

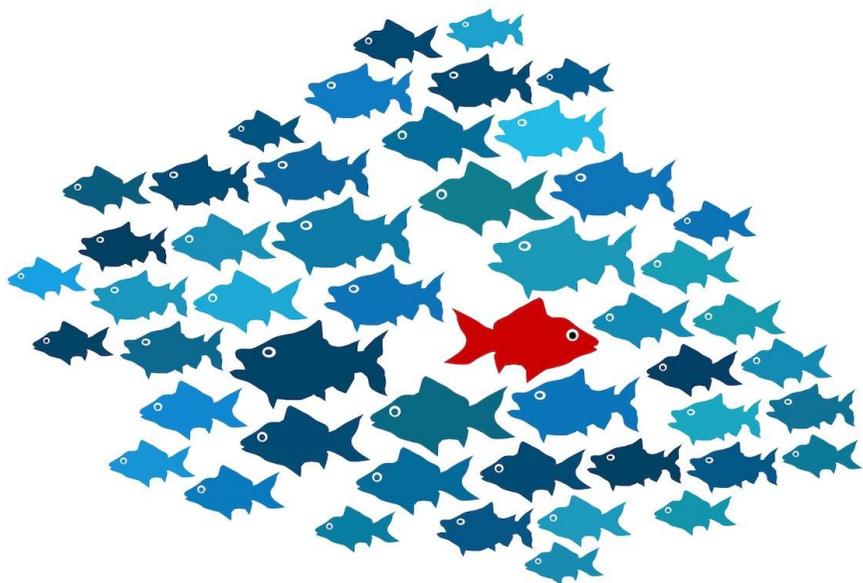
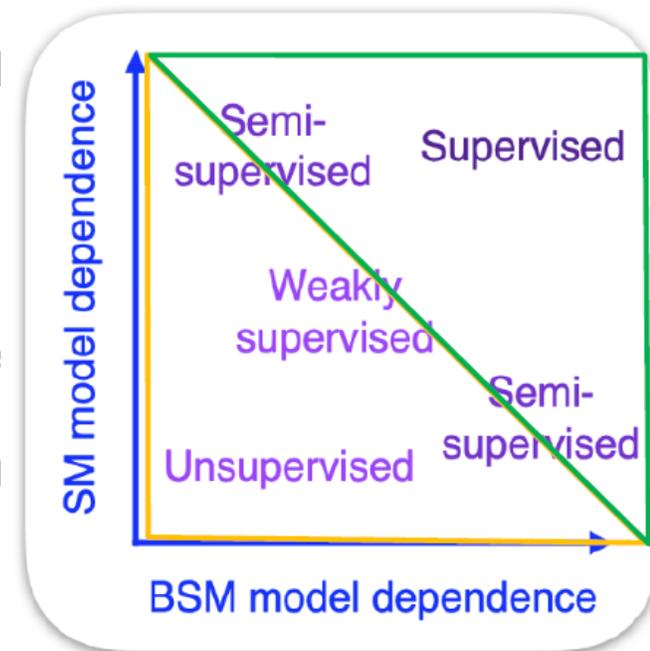


UNIVERSITÀ DEGLI STUDI DI NAPOLI  
FEDERICO II

APPROCCIO DI ANOMALY DETECTION PER LA RICERCA DI  
RISONANZE DIBOSONICHE AD ALTA MASSA IN STATI FINALI  
COMPLETAMENTE ADRONICI UTILIZZANDO COLLISIONI PP A  
 $\sqrt{s} = 13$  TEV CON IL RIVELATORE ATLAS

# INTRODUZIONE

- Attualmente ad LHC non sono stati osservati eventi di Fisica oltre il Modello Standard (BSM)
- La maggior parte delle ricerche utilizza un approccio **model-dependent**
  - E se ci fosse un segnale diverso dai modelli scelti?
- Negli approcci **model-independent** ci sono minime assunzioni sulle proprietà del segnale
  - Non ottimale come il model-dependent, ma garantisce una ricerca a più ampio spettro.
  - Utilizzo di Machine Learning non supervisionato!



- Nell'**Anomaly Detection** (AD) si usano architetture Machine Learning per identificare valori anomali in un insieme di oggetti standard
  - Nella Fisica delle Alte Energie quest'applicazione consiste nell'identificazione di caratteristiche negli eventi ricostruiti non consistenti con il fondo atteso.

# AD IN STATI FINALI COMPLETAMENTE ADRONICI

Molti processi di fisica BSM coinvolgono particelle massive le quali decadono in stati finali adronici, quindi jets.



Se  $m_{Z'} \gg m_X \wedge m_Y$



Allora X e Y si trovano in un regime di alto impulso trasverso

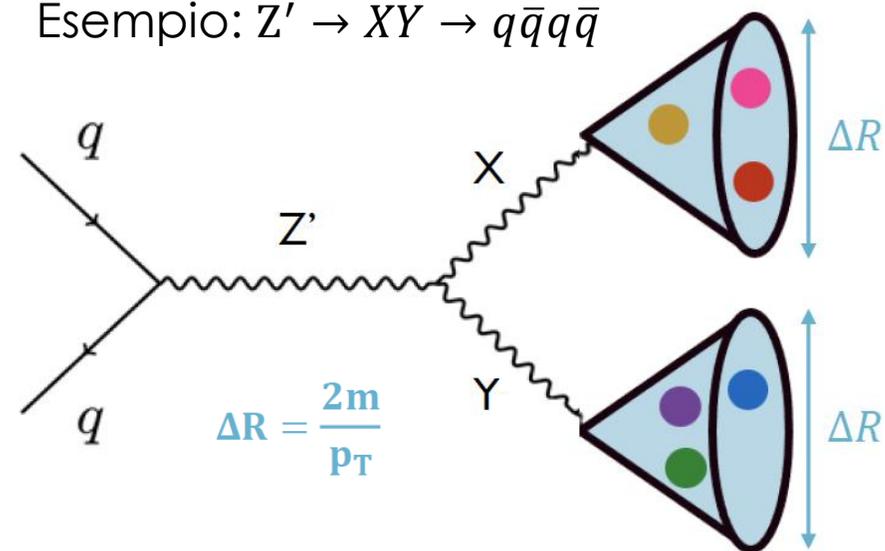


È vantaggioso ricostruire questo tipo di eventi come un unico jet

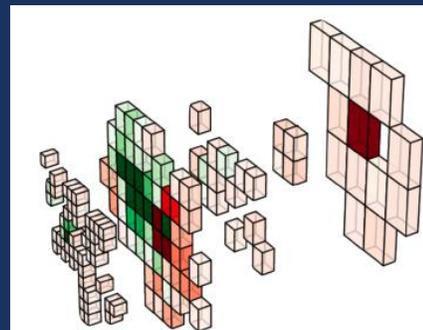


Ogni evento della nostra analisi deve contenere due large R-jets nello stato finale. (Anti-kT con  $R=1$ )

Esempio:  $Z' \rightarrow XY \rightarrow q\bar{q}q\bar{q}$



I costituenti sono i rilasci energetici dello sciame adronico generato da quark nel rivelatore, usati per la ricostruzione dei jet tramite algoritmo Anti-kT



Ricerca di una nuova risonanza  $Y \rightarrow XH \rightarrow q\bar{q}b\bar{b}$  con il dataset del Run 2:

- Modello Heavy Vector Triplet:
  - Range di massa:  $m_Y = 1 \div 6 \text{ TeV}$ ,  $m_X = 65 \div 3000 \text{ GeV}$ ;
- Identificazione dei candidati H e X:

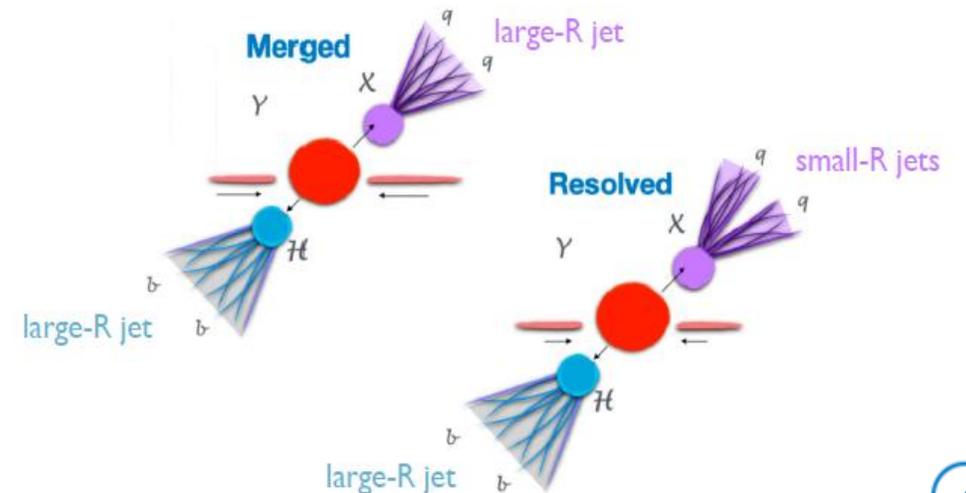
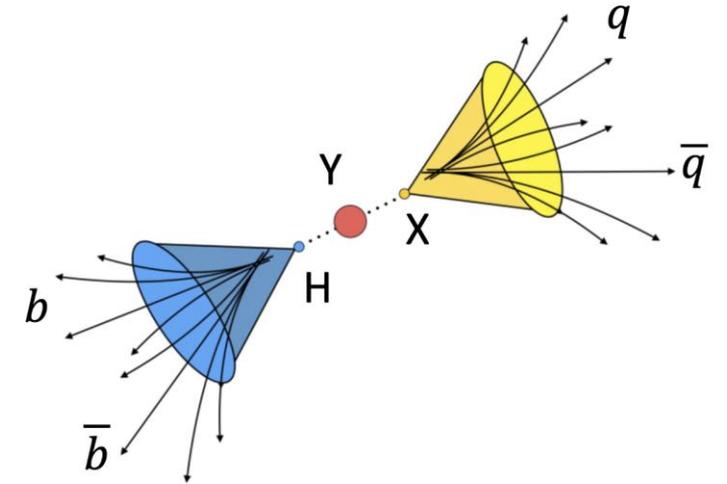
$H^0$  identificato con una DNN supervisionata allenata per ricercare risonanze in  $b\bar{b}$ .

$X$  identificato applicando diverse strategie di selezione:

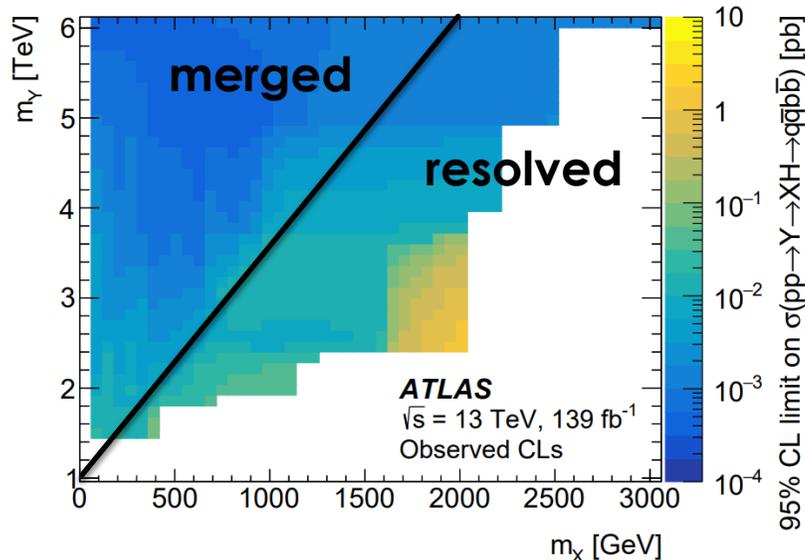
**Model-independent:** Score AD, in particolare una **VRNN**, che prende in input i primi 20 costituenti di ogni large-R jet con  $p_T > 1.2 \text{ TeV}$ .

