Graph Neural Networking challenge 2022:

Improving Network Digital Twins through Data-centric AI

https://bnn.upc.edu/challenge/gnnet2022

José Suárez-Varela

Barcelona Neural Networking center

Universitat Politècnica de Catalunya

May 27th 2022

What is the Graph Neural Networking challenge?

Graph Neural Networking challenge

https://bnn.upc.edu/challenge/gnnet2022

- **Series of annual competitions on Graph Neural Networks applied to Networking**

- Each edition brings a fundamental challenge on GNNs applied to Computer Networks:
  - Graph Neural Networking challenge 2020: Modeling QoS-aware queue scheduling policies at networks
  - Graph Neural Networking challenge 2021: Creating a Scalable Network Digital Twin

Graph Neural Networks are becoming a hot topic in networking!

It is the first (and the only) competition on GNNs applied to computer networks

ITU AI/ML in 5G challenge

https://aiforgood.itu.int/about/aiml-in-5g-challenge

- Organized as part of the **ITU AI/ML in 5G challenge**

- Several problem statements on AI/ML applied to networks, one of them is the Graph Neural Networking challenge

Graph Neural Networking challenge 2022

Problem statement:

Improving Network Digital Twins through Data-centric AI

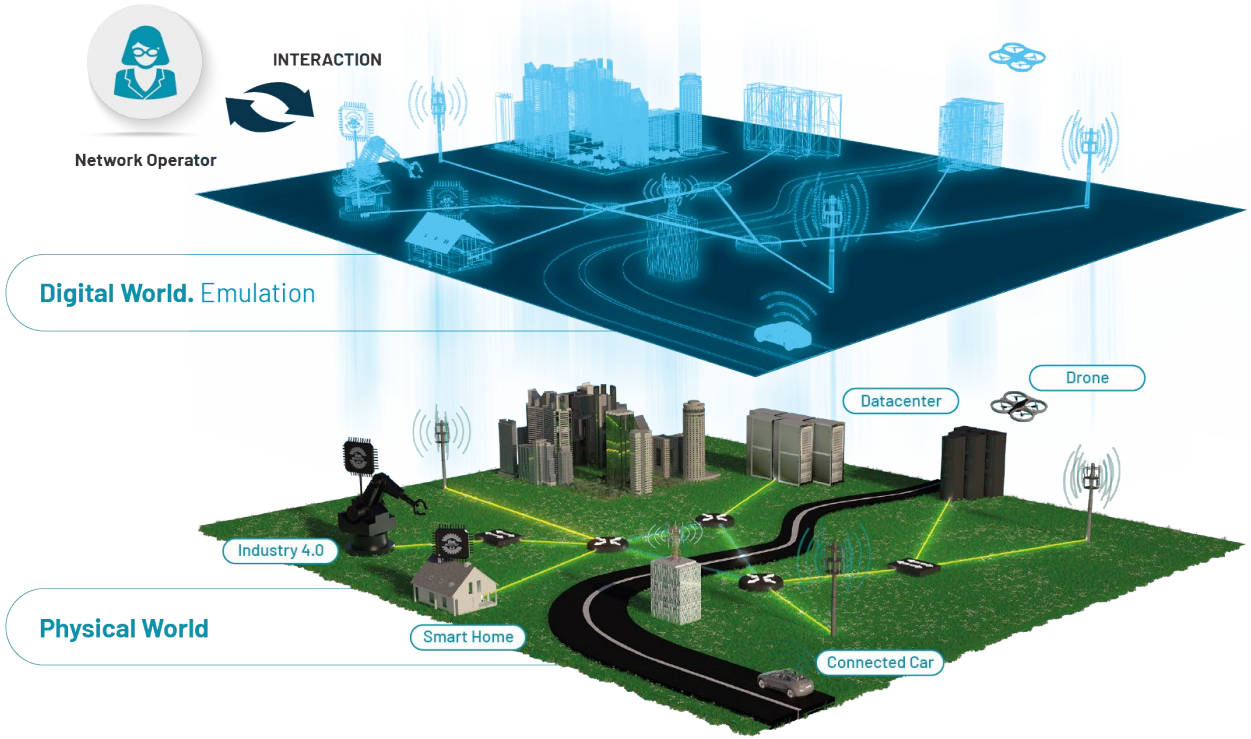
What is a Network Digital Twin?

What is a Digital Twin?

- A digital twin is a virtual replica of a physical object or process

- It permits to simulate the behavior of a physical system under certain input conditions:
  - What will happen if there is a specific failure? (e.g., in the electrical system)
  - What happen if I make a change in the object? (e.g., new wing design)

What is a Network Digital Twin?



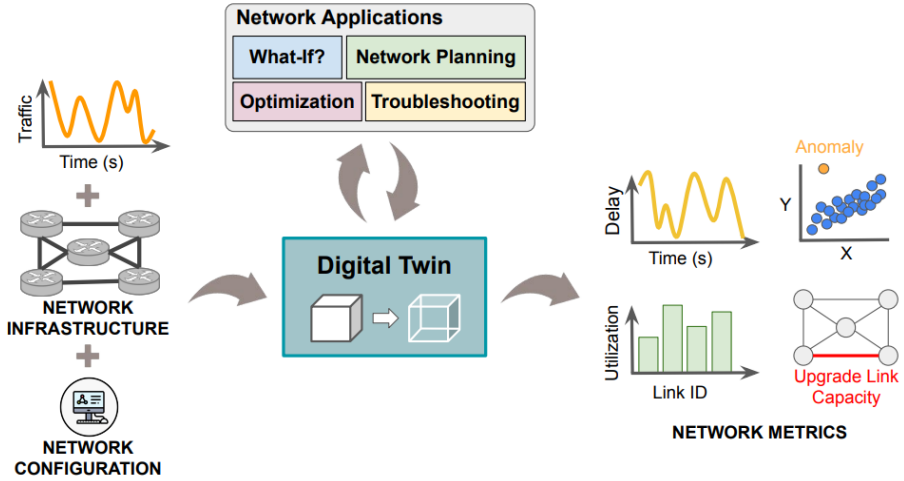
- A Network Digital Twin is a virtual replica of a physical network
- It enables to reproduce the behavior of the network under certain what-if scenarios:
  - What happens if I change the configuration?
  - What happens if there is a random failure?

Network Digital Twin: Use case examples

Digital Twins can be applied to many fundamental networking applications*

Network Optimization and What-if analysis

- What happens if we re-route traffic on another path? (Traffic Engineering)
- Can I support new user SLAs with the current resources?



Network Planning:

- Which is the best network upgrade within a limited budget?

Troubleshooting:

There was a temporary service disruption that affected some SLAs:

- What was the root cause?
- Can we find a way to prevent this in the future? (*e.g.*, add link redundancy)

