Graph Neural Network: Its applications to constrain BSM models and EFTs

Akanksha Bhardwaj

Oklahoma State University

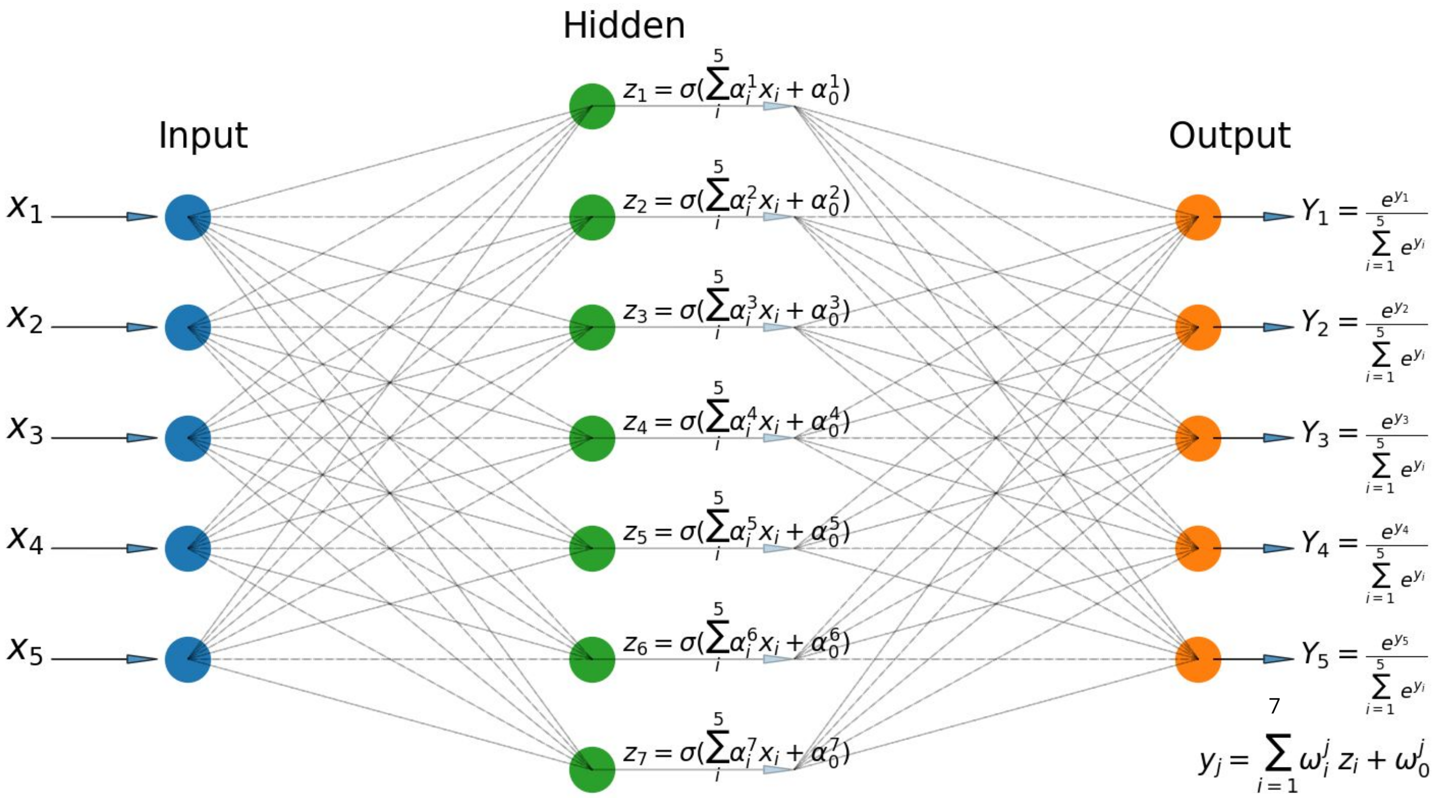
JHEP 08 (2021) 080 O. Atkinson, **A.B**, C. Englert, V.Ngairangbam, M. Spannowsky

JHEP 04 (2022) 137 O. Atkinson, **A.B**, S. Brown, C. Englert, D. Miller, P. Stylianou

Outline

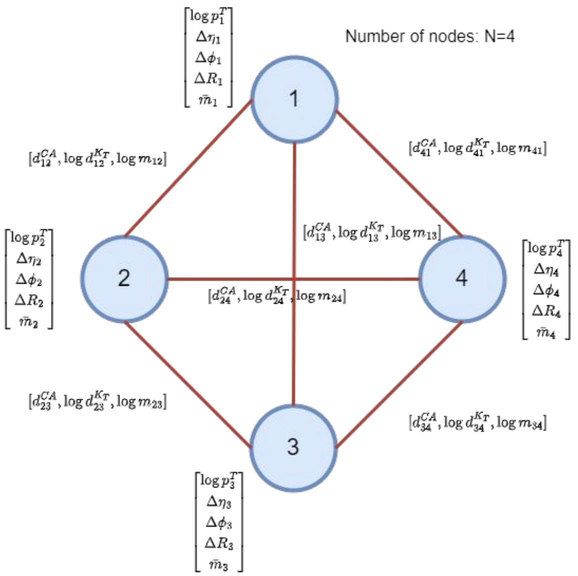
- Graph Neural Networks (GNNs) are powerful deep learning algorithm for automatic feature extraction on graph structured data
  - General enough for non-Euclidean physics data
  - Can encode pairwise relation through edge features
- Graph Autoencoder for unsupervised detection of non-QCD jets:
  - design a symmetric decoder capable of simultaneously reconstructing edge features and node features
- Application of GNNs in a supervised scenario : an SMEFT analysis of top-antitop pair production and its decay
  - Multiclass classification of thirteen independent Wilson coefficients switched on simultaneously

Artificial Neural Networks

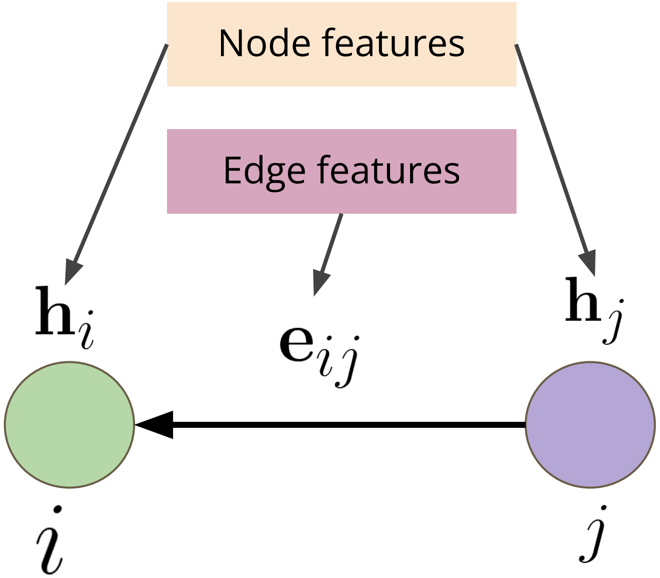


What are Graphs?

- A set of objects, and the relations between a pair of objects are naturally expressed as a *graph*.
- Graph Neural Networks (GNNs) operate on graph data
- To further describe each node, edge or the entire graph, we can store information in each of these pieces of the graph.
    - Vertex (or node) embedding
    - Edge (or link) attributes and embedding
    - Global (or graph) embedding



Message passing operation



1. Message Passing

$$\mathbf{m}_{ij} = \mathbf{M}(\mathbf{h}_i, \mathbf{h}_j, \mathbf{e}_{ij})$$

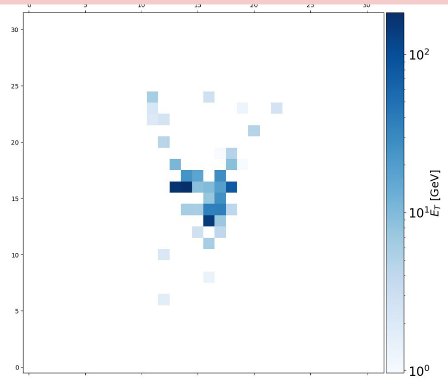
Neural Network Shared for all edges

2. Node Readout

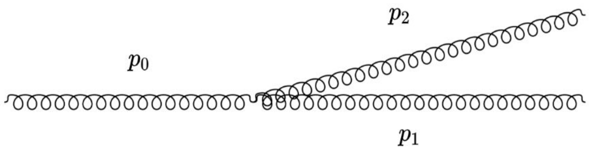
$$\mathbf{h}_i' = \mathbf{A}(\mathbf{h}_i, \{\mathbf{m}_{ik} \mid k \in \mathcal{N}(i)\})$$

Jet Substructure at LHC

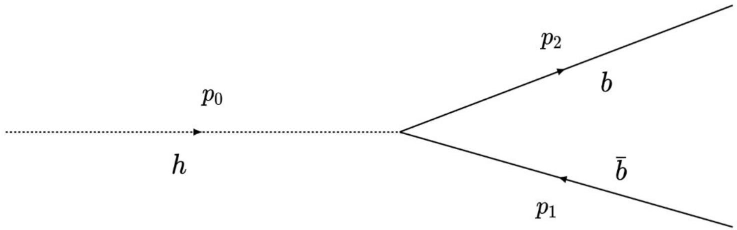
Jet Image of a boosted Top Quark



First splitting/decay at Parton-level



Dominantly Soft or collinear
$(z_2 << z_1 \text{ or } \theta_{12} \to 0)$



Democratic splitting
$(z_2 \sim z_1)$

QCD jets(Background)

