ITeDA

OBSERVATORIO PIERRE AUGER

KIT Karlsruher Institut für Technologie

Graph Neural Networks for photon search at tens of PeV

E. Rodriguez, N. González, M. Roth, B. Wundheiler

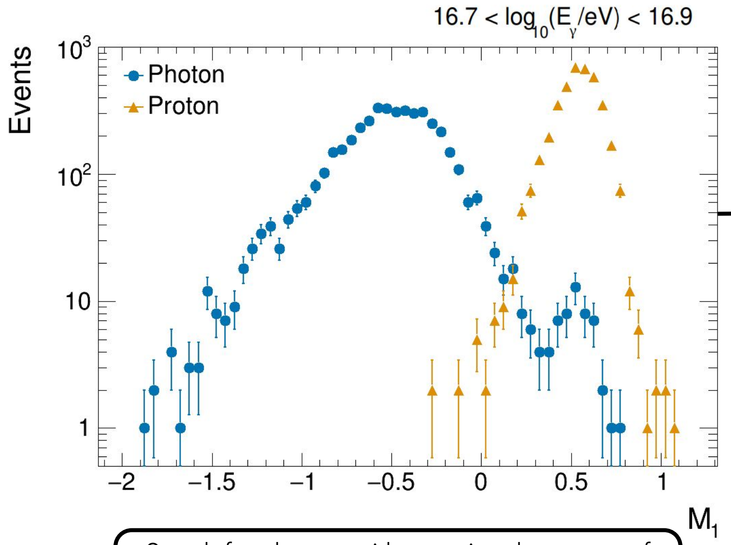
September 5, 2024

Auger Youngsters Meeting

Motivation

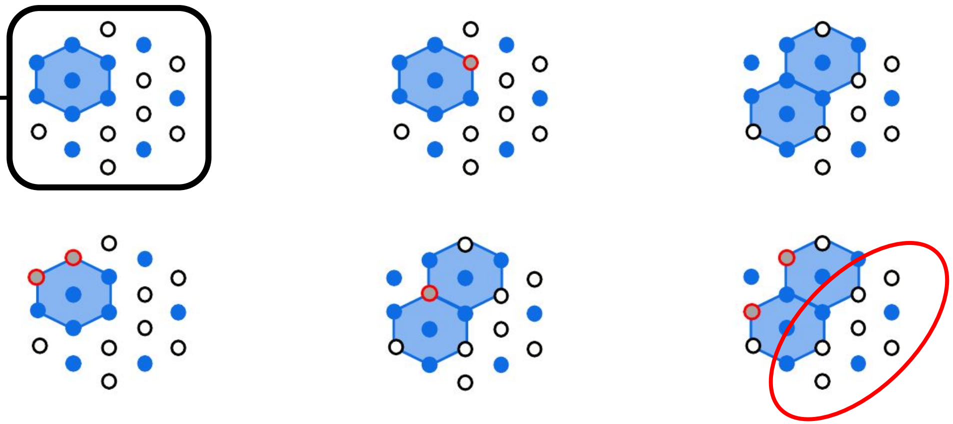
Our aim is to complement current SD433-UMD photon searches using with ML methods

- To enhance discrimination
- Relax data selection cuts → increase exposure



Search for photons with energies above tens of PeV at the Pierre Auger Observatory
Under Collaboration review.

SD433+UMD active station/counter configurations



Can we use these???

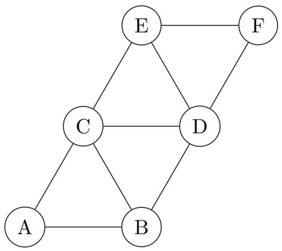
Why graphs neural networks?

Graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$

- Representation or embedding $\mathscr{F}$ — Feature Vector/Matrix $X \in \mathbb{R}^{N \times F}$
- Structure
  - Nodes $\mathscr{V}$ — Entities
  - Edges $\mathscr{E}$ — Relationships
  - $A \in \mathbb{R}^{N \times N}$

- Natural representation of the array
- Easy to include to off-grid stations
- Flexible representation of events

<u>Toy example</u>



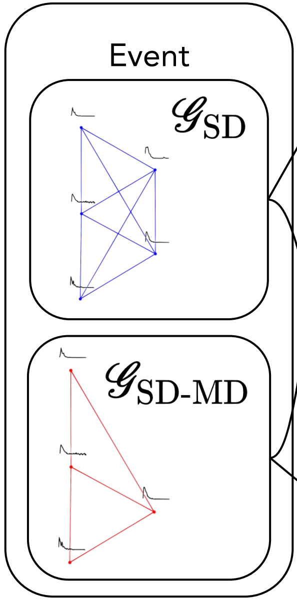
$\mathbf{A} =$

	A	B	C	D	E	F
A	0	1	1	0	0	0
B	1	0	1	1	0	0
C	1	1	0	1	1	0
D	0	1	1	0	1	1
E	0	0	1	1	0	1
F	0	0	0	1	1	0

$\mathbf{X} =$

Node	x	y	$\lg(S)$
A	-216.5	-374.98	-0.05
B	216.5	-374.98	-0.08
C	0	0	2.2
D	216.5	0	0.77
E	433	374.98	0.85
F	807.98	374.98	0.29

Hybrid SD-MD event representation with graphs



$$X_{\text{SD}} = \begin{array}{c|cccc} \text{Node} & x & y & \Delta t_{\text{hottest}} & n_{\text{PMTs}} \\ \hline \text{SD}_{\text{hottest}} & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{array}$$

$$X_{\text{traces}} = \begin{array}{c|cccc} \text{Node} & t_1 & t_2 & \cdots & t_{n_{\text{t-bins}}} \\ \hline \text{SD}_{\text{hottest}} & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{array}$$

$$X_{\text{SD-MD}} = \begin{array}{c|cccccc} \text{Node} & x & y & \Delta t_{\text{hottest}} & n_{\text{PMTs}} & \rho_\mu & A_{\text{eff}} \\ \hline \text{SD}_{\text{hottest}} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{array}$$

- Events are represented by 2 graphs
- Graphs are composed only by candidate stations/counters
- Connections up to 2nd neighbors

Flexible for all SD-MD configurations :)

Redundancy in the input :(

In this example:

X_{SD} : 5 rows, 4 cols

$X_{\text{SD-MD}}$: 4 rows, 6 cols

X_{traces} : 5 rows, $n_{\text{t-bins}}$ cols for $\mathscr{G}_{\text{SD}}$

X_{traces} : 4 rows, $n_{\text{t-bins}}$ cols for $\mathscr{G}_{\text{SD-MD}}$

$$\mathscr{V}_{\text{SD-MD}} \subseteq \mathscr{V}_{\text{SD}}$$
$$\mathscr{E}_{\text{SD-MD}} \subseteq \mathscr{E}_{\text{SD}}$$
$$\mathscr{F}_{\text{SD}} \subseteq \mathscr{F}_{\text{SD-MD}}$$

Dataset and augmentation

Models: EPOS-LHC; FLUKA INFN

Offline: v4r0p2-pre3

CORSIKA: 7.7xx

Energy range: lg(E/eV) ∈ (16.5-17.5)