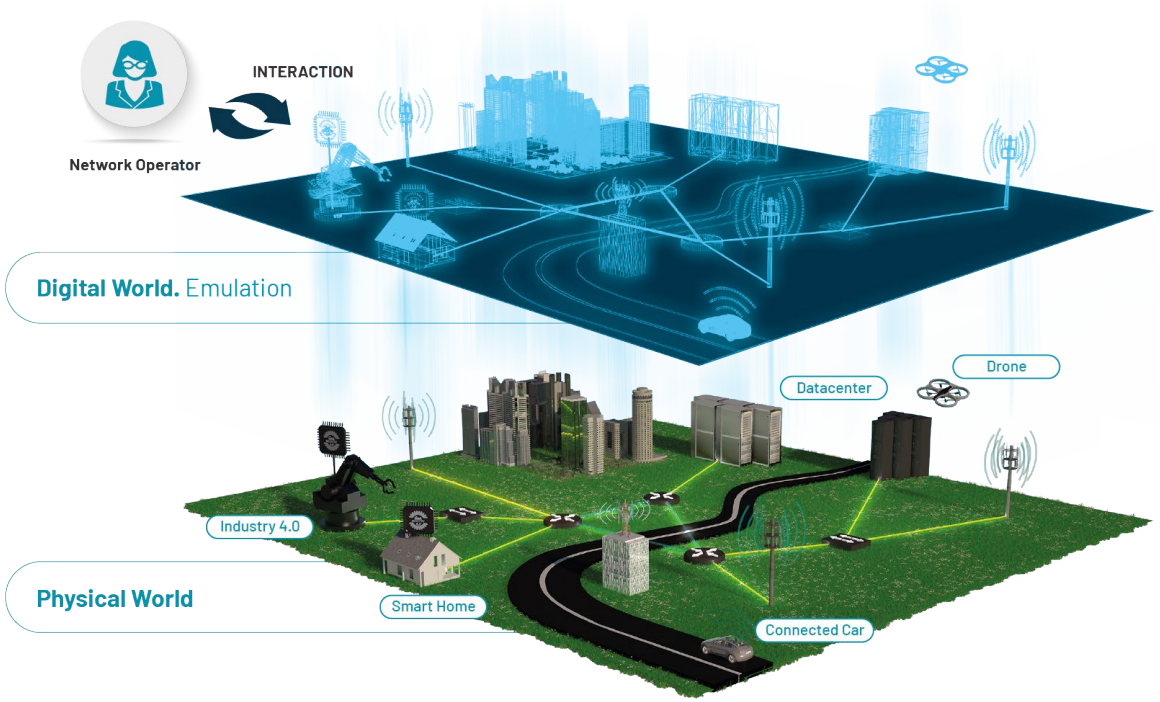
* P. Almasan, M. Ferriol-Galmés, J. Paillisse, J. Suárez-Varela, D. Perino, D. López, A. A. Pastor Perales, P. Harvey, L. Ciavaglia, L. Wong, V. Ram, S. Xiao, X. Shi, X. Cheng, A. Cabellos-Aparicio, P. Barlet-Ros, "Network Digital Twin: Context, Enabling Technologies and Opportunities," arXiv preprint arXiv:2201.01144, 2022

Network Digital Twin: Context

Is this a new concept?

What about the existing literature on network modeling?

(e.g., network simulators, analytical models)



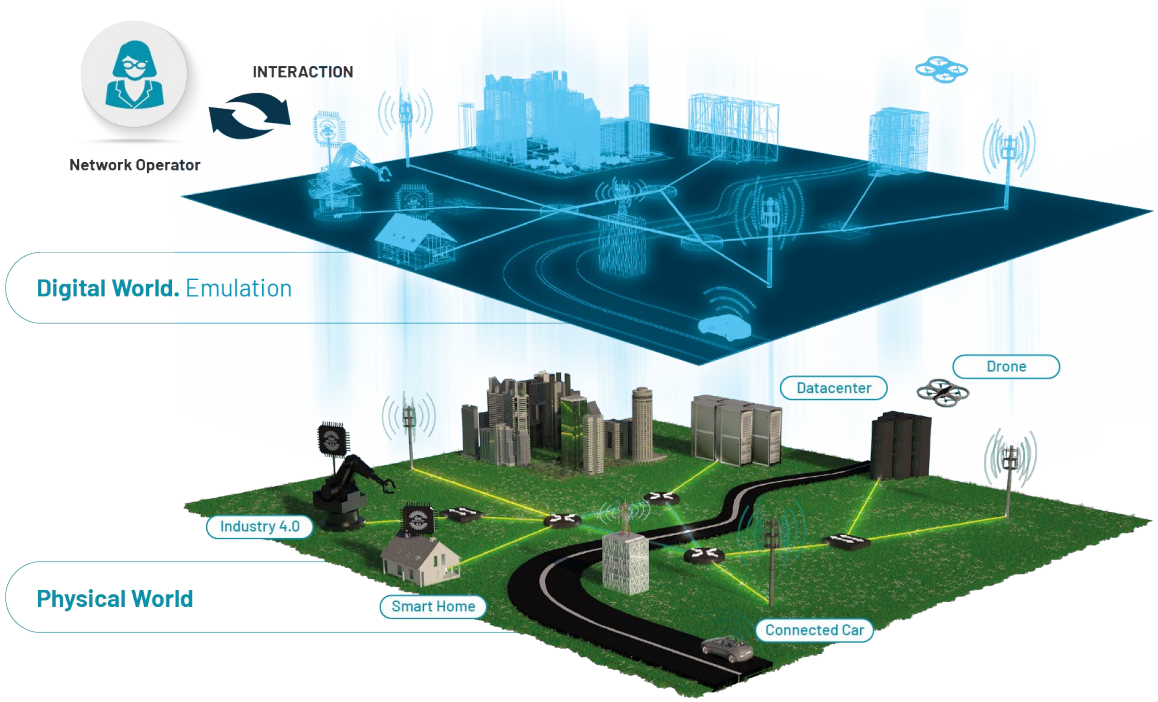
Network Digital Twin: Context

Is this a new concept?

What about the existing literature on network modeling?

(e.g., network simulators, analytical models)

- **Network Digital Twins**:
  - Renovated concept of classical network modeling with the ambition of achieving **accurate real-time digital replicas** of the network
  - **Machine learning (ML)** is promising for building **accurate and lightweight** data-driven network models

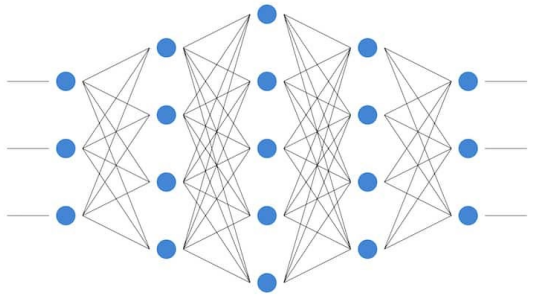


How can we build a ML-based Network Digital Twin?

How can we create a ML-based Network Digital Twin?

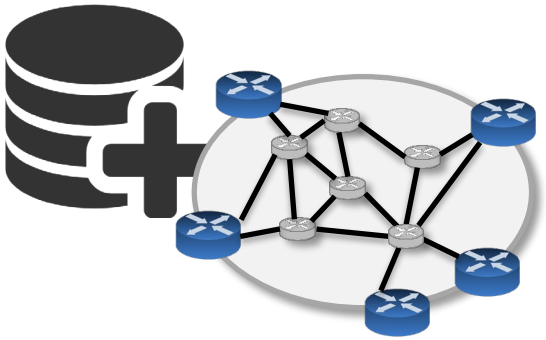
VENDOR'S NETWORKING LAB

ML model (e.g., Graph Neural Network)



+

Dataset



Network testbed

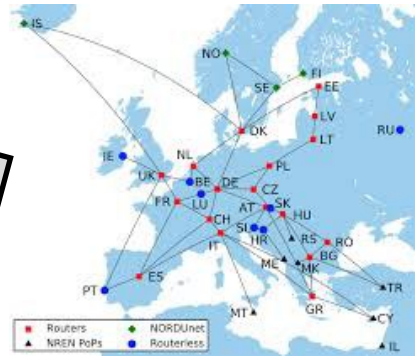
Training

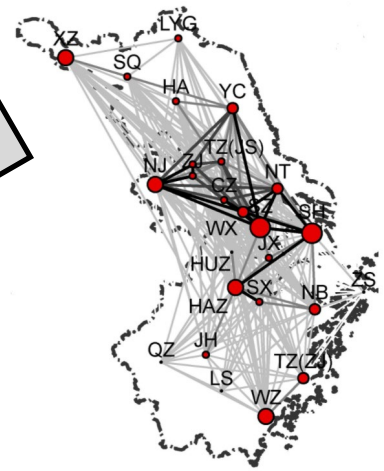
NETWORK DIGITAL TWIN

Deployment

Deployment

Customer's networks

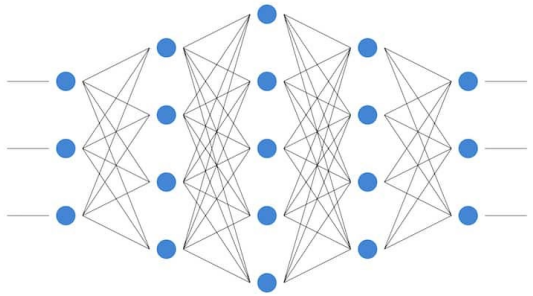




How can we create a ML-based Network Digital Twin?

