SafeStreets-LDN: Graph Transformers for Crash Risk in London

Md Abdul Raihan Tanzim [1]

Abstract. Globally, road traffic crashes cause nearly 1.3 million preventable deaths and an estimated 50 million injuries each year – making it the leading killer of children and young people worldwide. As things stand, they are set to cause a further estimated 13 million deaths and 500 million injuries during the next decade and hinder sustainable development. These unacceptable numbers, both in absolute and relative terms, have remained largely unchanged for the past 20 years, despite the painstaking work of the United Nations and other road safety bodies. Road safety is a crisis that has gone on far too long. No road deaths are necessary or acceptable. Cars have been around for over 120 years, and we know how to prevent these tragedies. Yet road crashes still claim more than two lives every minute, and nearly 1.2 million lives every year. United Nations Sustainable Development Goals and the UN Decade of Action for Road Safety 2021-2030, the world has set an ambitious target of halving road deaths worldwide by 2030 (United Nations General Assembly 2020).

This study investigates whether the crash risk of road intersections can be predicted using only static, readily available geospatial data. We construct a Greater London pipeline by combining OpenStreetMap road networks with collision records from 2020–2024, mapping crashes to nearest intersections and creating occurrence and severity-based risk labels. The road system is modeled as a graph with node features capturing intersection topology and road types, and edge attributes describing travel properties (optionally including direction and bearing). We evaluate graph-based models, especially TransformerConv GNNs, against XGBoost and trivial baselines under spatially separated train/test splits, producing city-wide risk estimates for safety screening.Safe roads power economies. Road deaths can cost countries around 3 to 5 percent of GDP, and ensuring more people can move safely to their jobs, schools and vital services drives development. are a natural choice.

1 Introduction

London experiences a high volume of road traffic and a substantial number of collisions each year. Anticipating where severe or frequent crashes are most likely to occur can support targeted infrastructure improvements, speed management, and enforcement strategies. Traffic accidents remain a severe global public health crisis, as emphasized by the WHO in its 2023 Global Status Report on Road Safety (World Health Organization 2023). Despite a slight reduction in overall fatalities to 1.19 million annually (see Figure 1), traffic-related incidents are still the leading cause of death for individuals aged 5 to 29 years old, underscoring the ongoing gravity of the issue.

1.1 Problem Statement

Can the riskiness of London's road intersections be predicted accurately using only static, open geospatial information and the topology of the road network—without relying on dense proprietary traffic counts, sensors, or real-time mobility data? Current practice still leans heavily on historical collision maps and blackspot analyses. These approaches treat locations in isolation, overlook how intersections are embedded within the wider network, and are inherently reactive: they flag danger only after serious crashes have already accumulated. This creates a gap for proactive, scalable safety screening that leverages road geometry and network structure to identify risky intersections before severe incidents occur.

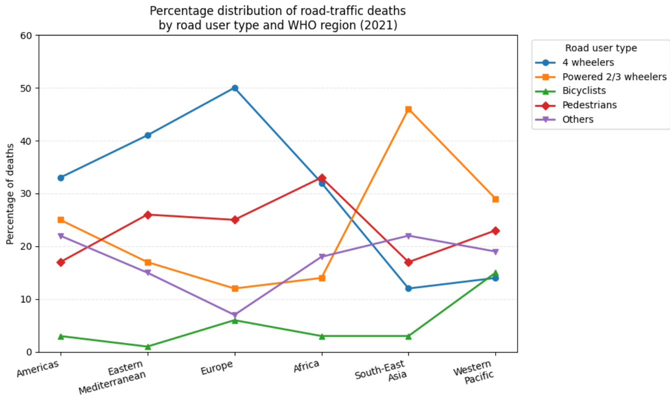

Figure 1: Percentage distribution of road-traffic deaths by road user type and WHO region, 2021.

For these limits for proactive safety interventions. There are some methods to help it out like Predict where accidents are likely to occur, Operate at fine spatial granularity, Generalise to locations with limited or no historical collision counts and Use data sources that are scalable and publicly accessible.Road traffic accidents in London remain a persistent threat to public safety, causing fatalities, serious injuries, economic loss, and congestion.

