Heterogeneous Graph Representation for Identifying Hadronically Decayed Tau Leptons at the High Luminosity LHC

Andris Huang[1], Xiangyang Ju[2], Jacob Lyons[1], Daniel Murnane[2], Mariel Pettee[2], Landon Reed[3]

[1]*UC Berkeley*, [2]*LBNL*, [3]*University of Minnesota Duluth*

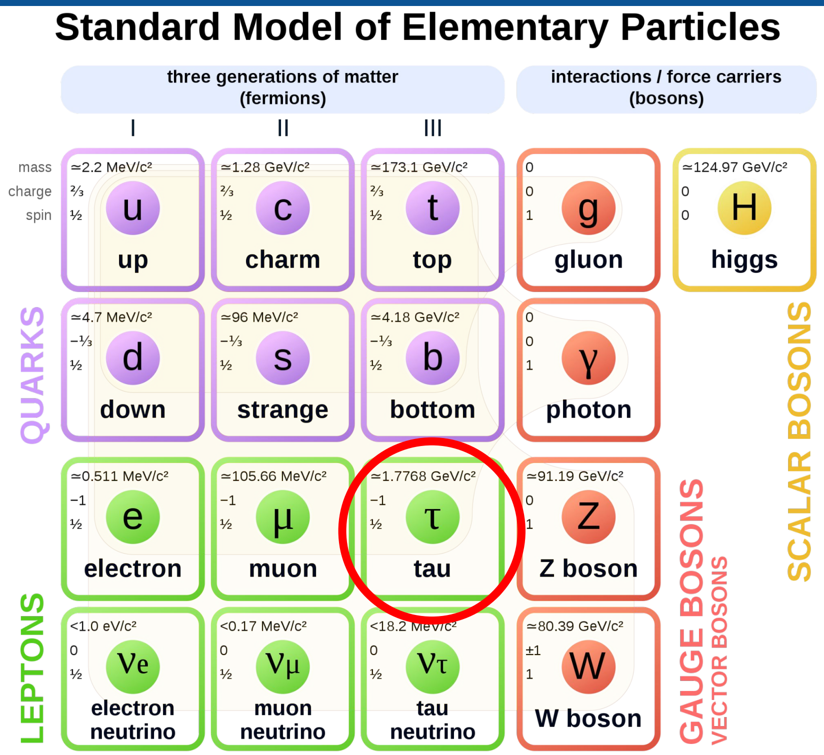
ML4Jets2022
November 3rd, 2022

Outline

- Background Information on Tau Leptons
- Data Simulation & Feature Selection
- Graph Neural Network (GNN) Architectures
- Heterogeneous Representations
- Results & Discussion

Background Information

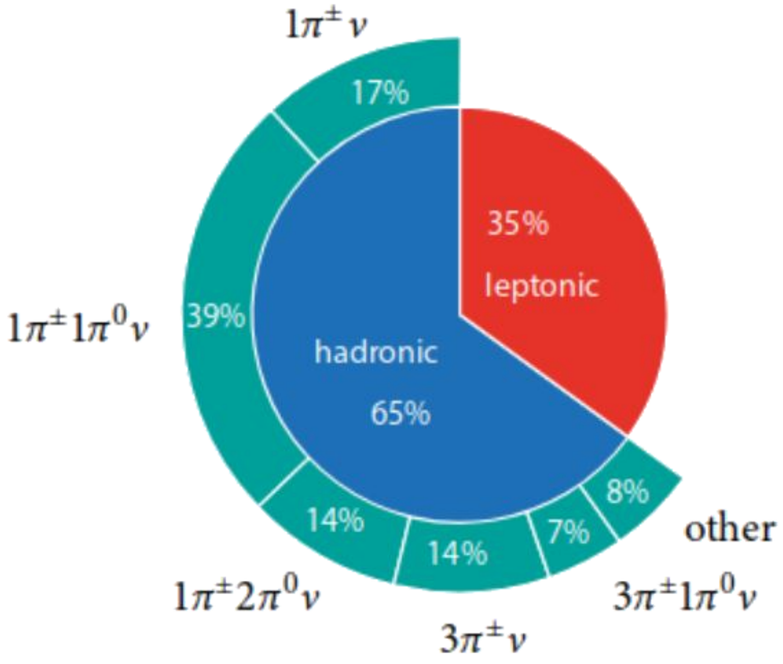
- Third generation fermion
- Mass of 1777 MeV → most massive lepton
- Decays with a mean lifetime of 290 ps and a mean flight distance of 49 μm/GeV
- Motivation: Used for making standard model measurements (H → ττ) & searching for new ditau resonances