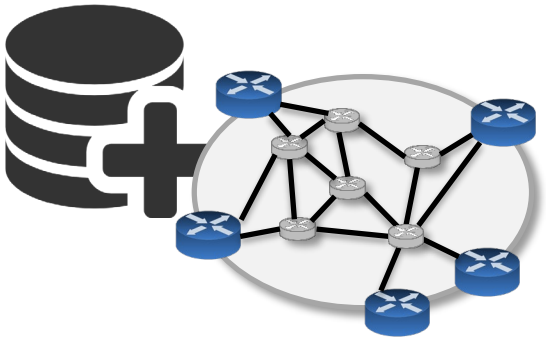
VENDOR'S NETWORKING LAB

ML model (e.g., Graph Neural Network)



+

Dataset



Network testbed

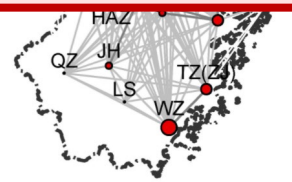
Training

NETWORK DIGITAL TWIN

Deployment

Deployn

Customer's networks

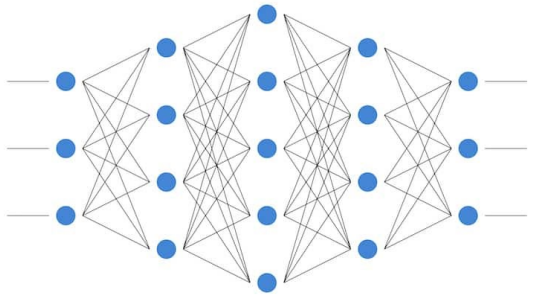


Training with data from real networks is not feasible, as it requires to cover edge cases (e.g., generate link failures, high congestion)

How can we create a ML-based Network Digital Twin?

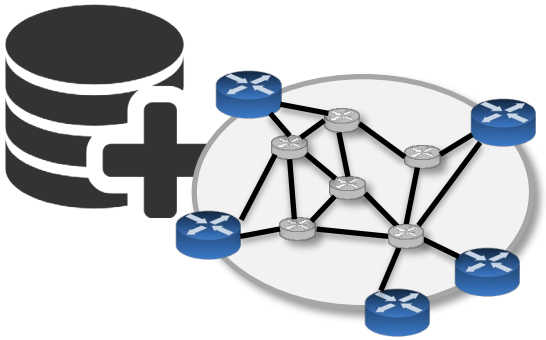
VENDOR'S NETWORKING LAB

ML model (e.g., Graph Neural Network)



+

Dataset



Network testbed

Training

NETWORK DIGITAL TWIN

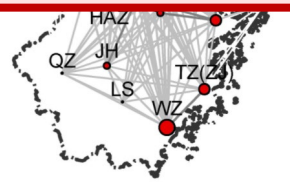
*Need for ML models that can **generalize to other networks** not seen during training*

Deployment

Deployn

Customer's networks

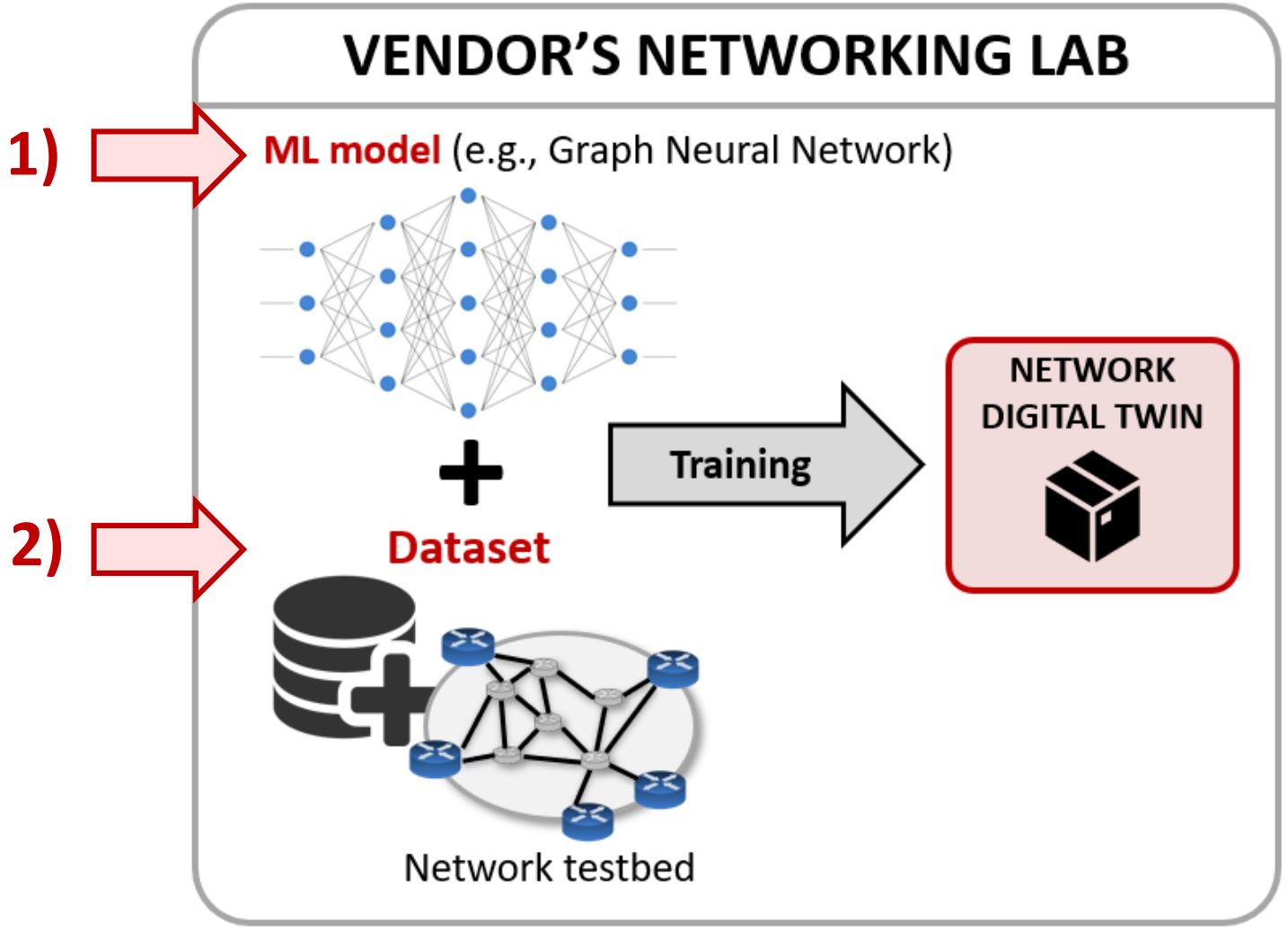
Training with data from real networks is not feasible, as it requires to cover edge cases (e.g., generate link failures, high congestion)



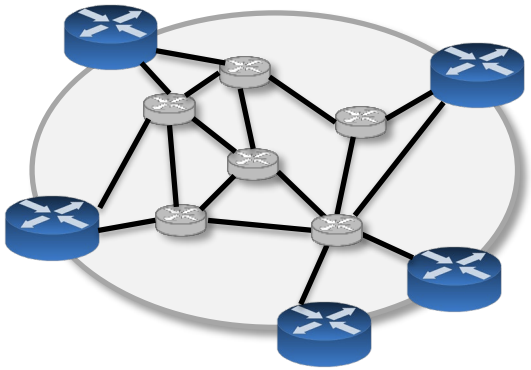
Let’s go into the vendor’s lab…

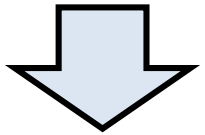
Creating the Network Digital Twin at the lab

