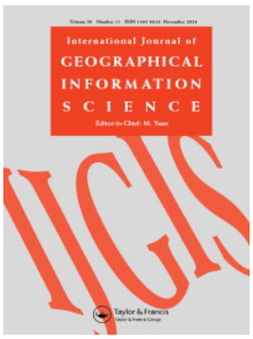


International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tgis20

Learning spatial interaction representation with heterogeneous graph convolutional networks for urban land-use inference

Zhaoya Gong, Chenglong Wang, Yuting Chen, Bin Liu, Pengjun Zhao & Zhengzi Zhou

To cite this article: Zhaoya Gong, Chenglong Wang, Yuting Chen, Bin Liu, Pengjun Zhao & Zhengzi Zhou (2024) Learning spatial interaction representation with heterogeneous graph convolutional networks for urban land-use inference, International Journal of Geographical Information Science, 38:11, 2235-2271, DOI: 10.1080/13658816.2024.2379473

To link to this article: https://doi.org/10.1080/13658816.2024.2379473

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
2024, VOL. 38, NO. 11, 2235–2271
https://doi.org/10.1080/13658816.2024.2379473

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Learning spatial interaction representation with heterogeneous graph convolutional networks for urban land-use inference

Zhaoya Gong[a,b]*, Chenglong Wang[a,b]*, Yuting Chen[a,b]*, Bin Liu[a,b], Pengjun Zhao[a,b] and Zhengzi Zhou[a,b]

[a]School of Urban Planning and Design, Peking University Shenzhen Graduate School, Shenzhen, China; [b]Key Laboratory of Earth Surface System and Human-Earth Relations of Ministry of Natural Resources of China, Peking University Shenzhen Graduate School, Shenzhen, China

ABSTRACT

Urban land use is central to urban planning. With the emergence of urban big data and advances in deep learning methods, several studies have leveraged graph convolutional networks (GCNs) with local functional characteristics from points of interest data and spatial features from flow data to infer urban land use. However, these studies cannot distinguish spatial interaction and spatial dependence in terms of conceptualization and modeling mechanisms and overlook the inadequacy of GCNs in modeling spatial interaction. This study proposes a novel framework—a heterogeneous graph convolutional network (HGCN)—to explicitly account for the spatial demand and supply components embedded in spatial interaction data. Several experiments, including 19 different models and datasets from Shenzhen and London, were conducted to validate the proposed framework and its generalizability within the same and different spatial contexts. The HGCN can distinguish heterogeneous mechanisms in supply- and demand-related modalities of spatial interactions, incorporating both spatial interaction and spatial dependence for urban land-use inference. Empowered by HGCN, we found that spatial interaction features play a distinctively crucial role in urban land-use inference compared to local attributes and spatial dependence features. In addition, our findings highlight the superiority of HGCN-based models in boosting performance and enhancing model transferability.

ARTICLE HISTORY
Received 17 November 2022
Accepted 8 July 2024

KEYWORDS
Heterogeneous graph convolutional networks; spatial interaction; urban land-use modeling

1. Introduction

Urban land use is a central theme in urban geography, planning, and transportation. Identifying the functional uses of urban lands is crucial for understanding urban spaces and evaluating planning strategies (Zhang *et al*. 2018, Chen *et al*. 2021).

CONTACT Zhaoya Gong z.gong@pku.edu.cn; Pengjun Zhao pengjun.zhao@pku.edu.cn
*These authors contribute equally to this paper.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/13658816.2024.2379473.

Conventionally, understanding the functional uses of urban lands has relied heavily on survey data, which are costly to collect and difficult to update promptly, rendering it challenging to gain an in-depth understanding of the dynamics and vibrancy of urban functions (Cheng *et al*. 2006). Remote sensing data have been widely used in land-use and land-cover change studies (Seto and Fragkias 2005, Ma *et al*. 2019). This approach functions relatively effectively in detecting land cover, defined by the physical features of the Earth's land surface and immediate subsurface, including the biota, soil, topography, water bodies, and built-up areas; however, its performance is highly limited in distinguishing between different functional land uses (commercial, industrial, public facilities, residential, etc.) within urban areas (Lambin and Geist 2008).

Emerging sources of urban big data, such as social media data, points of interest (POI) data, and mobile phone data, help elucidate the use of urban spaces at a fine-grained spatiotemporal resolution (Goodchild 2007, Crooks *et al*. 2015). A large body of literature utilizes the local characteristics of urban places, generally represented by POI data, to infer urban land use based on the assumption that the function of a place depends on the human activity that occurs there (Jiang *et al*. 2015, Gao *et al*. 2017, Yan *et al*. 2017, Zhai *et al*. 2019, Niu and Silva 2021). However, a city is not a collection of isolated areas but a network of connected functional places where people travel. From a supply-demand perspective, when the supply of certain functions in one place is insufficient, individuals travel to other places to fulfill their demands. This functional complementarity results in spatial interactions (Wang 2016). Therefore, the function of one place not only depends on its local characteristics and similar functions of nearby places but also interacts with the functions of other places with different local characteristics. In this study, we define spatial interaction as social and economic linkages driven by the spatially distributed supply and demand of land-use functions between places, commonly realized by commuting, shopping, school trips, etc. In contrast, spatial dependence, as defined in this study, only reflects spatial (auto)correlation; spatially proximate units tend to have more similar attributes (Anselin 2013).

Utilizing the rapid advancements in deep learning, several recent studies have combined different sources of urban big data, such as urban imagery data and human movements from mobile phones or GPS data, and have achieved improvements in analyzing urban functions in various scenarios (Liu *et al*. 2016, Zhu *et al*. 2020, Deng *et al*. 2022, Xu *et al*. 2022). One of the most widely used deep neural network models, graph convolutional networks (GCNs), has been employed in numerous studies to manage movement (or flow)-based spatial interaction data. However, these studies have two limitations. In general, the distinction between spatial interaction and spatial dependence has not been explicitly addressed in either conceptualization or modeling. In particular, the inability of GCNs to sufficiently model the demand and supply components of spatial interaction has not been recognized despite its adequacy in considering the spatial dependence of features.

To address these issues, this study proposes a novel framework called heterogeneous graph convolutional networks (HGCN) to explicitly account for the spatial demand and supply components embedded in spatial interaction data, in addition to considering local attributes and spatial dependence features, for urban land-use inference. This framework can distinguish between the underlying mechanisms of spatial

interaction and spatial dependence features and has the capacity to incorporate both mechanisms for urban land-use inference. By learning a heterogeneous representation, the proposed HGCN captures (1) the multimodal nature of places (nodes) as demanders or suppliers of urban functions and (2) the asymmetry of spatial interactions (links) between places from either a flow generation or flow attraction perspective. This is achieved by innovatively devising a convolution operator based on spatial accessibility. A series of carefully designed experiments (including 19 different models) were conducted on datasets from two cities, Shenzhen and London, to validate the proposed framework.

The remainder of this paper is organized as follows: In Section 2, we review the relevant literature and identify the drawbacks of the existing research. In Section 3, the HGCN framework is proposed with methodological details. In Section 4, we describe the experiments and report the results of the proposed models and baselines. In Section 5, we discuss these findings. In Section 6, conclusions are drawn, and future work is presented.

2. Literature review

2.1. Theoretical foundations of urban land-use functions

Urban land use is urban planning-oriented and acts as a central vehicle for socioeconomic activities occurring on the land surface, such as residence and employment (Fresco et al. 1994). Urban land-use systems, as one of the most striking forms of human-nature interactions, are shaped by the demands of human activities, the physical characteristics of land resources that provide specific functions, and the spatial interactions between them (Liu *et al*. 2016). As Ullman (1954) argues, complementarity is a precondition for spatial interaction. Classical urban theories have revealed the spatial interactions induced by a supply-demand mechanism for certain functions, such as agricultural production (Von Thünen 2018), transportation functions (Hansen 1959), and occupational and residential functions (Alonso 1960).

Three key aspects are involved in this mechanism. One is 'demand,' referring to the demand of places for human activities consuming goods or services, which further leads to the location choice of places whose land uses supply the urban functions demanded by human activities. The second is 'supply,' which refers to the products or services supplied by lands that can fulfill the demands of certain human activities, such as commerce or residence. The third is the linkage between the spatially distributed supplies of urban functions and the demands of human activities, represented by certain types of interactions between places, such as the movement of individuals, goods, or information. The spatial interaction theory, which accounts for the supply and demand of urban functions, has been incorporated into urban land-use modeling to highlight the spatial processes involved and enhance the explainability of models from an explicit spatial perspective. This is achieved through developing integrated land-use and transport models (Hunt *et al*. 2005, Zhong *et al*. 2022) and disaggregated methods empowered by the complexity theory of urban systems. Examples include agent-based urban simulation models (Parker *et al*. 2003).

