# Graph-Convolutional Neural Networks for Large-scale structure clustering

**Farida Farsian**, Federico Marulli, lauro Moscardini, Carlo Giocoli

Bayesian Deep Learning for Cosmology and Time Domain Astrophysics workshop

Paris, June 2022

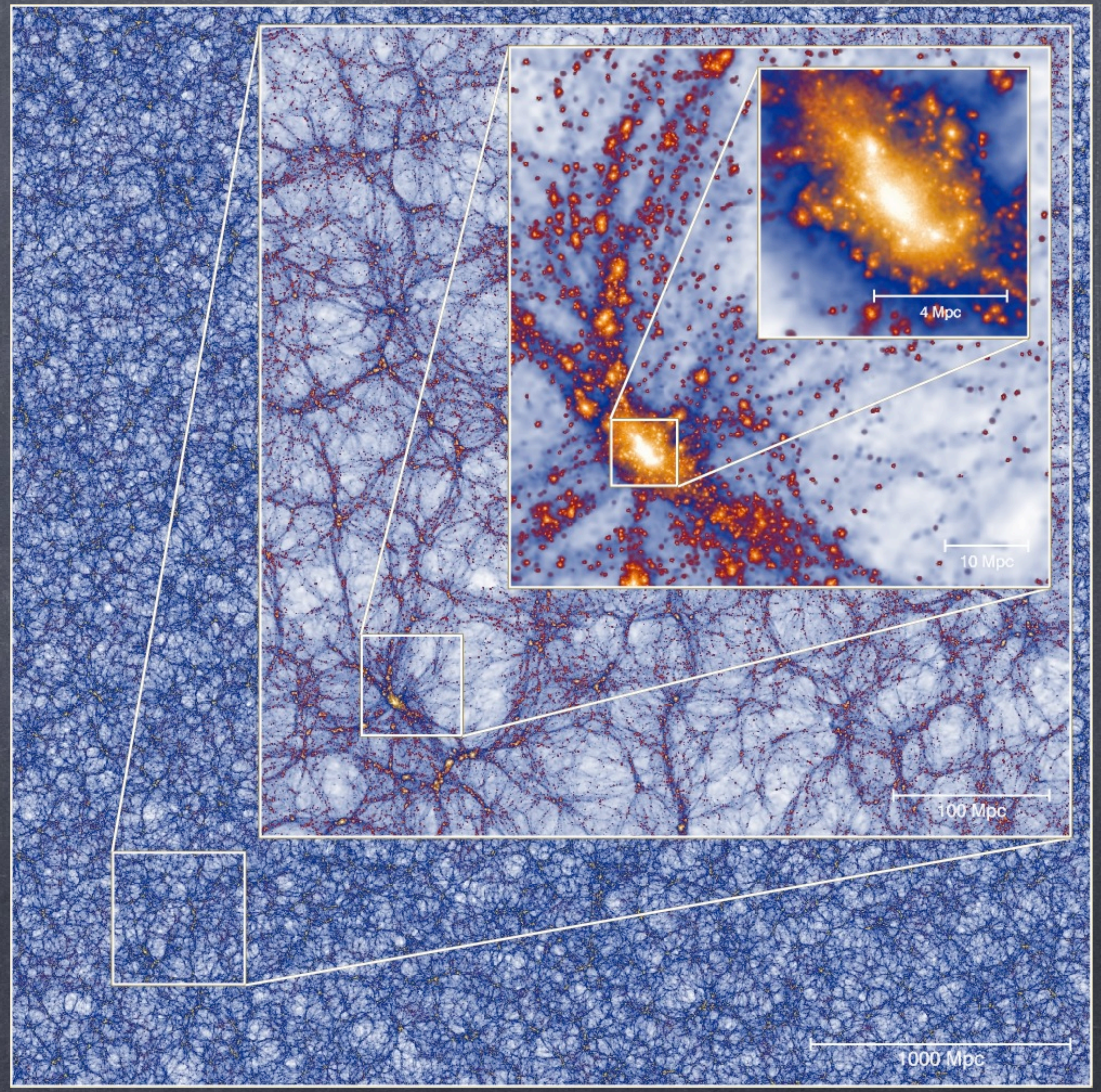
# Motivation

- Standard cosmological analyses based on abundances, two-point and higher-order statistics, for extracting the information encoded in the Large Scale Structure (LSS), have been widely used up and

- They can only exploit a sub-set of the whole information content available.

- For Euclid like surveys, many Dark Energy and Modified Gravity models have to be explored.

- The need of extracting maximum information from Dark Matter (DM) halo spatial distribution without using compressed statistics.