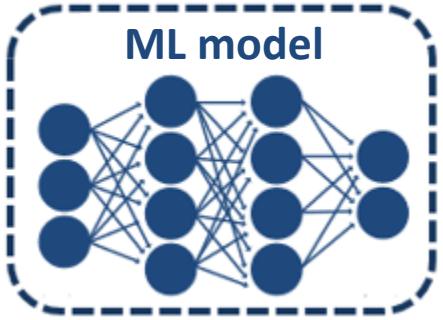
There are two fundamental components:



Creating the Network Digital Twin at the lab

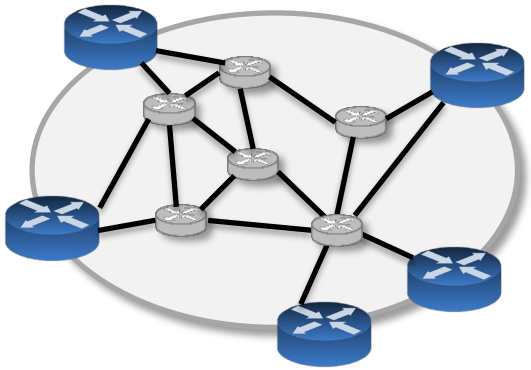


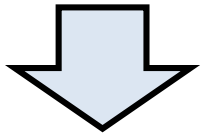


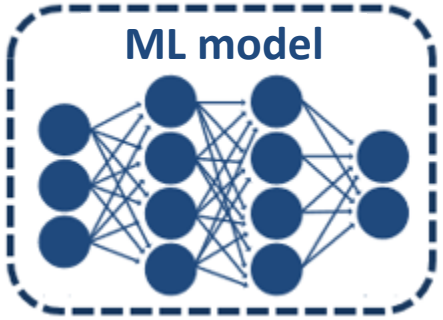


ML model:

Creating the Network Digital Twin at the lab



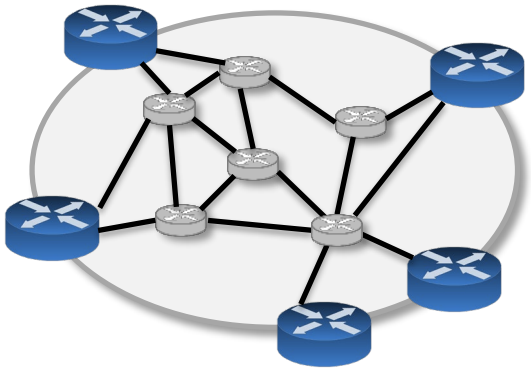


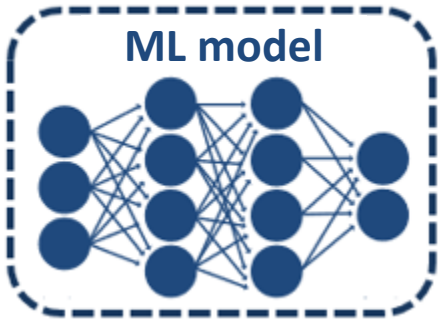


ML model:

Need for ML-based models with strong generalization capabilities over network data

Creating the Network Digital Twin at the lab



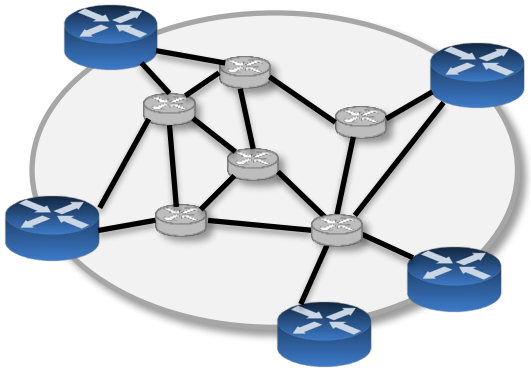


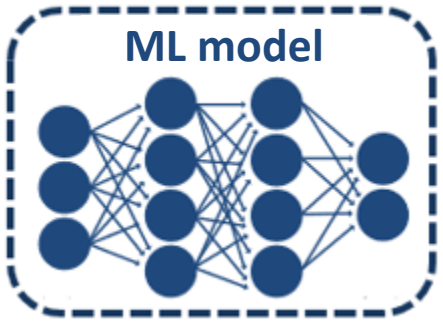
ML model:

Need for ML-based models with strong generalization capabilities over network data

The two first editions of the Graph Neural Networking challenge have focused on building Graph Neural Networks (GNN) that generalize over networks

Creating the Network Digital Twin at the lab





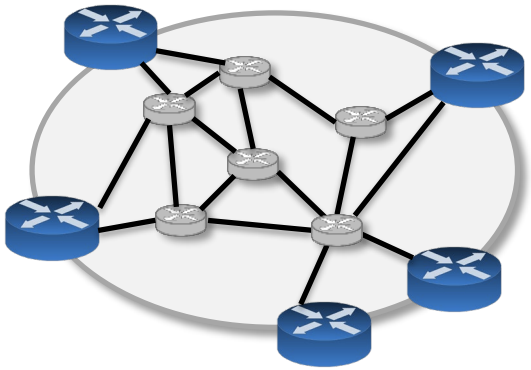
ML model:

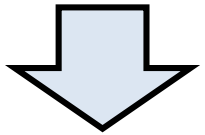
Need for ML-based models with strong generalization capabilities over network data

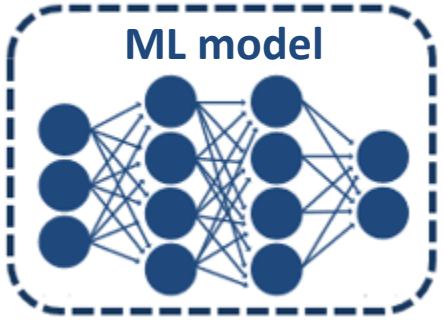
The two first editions of the Graph Neural Networking challenge have focused on building Graph Neural Networks (GNN) that generalize over networks

GNNs have already shown strong generalization capabilities over network data (e.g., scalability, routing, queue scheduling, traffic models)

Creating the Network Digital Twin at the lab



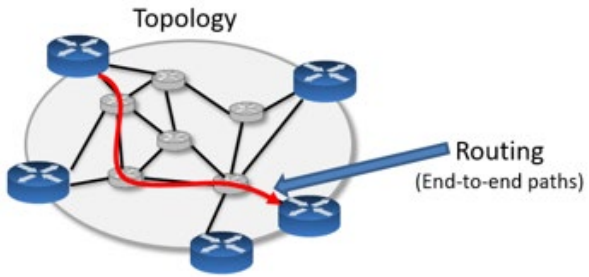




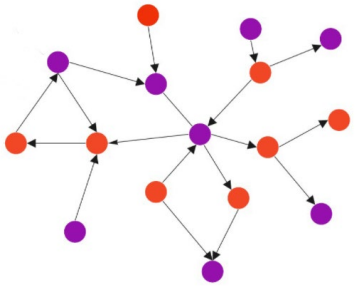
ML model:

- **Communication networks** comprise relational information at many different levels (*i.e.*, **graphs**):
  - Topology
  - Routing
  - User connections
  - Signal Interference
  - Flow inter-dependencies
  - …