**Model-dependent:**

- Resolved  $\frac{m_X}{m_Y} > 0.3 \Rightarrow$  2 small-R jets;
- Merged  $\frac{m_X}{m_Y} < 0.3 \Rightarrow$  1 large-R jet;

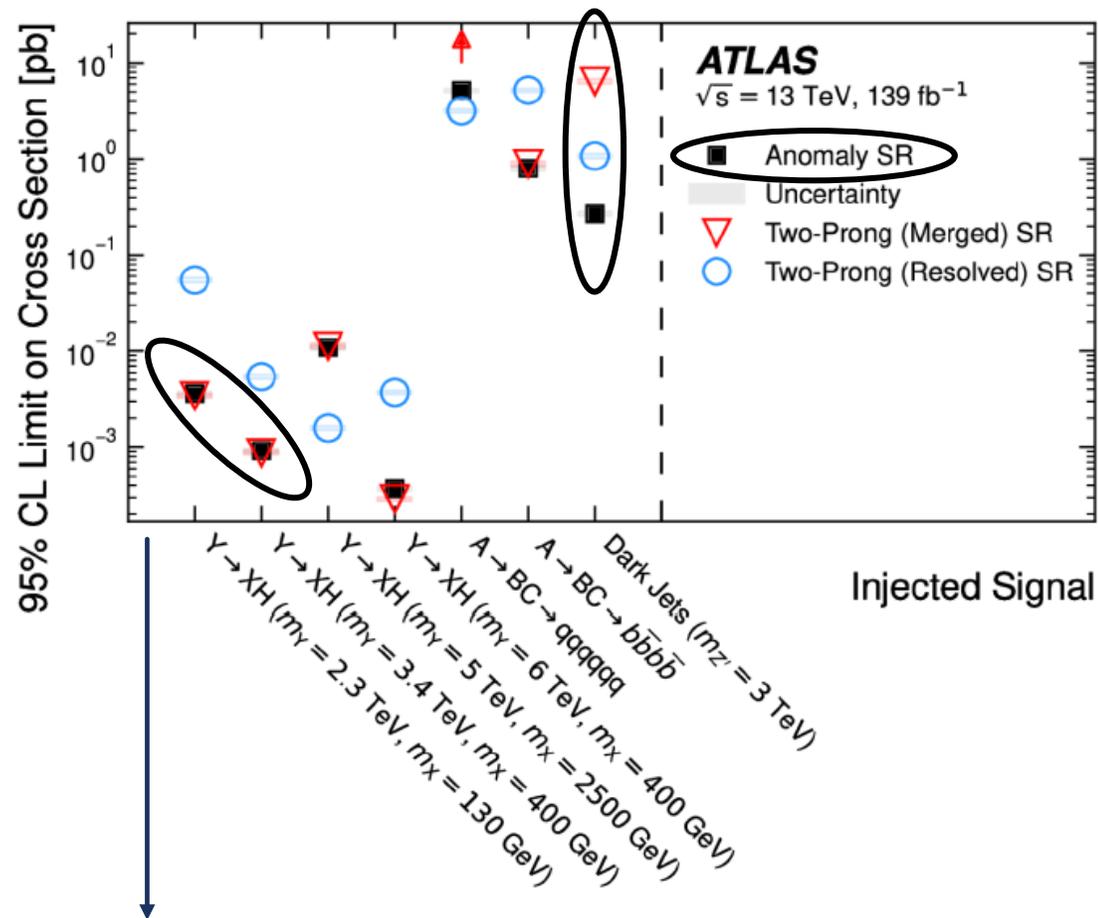
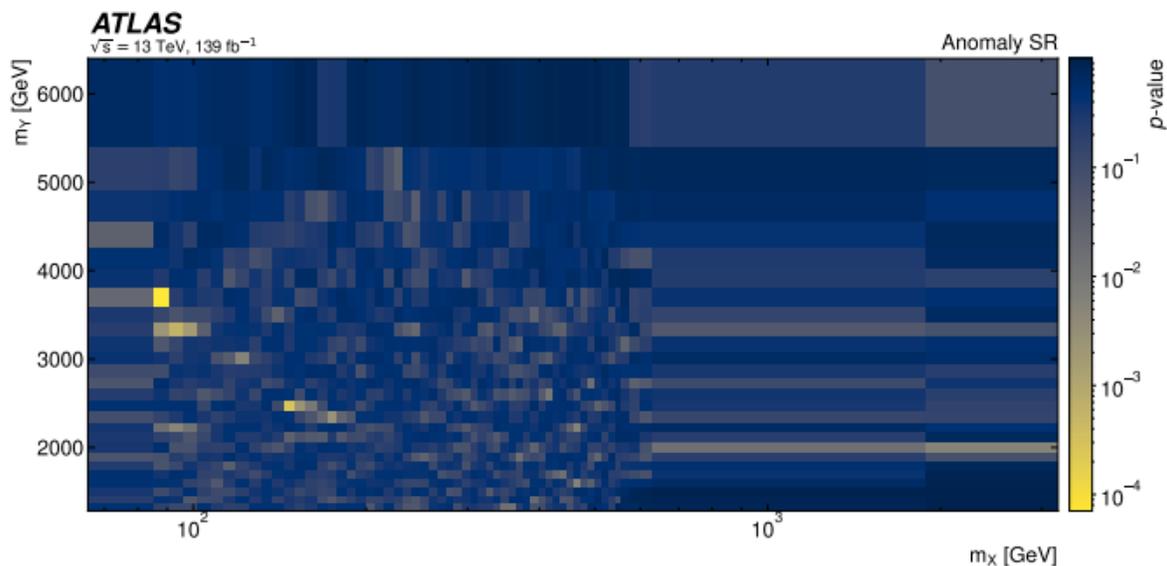


Regioni merged e resolved combinate:  
Fit in ipotesi di solo fondo e di fondo + segnale sul rapporto di verosimiglianza.



## Regione Anomaly detection:

Calcolo p-valori in ipotesi di solo fondo con BumpHunter.



I risultati finali mostrano che la AD è competitiva rispetto alla controparte model-dependent. Nel caso di Dark Jets si ottengono addirittura risultati migliori.

# GRAFI E JET

- I **grafi** sono un insieme di elementi detti nodi che possono essere collegati fra loro da linee chiamate archi (edges)
- A partire dai grafi è possibile definire grandezze come:

**Coefficiente di clustering** → misura della tendenza dei nodi a raggrupparsi insieme;

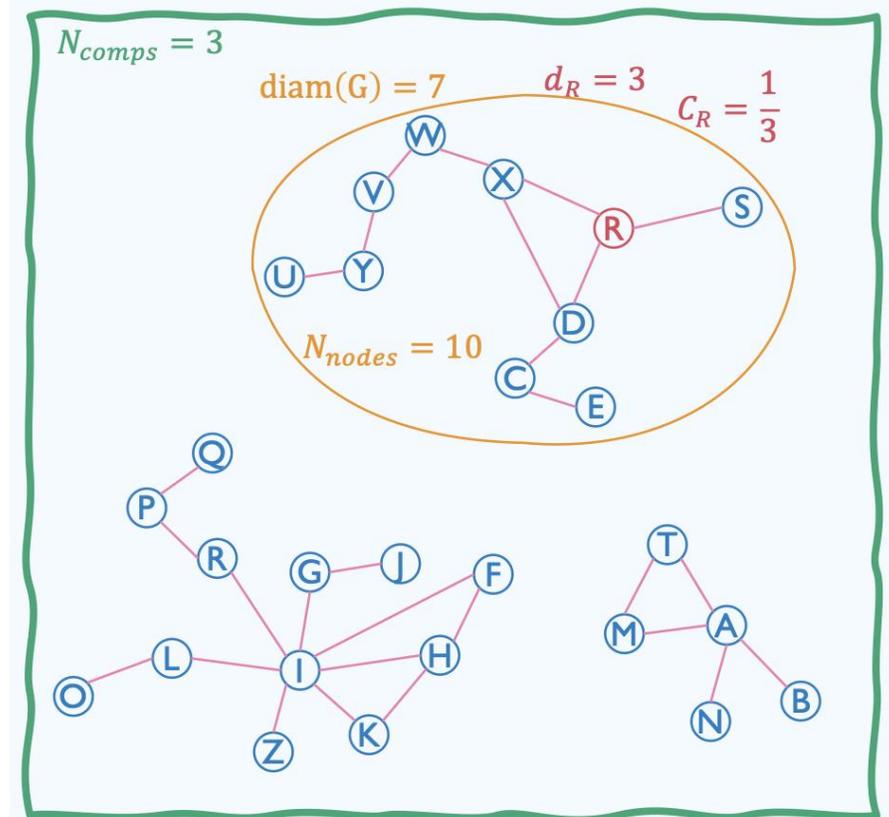
**Mean degree** e **diametro** del grafo

**Il numero dei nodi** e delle componenti connesse

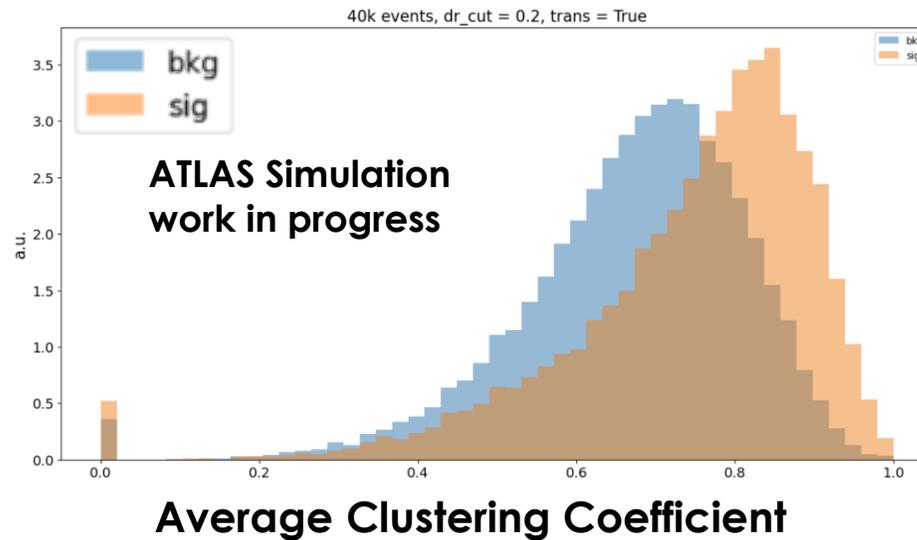
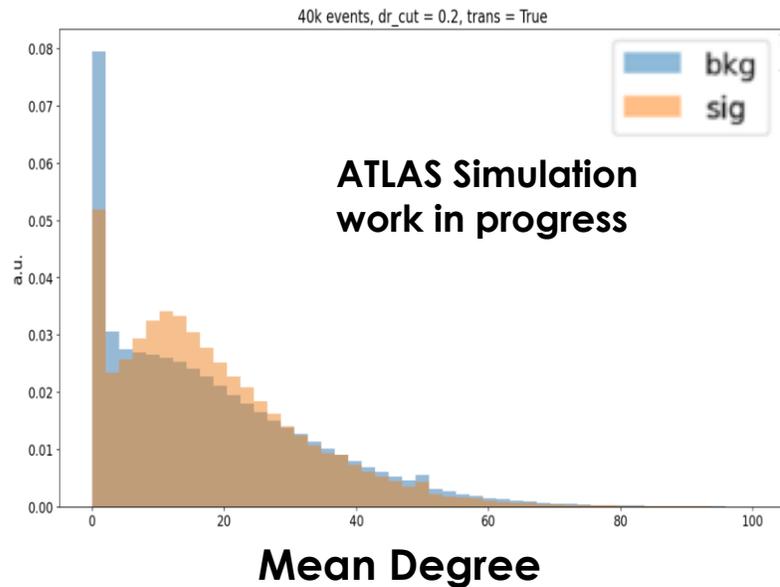
I jet hanno una struttura interna sparsa!