1.2 Project Goals

This project aims to build and evaluate a graph-based accident risk prediction framework for Greater London using GNNs and classical baselines. The goals are to: (i) represent the London road network as a graph from OSM;(OpenStreetMap contributors 2024) (ii) engineer meaningful node/edge features (connectivity, road type, speed, length, direction); and (iii) train models for two node-level tasks—accident occurrence (binary) and accident severity (multi-class risk bins). We compare GNN variants (TransformerConv, edge-feature ablations, multi-task learning) against baselines (XGBoost, count models, trivial predictors) using spatial splits/CV to test generalisation to unseen areas. The final output is an interpretable London crash-risk map to identify high-risk junctions and corridors, while noting limits from data quality and class imbalance and suggesting extensions with richer temporal/exposure data.

2 Background

2.1 Baseline Models

2.1.1 *Trivial baselines.*

For accident occurrence (binary classification), two simple predictors are included: (i) a *majority-class baseline* that assigns every node the most common training label, and (ii) a *random baseline* that predicts collisions according to the observed positive rate in the training set.

$$\hat{y}_i = \arg\max_c \sum_{j \in \mathcal{T}} \mathbb{I}(y_j = c). \tag{1}$$

[1] School of Computing and Mathematical Sciences, University of Greenwich, London SE10 9LS, UK, email: mt9498y@gre.ac.uk

These baselines provide a low-performance floor under class imbalance, showing task difficulty versus naive guessing. For severity, matching majority and random baselines are used from the training class distribution.(McBride Ellis 2024)

2.1.2 *XGBoost baseline.*

As a strong non-graph machine-learning comparator, gradient-boosted decision trees (XGBoost) are trained on the same node feature set.

$$\hat{f}(\mathbf{x}_i) = \sum_{m=1}^{M} \eta_m \, h_m(\mathbf{x}_i), \qquad \hat{y}_i = g\Big(\hat{f}(\mathbf{x}_i)\Big), \tag{2}$$

XGBoost handles nonlinear tabular patterns well. Binary and multi-class variants are trained for occurrence and severity, with SHAP used for feature interpretation and comparison to GNN structure effects.(Chen & Guestrin 2016)

Overall, these baselines span increasing levels of complexity and provide a rigorous reference against which the benefits of graph-based modelling and spatial generalisation can be assessed.

2.2 Node Features and Edge Structure

2.2.1 *Topological and Road Features*

Table 1: Summary of node-level features used in the implemented London road-graph models.

Feature Group	Description (as implemented)
Topological feature	**Degree**: node degree computed on the undirected version of the OSM drive graph, capturing local junction connectivity.
Aggregated incident-edge features	**Mean edge length** and **mean estimated speed**: for each node, incident road segments are aggregated by averaging OSM edge attributes `length` and `speed_kph` across adjacent edges.
Road category indicators	**Highway type one-hots**: the ten most frequent OSM `highway` categories in London (e.g., `primary`, `secondary`, `tertiary`, `residential`, `motorway_link`) are encoded as binary indicators per node, derived from incident edges.
Preprocessing	Continuous features are filled for missing values and standardised with `StandardScaler`; the resulting node feature vector is $\mathbf{x}_i \in \mathbb{R}^F$.

2.2.2 *Graph Structure and Edge Attributes*

Table 2: Graph representation and edge features used by the GNN models.

Component	Description (as implemented)
Nodes	Intersections and dead-ends from the OSM `drive` network. OSM node IDs (`osmid`) are mapped to a contiguous index `node_idx` for tensor construction.
Directed edges (u, v)	Edges follow OSM road geometry and one-way constraints in the directed multigraph. Each road segment between indexed nodes contributes a directed edge.
`edge_index`	Sparse COO edge list of shape `[2, E]` storing source and target node indices for all directed edges, as required by PyTorch Geometric.(Fey & Lenssen 2019)
Basic edge attributes	Edge feature matrix $\mathbf{e}_{uv} \in \mathbb{R}^3$ containing OSM-derived `length` (m), `speed_kph`, and `travel_time` (s).
Full edge attributes (ablation)	Extended edge feature matrix $\mathbf{e}_{uv} \in \mathbb{R}^7$ adding: (i) direction unit vector (`dir_x`, `dir_y`) from node coordinate differences, and (ii) bearing encoding $(\cos\theta, \sin\theta)$ from OSM edge bearings.
Use in models	TransformerConv consumes $\mathbf{X}$, `edge_index`, and either basic or full `edge_attr`, enabling attention to be modulated by road-segment properties.

