the observed particles and interactions. It is also built to be renormalizable - divergence in the calculation can be "regularized" to yield finite predictions for the physical observable. The particles in the SM are first divided into fermions (odd half-integer spins)

and bosons (integer spin). Fermions, the basic building blocks of matter in the SM of particle physics, are divided into two primary categories: leptons and quarks. Both leptons and quarks are further divided into three generations, each progressively heavier than the previous.

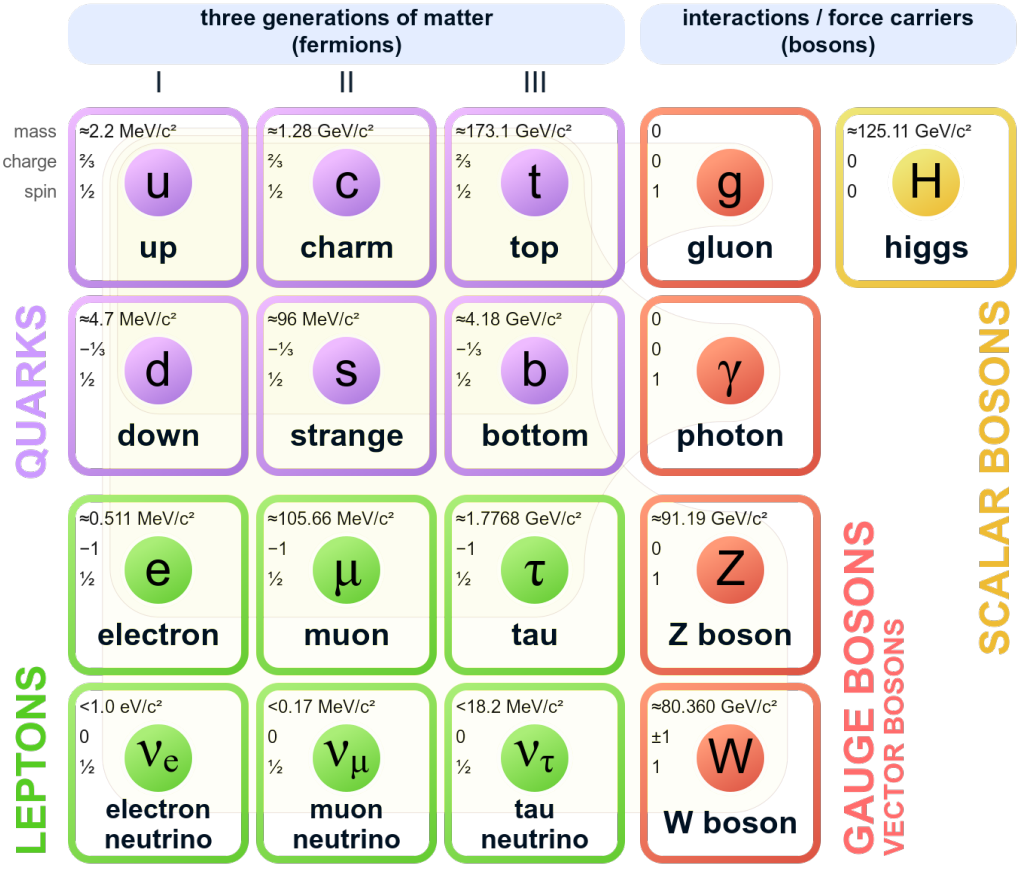


Figure 2.1: SM particles classification. Particles are classified into fermions and bosons based on the spin. Fermions are further grouped into three mass-based generations which are divided into quarks and leptons. Interaction of the elementary particle with the Higgs boson gives them mass while other bosons mediate fundamental interactions. Image from Ref. [1].

Quarks are fundamental constituents of hadrons, which include particles like protons and neutrons. They engage in strong interactions due to their unique property known as color charge. Quarks come in six different types or "flavors". As shown in Figure 2.1, the first generation is composed of the up (u) and down (d) quarks, which are the lightest and form the building blocks of protons and neutrons. The second generation features the charm

(c) and strange (s) quarks, while the third generation contains the heaviest quarks: top (t) and bottom (b). Each type of quark possesses a distinct electric charge. The u, c and t quarks each have a charge of $+\frac{2}{3}$, while the d, s, and b quarks each carry a charge of $-\frac{1}{3}$. Additionally, each quark carries a "color charge", labeled as red, blue, or green. However, these labels are abstract and don't relate to the actual colors we see. The principle of color confinement in quantum chromodynamics (QCD) ensures that observable particles must be color-neutral, implying quarks combine in ways that their color charges neutralize each other, resulting in colorless particles. Also, for every quark, there exists a corresponding anti-quark. These anti-quarks have opposite electric charges to their quark counterparts but carry the same absolute value of the charge. The color charge of an anti-quark is the anti-color of its corresponding quark, ensuring that when quarks and anti-quarks combine, they form color-neutral particles.

Leptons have electric charges, and they interact through electromagnetic and weak forces. There are six distinct types of leptons and are also divided into three generations. In the first generation, there is the electron e^- and its associated neutrino, the electron neutrino ν_e. The electron has a negative electrical charge. The electron neutrino, on the other hand, is electrically neutral. The second generation consists of the muon μ^- and the muon neutrino ν_μ. Muons are heavier counterparts to electrons, sharing the same electric charge but with a greater mass. Like the electron neutrino, the muon neutrino is electrically neutral. The third generation has the tau τ^- and the tau neutrino ν_τ. The tau is the heaviest among the electron and muon, possessing the same negative electric charge but with a much larger mass. The tau neutrino is electrically neutral. Neutrinos only interact via the weak force and are massless in the SM but experimental evidence has shown that they are massive. Also, for each lepton, there exists a corresponding anti-lepton. Anti-leptons have electric charges opposite to their lepton counterparts. For instance, the positron e^+ is the anti-particle of an electron.

Bosons are particles with integer spin values, and act as force carriers that mediate interactions between other particles. In total, there are 12 distinct bosons in the SM: 1 photon, 2 W bosons, 1 Z boson, 8 gluons, and 1 Higgs boson. Gluons are responsible

for the strong force, which binds quarks together within hadrons. Gluons themselves are unique because they also carry the 'color charge' and can thus interact with other gluons. The photon (γ) is massless and mediates the electromagnetic interactions between charged particles. Photons play a crucial role in many physical phenomena, from reflection and refraction to the behavior of electromagnetic waves. The $W$and Z bosons are responsible for mediating the weak nuclear force, which governs processes like beta decay in atomic nuclei. The W bosons are charged, with types W^+ and W^-, whereas the Z^0 boson is neutral. These bosons possess significant mass; the W bosons are approximately 80 times the proton's mass, and the Z^0 boson is slightly heavier. The Higgs boson is a scalar, neutral boson whose associated field gives mass to elementary particles. When particles move through the Higgs field, they acquire mass in proportion to their interaction strength with the Higgs field. In 2012, at the CERN's LHC, the Higgs boson was discovered [2, 3].

As mentioned, these particles in QFT are described by local field $\psi(x,t)$. The SM Lagrangian governing the evolution of the field can be broken down as follows.

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}$$

The $\mathcal{L}_{\text{QCD}}$ is the Lagrangian for Quantum Chromodynamics (QCD), which describes the strong interactions between quarks and gluons. The $SU(3)$ gauge symmetry corresponds to the strong interactions. There are eight generators of the $SU(3)$ group which are 3×3 complex matrices. This leads to quarks having 3-color charges where the matrix representation corresponds to 8 gluons. As the $SU(3)$ is non-abelian, gluons can interact with themselves. The electroweak Lagrangian describes both the electromagnetic and weak interactions, which are unified into a single theoretical framework under the gauge group $SU(2)_L \times U(1)_Y$. The Higgs Lagrangian describes the Higgs field's dynamics and potential, which is responsible for the spontaneous breaking of the electroweak symmetry. The Yukawa Lagrangian captures the interaction between the Higgs field and the fermions. After the spontaneous symmetry breaking by the Higgs mechanism, it gives rise to the masses of the fermions.

Proton-proton collisions

The research in this thesis is based on the proton-proton collisions at the LHC. Protons are bound states of two up quarks and one down quark called valance quarks. These quarks interact with each other via the exchange of gluons. The proton also contains the sea of virtual quark and anti-quark pairs produced from the gluon. Collectively these constituents of the proton are called partons. Therefore, in a 13.6 TeV proton-proton collision at the Large Hadron Collider (LHC), the interaction occurs between partons.

To predict the rate of a process occurring in the LHC after the proton-proton collisions – such as the production of a yet-unknown particle, it is important to understand the dynamic structure of the proton. The parton distribution function (PDF), denoted as $f_a(x, \mu^2)$, describes the probability of finding a parton of type a carrying a fraction x of the proton's overall momentum, evaluated at the energy scale μ^2. The PDF cannot be calculated from first principles but can be obtained from experimental measurements [4]. Finding PDFs will aid in determining the cross-sections for high-energy processes from the proton-proton collisions at the LHC.

At the sufficiently high energy scale, μ^2, the proton-proton collisions are inelastic and interaction takes place between the partons. This scattering process in high-energy colliders such as the LHC can be categorized into two types - hard or soft. The majority of the proton-proton collisions are soft i.e. they produce potentially new particles with low transverse momentum. Hard scattering refers to interactions between partons that occur at short distances and involve large momentum transfers. Most physics analysis involves the study of the hard scattering process. One of the outcomes of the hard scattering process is the production of "jets", described in the next section. The complete picture of collision phenomena is more complex as it involves several additional QED and QCD processes.

Hadronic Jet Production

A collimated stream or **jet** of particles is produced in three steps: hard scattering process, parton shower, and hadronization. After the initial hard scattering, the outgoing quarks and gluons have a lot of energy. As they move away from the interaction point, they ra-

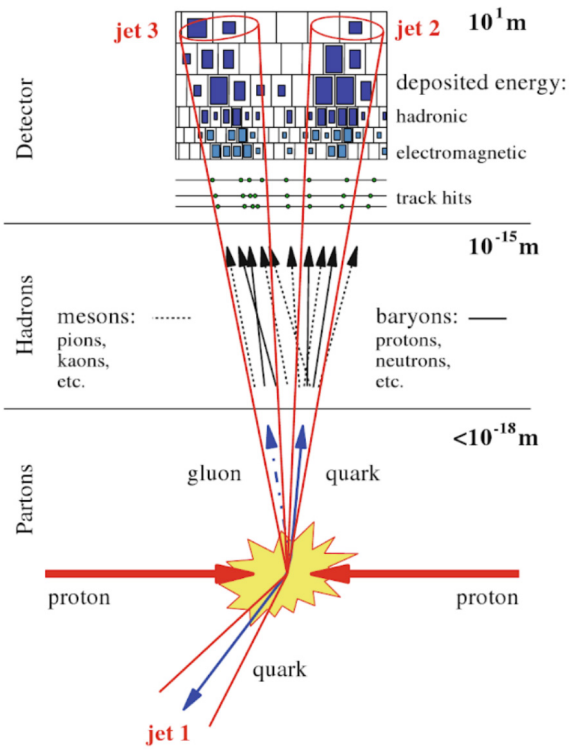


Figure 2.2: Jet production at the detectors such as ATLAS shown in Parton, Hadrons, and detector scale. Image taken from Ref. [5].

diate off additional gluons due to the strong force. These gluons can, in turn, produce quark-antiquark pairs. This cascade of particles is called a "parton shower". However, the particles in the shower cannot exist individually and start forming color-neutral hadrons by creating quark-antiquark pairs from the vacuum to combine with - a phenomenon known as hadronization. As the perturbation series expansion of QCD breaks down at low energy, direct calculations are challenging, leading to the development of models like the Lund String Model to explain hadronization.

The basic idea behind the Lund String Model is to visualize the force between a quark and an anti-quark (or between a quark and a gluon, or between two gluons) as a "string" of a gluonic field. As the quarks move apart after a high-energy collision, this string stretches between them. The energy of the string increases as the quarks move further apart due to the strong force. When the energy in a portion of the string becomes sufficiently large, comparable to the mass of a quark-antiquark pair, the string can "break", and a new quark-

antiquark pair is created from the energy of the string. This is in line with the principle of energy-mass equivalence from relativity. The original quarks and the newly produced quark-antiquark pairs combine to form hadrons. For instance, an original quark might combine with one of the newly created antiquarks to form a meson. This process continues, leading to the production of multiple hadrons until the energy of the string is used up. This cascade of particles is a jet as shown in Figure 2.2

The Lund String Model, along with its detailed dynamics, forms the basis of the PYTHIA [6] event generator, a widely used tool in particle physics to simulate high-energy collisions and the subsequent jet production as shown in Figure 2.2.

2.1.1 Shortcomings of Standard Model

While the SM has been remarkably successful in describing the fundamental particles and forces that govern the universe, it has major limitations. First, it does not explain fundamental phenomena observed in nature such as the nature of dark matter, the matter-antimatter asymmetry observed in the universe, gravity, and neutrino masses, among others. Also, the SM has theoretical inconsistencies such as hierarchy problems, referring to the large discrepancy between the scales of electroweak force and gravity; and Landau pole [7] problems, which are characterized by the momentum (or energy) scale at which the coupling constant (interaction strength) of a quantum field theory becomes infinite, among others.

Dark Matter(DM)

Several astrophysical observations indicate the presence of a form of non-baryonic matter that the SM does not explain. The first of these observations was provided in the early 1930s by rotational curves of spiral galaxies that could not be explained by the luminous mass alone. If only the luminous mass were all matter, it is expected that the rotation velocity will decrease as one moves further from the galactic center, as described by

$$v^2 = \frac{G \times M}{r},$$

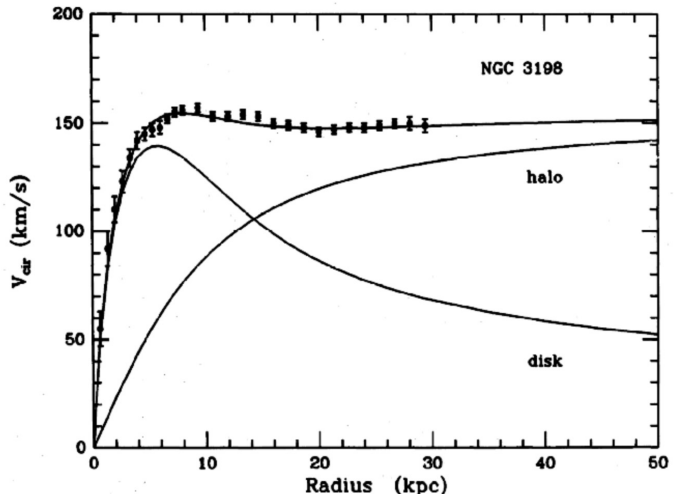


Figure 2.3: Illustration of the rotation curve of the spiral galaxy NGC 3198. The bottom curve represents the predicted motion, while the top curve shows what is observed. The difference between these two curves is attributed to dark matter (DM). Image taken from Ref. [8].

for circular orbits where M represents the mass within a radius r and G represents gravitational constant. For spiral galaxies such as NGC 3198, the predicted rotational curve is shown in Figure 2.3 labeled as "disk". Yet, actual observations of these galaxies revealed that the rotational speeds remain largely constant, even as their distance from the center increases, contradicting the expected behavior based on visible matter alone. The line marked “halo” in Figure 2.3 is the extra matter needed to reproduce the observed rotation curve, attributed to DM - a non-baryonic matter that does not interact much with matter other than gravity.

Likewise, Einstein’s general relativity theory states that light can be bent or "lensed" by mass. Using this lensing effect, we can detect the presence of mass, even if it doesn’t produce any light. Observations using gravitational lensing have deduced that the universe contains about five times more DM than baryonic matter [9]. Beyond evidence from rotation curves and gravity lensing, there is a plethora of observational evidence such as the existence of hot gas in clusters, and the bullet clusters among others [10] that provide the evidence for the existence of dark matter.

2.2 Beyond Standard Model (BSM)

The fact that the SM does not contain DM candidates is compelling evidence to suggest that there is a more universal theory beyond the SM. BSM theories propose new particles, forces, and symmetries that can address the shortcomings of SM, including the nature and existence of DM, the unification of fundamental forces, and the asymmetry between matter and antimatter among others.

Examples of BSM theories encompass a diverse range of approaches. Top-down theories, such as supersymmetry, propose high-level, comprehensive frameworks that seek to provide a unified explanation for fundamental forces and particles. Supersymmetry, for example, states that there is an existence of symmetry between fermions and bosons. If the theory is experimentally verified, R-parity (also called matter parity) conserving supersymmetry could help explain phenomena, such as the nature of DM. However, this symmetry has not been observed in nature.

In contrast, bottom-up theories, like the hidden-valley models [11], take a more incremental approach by extending the SM to account for observed phenomena that cannot be explained within the existing SM framework. These approaches are phenomena driven to build simplified models that give unique signatures in the particle collider detectors instead of building UV-complete models of physics that remain consistent and free from physical or mathematical inconsistencies at all energy scales, including extremely high energies approaching or reaching the ultraviolet (UV) cutoff. Also, considering these simplified models facilitates signature-driven searches for new physics which also remain sensitive to a wide range of top-down BSM scenarios. These BSM theories collectively aim to address fundamental questions about the universe's composition, forces, and unexplained phenomena.

2.2.1 Extending the SM with a Dark Sector

The simplest of theoretical approaches to explain the existence of DM, without modifying the SM is by extending it to require the existence of a "dark sector". The dark sector (also called hidden sectors) is the hypothetical collection of quantum fields and their corresponding unobserved particles [12].

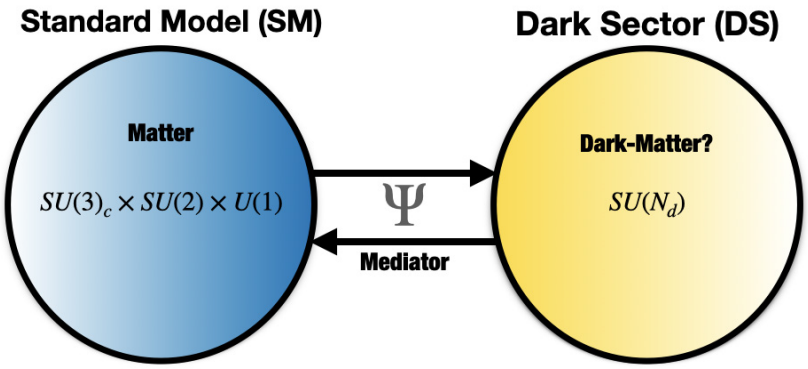


Figure 2.4: Illustration of extension of SM with DS, where mediator Ψ facilitates interaction between them.

In these models, the dark sector is connected to the visible sector (SM physics) via mediators Ψ (as shown in Figure 2.4), termed portals which allow for processes involving new particles to be observed at colliders such as the LHC. The nature of interactions between the two sectors is influenced by several factors, including the mediator's spin and parity, among other characteristics. There are four main types of portals considered generally in the literature. These include vector portals [13] (vector), Higgs portals [14] (scalar), neutrino portals [15] (fermionic) and axion portals [16] (pseudo-scalar).

