UNIVERSITY OF ZAGREB
Faculty of Electrical Engineering and Computing

TECHNISCHE UNIVERSITÄT WIEN

Technische Universität Berlin

AIoTwin

Twinning action for spreading excellence in Artificial Intelligence of Things

Semi-decentralized Training of Spatio-Temporal Graph Neural Networks for Traffic Prediction

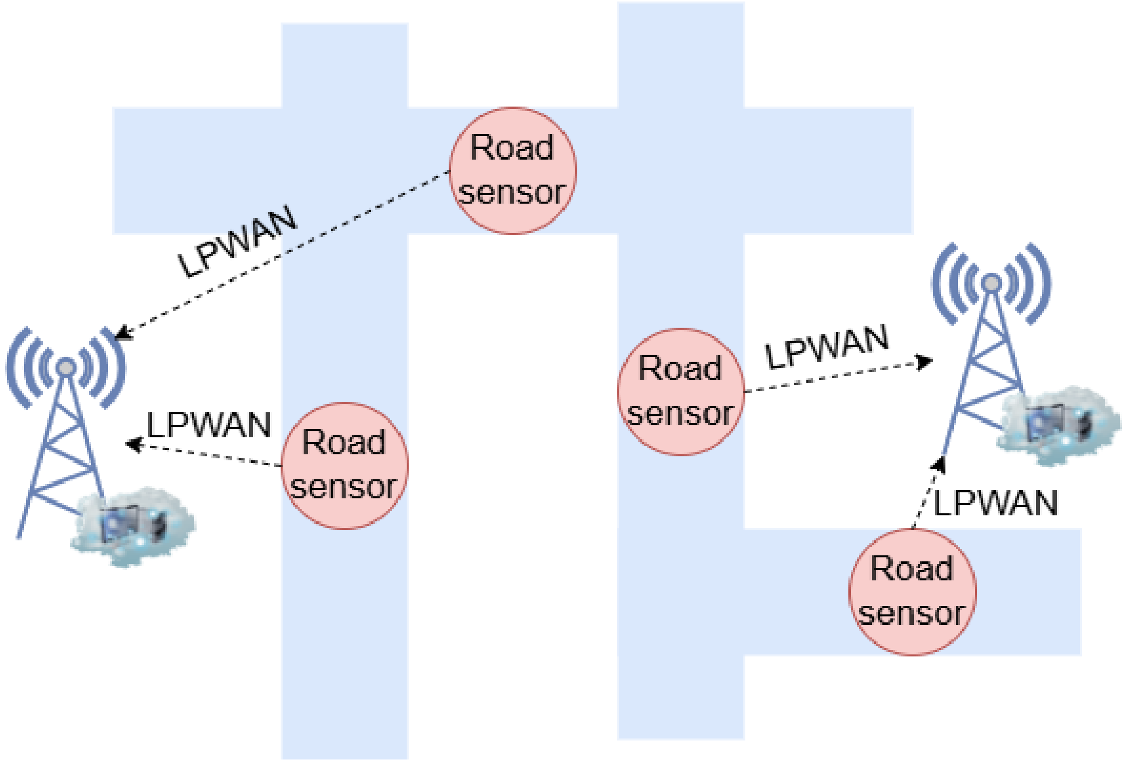
Ivan Kralj, mag. ing.

European Commission

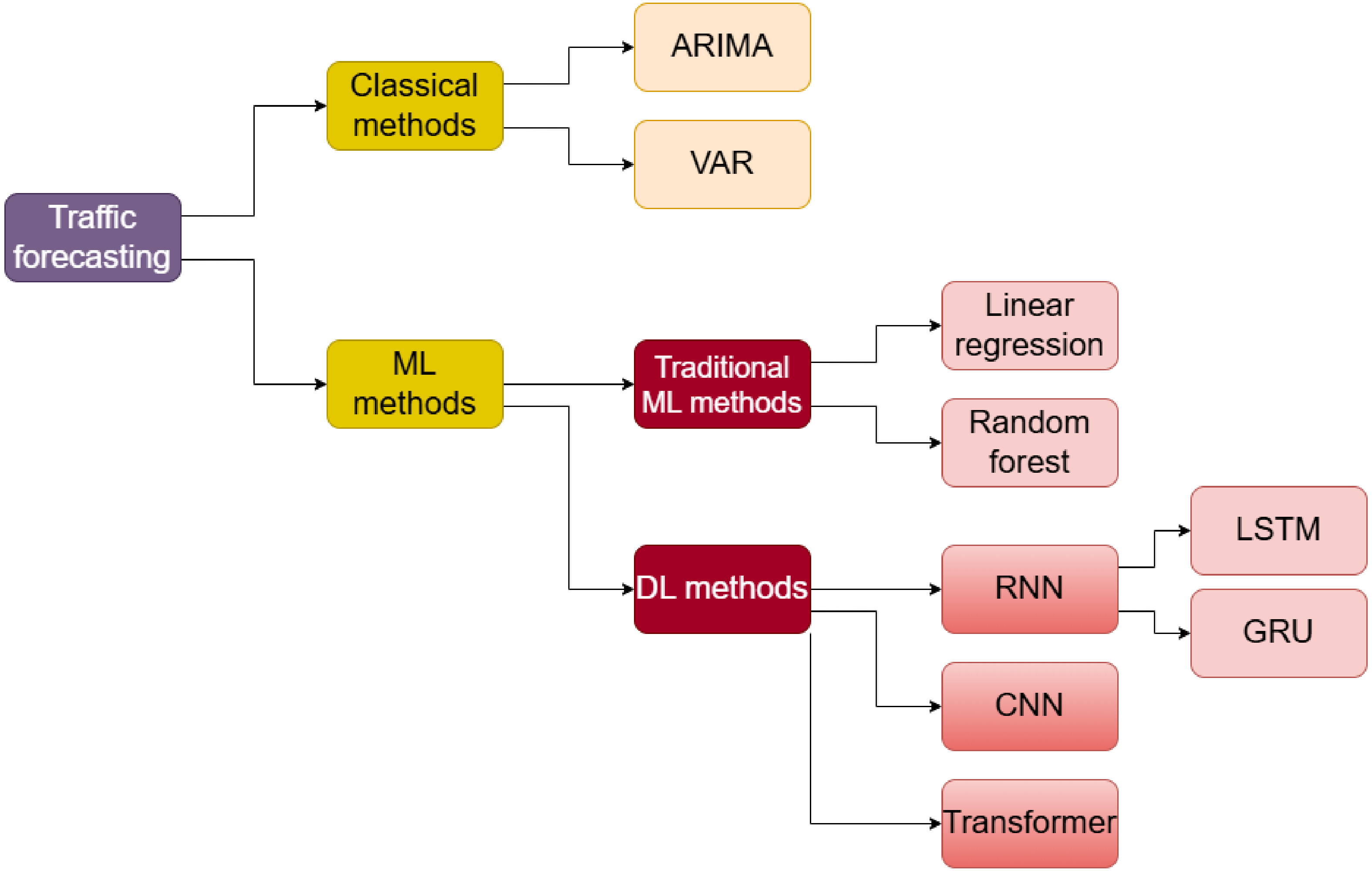
Grant agreement ID: 101079214

Introduction

- Traffic prediction
  - Speed, volume, density…
  - Forecast horizons: short-term (15 min), mid-term (30-min) and long-term (60-min)