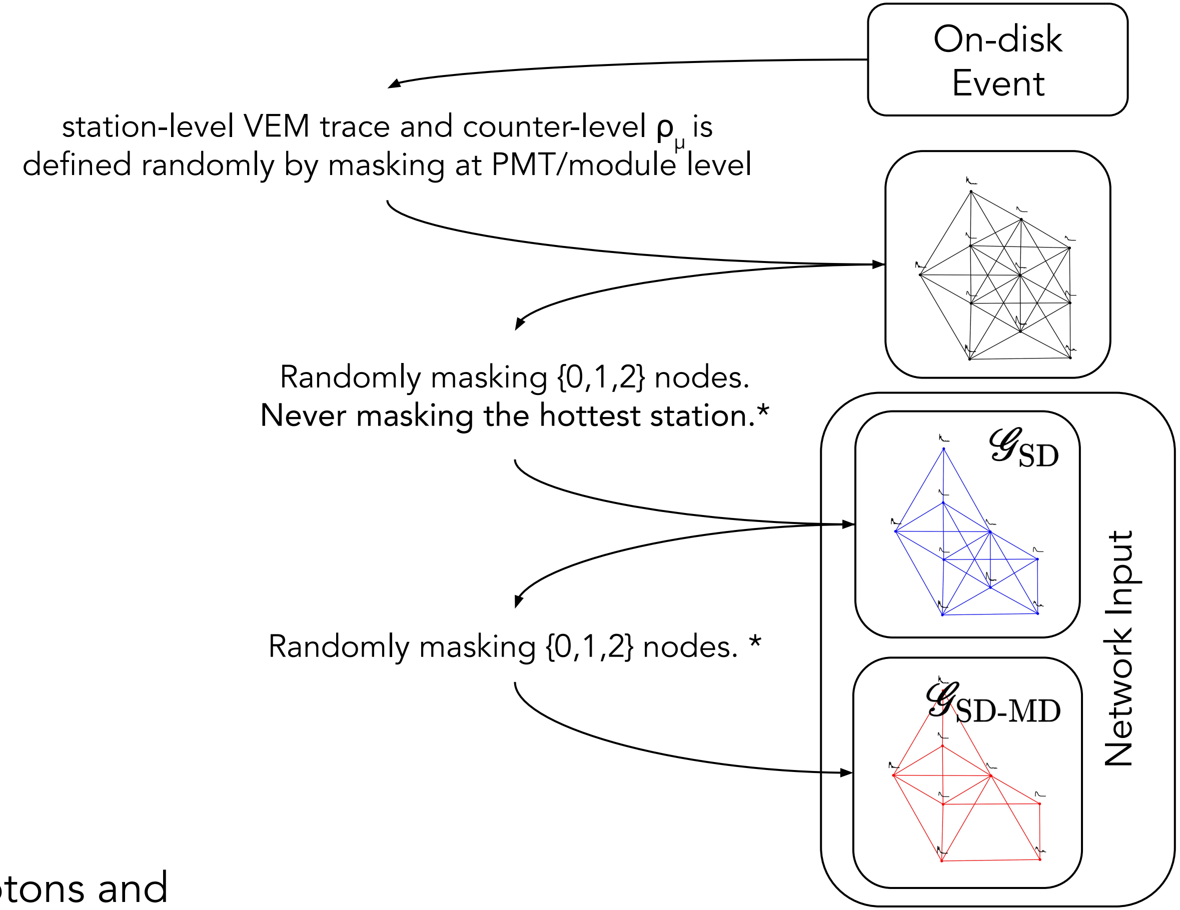
After quality cuts
105k protons / 103k photons

Train: 118k (57%)

Val: 40k (18%)

Test: 50k (25%)

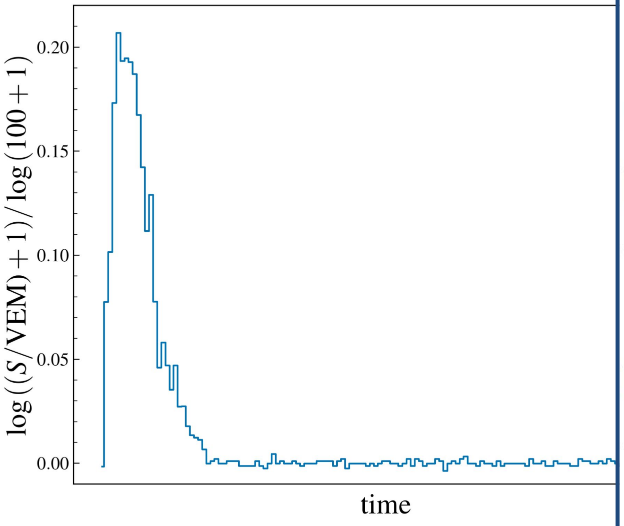
All events were both reconstructed as protons and photons (N. González & I. Maris, [GAP2021-056](GAP2021-056))

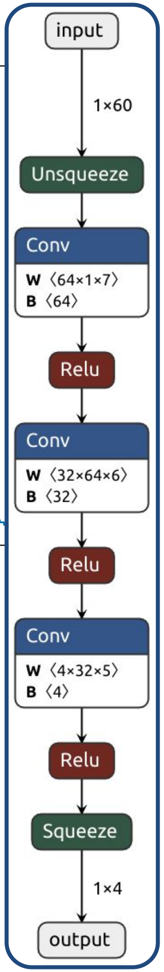


* No masking is done for graphs with 3 nodes.

Trace Analyzer

- First 60 bins of the averaged VEM trace
- Randomly masking 0, 1, or 2 PMTs
- Batch normalization helps A LOT
- Only 4 features as output → Thanks Fiona!
- One Trace Analyzer for the whole network





For each SD station, the average trace is processed by the Trace Analyzer and appended as node features dynamically.

Graph Convolutional Networks (GCN)

new representation

linear combination of features

$$X^{k+1} = \sigma(\underbrace{\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}}\overbrace{X^{k}W^{k}})$$

activation function

$$\tilde{A} = A + \mathbf{I}$$

$$\tilde{D}_{ii} = \sum_{j} \tilde{A}_{ij}$$

learnable parameters

weighted based on neighborhood and node degree

Conceptually: fancy linear combination

In practice: Forward Pass is just matrix multiplication

Geometry Analyzer using Graph ATtention (GATs)

- Using 3 Graph Attention layers
- 3 attention heads per layer
- 3 Geometry Analyzers
- Batch normalization here too

