A Time-Window GNN-Based Network Partition for Identifying High- and Low-Risk Nodes in Road Networks

Simone Paradiso

IVORY

PhD Candidate and Researcher

oseven

Together with:
Apostolos Ziakopoulos, Haris Sideris and George Yannis

Department of Transportation Planning and Engineering
National Technical University of Athens

WCTR World Conference on Transport Research

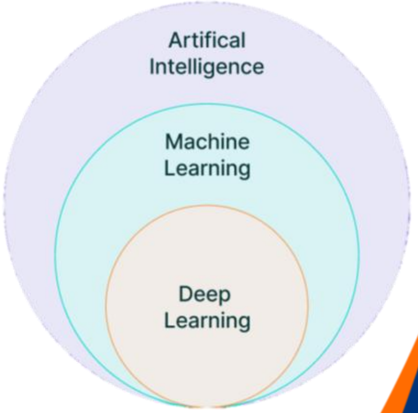
World Conference of Transport Research, Toulouse 2026 (WCTR 2026 Toulouse)
6-10 July 2026, Toulouse, France

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590

Introduction

- Road crashes claim 1.3 million lives annually, the leading cause of death for those under 29 and among the top 10 globally.
- IVORY framework.
  - European Union's Horizon Europe research and innovation programme Marie Skłodowska-Curie Industrial Doctorates (grant No 101119590).
  - It develops fair and explainable Artificial Intelligence (AI) to enhance road safety while sharing knowledge.
  - DC9 focuses on creating an AI framework to analyze road safety KPIs and explore multiscale approaches within the framework.
- Traditional crash prediction relies on econometrics; now enhanced by Machine Learning (ML) & Deep Learning (DL), with Graph Neural Networks (GNNs) extending DL to graph-structured data.