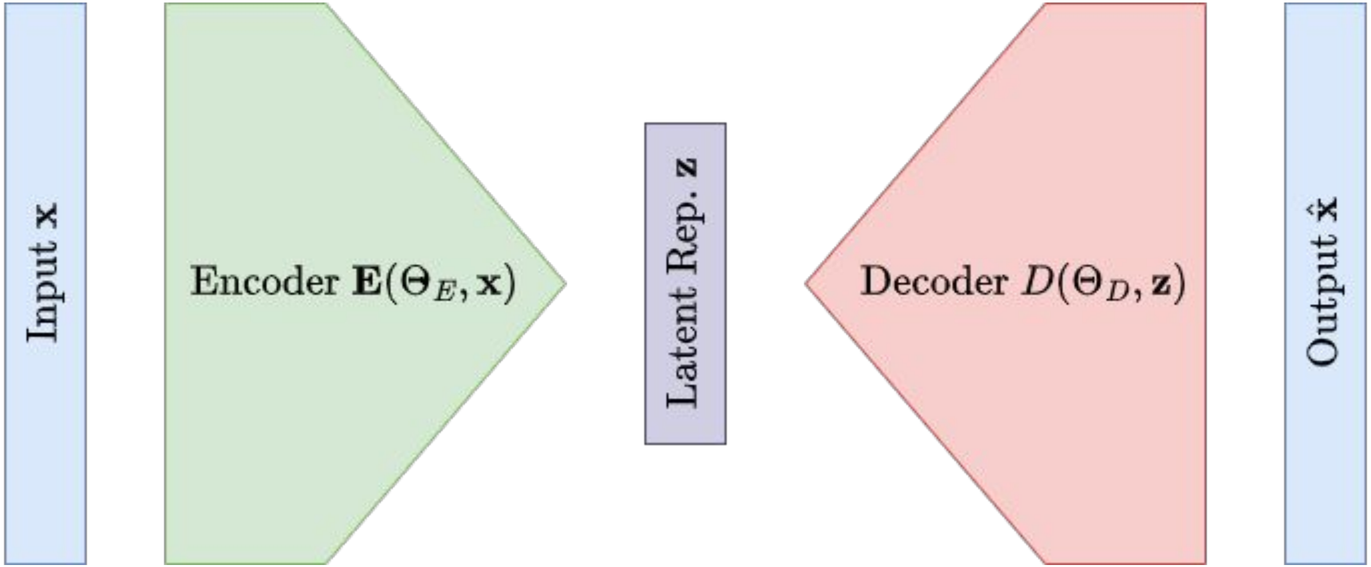
Higgs jets (2-prong Signal)

z_i is relative hardness

for hadronic colliders: $z_i = \dfrac{p_T^i}{\sum_{j=1}^{n} p_T^j}$

$$\hat{p}_i = (y_i, \phi_i)$$

Autoencoders

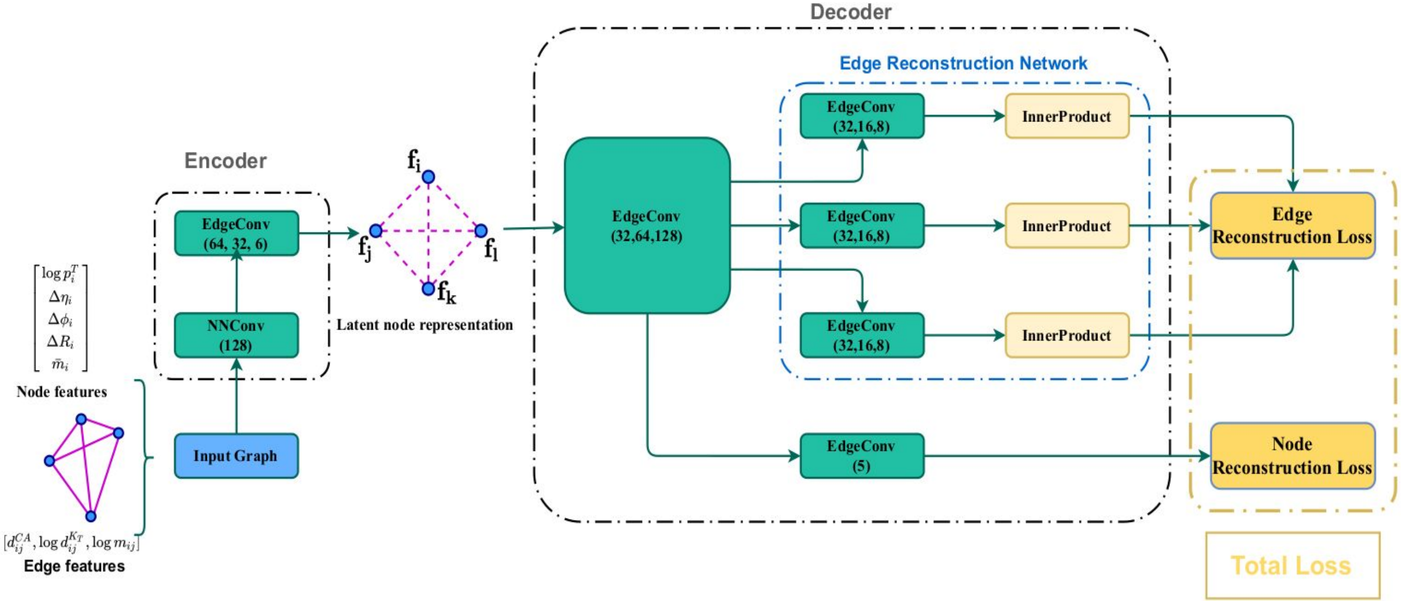


Autoencoders are neural networks that map an input space to a bottleneck dimension (the latent dimension) and then back again to a space identical to the input.

Train on background (most likely events)

=> higher reconstruction error on previously unseen events **= Anomaly**

Graph Autoencoder



Adjacency matrices:

$$A_{ij}^a = A_{ji}^a = \begin{cases} e_{ij}^a & \text{if } i \neq j \\ 1 & \text{otherwise} \end{cases} ,$$

e_{ij}^a: components of Edge feature vectors.

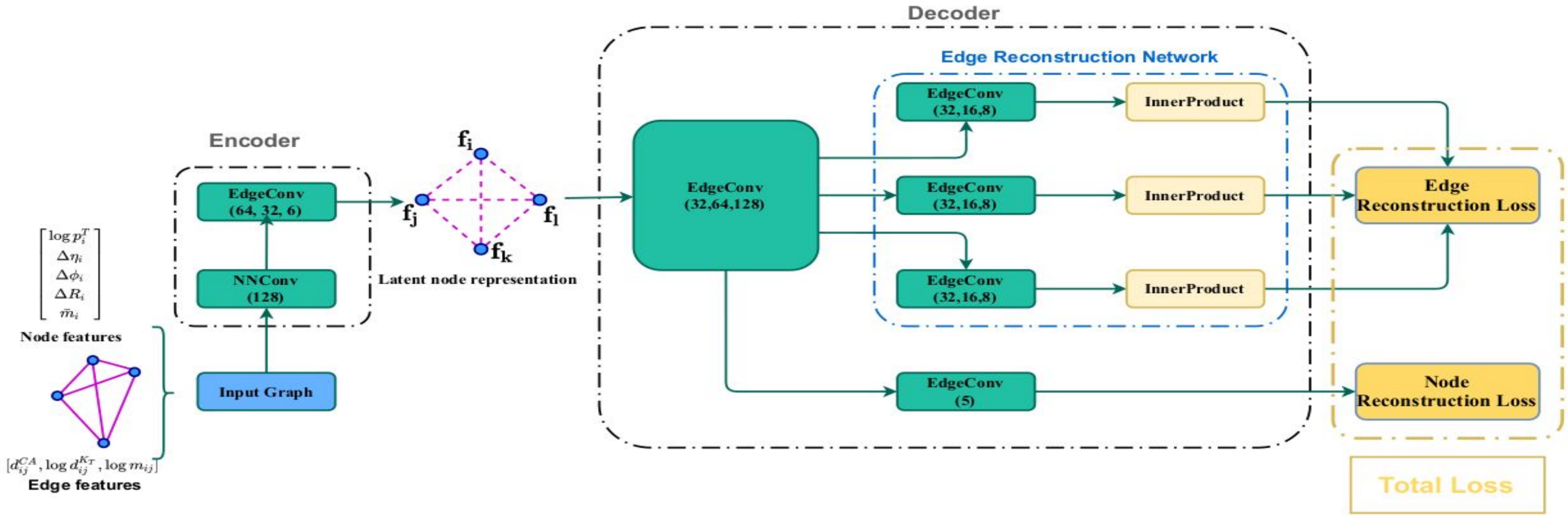
Inner-product layer:

$$\hat{A}_{ij} = \vec{h}_i \cdot \vec{h}_j ,$$

$\vec{h}_i$: Node-features of preceding layer

Network also reconstructs three $N \times N$ adjacency matrices for each graph.

Graph Autoencoder



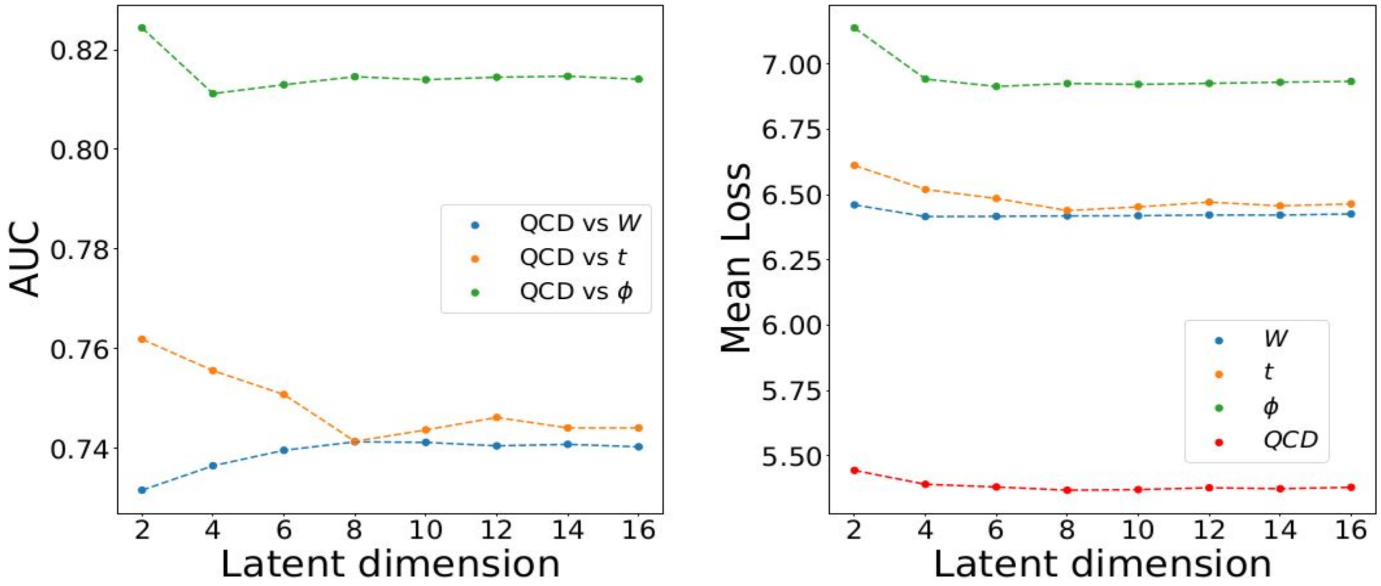
Loss-function:

