Enhanced GNN generalization to Real Network Dataset by Attention Mechanism

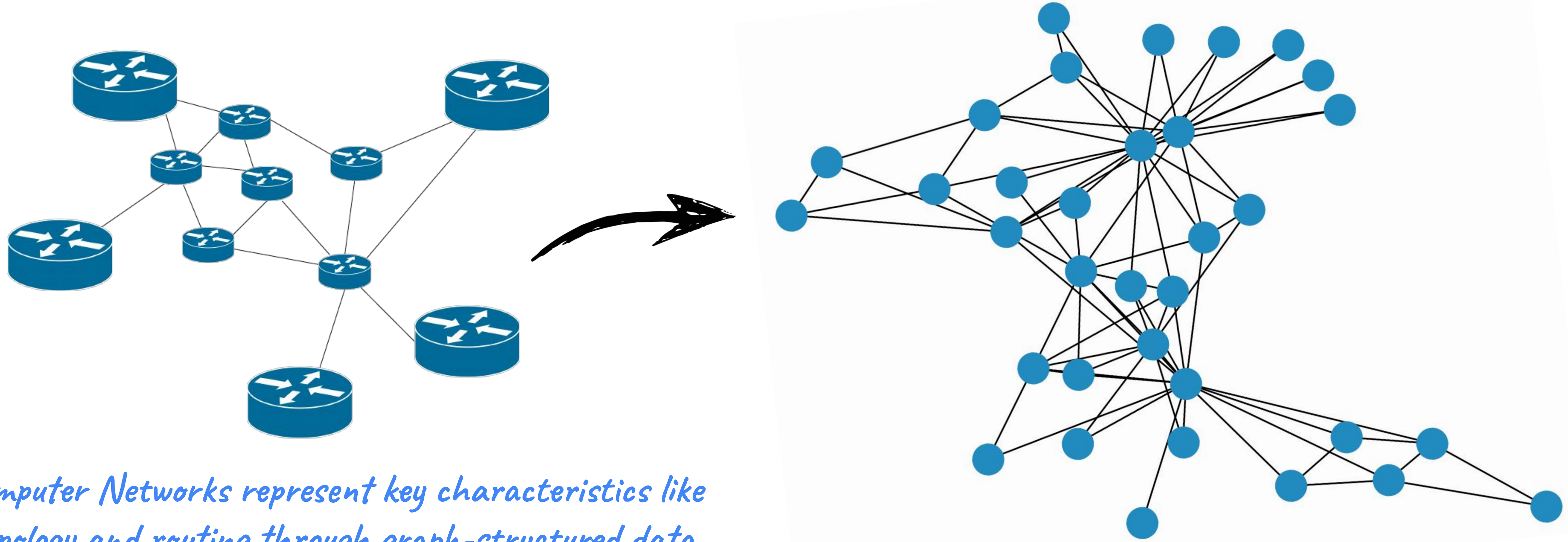
Cláudio Matheus Modesto, Rebecca Aben-Athar, Andrey Silva, Silvia Lins

5G & IoT Research Group - Telecommunications, Automation and Electronics Research and Development Center (LASSE) Universidade Federal do Pará (UFPA)