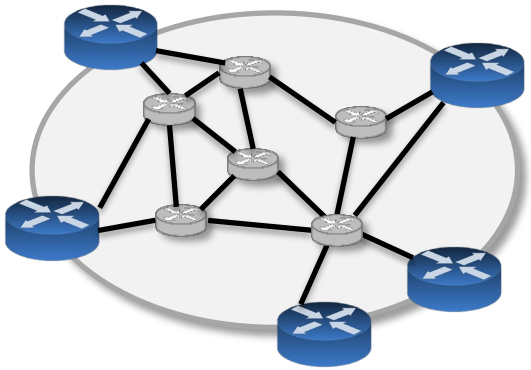
Network scenario

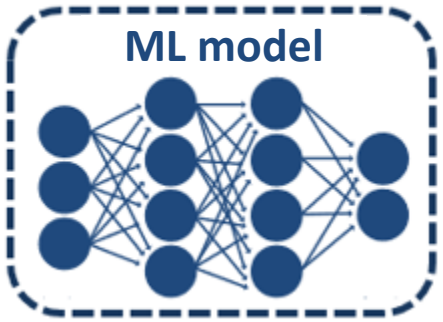


Graph representation



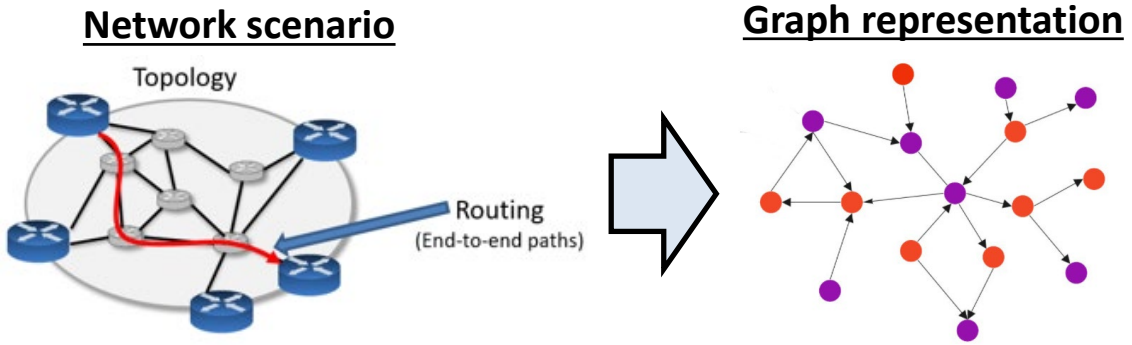
Creating the Network Digital Twin at the lab





ML model:

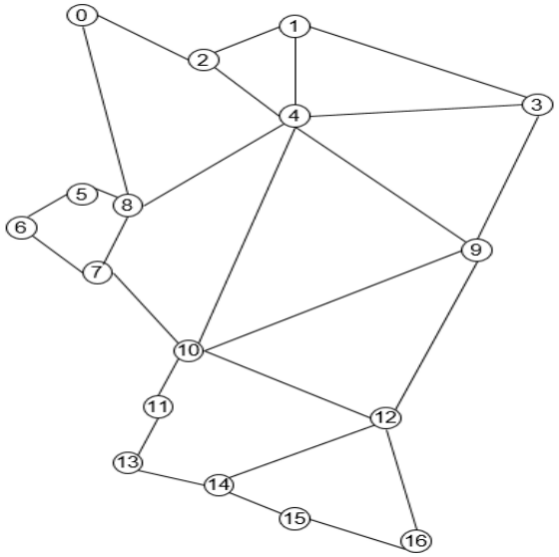
- **Communication networks** comprise relational information at many different levels (*i.e.*, **graphs**):
  - Topology
  - Routing
  - User connections
  - Signal Interference
  - Flow inter-dependencies
  - …



- **GNN is the most suitable ML technique to process, model and generalize over this graph-structured data**

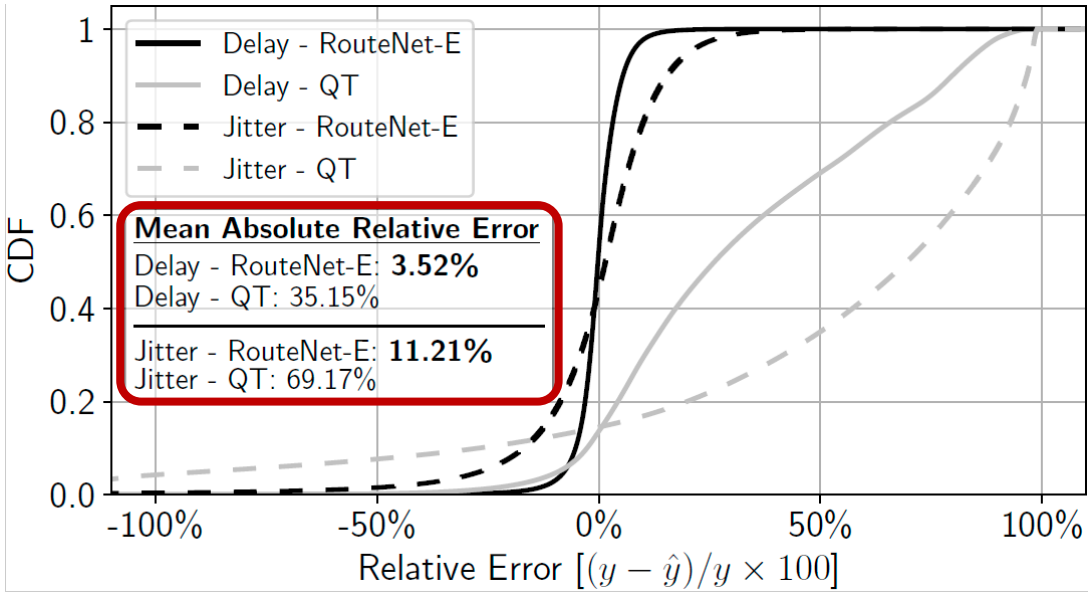
Evaluation: RouteNet-Erlang

- **RouteNet-Erlang** (IEEE INFOCOM 2022)*
- Evaluated in 50,000 samples of an **unseen topology** (GBN)