2.2. Urban function inference based on local attributes

POI data are generally compiled using online digital maps to represent urban places with functional properties (e.g. residential, commercial, and public facilities) as geographic points for mapping or navigation purposes. Several studies have used POI data to extract urban features (Jiang *et al*. 2015, Gao *et al*. 2017, Yan *et al*. 2017, Zhai *et al*. 2019, Niu and Silva 2021). However, despite their richness in reflecting urban functional characteristics, the majority of POI data suffer from representativeness issues. For example, service-related POIs generally have better representativeness than industry-related POIs because the former are requested more frequently by users. Recently, new methods have been used to elucidate the internal structure of cities by developing quantitative methods that utilize POI data. These predominantly fall into two categories: First, the density-based approach applies natural language processing methods (such as latent Dirichlet allocation, probabilistic latent semantic analysis, and term frequency-inverse document frequency) to extract the density distribution of urban functions (Yuan *et al*. 2012). It regards regions as documents, POIs as words, and urban functions as topics. However, the density-based approach focuses only on the spatial distribution of different urban functions and disregards the spatial relationships between them.

Second, semantic-based word embedding methods from recent deep learning methods have been introduced to capture the spatial context of POIs (Yan *et al*. 2017, Niu and Silva 2021). This approach considers the relationship between the type of the target POI and the types of its surrounding POIs as analogous to the relationship between a word and the word sequence (textual context) containing that word. Embeddings for POIs, that is, numerical vectors representing POI types, can be constructed by extending textual contexts to spatial contexts. Yao *et al*. (2017) used the Word2Vec model (Mikolov *et al*. 2013) to extract high-dimensional vectorized features from POI sequences constructed based on spatial proximity and captured contextual information. Several studies have subsequently considered spatial proximity and extended the Word2Vec model to the Place2Vec model (Yan *et al*. 2017, Zhai *et al*. 2019). Recognizing the inadequacy of Word2Vec in capturing spatial heterogeneity among urban areas, Niu and Silva (2021) applied the Doc2Vec model (Le and Mikolov 2014) to directly train vector representations of urban areas and substantially improve the identification of urban functional areas.

2.3. Graph neural network models and spatial interaction

The emergence of graph-based deep learning neural networks (GNNs) has provided a new perspective for spatial interaction modeling, which considers the interactions between places as a spatially embedded graph or network, where areal units are represented as nodes and interacting flows are represented as weighted edges (Liu *et al*. 2016, Zhu *et al*. 2020, Xu *et al*. 2022). However, the majority of existing spatially explicit GNN models consider the spatial dependence (functional similarity) of urban functions nearby, according to Tobler's first law of geography. This is typically performed in three forms: adjacency-based, distance-based, and link-based. The first two forms

involve contiguity-based (Guan *et al*. 2023) and distance-decay-based (Xu *et al*. 2022) neighborhood configurations, respectively. The third form uses flow data as links to define the topological neighborhood (Liu *et al*. 2016), which is a network variant of the first two forms. Consequently, none of the three forms represents the underlying mechanism of spatial interaction theorized by Ullman (1954), that is, the function complementarity or the supply-demand relationship between urban places, which is also the definition used in this study.

Specifically, the spatial interaction between land-use functions requires the recognition of different functional roles of lands, that is, the role supplying land-use functions and the role demanding them, which are captured by the heterogeneity of the direction and magnitude of spatial interaction flows, such as human mobility trips. To the best of our knowledge, no GNN-based models have been developed for this purpose. Therefore, to utilize the great potential of GNNs, the distinction between spatial interaction and spatial dependence must be explicitly treated in both conceptualization and modeling.

3. Methodology

3.1. Proposed framework

3.1.1. Urban land-use inference problem

Given a set of spatial units U and their local attributes S relevant to certain urban functions, we consider two types of spatial relationships, namely spatial dependence R_D and spatial interaction R_I, between spatial units in the inference of land use Y. That is, to determine function f :

$$f(S, R_D, R_I) \rightarrow Y, \tag{1}$$

where $y_i = \left[t_i^1, \ldots, t_i^m\right] \in Y$, $\sum_{k=1}^{m} t_i^k = 1$ represents the composition of land uses within a spatial unit u_i and m indicates the number of land-use types. For any spatial unit $u_i \in U$, a set of POIs located within it, denoted as $s_i = \{p_i^1, \ldots, p_i^{|s_i|}\} \in S$, can be used to characterize the local attributes.

Given the urban land-use inference problem, we propose a novel framework, the HGCN, which integrates spatial interaction theory with a carefully designed graph neural network to account for the heterogeneity of spatial relationships (R_D and R_I). Specifically, a multigraph representation, $G = (V, E)$, $|V| = |U|$, is employed to allow for multiple edges with different properties between the same pair of nodes (Gjoka *et al*. 2011). To recognize the different roles of spatial units in spatial interaction and dependence relationships, we define three heterogeneous edges (meta-paths) $E = \{E_{in}\} \cup \{E_{out}\} \cup \{E_{adj}\}$: supply, demand, and proximity; accordingly, we further define three modes—production, attraction, and adjacency—for a spatial unit of land (node), which shifts modes adaptively conditioned on the meta-path type (Figure 1). This draws inspiration from the meta-path method (Sun *et al*. 2011), a composite relation scheme containing prior knowledge that utilizes message-passing to capture rich semantic information from distinct and heterogeneous edges. Therefore, this framework can distinguish between the underlying

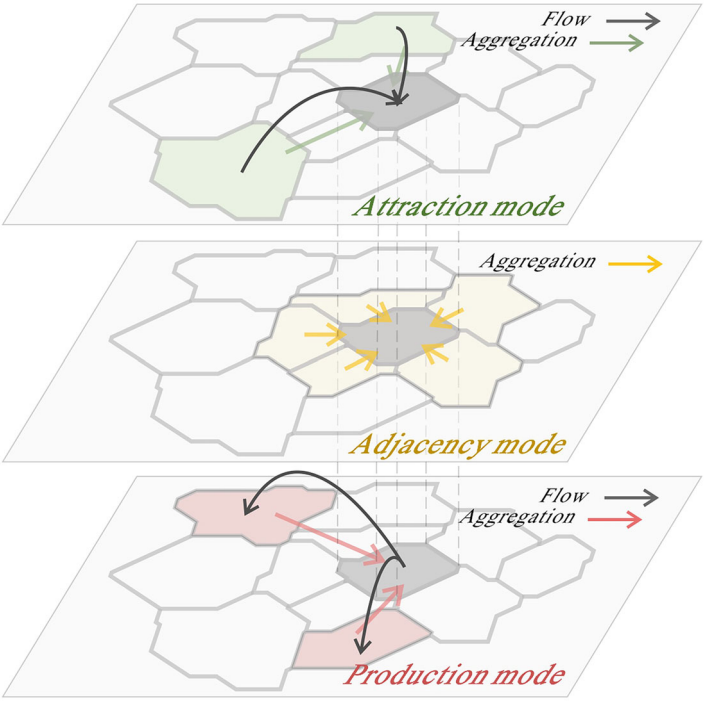


Figure 1. Three meta-paths centered around a specific node. These meta-paths reflect the multi-modal nature of nodes (attraction/production/adjacency).

mechanisms of spatial interaction and spatial dependence and combine both mechanisms in land-use inference.

Specifically, when one unit is in the attraction mode, this framework captures the commonality and diversity of the spatial demands that this unit supplies through its inflows from other units that are in the production mode. When one unit is in the production mode, it captures the commonality and diversity of spatial supplies that this unit demands through its outflows to other units that are in the attraction mode. The multimodal nature of spatial units determines the type and asymmetry of their spatial interactions with other units. When a unit is in the adjacency mode, it captures the spatial dependence feature of its neighborhood, defined by the proximity relation of the linkages.

Figure 2 illustrates the overall workflow. (1) During the pre-processing stage, we use the Doc2Vec model to learn node embeddings (local attribute features) of each spatial unit from POI data (Section 3.2); (2) In the HGCN framework, given the node embeddings of spatial units, the framework explicitly captures spatial dependence and spatial interaction features between spatial units based on adjacency and spatial flows. Finally, we combine different features as combined node embeddings for each spatial unit (Section 3.3); (3) In the function recognition stage, a position-wise feed-forward network (FFN) (Vaswani *et al.* 2017) is employed to predict the proportion of various urban land uses for each spatial unit based on the combined node embeddings (Section 3.3).

