Università degli Studi di Napoli Federico II

Scuola Politecnica e delle Scienze di Base

Area Didattica di Scienze Matematiche Fisiche e Naturali

Dipartimento di Fisica "Ettore Pancini"

Machine Learning analysis of $Y \rightarrow XH \rightarrow \overline{q}q\,\overline{b}b$ channel in ATLAS experiment using Graph Representation approach

Relatori:

Prof.ssa Elvira **Rossi**

Prof. Francesco **Conventi**

Dott. Francesco **Cirotto**

Candidato:

Antonio **Corvino**

Matr. N95000**648**

A.A. 2022/2023

Summary:

Motivations;

Machine learning and Graph Theory;

Analysis of fully hadronic final state $(\mathrm{Y} \to \mathrm{XH} \to \mathrm{qqbb})$ with ATLAS data;

Results.

Goal:

To evaluate the effectiveness of the graph representation of jets to train a neural network for signal/background classification.

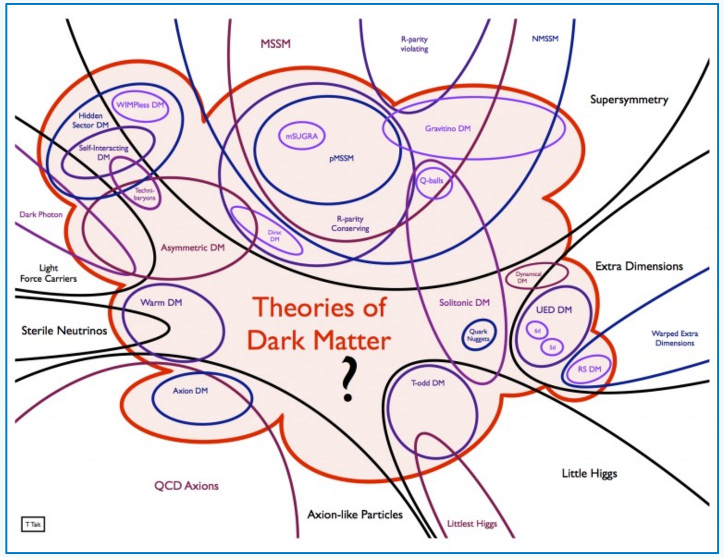
ANOMALY DETECTION (AD) OVERVIEW

Standard Model describes all fundamental particles. However, there are still some questions that remain unanswered (dark matter, neutrino masses, etc.).

Some extensions, called **Beyond Standard Model** (BSM) theories, have been introduced to solve these limits.

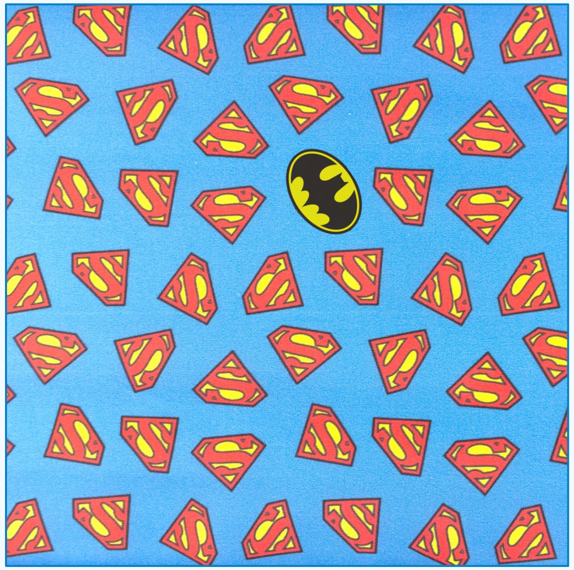
AD techniques can be designed to do experiments with minimal assumptions (model-independency) for BSM searches.

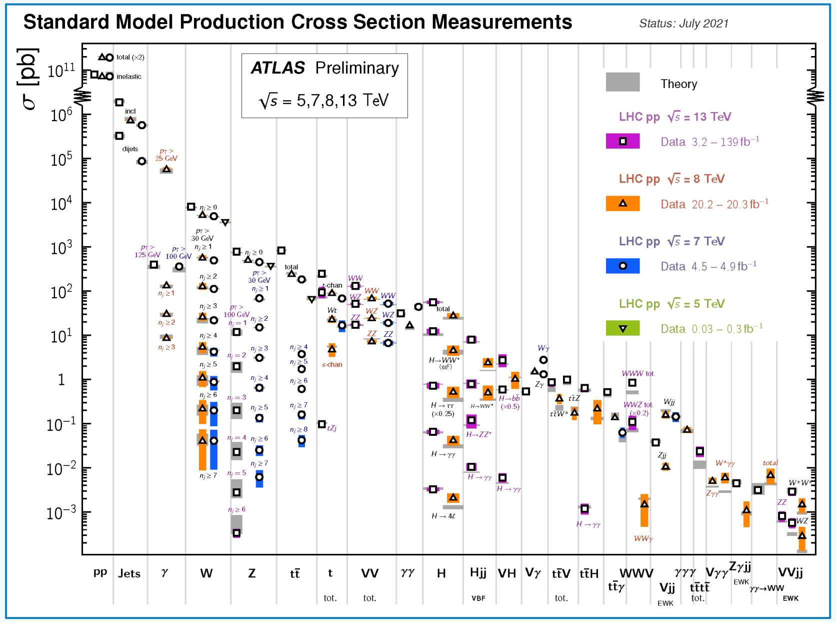
AD aims to identify **unexpected deviations** or **unusual patterns** in data, potentially indicating new information or anomalous behavior.

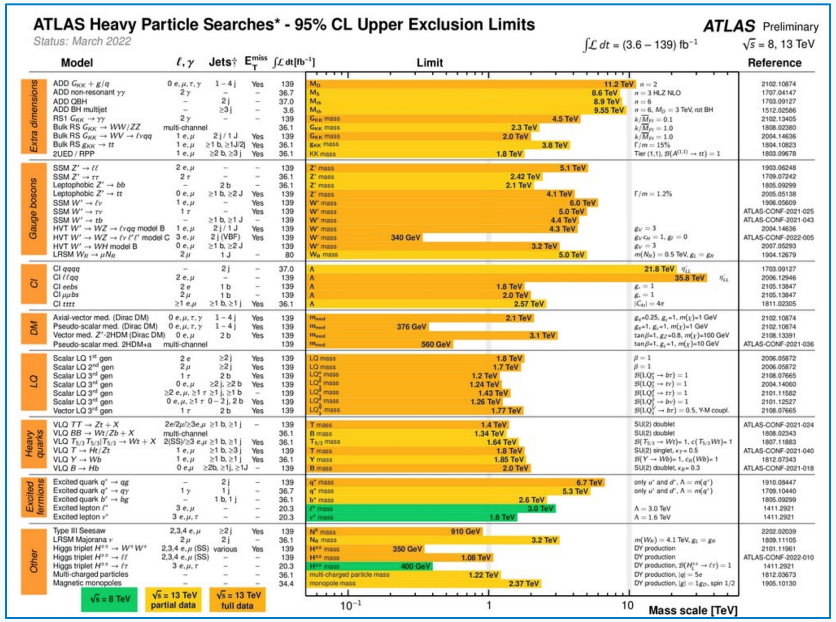


Measurements

Searches



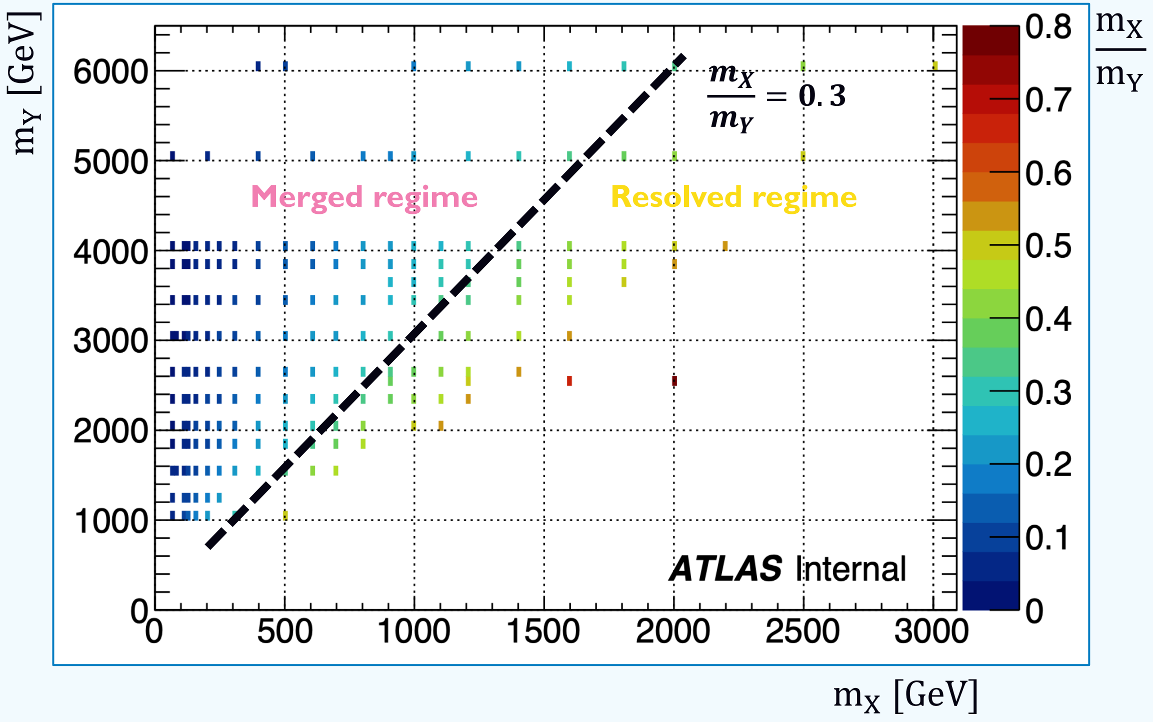


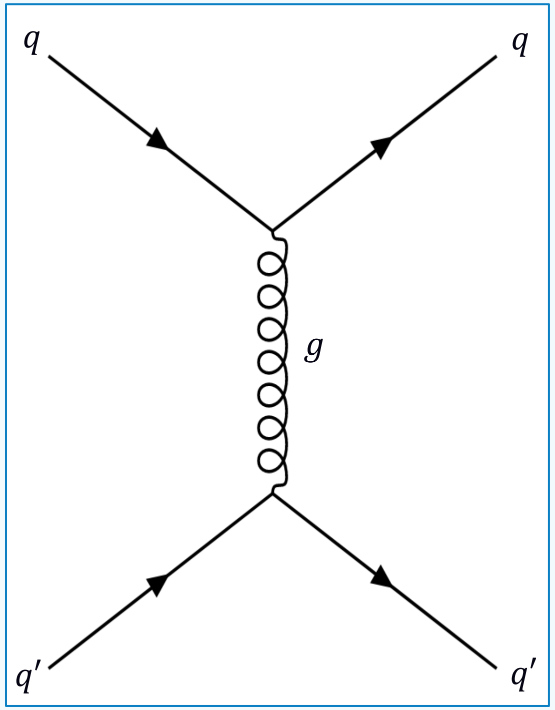


$Y \rightarrow XH$ IN HADRONIC FINAL STATES

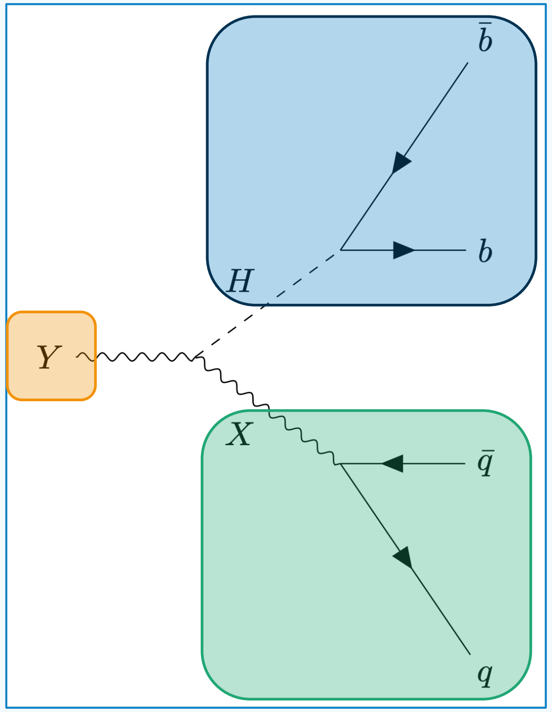
Phys. Rev. D 108, 052009 – Published 18 September 2023

- First Anomaly Detection application in unsupervised approach in ATLAS
- **Heavy Vector Triplet** model-based
- Analysis performed on full Run-2 dataset ($\mathrm{L} = 139\ \mathrm{fb}^{-1}$) with data collected at $\sqrt{s} = 13\ \mathrm{TeV}$ collisions with the ATLAS detector.
- Boosted and Resolved regime depending on mass ratio: $\frac{m_X}{m_Y}$



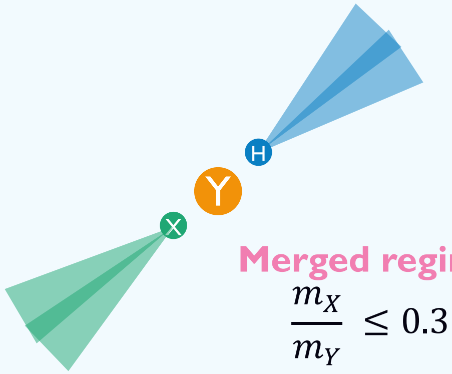


QCD di-jets (~97% background)



YXH (signal)