$$L_{node} = \sqrt{\sum_{ia} \frac{(\hat{x}_i^a - x_i^a)^2}{N \times 5}}, \quad L_{edge} = \sum_a \sqrt{\sum_{ij} \frac{(\hat{A}_{ij}^a - A_{ij}^a)^2}{N \times N}},$$

$$L_{auto} = \lambda_{node}\ L_{node} + \lambda_{edge}\ L_{edge}$$

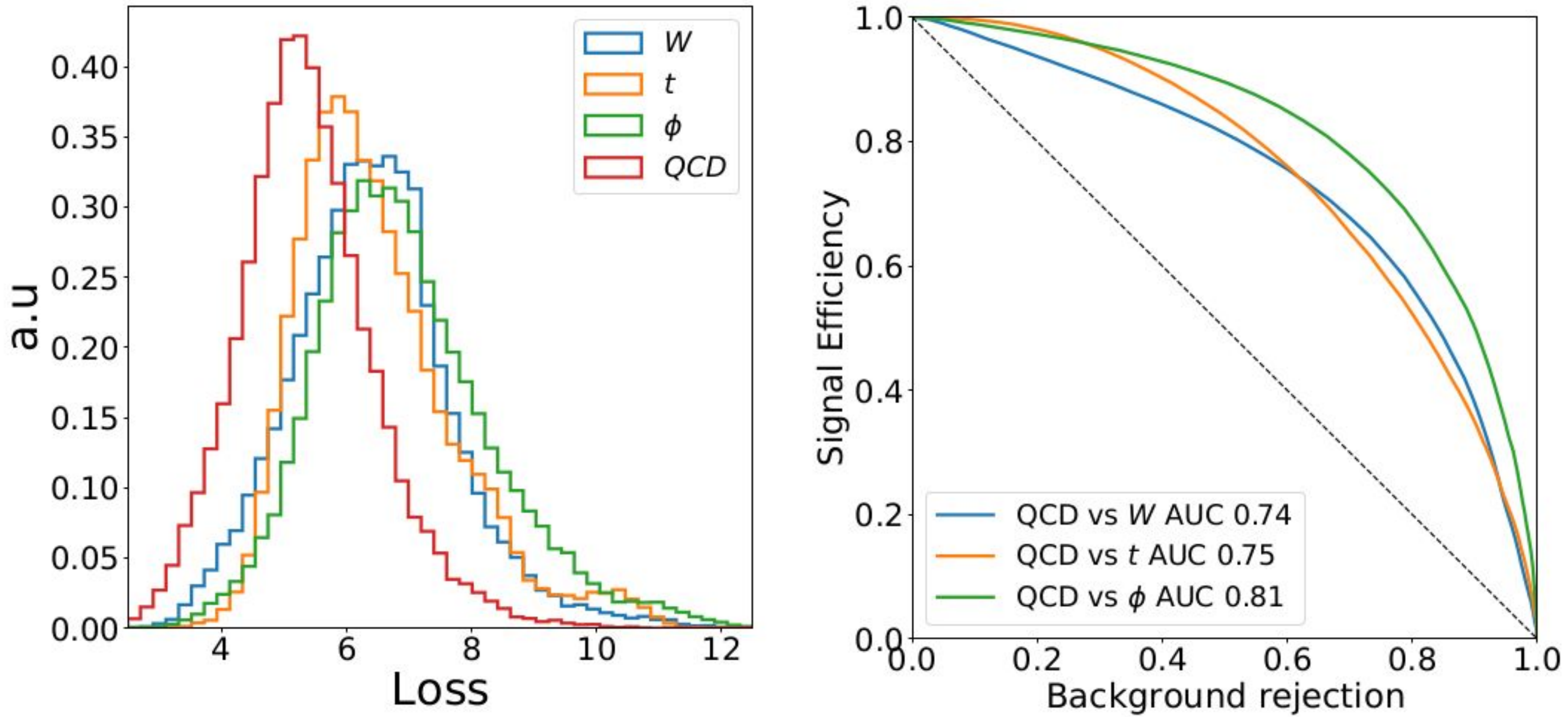
$\lambda_{node} = 0.3$ and $\lambda_{edge} = 1$.

Latent dimension scan

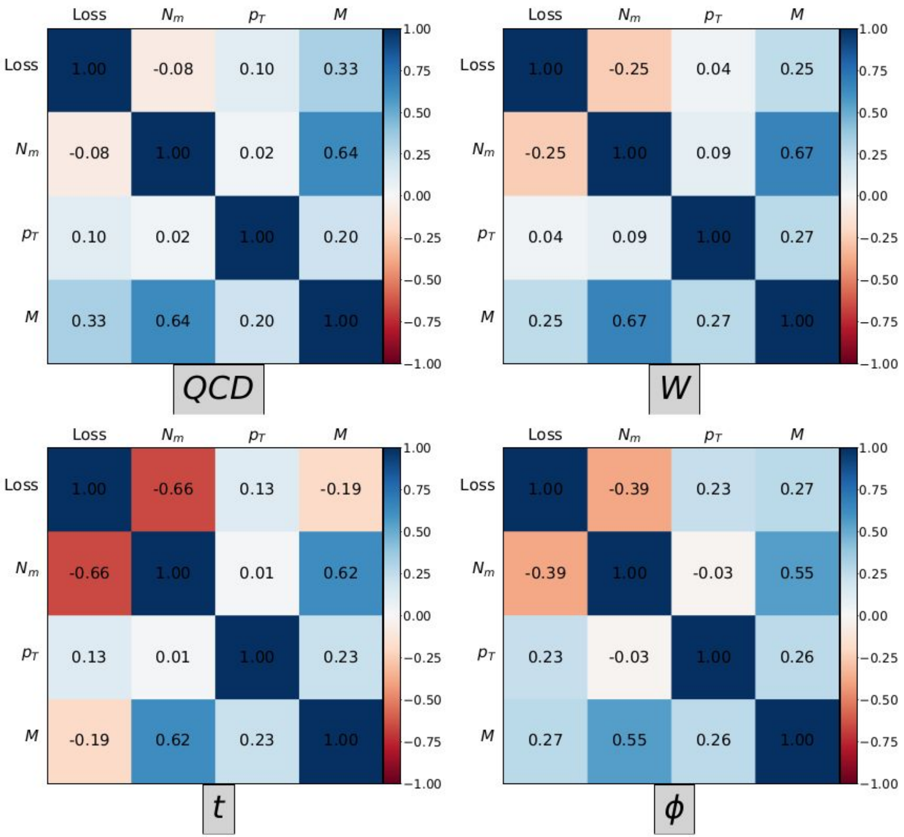


Choose latent dimension $= 6$

Loss distribution for latent dimension 6



Loss correlations

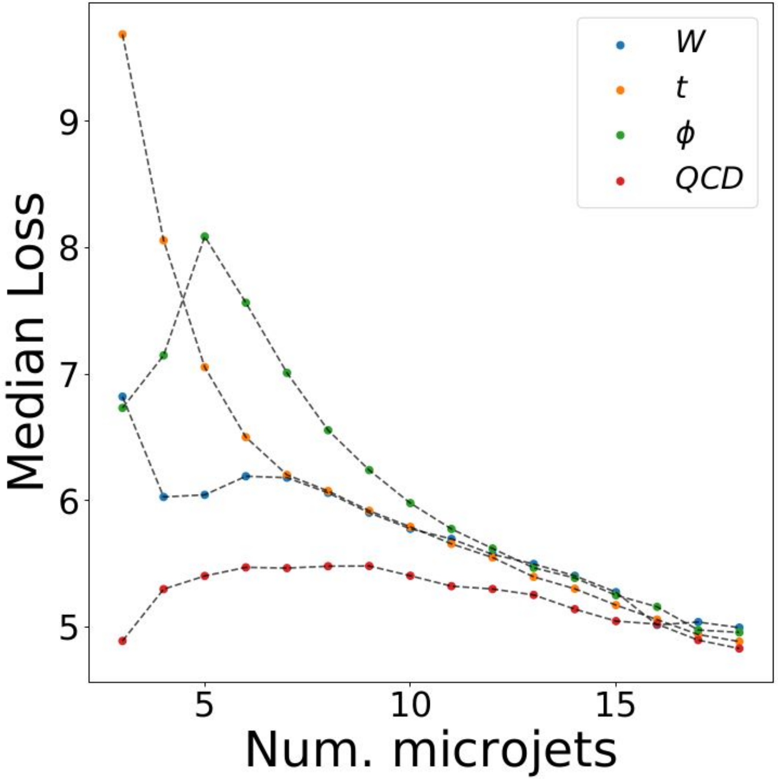


Shared weights for all edges per layer

QCD jets:

$\Longrightarrow$ Learning a uniform structure regardless of the number of nodes?

Loss correlation



Signal Jets: Extra radiation other than their multiplicities arise from QCD radiation

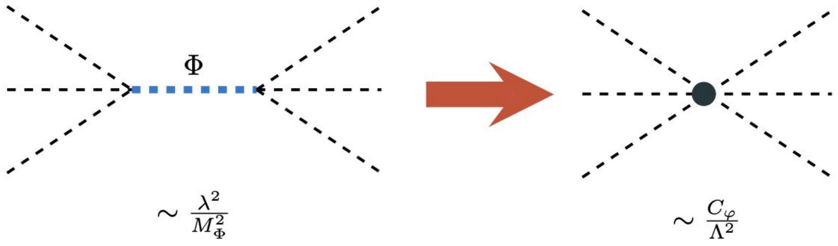
Graph Neural Network : Application to EFT in supervised learning framework

Effective Field Theory

- Lack of experimental evidence of new physics may indicate a mass gap between SM and BSM scales.
- SMEFT Lagrangian

$$\mathcal{L}_{\mathrm{SMEFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_{i>4}\sum_{k}\frac{C_k^{(i)}}{\Lambda^{i-4}}\mathcal{O}_k^{(i)}$$

  - Top-down:
    - Integrate out heavy BSM states of UV-complete theory.
    - Match Wilson Coefficients (WCs) to variables of full theory



Effective Field Theory

- Bottom-up:
  - Agnostically include all allowed operator deformations as an expansion around Λ^{-1}.
  - Truncate the series to and interpret LHC results as bounds on Wilson Coefficients $C_k^{(i)}$ of the operators.
- Only one $\mathcal{O}^{(5)}$ which is relevant for neutrino physics.
- With minimal flavour violation and baryon number conservation there are 59 operators $\mathcal{O}_k^{(6)}$.
- 16 dimension-6 operators relevant for top physics.
- Keep terms in Lagrangian only up to Λ^{-2}.

