# Path Representation Learning in Road Networks

**Sean Bin Yang**

**AAU Thesis Defense**

Supervisor: Prof. Bin Yang

Co-Supervisor: Assoc. Prof. Jilin Hu

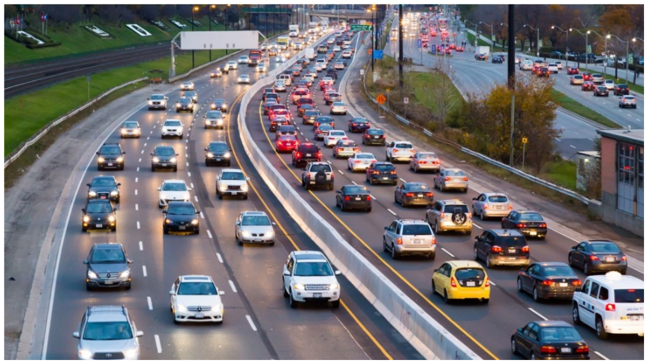
Department of Computer Science

Technical Faculty of IT and Design, Aalborg University

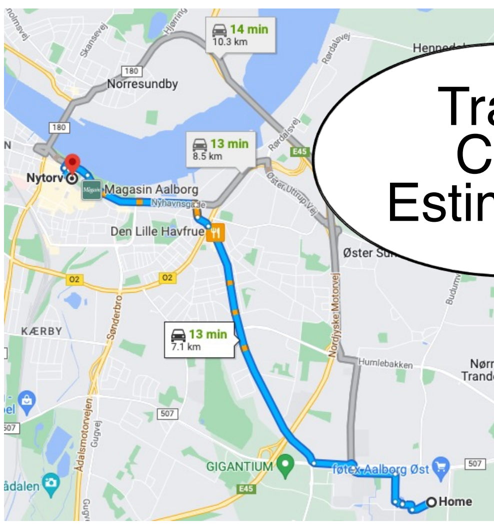
# AGENDA

# AGENDA

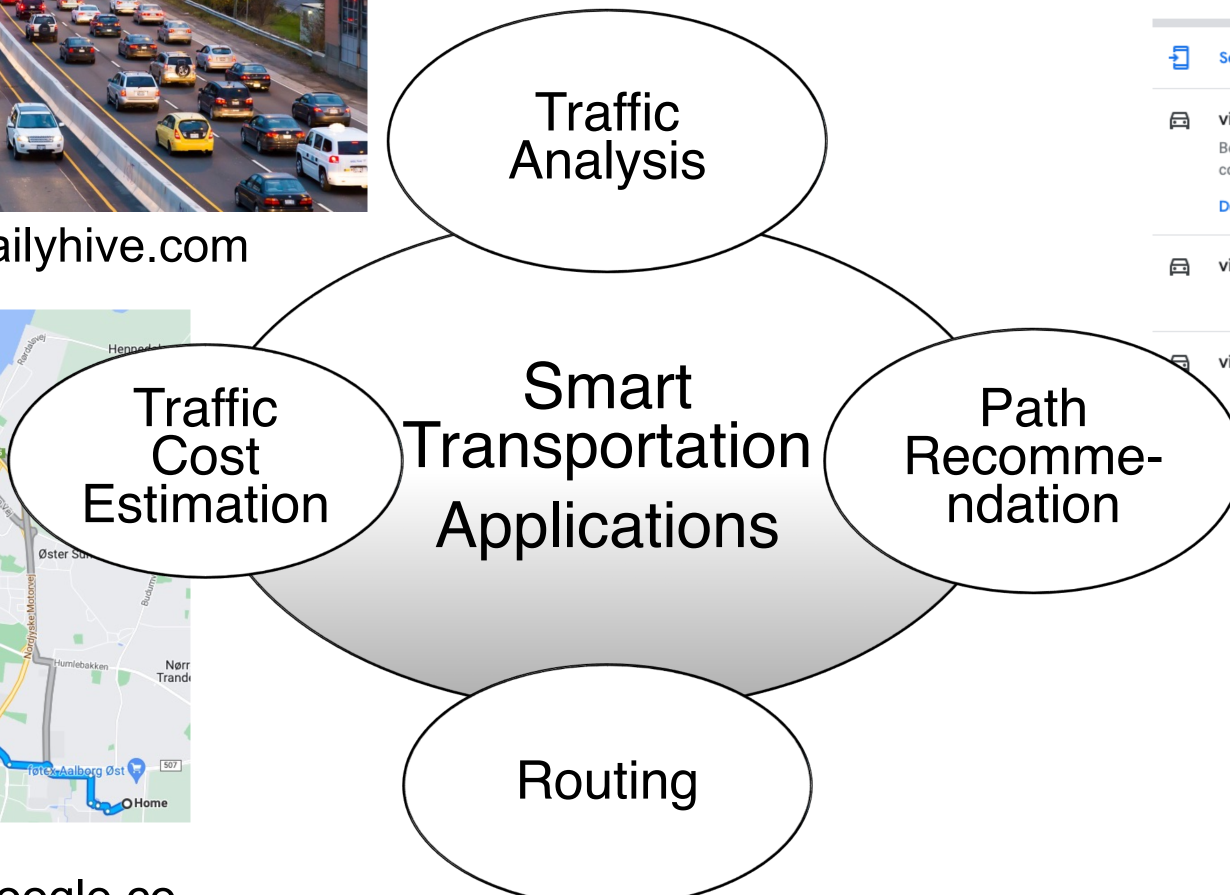
# Smart Transportation Applications

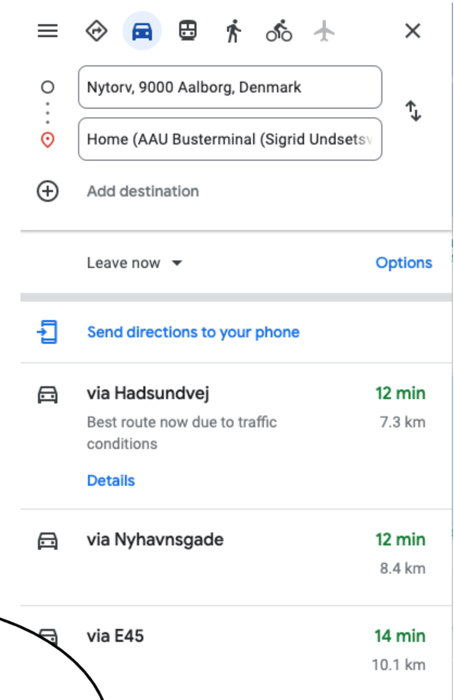


https://dailyhive.com




https://www.google.com/maps

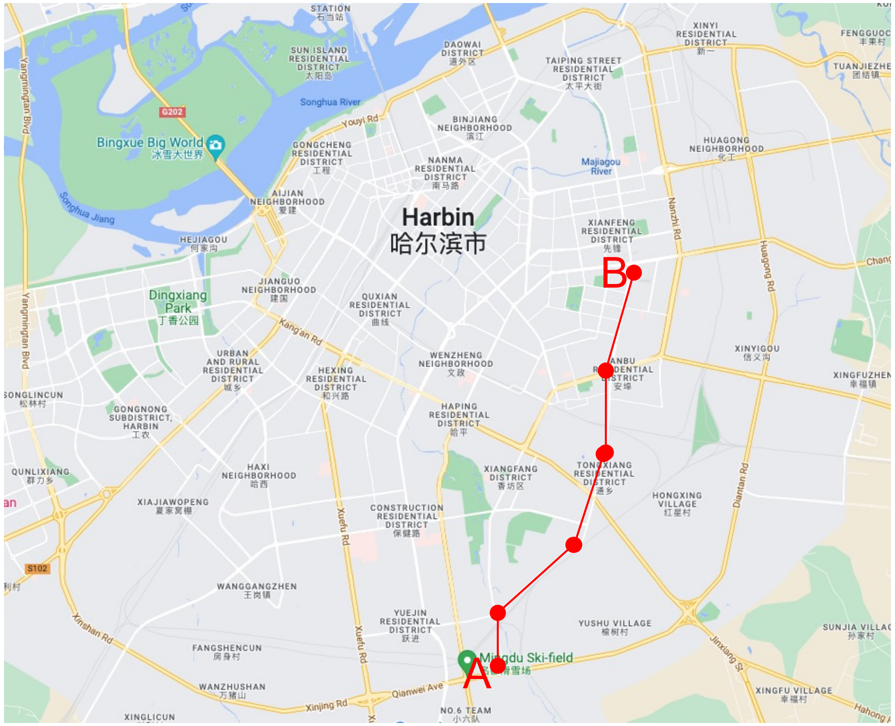



https://www.google.com/maps

# Trajectory vs. Path

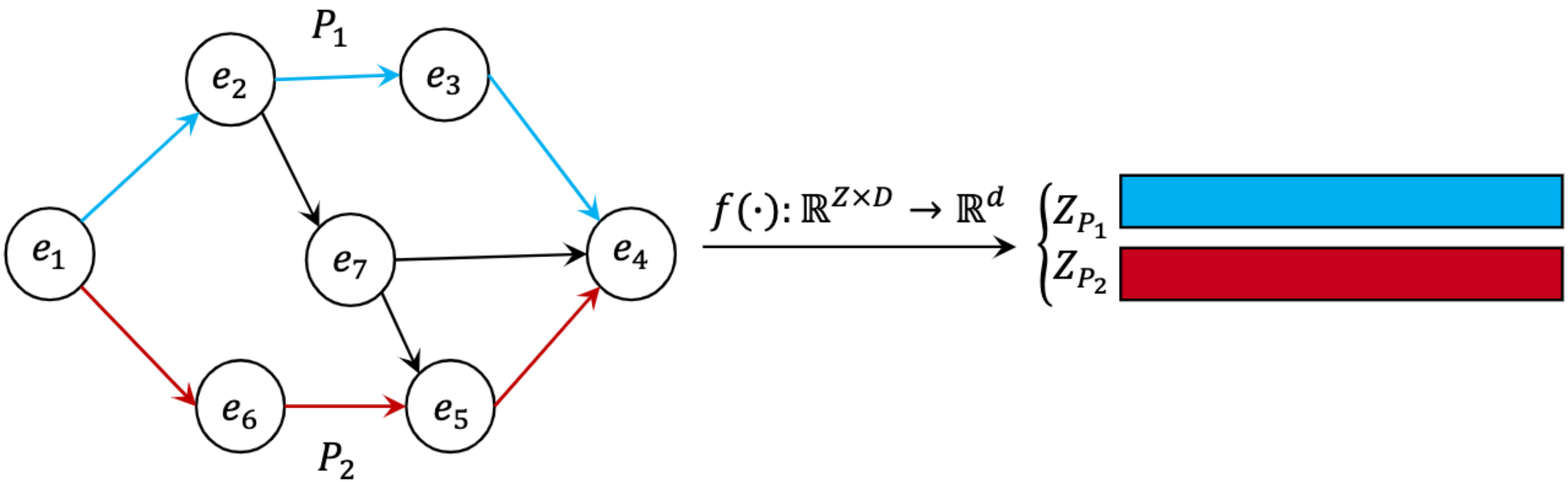
Trajectory, Harbin (Li et al., 2019)




Path, Road Network, Harbin

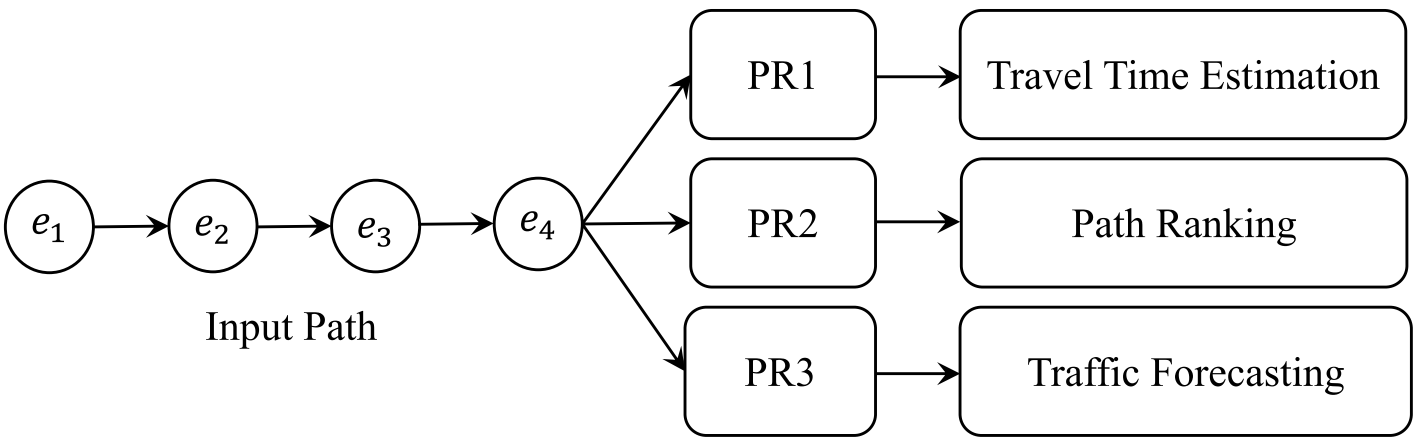
# Path Representation Learning

- What is Path Representation Learning?