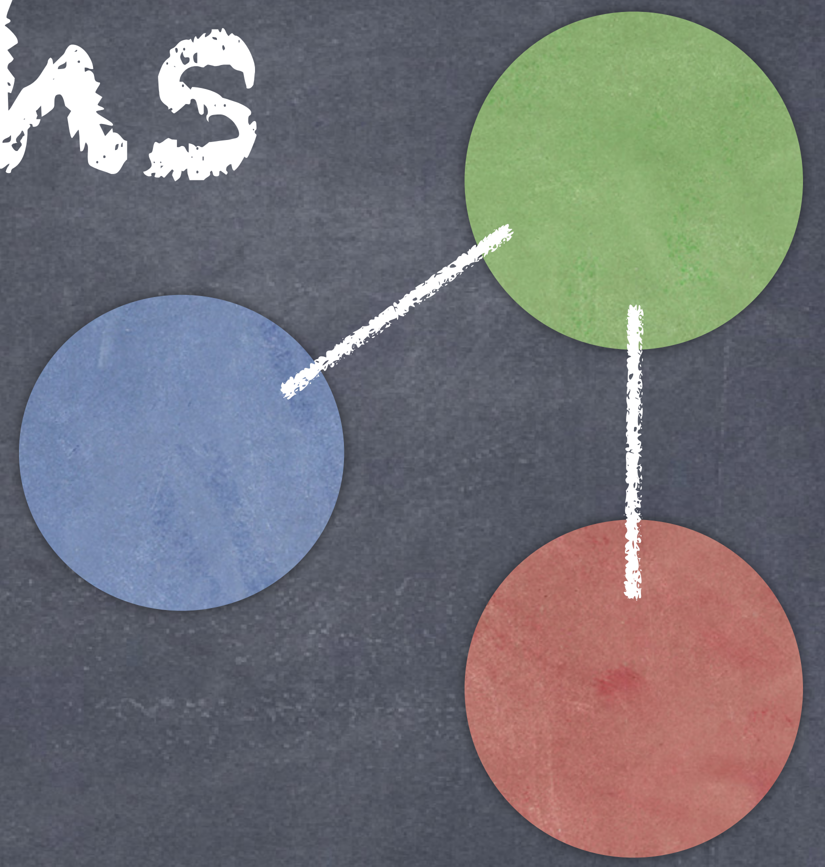


Question:
How to extract the whole information available without compressing it?

Possible Answer:
using the row data in the halo/galaxy catalogue

Challenge:
facing a sparse data, grasping larger scale statistics (the relations between halos/galaxies)

# Proposed solution: Graphs



- Considering row information of DM field, such as mass and coordinates of halos.

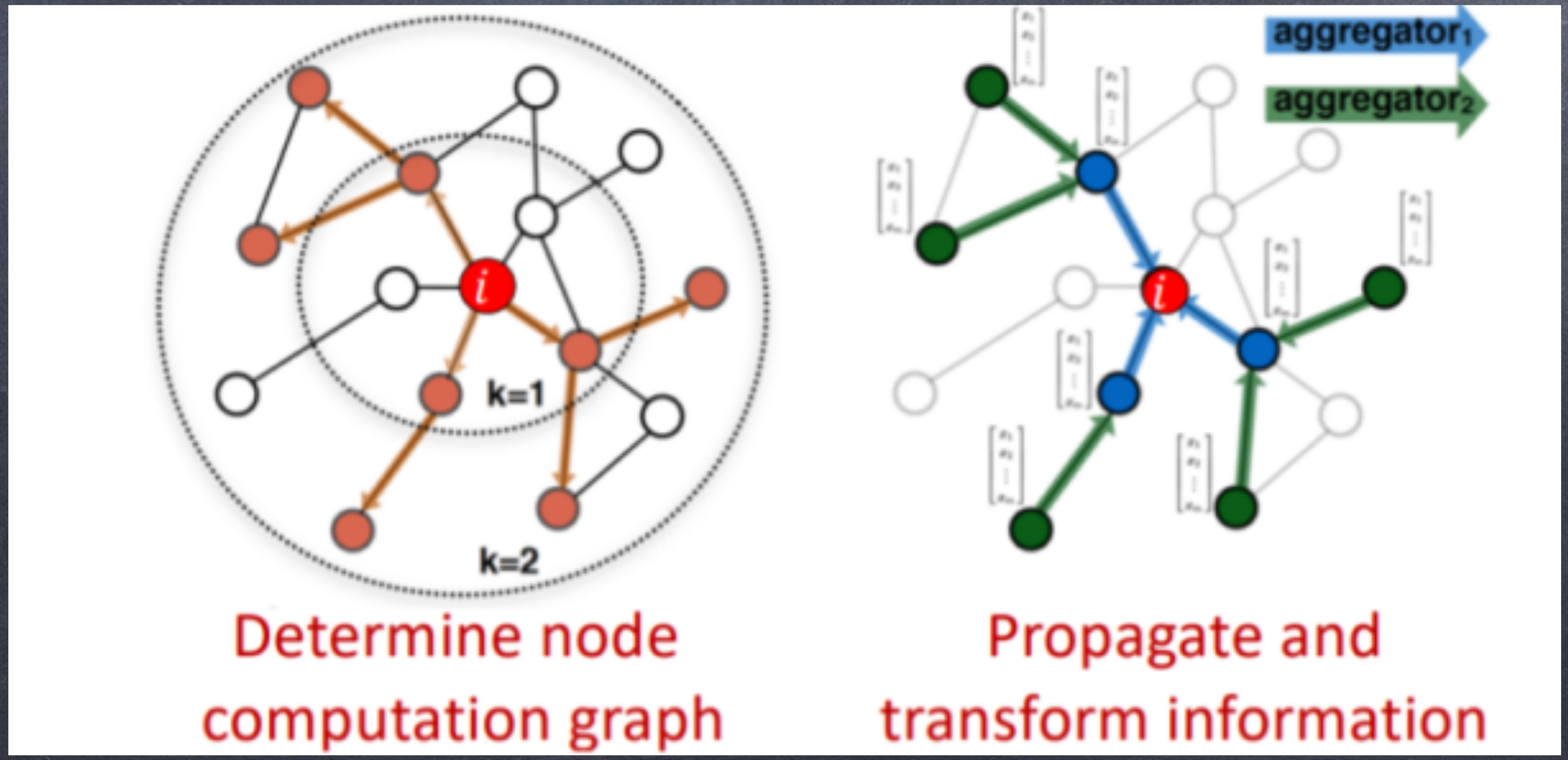
- Representation of cosmic web data in the form of graphs contains the clustering information automatically.