Ericsson Research

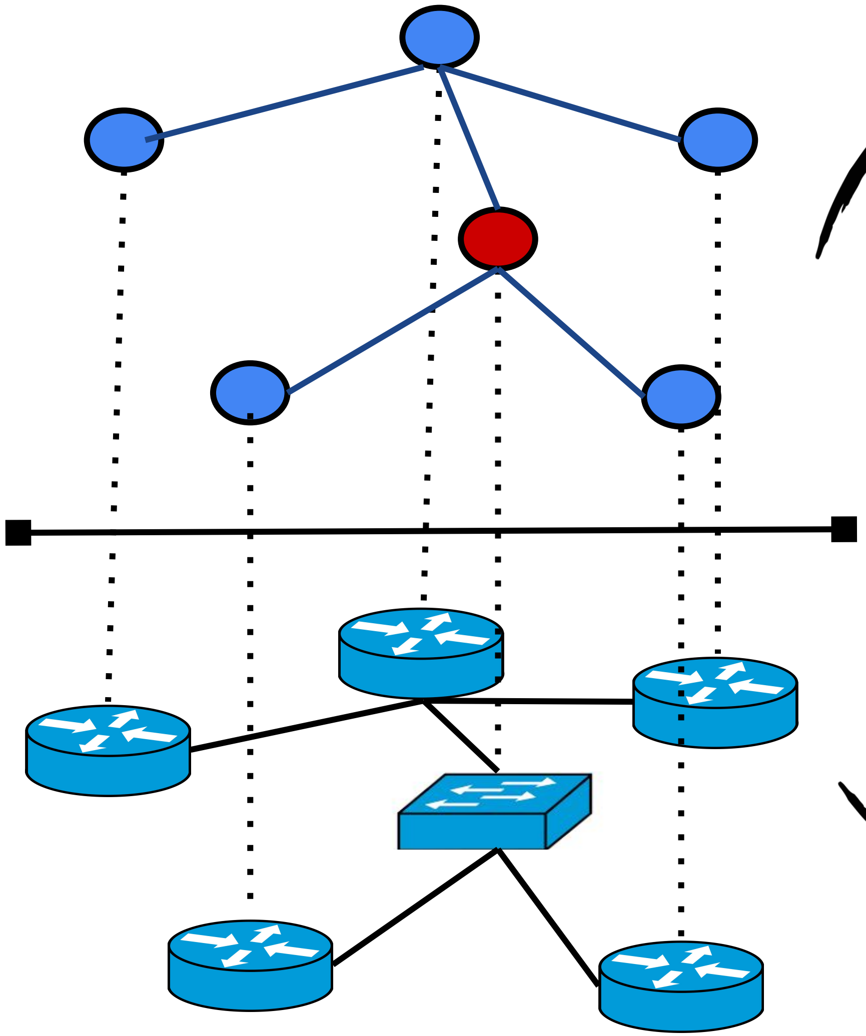
Motivation

Graph Neural Networks (GNNs) are a central tool for applying AI/ML techniques to networking applications **due to their unique ability to generalize over graph data**.



Computer Networks represent key characteristics like topology and routing through graph-structured data

Motivation



Digital Twin

Digital Twin

Enables to mimic complex environments and to answer what / if questions like:

- What happens if a specific node becomes unreachable?
- In terms of QoS parameters, *what* is the impact?
- Similarly, if a random node undergoes a failure?

Physical Environment

Problem Statement

- **Given a real network database**

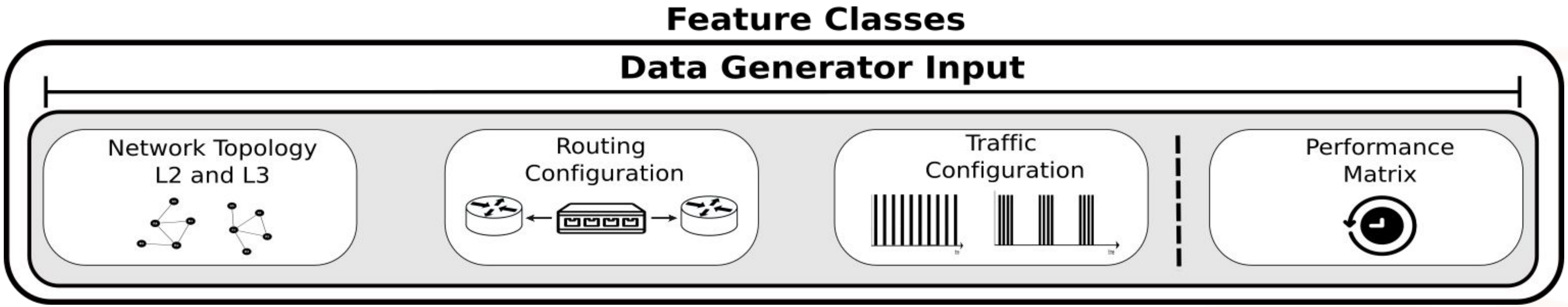


Figure 1: Structure of the database

This challenge aims to create a Network Digital Twin based on neural networks that can accurately estimate **QoS performance metrics** given the network state and the input traffic.