I costituenti calorimetrici possono essere utilizzati per creare un oggetto di tipo grafo.

- I grafi costruiti per la ricerca sono pesati e hanno il self-loop:
  - I nodi sono i costituenti con features: [frazione di  $p_T$ ,  $\eta$ ,  $\phi$ ]
  - Gli edges rappresentano la relazione tra i nodi:  $1/\Delta R$ , presente se  $\Delta R < 0.2$ , con  $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$



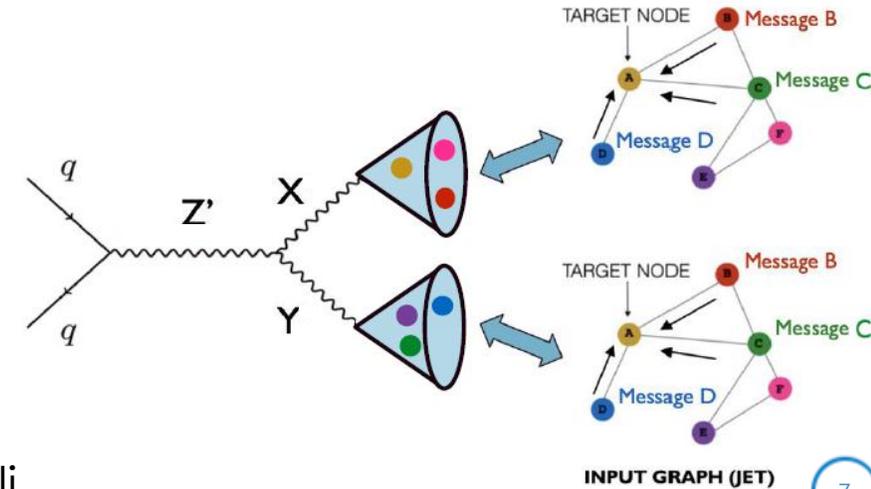
# JET-GRAPHS



R&D sul [LHC Olympics2020](#) dataset: è un dataset pubblico contenente simulazioni fondo + segnale utilizzato per studi di ML

- Eventi QCD dijet di fondo;
- $Z' \rightarrow XY \rightarrow q\bar{q}q\bar{q}$  segnale, con  $m_{Z'} = 3.5 \text{ TeV}$ ,  $m_X = 500 \text{ GeV}$  e  $m_Y = 100 \text{ GeV}$ ;

- Costruiamo per ogni jet un grafo a partire da tutti i suoi costituenti. La rete di grafi (GNN) ci restituirà un Anomaly Score (AS) per ogni jet.
- L'allenamento con validazione della GNN viene svolto sul fondo e poi testato su un dataset fondo + segnale. Ci aspettiamo di ottenere AS maggiori per i jet del segnale.
- È possibile definire un AS per l'evento combinando gli AS dei singoli jet.



# MECCANISMO DI FUNZIONAMENTO GNN

Le rappresentazioni di ogni nodo ad ogni layer sono aggiornate aggregando l'informazione passata tra il nodo in questione e tutti i nodi a lui più vicini ↔ «Message Passing»

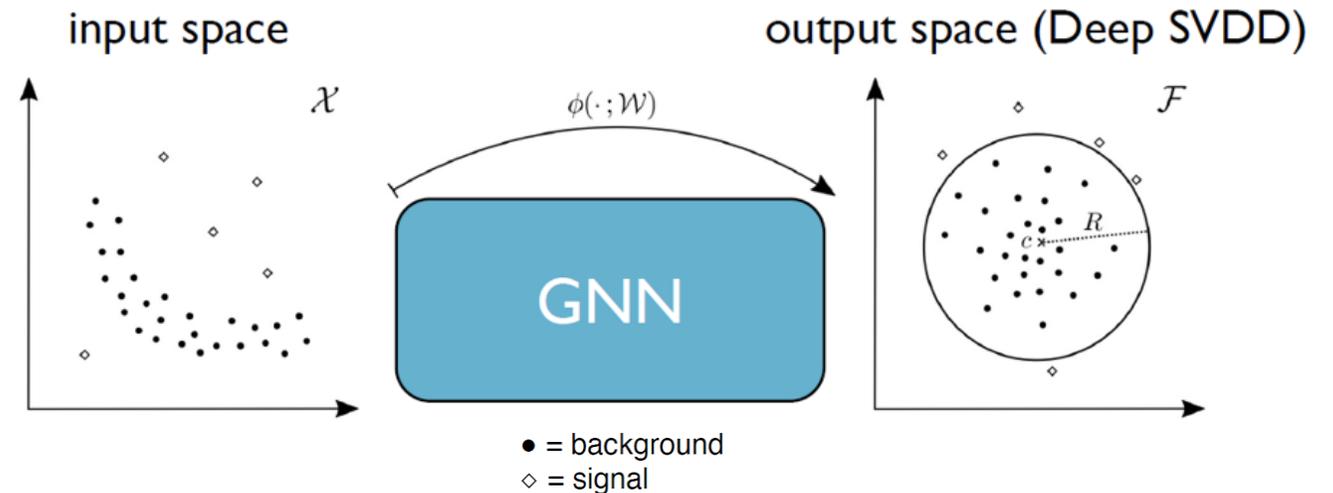
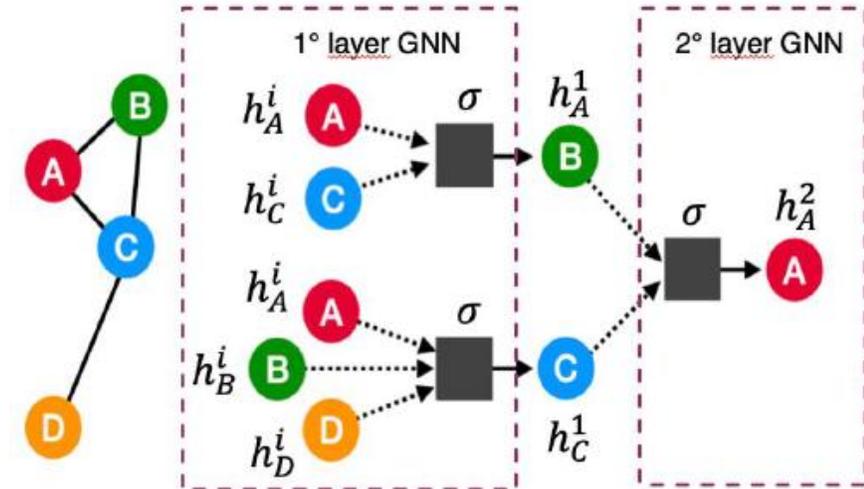
Al passare di ogni layer la rappresentazione del nodo sarà arricchita da informazioni provenienti da nodi sempre più distanti nel grafo!

Due architetture provate:

1. EGAT, Edge Graph Attention Network: meccanismo di attenzione + aggiornamento edges
2. GIN, Graph Isomorphism Network: Multi Layer Perceptron Networks sulle rappresentazioni. Sono le GNN più espressive per questo motivo.

$$AS = \|\phi(G_i; W^*) - c\|^2$$

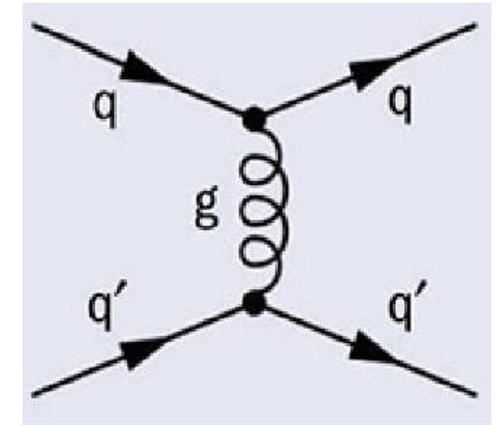
Generico esempio di layer GNN:



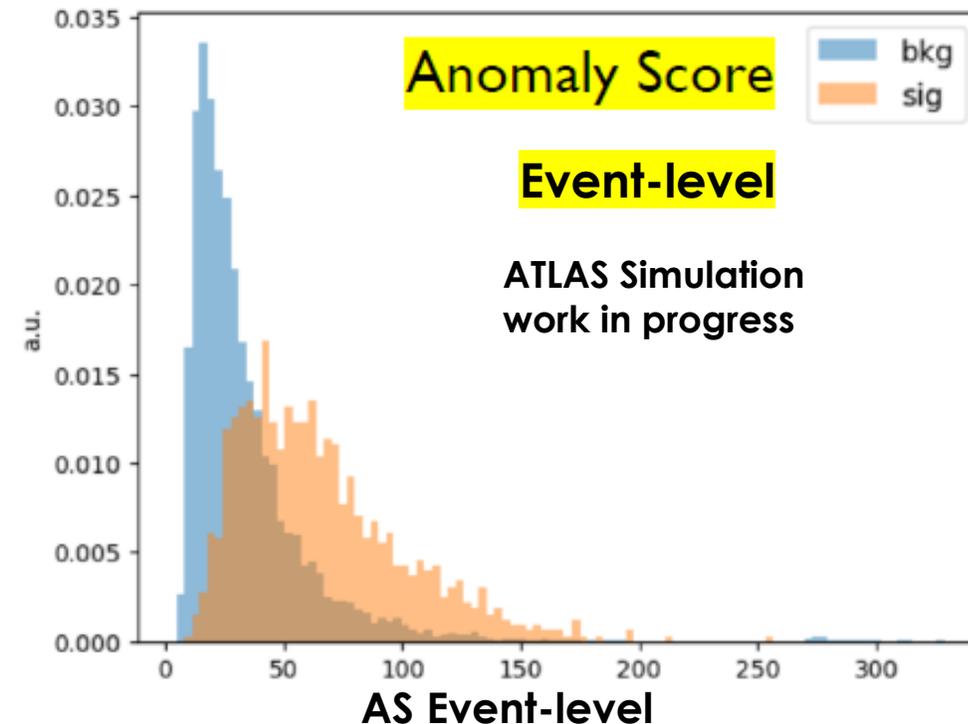
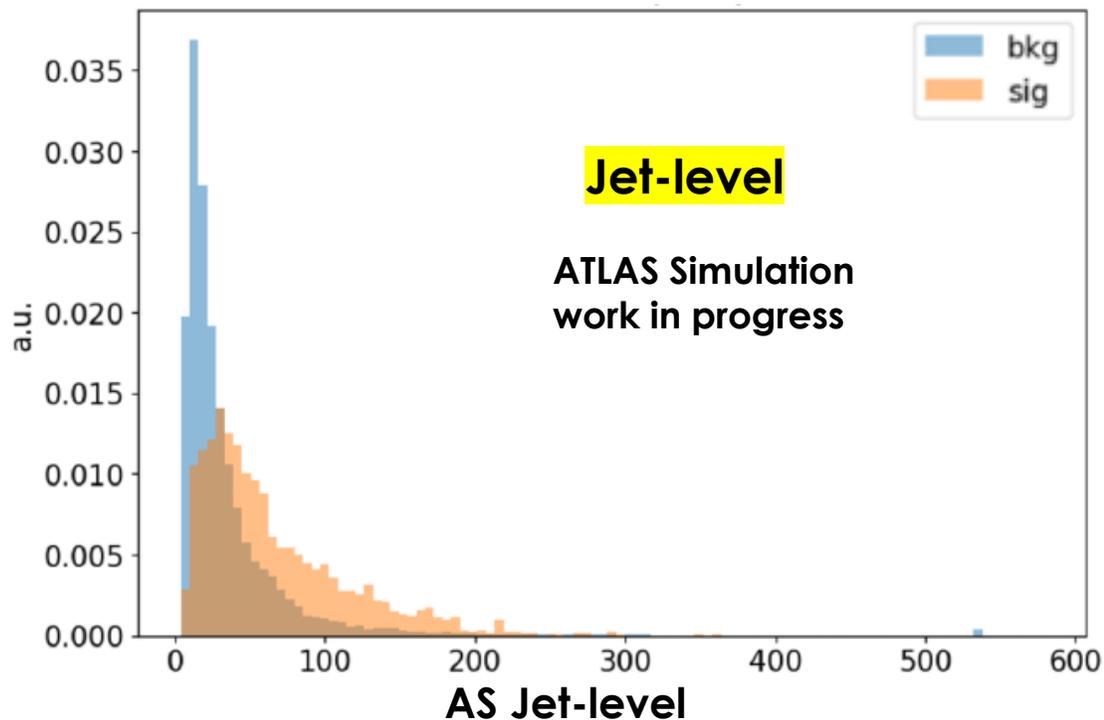
# RISULTATI LHC OLYMPICS

