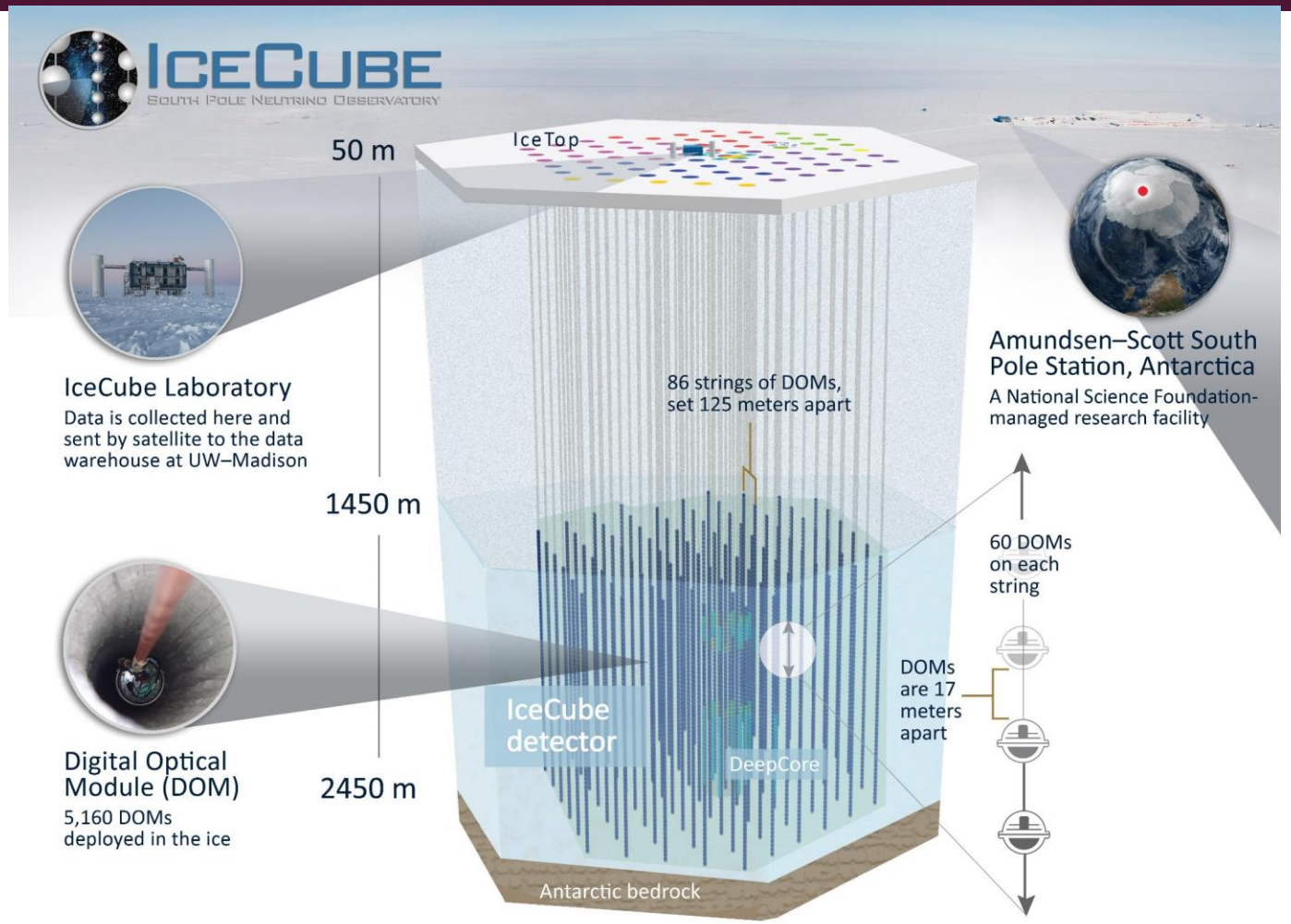
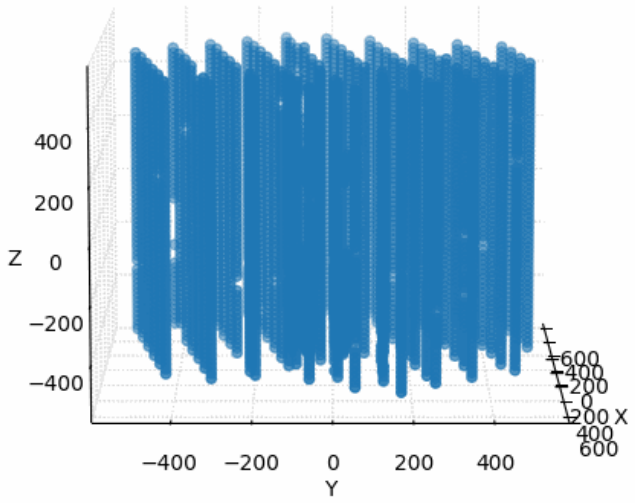
GRAPH NEURAL NETWORK ON ICECUBE EXPERIMENT