- Using Graph Neural Network to capture the graph structure of data.

- Such method can be used to extract cosmological parameters in likelihood-free manner.

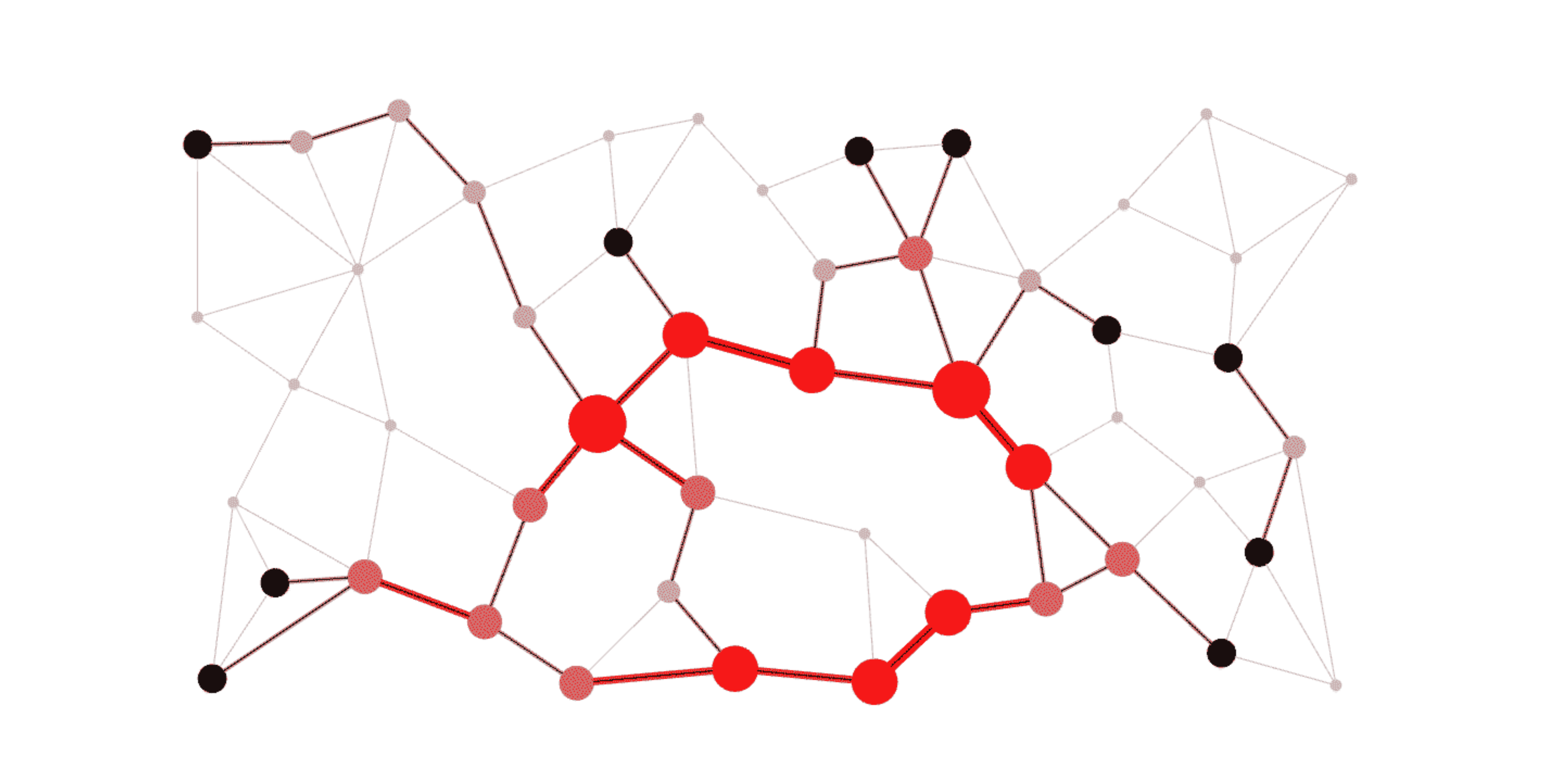
# Graph Neural Networks

- GNNs are a class of deep learning methods designed to perform inference on data described by graphs.
- The information can be extracted in different leve:
  - Node level, - Edge Level, - Graph level



