



IRC-Safe Graph Autoencoder for Unsupervised Anomaly Detection

Oliver Atkinson¹, Akanksha Bhardwaj^{1*}, Christoph Englert¹, Partha Konar², Vishal S. Ngairangbam^{2,3} and Michael Spannowsky⁴

¹ School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom, ² Theoretical Physics Division, Physical Research Laboratory, Shree Pannalal Patel Marg, Ahmedabad, India, ³ Discipline of Physics, Indian Institute of Technology, Palaj, India, ⁴ Department of Physics, Institute for Particle Physics Phenomenology, Durham University, Durham, United Kingdom

Anomaly detection through employing machine learning techniques has emerged as a novel powerful tool in the search for new physics beyond the Standard Model. Historically similar to the development of jet observables, theoretical consistency has not always assumed a central role in the fast development of algorithms and neural network architectures. In this work, we construct an infrared and collinear safe autoencoder based on graph neural networks by employing energy-weighted message passing. We demonstrate that whilst this approach has theoretically favorable properties, it also exhibits formidable sensitivity to non-QCD structures.

OPEN ACCESS

Edited by:

Caterina Doglioni,
Lund University, Sweden

Reviewed by:

Baptiste Ravina,
University of Göttingen, Germany
Nadezda Chernyavskaya,
European Organization for Nuclear
Research (CERN), Switzerland

*Correspondence:

Akanksha Bhardwaj
akanksha.bhardwaj@glasgow.ac.uk

Specialty section:

This article was submitted to
Big Data and AI in High Energy
Physics,
a section of the journal
Frontiers in Artificial Intelligence

Received: 13 May 2022

Accepted: 23 June 2022

Published: 22 July 2022

Citation:

Atkinson O, Bhardwaj A, Englert C,
Konar P, Ngairangbam VS and
Spannowsky M (2022) IRC-Safe
Graph Autoencoder for Unsupervised
Anomaly Detection.
Front. Artif. Intell. 5:943135.
doi: 10.3389/frai.2022.943135

Keywords: anomaly detection, graph neural network, high energy physics, IRC safety, anomalous jets

1. INTRODUCTION

New physics searches at the high-energy frontier of the Large Hadron Collider (LHC) have so far not resulted in any significant deviation of experimental results from the Standard Model (SM) expectation. However, with a growing dataset of these high energy measurements, the pressure mounts for theoretically motivated scenarios of beyond the SM (BSM) physics which have been devised to tackle known shortcomings of the SM. So what are the ways out of this juxtaposition of experimental agreement with the SM and its failure to describe established physics at small and large distances?

On the one hand, there is an increasing emphasis on theoretically as-model-independent-as-possible approaches based on effective field theory (EFT) (Weinberg, 1979). EFT navigates QFT correlations away from the SM prediction in any possible direction given the SM symmetry and particle content, thus avoiding UV model biases. Alas, such an approach poses its own challenges: looking for deviations from the SM expectation along these lines involves many *ad-hoc* interactions. Concrete models will typically only source a subset of relevant interactions, e.g., (Englert et al., 2019; Das Bakshi et al., 2021; Bakshi et al., 2022). There has been great progress to facilitate matching calculations (Carmona et al., 2021), however, depending on the new physics scenario, this can create a significant overhead that must be included in the parameter fitting procedure itself (Freitas et al., 2016; Englert et al., 2020).

On the other hand, we can look for phenomenological deviations from specific SM signatures directly in collider results without any new physics bias. Under the assumption that collider data can be modeled sufficiently adequately, we can employ the SM expectation to identify regions where measurements do not follow the SM expectation. This anomaly detection has emerged as a powerful tool to look for any hidden signature of new physics in the data. Recently, a range of state-of-the-art

methods for anomaly detection (Collins et al., 2018, 2021; Aaboud et al., 2019; Blance et al., 2019; De Simone and Jacques, 2019; Blance and Spannowsky, 2020; Cheng et al., 2020; Hajer et al., 2020; Nachman, 2020; Nachman and Shih, 2020; Araz and Spannowsky, 2021; Atkinson et al., 2021b; Hallin et al., 2021; Canelli et al., 2022) using deep learning have been designed.

Theoretical consistency when confronting collider data with theoretical expectations is pivotal. The formulation of infrared and collinear (IRC)-safe observables is necessary to guarantee the comparability of experimental measurements and theoretical predictions to all orders in perturbation theory employing the Kinoshita-Lee-Nauenberg (KLN) theorem (Kinoshita, 1962; Lee and Nauenberg, 1964) and collinear factorization (Collins et al., 1989) of parton distributions. Any sensitivity enhancement observed by algorithms that inadvertently employ IRC-unsafe information will be critically assessed in subsequent studies, with a potentially meaningless theoretical outcome. This is a tedious task all too familiar from the use of IRC-unsafe jet clustering algorithms (e.g., iterative cone algorithms) by the CDF and D0 experiments (Kilgore and Giele, 1997), which was later only partially addressed with the midpoint algorithm during Tevatron Run II, until fully IRC-safe algorithms (Catani et al., 1993; Cacciari et al., 2008) were established as the only theoretically meaningful community consensus.

It seems prudent to avoid mistakes of the past: in this paper, we devise an IRC-safe Graph Neural Network (GNN) autoencoder algorithm, employing an Energy-Weighted Message-Passing Network (EMPN) (Konar et al., 2022) for unsupervised anomaly detection. While the IRC-safe loss function is the primary observable of our autoencoder, we also study the latent space (graph) representation structure as a motivating tool for new physics discrimination (Atkinson et al., 2021b; Dillon et al., 2021) and highlight the relations to known and more “traditional” IRC-safe observables. This paper is organized as follows: in Section 2, we outline our EMPN approach and detail our IRC-safe graph construction before we introduce the IRC-safe autoencoder architecture and simulation framework in Section 3. Section 4 is devoted to the discussion of the sensitivity performance of the autoencoder; we also highlight the correlation of sensitivity with more traditional jet-based observables. We conclude in Section 5.

2. A BRIEF OUTLINE OF ENERGY-WEIGHTED MESSAGE PASSING ALGORITHM

This section presents a brief overview of the IRC safe Energy-weighted Message passing algorithm (Konar et al., 2022). It generalizes Energy Flow Networks (Komiske et al., 2019; Dolan and Ore, 2021), an IRC safe feature extraction on point clouds, by learning relational information between two elements (nodes) by constructing a graph out of the point cloud. This procedure is similar to message-passing networks like the Dynamic Graph Convolutional Neural Network (DGCNN) (Wang et al., 2019) that extract local features beyond the global feature extraction *via* point-cloud-based architectures such as deep-sets (Zaheer et al.,

2017) and PointNet (Charles et al., 2017; Qi et al., 2017). The algorithm consists of two necessary ingredients:

- An IRC safe prescription for constructing graphs which guarantees that the graph is invariant under soft and collinear splittings;
- An energy-weighted summed aggregation of messages (and node features after the final message-passing layer) taking the directional (unit vectors or angles) inputs \hat{p}_i and \hat{p}_j of the nodes connected by an edge (j, i) at the initial layer.

In the following, we discuss these two elements separately.

2.1. IRC Safe Graph Construction

The inductive biases that a message-passing algorithm imposes on its input data are highly dependent on the graph structure. For instance, the neighborhood sets (the set of particles a node is connected with) determine the local connectivity of the nodes. Thus, the graph construction algorithm from a point cloud gives a strong indication that a graph neural network is the best avenue to pursue. The IRC safety of a message-passing algorithm also depends on the graph construction, and we highlight such an IRC safe graph construction algorithm in this section.