https://en.wikipedia.org/wiki/Standard_Model

Decay Modes

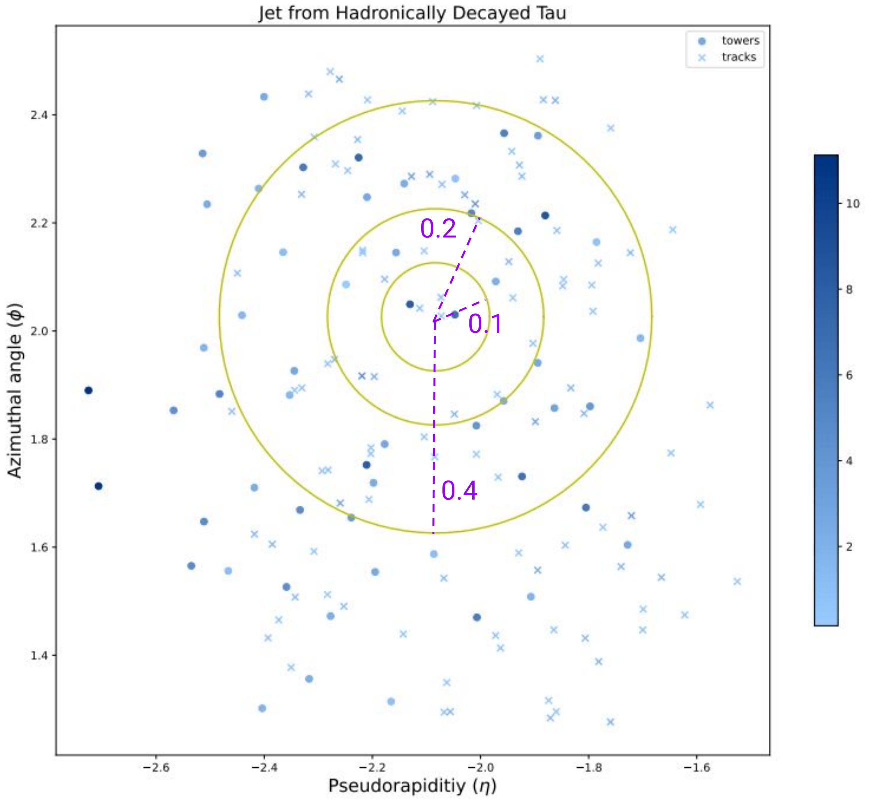
- Decays either leptonically or hadronically
- Our focus: the hadronic decay modes
    - Mostly include 1 or 3 charged pions
    - Visible signature appears as a jet with either 1 or 3 tracks
- Challenges
    - Neutrinos escape the detector, carrying away a fraction of energy
    - Dense environment at the HL-LHC



https://www.lhc-ilc.physik.uni-bonn.de/ergebnisse/dateien/t00000078.pdf?c=t&id=78

Data Simulation

- Proton-proton collisions
  - Center-of-mass energy: 13 TeV
  - HL-LHC: With an average 200 additional pp collisions
  - Jet reconstructed using the anti-k_t algorithm
- Signals
  - Hadronic Tau leptons from $\gamma^* \rightarrow \tau\tau$ processes
- Backgrounds
  - Jets from the QCD processes
- *Low-level* particle-flow kinematics
  - $p_T^{\text{track}}, \eta^{\text{track}}, \phi^{\text{track}}, d_0, z_0$
- *Low-level* tower kinematics
  - $E_T^{\text{tower}}, \eta^{\text{tower}}, \phi^{\text{tower}}$
- *Jet-level* kinematics
  - $p_{\text{T}}^{\text{jet}}, \eta^{\text{jet}}, \phi^{\text{jet}}$
- *High-level* variables
  - ATL-PHYS-PUB-2019-033



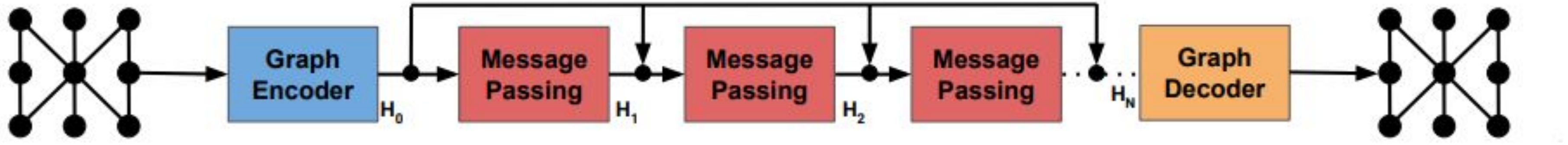
Feature Selections

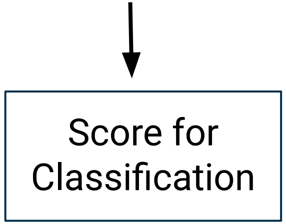
	G1	G2
Track nodes	$p_T^{track}, \eta^{track}, \phi^{track}, d_0, z_0$	$p_T^{track}, \eta^{track}, \phi^{track}, d_0, z_0, p_T^{jet}$
Tower nodes	$E_T^{tower}, \eta^{tower}, \phi^{tower}, 0, 0$	$E_T^{tower}, \eta^{tower}, \phi^{tower}, 0, 0, p_T^{jet}$
Graph-level	None	None