Problem Statement

Input/Output relation. A possible example of the model in the problem

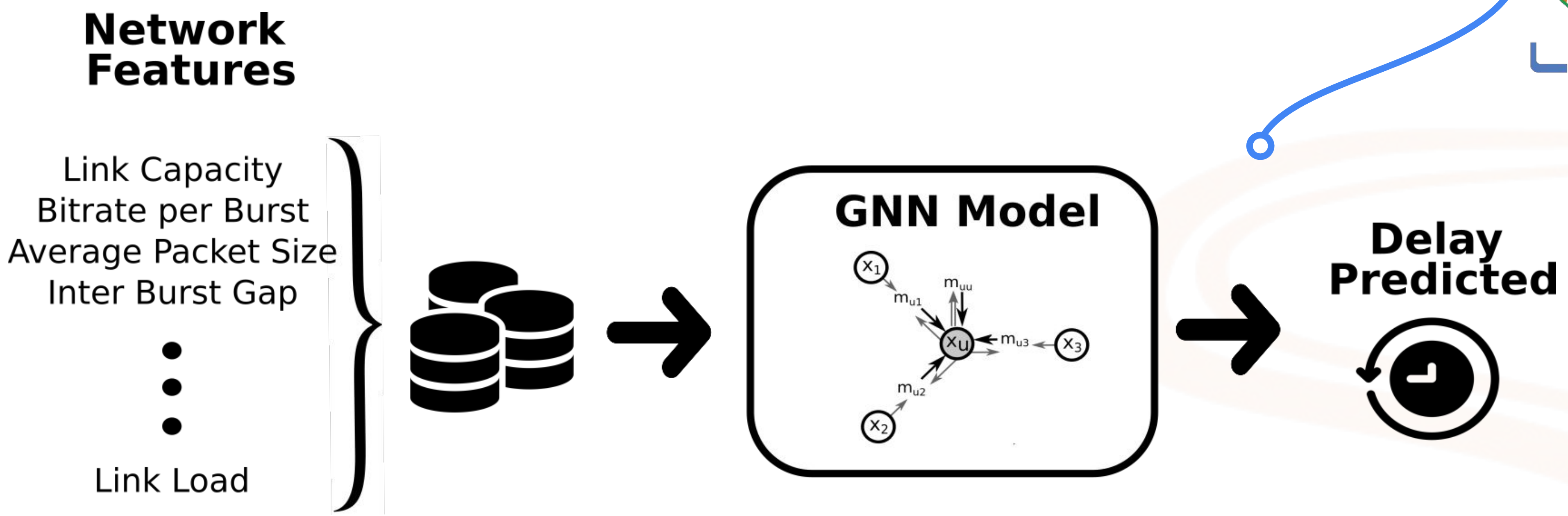


Figure 2: Schematic representation of GNN Model called RouteNet

Given the database, we should apply this in a neural network model.
Or a GNN model such as **RouteNet**.

Problem Statement

Using the predicted mean per-flow delay and the label extracted from the performance matrix

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right|$$

Any metric could be used to train and evaluate the model performance, but the **challenge adopted the MAPE** for comparison purposes.

m0b1us solution

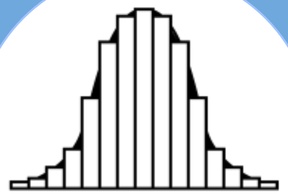
Feature Selection

Choose the most consistent, non-redundant, and relevant features across large datasets

MAPE: 44.77%

Baseline MPNN

01



Z-Score Normalization

Different variables are on a standard scale

02

03

Attention Mechanism

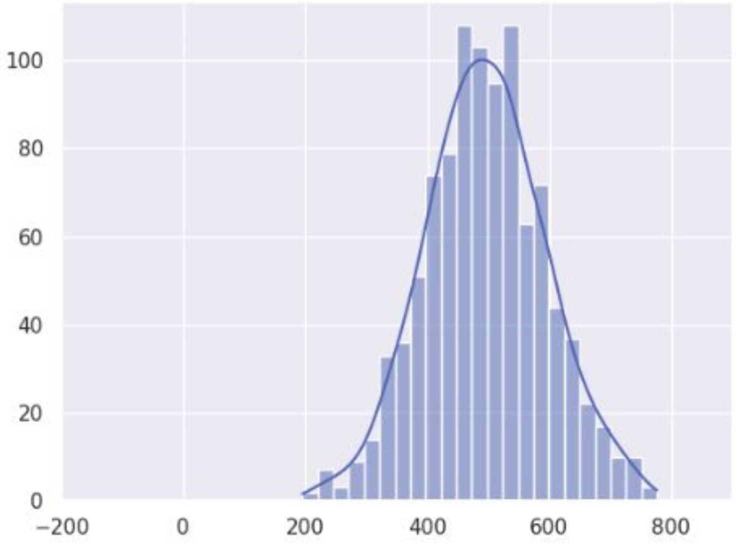
Weight a given set of features according to their importance