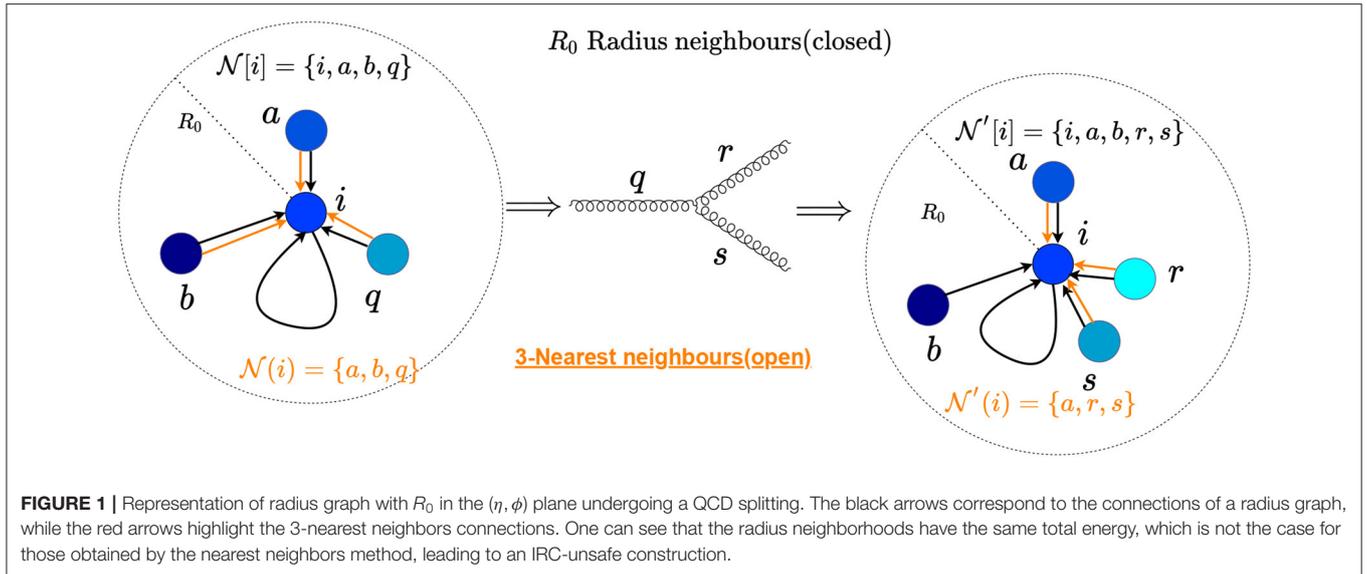
Let $\mathcal{S} = \{p_1, p_2, p_3, \dots, p_N\}$ be the set of four-vectors of the particles within a jet, while $\mathcal{S}' = \{p_1, p_2, \dots, p_{N+1}\}$ is the same set in the presence of an additional splitting. The collinear limit is when the emitted particles r and s with the angular separation Δ_{rs} tending to zero, while the soft limit refers to the case when one of the particle’s energy tends to zero. These four-vectors can be written as

$$p_i = (z_i, \hat{p}_i), \text{ with } z_i = \frac{p_T^i}{\sum_{j \in \mathcal{S}} p_T^j} \text{ and } \hat{p} = (\eta, \phi), \quad (1)$$

for hadron colliders, with the separation in the $\eta - \phi$ plane between two particles i and j , defined as $\Delta R_{ij} = \sqrt{\Delta \eta_{ij}^2 + \Delta \phi_{ij}^2}$ denoting the quantity analogous to Δ_{ij} . Since we will be taking directed edges, the neighborhood set of a node i will be the set of all nodes with incoming connections to i . For all particles i in \mathcal{S} or \mathcal{S}' , a graph construction algorithm will construct neighborhood sets $\mathcal{N}[i]$ and $\mathcal{N}'[i]$, respectively. We will use a “closed” neighborhood with $i \in \mathcal{N}[i]$ instead of an “open” neighborhood $i \notin \mathcal{N}(i)$, since the second choice will always be IRC unsafe when the node i splits. To illustrate this, we show the radius graph with R_0 in the (η, ϕ) plane in **Figure 1**, where the node q undergoes a splitting. The black arrows highlight the connections of the radius graph. **Figure 1** also demonstrates a nearest neighborhood connection as an example of an IRC unsafe graph construction.

To formalize the graph construction algorithm in terms of the four-vectors of the particles, we define a decision function $\mathbf{D}(p_i, p_j)$ and a threshold function $\mathbf{T}(p_i, p_j)$, such that any particle j with four-vector p_j will be assigned to the neighborhood of particle i with four-vector p_i if $\mathbf{D}(p_i, p_j)$ is less-than or equal-to $\mathbf{T}(p_i, p_j)$. This can be summarized as

$$\mathbf{D}(p_i, p_j) \leq \mathbf{T}(p_i, p_j) \implies j \in \mathcal{N}[i]. \quad (2)$$



Since we are interested in the soft and collinear limits, constructing an IRC safe graph requires putting conditions on these functions in the respective kinematical configurations.

The required condition on these functions for a “parent” splitting $q \rightarrow r + s$ when the “daughters” r, s become collinear is

$$\begin{aligned}
 \mathbf{D}(p_i, p_r + p_s) \leq \mathbf{T}(p_i, p_r + p_s) &\Leftrightarrow \mathbf{D}(p_i, p_r) \leq \mathbf{T}(p_i, p_r) \wedge \\
 &\mathbf{D}(p_i, p_s) \leq \mathbf{T}(p_i, p_s), \\
 \mathbf{D}(p_r + p_s, p_i) \leq \mathbf{T}(p_r + p_s, p_i) &\Leftrightarrow \mathbf{D}(p_r, p_i) \leq \mathbf{T}(p_r, p_i) \wedge \\
 &\mathbf{D}(p_s, p_i) \leq \mathbf{T}(p_s, p_i),
 \end{aligned} \tag{3}$$

where the second condition arises since the nodes q, r or s can also be the node whose neighborhood is being determined. The only requirement in the IR limit for a daughter particle is that all the particles in the set $\mathcal{N}[i]$ are also present in $\mathcal{N}'[i]$, with the only potential addition of a soft particle. This is guaranteed by the form of Equation (2), since both functions depend only on the four-vector of the two nodes of interest¹. The conditions (c.f. Equation 3) are satisfied in the collinear limit $\Delta_{rs} \rightarrow 0$ if

$$\mathbf{D} = \mathbf{D}(\hat{p}_i, \hat{p}_j), \quad \mathbf{T} = \mathbf{T}(\hat{p}_i, \hat{p}_j), \tag{4}$$

employing the definitions Equation (1). Therefore, graphs formed by connecting particles within a constant radius R_0 in the $\eta - \phi$ plane are IRC safe when the decision and threshold functions take the form

$$\mathbf{D} = \Delta R_{ij}, \quad \mathbf{T} = R_0. \tag{5}$$

Note that these choices of \mathbf{D}, \mathbf{T} yield closed neighborhoods without additional requirements. We will use these graphs in the remainder of this paper; the neighborhood of a particle of such a radius graph is shown in **Figure 1**.

¹This is not the case for popular graph construction algorithms like k -nearest neighbors, for which the decision and threshold has a complicated dependence on the distance of the primary node i with every other particle in the graph, and on the number of elements in the neighborhood set.

