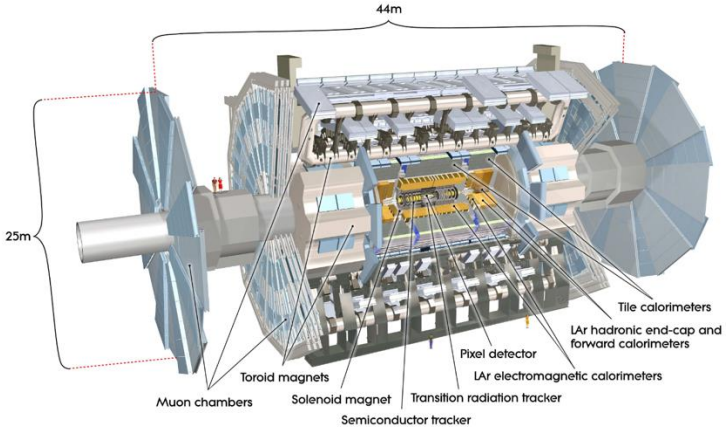
UNSUPERVISED MACHINE LEARNING FOR ANOMALY DETECTION IN LHC COLLIDER SEARCHES

ANTONIO D'AVANZO, ON BEHALF OF THE ATLAS COLLABORATION

EuCAIFCon 2025, Cagliari, 18/06/2025

Motivation: Beyond Standard Model physics in ATLAS

- **ATLAS** (**A** **T**oroidal **L**HC **A**pparatu**S**) is a general-purpose detectors at the Large Hadron Collider (LHC) at CERN
  - Designed to exploit its full discovery potential due to **Standard Model open questions**, i.e. dark matter, gravity ecc.

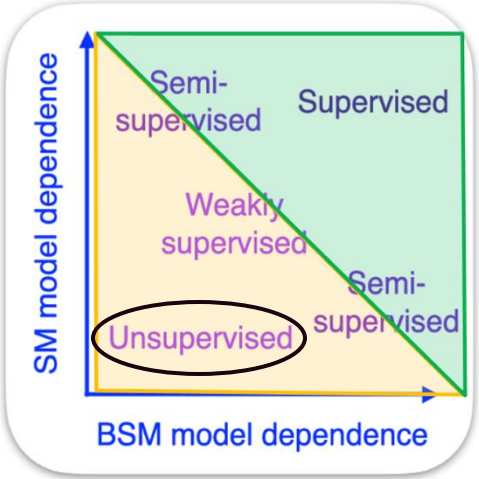


Classic ATLAS search: model dependent

1. A new well motivated physics-scenario is chosen
2. Event selection based on physics signatures (reduce background - keep signal with aid from MonteCarlo)
3. Discovery or constraints are set on process
   - Unlikely to be sensitive to different processes

Recently popular ATLAS searches: model independent

1. Minimal assumptions of signal properties
2. Check for deviations from background-only hypothesis (**unsupervised task using only data!**)
   - Not sensitive as model-dependent, but wider scope

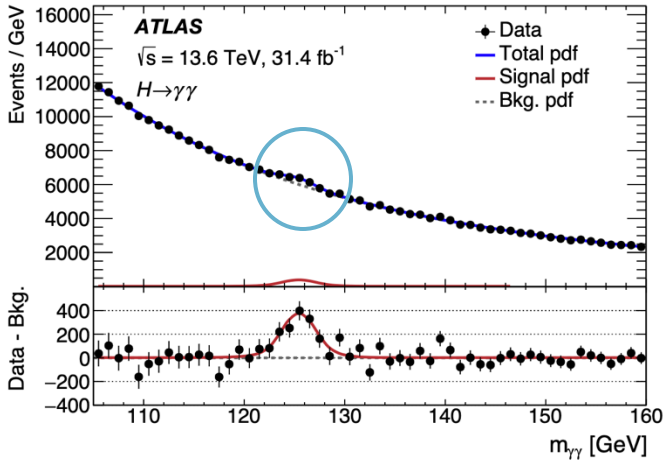


Unsupervised ML: Anomaly Detection

- Anomaly Detection (AD) refers to Machine Learning (ML) techniques used to spot outliers in a dataset.

- Identification of features of detector data inconsistent with the expected background.
  - Seminal work in ATLAS: Phys. Rev. D 108, 052009

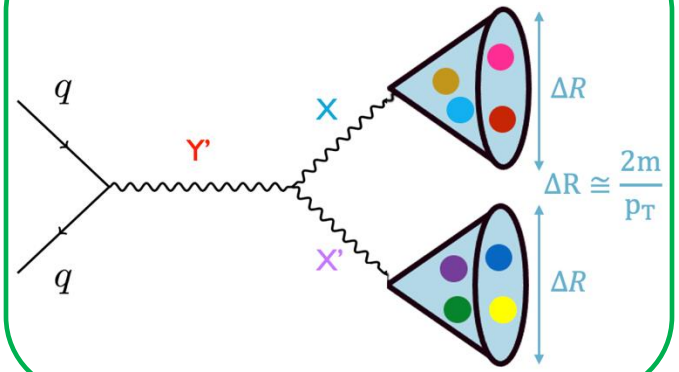
Eur.Phys.J.C 84 (2024) 1, 78



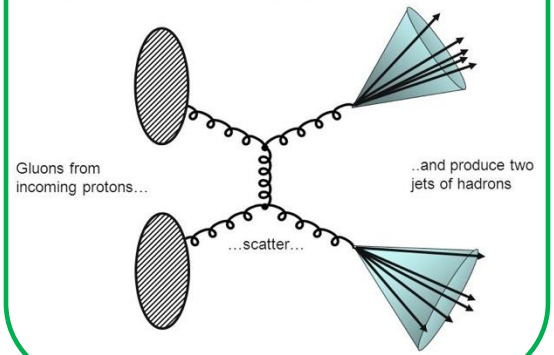
Physical processes treated as anomaly in this talk

- Decay of new particles in **fully hadronic** final states from pp collisions
  - Jets defined from calorimeter energy deposits (constituents)

Targeted signal



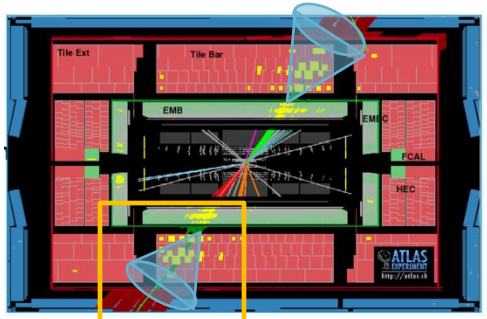
Expected background

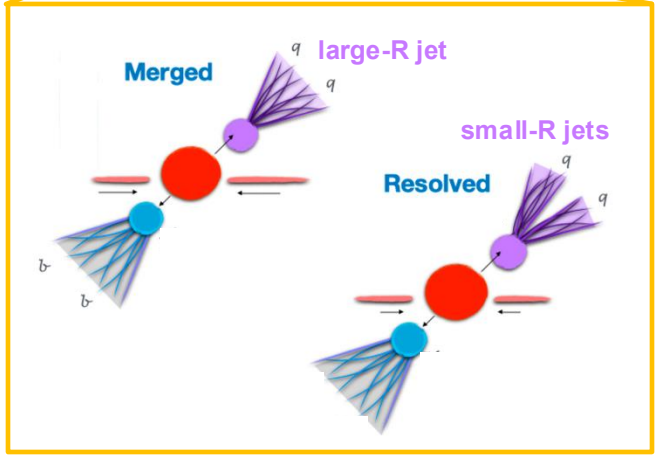


Y' mass (~ TeV) >> X – X' masses (~ O(10^2) GeV)