R&D sul LHC Olympics dataset: [LHC Olympics 2020](#) è un dataset pubblico contenente simulazioni fondo + segnale utilizzato per studi di ML

- Eventi QCD dijet di fondo; 
- $Z' \rightarrow XY \rightarrow q\bar{q}q\bar{q}$  segnale, con  $m_{Z'} = 3.5 \text{ TeV}$ ,  $m_X = 500 \text{ GeV}$  e  $m_Y = 100 \text{ GeV}$ ;
- Preselezione a  $p_T > 1.2 \text{ TeV}$  e  $|\eta| < 2.5$ .

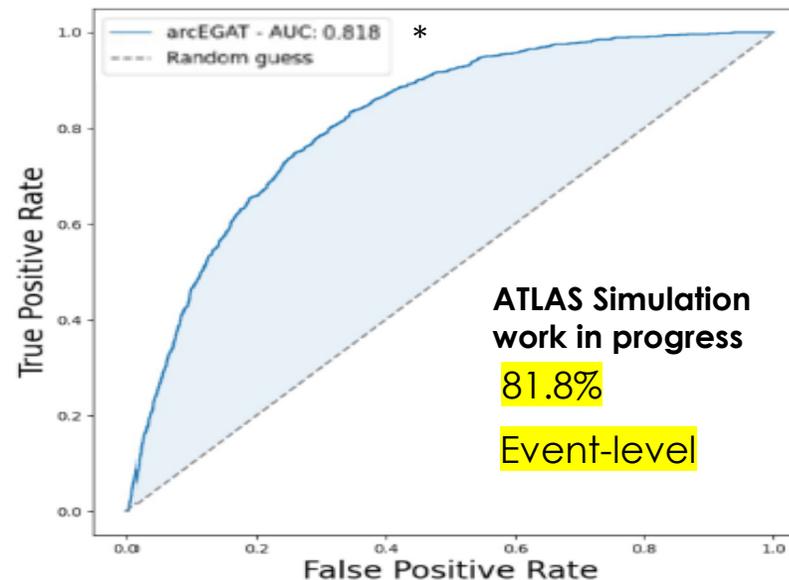
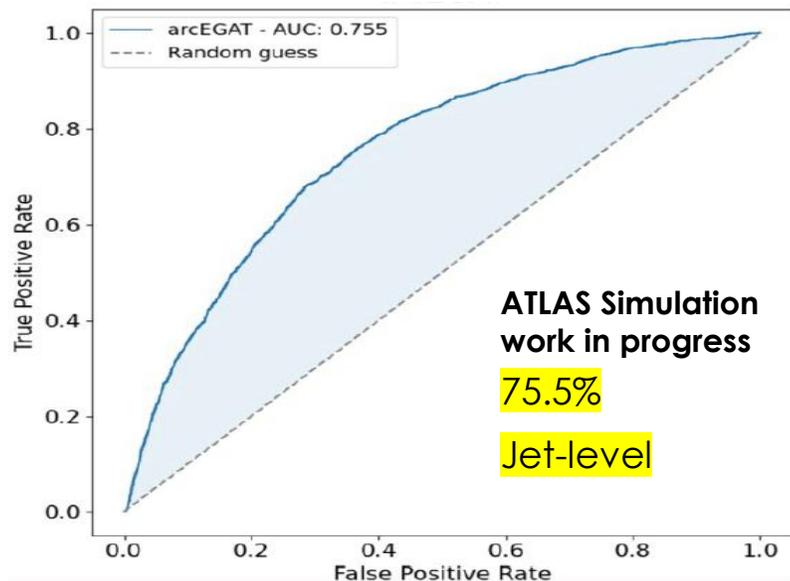


Definiamo AS per un evento come la media degli AS dei i due jet.



# RISULTATI LHC OLYMPICS

Usiamo l'AUC come metrica per valutare la capacità della rete di separare il segnale dal fondo.

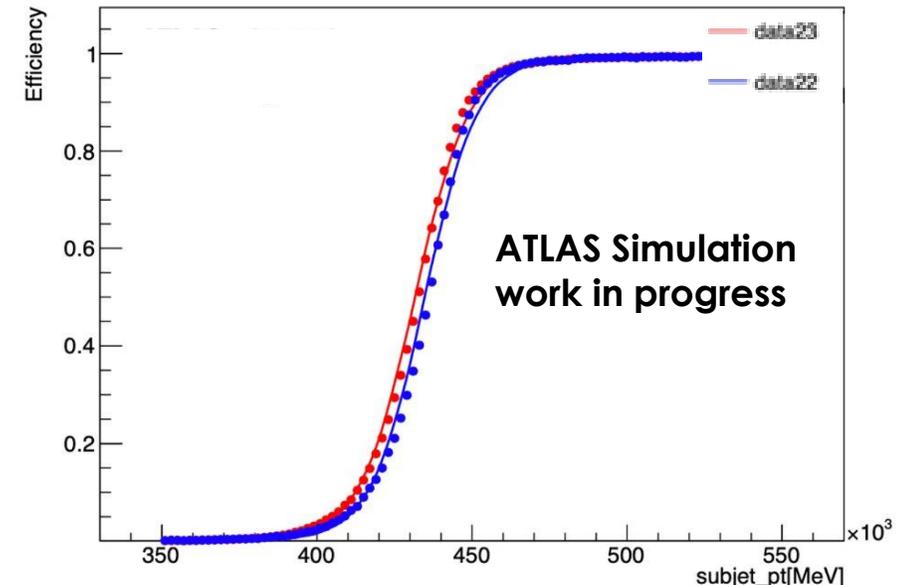


\* AUC = Area Under the ROC Curve

Model	GIN supervised	EGAT supervised	GIN unsupervised	EGAT unsupervised
loss	CrossEntropy	CrossEntropy	DeepSVDD	DeepSVDD
AUC jet-level 2-prong	90.2%	89.9%	73.7%	75.5%
AUC event-level 2-prong	96.5%	96.5%	79.6%	81.8%

# ANOMALY DETECTION IN STATI FINALI COMPLETAMENTE ADRONICI IN ATLAS

- Ricerca di nuove risonanze in stati finali completamente adronici in ATLAS con tecniche AD
- Tipo di analisi: completamente data driven
- Dataset:
  - Dati raccolti da ATLAS durante il Run-3 (2022-2024)
  - Fondi simulati Monte Carlo: QCD dijet (dominante), top-antitop, V+jets;
  - Segnali utilizzati come benchmark (HVT YXH, HVT W'WW, Dark Jets).
- Regione di segnale definibile con un taglio sull'Anomaly Score
- Selezione degli eventi: 2 large-R jet per evento, preselezione e richiesta di trigger applicate
  - Possibilità di studiare i nuovi item di trigger basati su large R-jet, che permettono di recuperare eventi a bassa massa del sistema dijet ( $1 \leq m_{jj} \leq 1.3 \text{ TeV}$ )



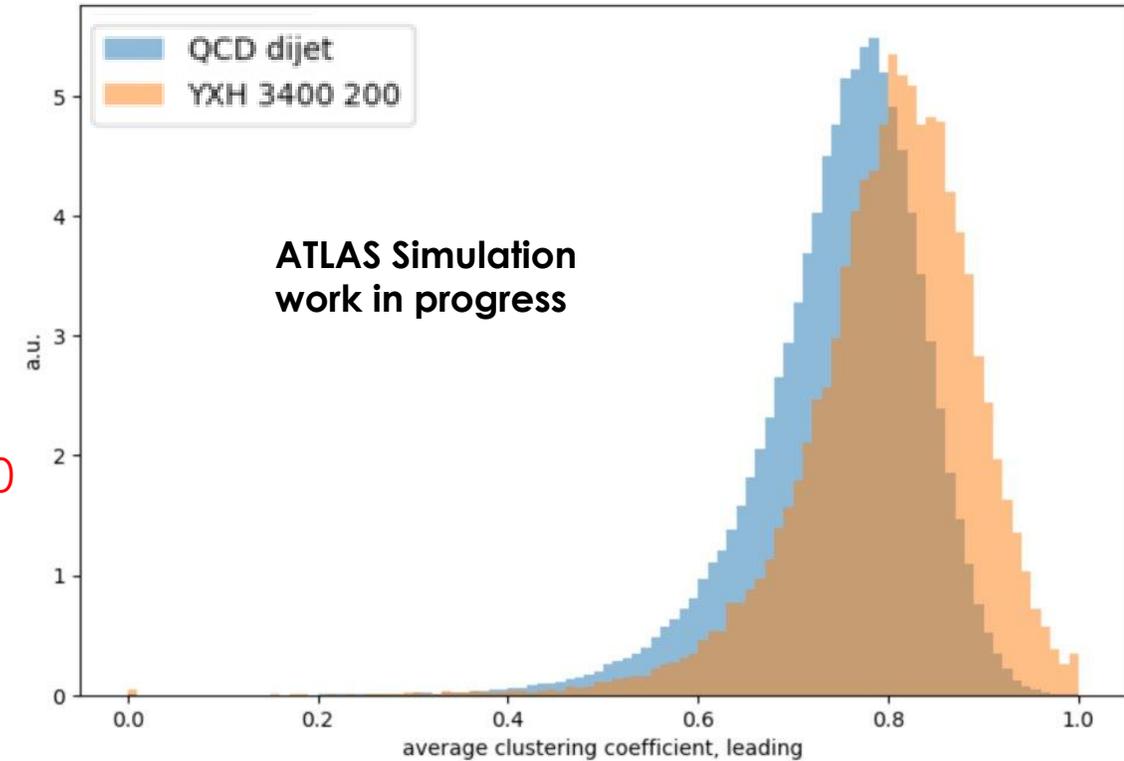
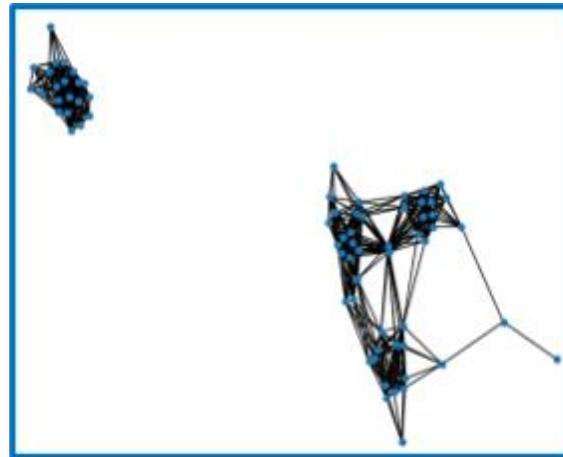
# ANOMALY DETECTION IN STATI FINALI COMPLETAMENTE ADRONICI IN ATLAS

- Introdotte migliorie:
  - Aggiunte ai nodi dei grafi extra features tra cui la massa dei costituenti;
  - Feature degli edges modificati come  $e^{-\Delta R}$ ;
- Nuove definizioni di grafo in architettura EGAT:
  - Jet-level come per LHC Olympics;
  - Event-level creando un grafo singolo per ogni coppia di jet dell'evento;
- Risultati preliminari sul fondo Monte Carlo QCD
  - Confronto tra gli score Event-level ottenuti con le due definizioni di grafo.