2.2. Energy-Weighted Message Passing

We detail the IRC safe message passing operation in this section. Before doing so, we summarize the general definition of message passing operation in the following steps. The first step, the message-passing stage, involves calculating the messages for all edges present in the graph. The message function, parameterized as a multilayer perceptron shared for all edges, takes the node features of the two nodes connected by an edge and evaluates the message. Since the message function does not need to be symmetric for the two node features, a direction convention is necessary for the second phase. In our convention, the message originates from all nodes in the neighborhood $\mathcal{N}[i]$ and flows toward the particle i . The second step, the node-readout stage, updates the node features of each node in the graph as a permutation-invariant function of all incoming messages.

IRC safety of the updated node features after a message-passing operation is crucially dependent on the nature of the node readout. A readout based on the maximum or minimum value of the node features depends on a single node feature in the neighborhood, and a soft or collinear splitting of this particular node would render the updated node feature IRC-unsafe. This is ultimately related to identifying a specific node in the neighborhood as special², which impedes KLN cancellations. A mean readout, on the other hand, explicitly depends on the cardinality of the neighborhood sets $\mathcal{N}[i]$ which is not a well-defined QCD quantity either since there can be an arbitrary but finite amount of resolvable emissions in the enhanced collinear or soft regions of phase space. Thus we use a summed readout, which will inclusively take all the particles in the neighborhood into account and will not explicitly depend on their size.

An IRC safe graph construction algorithm ensures two things: the equality of the sum of energy (transverse energy in the case of hadron colliders) of all particles in either neighborhood sets

²This is also the reason for using closed neighborhoods $\mathcal{N}[i]$, as an open neighborhood $\mathcal{N}(i)$, would give a special status to the node i .

and the presence of both collinear daughters in $\mathcal{N}'[i]$ if the parent is present in $\mathcal{N}[i]$. Defining a scope-dependent energy weight-factors analogous to z_i as

$$\omega_j^{(\mathcal{K})} = \frac{p_T^j}{\sum_{k \in \mathcal{K}} p_T^k},$$

with \mathcal{K} denoting the set of particles in the particular readout operation, any message passing of the form

$$\mathbf{h}_i^{(l+1)} = \sum_{i \in \mathcal{N}[i]} \omega_j^{(\mathcal{N}[i])} \hat{\Phi}^{(l)}(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}), \quad (6)$$

with $\mathbf{h}_i^{(0)} = \hat{p}_i$ and $\mathbf{h}_i^{(l)}$ denoting the updated node-features after l message-passing operations satisfies IRC safety; in the infrared limit, it is straightforward to see that any soft particle with $z_r \rightarrow 0 \implies \omega_r^{(\mathcal{N}[i])} \rightarrow 0$ for any node i . The splitting $q \rightarrow r + s$ for IRC-safe graphs therefore yields

$$\omega_q^{(\mathcal{N}[i])} = \omega_r^{(\mathcal{N}[i])} + \omega_s^{(\mathcal{N}[i])}. \quad (7)$$

In the collinear limit with $\hat{p}_q = \hat{p}_r = \hat{p}_s$ we have $\hat{\Phi}^{(0)}(\hat{p}_i, \hat{p}_q) = \hat{\Phi}^{(0)}(\hat{p}_i, \hat{p}_r) = \hat{\Phi}^{(0)}(\hat{p}_i, \hat{p}_s)$. Combining this with Equation (7), we obtain (for $l = 0$)

$$\omega_q^{(\mathcal{N}[i])} \hat{\Phi}^{(0)}(\hat{p}_i, \hat{p}_q) = \omega_r^{(\mathcal{N}[i])} \hat{\Phi}^{(0)}(\hat{p}_i, \hat{p}_r) + \omega_s^{(\mathcal{N}[i])} \hat{\Phi}^{(0)}(\hat{p}_i, \hat{p}_s).$$

When evaluating Equation (6) for the neighborhood of a node i , the terms on the RHS and LHS of this expression are the only ones which will not be common between $\mathcal{N}[i]$ and $\mathcal{N}'[i]$, due to the IRC safe graph construction. The same expression is followed when $i = q$ on the left, and $i = r$ or $i = s$ on the right, since $\{r, s\} \subset \mathcal{N}'[s]$ and $\{r, s\} \subset \mathcal{N}'[r]$, with all three neighborhoods (including $\mathcal{N}[q]$) containing the same particles except for q , r , and s . Therefore, from Equation (6), we have $\mathbf{h}_q^{(1)} = \mathbf{h}_r^{(1)} = \mathbf{h}_s^{(1)}$ for collinear splittings. On the other hand, for a soft daughter, say r , we have $\mathbf{h}_q^{(1)} = \mathbf{h}_s^{(1)}$, but $\mathbf{h}_r^{(1)} \neq \mathbf{h}_q^{(1)}$, with $\mathbf{h}_r^{(1)}$ not necessarily zero. The presence of the node features of the daughter particles, even in the soft or collinear limit, impedes an IRC safe examination of the full jet graph unless observables are specifically designed to be insensitive to their presence in the IRC limit. The procedures to take care of these additional nodes are explained in the following sections, which are different for supervised and unsupervised methods. Since the above derivation used the collinearity of q , r , and s , for IRC safe neighborhoods, for the same neighborhoods and any successive application of an energy-weighted message passing of the form Equation (6), we have $\mathbf{h}_q^{(l)} = \mathbf{h}_r^{(l)} = \mathbf{h}_s^{(l)}$ for any l .

3. IRC-SAFE GRAPH AUTOENCODER

In a supervised machine learning scenario, the IRC-safe graph readout acting on the node features of the final message-passing operation gives an IRC-safe graph representation, and one loses the graph's structure. The graph representation, a fixed-length

vector obtained after applying a permutation invariant function on the node features for any variable-length graph, feeds into the downstream network. Therefore, training a classifier on the loss function defined with the downstream network's output proceeds without any complications from the presence of additional soft or collinear nodes. On the other hand, a graph autoencoder similar to the one proposed in Atkinson et al. (2021a) preserves the graph structure until the output. Therefore, the autoencoder's output graph will have additional nodes in the soft and collinear limits in the case of extra emissions. Since the observable of interest for anomaly detection with an autoencoder is the loss function, we need to ensure its IRC safety. In this section, we first devise an IRC safe loss function and give details of the network architecture and training.

3.1. An IRC-Safe Loss Function

The definition of the loss function involves input which changes with a soft or collinear splitting. Therefore, the loss which is normally used as an observable in anomaly detection, needs to be IRC-safe. A simple IRC-safe loss function for a jet with constituent set \mathcal{G} is of the form

$$\mathcal{L}_{\mathcal{G}} = \sum_{i \in \mathcal{G}} z_i d(\hat{p}_i, \hat{p}_i). \quad (8)$$

The barred quantities are the output of the network, while the unbarred quantities are the inputs to the network. The function $d(\hat{p}_i, \hat{p}_i) \geq d_0$ denotes a well-behaved metric (one-to-one) between the input and the output space, with $d(\hat{p}_i, \hat{p}_i) = d_0$. We now show that this is indeed an IRC safe choice:

Any soft particle s , will not contribute to the sum since $z_s \rightarrow 0$, and hence it is IR safe. For the splitting $q \rightarrow r + s$ we have

$$\begin{aligned} \mathcal{L}_S &= \dots + z_q d(\hat{p}_q, \hat{p}_q) + \dots \\ \mathcal{L}_{S'} &= \dots + z_r d(\hat{p}_r, \hat{p}_r) + z_s d(\hat{p}_s, \hat{p}_s) + \dots \end{aligned}$$