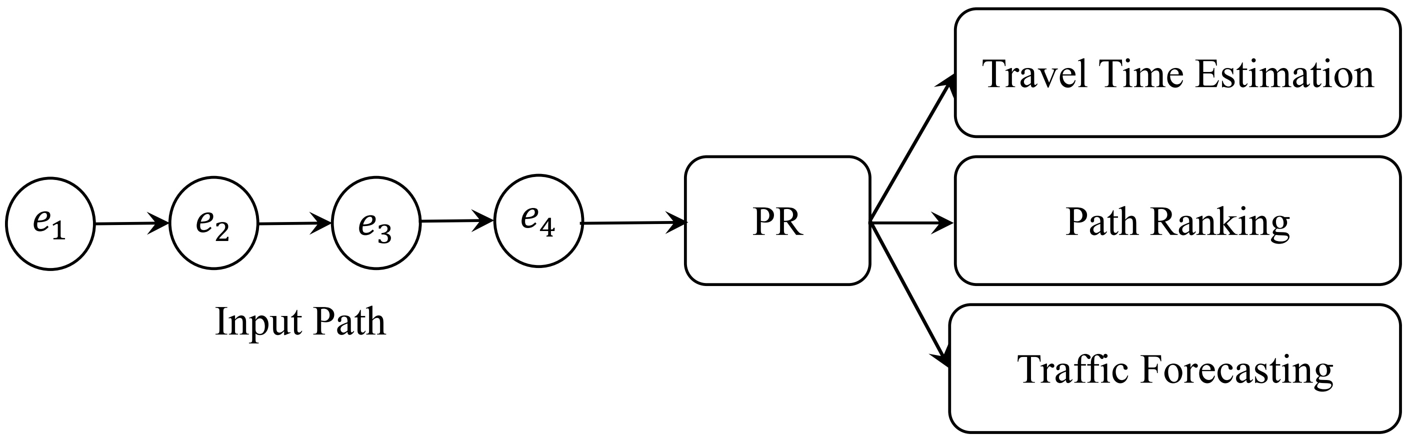
# Task-specific Path Representation

- **Supervised Learning** is a **task-specific** learning procedure.
  - Large amounts of labeled dataset



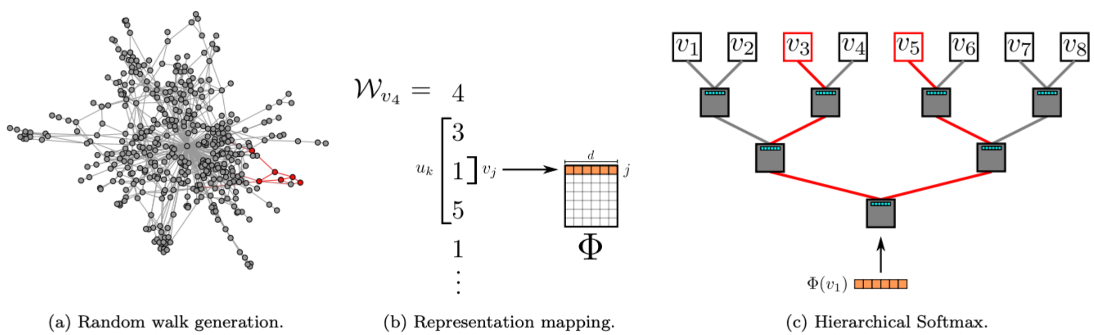
# Why task-unspecific Path Representation?

- How to learn a **task-unspecific path representation** is nontrivial.
  - Taking advantage of a large amount of unlabeled data.

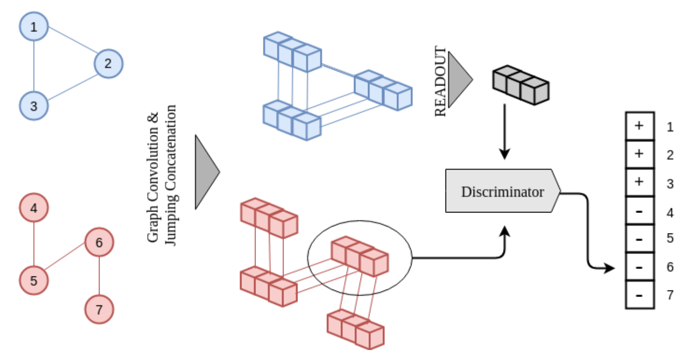


# Task-Unspecific Representation Learning

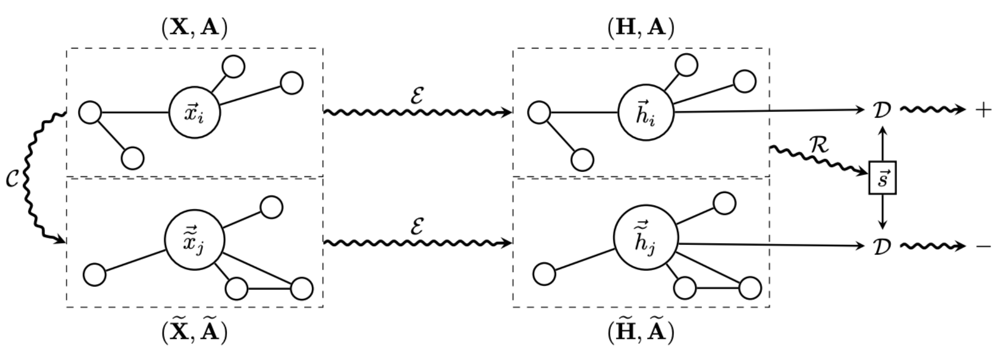
- **Unsupervised learning** learns a task-unspecific representation, but mainly focus on graph data analysis.



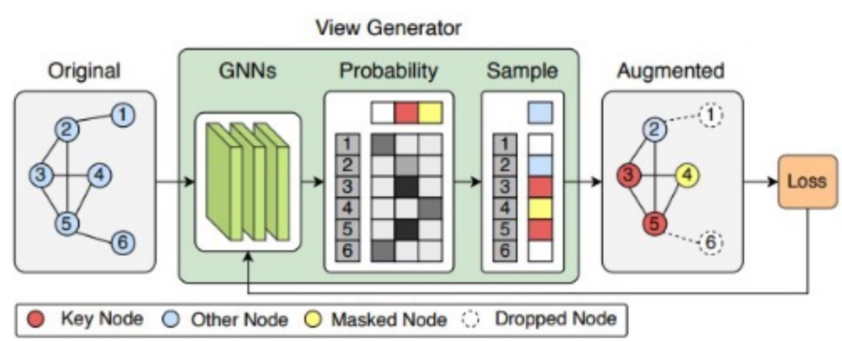

DeepWalk (Perozzi et al., 2014)




InfoGraph (Sun et al., 2020)

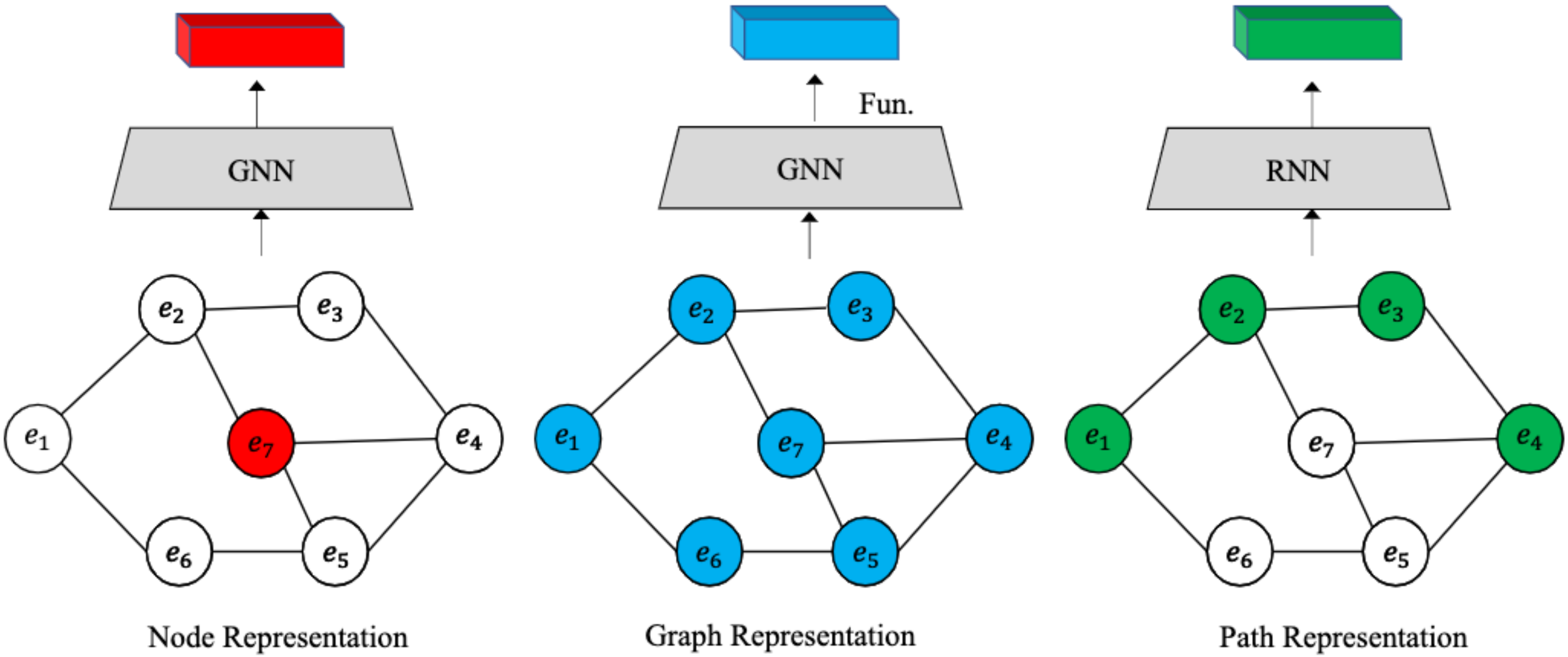


Deep Graph InfoMax (Velickovic et al., 2019)