Grafo evento QCD dijet



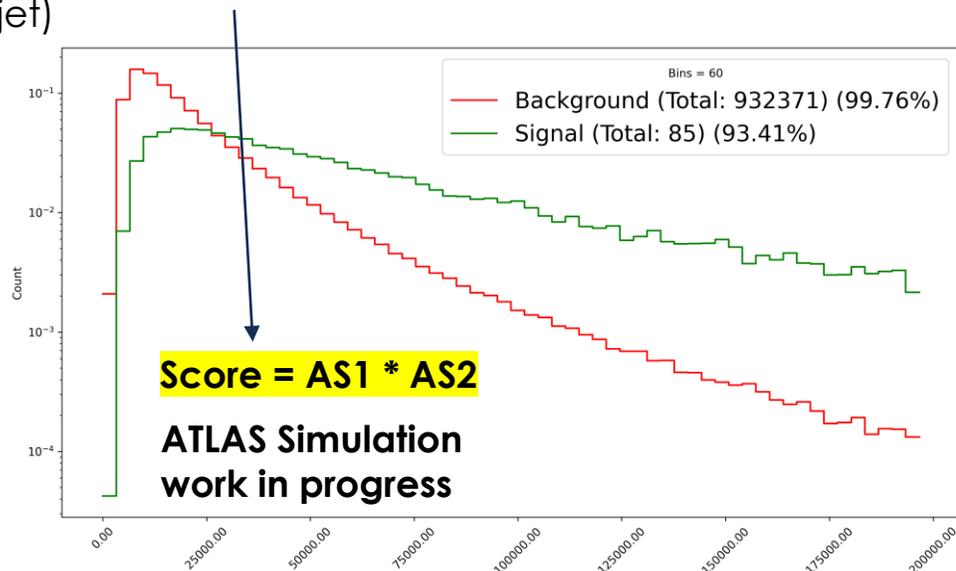
Grafo evento YXH 3400 200



# RESULTATI PRELIMINARI

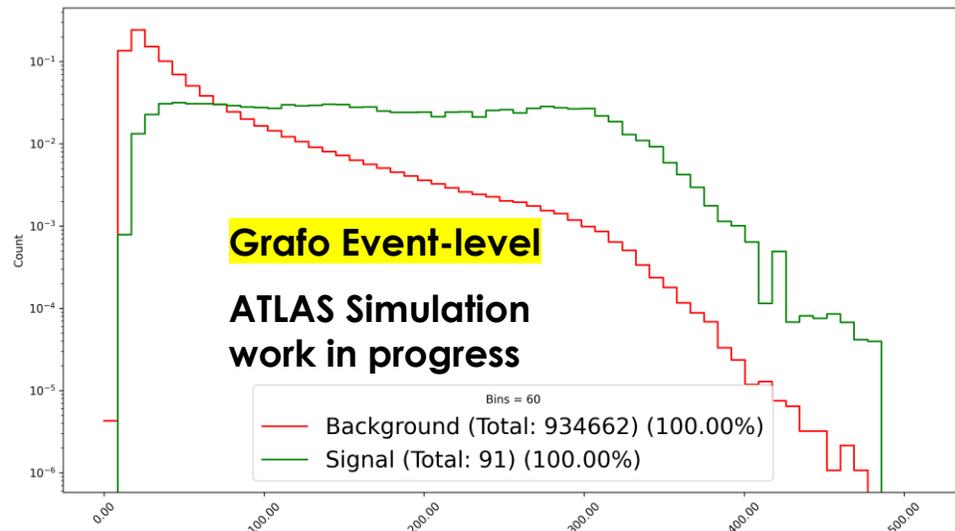
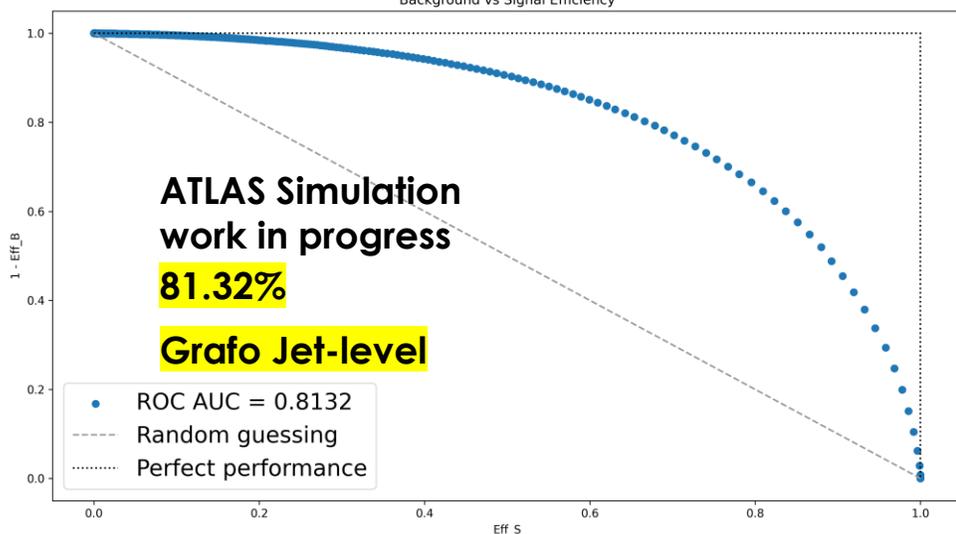
Lo score Event-level è ottenuto moltiplicando gli score dei singoli jet (un grafo per jet)

YXH con  $m_Y = 3400 \text{ GeV}$  e  $m_X = 200 \text{ GeV}$   
Fondo QCD MC di circa 1M di eventi



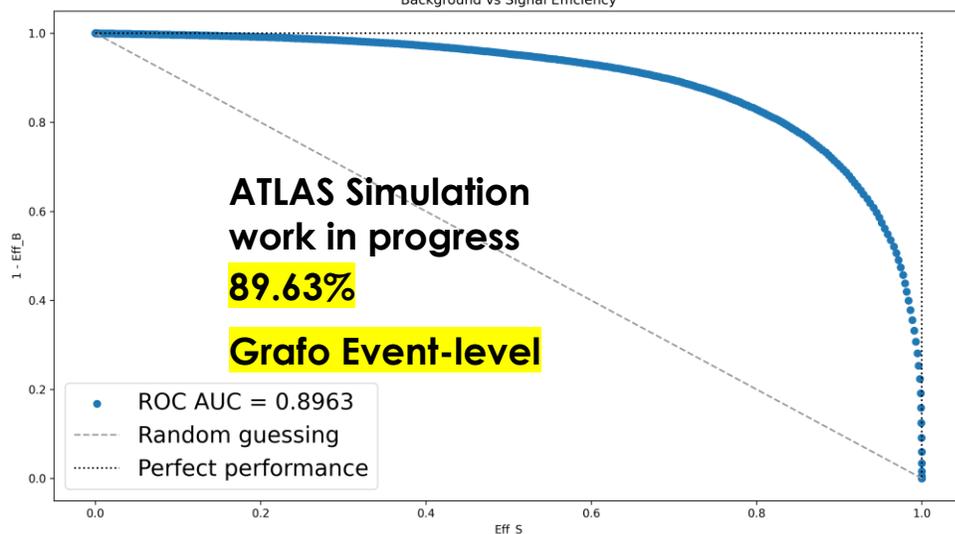
Event-level AS

Background vs Signal Efficiency



Event-level AS

Background vs Signal Efficiency



# CONCLUSIONI

---

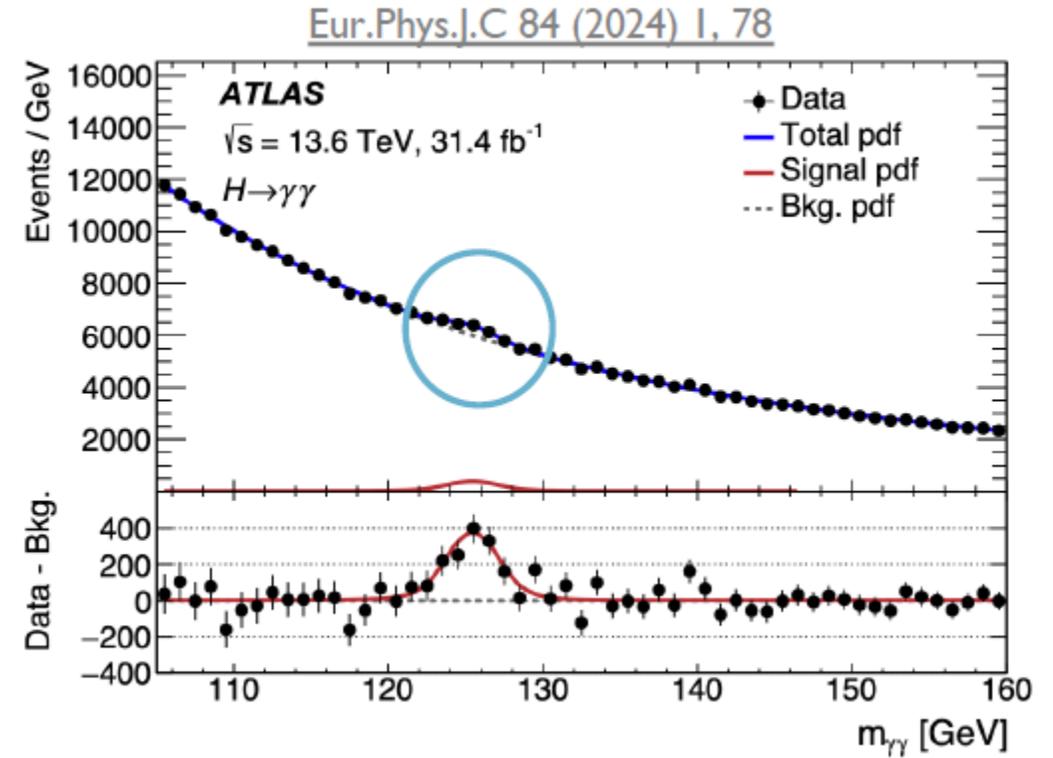
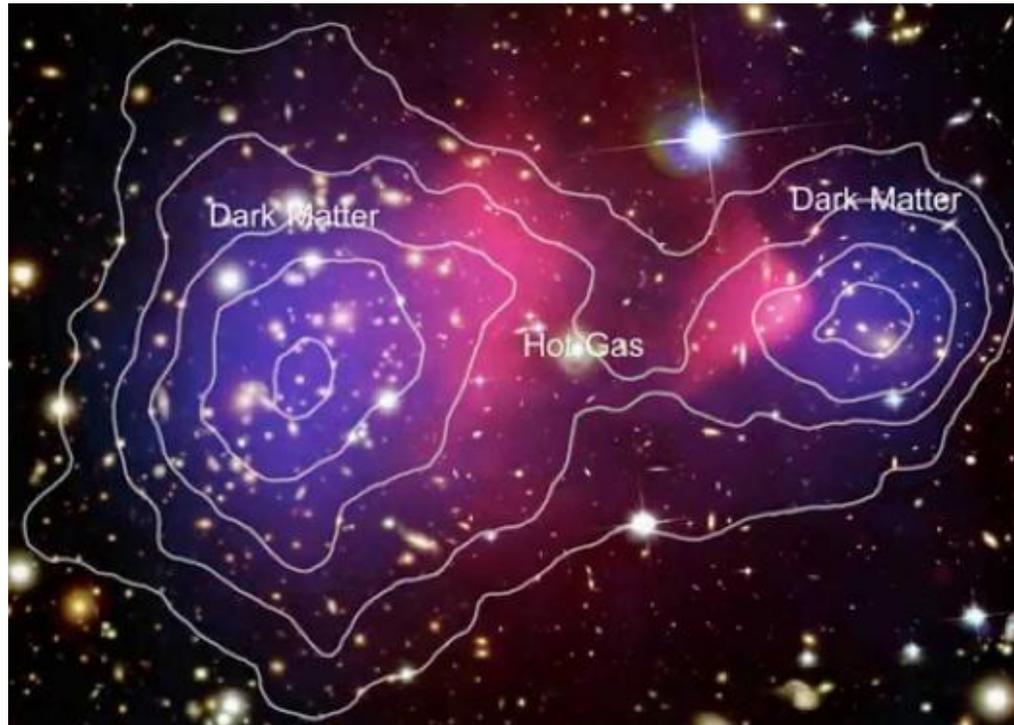
- Le tecniche **Anomaly detection** sono diventate parte integrante in HEP e vengono utilizzate su più fronti: *ricerche BSM, identificazione e ricostruzione, metodi indipendenti dal modello, data quality monitoring*
- I risultati ottenuti con la prima analisi **AD in ATLAS (Y->XH)** per la ricerca BSM molto incoraggianti
- L'utilizzo di tecniche ML avanzate come **GNN** rende l'applicazione in ambito di ricerche unsupervised BSM molto promettente ma allo stesso tempo molto complesso e impegnativo sotto diversi aspetti.
- Sviluppo in ATLAS di **un nuovo algoritmo AD nelle ricerche BSM** in stati finali completamente adronici:
  - *R&D sul LHC Olympics dataset mostra risultati promettenti per la performance della GNN per la AD;*
  - *Analisi GNN per AD sul Run 3 dataset (analisi e risultati attesi entro il 2026)*