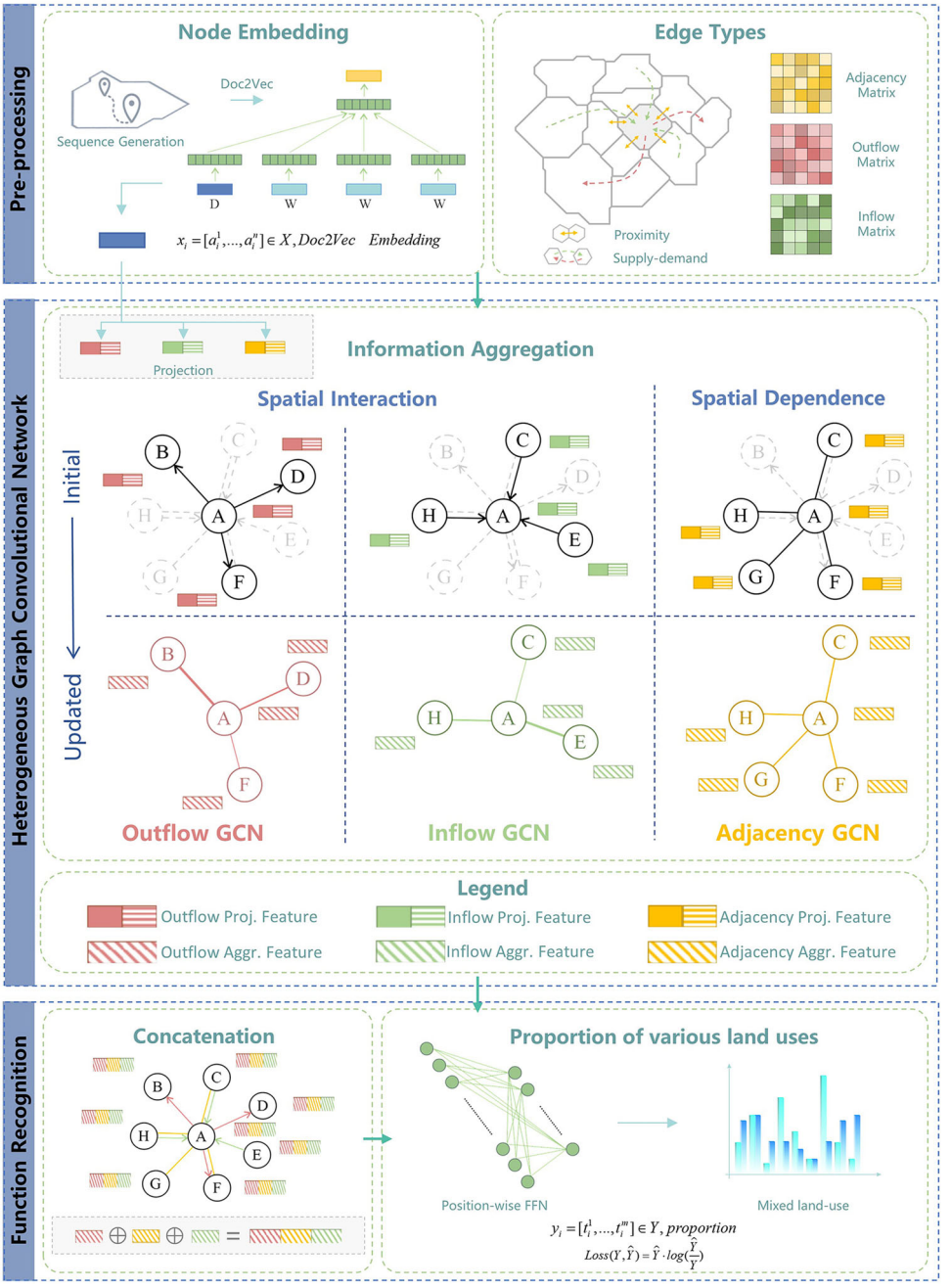


Figure 2. Proposed workflow.

3.2. Learning representation of local attribute features

POIs exist in online digital maps, often represented by point symbols indicating functional elements such as shops, parks, and residences. Due to their strong timeliness, accessibility, and rich semantics, we utilize them to characterize the function-related

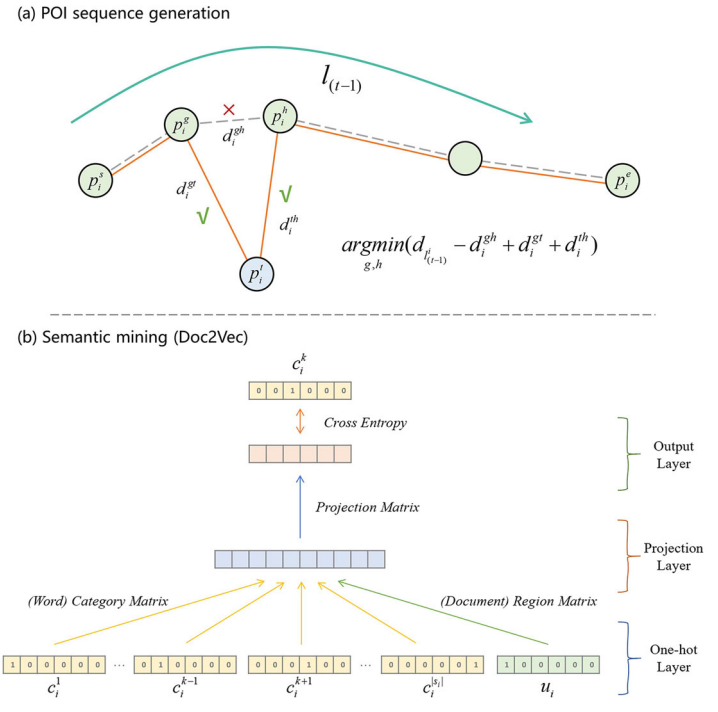


Figure 3. The representation of function-related local attributes of spatial units.

local attributes of spatial units. Technically, this process, as illustrated in Figure 3, can be divided into two major steps:

*Step 1: POI sequence generation (*Yao *et al.* 2017*)*. We begin by obtaining the set of POIs $s_i = \{p_i^1, \ldots, p_i^{|s_i|}\} \in S$ within a given spatial unit $u_i \in U$. Afterward, we calculate the spatial distances between each pair of POIs and designate the two with the farthest distance as the initial sequence $l_{(0)}^i = [p_i^s, p_i^e]$. Subsequently, we iteratively add POIs to the sequence until all within the unit are included. Suppose, at a given timestamp t, we need to insert a new POI p_i^t into the sequence, we can identify a suitable position between two existing POIs p_i^g, p_i^h, satisfying the following condition:

$$\underset{g,h}{\arg\min} \left(d_{l_{(t-1)}^i} - d_i^{gh} + d_i^{gt} + d_i^{th} \right), \tag{2}$$

where $d_{l_{(t-1)}^i}$ is the length of the sequence at timestamp $t-1$ and $d_i^{gh}, d_i^{gt}, d_i^{th}$ represents the distances between the POI p_i^g and p_i^h, p_i^g and p_i^t, p_i^t and p_i^h, respectively. After acquiring this POI sequence, we proceed to retrieve the level-2 category information for each POI, constructing a textual sequence $[c_i^1, \ldots, c_i^{|s_i|}]$, e.g. [recreational, recreational, …, foodstuffs, organization, accommodation]. This sequence effectively captures the functional characteristics of each spatial unit.

Step 2: Semantic mining. The Doc2Vec model extends the idea of Word2Vec to produce fixed-length vectors for documents (Le and Mikolov 2014). To capture semantic meaning and relationships in a vector space, it trains a neural network to predict the

context (words) of a given document. Given the similarity between 'words-documents' and 'POIs-spatial units', we employ this approach to encode textual sequences into vectors. In this approach, a POI class is represented as a vector in a shared vector space. POI classes with similar contexts are located near each other in space, whereas POI classes that only occur in different and disconnected contexts are placed further away. Doc2Vec aggregates the embeddings of POIs, captures the spatial proximity and context of POIs in a spatial unit, and produces a node embedding $x_i = [a_i^1, \ldots, a_i^n] \in X$ for each unit $u_i \in U$.

3.3. HGCN

To incorporate the three different node modes (production, attraction, and adjacency) and three types of meta-paths (supply, demand, and proximity) defined in Section 3.1, we devise an HGCN layer based on conventional GCN layers (Kipf and Welling 2016), as follows:

$$H^t = HGCN(H^{t-1}, L, I) = Concat(Feat), Feat \subseteq \left\{Adj^t, Attr^t, Prod^t\right\}, \tag{3}$$

$$Adj^t = GCN\left(H^{t-1}W_{adj}^t, L\right), \tag{4}$$

$$Attr^t = GCN\left(H^{t-1}W_{in}^t, I^T\right), \tag{5}$$

$$Prod^t = GCN\left(H^{t-1}W_{out}^t, I\right), \tag{6}$$

where H^t is the hidden state output by the t−th HGCN layer (H^0 is equivalent to the local attributes X), L is the adjacency matrix with the addition of self-loops, I is the normalized origin-destination matrix, and $W_{adj}^t, W_{in}^t, W_{out}^t$ are matrices of trainable parameters (weights) in the neural networks. We use GCN layers $GCN(\cdot)$ to aggregate the features pertaining to the three node modes (Adj^t, $Attr^t$, and $Prod^t$). $Concat(\cdot)$ is a function that combines different features by concatenating vectors along the feature dimension. $Feat$ is a subset of the features $\left\{Adj^t, Attr^t, Prod^t\right\}$, indicating that it can be a set of any combination of different features in $\left\{Adj^t, Attr^t, Prod^t\right\}$. This process is similar to the multi-head attention mechanism in the Transformer architecture (Vaswani *et al*. 2017). However, the difference is that HGCN explicitly acquires spatial interaction and dependence features from spatial flow data and spatial proximity, whereas the Transformer architecture learns features implicitly from the data. In the following, we introduce four specific steps for each HGCN layer:

Step 1: Three sets of trainable parameters $W_{adj}^t, W_{in}^t, W_{out}^t$ are used to project H^{t-1} into three feature spaces pertaining to the production, attraction, and adjacency modes, where the first two modes capture spatial interaction features, and the third captures spatial dependence features.