P. Veličković et al. Graph attention networks. *arXiv preprint* arXiv:1710.10903, 2017.

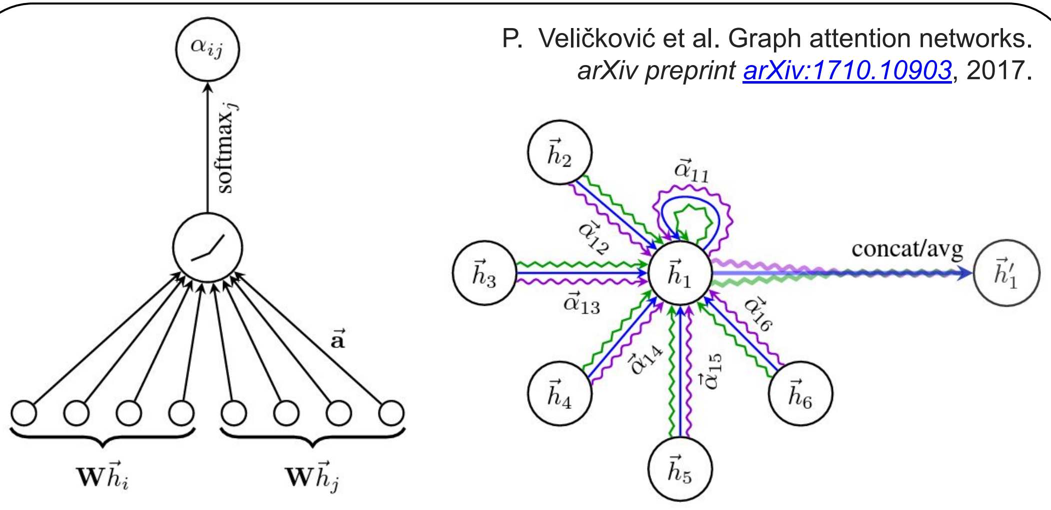
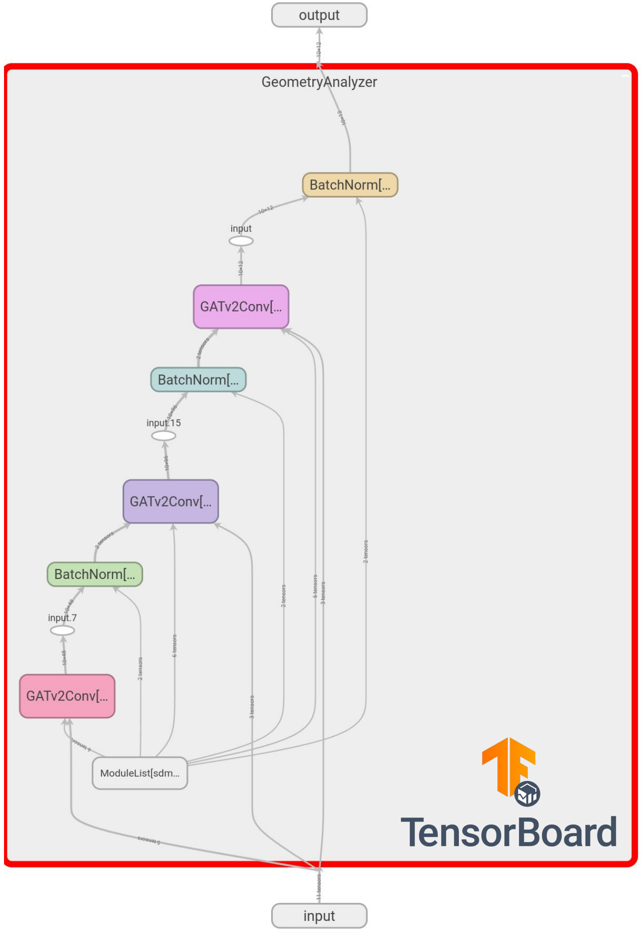
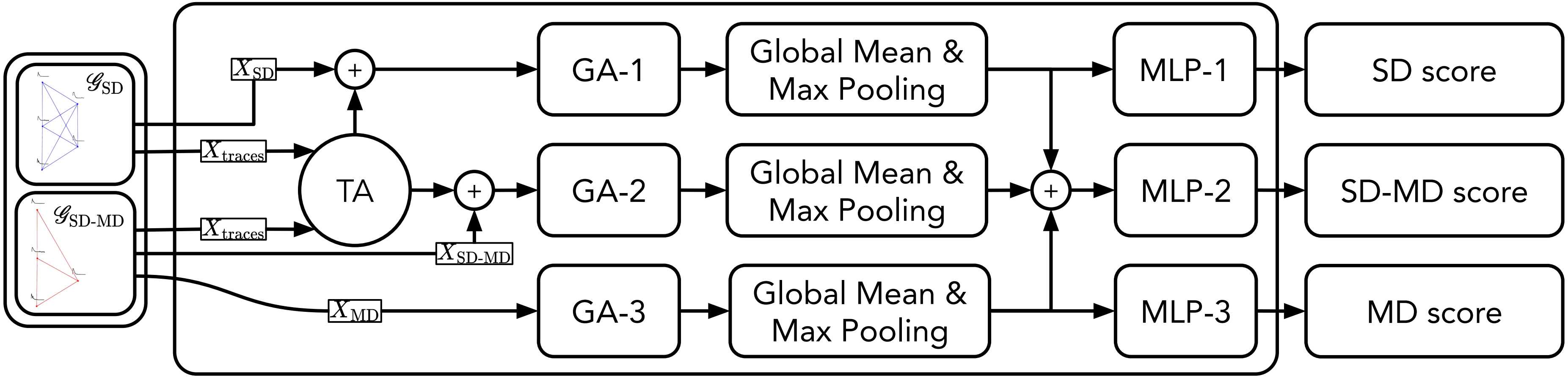


Figure 1: **Left:** The attention mechanism $a(\mathbf{W}\vec{h}_i, \mathbf{W}\vec{h}_j)$ employed by our model, parametrized by a weight vector $\vec{\mathbf{a}} \in \mathbb{R}^{2F'}$, applying a LeakyReLU activation. **Right:** An illustration of multi-head attention (with $K = 3$ heads) by node 1 on its neighborhood. Different arrow styles and colors denote independent attention computations. The aggregated features from each head are concatenated or averaged to obtain $\vec{h}'_1$.



