



Full length article

Heterogeneous graph neural networks with post-hoc explanations for multi-modal and explainable land use inference

Xuehao Zhai ^{a,*,}, Junqi Jiang ^{b,}, Adam Dejl ^{b,}, Antonio Rago ^{b,c,}, Fangce Guo ^{a,},
Francesca Toni ^{b,}, Aruna Sivakumar ^{a,}

^a Imperial College London, Department of Civil and Environmental Engineering, Exhibition Rd, South Kensington, London, SW7 2AZ, UK

^b Imperial College London, Department of Computing, South Kensington Campus, London, SW7 2AZ, UK

^c King's College London, Department of Informatics, Strand Campus, 30, Aldwych, London, WC2B 4BG, UK

ARTICLE INFO

Keywords:

Urban information fusion

Heterogeneous graph neural networks

Explainable AI

ABSTRACT

Recently, the increased use of sensor and location technologies has facilitated the collection of multi-modal mobility data, offering valuable insights into daily activity patterns. Many studies have adopted advanced data-driven techniques to explore the potential of these multi-modal mobility data in land use inference. However, existing studies often process samples independently, ignoring the spatial correlations among neighbouring objects and heterogeneity among different services. Furthermore, the inherently low interpretability of complex deep learning methods poses a significant barrier in urban planning, where transparency and extrapolability are crucial for making long-term policy decisions. To overcome these challenges, we introduce an explainable framework for inferring land use that synergises heterogeneous graph neural networks (HGNs) with Explainable AI techniques, enhancing both accuracy and explainability. We evaluate the proposed approach on three cities with different urban layout and mobility combinations: London (with tube, bus and bike sharing data sources), San Francisco (parking and bike sharing), and New York City (metro and bike sharing). The empirical experiments demonstrate that the proposed HGNs significantly outperform baseline graph neural networks for all six land use indicators, especially in terms of 'office' and 'sustenance'. We then deploy feature attribution (at both temporal and spacial levels) and counterfactual explanations, which shed light on several important findings. The node-based feature attribution explanations show that the symmetrical nature of the 'residence' and 'work' categories predicted by the framework aligns well with the commuters' 'work' and 'recreation' activities. Meanwhile, the spatial feature attribution explanations indicates that the heightened central importance of commercial categories and the dominance of residential influences in outer zones align closely with typical urban structures. Finally, the counterfactual explanations reveal that variations in node features and types are primarily responsible for the differences observed between the predicted land use distribution and the ideal mixed state. These analyses demonstrate that the proposed HGNs can suitably support urban stakeholders in their urban planning and policy-making. The source code is available at <https://github.com/xuehao0806/GNN-land-use>.

1. Introduction

Land use is a term used by urban planners to describe the function of human-made spaces in which people live, work, and recreate on a day-to-day basis. Typically, city planners are responsible for determining land use plans, shaping a city's layout by dictating what types of buildings can be located in a given space. However, due to the complex and dynamic nature of the urban environment, the actual function – both activity pattern and intensity – may deviate from the intentions of land use planning [1–3]. For instance, an area initially planned as a commercial zone might gradually transition into a mixed-use area

because of population growth and changes in housing demands. As the cities dynamically evolve, this deviation between land use planning and actual usage may widen, leading to inefficient use of resources, congestion, community dissatisfaction and social inequality [4,5]. To support sustainable urban planning and related policies, stakeholders need accurate and effective tools to detect the spatial distribution of human activities.

Urban land use inference has been a challenging task for decades. Early studies primarily used simple statistical methods (e.g., Linear

* Corresponding author.

E-mail address: x.zhai20@imperial.ac.uk (X. Zhai).

<https://doi.org/10.1016/j.inffus.2025.103057>

Received 25 October 2024; Received in revised form 21 February 2025; Accepted 25 February 2025

Available online 8 March 2025

1566-2535/Crown Copyright © 2025 Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

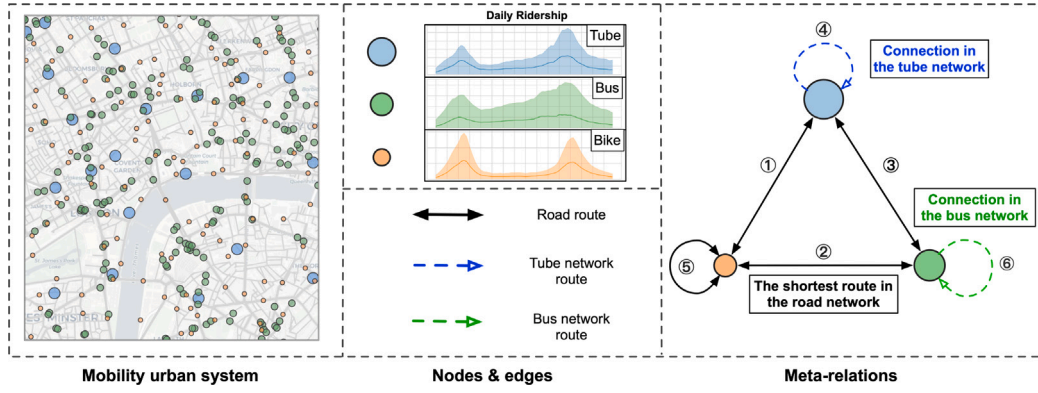


Fig. 1. This figure depicts a heterogeneous mobility network featuring meta-relations that highlight various connection types. The left panel displays a central London area with nodes coloured by different types of mobility services, representing tube stations (blue), bus stops (green), and bike stations (orange). The middle panel presents a legend for road routes (solid black lines), tube routes (dashed blue lines), and bus routes (dashed green lines), accompanied by daily ridership graphs for each mode of transport. The right panel outlines meta-relations with nodes and their connections, illustrating the complexity of information in a multi-modal mobility network.

Regressions) [6,7] and simulation-based methods (e.g., Cellular Automata) [8,9] with survey data to explore urban growth patterns and dynamics with rule-based interactions. However, these traditional methods rely on over-simplified assumptions, struggling to accurately reflect complex real-world scenarios, especially for large metropolitan areas. In recent years, the advent of the era of big data has provided a new paradigm for this problem. Emergent sensor and location technologies have facilitated the collection of human activity-based data, such as mobile phone and mobility records. These passive data sources offer alternative ways for monitoring people's daily activities, facilitating the description of the urban space function. Furthermore, advanced AI models have shown significant promise in handling multi-modal and non-linear data in urban systems [10,11]. These opportunities provide a strong foundation for quantitatively detecting urban land use. In the past decade, 23% (38 out of 165) of ML-based research papers in urban spatial studies focused on inferring land use and urban form [12], which shows that this problem has gradually become one of the main branches of AI applications in the field of urban studies.

Despite these recent advances, inferring land use in an accurate, explainable and thus trustworthy manner remains difficult in practice. The challenges in existing research can be summarised from three perspectives: data selection, methodology and post-hoc analysis.

A primary challenge lies in determining what type of activity-based data is efficient and how to fuse multi-modal data reasonably. Many activity-based datasets [1,13] have been used in urban land use inference. However, inherent bias exists in many types of data when used for the representation of human activity. For instance, the number of mobile calls or messages occurring in an area could be affected by factors such as user occupation types and user communication preferences, potentially skewing observed usage intensity. Similar problems exist with social media data, which is strongly related to the user's age. Moreover, data involves user privacy, which is highly restricted for commercial use in many countries, making these types of data unavailable to many practical projects. Mobility data, on the other hand, offers low-cost, real-time, and objective insights into people's daily movements. Meanwhile, travel demand has been systematically proven to be closely related to land use patterns [14,15]. For example, a higher volume of bus passengers recorded during weekday peak hours suggests that the land use surrounding this station is work-related or residential. While mobility data can provide insights into daily human movements and associated land use patterns, a few questions remain about (a) which modes of transportation data are most indicative of land use; (b) how to aggregate this data across time and space effectively; and (c) how to integrate data from various transportation modes to form a comprehensive analysis.

Another challenge lies in determining how to address the spatial dependency inherent in urban activities and the heterogeneity among

different mobility services. Urban activities tend to cluster, exhibiting continuity along road networks and potential disruptions by physical barriers like rivers, railways and highways [16]. Moreover, different types of mobility data contribute to different degrees of understanding surrounding land uses. Large transit hubs primarily facilitate transfers and are closely tied to adjacent transport services. In contrast, smaller transit facilities are crucial for serving 'first or last mile' travel and are strongly linked to local residential and work activities [3]. Most deep learning models process samples independently, overlooking the spatial dependency between samples and their surrounding objects. Although recent studies [17,18] try to solve this problem by employing homogeneous graph networks, these methods often fail to encapsulate the multi-level nodes (e.g., bike stops, bus stations) and edges (e.g., road networks, tube networks) in complex urban areas.

A third challenge lies in the fact that establishing explainability is also essential in policy and land use studies in order to build trust among government planners and policymakers. Effective land use models must be accurate, reliable, accountable for anomalies, and capable of extrapolation [19]. However, the black-box nature of deep learning models often hampers their trustworthiness in real-world applications. Therefore, a (black-box) model's credibility with planners needs to be established by systematically proving these properties through post-hoc analysis, thus reinforcing their utility in practical urban planning contexts.

Motivated by these challenges, this paper makes contributions in three key aspects: integrating mobility data, fusing multiple sources of mobility data from different travel modes using heterogeneous graph neural networks, and enhancing model explainability by using post-hoc explainable AI methods. Specifically, we make the following contributions:

- We have developed a novel data pre-processing pipeline to convert multi-modal mobility data sources into a heterogeneous graph structure, including bus, tube (i.e. London underground) and bike sharing, as well as topology network data and road types, as shown in Fig. 1. These rich information sources enable deep learning models to learn spatial representations that capture the heterogeneity of objects within the urban transport system. Readers can refer to the workflow shown in Fig. A.13, along with the accompanying open-source code, to construct a similar graph dataset structure for their regions of interest.
- We pioneer the application of Heterogeneous Graph Networks (HGNs) for data fusion and land use inference tasks, addressing the spatial dependency of land use features and the heterogeneity among different types of nodes and edges (Fig. 2). Our findings indicate that the proposed method markedly enhances predictive performance over traditional neural networks and homogeneous graph neural networks.

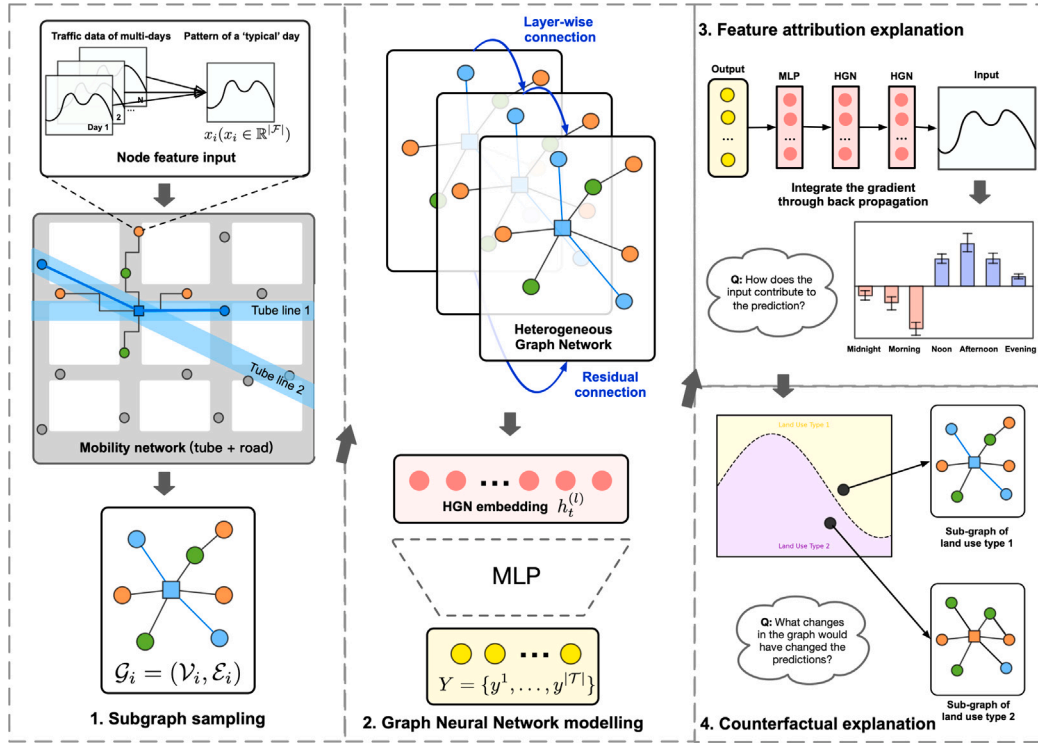


Fig. 2. Overview of our framework for land use inference based on multi-mobility systems.

- We pioneer the usage of feature attribution explanations in land use inference to empower explainability, addressing stakeholders' concerns as to how mobility distribution in a day could affect land use predictions. By highlighting the importance of different segments of the temporal patterns of human activity, we demonstrate that our model's behaviour aligns with a common understanding of commuters' daily movement patterns. We then extend the explainability analysis to spatial (subgraph-level) feature attribution, enabling policymakers to capture broader spatial patterns beyond individual nodes and identify localised anomalies that inform targeted urban interventions.
- We also define a new counterfactual explanation notion specifically suitable for explaining HGNs in land use inference tasks. We demonstrate through example analysis that our counterfactual explanations can be used to reveal changes required for the input to achieve the ideal mixed land use prediction. Through experiments, we also use such explanations to investigate the contributions of different elements in the input data to the differences between the model's predictions and the ideal mixed land use states.

This paper presents a pipeline for processing mobility data into heterogeneous graph data and utilises HGNs and explainability techniques to achieve explainable land use inference. Our research used mobility datasets from three cities: London, San Francisco, and New York City. While we focus our analysis of explanations on London, the proposed pipeline has significant potential for application to mobility data from other cities.

The remainder of this paper is organised as follows: Section 2 presents a literature review of relevant research on land use analysis, graph deep learning, and explainable AI; Section 3 describes the data collection and pre-processing steps; Section 4 outlines the methodology for Heterogeneous Graph Modelling framework and post-hoc explanations methods; Section 5 reports the results and explainability analysis; Section 6 discusses the implications of our findings, limitations, and recommendations for future research.

2. Literature review

In this section, we first overview related work in land use inference and identify key research gaps—namely, the lack of methods for urban multi-modal data fusion, insufficient consideration of spatial heterogeneity, and the absence of explainability analysis. Recognising these gaps, we examine methodologies in multi-modal data fusion – which enable the integration of diverse mobility and infrastructure data – and explainable artificial intelligence (XAI) from other domains, which enhance the explainability of complex models. Building upon insights from this comprehensive review, we lay the groundwork for a novel pipeline that fully utilises these multimodal data in an accurate and explainable manner.

2.1. Background and gaps for land use inference

Land use inference (also known as region function identification) is the process of determining the actual function and intensity of use of urban spaces [4,20]. This paper focuses on the problem of land use inference in urban areas based on people's activity data. It is an important research topic because it helps identify gaps between planned land use and how people actually use city spaces, which can lead to challenges like traffic congestion and inadequate infrastructure. By inferring land use based on current human activity, governments can make better policy decisions, optimise facilities, and help businesses distribute more effectively, thereby improving the urban economy [1,3].

In the literature on land use, early research used simple statistical approaches (e.g., linear regressions) or simulation-based models (e.g., cellular automata) to understand urban growth patterns. Due to the development of sensor and location technologies, land use modelling has evolved from relying on survey data to incorporating activity-based big data (e.g., social media data, mobile data and transport data), enhancing understanding of urban dynamics in a fine-grained manner. In line with this change, there has been a significant shift from traditional model-driven to advanced data-driven methods [1,17,21]. For example, Toole et al. [1] applied call frequency data and correlation

Table 1
Summary of papers on data-driven land use inference.

Paper	Scenario	Data	Method	Land use types
[1]	Boston, USA	Mobile phone data	Random forests	Residential, commercial, industrial, parks
[23]	Harbin, China	Call frequencies	Correlation analysis	Work, residential
[22]	Wuhan, China	Mobile phone calls	Deep Convolutional Autoencoder	Residential, commercial, business, industrial, mixed
[2]	Beijing, China	GPS Trajectory	Topic-based model	Education, residential, entertainment, mixed, tourist interest
[24]	Shenzhen, China	Ridership of metro, bus, and cab	Spectral clustering framework	Residential, Commercial, Transport
[3]	San Francisco, USA	Shared cycling ridership	Neural network	Residential, work, consumption, transport
[25]	Hangzhou, China	Public bike ridership, Taxi GPS	Latent semantic classification	Residential, business, industrial, administration
[26]	Beijing, China	GIS information	Graph Convolutional Neural Network	Business, Transport, Residence, Recreation
[17]	Beijing, China	Vehicle GPS and vehicle trajectory	Graph Convolutional Neural Network	Commercial, public, traffic
[27]	Beijing, China	Taxi trajectory and bus ridership	Latent clustering; Noise reduction	Residential, commercial, mixed
[28]	Chengdu, China	Online Car-Hailing; POI	Ensemble empirical modal decomposition	Business, office, residential, tourist interest
[18]	Seoul, Korea	Taxi ridership	Statistical methods	Residential, industrial, commercial, green
[29]	Shanghai, China	Metro ridership	Spatiotemporal similarity	Residential, employment, transport, mixed
[30]	Calgary, Canada	GTFS and APC data	Statistical methods	Residential, commercial, recreational, industrial, institutional
[31]	Shenzhen, China	Surround buildings and activities data	Transformer	Residential, commercial, recreational, industrial, institutional

analysis to infer land use types in Boston, USA. The latest trend in urban land use inference incorporates big data and machine learning techniques to provide highly accurate insights and dynamic forecasts of land use. Hu et al. [17] employed vehicle GPS and trajectory data, using Graph Convolutional Neural Networks, to categorise road network patterns in Beijing into commercial, public, and traffic types. Similarly, Sun et al. [22] applied mobile phone data and deep convolutional autoencoders to deduce various land use types – residential, commercial, business, and industrial – in Wuhan, China. A specific summary of activity data-based research in this domain is given in Table 1 below, where the first three studies use mobile data, and the subsequent ones use mobility data.

A recent review paper found that 23% (38 out of 165) of AI-based research papers in urban spatial studies focused on inferring land use and urban form [12], which showed that inferring land use has become one of the main branches of AI application in the field of city planning and the urban geoscience domain [32].