---

# BACKUP

# INTRODUZIONE

- Il Modello Standard ha avuto un enorme successo sperimentale.
- Scoperta del Bosone di Higgs  $H^0$  nel 2012 da parte di ATLAS e CMS.



- Tante evidenze sperimentali che necessitano ancora di una spiegazione.
- Ad LHC non è stata osservata fisica che va oltre il Modello Standard (BSM).

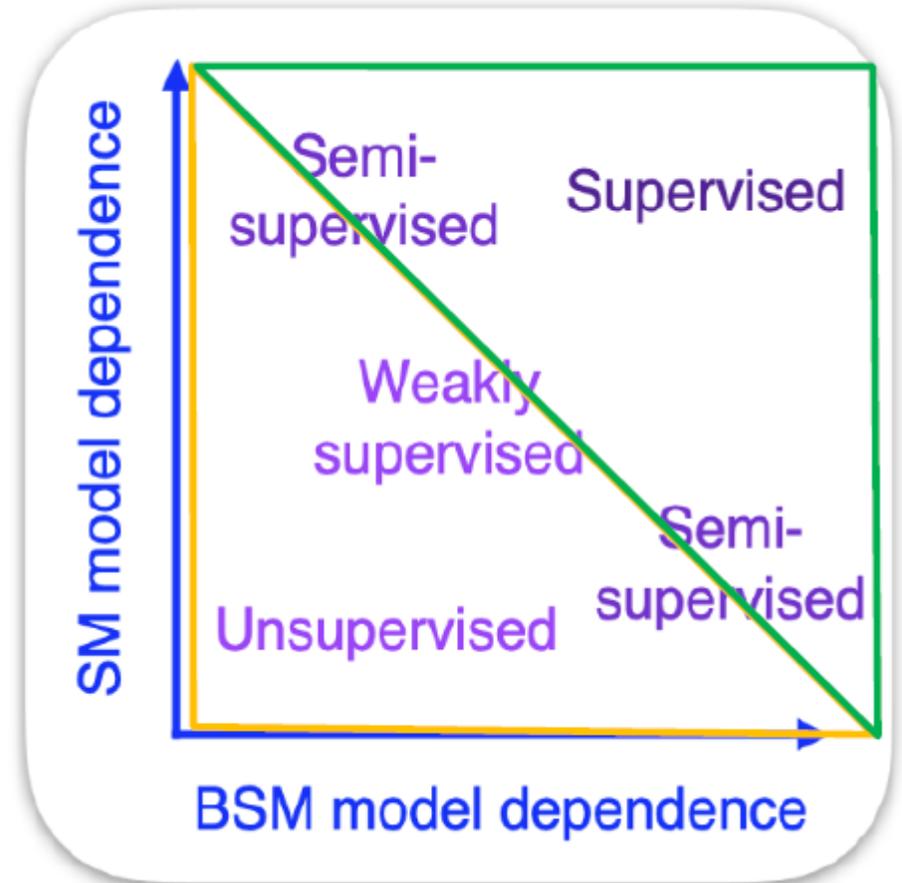
# INTERPRETAZIONI DEI RISULTATI DI UNA RICERCA

## Approccio model-dependent:

- Ricerca di uno specifico segnale;
- Selezione di eventi ottimizzata utilizzando le proprietà del segnale conosciute;
- La sensibilità diminuisce per processi con segnale diverso dal modello scelto.

## Approccio model-independent:

- Minime assunzioni sulle proprietà del segnale;
- Non ottimale come il model-dependent segnale per segnale, ma garantisce una ricerca a più ampio spettro.

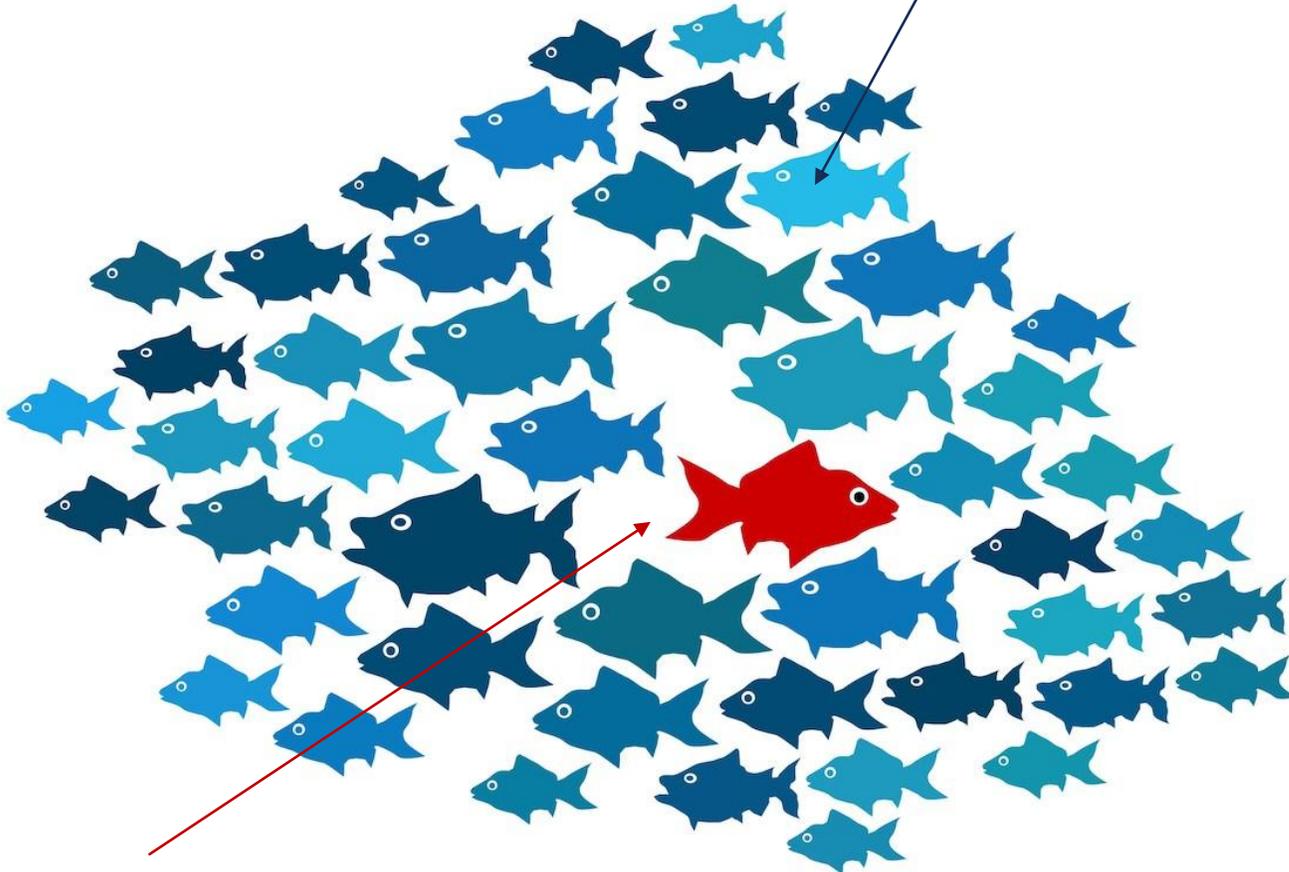


**Machine Learning non supervisionato**

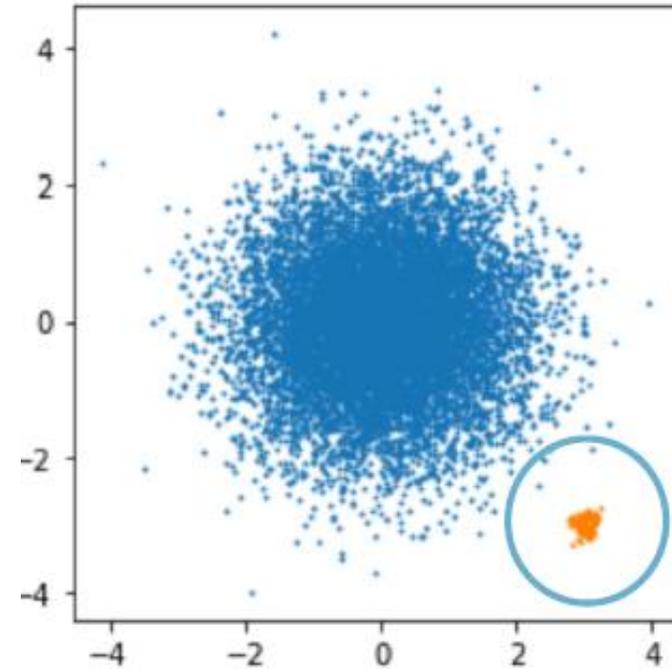
# ANOMALY DETECTION (AD)

La generalità della predizione sarà dipendente dall'inverso del livello di supervisione dell'approccio

Non hanno un'etichetta !