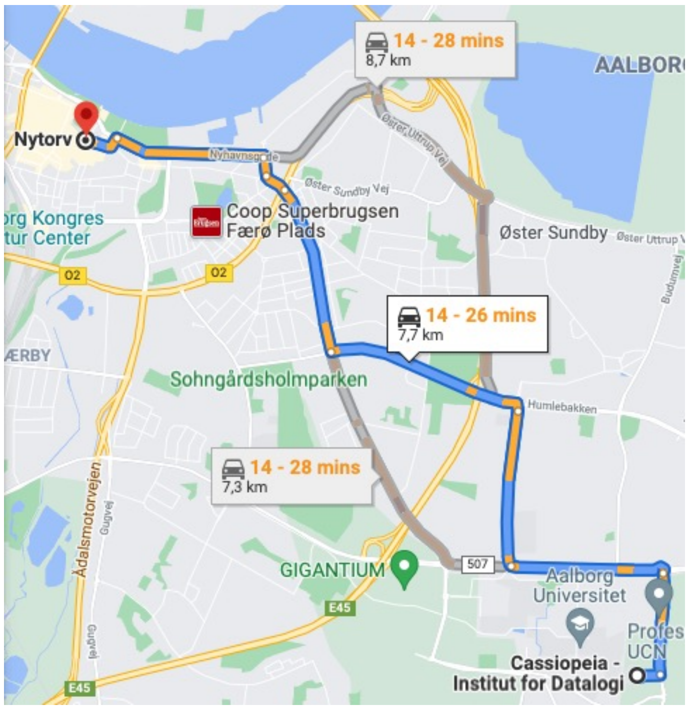

AutoGCL (Yin et al., 2022)

# Path vs. Node/Graph Representation

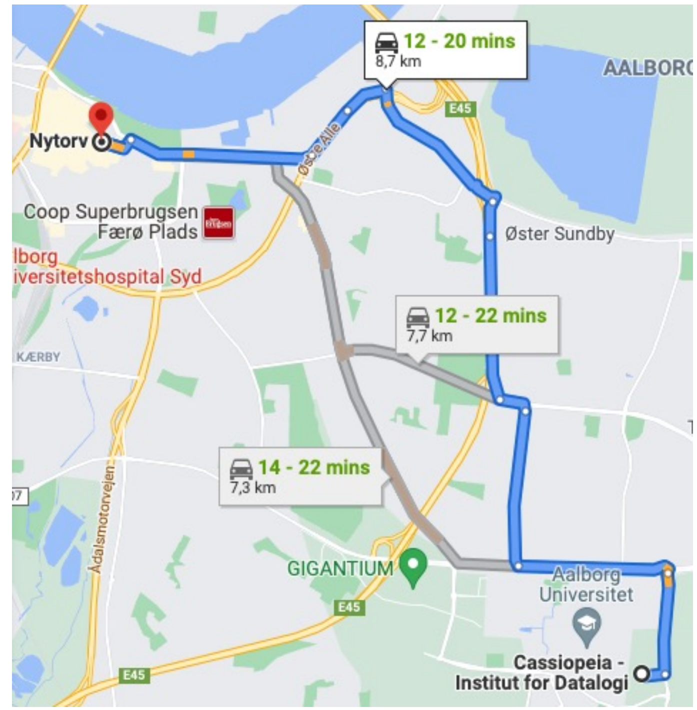


# How to Consider Temporal Information?

- Importance of **temporal information**



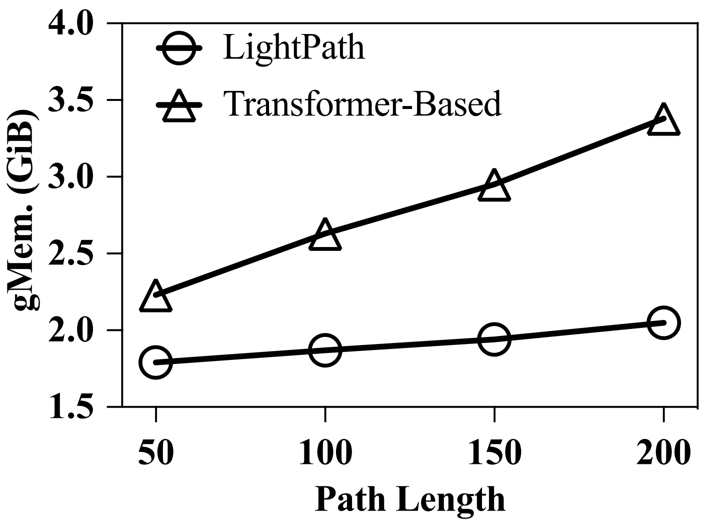

(a) 8:00 a.m.



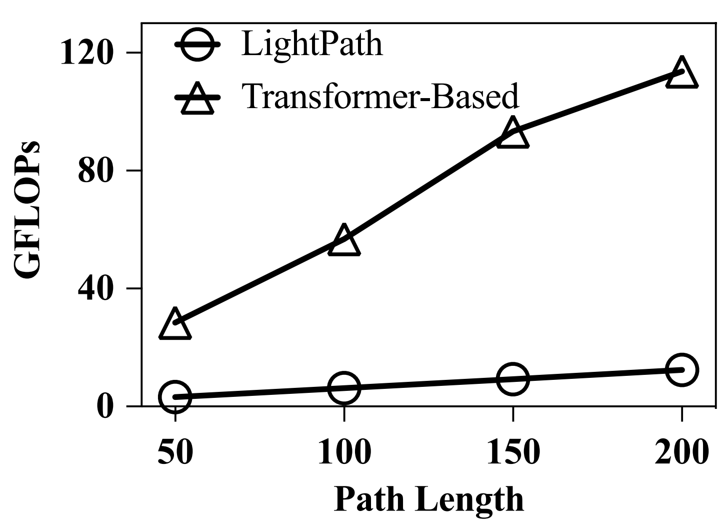

(b) 10:00 a.m.

# Why Lightweight and Scalable PRL?

- Poor Scalability w.r.t. Path Length




(a) GPU Memory




(b) GFLOPs

- Mode size with different encoder layers

| Encoder Layers L | 12 | 24 | 48 | 96 |
|---|---|---|---|---|
| Parameters (Millions) | 29.85 | 55.07 | 105.51 | 206.40 |

# Publications

A. **Sean Bin Yang**, Bin Yang, “Learning to Rank Paths in Spatial Networks,” ICDE 2020: 2006-2009.
B. **Sean Bin Yang**, Chenjuan Guo, Bin Yang, “Context-Aware Path Ranking in Road Network,” IEEE Trans. Knowl. Data Eng. 2022, 34(7): 3153-3168 (2022).
C. **Sean Bin Yang**, Chenjuan Guo, Jilin Hu, Jian Tang, Bin Yang, “Unsupervised Path Representation Learning with Curriculum Negative Sampling,” IJCAI 2021: 3286-3292.
D. **Sean Bin Yang**, Chenjuan Guo, Jilin Hu, Bin Yang, Jian Tang, Christian S. Jensen, “Weakly-supervised Temporal Path Representation Learning with Contrastive Curriculum Learning,” ICDE 2022: 2873-2885.
E. **Sean Bin Yang**, Jilin Hu, Chenjuan Guo, Bin Yang, Christian S. Jensen, “LightPath： Lightweight and Scalable Path Representation Learning,” 2022 (VLDB, in Submission).

# Challenges Addressed

Task-specific
Path Representations

Paper A & B

Task-unspecific
Path Representations

Paper C

Paper D

Paper E

**Motivations**

**How to learn context aware task-specific model for path ranking ?**

**How to learn a path representation in an unsupervised manner such that it can be used for different downstream tasks?**

**How to learn a temporal path representation by leveraging the information from weak lable?**

**How to design a lightweight and scalable path representation framework by considering resource-constrained environments?**

# Challenges Addressed

Task-specific
Path Representations

Paper A & B

Task-unspecific
Path Representations

Paper C

Paper D

Paper E

**Motivations**

**How to learn context aware task-specific model for path ranking ?**

**How to learn a path representation in an unsupervised manner such that it can be used for different downstream tasks?**

**How to learn a temporal path representation by leveraging the information from weak lable?**

**How to design a lightweight and scalable path representation framework by considering resource-constrained environments?**

# Challenges Addressed

## Motivations

Task-specific Path Representations

Paper A & B → How to learn context aware task-specific model for path ranking ?

Task-unspecific Path Representations

Paper C → How to learn a path representation in an unsupervised manner such that it can be used for different downstream tasks?

Paper D → **How to learn a temporal path representation by leveraging the information from weak lable?**

Paper E → How to design a lightweight and scalable path representation framework by considering resource-constrained environments?

# Challenges Addressed

Task-specific
Path Representations

Paper A & B

Task-unspecific
Path Representations

Paper C

Paper D

Paper E

**Motivations**

**How to learn context aware task-specific model for path ranking ?**

**How to learn a path representation in an unsupervised manner such that it can be used for different downstream tasks?**

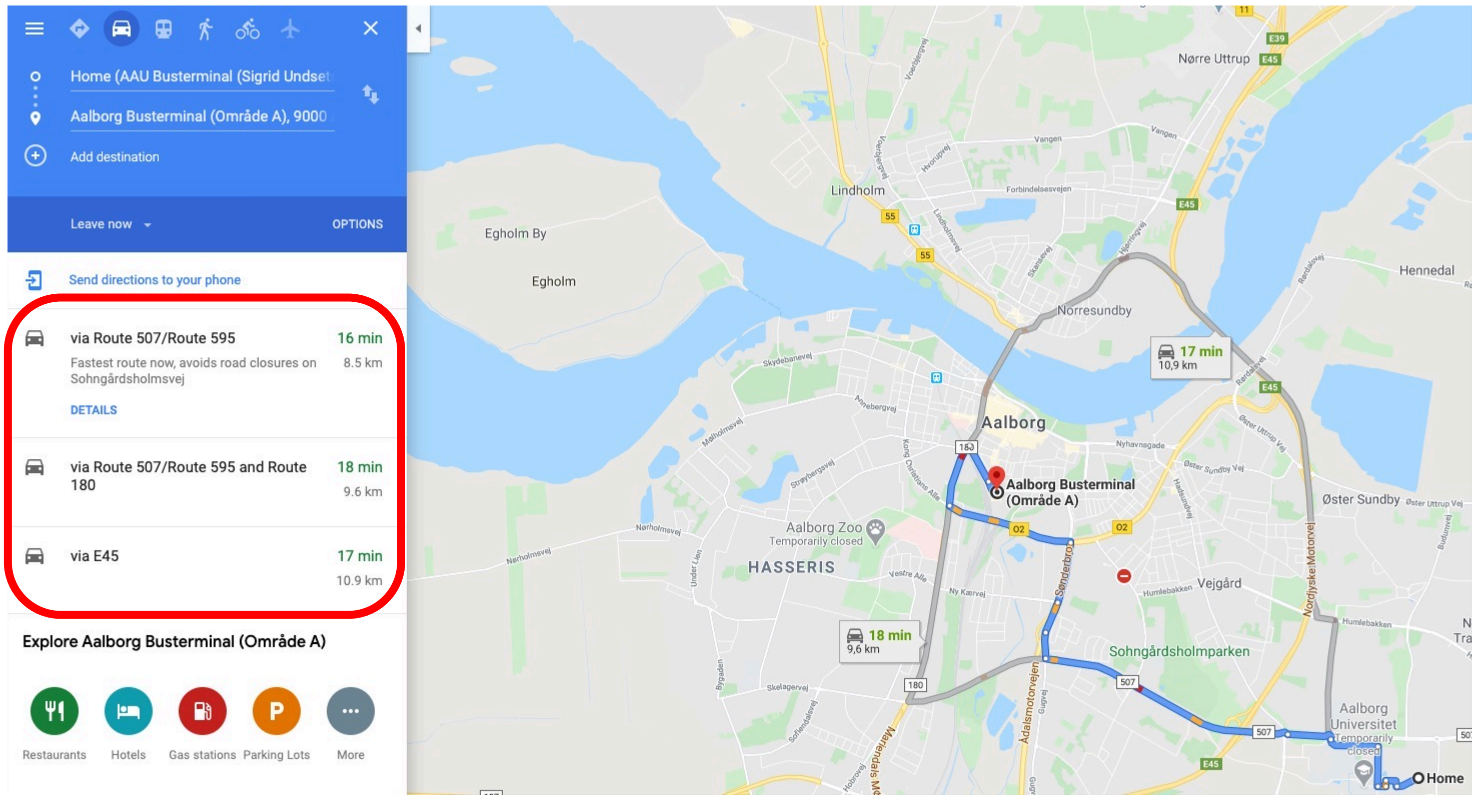
**How to learn a temporal path representation by leveraging the information from weak lable?**

**How to design a lightweight and scalable path representation framework by considering resource-constrained environments?**

# AGENDA

# Intuition

- Local Drivers often choose paths that are neither shortest nor fastest.
- Google Maps and TomTom, etc.