Cross Section from SMEFT

Any differential cross section follows:

$$d\sigma = d\sigma_{\mathrm{SM}} + \frac{C_i}{\Lambda^2} d\sigma_i + \frac{C_i C_j}{\Lambda^4} d\sigma_{ij}$$

Interference

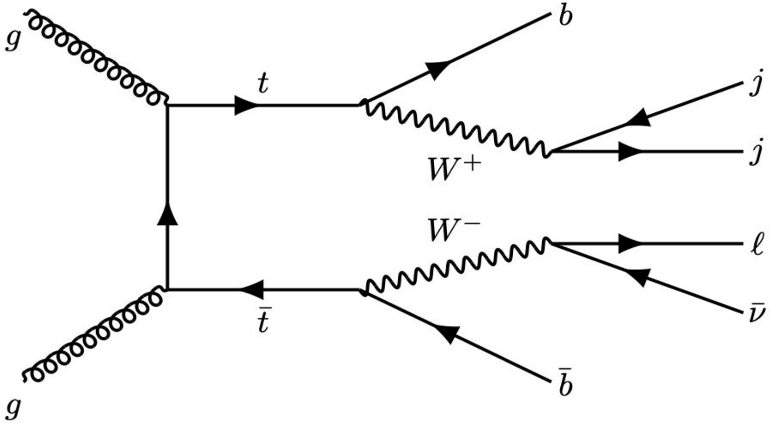
Cross terms

- Λ^{-4} terms are suppressed, truncate series at Λ^{-2}
- Differential distributions can be used to quantify allowed range on WCs.
- Optimised selection of signal region can result in improved bounds.
- Any improvement for linear case should generalise to Λ^{-4} terms.

Improving SMEFT results with GNNs

- Focus on process $pp \to t\bar{t} \to \ell b\bar{b}jj + \not{E}_T$
- 13 relevant SMEFT operators in this process.
  - $\boldsymbol{X^3}$: $\mathcal{O}_G$
  - $\boldsymbol{\psi^2 X\varphi}$: $\mathcal{O}_{uG}^{33}$, $\mathcal{O}_{uW}^{33}$
  - $\boldsymbol{\psi^2\varphi^2 D}$: $\mathcal{O}_{\varphi q}^{(3)33}$
  - $\boldsymbol{\psi^4}$: $\mathcal{O}_{qq}^{(1)i33i}$, $\mathcal{O}_{qq}^{(3)i33i}$, $\mathcal{O}_{qq}^{(3)ii33}$, $\mathcal{O}_{qu}^{(8)33ii}$, $\mathcal{O}_{qu}^{(8)ii33}$, $\mathcal{O}_{ud}^{(8)33ii}$,

    $\mathcal{O}_{uu}^{i33i}$, $\mathcal{O}_{qd}^{(8)33ii}$, $\mathcal{O}_{lq}^{(3)ii33}$ $\longrightarrow$ lepton-quark operator

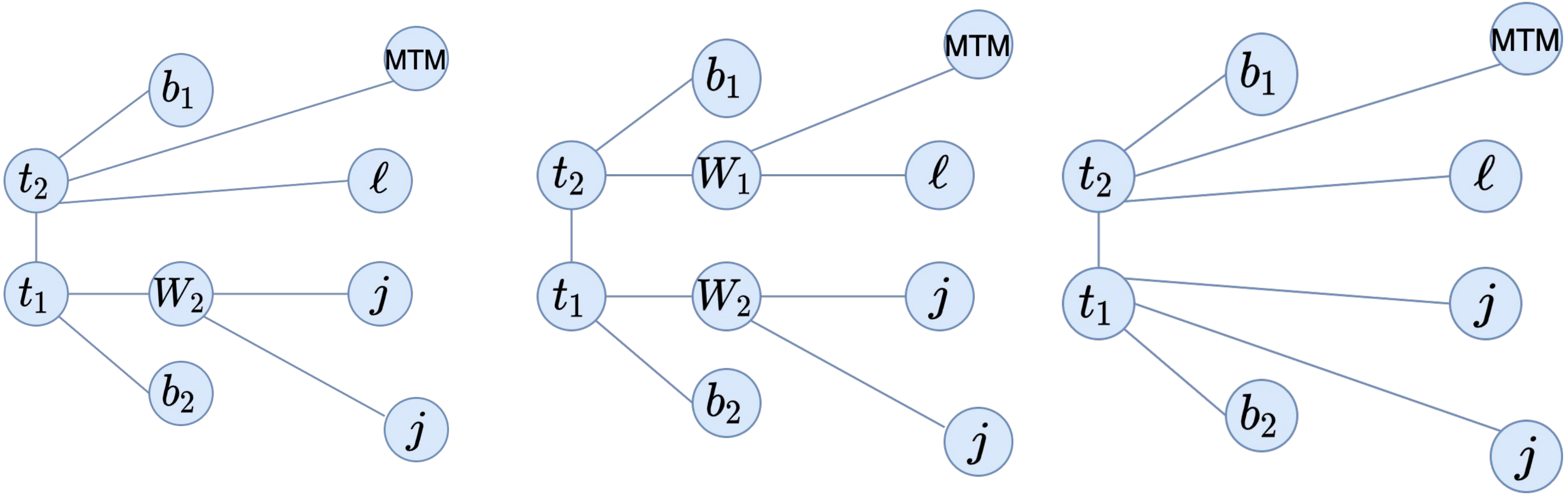


Particle event as a graph

1. Usually graphs are constructed either w.r.t. the distance between the nodes or fully connected.
2. Here instead attempt to use systematic procedure to construct physics-inspired graphs from final states :
   a. Impose basic selection criteria on jets, leptons and b-quarks
   b. Create nodes for jets, lepton, b-quarks and missing transverse momentum (MTM).
   c. Attempt to reconstruct invariant mass of the two W bosons from lepton-MTM and dijet.
   d. If reconstructed masses are relatively close to actual W mass create node.
   e. Attempt to reconstruct top quarks and create nodes.

Particle events as Graph

Edges are connected according to the decay from their parent particles



Graph Features

Node Features

- Transverse momentum
- Azimuthal angle
- Pseudorapidity
- Energy
- Invariant mass
- Particle Identity

No Edge feature in our set-up

Goal: GNN should utilize the features and graph structure to classify events according to the operator that gave rise to it or as purely-SM → Multi-Class Setup

GNN Architecture

Edge Convolution:

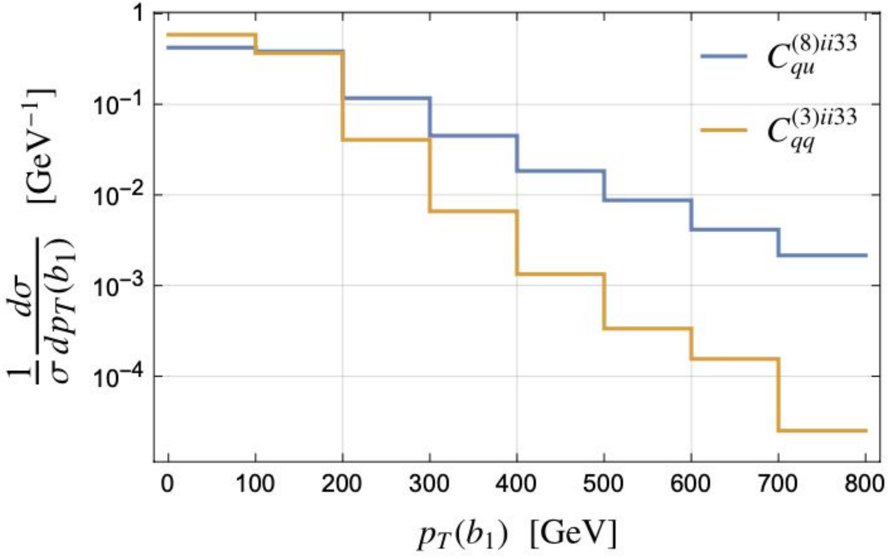
$$\vec{x}_i^{(l+1)} = \underbrace{\frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)}}_{\text{Aggregation}} \overbrace{\textsc{ReLU}\left(\Theta\cdot\left(\vec{x}_j^{(l)}-\vec{x}_i^{(l)}\right)+\Phi\cdot\left(\vec{x}_i^{(l)}\right)\right)}^{\text{Message}},$$

- Updates features of vertex i using other nodes in its neighbourhood, N{i}
- Node aggregation is done by taking mean
- Θ and Φ are linear layer
- Graph readout done by taking mean