**ANOMALIA !**



AD in fisica delle particelle: identificazione, nei dati raccolti, di eventi con features inconsistenti con gli eventi di fondo

Prima di costruire i grafi viene eseguita una trasformazione ai jet:

---

## Algorithm 1: Jet Alignment

---

### Start

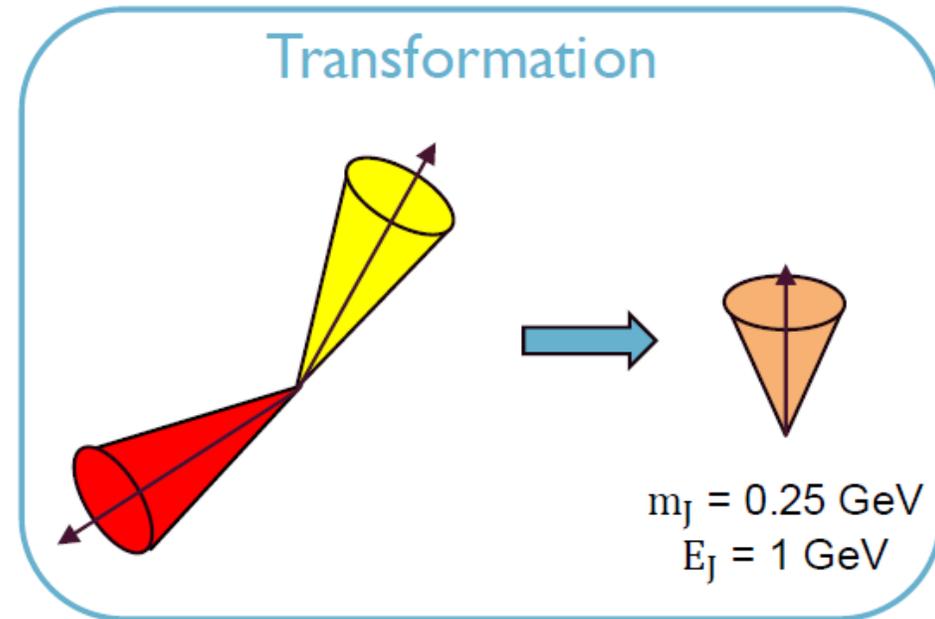
Boost jet in  $z$  direction until  $\eta_{Jet} = 0$

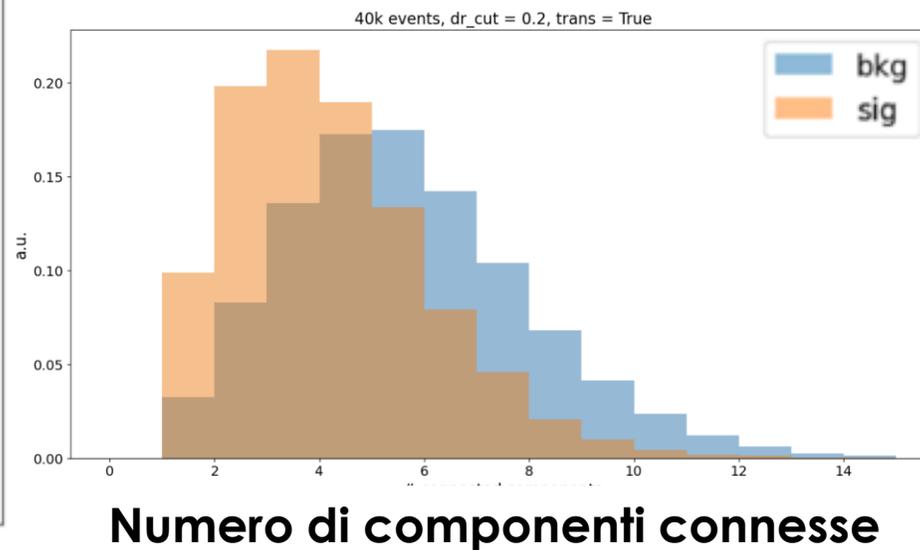
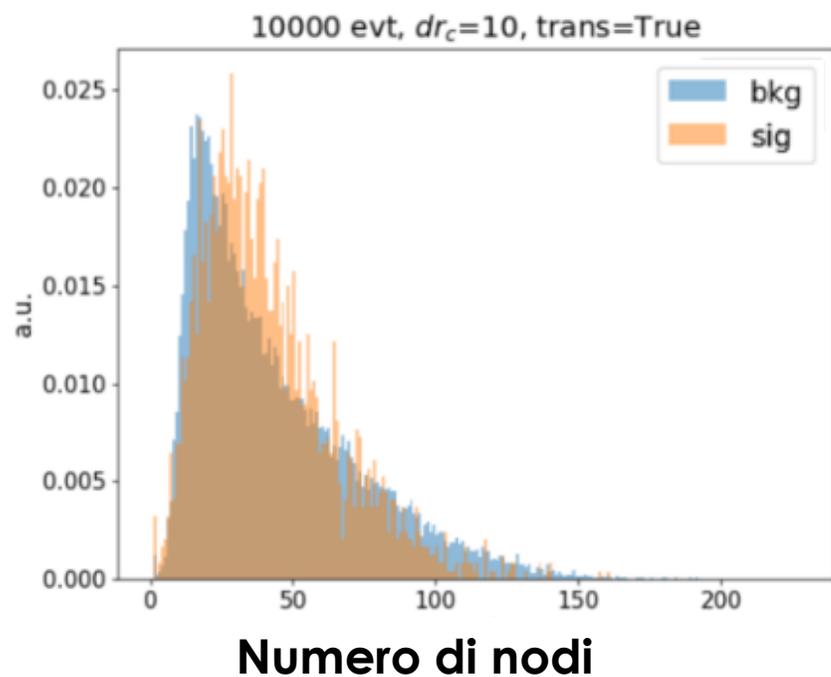
Rotate jet about  $z$  axis until  $\phi_{Jet} = 0$

Rescale jet four-vector such that  $m_{Jet} = 0.25 \text{ GeV}$

Boost jet along its axis until  $E_{Jet} = 1 \text{ GeV}$

Rotate jet about  $x$  axis until hardest constituent has  $\eta_1 = 0, \phi_1 > 0$





R&D sul [LHC\\_Olympics2020](#) dataset: è un dataset pubblico contenente simulazioni fondo + segnale utilizzato per studi di ML

- Eventi QCD dijet di fondo;
- $Z' \rightarrow XY \rightarrow q\bar{q}q\bar{q}$  segnale, con  $m_{Z'} = 3.5 \text{ TeV}$ ,  $m_X = 500 \text{ GeV}$  e  $m_Y = 100 \text{ GeV}$ ;

## Architettura EGAT:

$$h_i^{(l+1)} = \sigma \left( \sum_{j \in N(i)} \alpha_{ij}^{(l)} W^{(l)} h_j^{(l)} + \overbrace{B^{(l)} h_i^{(l)}}^{\text{Termine di bias}} \right), \forall l \in \{0, \dots, L-1\}$$

$$\alpha_{ij}^{(l)} = \text{softmax}_i \left( e_{ij}^{(l)} \right)$$

Matrice dei pesi  
Funzione di attivazione

$$e_{ij}^{(l)} = \vec{F}(f'_{ij}) \quad f'_{ij} = \text{LeakyReLU}(A[h_i \| f_{ij} \| h_j])$$

Attention scores normalizzati  
Edge features

## Architettura GIN:

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

MLP

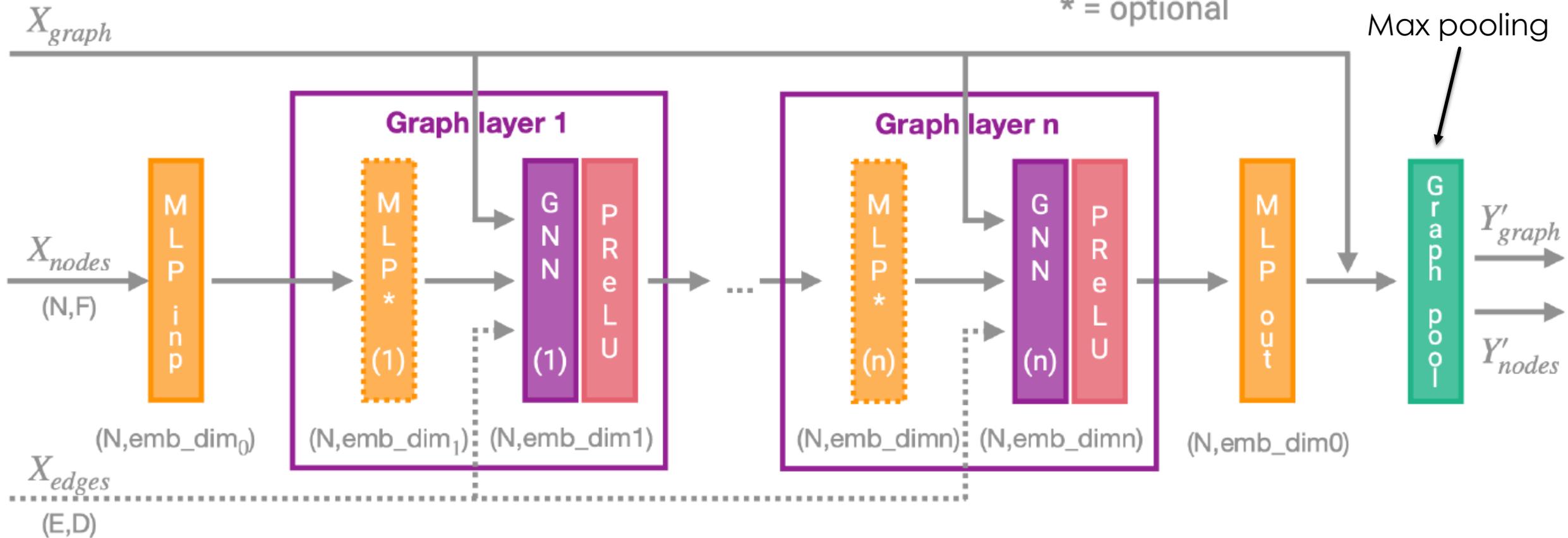
Somma, media, massimo, etc.

Iperparametro

$h_i^{(l)}$  = embedding delle features del nodo  $i$  al layer  $l$ -esimo

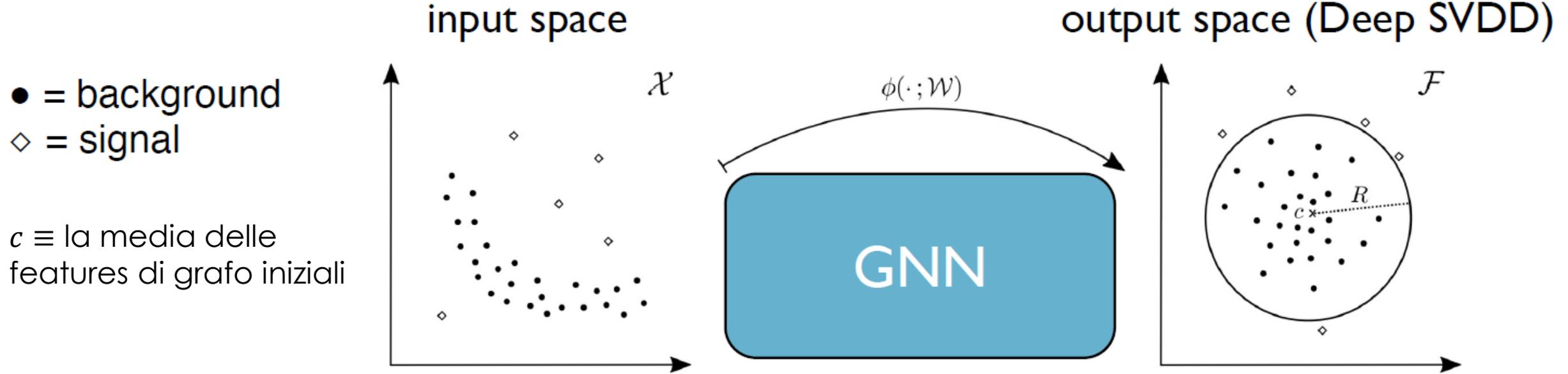
# ARCHITETTURA DELLA RETE

N = number of nodes  
F = number of node features (3)  
E = number of edges  
D = number of edge features (1)  
\* = optional