While the field of land use inference has made progress, it still faces significant research gaps that hinder its development. First, despite the availability of diverse mobility datasets, there is a notable lack of methods for effectively fusing multi-modal data in land use inference. As summarised in Table 1, existing studies often rely on a single type of data, such as trajectory data or ridership data, each with inherent limitations. Trajectory data is a good source for capturing road-level usage intensity but may not accurately represent the land use functions of surrounding areas [2,33]. Conversely, ridership data, which measures the flow of passengers into and out of stations, provides a more intuitive reflection of travel related to human activities. However, relying on a single type of transit facility does not provide comprehensive coverage of the entire urban area. Data fusion can be an effective way to address the lack of spatial coverage and potential biases. Yet, only a few studies [27,33] have performed data fusion (e.g., ensemble methods at the output level) to merge different datasets, which limits the exploration of complementary relationships within the mobility system.

Another critical research gap is the insufficient consideration of spatial heterogeneity inherent in urban environments. While recent research has begun utilising Graph Neural Networks (GNNs) [17,22,24,26] to model spatial relationships, these studies predominantly employ homogeneous graphs that do not distinguish between different types of nodes and edges. This simplification overlooks the diverse interactions of various mobility services and infrastructural elements within cities. Different nodes – such as bike stations, bus stops, and transit hubs – and edges like road networks and subway lines have unique characteristics and contribute differently to land use patterns. For instance, large transit hubs are primarily associated with transfers and connectivity to other transport services, while smaller facilities are integral for first- and last-mile connectivity, closely tied to local residential and work activities [3]. The reliance on homogeneous graphs fails to capture these nuances, underscoring the need for methodologies that incorporate heterogeneous graphs to more accurately represent the multi-level interactions in complex urban areas.

The final research gap lies in the absence of explainability analysis in land use inference models. The successful application of AI in critical domains like urban planning and policy-making hinges on models being explainable and trustworthy. The black-box nature of complex data-driven models often hampers their adoption, as stakeholders demand transparency and accountability [34]. Besides, stakeholders anticipate that the model can go beyond the prediction of the outcome; they also expect it to be able to respond to hypothetical or “what-if” scenarios, like “what efforts can be made to achieve sustainable land use planning”. However, few land use inference studies have provided explainable modelling or post-hoc explanations for their proposed models.

2.2. Opportunities from related domains

To address the above-mentioned research gaps, we conduct a literature investigation on multi-modal data fusion methodologies, graph modelling techniques, and XAI approaches from related domains to

identify feasible solutions. These are discussed in detail in the following Sections 2.2.1 and 2.2.2.

2.2.1. Multi-modal data fusion and graph modelling

The need for fusion of multi-modal data for better accuracy and usability is not unique to land use inference but also exists in many efforts for the development of smart and sustainable cities, such as traffic prediction [35,36], energy management [37,38], and healthcare [39]. In urban studies, commonly used fusion strategies can be divided into the following categories: fuzzy logic-based [11,40], Bayesian-based [41,42], ensemble-based [43,44], and graph-based [11,45,46]. For the mobility data discussed in this paper, graph-based methods like GraphSages [47] and Graph Attention Networks [48] are particularly effective due to the strong ability to process spatial dependency among the facilities in a large-scale graph form. For example, Hu et al. [17] introduced a Graph Convolutional Network (GCN) to understand the social function of road segments, showing significant accuracy compared with traditional machine learning.

Building upon the effectiveness of graph-based methods, advanced techniques such as Heterogeneous Graph Networks (HGNs) have emerged to better capture the complexity of mobility systems. As an extension of Graph Neural Networks (GNNs), HGNs are designed to handle graph structures with multiple types of nodes and edges, allowing for a more nuanced representation of spatial heterogeneity and hierarchical relationships. Studies have shown that HGNs excel in fusing multi-modal data and modelling complex graph structures in various domains, including recommendation systems [49], biomedical research [50], and social networks [51]. In urban environments, HGNs have been successfully applied to challenges like traffic prediction and signal control, achieving notable performance improvements [36,52]. Despite these successes, HGNs have not yet been explored for land use inference, presenting a promising opportunity to address existing gaps in the field.

2.2.2. Explainable AI

In the field of Explainable AI (XAI), significant progress has been made in explaining complex machine learning models. This is especially important in urban planning and land use inference, where stakeholders require clear insights into model predictions. Since our research involves Graph Neural Networks (GNNs) to model spatial dependencies, it is essential to explore XAI methods specifically designed for graph-based models. Recent advancements in graph-based explanations provide promising tools for understanding the complex relationships within heterogeneous mobility graphs. Specifically, we focus on three types of analytic XAI methods: feature ablation analysis, feature attribution methods, and counterfactual explanations.

- **Feature ablation analysis** assesses the significance of individual input features or components by systematically removing or altering them and observing the impact on the model's predictions. This method is a straightforward explanation method that has been commonly utilised to understand the function of elements in models [53–55]. In graph neural networks, feature ablation has been applied to interpret the roles of nodes and edges, aiding in the comprehension of models in domains such as molecular chemistry [56], medical diagnosis [57] and social network analysis [58,59].
- **Feature attribution methods** quantify the approximate importance of individual input features on a particular model prediction through an attribution value. These techniques, such as Integrated Gradients [60] and SHapley Additive exPlanations (SHAP) [61], clarify the roles of input features in the prediction of the outcome and have been successfully used in understanding graph elements in many critical domains, such as molecular science [62,63] and Internet of Things [64].

- **Counterfactual explanations** demonstrate which changes are needed in the factual input to achieve a different output prediction [65]. These methods have been used in domains like molecular science [66] and recommendation systems [67] to understand how potential changes in the graph might result in desired outcomes.

While the above-mentioned methods are model-agnostic and theoretically can quantify the contributions of different node features in graph models, practical limitations in current implementations (e.g., Captum) hinder their direct application to HGNs. Moreover, these tools typically do not assess the influence of neighbours and edges or perform localised reasoning for state-of-the-art models like the HGNs. This underscores the need for developing specialised explainability techniques that can handle the complexity of HGNs, thereby enhancing the interpretability of land use inference models and facilitating their adoption in urban planning.

2.3. Summary and proposed framework

In summary, to address the identified research gaps and leverage opportunities from related domains, we propose a novel modelling framework for land use inference. This framework integrates multi-modal data fusion through the construction of heterogeneous graphs that incorporate various urban network structures and mobility data. By employing a Heterogeneous Graph Transformer for land use prediction, we effectively capture spatial heterogeneity and complex interactions within urban environments. Furthermore, we incorporate three post-hoc explainability methods – ablation analysis (AA), feature attribution (FA), and counterfactual explanations (CE) – to enhance the interpretability and trustworthiness of our model.

Given that our approach utilises a heterogeneous graph model, we set homogeneous graph models such as Graph Convolutional Network (GCN), GraphSage, and Graph Attention Network (GAT) as baselines for comparison. Although we validate our approach using London as a case study due to its diverse transportation modes and varied land use patterns, our framework is designed to be generalisable and can be applied to any city with available mobility and infrastructure data. By leveraging adaptable modelling techniques and widely accessible data sources, our method provides a versatile tool for urban planners and policymakers seeking to improve land use inference in various urban contexts.

3. Preliminaries

This section introduces the research problem and the basic concepts for building heterogeneous graph models. The notation used throughout this paper is shown in Table 2.

We begin by defining heterogeneous graphs.

Definition 1 (Graph). A graph $\mathcal{G}$ is represented as $(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a set of nodes (or vertices), i.e., $\mathcal{V} = \{v_1, \dots, v_{|\mathcal{V}|}\}$, and $\mathcal{E}$ denotes a set of edges connecting nodes, i.e., $\mathcal{E} = \{e_1, \dots, e_{|\mathcal{E}|}\}$. Each edge $e \in \mathcal{E}$ is a pair of nodes (v_i, v_j) iff v_i and v_j are connected.

In a heterogeneous graph nodes and/or edges belong to different types, as follows:

Definition 2 (Heterogeneous Graph). Let $\mathcal{A}$ denote a set of node types and $\mathcal{R}$ denote a set of edge types, such that $|\mathcal{A}| + |\mathcal{R}| > 2$. Then a heterogeneous graph $\mathcal{G}$ is $(\mathcal{V}, \mathcal{E})$ such that there exists a function $\phi : \mathcal{V} \rightarrow \mathcal{A}$ that maps each node with its corresponding node type and a function $\psi : \mathcal{E} \rightarrow \mathcal{R}$ that maps each edge with its corresponding edge type.

Definition 3 (Meta Relation). The meta relation for an edge $e = (v_s, v_t)$ linked from source node v_s to target node v_t is denoted as $\langle \phi(s), \psi(e), \phi(t) \rangle$.

Table 2

Notation and Descriptions. The table is divided into three sections: the first section includes general operations; the second section includes components of the graph; and the third section includes elements related to the learning problem.

Notation	Description
$\oplus$	A concatenation operation
$ \cdot $	The cardinality of a set
$\mathcal{G}$	A graph
$\mathcal{V}$	A set of nodes
$\mathcal{E}$	A set of edges
$\mathcal{A}$	A set of node types
$\mathcal{R}$	A set of edge types
$v_i \in \mathcal{V}$	A single node with subscript i (also referred to simply as i)
$e_{ij} \in \mathcal{E}$	An edge representing a pair of nodes (v_i, v_j) iff v_i and v_j are connected
ϕ	A node-to-type mapping
ψ	An edge-to-type mapping
$\mathcal{F}$	A set of features of each node
$\mathcal{T}$	A set of land use targets
$x_{i,j}$	The j th feature for node i
$h_i^{(0)}$	The initial node representation after the feature transformation
$h_s^{(l)}$	The updated representation of target node s at layer l
Θ	The set of parameters for training
$W_{\phi(i)}^P$	The initial projection matrix for node type $\phi(i)$
$W_{\tau, \phi(s)}^{MH}$	The multi-head matrix for source node type in head τ
$W_{\psi(e)}^{MP}$	The message passing matrix for edge type $\psi(e)$
$W_{\tau}^Q, W_{\tau}^K, W_{\tau}^A$	The query, key and attention matrices for head τ
$W_{\phi(i)}^{AG}$	The aggregation matrix for target node type $\phi(i)$
$m^{(l)}(s, e, i)$	The message representation from v_s to v_i through edge e at layer l
$\alpha^{(l)}(s, e, i)$	The attention weights from v_s to v_i through edge e at layer l
ξ	The scaling factor in the attention mechanism
θ	The number of heads in the multi-head mechanism

Urban mobility systems can be conceptualised as heterogeneous graph structures with multiple types of service facilities in the form of nodes (e.g., tube and bus stations) and spatial connections among them in the form of edges (e.g., transit routes). Specifically, we consider two categories of edges: edges in the road network and edges in the transit networks. When travellers transfer between different types of stations, they usually need to walk. Therefore, as shown in Fig. 1, we regard the connections across various types of nodes as the shortest path on the road network (①, ②, ③). For edges between nodes of the same type, there may be shortest routes in a road network (④) or connecting paths in a specific network, such as a tube network or a bus network (⑤ and ⑥).

Given the complexities in the data, this paper aims to build an explainable heterogeneous graph-based framework for the land use inference problem. To begin with, the research is framed as a multi-task regression problem. This setting stems from practical considerations: compared with merely categorising land use types, it is more useful for stakeholders to obtain the intensities of different land use because of the need to analyse diversity [3].

Mathematically, we build a multivariate mapping function $\mathcal{M}$ that regresses land use intensity based on daily mobility distribution and urban heterogeneous graph structure, shown as follows:

$$Y = \mathcal{M}_{\Theta}(\mathcal{G}; X) \quad (1)$$

where the parameters set for training are Θ , the nodes feature $X \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{F}|}$ ($\mathcal{F}$ is the set of features of each node¹) represents the mobility inbound or outbound variables at different time intervals, such as ‘15-min passengers inflow in a typical day’² for all nodes in $\mathcal{G}$. $Y \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{T}|}$ ($\mathcal{T}$ is the set of targets³) denotes the multi-regression targets. The goal

of land use inference is to find an optimised set of Θ^* to minimise the loss between the estimated land use indicators and land use labels:

$$\Theta^* = \arg \min_{\Theta} \frac{1}{n} \sum_{i=1}^n \text{loss}(\mathcal{F}_{\Theta}(\mathcal{G}; X), \hat{Y}) \quad (2)$$

where $\hat{Y} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{T}|}$ refers to the land use ground-truth labels and n refers to the number of samples in the training dataset.

4. Methodology

Fig. 2 illustrates the framework of this methodology, divided into three main parts: (a) Graph Building and Sampling: synthesising traffic data into typical daily patterns, collecting edge information from a hierarchical urban mobility network, and sampling heterogeneous subgraphs; (b) Graph Neural Network Modelling: processing these subgraphs through a HGN with layer-wise and residual connections, embedding the output into a feature vector, and using a fully connected linear layer to predict land use targets; (c) Explanation Techniques: using feature attribution explanations to determine the contribution of input features to the predictions and counterfactual explanations to explore the impact of changes in the graph structure on prediction outcomes.

4.1. Heterogeneous attention embedding

Given the rich mobility information and road topology information, our primary focus is on effectively embedding structural data and node-level mobility features within urban mobility networks. Inspired by the recent progress of HGNs in social network analysis [68], we propose a Heterogeneous Graph Encoder based on HGNs. The HGN module aims to learn node representations through predetermined meta-relations. To model the influence of various node types, we utilise a multi-head attention mechanism.

Firstly, we define a node feature transformation function that aligns features x_i of nodes v_i to a unified feature space, based on their node types $\phi(i)$. This is formulated as:

$$h_i^{(0)} = W_{\phi(i)}^P \cdot x_i, \quad (3)$$

¹ Concretely, we denote nodes features as the daily passenger inbound data of each station (15-minute interval ridership from 6:00 to 22:00, forming 64 input features).

² A ‘typical day’ refers to the mean values of data collected over multiple statistical periods (weeks, months) under normal, non-exceptional conditions [3].

³ Concretely, we use, as $\mathcal{T}$, the land use indicators in this task (e.g., office and residence).

where W^P is a type-specific linear projection matrix. Through this, we derive the initial node representation $h_i^{(0)}$ for node i .

Subsequently, we implement a multi-head mechanism to compute the message representation $m^{(l)}(s, e, t)$ at layer $l \geq 0$ from source nodes v_s to target nodes v_t . This process accommodates the types of nodes $\phi(s)$, $\phi(t)$ and edges $\psi(e)$ in the heterogeneous graph, as expressed:

$$m^{(l)}(s, e, t) = \bigoplus_{\tau \in [1, \theta]} \left(\left(W_{\tau, \phi(s)}^{\text{MH}} h_s^{(l-1)} \right) W_{\psi(e)}^{\text{MP}} \right) \quad (4)$$

Here, $\bigoplus$ represents the concatenation operation, and θ the total number of heads, while $W_{\tau, \phi(s)}^{\text{MH}}$ and $W_{\psi(e)}^{\text{MP}}$ are matrices designated for processing based on node and edge types.

Next, we apply a multi-head attention framework to derive attention weights $\alpha^{(l)}(s, e, t)$, facilitating the interaction between source v_s and target v_t nodes via edges $\psi(e)$, described as follows:

$$\text{head}_{\tau}^{(l)}(s, e, t) = \left(\left(W_{\tau}^Q h_s^{(l-1)} \right)^T W_{\tau}^K \left(W_{\tau}^K h_t^{(l-1)} \right)^T \right) \frac{1}{\sqrt{\xi}}, \quad (5)$$

$$\alpha^{(l)}(s, e, t) = \text{Softmax} \left(\bigoplus_{\tau \in [1, \theta]} \text{head}_{\tau}^{(l)}(s, e, t) \right), \quad (6)$$

where W_{τ}^Q , W_{τ}^K , and W_{τ}^A are the linear projection matrices per head τ , tailored according to node and edge types, and ξ is an adaptive scaling factor for the attention mechanism.

Finally, the information is aggregated from the neighbours of the target node with attention weights to update the representation of the target node, expressed as:

$$h_t^{(l)} = W_{\phi(t)}^{\text{AG}} \left(\sum_{v_s \in \mathcal{N}(v_t)} \left(\alpha^{(l)}(s, e, t) \cdot m^{(l)}(s, e, t) \right) \right) + h_t^{(l-1)}, \quad (7)$$

with W^{AG} as the aggregation-specific linear matrix considering the type of the target node. Here, $\mathcal{N}(v_t)$ denotes the set of neighbour nodes of the target node v_t .

4.2. Explainable AI methods for post-hoc explanations

This section discusses the XAI methods proposed to enhance explainability in land use inference: feature attribution analysis and counterfactual analysis. These methods are motivated by potential trust issues among stakeholders, particularly in urban planning and development, regarding predictive models, especially when dealing with complex input-output relationships. Feature attribution explanations help stakeholders better understand the importance of specific elements within the models. Additionally, counterfactual explanations enable stakeholders to explore minimal changes that could lead to desired outcomes, such as balanced land use. We demonstrate how these XAI methods can be applied in the land use inference context, which could potentially be useful for improving decision-making and fostering trust in model predictions.

4.2.1. Feature attribution methods

The first class of techniques we use to explain the predictions of our model are *feature attribution methods*. These techniques quantify the approximate importance of the individual input features on a particular model prediction through linear *feature attribution scores*. Such scores can provide invaluable insights regarding the reasoning of deep learning models and uncover possible issues or biases [69,70].

In the context of explaining an HGN model $\mathcal{M}$ applied to a heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, we can formalise the feature attribution scores as follows. Given a target node in the graph $v_i \in \mathcal{V}$ and a target output $y_{i,j}$, our goal is to compute the attribution scores $S \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{F}|}$ estimating the importance of every node and feature combination for the model output $y_{i,j}$. Intuitively, the element $S_{k,l}$ of the attribution score matrix provides an estimate of how much the l th feature of the k th node, $x_{k,l}$, affected the j th output at the i th node, $y_{i,j}$. To get a more holistic view of the importance of the different features, we will also consider

the aggregated attribution scores $a \in \mathbb{R}^{|\mathcal{F}|}$ where $a_k = \sum_{i=0}^{|\mathcal{V}|} S_{i,k}$. Each element of the aggregated attribution scores vector estimates how much a particular input feature affected the target output when considering its combined effect from all the nodes in the graph.