The network



Selected inside the network

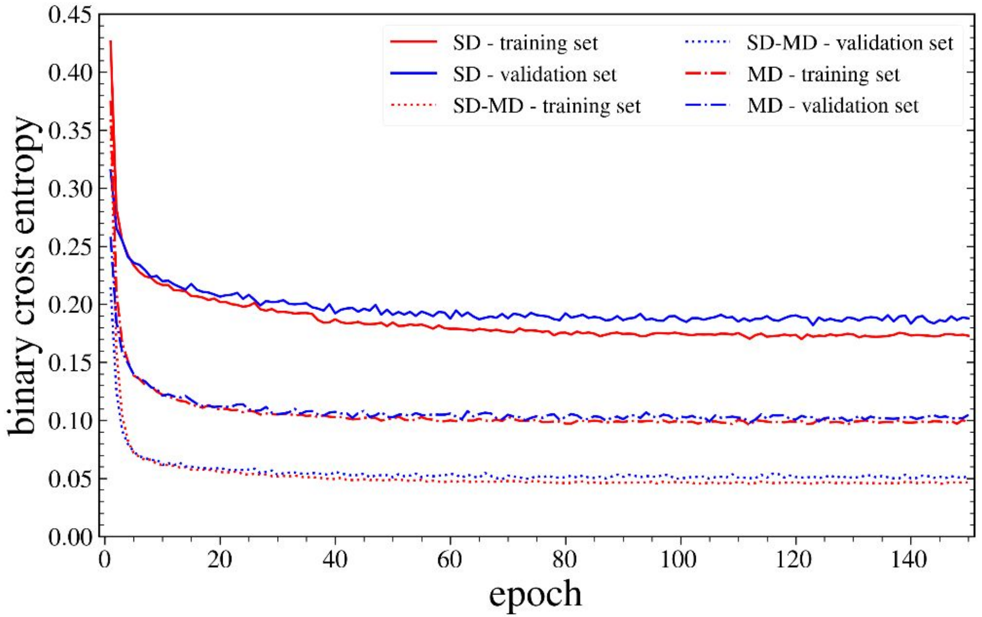
$X_{\mathrm{MD}} =$

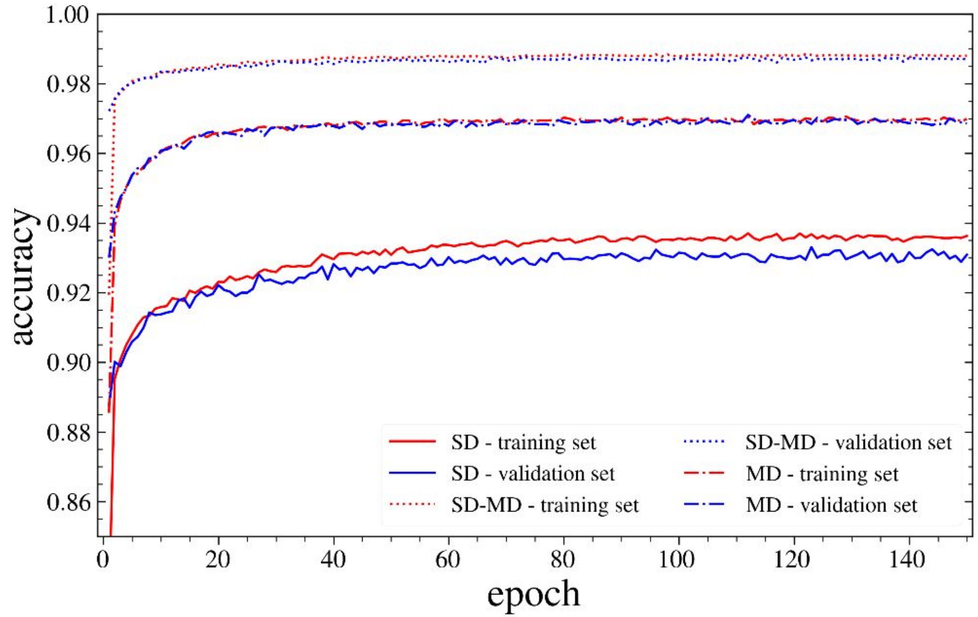
Node	x	y	$\Delta t_{\mathrm{hottest}}$	ρ_μ	A_{eff}
$\mathrm{SD}_{\mathrm{hottest}}$	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...

MLP $(N_{\mathrm{inputs}}, 16) \rightarrow (16, 8) \rightarrow (8, 1)$

Training

- Binary cross entropy as loss (global loss is the sum of the 3 losses)
- Adam optimizer
- Learning rate decreased when the global loss plateaus

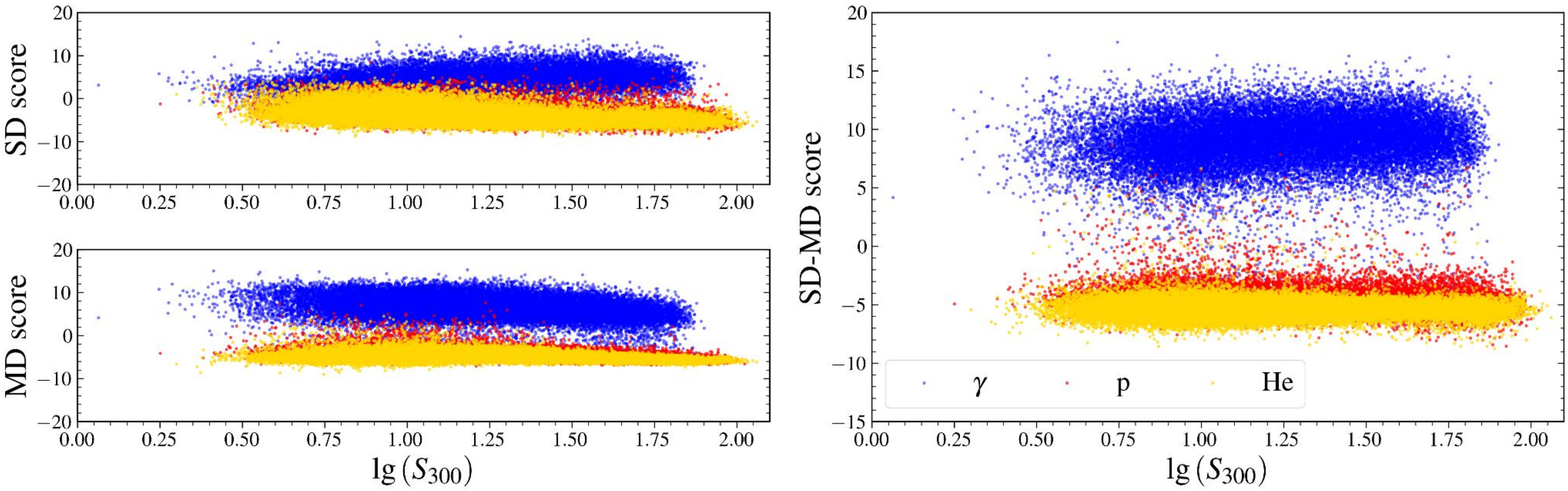




Performance

- No masked stations/counters
- No masking for PMTs/Modules
- Quality cuts: successful reconstruction and zenith angle below 45 deg

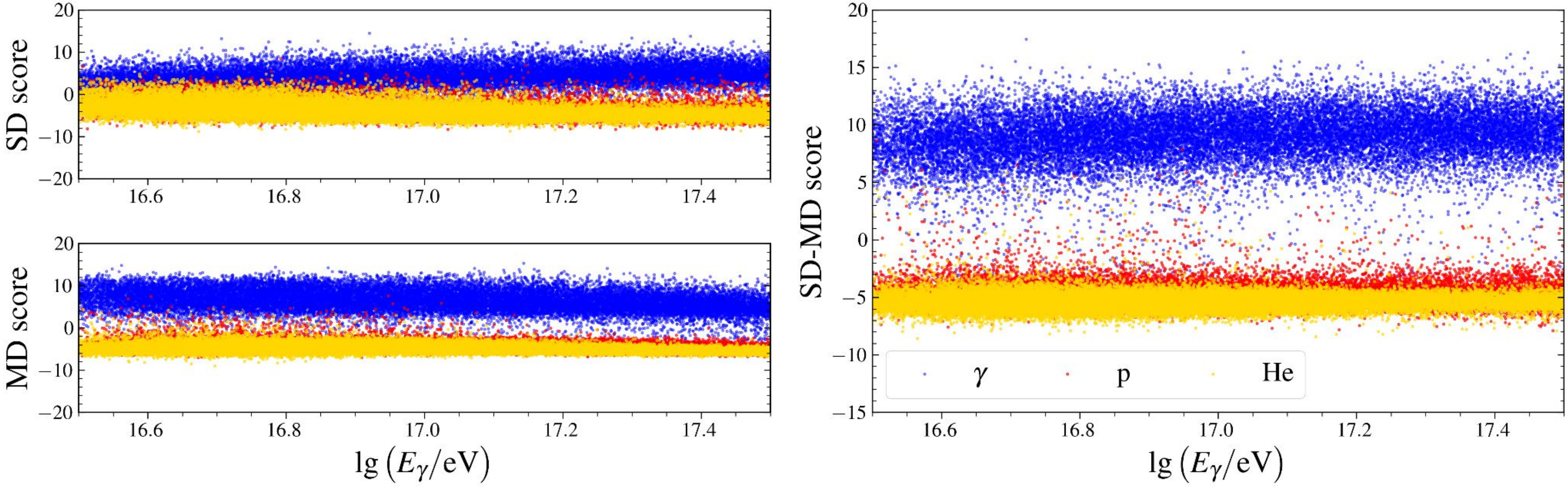
The network has never seen the shower size and wasn't trained on helium.