Two Operator scenario

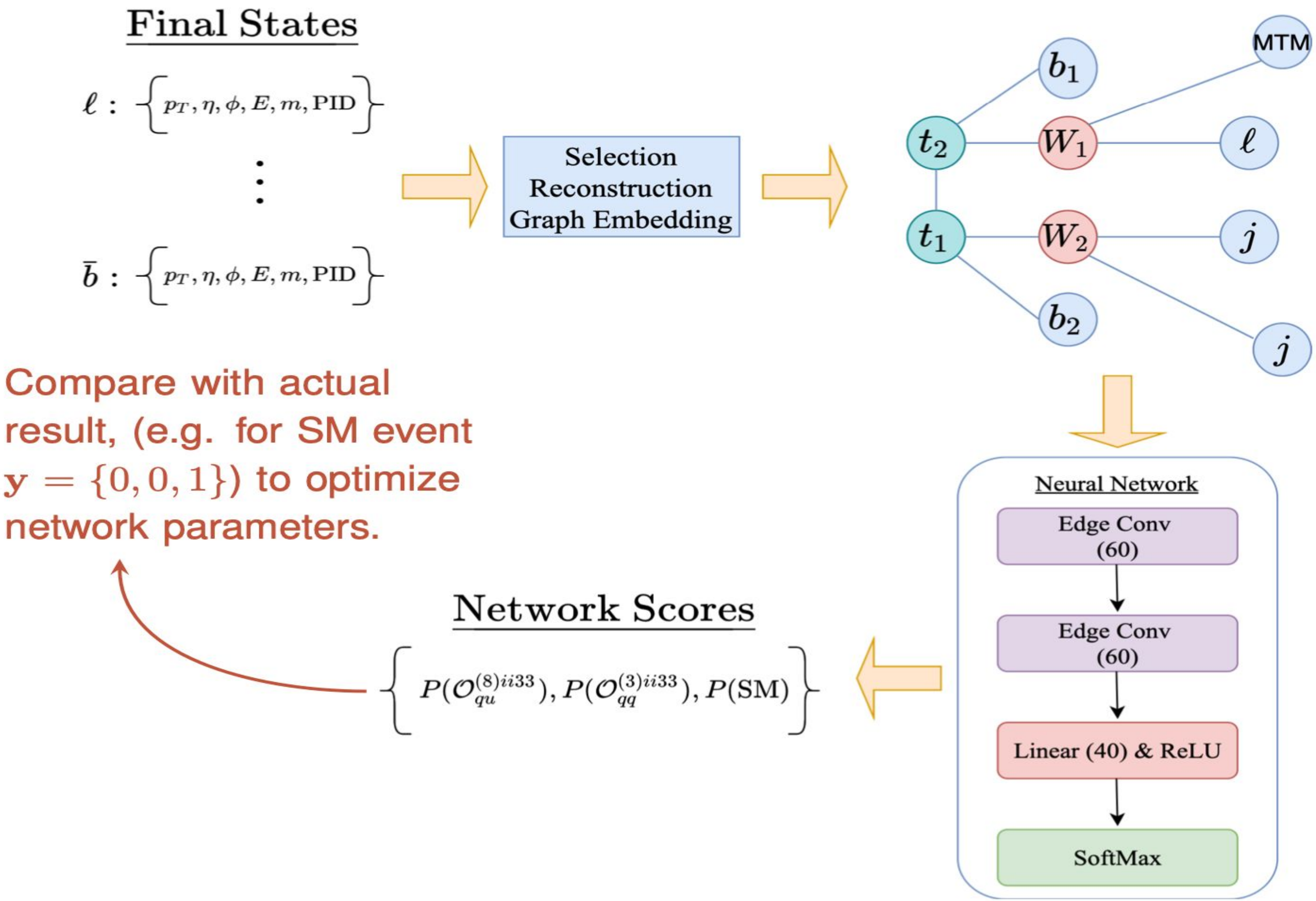
- Test a simple scenario with only two four fermion operator

$$\mathcal{O}_{qu}^{(8)ii33} = (\bar{q}_i \gamma_\mu T^A q_i)(\bar{u_3} \gamma^\mu T^A u_3)\,,$$
$$\mathcal{O}_{qq}^{(3)ii33} = (\bar{q}_i \gamma_\mu \tau^I q_i)(\bar{q_3} \gamma^\mu \tau^I q_3)\,.$$



Normalized $p_T(b_1)$ distributions at the 13 TeV LHC for the two operators used in the three-class example.

Schematics

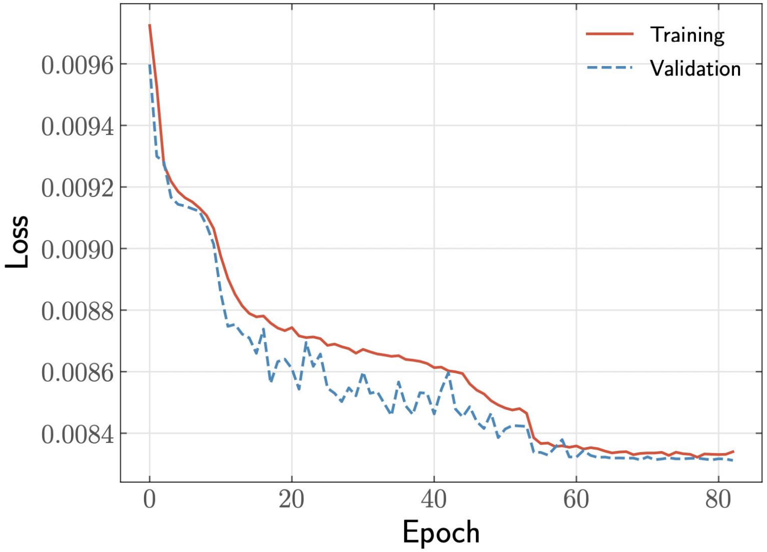

Final States
ℓ : $\{p_T, \eta, \phi, E, m, \mathrm{PID}\}$
$\bar{b}$: $\{p_T, \eta, \phi, E, m, \mathrm{PID}\}$
Selection
Reconstruction
Graph Embedding
t_2
t_1
b_1
b_2
W_1
W_2
MTM
ℓ
j
j
Neural Network
Edge Conv (60)
Edge Conv (60)
Linear (40) & ReLU
SoftMax
Network Scores
$\{P(\mathcal{O}_{qu}^{(8)ii33}), P(\mathcal{O}_{qq}^{(3)ii33}), P(\mathrm{SM})\}$
Compare with actual result, (e.g. for SM event $\mathbf{y} = \{0, 0, 1\}$) to optimize network parameters.

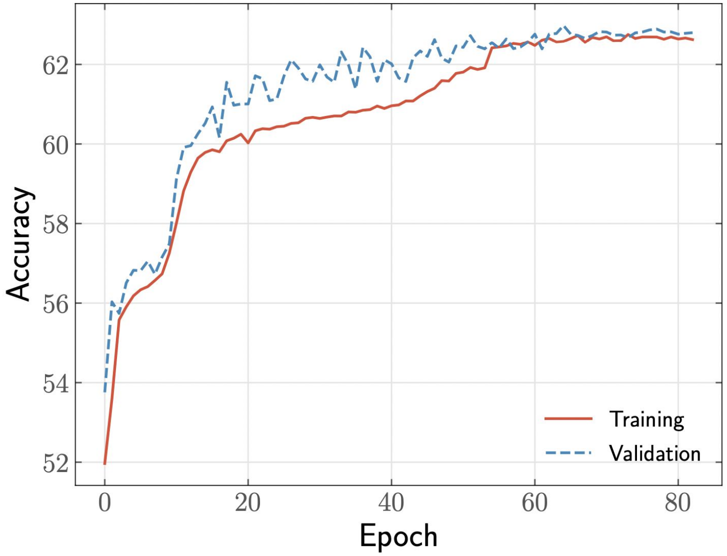
Training the GNN

Optimize parameters by minimizing categorical cross-entropy function

$$\mathcal{L} = -\sum_{i=1}^{N_{\text{classes}}} y_i \log \hat{y}_i \,.$$

Optimized hyperparameters & architecture to achieve better convergence.

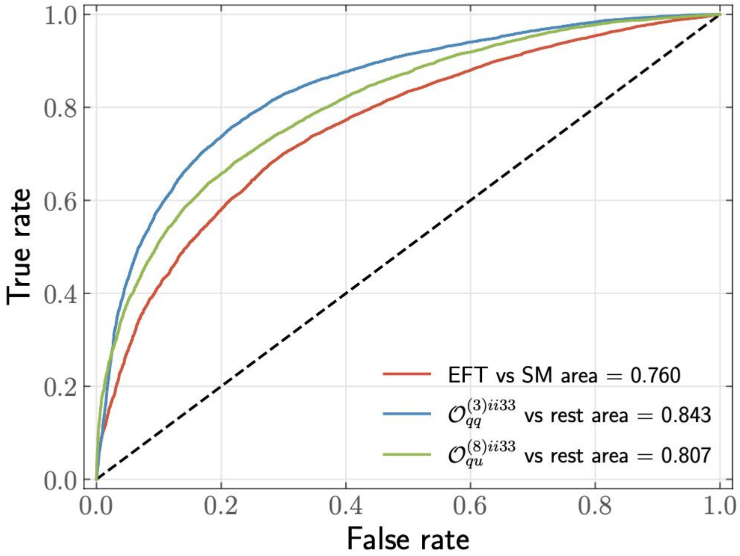




Performance of GNN via ROC curve

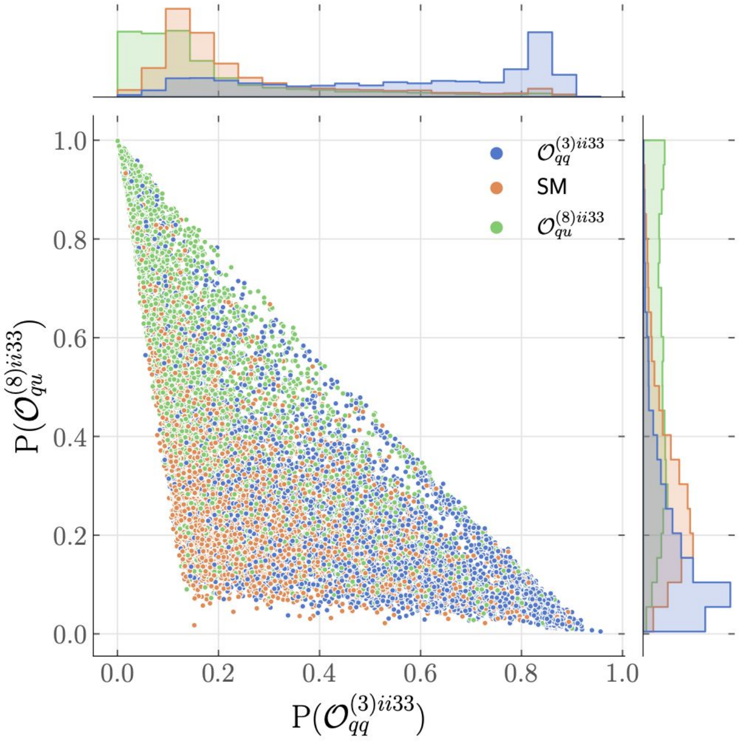
- Different architectures and embeddings were compared with Receiver Operating Characteristic curves.
- ROC curves calculated for each operator in a one-vs-rest scheme.
- Additionally calculate curve for overall EFT performance using

$$\mathrm{P}(\mathsf{BSM}) = \mathrm{P}(\mathcal{O}_{qq}^{(3)ii33}) + \mathrm{P}(\mathcal{O}_{qu}^{(8)ii33}) = 1 - \mathrm{P}(\mathsf{SM}) \ .$$



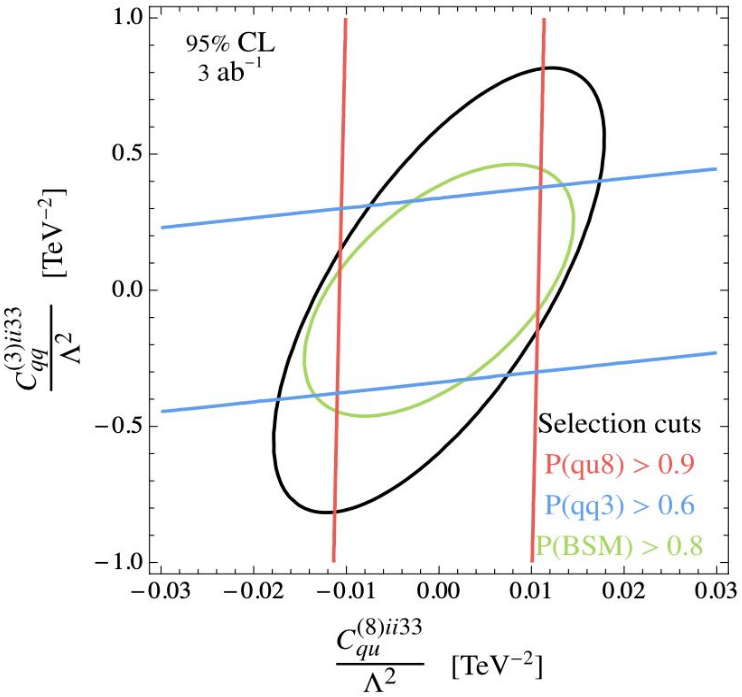
Probability distribution of the network output

The network essentially separates events into three different regions.