2.3 Methodology

2.3.1 *Data Handling*

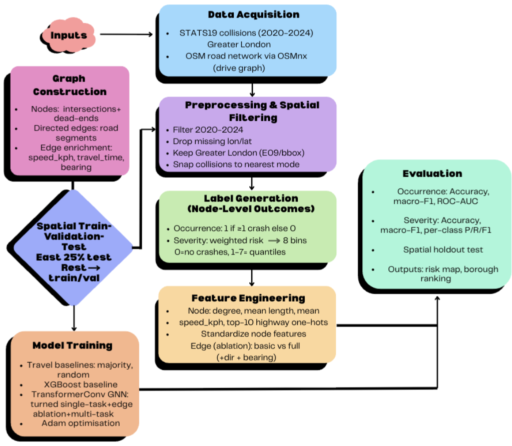

Figure 2: : End-to-End Pipeline for Node-Level Road-Crash Risk Prediction on the London Network.

2.3.2 *Feature Selection/Reduction*

OpenStreetMap tags are high-dimensional and sparse, so a compact feature set is constructed to improve stability and reduce noise. The `highway` attribute is reduced by keeping only the ten most frequent road types in London and encoding them as binary indicators, while rare categories are dropped. Edge-level signals are summarised at nodes via mean incident-edge length and mean estimated speed, yielding compact road-context features per intersection. Missing values are imputed with neutral defaults, and continuous features are standardised using z-scores, $x'_{ij} = (x_{ij} - \mu_j)/\sigma_j$, to align scales across inputs. Finally, SHAP importance from the XGBoost baseline is used as a post-hoc check that retained features contribute meaningfully(Lundberg & Lee 2017), without further pruning.

2.3.3 *Proposed model: Transformer-based Graph Convolutions (TransformerConv)*

TransformerConv applies multi-head self-attention for graph message passing. For each node v, neighbour features $u \in \mathcal{N}(v)$ are weighted by learned attention scores $\alpha_{vu}^{(h)}$ and aggregated across heads, allowing different relational patterns to be captured. Unlike standard GCN/GAT, attention can incorporate edge attributes $\mathbf{e}_{uv}$ (e.g., road length, speed, travel time, direction), so updates depend on both connectivity and road-segment properties. This enables risk information to propagate along connected streets(Shi et al. 2020), supporting network-aware intersection risk prediction **?**.

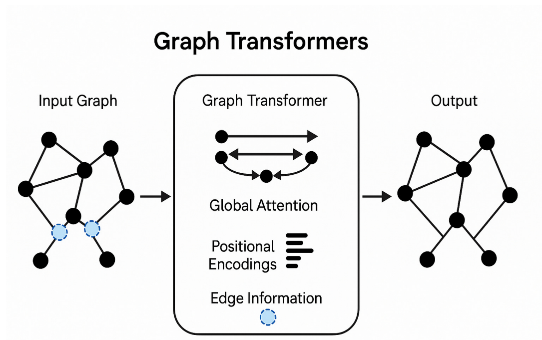

Figure 3: : Tranformer Model

2.3.4 *Transformer-based Graph Convolutions Fitting Process*

TransformerConv GNN is trained on the Greater London road graph with standardised node features and OSM-derived edge attributes. Collisions are mapped to nearest graph nodes to form occurrence labels (0/1) and 8-level severity labels from weighted risk scores. Spatial generalisation is enforced via a longitude-based split: eastern 25 percentage of nodes reserved for testing, remainder split into train/-validation. Optimisation uses Adam; focal loss (Lin et al. 2017) handles occurrence imbalance, while cross-entropy is applied for severity classification(Kingma & Ba 2015). Edge-feature ablations compare basic attributes against a full set including direction and bearing, alongside a shared multi-task encoder with two heads. Optuna tunes hidden size, attention heads, dropout, learning rate, and weight decay, after which the best model yields risk maps and final test metrics.