MAPE: 20.00%

Final Prediction

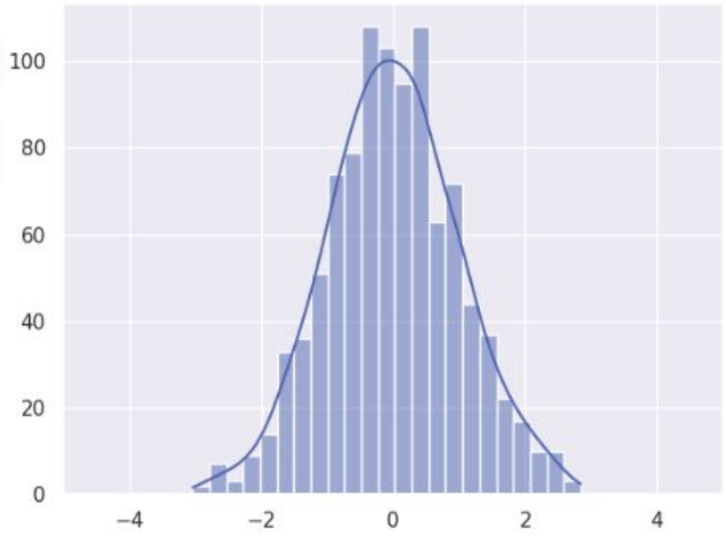
Z-Score Normalization

The Z-score normalization was used in this work instead of the default feature normalization found in the model baseline, that is, min-max normalization.

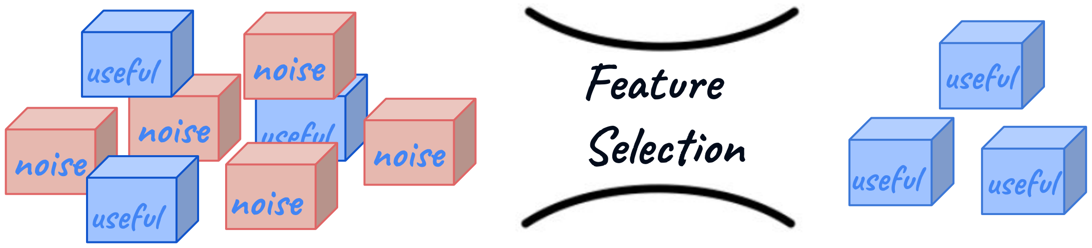


Z - Score

$$z = \frac{x - \mu}{\sigma}$$



Feature Selection



Advanced Techniques for Supervised Learning

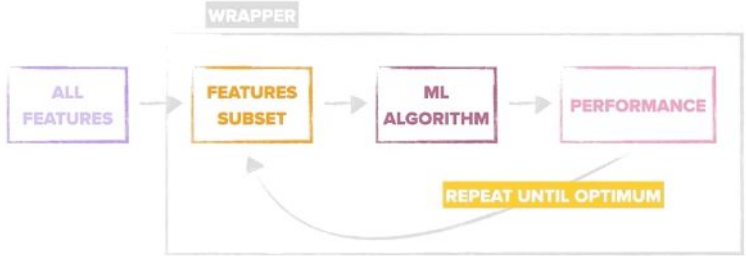
Filter-Based

- Ranking the features based on their score
- Computationally fast
- Multi Information (MI) Algorithm

$$MI(x,y) = \sum p(x_i, y_j) \log \left(\frac{p(x_i, y_j)}{p(x_i)p(y_j)} \right)$$