One such extension to the SM gauge group, relevant to this thesis, can be more concretely expressed as:

$$G_{SM} \times SU(N_d)$$

where $G_{SM} = SU(3)_c \times SU(2) \times U(1)$ and $N_d \geq 2$ is the number of dark colors [17]. In this context, it is relevant to consider a neutral vector mediator Z' [18] which couples with both SM quark and dark quark pairs. Using this as a starting point, particle phenomenology is explored for unique experimental signatures.

The phenomenology of the model considered above is governed by the following parameters: the Z' mass, it's coupling to visible and dark quarks g_q and g_{qD}, the number of dark colors, dark flavors, the scale of the dark sector confinement Λ_d and the mass scales of the dark hadrons m_D [18].

Another important point to note is that, as the portal couples the dark sector to SM quarks, observable of interest for studying these models will be showers of quarks and gluons called jets. Depending upon the phenomenological consideration of parameters described above, these models can lead to non-conventional jet signatures such as semi-visible jets and **emerging jets** in the detectors of hadron colliders [19, 17]. Semi-visible jets are the spray of stable invisible dark pions along with unstable states that decay back to the SM. Being able to identify the emerging jet (described in Section 2.4) signatures is of special interest in this thesis as they are BSM long-lived particle decay signatures. A description of long-lived particles is included below.

2.3 Long Lived Particles (LLPs)

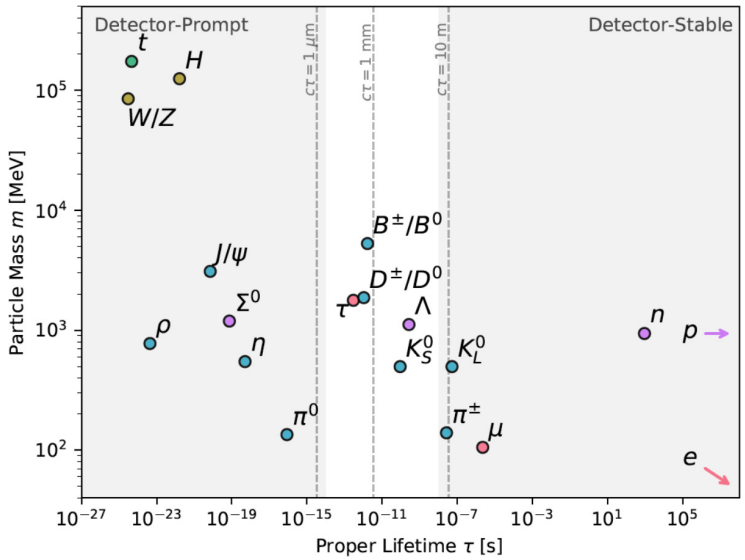


Figure 2.5: The masses and proper lifetimes of some particles in the SM. The shaded regions on the left represent regions where the particles are considered prompt. The shaded region on the right represents LLPs but is detector-stable. The un-shaded region in the middle contains particles that have a significant probability of decaying within the ATLAS detector. Image from Ref. [20].

As there have been no strong indications of new physics at the collider experiments, this would indicate two possibilities. Either the new physics is above the energy scale accessible to LHC - the most powerful particle collider, - or we have been looking at the "wrong places".

At the beginning of the LHC era, the general approach in the hunt for new physics was to search for the simplest scenarios first. But, with years rolling by and no tangible signs of new physics, physicists have begun their search attempts in significantly challenging but unexplored phase spaces, such as **long-lived particle**(LLP) signatures.

LLPs are particles that travel an observable distance from the primary $p-p$ collision point in the particle detectors before decaying and will have macroscopic proper lifetimes - the amount of time for which they remain stable when observed from their reference frame. Most BSM physics searches have been performed with the assumption that the particles decay promptly near the primary interaction point. Nevertheless, within the Standard Model (SM), the lifetimes of particles vary widely across several orders of magnitude, as shown in Figure 2.5. A proper lifetime of a particle is inversely proportional to the decay width Γ,

$$\Gamma = \frac{1}{2m_\chi} \int d\Pi_f \left|\mathcal{M}\left(m_\chi \to \{p_f\}\right)\right|^2 .$$

Here, $\mathcal{M}$ represents the matrix element corresponding to the decay of the particle into its final state products p_f, while m_χ denotes the particle's mass, and $d\Pi_f$ represents the decay's phase space. The matrix element encapsulates information about the coupling strength among others. Based on this, a particle can have a long lifetime due to limited available phase space, smaller couplings to its decay products, or when the coupling constant is suppressed if the decay is mediated through heavy particles.

Figure 2.5 shows the proper lifetime of SM particles against their mass and divides the plot into three distinct regions based on their proper lifetime. The left-most shaded region represents particles whose lifetimes are short enough that the distance between their production and decay is less than the detector's spatial resolution. Such particles are termed as having a "prompt" decay within the detector. The right-most shaded section of Figure 2.5 includes particles whose lifetimes enable them to traverse beyond the ATLAS detector before undergoing decay, and these are labeled as detector-stable. The central, unshaded region represents particles with lifetimes that have a high chance to decay within the detector but are long-lived enough to set themselves apart from the promptly decaying ones.

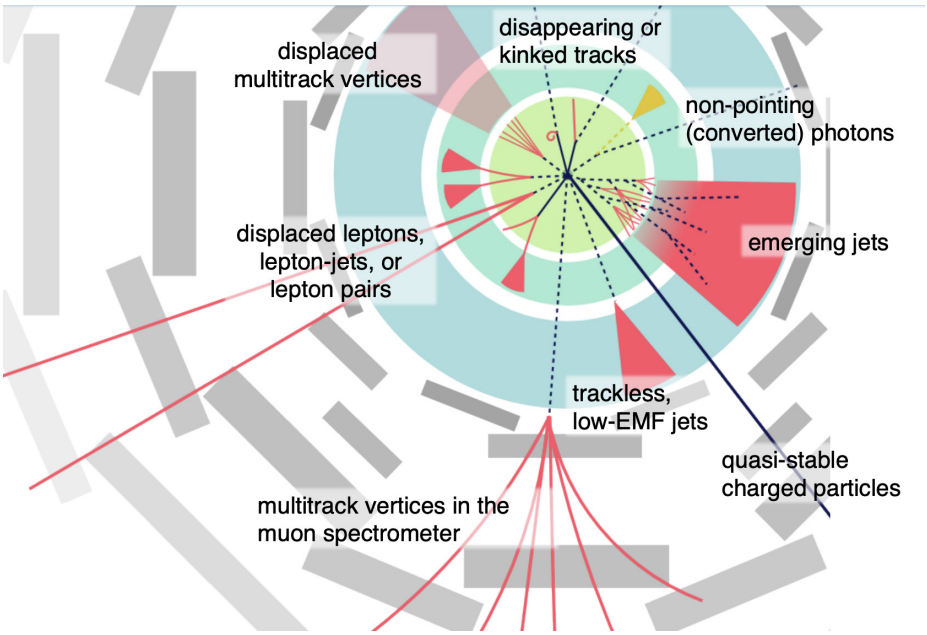


Figure 2.6: Complex experimental signatures that can arise from BSM LLPs within LHC detectors. Displayed is a cross-sectional view in the azimuthal angle,ϕ, typical of general-purpose detectors like ATLAS or CMS with different possibilities of signatures from BSM LLP decays. Towards the right is the emerging-jet signature comprising many particle trajectories that are displaced from the primary interaction point. This signature is of special significance to this thesis. Image from Ref. [21].

Thus, there is no reason to exclude the possibility of LLP signatures in BSM physics. These also exist in SM! Also, LLPs are motivated by top-down theories discussed in Section 2.2 and even from bottom-up consideration which will be elaborated in the following section. The detectors at the LHC are currently used to search for various long-lived signatures, as illustrated in Figure 2.6. These include disappearing tracks, multiple displaced jet scenarios, and notably, **emerging-jets** [22].

2.4 Emerging Jets

EJs are jets with many *displaced tracks* and *displaced vertices*, whose phenomenology is described below based on the description in Ref. [17].

As we learned in the Section 2.2.1, the phenomenology of an extended SM with dark sector (DS) extension $G_{SM} \times SU(N_d)$, with a neutral vector mediator(Z') acting as a "portal" between the SM and the DS physics, can lead to a range of un-conventional jet signatures

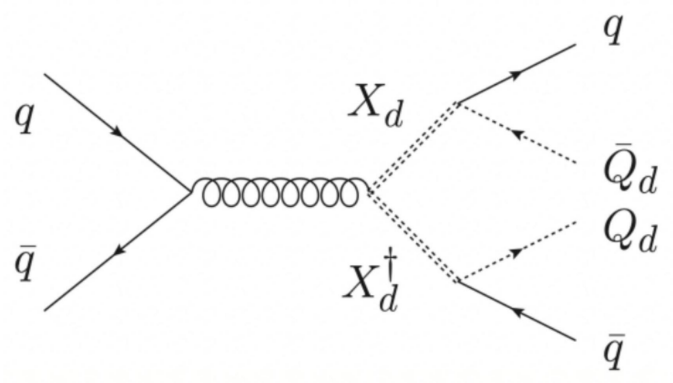


Figure 2.7: Feynman diagrams for EJs production mediated by $Z^{'}$ - hypothesized heavy vector mediator. The jet (cone) is an emerging jet.

depending upon parameters selected. A well-motivated parametric consideration is detailed which then leads to the emerging jet phenomenology as explained below.

The lightest baryon of the hidden sector can be a good DM candidate as it is stable analogous to the proton. As gravity lensing observations have provided evidence that the universe contains five times more DM than baryonic matter (Section 2.1.1), the lightest baryon of the hidden sector π_d, should be in the order of $5 \times m_{\text{proton}}$ to address the asymmetry. Also, Λ_d is the approximate mass of most of the dark mesons and baryons. Therefore, DM confinement scale Λ_d is considered to be $\mathcal{O}(1-10)$ GeV.

In the composite DM model under discussion, which mirrors QCD-like behavior, the annihilation of dark baryons with their anti-baryon counterparts into dark pions is highly effective. These dark pions can subsequently decay back into particles of the SM. These decays, however, should not interfere with the Big Bang Nucleosynthesis (BBN). BBN is a critical period in the early universe, a few seconds after the Big Bang, during which light atomic nuclei (like hydrogen and helium) were formed. Any new physics or particles (like dark pions) should not challenge the BBN as it is a well-established theory. As a result, the "lifetime" of dark pions is constrained to be less than one second. This implies an upper bound on the mediator Z' mass, approximately around 100 TeV. Nonetheless, mediators with a lower mass were more likely to lead to dark QCD confinement closer to the QCD scale. This led to the selection of TeV scale mediators as the model's parameter.

Also, Ref. [17] shows that it is well motivated for the dark pions decay of the GeV scale with TeV scale mediators to traverse centimeters to meter lengths before decaying back to SM particles. Thus, this results in BSM LLP signatures.

These parametric considerations lead to a very interesting phenomenology as shown in Figure 2.7. First, the production of the Z' boson of the TeV scale that decays to two dark quarks is considered. The energy of these dark quarks will be much larger than the dark confinement scale Λ_d, mentioned above. Consequently, these dark quarks will undergo showering and hadronization, resulting in the production of many dark mesons, represented by the dotted lines in Figure 2.7. Based on the previously noted decay lengths for dark pions, which range from centimeters to meters, these dark jets traverse a distance of centimeters as an LLP before they decay to the visible sector. The probability that a particle survives for time t before decaying, however, is an exponential distribution and is governed both by proper lifetime of a particle, τ, and the Lorentz factor γ related by

$$P(t) = e^{\frac{-t}{(\gamma\tau)}}$$

[23]

Thus, each dark pion decays uniquely at different distances, as illustrated by the origin points of the red lines in Figure 2.7. This gives rise to the distinct "emerging jets" signature. In detectors such as ATLAS, this signature could be seen as sprays of many *displaced tracks* - particle trajectories that do not originate from the main proton-proton interaction point, and many *displaced vertices* - the common origin of tracks that are away from the primary proton-proton interaction point.

Chapter 3

Physics Experiment Introduction

The real voyage of discovery consists not in seeking new landscapes, but in having new eyes.

– Marcel Prosut

3.1 The ATLAS Detector

ATLAS (A Toroidal Large Hadron Collider Apparatus) is a cylindrical multi-purpose particle detector at the LHC. It is 46 meters long, 25 m high, and 25 m wide, weighs 7000 tonnes, and is one of the largest particle detectors ever built. When protons in the LHC are accelerated almost at the speed of light and collide at the center of the ATLAS detector, they generate collision remnants, resulting in the creation of new particles. These newly formed particles disperse in all directions from the collision point. The ATLAS detector is composed of three subsystems known as subdetectors, each designed to specialize in identifying particular types of particles generated during proton-proton (pp) collisions. Also, the detector has a huge system of magnets that bend the trajectory of charged particles to determine the charge and momentum of the particles. Two distinct superconducting magnet systems are used: solenoidal and toroidal. These systems are cooled to approximately 4.5 K to produce the powerful magnetic fields in the detector.

The Co-ordinate System

In the ATLAS detector, a right-handed coordinate system is used, centered where particle collisions occur. The $\hat{z}$ direction aligns with the beam-line, while $\hat{y}$ points upward towards

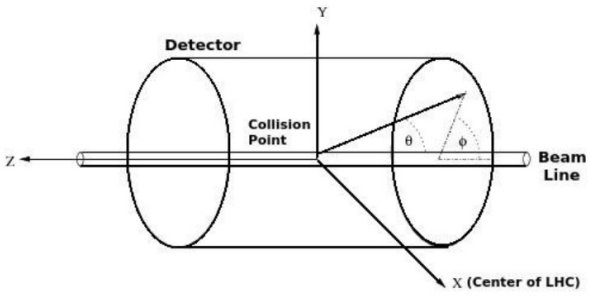


Figure 3.1: Coordinate system of the ATLAS detector.

the ground's surface and the $\hat{x}$ direction is oriented toward the LHC's center, as shown in Figure 3.1. The plane formed by the x and y axes is referred to as the "transverse plane." Angles in this system are defined with the azimuthal angle ϕ measured around the beam direction and the polar angle θ taken from the beam axis itself. Momentum can be divided into a component in the transverse plane with the notation p_T and in the longitudinal plane p_z. Useful kinematic variables can be defined to be invariant under boosts along the longitudinal axis. For example, rapidity is defined as

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}$$

where E is the energy of the incoming particle. However, measuring p_z in high-energy collisions can be challenging because the momentum observed along the direction of the beam might simply be residual from the original beam particles. To address this, a valid estimate of rapidity in the $\beta \to 1$ limit termed pseudorapidity is defined as:

$$\eta = -\ln \tan((\frac{\theta}{2}))$$

3.1.1 Inner Detector and Tracking

The Inner Detector (ID) of the ATLAS experiment is the innermost subsystem and is responsible for tracking charged particles emerging from proton-proton collisions. The ID is designed to provide high-resolution measurements of the trajectories of charged particles over a wide range of pseudorapidity, ensuring efficient track reconstruction. The ID operates within a 2 Tesla magnetic field provided by a superconducting solenoid. The ID consists of

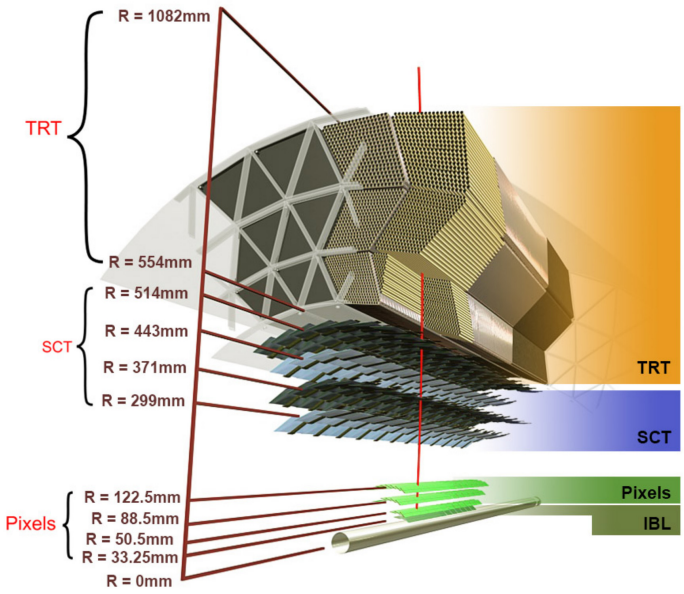


Figure 3.2: Cutaway view of the ATLAS Inner Detector showcasing its intricate layers. From the innermost region, the four-pixel layers can be seen followed by 5 layers of the silicon strip. across a larger volume. The outermost layer is the TRT stacked with a tube of 300,000 thin-walled drift tubes. Image taken from Ref. [24].

three major components: the Pixel Detector, the Semiconductor Tracker, and the Transition Radiation Tracker as shown in Figure 3.2. These detectors are stacked in layers also shown in Figure 3.2.

Pixel Detector

The Pixel Detector is the innermost component of the ID and is located closest to the beamline. It comprises four cylindrical layers around the beam pipe and three layers of disks in the forward regions. Each pixel acts as an individual microdetector. When a charged particle passes through, it produces a charge in the silicon. This charge is then collected, producing a signal that gives a precise position measurement. Due to its proximity to the collision point, the Pixel Detector faces a high particle flux and thus requires a fine granularity to distinguish tracks from nearby particles.

Semiconductor Tracker (SCT)

Located outside the Pixel Detector, the SCT is made up of a silicon microstrip. The SCT has four layers of cylindrical barrels around the beam pipe and nine disks on each end. Similar

to the Pixel Detector, charged particles passing through the silicon strips generate a charge, which is collected and gives a positional measurement. While it doesn't have the same spatial resolution as the Pixel Detector, the SCT significantly extends the tracking coverage in both the radial and longitudinal directions. The decision to use a Silicon Tracker (SCT) instead of extending the Pixel Detector throughout the Inner Detector in the ATLAS experiment arises from a balance of factors including resolution, coverage, cost, and technological constraints. The SCT provides additional precision track points, aiding in momentum resolution and track reconstruction.

Transition Radiation Tracker (TRT)

The outermost component of the ID, the TRT, is different in design compared to the Pixel and SCT. It consists of many straw tubes filled with a gas mixture. As charged particles pass through these tubes, they ionize the gas, creating a detectable signal. The TRT is named for its ability to detect transition radiation: as high-energy particles move from one medium to another, they emit X-rays. This is particularly useful for distinguishing between electrons and pions. The long straws of the TRT provide many position measurements along a particle's path, improving the tracking efficiency and momentum resolution.