# Graph Neural Network

- Able to capture the graph structure of data which is often very rich
- Able to apprehend global permutation invariant quantities
- Suitable to deal with irregular and sparse data

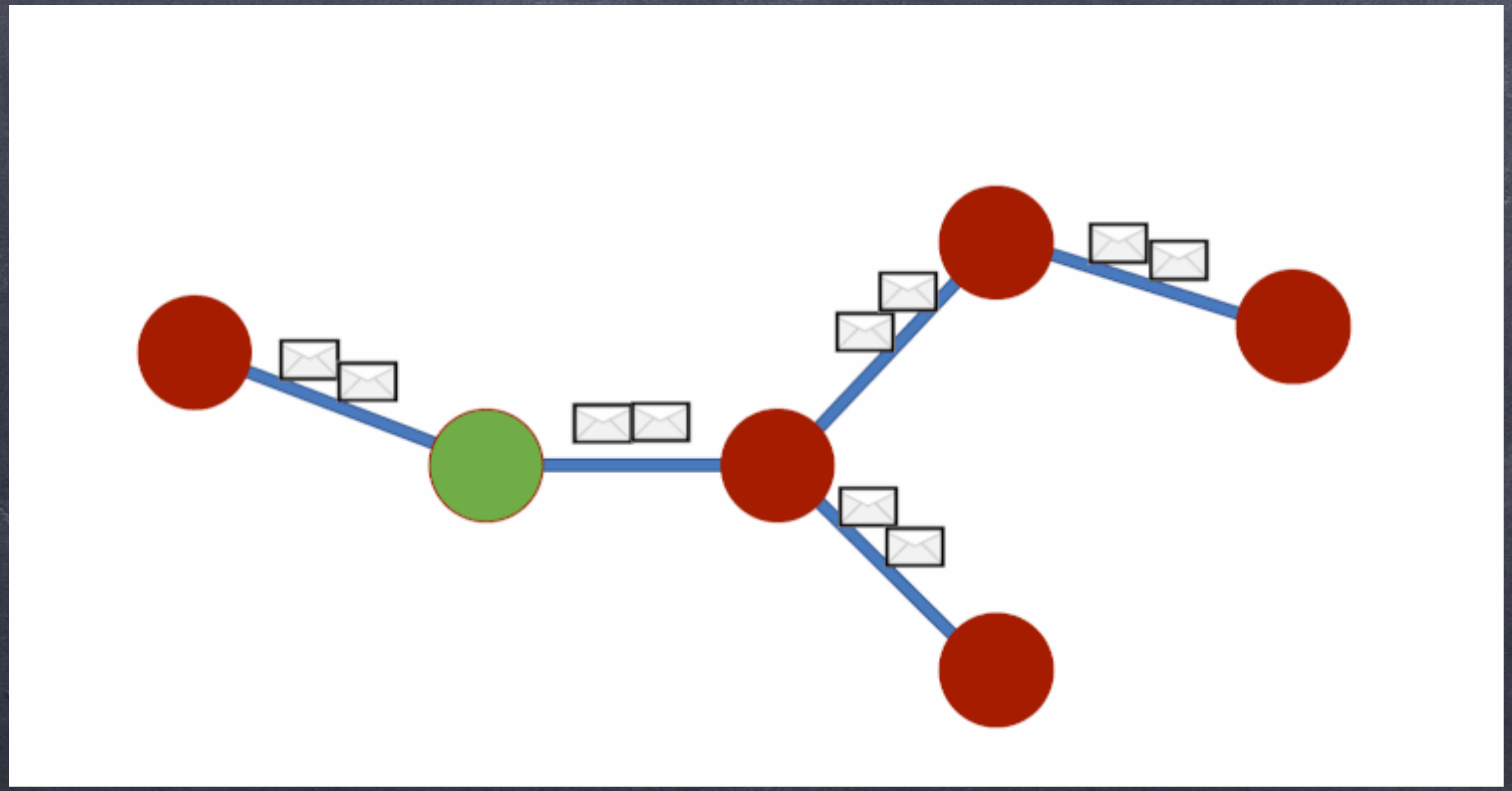


# Message passing schema

1. for each node in the graph, all the neighboring node messages are gathered;