Boosted topology

Hadronic calorimeter

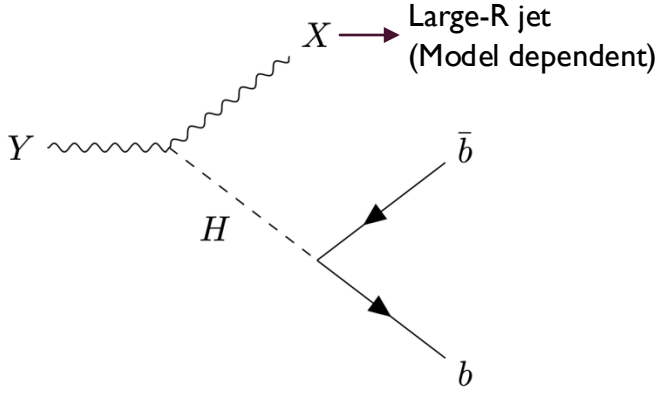


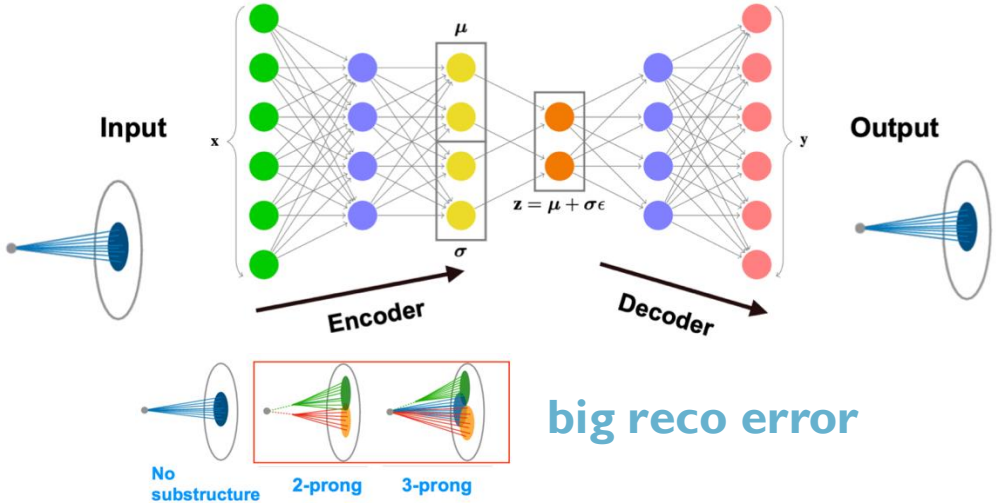


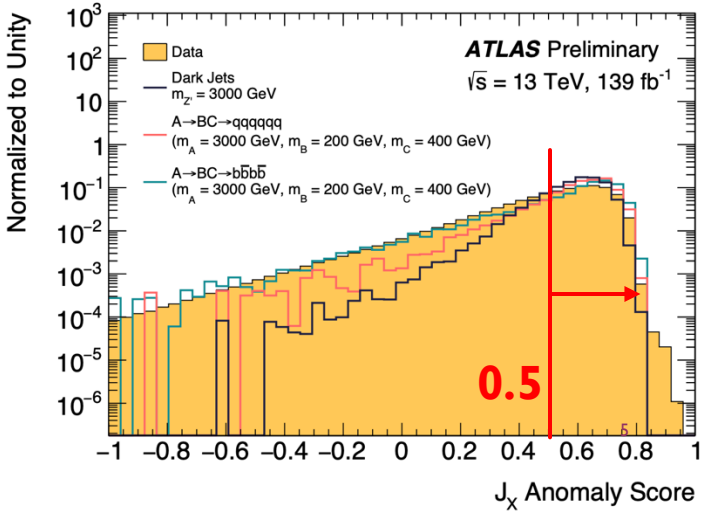
Seminal work in ATLAS: Y → XH SEARCH

Phys.Rev.D 108 (2023) 052009

- Search for a heavy-mass resonance **Y** decaying in a Higgs boson ($\mathbf{H \to b\bar{b}}$) and a new particle X in the **fully hadronic** channel
- First ATLAS publication with a fully unsupervised (model-independent) Anomaly Detection approach based on Variational Recurrent Neural Network (VRNN)
  - Trained on data-only over **contituents** (modeled as a sequence of four-vectors) of jets with p_T > 1.2 TeV
- Anomaly score computed from **VRNN output** to select X boson
  - **Sensitive to alternative X decay hypothesis other than 2-prong** (e.g. three-prong and dark jet)



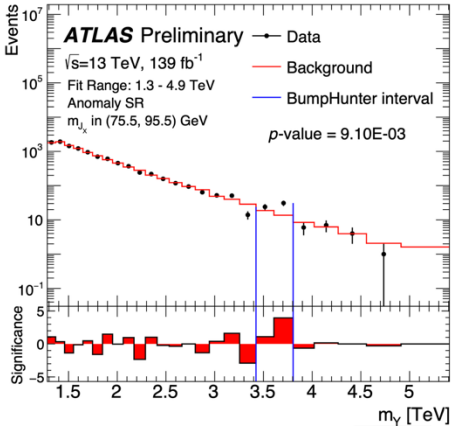


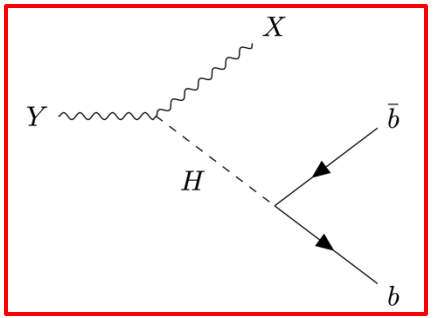


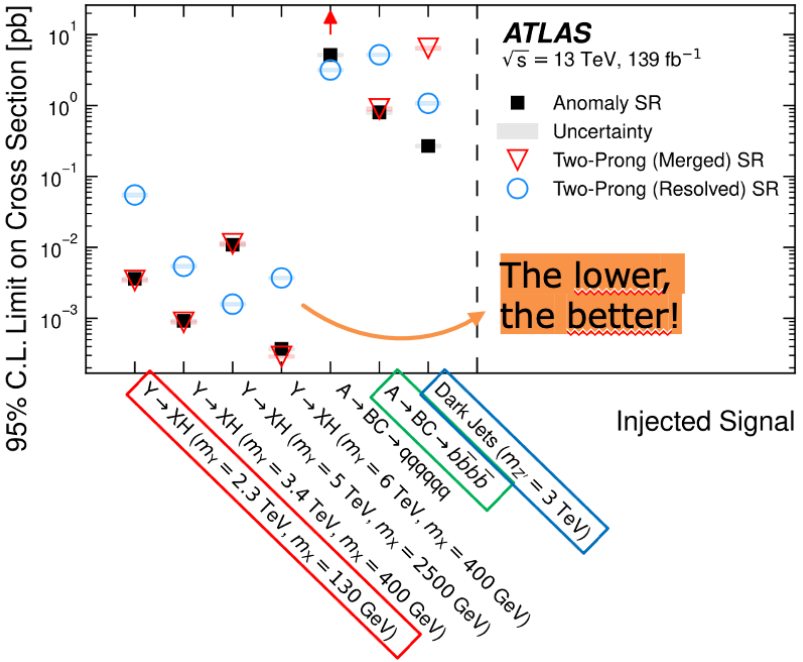
Results

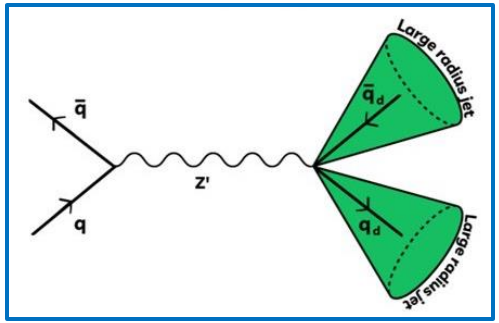
Phys.Rev.D 108 (2023) 052009

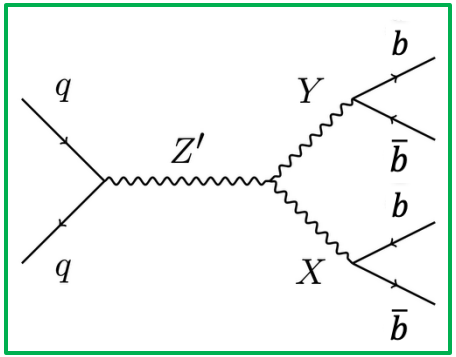
- Final fit to look for excess performed on final state invariant mass distribution m_{jj} of data, repeated several times in overlapping bins of the X candidate mass
  - No excess found, constraints on production cross section $\sigma(Y \rightarrow XH \rightarrow qqbb)$ set