Since, by construction, a GNN's node output after L total message-passing operations $\mathbf{h}_i^{(L)} = \hat{p}_i$, is a function of the input four-vectors $\{p_1, p_2, p_3, \dots, p_N\}$, in general, they can have a very complicated dependence on all the input node features. However, due to the IRC safety of the EMPN, we have

$$\mathcal{L}_{S'} - \mathcal{L}_S = z_r d(\hat{p}_r, \hat{p}_r) + z_s d(\hat{p}_s, \hat{p}_s) - z_q d(\hat{p}_q, \hat{p}_q). \quad (9)$$

In the collinear limit with $\hat{p}_q = \hat{p}_r = \hat{p}_s \implies \hat{p}_q = \hat{p}_r = \hat{p}_s$, we therefore have (since $z_q = z_r + z_s$),

$$z_q d(\hat{p}_q, \hat{p}_q) = z_r d(\hat{p}_r, \hat{p}_r) + z_s d(\hat{p}_s, \hat{p}_s) \implies \mathcal{L}_{S'} - \mathcal{L}_S = 0, \quad (10)$$

i.e., collinear safety. In the following analysis of the EMPN autoencoder we will use mean-squared error between the input and output node features for $d(\hat{p}_i, \hat{p}_i)$.

3.2. Jet Graph Definition

To demonstrate the performance of the described algorithm, we use the publicly available top-tagging dataset of Butter et al.

(2018) and Kasieczka et al. (2019). The dataset contains a training, validation and testing set of 600k, 200k, and 200k QCD jets, respectively. The training and validation are done only with the background QCD samples since the total cross-section of their production would be orders of magnitude higher than most probable signals. Although the dataset has the same number of top jets for each of the three analysis stages, we use the 200k top jets of the test dataset as a benchmark signal scenario. These jets are simulated using Pythia8 (Sjostrand et al., 2008; Sjöstrand et al., 2015) and passed through Delphes3 (de Favereau et al., 2014) for the detector simulations using the default ATLAS parameter card. Jets are clustered from particle flow (Eflow) constituents with a distance parameter $\Delta R = 0.8$ using the anti- k_t algorithm (Cacciari et al., 2008). The transverse momentum of the jets is in the range $p_T \in [550, 650]$ GeV.

Using the constituents of these jets, we construct the radius graphs which serve as the input to the IRC safe graph network. To construct the jet radius graph, we first calculate the inter-particle distance ΔR_{ij} in the (η, ϕ) plane. Next, we define a set of all the particles i as the neighborhood set $\mathcal{N}[i]$ such that $\Delta R_{ij} \leq R_0$, where R_0 is an external tunable parameter. Each node is associated with three node features

$$\mathbf{h}_i^0 = (\Delta\eta_i, \Delta\phi_i, \Delta R_i), \quad (11)$$

where $\Delta\eta_i$, $\Delta\phi_i$, ΔR_i are calculated with respect to the jet axis. For the network analysis, we choose $R_0 = 0.3$. Since the dependence of the classification power on R_0 for the supervised case was found to be mild (Konar et al., 2022), with the AUC values changing in the third decimal value for different values of R_0 between 0.1 and 0.5, we restrict ourselves to a single value in the intermediate range. The final node vectors contain information about the L -hop neighborhood with an effective radius of $R_0 \times L$. On the other hand, the primary region of activity for the one-prong QCD jets used to train the network lies in a relatively small central region of the total jet of radius $\Delta R = 0.8$. Therefore, the features learnt by the autoencoder would be weakly dependent on R_0 , once the effective radius covers a significant portion of the central region.

3.3. Network Architecture and Training

Now that we have described the construction of the jet graphs, we discuss the details of the network architecture and training in this section. Follow from **Figure 2** where we sketch a schematic diagram of an IRC safe graph-autoencoder. The encoder consists of three edge convolution operations with output dimensions of 128, 64 and 2, which is the dimension of the latent representation. Since we take three-dimensional node features, we restrict ourselves to a 2-dimensional latent space (g_1, g_2) to induce an information bottleneck³. The decoder also has three edge convolution operations, with the first two dimensions

³The effective number of inputs to a message function could be twice the number of input features—one each for the two nodes connected by an edge. However, a concrete understanding of the universal approximation properties of graph neural networks (Brüel Gabriësson, 2020) is yet to be achieved, making it difficult to precisely determine the actual input dimensions when looking at the complete graph neural network.

mirroring the encoder network dimensions (excluding the latent dimension). Finally, the last edge convolution operation maps the 128-dimensional node vectors at the penultimate message passing layer to a three-dimensional space to reconstruct the input node features.

We take $\hat{\Phi}^{(l)}$ at each message-passing layer to be a multilayer perceptron (MLP). For an edge convolution operation, we have for two node features $\mathbf{h}_i^{(l)}$ and $\mathbf{h}_j^{(l)}$ connected by an edge in Equation (6),

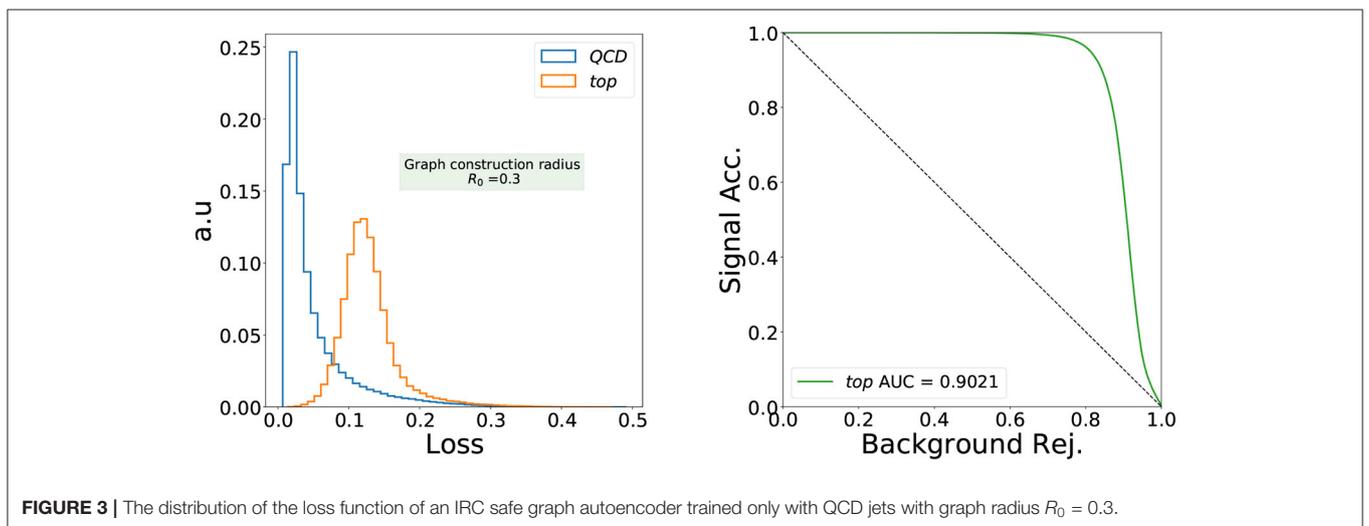
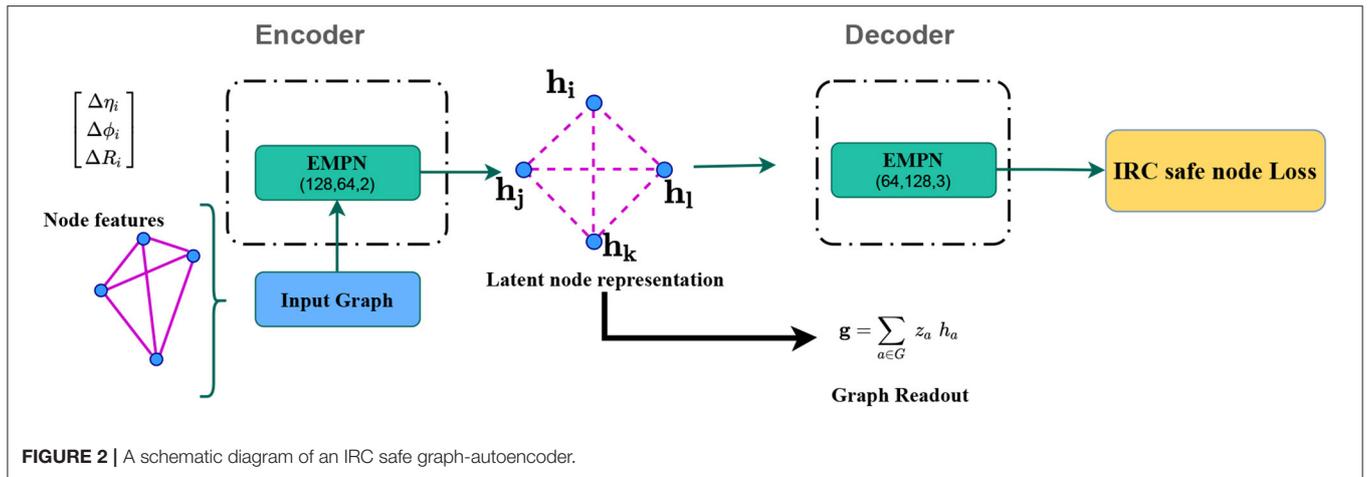
$$\hat{\Phi}^{(l)}(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}) = \hat{\Phi}^{(l)}(\mathbf{h}_i^{(l)} \oplus [\mathbf{h}_j^{(l)} - \mathbf{h}_i^{(l)}]).$$