# STRATEGIA ANOMALY DETECTION

La rete si allena seguendo l'obiettivo del Deep Support Vector Data Description:



La distanza dal centro della sfera così costruita è la nostra Anomaly Score (AS) =  $\|\phi(G_i; W^*) - c\|^2$

Ottimizzazione:

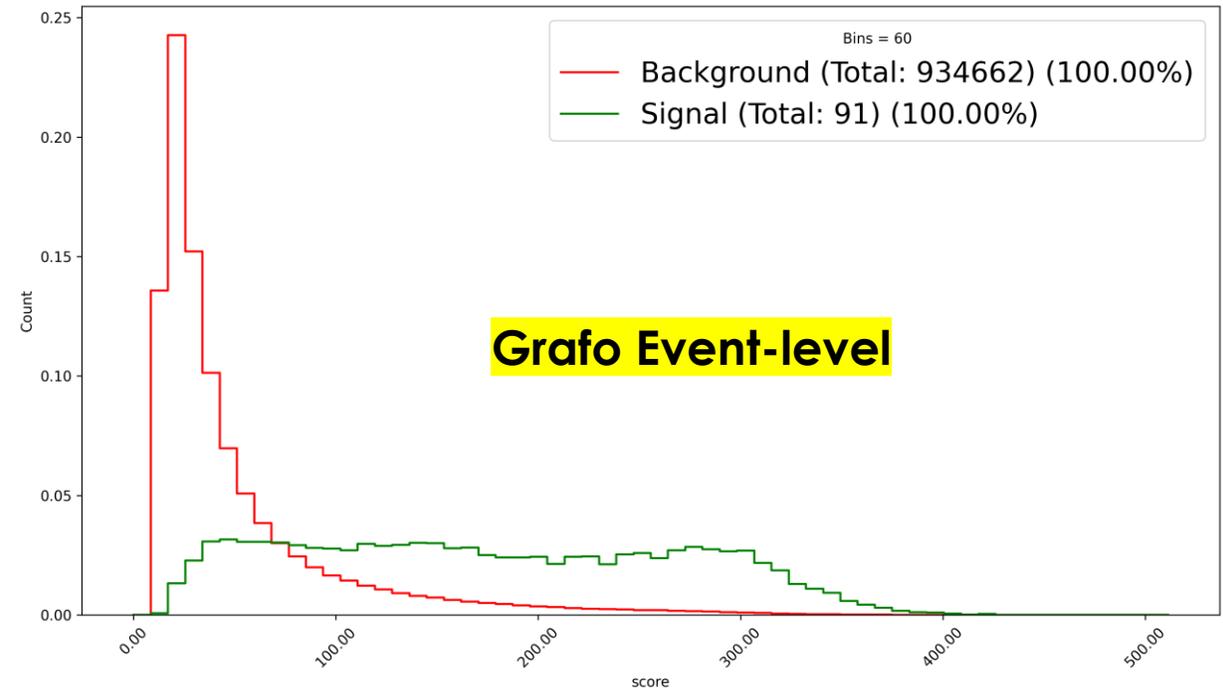
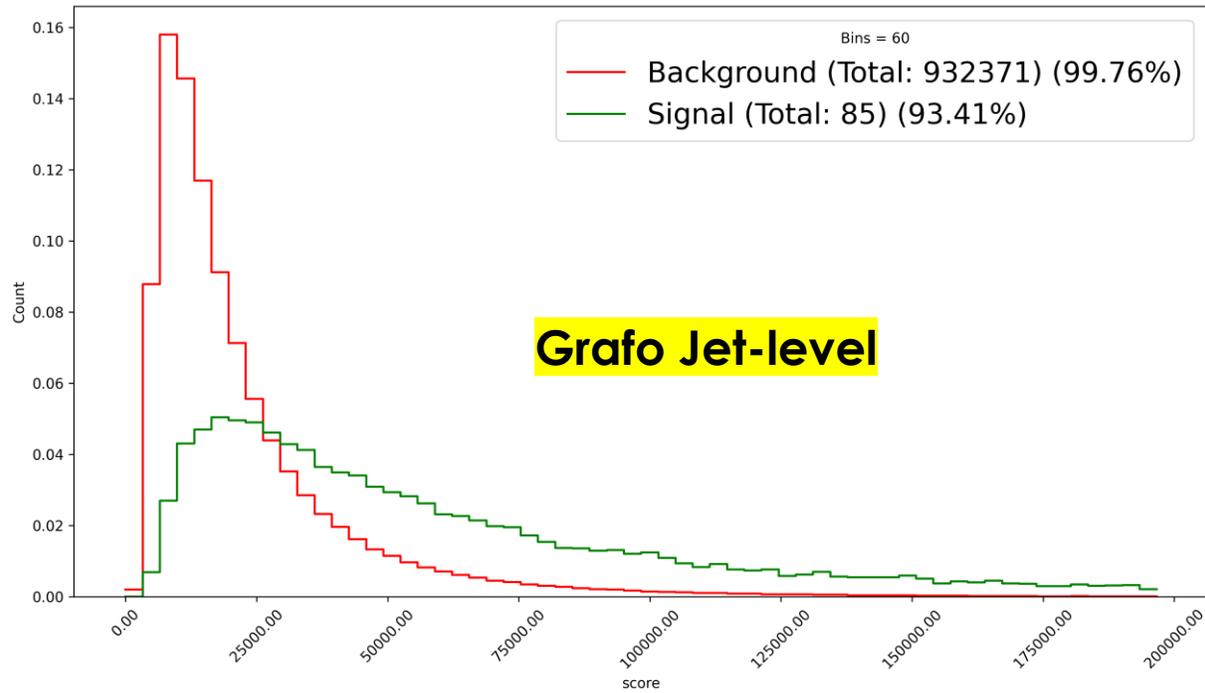
Per GIN:

$$Loss = \frac{1}{N} \sum_{i=1}^N \|\phi(G_i; W) - c\|^2 + \frac{\lambda}{2} \sum_{l=1}^L \|W^l\|_F^2$$

Per EGAT:

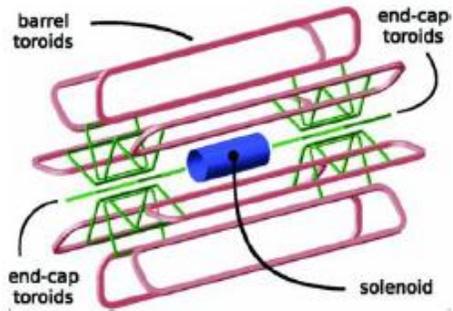
$$Loss = \frac{1}{N} \sum_{i=1}^N \|\phi(G_i; W) - c\|^2$$

# RISULTATI PRELIMINARI



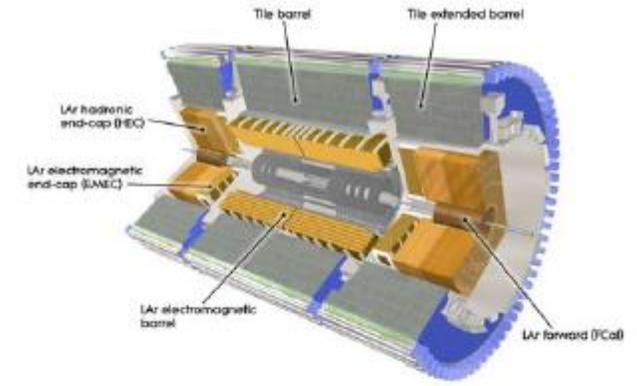
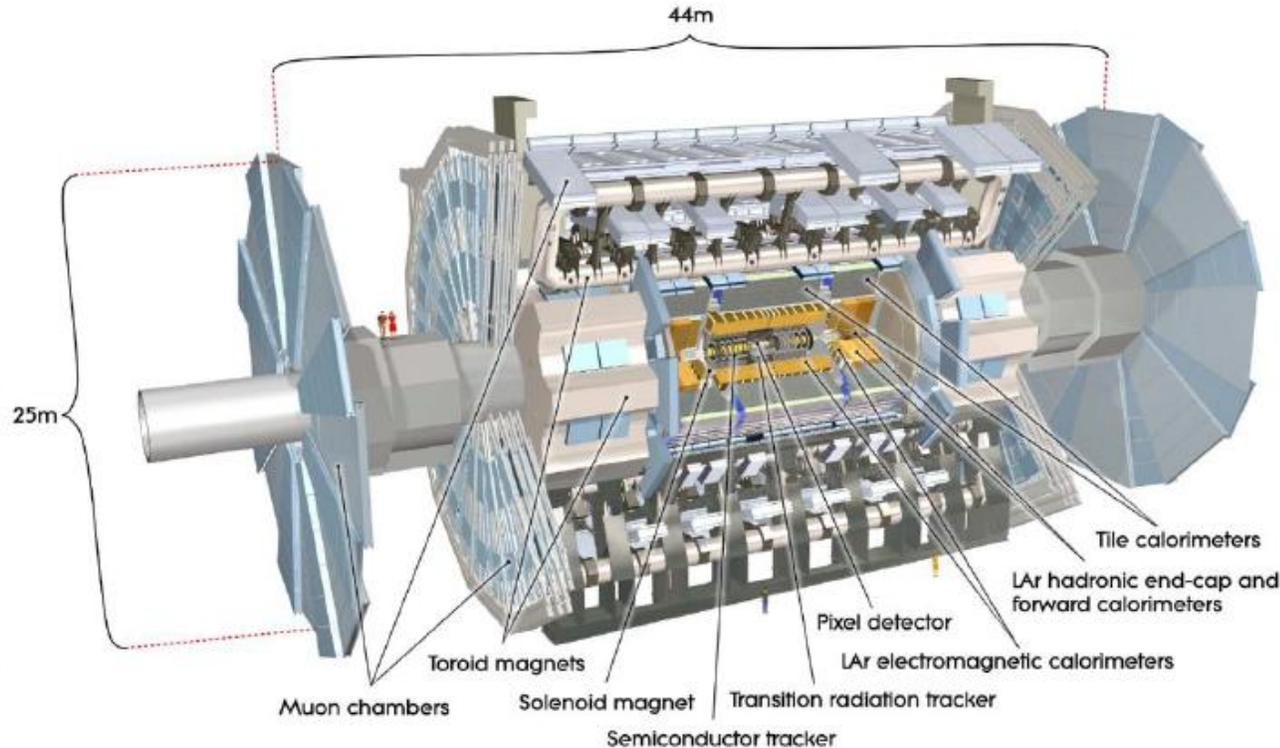
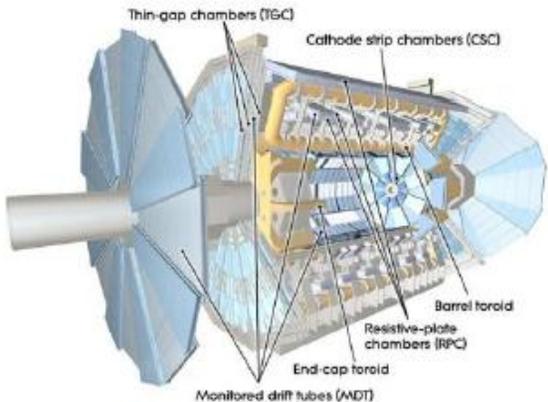
# ATLAS: A Toroidal LHC ApparatuS

L'invariante relativistico  $\eta = -\ln \left[ \text{tg} \left( \frac{\theta}{2} \right) \right]$   
 «pseudo-rapidità» è utilizzato al posto della  $\theta$



## Superconductive Magnetic System

## Muon Spectrometer



## Calorimeter System

## Inner Detector