ANA-IULIA HUTU, THOR KONGSTAD MADSEN*

*all members contributed equally


IceCube
South Pole Neutrino Observatory
50 m
IceTop
IceCube Laboratory
Data is collected here and sent by satellite to the data warehouse at UW–Madison
1450 m
86 strings of DOMs, set 125 meters apart
Amundsen–Scott South Pole Station, Antarctica
A National Science Foundation-managed research facility
60 DOMs on each string
DOMs are 17 meters apart
IceCube detector
DeepCore
Digital Optical Module (DOM)
5,160 DOMs deployed in the ice
2450 m
Antarctic bedrock



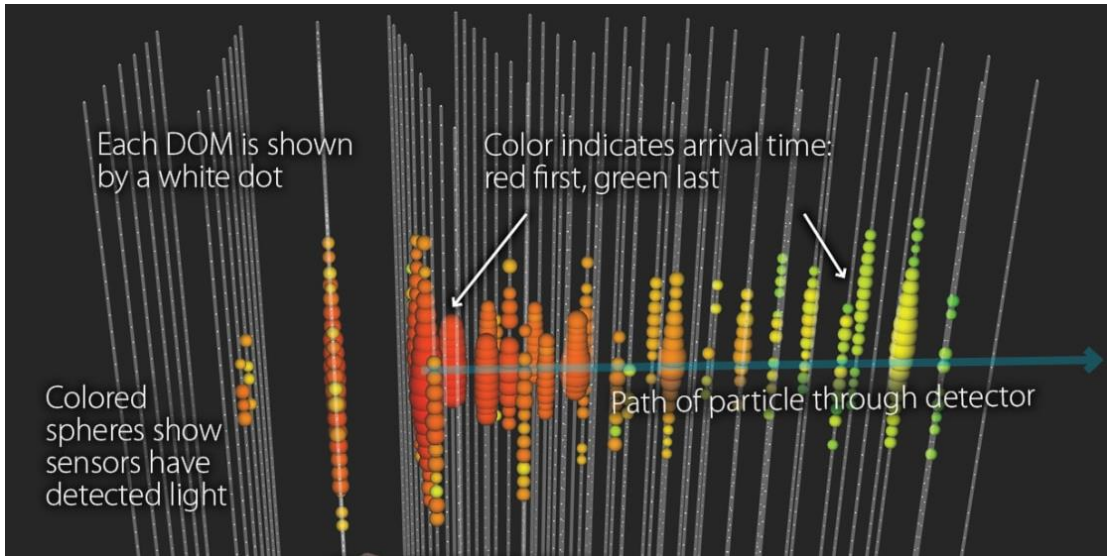
Dataset (simulation generated):