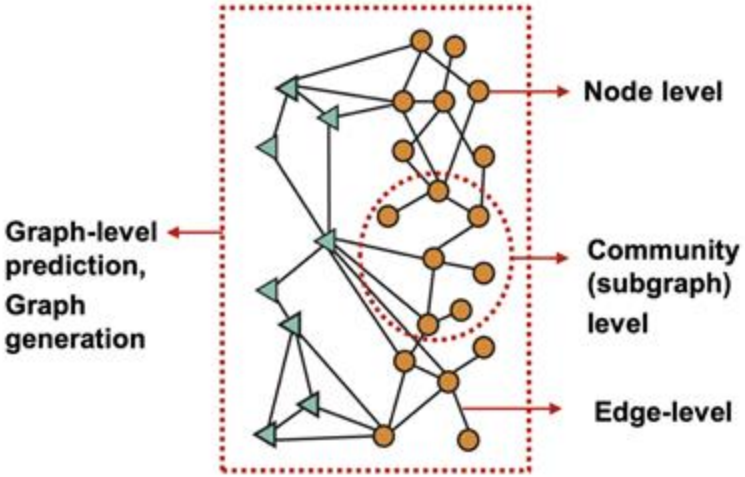


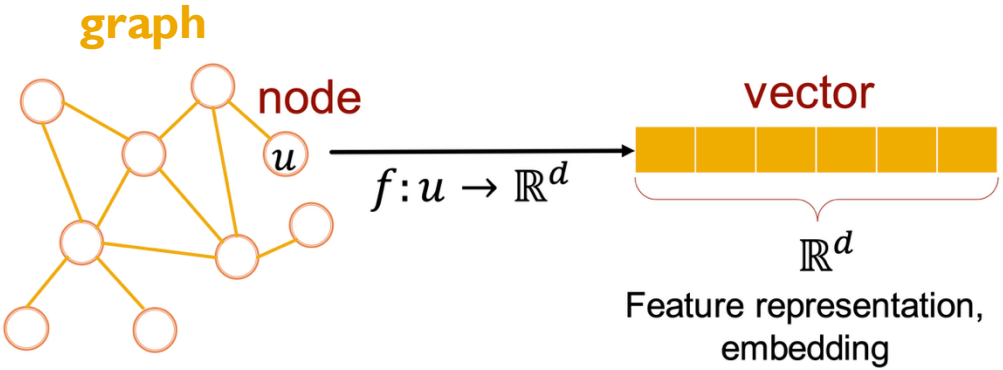


New approach in ATLAS: Graphs

- Data should sometimes be arranged in other forms other than vectorial (protein chains, social networks ecc.)
  - **Graph representation**: nodes (entities) and edges (connections)
  - Nodes and edges typically contain **features** specific to each element and each pair

- **Graph Neural Networks** (GNNs) are ML architectures built specifically to make predictions on graphs, exploiting their relational nature.
  - Training used to learn the vector representation (embedding h_v) of each node of the input graphs by a **message passing** mechanism.

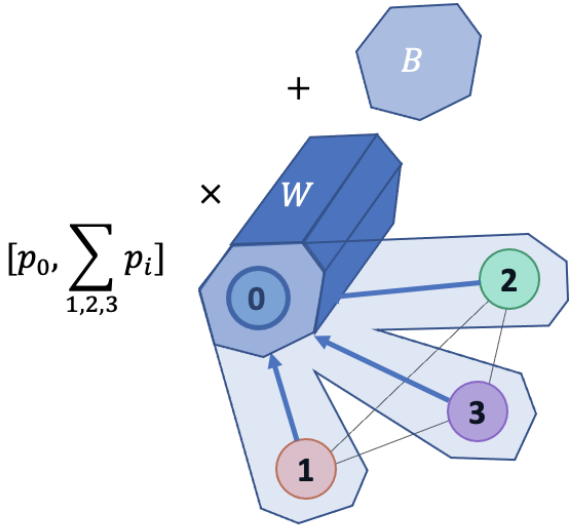


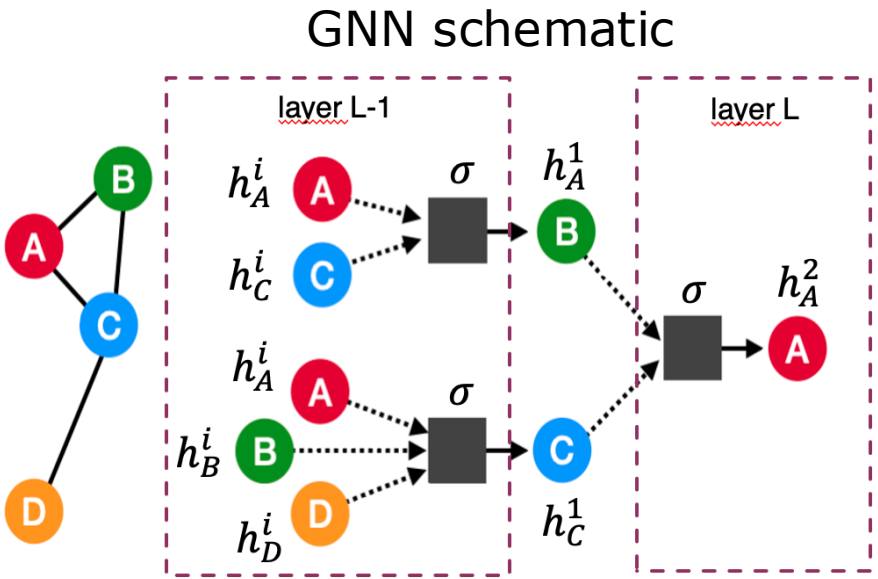


Several task levels, carried out by processing the final node embeddings in certain ways.

Graph Neural Networks (GNNs)

- The embeddings are updated at each layer by aggregating the information passed between the **target node** and the nodes from its closest neighbourhood → message passing



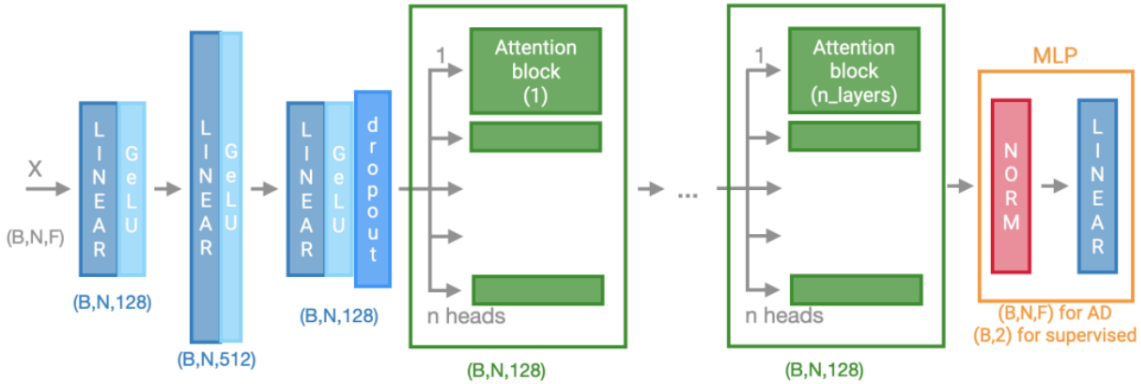


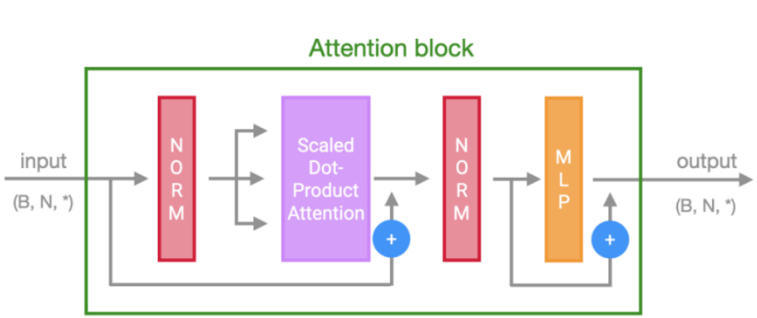
- G embedding is obtained by pooling the nodes embedding at the final layer into one global representation
  - Global sum pooling: $h_G = \mathrm{Sum}(\{h_v^L \in \mathbb{R}^d, \forall v \in G\})$
  - Global mean pooling: $h_G = \mathrm{Mean}(\{h_v^L \in \mathbb{R}^d, \forall v \in G\})$
  - Global max pooling: $h_G = \mathrm{Max}(\{h_v^L \in \mathbb{R}^d, \forall v \in G\})$

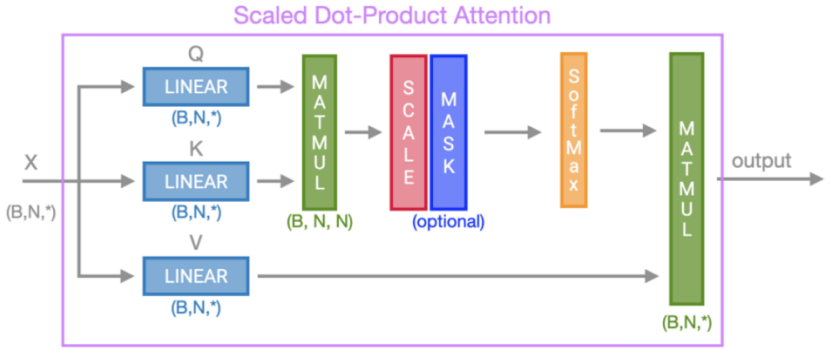
Transformer

- Deep Learning architectures that require classical vectorial input of size (B,N,F)
  - **Equivalent to fully connected graph input to GNN!**
- Based on Attention Mechanism, robust and fast to train