Tracking

A track is a physics object reconstructed from the signals of the inner detector and represents the trajectory of charged particles as they move through the ID. Tracks are described by five sets of parameters and a reference frame using the perigee representation. The prime objective of the track reconstruction is to then compute the value of these parameters. The description of the track parameters is listed in Table 3.1

The track reconstruction algorithm in ATLAS typically operates through a two-step process. Initially, an 'inside-out' tracking is executed, as depicted in Figure 3.3. Hits from the Pixel and SCT are clustered into space points. These space points are seeded into groups of three which then must meet some d_0 and p_T requirements. The combinatorial Kalman filter algorithm [26] is then utilized to generate track candidates from the accepted seeds by extending the seeds with additional space points (usually from the SCT) that

Parameter	Description
d_0	distances of closest approach between the track and beamline in the transverse plane.
z_0	distances of closest approach between the track and the collision center in the longitudinal plane
ϕ	azimuthal angle of the track's momentum in the transverse plane
θ	polar angle of the track momentum
q/p	charge of the particle divided by its momentum, indicating the curvature of the track in the magnetic field.

Table 3.1: Five key track parameters used in ATLAS tracking system.

match the trajectory of the initial seed as shown in the Figure 3.3. This generates many track candidates, all of which are scored using different metrics taking into account quality criteria such as the number of hits. Candidates with higher scores are likely to accurately represent the trajectory of a charged particle, while those with lower scores—often merely combinatorial aggregations of hits—are discarded. Following this, the track candidates are extrapolated to the TRT, as indicated by the purple dotted line in Figure 3.3. The track then undergoes a refit, incorporating all its points. This refitted track is assessed once more using analogous scoring mechanisms, and any sub-optimal tracks are discarded.

The next step is the outside-in algorithm. Here, the hits from the TRT are integrated in a reverse direction, being extrapolated back towards the silicon detectors using a back-tracking algorithm. This method identifies tracks that didn't emerge from the primary $p-p$ interaction point of the ATLAS detector. The final collection of tracks from both the inside-out and outside-in algorithms is called *standard* tracks.

However, standard tracking faces a significant hurdle. The efficiency of track reconstruction is greatly reduced for tracks with impact parameter $d_0 > 5$ mm. Dark pions, considered in the Section 2.4, that lead to emerging jet phenomenology, have long lifetimes before they decay into SM particles. Therefore the particle tracks from these decays may have large impact parameters ($d_0 > 5$ mm). To improve the reconstruction efficiency for these kinds of signatures, a *Large Radius Tracking* [27] algorithm is used. This algorithm is run after the standard tracking using the detector hits not already assigned as standard tracks. It

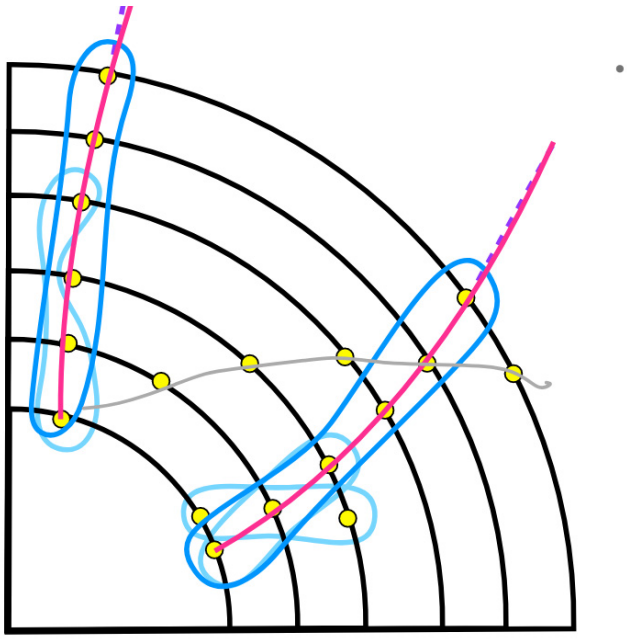
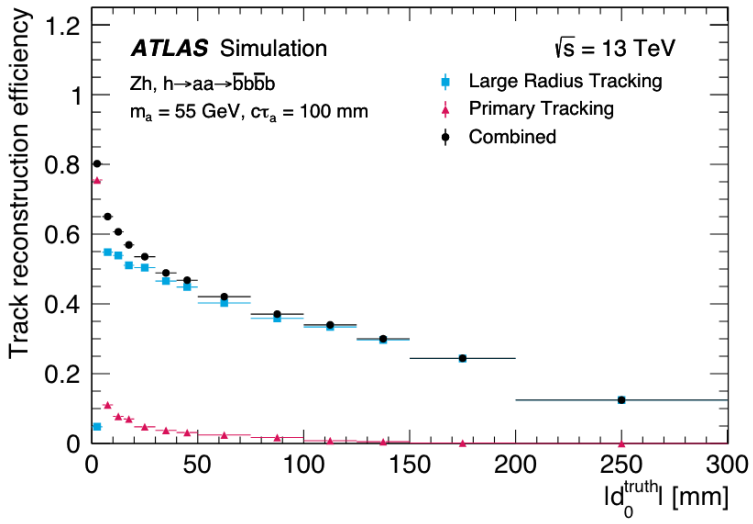


Figure 3.3: Illustration of inside-out track reconstruction using pixel hits. First seeds are found using 3-hit groups from the Pixel Detector (light blue). The seeds then are extended using a combinatorial Kalman filter to strips (dark blue). The track candidates are evaluated against different metrics and sub-optimal ones are discarded. After this, the seeds are extended to TRT. Finally, the track is re-fitted with all points (pink) and is evaluated for its quality to get the best track candidates. Image from Ref. [25].

follows the same inside-out procedure as explained above but is specifically optimized for LLP signatures. The performance of the algorithm is evaluated on the simulated sample with $H \to aa \to \bar{b}\bar{b}\bar{b}\bar{b}$ events, i.e exotic decays of Higgs boson into pairs of long-lived pseudo scalar (a), each of which decays into bottom quark pairs as shown in Figure 3.4. Beyond $|d_0| > 5$ mm, the efficiency of the primary (standard) tracking algorithm diminishes noticeably. Yet, the LRT algorithm manages to maintain a substantial efficiency up to $|d_0| < 300$ mm [27]. Thus, the LRT algorithm plays a crucial role in the work presented in this thesis.

3.1.2 Calorimeters and Jets

The ATLAS calorimetry system is composed of two distinct subsystems and is designed to measure and absorb the energy of both charged and electrically neutral particles. The electromagnetic calorimeter (ECAL) measures the energy of photons and electrons while the hadronic calorimeter (HCAL) measures the energy of hadronic particles which do not deposit all their energy in ECAL. Both of these calorimeters are made up of alternating



(a)

Figure 3.4: The primary (standard), LRT, and combined efficiencies for reconstructing displaced tracks produced from the decay of a long-lived pseudo-scalar a associated with Higgs decay for simulated samples. The track reconstruction efficiency is shown as a function of $|d_0|$ Image from Ref. [27].

"passive" and "active layers". Passive layers absorb the particle and initiate electromagnetic or hadronic showers while active layers measure the energy of those showers. Calorimeter signals are used to reconstruct physics objects called *jets* among others.

Electromagnetic Calorimeter(ECAL)

The ECAL immediately follows the ID and is also positioned concentrically around it. It is an accordion-shaped calorimeter with alternating layers of lead and liquid argon. The electromagnetic shower is initiated by e^-, e^+, or γ after they interact with the heavy nucleus of the lead. In the context of high-energy collisions, this shower is dominated by two radiative processes: photons engage in pair production, as shown in Figure 3.5, while electrons undergo bremsstrahlung. The shower then is sampled in the active layer made up of liquid argon, where charged particles passing through create a signal via ionization.

Hadronic Calorimeter (HCAL)

The HCAL encircles ECAL and measures the energy of hadronic particles, which aren't fully absorbed by ECAL. This calorimeter comprises alternating steel and plastic scintil-

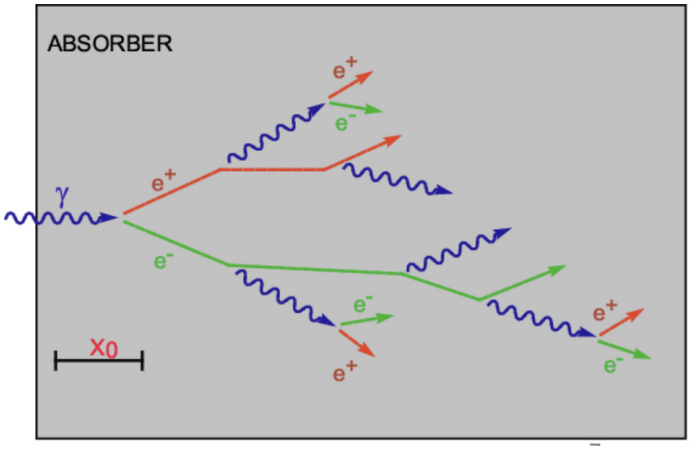


Figure 3.5: Example of an electromagnetic shower from pair production after the photon interacts with the absorbing layer. Image taken from Ref. [28].

lating tiles. When particles strike the "passive" steel layers, they produce hadronic showers. However, particles that decay electromagnetically such as η and π^0 are also produced in hadronic showers, which in turn leads to electromagnetic showers (for example, $\pi^0 \to \gamma\gamma$). An illustration of particle showers in the hadronic calorimeter is shown in Figure 3.6 where the blue-coloured particle showers are from hadronic interactions while the red-coloured particles are from electromagnetic showers. Subsequently, these showers hit the "active" plastic tiles which in turn emit photons as a response. These photons are then transformed into an electrical signal to help measure the energy of incoming particles.

Jets

Jets are physics objects constructed from clusters of energy deposits in the calorimeters. More specifically, the calorimeter cells are grouped into three-dimensional clusters (topo-clusters) using the nearest-neighbor algorithm. After additional corrections to the topo-clusters, jets are reconstructed using the anti-k_t algorithm with one configurable parameter R [29]. The anti-k_t is a sequential recombination algorithm taking in topo-cluster inputs (as pseudojets). At its core, it calculates two types of distances: the distance between two objects (or pseudo jets)

$$d_{ij} = \min\left(p_{Ti}^{-2}, p_{Tj}^{-2}\right) \frac{\Delta_{ij}^2}{R^2}$$

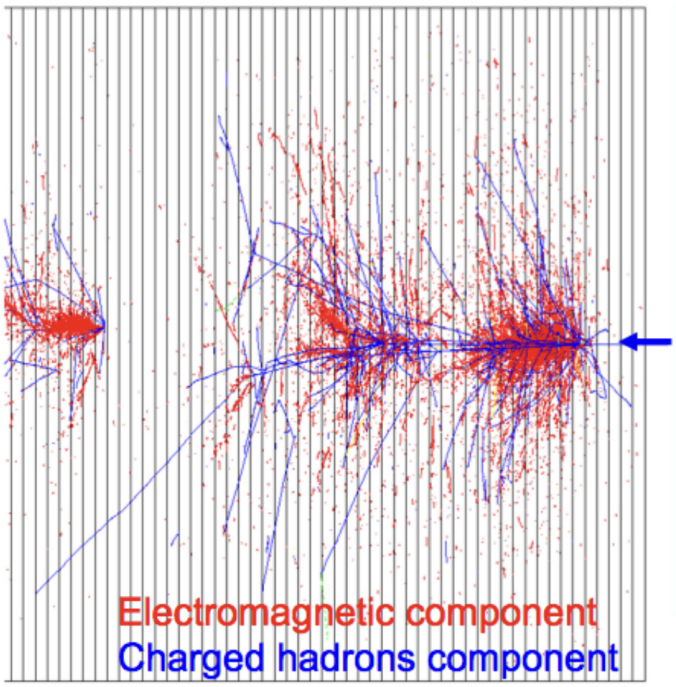


Figure 3.6: Illustration of particle shower in hadronic calorimeters. The incoming particle interacts via the strong force with the absorbing layer made up of steel to produce hadronic showers. However, particles such as π^0 are also produced in the hadronic shower which in turn leads to electromagnetic shower. Thus, particle showers in hadronic calorimeters are made up of charged hadron components and electromagnetic components. Image taken from Ref. [28].

and the distance between an object (or pseudojet) and the beam axis

$$d_{iB} = p_{Ti}^{-2}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and p_{Ti}, y_i and ϕ_i are the transverse momentum, rapidity, and azimuth of object i. The algorithm iteratively clusters pairs of objects that have the smallest distance between them in an event via the following steps.

1. Compute the distance d_{iB} for the i^{th} object, and d_{ij} between i^{th} object and all other object j from the topo-cluster.

2. Identify the smallest of these distances

3. If $d_{iB} < \min(d_{ij})$, then the i^{th} object is removed from the set of constituents and is considered as a complete jet. Otherwise, combine the objects i and j into a single pseudojet and remove the original i and j objects from the list.

4. Repeat the above steps until no objects remain.

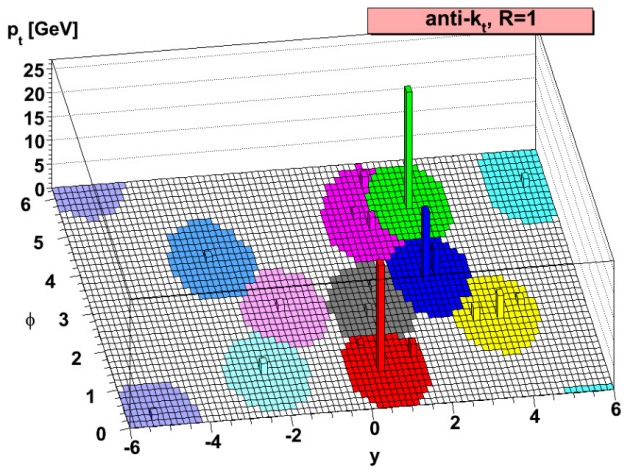


Figure 3.7: A sample event at the parton level which was clustered using the anti-kt algorithm to reconstruct jets. Image from Ref. [30].

After the jet reconstruction, the four momenta of the jets are calculated by summing the energy and momenta of topo-clusters used to create them. However, the energy of the final jets needs to be calibrated to better model the jet energy at the particle level. The calibration process for jets is designed to adjust the observed energy to compensate for factors like pileup impacts, geometrical influences, inconsistencies in the calorimeter's response, and discrepancies between real and simulated events.

The standard value of the R parameter used in the ATLAS experiment is 0.4. However, as EJs are formed after showering of dark quarks as discussed in Section 2.4, the value of the radius parameter used for the work relating to this thesis is R=1 as shown in the Figure 3.7. The reconstructed jets are then called fat jets. Being able to discriminate these fat jets having emerging jet topology is central to the work of this thesis.

3.1.3 Muon Spectrometer(MS)

The Muon Spectrometer consists of separate high-precision tracking subsystems for muons. Muons are 207 times heavier than electrons and can traverse through other sub-detectors without depositing much energy, unlike most of the other particles that are absorbed by the calorimeter systems. It is the outermost part of the ATLAS detector, encircling other components like the inner detector and the calorimeters. It is designed to function in the high magnetic fields generated by the toroidal magnets, allowing it to precisely measure the momentum of high-energy muons.

The MS utilizes a combination of different technologies. Monitored Drift Tubes (MDTs) and Cathode Strip Chambers (CSCs) are employed for precision measurement of the muon tracks. To provide fast trigger decisions, Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs) are used. To bend the paths of the muons and thus measure their momenta, the Muon Spectrometer utilizes a magnetic field generated by three large air-core toroid magnets. These magnets create a magnetic field orthogonal to the muon path, enabling a precise momentum measurement.

3.2 Trigger System

The interactions within the ATLAS detectors generate a massive amount of data. However, most *pp* collisions are soft parton-scattering events that lack significant physics phenomena. ATLAS employs a two-level "trigger" system that tells the detector which events to capture and which to disregard.

The first level of trigger (L1) operates in real-time and is hardware-based. Its primary responsibility is to rapidly assess the raw data coming directly from the detector and decide, in a matter of mere microseconds, whether the event in question is potentially interesting and warrants further analysis. It does so by examining coarse data from the calorimeters and the muon system to identify signatures of high-energy particles, such as electrons, photons, muons, and jets. Given the LHC's prolific collision rate, the L1 Trigger plays a crucial role in reducing the data rate from around 40 million events per second to a more manageable 100,000 events per second.

The data selected by the L1 trigger is further passed onto a software-based trigger system known as a high-level trigger (HLT) which performs more detailed filtering of the events. Long-lived particle events are notably difficult to trigger. The Run 3 EJs analysis uses a dedicated emerging-jets trigger which will be explained in the following section. Sophisticated data acquisition and computing systems are then utilized to analyze the recorded collision events.

3.3 Emerging Jet Searches at the ATLAS

Two distinct searches for emerging jet signatures are currently in progress in ATLAS: the Run-3 EJ analysis and the Run-2 EJ analysis. These analyses focus on different production modes, and data-taking periods, and utilize varying analytical approaches. This section briefly discusses the Run-3 EJ analysis. My work for this thesis was as a member of EJ Run-03 analysis.

3.3.1 Run-3 Emerging Jets analysis

The Run-3 EJs search targets s-channel production of dark quarks via the $Z^{'}$ mediator. This analysis is conducted using data collected during the third LHC data-taking period, beginning in 2022. The final state has two EJs as shown in Figure 2.7.

Contrasting with conventional searches conducted at ATLAS, the analysis of EJs presents distinctive challenges. The design of the ATLAS detector is primarily focused on reconstructing particles originating in proximity to the primary proton-proton (pp) interaction point. However, the recent integration of *Large Radius Tracking (LRT)*, discussed in Section 3.1.1 allows for the reconstruction of particle trajectories (with good efficiency) characterized by longer lifetimes. As the dark-pions that decay into the SM particles have a macroscopic lifetime because of phenomenological considerations mentioned in section Section 2.4, LRT plays a pivotal role in facilitating this analysis.

Another notable distinction in the emerging jet analysis, compared to standard ATLAS searches, pertains to the utilization of large-R jets reconstruction algorithm. Emerging jet signatures are distinct in that they have a large jet cone because of the dark shower and hadronization before the dark pions decay to SM jets, as shown in Figure 3.8. Thus, to

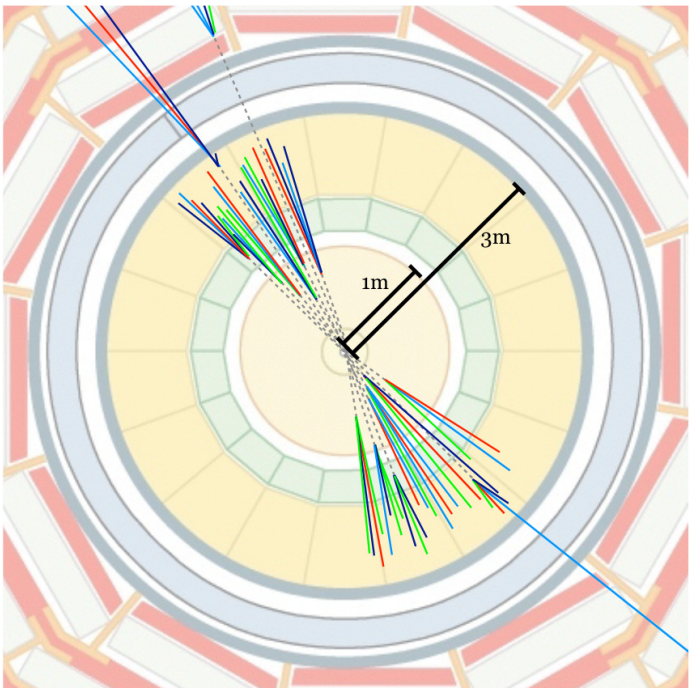


Figure 3.8: A visual representation illustrating the creation of two EJs through the pair production of dark quarks at the ATLAS detector. The diagram depicts a cross-section in the x-y plane of a detector, with the beam pipe receding into the page. The tracker and calorimeter radii are approximately indicated. The dashed lines are dark mesons that are not "visible" to the detector. As the dark pions travel a certain distance, they each decay into SM particles, forming a large cluster of displaced tracks and vertices - EJs. Image from Ref. [17].

computationally reconstruct these jets effectively, this analysis makes use of large R jets (radius parameter $R = 1$) as discussed in Section 3.1.2.