In our work, we use two different feature attribution methods for explaining our HGN model—Gradient · Input [71] and Integrated Gradients [72]. Gradient · Input is a simple gradient-based attribution method defined as follows, in our setting:

Definition 4 (Gradient · Input Explanation). Given an explained model $\mathcal{M}$, an input graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a target node $v_i \in \mathcal{V}$ and a target output $y_{i,j}$, the Gradient · Input feature attributions scores are defined as:

$$S_{k,l}^{G \cdot I} = x_{k,l} \cdot \frac{\partial y_{i,j}}{\partial x_{k,l}} = x_{k,l} \cdot \frac{\partial \mathcal{M}(\mathcal{G})_{i,j}}{\partial x_{k,l}}$$

The advantage of Gradient · Input is its computational efficiency, its conceptual simplicity as well as its good ability to faithfully capture the behaviour of the model in the small region around the considered input [73]. However, the attributions produced by Gradient · Input may also be unreliable, particularly if the function of the considered model is highly irregular (resulting in a problem known as shattered gradients [74]) or saturates for the given input (e.g., due to the predicted output already being sufficiently high). These limitations are largely addressed by Integrated Gradients explanations, which are defined as follows, in our setting:

Definition 5 (Integrated Gradients Explanation). Given an explained model $\mathcal{M}$, an input heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a target node $v_i \in \mathcal{V}$, a target output $y_{i,j}$ and a reference graph $\bar{\mathcal{G}} = (\bar{\mathcal{V}}, \bar{\mathcal{E}})$, the Integrated Gradients feature attributions scores are defined as:

$$S_{k,l}^{IG} = (x_{k,l} - \bar{x}_{k,l}) \cdot \int_{\alpha=0}^1 \frac{\partial \mathcal{M}(\bar{\mathcal{G}} + \alpha \cdot (\mathcal{G} - \bar{\mathcal{G}}))_{i,j}}{\partial x_{k,l}} d\alpha$$

In contrast with Gradient · Input, which only considers the gradient at the explained input instance, Integrated Gradients are based on an aggregate gradient on the path from some suitable reference to the currently considered input. This helps to naturally smooth out any local irregularities in the gradient and to capture the influence of the input features even if their effect already saturated close to the currently considered input [72]. In our experiments, we will consider the neutral reference graph $\bar{\mathcal{G}}$ with the same structure as the input graph $\mathcal{G}$, but all the features on the nodes and edges set to zero. Using zero feature values has been advocated as a reasonable default choice for the reference input [72]. We must also note that Integrated Gradients in the pure form defined above are typically intractable for complex deep learning models. Thus, in practice, the integrated gradient scores are computed, in our setting, via the following approximation:

Definition 6 (Approximate Integrated Gradients Explanation). Given an explained model $\mathcal{M}$, an input heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a target node $v_i \in \mathcal{V}$, a target output $y_{i,j}$, a reference graph $\bar{\mathcal{G}} = (\bar{\mathcal{V}}, \bar{\mathcal{E}})$ and a number of steps n , the approximate Integrated Gradients feature attributions scores are defined as:

$$S_{k,l}^{IG-A} = (x_{k,l} - \bar{x}_{k,l}) \cdot \sum_{k=0}^n \frac{1}{n} \cdot \frac{\partial \mathcal{M}(\bar{\mathcal{G}} + \frac{k}{n} \cdot (\mathcal{G} - \bar{\mathcal{G}}))_{i,j}}{\partial x_{k,l}}$$

Throughout this paper, when considering Integrated gradient explanations, we will use the above approximation with $n = 50$ as the number of steps.

In addition to the above gradient-based methods, we also initially considered perturbation approaches based on Shapley Values [61]. However, given the large size of the considered input graph, we found these to be highly computationally inefficient, taking several hours to compute the explanation for a single output. We also note that there are several bespoke methods specifically focused on explaining

GNN models, such as GNN-Explainer [58]. Unfortunately, the available implementations for these methods were not compatible with the library for HGNs that we used, which is why we opted to use the above methods instead. Extending GNN-Explainer and other GNN-specific explainability methods to our heterogeneous model poses an interesting line of future work, but is outside the scope of this paper.

4.2.2. Counterfactual explanations

In contrast to attributing model output predictions to parts of input features, Counterfactual Explanations (CEs) reveal the minimal changes needed from the factual input to achieve a different and more desirable output prediction [65]. In this section, we define CEs suitable for the land use inference task with heterogeneous graphs. For graph data, CEs often take the form of subgraphs or modified graph inputs [75] due to the connectivity between the nodes, which is especially the case in land use inference since our interest is always on an area [76]. We start by defining the subgraph consisting of a target node and its neighbouring nodes.

Definition 7 (1-Hop Complete Subgraph). Given an input heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a target node $v_i \in \mathcal{V}$, the 1-hop complete subgraph of v_i is defined as $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$, where $\mathcal{V}_i = \{v_j \mid e_{i,j} \in \mathcal{E} \vee e_{j,i} \in \mathcal{E}\} \cup \{v_i\}$ and $\mathcal{E}_i = \{e_{j,k} \mid v_j, v_k \in \mathcal{V}_i, e_{j,k} \in \mathcal{E}\}$.

Another concept central to CEs for graph data is the dissimilarity measure between two subgraphs, which allows comparisons of candidate CEs. Specifically, suppose we have two candidate subgraph CEs; that which is more similar to the input is considered the better candidate, because it would require less effort to change from the input [65]. Given the multi-modal nature of our data, we design a dissimilarity measure covering the differences in node features, node types, edge types, and (sub)graph structure as follows:

Definition 8 (Local Subgraph Dissimilarity). Given an input heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, two 1-hop complete subgraphs $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$ and $\mathcal{G}_j = (\mathcal{V}_j, \mathcal{E}_j)$ for nodes $v_i, v_j \in \mathcal{V}$, respectively, the local subgraph dissimilarity between $\mathcal{G}_i$ and $\mathcal{G}_j$ is:

$$d(\mathcal{G}_i, \mathcal{G}_j) = d_{\mathcal{V}}(\mathcal{G}_i, \mathcal{G}_j) + d_{\mathcal{A}}(\mathcal{G}_i, \mathcal{G}_j) + d_{\mathcal{R}}(\mathcal{G}_i, \mathcal{G}_j) + d_{\mathcal{G}}(\mathcal{G}_i, \mathcal{G}_j)$$

where:

$$d_{\mathcal{V}}(\mathcal{G}_i, \mathcal{G}_j) = \frac{1}{|\mathcal{V}_j|} \sum_{v_k \in \mathcal{V}_j} \frac{1}{2} (L2_dist(x_i, x_k) + \cosine_dist(x_i, x_k))$$

is the node feature dissimilarity, $L2_dist$ and $\cosine_dist$ are L2 (Euclidean) distance and cosine distance respectively,

$$d_{\mathcal{A}}(\mathcal{G}_i, \mathcal{G}_j) = 1 - \frac{\sum_{k=1}^{|\mathcal{A}|} \min(\overline{\phi(\mathcal{V}_i)_k}, \overline{\phi(\mathcal{V}_j)_k})}{\sum_{k=1}^{|\mathcal{A}|} \max(\overline{\phi(\mathcal{V}_i)_k}, \overline{\phi(\mathcal{V}_j)_k})}$$

is the node type dissimilarity, ϕ maps nodes to its corresponding type, $\phi(\mathcal{V}_i)$ is the node types of subgraph nodes $\mathcal{V}_i$, $\overline{\phi(\mathcal{V}_i)} \in \mathbb{R}^{|\mathcal{A}|}$ represents the counts of each node type in $\mathcal{A}$,

$$d_{\mathcal{R}}(\mathcal{G}_i, \mathcal{G}_j) = abs(\frac{1}{|\mathcal{E}_i|} \sum_{e_{k,l} \in \mathcal{E}_i} \psi(e_{k,l}) - \frac{1}{|\mathcal{E}_j|} \sum_{e_{m,n} \in \mathcal{E}_j} \psi(e_{m,n}))$$

is the edge type dissimilarity, ψ maps edges to its corresponding type,

$$d_{\mathcal{G}}(\mathcal{G}_i, \mathcal{G}_j) = ||\mathcal{E}_i| - |\mathcal{E}_j||$$

is the graph structure dissimilarity, $abs(*)$ takes the absolute value of a real number.

This score combines four sources of dissimilarities. Firstly, $d_{\mathcal{V}}$ denotes the node features differences where L2 captures the magnitude differences, while cosine distance captures the traffic pattern differences. $d_{\mathcal{A}}$ calculates the node type differences between the nodes in the two subgraphs using the multiset Jaccard distance [77]. $d_{\mathcal{R}}$ calculates the absolute differences between the average road types, obtained by

the edge types in both subgraphs. Finally, $d_{\mathcal{G}}$ represents the graph structure differences expressed as the absolute difference between the numbers of edges, which is often considered as sparsity in graph CEs [75]. We now use Local subgraph dissimilarity as a core heuristic⁴ to guide the search for CEs:

Definition 9 (Counterfactual Explanation). Given an explained model $\mathcal{M}$, an input heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a target node $v_i \in \mathcal{V}$ with predicted output y_i , a desirable prediction result $y' \neq y_i$, the counterfactual explanation for v_i 's prediction is $\mathcal{G}_j = (\mathcal{V}_j, \mathcal{E}_j)$, the 1-hop complete subgraph for node v_j , which is:

$$\underset{v_j \in \mathcal{V} | y_j = y'}{argmin} d(\mathcal{G}_i, \mathcal{G}_j)$$

Our definition finds a 1-hop complete subgraph from the given heterogeneous graph data as the CE for a given target node. It is the closest (measured by local subgraph dissimilarity) subgraph to the 1-hop complete subgraph of the target node, among all the subgraphs of nodes which are predicted with a pre-specified desirable prediction by our explained model. Our method has the same spirit as the nearest-neighbour CEs method for tabular data, which finds the closest point in the training dataset as the CE for an input point [78]. Such CEs taken from the dataset have the advantage of lying within the data manifold, therefore they are realistic data and not algorithmic artefacts [79]. Additionally, they are more robust to perturbations that could occur in the inference pipeline [80].

Practically, the CEs can be computed as the following. First, initialise the 1-hop complete subgraph for the target node. Then, filter out the nodes predicted with the desirable label. In an iterative procedure, for each of these nodes, take their 1-hop complete subgraph, and compute the local subgraph dissimilarity between the subgraph and the input node's subgraph. The subgraph with the lowest dissimilarity is returned as the CE. As will be seen in Section 5.7.3, we specify a mixed prediction and use CEs to investigate what changes are needed to achieve a community with balanced land use [81,82].

5. Empirical experiments

5.1. Dataset description

To achieve comprehensive coverage and enable fine-grained analysis with varied urban layout and distinct mobility modes, we selected four distinct areas for building datasets: Inner London, Outer London, San Francisco, and New York City. The specific cities and the mobility modes used (and the corresponding open source data) are listed in Table 3.

Specifically, the spatial distributions of these datasets are illustrated in Fig. 3—London (top row), San Francisco (middle row), and New York City (bottom row).

Among the cities, Inner and Outer London were chosen as the study areas for post-hoc explanation due to their complex and diverse land use. The inner regions feature dense mobility networks, whereas the outer regions exhibit contrasting patterns and structures. This dual-dataset approach ensures that our analysis captures the full spectrum of urban mobility characteristics across both densely populated and suburban areas of the city.

As mentioned, this paper presents a pipeline that processes mobility data into heterogeneous graph data and employs GNNs combined with explainability techniques for land use inference. Although we primarily focus on the four chosen regions, the pipeline is flexible, enabling the construction of heterogeneous graph datasets for other area tailored to specific needs and available mobility data.

⁴ Note that our dissimilarity notion is not a distance, e.g., it does not satisfy symmetry, as $d(\mathcal{G}_i, \mathcal{G}_j) \neq d(\mathcal{G}_j, \mathcal{G}_i)$.

Table 3
Comparison of datasets across different urban areas.

Name	Residents' Travel Behaviour	Mobility Modes
Inner London	High population density with short commuting distances; extensive use of public transport and bike-sharing	Bus, Tube, bike sharing
Outer London	Primarily suburban environment with longer commute distances; multiple public transport options	Bus, Tube, bike sharing
San Francisco	Central areas rely on car travel supplemented by bike-sharing; peripheral districts are more dispersed	Parking, bike sharing
New York	Extensive public transport coverage; strong demand for both metro and bike-sharing	Metro, bike sharing

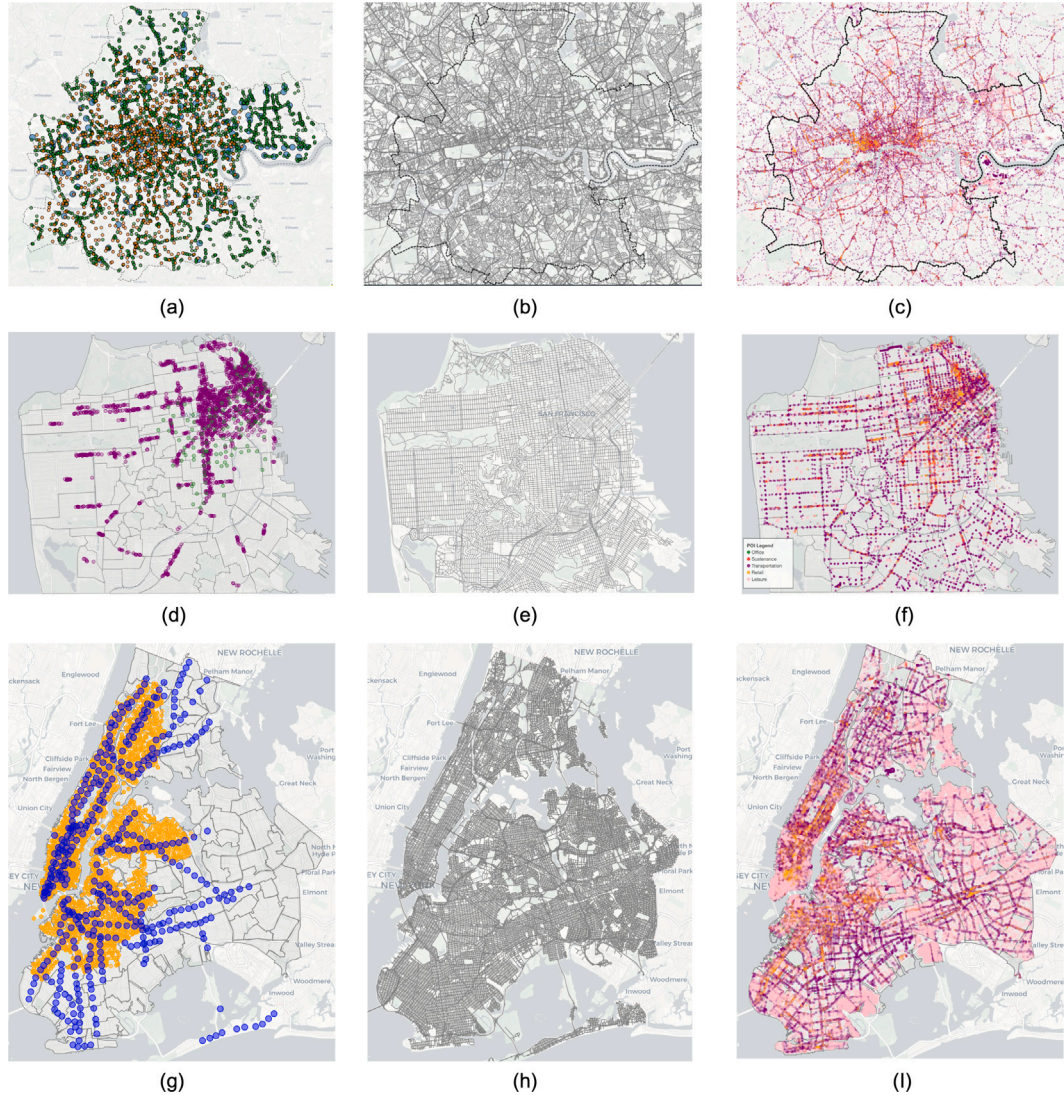


Fig. 3. Various urban infrastructure and POI distributions across three cities: (a) locations of London's tube (blue), bus (green), and cycling (orange) stations; (b) London's road network; (c) distribution of five categories of POIs in London—office (green), sustenance (red), transport (purple), retail (orange), and leisure (pink); (d) locations of San Francisco's parking meters (purple) and cycling stations (green); (e) road network of San Francisco; (f) distribution of five categories of POIs in San Francisco with same colour set of the London one; (g) locations of New York City's metro stations (blue) and cycling stations (orange); (h) road network of New York City; and (i) distribution of five categories of POIs in New York City with same colour set of the London one.

5.2. Dataset processing

The constructed graph datasets comprise three main parts: node features, edge features, and labels. In addition, we set the catchment area (the geographical zone surrounding each node that reflects its area of influence or service, typically defined by a specific radius or travel

distance) to determine the spatial extent of the land use measurement corresponding to each graph node. These elements are introduced on below:

- **Node features.** For any mobility station, we use its 15-minute-boarding distribution on a 'typical weekday' (as averaged over

the weekdays) as the input nodes feature.⁵ The daily passenger inbound data of each station is processed into 15-minute interval ridership from 6:00 to 22:00 (64 input features). The specific pipeline for feature pre-processing is given in [Appendix A](#).

- **Edge features.** The edges are based on comprehensive urban road networks, as shown in [Fig. 3\(b\)](#), (e), and (h). There are specifically five classes of road edge categories involved: Primary, secondary, tertiary, residential, and unclassified (a detailed introduction of this information can be found at [wiki of OSM](#)). These types of edge-level geography information are collected from the Open Street Map and encoded into one-hot vectors to represent the edge-level information. We do not use connectivity in the transit network because the available data does not distinguish the direction of routes among the bus stops. Although including this information would further enrich the information contained in the heterogeneous graph, it is beyond the scope of this research.
- **Labels.** In the land use inference problem, there is no established paradigm for labelling. We adhere to the recommended practices of the most cited foundational studies [1] – using the density of Points of Interests (POIs, locations with specific significance such as restaurants, landmarks, or hospitals) to characterise the intensity of specific land use characteristics in a region. To ensure comprehensiveness and objectivity, we selected six POI categories from the OSM database: office, sustenance, transport, retail, leisure and residence. Due to the restriction of OSM's API on the quantity of data collected, it is difficult to obtain complete POI data. Consequently, we chose a few major tags to represent each land use category, as detailed in [Table 4](#). The spatial distributions of these labels are displayed in [Fig. 3\(c\)](#), (f), and (i).
- **Catchment area.** We establish a criterion that assigns each node a catchment area—a circular region centred on the node with a predetermined radius. Each node is labelled by the land use indicators within this catchment area. This process considers two types of label tags:
 - For point tags (e.g., locations of different services), count the number of target types within the catchment area.
 - For polygon tags (e.g., parks, buildings), calculate the proportion of the catchment area that these tags cover.

We have used a 1 km radius for the catchment area in our analysis. As our focus is on demonstrating the use of heterogeneous graph models, a comparison of different catchment radii and their impacts is left to future work.