B = batch size
N = number of objects
F = number of features





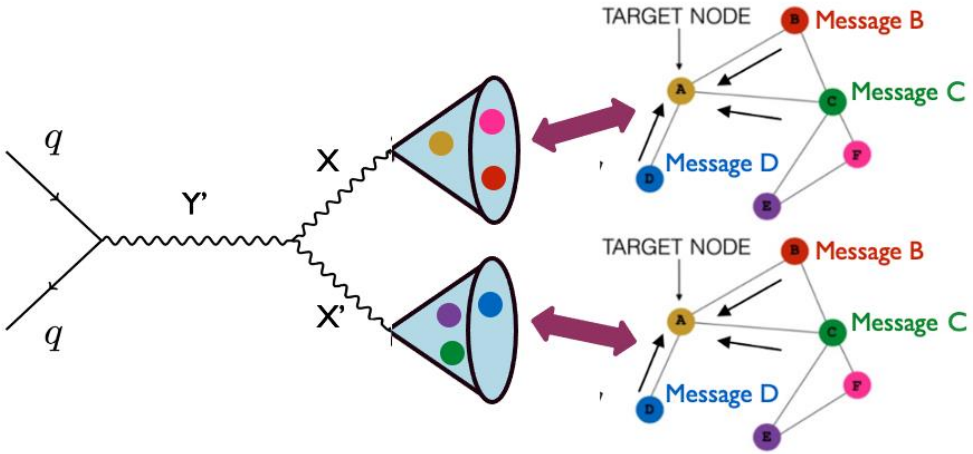


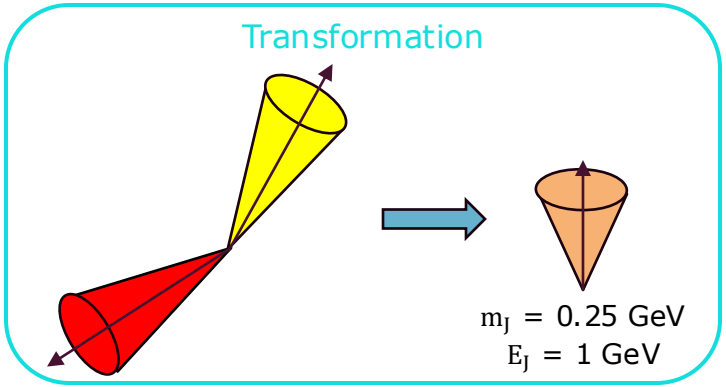
The idea: graphs are the new jets

- Model agnostic search for new physics in fully hadronic final states with the ATLAS detector using graph neural networks (GNNs)

- Only signal assumption: 2 boosted Large-R jets per event (Anti-kT algorithm with R = 1)
  - **Jets have sparse structure, suitable for graph representation exploiting low level features! (jets constituents)**

- Graphs have messages:
  - **Nodes** ≡ constituents → [pT fraction, η, ϕ] features
  - **Edges** ≡ relations → 1/ΔR features, exist if ΔR < 0.2





Graph-level embedding by GNN/Transformer with propagated messages
- Data augmented for mass decorrelation (transformed constituents)

Anomaly Detection strategy

- **Key concept**: Unsupervised training on data (mostly QCD background)
- Form of discriminant Anomaly Score s(x) per jet depends on **the considered ML architecture**

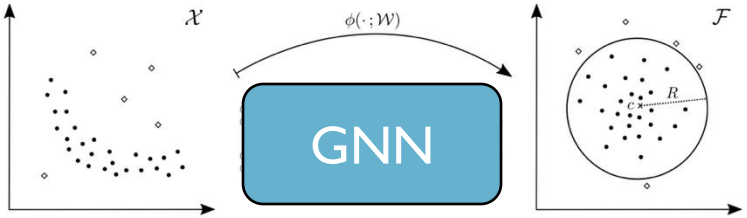
Graph Neural Network

input space

● = background
◇ = signal

C = average of initial graph features vectors

output space (Deep SVDD)



GNN maps graph features from parameters space X → F by Deep Support Vector Data Description (SVDD) objective

$$\min_{W} \frac{1}{N}\sum_{i=1}^{N} \|\mathrm{GIN}(G_i;W)-\mathrm{c}\|^2 \qquad s(\boldsymbol{x}) = \|\phi(\boldsymbol{x};\mathcal{W}^*)-\boldsymbol{c}\|^2$$

Transformer

Employes simple MSE loss minimization

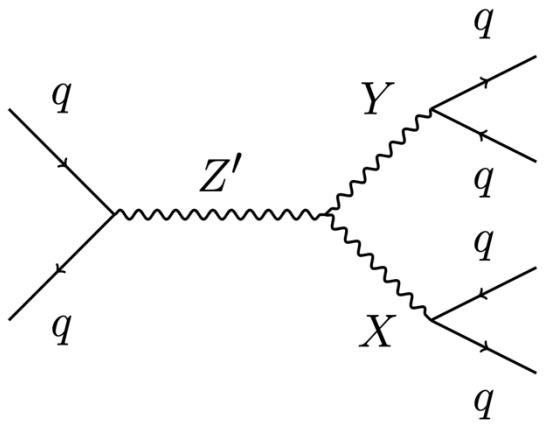
$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2$$

$$s(x) = \|y(x;W^*) - x\|$$

R&D results on open data

https://lhco2020.github.io/homepage/

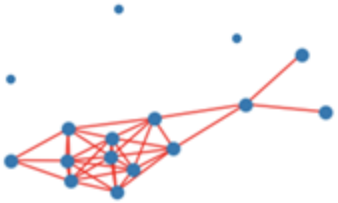
2 prong signal benchmark



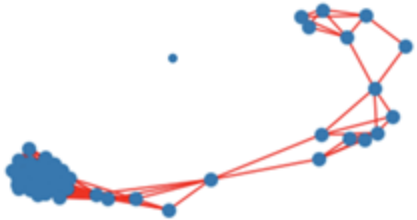
Toy model

R&D LHC Olympics dataset

- QCD dijet events as background
- $Z' \rightarrow XY \rightarrow qqqq$ signal events
- $m_{Z'} = 3.5$ TeV, $m_X = 500\ GeV$, $m_Y = 100\ GeV$
- Reconstructed with anti-k_T with R = 1.0