Marie Skłodowska-Curie Actions



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590

IVORY oseven

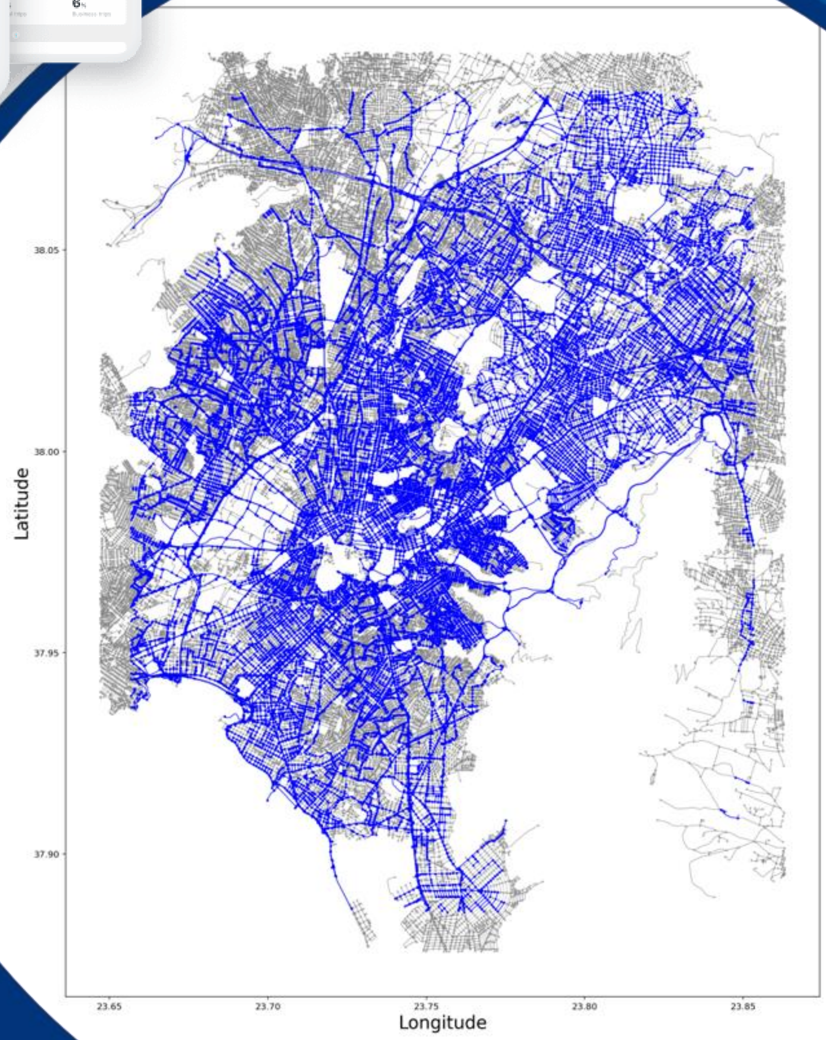
Data Sources

- OSeven Telematics provided **telematics data** collected via smartphone hardware sensors, anonymized and compliant with Greek and European personal **data protection** regulations (GDPR).
  - Raw data are processed by **proprietary machine learning** algorithms.
  - Reliability is validated against OBD data, on-road tests, simulators, and literature benchmarks.

GDPR COMPLIANT

- OpenStreetMap is a free, editable **global map** created by volunteers and released under an open-content license.

OpenStreetMap



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590

IVORY oseven

Telematics Aggregation

- Based on the telematics data, a coordinate **bounding box** was defined and used to extract a structured **graph** from OpenStreetMap via the **OSMnx** Python library.
- From the graph, **node** and **edge** features were stored in two different datasets.

Node feature aggregation

- Data points **within a 50m buffer radius** around each node were used to characterize it, limited to those **on its connected edges**.

Edge feature aggregation

- Each raw data point was matched to its **nearest edge**.

Features	Description
Street_Count	Number of streets connected to the intersection
SmoothenedSpeed	Average speed of vehicles near the node
SpeedingFlag	Count of speeding events near the node
Mobile_usage	Number of instances of phone usage near the node
Harsh_acc	Number of harsh acceleration events near the node
Harsh_brk	Number of harsh braking events near the node
Event_intensity	Average intensity of harsh events near the node
Trips_count	Number of trips recorded near the node

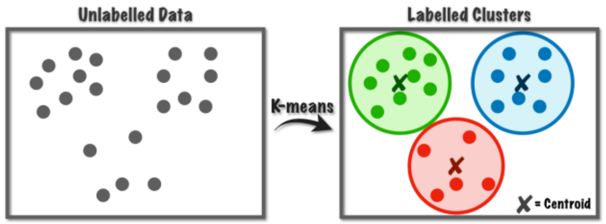
This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590

K-Means Clustering

- Given a dataset of points $\{x_1, x_2, \ldots, x_n\}$, K-Means aims to partition them into K, clusters $\{C_1, C_2, \ldots, C_k\}$, by minimizing the total within-cluster sum of squared distances:

$$\underset{C}{arg\,min} \sum_{k=1}^{K} \sum_{x_i \in C_k} ||x_i - \mu_k||^2$$

Where μ_k is the centroid (mean) of cluster C_k, and $||x_i - \mu_k||^2$ is the squared Euclidean distance.



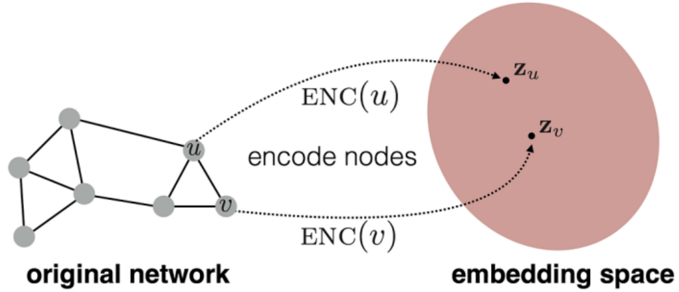
- K-Means: efficient and widely used.
  - Linear time complexity, fast convergence, low memory use.
  - Easy to implement and interpret, using clear centroids and simple assignments.
  - It is sensitive to initial centroids (can converge to local minima).