Therefore the input vector to the MLP has twice the node-feature's dimensions, since the direct sum $\mathbf{h}_i^{(l)} \oplus (\mathbf{h}_j^{(l)} - \mathbf{h}_i^{(l)})$, is a concatenation of the two vector quantities of equal dimensions. The dimension of the MLP's output is the same as the output dimension of the message passing operations and has a linear activation. We fix the MLP to have two hidden layers with ReLU activation and the same number of nodes as the output dimension. The network is implemented using the Pytorch-Geometric (Fey and Lenssen, 2019) package. Note that we have not performed any hyperparameter scan as part of this present, proof-of-concept study. We train the network for fifty epochs with a learning rate of 0.001 using the Adam (Kingma and Ba, 2014) optimiser. The training and validation losses are compared after each epoch to ensure that there is no overfitting or a premature termination of training. The epoch with minimum validation loss is used to infer the anomaly detection on the test dataset.

4. ANOMALY DETECTION PERFORMANCE AND RESULTS

We now discuss the performance of the designed IRC safe loss function in detecting anomalous jets when the network is trained only on the QCD background. We choose boosted top jets from the aforementioned public dataset as our benchmark. In **Figure 3** (left), we show the distribution of the loss function for the QCD and top jets (our inputs are the node features given in Equation 11). As can be seen, the distributions of the loss function values for the QCD and top jets are significantly different, highlighting the capability of the architecture to detect anomalous jets in an IRC-safe way. The Receiver-Operator-Characteristic (ROC) curve and the Area Under the Curve (AUC) of 0.902 shown in **Figure 3** (right) confirm the good separation shown in the loss distribution, rivaling convolutional autoencoders (Heimel et al., 2019; Roy and Vijay, 2019; Farina et al., 2020; Finke et al., 2021) which also have AUCs close to such values [up to 0.93 (Heimel et al., 2019) and 0.91 (Finke et al., 2021)] on the same dataset. Although we did not perform a hyperparameter scan for this study, we observed a decrease in performance for a one-dimensional latent space.

Top jets possess a different and hard kinematical structure that is typically not present in QCD jets. The ability to look



into the soft and collinear splittings from the QCD shower evolution in an IRC safe way enables the network to access such information and the hard radiation pattern in a theoretically meaningful way. Modifications of the soft and collinear radiation patterns that are seeded by novel hard scales (see, e.g., Englert et al., 2011; Soper and Spannowsky, 2011, 2014; Gerwick et al., 2012a,b; Prestel and Spannowsky, 2019 for a more traditional jet-based approach to this) are therefore consistently included in the anomaly detection performance. Therefore, when such non-QCD structures are present, the anomaly detection performance is considerably improved.

In light of these results, it is worthwhile to compare our autoencoder results to phenomenological aspects of QCD in jet substructure analyses. From the point of view of soft and collinear features, Energy Correlation Functions (ECF) (Larkoski et al., 2013) are particularly relevant for such a comparison as we will motivate below. Furthermore, given that our autoencoder

condenses the QCD information into the latent space in an IRC-safe way, it is interesting to see how it correlates with ECF observables. To this end, we define

$$\mathbf{g} = \sum_{a \in \mathcal{G}} z_a \mathbf{h}_a, \quad (12)$$

where \mathbf{h}_a are the latent node features. Similar to the graph readout in a classification scenario (Konar et al., 2022), this is an IRC safe representation of the jet. The distribution of the individual components of the two-dimensional graph representation are shown in **Figure 4**. The good performance of the autoencoder is reflected in the good separation in the latent space. The two latent space directions are, however, completely anti-correlated; see **Figure 5** (they are also highly correlated with the loss). Thus, restricting ourselves to any of these three variables would be sufficient for the anomaly detection problem studied in this work.