2.3.5 *Proposed system model*

TransformerConv extends multi-head self-attention to graphs, optionally using edge features, enabling richer, direction- and context-aware interactions beyond fixed-hop convolutions. To capture spatial dependence along connected roads, there is a use of a Transformer-based GNN with attention conditioned on edge attributes.

$$q_v^{(h)} = W_Q^{(h)} h_v^{(k)}, \quad k_u^{(h)} = W_K^{(h)} h_u^{(k)}, \quad v_u^{(h)} = W_V^{(h)} h_u^{(k)} \tag{3}$$

$$\alpha_{vu}^{(h)} = \mathrm{softmax}_{u \in \mathcal{N}(v)} \left(\frac{q_v^{(h)\top} k_u^{(h)}}{\sqrt{d_h}} \right) \tag{4}$$

$$h_v^{(k+1)} = \Big\Vert_{h=1}^{H} \sum_{u \in \mathcal{N}(v)} \alpha_{vu}^{(h)} v_u^{(h)} \tag{5}$$

Performs multi-head self-attention on graphs via Q,K,V to aggregate context- and edge-aware neighbor signals.(Vaswani et al. 2017)

3 Experiments and results

3.1 Experiments

3.1.1 *Data Collection*

Accident data come from UK police STATS19 records (Jan 2020–Dec 2024), including time, longitude/latitude, and severity (fatal/serious/slight), plus contextual fields like road layout, speed limit, lighting, weather, and surface conditions(Department for Transport 2025). Only the spatial and severity information is used to define risk labels; contextual variables are left for future extensions.

Records are filtered to the 2020–2024 period and entries with missing coordinates are removed. The dataset is restricted to Greater London using local authority district codes (ONS codes beginning with `E09`); where these codes are unavailable, a geographic bounding-box covering London is applied. Each collision is then snapped to its nearest road-network node (intersection or dead-end) using `osmnx.distance.nearest_nodes`(Boeing 2017), enabling aggregation of incidents to the intersection level.

3.1.2 *Hyperparameter Tuning*

TransformerConv hyperparameters are tuned with Optuna to maximise validation macro-F1 under the spatial split, using full edge features(Akiba et al. 2019). For each task, 35 trials are run; each trial trains for 40 epochs and is evaluated on the validation mask. The search space covers hidden size $h \in \{32, 64, 128, 256\}$, heads $k \in \{2, 4, 8\}$, dropout $p \in [0.2, 0.7]$, learning rate $\eta \in [5{\times}10^{-4}, 5{\times}10^{-3}]$ (log-uniform), and weight decay $\lambda \in [10^{-6}, 5{\times}10^{-4}]$ (log-uniform). Occurrence uses focal loss, severity uses cross-entropy.

3.2 Results

Table 3: Test performance on the eastern London spatial holdout.

Model	Occurrence	Severity
Majority baseline	0.853	0.853
Random baseline	0.711	0.717
XGBoost	0.840	0.839
TransformerConv (tuned)	**0.854**	**0.840**

3.2.1 *Results Analysis*

Figure 3 maps all recorded road traffic accidents across Greater London onto the road network, revealing clear spatial clustering rather than a uniform distribution. High densities of crashes appear along major arterials and around complex junctions, motivating network-aware modelling of intersection risk.