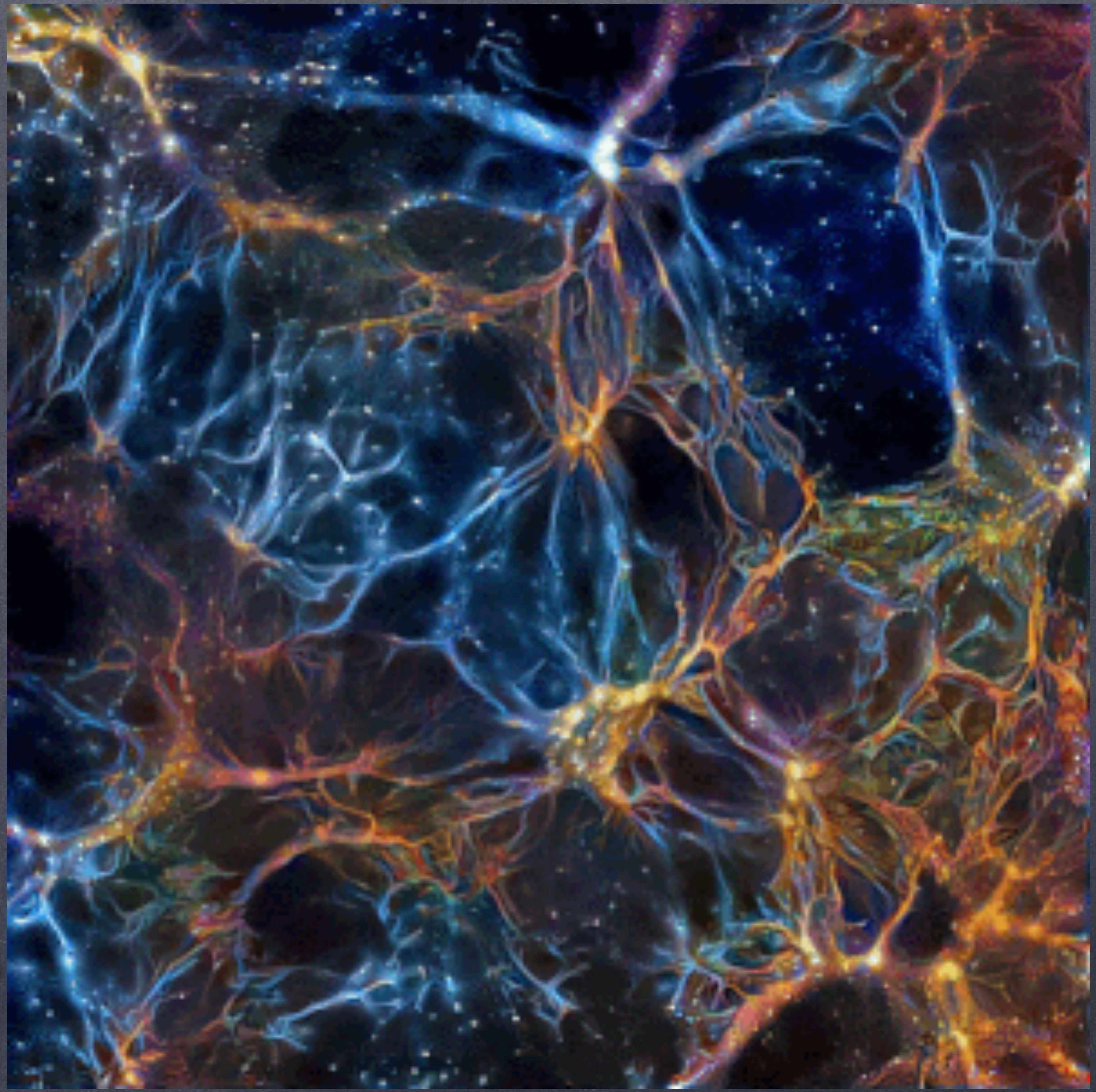
2. then, all messages are aggregated via an aggregate function (like sum).

3. lastly, all pooled messages are passed through an update function, usually a learned neural network