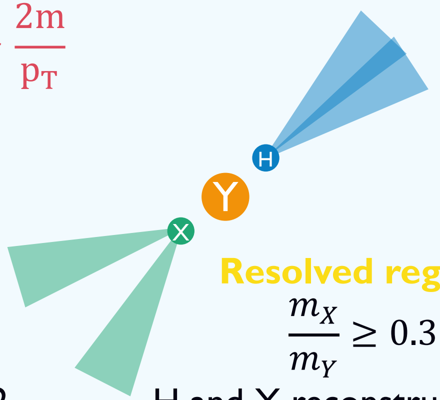
$$\Delta R \sim \frac{2m}{p_T}$$



Merged regime

$\frac{m_X}{m_Y} \leq 0.3$

H and X reconstructed as 2 Large-R jets



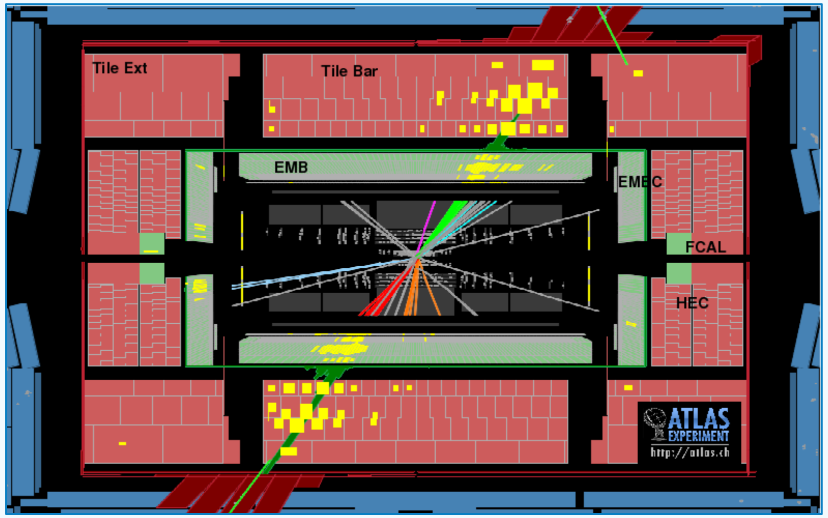
Resolved regime

$\frac{m_X}{m_Y} \geq 0.3$

H and X reconstructed as 1 Large-R jet and 2 Small-R jets

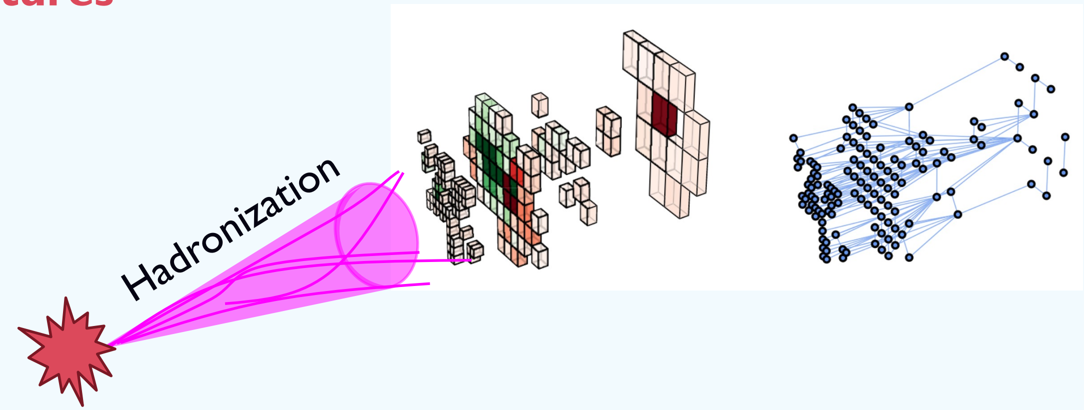
MOTIVATIONS

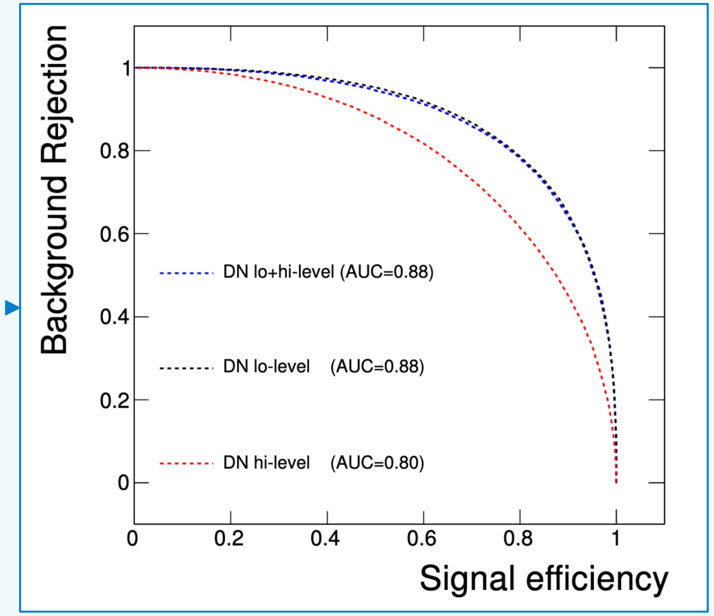
[Searching for Exotic Particles in High-Energy Physics with Deep Learning]



A set of features with basic information (low-level) such as information coming directly from the detectors implies better performances wrt features built combining basic information (high-level).

Goal: to study graph representation of low-level features (variables) such as jets constituents.





Why graphs?

Traditional architectures, assume a geometrically stable data organization.

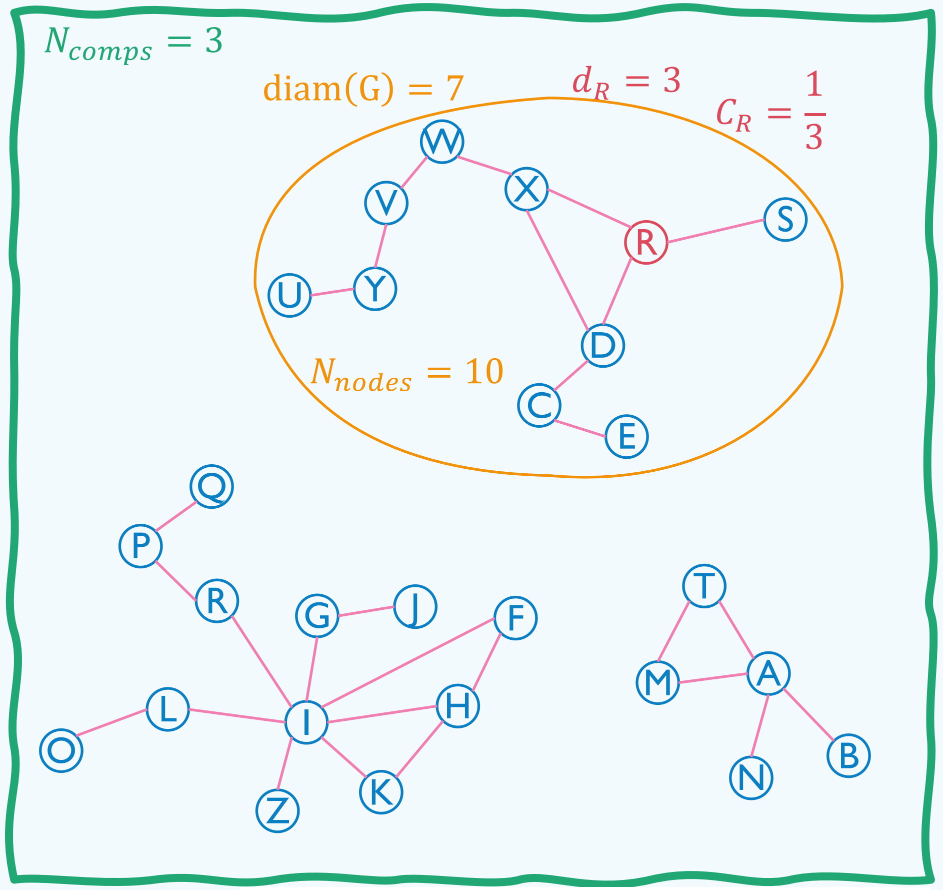
Energy deposits of the constituents in calorimeters often exhibit sparse data characteristics.

Geometrical deep learning architectures, like GNNs, have demonstrated enhanced learning capabilities and performance on such data type.

GRAPH THEORY

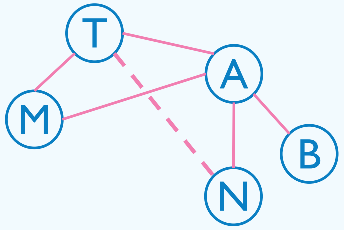
Graphs $\mathrm{G} = (\mathrm{V}, \mathrm{E})$ consist of vertices (**nodes** $\mathrm{v} \in \mathrm{V}$) that represent entities and **edges** (connections $\mathrm{e} \in \mathrm{E}$) that represent relationships between the vertices.

- Clustering Coefficient C_i: measure of the tendency of nodes to cluster together;
- Degree d_R number of edges;
- Diameter $\mathrm{diam}(\mathrm{G})$: the length of the shortest path between the most distanced nodes;
- Connected components $\mathrm{N_{comps}}$: subgraphs that we can isolate in a graph
- Number of nodes $\mathrm{N_{nodes}}$: the actual number of vertices of a graph or a component

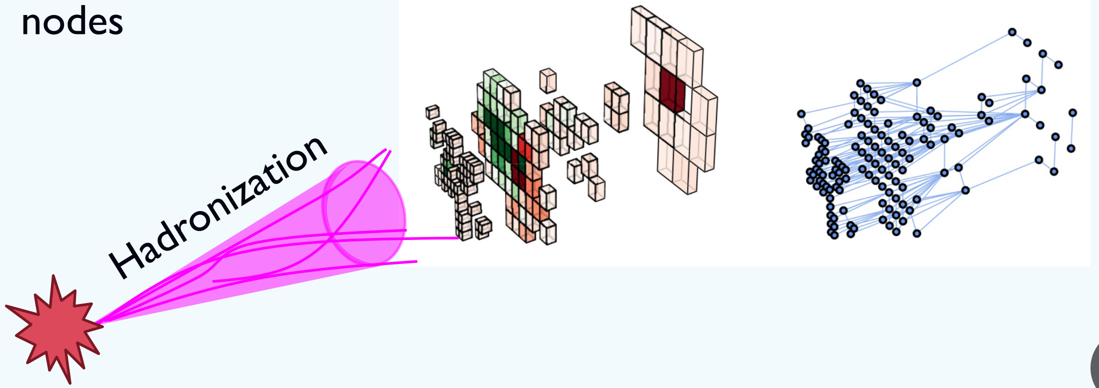


Graphs created using **up to 40 jet constituents as nodes**:

- $(\mathrm{p_T^{frac}}, \eta, \phi)_\mathrm{i}$ used as node features for constituent i $\rightarrow \mathrm{p_T^{frac}} = \frac{\mathrm{p_T^i}}{\mathrm{p_T}}$
- Edges defined with a criteria on distance $\Delta\mathrm{R}$ between two nodes $\Delta\mathrm{R}(\mathrm{const_i}, \mathrm{const_j})$



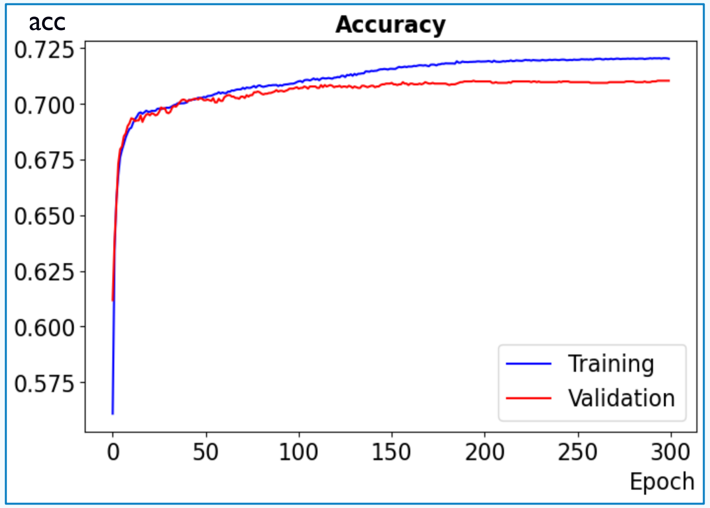
$$\Delta\mathrm{R_{max}} = 0.1$$
$$\Delta\mathrm{R(T,N)} = 0.25 > \Delta\mathrm{R_{max}} \rightarrow \textbf{NO}\text{ edge}$$
$$\Delta\mathrm{R(A,N)} = 0.07 < \Delta\mathrm{R_{max}} \rightarrow \text{edge}$$

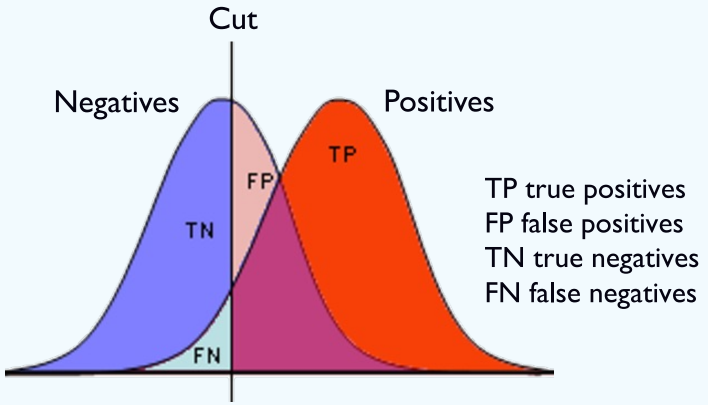


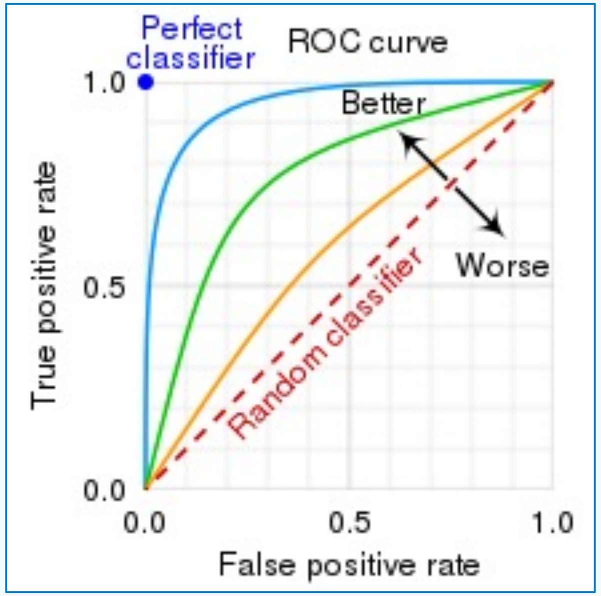
DNN AND GRAPH FEATURES

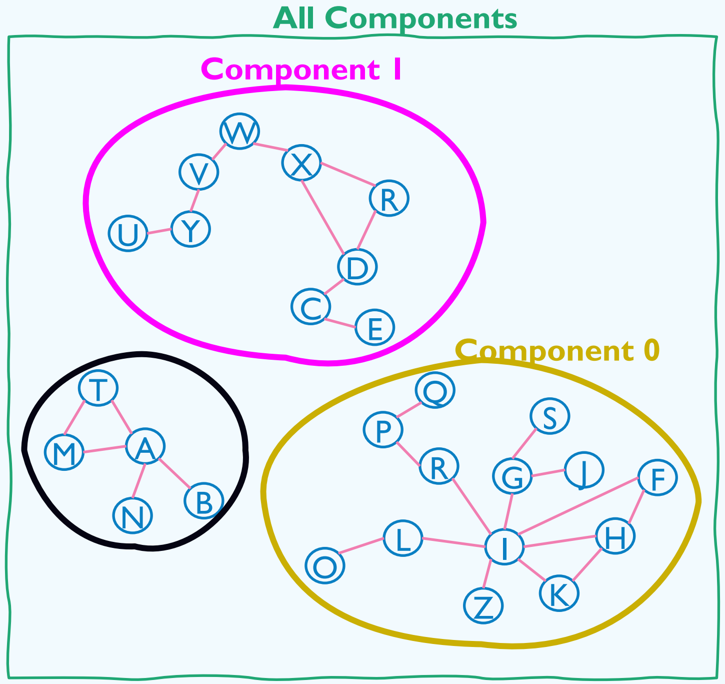
Supervised approach: event labels aware

Area Under the ROC Curve and Accuracy score used as metrics in classification task with DNN.









Different feature combinations using kinematical and geometrical features.
0 and **1** refers to the **first** and the **second** component of the graph, ordered by number of nodes.

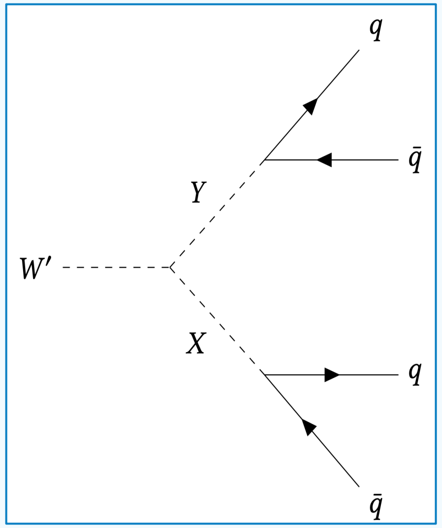
- **Kin**: kinematical variables
- **Geo01**: graph variables for connected components
- **GeoAll01**: graph variables for the entire graph and connected components
- **KinGeoAll01**: all of the above

BENCHMARK RESULTS

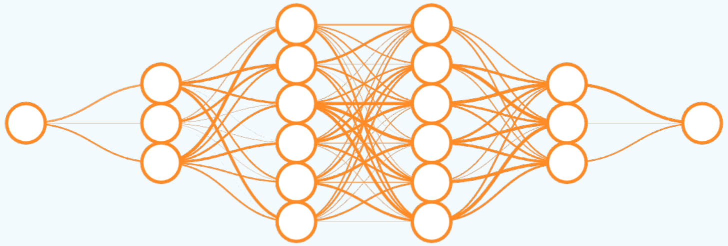
[Welcome to the home of the LHC Olympics 2020!]