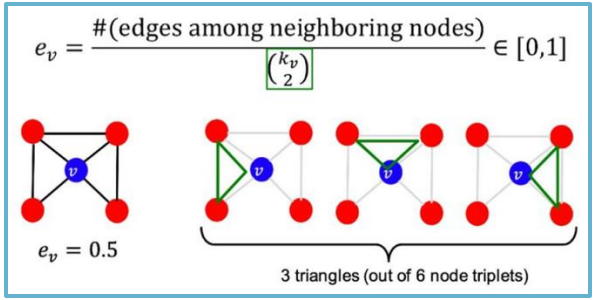
QCD dijet

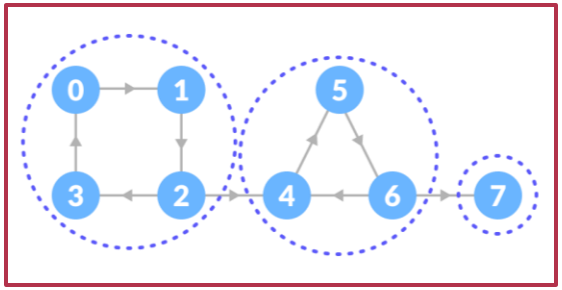


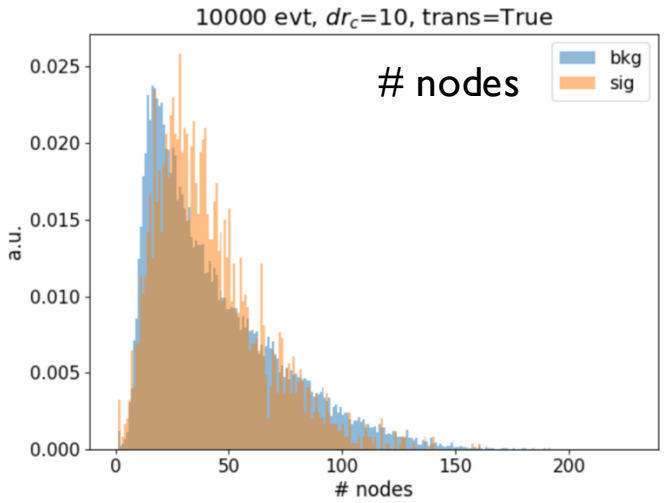
$Z' \rightarrow XY \rightarrow qqqq$

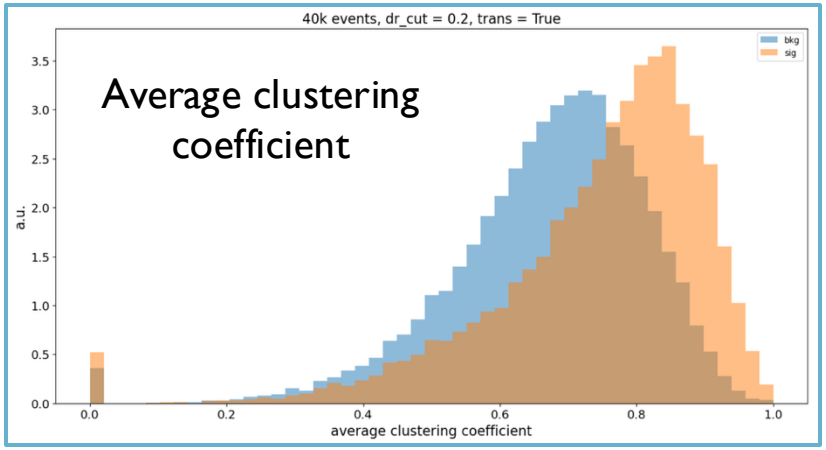
Graphs characteristics

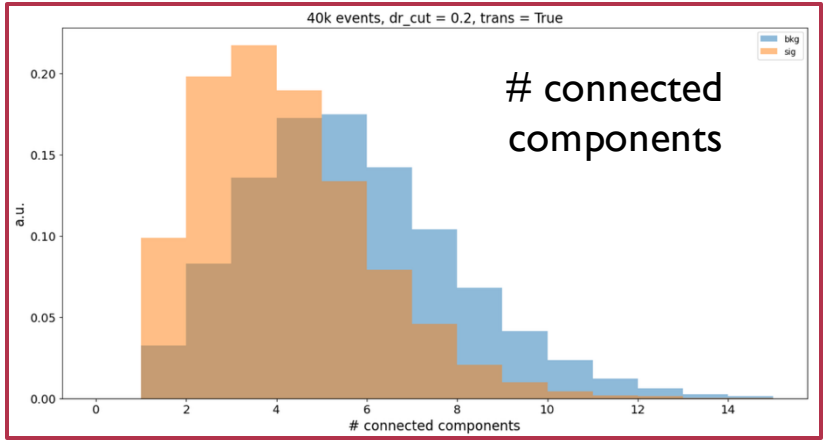
- Graphs are very powerful to understand and relate structure quantities to **Z' signal** and **QCD background** jets











Technical infrastructure

- **Graph Neural Network** trained on INFN Naples IBiSCo (**I**nfrastructure for **BI**g data and **S**cientific **CO**mputing) GPUs cluster
  - Provided with a total of 6 (nodes) x 2 GPUs
  - CUDA support allows the use of Pytorch tensors allocated on GPU, thus speeding up the training process

- **Transformer** trained on University of Rome La Sapienza computing resources, also provided with GPUs
  - Similar architecture to GNN, only different number of layers (3)
  - Training time: ~ 50s per epoch

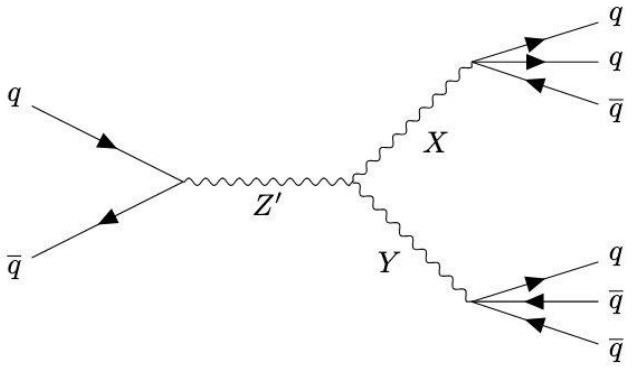
Graph Neural Network

Architecture	1 MLP (3 layers) → 5 layers GNN → 1 MLP (3 layers)
Loss	DeepSVDD
Layer dimension	128
Dataset size	1.1 M (1M background : 100k signal)
Dataset split	Training: 20% (background only), Validation: 1%, test: 79%
Batch size	1024
Training time	~ 100s/epoch
Output level	jet-event

Preliminary results on open data

other benchmark models

- 3-prong signals with same masses
- Anomaly detection event score computed as the mean value of AS(Jet1) and AS(Jet2)

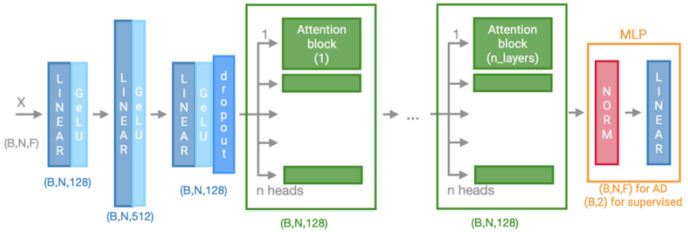


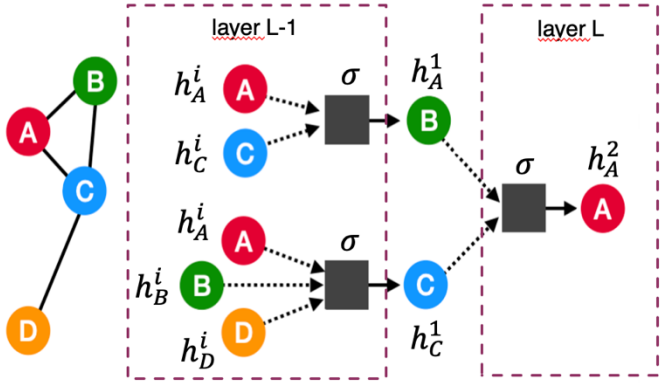
Model	Transformer supervised	GIN supervised	EGAT supervised	Transformer *un*supervised	GIN *un*supervised	EGAT *un*supervised
loss	CrossEntropy	CrossEntropy	CrossEntropy	MSE	DeepSVDD	DeepSVDD
AUC jet-level 2prong	91.3%	90.2%	89.9%	75.5%	73.7%	75.5%
AUC event-level 2prong		96.5%	96.5%		79.6%	81.8%
AUC jet-level 3prong	86.8%	75.5%	84.8%	69.1%	52.6%	67.2%
AUC event-level 3prong		84.1%	92.4%		54%	74.3%

Preliminary results on open data

other benchmark models

- 3-prong signals with same masses
- Anomaly detection event score computed as the mean value of AS(Jet1) and AS(Jet2)





Model	Transformer supervised	GIN supervised	EGAT supervised	Transformer *un*supervised	GIN *un*supervised	EGAT *un*supervised
loss	CrossEntropy	CrossEntropy	CrossEntropy	MSE	DeepSVDD	DeepSVDD
AUC jet-level 2prong	91.3%	90.2%	89.9%	75.5%	73.7%	75.5%
AUC event-level 2prong		96.5%	96.5%		79.6%	81.8%
AUC jet-level 3prong	86.8%	75.5%	84.8%	69.1%	52.6%	67.2%
AUC event-level 3prong		84.1%	92.4%		54%	74.3%

GIN (Graph Isomorfism Network) → Most possibly expressive GNN

EGAT (Edge Graph Attention Network) → GNN with attention mechanism and edge weights updating

Preliminary results on open data

other benchmark models

- 3-prong signals with same masses
- Anomaly detection event score computed as the mean value of AS(Jet1) and AS(Jet2)

Model	Transformer supervised	GIN supervised	EGAT supervised	Transformer *un*supervised	GIN *un*supervised	EGAT *un*supervised
loss	CrossEntropy	CrossEntropy	CrossEntropy	MSE	DeepSVDD	DeepSVDD
AUC jet-level 2prong	91.3%	90.2%	89.9%	75.5%	73.7%	75.5%
AUC event-level 2prong		96.5%	96.5%		79.6%	81.8%
AUC jet-level 3prong	86.8%	75.5%	84.8%	69.1%	52.6%	67.2%
AUC event-level 3prong		84.1%	92.4%		54%	74.3%

Cross Entropy

$$BCE = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)$$

Mean Squared Error

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2$$

Deep SVDD

$$\min_{W} \frac{1}{N}\sum_{i=1}^{N} \|\mathrm{GIN}(G_i; W) - \mathrm{c}\|^2$$

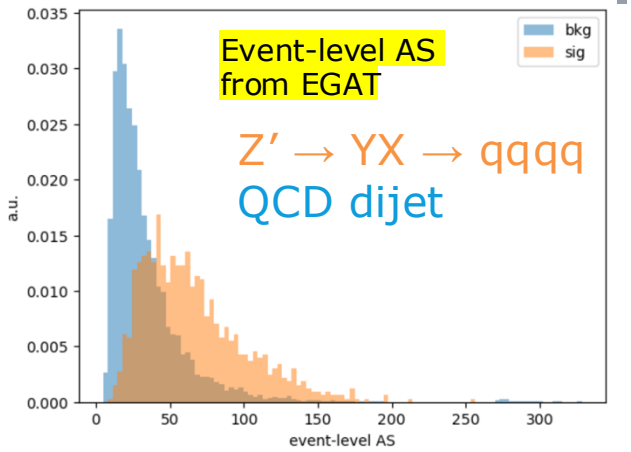
Preliminary results on open data

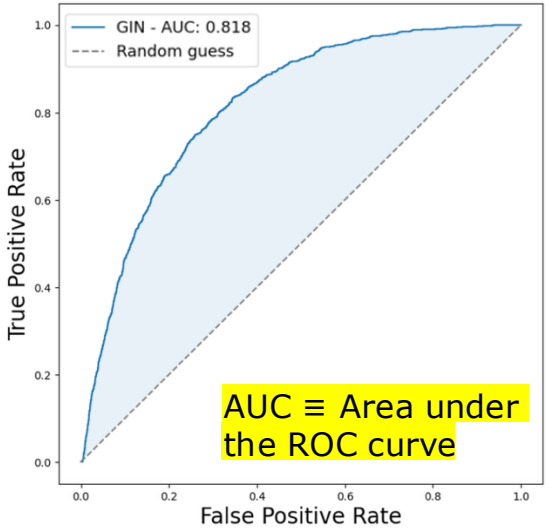
Work in progress

other benchmark models

- 3-prong signals with same masses
- Anomaly detection event score computed as the mean value of AS(Jet1) and AS(Jet2)