Performance

- No masked stations/counters
- No masking for PMTs/Modules
- Quality cuts: successful reconstruction and zenith angle below 45 deg

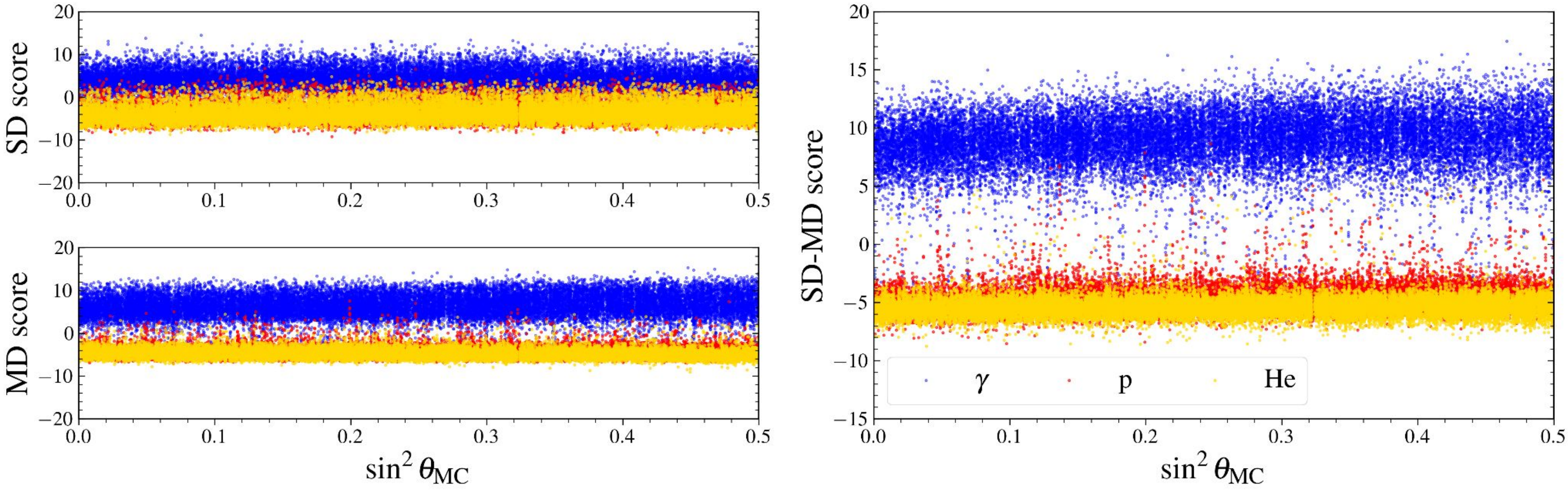
The network has never seen the energy and wasn't trained on helium.



Performance

- No masked stations/counters
- No masking for PMTs/Modules
- Quality cuts: successful reconstruction and zenith angle below 45 deg

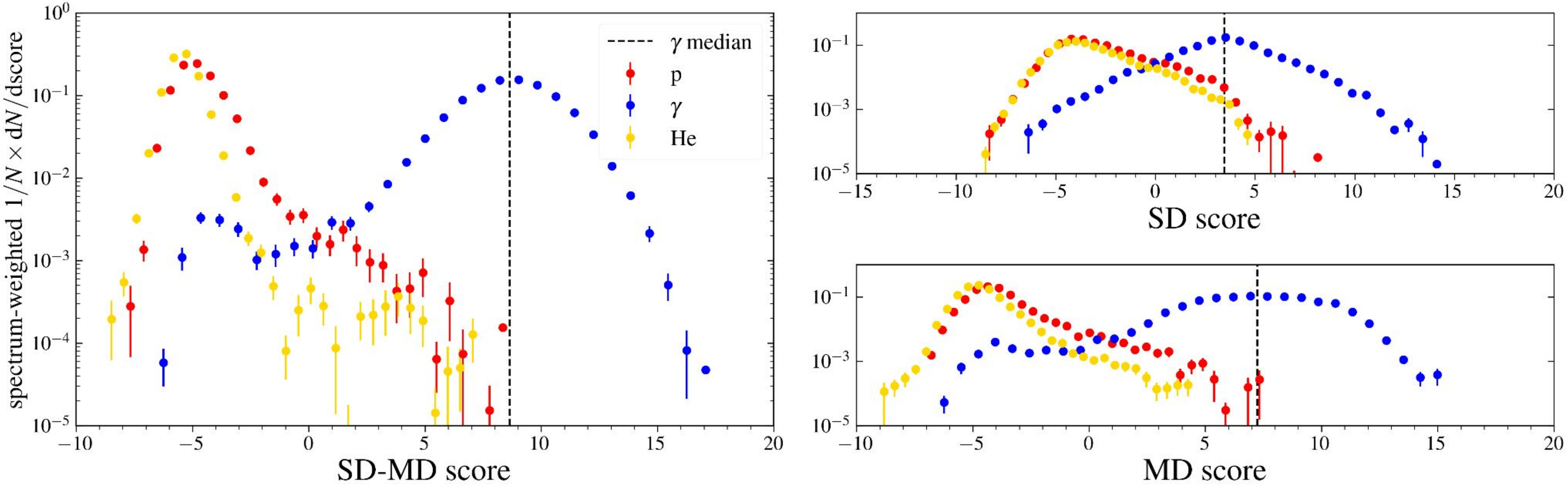
The network has never seen the zenith angle and wasn't trained on helium.