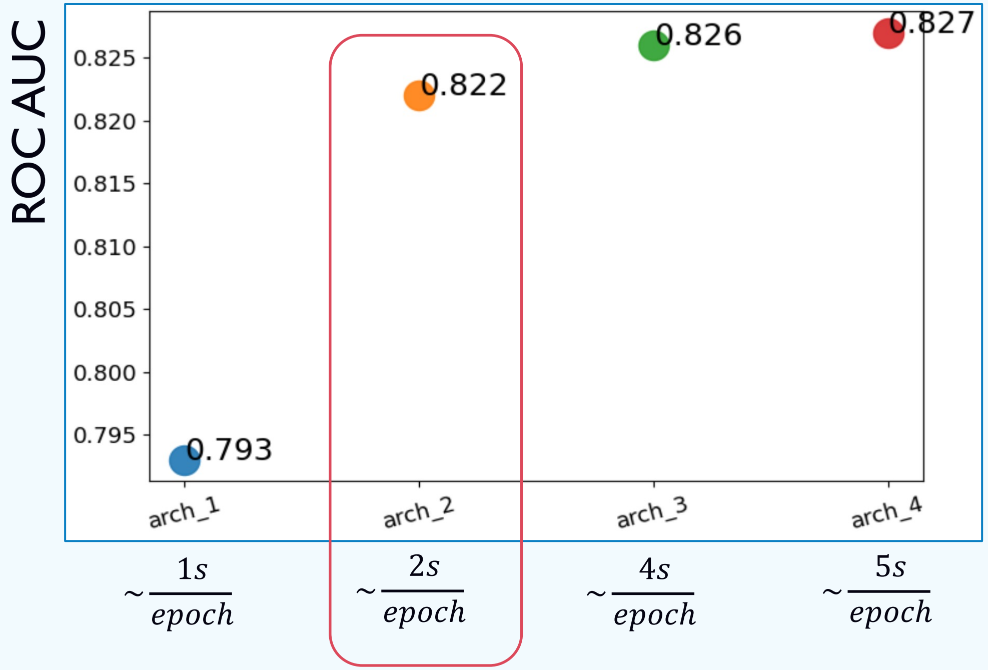
Preliminary test on graph representation features using a dataset with QCD **background** and full hadronic final state as **signal**

4 NN architectures that differ from each other by complexity.



Hyperparameters	
Parameter	**Value**
Optimizer	ADAM
Loss function	Binary Cross Entropy
Learning rate	10^{-6}
Batch size	32
Epochs	300
Validation Fraction	15%





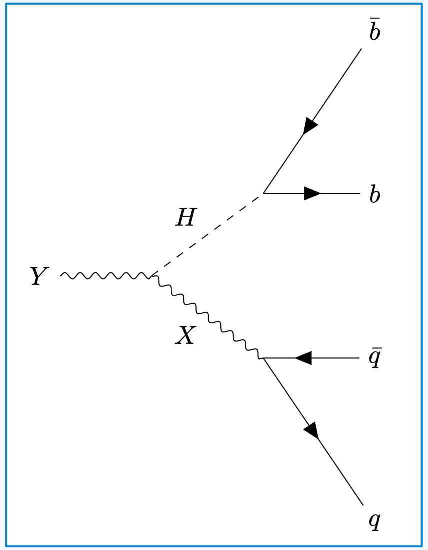
Results:

1. A graph representation of jets can be useful to perform signal/background discrimination for the kinds of BSM processes treated.
2. The best architecture has been chosen with the best compromise in performance (ROC AUC) and time per epoch

$Y \to XH \to q\bar{q}\, b\bar{b}$

	Type	Process	Events
Data	ATLAS data	QCD dijet	50 k
Signal	MC simulation ($36.1\ fb^{-1}$)	$Y \to XH \to \bar{q}q\bar{b}b$	17 k

Preselection over the events	
m_{j_1}, m_{j_2}	$> 50\ GeV$
Leading large-R jet p_T	$> 500\ GeV$
m_{jj}	$> 1300\ GeV$



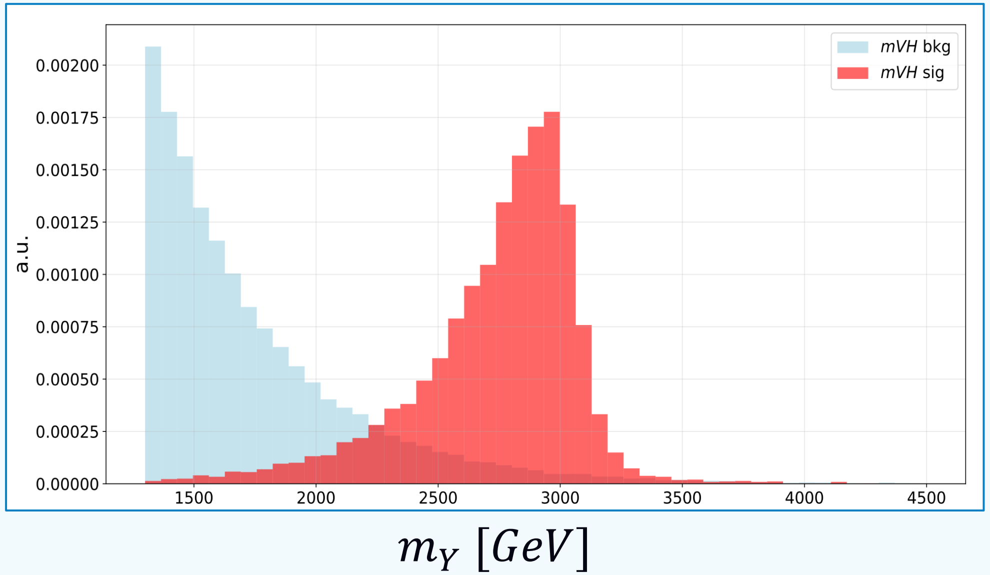
Training sample

$m_Y = 3\ TeV$

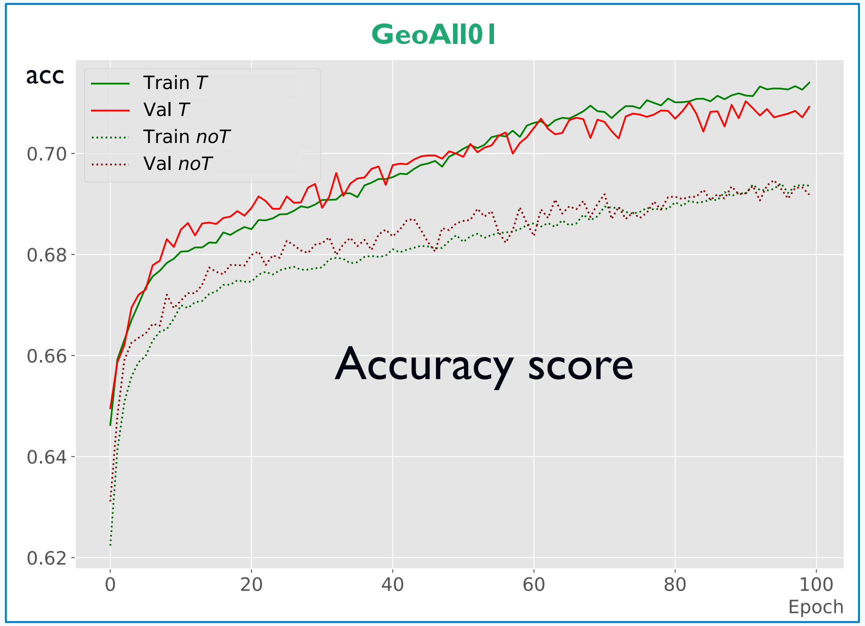
$m_X = 300\ GeV$

$\frac{m_X}{m_Y} = 0.1$

Merged regime



Graphs created with a **jet transformation**. Deep learning techniques are capable to learn features with a **large correlation in jet masses** and QCD background have a wide spread distribution over the mass → bias over the mass

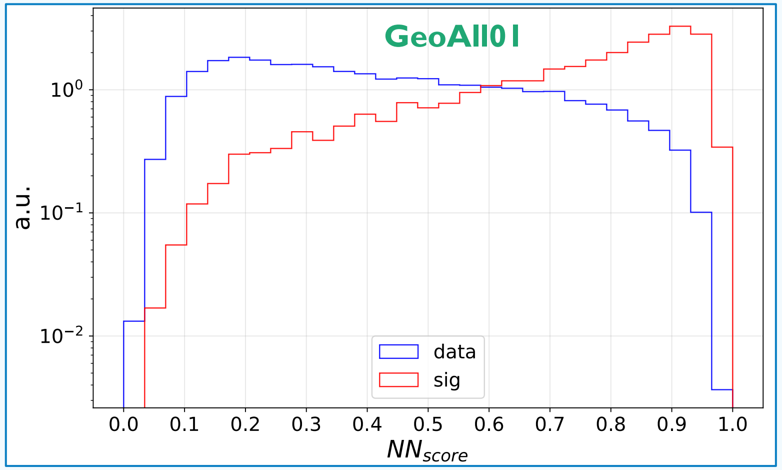


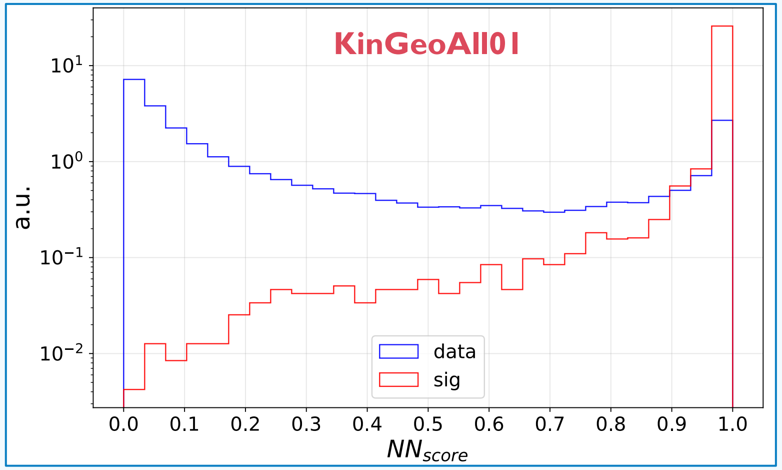
FEATURE COMBINATION RESULTS

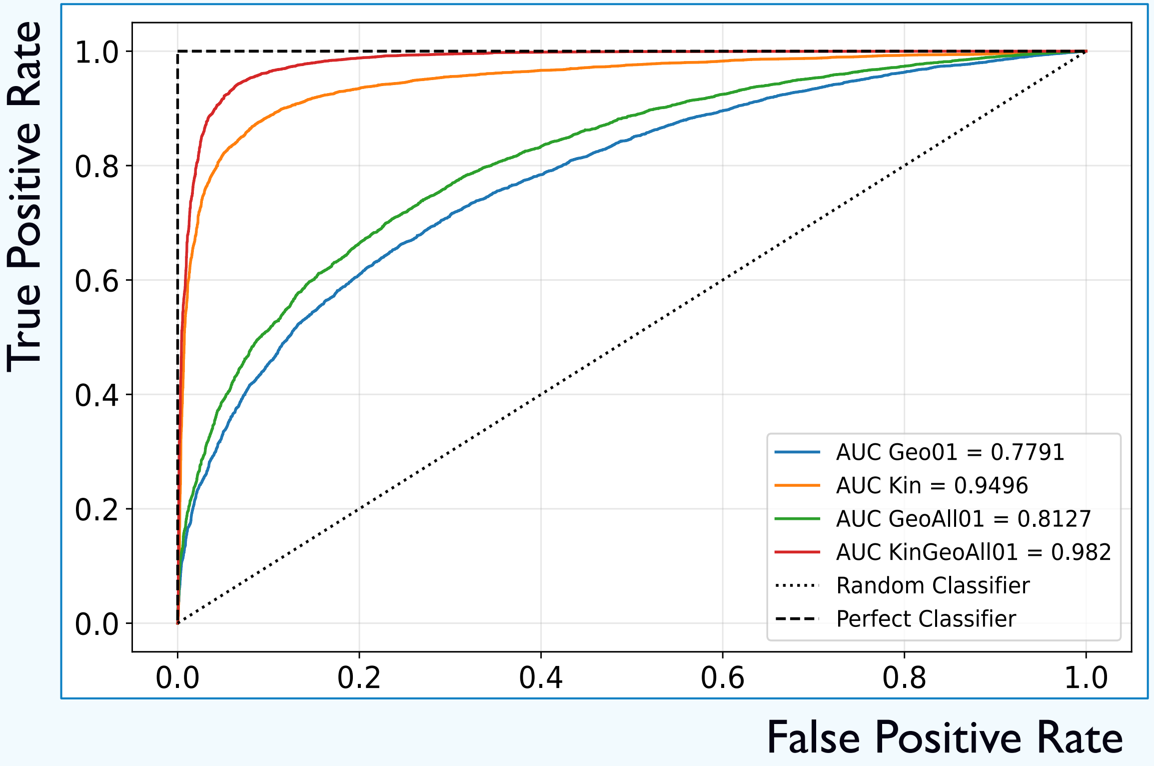
Using only geometric features show good results in deep neural network performance.

As expected, using kinematic features show good discriminative power.

Adding the geometric variables to the kinematic ones slightly improves the discriminant power.







Geo01: graph variables for connected components

Kin: kinematical variables

GeoAll01: graph variables for the entire graph and connected components

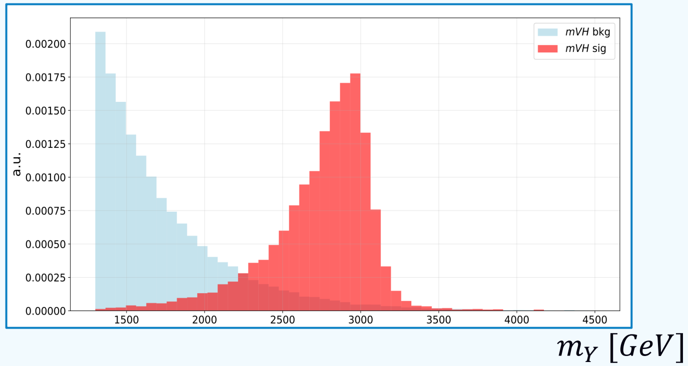
KinGeoAll01: all of the above

NN CUT CRITERIA: SIGNIFICANCE GAIN

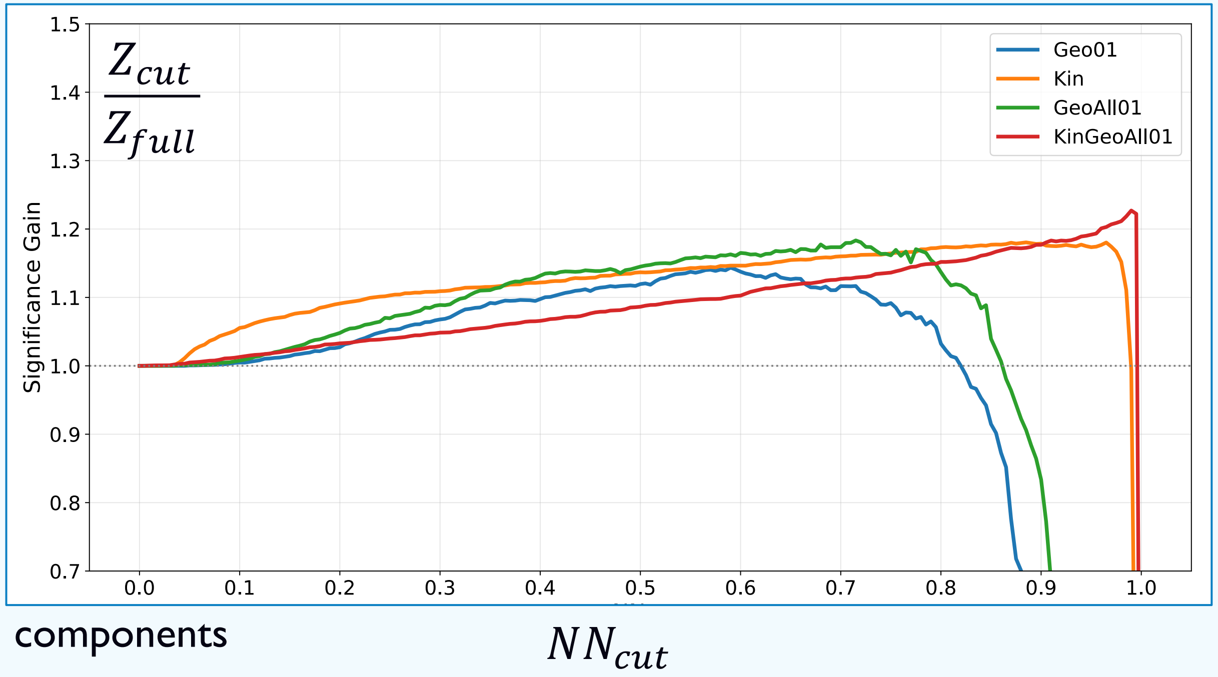
- Optimization made on signal with $\mathrm{m_X} = 300\ \mathrm{GeV}$ and $\mathrm{m_Y} = 3\ \mathrm{TeV}$
- Significance σ_i computed on each bin of invariant mass m_Y distribution

$$\sigma_i = \sqrt{2\left((s_i + b_i)\ln\left(1 + \frac{s_i}{b_i}\right) - s_i\right)}$$