Step 2: To capture the spatial dependence features, the subgraph $G_{adj} = (V, E_{adj})$ is used, and the GCN message-passing mechanism is applied to synthesize the neighborhood information to represent the features of the adjacency mode (Equation (4)). The neighborhood setting L can be further specified by adjacency ($L_{i,j}^{adj}$) and distance-decay ($L_{i,j}^{dd}$). The former represents the first-order queen neighborhood, and the latter is a neighborhood that includes all other units but is weighted by their inverse

distance to the target unit (Equation (7)), reflecting the two extremes of Tobler's First Law of Geography. The specific formula for the latter is as follows:

$$L_{i,j}^{dd} = \ln\left(\frac{1+\max(d)^{\varepsilon}}{1+d_{i,j}^{\varepsilon}}\right), \tag{7}$$

where $d_{i,j}$ represents the distance between two spatial units, and ε is the distance-decay coefficient.

Step 3: Obtain spatial interaction information using inflow and outflow data. As mentioned in Section 3.1, spatial units have two crucial modes: production and attraction. The representations for the attraction and production modes are as follows:

$$\begin{aligned} Attr_j &\propto \frac{1}{D_j}\cdot\sum_i \frac{T_{ij}}{O_i}\cdot V_i, \\ Prod_i &\propto \frac{1}{O_i}\cdot\sum_j \frac{T_{ij}}{D_j}\cdot S_j, \end{aligned} \tag{8}$$

s.t.

$$\begin{cases} O_i = \sum_j T_{ij} \\ D_j = \sum_i T_{ij} \end{cases},$$

where T_{ij} is the flow from units i to j; O_i and D_j are the marginal total flows at units i and j, respectively; and V_i and S_j represent the vector representation of the demand and supply of spatial units, respectively, and they are from the projected feature spaces of $H^{t-1}W_{in}^{t}$ and $H^{t-1}W_{out}^{t}$, respectively. According to the spatial interaction theory, specifically Alonso's model (Alonso 1960), the flow T_{ij} between i and j should depend on the characteristics of the localities of the origin and destination, the alternative opportunities available from that origin, and the degree of competition existing at that destination:

$$T_{ij} = \frac{A_i^{1-\alpha}V_i B_j^{1-\beta}S_j}{c_{ij}}, \tag{9}$$

where parameters α and β are values between zero and one, and c_{ij} represents the spatial cost. Opportunities and competition are represented by A_i^{-1} and B_j^{-1}, respectively, and are defined as follows:

$$A_i^{-1} = \sum_j \frac{B_j^{1-\beta}S_j}{c_{ij}}, \tag{10}$$

$$B_j^{-1} = \sum_i \frac{A_i^{1-\alpha}V_i}{c_{ij}}. \tag{11}$$

Inserting Equations (9)–(11) into Equations (8), we can derive the production and attraction as the following accessibility-based operators (Wang 2021):

$$Attr_j \propto \frac{1}{B_j^{-1}} \cdot \sum_i \left(\frac{V_i}{c_{ij}} \cdot \frac{1}{A_i^{-1}} \right) = B_j \cdot \sum_i A_i \frac{V_i}{c_{ij}}, \tag{12}$$

$$Prod_i \propto \frac{1}{A_i^{-1}} \cdot \sum_j \left(\frac{S_j}{c_{ij}} \cdot \frac{1}{B_j^{-1}} \right) = A_i \cdot \sum_j B_j \frac{S_j}{c_{ij}}, \tag{13}$$

where the attraction of a spatial unit of land is related to the demand of other units, weighted by the inverse of spatial costs, with the opportunity effect normalized on each i side and the competition effect normalized on the j side after the summation. However, the production of a spatial unit is related to the supply of other units it can access, weighted by the inverse of spatial costs, with the competition effect normalized on each j side and the opportunity effect normalized on the i side after the summation. Equation (8) can be integrated into the graph convolutional framework by extending the message-passing mechanism. Technically, we need only extract a subgraph $G_{flow} = (V, E_{flow} = \{E_{in} \cup E_{out}\})$ and use the message-passing mechanism to synthesize the supply and demand information.

Step 4: After obtaining the spatial dependence and spatial interaction features, their vectors are concatenated along the feature dimension as node embeddings (Equation (3)). Note that for each node, we use the same set of trainable parameters (neural network weights), which are similar to the filters of convolutional neural networks (Krizhevsky *et al*. 2012). In addition, this model still has node-order equivariance, which is significant in graph models.

3.3.1. Function recognition

After obtaining the node embeddings H^t, we can predict the proportion of various land uses within spatial units using a simple position-wise *FFN* (Vaswani *et al*. 2017):

$$Y = FFN(H^t). \tag{14}$$

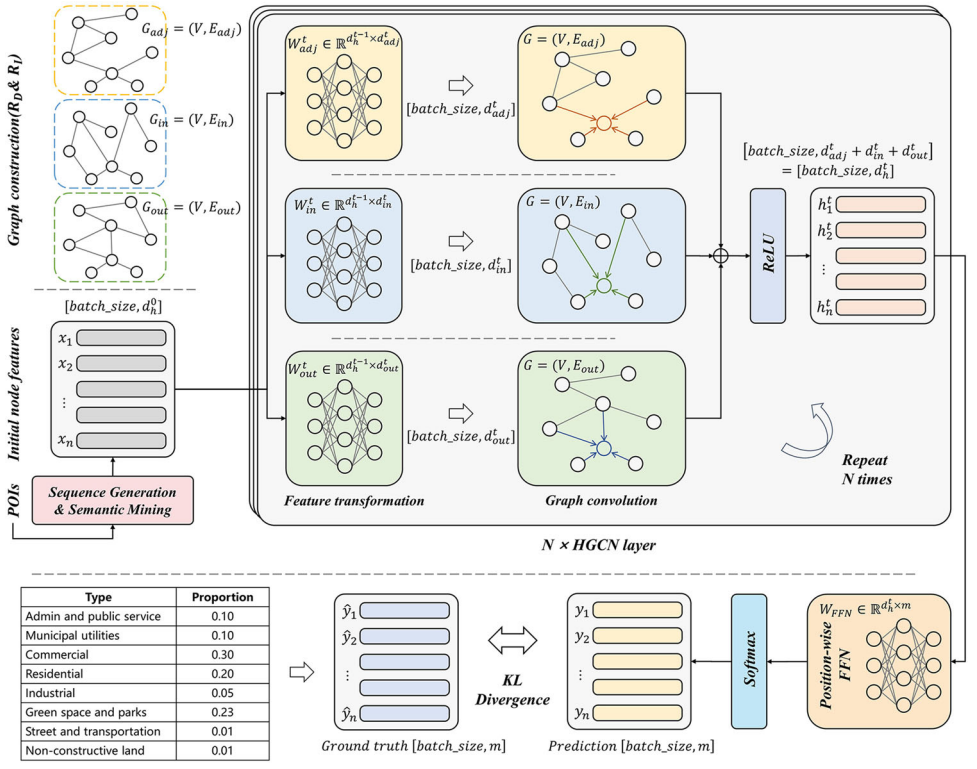
Because all the embedding computations introduced above are completely differentiable, the HGCN framework and urban function recognition stage can be trained in an end-to-end manner once a suitable loss function is defined. In this model, our goal is to minimize the difference between the predicted distribution and the actual distribution. To achieve this goal, we select the KL Divergence as our loss function, defined as:

$$Loss(Y, \hat{Y}) = \hat{Y} \cdot \log\left(\frac{\hat{Y}}{Y}\right), \tag{15}$$

where $\hat{Y}$ is the ground truth data of the proportion of various land uses within spatial units. The underlying network architecture is shown in Figure 4.

3.4. Data and study areas

To demonstrate the feasibility of our framework in practice, a series of experiments were conducted in Shenzhen, China. Also, to show that our framework can be applied to other cities, selective experiments were conducted in London, UK (Figure 5). Both Shenzhen and London are large cities with high population densities, diverse urban



Type	Proportion
Admin and public service	0.10
Municipal utilities	0.10
Commercial	0.30
Residential	0.20
Industrial	0.05
Green space and parks	0.23
Street and transportation	0.01
Non-constructive land	0.01

Figure 4. The underlying network architecture.

functions, and complex mobility patterns. We collected data on land use, POI, and human mobility data from both areas. For Shenzhen, the spatial unit was defined as a regular grid of 250 m × 250 m, while for London, the spatial units were delineated coverage areas based on bike docking stations (see details of area delineation in Appendix A). Land use and POI data were aggregated into spatial units for the two cities.

POI data: We used POI data to quantify the urban functional features of the spatial units (detailed in Section 3.2). The POI data for Shenzhen in 2018 were sourced from AMap (https://ditu.amap.com/), comprising a total of 1,577,354 records. Each record contains attributes of the corresponding POI, such as longitude, latitude, name, address, phone number, and a classification code with a three-level structure indicating the specific functional type (with 245 unique categories for level-2). London's POI data for 2020 were obtained from the Ordnance Survey, with records containing identical information to that for Shenzhen. It encompasses 490,626 records, classified into 52 unique level-2 categories.