	G3	G4
Track nodes	$p_T^{track}, \eta^{track}, \phi^{track}, d_0, z_0, p_T^{jet}$	$p_T^{track}, \eta^{track} - \eta^{jet}, \phi^{track} - \phi^{jet}, d_0, z_0, p_T^{jet}$
Tower nodes	$E_T^{tower}, \eta^{tower}, \phi^{tower}, 0, 0, p_T^{jet}$	$E_T^{tower}, \eta^{tower} - \eta^{jet}, \phi^{tower} - \phi^{jet}, 0, 0, p_T^{jet}$
Graph-level	$p_T^{jet}, \eta^{jet}, \phi^{jet}$	None

	G5	G6
Track nodes	$p_T^{track}, \eta^{track} - \eta^{jet}, \phi^{track} - \phi^{jet}, d_0, z_0, p_T^{jet}$	$p_T^{track}, \eta^{track} - \eta^{jet}, \phi^{track} - \phi^{jet}, d_0, z_0, p_T^{jet}$
Tower nodes	$E_T^{tower}, \eta^{tower} - \eta^{jet}, \phi^{tower} - \phi^{jet}, 0, 0, p_T^{jet}$	$E_T^{tower}, \eta^{tower} - \eta^{jet}, \phi^{tower} - \phi^{jet}, 0, 0, p_T^{jet}$
Graph-level	$p_T^{jet}, \eta^{jet}, \phi^{jet}$	$p_T^{jet}, \eta^{jet}, \phi^{jet}$, High-level Variables

GNN Architecture





Encoder

- Maps input graph into a hidden representation

Message Passing

- Update edges based on neighboring nodes and globals: $e'_{ij} \leftarrow \phi^{e}(e_{ij}, v_i, v_j, u)$
- Update nodes by aggregating edge information: $v'_j \leftarrow \phi^{v}\left(E'_j, v_j, u\right)$
- Update globals by aggregating nodes and edges: $u' \leftarrow \phi^{u}\left(E', V', u\right)$

Decoder

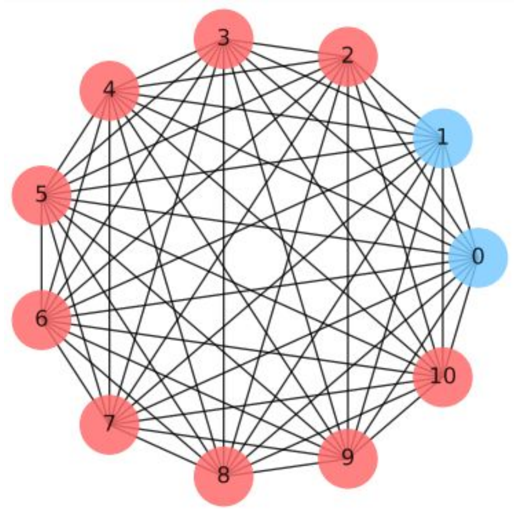
- Update nodes, edges, globals independently
- Apply sigmoid function on globals to produce a score

Heterogeneous Representation

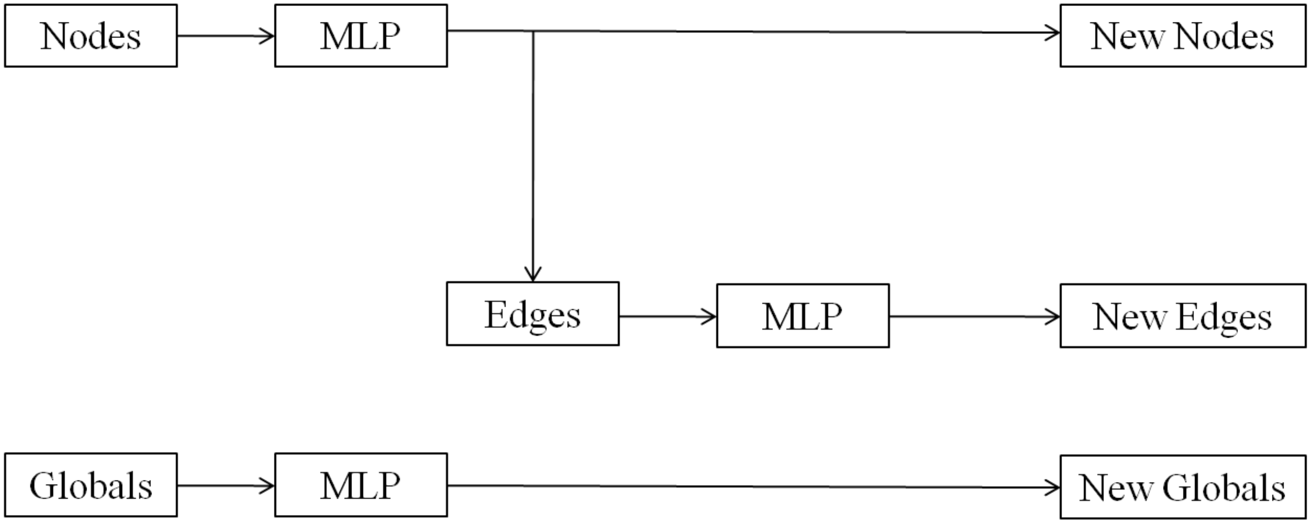
Guiding Question: Should different objects in the same graph be treated in the same way?