- Global significance $Z = \sqrt{\sum_i \sigma_i^2}$



Significance Gain



	Cut	Gain
Geo01	0.6	1.1
Kin	0.9	1.2
GeoAll01	0.7	1.2
KinGeoAll01	0.9	1.2

Geo01: graph variables for connected components

Kin: kinematical variables

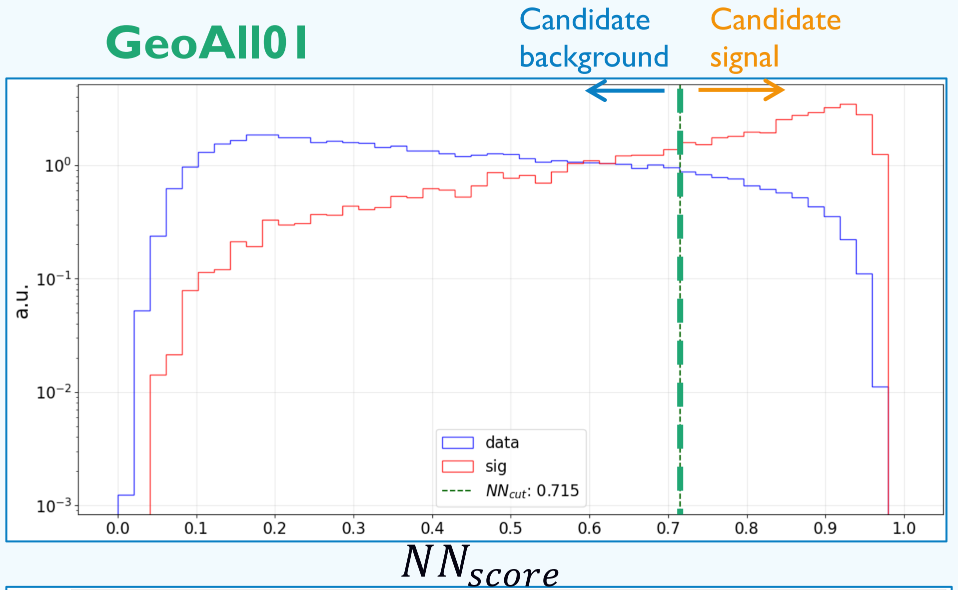
GeoAll01: graph variables for the entire graph and connected components

KinGeoAll01: all of the above

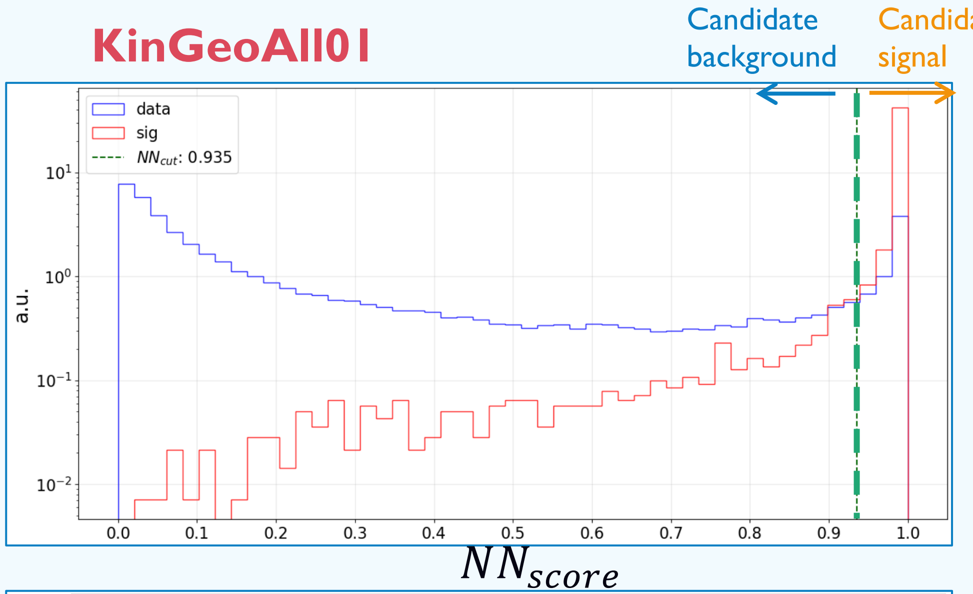
NN_{score} AND m_Y DISTRIBUTION WITH CUT IN NN_{score}

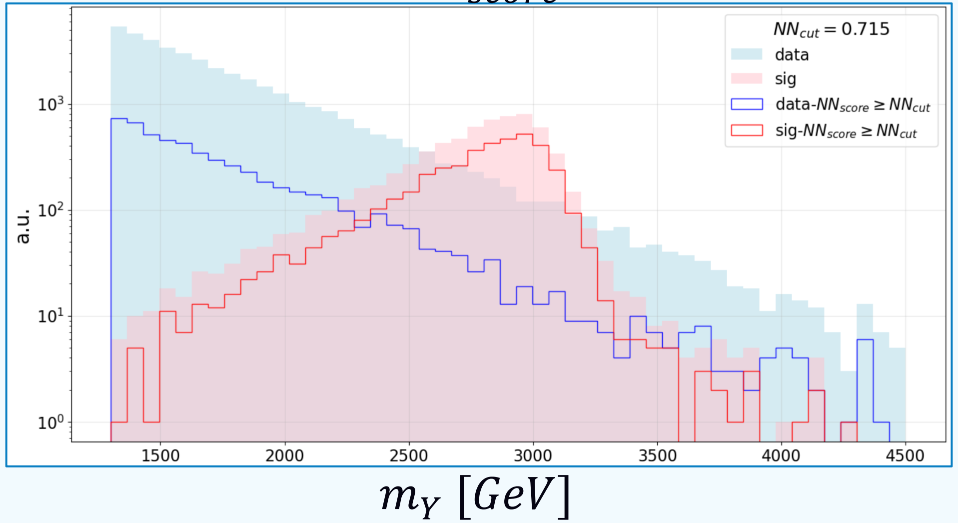
The plot displays the distribution of ATLAS background and the signal sample ($m_X = 300\ GeV, m_Y = 3\ TeV$) using the cut on the score associated with various combinations of features.

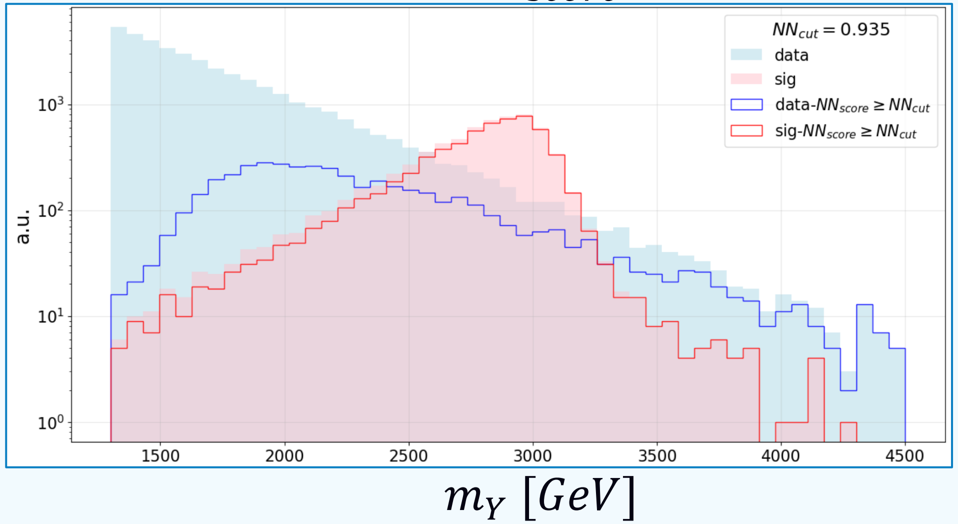
GeoAll01



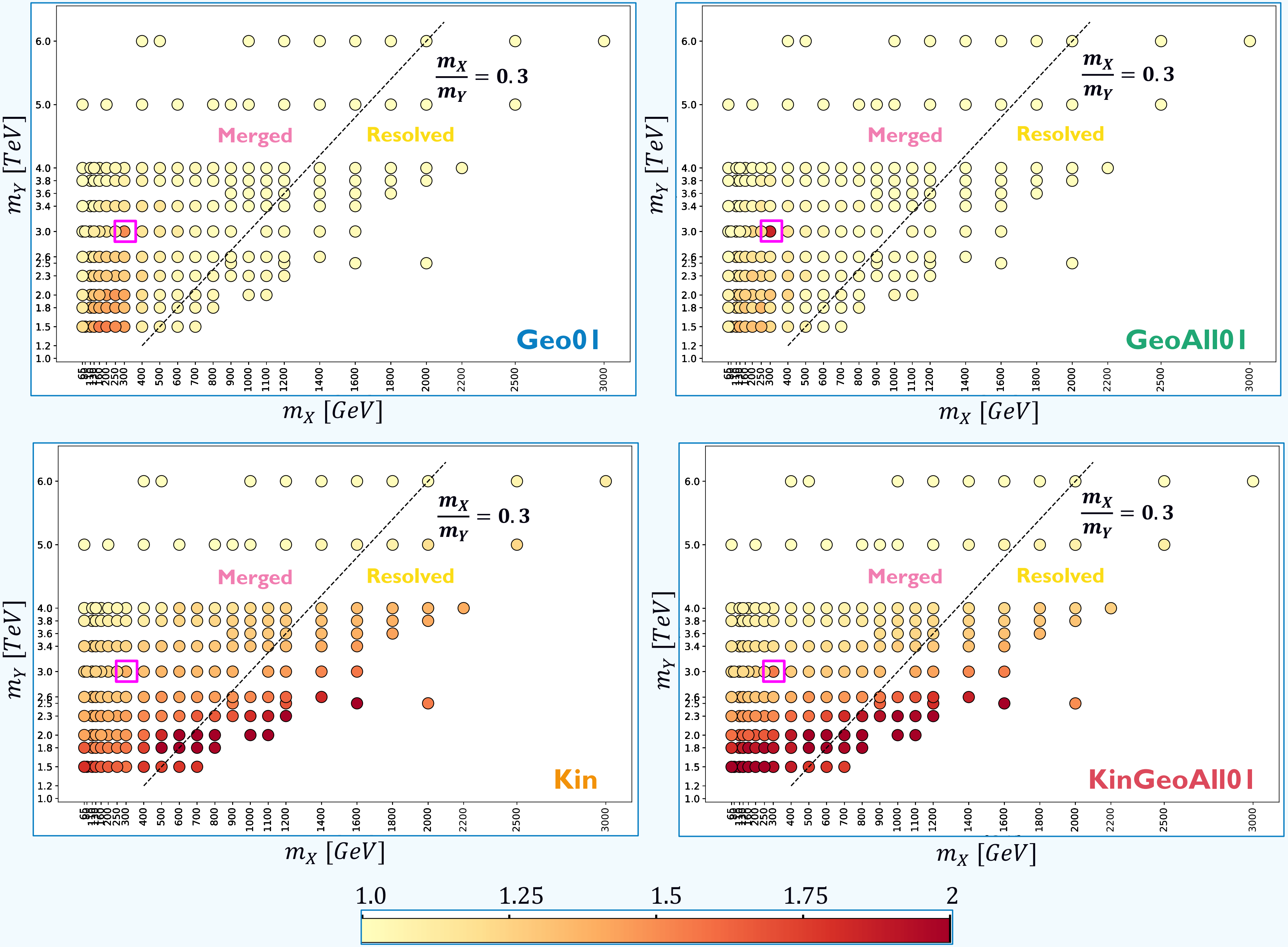
KinGeoAll01







EXPECTED YXH UPPER LIMIT CROSS SECTION RATIO AT 95% CONFIDENCE LEVEL



Ratio between the expected upper limit cross section **without** any cut and expected upper limit cross section **with** best cut

$$\tilde{\sigma} = \frac{\sigma_{full}}{\sigma_{cut}}$$

□ Training signal point

Kin: kinematical variables

Geo01: graph variables for connected components

GeoAll01: graph variables for the entire graph and connected components

KinGeoAll01: all of the above

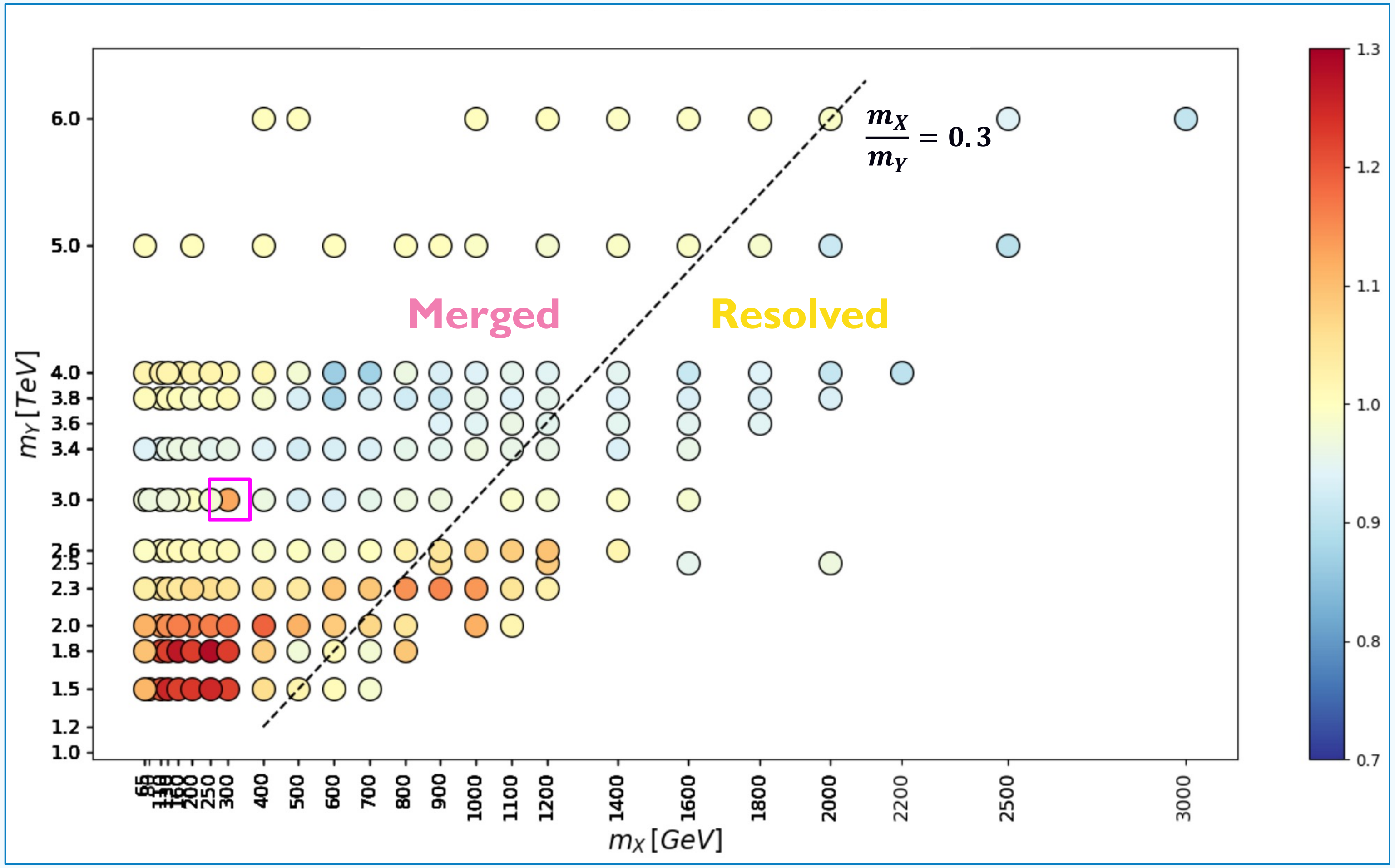
EXPECTED YXH UPPER LIMIT CROSS SECTION GAIN AT 95% CL – EFFECT OF GEOMETRIC FEATURES

Ratio between DNN with both kinematical and graph features and DNN with kinematical features expected upper limit cross section.