# Simulation

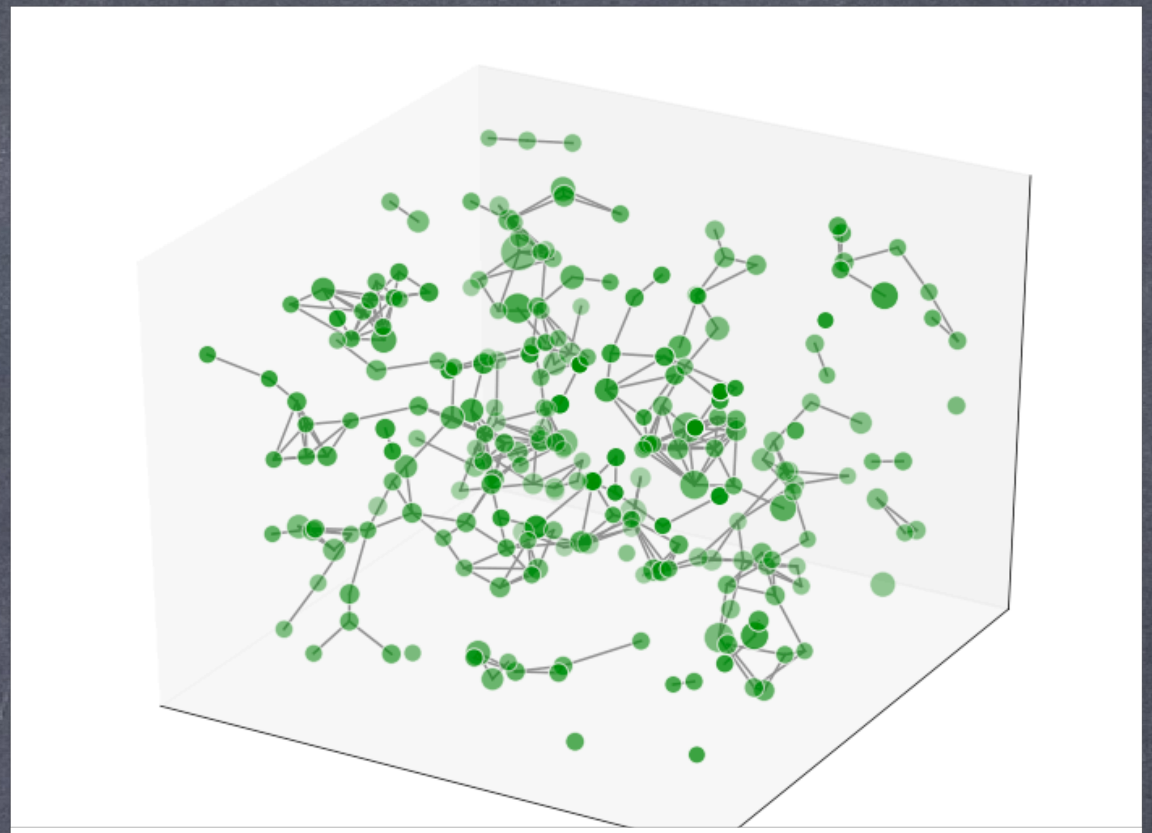
- Quijote simulation, full N-body, Boxes of 1 Gpc/h
- 500 realizations at z= 0.
- The DE parameter, w0, changes in the range of [-1.05, -1, -0.95].