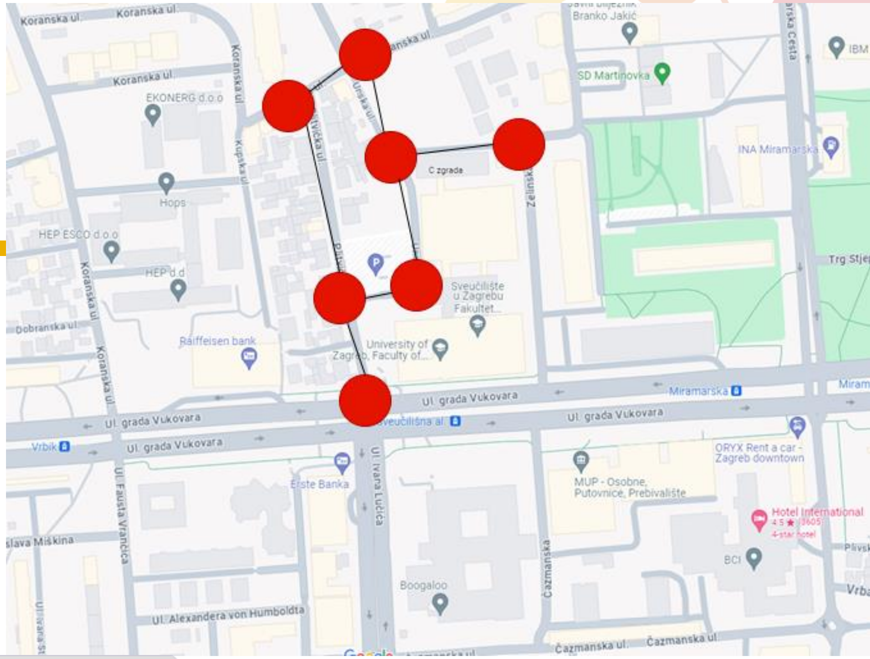


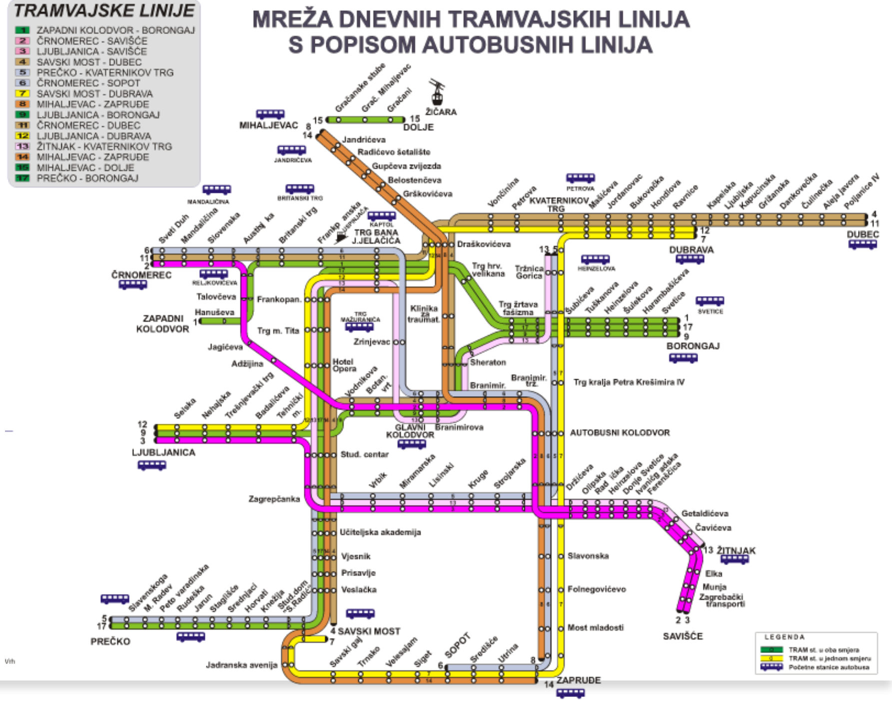
Introduction



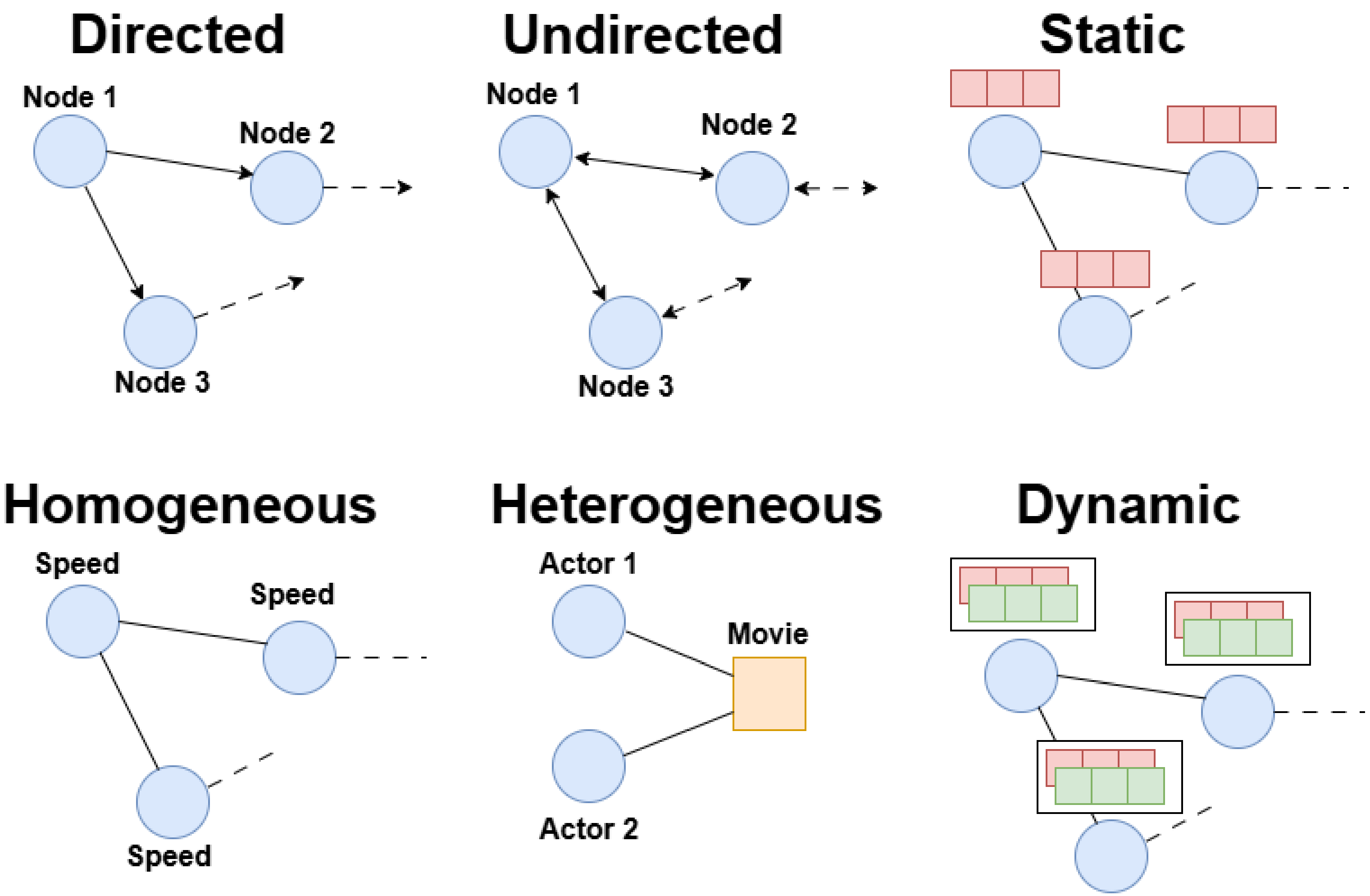
What are graphs?

- In general: G = (V, E)
  - V = $\{v_1, \ldots, v_{|V|}\}$ – set of nodes
  - $E = \{e_1, \ldots e_{|V|}\}$ – set of edges
    - $e = (v_i, v_j) \in E$ – connection relationship between 2 nodes
- In ML: G = (X, A)
  - $X \in \mathbb{R}^{|V| x\ M}$ - node feature matrix
    - M – number of node features
  - $A \in \mathbb{R}^{|V| x\ |V|}$ - adjacency matrix
    - $A_{ij} = 1\ if\ (v_i, v_j) \in E, otherwise\ A_{ij} = 0$