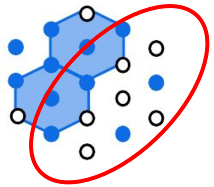
Performance

- No masked stations/counters
- No masking for PMTs/Modules
- Quality cuts: successful reconstruction and zenith angle below 45 deg

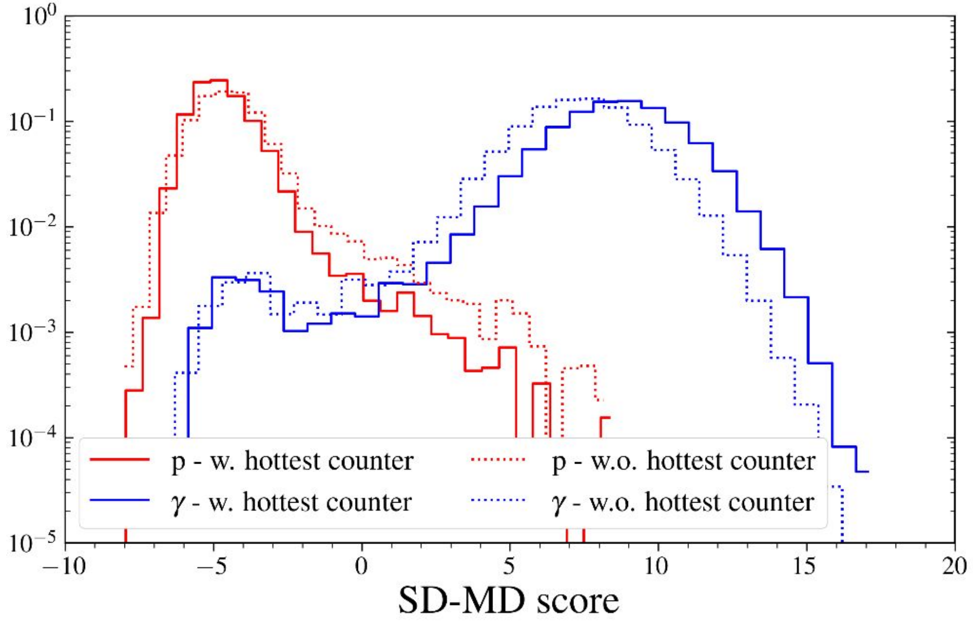
Photons are weighted by $\propto E^{-2}$ and hadrons $\propto E^{-3}$.

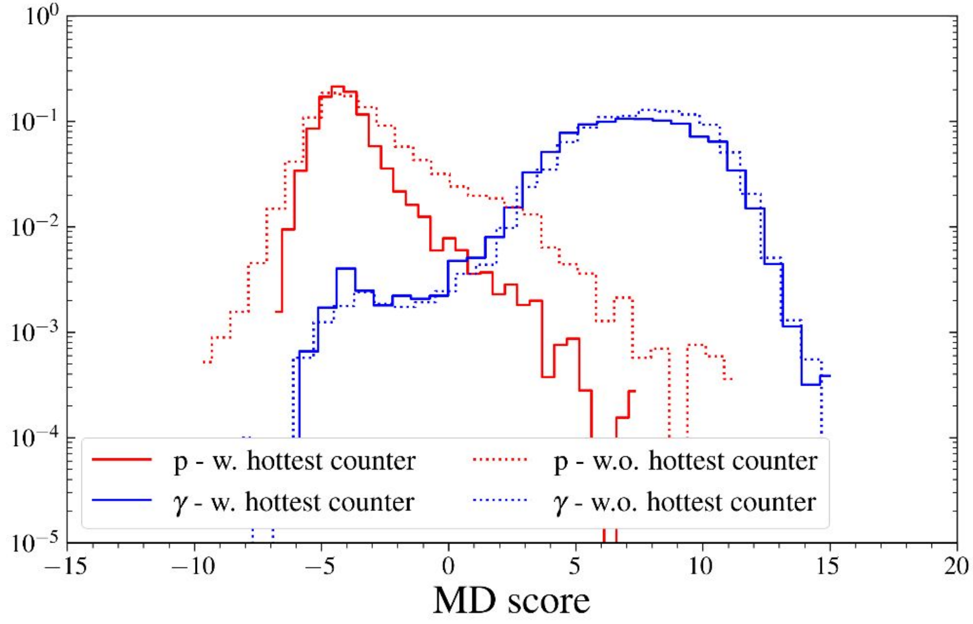


Impact of missing the counter from the hottest station



- Removing the hottest counter reduces the discrimination potential
- The SD-MD score still shows the capacity to discriminate → increase exposure





Outlook and summary

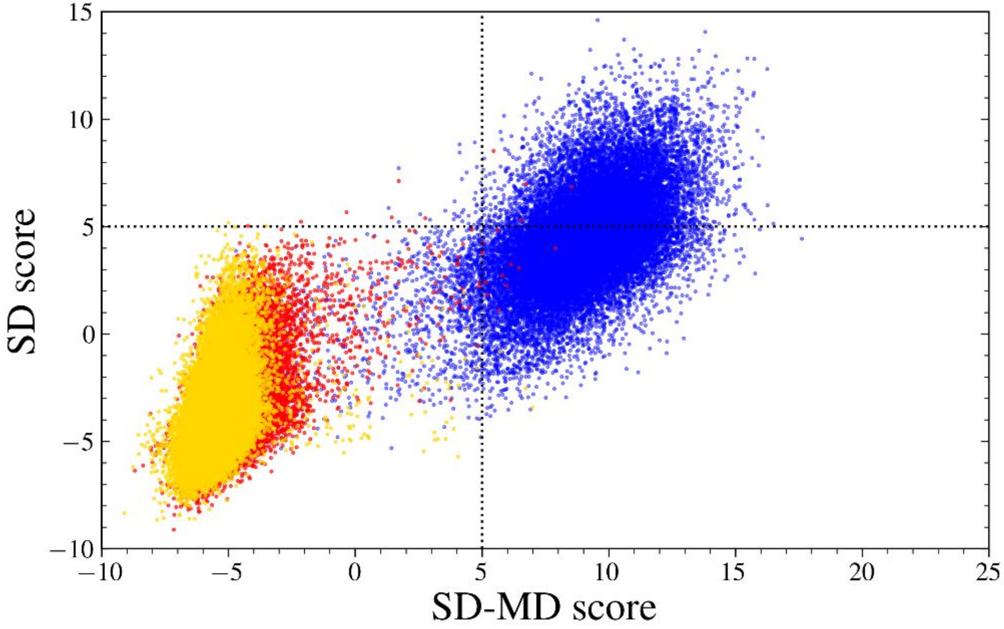
- Networks are able to separate photons from hadrons
    - Separation is independent on shower size and zenith angle
    - Cross-check with helium shows the network separates heavier primaries
- Networks should be re-trained based on more realistic input !!!
- Final graph representation and network architecture is still to be decided :(