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590

IVORY oseven

Graph Neural Networks: Introduction to GATs

- GNNs learn compact vector representations of nodes that capture node features, structural roles and neighborhood context.



- Advancements include Graph Convolutional Networks (GCN) using convolution operations, and Graph Attention Networks (GAT) applying attention mechanisms.
  - GATv1 — computes attention using transformed node features before interaction

$$\alpha_{ij} = softmax_j\left(LeakyReLU\left(\bar{a}^T[\boldsymbol{W}\bar{h}_i||\boldsymbol{W}\bar{h}_j]\right)\right)$$

  - GATv2 — computes attention after combining node features

$$\alpha_{ij} = softmax_j\left(\bar{a}^T LeakyReLU\left(\boldsymbol{W}[\bar{h}_i||\bar{h}_j]\right)\right)$$

  - Attention coefficients (α) let nodes weigh neighbors unequally and can also incorporate edge features.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590

Self-Supervised Node Embedding with GATv2

Model Architecture

- Using two GATv2 layers, each followed by LeakyReLU.
- Using a multi-head attention (3) to stabilize training and improve accuracy.

Training Strategy

- Self-supervised contrastive learning objective to learn meaningful node embeddings.
  - Maximize similarity to true neighbors while minimizing similarity to sampled non-neighbors.
- Each weekly model is trained for 10 epochs, optimized with Adam.
  - Generating weekly node embeddings.

1. For each node $i = 1$ to N.
   a. Calculate the cosine similarity with its neighbours ${S_{ij}}^{+}$ and with a randomly sampled set of non-neighbors ${S_{ij}}^{-}$.
   b. Scale these similarity scores using a temperature parameter and apply the exponential function to the scaled values to obtain $e^{\frac{{S_{ij}}^{+}}{\tau}}$ and $e^{\frac{{S_{ij}}^{-}}{\tau}}$.
   c. Sum the values over neighbours $\sum_{j\in P_i} e^{\frac{{S_{ij}}^{+}}{\tau}}$ and non-neighbours separately $\sum_{j\in N_i} e^{\frac{{S_{ij}}^{-}}{\tau}}$.
   d. Compute the ratio of neighbour sum to total sum, then apply logarithm:

$$\mathcal{L}_i = -\log\left(\frac{\sum_{j\in P_i} e^{\frac{{S_{ij}}^{+}}{\tau}}}{\sum_{j\in P_i} e^{\frac{{S_{ij}}^{+}}{\tau}} + \sum_{j\in N_i} e^{\frac{{S_{ij}}^{-}}{\tau}}}\right).$$

2. Average over all nodes $\mathcal{L}_{TOT} = \frac{\sum_i \mathcal{L}_i}{N}$.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590