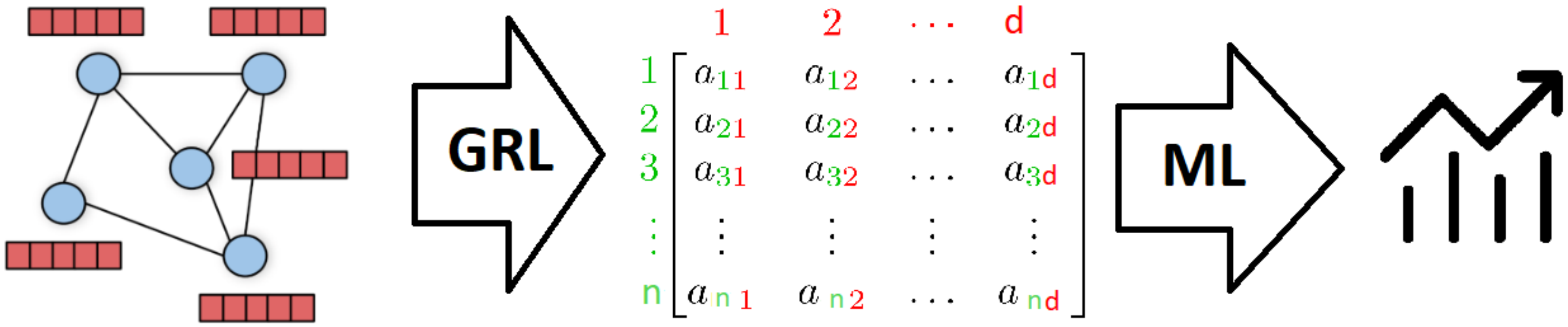


Types of graphs



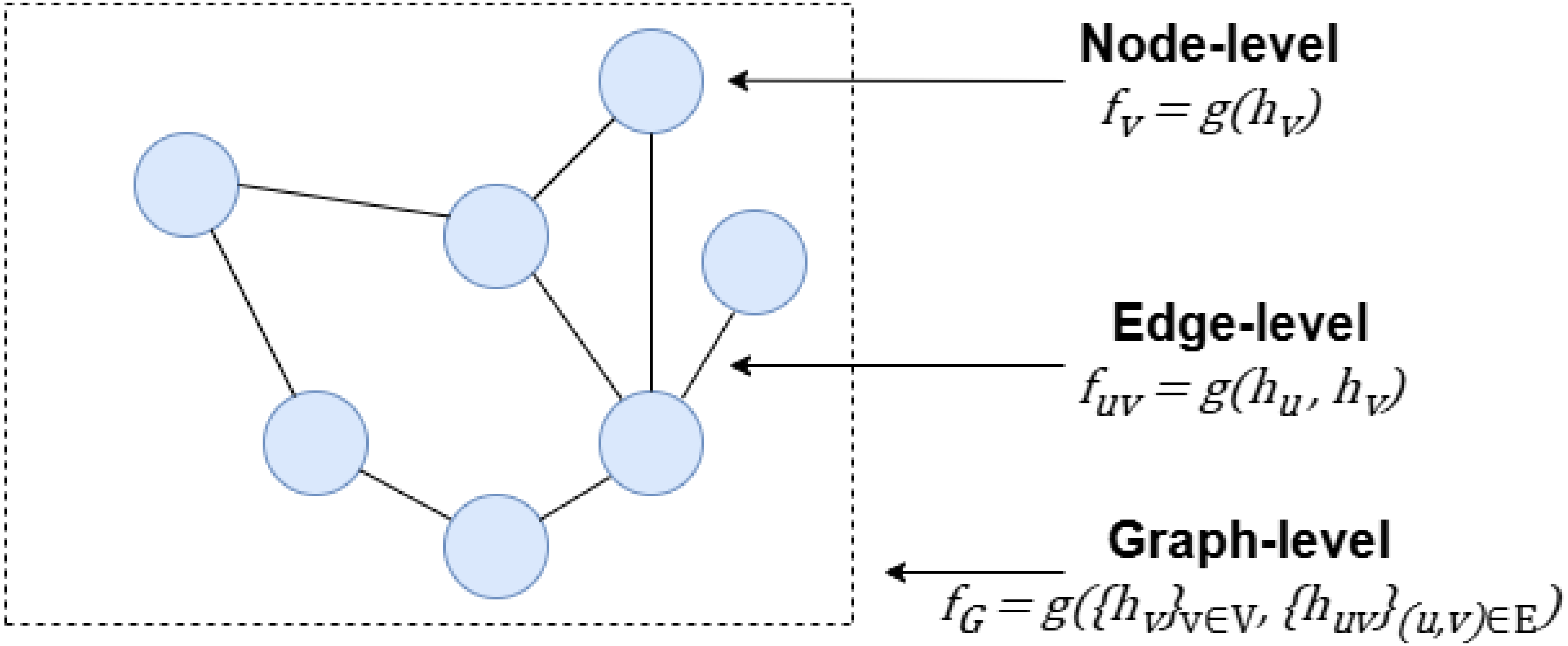
Graph Representation Learning

- Bridge between original input data and task objective in the graph
  - Learn embedded representation of nodes or entire graph from the input
  - Apply embedded representations to downstream related tasks
    - Node classification, graph classification, link prediction, etc.
- Node embedding
  - $H \in \mathbb{R}^{NxD}$

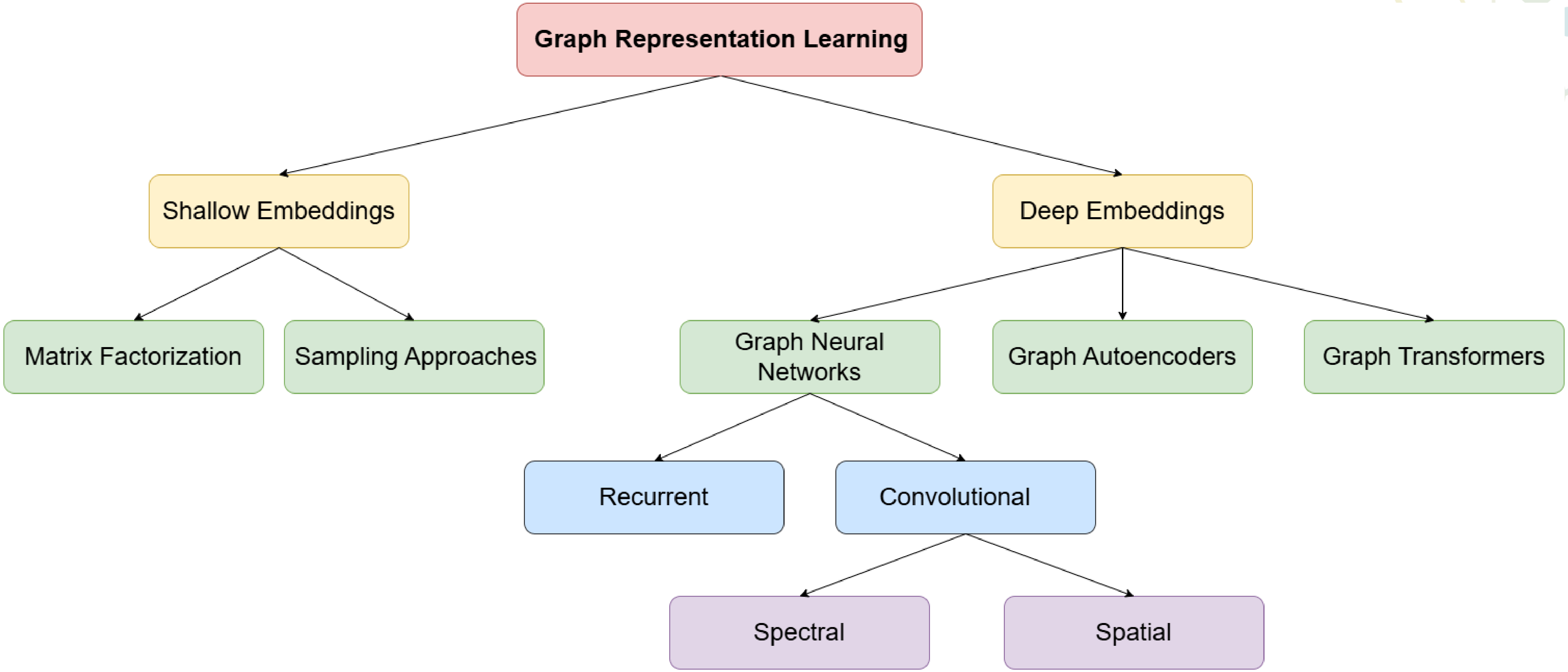


$$
\begin{array}{c|cccc}
 & 1 & 2 & \cdots & d \\
\hline
1 & a_{11} & a_{12} & \cdots & a_{1d} \\
2 & a_{21} & a_{22} & \cdots & a_{2d} \\
3 & a_{31} & a_{32} & \cdots & a_{3d} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
n & a_{n1} & a_{n2} & \cdots & a_{nd}
\end{array}
$$