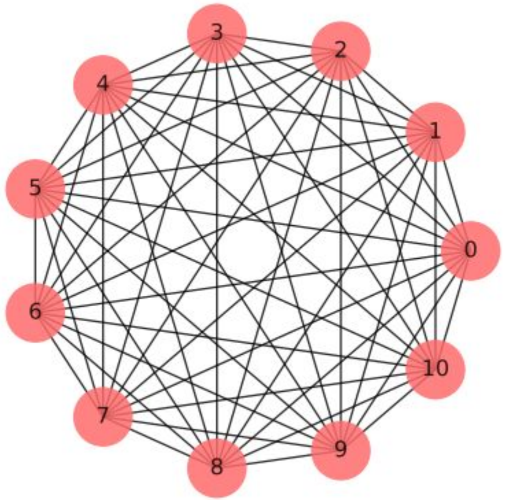
Encoding Schemes:

- Homogeneous Encoding
- Heterogeneous Node Encoding
  - Two types of nodes → two distinct neural network functions
- Heterogeneous Edge & Node Encoding
  - Three types of edges → three distinct NN functions
- Recurrent Encoder (inspired from the RNN architecture)
  - Encode nodes as sequences, no edge encodings

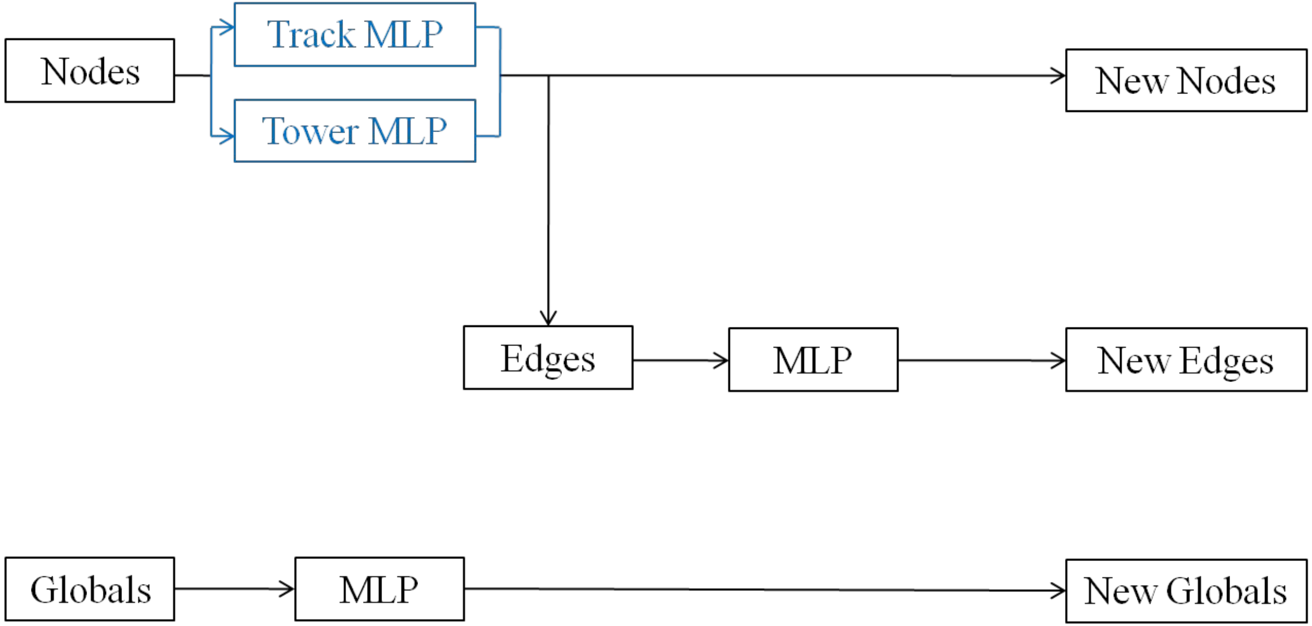


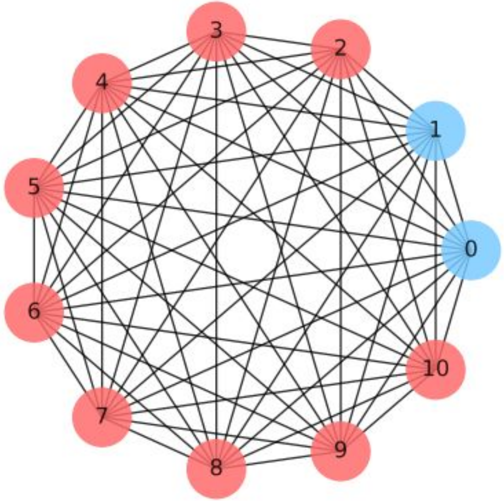
Homogeneous Encodings



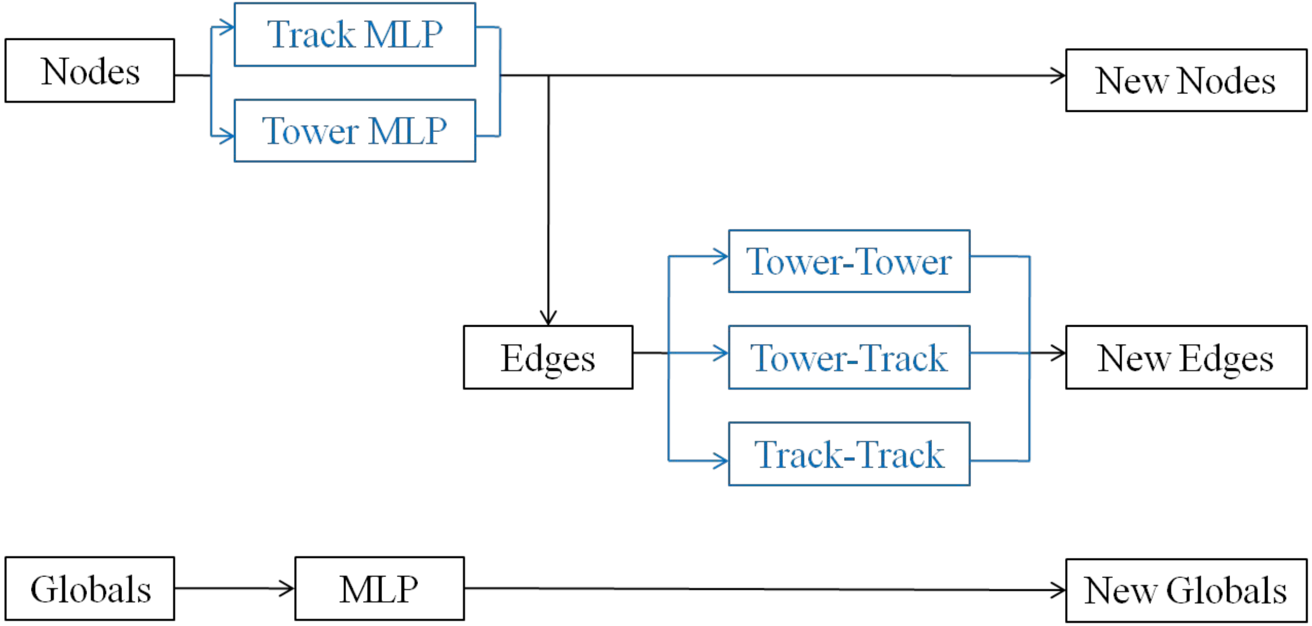


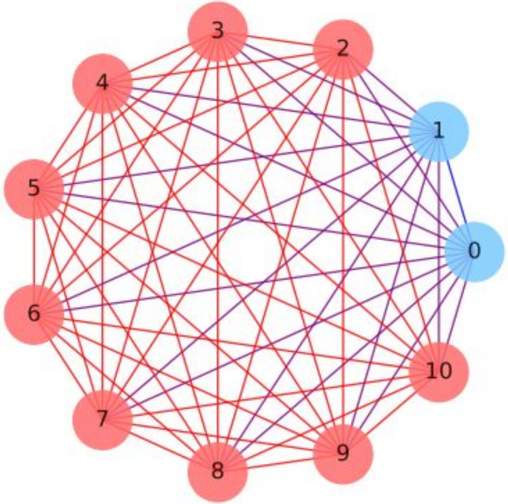
Heterogeneous Node Encodings



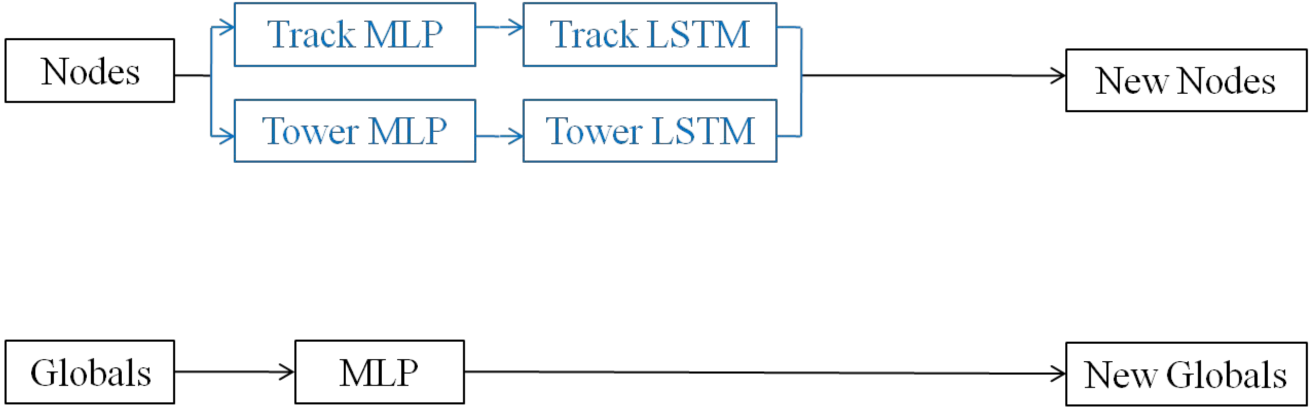


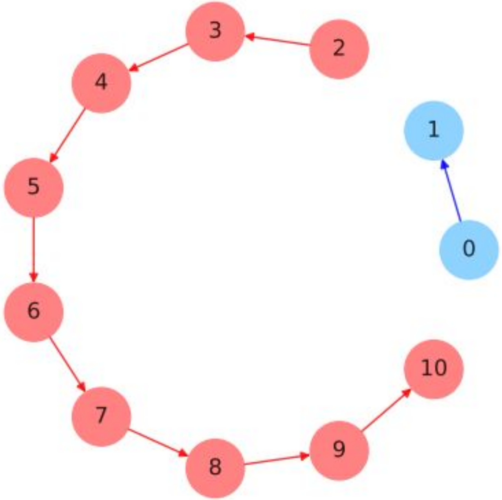
Heterogeneous Node & Edge Encodings





LSTM Encodings