Inputs to K-Means Clustering

Two possible representations of road network nodes

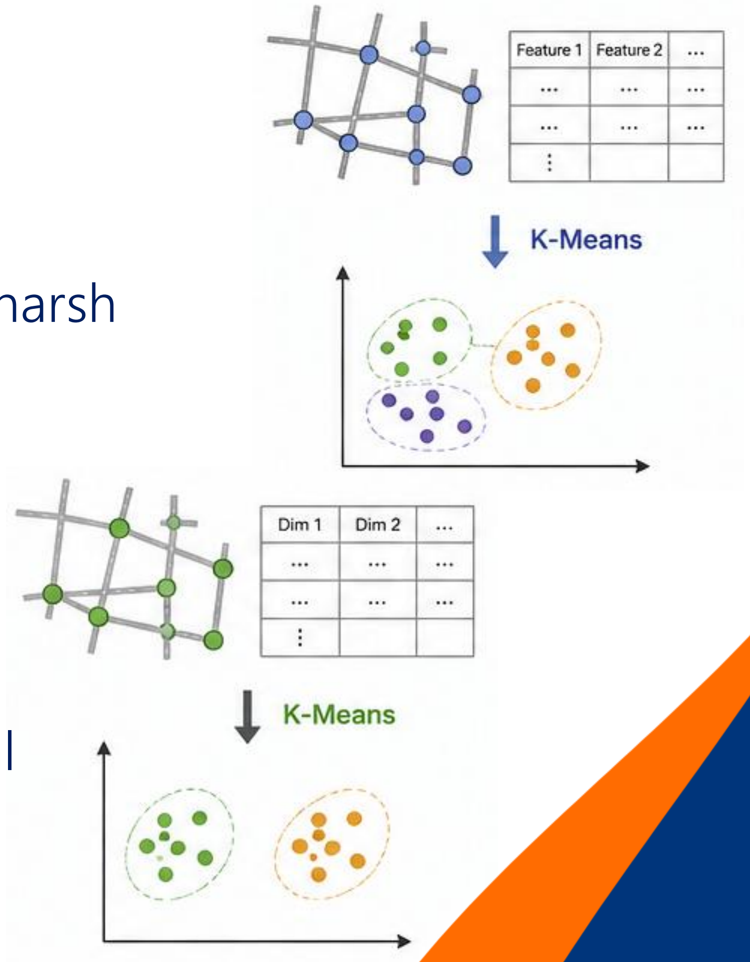
It is possible to cluster nodes using either:

(i) Raw node features

- Direct attributes of each node (e.g., speed, street number, harsh brakings)
- No graph structure considered

(ii) Learned node embeddings

- Low-dimensional vectors learned from the graph via GNN
- Encode connectivity, spatial/graph structure, and functional similarity



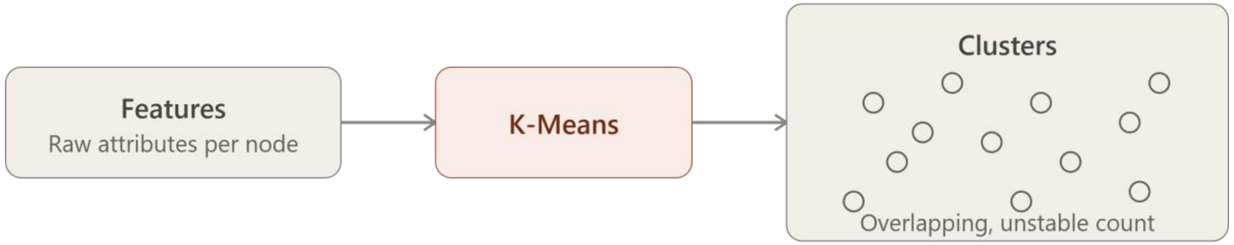
This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590

IVORY

oseven

Clustering Stability — Raw Features

- Raw features produce unstable, lower-quality clusters:
  - Unstable optimal K across weeks: values jump between 2, 3, 7, 9, 10
  - Volatile, lower silhouette scores: ranging ~0.37–0.65.