Graph Representation Learning



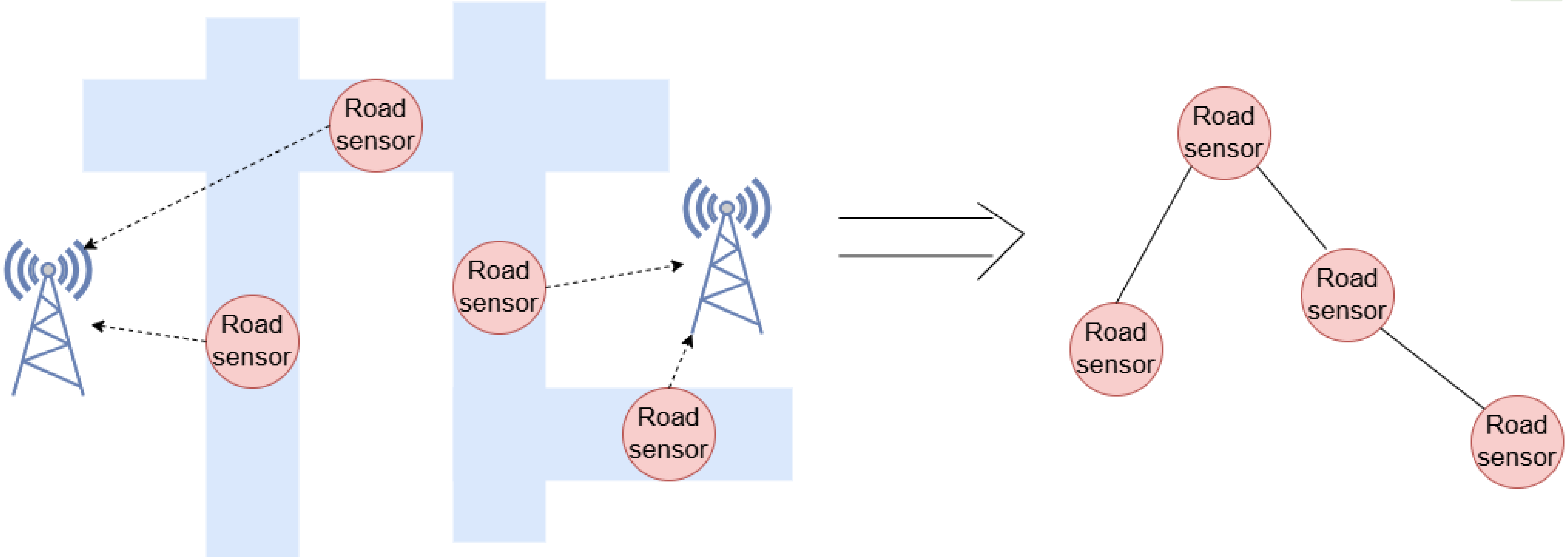
GRL Methods



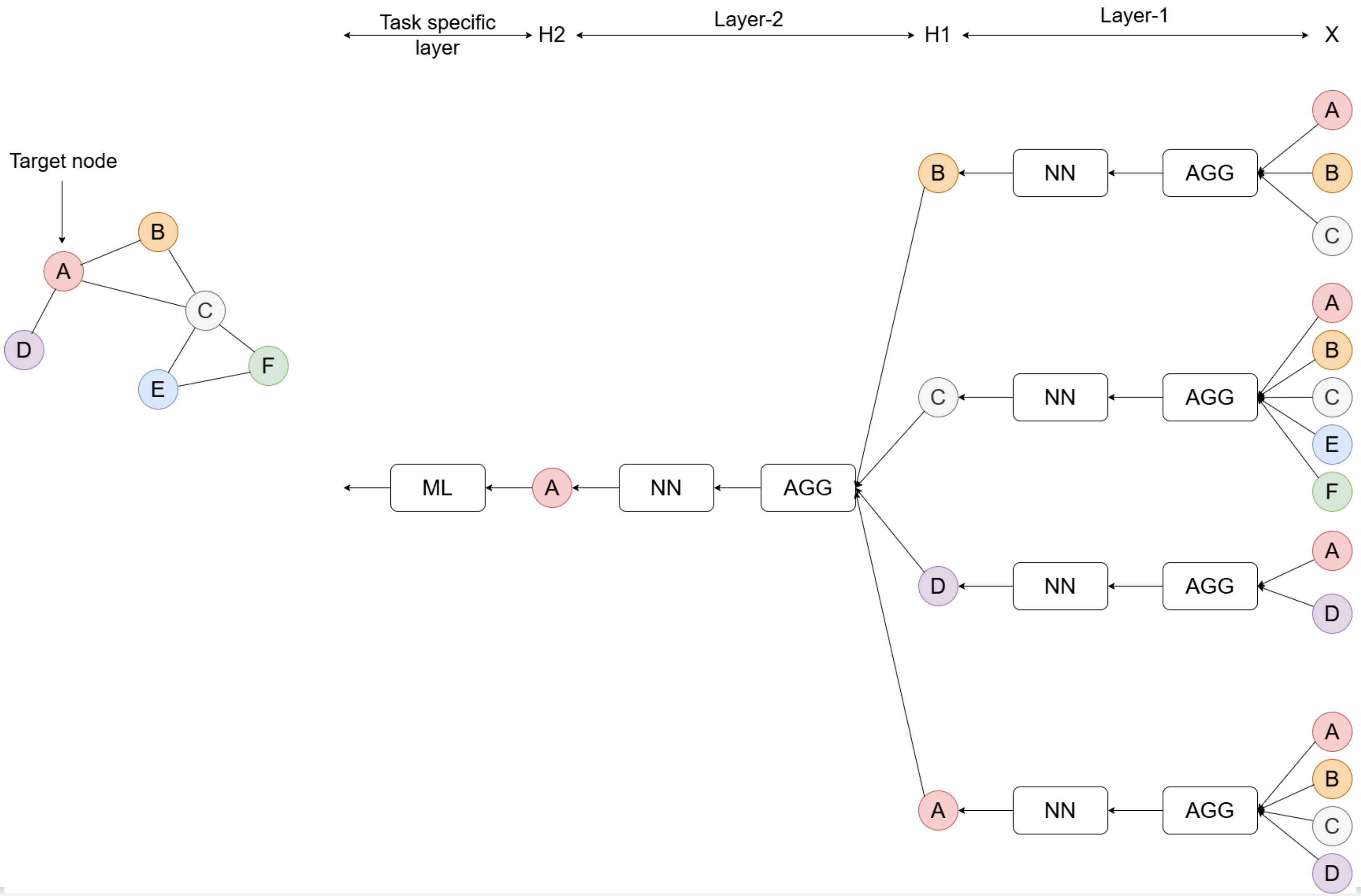
Graph Representation Learning
Shallow Embeddings
Deep Embeddings
Matrix Factorization
Sampling Approaches
Graph Neural Networks
Graph Autoencoders
Graph Transformers
Recurrent
Convolutional
Spectral
Spatial

Graph Neural Networks (GNNs)

- A type of neural network for graph-structured data



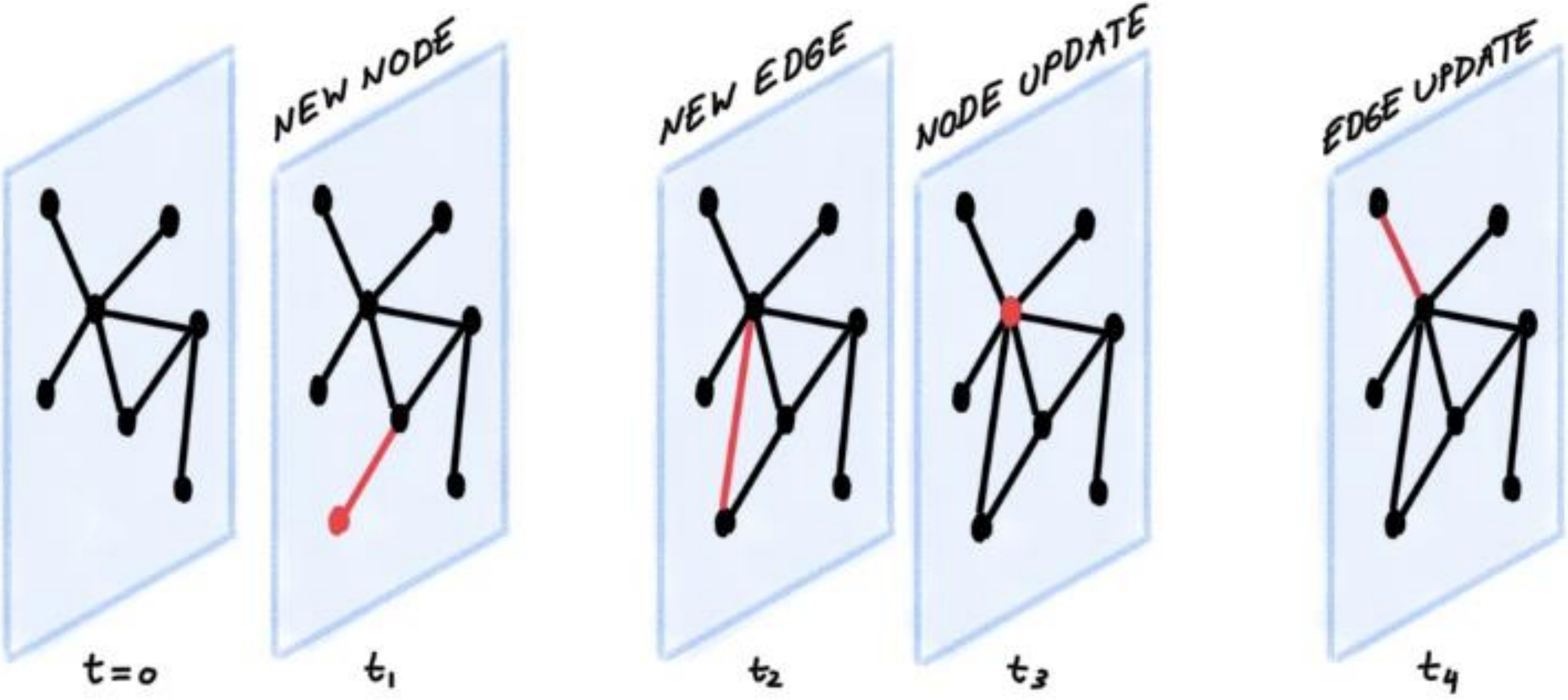
Graph Neural Networks (GNNs)