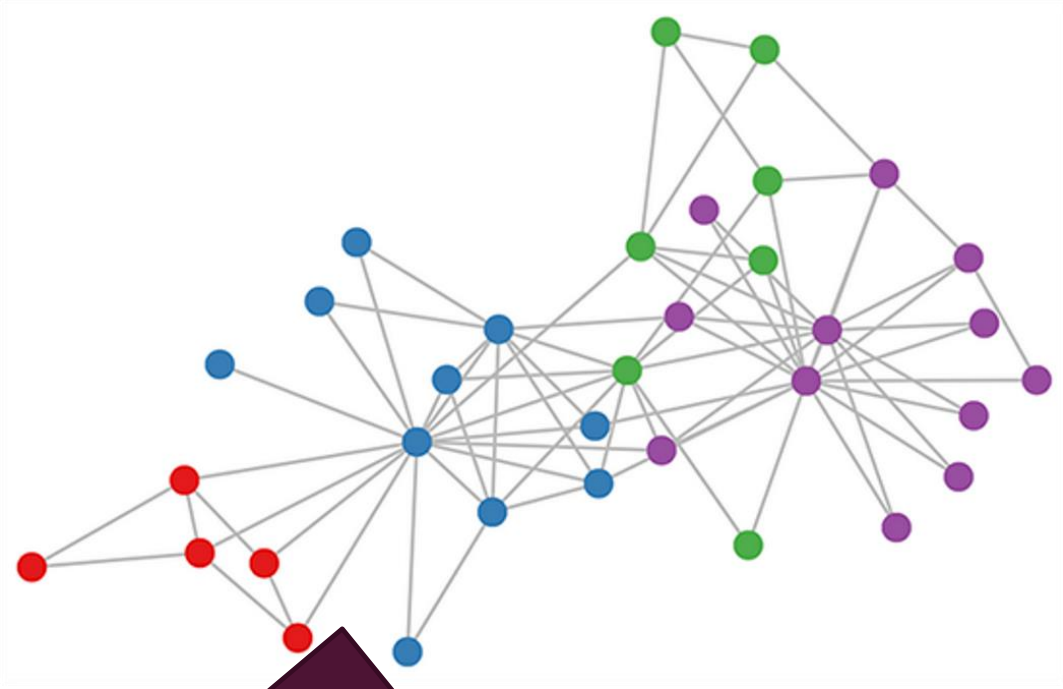
Graph definition:
DOMs: 5074
DOM strings: 144

Features/events:
14.213.122 DOM activations:
DOM positions, DOM time, charge

Truth values:
200.000 events: energy, particle id, and many more (total of 16)

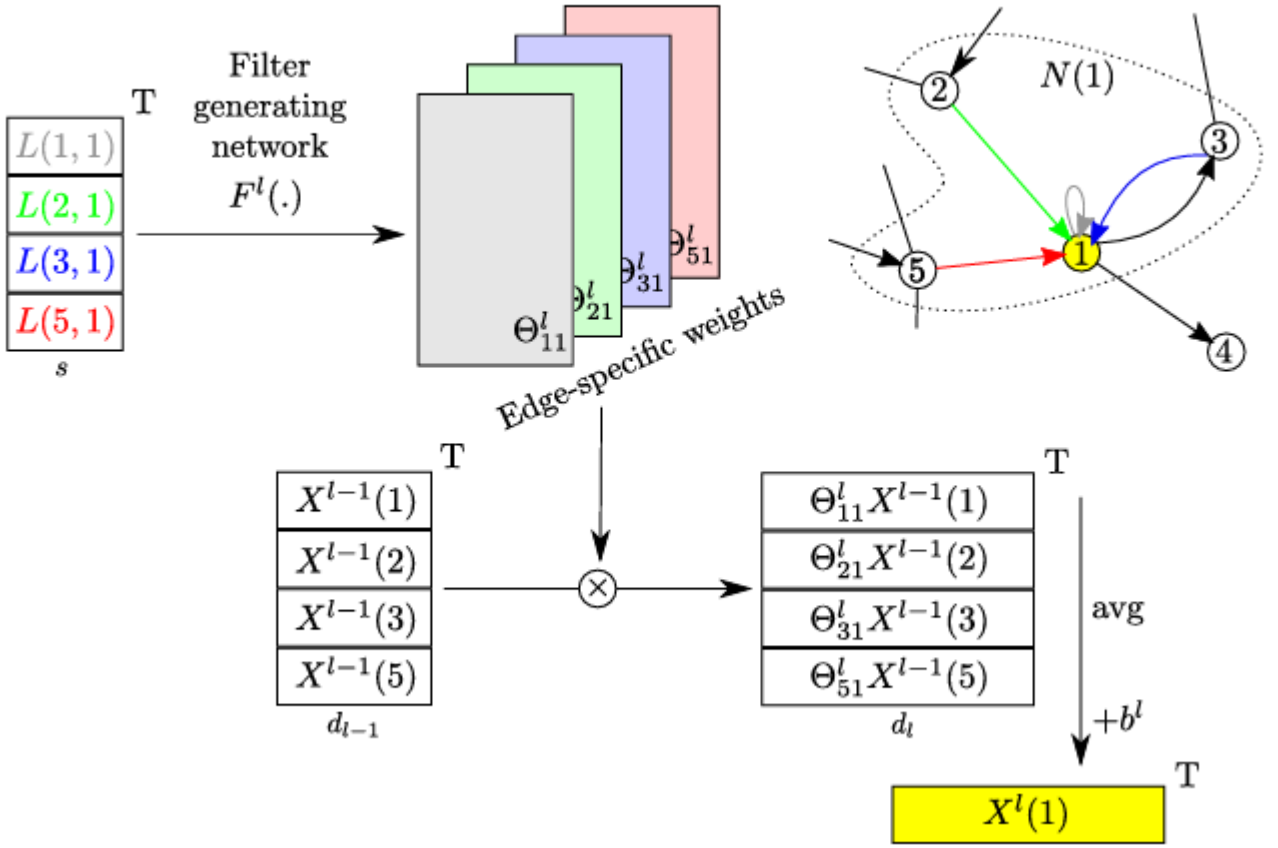
GRAPH NEURAL NETWORK (GNN) ON ICECUBE DATA