Model	Transformer supervised	GIN supervised	EGAT supervised	Transformer *un*supervised	GIN *un*supervised	EGAT *un*supervised
loss	CrossEntropy	CrossEntropy	CrossEntropy	MSE	DeepSVDD	DeepSVDD
AUC jet-level 2prong	91.3%	90.2%	89.9%	75.5%	73.7%	75.5%
AUC event-level 2prong		96.5%	96.5%		79.6%	81.8%
AUC jet-level 3prong	86.8%	75.5%	84.8%	69.1%	52.6%	67.2%
AUC event-level 3prong		84.1%	92.4%		54%	74.3%

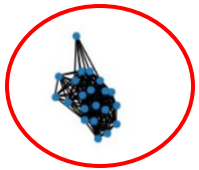


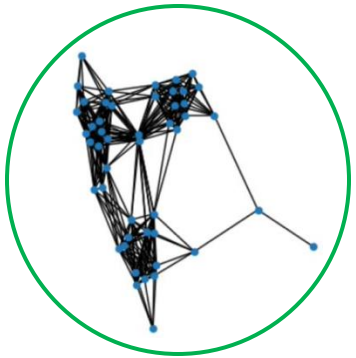


Prospects on ATLAS Run 3 data

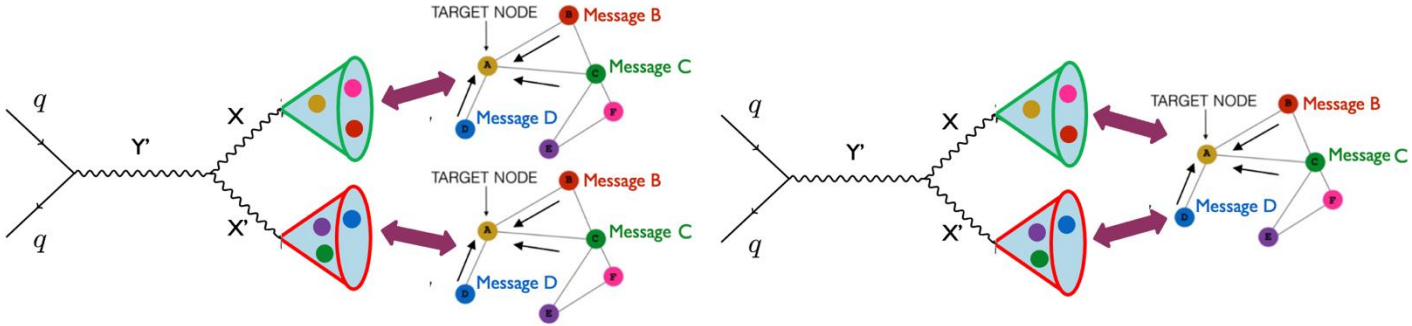
- **High mass resonance search in fully hadronic final states with ATLAS data collected during LHC Run 3 period (2022-2026)**
  - Completely model agnostic
  - 2 large-R jets expected per event, preselection applied to assure optimal trigger efficency
  - Signal region based on Anomaly Score selection
  - More ATLAS oriented input features

Simulated signal event





LHCOlympics R&D archetecture, two approaches for graphs definition



- **Background estimation** with data-driven procedure
  - Functional fit in control region
- **Statistical analysis**
  - Fit on final observable: distribution of invariant mass of dijet system

Conclusions

- No new interactions and particles since the Higgs boson's discovery → more generic searches opposed to the existing model-dependent analysis standard

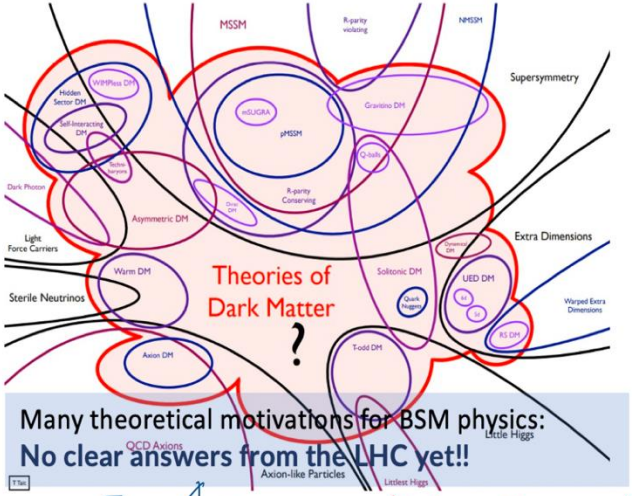
- Model agnostic searches with jets in final state becoming a main topic in the ATLAS collaboration
  - Our effort: **Anomaly Detection with Graph Neural Networks in Run 3**
  - R&D shows promising results on LHCOlympics for AD performance

- **Current effort: ATLAS data!**
  - Our R&D proved to be fruitful when applied on actual Run 3 data
  - Background estimation is the next step
  - Timeline for finalization of statistical inference: mid 2026

Thank you for your attention!

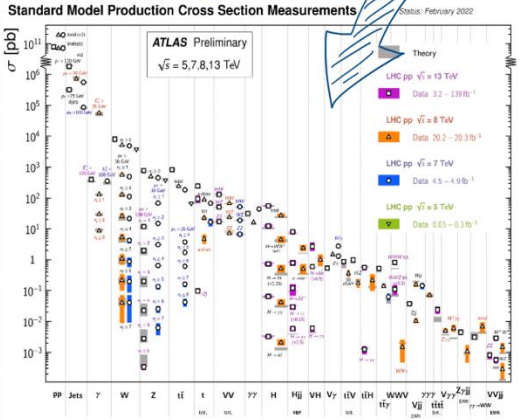
BACKUP

Motivation: Beyond Standard Model physics

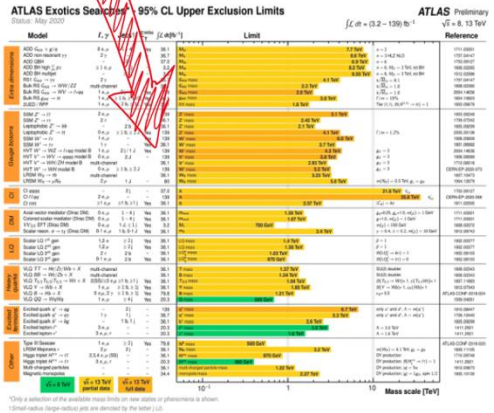


- Standard Model (SM) remarkably predictive of experimental results
  - Discovery of the Higgs boson in 2012 by ATLAS and CMS
- Open questions: dark matter, gravity, hierarchy problem, matter – antimatter asymmetry ecc.

measurements



Searches



Where is new physics? Not trivial

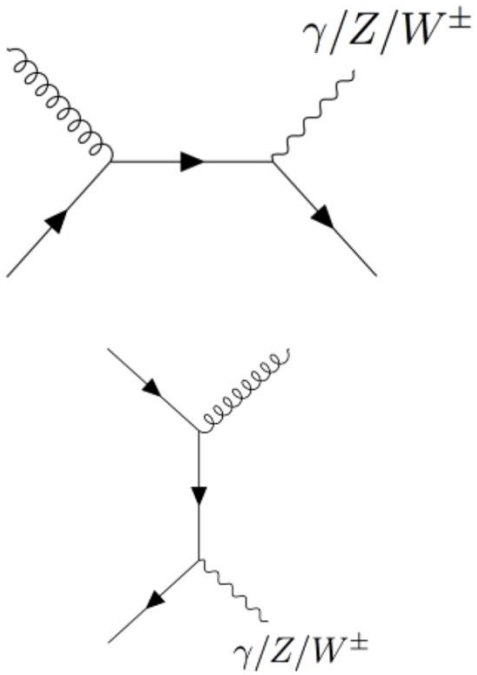
Some hints from previous searches, but no clear direction

Maybe we are looking at the wrong directions, many places to look and time needed to do so