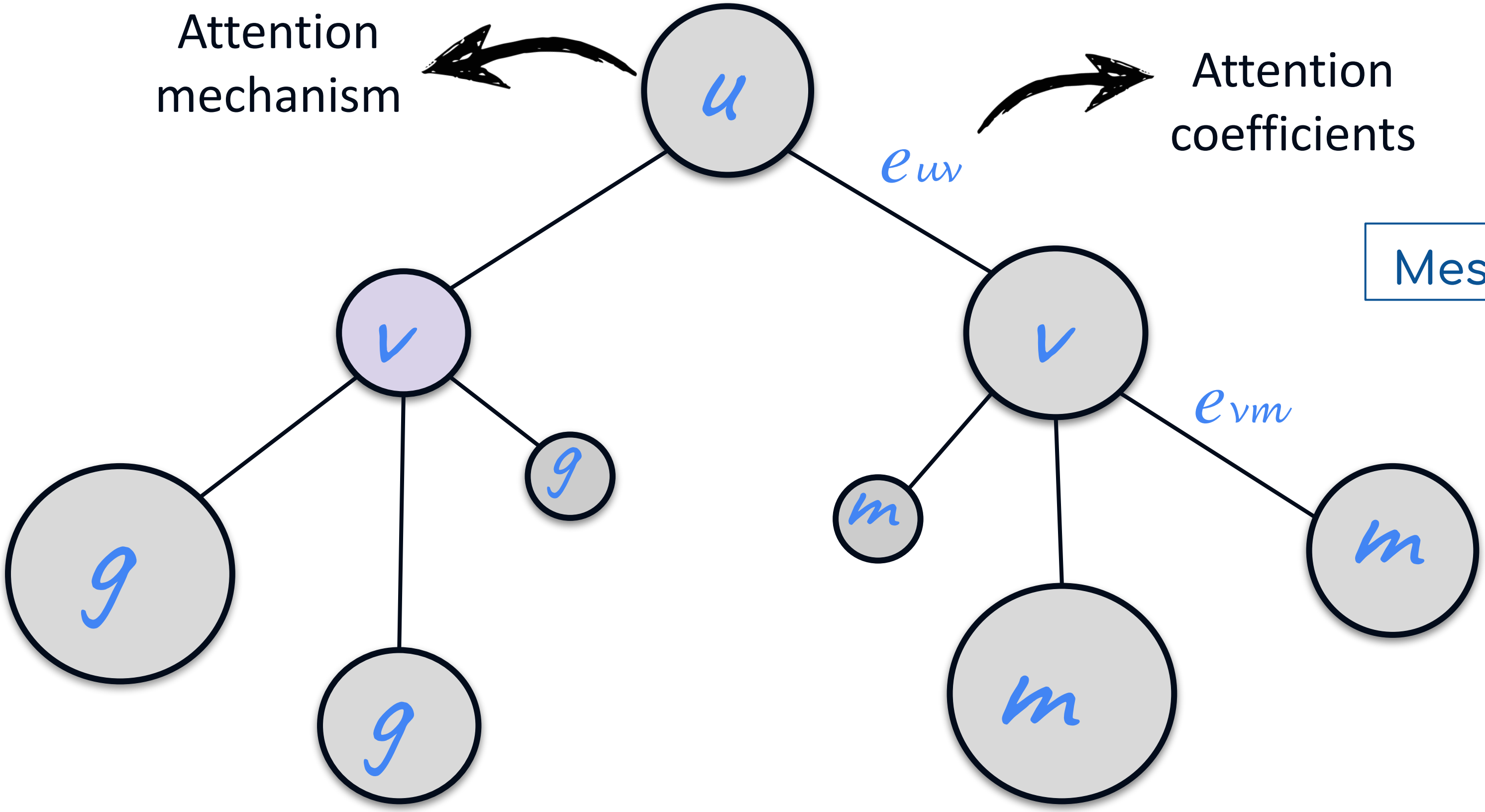
Wrapper-Based

- Analyzes the feature set based on ML algorithm
- Computationally expensive
- Exhaustive Algorithm



Attention Mechanism

The attention mechanism would be another layer that provides a fine-tuning to identify the importance of selected features, but in this case, not eliminating definitively, but **decreasing** or **increasing** their relevance from scores.



Message-Passing ⊇ Attention ⊇ Convolution

The weight depends on anything the embeddings capture, such as features or their local structure!