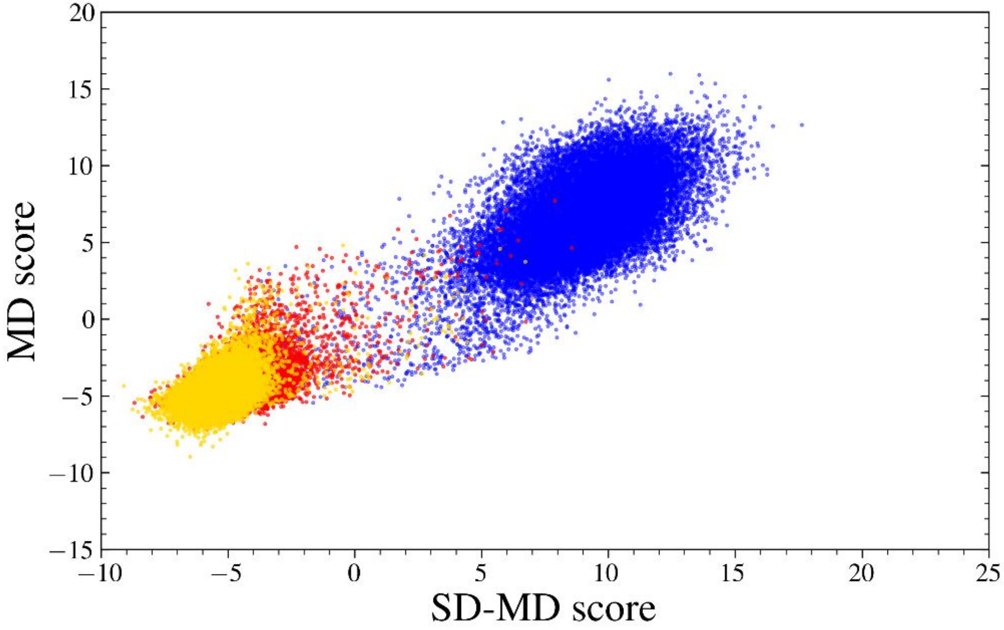
  BUT GNNs work :)
- Performance and systematics will be explored after re-training

Backup

Score correlation and candidate cut outlook

- Criteria for photon search could consider more than 1 score
- Including another observable could also be considered

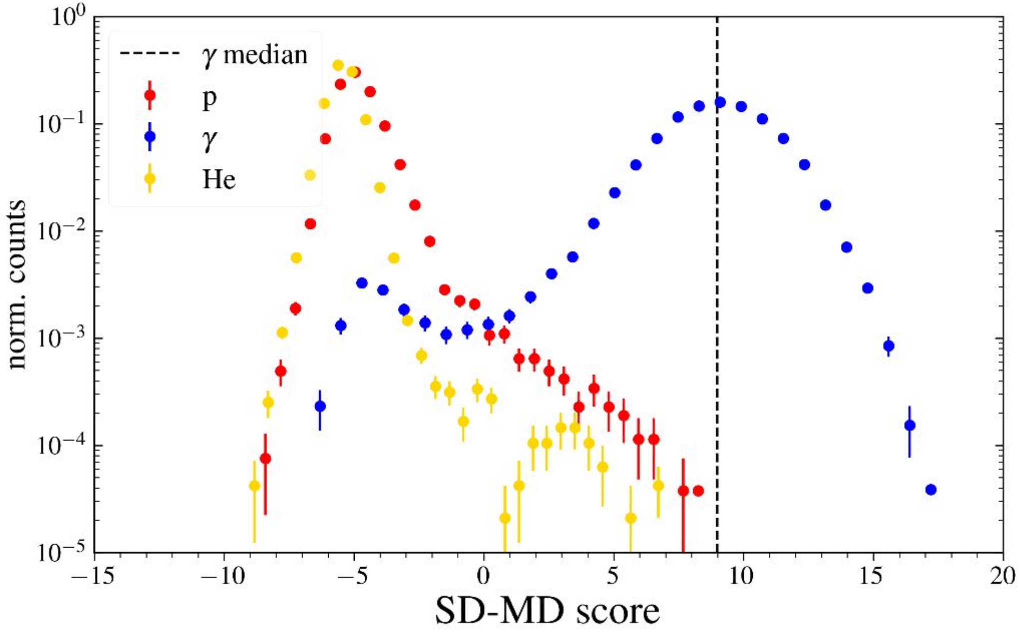


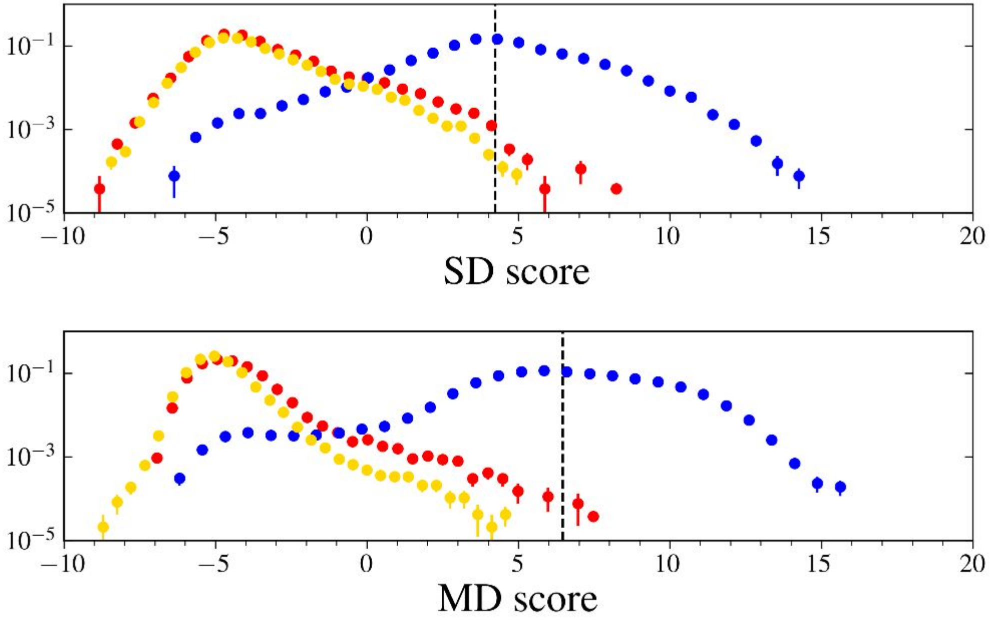


Tuning simulations

- Treat start time correctly with PMT masking
- From data:
  - Saturation threshold histogram - How to deal with saturation?
  - Black tanks/ bad PMTs distributions
  - Bad/Missing counters/modules distributions
- VEM traces - Peak or Charge? Rescaling?
- How to deal with ageing? Is this possible for saturated stations?