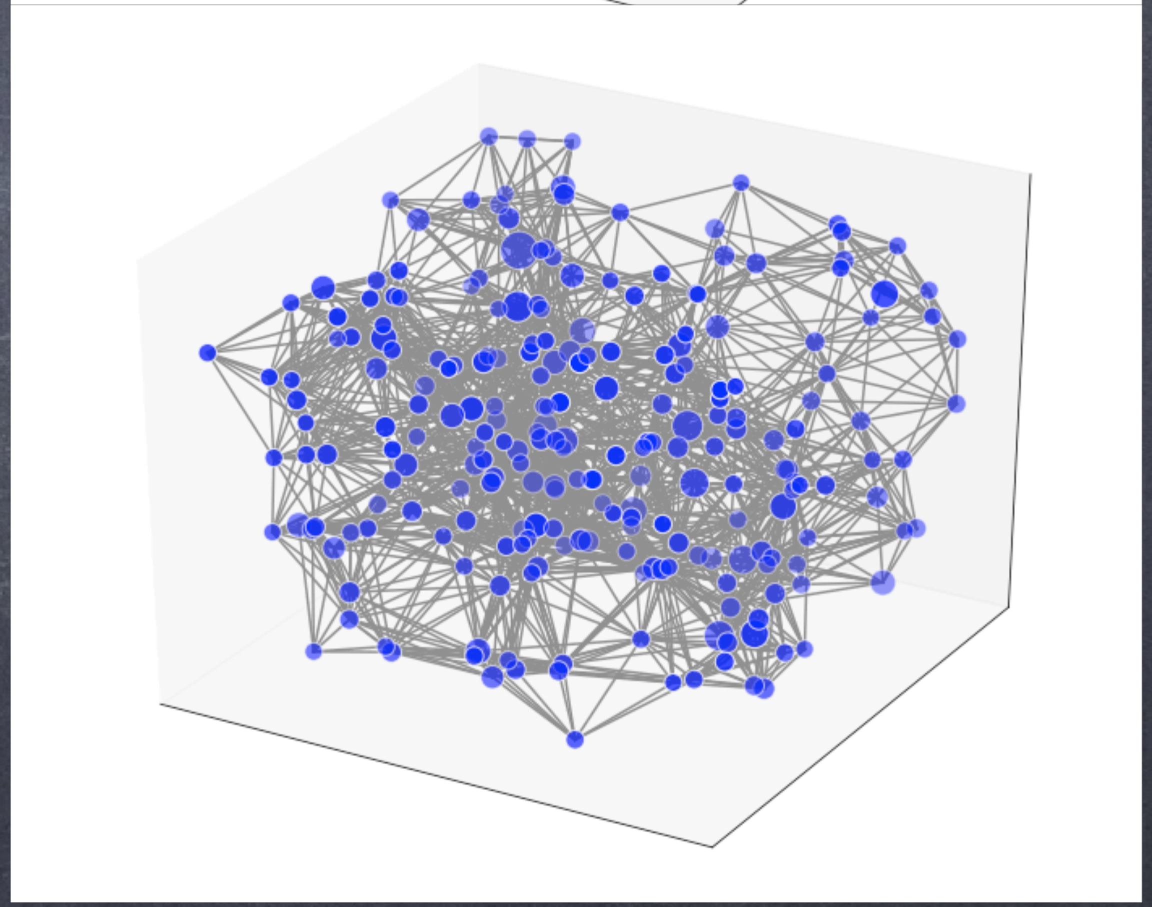


Quijote, arXiv:1909.05273 FVN et al 2020

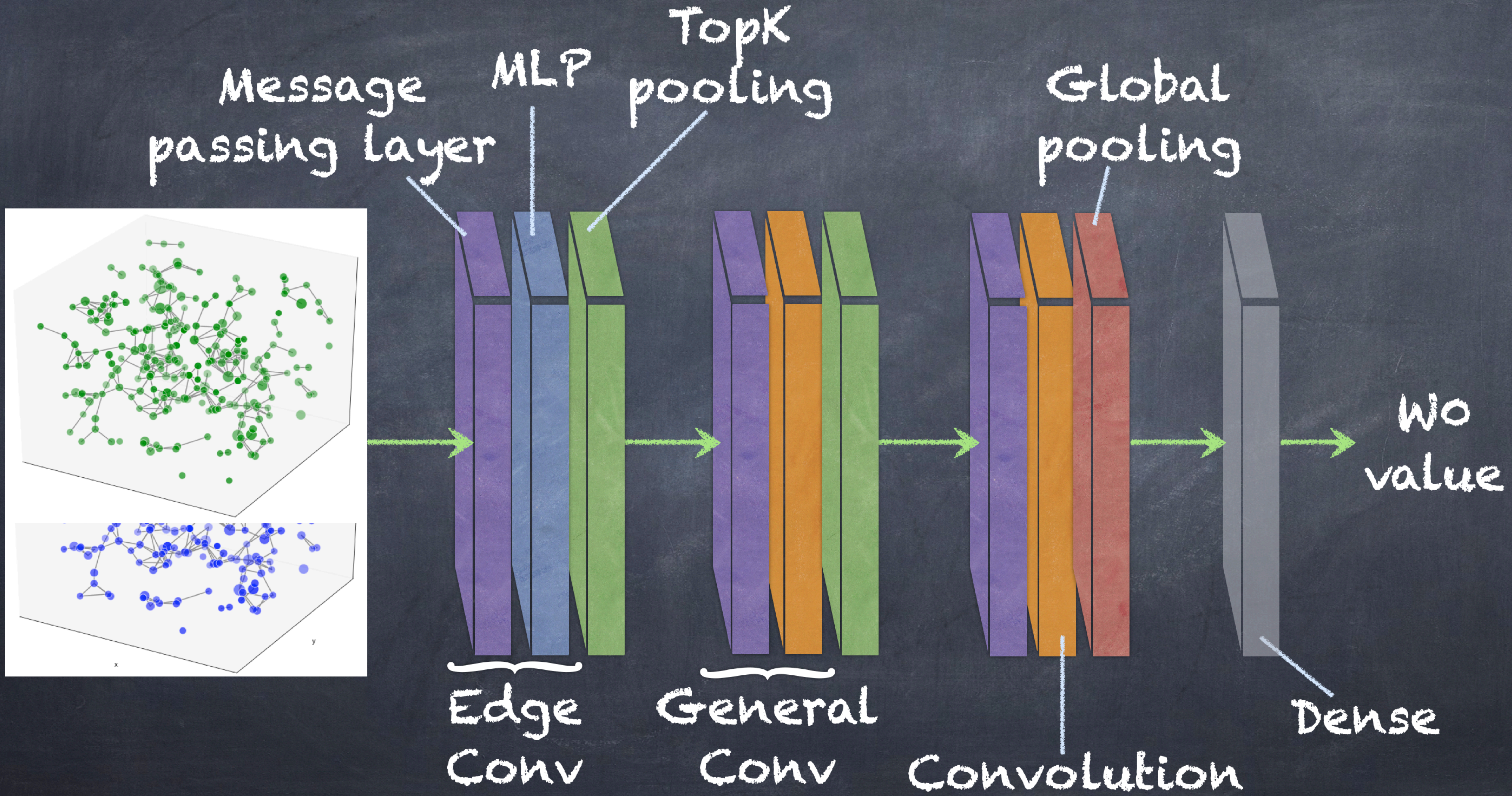
# Data prepration

- Only mass and coordinates of halos are give to the network as features.
- We have applied mass cut of $7 \times 10^{14}$ for each catalogue.
- Two nodes i and j are connected by an edge if they are c.loser than a certain distance r
- In our analysis r is a free parameter, the results presented here are based on r = 100 Mpc





# Network architecture




Yue Wang et al. 2018 Jiaxuan You et al. 2020

# Other setup of the project:

Graph-level analysis is done for the DM halo catalogue

360 realizations dedicated to training set, 40 validation set, 100 training set

Using Spektral package, in Tensorflow

Optimizer: ADAM

Activation function: Relu

Loss function: MSE

# Results: Classification

- 99% of accuracy for Binary classification, to distinguish between w0 = -1.05 & w0 = -0.95.