Graph convolutions…

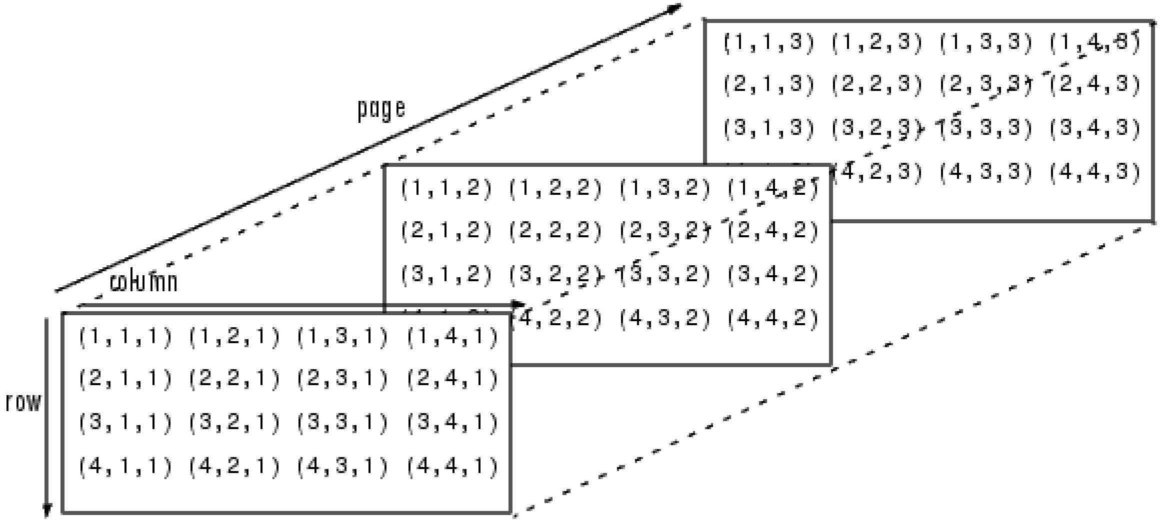
- Graph Convolutional Networks (GCN)
- Chebyshev Graph Convolutional Neural Networks (ChebConv)
- Graph Sample and Aggregation (GraphSAGE)

However…



Spatio-Temporal Graph Neural Networks

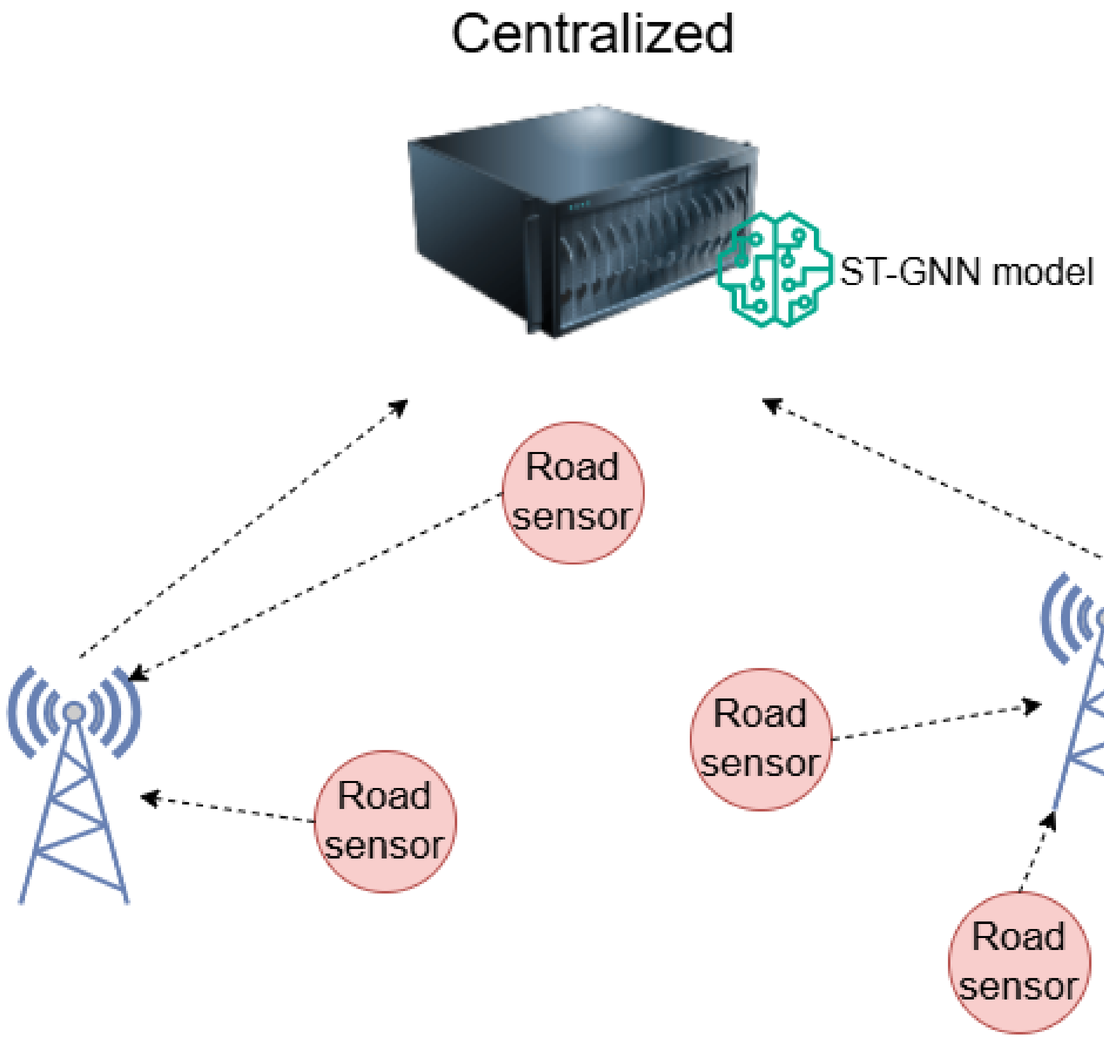
- $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t, W)$
  - $\mathcal{V}_t$ - set of verteces at time step t
  - $\mathcal{E}_t$– set of edges at time step t
  - $W \in \mathbb{R}^{n\ x\ n}$ - weighted adjacency matrix



Spatio-Temporal Graph Neural Networks

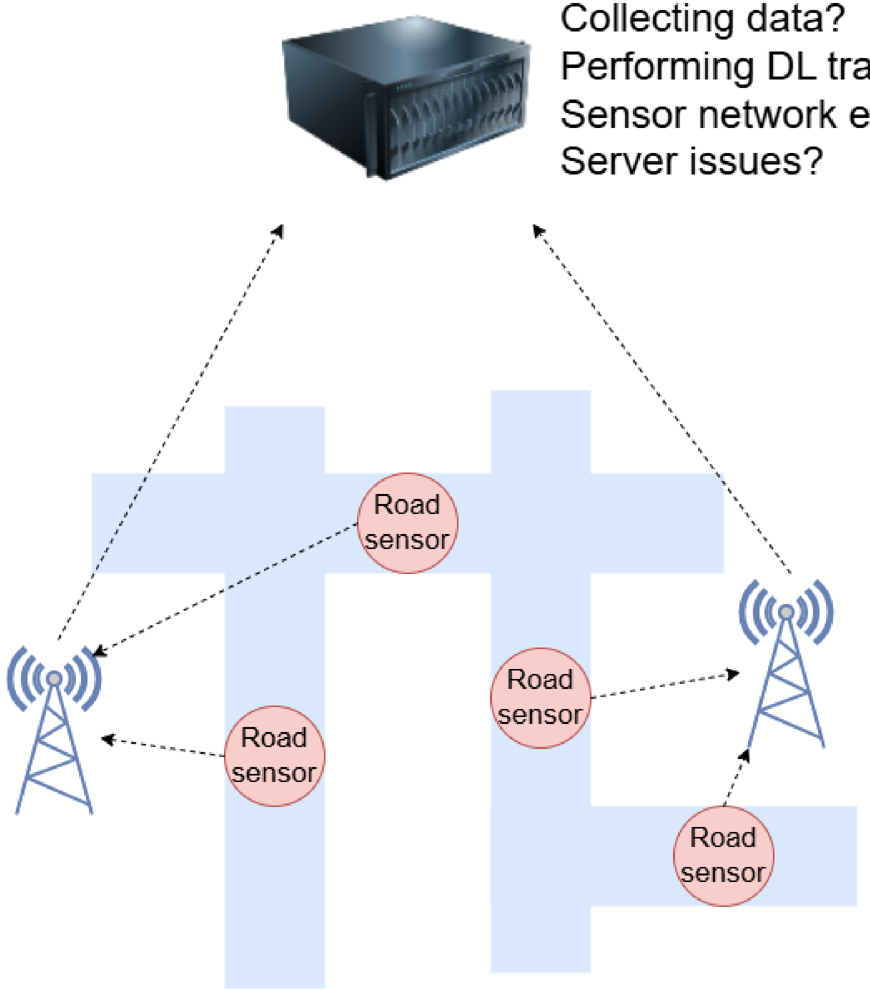
- **Spatio-Temporal Graph Convolutional Network (ST-GCN)**
- Diffusion Convolutional Recurrent Neural Network (DCRNN)
- Spatio-Temporal Autoregressive Model (ST-AR)

Centralized



Limitations

Collecting data?
Performing DL training & inference?
Sensor network expands?
Server issues?