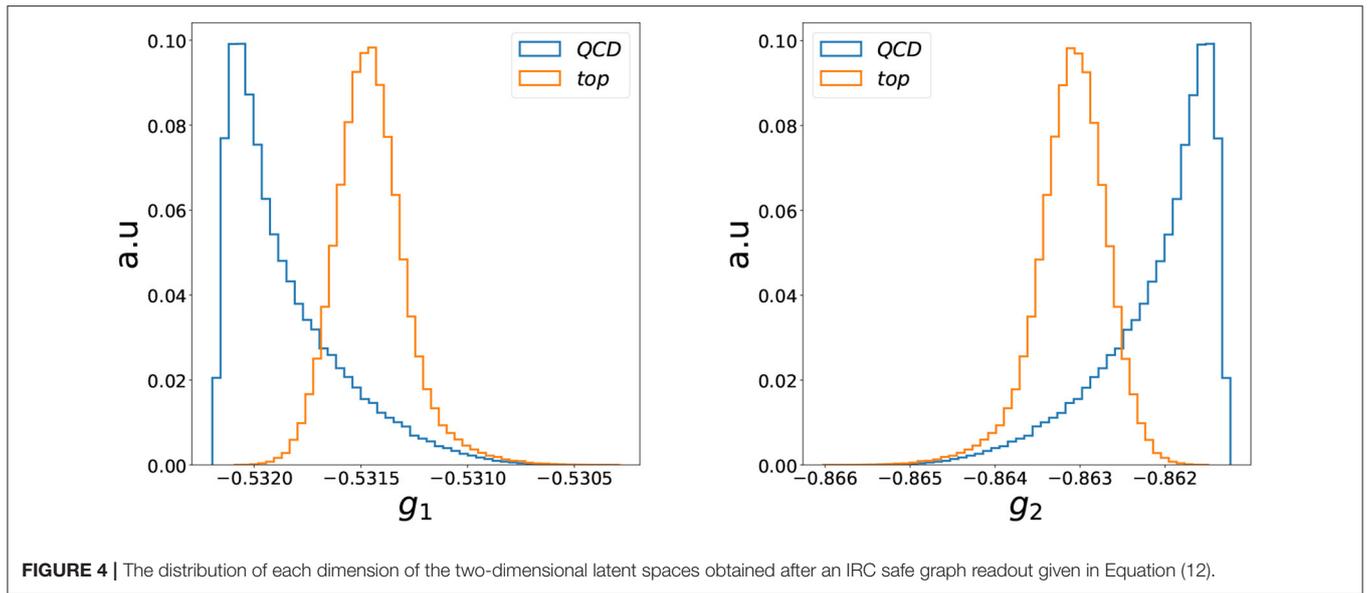


FIGURE 4 | The distribution of each dimension of the two-dimensional latent spaces obtained after an IRC safe graph readout given in Equation (12).

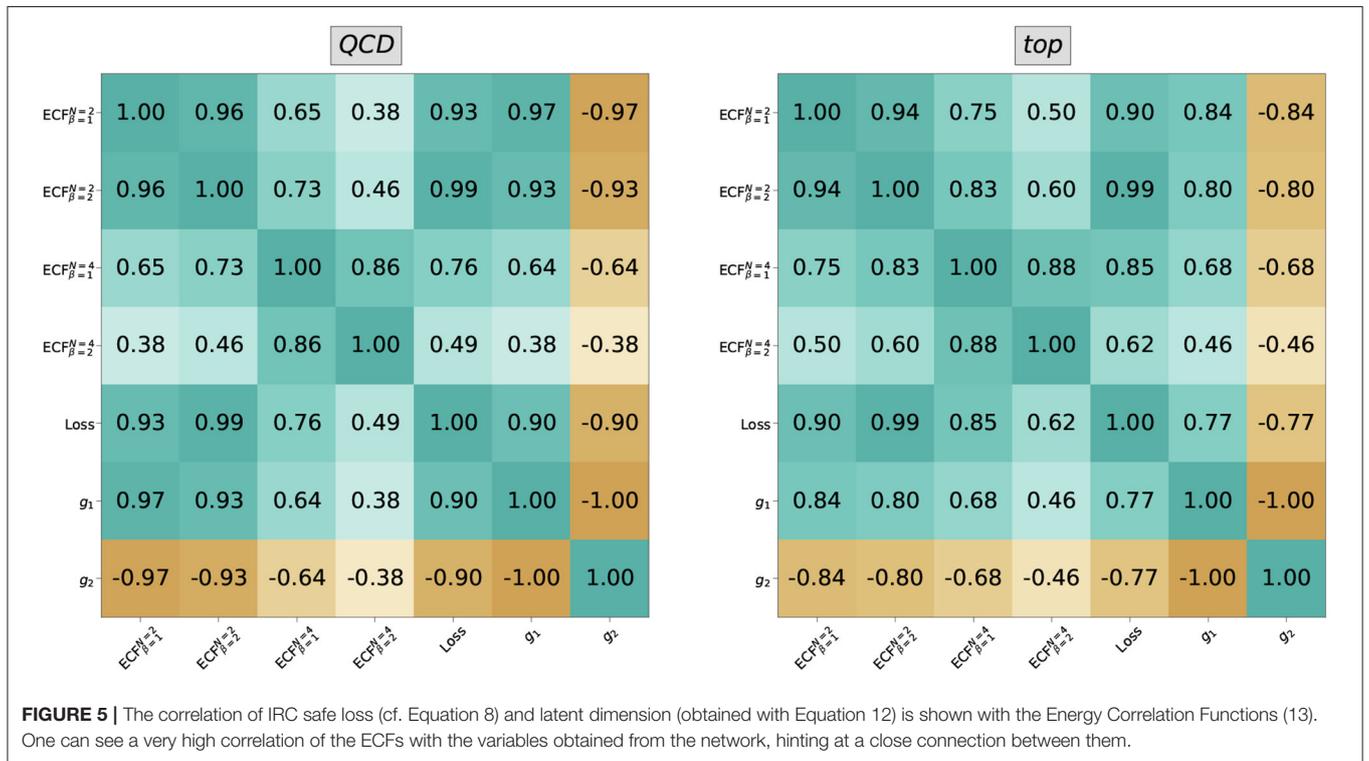


FIGURE 5 | The correlation of IRC safe loss (cf. Equation 8) and latent dimension (obtained with Equation 12) is shown with the Energy Correlation Functions (13). One can see a very high correlation of the ECFs with the variables obtained from the network, hinting at a close connection between them.

The loss would most likely be a better choice when one focuses on anomaly detection capabilities since it condenses the information of the two-dimensional latent space into a single quantity. On the other hand, any latent feature would be more suitable for

applications demanding lower execution times, since in this case only the encoder needs to be evaluated during inference.

Moving on to the relation of the learned information with ECFs, we first define the ECFs as

$$\text{ECF}(N, \beta) = \sum_{i_1 < i_2 < \dots < i_{N-1} < i_N \in J} \left(\prod_{a=1}^N z_{i_a} \right) \left(\prod_{b=1}^{N-1} \prod_{c=b+1}^N \Delta R_{i_b i_c}^\beta \right). \quad (13)$$

Focussing specifically on the case $N = 2$, we obtain

$$\text{ECF}(2, \beta) = \sum_{j < i \in I_J} z_i z_j \Delta R_{ij}^\beta = \sum_{i=1}^{|J|} z_i \sum_{j=i+1}^{|J|} z_j \Delta R_{ij}^\beta, \quad (14)$$

where I_J is the index set of the constituent set J , and $z_i = p_{T^i}^i / (\sum_{k \in I_J} p_{T^k}^k)$. We can rewrite the expression as

$$\text{ECF}(2, \beta) = \sum_{i=1}^{|J|} z_i H_i, \quad H_i = \sum_{j=i+1}^{|J|} z_j \Delta R_{ij}^\beta. \quad (15)$$

Therefore, the quantity H_i can be regarded as a scalar node feature obtained from the message function $\Delta R_{ij}^\beta = \Delta R_{ij}^\beta(\hat{p}_i, \hat{p}_j)$, with a weighted (by z_j) summed readout, while the sum over i to get the ECF is similar to a graph readout operation on all the nodes (or constituents) of the jet. Although the graph structure in the current case is the 2-combinatorial graph, such an analogy suggests that the features extracted by the EMPN are closely connected to ECFs.

This expectation is analyzed in more detail in **Figure 5**, where we show the correlation of different order ECFs with each dimension g_i of the latent graph readout and the loss function. There is a strong correlation between the 2-point ECFs and the network outputs, which decreases when considering the 4-point ECFs. This difference illustrates the close relation of the message passing architecture to the 2-point ECFs. The latent dimensions show a higher correlation for $\beta = 1$ than $\beta = 2$, while the opposite holds for the loss function. This may be due to the ReLU activation, which is essentially a linear function for all positive arguments, while the loss function's higher correlation to the quadratic ECFs may be due to the usage of the mean-squared error as $d(\hat{p}_i, \hat{p}_i)$.