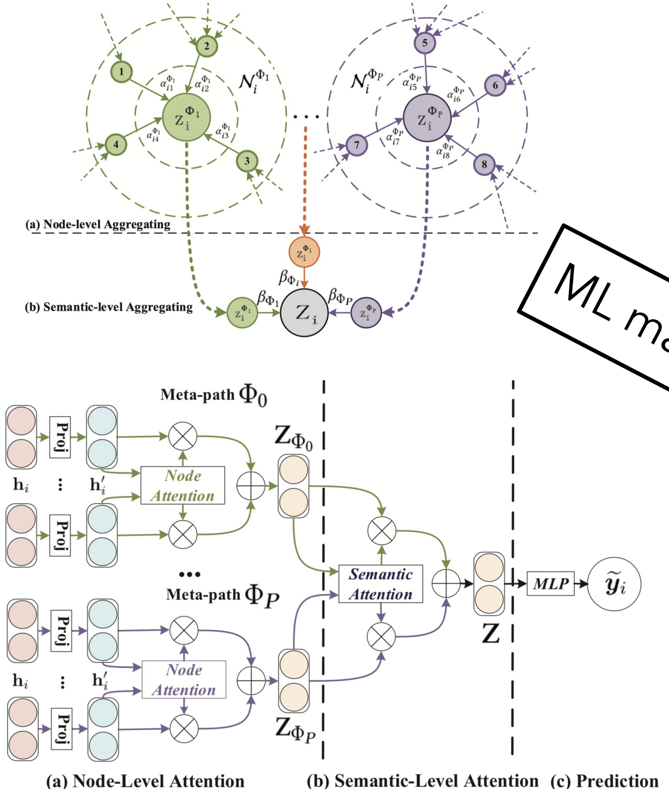
Performance





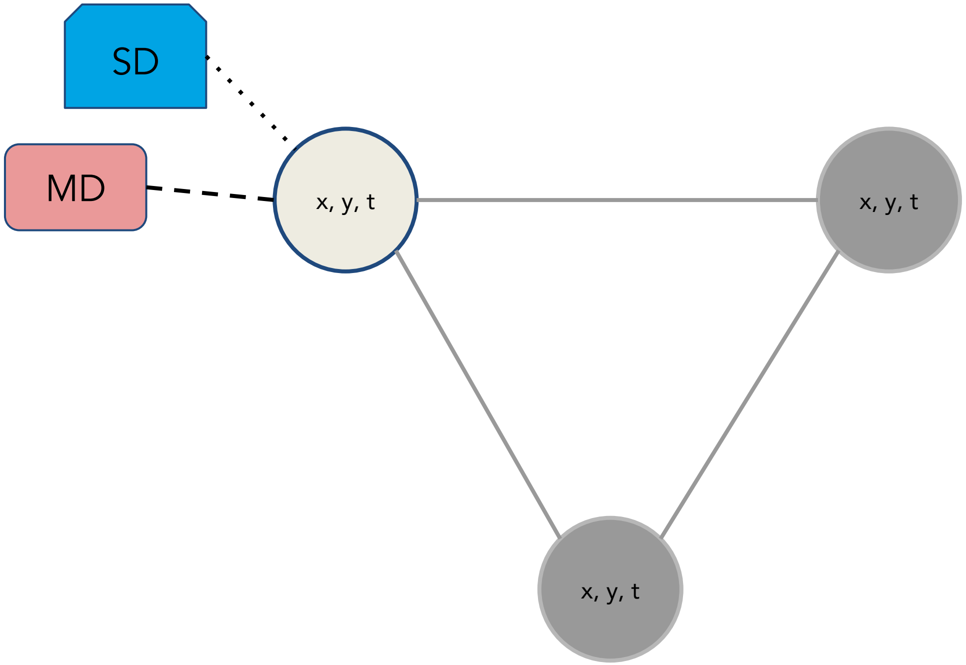
Going heterogeneous?

Wang, Xiao, et al. "Heterogeneous graph attention network." *The world wide web conference*. 2019.



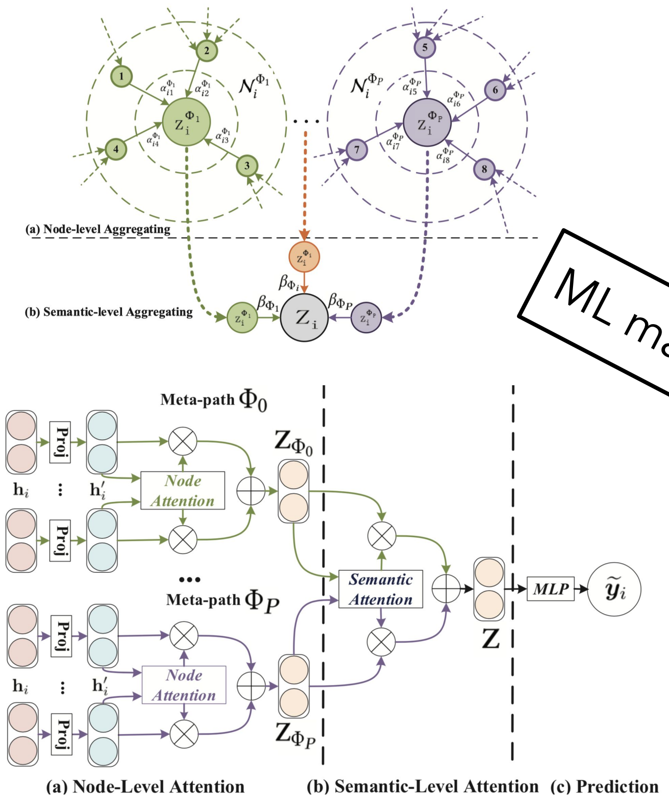
ML magic

Put simply, we could have heterogeneous graphs such that …

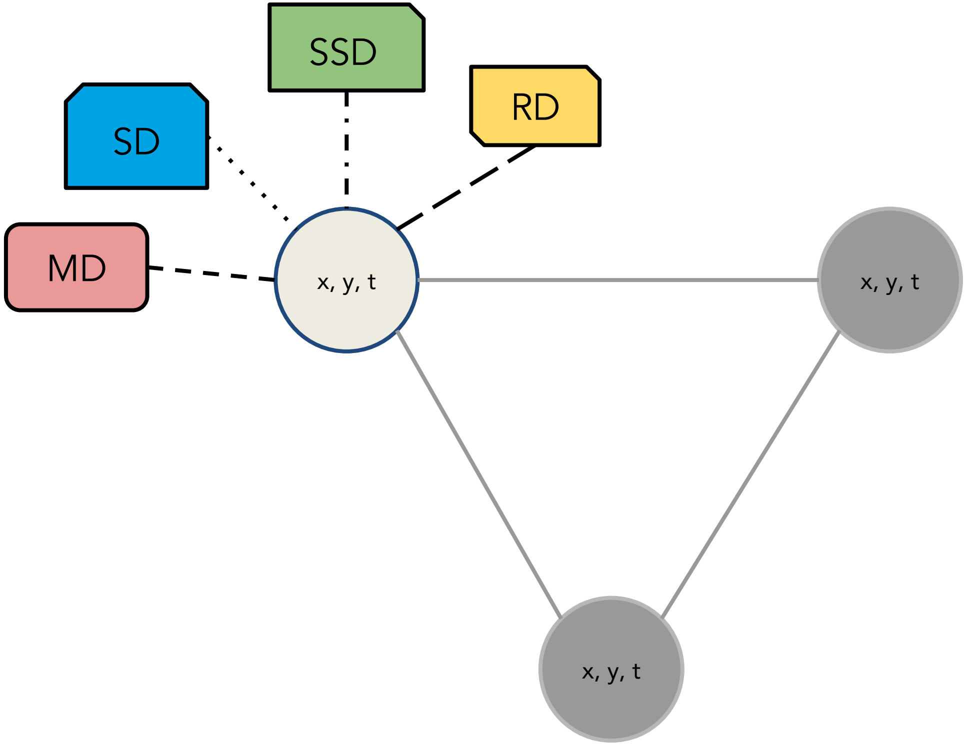


Going heterogeneous?

Wang, Xiao, et al. "Heterogeneous graph attention network." *The world wide web conference*. 2019.



ML magic



Should be straightforward to include information from other detectors for other analyses