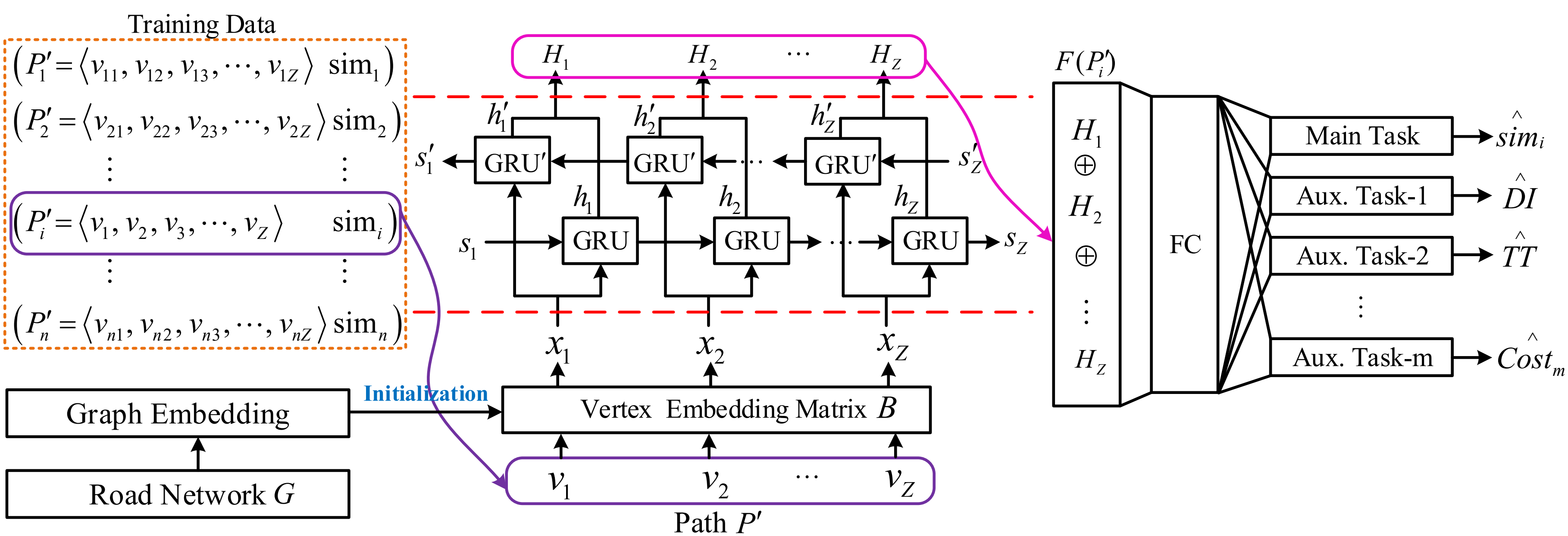

V. Ceikute and C. S. Jensen, "Routing service quality - local driverbehavior versus routing services," in2013 IEEE 14th International Conference on Mobile Data Management, Milan, Italy, June 3-6, 2013 -Volume 1, 2013, pp. 97–106.

# Problem Definition

- Given a set of N candidate paths **P** that connect the same source and destination and optimal contexts such as a departure time and driver identifier, we aim at:
  - Estimating a ranking score $sim(P, P')$ for each candidate path $P_i' \in \boldsymbol{P}$, $P$ denotes the trajectory path;
  - Providing a ranked list of the candidate paths $< P_1', P_2', \cdots, P_\mathrm{N}' >$, such that $sim(P, P_i') > sim\left(P, P_j'\right)$ when $1 \leq i < j \leq N$.

# PathRank

- Training Data Generation
- Vertex embedding
- Path Representation Learning
- Multi-task learning




$\widehat{sim}_i$: Ranking Score Estimation

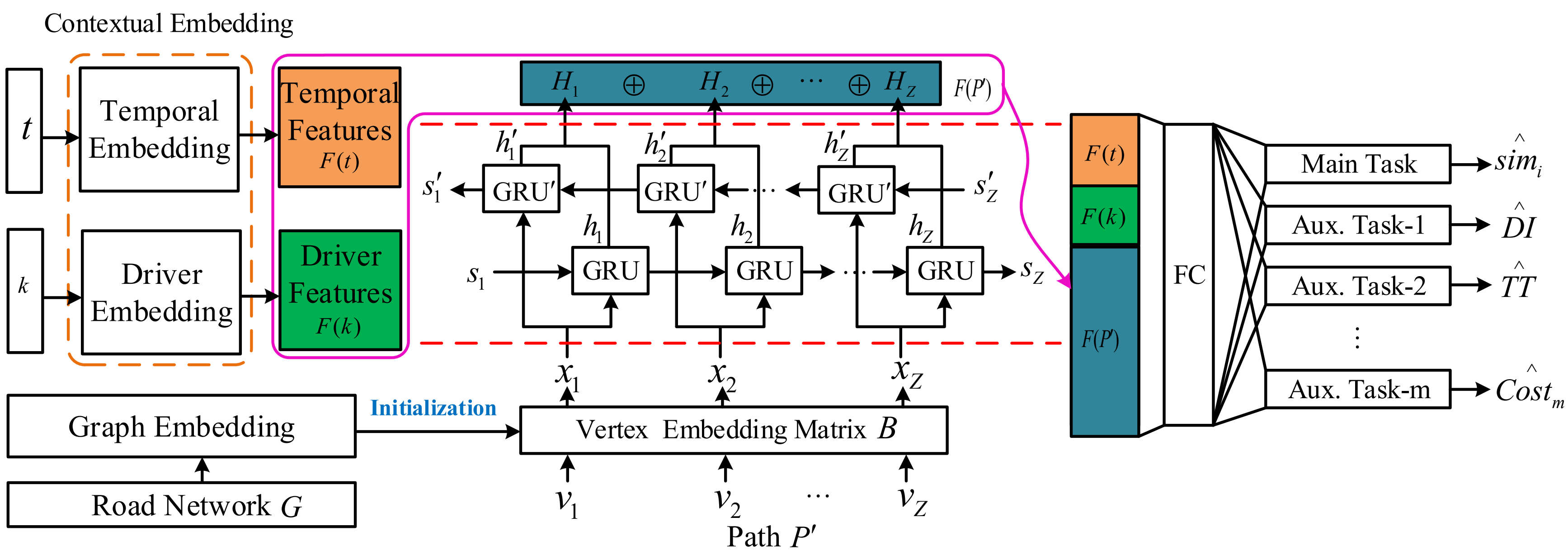
$\widehat{DI}$ : Travel Distance Estimation

$\widehat{TT}$ : Travel Time Estimation

$\widehat{Cost}_m$: Travel Cost Estimation

# Context Aware PathRank

- Departure time
- Driver ID



# Experiments

- Dataset
  - Aalborg, Denmark
- Evaluation Metrics
  - MAE, MARE
  - Kendall/Spearman's Rank Correlation Coefficient ($\tau, \rho$)
- Baselines
  - Linear Regression (LR)
  - Lasso Regression
  - Support Vector Regression
  - Decision Tree Regression
  - Decision Tree Regression with AdaBoost
  - LSTM

# Experiments

- Comparison with Regression Baselines

| | Methods | MAE | MARE | $\tau$ | $\rho$ |
|---|---|---|---|---|---|
| BF | **LR** | 0.2640 | 0.4012 | 0.6879 | 0.7150 |
| | **Lasso** | 0.2876 | 0.4371 | 0.6245 | 0.6678 |
| | **SVR** | 0.2390 | 0.3632 | 0.6543 | 0.6683 |
| | **DT** | 0.2516 | 0.3824 | 0.6530 | 0.6777 |
| | **DTA** | 0.2686 | 0.4082 | 0.6784 | 0.7135 |
| AF | **LR** | 0.3430 | 0.5213 | 0.0864 | 0.0854 |
| | **Lasso** | 0.2955 | 0.4484 | 0.6260 | 0.6686 |
| | **SVR** | 0.3369 | 0.5120 | 0.0857 | 0.0846 |
| | **DT** | 0.4141 | 0.6284 | 0.4050 | 0.0693 |
| | **DTA** | 0.4301 | 0.6527 | 0.0812 | 0.0395 |
| Deep Learning | **LSTM** | 0.2682 | 0.4076 | 0.4569 | 0.4619 |
| | **PRC** | 0.0611 | 0.0929 | 0.8178 | 0.8454 |

# AGENDA

# Motivation-Unsupervised Path Representation

- Why Un-Supervised Path Representation Learning?
  - There are not many labelled data available
  - Labelling data is expensive and time consuming
  - Taking advantage of a large amount of unlabeled data
  - Enhancing the supervised-learning using pre-training

# Motivation-Un-supervised Path Representation

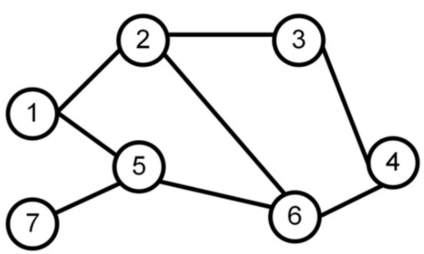
- Problem Definition
  - Given a set of path P in graph G, Path Representation Learning (PRL) aims at learning a path representation vector $\mathrm{p_i} \in R^{D'}$ for each path $P_i \in P$.

$$PE_{\psi}: X_i \in R^{Z \times D} \rightarrow p_i \in R^{D'}$$

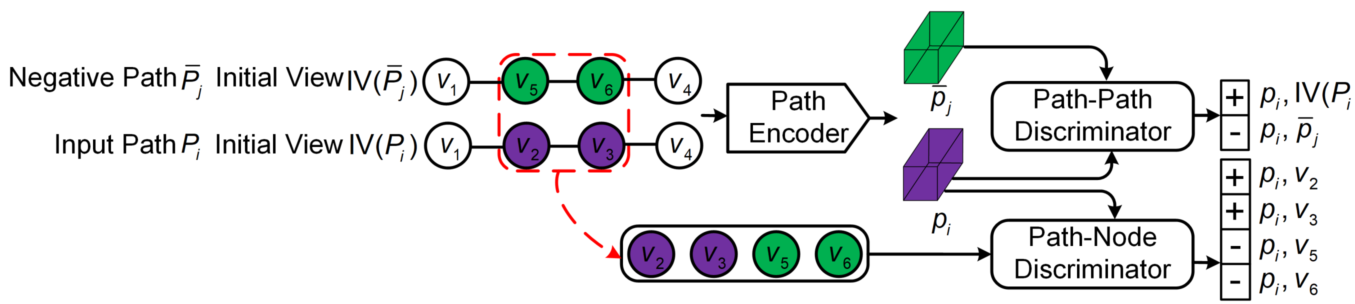
    - $\psi$: learnable parameters for the path encoder
    - $Z$: the length of path $P_i$
    - $D'$: dimension of the learned path representation vector $p_i$

# Path InfoMax (PIM)

- PIM
  - Path Encoder
  - Path-Path Discriminator
  - Path-Node Discriminator




Road Network




Path InfoMax Framework

$v_i$ : Node representation (Node2vec, Grover et al., 2016).