* LSTM requires a fixed-length inputs ⇒ only use first 10 tracks & 6 towers, sorted by decreasing P_T and E_T

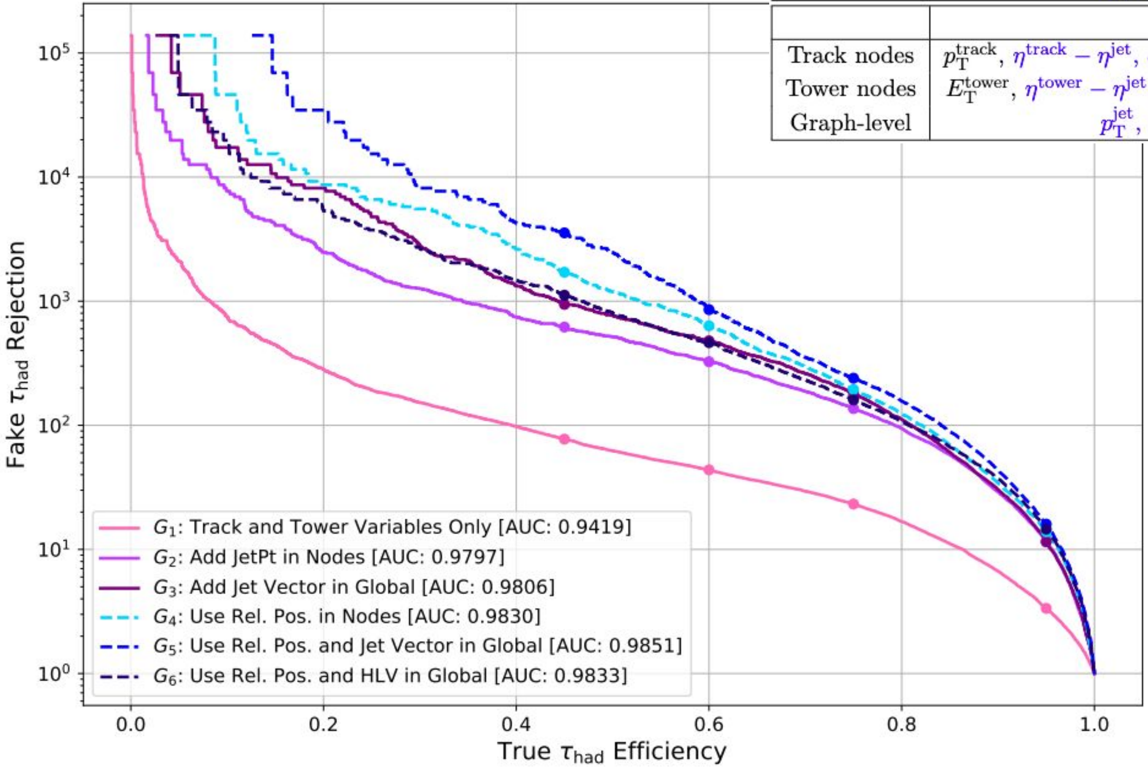
Results

Feature Selections

	G1	G2
Track nodes	p_T^{track}, η^{track}, ϕ^{track}, d_0, z_0	p_T^{track}, η^{track}, ϕ^{track}, d_0, z_0, p_T^{jet}
Tower nodes	E_T^{tower}, η^{tower}, ϕ^{tower}, 0, 0	E_T^{tower}, η^{tower}, ϕ^{tower}, 0, 0, p_T^{jet}
Graph-level	None	None