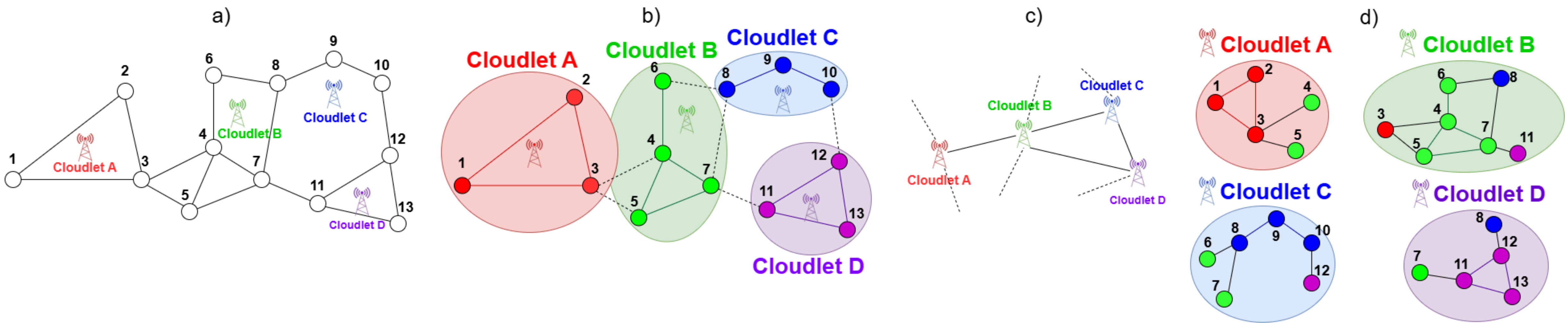


Previous works

- M. Nazzal et. al. [1]
  - Enabled semi-decentralized ***inference*** by deploying a pre-existing model, with ***no support for semi-decentralized training***
- L. Giaretta et. al. & R. Olshevskyi et. al. [2, 3]
  - Explored the challenges and solutions associated with fully decentralized training of GNNs, with their solution ***not tailored to ST-GNNs***

Semi-decentralized training of ST-GNNs

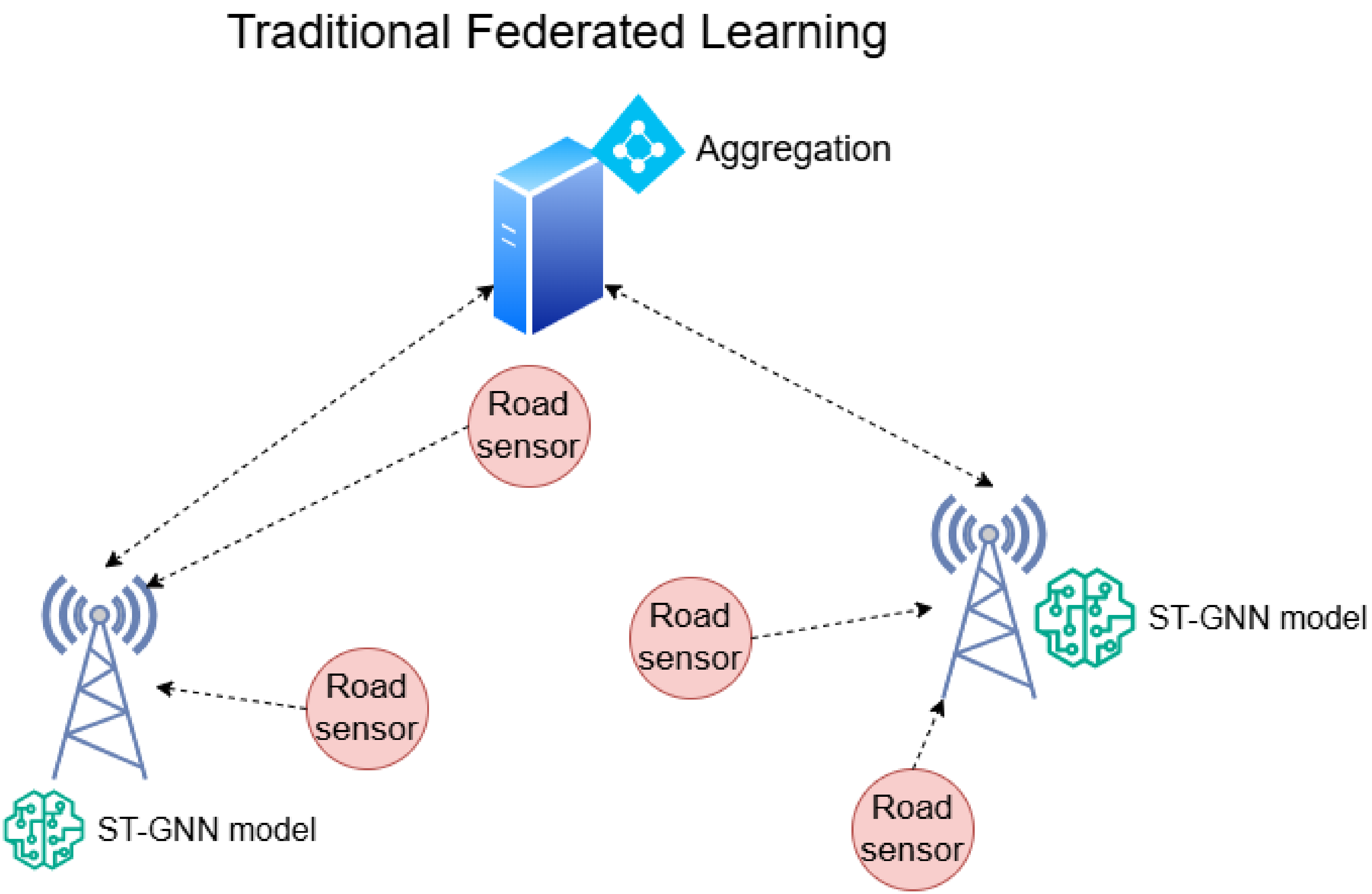
- Semi-decentralized ST-GNN architecture inspired by Nazzal et. al. [1]