: Path representation.

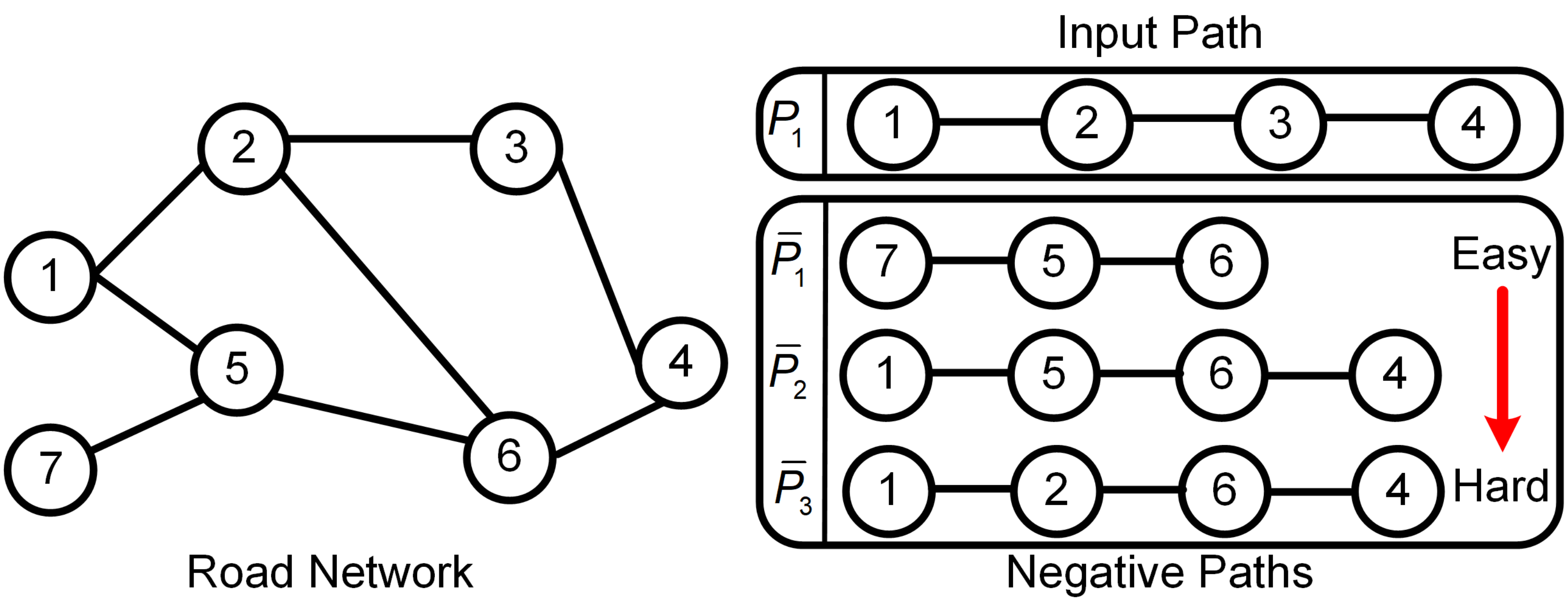
Path Encoder : Take a sequence of nodes as input, output path representation.

Path-Path Discriminator : Take a negative and positive path representation pair (i.e. $\langle (p_i, \bar{p}_j), - \rangle$) as input and output the corresponding label.

Path-Node Discriminator : Take a negative and positive pair as input and output the corresponding label.

# Path InfoMax (PIM)

- Curriculum Negative Sampling
  - Diversified Top-k shortest path algorithm is used to generate path from same source and destination.



# Experiments

- Datasets
  - Aalborg, Denmark
  - Harbin, China
- Downstream Tasks
  - Path Travel Time Estimation
    - MAE, MARE, MAPE
  - Path Ranking
    - MAE, Kendall / Spearman's Rank Correlation Coefficient ($\tau, \rho$)
- Regression Model
  - Gaussian Process Regressor (GPR)

# Experiments

- BaseLines
  - Node2vec (Grover et al., 2016),
  - Deep Graph InfoMax (DGI) (Velickovic et al., 2019)
  - Graphical Mutual Information Maximization(GMI) (Peng et al., 2020)
  - Memory Bank (MB) (Wu et al., 2018)
  - InfoGraph (Sun et al., 2020)
  - BERT (Devlin et al., 2019)
  - PathRank (Yang et al., 2020)

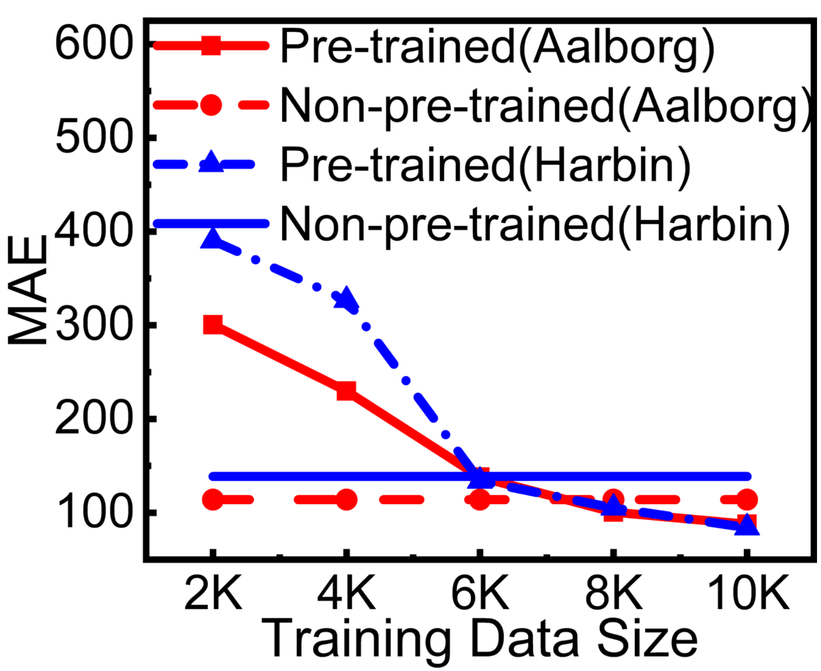
# Experiments

- Accuracy on Travel Time estimation and Ranking score estimation (Aalborg)

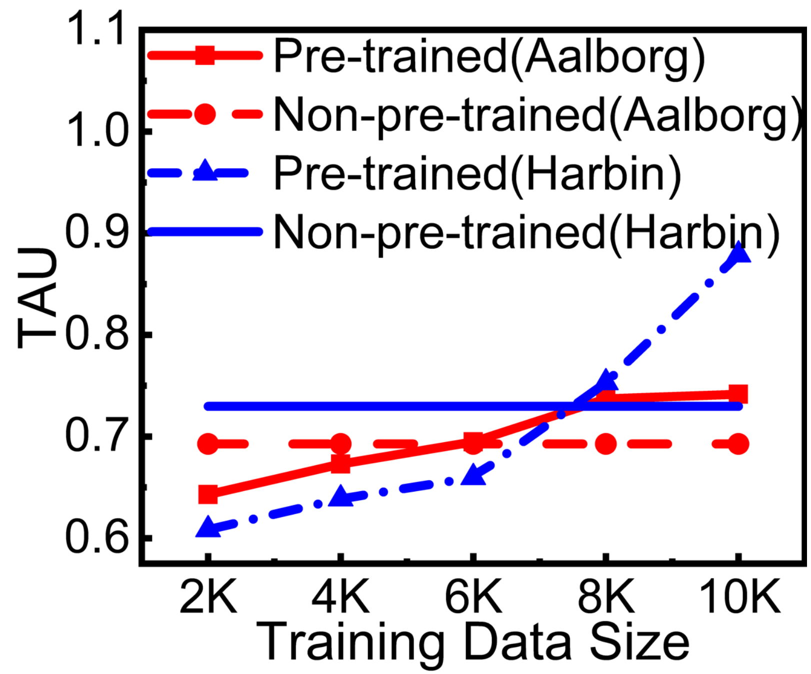
| Method | Travel Time Estimation | | | Path Ranking | | |
|---|---|---|---|---|---|---|
| | MAE | MARE | MAPE | MAE | $\tau$ | $\rho$ |
| *Node2vec* | 121.43 | 0.27 | 31.04 | 0.18 | 0.66 | 0.70 |
| *DGI* | 192.63 | 0.42 | 82.44 | 0.54 | 0.49 | 0.52 |
| *GMI* | 136.58 | 0.30 | 50.81 | 0.23 | 0.58 | 0.61 |
| *MB* | 243.97 | 0.53 | 84.17 | 0.35 | 0.34 | 0.38 |
| *BERT* | 254.17 | 0.54 | 61.61 | 0.36 | 0.38 | 0.39 |
| *InfoGraph* | 132.28 | 0.29 | 39.47 | 0.17 | 0.69 | 0.73 |
| *PIM* | 76.10 | 0.16 | 17.28 | 0.12 | 0.72 | 0.76 |

# Experiments

- Effects of Pre-training
  - Less data but same performance
  - Same data but better performance
  - Enhance the supervised learning




Travel Time Estimation

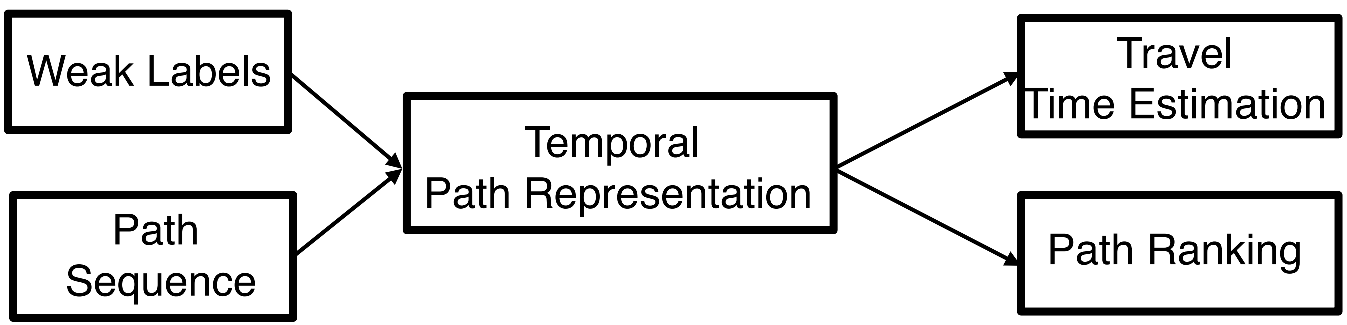



Path Ranking

# AGENDA

# Motivation-Temporal Path Representation

- Why weakly-supervised path representation learning?
  - Use weak labels
    - Inexpensive to achieve (compare to “strong” task specific labels)
    - Relevant to departure times
    - Relevant to various downstream tasks
    - Examples: peak vs. off-peak hours; traffic congestion indices