Many of the working points **below** $m_Y = 3000\ GeV$ show an improvement.

The improvement is poorer at **high** values of m_Y, as the DNN with graph variables helps remove the background in the region with **low** m_Y.

KinGeoAll01/Kin



CONCLUSIONS

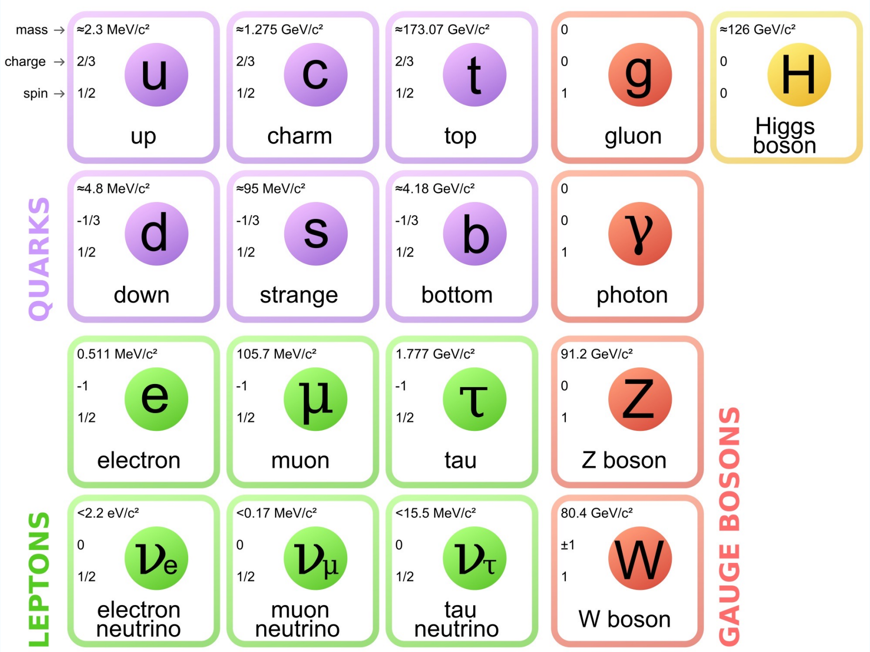
- Different DNN architectures tested using different combinations of kinematical and graph features.
- Geometric representation of jets as graphs, provide good discriminant power.
- Combination of kinematical and geometric information (KinGeoAll01) provide best results.
- A dedicated training on each point can probably lead to better performance
- Potential future applications involve the use of Graph Neural Network architectures with unsupervised approach.

GRAZIE PER L’ATTENZIONE

BACKUP

STANDARD MODEL

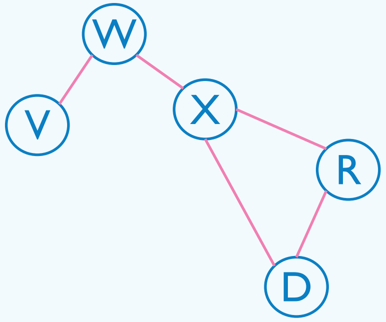
- Model of particle physics that best describes current observations
  - 6 quarks;
  - 3 leptons and 3 neutrinos;
  - 6 bosons.
  - …and respectively anti-particles
- Limits
  - It has 19 free parameters;
  - Do not take into account neutrinos masses or other cosmological phenomena



HEAVY VECTOR TRIPLETS

- Heavy Vector Triplets is a class of particle classified with a particularly high mass – at least $1.5\,TeV$ – described with a set of 3 vector, spin-1 bosons:
  - 2 charged
  - 1 neutral
- The properties of these particles are:
  - V_μ^a the field eigenstates, with $a = 1, 2, 3$
  - $V_\mu^\pm = \frac{V_\mu^1 \mp i V_\mu^2}{\sqrt{2}}$ and $V_\mu^0 = V_\mu^3$ as the charge eigenstates
- Note:
  - This can describe the system of W and Z as other set of particles
  - Field eigenstates are not mass eigenstates

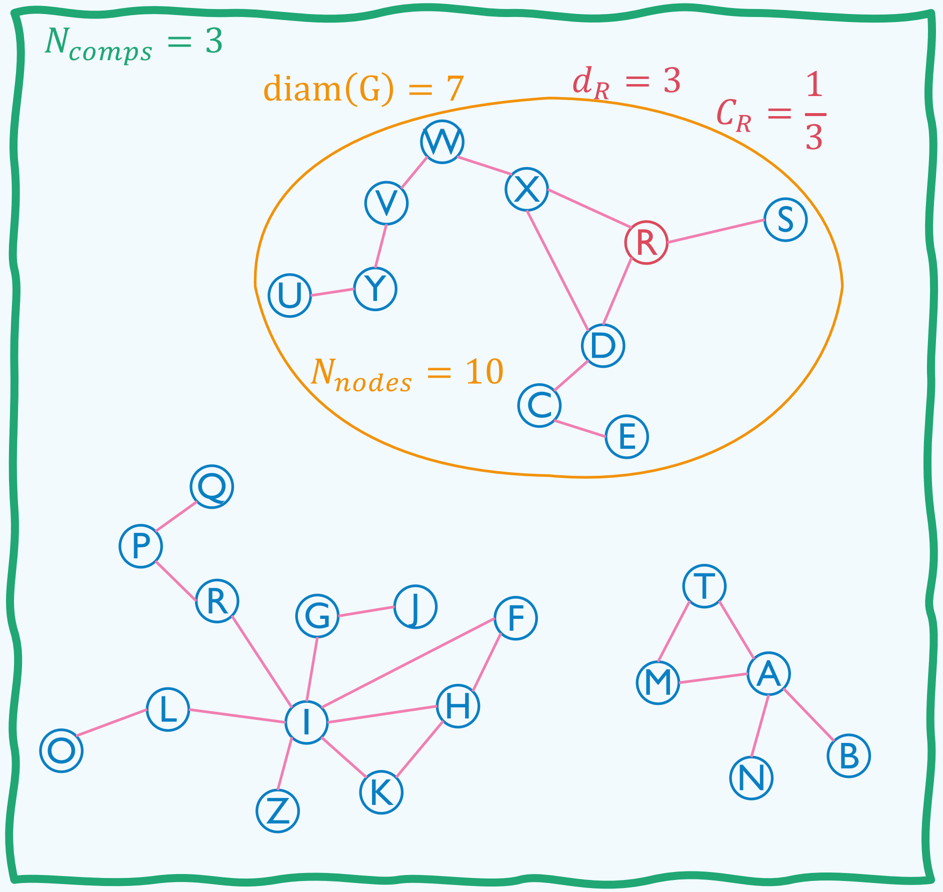
GRAPH THEORY



$$A = \begin{array}{c} \\ \\ \end{array}\begin{array}{cc} \begin{array}{ccccc} \mathrm{V} & \mathrm{W} & \mathrm{X} & \mathrm{R} & \mathrm{D} \end{array} & \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} & \begin{array}{c} \mathrm{V} \\ \mathrm{W} \\ \mathrm{X} \\ \mathrm{R} \\ \mathrm{D} \end{array} \end{array}$$

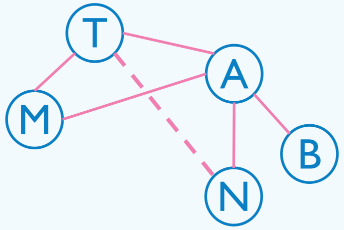
Graphs $G = (V, E)$ consist of vertices (**nodes** $v \in V$) that represent entities and **edges** (connections $e \in E$) that represent relationships between the vertices. The adjacency matrix A is used to describe a graph

- Clustering Coefficient $C_i = \frac{2|\{e_{jh}: u_j, u_h \in N_i, e_{jh} \in E\}|}{k_i(k_i - 1)}$
- Degree $d_u = \sum_{v \in V} A_{uv}$
- Diameter: the length of the shortest path between the most distanced nodes;
- Connected components N_{comps}: subgraphs that we can isolate in a graph
- Number of nodes N_{nodes}: the actual number of vertices of a graph or a component



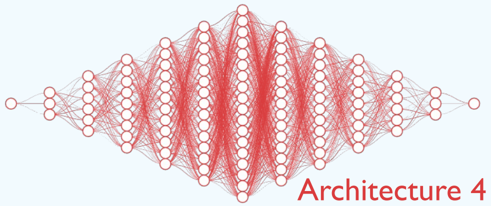
Graphs created using **jet constituents up to 40 as nodes**:

- $(p_T^{frac}, \eta, \phi)_i$ used as node features for constituent i $\rightarrow p_T^{frac} = \frac{p_T^i}{p_T^{tot}}$
- Edges defined with a criteria on distance ΔR between two nodes $\Delta R(\text{const}_i, \text{const}_j)$



$$\Delta R_{max} = 0.1$$
$$\Delta R(T, N) = 0.25 > \Delta R_{max} \rightarrow \text{NO edge}$$
$$\Delta R(A, N) = 0.07 < \Delta R_{max} \rightarrow \text{edge}$$

DEEP NEURAL NETWORK (DNN) SUPERVISED

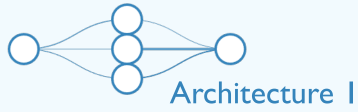


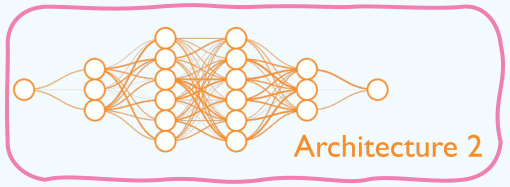
Artificial Neural Network with multiple hidden layers. The depth of the network enables it to capture complex patterns.

Goal of the training is to minimize the loss function that define the discrepancy between the real label and the prevision in a classifier task.

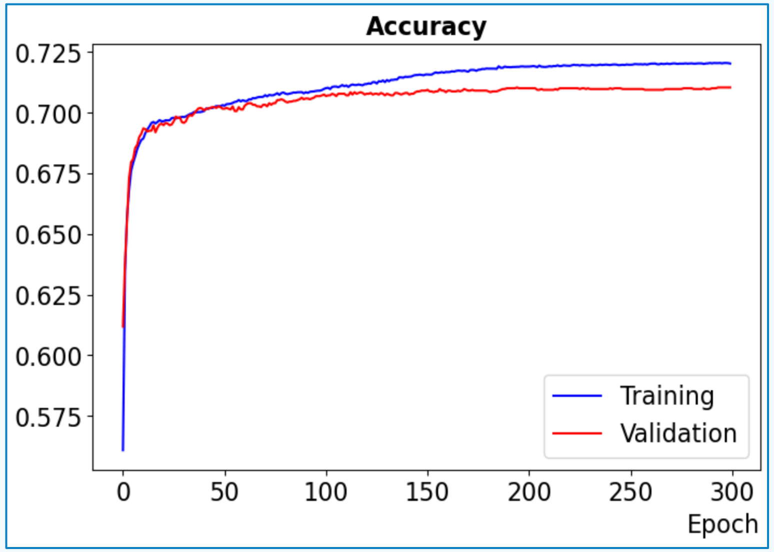
Area Under the ROC Curve and Accuracy score used as metrics in classification task.

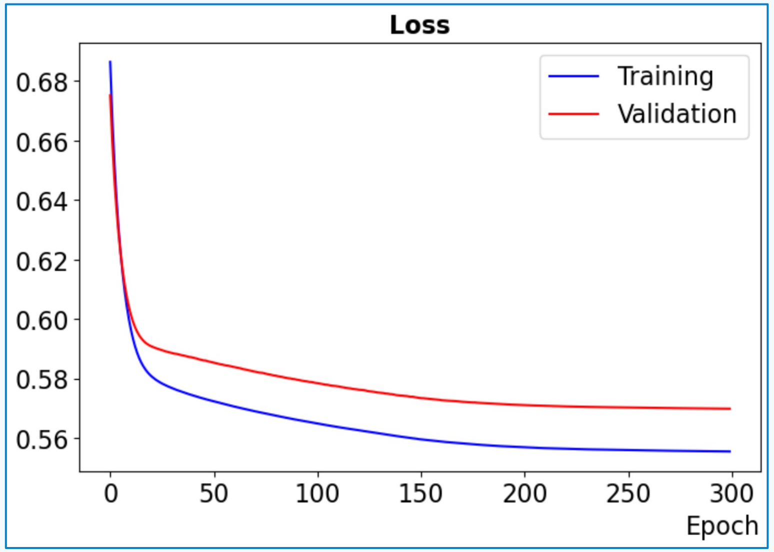
4 NN architectures that differ from each other by the number of nodes per layer and number of layers.