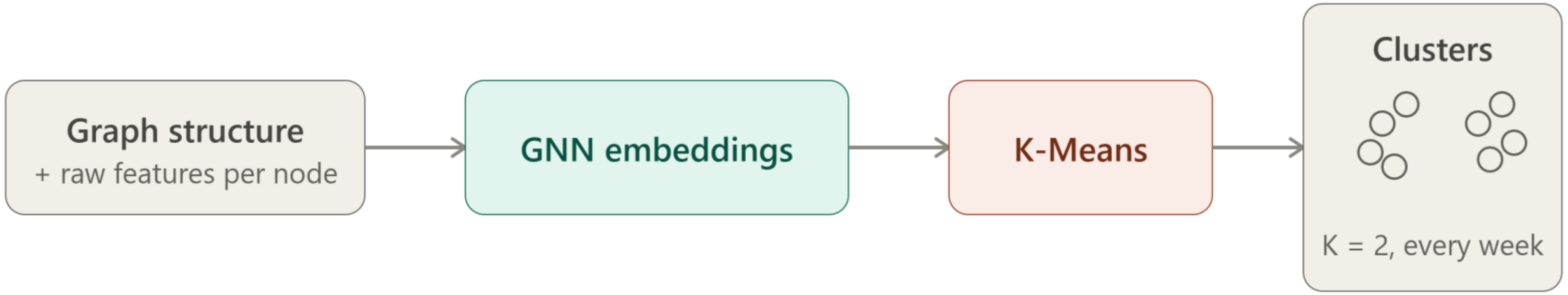


Week	Optimal K	Max Silhouette Score	Silhouette Score for K=2	Rank of K=2 in Silhouette Scores
2024-08-29	2	0.6116	-	-
2024-09-05	2	0.6079	-	-
2024-09-12	2	0.6042	-	-
2024-09-19	9	0.3905	0.2678	9
2024-09-26	10	0.3984	0.1890	9
2024-10-03	2	0.6264	-	-
2024-10-10	2	0.6101	-	-
2024-10-17	2	0.6227	-	-
2024-10-24	2	0.6137	-	-
2024-10-31	9	0.3940	0.2660	9
2024-11-07	2	0.6002	-	-
2024-11-14	3	0.5928	0.5871	2
2024-11-21	2	0.5758	-	-
2024-11-28	7	0.3715	0.2700	9
2024-12-05	2	0.6141	-	-
2024-12-12	2	0.5916	-	-
2024-12-19	2	0.6042	-	-
2024-12-26	2	0.6588	-	-

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590

Clustering Stability — Embeddings (GNN)

- Learned embeddings produce stable, high-quality clusters:
  - The optimal K is consistently 2 across all weeks.
  - Silhouette scores are generally higher and more stable (e.g., ~0.62–0.91 range, with a strong peak at 0.91).