# Problem Definition

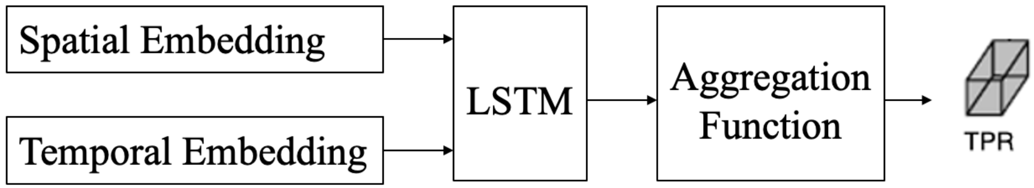
- Given a set of temporal paths TP = $\{tp_1, tp_2, \dots, tp_n\}$, where each temporal path $tp_i$ is with a weak label $y_i$ , temporal path representation learning (*TPRL*) aims to learn a temporal path representation $TPRL(tp_i)$ for each temporal path $tp_i \in TP$.

$$TPRL_{\psi}(tp_i): R^{d_{tem}} \times R^{M \times d} \rightarrow R^{d_h}$$

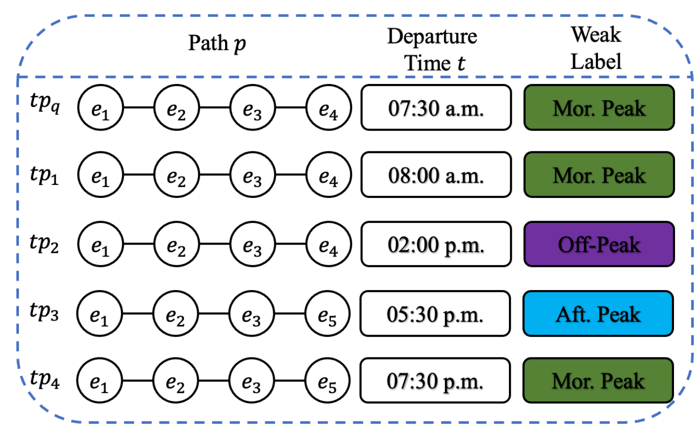
  - $\psi$: learnable parameters for the path encoder
  - $M$: is the total number of edges in the path
  - $d$, $d_{tem}$, $d_h$: edge feature dimension, a departure time embedding, and a resulted temporal path representation.
- We propose a basic framework and an advanced framework to solve the problem.

# Basic-Weakly-supervised Contrastive(WSC)

- Temporal Path Encoder

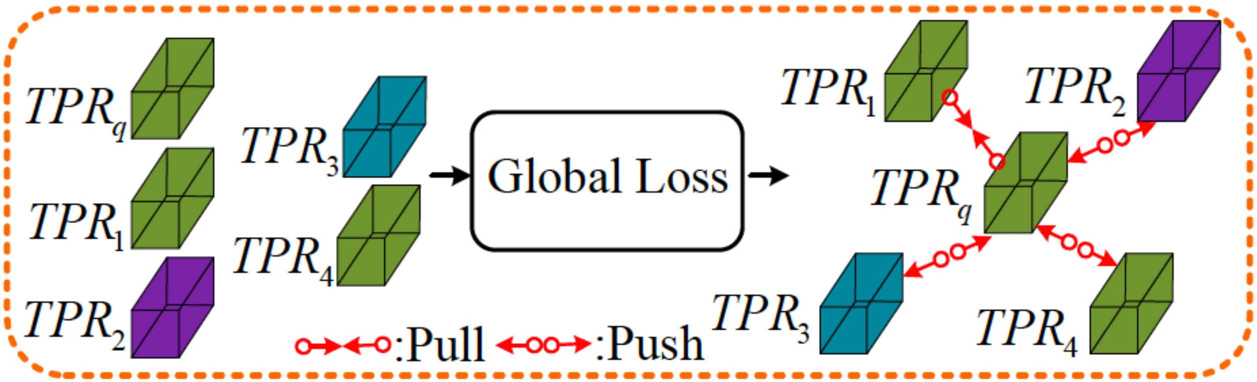



- Train the temporal path encoder using contrastive learning to obtain generic temporal path representations (TPRS).
  - Positive Samples
    - Positive temporal paths are not only different representations of the same temporal path, but also include temporal paths that traverse the same path with the same weak label.
    - $tp_1$ is the positive samples
  - Negative Samples
    - Same paths but different weak labels
    - Different paths but the same weak labels
    - Different paths and different weak labels
    - $tp_2$ , $tp_3$, $tp_4$ are the negative samples



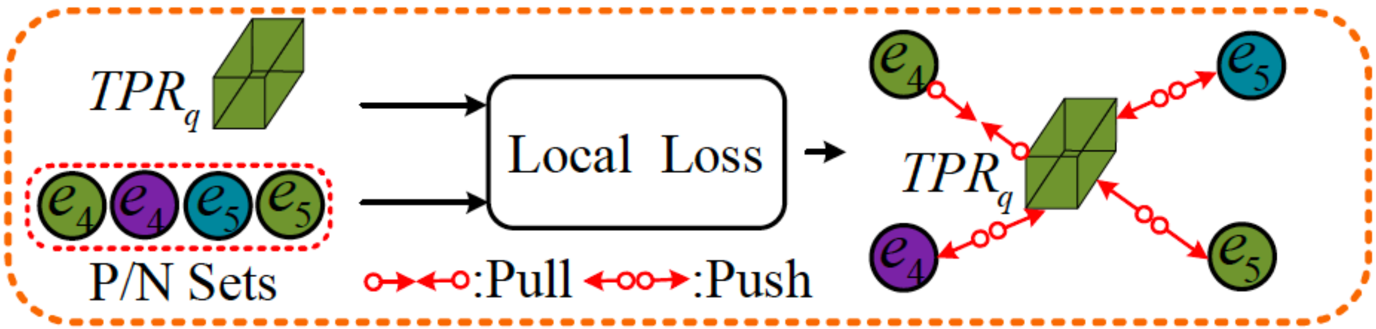
# Basic-Weakly-supervised Contrastive(WSC)

- Global contrastive loss
  - Pull closer TPRs with positive temporal path samples, while push away TPRs from negative temporal path samples.



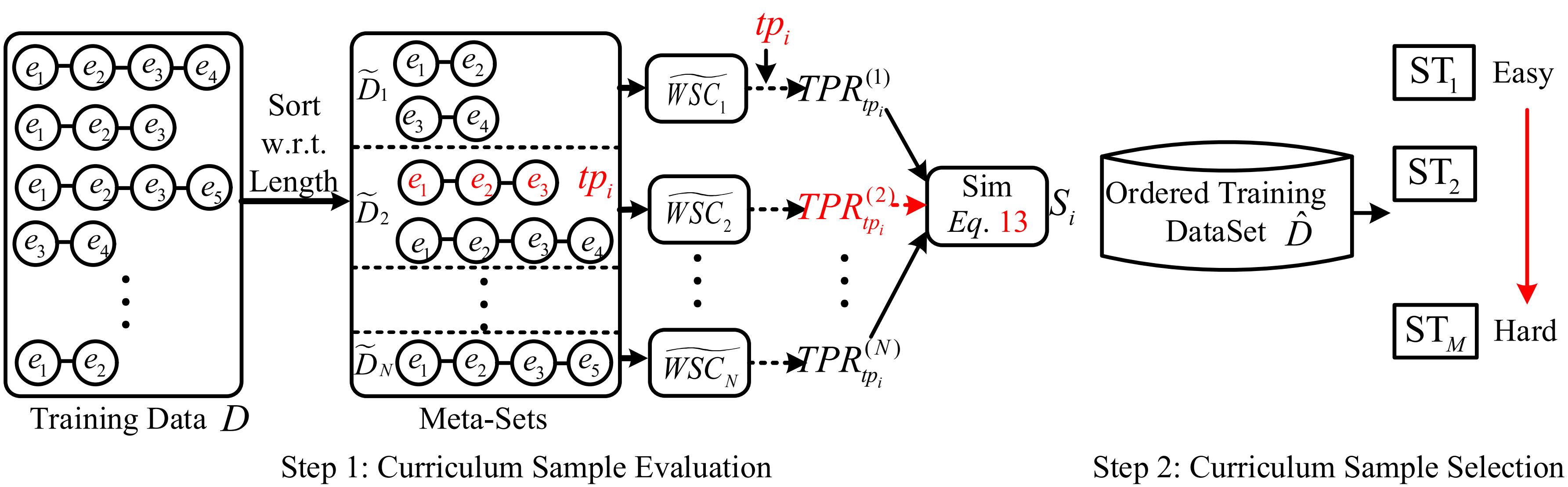
# Basic-Weakly-supervised Contrastive(WSC)

- Local contrastive loss
  - Pull closer TPRs with positive temporal edges samples, while push away TPRs from negative temporal edges samples.



# Advanced-Contrastive Curriculum Learning (WSCCL)

- Contrastive Curriculum Learning (CCL)



# Experiments

- Datasets
  - Aalborg, Denmark
  - Harbin, China
  - Chengdu, China
- Weak Labels
  - Peak hours and off-peak hours
  - Traffic congestion indices

# Experiments

- Downstream Tasks
  - Path Travel Time Estimation
    - MAE；MARE；MAPE
  - Path Ranking
    - MAE; Kendall/Spearman's Rank Correlation Coefficient ($\tau, \rho$)
  - Path Recommendation
    - Classification Accuracy (Acc.); Hit Rate (HR)
- Regression and Classification Model
  - Gradient Boosting Regressor (GBR)
  - Gradient Boosting Classifier (GBC)

# Experiments

- BaseLines
  - Unsupervised learning based
    - Node2vec (Grover et al., 2016)
    - Deep Graph InfoMax (DGI) (Velickovic et al., 2019)
    - Graphical Mutual Information Maximization(GMI) (Peng et al., 2020)
    - Memory Bank (MB) (Wu et al., 2018)
    - InfoGraph (Sun et al., 2020)
    - BERT (Devlin et al., 2019)
    - PIM (Yang et al., 2021)
  - Supervised learning based
    - DeepGTT (Li et al., 2019)
    - HMTRL (Liu et al., 2021)
    - PathRank (Yang et al., 2020)

# Experiments

- Performance Comparison (Aalborg)

| | Travel Time Estimation | | Path Ranking | | Path Recommendation | |
|---|---|---|---|---|---|---|
| | MAE | MAPE | $\tau$ | $\rho$ | Acc. | HR |
| *Node2vec* | 63.82 | 45.67 | 0.23 | 0.60 | 0.79 | 0.51 |
| *DGI* | 67.22 | 49.36 | 0.24 | 0.60 | 0.74 | 0.55 |
| *GMI* | 70.61 | 52.40 | 0.24 | 0.59 | 0.78 | 0.53 |
| *MB* | 57.32 | 39.37 | 0.23 | 0.62 | 0.67 | 0.48 |
| *BERT* | 71.96 | 45.42 | 0.26 | 0.49 | 0.60 | 0.43 |
| *InfoGraph* | 69.36 | 41.28 | 0.26 | 0.52 | 0.72 | 0.69 |
| *PIM* | 57.66 | 39.34 | 0.22 | 0.60 | 0.79 | 0.82 |
| *DeepGTT* | 44.78 | 26.53 | 0.39 | 0.12 | ---- | ---- |
| *HMTRL* | 40.59 | 21.81 | 0.17 | 0.65 | 0.80 | 0.86 |
| *PathRank* | 37.09 | 23.89 | 0.23 | 0.64 | 0.77 | 0.71 |
| *WSCCL* | 31.66 | 31.39 | 0.15 | 0.68 | 0.82 | 0.88 |