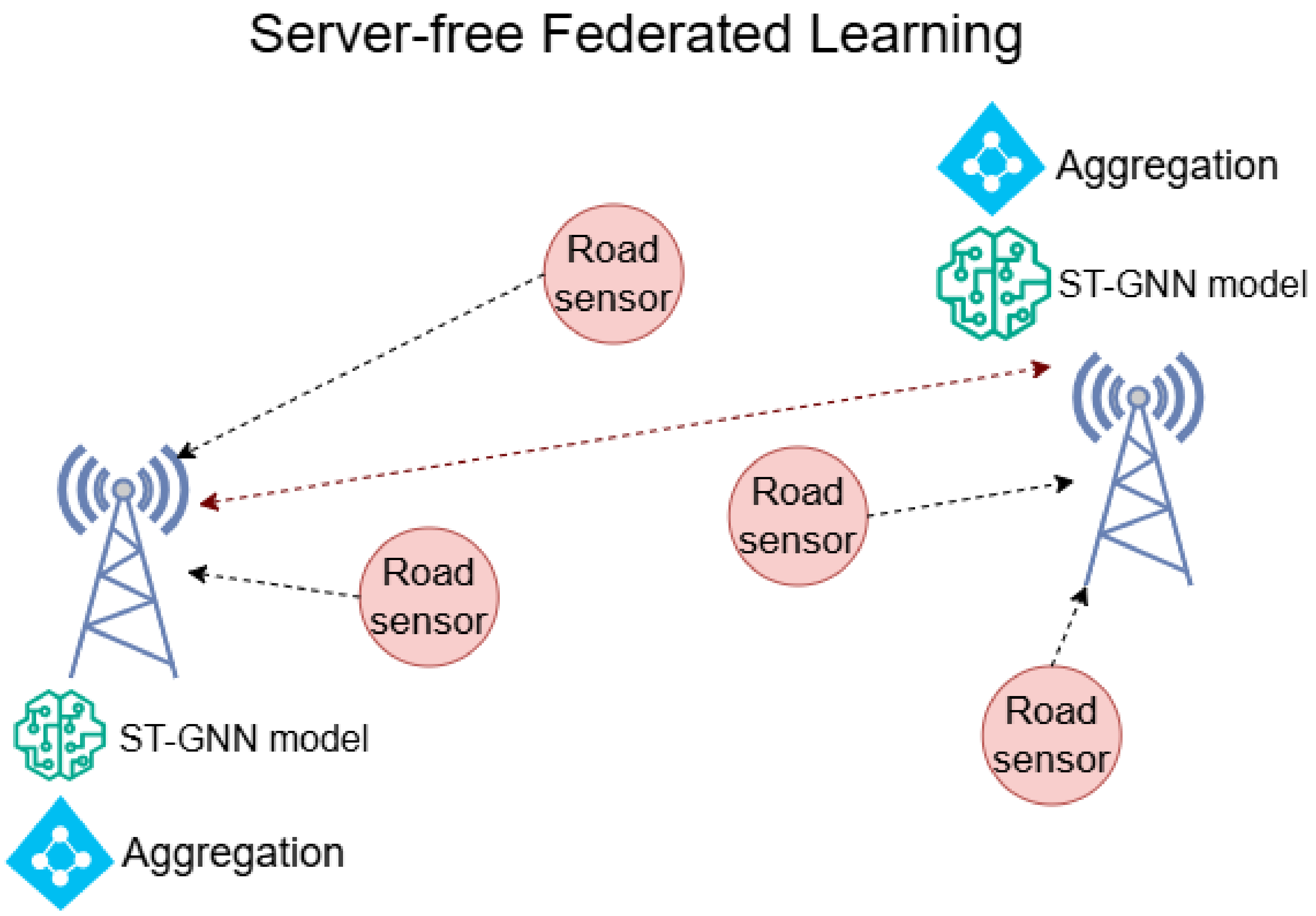
Training approaches

- Centralized - Base
- Traditional Federated Learning
- Server-free Federated Learning
- Gossip Learning

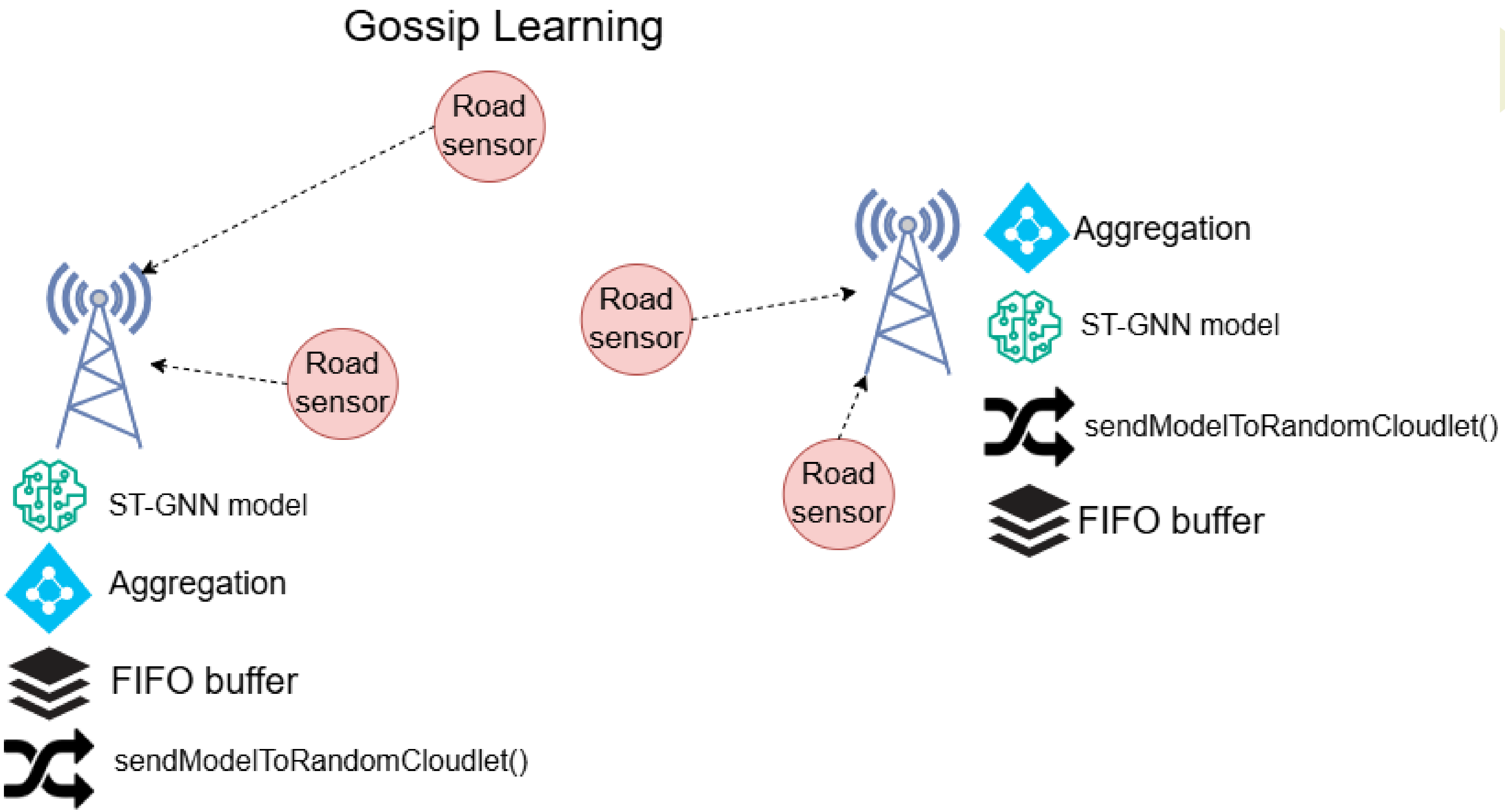
Traditional Federated Learning



Server-free Federated Learning



Gossip Learning



Testbed

EuroHPC
Joint Undertaking

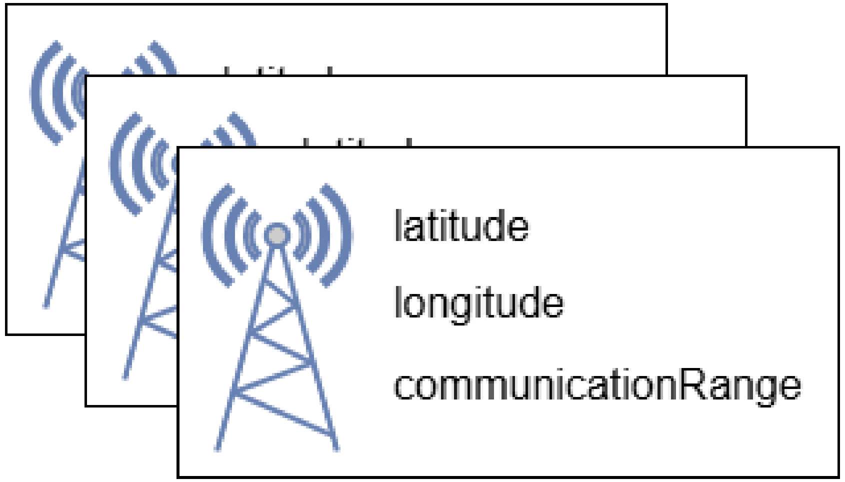
LEONARDO
CINECA

srce | Sveučilište u Zagrebu
Sveučilišni računski centar

Simulator

PyTorch

RAY



...