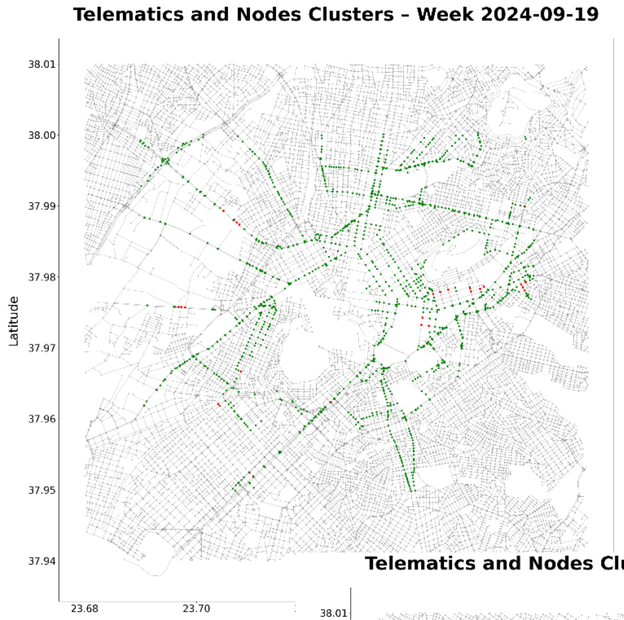


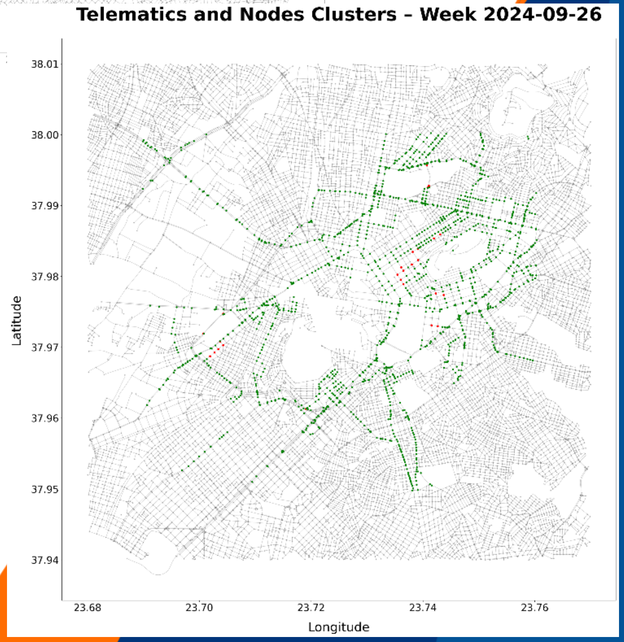
Week	Optimal K	Max Silhouette Score
2024-08-29	2	0.6271
2024-09-05	2	0.6425
2024-09-12	2	0.9137
2024-09-19	2	0.8520
2024-09-26	2	0.6506
2024-10-03	2	0.6323
2024-10-10	2	0.7660
2024-10-17	2	0.7175
2024-10-24	2	0.6463
2024-10-31	2	0.6222
2024-11-07	2	0.6028
2024-11-14	2	0.6362
2024-11-21	2	0.6482
2024-11-28	2	0.7543
2024-12-05	2	0.6695
2024-12-12	2	0.6176
2024-12-19	2	0.6915
2024-12-26	2	0.6728

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590

What Do the GNN Clusters Represent?

- Weekly **pseudo-cluster centroids** were reconstructed by **averaging** raw features within each cluster.
- Cluster 1 (Larger) — **Conservative drivers**
  - Stable driving behavior with lower speeds.
  - Minimal speeding and harsh braking/acceleration events.
  - Low phone usage while driving.
- Cluster 2 (Smaller) — **Aggressive drivers**
  - Higher speeds with frequent speeding events.
  - Elevated phone usage and harsh driving events.
  - Highly trafficked.