# Experiments

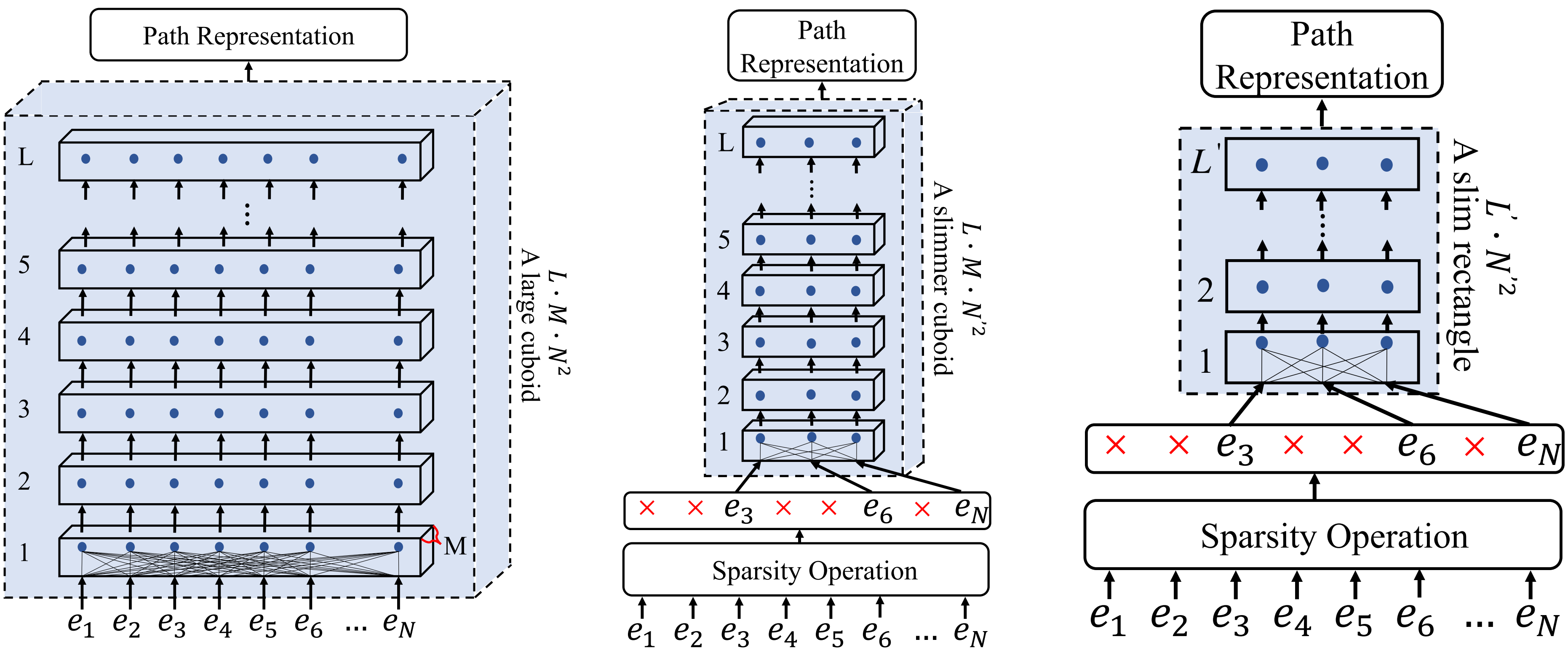
- Temporal Enhanced PIM (Aalborg)

| | **Travel Time Estimation** | | | **Path Ranking** | | |
|---|---|---|---|---|---|---|
| | MAE | MARE | MAPE | MAE | $\tau$ | $\rho$ |
| *PIM-Temporal* | 42.27 | 0.19 | 27.95 | 0.19 | 0.65 | 0.70 |
| *WSCCL* | 31.66 | 0.13 | 21.39 | 0.15 | 0.68 | 0.72 |

# AGENDA

# LightPath

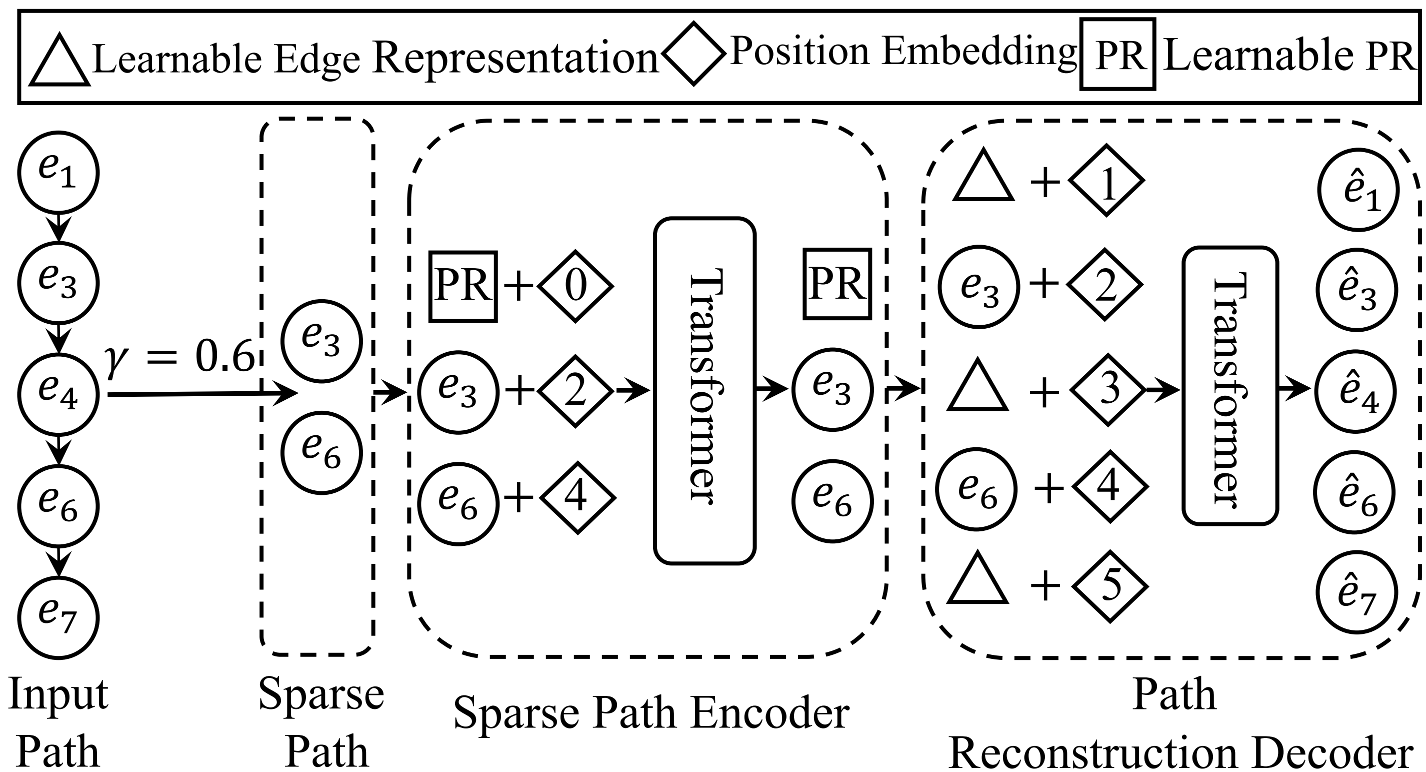
- Poor scalability w.r.t. path length
- Very large mode size




(a) Traditional Transformer Encoder (b) Sparse Transformer Encoder (c) LightPath

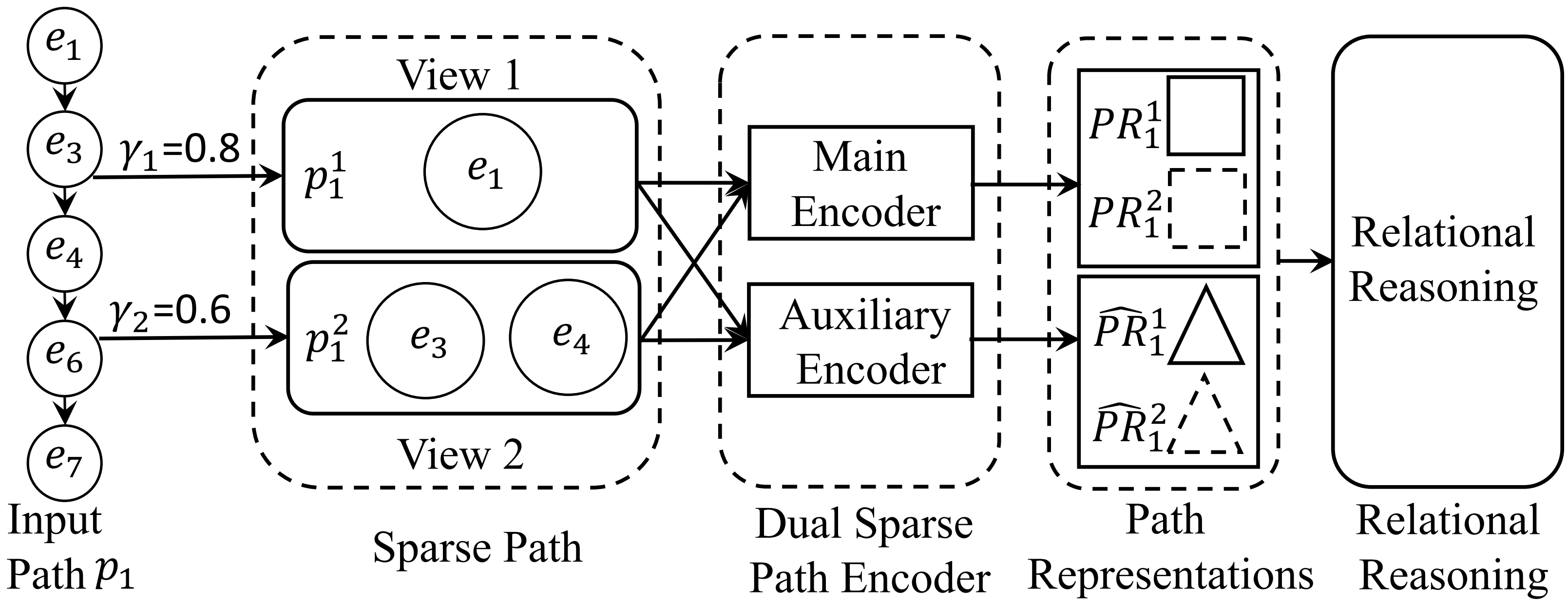
# Sparse Path Auto-Encoder

- Sparsity operation
- Learnable path representation
- Sparse path encoder
- Path reconstruction decoder