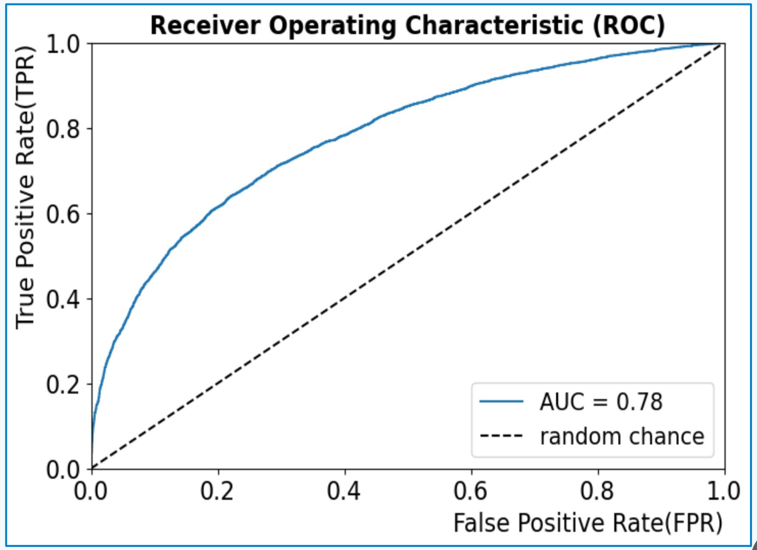




Input	Hidden 1	Hidden 2	Hidden 3	Hidden 4	Output
N_f	$N_f \times 3$	$N_f \times 6$	$N_f \times 6$	$N_f \times 3$	1
N_f	**$N_f \times 5$**	**$N_f \times 10$**	**$N_f \times 10$**	**$N_f \times 5$**	**1**
N_f	$N_f \times 8$	$N_f \times 16$	$N_f \times 16$	$N_f \times 8$	1





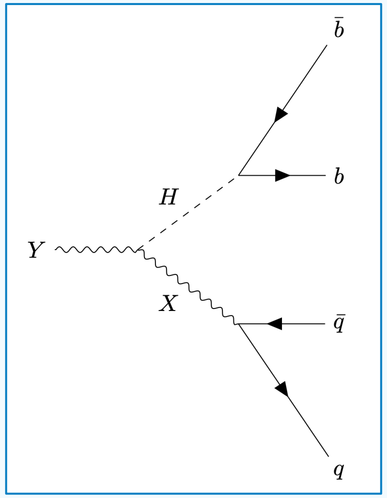


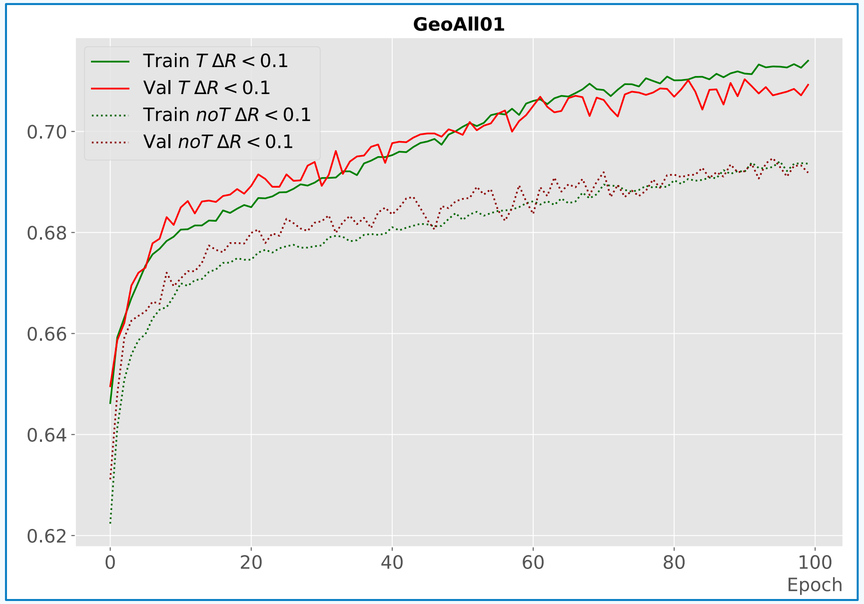
DNN TRAINING HYPERPARAMETERS

Hyperparameter Tuning

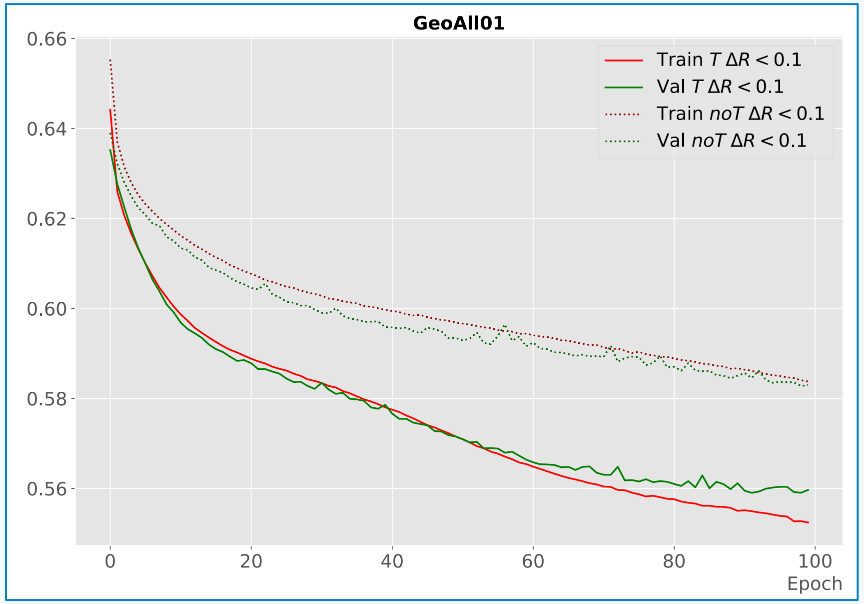
Parameter	Value
Optimizer	Adam
Jet Transformation	**Yes**, No
Loss function	Binary Cross Entropy
Learning rate	$10^{-3}, 10^{-4}, 10^{-5}, \mathbf{10^{-6}}, 10^{-7}$
Batch size	$4, 8, 16, \mathbf{32}, 64$
Epochs	$50, 100, 200, \mathbf{300}$
Validation Fraction	$10\%, \mathbf{15\%}, 20\%, 25\%, 30\%$

Jet Transformation to create graphs. Deep learning techniques are very capable to learn features with a large correlation in jet masses and QCD background have a wide spread distribution over the mass → bias over the mass





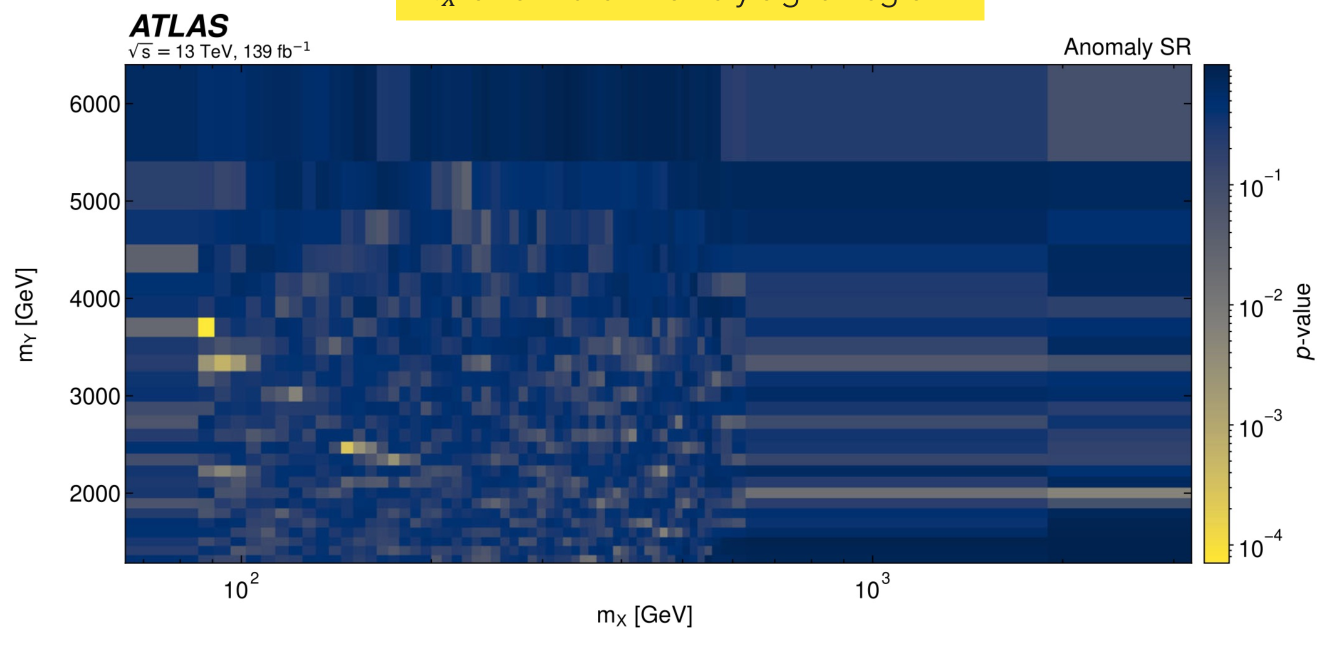
Comparison between the use of Jet Transformation on accuracy score and loss function

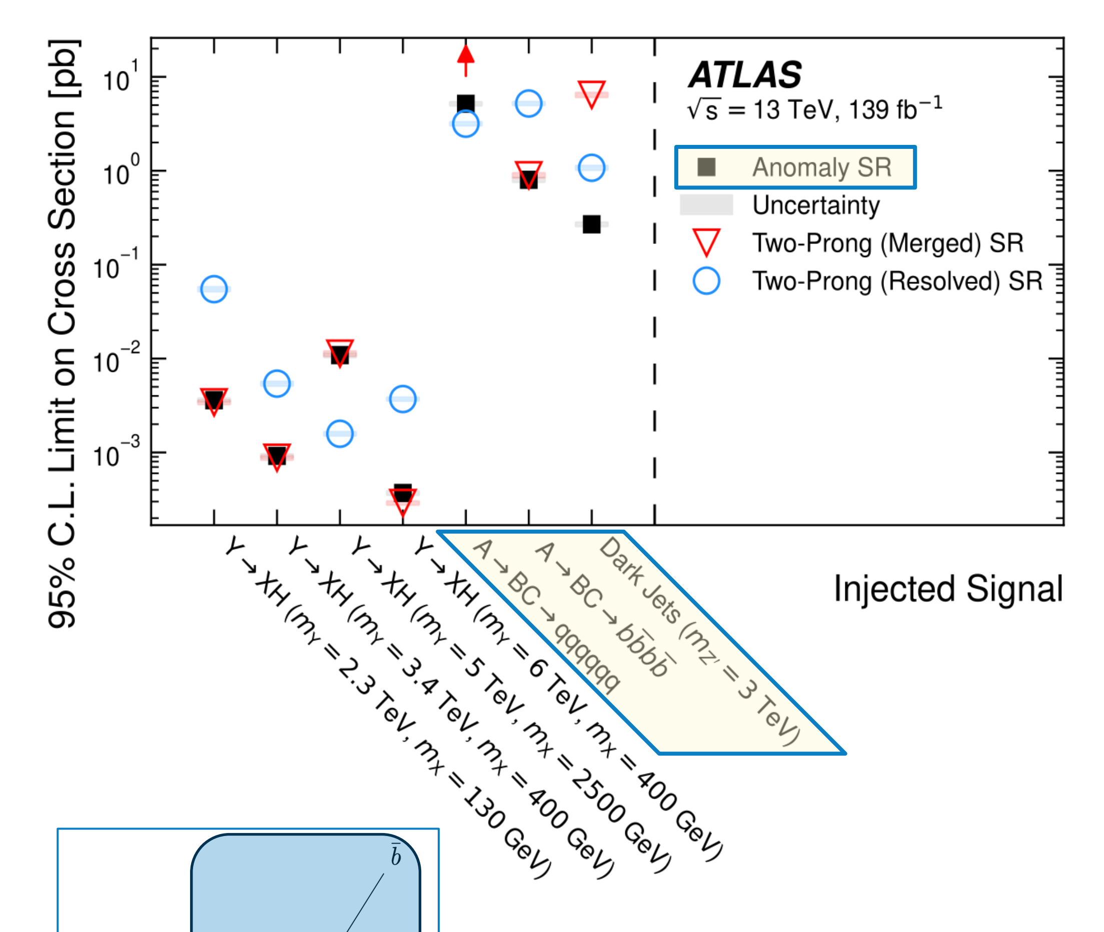


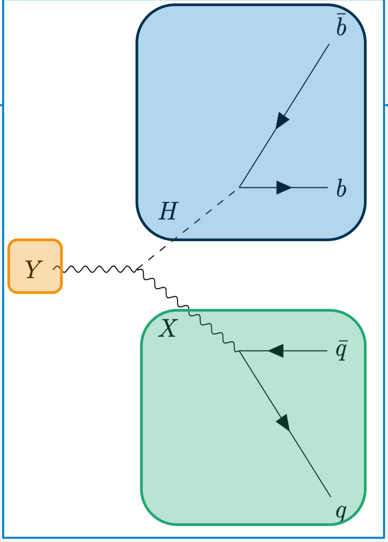
$Y \to XH$ IN HADRONIC FINAL STATES

- Higgs identification with $D_{H_{bb}}$ **NN score**
- Completely data-driven **Machine Learning technique** to estimate QCD background
- Anomaly detection **discovery region** introduced with novel data-driven anomaly score (AS) using **ML**

Observed p-values across all m_Y and m_X bins in the Anomaly signal region



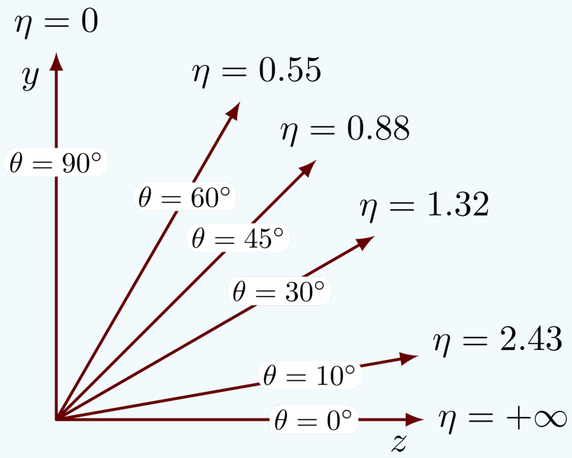




HADRONIC CALORIMETER (HCAL) AND JET RECONSTRUCTION

Jets reconstructed using **tracks** in ID, **calorimeter deposits** and **anti-k_T** algorithm.

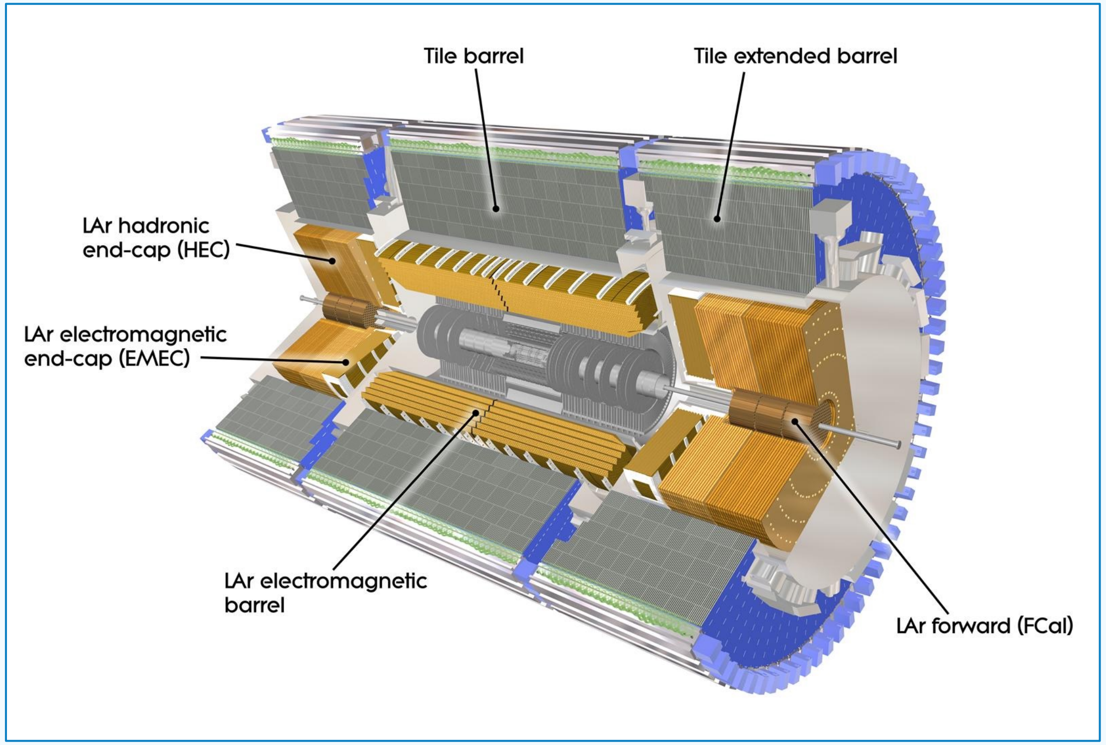
- **Tile HCAL**: 14 mm of iron absorber alternated to a 3 mm sparkling plates, in bunches;
- **Liquid Argon end-cap HCAL**: copper and tungsten as absorbers and LAr as active component.