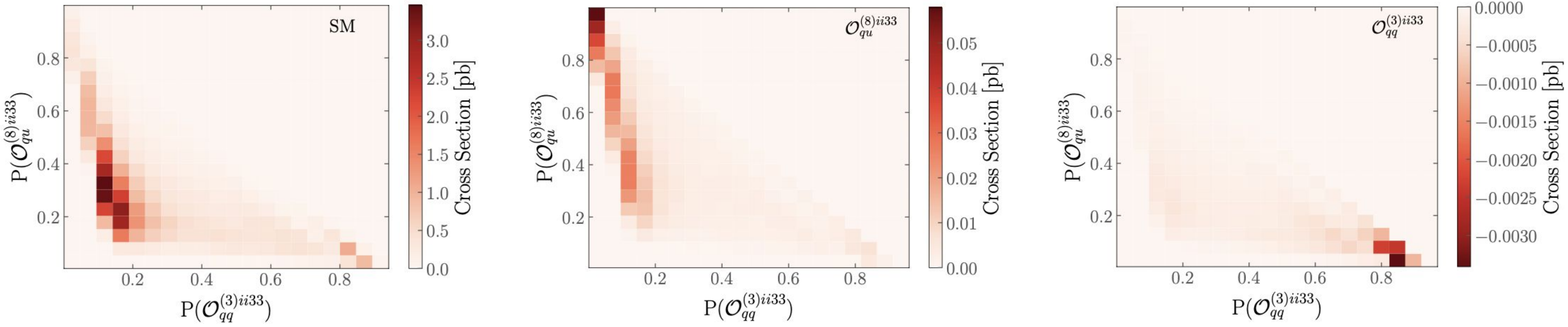
Fit

- Performed χ2 fit using pT (b1) distributions at an extrapolated luminosity 3/ab.
- Cuts on network scores result in improved performance over just the selection cuts.



Fit-alternative

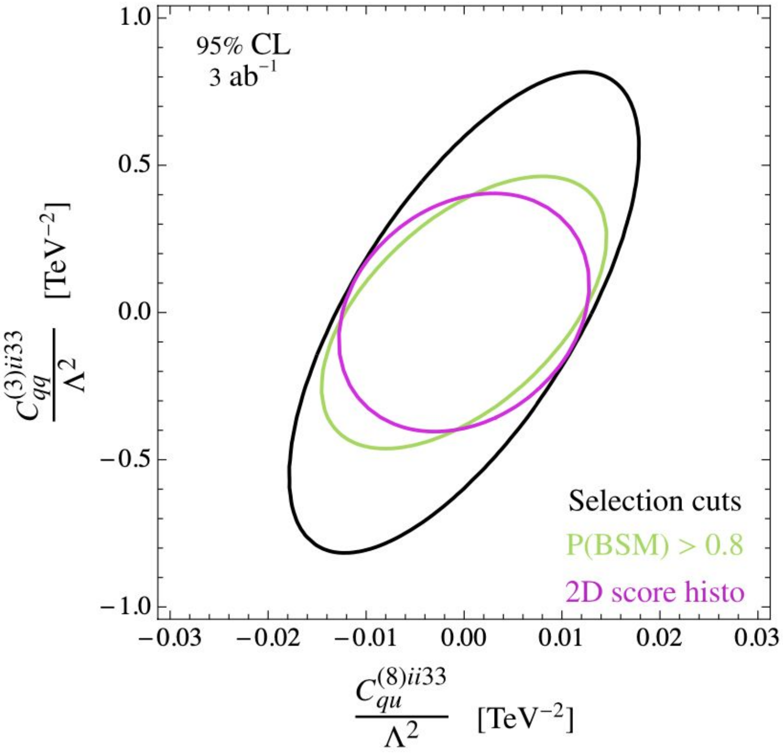
Can also perform fit on flattened two-dimensional histograms from scores.



Example two-dimensional histograms for each contribution, normalised to the cross-section rate.

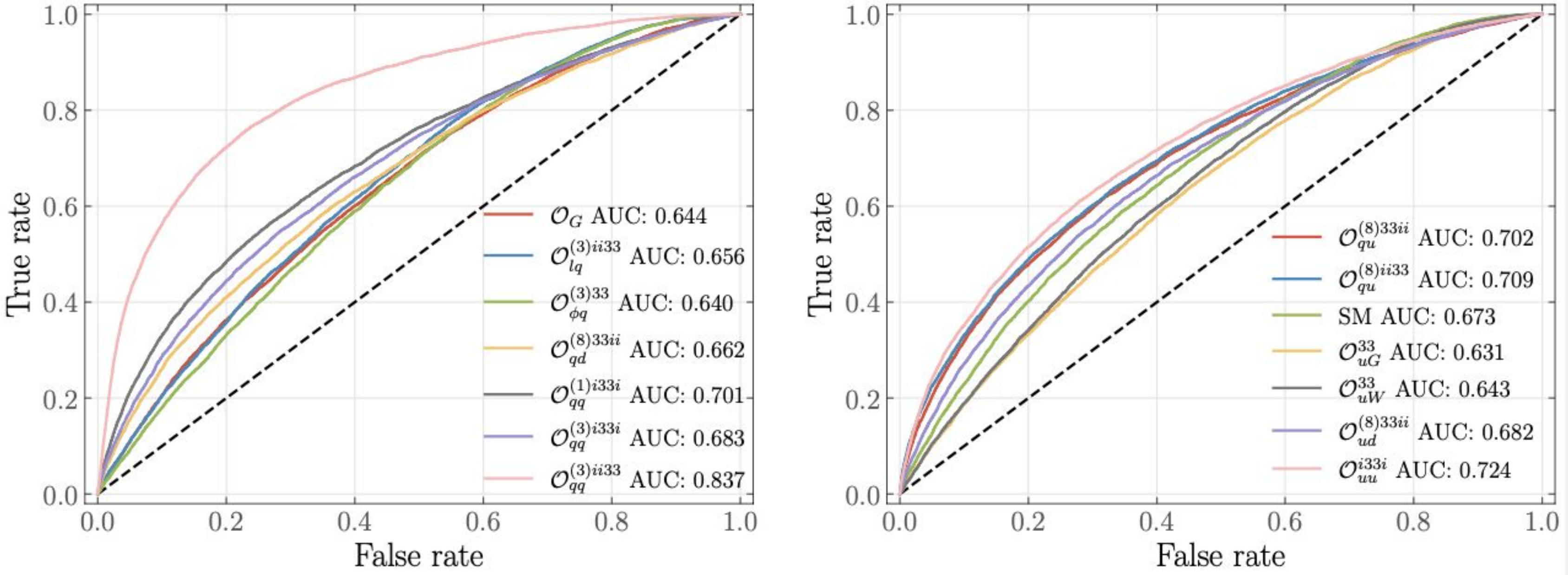
Contours obtained from 2D Histogram

Cutting on the BSM score provides similar contours with the 2D score histogram approach.



Full Fit constraints with GNN selection

- Extend setup to 13 SMEFT operators relevant to the process.
- ROC curves indicate the capability of the network to distinguish operators.



Baseline analysis

- Performed analysis of **CMS (1610.04191)** for comparison with results based on GNN score cut.

Distribution	Observable	Binning
$\frac{1}{\sigma}\frac{d\sigma}{d\lvert y_t^h\rvert}$	$\lvert y_t^h\rvert$	$[0.0, 0.2, 0.4, 0.7, 1.0, 1.3, 1.6, 2.5]$
$\frac{1}{\sigma}\frac{d\sigma}{d\lvert y_t^l\rvert}$	$\lvert y_t^l\rvert$	$[0.0, 0.2, 0.4, 0.7, 1.0, 1.3, 1.6, 2.5]$
$\frac{1}{\sigma}\frac{d\sigma}{d\lvert y_{t\bar{t}}\rvert}$	$\lvert y_{t\bar{t}}\rvert$	$[0.0, 0.2, 0.4, 0.6, 0.9, 1.3, 2.3]$
$\frac{1}{\sigma}\frac{d\sigma}{dp_\perp^{t,h}}$	$p_\perp^{t,h}$	$[0, 45, 90, 135, 180, 225, 270, 315, 400, 800]$ GeV
$\frac{1}{\sigma}\frac{d\sigma}{dp_\perp^{t,l}}$	$p_\perp^{t,l}$	$[0, 45, 90, 135, 180, 225, 270, 315, 400, 800]$ GeV
$\frac{1}{\sigma}\frac{d\sigma}{dm_{t\bar{t}}}$	$m_{t\bar{t}}$	$[300, 375, 450, 530, 625, 740, 850, 1100, 2000]$ GeV
$\frac{1}{\sigma}\frac{d\sigma}{d\lvert y_{t\bar{t}}\rvert d\lvert m_{t\bar{t}}\rvert}$	$\lvert y_{t\bar{t}}\rvert$	$[0.0, 0.2, 0.4, 0.6, 0.9, 1.3, 2.3]$
	$m_{t\bar{t}}$	$[300, 375, 450, 625, 850, 2000]$ GeV
$\frac{1}{\sigma}\frac{d\sigma}{dp_\perp^{t,h} d\lvert y_t^h\rvert}$	$p_\perp^{t,h}$	$[0, 45, 90, 135, 180, 225, 270, 315, 400, 800]$ GeV
	$\lvert y_t^h\rvert$	$[0.0, 0.5, 1.0, 1.5, 2.5]$

Table 1: Distributions included in the fit in this work.