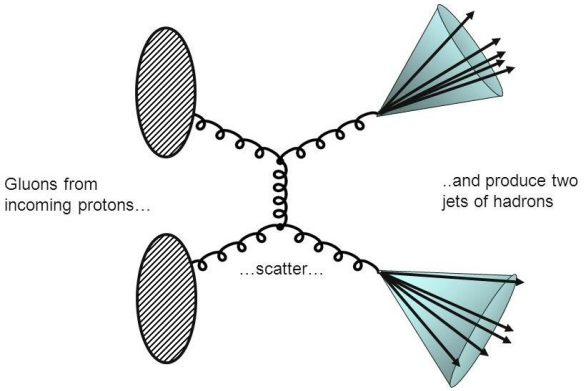
Expected background in pp collision at LHC

- New physics signals could be very difficult to separate from Standard Model background processes with a very similar experimental signature.
  - Montecarlo methods and/or data-driven methods are currently used at LHC to assess the background contribution

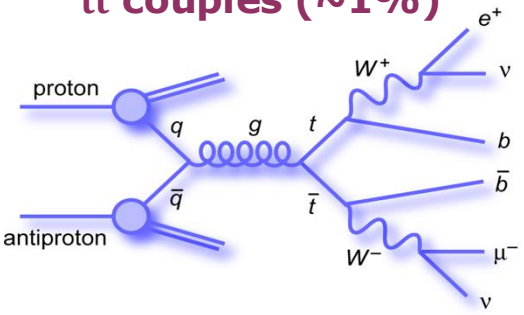
W/Z + jets (~2%)



QCD di-jet (dominating)



$t\bar{t}$ couples (~1%)

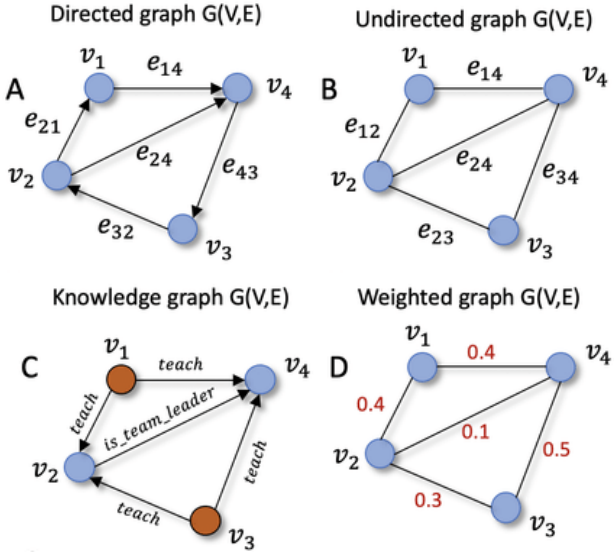


Another approach: graphs

Some data must be arranged in array-like objects in order to be processed by machine learning algorithms, but sometimes it just doesn't feel intuitive (protein chains, social networks between peope, ecc.)

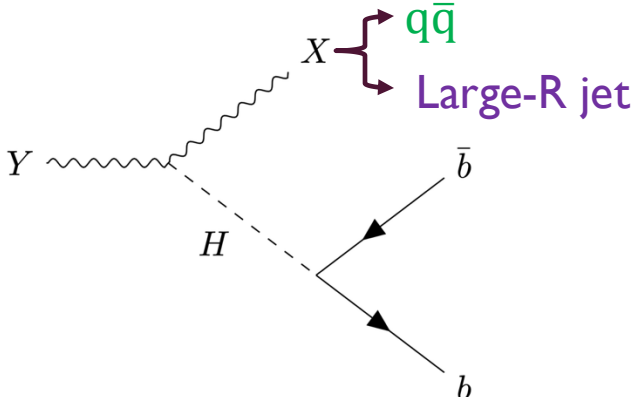
Graph representation!

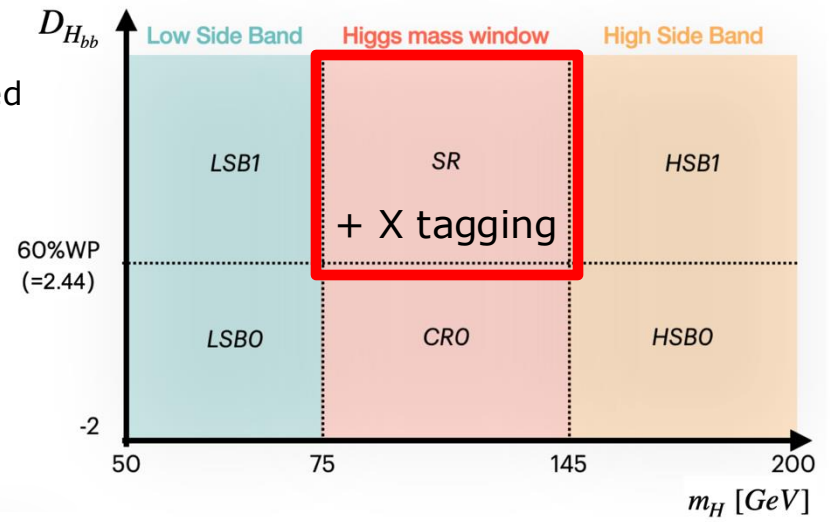
- Structured objects composed of entities used to describe and analyze relations and interactions (edges) between such entities (nodes).
  - Nodes and edges typically contain features specific to each element and each pair



SEMINAL WORK IN ATLAS: Y → XH SEARCH OVERVIEW

- Search for a heavy-mass resonance **Y** decaying in a Higgs boson ($H \to b\bar{b}$) and a new particle X in the **fully hadronic** channel
- Mass range: m_Y in 1 - 6 TeV range, m_X in 65 - 3000 GeV range → boosted regime for H boson
- Event selection on jets follows several steps:
  1. Event preselection
  2. Higgs candidate large-R jet assignment by Deep Neural Network H → bb tagger
  3. H candidate tagging
  4. X tagging, **two scenarios**:
     - Model dependent: 2-prong ($X \to q\bar{q}$) boosted ($m_X/m_Y < 0.3$) and resolved ($m_X/m_Y > 0.3$)
     - Model independent: anomalous X hadronic decay in large-R jet
- Background is estimated fully data-driven via Machine Learning approach using control regions

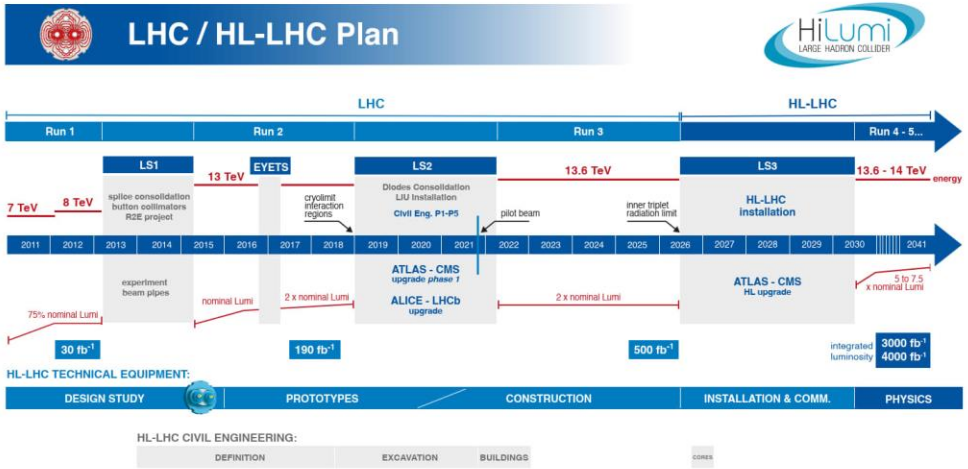




Phys.Rev.D 108 (2023) 052009

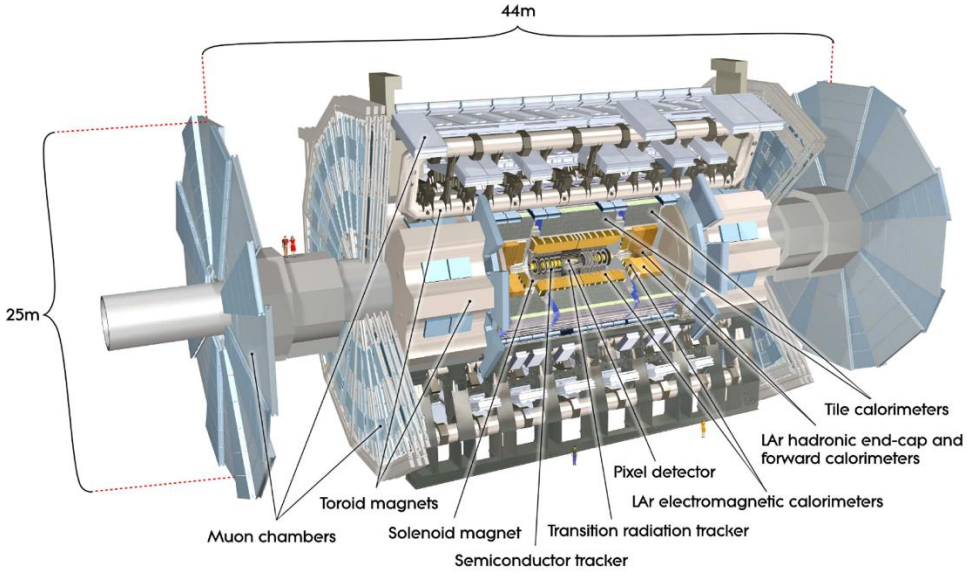
The Large Hadron Collider (LHC)

- The Large Hadron Collider is the world's biggest particle collider, situated at CERN in Geneva
  - p-p collision at $\sqrt{s} = 13.6\ \mathrm{TeV}$ for most of the time, also Pb-Pb collisions at the end of data-taking years
  - Run 3 phase of data taking



ATLAS: A Toroidal LHC ApparatuS

- **ATLAS** is one of the two general-purpose detectors at LHC alongside CMS.
  - With its 7000-tonnes of weight, it is designed to exploit the full discovery potential of the LHC.