	G3	G4
Track nodes	p_T^{track}, η^{track}, ϕ^{track}, d_0, z_0, p_T^{jet}	p_T^{track}, $\eta^{track}-\eta^{jet}$, $\phi^{track}-\phi^{jet}$, d_0, z_0, p_T^{jet}
Tower nodes	E_T^{tower}, η^{tower}, ϕ^{tower}, 0, 0, p_T^{jet}	E_T^{tower}, $\eta^{tower}-\eta^{jet}$, $\phi^{tower}-\phi^{jet}$, 0, 0, p_T^{jet}
Graph-level	p_T^{jet}, η^{jet}, ϕ^{jet}	None

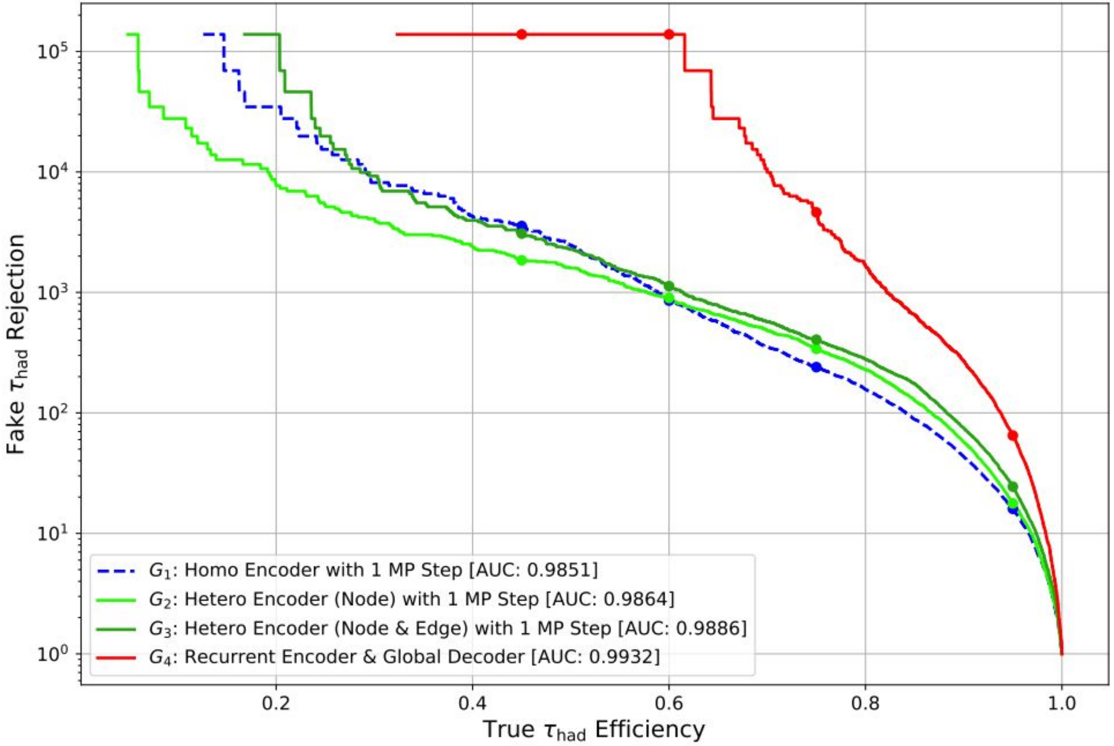
	G5	G6
Track nodes	p_T^{track}, $\eta^{track}-\eta^{jet}$, $\phi^{track}-\phi^{jet}$, d_0, z_0, p_T^{jet}	p_T^{track}, $\eta^{track}-\eta^{jet}$, $\phi^{track}-\phi^{jet}$, d_0, z_0, p_T^{jet}
Tower nodes	E_T^{tower}, $\eta^{tower}-\eta^{jet}$, $\phi^{tower}-\phi^{jet}$, 0, 0, p_T^{jet}	E_T^{tower}, $\eta^{tower}-\eta^{jet}$, $\phi^{tower}-\phi^{jet}$, 0, 0, p_T^{jet}
Graph-level	p_T^{jet}, η^{jet}, ϕ^{jet}	p_T^{jet}, η^{jet}, ϕ^{jet}, High-level Variables



Finding:

Jet-level information are essential for better performance

Heterogeneous Representations



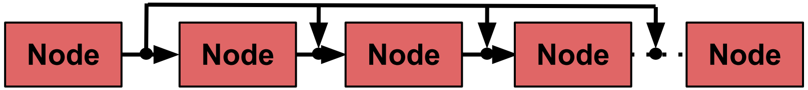
Findings:

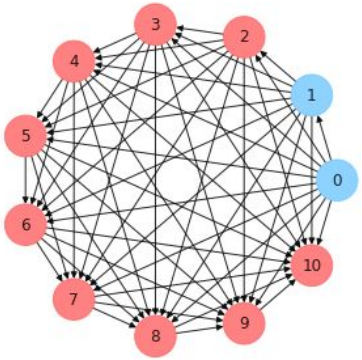
- Heterogeneous encodings
  - Better rejection for high efficiency
  - Similar rejection for low efficiency
- Sequentially biased encoding
  - Outperforms permutationally invariant encodings