We further normalise the heterogeneous graph dataset as the final step of preprocessing. We tested various normalisation methods and selected the optimal combination of z-score normalisation for node features, one-hot encoding for edge types and min-max normalisation for labels. Since this aspect is not the primary focus of our paper, we opt not to include a detailed discussion of this process. To assist readers in applying our work to their own urban areas of interest, we have provided a detailed workflow that includes the entire data acquisition and processing pipeline (as shown in [Fig. A.13](#)).

5.3. Baseline models for comparison

Since the proposed model in this paper is designed to fuse multi-modal mobility information within urban environments through heterogeneous graph modelling, our primary focus is on the significance of heterogeneity for land use inference. In light of this, we selected

baseline models that include not only basic neural network models but also a series of classical homogeneous graph models, as well as an early heterogeneous graph model. These models are widely used as baseline models in the fields of social network analysis [83], recommendation systems [47], and traffic flow prediction [84].

- **Neural Network (NN):** A basic neural network model, serving as a foundational benchmark against more complex graph-based models.
- **Graph Convolutional Network (GCN):** Utilises convolutional processes directly on the graph structure to capture neighbourhood information [83].
- **Graph Attention Network (GAT):** Incorporates attention mechanisms into graph neural networks, enabling nuanced weighting of node contributions from neighbours [48].
- **GraphSage:** A scalable graph neural network method that generates embeddings by sampling and aggregating features from a node's local neighbourhood [47].
- **Relational Graph Convolutional Network (RGCN):** Extends GCNs to manage graph heterogeneity by incorporating various relational types among nodes during the learning process [85].

Each of these models represents a unique approach to incorporating graph structure into the learning process, which makes them suitable baselines for assessing the performance of our proposed HGT model.

5.4. Experiment setting

To ensure fair comparisons against multiple baseline models, we used the same configuration to model, train, and test them, as detailed in [Table 5](#). The specific experimental details are given below:

System and python environment: The experiments were undertaken on a Linux Ubuntu system, utilising CUDA Version 11.8 with NVIDIA RTX A5000 GPU for deep learning model optimisation. Computations were executed on Python 3.10 within a dedicated virtual environment (Readers can check “requirement.txt” on the GitHub website of this project).

Training setting: For comprehensive model evaluation and testing, we randomly allocated 60% of the samples for training, with the remaining 20% for validation and 20% for testing. We choose the training epoch number and batch size as 200 and 64, respectively.

General hyper-parameters: The depth of the model (Hidden layer number) is often a trade-off between model complexity and performance. We take it as 2 to strike a balance by allowing the model to capture sufficient graph information without introducing excessive complexity or overfitting. For the activation function, ReLU was chosen for its computational efficiency and ability to avoid the vanishing gradient problem, enabling better training in deep networks. For hidden dimension, choice of optimiser and learning rate, we performed sensitivity analysis and chose the combination of compatible accuracy and computational efficiency as shown in [Table 5](#). The detailed analysis can be found in [Appendix C](#).

Graph model's hyper-parameters: As this paper focuses on processing a large graph, small batch data loading at the graph level is a crucial step for optimising performance and training speed. PyG offers several small-batch data-loading methods for homogeneous graphs. After comparison, we selected the best-performing neighbour sampling method for homogeneous graph models. For the HGT method, we used the HGTLoader. As the method of dividing a graph into small batches has been well studied [41,80] and is not the focus of this paper, we will not analyse this part in detail. For the graph hop number, we opted for a 2-hop setting, ensuring the model remains efficient and feasible to run. For the attention-head number in HGT, since this was not the focus of the study, we only tested the case where the number of heads is 2.

⁵ London's bus service records only boarding numbers, not alighting, and bike data for weekends is unavailable. Consequently, we used only weekday boarding flow as input to ensure fair comparisons among three mobility modes.

Table 4
POI categories and their specific tags for labelling land use.

Category	Tags	Data types
Office	Government, Association, Company, Consulting, Financial, Coworking, Lawyer, Estate, Insurance, Telecommunication	Points
Sustenance	Restaurant, Cafe, Bar, Pub, Fast Food	Points
Transport	Bicycle, Subway, Transport, Taxi, Parking, Ferry, Tram	Points and Polygons
Retail	Convenience, Supermarket, Mall, Department, Bakery, Butcher, Clothes, Hardware, Furniture, Electronics	Points
Leisure	Park, Sports, Playground, Swimming, Pitch, Track, Fitness, Garden, Nature	Points and Polygons
Residence	Residential, House, Detached, Apartments, Population	Points and Polygons

Table 5
Summary of Hyper-Parameter Settings.

Parameter	Value
Hidden Layer Numbers	2
Hidden Layer Dimension	128
Activation Function	ReLU
Attention-Head Numbers	2
Optimiser	AdamW [86]
Epoch Number	200
Batch-Size	64
Learning Rate	0.002
Sampling for Homo-Graph	neighbour Loader [47]
Sampling for Hetero-Graph	HGT Loader [87]
Graph Hop Number	2

5.5. Measurement metrics

To evaluate the performance of our regression models, we employ three widely recognised metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). These metrics are essential for assessing the accuracy and predictive quality of the models. They are defined as follows, where n is the number of observations in the test set, y_i represents the actual values, and $\hat{y}_i$ are the predicted values.

- **Mean Absolute Error (MAE):** MAE quantifies the average absolute difference between the predicted values and the actual values, giving an insight into the overall error magnitude without the influence of error direction. Lower MAE values indicate better model performance, as they signify smaller errors.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- **Root Mean Square Error (RMSE):** RMSE is a quadratic scoring rule that calculates the square root of the average of squared differences between the predicted and actual values. It emphasises larger errors, which can be particularly important in certain contexts. Like MAE, lower RMSE values reflect better predictive performance.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- **Coefficient of Determination (R-Squared):** R^2 is a statistical measure that indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. It is used to measure how well the model captures the variability of the dataset. Higher R^2 values indicate better model performance, as they represent a greater proportion of explained variance.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

where $\bar{y}$ is the mean of the actual values y_i .

5.6. Results

5.6.1. Overall accuracy

This section offers a comprehensive evaluation of various deep learning methods using the four chosen datasets. The models evaluated included the proposed HGT framework and were compared with the baseline methods mentioned in Section 5.2. We first compare the overall performance of our framework across the datasets in Table 6. The indicator (Impr.) is measured by improvement percentage over the baseline neural network (NN) model. Subsequently, in 7, we provide a more detailed analysis of the model's accuracy variations across six land use reference tasks.

Table 6 indicate a consistent improvement in performance when transitioning from general deep learning models to complex graph-based models among four datasets. The HGT model consistently delivers the highest improvement percentages across all categories, showing its capability to handle heterogeneous graph data. In addition, we observe that the Outer London dataset achieves the best prediction performance among the four, likely because its land use complexity is lower than that of the other regions. Furthermore, the two London datasets outperform those from San Francisco and New York City, possibly because London data incorporate three distinct mobility types, whereas the San Francisco and New York City datasets include only two. This richer and more diverse set of mobility types seems to provide the models with deeper insights into human activity patterns, thereby enhancing overall prediction accuracy in both Inner and Outer London.

7 specifically compare the performance of the models in different tasks in London. For the Inner London area, the 'Sustenance' indicator has substantial benefits from graph-based models, with the HGT achieving MAE improvements of 68.35%. The RGCN and HGT models consistently outperform others, probably due to their ability to capture heterogeneous relationships within the mobility and structure data. The smallest improvement is noted in the 'leisure' land use indicator, suggesting that there might be unobserved information exists to boost predictive accuracy.

In Outer London, the HGT model performs better than the baseline GNNs. A possible reason for this is that the sparser connectivity in Outer London decreases the performance of conventional GNNs, but has less impact on the HGT model because of the ability to comprehensively fuse multi-mode mobility information like people's movements and traffic facilities.

Detailed performance of models in San Francisco and New York City can be found in 8. It can be observed that, HGT demonstrates robust predictive accuracy in most examples and offers a notable improvement over the baseline models (NN and GCNs). Due to the relatively limited number of nodes, the accuracies in this experiment are slightly lower than those observed in London, yet significantly higher than the results from another study [3] in San Francisco.

5.6.2. Model performance evaluation

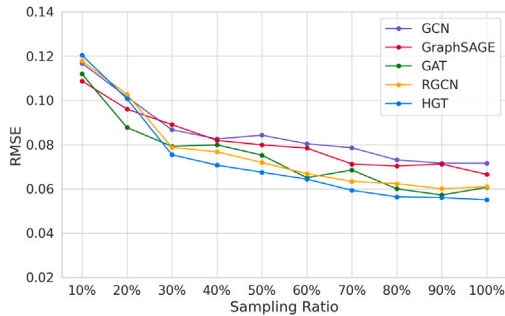
For the validation of the proposed model, this subsection evaluates the impact of the ratio of the training set, generalisation and Residual distribution.

Table 6
Comparison of overall performance across four datasets.

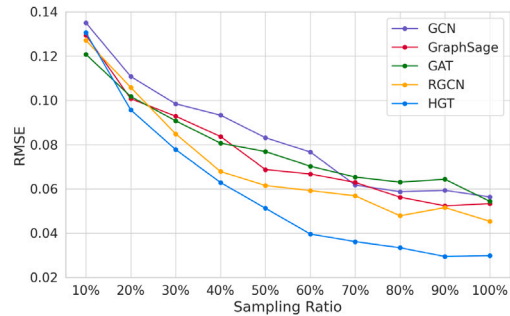
Dataset	Metrics	NN	GCN	GraphSage	GAT	RGCN	HGT
Inner London	MAE	0.099	0.051	0.048	0.046	0.042	0.036
	Impr.	–	48.48%	51.52%	53.54%	57.58%	63.64%
	RMSE	0.135	0.069	0.066	0.063	0.059	0.054
	Impr.	–	48.89%	51.11%	53.33%	56.30%	60.00%
Outer London	MAE	0.077	0.056	0.053	0.054	0.045	0.030
	Impr.	–	27.27%	31.17%	29.87%	41.56%	61.04%
	RMSE	0.111	0.081	0.076	0.075	0.064	0.055
	Impr.	–	27.03%	31.53%	32.43%	42.34%	50.45%
San Francisco	MAE	0.109	0.067	0.063	0.064	0.057	0.045
	Impr.	–	38.44%	42.20%	41.28%	47.71%	58.72%
	RMSE	0.159	0.097	0.091	0.091	0.074	0.072
	Impr.	–	38.99%	42.77%	42.77%	53.46%	54.72%
New York City	MAE	0.118	0.072	0.067	0.067	0.060	0.051
	Impr.	–	38.98%	43.22%	43.22%	49.15%	56.78%
	RMSE	0.166	0.100	0.094	0.094	0.082	0.077
	Impr.	–	39.76%	43.37%	43.37%	50.60%	53.61%

Table 7
Performance metrics and improvements across different land use targets in Inner London and Outer London.

Indicators	Metrics	Inner London						Outer London					
		NN	GCN	GraphSage	GAT	RGCN	HGT	NN	GCN	GraphSage	GAT	RGCN	HGT
Office	MAE	0.086	0.043	0.043	0.039	0.035	0.031	0.071	0.051	0.047	0.046	0.039	0.026
	Impr	–	50.00%	50.00%	54.65%	59.30%	63.95%	–	40.70%	45.35%	46.51%	54.65%	69.77%
	RMSE	0.124	0.060	0.058	0.054	0.051	0.047	0.112	0.080	0.072	0.069	0.058	0.052
	Impr	–	51.61%	53.23%	56.45%	58.87%	62.10%	–	35.48%	41.94%	44.35%	53.23%	58.06%
Sustenance	MAE	0.079	0.033	0.032	0.034	0.031	0.025	0.065	0.046	0.043	0.042	0.038	0.024
	Impr	–	58.23%	59.49%	56.96%	60.76%	68.35%	–	41.77%	45.57%	46.84%	51.90%	69.62%
	RMSE	0.116	0.046	0.045	0.049	0.044	0.039	0.099	0.066	0.061	0.058	0.054	0.046
	Impr	–	60.34%	61.21%	57.76%	62.07%	66.38%	–	43.10%	47.41%	50.00%	53.45%	60.34%
Transport	MAE	0.097	0.046	0.043	0.039	0.039	0.033	0.078	0.056	0.055	0.053	0.046	0.031
	Impr	–	52.58%	55.67%	59.79%	59.79%	65.98%	–	42.27%	43.30%	45.36%	52.58%	68.04%
	RMSE	0.139	0.063	0.059	0.054	0.054	0.050	0.111	0.079	0.075	0.073	0.064	0.055
	Impr	–	54.68%	57.55%	61.15%	61.15%	64.03%	–	43.17%	46.04%	47.48%	53.96%	60.43%
Retail	MAE	0.103	0.057	0.053	0.050	0.046	0.039	0.078	0.062	0.060	0.055	0.051	0.031
	Impr	–	44.66%	48.54%	51.46%	55.34%	62.14%	–	39.81%	41.75%	46.60%	50.49%	69.90%
	RMSE	0.134	0.077	0.070	0.068	0.065	0.058	0.111	0.084	0.080	0.075	0.070	0.056
	Impr	–	42.54%	47.76%	49.25%	51.49%	56.72%	–	37.31%	40.30%	44.03%	47.76%	58.21%
Leisure	MAE	0.106	0.064	0.060	0.058	0.051	0.044	0.089	0.068	0.060	0.071	0.050	0.037
	Impr	–	39.62%	43.40%	45.28%	51.89%	58.49%	–	35.85%	43.40%	33.02%	52.83%	65.09%
	RMSE	0.141	0.086	0.083	0.080	0.069	0.067	0.131	0.102	0.092	0.098	0.073	0.074
	Impr	–	39.01%	41.13%	43.26%	51.06%	52.48%	–	27.66%	34.75%	30.50%	48.23%	47.52%
Residence	MAE	0.123	0.062	0.058	0.054	0.051	0.042	0.078	0.055	0.055	0.059	0.048	0.030
	Impr	–	49.59%	52.85%	56.10%	58.54%	65.85%	–	55.28%	55.28%	52.03%	60.98%	75.61%
	RMSE	0.158	0.083	0.078	0.074	0.070	0.062	0.104	0.075	0.073	0.076	0.064	0.048
	Impr	–	47.47%	50.63%	53.16%	55.70%	60.76%	–	52.53%	53.80%	51.90%	59.49%	69.62%



(a) Analysis on Inner London.



(b) Analysis on Outer London.

Fig. 4. Comparison of different ratios of the training dataset.

Impact of training set size We randomly sampled the training set at ratios from 10% to 100%. The performance of the proposed HGT model,

in terms of average RMSE, compared with various GNNs, is depicted in Fig. 4.

Table 8

Performance metrics and improvements across different land use targets in San Francisco and New York City.

Indicators	Metrics	San Francisco						New York City					
		NN	GCN	GraphSage	GAT	RGCN	HGT	NN	GCN	GraphSage	GAT	RGCN	HGT
Office	MAE	0.108	0.061	0.057	0.057	0.049	0.041	0.120	0.065	0.060	0.058	0.050	0.045
	Impr	–	43.50%	46.91%	47.57%	54.29%	62.32%	–	45.83%	50.00%	51.67%	58.33%	62.50%
	RMSE	0.170	0.096	0.089	0.085	0.075	0.068	0.175	0.095	0.090	0.085	0.078	0.070
	Impr	–	43.45%	47.74%	50.23%	56.05%	59.75%	–	45.71%	48.57%	51.43%	55.43%	60.00%
Sustenance	MAE	0.106	0.056	0.053	0.053	0.048	0.045	0.110	0.060	0.057	0.056	0.050	0.048
	Impr	–	46.95%	49.69%	50.39%	54.75%	57.42%	–	45.45%	48.18%	49.09%	54.55%	56.36%
	RMSE	0.162	0.081	0.076	0.073	0.069	0.065	0.162	0.085	0.080	0.078	0.072	0.069
	Impr	–	49.69%	53.19%	54.67%	57.46%	59.84%	–	47.53%	50.49%	51.23%	55.56%	57.41%
Transport	MAE	0.091	0.065	0.061	0.062	0.055	0.040	0.100	0.070	0.067	0.065	0.058	0.050
	Impr	–	28.88%	33.15%	32.13%	39.68%	56.44%	–	30.00%	33.00%	35.00%	42.00%	50.00%
	RMSE	0.130	0.091	0.086	0.085	0.064	0.067	0.140	0.095	0.090	0.089	0.078	0.070
	Impr	–	29.65%	33.84%	34.39%	50.91%	48.64%	–	32.14%	35.71%	36.43%	44.29%	50.00%
Retail	MAE	0.095	0.071	0.069	0.064	0.060	0.040	0.110	0.075	0.072	0.070	0.065	0.052
	Impr	–	25.04%	27.59%	32.59%	36.60%	57.98%	–	31.82%	34.55%	36.36%	40.91%	52.73%
	RMSE	0.135	0.097	0.092	0.093	0.069	0.065	0.150	0.100	0.095	0.093	0.085	0.073
	Impr	–	28.31%	31.62%	31.01%	48.65%	51.93%	–	33.33%	36.67%	38.67%	43.33%	51.33%
Leisure	MAE	0.125	0.080	0.072	0.076	0.068	0.063	0.130	0.085	0.075	0.079	0.070	0.063
	Impr	–	35.83%	42.25%	39.04%	45.39%	49.74%	–	34.62%	42.31%	39.23%	46.15%	51.54%
	RMSE	0.184	0.119	0.109	0.116	0.100	0.101	0.190	0.121	0.113	0.119	0.104	0.102
	Impr	–	35.40%	40.56%	37.14%	45.63%	45.28%	–	36.32%	40.52%	37.37%	45.26%	46.32%
Residence	MAE	0.129	0.067	0.068	0.072	0.060	0.043	0.135	0.078	0.073	0.076	0.065	0.050
	Impr	–	48.13%	47.53%	44.48%	53.50%	66.85%	–	42.22%	45.93%	43.70%	51.85%	62.96%
	RMSE	0.172	0.097	0.092	0.093	0.069	0.065	0.178	0.105	0.098	0.100	0.072	0.078
	Impr	–	46.72%	48.19%	46.54%	53.01%	55.87%	–	41.08%	45.03%	43.86%	59.68%	56.19%

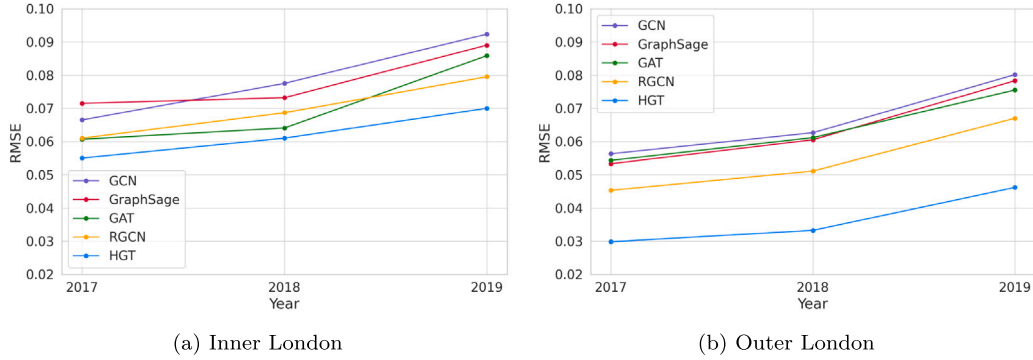


Fig. 5. Compared with external datasets collected in 2018 and 2019.