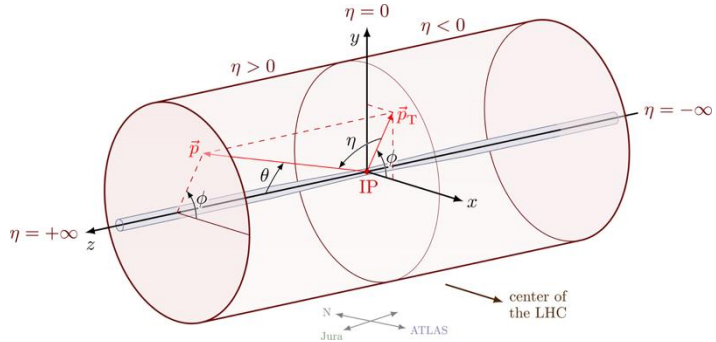


The Lorentz-invariant pseudo-rapidity is used instead of θ:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$$

Distance in the η-ϕ plane:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

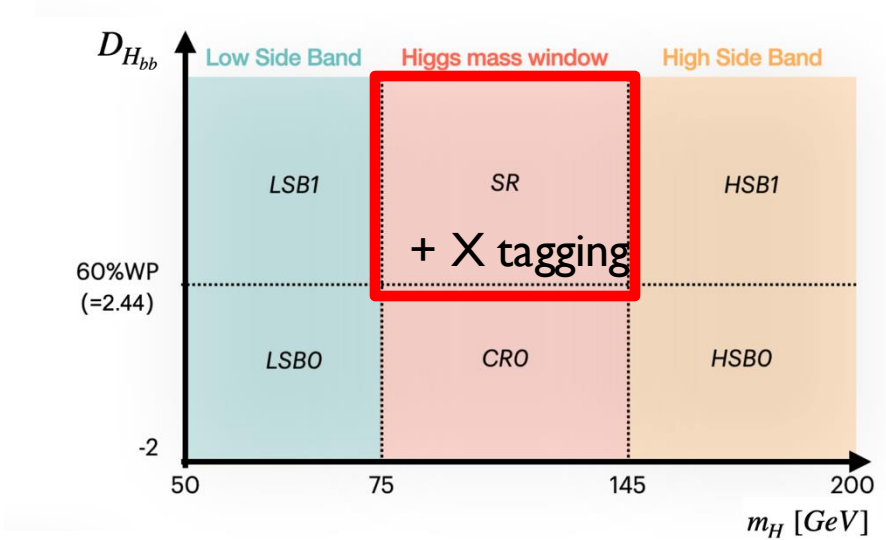


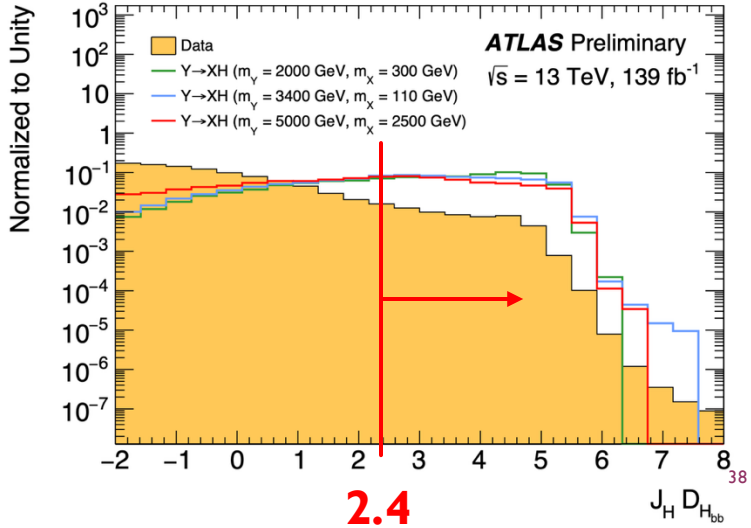
ATLAS NAPOLI

Model independent signal region

Phys.Rev.D 108 (2023) 05

- X and H candidate associated to pT-leading and –subleading jets, ambiguity resolved by $H \to b\bar{b}$ tagger based on Deep Neural Network
  - Discriminant $D_{H_{bb}}$ score computed from NN outputs per jet → H candidate chosen by highest score criteria
- H candidate is further tagged if $D_{H_{bb}} > 2.44$
- X candidate tagged with discriminant from fully data-driven anomaly detection





- EGAT extends on GAT model by implementing edge features in a different way and by allowing updating of the edge weights tensor between each layer of GNN (edge embedding).

GATConv

class **dgl.nn.pytorch.conv.GATConv**(*in_feats, out_feats, num_heads, feat_drop=0.0, attn_drop=0.0, negative_slope=0.2, residual=False, activation=None, allow_zero_in_degree=False, bias=True*) [source]

Bases: `torch.nn.modules.module.Module`

Graph attention layer from Graph Attention Network

$$h_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} \alpha_{i,j} W^{(l)} h_j^{(l)}$$

where α_{ij} is the attention score bewteen node i and node j:

$$\alpha_{ij}^{l} = \mathrm{softmax}_i(e_{ij}^{l})$$
$$e_{ij}^{l} = \mathrm{LeakyReLU}\left(\vec{a}^T [Wh_i \| Wh_j]\right)$$

Returns:

- *torch.Tensor* – The output feature of shape $(N, *, H, D_{out})$ where H is the number of heads, and D_{out} is size of output feature.
- *torch.Tensor, optional* – The attention values of shape $(E, *, H, 1)$, where E is the number of edges. This is returned only when `get_attention` is `True`.

EGATConv

class **dgl.nn.pytorch.conv.EGATConv**(*in_node_feats, in_edge_feats, out_node_feats, out_edge_feats, num_heads, bias=True*) [source]

Bases: `torch.nn.modules.module.Module`

Graph attention layer that handles edge features from Rossmann-Toolbox (see supplementary data)

The difference lies in how unnormalized attention scores e_{ij} are obtained:

$$e_{ij} = \vec{F}(f'_{ij})$$
$$f'_{ij} = \mathrm{LeakyReLU}\left(A[h_i \| f_{ij} \| h_j]\right)$$

where f'_{ij} are edge features, A is weight matrix and

Returns:

- *pair of torch.Tensor* – node output features followed by edge output features The node output feature of shape (N, H, D_{out}) The edge output feature of shape (F, H, F_{out}) where:

  H is the number of heads, D_{out} is size of output node feature, F_{out} is size of output edge feature.

- *torch.Tensor, optional* – The attention values of shape $(E, H, 1)$. This is returned only when :attr: *get_attention* is `True`.

- **Selfloop** is required because of how the node representation is updated.

GRAPH ISOMORPHISM NETWORK (GIN)

- GIN formulation employs both message passing and MLPs, making it the most expressive GNN:

$$\mathrm{MLP}_{\Phi}\left((1+\epsilon)\cdot \mathrm{MLP}_f(c^{(k)}(v)) + \sum_{u\in N(v)} \mathrm{MLP}_f(c^{(k)}(u))\right)$$

learnable parameter

$$c_{\square}^{(k)}(u) \leftrightarrow h_j^{(l)}$$

Embedding of node u (j) al layer k (l)

- This expression can be rewritten in a more general way, also allowing for edge weights to be considered in the graph convolution.

$$h_i^{(l+1)} = f_{\Theta}\left((1+\epsilon)h_i^l + \text{aggregate}\left(\{e_{ji}h_j^l, j\in\mathcal{N}(i)\}\right)\right)$$

- Aggregate can be any permutation invariant function (Sum, Mean, Max ecc.)