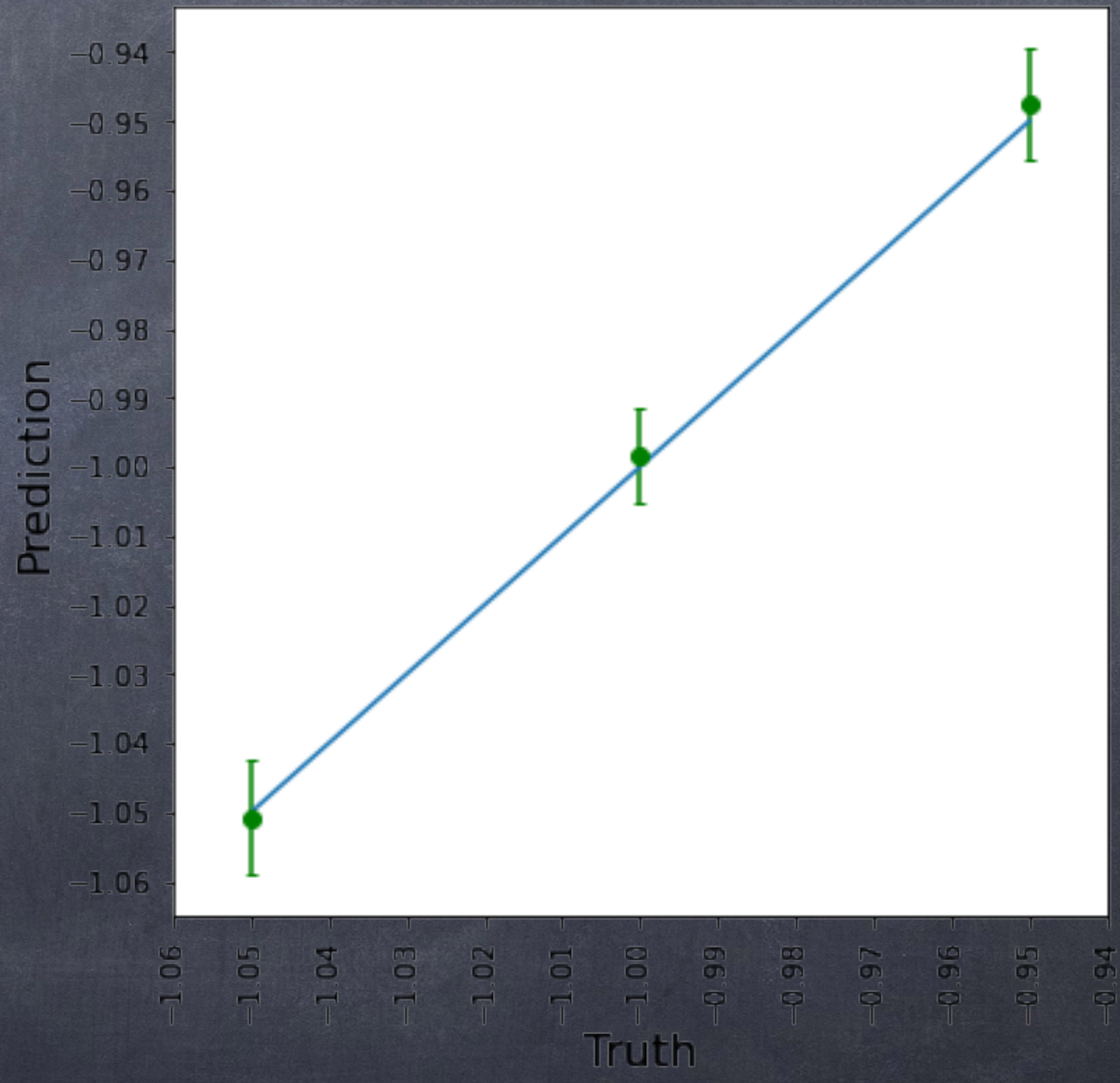
- 97% of accuracy for Multi-class classification, to distinguish between three values of w0.

| | Redshift | # halo (per realization) | Range | Train Acc (360 realizations) | Valid acc (40 realizations) | Test acc (100 realizations) |
|---|---|---|---|---|---|---|
| Binary classification | z= 0 | ~1000 | [0, 100] | 100% | 100% | 99% |
| Multi classification | z= 0 | ~1000 | [0, 100] | 99% | 99% | 97% |

F.Farsian et al. in prep

# Results: Regression

The GNN is able to predict the value of w0 correctly with only 2% error.




F.Farsian et al. in prep

# Future prospect

- Using Pinocchio simulation which has control on w0 and wa parameters and Applying the GNN on different DE models
- Provide a Constraint on modified gravity models and comparing with standard probes constraints
- Application of developed GNN on mock Galaxy catalogues, to provide forecasts for next-generation galaxy redshift surveys, such as Euclid and LSST.

# Conclusion

- GNN can be applied on any kind of astrophysical data which are characterised by point clouds.
- The built model is able to distinguish different dark energy models with very high accuracy in Binary classification (acc = 99%) and Multi-class classification (acc = 97%)
- The model is able to predict the value of w0 for the specific model with high precession.

Thank you for your attention