The results indicate that the HGT model's performance improves as the size of the training set increases. When trained with a limited sample size (10%–20% of the full dataset), the HGT model performs similarly to other graph neural networks. This may be because the larger number of parameters in the HGT model are not fully utilised with smaller datasets. However, as the training set size increases, the HGT model outperforms the other models, demonstrating a greater capacity to leverage more data for improved performance.

Generalisation To validate the model's performance with external datasets, we also collected the mobility data in 2018 and 2019 of the test data masks from the same data source. The difference in average RMSE for different models is shown in Fig. 5.

The figure illustrates the RMSE of different models in 2017, 2018, and 2019. The additional error introduced by the different years stems from changes in demand patterns due to the evolving functionality of urban areas. In a highly developed city, these changes occur gradually, so the difference in RMSE between 2017 and 2018 is minimal. However, as these changes accumulate over time, the error increases,

leading to a more significant marginal difference in RMSE observed in 2019 compared to 2018.

Residual distribution To better understand the function of heterogeneous graph models at the individual level, we compare the geographical distribution of residuals in the geometric map between GCN and HGT. Fig. 6 illustrates an example of one of the land use indicators – 'Residence', where the size of the circular markers indicates the absolute value of the residuals. The colour of each marker indicates whether the residual is positive or negative: red for over-predicted values and blue for underestimated values. Similar comparisons of other land use indicators can be found in Fig. D.15 in Appendix C.

A comparison in Fig. 6 reveals that the residuals in HGT are smaller overall than those in GCN. Additionally, a continuous distribution of red and blue colours in the GCN map suggests that the model consistently overestimates or underestimates certain areas or roads. This suggests unobserved heterogeneity is hindering the GCN's performance, while HGT mitigate this problem, displaying a more uniform colour distribution. According to Fig. D.15, the same observations applied in other land use indicators, such as 'leisure' and 'retail'.

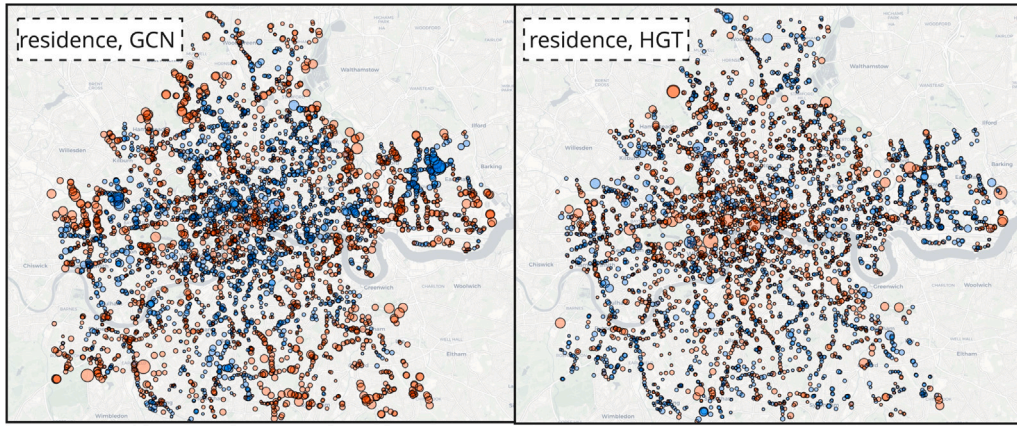


Fig. 6. Spatial distribution of residual of residence: GCN (left) and HGT (right).

5.7. Explainability analysis

As explained earlier, while existing feature attribution tools like Integrated Gradients [72] and DeepLIFT [71] can quantify the contributions of node features in graph models, directly applying them to heterogeneous graph models poses significant challenges. These methods are not specifically tailored to handle the node and edge types inherent in heterogeneous graphs. To address these limitations, we performed a comprehensive analysis using the XAI methods introduced in Section 4.2, specifically adapted for HGTs. This analysis includes three complementary approaches: ablation analysis, feature attribution analysis, and counterfactual analysis. Additionally, we utilised general rules of human activity as proxies to assess the alignment between the model's decision-making process and urban planners' understanding, thereby enhancing the model's credibility. The following sections explore each aspect of the explainability analysis in detail.

5.7.1. Ablation analysis on connectivity in the graph

Self-node contribution vs neighbour contribution Given the significant performance improvement of the proposed HGT model demonstrated in 7, an important question arises: how do the contributions of neighbour nodes compare to those of self-node information in enhancing model performance? To address this, we conducted a feature ablation analysis [88].

Specifically, we separately removed or masked the features of neighbour nodes and self-nodes, and then measured the impact on the model's predictive performance. By isolating and quantifying the specific contributions of each component, we gained deeper insights into how the model leverages neighbourhood information versus self-information to achieve its improved performance. The results are shown in Fig. 7.

Fig. 7 presents a histogram of the feature ablation results of neighbour nodes and self-nodes for predictions made by the HGT model. The results indicate that the features of neighbour nodes ("neighbour") have a significant impact on model predictions, with average importance values ranging from 0.08 to 0.30 across different indicators. In contrast, the self-node features ("Self") exhibit lower importance, with average values ranging between 0.04 and 0.24.

Specifically, the features of neighbour nodes contribute approximately 55.0% to 70.5% of the predictive power in the land use inference task. This finding underscores the critical role of neighbourhood structures in GNN-based models and highlights the necessity of incorporating neighbourhood feature information to enhance the accuracy of such models. Notably, integrating neighbour node features in this manner has not been widely explored in existing literature [88]. Our paper emphasises the importance of this aspect and suggests that future research should consider the significant contribution of neighbour nodes in graph-based learning tasks.

Contribution from different node types Another question lies in how to evaluate the role of different types of mobility nodes. Specifically, we constructed four scenarios by removing some nodes and connections: (a) all nodes and edges; (b) only bus and bike nodes with their edges; (c) only bus and tube nodes with their edges; (d) only bus nodes with their edges. Owing to the limited number of bike and tube nodes, these modes were not assessed as separate scenarios. We use R-square as the measurement for comparison as shown in Fig. 8.

According to Fig. 8, we can draw two conclusions. First, increased connectivity enhances performance across all metrics in all graph models. Specifically, moving from chart (a) to chart (d) reveals a noticeable decrease in performance metrics across domains like transport, retail, and residential areas. Second, the heterogeneous information acts as complementary for prediction, especially for areas under sparse connectivity. As noted previously, inferring land use for leisure is challenging because vehicular and cycling mobility data poorly represent non-commuting activities like leisurely travelling. However, we observed that heterogeneous graph models (HGT and RGCN) significantly improve the accuracy performance in the 'leisure' land use.

5.7.2. Feature attribution analysis

This section examines feature attributions—quantitative values representing the impact of the input features at each node (i.e., average passenger inbound data at different 15-minute intervals during the day) on the outcome using the Approximate Integrated Gradients index. We illustrate the distribution of this index across both the temporal (feature) dimension and the spatial dimension, shown as follows.

Temporal-level analysis The results are displayed in a heatmap format, where red indicates a positive correlation and blue a negative correlation between input features of different 15-minute intervals in a day and land use indicators, as shown in Fig. 9.

For Inner London, the feature attribution results in Fig. 9(a) closely align with the general understanding of human activities, illustrated in three categories with different characteristics. The first category, comprising **transport, office, and sustenance**, is primarily associated with work-related activities. This category shows a high sensitivity to commuter inflow during peak hours. Specifically, inflow is lower in the morning (10:00–11:00 AM) and significantly higher in the evening (5:00–6:00 PM). This indicates that areas with these land uses attract more commuters returning from work in the evening than departing in the morning. Besides, a subtler trend shows that Category 1 (work-related) stations experience reduced feature importance in the early evening (6:45–7:45 PM) and increased importance later at night (8:45–9:45 PM). This likely reflects a common urban lifestyle pattern, where individuals engage in leisure and entertainment activities after returning home from work. In contrast, the second category, consisting

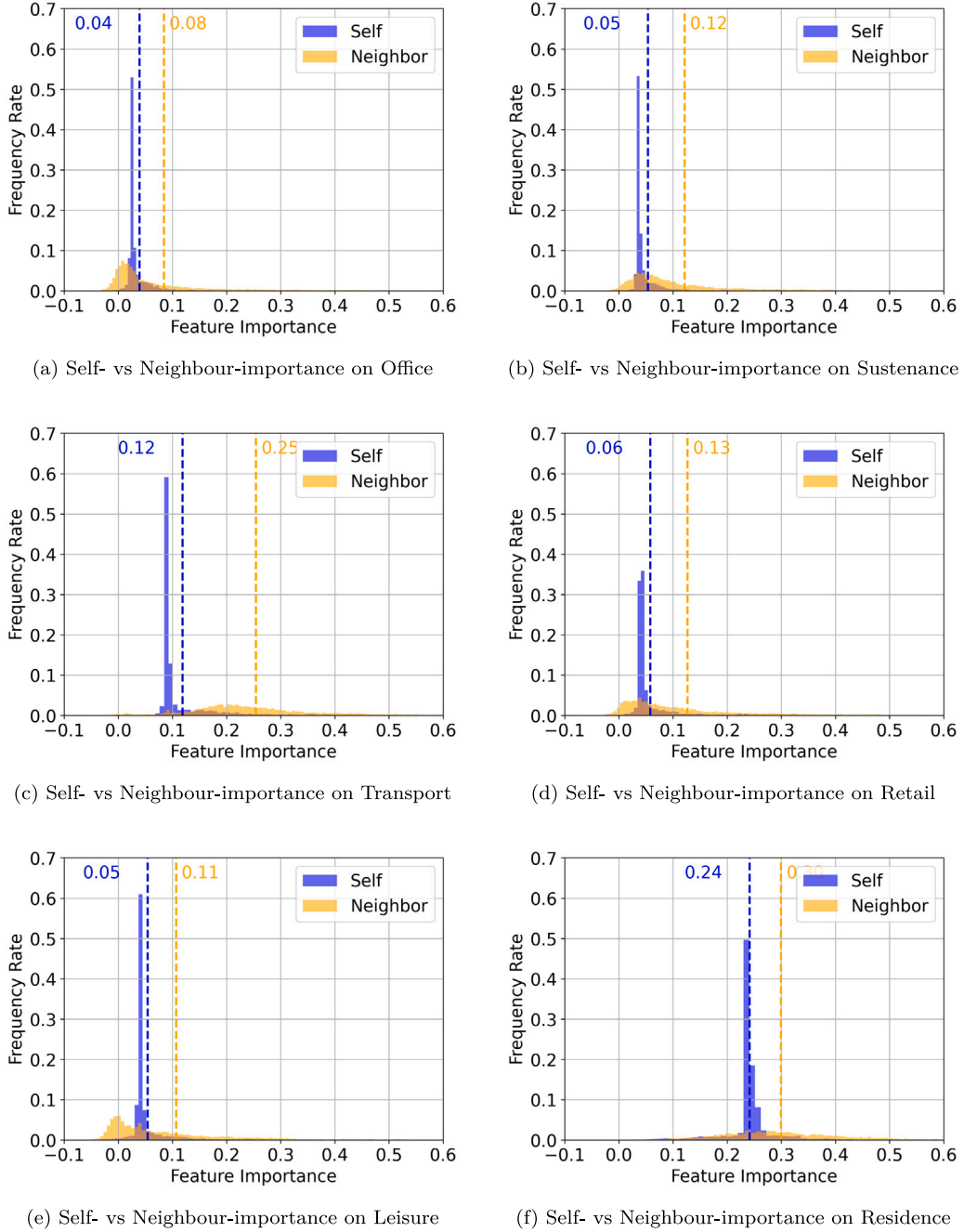


Fig. 7. Ablation feature importance on 6 land use types using HGT.

residential land use areas, exhibits a reverse trend with higher morning inflow and lower evening inflow. This reflects typical residential patterns where people leave for work in the morning and return in the evening. The third category, including **leisure and retail**, is less affected by inflows at peak hours. Notably, these areas show minimal variation in inflow during peak hours but have varied activity during other times, like late evenings, which aligns with shopping and leisure activities.

Fig. 9(b), which depicts the Outer London dataset, does not exhibit the same clear patterns observed in Inner London. The feature attribution values for various land uses in Outer London are more dispersed across different time intervals, lacking the pronounced peak-hour trends seen in Inner London. This likely reflects the diversity of activity patterns and the complexity of travel purposes in Outer

London areas. Specifically, areas associated with residential and work-related activities do not display the same morning and evening peak characteristics as observed in Inner London. This could be due to more dispersed commuting patterns in Outer London, where residents are influenced by a wider range of external factors, such as longer commuting distances and a greater reliance on private transportation.

It is important to note that various feature attribution methods might produce different outcomes. As shown in (Fig. 9(b)) and InputX-Gradient (Fig. D.14 in appendix), Integrated Gradients are generally considered more reliable because they account for changes over a broader range, offering a more comprehensive understanding of feature importance. Conversely, InputX-Gradient, focusing on a single point gradient, failed to align with the general context. As shown in this case, the importance of heatmap from the Integrated Gradient method aligns more closely with the characteristics of human activity in London,

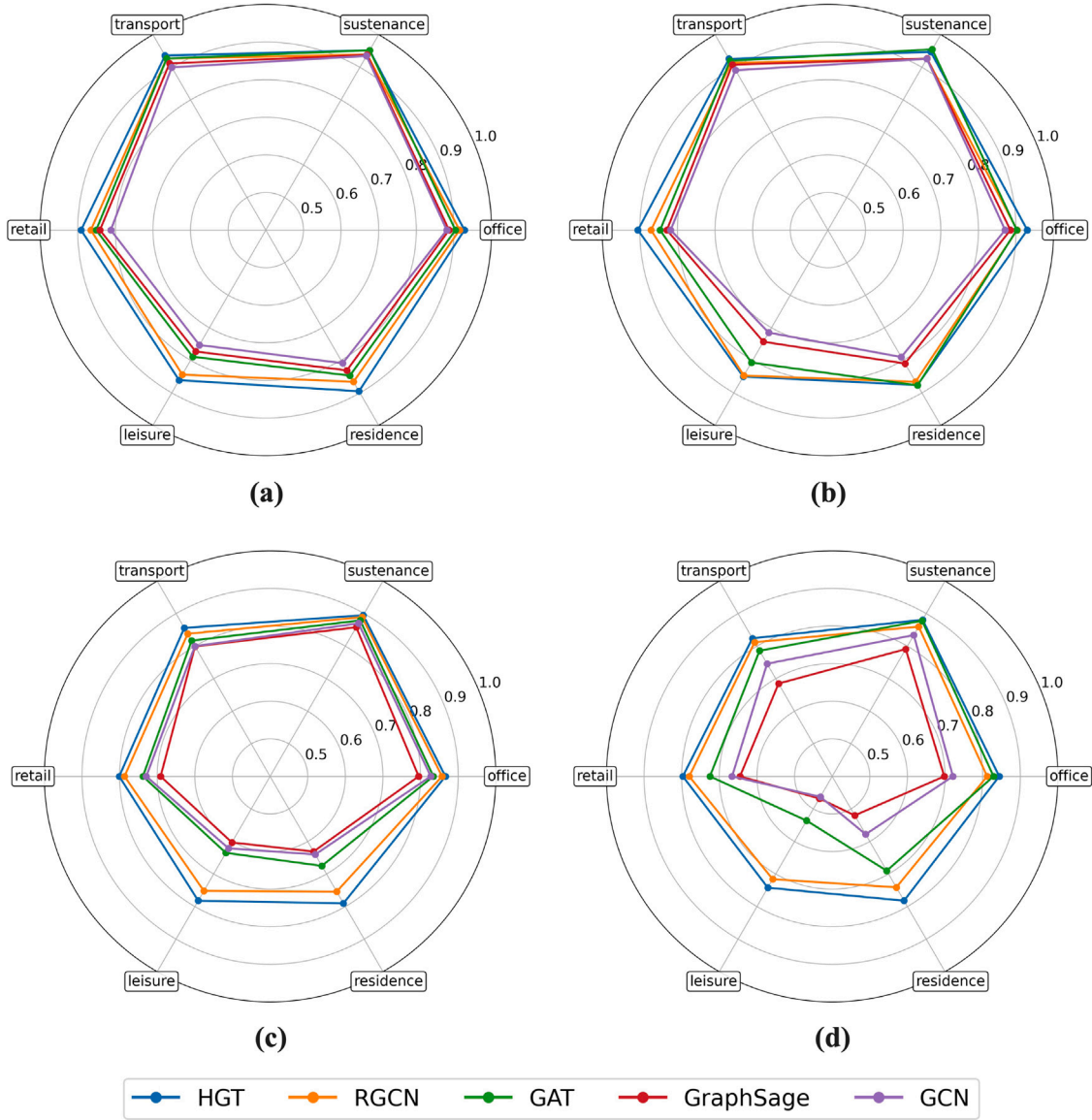


Fig. 8. R^2 of six land use indicators in four scenarios: (a) bus, bike, and tube; (b) bus and bike; (c) bus and tube; (d) bus.

reflecting our understanding of the comparative performance of the methods.

Spatial (node)-level analysis Variations in how mobility patterns influence the functioning of a local area are also a key concern for policy-makers. To address this, we aggregate the node-level feature attributions into subgraph-level attributions (using the mean of the contributions of all features of the nodes in the subgraph).