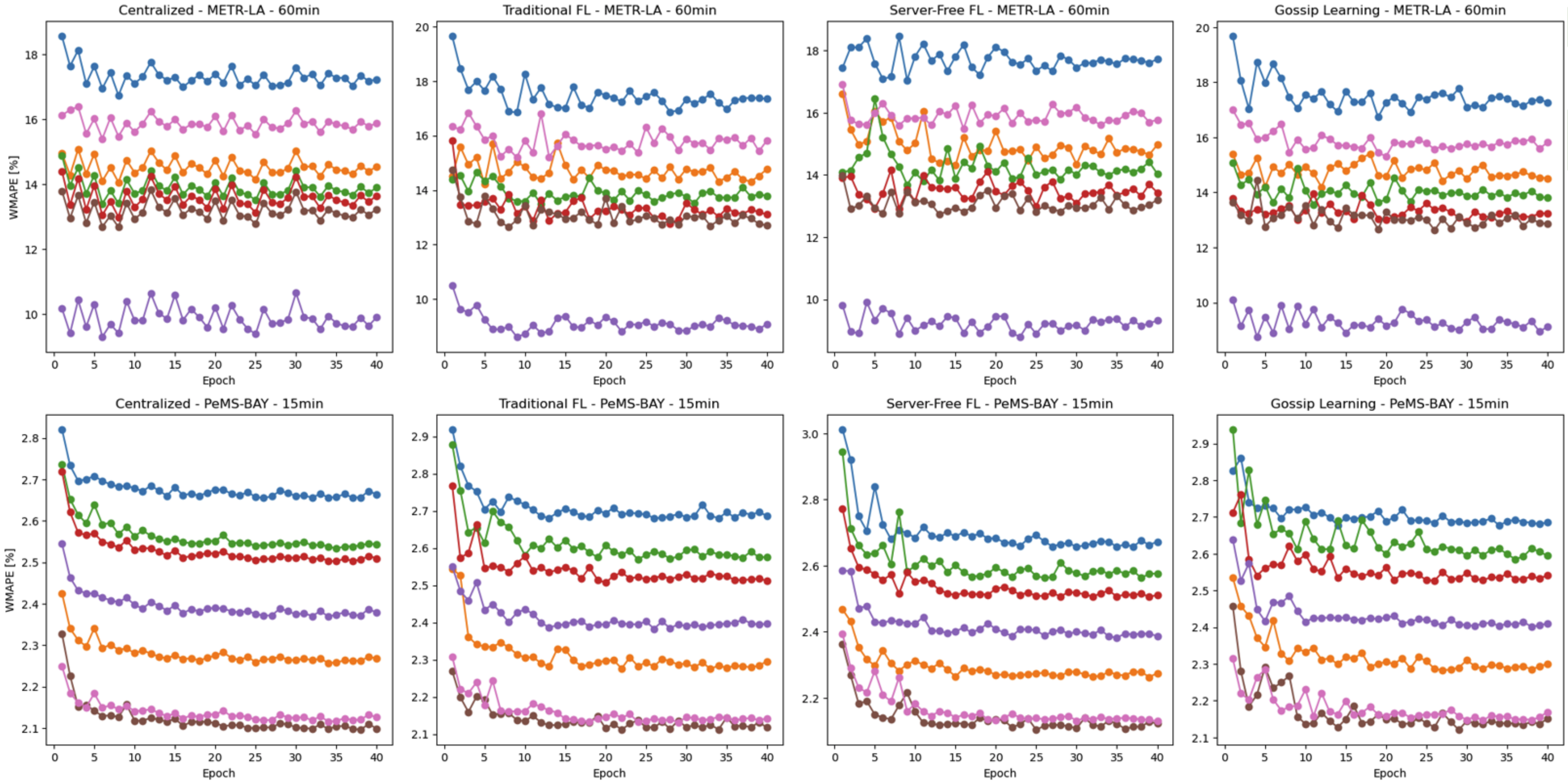
Experimental setup

- METR-LA & PeMS-BAY
- 3 horizons
- Evaluation metrics
- Sensors are distributed across 7 cloudlets based on proximity and communication range (8 km)

Results

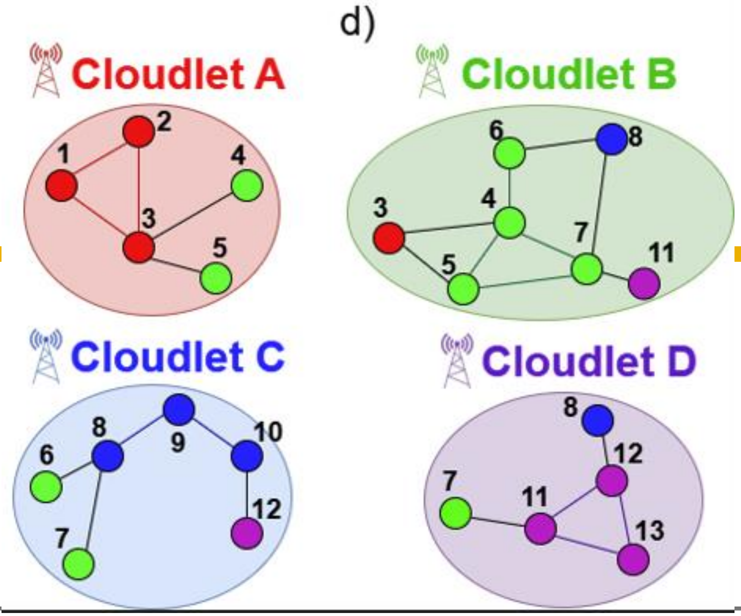
Dataset	Setups	15 min			30 min			60 min		
		MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE
METR-LA	Centralized	**3.78**	9.05	**7.45**	**5.14**	11.61	**10.12**	**7.35**	14.59	**14.47**
	Traditional FL	3.97	9.01	7.82	5.40	11.41	10.63	7.82	14.48	15.40
	Server-free FL	3.90	**8.98**	7.78	5.35	11.37	10.67	7.79	**14.41**	15.55
	Gossip Learning	3.88	9.04	7.74	5.43	**11.35**	10.85	7.56	14.42	15.10
PeMS-BAY	Centralized	**1.48**	**3.09**	**2.38**	**1.97**	**4.23**	**3.17**	**2.59**	5.41	**4.17**
	Traditional FL	1.50	3.12	2.42	2.03	4.29	3.27	2.67	5.51	4.29
	Server-free FL	1.50	3.12	2.42	2.01	4.25	3.23	2.65	5.44	4.28
	Gossip Learning	1.51	3.12	2.43	2.03	4.28	3.26	2.63	**5.39**	4.24

Results



Centralized - METR-LA - 60min
Traditional FL - METR-LA - 60min
Server-Free FL - METR-LA - 60min
Gossip Learning - METR-LA - 60min
Centralized - PeMS-BAY - 15min
Traditional FL - PeMS-BAY - 15min
Server-Free FL - PeMS-BAY - 15min
Gossip Learning - PeMS-BAY - 15min
WMAPE [%]
Epoch

Results



Dataset	Setups	Model [MB]/epoch	Training FLOPs/epoch	Aggregation FLOPs/epoch	Node feature [MB]
METR-LA	Centralized	-	1.68T	-	4.76
	Traditional FL	0.91 / 7	10.92T / 7	1.12M / 7	25.83 / 7
	Server-free FL	1.61 / 7	10.92T / 7	3.64M / 7	25.83 / 7
	Gossip Learning	0.91 / 7	10.92T / 7	2.94M / 7	25.83 / 7
PeMS-BAY	Centralized	-	4.06T	-	11.27
	Traditional FL	0.91 / 7	27.23T / 7	1.12M / 7	66.08 / 7
	Server-free FL	1.82 / 7	27.23T / 7	3.92M / 7	66.08 / 7
	Gossip Learning	0.91 / 7	27.23T / 7	2.94M / 7	66.08 / 7

Conclusion and future work

- Semi-decentralized setups are comparable to centralized approaches in performance metrics
- Highlight overlooked issues in existing literature for distributed ST-GNNs
  - Variation in model performance across different geographical areas
  - Communication overhead and computational cost due to large receptive field of GNNs

- Future work – focus on reducing communication overhead
  - Remove duplicated nodes and features

Literature

1. M. Nazzal, A. Khreishah, J. Lee, S. Angizi, A. Al-Fuqaha, M. Guizani, Semi-decentralized inference in heterogeneous graph neural networks for traffic demand forecasting: An edge-computing approach, IEEE Transactions on Vehicular Technology (2024).
2. L. Giaretta, S. Girdzijauskas, Fully-decentralized training of gnns using layer-wise self-supervision (2023).
3. Olshevskyi, Rostyslav, Zhongyuan Zhao, Kevin Chan, Gunjan Verma, Ananthram Swami, and Santiago Segarra. "Fully Distributed Online Training of Graph Neural Networks in Networked Systems." arXiv preprint arXiv:2412.06105 (2024).
4. Kralj, Ivan, et al. "Semi-decentralized Training of Spatio-Temporal Graph Neural Networks for Traffic Prediction." *arXiv preprint arXiv:2412.03188* (2024).

Training of GNNs

- Forward & backward pass
  - Goal: Update weights and gradients
- Loss function
  - Compare predicted values with ground truth
- Optimization algorithm
  - Minimize error in loss function