The positions and features of the activated DOMs
are turned into a higher dimensional graph.

DYNAMIC EDGE CONVOLUTION



IMPLEMENTATION

Preprocessing

- NaN values check
- Outliers: zscore
- Scaling: predefined through Graphnet
- Data splitting: split into train-test,
  split again train into train-validation(0.7/0.3)

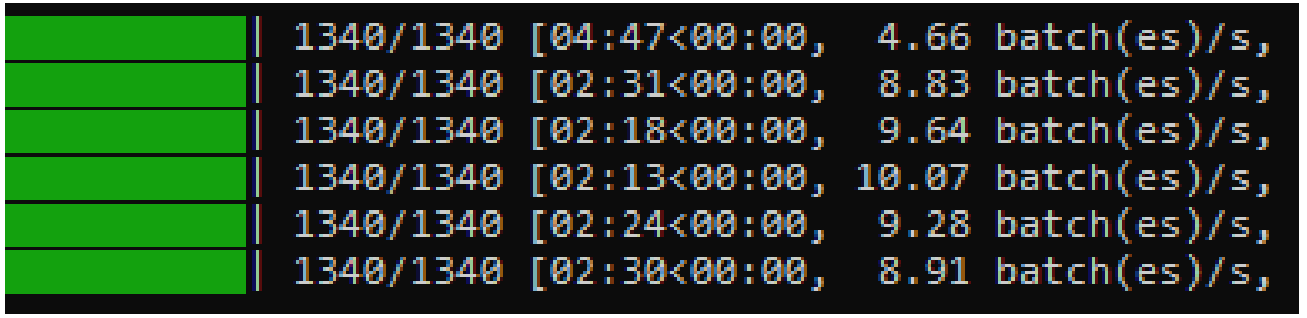
Processing

- Graphnet package
  - https://github.com/graphnet-team/graphnet
- Classification of events into neutrino/muon
  - Loss Function: Binary Cross Entropy Loss
- Regression to determine energy of event
  - Loss Function: LogCosh
- Optimizer: Adam

THE TROUBLES

- Expected: GPU-limiting timestep
- Unexpected: CPU-limiting timestep

- Course: Creating input graphs proved to be CPU-intensive, hindering the process
- Graphnet's implementation did not save input graphs between epochs
- Our contribution?



STILL CPU limited on 11th gen i7 core !!!!

THE TROUBLES CONTINUED
DATASET SIZE

- Started with a dataset of 100 000 events in training set
  - In 10 hours, we had built graphs for less than 10% of the first epoch
- Reduced to 15 000 events in training
  - An epoch takes 4 hours
- Training is a breeze, but dataloader is CPU intensive

REDUCED TO 2000 EVENTS

Predictions to come (hopefully)
Likely not

APPENDIX

- https://github.com/ThorKongstad/ml_finale_project_code