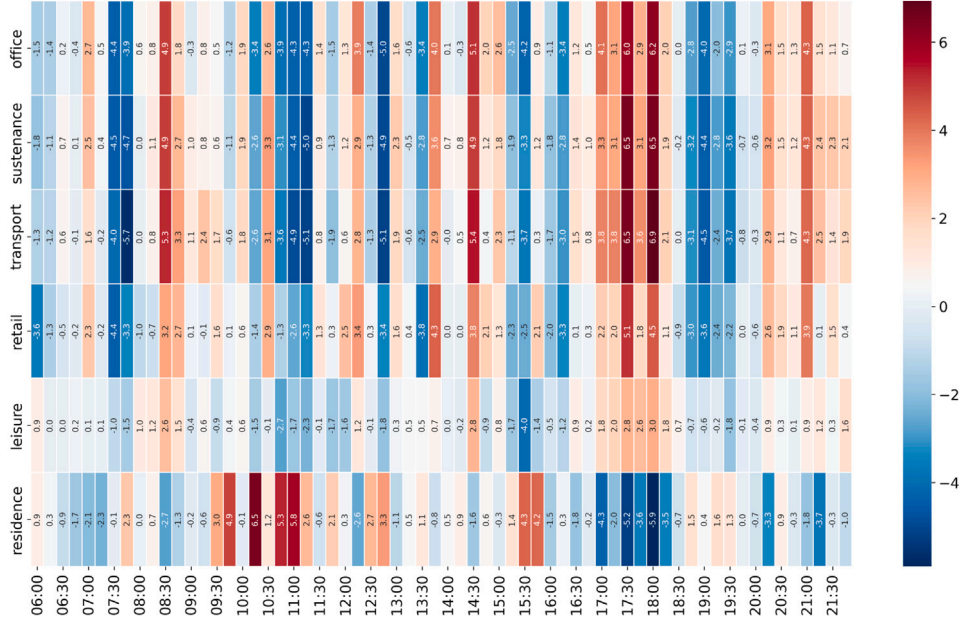
For better visualisation, we aggregate this combined attribution across urban administrative zones (wards of the Inner London), as shown in 10.

Given these two groups of heat maps shown in Fig. 10, we can draw several conclusions. Firstly, comparing peak with off-peak heat maps, it is evident that land use – particularly the land use associated with offices and restaurant services – is sensitive to peak-period ridership input compared with off-peak ridership. In contrast, during off-peak hours, the model appears relatively insensitive to land use factors. This finding aligns with well-established mobility patterns: during commuting periods, individuals are more likely to travel for work or use sustenance services.

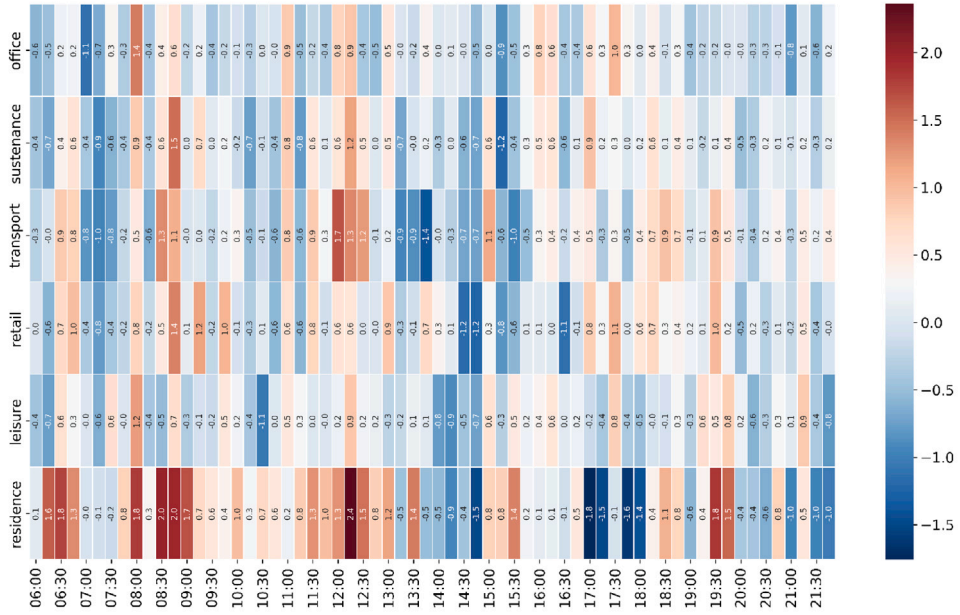
In terms of spatial distribution of regional importance, Office and Sustenance are notably sensitive in central areas, where employment

density and concentration of commercial activities tend to be highest. Conversely, these land uses are much less sensitive in peripheral or suburban zones. In contrast, transport, retail, and leisure demonstrate a more even, diffuse spatial pattern, indicating that the need for transport services, shopping areas, and leisure activities are not solely confined to central locations. Lastly, residence proves most significant for outer areas, which often serve as commuting origins; its impact diminishes in the city centre, where the daytime population is primarily composed of workers and visitors rather than residents. These observations are consistent with typical urban structures, where central areas host more employment and commercial mobility pattern, and residential neighbourhoods extend outward.

Notably, there were some anomalous areas. In our case, a few locations (e.g., Coleman Street in the central district, Weaver in the east, and Chelsea Riverside in the west) exhibit patterns distinct from neighbouring zones. These “pockets” may reflect specific local characteristics (e.g. exceptionally high employment density, major transport hubs, or specialised commercial corridors) that amplify their influence on peak ridership. Policymakers should be attentive to these areas when designing interventions, such as congestion management policies, fare adjustments, or land use regulations aimed at shaping travel behaviours.



(a) Integrated-gradients importance for Inner London dataset



(b) Integrated-gradients importance for Outer London dataset

Fig. 9. Comparison of the heatmap of the distribution of Integrated-gradients importance for Inner and Outer London datasets.

5.7.3. Counterfactual explanation analysis

In the previous section, we focused on the explainability of our model by analysing the feature attribution of node features. Although this analysis offered valuable insights into individual node attributes, it did not account for the importance of other critical factors such as node types, edge types, and the overall graph structure. To overcome this limitation, this section introduces counterfactual analysis for a more comprehensive understanding of our model. We set 'mixed' as an expected scenario. Mixed land use refers to an area with balanced facilities and diverse activities, crucial for enhancing urban vitality, reducing commute distances, and promoting sustainable development [82]. Mathematically, given the predicted land use intensity $y_{i,j}$ calculated from the proposed land use models, we applied the Shannon

diversity index to quantify the diversity of land uses within each area. The Shannon diversity (SD) is calculated as follows:

$$SD_i = - \sum_{j \in T} \frac{y_{i,j}}{\sum_{j \in T} y_{i,j}} \ln \left(\frac{y_{i,j}}{\sum_{j \in T} y_{i,j}} \right) \quad (8)$$

where $y_{i,j} / \sum_{j \in T} y_{i,j}$ represents the proportion of land use of type i in the area. This formula provides a diversity score for each area to reflect the variety of land uses present. After calculating the diversity scores for all areas, we ranked them in descending order. The areas with the highest diversity scores, representing the top 10% of all nodes, were classified as 'mixed' land use types. This approach ensures that we focus on regions that already demonstrate high land use diversity.

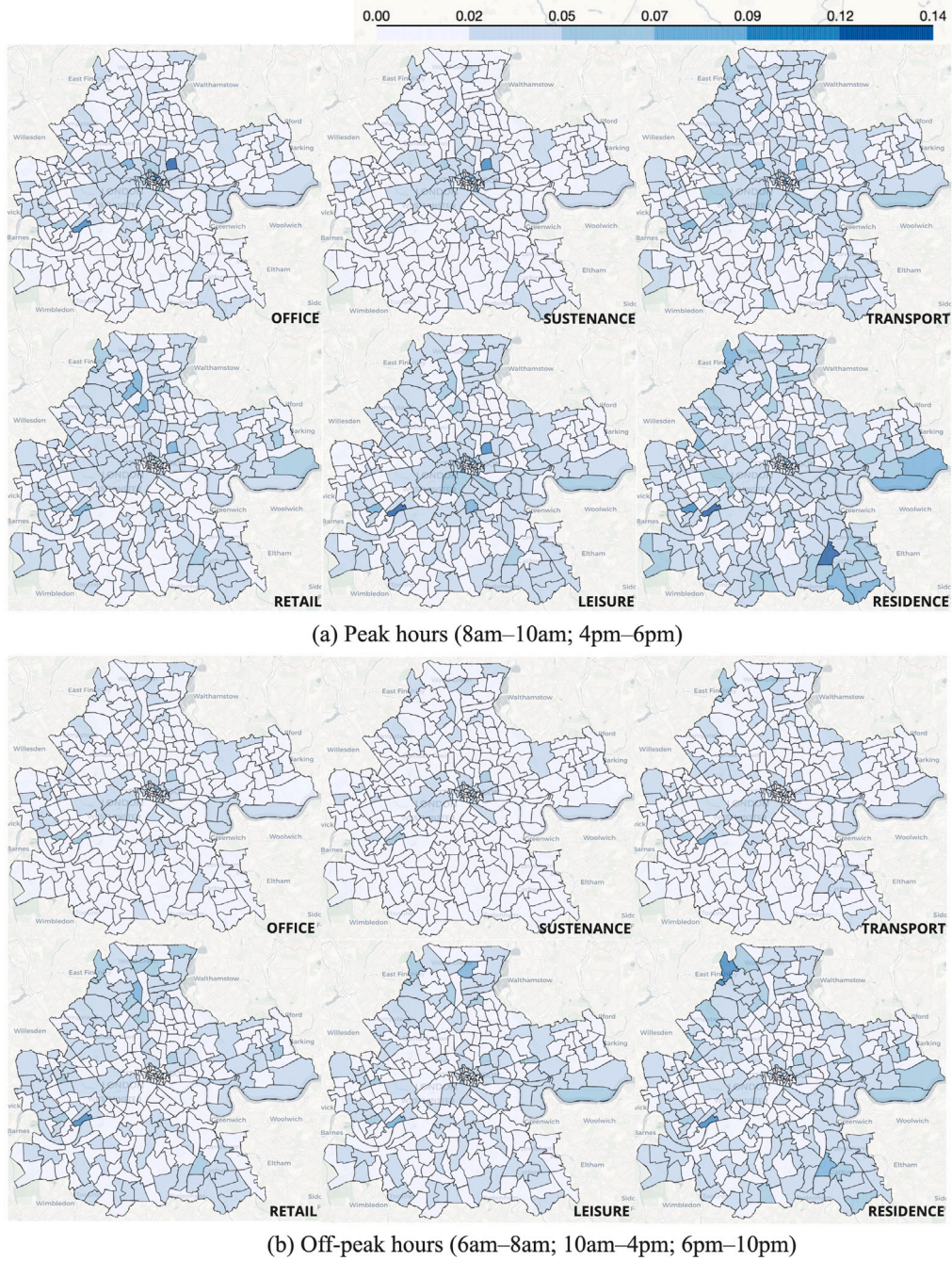


Fig. 10. Regional contribution (IG) of daily ridership in the inner London using model HGT.

Based on the Counterfactual Explanations with an alternative target class of mixed (CEs), this counterfactual analysis is conducted in two parts. Initially, we perform a local-level analysis to demonstrate the practical application of CEs. In this analysis, we select a specific input node and target a ‘mixed’ land use prediction as the desired outcome. We identify a CE for this node and compare it with the original input to underscore the necessary changes for achieving the target prediction. Secondly, we conduct an aggregated-level analysis to demonstrate how dissimilarity measures for CEs can illuminate the importance of factors such as node features, node types, edge types, and graph structure.

Local-level analysis We present CEs for two example input nodes which do not belong to the ‘mixed’ land use area; these are the inputs 1 and 2 which are located in Central and East London respectively (Fig. 11). We compute CEs for the two nodes which reveal the minimum

changes to make from the input (subgraph) in order to achieve the ‘mixed’ land use. The node feature differences between the input nodes and the CE subgraphs are illustrated in Fig. 12, and the node type, edge type, and graph structure differences are listed in Table 9. For input 1, as can be observed from the magnitudes in Fig. 12(a), the required changes are mostly in the node features, while the differences are small for the items in Table 9. Indeed, this intuition is demonstrated by the dissimilarity measures, which for the four data modalities are respectively 0.271, 0.056, 0.043, and 0.043. The high node feature dissimilarity emphasises the importance of changing node features for input 1. For input 2, however, the dissimilarities are 0.047, 0.0, 0.133, and 0.069, meaning that the change of edge (road) type is the primary factor in achieving the ‘mixed’ land use. The magnitudes of node feature changes are relatively small (Fig. 12(b)), while the edge type changes are more significant.

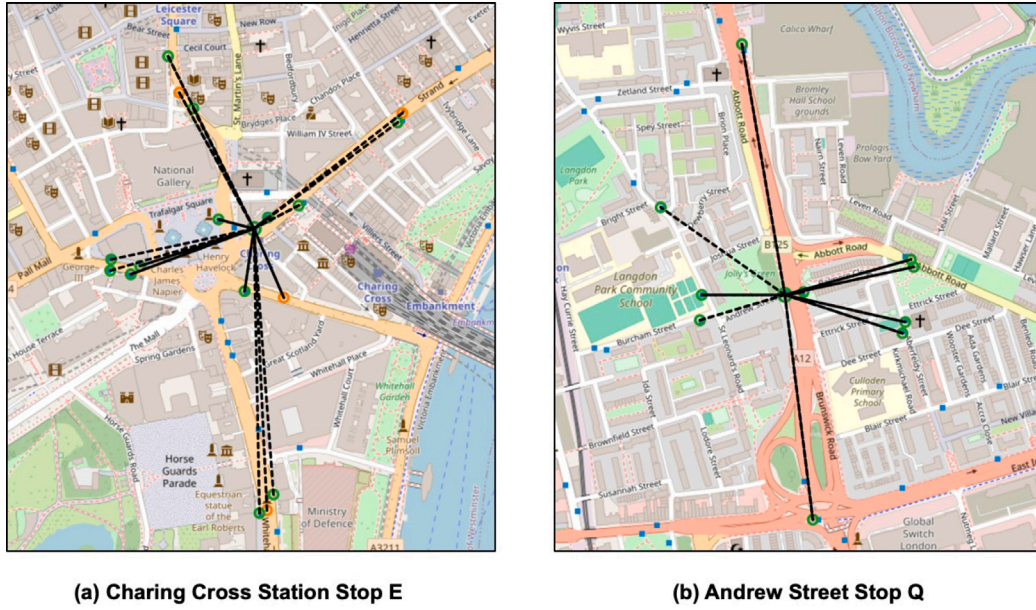


Fig. 11. Two local-level cases: (a) Input 1: a bus stop in the central London area with a high level of ‘transport’ prediction; (b) Input 2: a bus stop in the eastern area with a high level of ‘leisure’ prediction.

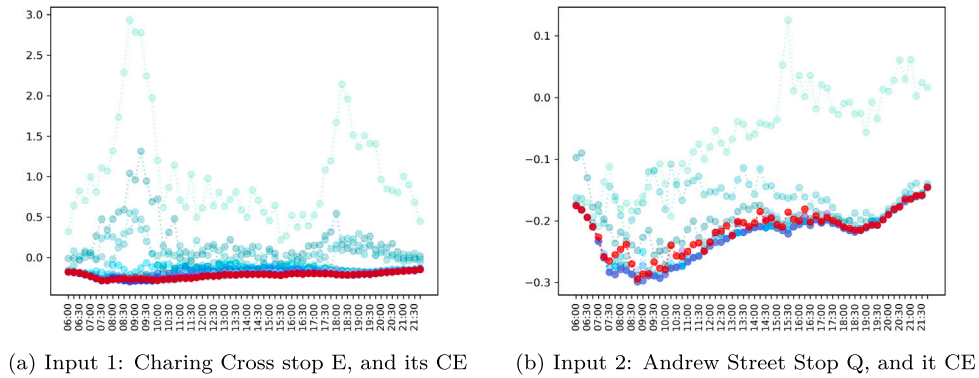


Fig. 12. Node feature differences. Each line (in different colours) is a vector of node features of a node, the line highlighted in red is for the input node, while the other lines in blue-like colours are for the nodes in the CE subgraph.

Table 9
Node type, edge type, graph structure differences between example inputs and their CE, both in the form of 1-hop complete subgraph.

Subgraph	Node type	Avg. Edge Type	Structure
Input 1	0 tube, 14 bus, 4 bike	0.52	141
CE 1	0 tube, 13 bus, 4 bike	0.70	135
Input 2	0 tube, 11 bus, 0 bike	1.68	72
CE 2	0 tube, 11 bus, 0 bike	1.15	67

Aggregated-level analysis Given a closed ‘mixed’ subgraph for each sample’s subgraph, we specify four types of differences: node feature, node type, edge type and graph structure. These differences can be regarded as relevant importance determining the minimal effort required to transition to a mixed land use type. All the values have been scaled into the 0–1 range, as shown in Table 10.

According to Table 10, In terms of the average values, node feature and node type are the primary sources of discrepancies, while edge type and graph structure have lower impact. Specifically, in leisure and residence tasks, the impact of node features is significantly reduced, which is aligned with the relatively low accuracy of these two tasks, as shown in Section 5.6.1. Conversely, these tasks slightly increase

Table 10
Counterfactual scores of different elements in a graph.

Indicators	Node feature	Node type	Edge type	Structure
office	0.637 ± 0.132	0.422 ± 0.139	0.030 ± 0.030	0.021 ± 0.023
sustenance	0.449 ± 0.212	0.457 ± 0.089	0.049 ± 0.045	0.039 ± 0.033
transport	0.287 ± 0.234	0.370 ± 0.121	0.038 ± 0.034	0.048 ± 0.052
retail	0.262 ± 0.302	0.399 ± 0.130	0.033 ± 0.021	0.042 ± 0.038
leisure	0.284 ± 0.260	0.269 ± 0.143	0.058 ± 0.044	0.048 ± 0.046
residence	0.216 ± 0.202	0.286 ± 0.141	0.045 ± 0.039	0.048 ± 0.052

the influence of edge types and graph structure. Additionally, the standard deviation indicates that node features contribute minimally to the difference for some sample regions, especially for indicators like ‘retail’, ‘leisure’ and ‘residence’. These patterns indicate that although node features are generally crucial, graph structural aspects and node relationships become more significant in leisure and residence contexts.

6. Conclusions

6.1. Summary

This paper builds an explainable graph-based framework for inferring land use from multi-modal mobility data in complex urban scenarios. In this framework, a Heterogeneous Graph Transformer model is introduced to fuse multi-modal mobility information for land use inference tasks. Next, a comprehensive explainable AI framework is built to enhance the models’ reliability for urban planners and provide insights for modellers.

The overall accuracy indicates that the use of HGT significantly improves accuracy. The analysis of model validation further confirms that the model generalises well not only on independent datasets from 2018 and 2019 but also across four different urban regions: Inner London, Outer London, San Francisco, and New York City, demonstrating generalisability in varied urban layout and mobility combination.