German Backbone Network

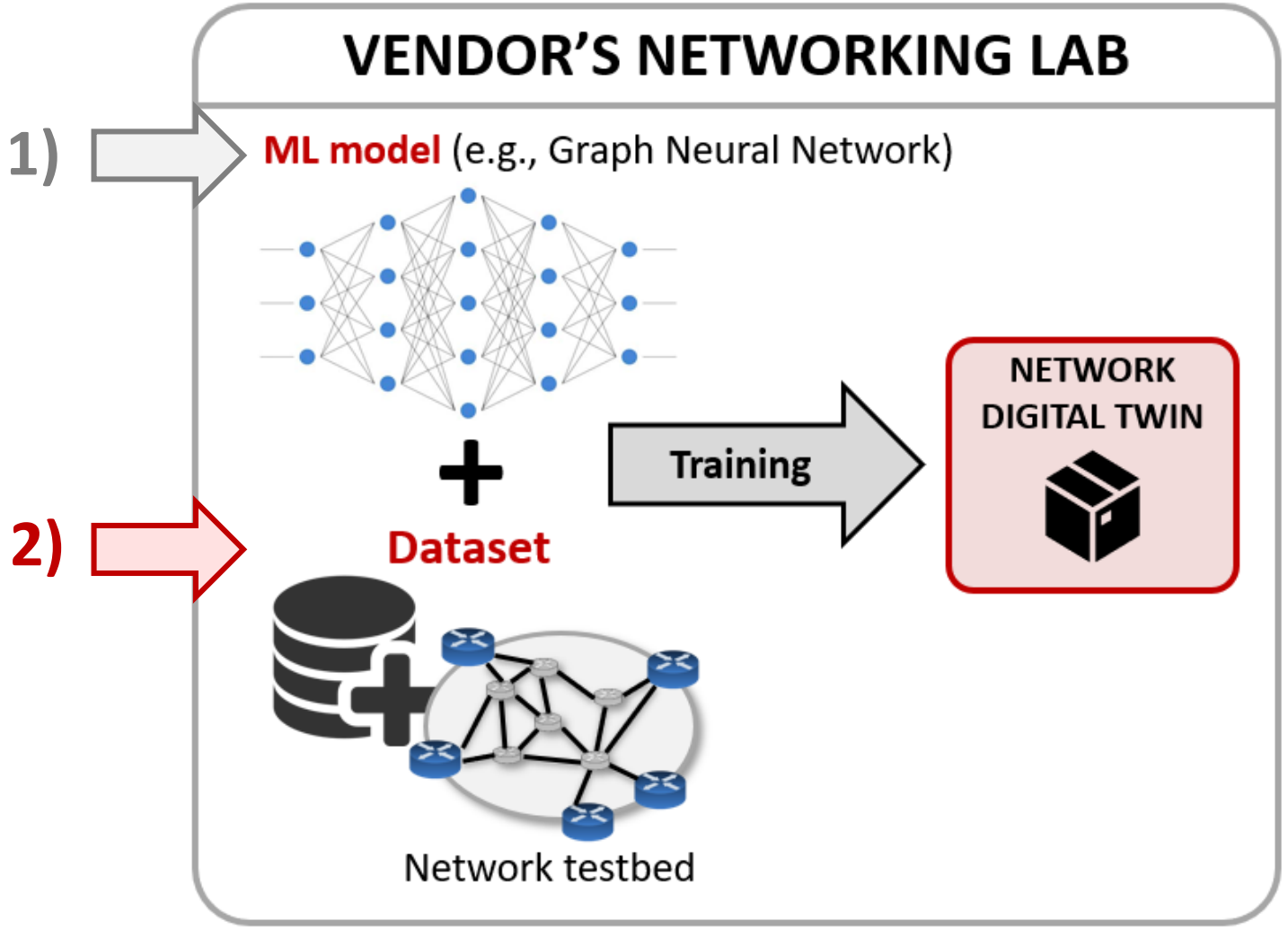
Mix of traffic models



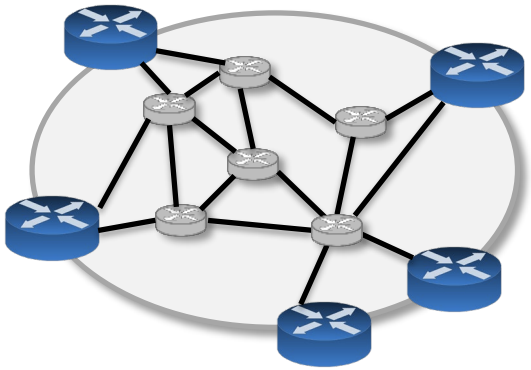
** M. Ferriol-Galmés, K. Rusek, J. Suárez-Varela, S. Xiao, X. Shi, X. Cheng, B. Wu, P. Barlet-Ros, A. Cabellos-Aparicio, "RouteNet-Erlang: A Graph Neural Network for Network Performance Evaluation", IEEE INFOCOM, 2022.*

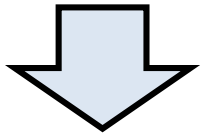
Creating the Network Digital Twin at the lab

There are two fundamental components:



Creating the Network Digital Twin at the lab



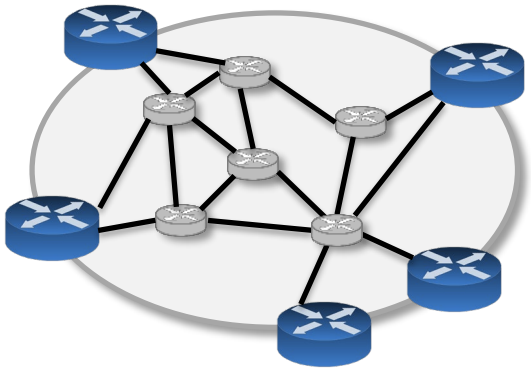


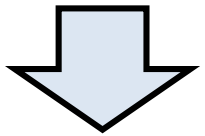
Dataset

Dataset:

There is no research on how to produce good datasets for ML models applied to networking!

Creating the Network Digital Twin at the lab





Dataset

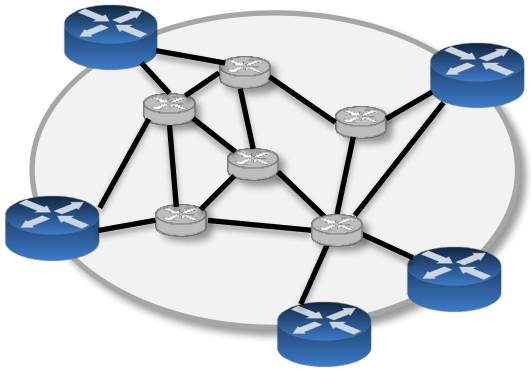
Dataset:

There is no research on how to produce good datasets for ML models applied to networking!

A good dataset requires…

- Domain expert knowledge → Understand what are relevant features to the ML model
- Good coverage of possible cases (e.g., congestion levels)
- Consider edge cases (e.g., failures)
- Avoid unambiguous labels (e.g., noise)
- Limited size! (cost of production)
- …

Creating the Network Digital Twin at the lab



Dataset

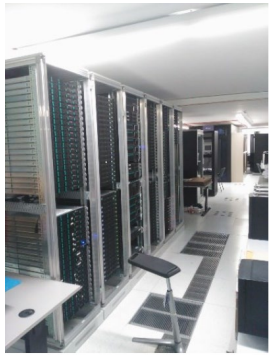
Dataset:

There is no research on how to produce good datasets for ML models applied to networking!

A good dataset requires…

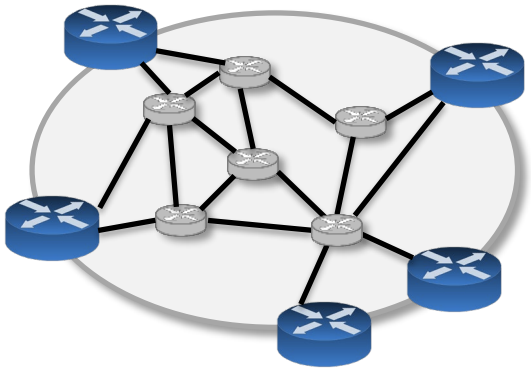
- Domain expert knowledge → Understand what are relevant features to the ML model
- Good coverage of possible cases (e.g., congestion levels)
- Consider edge cases (e.g., failures)
- Avoid unambiguous labels (e.g., noise)
- Limited size! (cost of production)
- …

In previous editions we have struggled to create datasets that had a good coverage of relevant cases
(e.g., 100k-400k training samples)



Costly simulations in our computing cluster!

Creating the Network Digital Twin at the lab

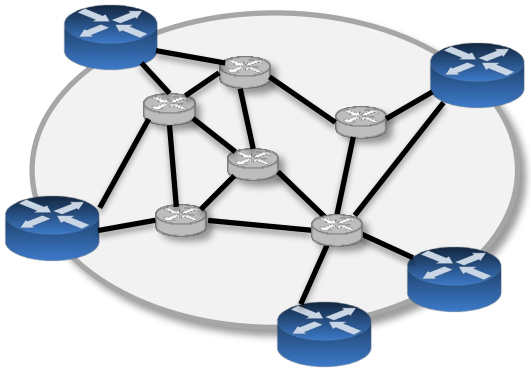


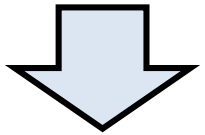
Dataset

Dataset:

It’s time to explore to Data-Centric AI for ML applied to networking

Creating the Network Digital Twin at the lab

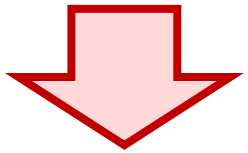




Dataset

Dataset:

It’s time to explore to Data-Centric AI for ML applied to networking



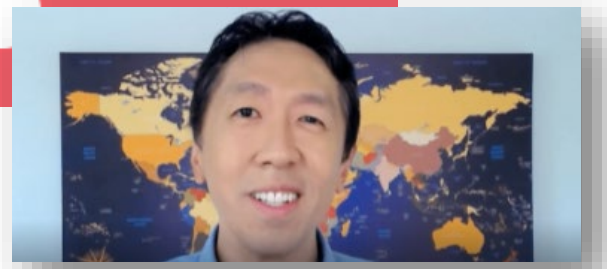
Potential benefits:

- **Large performance gains** (better coverage of important training samples)
- **Cost savings** (less training samples needed)

Data-Centric AI is a hot topic in the ML field!