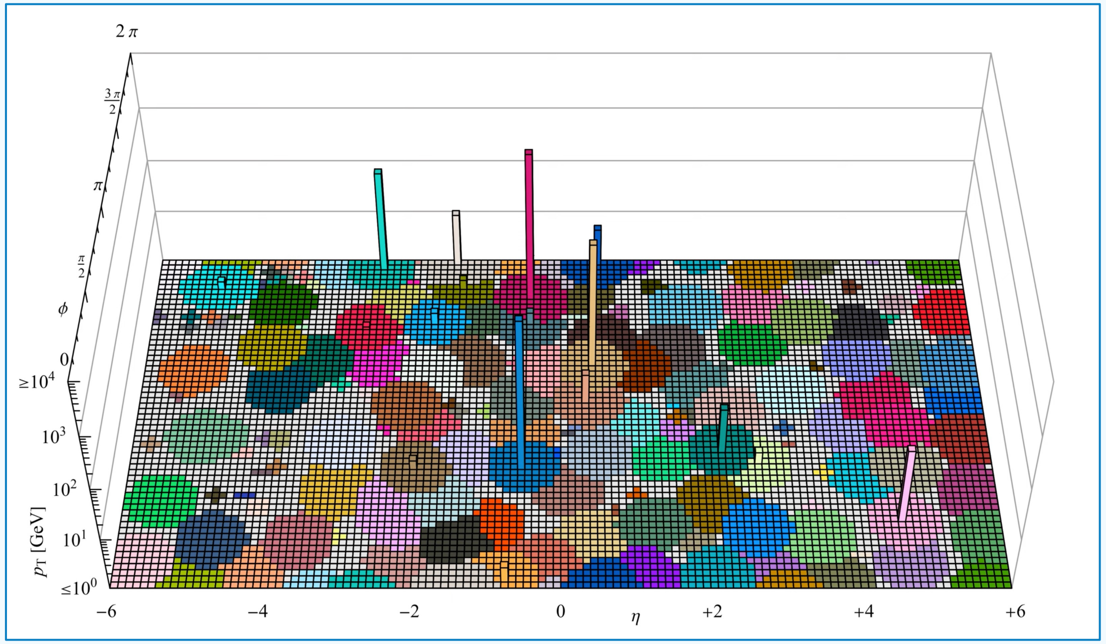


Energy resolution HCAL

Pseudorapidity range	Energy resolution $\frac{\sigma_E}{E}$
$\|\eta\| < 3.0$	$\frac{50\%}{\sqrt{E}} \oplus 3\%$
$3.0 < \|\eta\| < 4.9$	$\frac{100\%}{\sqrt{E}} \oplus 10\%$

Anti- k_T reconstruction algorithm takes **topoclusters** (clusters of energy deposits in the calorimeters) as input and combine them to form jet cones with characteristic radius R using a distance parameter.





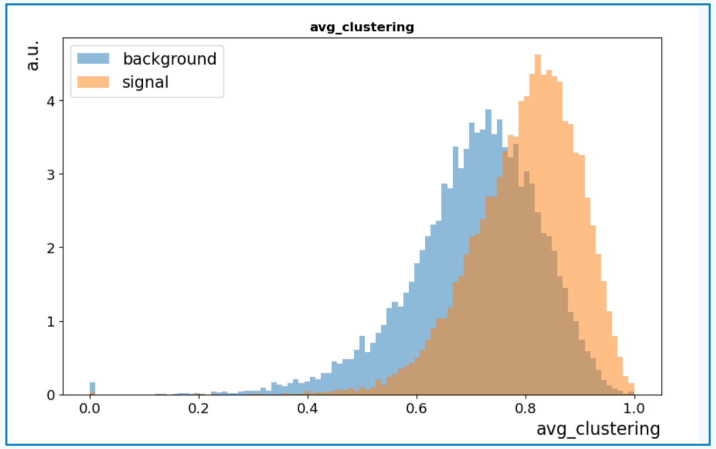
DNN WITH GRAPH FEATURES

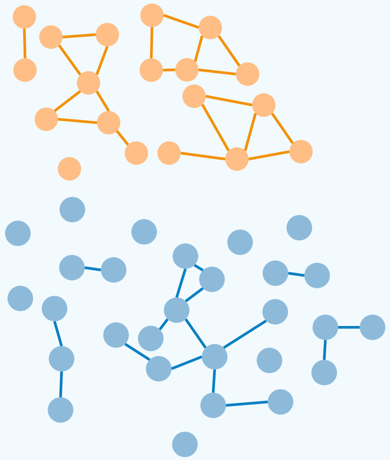
Different feature combinations using kinematical and geometrical features.

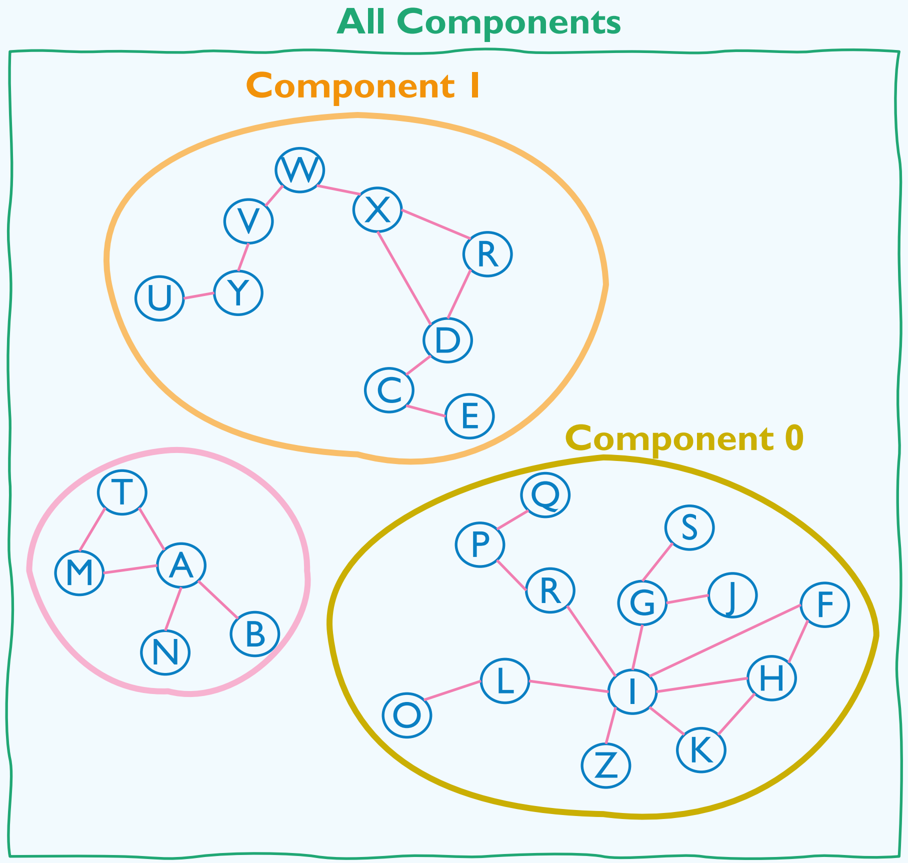
0 and **1** refers to the **first** and the **second** component of the graph, ordered by number of nodes.

<table>
<tr><td colspan="4">KinGeoAll01</td></tr>
<tr><td></td><td colspan="3">GeoAll01</td></tr>
<tr><td>Kin</td><td></td><td colspan="2">Geo01</td></tr>
<tr><th>KINEMATICS</th><th>GEOMETRIC ALL</th><th>GEOMETRIC 0</th><th>GEOMETRIC 1</th></tr>
<tr><td rowspan="2">p_T Jet</td><td colspan="3">Average clustering coefficient</td></tr>
<tr><td colspan="3">Number of nodes</td></tr>
<tr><td rowspan="2">η Jet</td><td colspan="3">Mean degree</td></tr>
<tr><td colspan="3">Diameter</td></tr>
<tr><td rowspan="2">ϕ Jet</td><td rowspan="2">Number of components</td><td colspan="2">Node fraction</td></tr>
<tr><td colspan="2">p_T fraction</td></tr>
</table>

LHC Olympics 2020







LHC OLYMPICS 2020

Welcome to the home of the LHC Olympics 2020!

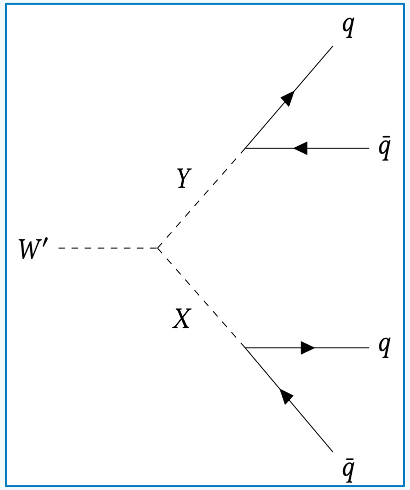
Preliminary test on graph representation features.

4 NN architectures that differ from each other by the number of nodes per layer and number of layers.

Architecture 2: best compromise between performance and use of computational resources.

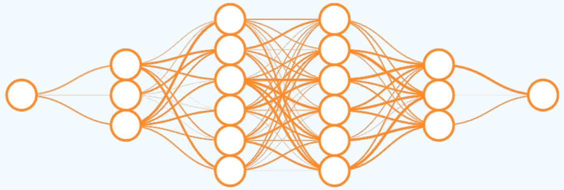
Type	Process
Background	QCD dijet
Signal	$W' \to XY \to \bar{q}q\ \bar{q}q$

Particle	Mass
W'	3.5 TeV
X	500 GeV
Y	100 GeV



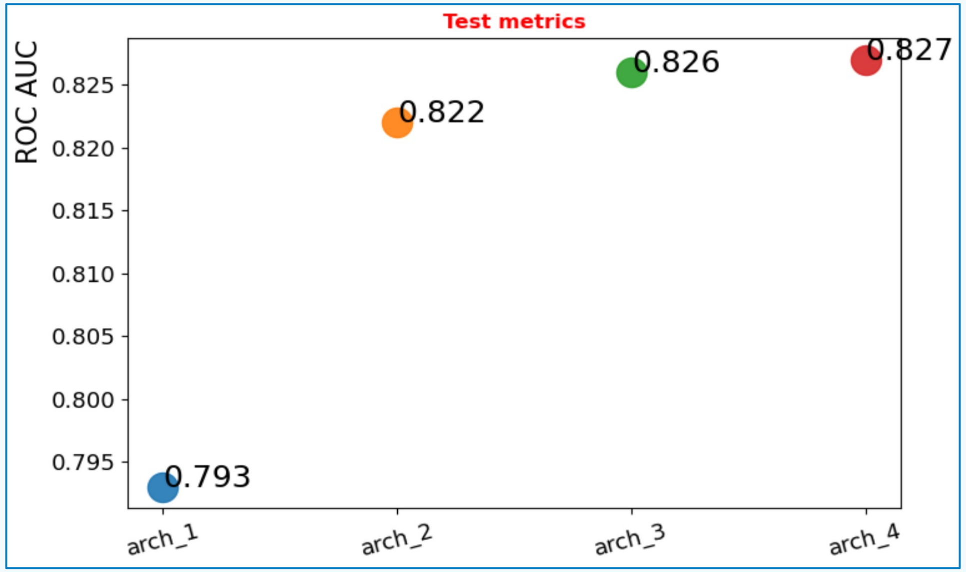
Number of events

Type	Training	Test
Data	10 603	143 806
Signal	10 601	2 650



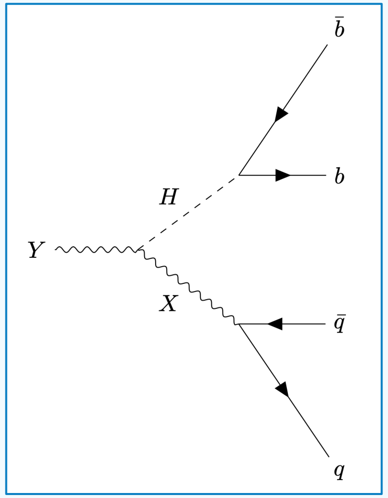
Result: a graph representation of jets can be useful to perform signal/background discrimination for the kinds of BSM processes treated.

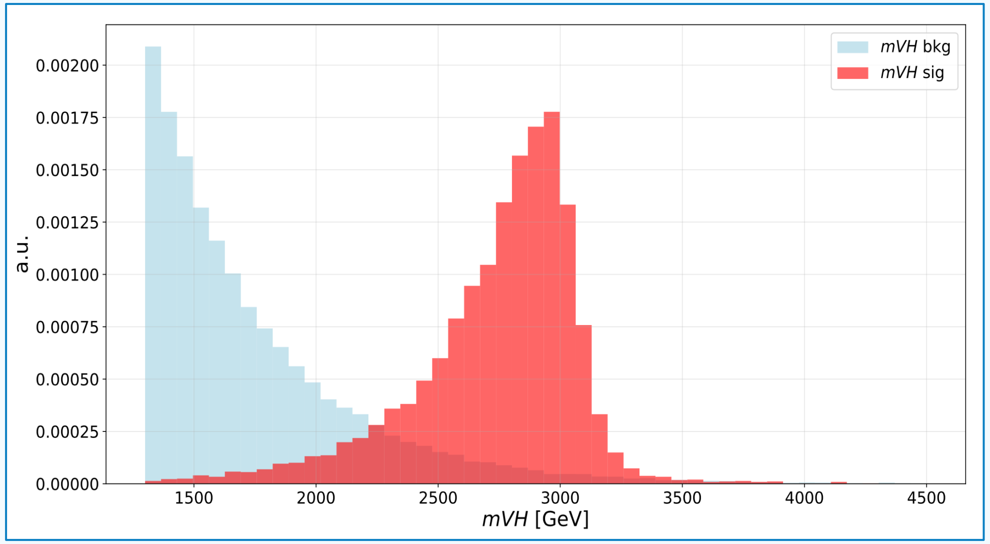
Architecture 2: best compromise between performance and use of computational resources.



YXH

	Type	Process
Data	ATLAS data	QCD dijet
Signal	MC simulation ($36.1\ fb^{-1}$)	$Y \to XH \to \bar{q}q\bar{b}b$

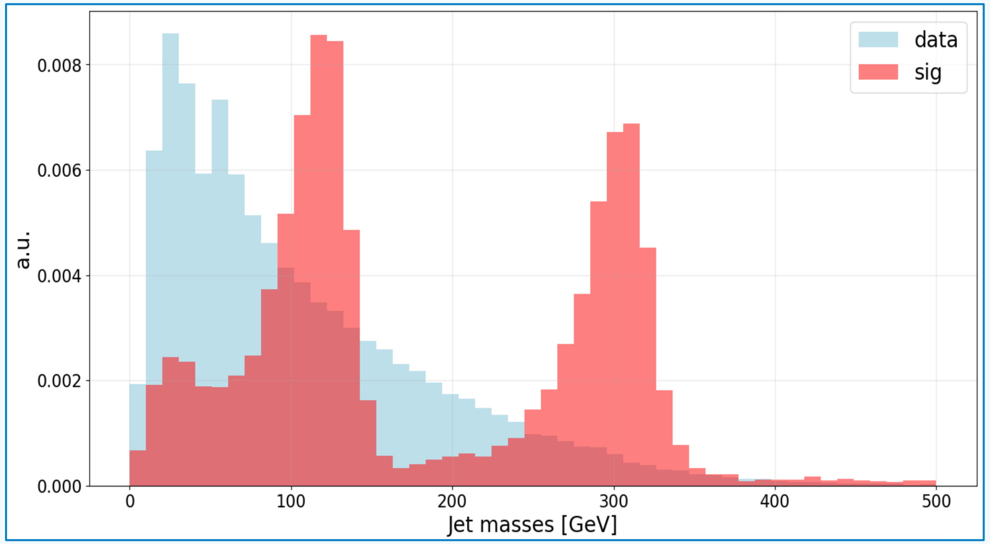




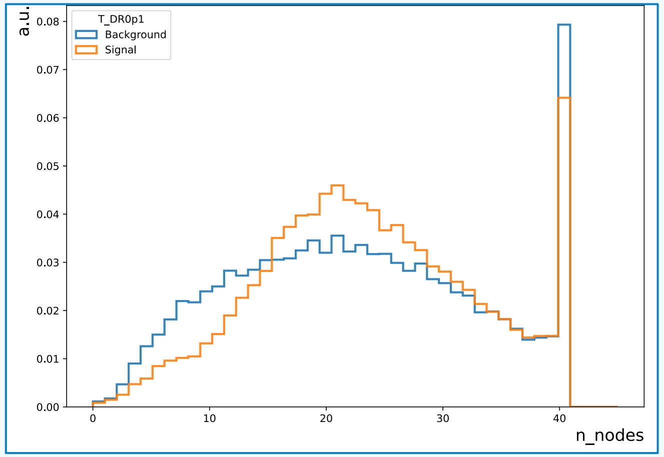
Data – $50k$ events (~0.7% of the available)

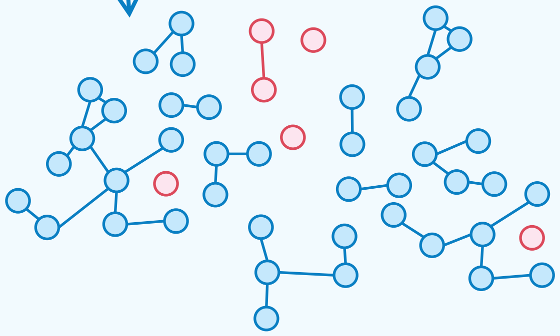
Signal – $17k$ events

- $m_Y = 3000\ GeV,\ m_X = 300\ GeV$
- Merged regime $\frac{m_X}{m_Y} = 0.1 < 0.3$ kinematic limit



Maximum number of nodes in a graph set to **40** – other nodes would not provide much more information and use computational time resources.