Spatial interaction data (human mobility data): For Shenzhen, mobile phone data for November 2018 were obtained from China UniCom, one of the largest communication operators in China, with more than 400 million users. The data records the daily movements of users between the predefined 250 m × 250 m grids for 30 days. There are over seven million records, and each record includes the origin, destination, date, and number of movements. The spatial interaction data of London were obtained

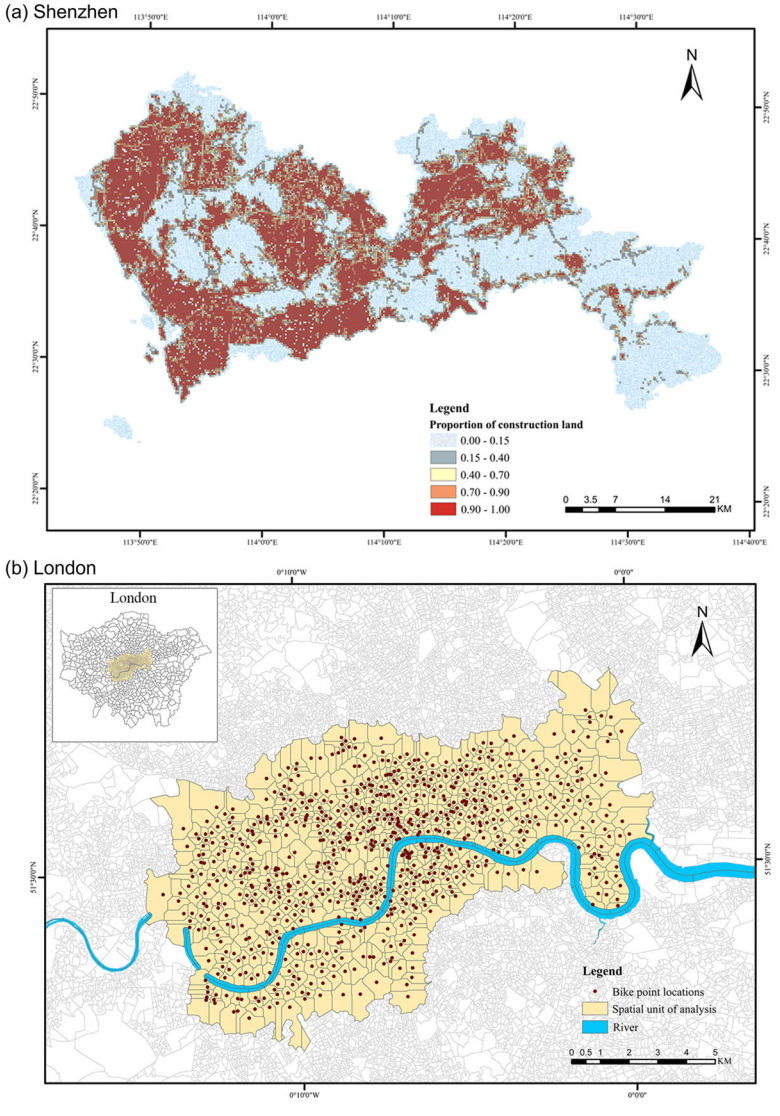


Figure 5. Study areas: Shenzhen (top) and London (bottom).

from the Transport for London Bicycle Flows dataset, which records the hourly movement of rental bicycles between parking stations in Central London. We used a full year of cycle flow data for 2020, including 11 million cycle flow records between 786 parking stations in Central London (refer to Appendix B for the validity of this dataset). Data pre-processing was performed to exclude distinct outliers in the speed and duration of use, and we made a daily average of the flows between any two spatial units in both areas. Detailed elaboration on raw spatial interaction data can be found in Appendix C.

Land use data: Land use data for both study areas were collected from an official land-use survey. Land use data for Shenzhen were collected from the 3rd National Land Resource Survey 'Classification of Land Use Status' (GBT 21010-2017) that was

launched by the State Council in September 2018, which includes 53 detailed land-use types. The data for London were derived from the Geomni UKMap Land Use dataset (https://digimap.edina.ac.uk/os), which includes 15 land functional use types. To ensure that the results were compatible with the two study areas, we aggregated data from both study areas into eight land-use categories: commercial (such as wholesale, retail, accommodation, catering, and financial land), industrial (such as industrial parks, factories, mineral extraction), residential, administration and public service (such as schools and hospitals), street and transportation (such as transport tracks and transport stations), municipal utilities (such as utility services and infrastructure), green space and parks, and non-constructive land (Table 1).

4. Experiments and results

4.1. Experiment design

Three sets of experiments were designed to answer the following questions (refer to the experimental configurations in Table 2):

Table 1. Mean and standard deviation (SD) of the percentage distribution across eight land uses.

	Shenzhen		London	
Category	Mean	SD	Mean	SD
Administration and public service	0.056	0.124	0.020	0.141
Municipal utilities	0.013	0.046	0.003	0.050
Commercial	0.088	0.160	0.242	0.428
Residential	0.261	0.273	0.463	0.499
Industrial	0.221	0.279	0.003	0.050
Green space and parks	0.048	0.131	0.062	0.242
Street and transportation	0.208	0.158	0.198	0.399
Non-constructive land	0.105	0.173	0.009	0.094

Table 2. Model specification with different features.

				Spatial dependence		Spatial interaction		
						Undirected	Directed	
GP	ID	Models	Local	Spatial adjacency	Distance decay	Average flow	Demand inflow	Supply outflow
A1	1	HGI	√					
	2	Semantic	√					
	3	Doc2Vec (L)	√					
A2	4	GLM	√					
	5	GMDA	√					
B	6	GCN_{adj}		√				
	7	GCN_{dd}			√			
C1	8	GCN_{avg}				√		
	9	GCN_{in}					√	
	10	GCN_{out}						√
	11	HGCN					√	√
C2	12	$L + GCN_{avg}$	√			√		
	13	$L + GCN_{in}$	√				√	
	14	$L + GCN_{out}$	√					√
	15	L + HGCN	√				√	√
C3	16	$GCN_{adj}+GCN_{avg}$		√		√		
	17	$GCN_{adj}+GCN_{in}$		√			√	
	18	$GCN_{adj}+GCN_{out}$		√				√
	19	$HGCN_{adj}$		√			√	√

Q1: Does land-use inference need to consider not only the local attributes of spatial units and the spatial dependence between local attributes but also the spatial interaction characteristics defined in the scope of this study?

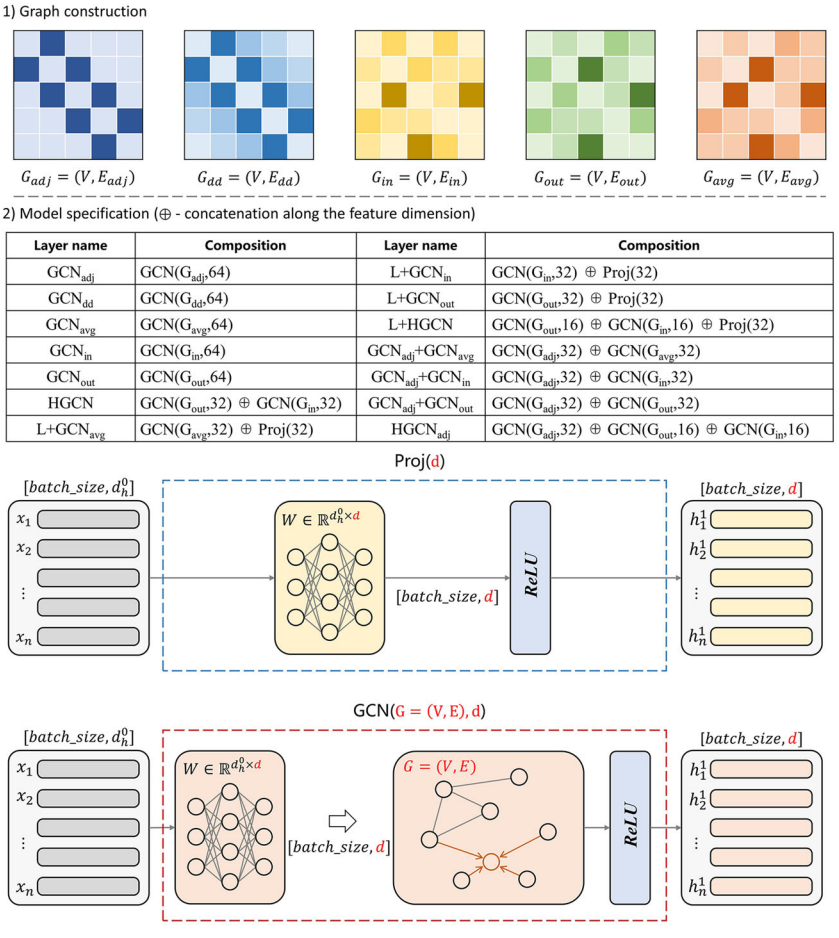
Q2: Are the proposed HGCN-based models superior to the GCN-based models in terms of capturing spatial interaction features?