E.g., Data-Centric AI Competition 2021 (by Andrew Ng)

https://https-deeplearning-ai.github.io/data-centric-comp/



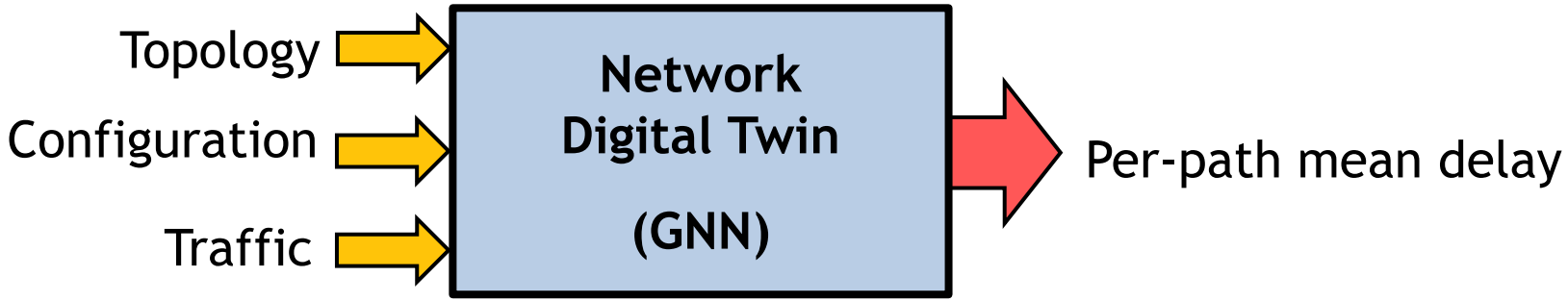
Computer vision
(handwritten digits)

Graph Neural Networking challenge 2022

Problem statement:

Improving Network Digital Twins through Data-centric AI

We provide a state-of-the-art model GNN model for Performance Evaluation

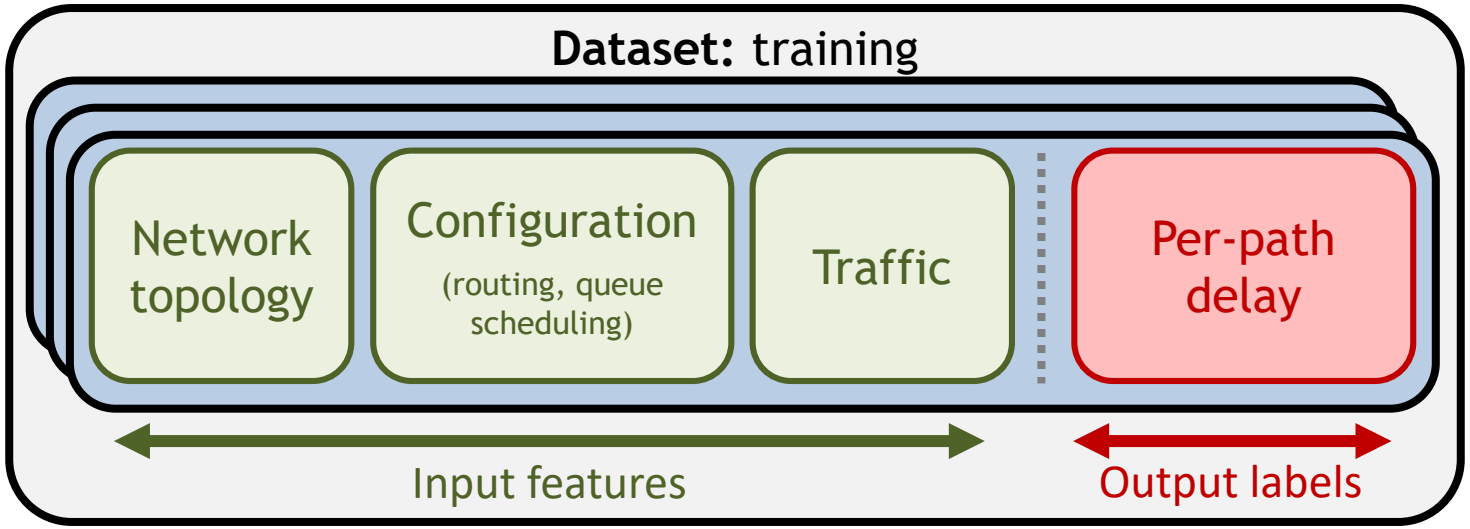


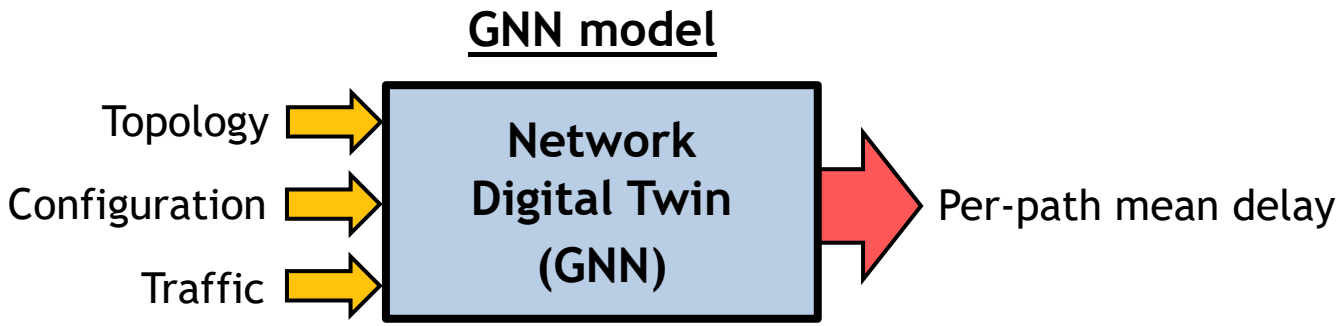
- **Input:**
  - Network topology
  - Configuration (routing, queue scheduling)
  - Traffic (different traffic models)

- **Output:**
  - Mean per-packet delay on each source-destination path

Problem statement:

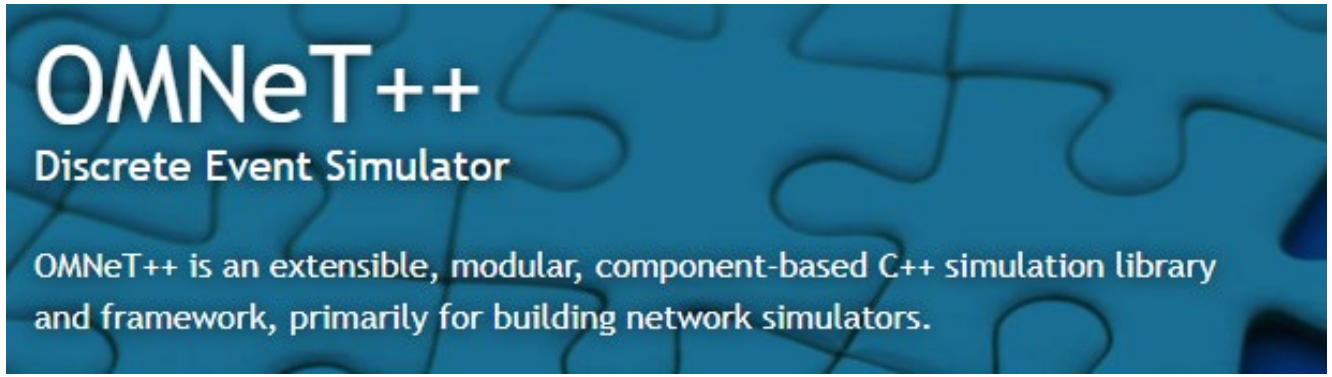
- Participants will be asked to generate the best dataset to train the GNN model





Graph Neural Networking challenge 2022

- To generate the dataset we provide a **discrete-event network simulator** based on OMNet++



https://omnetpp.org/

- **Simulators are very costly** → We have scaled down the problem to enable the generation of the dataset on commodity hardware (< 1h)