Likewise, this analysis uses a dedicated emerging-jets trigger that requires jet $p_T > 200$ GeV, $|\eta| < 1.8$, and prompt track fraction (PTF) < 0.08 where PTF is defined as:

$$PTF = \frac{\sum_{\text{trk} \in \text{jet}} p_{\text{T}}^{\text{trk}} \left(d_0 < 2.5\sigma\left(d_0\right)\right)}{p_{\text{T}}^{\text{jet}}}$$

Here, $\sigma(d_0)$ is the uncertainty in d_0.

Events selected by the trigger system can generally be contaminated by many background processes, obscuring the signal events featuring EJs, produced by the decay of dark mesons. As the dark mesons are *long-lived*, background events encompass a range of possibilities, ranging from QCD jets to other non-standard sources (that can't be precisely

simulated) like cosmic rays, beam effects, and random track crossings. Obtaining a discriminant for signal events is a complicated task due to the intricate topology of EJs in the signal events, demanding specialized algorithms and methodologies. Central to this work is the application of advanced machine learning techniques, particularly GNNs, to aid in the derivation of this discriminant. The GNN-based approach will be crucial in searching for signatures of LLP decays and thus will provide great value to the EJs Run-3 analysis.

Chapter 4

Machine Learning Introduction

> *Symmetry, as wide or as narrow as you may define its meaning, is one idea by which man through the ages has tried to comprehend and create order, beauty, and perfection.*
>
> – Herman Weyl

The data generated from particle collisions, such as those observed at the LHC, are inherently high-dimensional. Each collision event can produce hundreds of particles, each characterized by various parameters. This high dimensionality poses not only computational challenges but also complicates the extraction of useful information, a phenomenon often referred to as the "curse of dimensionality." ML-based algorithms excel in this regard and have become increasingly popular in HEP experiments in the recent decade with the advancements in the field of deep learning. The ML algorithms are trained on simulated data and then applied to LHC collision data. Unlike traditional ML algorithms, which might treat data entities independently, GNNs leverage the connections and interactions between data entities. Further details about the GNN are provided in the following section, whose discussion is influenced by the ideas and concepts presented in Ref. [31].

4.1 Graph Neural Networks (GNNs)

GNNs are special types of machine learning architectures that operate on graph-structured data. In the context of computer science, the graph is a type of *data-structure*. Graphs encode relations (edges) between different collections of entities (nodes) [32]. Graphs are

invariant under permutations, meaning that the order of nodes within a graph does not affect its overall representation. GNNs exploit this property by aggregating information from neighboring nodes, enabling them to create node embedding that captures the relational context. This makes GNNs effective tools for various tasks involving data that can be represented as graphs, including prediction and classification.

4.1.1 Graphs

an Mathematically, a *graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined as collection of *nodes* $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ and *edges* $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ between the pairs of nodes. An edge is *directed* if it is an ordered pair of distinct nodes, i.e. $\mathcal{E} = \{(v_i, v_j) \mid v_i, v_j \in \mathcal{V}\}$ and *un-directed* if it is an un-ordered pair of distinct nodes, i.e. $\mathcal{E} = \{\{v_i, v_j\} \mid v_i, v_j \in \mathcal{V}\}$. A more compact and structured definition of *graph-connectivity* is given by the *adjacency matrix*, $\mathbf{A}$. For a graph with n nodes, its $n \times n$ adjacency matrix is defined as:

$$A_{v_i v_j} = \begin{cases} 1 & (v_i, v_j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

For an undirected graph, $\mathbf{A}^\intercal = \mathbf{A}$ because the adjacency matrix is symmetric, reflecting the fact that if node i is adjacent to node j, then node j is also adjacent to node i. While for a directed graph, $\mathbf{A}^\intercal \neq \mathbf{A}$ because the edges, represented by ordered pairs, do not lead to symmetric adjacency relations. In the context of graph representation learning, nodes are often used to represent multi-dimensional (s-dimensions) *feature-vectors*. Feature vectors are numerical representations of data entities that capture various attributes and characteristics of those data entities and are denoted by $\mathbf{x_u}$ for $u \in \mathcal{V}$. A feature matrix $\mathbf{X}$ is obtained by stacking node features as rows of a $n \times s$ matrix:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1s} \\ x_{21} & x_{22} & \ldots & x_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{ns} \end{bmatrix}$$

Where x_{ij} represents the value of the jth feature for the ith data instance, n is the number of data instances, and s is the number of features.

Graphs are versatile data structures used across various fields and disciplines to model and analyze relationships, connections, and dependencies. Social networks are perhaps among the most commonly studied examples of graphs, where nodes represent users, edges correspond to friendship relations between them, and node features model user properties such as age, profile picture, etc. The friendship relationship for all the users can be represented by adjacency matrix ($\mathbf{A}$) as shown in Figure 4.1, where the colored region represents $a_{uv} = 1$, i.e the two users (represented by nodes) share an edge. Graphs are also commonly used as a representation of molecules, scientific citation networks, etc.

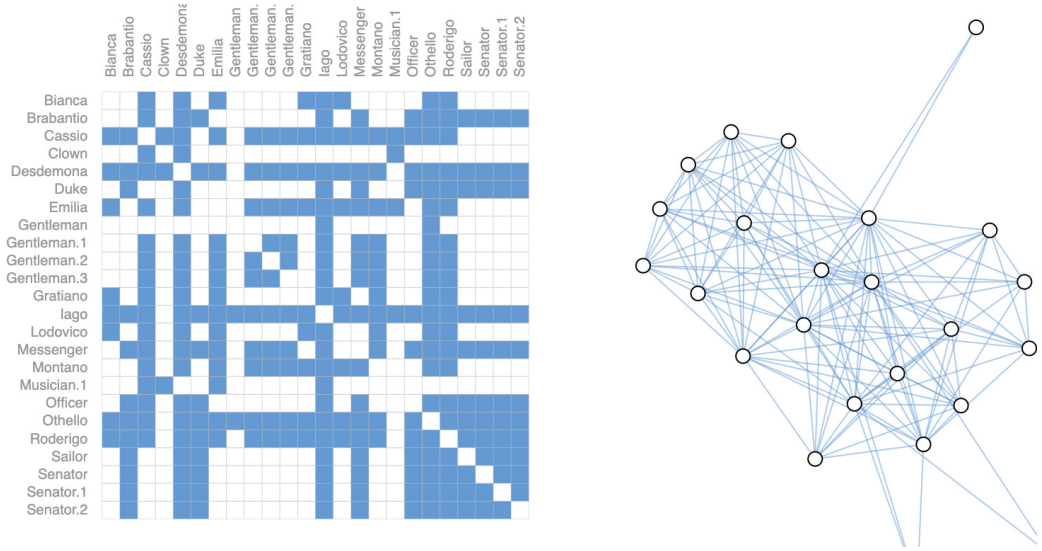


Figure 4.1: (Left) The adjacency matrix of the friendship between a group of people. (Right) Graph representation of these relations in the form of edges connecting different people - nodes. Image taken from Ref. [32].

In practice, unordered data sets with structural relationships are cast as graphs to learn about their relationships using advanced machine learning algorithms. However, data represented in grids and sets can also be viewed as a particular case of graphs, showcasing the extreme generality and powerful data representation capability of graph representations.

4.1.2 Symmetry of the Graph

The representation of the nodes $\mathcal{V}$ is a set without particular order. Thus, the graph exhibits structural symmetry of permutation invariance and can be modeled by a *group* of *permutations*.

A group is an abstract mathematical object formed from a collection of symmetries. A permutation group, $\mathfrak{G} = \sum_n$, is the group whose elements are all possible ordering of the node indices $\{1, ..., n\}$. Likewise, the *action* of permutation, $\mathfrak{g} \in \sum_n$, on a domain, D, is defined as mapping $(\mathfrak{g}, u) \rightarrow \mathfrak{g} \cdot u$ for $u \in D$. In machine learning, the input is typically represented as a 2-dimensional grid or a matrix. Therefore, when processing data by a machine learning algorithm, a node feature matrix is used instead of sets of node features. The action of a permutation group on a set of nodes reshuffles the rows of the input feature matrix $\mathbf{X}$. This transformation can be represented by a $n \times n$ matrix $\mathbf{P}$, in which each row and column contains precisely one '1', with all other entries being '0'.

4.1.3 Functions on a Graph

As graphs exhibit permutation symmetry, functions on them can display properties such as permutation *invariance* and *equivariance*, depending on their behavior under node permutations.

Invariant functions generate outputs that remain unaltered under group actions on the input domain. This concept is illustrated in Figure 4.2. As an example, consider the function:

$$f(\mathbf{X}) = \phi\left(\sum_{u \in \mathcal{V}} \psi(\mathbf{x_u})\right)$$

acting on a *set* of node features. Here, both ϕ and ψ are differentiable functions such as a Multi-Layer Perceptron (MLP), ψ transforms each node's features, while ϕ takes the sum of these transformed outputs over all nodes [33]. An MLP is a feedforward artificial neural network composed of multiple layers of nodes, where each node in a layer is connected to all nodes in the previous and subsequent layers. Because the summation remains consistent regardless of node label order, transforming $\mathbf{X}$ by $\mathbf{P}$ (i.e rearranging the node labels) will yield an unchanged result for the function i.e $f(\mathbf{PX}) = f(\mathbf{X})$, as shown in Figure 4.2.

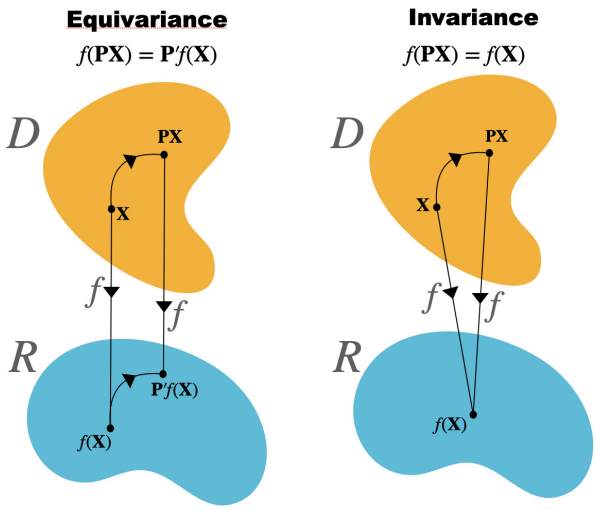


Figure 4.2: Differences between invariant and equivariant functions. Here, $f : D \to R$ is a mapping. $\mathbf{P}$ represents a permutation group whose action reshuffles rows of feature matrix $\mathbf{X}$, which is equivalent to relabeling the nodes of a graph. Invariant functions satisfy $f(\mathbf{PX}) = f(\mathbf{X})$, while equivariant functions satisfy $f(\mathbf{PX}) = \mathbf{P}f(\mathbf{X})$ for set of node information.

Conversely, equivariant functions exhibit a consistent transformation in their output as a result of the group's action on the input as shown in Figure 4.2. As an example, consider $f(\mathbf{X}) = \mathbf{H}$ acting on a *set* of input features, where $\mathbf{H}$ is stacks of latent node features $(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3..., \mathbf{h}_n)^T$. Here, 'latent' refers to features that have been transformed or generated from the original data $\mathbf{X}$, capturing underlying patterns or structures in the data. The latent features $\mathbf{h}_u$ are obtained after linearly transforming the input feature vectors $\mathbf{x}_u$. The order of the rows of $\mathbf{H}$ is tied to the order of the nodes of $\mathbf{X}$ so that we know which output feature a particular input feature corresponds to [31]. Formally, for such a function, we have $f(\mathbf{PX}) = \mathbf{P}f(\mathbf{X})$, showing its equivariant nature as illustrated in the Figure 4.2.

This discussion on equivariant and invariant functions focuses on the node information of a graph, particularly the set of node features. It doesn't account for the connectivity structure of the graph. In mathematical terms, a function f that incorporates the graph's connectivity through the adjacency matrix is *permutation invariant* if

$$f(\mathbf{PX}, \mathbf{PAP}^\intercal) = f(\mathbf{X}, \mathbf{A})$$

and a function f is *permutation equivariant* if

$$f(\mathbf{PX}, \mathbf{PAP}^\intercal) = \mathbf{P}f(\mathbf{X}, \mathbf{A})$$

[31].

Here, applying the permutation matrix $\mathbf{P}$ to the node feature matrix $\mathbf{X}$ implies applying to the rows and columns of adjacency matrix $\mathbf{PAP}^\intercal$ [31].

4.1.4 GNN

A GNN is an optimizable transformation on graph attributes (nodes, edges, and global context) that preserves graph symmetries [32]. Global context refers to the attributes applicable to the entire graph. The simplest way of designing a neural network architecture that can operate on a graph would be to use an adjacency matrix as an input to the MLP:

$$\mathbf{h}_\mathcal{G} = \mathbf{MLP}(\mathbf{A}[1] \oplus \mathbf{A}[2] \oplus ... \oplus \mathbf{A}[|\mathcal{V}|]),$$

where $\mathbf{A}[i] \in \mathbf{R}^{|\mathcal{V}|}$ is the row of an adjacency matrix and $\oplus$ denotes vector concatenation. However, this function does not exploit permutation symmetry, i.e. the neural network has to learn different weights if the nodes of the graphs are relabeled, even for the same graph! The most common GNN architectures are permutation equivariant functions constructed by applying a shared permutation invariant function $\phi(\mathbf{x_u}, \mathbf{X}_{\mathcal{N}(u)})$ over the local neighborhood i.e.

$$f(\mathbf{X}, \mathbf{A}) = \begin{bmatrix} -\phi(\mathbf{x_1}, \mathbf{X}_{\mathcal{N}_\mathbf{1}})- \\ -\phi(\mathbf{x_2}, \mathbf{X}_{\mathcal{N}_\mathbf{2}})- \\ \vdots \\ -\phi(\mathbf{x_n}, \mathbf{X}_{\mathcal{N}_\mathbf{n}})- \end{bmatrix}$$

[31].

Here, $\mathcal{N}_u = \{v : (u, v) \in \mathcal{E}\}$, is (un-directed) neighbourhood of a node u and $\mathbf{X}_{\mathcal{N}_u} = \{\{\mathbf{x}_v : v \in \mathcal{N}_u\}\}$ is a multi-set of neighbourhood features. Equivariance is ensured by

requiring ϕ to be independent of the order of the vertex in $\mathbf{X}_{\mathcal{N}_i}$, i.e ϕ is permutation invariant.

GNNs are the most general class of deep learning architecture where most of the other deep learning architectures (e.g. transformer style self-attention [34]) can be expressed as special cases of a GNN with additional geometric structure, via the graph network (GN) formalism [35]. Despite the generalization, most prevalent GNN architectures are derived from three *flavors*. Broadly, these flavors govern the extent to which ϕ transforms the neighborhood features [31].

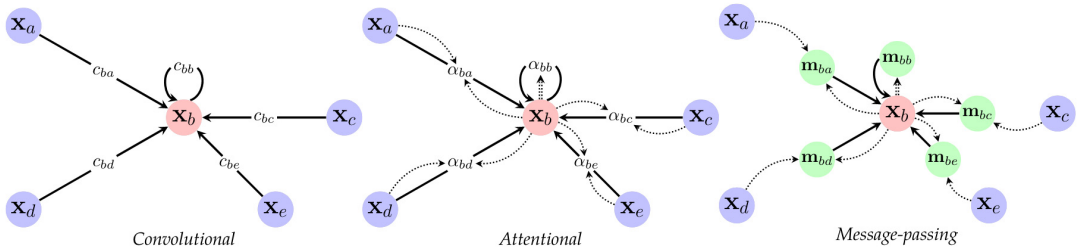


Figure 4.3: Data-flow visualization in three different flavors of GNN for node b from its neighboring nodes. (Left) In convolutional flavor, the neighborhood features are multiplied by a constant c_{uv} before being passed to node b. (Centre) In attention flavor, the weights are computed implicitly by attention mechanism $a(\mathbf{x}_u, \mathbf{x}_v)$. (Right) In message passing flavor, instead of having features of sending nodes multiplied with scalar weights, a vector-valued message is computed based on features of both sender and receiver nodes: $m_{uv} = \psi(\mathbf{x}_u, \mathbf{x}_v)$ Image from Ref. [31].

In the **convolutional flavor** [36, 37] of GNN, node features are updated by aggregating neighboring features with fixed weights c_{uv}. The aggregation is performed as:

$$\mathbf{h}_u = \phi\left(\mathbf{x}_u, \square_{v \in \mathcal{N}_u} c_{uv}\psi(\mathbf{x}_v)\right)$$

.

Here, c_{uv} is a constant that directly depends upon elements in $\mathbf{A}$. The symbol $\square$ represents a permutation invariant aggregation operation, such as sum, mean, or maximum. The functions ϕ and ψ are differentiable, such as the MLP. As shown in Figure 4.3, convolutional GNNs multiply neighborhood features (typically after being transformed by ψ)

with a predetermined weight c_{uv} for aggregation. This aggregated representation is then transformed by ϕ to produce the updated node vectors $\mathbf{h}_u$.

Attentional flavor [38, 39, 40] of GNNs use the attention mechanisms to weigh the importance of neighboring nodes when aggregating information. The node representation is transformed as:

$$\mathbf{h}_u = \phi\left(\mathbf{x}_u, \underset{v\in\mathcal{N}_u}{\square} a(\mathbf{x}_u, \mathbf{x}_v)\psi(\mathbf{x}_v)\right)$$

Here, $a(\mathbf{x}_u, \mathbf{x}_v)$ is generally a learnable single-layer function whose output is soft-max normalized giving a score of similarity/importance of neighboring node feature. As illustrated in Figure 4.3, the neighboring node features (potentially after transforming by ψ) are weighted by attention scores before being aggregated. It's worth noting that there exist various other methods to determine these *self-attention* scores.

The **message-passing flavor** [35, 41] of GNN involves exchanging and updating messages between neighbors.

$$\mathbf{h}_u = \phi\left(\mathbf{x}_u, \underset{v\in\mathcal{N}_u}{\square} \psi(\mathbf{x}_u, \mathbf{x}_v)\right)$$

Vector-valued messages $\psi(\mathbf{x}_u, \mathbf{x}_v)$ are generated by taking both the sending node and receiving node's features as shown in Figure 4.3. This *message* is then *passed* by aggregating it across all neighbors. Subsequently, it is transformed using ϕ to update the node representation to $\mathbf{h}_u$.