For modellers, the conclusions from the explainability analysis can directly guide the construction of an effective land use inference model based on multi-modal data fusion. Section 5.6.1 demonstrates that for a well-performed land inference model (HGT), the majority of the information (55.6% to 70.5%) is derived from neighbouring nodes and edges rather than from the node’s information. Further analysis in Section 5.6.1 reveals that the order of importance of the contributions from the different node types is: transit data, bicycle data, and subway data. Multiple comparisons between Inner and Outer London indicate that heterogeneous and homogeneous graph models perform more similarly in the densely connected areas of Inner London. However, in the Outer London region, where connectivity is sparser, the heterogeneous graph models (RGCN and HGT) are more effective.

Combined with our sensitivity analysis of the training sample size, we derive a rough value judgement curve: the smaller the training data size and the greater the connectivity, the more suitable it is to choose ordinary homogeneous graph models (GAT, GraphSage, and GCN). As the data volume increases and the connectivity becomes sparser, it becomes more suitable to choose heterogeneous graph models (RGCN, HGT).

For scenario users, such as planners or policymakers, relying solely on the external conclusions of a model that has not undergone rigorous user testing may not be the most prudent approach, even if the model achieves high accuracy. We cautiously propose that the current explainability analysis should serve as a usability check to reinforce model credibility. For example, as shown in the feature attribution analysis, the symmetrical nature of the ‘residence’ and ‘work’ categories aligns well with the commuter’s ‘work’; and ‘recreation’ activities in London, giving some confidence in the suitability of the proposed model for the land use setting. Additionally, subgraph-level (spatial) feature attribution analysis captures broader spatial patterns by aggregating

node-level contributions across urban zones. This reveals how land use importance varies between central areas (dominated by commercial and transport activities) and peripheral areas (where residential influences prevail), uncovering localised anomalies that guide more targeted policy interventions. These spatial insights further confirm the applicability of our approach in complex urban scenarios.

Meanwhile, counterfactual explanations highlight the efforts needed to alter any node’s prediction from a micro perspective, allowing experts to assess the GNNs’ explainability based on differences between subgraphs. The analysis of the counterfactual explanations in Section 5.7.3 reveals that in areas with different functions and historical backgrounds (e.g., Central London and East London), the different types of information in the heterogeneous graph data, which represent various elements in the sub-map, have significantly different forms of contribution. This is consistent with the planner’s understanding of these regions, thus validating the usefulness of fusing data from different sources. We posit that these insights will help to build user trust in GNNs by providing a clear, evidence-based rationale for each prediction, which is crucial for users in urban planning and management to trust the proposed methods.

6.2. Limitations

This proposed framework also has several key limitations that offer opportunities for future research. Firstly, the challenge of spatiotemporal matching of multivariate data presents significant constraints. Aligning diverse datasets across both space and time often leads to inconsistencies, especially when dealing with heterogeneous sources. This issue can impact the reliability of our findings and underscores the need for more comprehensive and well-matched data in future studies.

Secondly, the data requirements and model complexity also pose limitations. The use of heterogeneous graphs in our models introduces additional complexity, necessitating a larger volume of data to ensure accurate fitting. This increase in variables and parameters not only heightens data demands but also makes data processing more resource-intensive.

There are also some limitations to data problems. The nodes’ features are only represented by outflow rather than inflow and outflow, potentially limiting the comprehensiveness of our findings. Future research will benefit from the growing availability of open graph-structured urban data, enabling comparative studies to enhance our understanding of urban dynamics in different countries and societies. Cross-validation could potentially enhance this model’s performance; however, due to limited hardware computational capacity, only standard training methods were utilised in this paper. Further, incorporating multiple explainable AI techniques will increase the model’s transparency and strengthen support for urban planning decisions, offering a more trustworthy analysis of urban land use dynamics.

CRedit authorship contribution statement

Xuehao Zhai: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Junqi Jiang**: Writing – review & editing, Writing – original draft,

Visualization, Validation, Software, Methodology. **Adam Dejl:** Writing – review & editing, Visualization, Validation, Software, Methodology. **Antonio Rago:** Writing – review & editing, Supervision. **Fangce Guo:** Writing – review & editing, Supervision. **Francesca Toni:** Writing – review & editing, Supervision. **Aruna Sivakumar:** Writing – review & editing, Supervision.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used ChatGPT to improve readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Detailed workflow

A detailed workflow is provided including the entire data acquisition and processing pipeline (as shown in Fig. A.13 in Appendix A). For each step, readers can find the corresponding implementation Python program in our open-source [GitHub repository](#). For data acquisition, the workflow involves obtaining three main types of data: Urban Facilities POIs, vector maps of roadway networks, and multi-modal mobility demand. The first two types of data can be directly sourced from OSM's open-source GIS database, which is easily accessible using the **OSMNX** package [official documentation](#). Tutorials for collecting POI information and road networks can be found at this [tutorial repository](#). This ensures that, for any region covered by OSM, these data components are always available.

The mobility data, however, needs to be collected by the readers themselves, with two key constraints: (a) it is crucial to ensure that the multi-modal mobility data are collected within the same time period, ideally within the same year; and (b) regarding data quality, the data must be able to reflect daily patterns (ideally on both weekdays and weekends), with fine-grained time intervals (e.g., 15-minute intervals as used in this paper).

During data processing, we tested various normalisation methods and selected the optimal combination: z-score normalisation for inputs and 0–1 normalisation for labels. Since this aspect is not the primary focus of our paper, a more detailed discussion is omitted.

We also made compromises based on specific data conditions in London. For example, London's bus service records only boarding numbers, not alighting, and bike data for weekends is unavailable. Consequently, we used only weekday boarding flow as input to ensure fair comparisons between different transportation modes. Additionally, we found that residential POIs in London do not accurately reflect actual residential patterns due to privacy protection measures. We therefore replaced the residential POI data with population density data from the census. These adjustments may not apply to other cities' data and should be tested by readers as appropriate.

For the modelling part, we combine node features (mobility data), edge features (urban network data), and land use labels into a graph structure data format using the PyG package. This process is outlined in the `data_preparation.py` file within our open-source code. In the second step, we convert this data into a heterogeneous graph structure compatible with PyG to accommodate the HGT model. Finally, we construct the HGT model according to the methodology described and proceed with training, validation, and testing as shown in `main_hetero.py`.

Appendix B. Supplementary experiments

To better validate this framework, we conducted additional experiments in San Francisco, CA, USA and New York City, NY, USA.

San Francisco dataset The details of dataset in San Francisco are given below.

- **Mobility modes** The nodes input incorporate two mobility modes: vehicle parking and inbound and outbound volume of shared bikes, collected from 2020 Jan to 2020 April. These two types of data are first standardised into 15-mins level inbound and outbound volume at the node (station)-level.
- **Parking data** The original dataset of parking meter transaction records is provided by the [San Francisco Municipal Transportation Agency](#), including transaction ID, meter ID, position, session start time, and session end time. This dataset consists of the 1957 valid parking meters. Each row represents the hourly inbound and outbound parking records of a specific zone in a day.
- **bike sharing data** bike sharing users' trip from origin to destination parking stations was collected using the API from [Bay-wheels Company](#), a bike sharing system in San Francisco, including start time, end time start station ID and end station ID of each trip record. This dataset involves 353 valid shared bikes stations. The ridership is pre-processed into hourly inbound and outbound traffic volume.
- **Land use target, network and catchment rule** are in exactly the same way as in this manuscript. Specifically, the residence is measured by the population density. The specific POI distribution and road network are illustrated in Fig. 3(f) and 3(e), respectively. For the road network, the road level uses the local road type combination, including: motorway, trunk, primary, secondary, tertiary and unclassified.

New York City dataset We have also conducted a supplementary experiment in New York City, NY, USA, where the mobility inputs include two modes: (1) the metro system, and (2) a bike sharing system. All the data were collected over the year 2020. Following the same procedure for data preprocessing, the inbound and outbound flows were aggregated into an 15-min station-level format. The details of each dataset are outlined below.

- **Metro system:** The original user-level dataset is provided by the [NY State Government Open Data](#), which includes the following fields for each record: *Journey_id*, *Timestamp*, *Origin Station Complex ID*, *Origin Station Complex Name*, *Origin Latitude*, *Origin Longitude*, *Destination Station Complex ID*, *Destination Station Complex Name*, *Destination Latitude*, *Destination Longitude*, *Estimated Average Ridership*. We focus on 424 valid transit stations in New York City. As shown by the blue markers in Fig. 3(g), these stations are mostly located within the five boroughs of New York.
- **bike sharing system:** The open-source dataset is provided by Citibike, accessible at [Citi Bike Open data](#), containing *ride_id*, *rideable type*, *started at*, *ended at*, *start station name*, *start station id*, *end station name*, *end station id*, *start lat*, *start lng*, *end lat*, *end lng*. After cleaning and filtering, we identify a total of 4226 valid bike-sharing stations in New York City, marked in yellow in Fig. 3(g).
- **Graph structure and POI data:** As in the main manuscript, the road network (graph edges) and Points of Interest (POIs) for land use features are obtained from OpenStreetMap (OSM). We follow the same catchment-area approach to aggregate POIs into six land use categories. The data structures for edges and node attributes are consistent with those outlined in previous sections. Fig. 3(i) illustrates the POI distribution, while Fig. 3(h) shows the road network.

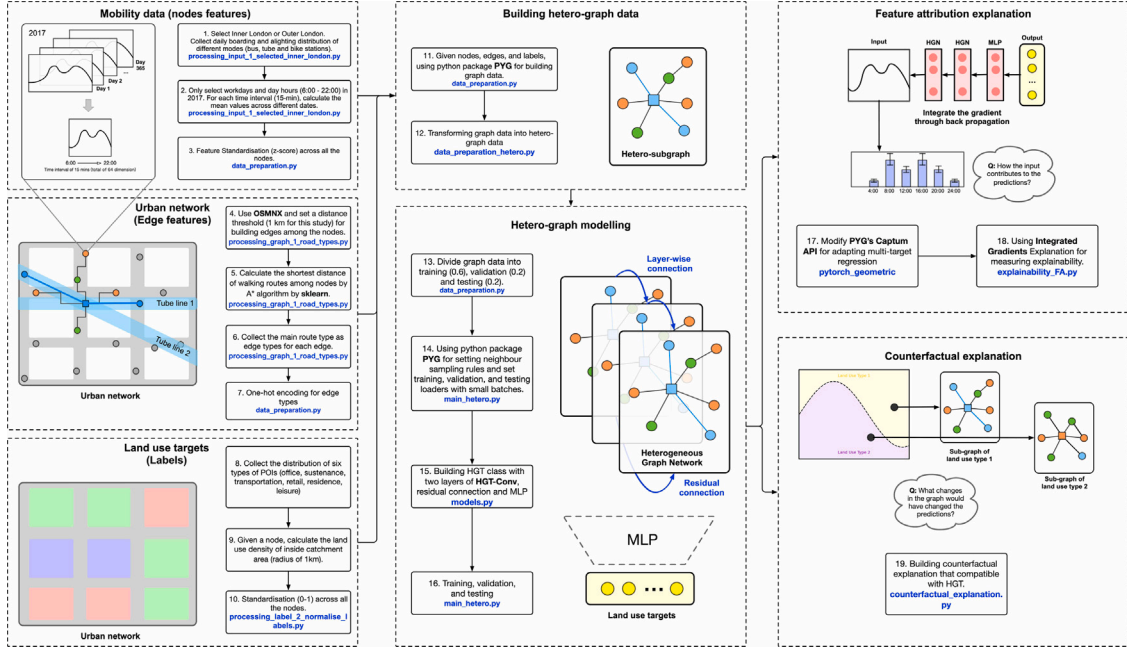


Fig. A.13. Detailed workflow of proposed study.

Table C.11

Average RMSE for Graph models with different hidden sizes.

Hidden_size	GraphSAGE	GCN	GAT	RGCN	HGT
32	0.096	0.099	0.109	0.087	0.073
64	0.085	0.089	0.078	0.076	0.061
128	0.074	0.080	0.073	0.065	0.056
256	0.066	0.074	0.080	0.064	0.095
512	0.066	0.069	0.124	0.075	0.726

Table C.12

Average RMSE for Graph models with different training optimiser (hidden-size 128).

Optimiser	GraphSAGE	GCN	GAT	RGCN	HGT
Adam	0.074	0.083	0.073	0.065	0.059
AdamW	0.075	0.082	0.073	0.063	0.056
RMSprop	0.077	0.082	0.075	0.065	0.058
SGD	0.146	0.140	0.139	0.151	0.162

Appendix C. Hyper-parameter calibration

We compared the results across different numbers of hidden layer neurons shown in Table C.11 for the Outer London dataset.

Based on the results, the HGT model achieves the best performance with the smallest error at a hidden size of 128, where the error is 0.056. Given this, HGT with a hidden size of 128 is selected as the optimal configuration for the task. As a result, we selected 128 as the number of hidden units.

Additionally, we compared the performance across four different learning optimisers: Adam, AdamW, RMSprop, and SGD, as shown in Table C.12.

The sensitivity analysis in Table C.12 shows that Adam, AdamW, and RMSprop yield similar performance for the HGT model, with only slight error-value differences. However, AdamW offers higher computational efficiency, making it the preferred choice for this task. Given

the usage of optimiser AdamW, we have done several training rounds and found that a learning rate of 0.002 provided a satisfactory balance between efficiency and accuracy, which is why it was chosen.

Appendix D. Supplementary results for explanation

See Figs. D.14 and D.15.

Data availability

I have shared the data source as well as the code repository in the experimental part of the article as well as in the appendix.

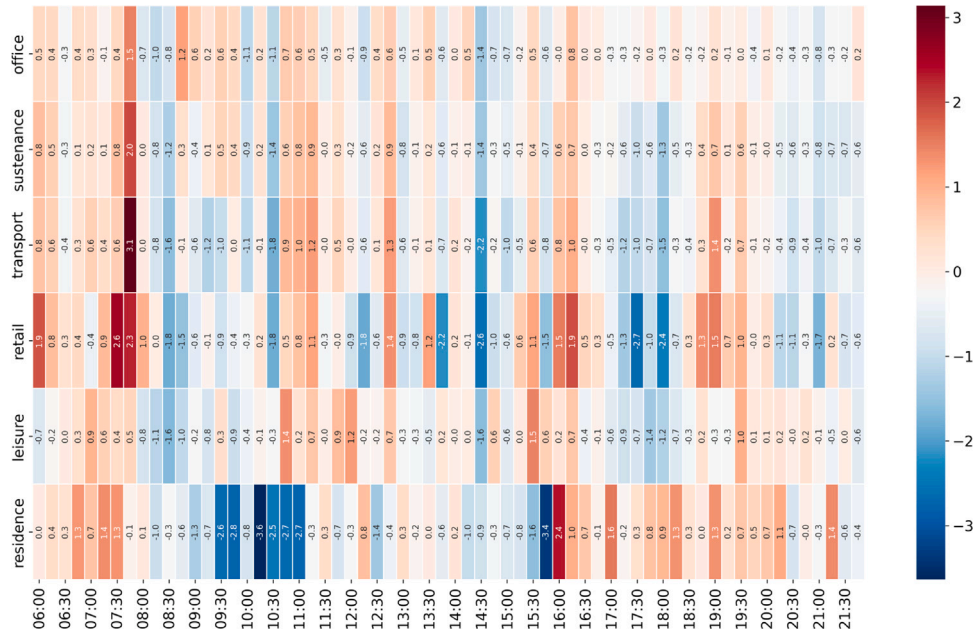


Fig. D.14. Heatmap of the distribution of InputX-gradient.

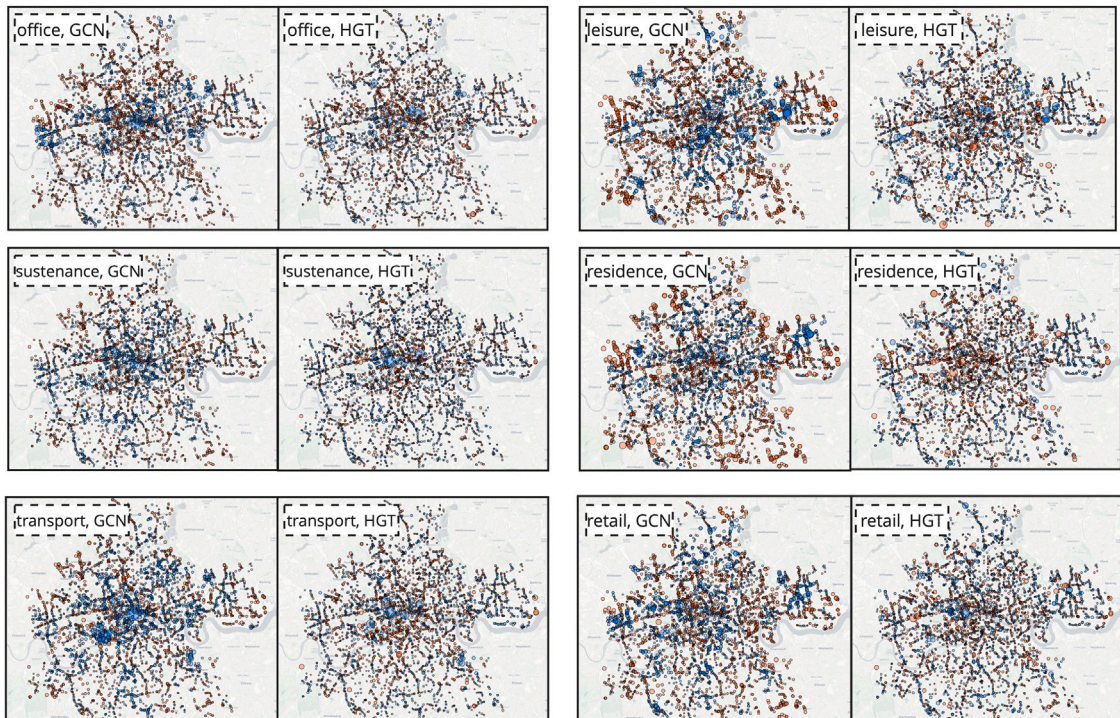


Fig. D.15. Special distribution of residual among six land use indicators.