Table 2 lists the three groups of models (A, B, and C) in terms of the types of features considered. Group A includes traditional statistical and neural network models that consider only the local attribute features. Specifically, Group A1 comprises three recently proposed multi-layer perceptron (MLP)-based models with different methods to learn POI embeddings, including Doc2vec, HGI (Huang *et al*. 2023), and Semantic (Huang *et al*. 2022). Group A2 includes two traditional statistical models, namely the general linear model (GLM) and Gaussian mixture discriminant analysis (GMDA), from the relevant literature (Ju *et al*. 2003). Group B includes GCN-based models that consider only the spatial dependence feature. GCN models were employed to incorporate the spatial dependence between local attributes and attributes in the spatial neighborhood based on adjacency and distance-decay settings, respectively.

Group C includes both GCN- and HGCN-based models that consider the spatial interaction feature with and without local attribute or spatial dependence features. Specifically, Group C1 includes one HGCN model and three GCN models using inflows (GCN_{in}), outflows (GCN_{out}), or averaged flows (GCN_{avg}) to construct graph-based spatial neighborhoods. In Group C2, the L indicates the selected Doc2Vec model using only local attribute features. The L + model indicates that local attribute features are combined with spatial interaction features extracted by GCN or HGCN mechanisms for land-use inference. In Group C3, the GCN_{adj}+model indicates that spatial dependence features are combined with spatial interaction features extracted by GCN or HGCN mechanisms for land-use inference. The $HGCN_{adj}$ is conceptually equivalent to the GCN_{adj}+HGCN. The structural specification of models in Table 2 is illustrated in Figure 6.

Comparison of the HGCN, L + HGCN, and $HGCN_{adj}$ in Group C to those in Groups A and B examines how different the impact of spatial interaction features is from those of local attribute and spatial dependence features on land-use inference, and thus it answers Q1. Moreover, comparison between HGCN-based and GCN-based models, namely HGCN, L + HGCN, and $HGCN_{adj}$ versus other models within Group C, examines the proposed advantage of HGCN over GCN, which responds to Q2.

In these experiments, we fixed the length (or dimension) of the vectors to 64, representing a hidden layer in the neural network model, which is a concatenation of different combinations of features, such as local attributes, spatial dependence, or spatial interaction, as specified in Table 2. When combined with other features (Groups C2 and C3, respectively), the vector lengths of both the local attribute and spatial dependence features were fixed at 32; otherwise, their lengths were 64 (Groups A and B respectively). In the former, the vector lengths of the other features were adjusted accordingly. When combined with other features, the vector length of the spatial interaction feature was fixed at 32; otherwise, it was set to 64 (Group C1). All HGCN-based models used only one HGCN layer for comparison. We adopted Adam as the optimizer and maintained a random seed (30), training epochs (400), and training/testing data (70%/30%) for all experiments. Results were obtained by calculating the mean and



Layer name	Composition	Layer name	Composition
GCN_{adj}	GCN(G_{adj},64)	L+GCN_{in}	GCN(G_{in},32) ⊕ Proj(32)
GCN_{dd}	GCN(G_{dd},64)	L+GCN_{out}	GCN(G_{out},32) ⊕ Proj(32)
GCN_{avg}	GCN(G_{avg},64)	L+HGCN	GCN(G_{out},16) ⊕ GCN(G_{in},16) ⊕ Proj(32)
GCN_{in}	GCN(G_{in},64)	GCN_{adj}+GCN_{avg}	GCN(G_{adj},32) ⊕ GCN(G_{avg},32)
GCN_{out}	GCN(G_{out},64)	GCN_{adj}+GCN_{in}	GCN(G_{adj},32) ⊕ GCN(G_{in},32)
HGCN	GCN(G_{out},32) ⊕ GCN(G_{in},32)	GCN_{adj}+GCN_{out}	GCN(G_{adj},32) ⊕ GCN(G_{out},32)
L+GCN_{avg}	GCN(G_{avg},32) ⊕ Proj(32)	$HGCN_{adj}$	GCN(G_{adj},32) ⊕ GCN(G_{out},16) ⊕ GCN(G_{in},16)

Figure 6. The structural specification of the models in Table 2.

standard deviation of the outcomes from 30 iterations for each model. We used PyTorch for model building and trained our models on a GeForce RTX 3090 GPU. Each model took approximately 10 minutes to train. In total, we conducted 19 experiments using the Shenzhen and London data.

As the inferred urban functional distributions are vectors of proportions, there are two general types of evaluation measures: distance and similarity metrics (Cha 2007, Geng 2016). Three metrics were selected to evaluate the model performance during the experiments:

$$L1\ \ distance = |Y - \hat{Y}|, \tag{16}$$

$$KL\ \ divergence = \hat{Y} \bullet \log\left(\frac{\hat{Y}}{Y}\right), \tag{17}$$

$$Cosine\ \ similarity = \frac{Y \bullet \hat{Y}}{||Y||_2 \bullet ||\hat{Y}||_2}, \tag{18}$$

where Y represents the estimated distribution of the functional proportions, and $\hat{Y}$ represents the corresponding ground truth proportions.

4.2. Comparison of model performance

Table 3 lists the average performance results of a series of models for Shenzhen, indicated by the 'mean ± standard deviation' of three evaluation metrics (Equations (16)–(18)) and ranked from the best to the worst based on these metrics, after 30 iterations for each model (excluding GMDA, which is a deterministic method).

Models considering only local attributes of spatial units: Group A includes a set of recent models that consider only the attribute features of spatial units for urban land-use inference. The best-performing model was selected from these models for experiments in Group C. Table 3 shows that, in general, the models in Group A performed the worst among all groups. Specifically, in Group A2, the performances of the two statistical methods, GLM and GMDA, ranked lowest among all models. The MLP-based models in Group A1 performed better in general than those in Group A2. Specifically, Doc2vec (L) surpassed other modern embedding techniques based on POIs and was selected for further experimentation in Group C.

Models considering only spatial dependence features: Group B includes a set of GCN models that fuse the attribute features of spatial units and their spatial dependence with other units. Table 3 shows that the performances of the models in Group B varied greatly. GCN_{dd}, with the distance-decay neighborhood, performed poorly compared to certain models in Group A1 (e.g. Semantic and Doc2Vec). In contrast, the performance of GCN_{adj} with the adjacency neighborhood ranked ninth among the nineteen models; thus, it was selected for further experimentation in Group C3.

Models considering spatial interaction features: Group C1 aimed to test if HGCN, which explicitly accounts for the demand (inflow) and supply (outflow) components of spatial interaction data, outperformed GCN using the same data. As illustrated in Table 3, the HGCN model (ranked 11th) exhibited superior performance. In comparison, the three GCN models had lower performance, i.e. GCN_{avg}, GCN_{in}, and GCN_{out} were ranked 12th, 14th, and 13th, respectively.

Table 3. Model performance for Shenzhen.

GP	Models	Evaluation metrics			Rank
		KL[Loss]↓	L1↓	Cosine↑	
A1	HGI	$0.7088_{\pm 0.0057}$	$0.1095_{\pm 0.0007}$	$0.7324_{\pm 0.0026}$	17
	Semantic	$0.6949_{\pm 0.0063}$	$0.1058_{\pm 0.0005}$	$0.7468_{\pm 0.0022}$	15
A2	**Doc2Vec (L)**	$\mathbf{0.5434_{\pm 0.0057}}$	$\mathbf{0.0926_{\pm 0.0006}}$	$\mathbf{0.7985_{\pm 0.0024}}$	**10**
	GLM	$0.9167\pm_{0.0061}$	$0.1398_{\pm 0.0005}$	$0.6310_{\pm 0.0028}$	19
	GMDA	0.7237	0.1235	0.6950	18
B	GCN_{dd}	$0.7075_{\pm 0.0065}$	$0.1139_{\pm 0.0006}$	$0.7106_{\pm 0.0028}$	16
	GCN_{adj}	$\mathbf{0.5259_{\pm 0.0042}}$	$\mathbf{0.0912_{\pm 0.0005}}$	$\mathbf{0.7987_{\pm 0.0021}}$	**9**
C1	GCN_{avg}	$0.6714_{\pm 0.0067}$	$0.1074_{\pm 0.0006}$	$0.7331_{\pm 0.0028}$	12
	GCN_{in}	$0.6820_{\pm 0.0073}$	$0.1087_{\pm 0.0007}$	$0.7278_{\pm 0.0029}$	14
	GCN_{out}	$0.6795_{\pm 0.0072}$	$0.1086_{\pm 0.0006}$	$0.7287_{\pm 0.0029}$	13
	HGCN	$0.6318_{\pm 0.0081}$	$0.1025_{\pm 0.0007}$	$0.7530_{\pm 0.0031}$	11
C2	$L + GCN_{avg}$	$0.5203_{\pm 0.0061}$	$0.0895_{\pm 0.0006}$	$0.8082_{\pm 0.0025}$	6
	$L + GCN_{in}$	$0.5241_{\pm 0.0063}$	$0.0900_{\pm 0.0006}$	$0.8065_{\pm 0.0027}$	8
	$L + GCN_{out}$	$0.5208_{\pm 0.0060}$	$0.0896_{\pm 0.0006}$	$0.8080_{\pm 0.0024}$	7
	L + HGCN	$0.5144_{\pm 0.0057}$	$0.0890_{\pm 0.0005}$	$0.8103_{\pm 0.0024}$	5
C3	$GCN_{adj}+GCN_{avg}$	$0.5005_{\pm 0.0053}$	$0.0877_{\pm 0.0006}$	$0.8112_{\pm 0.0022}$	2
	$GCN_{adj}+GCN_{in}$	$0.5025_{\pm 0.0052}$	$0.0880_{\pm 0.0005}$	$0.8103_{\pm 0.0022}$	4
	$GCN_{adj}+GCN_{out}$	$0.5012_{\pm 0.0053}$	$0.0878_{\pm 0.0005}$	$0.8108_{\pm 0.0021}$	3
	$HGCN_{adj}$	$0.4974_{\pm 0.0053}$	$0.0876_{\pm 0.0005}$	$0.8119_{\pm 0.0023}$	1

Note: Bold values indicate the best-performing models in groups A and B, respectively.