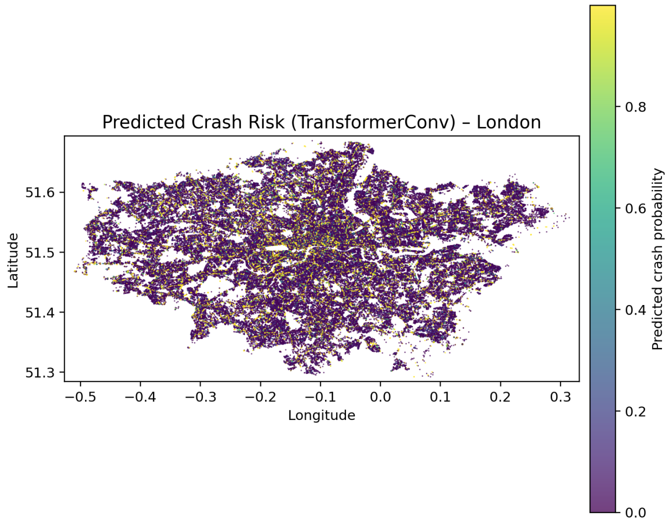

Figure 4: : Traffic accident locations of Greater London

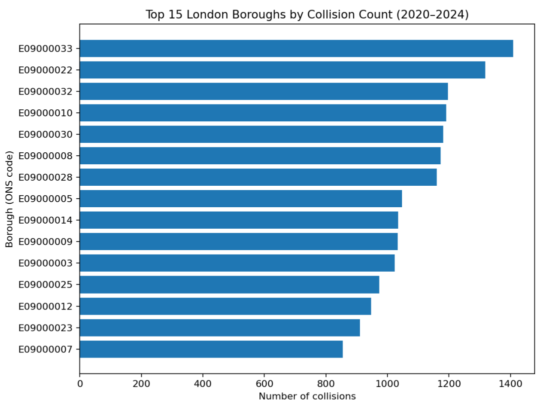

Figure 5: : Top 15 london Borough Collision count

Figure 5 reinforces the city's uneven risk landscape: collision counts are concentrated in a relatively small group of boroughs, mostly central and high-traffic areas, and this aligns with the dense hotspots visible in Figure 4. Model performance is summarised in Figures 6 and 7. For accident occurrence, the tuned TransformerConv yields the strongest results, clearly outperforming trivial baselines and performing on par with or slightly better than the XGBoost baseline. The improvement is most evident in macro precision/recall/F1, indicating better sensitivity to the minority crash class under spatial holdout testing. By contrast, severity prediction remains difficult: all methods achieve low macro scores because high-risk classes are rare and predictions are dominated by the zero-risk category. This suggests that network structure helps most for detecting whether crashes occur, while separating severity levels will likely require richer temporal or exposure features.

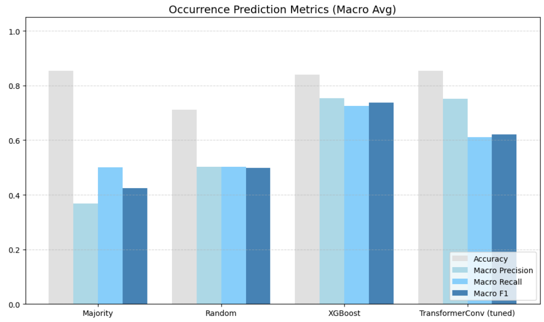

Figure 6: : Prediction Metrics

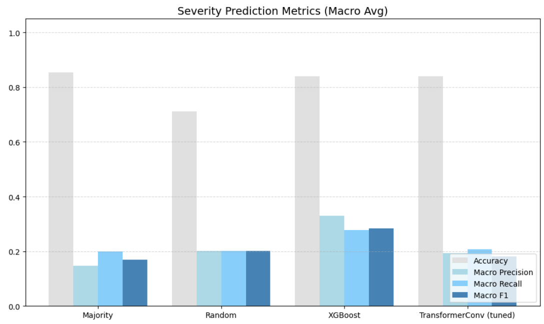

Figure 7: : Prediction Metrics

4 Discussion

4.1 Strengths

This work demonstrates several practical and methodological strengths. First, the framework relies entirely on open-access sources (STATS19 collision records and OpenStreetMap road geometry), making the pipeline transparent, reproducible, and transferable to other cities with minimal modification. Second, representing London's streets as a graph provides a natural way to encode spatial dependence: intersection connectivity, adjacency effects between nearby high-risk junctions, and functional road hierarchy are all captured directly through the network structure. Third, across both prediction tasks, TransformerConv-based GNNs provide consistent gains over non-graph baselines, indicating that message passing on the road network adds meaningful signal beyond tabular road attributes alone. Finally, the fitted models yield interpretable outputs (risk maps, attention weights, SHAP comparisons) that align with real policy needs by highlighting high-risk corridors and junctions rather than only producing abstract accuracy scores.

4.2 Limitations

Despite these strengths, several limitations should be noted. The approach depends on the completeness and positional accuracy of police-reported collisions; under-reporting, location noise, or inconsistencies in severity coding may propagate into node labels. Model inputs are largely static and structural, so temporal drivers of risk (time-of-day cycles, weekday/weekend effects, weather, seasonal variation, and long-term trend shifts) are not captured in the current setup. Geographic generalisation is evaluated using a single longitude-based holdout (east London versus the rest), which is a reasonable but coarse proxy for spatial transfer; multiple alternative splits or more rigorous spatial blocking could further stress-test robustness. Finally, discretising severity into eight risk bins simplifies evaluation but may blur subtle differences between moderately risky sites, and strong class imbalance (especially for higher-risk bins) can limit calibration quality and sensitivity to rare but critical locations.