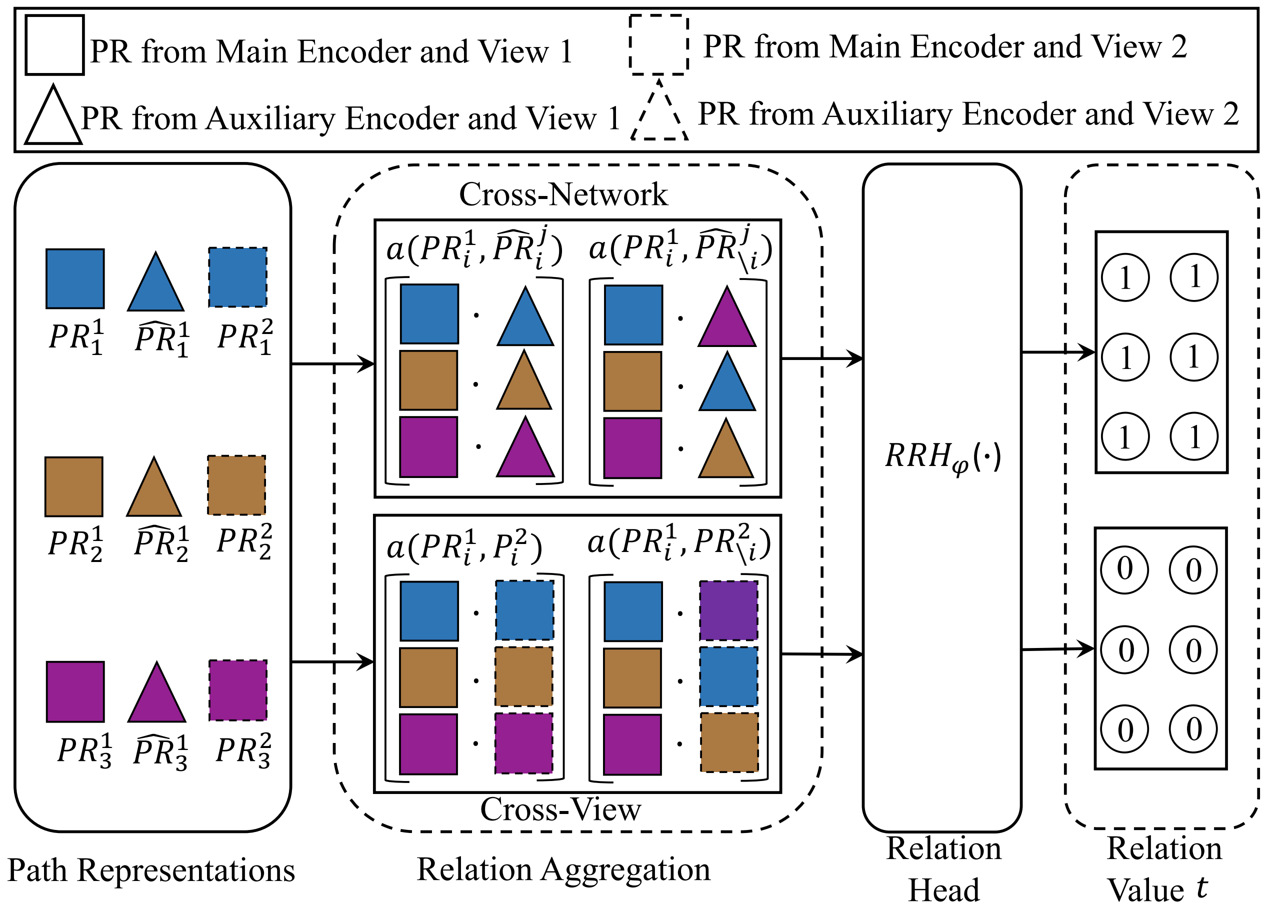
# Relational Reasoning PRL

- Dual Sparse Path Encoder
  - Construct Path Representations



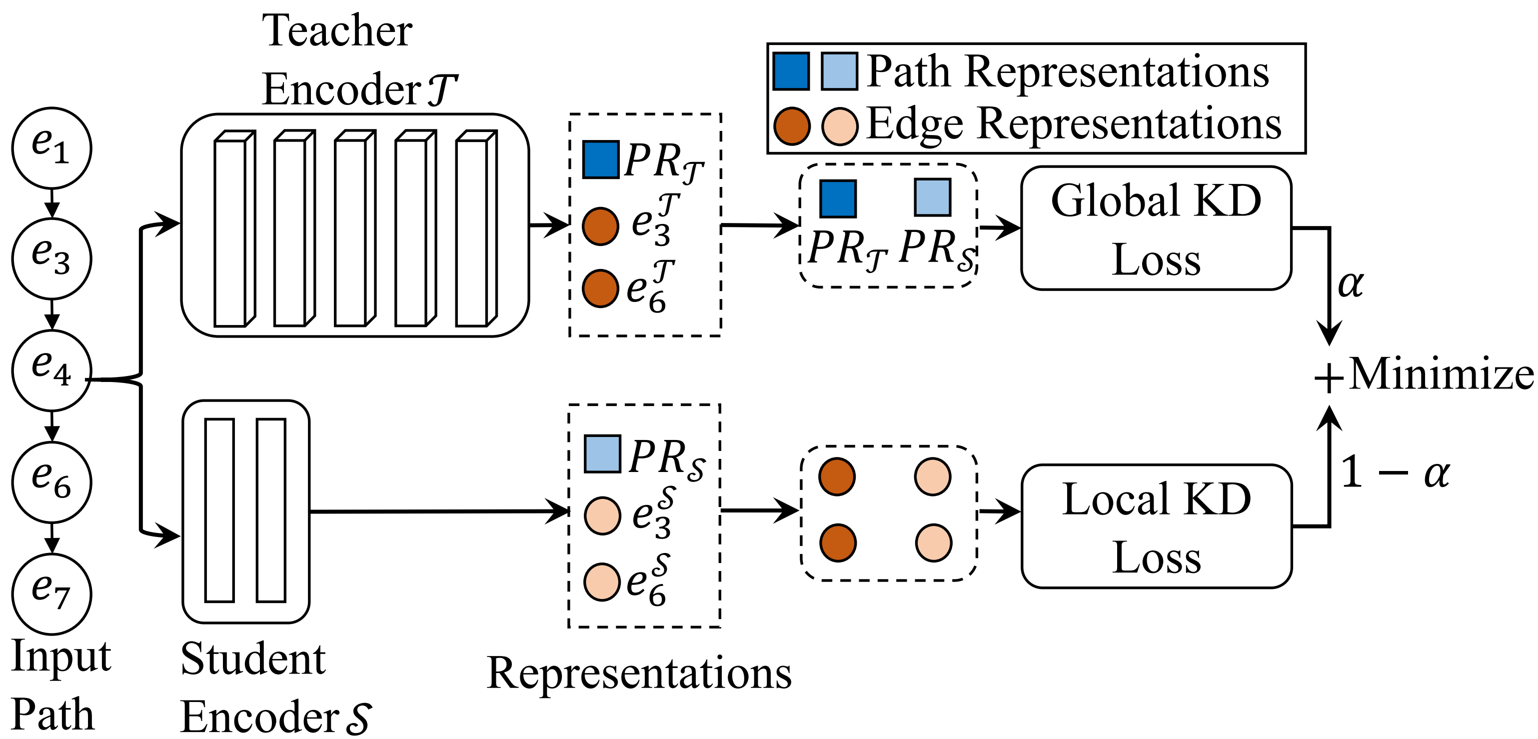
# Relational Reasoning PRL

- Relational Reasoning
  - Cross-Network Relational Reasoning
  - Cross-View Relational Reasoning



# Global Local Knowledge Distillation (GLKD)

- To reduce mode size
  - Global-path representation Distillation
  - Local-edge Correlation Distillation



# Experiments

- Datasets
  - Aalborg, Denmark
  - Chengdu, China
- Downstream Tasks
  - Path Travel Time Estimation (MAE, MARE, MAPE)
  - Path Ranking (MAE, Kendall/Spearman's Rank Correlation Coefficient ($\tau, \rho$))
- Regression Model
  - Gradient Boosting Regressor (GBR)
- Baselines
  - Node2vec,
  - MoCo
  - Toast; t2vec; NeuTraj
  - PIM
  - HMTRL; PathRank; LightPath-Sup

# Experiments

- Accuracy on Travel Time Estimation and Ranking Score Estimation

| Method | Travel Time Estimation | | | Path Ranking | | |
|---|---|---|---|---|---|---|
| | MAE | MARE | MAPE | MAE | $\tau$ | $\rho$ |
| *Node2vec* | 154.07 | 0.20 | 25.22 | 0.24 | 0.59 | 0.64 |
| *MoCo* | 146.29 | 0.19 | 21.60 | 0.25 | 0.53 | 0.57 |
| *Toast* | 137.27 | 0.17 | 20.43 | 0.24 | 0.59 | 0.63 |
| *t2vec* | 147.24 | 0.19 | 22.13 | 0.25 | 0.52 | 0.56 |
| *NeuTraj* | 117.06 | 0.15 | 18.09 | 0.25 | 0.60 | 0.64 |
| *PIM* | 102.09 | 0.14 | 14.92 | 0.20 | 0.63 | 0.67 |
| *HMTRL* | 101.81 | 0.13 | 14.51 | 0.17 | 0.68 | 0.72 |
| *PathRank* | 115.37 | 0.15 | 16.41 | 0.21 | 0.64 | 0.68 |
| *LightPath-Sup* | 105.51 | 0.15 | 16.35 | 0.14 | 0.68 | 0.72 |
| *LightPath* | 85.76 | 0.11 | 12.12 | 0.13 | 0.73 | 0.77 |

# Experiments

- Model Scalability vs. Reduction Ratio ($\gamma$) and Path Length (N)

| $N$ | *LightPath* | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\gamma = 0$ | | $\gamma = 0.1$ | | $\gamma = 0.3$ | | $\gamma = 0.5$ | | $\gamma = 0.7$ | | $\gamma = 0.9$ | | Para. |
| | GFLOPs | gMem. | GFLOPs | gMem. | GFLOPs | gMem. | GFLOPs | gMem. | GFLOPs | gMem. | GFLOPs | gMem. | |
| 50 | 8.01 | 1.47 | 7.48 | 1.39 | 6.44 | 1.37 | 5.39 | 1.36 | 4.34 | 1.34 | 3.18 | 1.33 | 1.570 |
| 100 | 15.77 | 1.60 | 14.72 | 1.50 | 12.62 | 1.48 | 10.52 | 1.46 | 8.43 | 1.44 | 6.23 | 1.43 | 1.570 |
| 150 | 23.53 | 1.72 | 21.95 | 1.64 | 18.81 | 1.60 | 15.66 | 1.58 | 12.52 | 1.55 | 9.27 | 1.52 | 1.570 |
| 200 | 31.29 | 1.90 | 29.19 | 1.81 | 24.99 | 1.77 | 20.80 | 1.73 | 16.61 | 1.68 | 12.31 | 1.65 | 1.570 |
| | *LightPath w/o KD* | | | | | | | | | | | | |
| | $\gamma = 0$ | | $\gamma = 0.1$ | | $\gamma = 0.3$ | | $\gamma = 0.5$ | | $\gamma = 0.7$ | | $\gamma = 0.9$ | | Para. |
| | GFLOPs | gMem. | GFLOPs | gMem. | GFLOPs | gMem. | GFLOPs | gMem. | GFLOPs | gMem. | GFLOPs | gMem. | |
| 50 | 33.68 | 1.78 | 30.64 | 1.70 | 22.55 | 1.61 | 18.47 | 1.53 | 12.39 | 1.47 | 5.70 | 1.41 | 5.525 |
| 100 | 66.60 | 2.53 | 60.52 | 2.37 | 48.36 | 2.11 | 36.19 | 1.86 | 24.03 | 1.72 | 11.26 | 1.58 | 5.525 |
| 150 | 99.53 | 3.44 | 90.41 | 3.23 | 72.16 | 2.76 | 53.91 | 2.35 | 35.65 | 2.03 | 16.82 | 1.82 | 5.525 |
| 200 | 132.54 | 4.74 | 120.29 | 4.30 | 95.96 | 3.53 | 71.64 | 2.94 | 47.31 | 2.43 | 22.37 | 2.10 | 5.525 |

# AGENDA

# Challenges Addressed

- Address path ranking in supervised multi-task learning.
- Learn task-unspecific path representations through mutual information maximization.
- Learn temporal path representation through weakly-supervised contrastive curriculum learning.
- Address the lightweight and scalable path representation learning based on relational reasoning and knowledge distillation.

# Future work

- Stochastic Path Representation Learning
- AutoPRL

Thanks!

Q&A