Group C2 aimed to test, while considering local attribute features (L), if L + HGCN, which explicitly accounts for the demand (inflow) and supply (outflow) components of spatial interaction data, consistently performed better than L + GCN using the same data. As presented in Table 3, the L + HGCN model (ranked 5th) outperformed its L + GCN counterparts (L + GCN_{avg}, L + GCN_{in}, and L + GCN_{out}), which were ranked from 6th to 8th, while local attributes were included.

Group C3 aimed to test, while considering spatial dependence features (GCN_{adj}), if $HGCN_{adj}$ (GCN_{adj}+HGCN), which explicitly accounts for demand (inflow) and supply (outflow) components of spatial interaction data, consistently outperformed GCN_{adj}+GCN using the same data. Consistent with the findings from Groups C1 and C2, Table 3 demonstrates that the $HGCN_{adj}$ model (ranked 1st) outperformed its GCN_{adj}+GCN counterparts (ranked from 2nd to 4th). We also visualized the features before input into the classifier for both $HGCN_{adj}$ and GCN_{adj}+GCN_{avg} (see Appendix D for details).

It is evident that spatial interaction is distinctively essential. L + HGCN in Group C2 and $HGCN_{adj}$ in Group C3 included Doc2Vec (L) and GCN_{adj}, respectively; they also incorporated spatial interaction features. The inclusion of spatial interaction information significantly improved the accuracy of model predictions. For example, the ranking improved from tenth to fifth when using L + HGCN, and from the ninth to the first when using $HGCN_{adj}$.

Notably, the models in Group C3 generally performed better than those in Group C2, which in turn performed better than those in Group C1. This is consistent with the finding that the best-performing model that considered the spatial dependence between local attributes (GCN_{adj}) had higher prediction power than the best-performing model that considered only local attributes (Doc2Vec). In addition, the $HGCN_{adj}$ model achieved the highest performance among all models. Selective experiments were also conducted for the London data, and the general findings were consistent with those found for the Shenzhen data (see Appendix Table A1 for details).

4.3. Model performance by land-use type

To compare the predictive performance of the models by land-use type, we selected the best-performing models in Groups A, B, and C and evaluated them using two error metrics: mean absolute error (MAE) and relative percentage difference (RPD). MAE was used to compare different models for the same land-use category by providing the magnitude of the absolute error, while RPD normalized the error on a relative scale, rendering the predictive performance of the models comparable across different land uses. Table 4 shows that the HGCN-based model, which considers both the spatial interaction and dependence features, outperformed the GCN-based model, which considers only the spatial dependence features, and the Doc2Vec model, which considers only the local attributes, for nearly all land-use categories, except for 'green space and parks,' where GCN_{adj} performed slightly better than $HGCN_{adj}$ according to the PRD (153.0% vs. 160.7%, respectively). This may be due to

Table 4. Prediction error of models by land use for the Shenzhen data.

	Models					
	Doc2Vec		GCN_{adj}		$HGCN_{adj}$	
Land uses	MAE	PRD	MAE	PRD	MAE	PRD
Administration and public service	**0.060**	136.1%	0.062	137.2%	**0.060**	**133.2%**
Municipal utilities	**0.021**	169.2%	**0.021**	170.4%	**0.021**	**168.2%**
Commercial	0.088	125.5%	0.087	122.6%	**0.084**	**122.3%**
Residential	0.145	81.0%	0.149	82.5%	**0.138**	**80.7%**
Industrial	0.141	117.3%	0.130	109.7%	**0.122**	**108.3%**
Green space and parks	0.062	163.8%	**0.058**	**153.0%**	**0.058**	160.7%
Street and transportation	0.107	59.6%	0.110	60.8%	**0.106**	**55.9%**
Non-constructive land	0.109	128.9%	0.105	126.5%	**0.100**	**119.3%**

Note: Bold values indicate the best MAE/PRD under different land uses.

the discrepancy between the large proportion of green space areas and the small number of POIs of this type.

Figure 7(a) presents a global visualization of the MAE for the HGCN-based model in Shenzhen, where red represents large errors compared to the ground truth. The region shown in Figure 7(b) illustrates a local estimation error (detailed in Table 5), indicating a significant phenomenon. In areas such as forests, where the ground truth value primarily consists of green spaces, the corresponding POIs are commercial and residential lands, such as small shops and scattered residential areas located within the parks. Consequently, the semantic information obtained from these POIs exhibits a substantial deviation from their ground truth values.

To compare model performance across different land-use categories and spatial effect patterns, in Figure 8, we visualize the relative performance improvement of the different models by subtracting the estimation error and displaying the difference in colors, with red indicating a great improvement. Tables 6 and 7 list the percentages of land-use types for each corresponding grid.

Figure 8(a) demonstrates the relative performance improvement of GCN_{adj} over Doc2Vec. As shown in Table 6, the area primarily comprises transportation and commercial land-use categories, with significant similarities between grids. Spatial dependence was the dominant spatial effect in this region. Therefore, GCN-based models constructed on the basis of topological adjacency effectively captured this spatial dependence and exhibited significant performance improvements.

Similarly, Figure 8(b) shows the performance improvement of $HGCN_{adj}$ over GCN_{adj}. As shown in Table 7, this area is characterized as a typical mixed land-use region with inherent spatial dependency. More significantly, functional complementarity exists between these spatial grids (e.g. grid 2 is dominated by residential use, grid 3 by commercial use, and grid 6 by industrial use), which provides a prerequisite for spatial interactions. $HGCN_{adj}$, which explicitly incorporates spatial interaction components, exhibited significant performance improvements in these mixed-use areas compared to that of GCN_{adj}. This is because human mobility flows often reflect socioeconomic linkages induced by the spatial distribution of the supply and demand of urban functions.