JET CONSTITUENTS TRANSFORMATION

A robust anomaly finder based on autoencoders

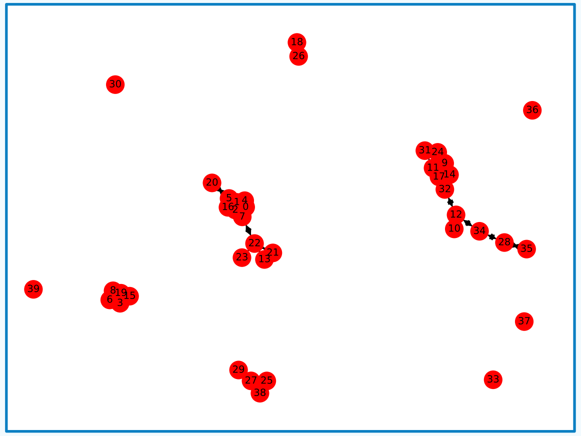
QCD-dijet data have a **wide spread distribution over their mass** and jets with a greater mass would be more important in the training, without any justified reason.

Transformation over jet constituents:

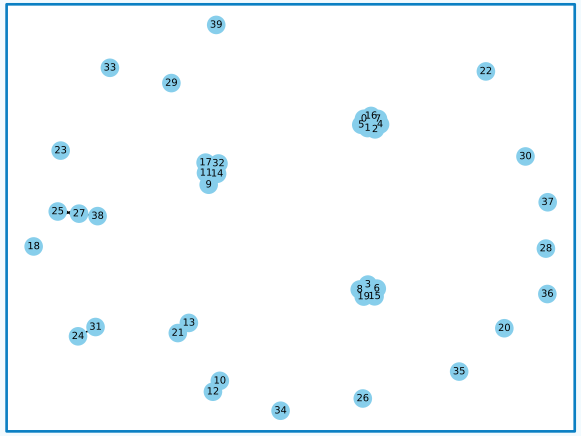
- **Rescaling the jets**: $m_j \rightarrow m_0 = 0.25\ \mathrm{GeV}$
- **Lorentz boost on jets**: $E_j \rightarrow E_0 = 1\ \mathrm{GeV}$
- **Rotation of constituents:** $\eta_j \rightarrow \eta_0 = 0$ and $\phi_j \rightarrow \phi_0 = 0$.

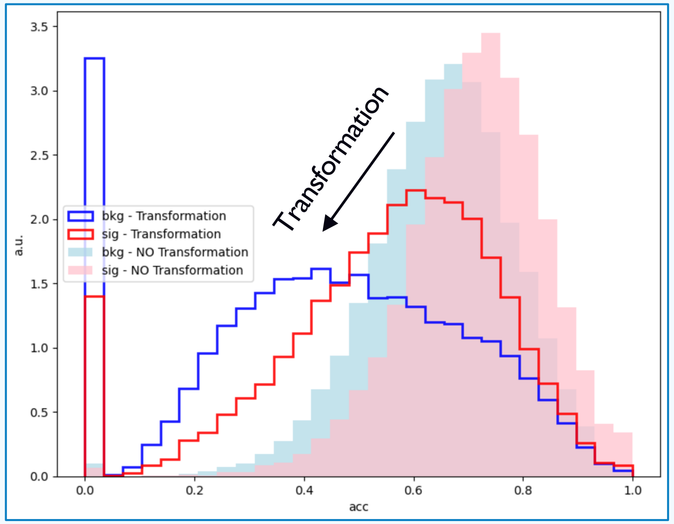
The effect of the jet constituents transformation is to modify graphs structure and, indirectly, to help with features separation improving the training performance.

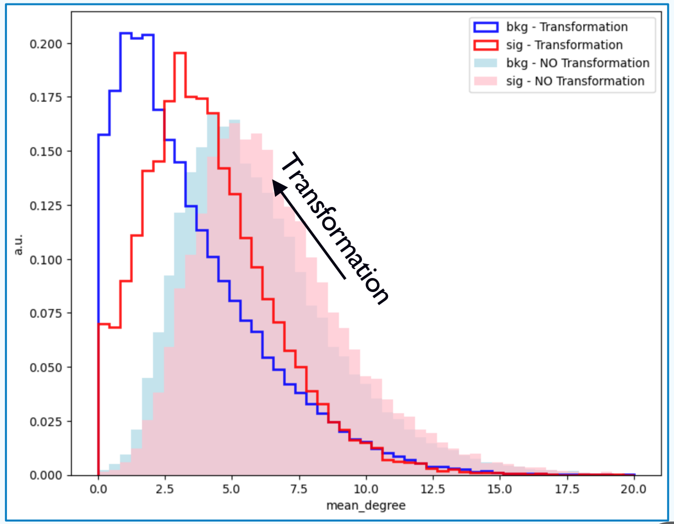
NO Transformation



Transformation

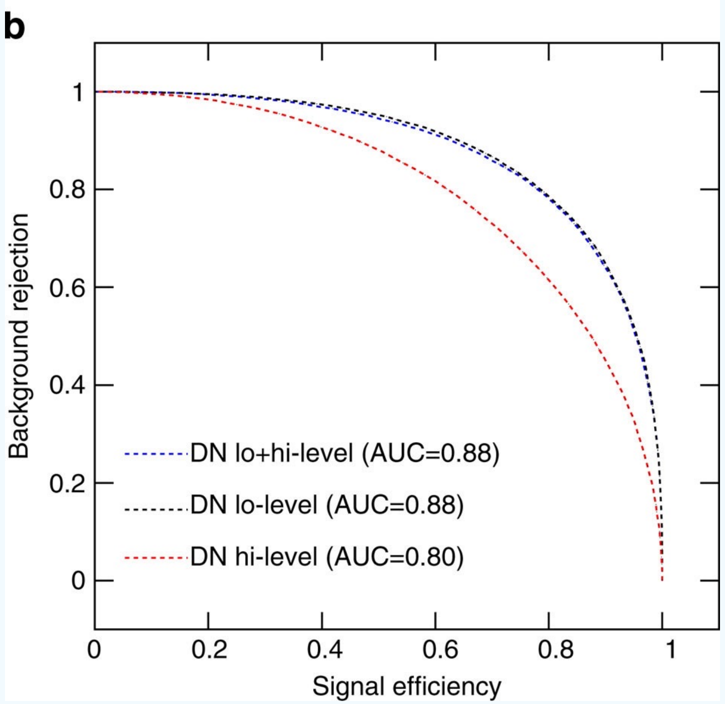


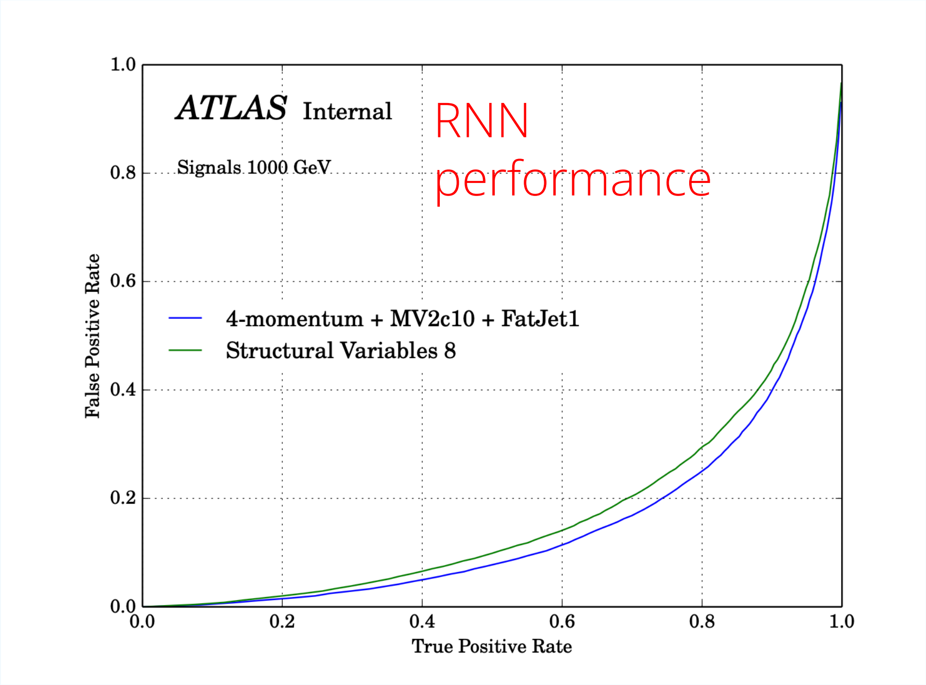




LOW-LEVEL VS HIGH-LEVEL FEATURES

"A set of features with basic information (low-level) such as information coming directly from the detectors, implies better performances with respect to features built combining basic information (high-level)"

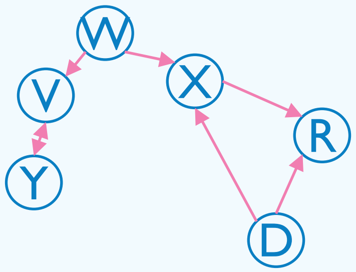




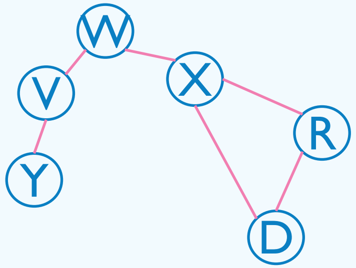
[Searching for exotic particles in high-energy physics with deep learning]

GRAPH THEORY

- Graphs are mathematical structures used to model pairwise relationships between objects.
- $G = (V, T, E)$
  - V set of nodes
  - T set of relations between edges
  - E set of edges
- Graphs consist of vertices (**nodes**) that represent entities or elements within a system and **edges** (connections) that represent relationships the vertices.
- Types of Graphs:
  - **Directed Graphs:** edges have a direction, indicating a one-way relationship between vertices.
  - **Undirected Graphs:** edges have no direction, indicating a symmetric relationship between vertices.
- Understanding graph theory provides valuable insights into the structure and relationships within complex systems, enabling the development of efficient algorithms and solutions across diverse domains.



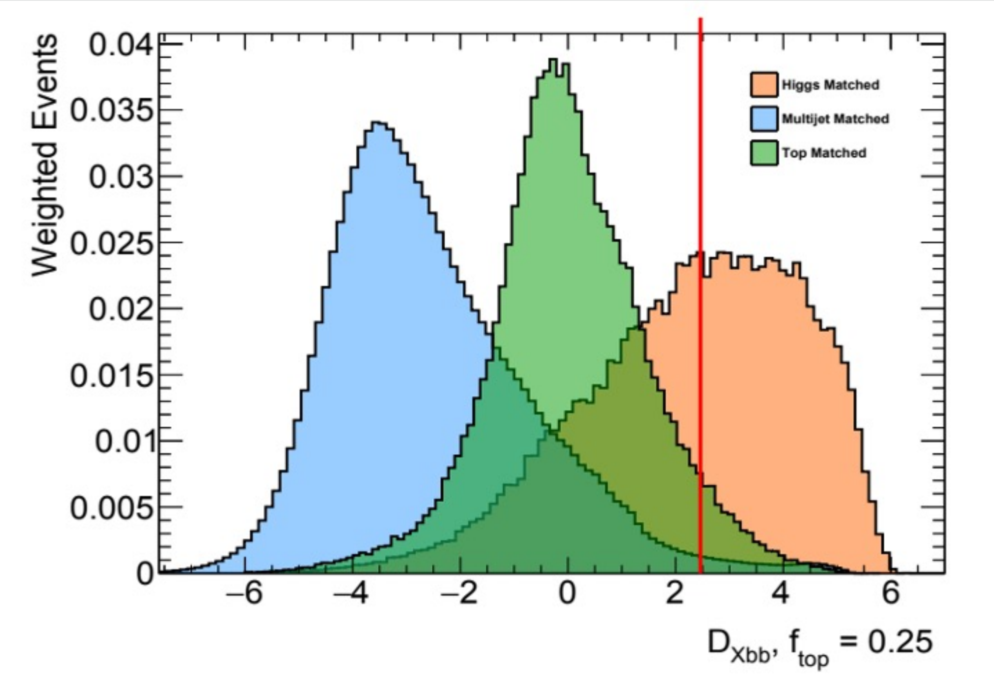
Directed



Undirected

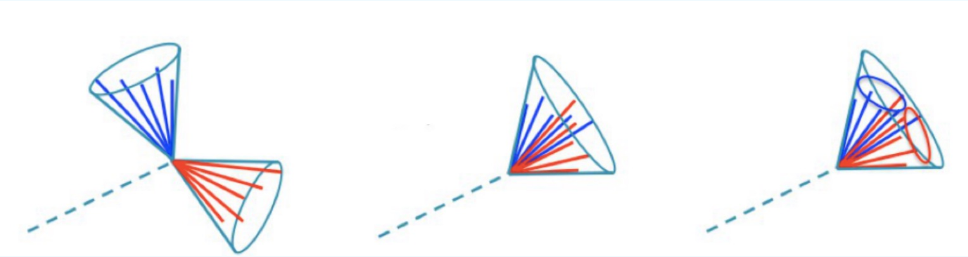
H BOSON SELECTION

- Based on the **output scores** (probabilities of tagging as Higgs, top or multijet) of a NN assigned to each jet.
  - p_{Higgs}
  - p_{top}
  - $p_{multijet}$
- Scores are combined in a unique value $D_{H_{bb}}$ for each jet.
  - $D_{H_{bb}} = \ln \frac{p_{Higgs}}{f_{top}\, p_{top} + (1 - f_{top})\, p_{multijet}}$;
  - f_{top} define the weight of the top background shape.
- Higher scores correspond to jets that are more likely to originate from Higgs to bb decays.



X BOSON SELECTION

- Once the H candidate is selected, the other of the two leading jet is automatically defined as **X candidate.** The X boson can be identified as one large-R jet or two small-R jet depending on masses ratio.
- **Merged**
  - If the mass ratio between X and Y resonances is small (<0.3) the X resonance is reconstructed via large-R jet
- **Resolved**
  - For larger masses ratio the decay products are no longer collimated and the resonance is reconstructed via small-R jets.
    - At least 4 small radius jets are required in the event;
    - Small jet pair with the minimum ΔR from the Higgs candidate (reconstructed as large-R jet) are discarded;
    - In the remaining jet collection, the X candidate is reconstructed taking the p_T leading and subleading.



$$R \sim \frac{2m}{p_T}$$