Message-passing neural networks provide a general framework where convolutional and attentional flavor can be thought of as special cases of message-passing GNNs [41]. In this context, the message takes in account only the sender's node features given by $\psi(\mathbf{x}_u, \mathbf{x}_v) = c_{uv}\psi(\mathbf{x}_v)$ for convolutional GNNs and $\psi(\mathbf{x}_u, \mathbf{x}_v) = a(\mathbf{x}_u, \mathbf{x}_v)\psi(\mathbf{x}_v)$ for attentional GNNs.

4.2 GNNs in particle physics

A type of particle physics experiment measures data generally by building large particle accelerators with complex detectors (such as ATLAS discussed in Section 3.1). The data in

these detectors is inherently sparse due to the design and geometry of the sensors and does not a-priory fit into a grid-like structured data [42].

The use of machine learning algorithms to determine the properties of particles at colliders such as LHC has become prevalent in recent years [43]. Some of the data from the ATLAS detector can be approximated as images and computer vision techniques can be applied to improve the performance of different analyses. For example, in Ref. [44], the calorimeter cells of the ATLAS detector are projected into a fixed grid of η and ϕ to get an image representation of calorimeter data. This image representation is then used to classify gluon vs quark-induced jets using Convolutional Neural Networks (CNNs). Nevertheless, image-based representations encounter challenges due to the irregular geometries of detectors and the sparse nature of the projections used. The inherent information loss in these representations can put a limit on the amount of information that can be extracted from these data.

Likewise, sequential data representations have been employed for various data-analysis tasks in the ATLAS experiment. For instance, in Ref. [45], a Recurrent Neural Network is used to determine the flavor of particle jets. To fit the data as a sequence, the tracks within a jet are ordered based on S_{d_0} - the lifetime signed transverse impact parameter significance [45]. However, this imposed ordering can potentially compromise learning performance. This limitation is demonstrated by Ref. [46] when comparing the results of the sequence-based algorithm with a set-based algorithm like Deep Sets [33].

Casting particle physics data as unordered sets with complex relations and interactions, i.e. *graphs,* offers a more natural approach with fewer constraints and limitations compared to the representations discussed above where it was often necessary to "force" the data into a particular representation. For instance, as shown in Figure 4.4, the tracking detector hits information that can be visualized as nodes in a graph, with their inherent relationships forming the edges. This graph-based representation is well-suited for GNNs to effectively predict relationships between different nodes and, in turn, reconstruct particle trajectories. Currently, GNNs are used for a wide variety of applications in particle physics such as jet

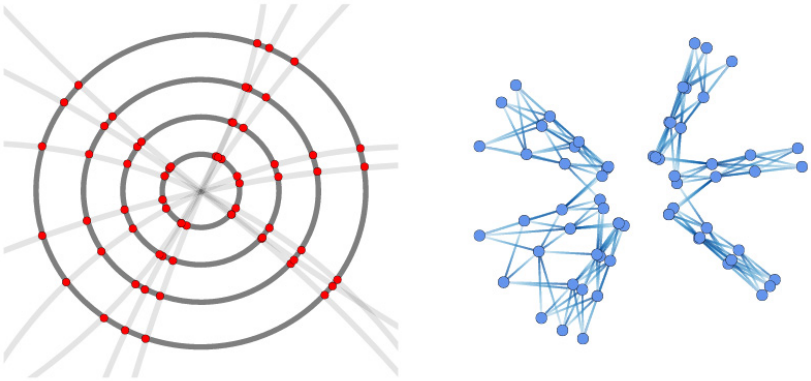


Figure 4.4: Clustering tracking detector hits into tracks using GNNs. Image from Ref [42].

classification [47], charged particle tracking [48], secondary vertex reconstruction [49], pileup mitigation [50], and event-classification [51] among others.

4.3 GNN in Flavour Tagging

The identification of jets originating from b and c quarks - flavor tagging, is a crucial component of physics analysis in the ATLAS experiment. It is important for both precise SM measurements and exploration of BSM physics scenarios.

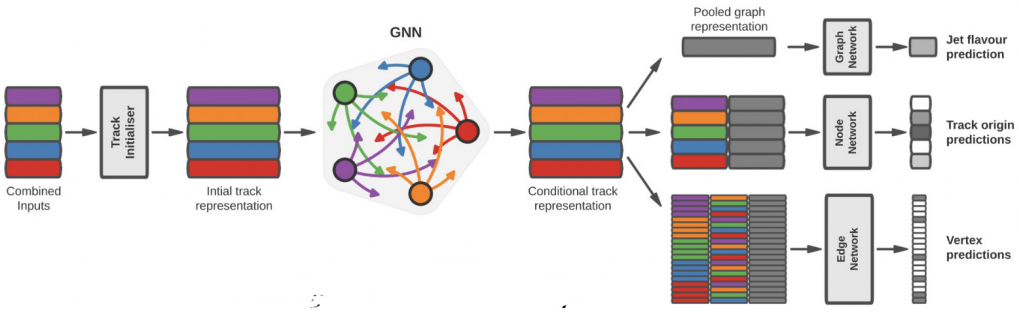


Figure 4.5: Architecture of for jet-flavor, track origin, and vertex-pair prediction. Combined jet and tack variables are fed into the GNN network. The output is a classification of jets, tracks and identification of tracks with common pairs of vertex. Image from Ref. [42].

ATLAS released a Graph Neural Network-based flavor tagging algorithm [52]. It uses individual track parameters, their associated uncertainties, and the jet's p_T - transverse momentum and η - signed jet pseudorapidity as inputs. In the network's graph, each node

represents a distinct track within the jet and is defined by a 23-dimensional feature vector ("combined inputs" as referenced in Figure 4.5) derived from the mentioned inputs. A more detailed discussion of the input variables can be found in Ref. [52].

The combined output is fed into a track initialization network as shown in Figure 4.5, comprising three hidden layers, producing an output size of 64 neurons. The track initialization network is similar to the Deep Sets model [33] but does not include an aggregation operation [52]. The network's output is used to populate the graph's nodes as feature vectors to the GNN. These input node features are then updated using three layers of a dynamic attention-based message passing neural network [53]. The final graph representation is then constructed using global attention pooling (GAP) [54]. Subsequently, the dense layer then predicts the jet flavor as shown in Figure 4.5. A dense node classification network, using features from a single updated node representation combined with the global jet representation, predicts the track truth origin (see Ref. [52] for more information) as shown in Figure 4.5. Finally, a dense edge classification network, utilizing the combined representations of track pairs and the overall jet representation, determines if the tracks in the pair share a common vertex.

The training of the GN1 architecture aims to minimize the composite loss function given by

$$L_{\text{Total}} = L_{\text{jet}} + \alpha L_{\text{track}} + \beta L_{\text{vertex}}$$

. Where:

- L_{jet} represents the categorical cross-entropy loss, averaged across all track pairs.
- L_{track} denotes the categorical cross-entropy loss associated with predicting the track origin.
- L_{vertex} is derived from the binary cross-entropy loss for track-pair compatibility, averaged over all track pairs.

Here, α and β serve as weighting parameters, ensuring that the categorical loss from jets has a more significant impact than both track identification and vertex identification.

It has been demonstrated that GN1 shows superior performance when compared to the previous ATLAS flavor tagging approaches. For jets coming from $t\bar{t}$ decays with transverse momentum $20 < p_T < 250$ GeV, it was shown that GN1 increases the light (c)-jets rejection by a factor of $\sim 1.8(\sim 2.1)$, for a b-jet efficiency of 70% [52]. Likewise, for jets from Z' decays with transverse momentum $250 < p_T < 5000$ GeV, it was demonstrated that GN1 improves the light(c)-jet rejection by a factor if $\sim 6(\sim 2.8)$, for 30% b-jet efficiency [52].

4.4 GNNs for tagging Emerging Jets

Building on the success of the GNN tagger in ATLAS flavor tagging, this thesis predominantly focuses on repurposing the GNN algorithm for "emerging-jets" tagging, which is integral to the EJs Run 3. EJs, with their intricate topology, require sophisticated tools for precise identification and study. Using the GNN-based flavor tagger's algorithm, a GNN model can be trained to identify these EJs. This section provides a discussion of the training samples, input parameters, and the GNN architecture used for EJ classification.

Training sample

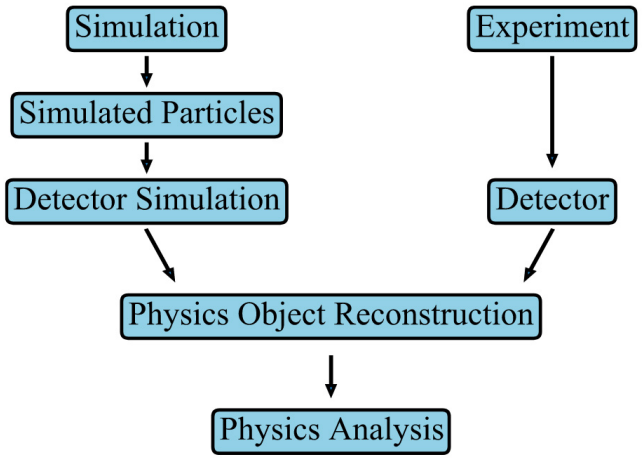


Figure 4.6: ATLAS uses simulation to create a "true" record of events in the detector. Machine learning algorithms are trained based on these simulated samples. Later, the models from machine learning algorithms are used to infer the physics of detector-based data. Image from Ref. [42].

The training of the GNN uses simulated datasets: SM events with QCD jets served as the background, while BSM events, characterized by dark quark pair production with

EJ signatures, are used as signals. In HEP experiments, simulation is used to establish a 'truth record' of the physics event responsible for a given detector outcome as shown in the Figure 4.6. This 'truth record' helps to train machine learning algorithms like GN1 to classify physics objects reconstructed from the ATLAS detector. These events were initiated by proton-proton collisions at a center of mass energy of $\sqrt{s} = 13.6$ TeV. The parameters considered for the simulation of the signal sample are listed in Table 4.1 and are based on models discussed in Section 2.2.1 and Section 2.4. The training used a total of ~2 million jets, with an equal split between jets from signal (BSM) and background (SM event).

	Model A	**Model B**	**Model C**
m_{π_d} (GeV)	5	10	20
$c\tau_{\pi_d}$ (mm)	5-50		
$m_{Z'}$ (GeV)	600-1500-3000		
Decay in dark sector	Dark rhos to dark pions		
Decay to SM	Dark pions to quarks		

Table 4.1: Parameters for benchmark models used to simulate signal samples.

GNN Architecture

Name	**GN1**	**GN2**
Trainable parameters	0.8M	1.5M
GNN Layers	3	6
Attention Heads	2	8
Embed dimension	128	129
Layer Normalization + Dropout	No	Yes
Dense update	No	Yes
Attention type	GATv2	Scaled Dot Product
Learning rate	$1e-3$	Once Cycle LRS (max LR $4e-5$)
Separate value projection	No	Yes

Table 4.2: Modified from GN1, GN2 now adopts a transformer-style architecture.

Recent developments by the ATLAS flavor tagging group have led to the extension of the features of GN1 leading to GN2 where the majority of the changes are optimizations of the model hyper-parameters. GN2 uses transformer-like circuits [34]. For example, layer normalization, first introduced in Ref. [34], has been shown to improve the expressivity (ability of NN to approximate functions) of the network [55]. The GN2 architecture also introduces dropout which helps with over-fitting. The type of attention mechanism has

been modified from dynamic graph attention [53] to scaled dot product introduced in the original transformer paper [34] which enhances training efficiency without compromising the physics performance. In this configuration, the GN2 architecture distinctly computes attention weights and updates node representations, with a dense layer present between the attention layers. Additionally, a one-cycle learning rate scheduler [56] is used. A full list of modifications is listed in Table 4.2. Initial studies for the EJs classification were based on GN1 but later updated to GN2. The outputs are soft-max probabilities for 2 classes of jets, 4 classes of tracks, and a track-pair compatibility score described in the section below.

Training Labels

To train the GNN, the dataset is classified into groups, first associating them with the training labels.

Jet Labels: The ATLAS simulated dataset discussed above does not have a "true" label for a jet that can distinguish between the signal jets and the background jets discussed in the section. Therefore, each jet in the training dataset has an "isDisplaced" binary label used for training. A jet is 'displaced" if

$$\sum_{\text{Displaced tracks}} p_T \text{ fraction} \geq 0$$

, where p_T fraction is the fraction of jet transverse momentum carried by tracks inside a jet and displaced tracks are the tracks reconstructed by the LRT algorithm discussed in Section 3.1.1. Signal samples with EJs are displaced jets while background samples with QCD jets are prompt (ie. isDisplaced = 0).

Track Labels: Tracks are labeled based on their truth information into four classes - pileup, fakes, primary, and displaced. Pile-up tracks arise from additional proton-proton interactions that occur within the same bunch crossing as the primary interaction. These interactions can produce secondary tracks that overlay the primary collision data and contaminate the signal jets. Such tracks are assigned a training **label "0"**.

As discussed in Section 3.1.1, tracks are scored by the tracking algorithm. Tracks that are from purely combinatorial collections of hits are fakes and are given a training **label "1"**

Primary tracks are those directly associated with the primary proton-proton interactions in the ATLAS detector. These tracks originate from the decay of particles produced in the initial collision and are prompt. They are assigned training **label "2"**.

Displaced tracks are tracks originating from secondary vertices that are displaced from the primary interaction point and are reconstructed by the LRT algorithm. These tracks are associated with the decay of long-lived particles, such as dark mesons, which travel a measurable distance before decaying. These tracks are assigned **label "3"**. Detecting and identifying displaced tracks are crucial for studying the properties of the potential new physics.

Additionally, vertex truth information is also included in tracks, i.e. a set of tracks inside a jet has the same discrete label assigned to it if they have the same decay vertex.

Input Variables

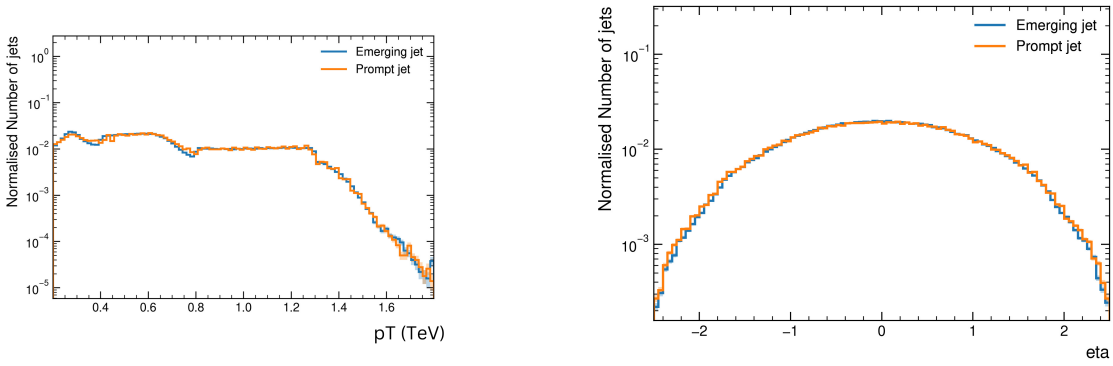


Figure 4.7: Distribution of jet-related Input Variables: p_T on the left and η on the right

.

The inputs to the GNN are two jet variables that constitute the basic kinematics of a jet - p_T, η, combined with 16 other track variables containing information about the reconstructed track parameters and constituent hits, detailed in Table 4.3. The distribution of jet variables p_T and, η is shown in Figure 4.7. These are "resampled" distributions to

Table 4.3: List of input variables for training with their respective descriptions.

Category	**Variable**	**Description**
Jet	p_T	Jet transverse momentum
	η	Signed jet pseudorapidity
Track	d_0	Distances of closest approach between the track and beamline in the transverse plane
	$z_0 \sin\theta$	Closest distance from the track to the primary interaction point in the transverse plane
	d_ϕ	Azimuthal angle of the track, relative to the jet ϕ
	dη	Pseudorapidity of the track, relative to the jet η
	$\frac{q}{p}$	Track charge divided by momentum (measure of curvature)
	$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
	$\sigma(\theta)$	Uncertainty on track polar angle θ
	$\sigma(\frac{q}{p})$	Uncertainty on $\frac{q}{p}$
	nPixHits	Number of pixel hits
	nSCTHits	Number of SCT hits
	nPixShared	Number of shared pixel hits
	nSCTShared	Number of shared SCT hits
	nPixHoles	Number of pixel holes
	nSCTHoles	Number of SCT holes
	IP3D_signed_d0_significance	Ratio of d_0 and $\sigma(d_0)$ defined for both positive and negative scale with reference to the primary interaction point.
	IP3D_signed_z0_significance	Ratio of $z_0 \sin(\theta)$ and $\sigma(z_0 \sin(\theta))$ defined for both positive and negative scale with reference to the primary interaction point

ensure uniformity in the kinetic distribution which is crucial to avoid kinematic biases in the tagging performance. This resampling is performed by using UMAMI, a software developed by the ATLAS flavor tagging performance group. Input track variable distributions are shown in Figure 4.8, Figure 4.9 and Figure 4.10. The low-level hit-related information from the Semiconductor Tracker (SCT) and the Pixel Detector of the ATLAS is also used as the input feature for each track. A "hit" refers to the interaction of a particle with the detector material that results in a measurable signal. Likewise, a hole refers to a missing expected

hit in a track. If two or more particles use the same hit information to reconstruct the track, then it is referred to as the "shared hit". Input features for the track are then the number of hits, shared hits, and holes from each of Silicon Tracker and Pixel Detector of the ATLAS. From the distributions, it seems that the most discriminatory variables for the tagger were d_0, $\frac{q}{p}$ and IP3D_signed_d0_significance.