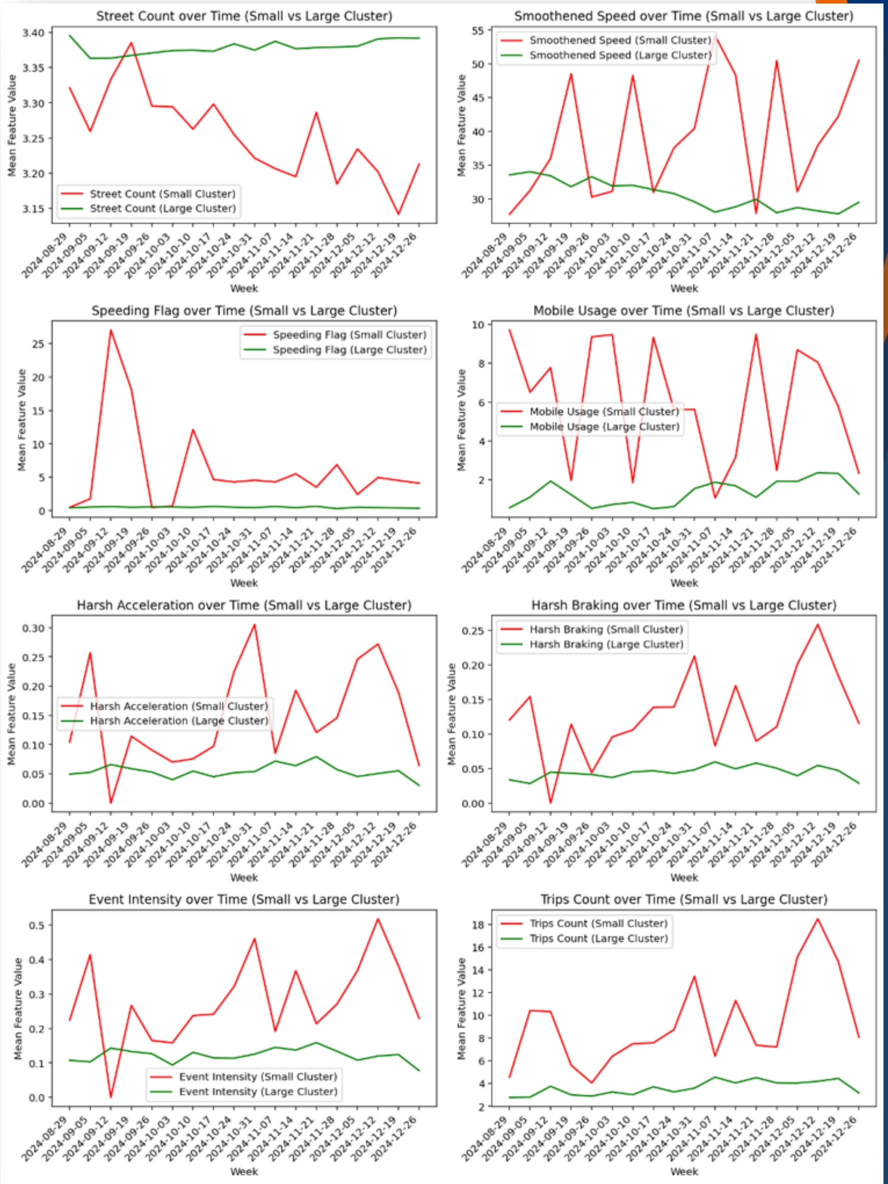


This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590

IVORY oseven

Weekly Changes in Cluster Behavior

- Larger cluster ↦ stable driving behaviors patterns across weeks near these intersections.
- Smaller clusters ↦ Greater week-to-week behavioral variability.
- Risk shifts are driven by different hazardous behaviors over time:
  - Sep 12 — Speeding flag is the differentiator.
    - The smaller cluster has fewer harsh events but far more speeding, signaling elevated risk despite a cleaner harsh-event record.
  - Nov 28 — Speed is the signal.
    - Phone use is similar across clusters; higher speeds define the riskier group.
  - Dec 5 — Phone use is the differentiator.
    - Speeds are comparable across clusters, but higher phone use stands out for the smaller cluster.



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590

IVORY oseven

Discussion

- Spatial road safety analyses can leverage **telematics data** and **clustering techniques** to characterize unsafe driving behavior across locations.
- GNN-based **node embeddings** capture network structure and neighborhood context, leading to **improved clustering** and **meaningful road network characterization**.
- Weekly segmentation captures **week-to-week changes** in hazardous events.
- The framework is **potentially transferable** to other study areas but may require **adaptation** to different telematics data.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590

IVORY

oseven

Potential Applications

➢ This approach enables proactive spatial risk analysis with node-based insights that show where to focus safety efforts and resources.

A diagnostic tool for road safety

- Shifts the paradigm from reactive crash reporting to proactive identification of unsafe locations.

Supports efficient resource allocation

- Larger cluster — stable behavior over time → lower intervention priority
- Smaller cluster — risk is alternately associated with harsh events, speeding, or mobile phone use → targeted mitigation strategies tailored to the dominant behaviour and the time window.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590

IVORY oseven

Conclusions

Key findings

- GNN embeddings improve clustering by capturing spatial context and network structure, producing more stable clusters with a consistent optimal cluster count over time.
- Clustering on node embeddings effectively separates spatial entities within the road network, better identifying locations where drivers behave more aggressively or carelessly
- Weekly segmentation reveals temporal trends, with different hazardous behaviors dominating the cluster composition from week to week.

Future Directions

- Test different GNN architectures and clustering techniques.
- Explore alternative temporal resolutions and buffer sizes.
- Normalize features by trip count to reduce traffic volume bias and better isolate behavioral patterns.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590

A Time-Window GNN-Based Network Partition for Identifying High- and Low-Risk Nodes in Road Networks

Simone Paradiso

IVORY

PhD Candidate and Researcher

oseven

Together with:
Apostolos Ziakopoulos, Haris Sideris and George Yannis

Department of Transportation Planning and Engineering
National Technical University of Athens

WCTR World Conference on Transport Research

World Conference of Transport Research, Toulouse 2026 (WCTR 2026 Toulouse)
6-10 July 2026, Toulouse, France

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590