References

- [1] J.L. Toole, M. Ulm, D. Bauer, M.C. Gonzalez, Inferring land use from mobile phone activity, in: Proceedings of the Acm Sigkdd International Workshop on Urban Computing, 2012.
- [2] J. Yuan, Y. Zheng, X. Xie, Discovering regions of different functions in a city using human mobility and POIs, in: Acm Sigkdd International Conference on Knowledge Discovery & Data Mining, 2012, p. 186.
- [3] J. Zhao, W. Fan, X. Zhai, Identification of land-use characteristics using bicycle sharing data: A deep learning approach, J. Transp. Geogr. 82 (2020).
- [4] S. Gao, K. Janowicz, H. Couclelis, Extracting urban functional regions from points of interest and human activities on location-based social networks, Trans. GIS 21 (3) (2017) 446–467.
- [5] X. Zhang, W. Li, F. Zhang, R. Liu, Z. Du, Identifying urban functional zones using public bicycle rental records and point-of-interest data, ISPRS Int. J. Geo-Inf. 7 (12) (2018) 459.
- [6] F. Yuan, K.E. Sawaya, B.C. Loeffelholz, M.E. Bauer, Land cover classification and change analysis of the twin cities (Minnesota) metropolitan area by multitemporal landsat remote sensing, Remote Sens. Environ. 98 (2–3) (2005) 317–328.

- [7] B. Zhou, K.M. Kockelman, Neighborhood impacts on land use change: A multinomial logit model of spatial relationships, *Ann. Reg. Sci.* 42 (2008) 321–340.
- [8] R.B. Matthews, N.G. Gilbert, A. Roach, J.G. Polhill, N.M. Gotts, Agent-based land-use models: a review of applications, *Landsc. Ecol.* 22 (2007) 1447–1459.
- [9] C.M. de Almeida, M. Batty, A.M.V. Monteiro, G. Câmara, B.S. Soares-Filho, G.C. Cerqueira, C.L. Pennachin, Stochastic cellular automata modeling of urban land use dynamics: empirical development and estimation, *Comput. Environ. Urban Syst.* 27 (5) (2003) 481–509.
- [10] M.A. Fadhel, A.M. Duhaim, A. Saihood, A. Sewify, M.N. Al-Hamadani, A. Albahri, L. Alzubaidi, A. Gupta, S. Mirjalili, Y. Gu, Comprehensive systematic review of information fusion methods in smart cities and urban environments, *Inf. Fusion* (2024) 102317.
- [11] V. Snášel, M. Štěpnička, V. Ojha, P.N. Suganthan, R. Gao, L. Kong, Large-scale data classification based on the integrated fusion of fuzzy learning and graph neural network, *Inf. Fusion* 102 (2024) 102067.
- [12] Y. Casali, N.Y. Aydin, T. Comes, Machine learning for spatial analyses in urban areas: a scoping review, *Sustain. Cities Soc.* 85 (2022) 104050.
- [13] J. Yuan, Y. Zheng, X. Xie, Discovering regions of different functions in a city using human mobility and POIs, in: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2012, pp. 186–194.
- [14] C.R. Bhat, S. Sen, N. Eluru, The impact of demographics, built environment attributes, vehicle characteristics, and gasoline prices on household vehicle holdings and use, *Transp. Res. Part B: Methodol.* 43 (1) (2009) 1–18.
- [15] R. Ewing, R. Cervero, Travel and the built environment: A meta-analysis, *J. Am. Plan. Assoc.* 76 (3) (2010) 265–294.
- [16] T. Yigitcanlar, N. Sipe, R. Evans, M. Pitot, A GIS-based land use and public transport accessibility indexing model, *Aust. Plan.* 44 (3) (2007) 30–37.
- [17] S. Hu, S. Gao, L. Wu, Y. Xu, Z. Zhang, H. Cui, X. Gong, Urban function classification at road segment level using taxi trajectory data: A graph convolutional neural network approach, *Comput. Environ. Urban Syst.* 87 (2021) 101619.
- [18] J. Choi, W. No, M. Park, Y. Kim, Inferring land use from spatiotemporal taxi ride data, *Appl. Geogr.* 142 (2022) 102688.
- [19] J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, S. Mullainathan, Human decisions and machine predictions, *Q. J. Econ.* 133 (1) (2018) 237–293.
- [20] H. Ritchie, M. Roser, Land use, 2013, Our World Data.
- [21] J. Yao, B. Zhang, C. Li, D. Hong, J. Chanussot, Extended vision transformer (ExViT) for land use and land cover classification: A multimodal deep learning framework, *IEEE Trans. Geosci. Remote Sens.* 61 (2023) 1–15.
- [22] Z. Sun, Z. Peng, Y. Yu, H. Jiao, Deep convolutional autoencoder for urban land use classification using mobile device data, *Int. J. Geogr. Inf. Sci.* 36 (11) (2022) 2138–2168.
- [23] Y. Yuan, M. Raubal, Y. Liu, Correlating mobile phone usage and travel behavior – A case study of Harbin, China, *Comput. Environ. & Urban Syst.* 36 (2) (2012) 118–130.
- [24] M. Yue, C. Kang, C. Andris, K. Qin, Y. Liu, Q. Meng, Understanding the interplay between bus, metro, and cab ridership dynamics in Shenzhen, China, *Trans. GIS* 22 (3) (2018) 855–871.
- [25] Z. Du, X. Zhang, W. Li, F. Zhang, R. Liu, A multi-modal transportation data-driven approach to identify urban functional zones: An exploration based on Hangzhou City, China, *Trans. GIS* 24 (1) (2020) 123–141.
- [26] D. Zhu, F. Zhang, S. Wang, Y. Wang, X. Cheng, Z. Huang, Y. Liu, Understanding place characteristics in geographic contexts through graph convolutional neural networks, *Ann. Am. Assoc. Geogr.* 110 (2) (2020) 408–420.
- [27] Q. Liu, W. Huan, M. Deng, X. Zheng, H. Yuan, Inferring urban land use from multi-source urban mobility data using latent multi-view subspace clustering, *ISPRS Int. J. Geo-Inf.* 10 (5) (2021) 274.
- [28] Z. Deng, X. You, Z. Shi, H. Gao, X. Hu, Z. Yu, L. Yuan, Identification of urban functional zones based on the spatial specificity of online car-hailing traffic cycle, *ISPRS Int. J. Geo-Inf.* 11 (8) (2022) 435.
- [29] H. Jiao, S. Huang, Y. Zhou, Understanding the land use function of station areas based on spatiotemporal similarity in rail transit ridership: A case study in Shanghai, China, *J. Transp. Geogr.* 109 (2023) 103568.
- [30] M.G. Demissie, L. Kattan, Understanding the temporal and spatial interactions between transit ridership and urban land-use patterns: an exploratory study, *Public Transp.* 14 (2) (2022) 385–417.
- [31] X. Zhang, X. Liu, K. Chen, F. Guan, M. Luo, H. Huang, Inferring building function: A novel geo-aware neural network supporting building-level function classification, *Sustain. Cities Soc.* 89 (2023) 104349.
- [32] T. Zhao, S. Wang, C. Ouyang, M. Chen, C. Liu, J. Zhang, L. Yu, F. Wang, Y. Xie, J. Li, et al., Artificial intelligence for geoscience: Progress, challenges and perspectives, *Innov.* (2024).
- [33] T. Qian, Y. Chen, G. Cong, Y. Xu, F. Wang, AdapTraj: a multi-source domain generalization framework for multi-agent trajectory prediction, in: *2024 IEEE 40th International Conference on Data Engineering, ICDE, IEEE, 2024*, pp. 5048–5060.
- [34] T. Papadakis, I.T. Christou, C. Ipektsidis, J. Soldatos, A. Amicone, Explainable and transparent artificial intelligence for public policymaking, *Data Policy* 6 (2024) e10.
- [35] C. Jin, T. Ruan, D. Wu, L. Xu, T. Dong, T. Chen, S. Wang, Y. Du, M. Wu, HetGAT: a heterogeneous graph attention network for freeway traffic speed prediction, *J. Ambient. Intell. Humaniz. Comput.* (2021) 1–12.
- [36] G. Li, Z. Zhao, X. Guo, L. Tang, H. Zhang, J. Wang, Towards integrated and fine-grained traffic forecasting: A spatio-temporal heterogeneous graph transformer approach, *Inf. Fusion* 102 (2024) 102063.
- [37] K.T. Chui, B.B. Gupta, R.W. Liu, P. Vasant, Handling data heterogeneity in electricity load disaggregation via optimized complete ensemble empirical mode decomposition and wavelet packet transform, *Sensors* 21 (9) (2021) 3133.
- [38] D. Wang, D. Zhong, A. Souri, Energy management solutions in the internet of things applications: Technical analysis and new research directions, *Cogn. Syst. Res.* 67 (2021) 33–49.
- [39] N. Schwalbe, B. Wahl, Artificial intelligence and the future of global health, *Lancet* 395 (10236) (2020) 1579–1586.
- [40] Y. Zheng, Z. Xu, X. Wang, The fusion of deep learning and fuzzy systems: A state-of-the-art survey, *IEEE Trans. Fuzzy Syst.* 30 (8) (2021) 2783–2799.
- [41] L. Du, R. Gao, P.N. Suganthan, D.Z. Wang, Bayesian optimization based dynamic ensemble for time series forecasting, *Inform. Sci.* 591 (2022) 155–175.
- [42] Y. Wang, Y. Liu, Bayesian entropy network for fusion of different types of information, *Reliab. Eng. Syst. Saf.* 195 (2020) 106747.
- [43] M.A. Ganaie, M. Hu, A.K. Malik, M. Tanveer, P.N. Suganthan, Ensemble deep learning: A review, *Eng. Appl. Artif. Intell.* 115 (2022) 105151.
- [44] Y. Ren, L. Zhang, P.N. Suganthan, Ensemble classification and regression-recent developments, applications and future directions, *IEEE Comput. Intell. Mag.* 11 (1) (2016) 41–53.
- [45] Y. Liang, Z. Zhao, F. Ding, Y. Tang, Z. He, Time-dependent trip generation for bike sharing planning: A multi-task memory-augmented graph neural network, *Inf. Fusion* (2024) 102294.
- [46] F. Zhou, T. Wang, T. Zhong, G. Trajcevski, Identifying user geolocation with hierarchical graph neural networks and explainable fusion, *Inf. Fusion* 81 (2022) 1–13.
- [47] W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [48] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, 2017, arXiv preprint arXiv:1710.10903.
- [49] C. Shi, B. Hu, W.X. Zhao, S.Y. Philip, Heterogeneous information network embedding for recommendation, *IEEE Trans. Knowl. Data Eng.* 31 (2) (2018) 357–370.
- [50] Y. Wu, Y. Chen, Z. Yin, W. Ding, I. King, A survey on graph embedding techniques for biomedical data: Methods and applications, *Inf. Fusion* 100 (2023) 101909.
- [51] X. Wang, D. Bo, C. Shi, S. Fan, Y. Ye, S.Y. Philip, A survey on heterogeneous graph embedding: methods, techniques, applications and sources, *IEEE Trans. Big Data* 9 (2) (2022) 415–436.
- [52] S. Yang, B. Yang, An inductive heterogeneous graph attention-based multi-agent deep graph infomax algorithm for adaptive traffic signal control, *Inf. Fusion* 88 (2022) 249–262.
- [53] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part I* 13, Springer, 2014, pp. 818–833.
- [54] J. Li, W. Monroe, D. Jurafsky, Understanding neural networks through representation erasure, 2016, arXiv preprint arXiv:1612.08220.
- [55] R. Meyers, M. Lu, C.W. de Puiseau, T. Meisen, Ablation studies in artificial neural networks, 2019, arXiv preprint arXiv:1901.08644.
- [56] Ł. Maziarka, T. Danel, S. Mucha, K. Rataj, J. Tabor, S. Jastrzębski, Molecule attention transformer, 2020, arXiv preprint arXiv:2002.08264.
- [57] S. Yin, Q. Peng, H. Li, Z. Zhang, X. You, H. Liu, K. Fischer, S.L. Furth, G.E. Tasian, Y. Fan, Multi-instance deep learning with graph convolutional neural networks for diagnosis of kidney diseases using ultrasound imaging, in: *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging and Clinical Image-Based Procedures: First International Workshop, UNSURE 2019, and 8th International Workshop, CLIP 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Proceedings* 8, Springer, 2019, pp. 146–154.
- [58] Z. Ying, D. Bourgeois, J. You, M. Zitnik, J. Leskovec, GNNExplainer: Generating explanations for graph neural networks, in: H.M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E.B. Fox, R. Garnett (Eds.), *Advances in Neural Information Processing Systems* 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8–14, 2019, Vancouver, BC, Canada, 2019, pp. 9240–9251, URL <https://proceedings.neurips.cc/paper/2019/hash/d80b7040b773199015de6d3b4293c8ff-Abstract.html>.
- [59] E. Rossi, B. Chamberlain, F. Frasca, D. Eynard, F. Monti, M. Bronstein, Temporal graph networks for deep learning on dynamic graphs, 2020, arXiv preprint arXiv:2006.10637.
- [60] M. Sundararajan, A. Taly, Q. Yan, Axiomatic attribution for deep networks, in: *International Conference on Machine Learning, PMLR*, 2017, pp. 3319–3328.
- [61] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [62] Z. Yan, W.L. Hamilton, M. Blanchette, Graph neural representational learning of RNA secondary structures for predicting RNA-protein interactions, *Bioinformatics* 36 (Supplement_1) (2020) i276–i284.

- [63] Y. Bi, D. Xiang, Z. Ge, F. Li, C. Jia, J. Song, An interpretable prediction model for identifying N7-methylguanosine sites based on XGBoost and SHAP, *Mol. Therapy-Nucleic Acids* 22 (2020) 362–372.
- [64] M. Dobrojevic, M. Zivkovic, A. Chhabra, N.S. Sani, N. Bacanin, M.M. Amin, Addressing internet of things security by enhanced sine cosine metaheuristics tuned hybrid machine learning model and results interpretation based on SHAP approach, *PeerJ Comput. Sci.* 9 (2023) e1405.
- [65] S. Wachter, B.D. Mittelstadt, C. Russell, Counterfactual explanations without opening the black box: Automated decisions and the GDPR, *Harv. J.L. Tech.* 31 (2017) 841.
- [66] D. Numeroso, D. Bacciu, Meg: Generating molecular counterfactual explanations for deep graph networks, in: 2021 International Joint Conference on Neural Networks, IJCNN, IEEE, 2021, pp. 1–8.
- [67] Z. Chen, F. Silvestri, J. Wang, Y. Zhang, Z. Huang, H. Ahn, G. Tolomei, Grease: Generate factual and counterfactual explanations for gnn-based recommendations, 2022, arXiv preprint arXiv:2208.04222.
- [68] X. Jiang, Y. Shen, Y. Wang, H. Shen, C. Xu, S. Ma, Meta-path based social relation reasoning in a deep and robust way, in: International Conference on Database Systems for Advanced Applications, Springer, 2023, pp. 3–20.
- [69] P. Pezeshkpour, S. Jain, S. Singh, B.C. Wallace, Combining feature and instance attribution to detect artifacts, 2021, CoRR abs/2107.00323, arXiv:2107.00323.
- [70] C. Meng, L. Trinh, N. Xu, J. Enouen, Y. Liu, Interpretability and fairness evaluation of deep learning models on MIMIC-IV dataset, *Sci. Rep.* 12 (2022).
- [71] A. Shrikumar, P. Greenside, A. Shcherbina, A. Kundaje, Not just a black box: Learning important features through propagating activation differences, 2016, arXiv preprint arXiv:1605.01713.
- [72] M. Sundararajan, A. Taly, Q. Yan, Axiomatic attribution for deep networks, in: D. Precup, Y.W. Teh (Eds.), Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6–11 August 2017, in: Proceedings of Machine Learning Research, vol. 70, PMLR, 2017, pp. 3319–3328, URL <http://proceedings.mlr.press/v70/sundararajan17a.html>.
- [73] M. Ancona, E. Ceolini, C. Öztireli, M. Gross, Towards better understanding of gradient-based attribution methods for deep neural networks, in: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings, OpenReview.net, 2018, URL <https://openreview.net/forum?id=Sy21R9JAW>.
- [74] D. Balduzzi, M. Frean, L. Leary, J. Lewis, K.W.-D. Ma, B. McWilliams, The shattered gradients problem: If resnets are the answer, then what is the question? in: International Conference on Machine Learning, PMLR, 2017, pp. 342–350.
- [75] M.A. Prado-Romero, B. Prenkaj, G. Stilo, F. Giannotti, A survey on graph counterfactual explanations: Definitions, methods, evaluation, and research challenges, *ACM Comput. Surv.* 56 (7) (2024) 171:1–171:37, <http://dx.doi.org/10.1145/3618105>.
- [76] Y. Casali, N.Y. Aydin, T. Comes, Machine learning for spatial analyses in urban areas: a scoping review, *Sustain. Cities Soc.* 85 (2022) 104050.
- [77] L.d.F. Costa, Further generalizations of the jaccard index, 2021, arXiv preprint arXiv:2110.09619.
- [78] D. Brughmans, P. Leyman, D. Martens, Nice: an algorithm for nearest instance counterfactual explanations, *Data Min. Knowl. Discov.* (2023) 1–39.
- [79] R. Poyiadzi, K. Sokol, R. Santos-Rodríguez, T.D. Bie, P.A. Flach, FACE: feasible and actionable counterfactual explanations, in: A.N. Markham, J. Powles, T. Walsh, A.L. Washington (Eds.), AIES '20: AAAI/ACM Conference on AI, Ethics, and Society, New York, NY, USA, February 7–8, 2020, ACM, 2020, pp. 344–350, <http://dx.doi.org/10.1145/3375627.3375850>.
- [80] J. Jiang, F. Leofante, A. Rago, F. Toni, Robust counterfactual explanations in machine learning: A survey, in: Proceedings of the 33rd International Joint Conference on Artificial Intelligence, IJCAI, 2024.
- [81] D. Farr, Sustainable Urbanism: Urban Design with Nature, John Wiley & Sons, 2011.
- [82] S. Lehmann, Sustainable urbanism: towards a framework for quality and optimal density? *Futur. Cities Environ.* 2 (2016) 1–13.
- [83] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, 2016, arXiv preprint arXiv:1609.02907.
- [84] S. Guo, Y. Lin, N. Feng, C. Song, H. Wan, Attention based spatial-temporal graph convolutional networks for traffic flow forecasting, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, (no. 01) 2019, pp. 922–929.
- [85] M. Schlichtkrull, T.N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, M. Welling, Modeling relational data with graph convolutional networks, in: The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15, Springer, 2018, pp. 593–607.
- [86] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, 2017, arXiv preprint arXiv:1711.05101.
- [87] Z. Hu, Y. Dong, K. Wang, Y. Sun, Heterogeneous graph transformer, in: Proceedings of the Web Conference 2020, 2020, pp. 2704–2710.
- [88] H. Yuan, H. Yu, S. Gui, S. Ji, Explainability in graph neural networks: A taxonomic survey, *IEEE Trans. Pattern Anal. Mach. Intell.* 45 (5) (2022) 5782–5799.