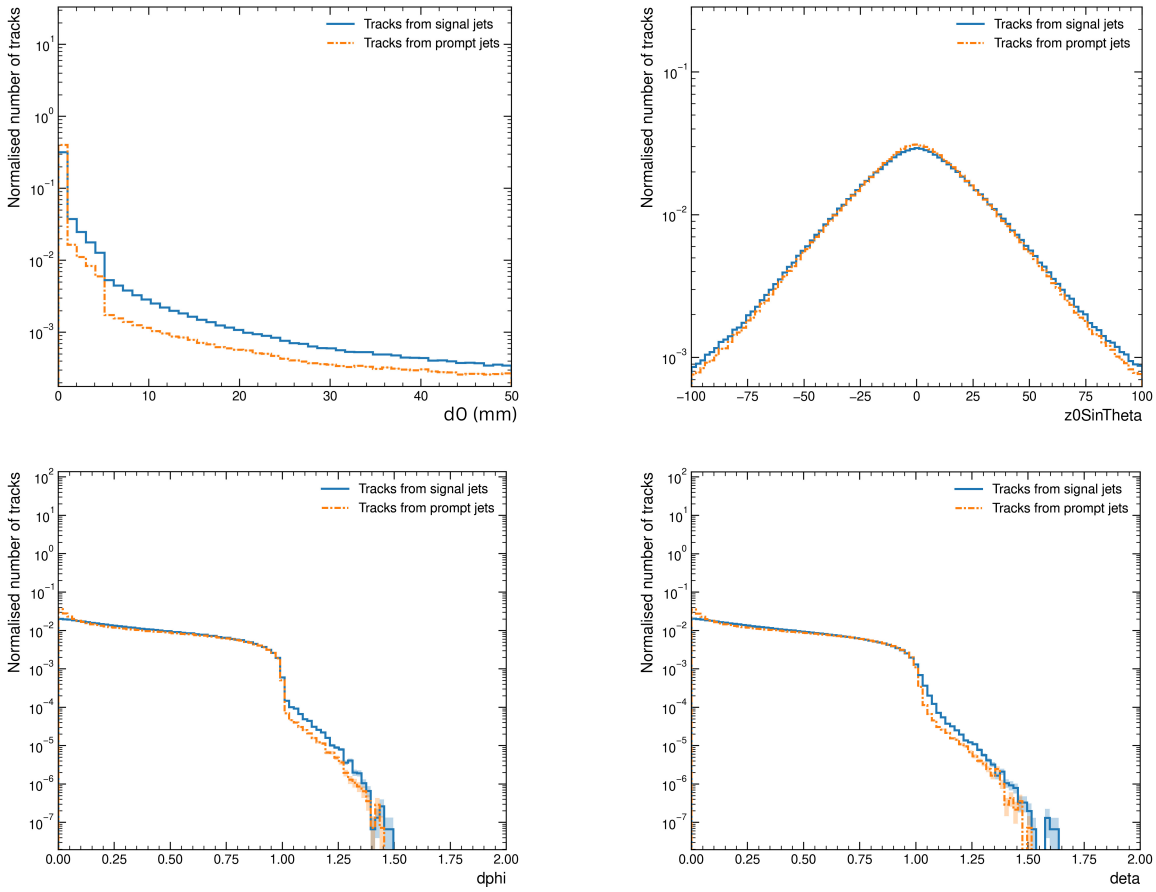


Figure 4.8: Distribution of track-related variables d_0, $z_0 \sin\theta$, d_ϕ and d_η

.

This model underwent training for 20 epochs for 3 classification tasks - jet classification into signal or background jets, track classification into the four-track classes discussed above, and vertex prediction task which is explained in the following section. Also, a comprehensive assessment of its performance is detailed.

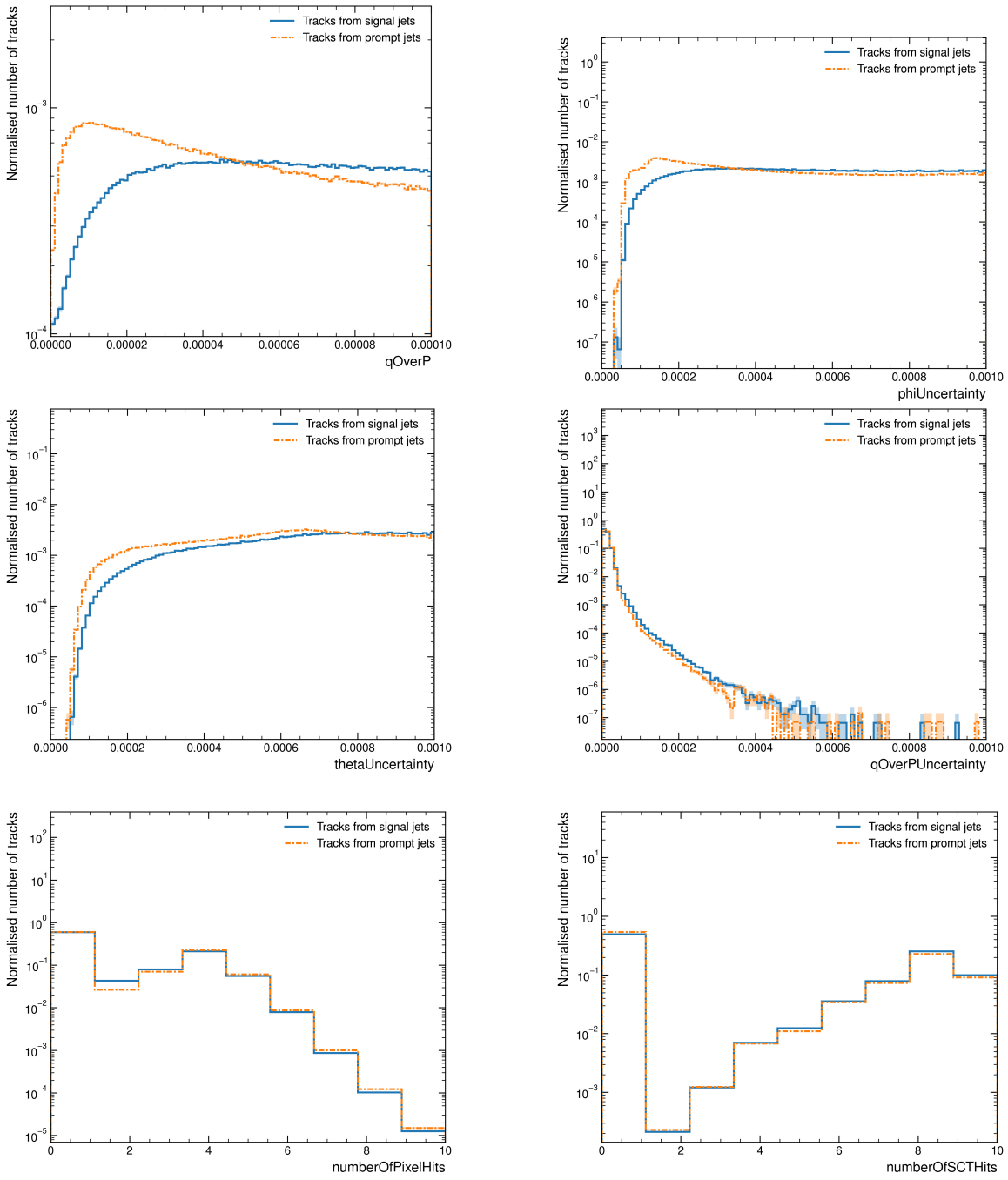


Figure 4.9: Distribution of track related variables: $\frac{q}{p}$, $\sigma(\phi)$, $\sigma(\theta)$, $\sigma(\frac{q}{p})$, number of pixel hits and number of SCT hits.

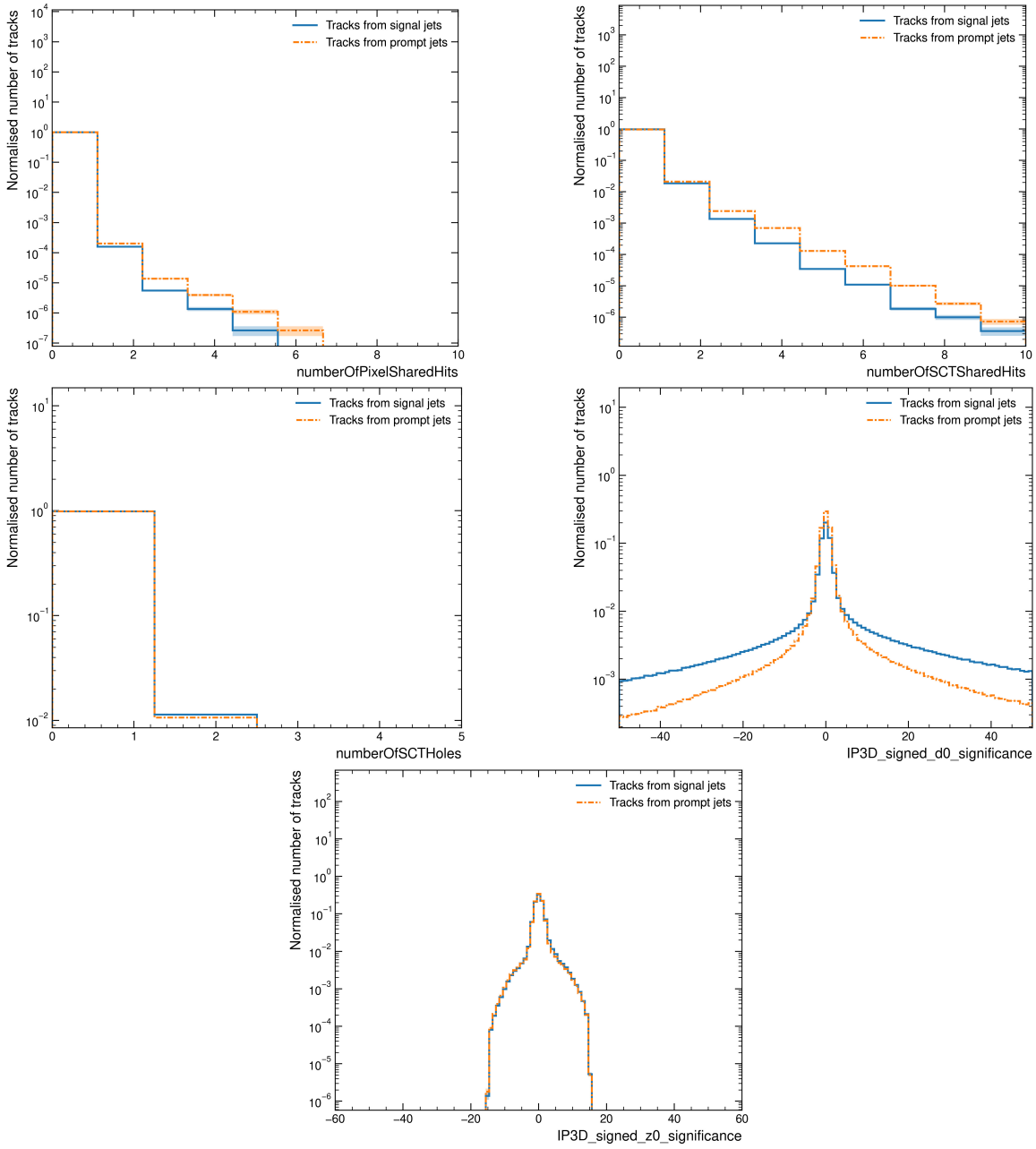


Figure 4.10: Distribution of track related variables: number of shared pixel hits, number of shared SCT hit, number of SCT holes, IP3D_signed_d0_significance and IP3D_signed_z0_significance.

Chapter 5

Performance of Graph Neural Network

Life is a constant process of dyeing.

– Arthur Schopenhauer

In the final chapter, a comprehensive evaluation of the effectiveness of GNNs within the context of the study on EJs is provided. The chapter begins with an evaluation of GNNs' performance in vertex identification, examining their efficiency, purity, and ability to capture the intricate topology of EJs, especially concerning vertex distributions. Then, the performance of GNNs in the track classification is also presented. Finally, GNNs' performance in EJ tagging is detailed along with its application in the EJ Run-3 analysis.

5.1 Vertex Identification

To investigate the characteristics of dark particles leading to EJ signatures, it's crucial to initially identify the collection of decayed particle tracks that share the same origin (referred to as *vertex finding*) and then the decay location (referred to as secondary "*vertex fitting*"). By using the particle physics dataset's *jet-graph representation*, where individual graph nodes represent characteristics of individual tracks and their associated particle jet, it becomes feasible to train a GNN model for determining if two nodes share an edge. This edge classification task serves as an advanced algorithm to find if sets of tracks share a common displaced vertex.

The performance of the GNN model's **secondary vertex** finding task for the tracks associated with EJs is detailed below using different evaluation metrics such as efficiency, purity, and number of identified vertex distribution. To offer a more insightful perspective on its performance, these performance metrics are compared with the current ATLAS secondary vertex finding algorithm, known as VSI (VrtSecInclusive).

5.1.1 Efficiency

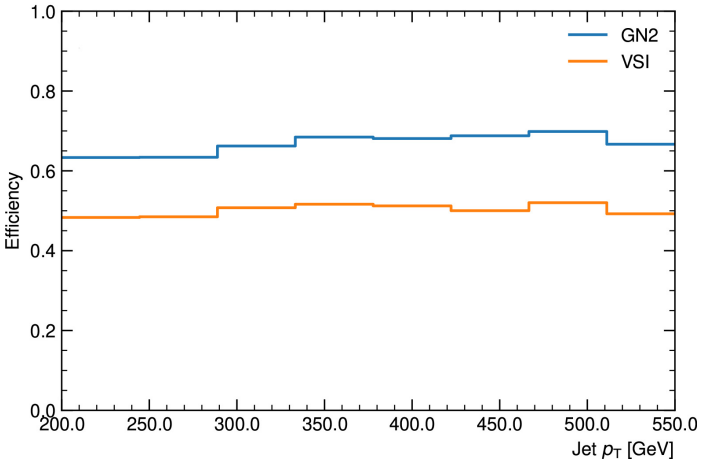


Figure 5.1: Efficiency comparison of GNN model (GN2) with VSI.

Efficiency is a crucial metric that quantifies how often the GNN accurately identifies the displaced vertices. For a more fine-grained notion of efficiency in the context of this analysis, efficiency is defined as the per-vertex fraction of tracks in the truth-vertex that are included in a common reconstructed vertex, i.e.

$$\text{Efficiency} = \frac{\#\text{tracks in predicted vertex that are also in truth vertex}}{\#\text{tracks in the true vertex}}$$

Consider the following example, the track IDs of the tracks in the truth vertex is: $[1001, 1002, 1003, 1004, 1005]$ and the GNN predicts the groups of tracks into vertex with IDs: $[1001, 1002, 1003]$. Then, three out of five tracks in the true vertex were efficiently identified with an efficiency score of $\frac{3}{5}$.

Figure 5.1 illustrates the efficiency averages as a function of p_T for the associated vertices. A high-efficiency level reflects the GNN algorithm's effectiveness in identifying the majority of authentic vertices that share common sets of tracks.

It was observed that the efficiency of the GNN-based vertex finding algorithm, when compared to VSI shows an improvement in identifying secondary vertices associated with EJs across the full range of p_T values. This highlights the superior performance of the GNN approach in comparison to a more traditional probabilistic approach.

5.1.2 Purity

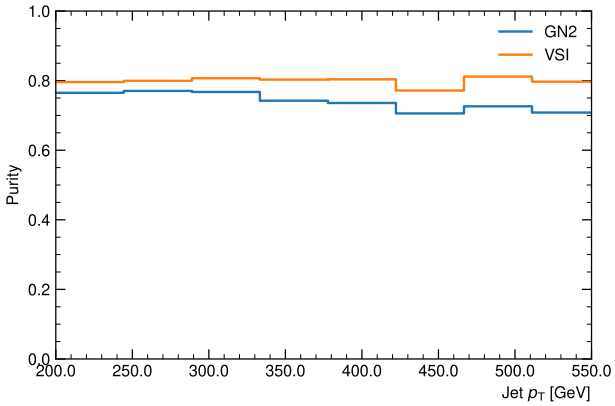


Figure 5.2: Purity comparison of GNN (GN2) with VSI.

Purity is another metric used to evaluate the quality of the vertex that GNN identifies. In the context of this analysis, purity is defined as the per-vertex fraction of tracks in the reconstructed vertex that are from the same truth vertex, i.e,

$$\text{Purity} = \frac{\#\text{tracks in true vertex that are also in predicted vertex}}{\#\text{tracks in the predicted vertex}}$$

Like in the case of efficiency, consider the example where the track IDs of the tracks in the truth vertex is: $[1000, 1001, 1002, 1003, 1004, 1005]$ and the GNN predicts the groups of tracks into vertex with IDs: $[1001, 1002, 1003]$. In this case, all three tracks that belong to the predicted vertex belong to the truth vertex. Thus, the predicted vertex is pure - with a purity score of 1.

A high purity value implies that the algorithm has successfully identified genuine vertices while minimizing the inclusion of false or spurious tracks. Purity is essential in ensuring the fidelity of the data analysis, as it directly impacts the precision of results derived from the reconstructed vertices.

GNNs and VSI show similar purity scores across most p_T ranges for vertices associated with jets as shown in Figure 5.2. VSI does have a slightly better purity for jets with higher p_T values. Overall, GNN's reconstructed vertices are very reliable!

5.1.3 Vertex Distribution

Capturing the complexity of the EJ accurately is essential for the analysis. The number of vertices per jet distribution provides insight into the performance of GNN in capturing the intricate jet topology.

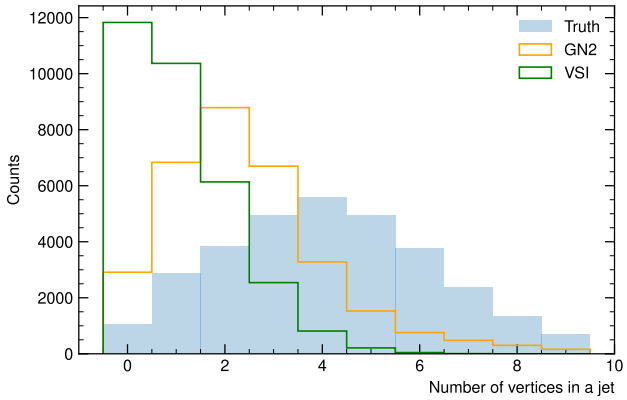


Figure 5.3: Number of vertex per jet distribution based on truth labels and predicted labels from VSI and GNN.

Truth distribution, in Figure 5.3, represents the actual distribution of vertices in a jet based on vertex truth labels of the simulated sample as discussed in Section 4.4. The distribution peaks around 4, reflecting the typical vertex count in these jets. The GNN-identified vertex's distribution shows a peak at 2, a slight deviation from the truth distribution. Nevertheless, the GNN's distribution is closer to the truth distribution. In contrast, the VSI's vertex distribution peaks at 0.

This shows that GNNs can reconstruct multiple vertices in a jet. While the GNN-generated distribution may not perfectly match the truth distribution, it captures the emerging jet topology much better than the existing VSI algorithm.

5.2 Track Identification

Ultimately, studying the properties of potential new physics requires precise identification of the origin of tracks associated with EJs. Utilizing the jet-graph representation of the particle physics dataset, it becomes feasible to employ GNNs for node classification through the process of node representation learning. In this study, tracks were classified into four distinct groups: primary, displaced, fake, and pileups as described in Section 4.4.