Baseline Analysis bounds

	2.3 fb^{-1}		3 ab^{-1}	
	Individual	Profiled	Individual	Profiled
$\bar{C}_G$	$(-0.0543, 0.0535)$	$(-0.1736, 0.1731)$	$(-0.0015, 0.0015)$	$(-0.0046, 0.0046)$
$\bar{C}_{\varphi q}^{(3)33}$	$(-0.0317, 0.0326)$	$(-0.0781, 0.0736)$	$(-0.0009, 0.0009)$	$(-0.0021, 0.0021)$
$\bar{C}_{uG}^{33}$	$(-0.0253, 0.0247)$	$(-0.0602, 0.0635)$	$(-0.0007, 0.0007)$	$(-0.0017, 0.0017)$
$\bar{C}_{uW}^{33}$	$(-0.0234, 0.0228)$	$(-0.0527, 0.0562)$	$(-0.0006, 0.0006)$	$(-0.0015, 0.0015)$
$\bar{C}_{qq}^{(1)i33i}$	$(-0.0202, 0.0204)$	$(-0.048, 0.0469)$	$(-0.0006, 0.0006)$	$(-0.0013, 0.0013)$
$\bar{C}_{qq}^{(3)i33i}$	$(-0.0101, 0.0102)$	$(-0.024, 0.0234)$	$(-0.0003, 0.0003)$	$(-0.0007, 0.0007)$
$\bar{C}_{qq}^{(3)ii33}$	$(-3.2964, 3.3259)$	–	$(-0.0917, 0.0917)$	$(-0.2982, 0.2955)$
$\bar{C}_{qu}^{(8)33ii}$	$(-0.0867, 0.0875)$	$(-0.2063, 0.2015)$	$(-0.0024, 0.0024)$	$(-0.0056, 0.0056)$
$\bar{C}_{qu}^{(8)ii33}$	$(-0.0577, 0.0583)$	$(-0.1373, 0.1342)$	$(-0.0016, 0.0016)$	$(-0.0038, 0.0038)$
$\bar{C}_{ud}^{(8)33ii}$	$(-0.1598, 0.1613)$	$(-0.3802, 0.3714)$	$(-0.0044, 0.0044)$	$(-0.0104, 0.0104)$
$\bar{C}_{uu}^{i33i}$	$(-0.0225, 0.0228)$	$(-0.0536, 0.0524)$	$(-0.0006, 0.0006)$	$(-0.0015, 0.0015)$
$\bar{C}_{lq}^{(3)ii33}$	–	–	$(-0.3289, 0.3288)$	$(-1.8400, 1.8832)$

Table 2: Baseline 2σ bounds for different luminosities with $\bar{C}_i = C_i \frac{v^2}{\Lambda^2}$.

Improvements on bounds

	2.3 fb^{-1}		3 ab^{-1}	
	Individual	Profiled	Individual	Profiled
$\bar{C}_{G}$	0.07%	14.53%	0.07%	11.72%
$\bar{C}_{\varphi q}^{(3)33}$	33.74%	34.16%	33.73%	33.82%
$\bar{C}_{uG}^{33}$	28.29%	32.12%	28.28%	30.76%
$\bar{C}_{uW}^{33}$	34.86%	35.36%	34.85%	35.57%
$\bar{C}_{qq}^{(1)i33i}$	3.50%	3.52%	3.50%	3.23%
$\bar{C}_{qq}^{(3)i33i}$	4.35%	4.31%	4.35%	5.01%
$\bar{C}_{qq}^{(3)ii33}$	63.83%	–	63.83%	72.06%
$\bar{C}_{qu}^{(8)33ii}$	3.45%	3.45%	3.45%	3.39%
$\bar{C}_{qu}^{(8)ii33}$	3.74%	3.80%	3.74%	3.77%
$\bar{C}_{ud}^{(8)33ii}$	4.62%	4.63%	4.62%	4.64%
$\bar{C}_{uu}^{i33i}$	3.38%	3.41%	3.38%	3.83%
$\bar{C}_{lq}^{(3)ii33}$	–	–	10.57%	40.26%

Table 3: Maximum improvements in 2σ bounds via a cut on the ML score.

Comments on the results

- GNN performs well in discriminating non-resonant top decay contributions.
- Sizeable improvement when momentum enhancement is present.
- Operators with small improvements are relatively under control via the inclusive rate and baseline selection.
- Improvements on profiled bounds can be greater than individual ones since a cut on the EFT score can select a region where the impact of other operators is reduced.
- Improvements should generalise to Λ^{-4} terms of cross-section expansion.

Conclusion

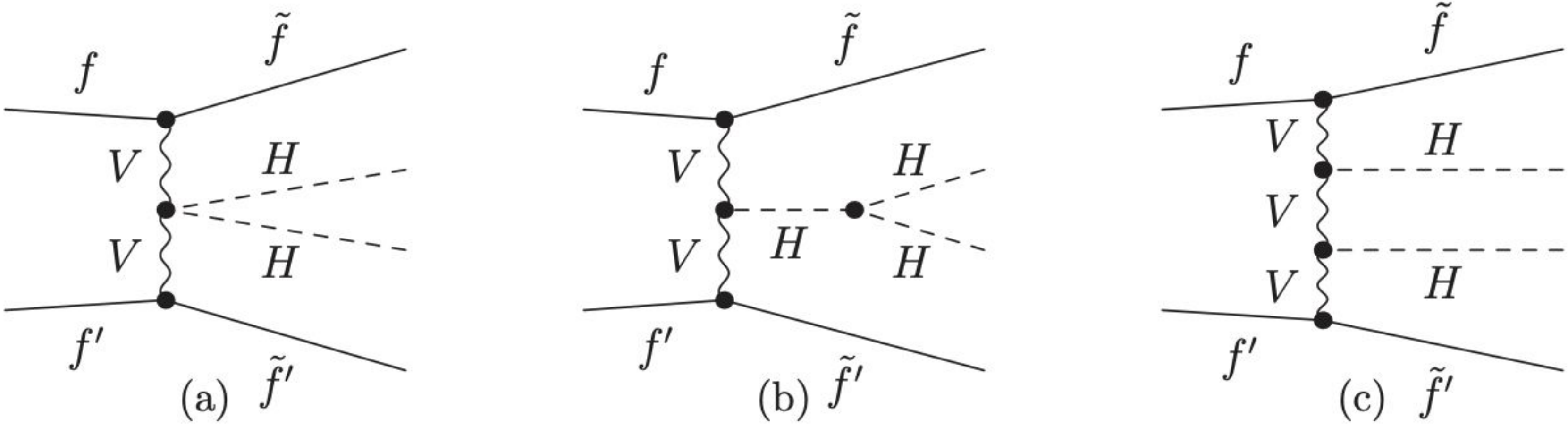
- Highly non-trivial task to design representation/algorithms which would achieve optimal knowledge of the background
- Graphs are an efficient way to represent jets, with the ability to incorporate relational (via edge-features) information between constituents
- Graph autoencoders can learn both local as well as global features(via edge-reconstruction) of QCD jets thereby making it a “promising candidate”
- Shown to be robust to complexity bias with the added benefit of an efficient representation with no inherent Euclidean bias
- Integration of physics-knowledge very much important to achieve the goal of learning non-trivial topologies of collider events

Thank you

Back up

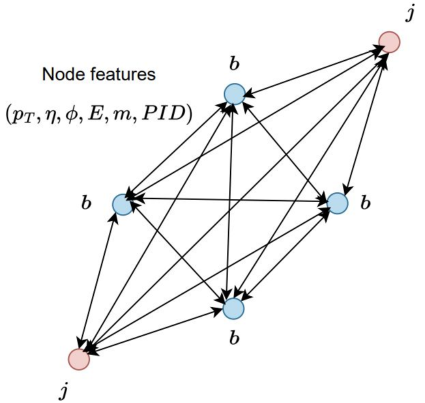
Applications to HEFT:

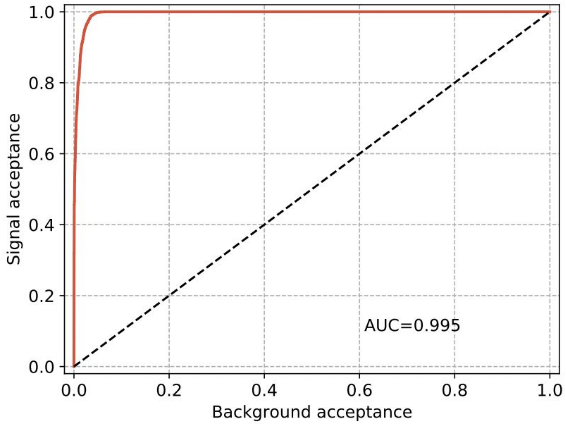
- Quartic Gauge-Higgs couplings:

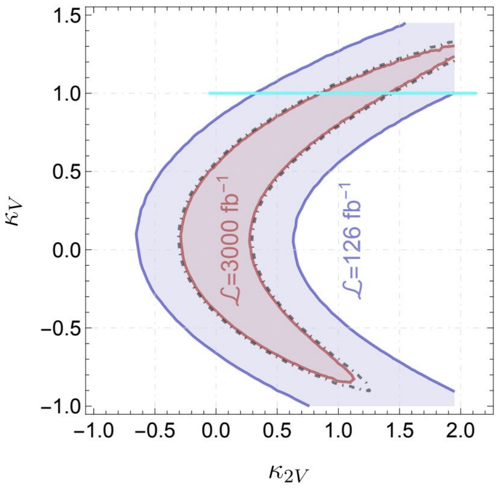


Feynman diagrams contributing to WBF-production of di-Higgs

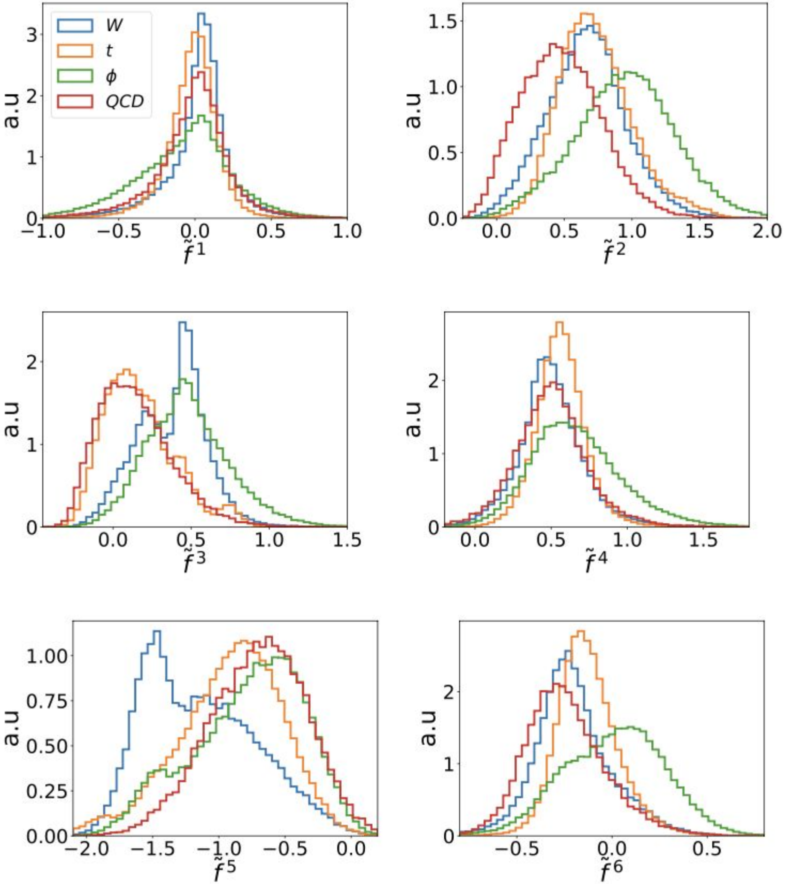
Fully connected Graph : constrainting the κ_{2v}