5. CONCLUSIONS

Infrared and collinear safety is not a luxury but an essential requirement to guarantee the theoretical consistency of particle physics collider data interpretations. The emerging and fast-developing area of anomaly detection should therefore incorporate IRC safety when analyzing data at the LHC where QCD activity plays a dominant role. New heavy physics significantly deviates from QCD phenomenology, predominantly characterized by soft and collinear emissions. Reflecting the QCD expectation adequately helps isolate anomalies further; the ability to meaningfully interpolate into the soft and collinear regime

is crucial for extending the reach of such techniques to lower scales. Despite this, IRC safety has not played an essential role in the implementation of anomaly detection. In this paper, we have placed IRC safety at the heart of anomaly detection for the first time by constructing a graph neural network autoencoder that employs Energy-Weighted Message-Passing, which gives rise to an IRC-safe architecture (Konar et al., 2022).

Graph neural networks are well-adapted approaches for isolating tell-tale correlations of final states (Atkinson et al., 2021a; Dreyer and Qu, 2021) and we find that our algorithm shows a high anomaly detection capability whilst having theoretically appealing properties. We have demonstrated this by injecting top jets as an anomaly and finding excellent discriminating sensitivity. While this partly results from the direct presence of a novel hard scale in the jet's substructure, additional sensitivity is accessed from a different soft and collinear shower pattern that accompanies the hard scale. To highlight this relation to well-studied observables in QCD phenomenology, we have shown a strong relation of the information encoded in our autoencoder's latent space with energy correlation functions. This motivates extending anomaly detection analyses using our framework to new physics scenarios of lighter BSM degrees of freedom, which we leave for future work.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://zenodo.org/record/2603256>.

AUTHOR CONTRIBUTIONS

OA, AB, and VN implemented the code and dataset used in the analysis. AB, CE, and VN wrote the first draft of the manuscript. All authors contributed to the conception and design of building the IRC safe graph autoencoder and contributed to manuscript revision, read, and approved the submitted version.

FUNDING

OA was supported by the UK Science and Technology Facilities Council (STFC) under Grant ST/V506692/1. AB and CE were supported by the STFC under Grant ST/T000945/1. CE was also supported by the Leverhulme Trust under Research Project Grant RPG-2021-031 and the IPPP Associate Scheme. PK and VN were supported by the Physical Research Laboratory (PRL), Department of Space, Government of India. MS was supported by the STFC under Grant ST/P001246/1. Part of the computational work detailed in this paper was performed using the HPC resources (Vikram-100 HPC) and TDP project at PRL.