Attention Mechanism

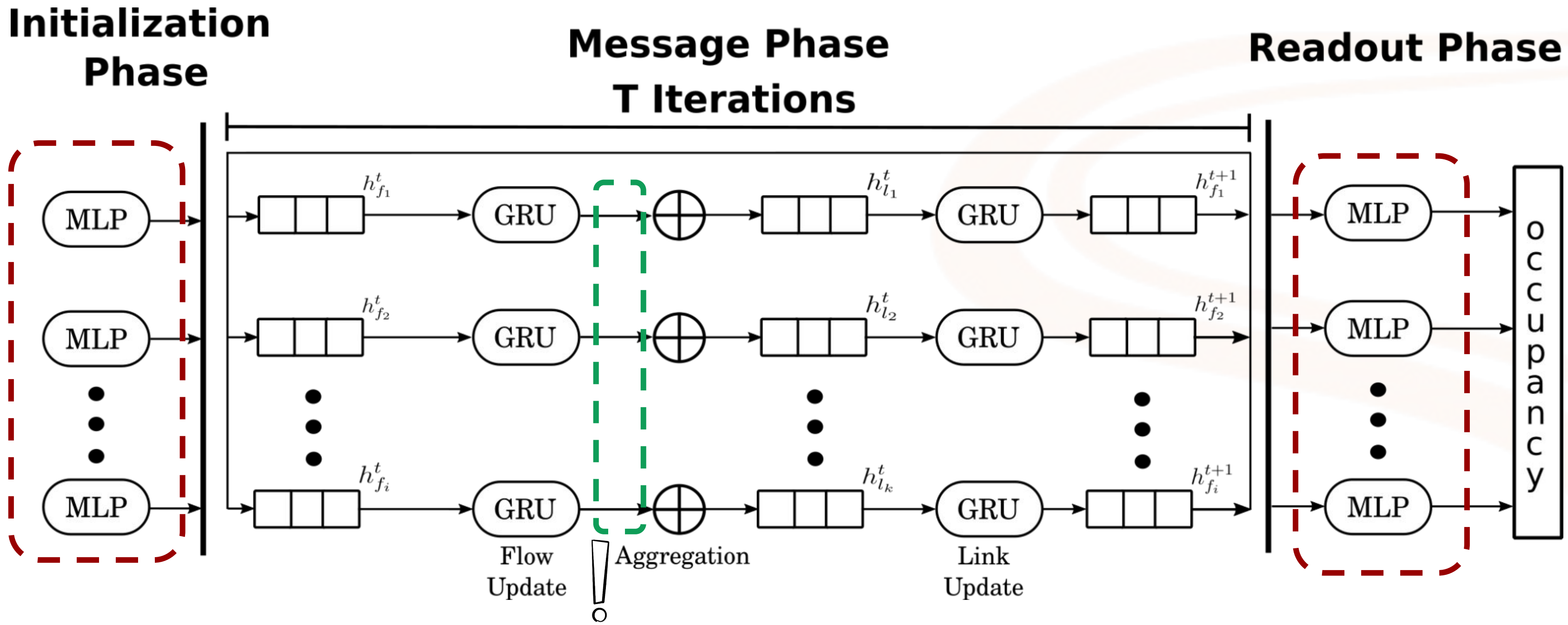
Our idea focused on exploring the attention layer simultaneously with the structure of multiple-stage message passing present in the baseline.

SeLU

Activation Function

LeCun Uniform

Weight Initialization



Attention Coefficient

$$e_{uv} = a^T \text{LeakyReLU}(\mathbf{W}[h_u || h_v])$$

Normalization

$$\alpha_{uv} = \text{softmax}(e_{uv})$$

Normalized Attention Coefficient

$$\boxed{\alpha_{uv} = \frac{\exp(\mathbf{a}^T \text{LeakyReLU}(\mathbf{W}[\mathbf{h}_u || \mathbf{h}_v]))}{\sum_{k \in \mathcal{N}_u} \exp(\mathbf{a}^T \text{LeakyReLU}(\mathbf{W}[\mathbf{h}_u || \mathbf{h}_k]))}}$$

Results and Discussion

Given the variety of tools mentioned, the following table summarizes the model performance on each of them, showing metrics from the training, validation, and test dataset

ID	Experiment	Iterations	Flow Emb. Length	Link Emb. Length	Features	Type of Feature Normalization	MAPE Loss Function		
							Train	Validation	Test
1	Baseline	8	64	64	Average bandwidth, Average packet size, Packets generated, Capacity, Load	Min - Max	**28.04%**	**32.60%**	**41.42%**
2	MSMP + SeLU	8	64	64	Average bandwidth, Average packet size, Packets generated, Load, Capacity, Bitrate per burst, Inter-packet gap mean Inter-packet gap, variance, Packets per burst, Normalized load, Flow length, 90th Packet size percentile	Min - Max	**18.20%**	**22.09%**	**26.50%**
3	MSMP + Attention	12	16	16		Min - Max	**16.99%**	**17.41%**	**24.02%**
4	MSMP + SeLU + Attention	12	16	16		Standardization (Z - Score)	**14.83%**	**15.71%**	**20.00%**

Our winner solution

Cláudio Barata

claudio.barata@itec.ufpa.br

Cláudio Matheus

Rebecca Aben-Athar

rebecca.athar@itec.ufpa.br

Rebecca Aben-Athar

Andrey Silva

andrey.silva@ericsson.com

Silvia Lins

silvia.lins@ericsson.com

Acknowledgments

Prof. Aldebaro Klautau

Prof. Glauco Gonçalves