Datasets

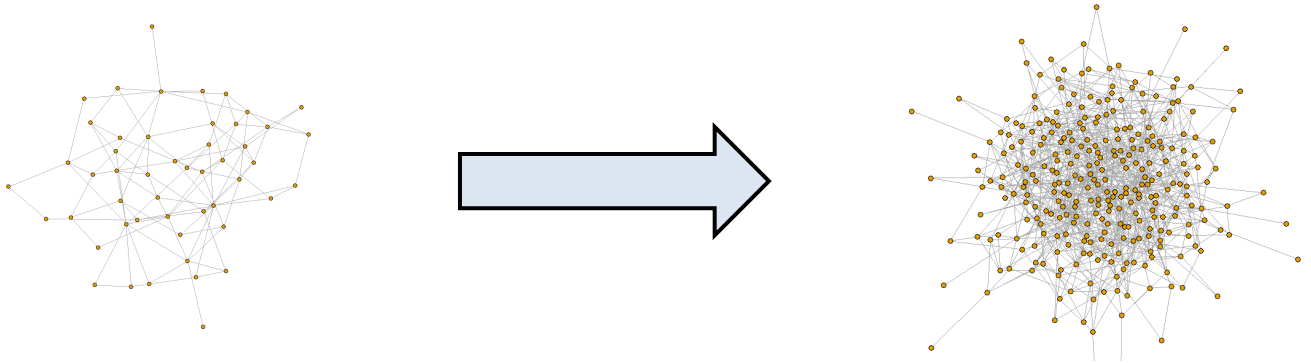
We provide a validation dataset:

- Samples in networks up to **300 nodes** (large)

Participants generate a training dataset:

- Constraints:
  - **Maximum 100 samples** (very limited dataset)
  - Samples must be from **networks up to 10 nodes (small)**

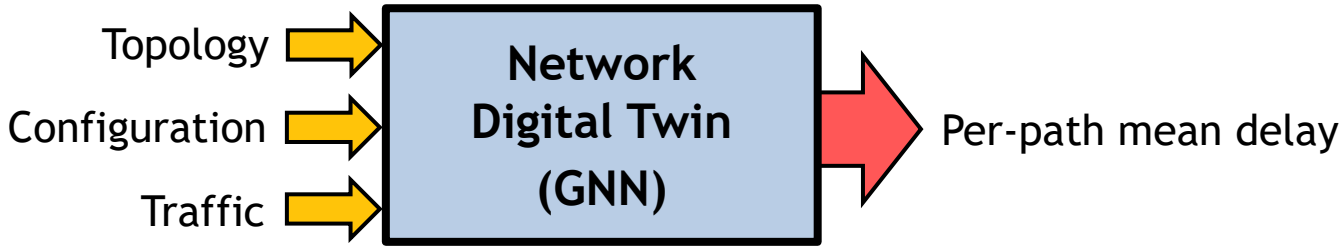
- Participants can train the GNN model and check the performance in the validation dataset (training < 1h)



Training (small networks)

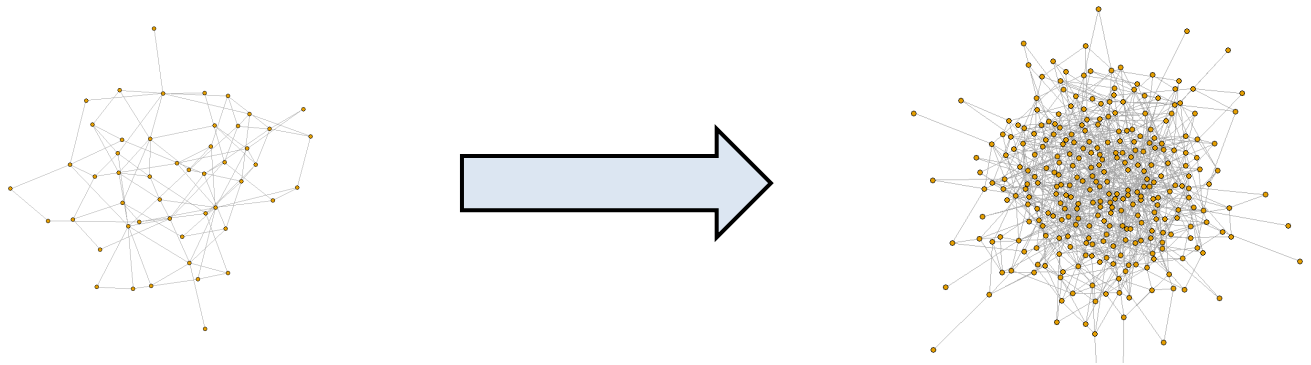
Validation (large networks)

Datasets



The training dataset should be designed to **help the GNN model generalize to**:

- **Larger topologies** (larger link capacities and traffic aggregates)
- **Configurations** (routing, queue scheduling)
- **Traffic** (different models and load levels)



Training (small networks)

Validation (large networks)

Evaluation

- At the end of the challenge (Oct 1st), we will evaluate participants' solutions on a **test dataset**
- The test dataset will follow similar distributions to the validation dataset (released at the beginning)
- **The evaluation phase lasts 15 days**, and is made automatically in our evaluation platform
- The GNN model will be trained with the training dataset of participants, and we will evaluate the prediction accuracy of the trained GNN models
- **Participants will see the ranking in real-time**

Evaluation

After the evaluation…

- **Provisional ranking** with the scores of all teams

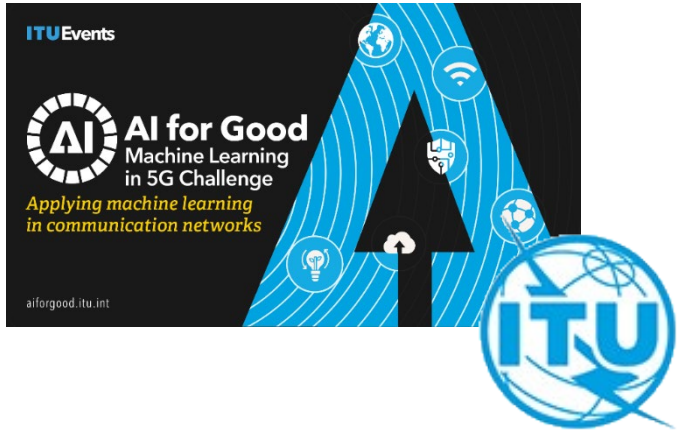
- We will ask top-5 teams for:
  - **Training datasets**
  - **Script to generate the datasets**
  - **A short report** describing their solution (1-3 pages)

- We will **reproduce and validate the training of top-5 solutions** to check that they comply with all the rules (e.g., max. 100 samples, networks up to 10 nodes)

Quick summary

Main resources:

- Network simulator based on OMNet++ (Docker image)
- GNN model
- Validation dataset
- Quick-start tutorial on Jupyter notebooks
- Mailing list for Q&A from participants (support from organizers)



Final outputs from participants:

- Training datasets
- Methods to create those datasets

Expected outcomes:

- Advance the state of the art on how to produce good datasets for ML applied to networking (**Data-centric AI for networking**)

Graph Neural Networking challenge 2022:

Improving Network Digital Twins through Data-centric AI

https://bnn.upc.edu/challenge/gnnet2022

Timeline (tentative)**:**

- **Challenge duration:** May-Nov 2022
- **Open registration:** May 27th-Sep 31st
- **Release of tools and validation dataset:** End of June
- **Evaluation phase:** Oct 1st-Oct 15th 2022
- **Top-5 teams submit the dataset, code and documentation:** Oct 31st 2022
- **Final ranking and official announcement of the winners:** Nov 2022

Check the website!

Graph Neural Networking challenge 2022:

Improving Network Digital Twins through Data-centric AI

https://bnn.upc.edu/challenge/gnnet2022

José Suárez-Varela

Barcelona Neural Networking center

Universitat Politècnica de Catalunya

May 27th 2022