Track Origin ROC

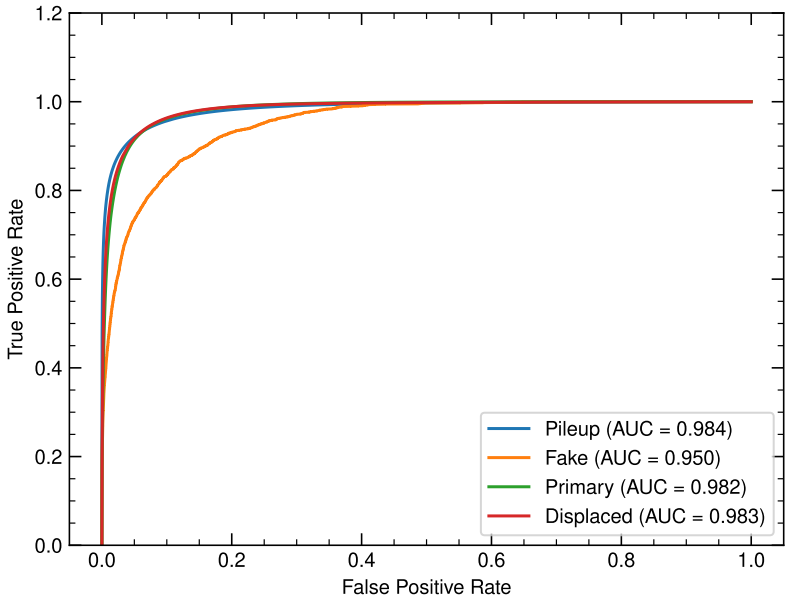


Figure 5.4: ROC curve for track origin classification using GNN model with its associated area under the curve (AUC).

The Receiver Operating Characteristic (ROC) curve serves as a fundamental tool in evaluating machine learning classification models. The ROC curve construction involves plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) across vari-

ous threshold settings. In this specific application, ROC curves for each tracking category are generated, and the Area Under the Curve (AUC) metric is used to quantitatively measure classification performance. A higher AUC score implies that the model has a high discriminating power across different classes, can handle imbalanced datasets, and is reliable and robust in correctly identifying tracks originating from primary interactions, displaced vertices, and pileup interactions.

The Area Under the Curve (AUC) values were calculated for each track category, yielding high AUC scores of 0.984 for pileup tracks, 0.982 for primary tracks, and 0.983 for displaced tracks as shown in Figure 5.4 Collectively, these high AUC values show the classification model's robustness when dealing with various track categories within EJs, and its adeptness at managing imbalanced datasets and accurately classifying tracks!

Track Origin Confusion Matrix

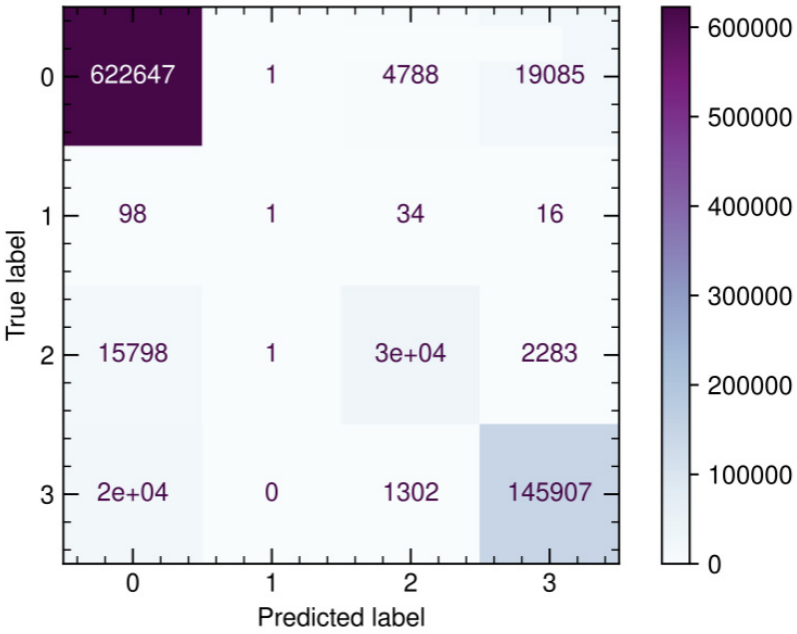


Figure 5.5: Confusion Matrix

The confusion matrix offers a visual representation of the model's predicted labels compared to the truth labels. It lays out a detailed breakdown of the model's predictions,

allowing for an evaluation of correct and incorrect classifications across each track category: primary, displaced, pileup, and fake tracks.

Each row of the confusion matrix represents the true track category based on truth labels discussed in Section 4.4, while each column corresponds to the predicted category based on the GNN model. Also presented in the Section 4.4, a track label of "0" represents pileup, "1" represents fake tracks, "2" represents primary tracks, and "3" represents displaced tracks. The diagonal elements of the matrix represent the correct classifications, indicating how many tracks were accurately identified within their respective categories. Off-diagonal elements represent misclassifications, offering insights into which categories tend to be confused with each other by the GNN model.

Equal instances of displaced tracks are misclassified as pileups and vice versa as shown by the shaded regions on the lower-left and the top-right box of the Figure 5.5. Also demonstrated by the darkly shaded top-left and bottom-right of the is the model's strong ability to correctly classify displaced tracks and pileup tracks, with a relatively low number of misclassifications, demonstrating the reliability and effectiveness of GNNs in handling multi-class classification tasks.

5.3 Jet Classification

The task of graph classification enables the prediction of a jet's class category, which, in this context, is either signal jets or background jets. This classification is based on the utilization of jet-graph representations. The performance of this machine learning task is demonstrated using probability separation distributions and ROC showing background rejection versus signal efficiency.

Probability Separation

The x-axis of Figure 5.6 represents the GNN scores attributed to each jet, which is the probability that a jet is classified as a signal jet. As shown in Figure 5.6, background (prompt) jets are denoted by red, while signal (displaced) jets are represented by blue. Two distinct peaks can be seen in the first and last bin. The prominent peak in the first bin for the background jets suggests that the majority of these jets have GNN scores within the

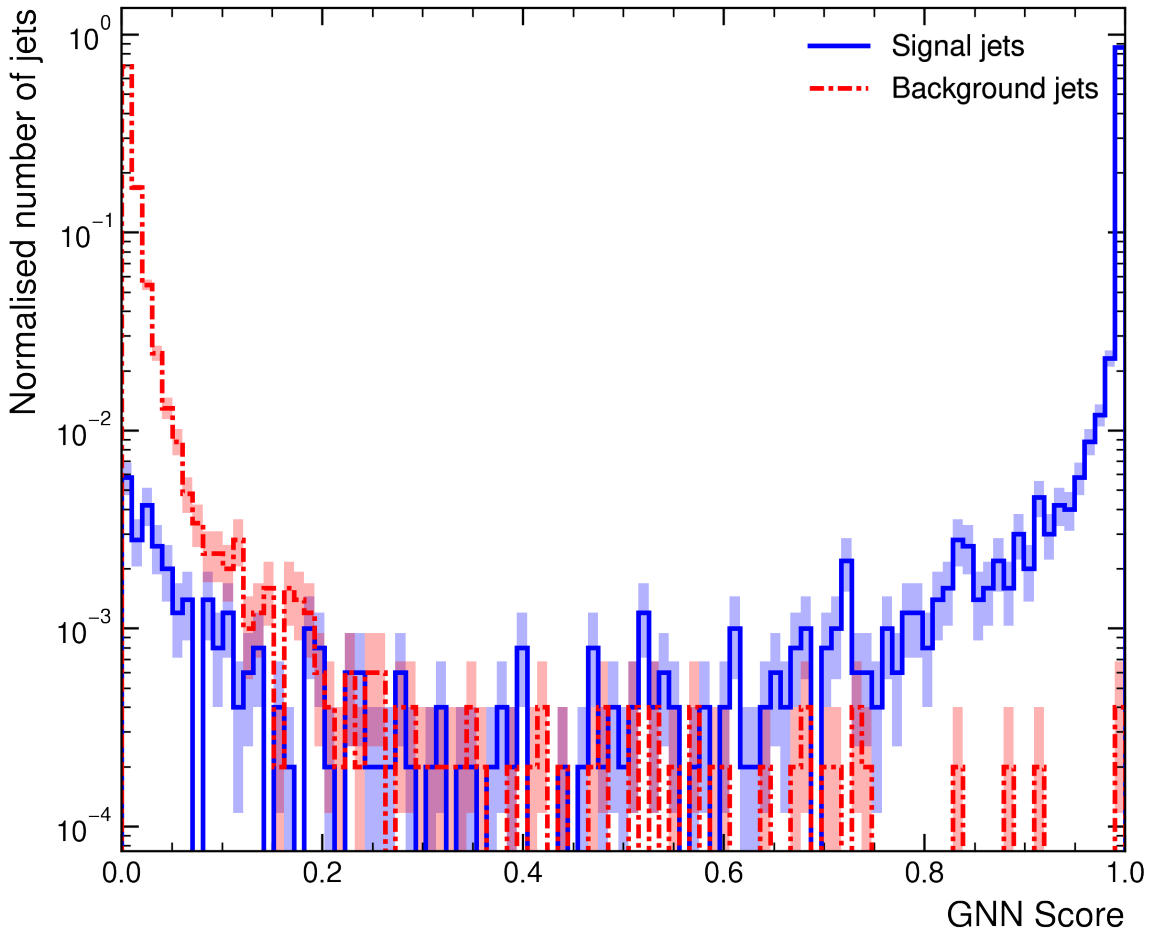


Figure 5.6: Probability distribution of GNN scores for jets. The scores are soft-max probabilities for each class - displaced (signal) jets and prompt (background) jets.

range of 0-0.01. This indicates an extremely low likelihood of these background jets being classified as signal jets. Similarly, the peak at the far right end of the histogram shows that the majority of signal jets are correctly classified with a high likelihood, falling within the 0.99-1.0 range. Also, the limited overlap between these two distributions highlights the effective separation between signal and background jets.

This observation demonstrates the efficiency of our model, as it effectively distinguishes between signal and background jets, providing a clear separation between the two jet classes.

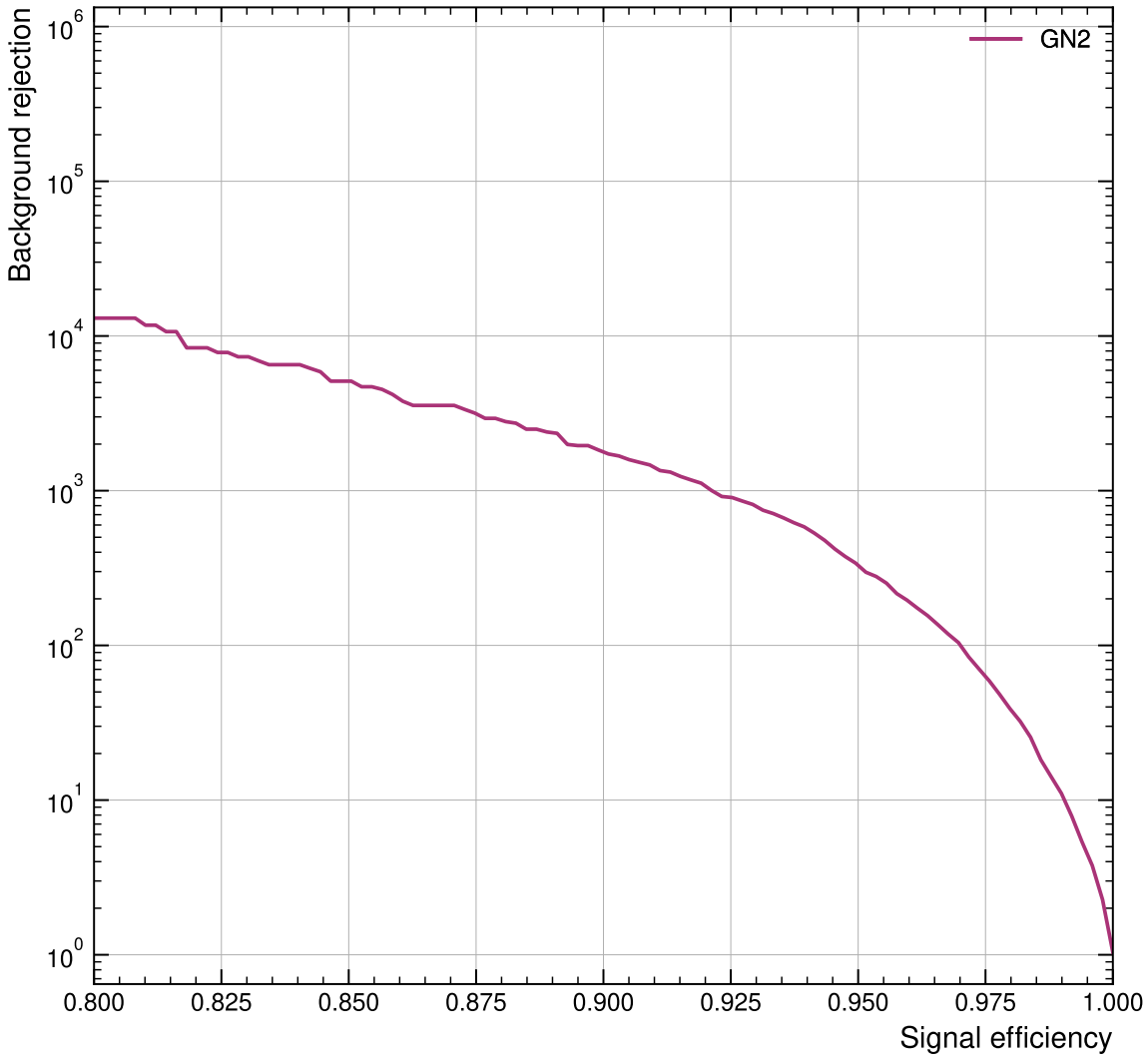


Figure 5.7: Background rejection vs signal efficiency.

Background Rejection vs Signal Efficiency (ROC)

In the context of this model's evaluation on jet classification, a specialized representation of the ROC curve is used, where the y-axis represents background rejection ($\frac{1}{\text{background efficiency}}$), plotted on a logarithmic scale and the x-axis represents signal efficiency from 0.8-1.0 scale.

Having a high background rejection is crucial to increase the sensitivity of the analysis. By focusing on the high signal efficiency region (0.8 to 1.0), this representation emphasizes the classifier's ability to retain the majority of signal events, which is critical as maximizing signal retention for EJs is essential. This representation directly relates to the sensitivity of the analysis, i.e., its ability to distinguish between signal and background. Figure 5.7 shows

that background rejection is greater than 10^4 for signal efficiency of 0.8. This implies that within a threshold where 80% of the signal jets are accurately identified, there is a solitary misclassification of background jets for every $\sim 10^4$ background jets. The exceptionally high level of background rejection demonstrated in this range of signal efficiency shows the remarkable efficacy of the GNN model in maintaining a low false positive rate while accurately classifying a high proportion of signal jets.

5.4 Jet View

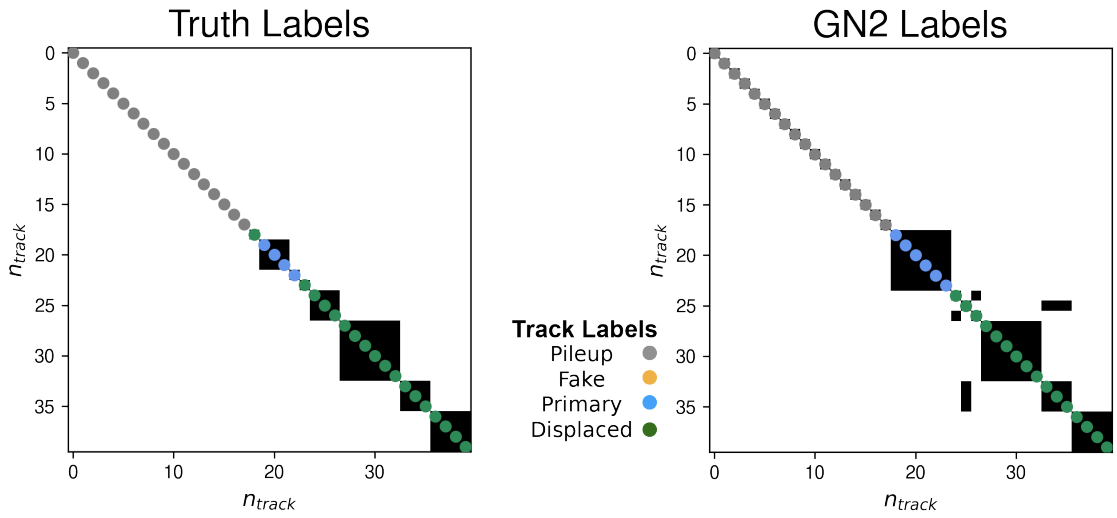


Figure 5.8: Jet view showing the topology of the jet with true labels (left) and predicted labels (right).

This visual representation demonstrates the outcomes of the classification of tracks and vertex predictions in signal samples containing EJs, with a focus on the jet's structural components, particularly the tracks. The illustration integrates vertex and track predictions for a single jet (right of Figure 5.8) from the GNNs and contrasts them with 'truth' labels from the simulated sample (left of Figure 5.8).

Every within the representation is a 40×40 matrix, representing the 40 tracks in a specific jet, sorted by the *truthVertexLabel*. This is an integer label assigned based on truth values from the simulation sample, ensuring that tracks from a common vertex share the same index as discussed in Section 4.4. When $i \neq j$, a matrix element $m_{ij} = 1$ is indicated by a black box, signifying that the pair of tracks share a common vertex; otherwise, $m_{ij} = 0$, represented

by a white box. For diagonal elements where $i = j$, m_{ij} corresponds to *trackOriginLabels*, depicted by colored circles by the label above, denoting the classification of the tracks.

5.5 GNN Validation

To assess whether GNN models are robust for distinguishing between background and signal, it's essential to validate them against the real ATLAS dataset, especially since the models were developed using MC samples.

For accurate validation, the selected region must be orthogonal to the signal-rich region and should predominantly contain background events. In this context, a single muon region is chosen for validation due to its orthogonality with the signal and high background content. This region is populated with events that contain charm, light, bottom jets, and a W boson in each of them, and it also requires more than 2 jets in the event. The full 2022 ATLAS data is considered for validation.

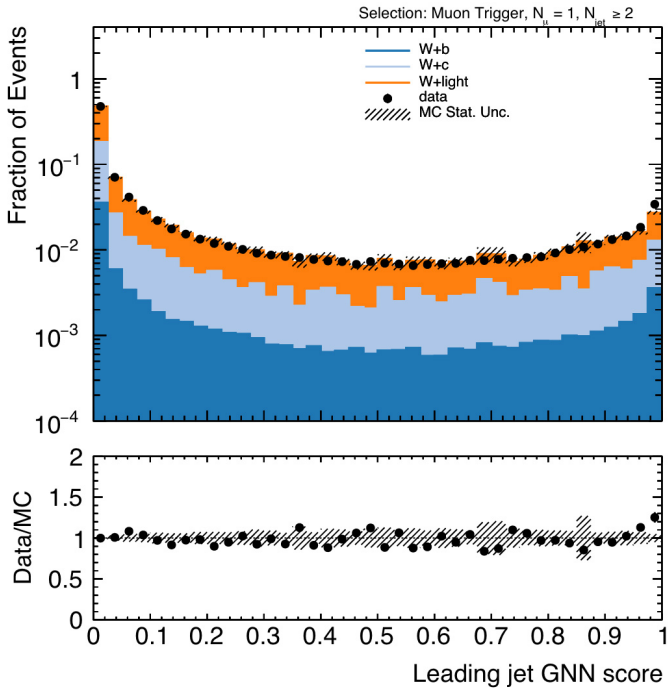


Figure 5.9: GNN model validation with 2022 data.