REFERENCES

- Aaboud, M., Aad, G., Abbott, B., Abidin, O., Abeloos, B., Abidi, S. H., et al. (2019). A strategy for a general search for new phenomena using data-derived signal regions and its application within the ATLAS experiment. *Eur. Phys. J. C* 79, 120.
- Araz, J. Y., and Spannowsky, M. (2021). Combine and conquer: event reconstruction with Bayesian ensemble neural networks. *J. High Energy Phys.* 4, 296. doi: 10.1007/04(2021)296
- Atkinson, O., Bhardwaj, A., Brown, S., Englert, C., Miller, D. J., and Stylianou, P. (2021a). Improved constraints on effective top quark interactions using edge convolution networks. *J. High Energy Phys.* 2022:137. doi: 10.1007/04(2022)137
- Atkinson, O., Bhardwaj, A., Englert, C., Ngairangbam, V. S., and Spannowsky, M. (2021b). Anomaly detection with convolutional graph neural networks. *J. High Energy Phys.* 8, 80. doi: 10.1007/08(2021)080
- Bakshi, S. D., Chakraborty, J., Englert, C., Spannowsky, M., and Stylianou, P. (2022). Landscaping CP-violating BSM scenarios. *Nucl. Phys. B* 975, 115676. doi: 10.1016/j.nuclphysb.2022.115676
- Blance, A., and Spannowsky, M. (2020). Unsupervised event classification with graphs on classical and photonic quantum computers. *J. High Energy Phys.* 21, 170. doi: 10.1007/08(2021)170
- Blance, A., Spannowsky, M., and Waite, P. (2019). Adversarially-trained autoencoders for robust unsupervised new physics searches. *J. High Energy Phys.* 10, 047. doi: 10.1007/10(2019)047
- Brüel Gabrielson, R. (2020). "Universal function approximation on graphs," in *Advances in Neural Information Processing Systems*, 33, eds H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin (Curran Associates, Inc.), 19762–19772.
- Butter, A., Kasieczka, G., Plehn, T., and Russell, M. (2018). Deep-learned top tagging with a Lorentz layer. *SciPost Phys.* 5, 28. doi: 10.21468/SciPostPhys.5.3.028
- Cacciari, M., Salam, G. P., and Soyez, G. (2008). The anti- k_t jet clustering algorithm. *J. High Energy Phys.* 4, 63. doi: 10.1088/1126-6708/2008/04/063
- Canelli, F., de Cosa, A., Pottier, L. L., Niedziela, J., Pedro, K., and Pierini, M. (2022). Autoencoders for semivisible jet detection. *J. High Energy Phys.* 2, 74. doi: 10.1007/02(2022)074
- Carmona, A., Lazopoulos, A., Olgoso, P., and Santiago, J. (2021). Matchmakereft: automated tree-level and one-loop matching. *SciPost Phys.* 12, 198. doi: 10.21468/SciPostPhys.12.6.198
- Catani, S., Dokshitzer, Y. L., Seymour, M. H., and Webber, B. R. (1993). Longitudinally invariant K_t clustering algorithms for Hadron Hadron collisions. *Nuclear Phys. B* 406, 187–224. doi: 10.1016/0550-3213(93)90166-M
- Charles, R. Q., Su, H., Kaichun, M., and Guibas, L. J. (2017). "PointNet: deep learning on point sets for 3D classification and segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Honolulu, HI: IEEE Computer Society), 77–85. doi: 10.1109/CVPR.2017.16
- Cheng, T., Arguin, J.-F., Leissner-Martin, J., Pilette, J., and Golling, T. (2020). Variational autoencoders for anomalous jet tagging. *arXiv preprint arXiv:2007.01850*. doi: 10.48550/arXiv.2007.01850
- Collins, J. C., Soper, D. E., and Sterman, G. F. (1989). Factorization of hard processes in QCD. *Adv. Ser. Direct. High Energy Phys.* 5, 1–91. doi: 10.1142/9789814503266_0001
- Collins, J. H., Howe, K., and Nachman, B. (2018). Anomaly detection for resonant new physics with machine learning. *Phys. Rev. Lett.* 121, 241803. doi: 10.1103/PhysRevLett.121.241803
- Collins, J. H., Martin-Ramiro, P., Nachman, B., and Shih, D. (2021). Comparing weak- and unsupervised methods for resonant anomaly detection. *Eur. Phys. J.* 81:617. doi: 10.1140/epjc/s10052-021-09389-x
- Das Bakshi, S., Chakraborty, J., Englert, C., Spannowsky, M., and Stylianou, P. (2021). CP violation at ATLAS in effective field theory. *Phys. Rev. D* 103, 055008. doi: 10.1103/PhysRevD.103.055008
- de Favereau, J., Delaere, C., Demin, P., Giammanco, A., Lemaetre, V., Mertens, A., et al. (2014). DELPHES 3, a modular framework for fast simulation of a generic collider experiment. *J. High Energy Phys.* 2, 57. doi: 10.1007/02(2014)057
- De Simone, A., and Jacques, T. (2019). Guiding new physics searches with unsupervised learning. *Eur. Phys. J. C* 79, 289. doi: 10.1140/epjc/s10052-019-6787-3
- Dillon, B. M., Plehn, T., Sauer, C., and Sorrenson, P. (2021). Better latent spaces for better autoencoders. *SciPost Phys.* 11, 61. doi: 10.21468/SciPostPhys.11.3.061
- Dolan, M. J., and Ore, A. (2021). Equivariant energy flow networks for jet tagging. *Phys. Rev. D* 103, 074022. doi: 10.1103/PhysRevD.103.074022
- Dreyer, F. A., and Qu, H. (2021). Jet tagging in the Lund plane with graph networks. *J. High Energy Phys.* 3, 52. doi: 10.1007/03(2021)052
- Englert, C., Galler, P., Pilkington, A., and Spannowsky, M. (2019). Approaching robust EFT limits for CP-violation in the Higgs sector. *Phys. Rev. D* 99, 095007. doi: 10.1103/PhysRevD.99.095007
- Englert, C., Galler, P., and White, C. D. (2020). Effective field theory and scalar extensions of the top quark sector. *Phys. Rev. D* 101, 035035. doi: 10.1103/PhysRevD.101.035035
- Englert, C., Plehn, T., Schichtel, P., and Schumann, S. (2011). Jets plus missing energy with an autofocus. *Phys. Rev. D* 83, 095009. doi: 10.1103/PhysRevD.83.095009
- Farina, M., Nakai, Y., and Shih, D. (2020). Searching for new physics with deep autoencoders. *Phys. Rev. D* 101, 075021. doi: 10.1103/PhysRevD.101.075021
- Fey, M., and Lenssen, J. E. (2019). "Fast graph representation learning with PyTorch Geometric," in *ICLR Workshop on Representation Learning on Graphs and Manifolds* (New Orleans, LA).
- Finke, T., Krämer, M., Morandini, A., Mück, A., and Oleksiyuk, I. (2021). Autoencoders for unsupervised anomaly detection in high energy physics. *J. High Energy Phys.* 6, 161. doi: 10.1007/06(2021)161
- Freitas, A., López-Val, D., and Plehn, T. (2016). When matching matters: loop effects in Higgs effective theory. *Phys. Rev. D* 94, 095007. doi: 10.1103/PhysRevD.94.095007
- Gerwick, E., Plehn, T., and Schumann, S. (2012a). Understanding jet scaling and jet vetos in Higgs searches. *Phys. Rev. Lett.* 108, 032003. doi: 10.1103/PhysRevLett.108.032003
- Gerwick, E., Plehn, T., Schumann, S., and Schichtel, P. (2012b). Scaling patterns for QCD jets. *J. High Energy Phys.* 10, 162. doi: 10.1007/10(2012)162
- Hajer, J., Li, Y.-Y., Liu, T., and Wang, H. (2020). Novelty detection meets collider physics. *Phys. Rev. D* 101, 076015. doi: 10.1103/PhysRevD.101.076015
- Hallin, A., Isaacson, J., Kasieczka, G., Krause, C., Nachman, B., Quadfasel, T., et al. (2021). Classifying Anomalies Through Outer Density Estimation (CATHODE). *arXiv preprint arXiv:2109.00546*. doi: 10.48550/arXiv.2109.00546
- Heimel, T., Kasieczka, G., Plehn, T., and Thompson, J. M. (2019). QCD or what? *SciPost Phys.* 6, 30. doi: 10.21468/SciPostPhys.6.3.030
- Kasieczka, G., Plehn, T., Thompson, J., and Russel, M. (2019). Top quark tagging reference dataset. Available online at: <https://zenodo.org/record/2603256#.YsWdc9JBwsc>
- Kilgore, W. B., and Giele, W. T. (1997). Next-to-leading order gluonic three jet production at hadron colliders. *Phys. Rev. D* 55, 7183–7190. doi: 10.1103/PhysRevD.55.7183
- Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980v9*. doi: 10.48550/arXiv.1412.6980
- Kinoshita, T. (1962). Mass singularities of Feynman amplitudes. *J. Math. Phys.* 3, 650–677. doi: 10.1063/1.1724268
- Komiske, P. T., Metodiev, E. M., and Thaler, J. (2019). Energy flow networks: deep sets for particle jets. *J. High Energy Phys.* 1, 121. doi: 10.1007/01(2019)121
- Konar, P., Ngairangbam, V. S., and Spannowsky, M. (2022). Energy-weighted message passing: an infra-red and collinear safe graph neural network algorithm. *J. High Energy Phys.* 2, 60. doi: 10.1007/02(2022)060
- Larkoski, A. J., Salam, G. P., and Thaler, J. (2013). Energy correlation functions for jet substructure. *J. High Energy Phys.* 6, 108. doi: 10.1007/06(2013)108
- Lee, T. D., and Nauenberg, M. (1964). Degenerate systems and mass singularities. *Phys. Rev.* 133, B1549–B1562. doi: 10.1103/PhysRev.133.B1549
- Nachman, B. (2020). Anomaly detection for physics analysis and less than supervised learning. *arXiv preprint arXiv:2010.14554*. doi: 10.48550/arXiv.2010.14554
- Nachman, B., and Shih, D. (2020). Anomaly detection with density estimation. *Phys. Rev. D* 101, 075042. doi: 10.1103/PhysRevD.101.075042

- Prestel, S., and Spannowsky, M. (2019). HYTREES: combining matrix elements and parton shower for hypothesis testing. *Eur. Phys. J. C* 79, 546. doi: 10.1140/epjc/s10052-019-7030-y
- Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017). "PointNet++: deep hierarchical feature learning on point sets in a metric space," in *Advances in Neural Information Processing Systems*, eds I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Long Beach, CA), 30.
- Roy, T. S., and Vijay, A. H. (2019). A robust anomaly finder based on autoencoders. *arXiv preprint arXiv:1903.02032*. doi: 10.48550/arXiv.1903.02032
- Sjöstrand, T., Ask, S., Christiansen, J. R., Corke, R., Desai, N., Ilten, P., et al. (2015). An introduction to PYTHIA 8.2. *Comput. Phys. Commun.* 191, 159–177. doi: 10.1016/j.cpc.2015.01.024
- Sjostrand, T., Mrenna, S., and Skands, P. Z. (2008). A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.* 178, 852–867. doi: 10.1016/j.cpc.2008.01.036
- Soper, D. E., and Spannowsky, M. (2011). Finding physics signals with shower deconstruction. *Phys. Rev. D* 84, 074002. doi: 10.1103/PhysRevD.84.074002
- Soper, D. E., and Spannowsky, M. (2014). Finding physics signals with event deconstruction. *Phys. Rev. D* 89, 094005. doi: 10.1103/PhysRevD.89.094005
- Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., and Solomon, J. M. (2019). Dynamic graph CNN for learning on point clouds. *ACM Trans. Graph.* 38, 1–12. doi: 10.1145/3326362
- Weinberg, S. (1979). Phenomenological Lagrangians. *Phys. A* 96, 327–340. doi: 10.1016/0378-4371(79)90223-1
- Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R. R., and Smola, A. J. (2017). "Deep sets," in *Advances in Neural Information Processing Systems* (La Jolla, CA), 30.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor CD is currently organizing a Research Topic with the authors MS.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Atkinson, Bhardwaj, Englert, Konar, Ngairangbam and Spannowsky. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.