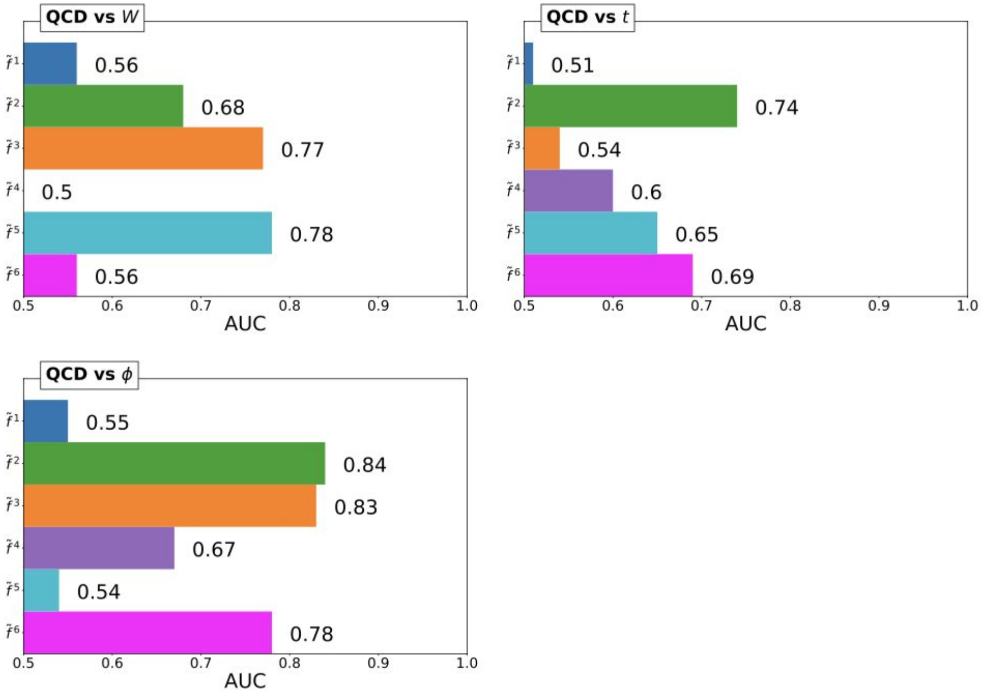


Latent graph-representation



$$\tilde{f}^a = \frac{1}{N} \sum_{i \in G} f_i^a$$

Latent graph-representation



Jet dataset details

$\sqrt{s} = 13$ TeV pp collisions, `MadGraph5`

QCD jets (Training and validation): dijet events

Signal benchmarks (Testing):

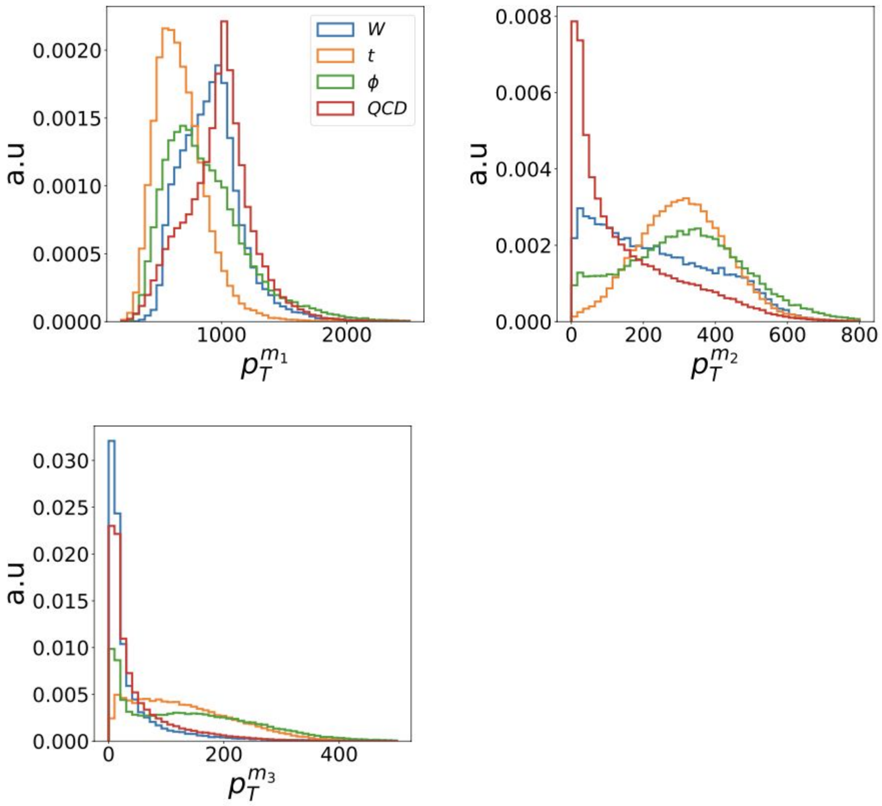
(i) boosted hadronically-decaying W bosons
(ii) boosted hadronically-decaying top quarks
(iii) a boosted scalar ϕ decaying as $\phi \to W^+W^- \to 4j$, with $m_\phi = 700$ GeV

Jets definition and cuts:

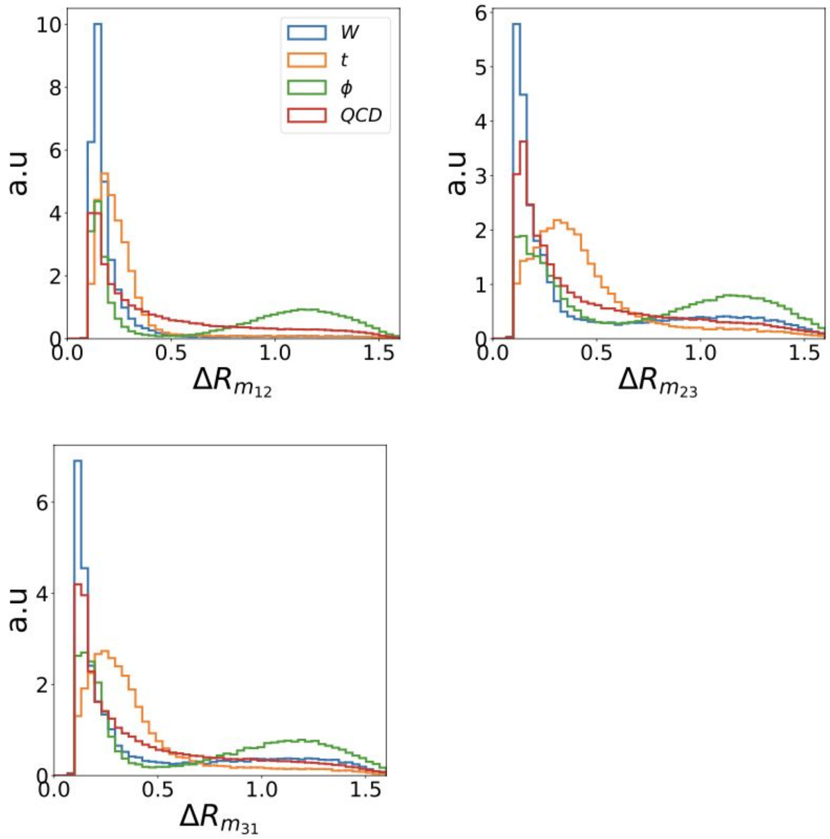
- anti-k_t algorithm with $R = 1.5$ with `FastJet`
- use final state particles after showering and hadronization with `Pythia8`
- Require $|y| < 2.5$ and $p_T > 1$ TeV
- Select hardest p_T jet from each event.

Final input for graph construction: reclustered with anti-k_T jet algorithm into microjets with $R = 0.1$ and $p_T \geq 5$ GeV

Feature distribution of Microjets



Feature distribution of Microjets



NN Conv [1704.01212]

Message-passing:

$$ {}^{ab}m_{ij}^{(1)} = {}^{ab}F_w(\vec{e}_{ij}) \times {}^{ab}\tilde{h}_j^{(0)} \,, $$

${}^{ab}\tilde{h}_j^{(0)}$ is formed by repeating $\vec{h}_j^{(0)}$ n times.
F_w=Edge-function(a Neural network)

Node-readout: Takes the mean of ${}^{ab}m_{ij}^{(1)}$ over all neighbouring nodes j, and then sums over the a index of the matrix:

$$ {}^{b}h_i^{(1)} = \sum_a \mathrm{mean}_{j\in\mathcal{N}(i)} \left(\left\{ {}^{ab}m_{ij}^{(1)} \right\}\right) \,, $$

Edge Conv [1704.06199]

Message-passing:

$$\vec{m}_{ij}^{(l)} = \Theta_w.(\vec{h}_j^{(l)} - \vec{h}_i^{(l)}) + \Phi_w.\vec{h}_i^{(l)} \,,$$

Θ_w and Φ_w are weights

Node-readout:

$${}^a h_i^{(l+1)} = \max_{j \in \mathcal{N}(i)} \{ {}^a m_{ij}^{(l)} \} \,,$$