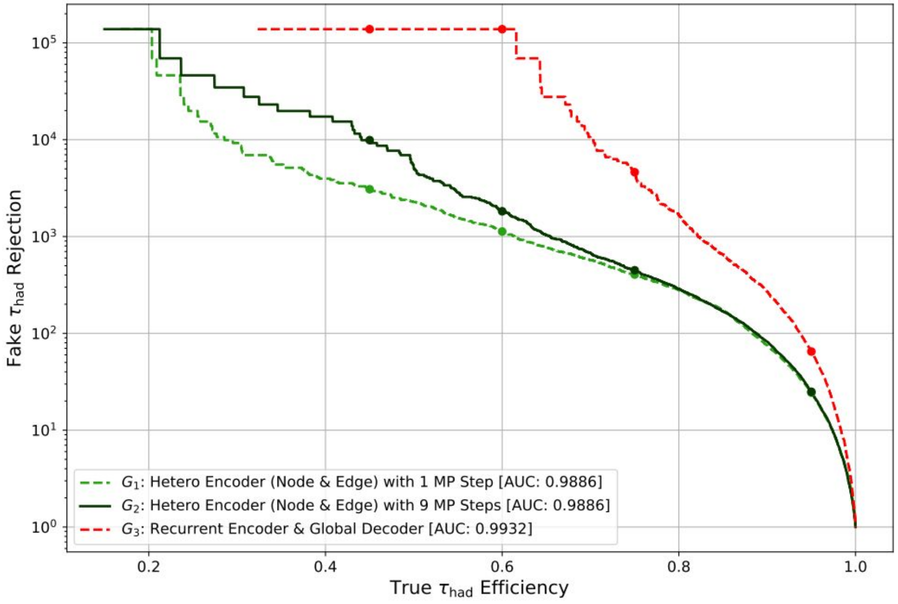
Discussion: More Message Passing Steps

Guiding Question: Why is recurrent encoding more powerful?

- Potential Reason 1: The final node is receiving an aggregated information from ALL previous nodes
  - Improvement on GNN: Large message passing steps
- Potential Reason 2: Sequential Bias

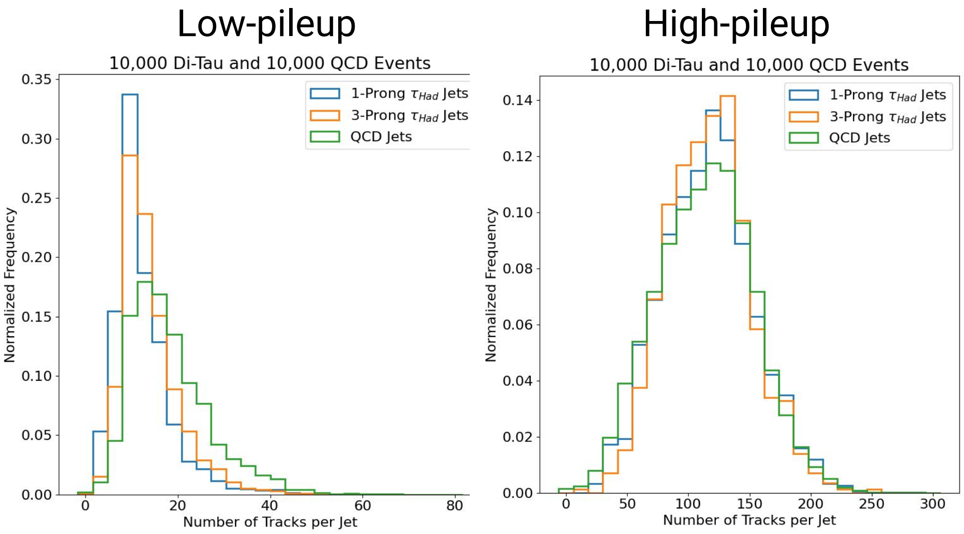






Discussion: Effects of Pileup

Inference Dataset: μ = 200			
Model	Training Dataset	AUC	Rejection at 75% Efficiency
Heterogeneous Node & Edge Encoder	μ = 200	0.9886	448.5
	μ = 40	**0.9614**	**32.8**
	Downgrade	0.0272 (2.75%)	415.7 (92.67%)
Recurrent Encoder	μ = 200	0.9932	4616.7
	μ = 40	**0.9722**	**117.3**
	Downgrade	0.0210 (2.11%)	4499.4 (97.46%)



Summary

- GNN architecture with fully-connected graphs for tau identification
- Feature Selections
  - Jet-level information are essential for better performance
- Heterogeneous Representation
  - Heterogeneous models yield better rejection for high efficiency and similar rejection for low efficiency than homogeneous model
  - Sequentially biased encoding outperforms permutationally invariant encodings
  - More message passing steps tends to improve performance

Rejection Curve for Effects of Pileup