Figure 5.9 indicates that for most parts of the distribution, the data and MC are in good agreement. However, there's a noticeable difference in the last bin, hinting at possible

systematic uncertainties that warrant further investigation and adjustment of the model to account for these differences. This suggests that the GNN model, initially built on simulated datasets, holds promise in dealing with actual data, retaining its precision and reliability. The agreement between real and simulated data confirms the employability of the GNN model to differentiate between signal and background in actual ATLAS datasets.

5.6 GNN Score in EJ Analysis

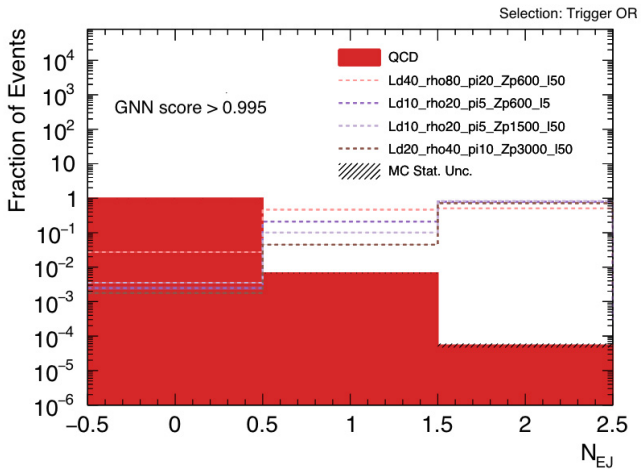


Figure 5.10: Requiring that the jets have GNN score 0.995 and that there are two jets in an event leads to significant background suppression and can be seen in the last bin.

After establishing the GNN model's ability to accurately distinguish between signal and background jets, and its usability in the ATLAS dataset, the next step is to make use of it to separate signal events from background events. One of the ways it can be done, for example, is to require that an event contains 2 jets with a GNN score greater than 0.995 as shown in the Figure 5.10. The last bin of this figure, where the shaded bins represent background events and non-shaded bins represent signal events, shows that the use of the GNN score can help in extremely high background suppression.

While the demonstrates the ability to distinguish signal events, the emerging jets analysis requires a method to estimate the number of background events. EJ signature is an LLP particle decay signature and therefore requires a data-driven method for background estimation.

ABCD Method

The ABCD method is a data-driven approach to estimating background contributions when the background includes a range of possibilities, ranging from QCD jets to other non-standard sources. This method involves dividing a $2D$ plane, identified by two independent variables, into four distinct regions: A - signal region, and B, C, D - control regions. By counting events in regions B, C, and D, one can predict the expected background in the signal region using the formula $N_A = \left(\frac{N_D}{N_C}\right) \cdot N_B$. It's crucial to note that this method assumes that the selected variables are independent and that events are uniformly populated across the regions.

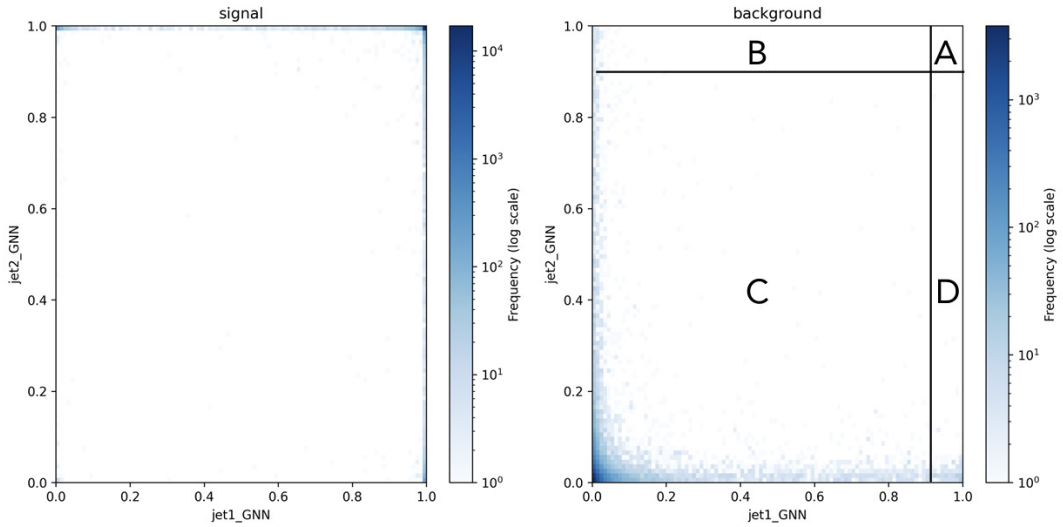


Figure 5.11: Creating ABCD plane. The left plot is populated by signal events and the right plot is populated by background events.

In our approach to the ABCD method, the current effort is on selecting two independent variables that are effective in separating the signal from the background, creating areas that are primarily filled with signal events. Figure 5.11 shows one idea we're exploring: using the GNN scores of the two leading jets. The demonstrates the creation of an ABCD plane, wherein region **A** is characterized by both GNN jet scores exceeding the threshold. Using a threshold of GNN score > 0.95, the ABCD method tested on simulated samples gives a

background event number estimate of 4.77 ± 0.22 in the signal region, where the observed number of background events was 2 ± 1.7.

However, before progressing with the implementation of ABCD planes as a reliable framework for background estimation, it is crucial to verify that the two variables employed are independent and uncorrelated. This aspect is under active investigation at the time of composing this thesis. Nonetheless, it is worth noting that the utilization of GNN-based jet tagging scores plays a pivotal role in this analysis.

Chapter 6

Conclusion

Somewhere, something incredible is waiting to be known.

– Carl Sagan

Extensions to the Standard Model predict the presence of a dark sector, which couples weakly with the Standard Model. This sector could produce unstable dark states, leading to the production of LLPs. These LLPs may generate atypical signatures in detectors, presenting significant challenges for discovery efforts. As a result, they might have remained undetected so far.

This thesis presents an approach for detecting these signatures of LLPs: Emerging Jets using GNNs. These signatures have an intricate topology consisting of displaced tracks and containing many displaced vertices within the jet cone. GNNs by their nature, are well-suited for identifying patterns and relationships within data that traditional analysis methods might overlook. In this thesis, a particle physics dataset is cast as a jet-graph representation, where each node of a graph represents a track, and then GNNs are used to aggregate information from the neighboring nodes, enabling them to create node embedding that captures the relational context, which then is used for different physics object predictions.

Utilizing this approach, GNNs have demonstrated high precision in tagging emerging jets, while effectively rejecting the QCD background. This development is crucial in the search for BSM physics, as it equips us with the capability to detect the signatures of long-lived BSM particles.

Furthermore, the GNN architecture has proven efficient in classifying displaced tracks as well as the identification of displaced vertices within the jet cone. In Section Section 5.2, the efficiency of the GNN architecture in classifying displaced tracks and identifying displaced vertices within the jet cone is demonstrated. This effectiveness is particularly important in the study of potential new physics, which needs the precise identification of the origins of tracks associated with emerging jets. Similarly, Section 5.1 showcases the GNN algorithm's proficiency in identifying a majority of the authentic vertices that share common sets of tracks. The application of GNNs in vertex identification can be important for investigating the properties of dark particles, which are hypothesized to lead to emerging jet signatures.

The model developed in this research is set for deployment in the ATLAS experiment as part of the Run-3 EJ analysis search. Its primary role will be to assist in the search for LLP signatures originating from the dark sector. Moreover, the approaches described in this thesis can be applied to construct additional GNN-based models tailored for various topologies beyond emerging jets. These model's ability to distinguish complex signal signatures can be instrumental in directing future searches at the LHC, potentially leading to significant discoveries.

Bibliography

[1] Wikipedia Contributors. *Standard Model - Wikipedia, the free encyclopedia*. https://en.wikipedia.org/wiki/Standard_Model. Accessed: 2024-04-25. 2022.

[2] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. URL: https://doi.org/10.1016%2Fj.physletb.2012.08.020.

[3] ATLAS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. URL: https://doi.org/10.1016%2Fj.physletb.2012.08.021.

[4] Ringaile Placakyte. *Parton Distribution Functions*. 2011. arXiv: 1111.5452 [hep-ph].

[5] Thomas Schörner-Sadenius, ed. *The Large Hadron Collider*. Berlin: Springer, 2015. ISBN: 9783319150000, 9783319150017. DOI: 10.1007/978-3-319-150017. URL: http://www.springer.com/gb/book/9783319150017.

[6] Torbjörn Sjöstrand et al. "An Introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191 (2015), pp. 159–177. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2015.01.024. URL: https://www.sciencedirect.com/science/article/pii/S0010465515000309.

[7] David J. E. Callaway. "Triviality Pursuit: Can Elementary Scalar Particles Exist?" In: *Phys. Rept.* 167 (1988), p. 241. DOI: 10.1016/0370-1573(88)90008-7.

[8] T. S. van Albada et al. "Distribution of dark matter in the spiral galaxy NGC 3198." In: 295 (Aug. 1985), pp. 305–313. DOI: 10.1086/163375.

[9] Richard Massey, Thomas Kitching, and Johan Richard. "The dark matter of gravitational lensing". In: *Reports on Progress in Physics* 73.8 (July 2010), p. 086901. DOI: 10.1088/0034-4885/73/8/086901. URL: https://doi.org/10.1088%2F0034-4885%2F73%2F8%2F086901.

[10] K. Freese. "Review of Observational Evidence for Dark Matter in the Universe and in upcoming searches for Dark Stars". In: *EAS Publications Series* 36 (2009), pp. 113–126. DOI: 10.1051/eas/0936016. URL: https://doi.org/10.1051%2Feas%2F0936016.

[11] Matthew J. Strassler and Kathryn M. Zurek. "Echoes of a hidden valley at hadron colliders". In: *Physics Letters B* 651.5-6 (Aug. 2007), pp. 374–379. DOI: 10.1016/j.physletb.2007.06.055. URL: https://doi.org/10.1016%2Fj.physletb.2007.06.055.

[12] R. Foot and S. Vagnozzi. "Dissipative hidden sector dark matter". In: *Physical Review D* 91.2 (Jan. 2015). DOI: 10.1103/physrevd.91.023512. URL: https://doi.org/10.1103%2Fphysrevd.91.023512.

[13] Jim Alexander et al. *Dark Sectors 2016 Workshop: Community Report*. 2016. arXiv: 1608.08632 [hep-ph].

[14] Giorgio Arcadi, Abdelhak Djouadi, and Martti Raidal. "Dark Matter through the Higgs portal". In: *Physics Reports* 842 (Feb. 2020), pp. 1–180. DOI: 10.1016/j.physrep.2019.11.003. URL: https://doi.org/10.1016%2Fj.physrep.2019.11.003.

[15] Brian Batell, Maxim Pospelov, and Adam Ritz. "Exploring portals to a hidden sector through fixed targets". In: *Physical Review D* 80.9 (Nov. 2009). DOI: 10.1103/physrevd.80.095024. URL: https://doi.org/10.1103%2Fphysrevd.80.095024.

[16] Yasunori Nomura and Jesse Thaler. "Dark matter through the axion portal". In: *Physical Review D* 79.7 (Apr. 2009). DOI: 10.1103/physrevd.79.075008. URL: https://doi.org/10.1103%2Fphysrevd.79.075008.

[17] Pedro Schwaller, Daniel Stolarski, and Andreas Weiler. "Emerging jets". In: *Journal of High Energy Physics* 2015.5 (May 2015). DOI: 10.1007/jhep05(2015)059. URL: https://doi.org/10.1007%2Fjhep05%282015%29059.

[18] Guillaume Albouy et al. "Theory, phenomenology, and experimental avenues for dark showers: a Snowmass 2021 report". In: *The European Physical Journal C* 82.12 (Dec. 2022). DOI: 10.1140/epjc/s10052-022-11048-8. URL: https://doi.org/10.1140%2Fepjc%2Fs10052-022-11048-8.

[19] Timothy Cohen, Mariangela Lisanti, and Hou Keong Lou. "Semivisible Jets: Dark Matter Undercover at the LHC". In: *Physical Review Letters* 115.17 (Oct. 2015). DOI: 10.1103/physrevlett.115.171804. URL: https://doi.org/10.1103%2Fphysrevlett.115.171804.

[20] Lawrence Lee et al. "Collider searches for long-lived particles beyond the Standard Model". In: *Progress in Particle and Nuclear Physics* 106 (May 2019), pp. 210–255. DOI: 10.1016/j.ppnp.2019.02.006. URL: https://doi.org/10.1016%2Fj.ppnp.2019.02.006.

[21] Heather Russel. *LLP Overview*. URL: https://hrussell.web.cern.ch/hrussell/images/graphics/LLP_overview.png.

[22] Jackson Carl Burzynski. "A Search for Exotic Higgs Decays or: How I Learned to Stop Worrying and Love Long-lived Particles"". Simon Fraser University (CA), 2021. URL: https://cds.cern.ch/record/2780737 (visited on 09/26/2023).

[23] Emily Anne Thompson. "Search for Long-Lived Supersymmetric Particles Using Displaced Vertices with the ATLAS Detector at the LHC". In: ().

[24] ATLAS Collaboration. *Cutway view of ATLAS detector*. CERN. 2020. URL: https://cds.cern.ch/images/ATLAS-PHOTO-2020-018-1.

[25] Kate Panchel. *Hunting for new physics using long-lived particles*. URL: https://indico.cern.ch/event/954239/contributions/4009464/.

[26] Rudolph Emil Kalman et al. "A new approach to linear filtering and prediction problems". In: *Journal of basic Engineering* 82.1 (1960), pp. 35–45.

[27] ATLAS Collaboration. *Performance of the reconstruction of large impact parameter tracks in the inner detector of ATLAS*. 2023. arXiv: 2304.12867 [hep-ex].

[28] Paolo Iengo. *The Atlas Calorimeters*. URL: https://people.na.infn.it/~elly/TesiAtlas/SeminariAtlas/Detector/IengoAtlasCalo.pdf.

[29] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. "The anti-ktjet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/04/063. URL: http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[30] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. "The anti-ktjet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/04/063. URL: http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[31] Michael M. Bronstein et al. *Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges*. 2021. arXiv: 2104.13478 [cs.LG].

[32] Benjamin Sanchez-Lengeling et al. "A Gentle Introduction to Graph Neural Networks". In: *Distill* (2021). https://distill.pub/2021/gnn-intro. DOI: 10.23915/distill.00033.

[33] Manzil Zaheer et al. *Deep Sets*. Apr. 14, 2018. arXiv: 1703.06114 [cs, stat]. URL: http://arxiv.org/abs/1703.06114 (visited on 12/11/2022). preprint.

[34] Ashish Vaswani et al. *Attention Is All You Need*. Dec. 5, 2017. DOI: 10.48550/arXiv.1706.03762. arXiv: 1706.03762 [cs]. URL: http://arxiv.org/abs/1706.03762 (visited on 05/17/2023). preprint.

[35] Peter W. Battaglia et al. *Relational inductive biases, deep learning, and graph networks*. 2018. arXiv: 1806.01261 [cs.LG].

[36] Felix Wu et al. *Simplifying Graph Convolutional Networks*. 2019. arXiv: 1902.07153 [cs.LG].

[37] Thomas N. Kipf and Max Welling. *Semi-Supervised Classification with Graph Convolutional Networks*. 2017. arXiv: 1609.02907 [cs.LG].

[38] Petar Veličković et al. *Graph Attention Networks*. Feb. 4, 2018. arXiv: 1710.10903 [cs, stat]. URL: http://arxiv.org/abs/1710.10903 (visited on 06/06/2023). preprint.

[39] Jiani Zhang et al. *GaAN: Gated Attention Networks for Learning on Large and Spatiotemporal Graphs*. 2018. arXiv: 1803.07294 [cs.LG].

[40] Federico Monti et al. *Geometric deep learning on graphs and manifolds using mixture model CNNs*. 2016. arXiv: 1611.08402 [cs.CV].

[41] Justin Gilmer et al. *Neural Message Passing for Quantum Chemistry*. 2017. arXiv: 1704.01212 [cs.LG].

[42] Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. "Graph Neural Networks in Particle Physics". In: *Machine Learning: Science and Technology* 2.2 (Jan. 8, 2021), p. 021001. ISSN: 2632-2153. DOI: 10.1088/2632-2153/abbf9a. URL: https://iopscience.iop.org/article/10.1088/2632-2153/abbf9a (visited on 07/08/2022).

[43] Alexander Radovic et al. “Machine Learning at the Energy and Intensity Frontiers of Particle Physics”. In: *Nature* 560.7716 (Aug. 1, 2018), pp. 41–48. ISSN: 1476-4687. DOI: 10.1038/s41586-018-0361-2. URL: https://doi.org/10.1038/s41586-018-0361-2.

[44] ATLAS Collaboration. “Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector”. In: (July 2017).

[45] ATLAS Collaboration. *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2255226.

[46] ATLAS Collaboration. “Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS”. In: (2020).

[47] Isaac Henrion et al. “Neural Message Passing for Jet Physics”. In: 2017. URL: https://api.semanticscholar.org/CorpusID:39724044.

[48] Steven Farrell et al. *Novel deep learning methods for track reconstruction*. 2018. arXiv: 1810.06111 [hep-ex].

[49] Hadar Serviansky et al. *Set2Graph: Learning Graphs From Sets*. 2020. arXiv: 2002.08772 [cs.LG].

[50] J. Arjona Martínez et al. “Pileup mitigation at the Large Hadron Collider with Graph Neural Networks”. In: *Eur. Phys. J. Plus* 134.7 (2019). 12 pages, 11 figures, p. 333. DOI: 10.1140/epjp/i2019-12710-3. arXiv: 1810.07988. URL: https://cds.cern.ch/record/2646050.

[51] Murat Abdughani et al. “Probing stop pair production at the LHC with graph neural networks”. In: *Journal of High Energy Physics* 2019.8 (Aug. 2019). DOI: 10.1007/jhep08(2019)055. URL: https://doi.org/10.1007%2Fjhep08%282019%29055.

[52] ATLAS Collaboration. *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*. Tech. rep. Geneva: CERN, 2022. URL: https://cds.cern.ch/record/2811135.

[53] Shaked Brody, Uri Alon, and Eran Yahav. *How Attentive are Graph Attention Networks?* 2022. arXiv: 2105.14491 [cs.LG].

[54] Yujia Li et al. *Gated Graph Sequence Neural Networks*. 2017. arXiv: 1511.05493 [cs.LG].

[55] Shaked Brody, Uri Alon, and Eran Yahav. *On the Expressivity Role of LayerNorm in Transformers' Attention*. 2023. arXiv: 2305.02582 [cs.LG].

[56] Leslie N. Smith. *A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay*. 2018. arXiv: 1803.09820 [cs.LG].