5 Conclusion and Future Work

5.1 Conclusion

This project introduced a graph-based framework for predicting crash risk across Greater London by integrating STATS19 collision data with an OpenStreetMap-derived road network. A directed intersection graph was constructed, collisions were aggregated to nodes, and two node-level tasks were modelled: accident occurrence and discretised severity risk. TransformerConv GNNs, which combine message passing with attention conditioned on road-segment attributes, were evaluated against trivial predictors, count-regression models, and XGBoost baselines under a spatially stratified split. Results show that incorporating road-network structure improves predictive performance relative to non-graph approaches, especially for identifying high-risk junctions. The final outputs include an interpretable, intersection-level risk map and borough rankings, supporting targeted road-safety analysis using fully open and reproducible data sources.

5.2 Future Work

Several extensions could strengthen both realism and generalisation. First, temporal and contextual dynamics (hour-of-day patterns, weekday effects, weather, seasonal drift) could be incorporated through time-aware features or spatio-temporal GNNs rather than relying on static structure alone. Second, exposure information such as traffic volumes, pedestrian flows, or mobility proxies would allow modelling of risk per unit exposure instead of raw collision counts, improving fairness across regions with different usage levels. Third, spatial robustness could be enhanced using multiple blocked splits, larger-scale spatial cross-validation, or inductive transfer tests on other UK cities. Fourth, severity modelling could move beyond discretised bins toward ordinal or continuous risk regression, paired with calibration methods and uncertainty estimation to better support decision-making under rare-event imbalance. Finally, richer feature sets (lane counts, signal presence, land use, speed enforcement locations) and multimodal fusion with satellite or street imagery may further improve localisation of hazardous corridors and enable more actionable safety insights.

ACKNOWLEDGEMENTS

First of all I want to thank you to my tutor, he is the best mentor that I got it in my life. I express my gratitude to my family, friends and coworkers for their valuable contributions to our effort. In addition, I express deep appreciation to the writers whose references I used for my study, as their expertise has greatly assisted me in doing this research.

REFERENCES

Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. (2019), Optuna: A next-generation hyperparameter optimization framework, *in* 'Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Workshops'. arXiv:1907.10902.

Boeing, G. (2017), 'Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks', *Computers, Environment and Urban Systems* **65**, 126–139.

Chen, T. & Guestrin, C. (2016), Xgboost: A scalable tree boosting system, *in* 'Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', pp. 785–794.

Department for Transport (2025), 'Road safety data', `https://www.gov.uk/government/statistics/road-safety-data`. Published 25 September 2025. Accessed 21 November 2025.

Fey, M. & Lenssen, J. E. (2019), 'Fast graph representation learning with pytorch geometric', arXiv:1903.02428.

Kingma, D. P. & Ba, J. (2015), 'Adam: A method for stochastic optimization', International Conference on Learning Representations (ICLR). arXiv:1412.6980.

Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. (2017), Focal loss for dense object detection, *in* 'Proceedings of the IEEE International Conference on Computer Vision (ICCV)', pp. 2980–2988.

Lundberg, S. M. & Lee, S.-I. (2017), A unified approach to interpreting model predictions, *in* 'Advances in Neural Information Processing Systems (NeurIPS 2017)', Vol. 30, pp. 4765–4774.

McBride Ellis, C. (2024), 'The orange book of machine learning', Online lecture notes. Section 7.4: Classification baseline (majority/ZeroR predictor).

OpenStreetMap contributors (2024), 'Openstreetmap', `https://www.openstreetmap.org`. Accessed 21 November 2025.

Shi, Y., Huang, Z., Feng, S., Wang, H. & Sun, Y. (2020), 'Masked label prediction: Unified message passing model for semi-supervised classification', arXiv:2009.03509. Introduces the TransformerConv operator used in PyTorch Geometric.

United Nations General Assembly (2020), 'Improving global road safety', Resolution A/RES/74/299. Proclaims the Second Decade of Action for Road Safety 2021–2030.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017), Attention is all you need, *in* 'Advances in Neural Information Processing Systems (NeurIPS)', Vol. 30, pp. 5998–6008. arXiv:1706.03762.

World Health Organization (2023), 'Global status report on road safety 2023'.