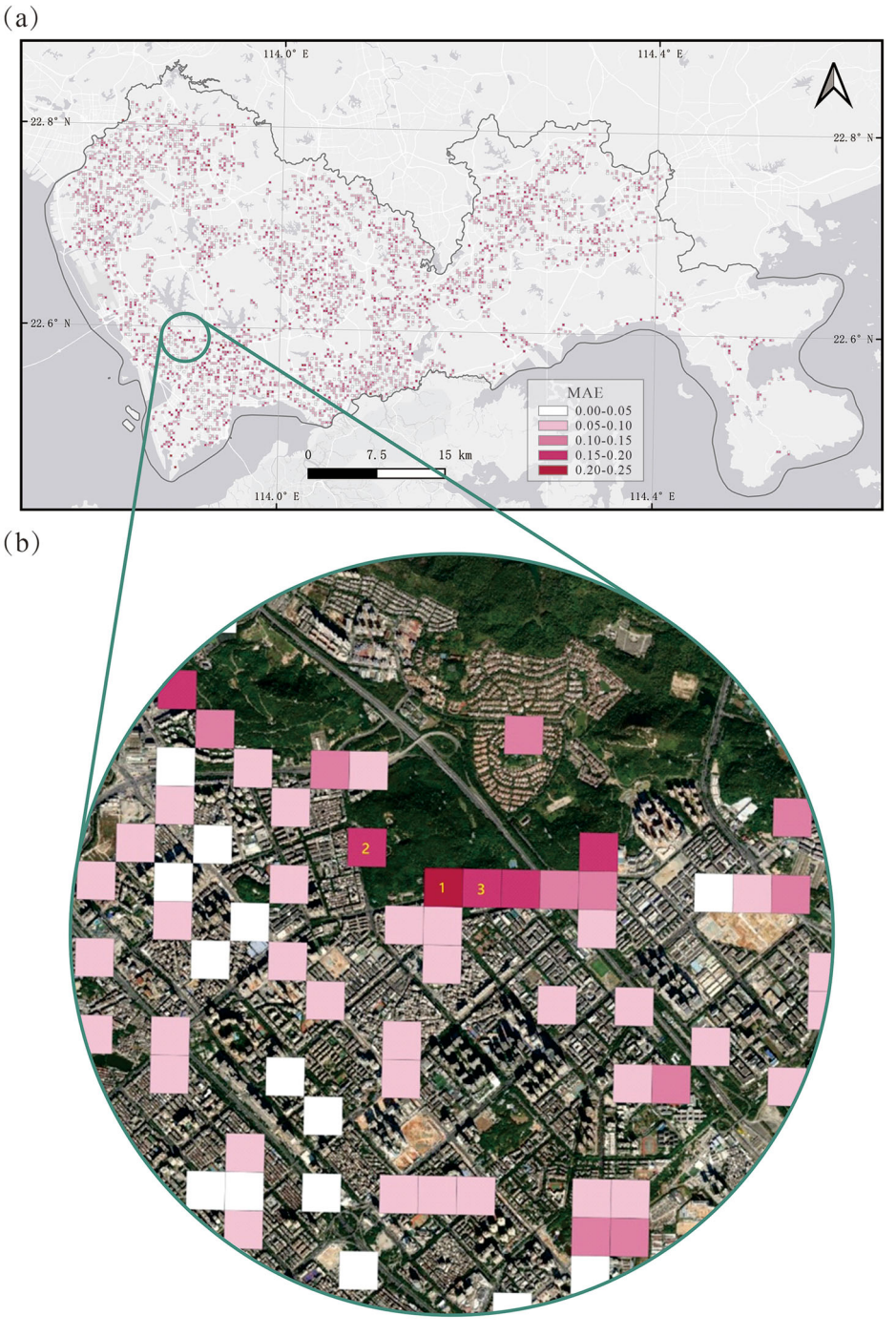


Figure 7. Global (a) and local (b) visualizations of MAE for the HGCN-based model.

Table 5. Proportions of different land-use categories in spatial grids (Figure 7(b)).

ID	1	2	3
Administration and public service	0.137	0.040	0.061
Municipal utilities	0.026	0.026	0.029
Commercial	0.048	0.129	0.033
Residential	0.293	0.132	0.128
Industrial	0.082	0.134	0.128
Green space and parks	**0.892**	**0.65**	**0.763**
Street and transportation	0.182	0.094	0.173
Non-constructive land	0.124	0.095	0.210

Note: Bold values indicate the percentage of the dominant land-use type in each spatial unit.

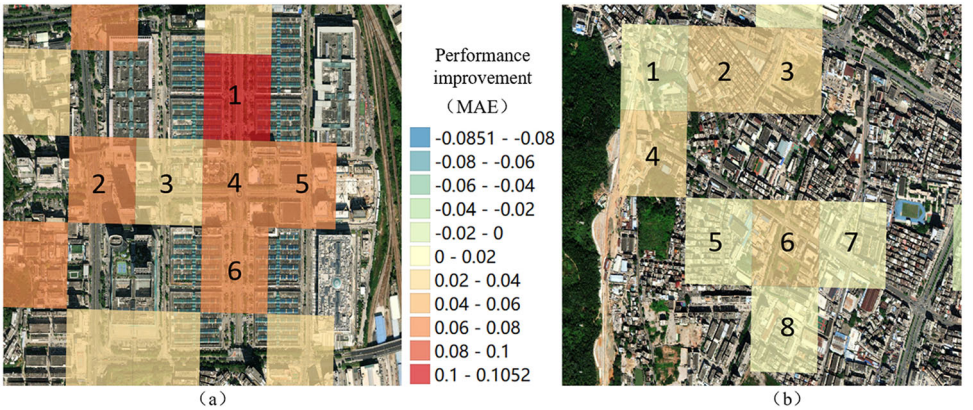


Figure 8. Performance improvement between different models: (a) GCN_{adj} over Doc2Vec, (b) $HGCN_{adj}$ over GCN_{adj}.

Table 6. Proportions of different land-use categories in spatial grids (Figure 8(a)).

ID	1	2	3	4	5	6
Administration and public service	0	0	0	0	0	0
Municipal utilities	0	0	0	0	0	0
Commercial	**0.694**	**0.559**	**0.718**	**0.579**	**0.770**	**0.695**
Residential	0	0.054	0	0	0	0
Industrial	0	0.022	0	0	0	0
Green space and parks	0	0	0	0	0	0
Street and transportation	0.306	0.364	0.282	0.421	0.230	0.305
Non-constructive land	0	0.001	0	0	0	0

Note: Bold values indicate the percentage of the dominant land-use type in each spatial unit.

5. Discussions

5.1. Local attribute, spatial dependence, and spatial interaction features

***Q1**: Does land-use inference need to consider not only the local attributes of spatial units and the spatial dependence between local attributes but also the spatial interaction characteristics defined in the scope of this study?*

The comparison between Groups C1, C2, and C3 demonstrates that models accounting for spatial interaction features alone (Group C1) have a certain predictive power. Models combining spatial interaction features with local attributes (Group C2) and spatial

Table 7. Proportions of different land-use categories in spatial grids (Figure 8(b)).

ID	1	2	3	4	5	6	7	8
Administration and public service	0	0.082	0	0	0	0.158	0.046	0.036
Municipal utilities	0.255	0.029	0	0.120	0	0	0	0
Commercial	0.005	0.132	**0.643**	0.018	0.019	0.041	**0.330**	0.050
Residential	**0.446**	**0.702**	0.221	0.099	**0.590**	0.130	0.147	0.295
Industrial	0.201	0.020	0.033	**0.383**	0.310	**0.482**	0.264	**0.565**
Green space and parks	0.024	0	0	0.072	0	0	0	0
Street and transportation	0.049	0.035	0.102	0.145	0.045	0.132	0.213	0.040
Non-constructive land	0.020	0	0.001	0.163	0.036	0.057	0	0.014

Note: Bold values indicate the percentage of the dominant land-use type in each spatial unit.

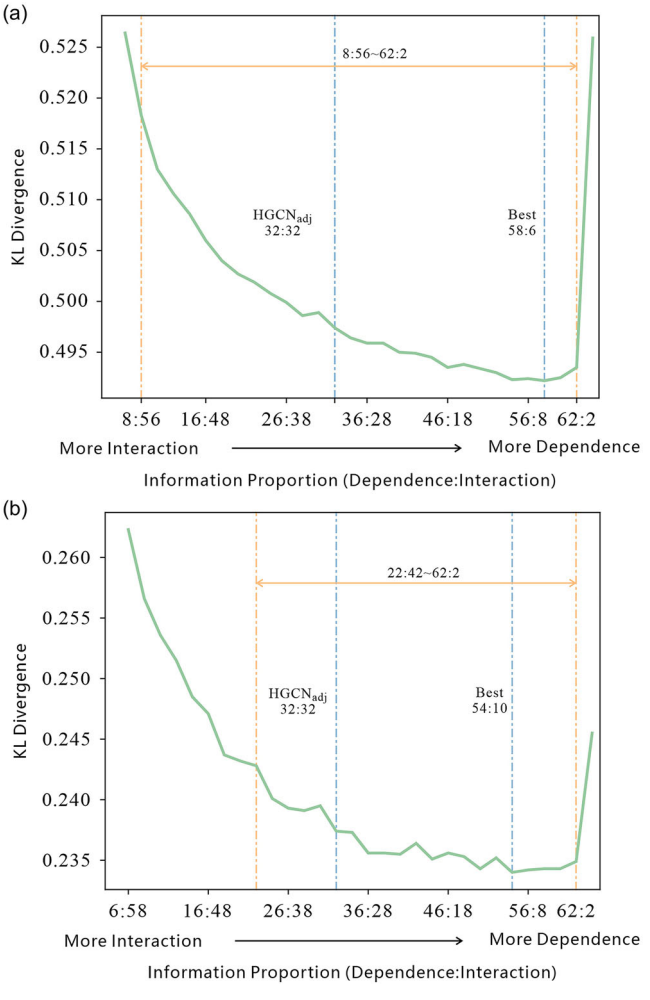


Figure 9. Spatial dependence vs. spatial interaction in the HGCN$_{adj}$ model for the Shenzhen (a) and London (b) datasets.

dependence features (Group C3) exhibit remarkable performance improvements. This indicates that spatial interaction features have a distinctly crucial role in contributing to urban land-use inference compared to that of local attributes and spatial dependence features.

Second, the results reported in Section 4 indicate that the local attribute features of spatial units have a certain predictive power in the urban land-use inference task, which has been recognized in the literature (Jiang *et al*. 2015, Gao *et al*. 2017, Yan *et al*. 2017, Zhai *et al*. 2019, Niu and Silva 2021). From a modeling perspective, machine learning models (HGI, Semantic, and Doc2Vec) exhibit general superiority over statistical methods, confirming the known advantages of machine learning, such as high nonlinearity, nonparametric specification, and flexible functional forms (Fischer and Gopal 1994). The results also indicate that the best-performing GCN-based model considering spatial dependence (GCN_{adj}) is superior to the best-performing model considering only local attribute features (Doc2vec), although the former requires an adequate neighborhood configuration (adjacency in this scenario) to optimize the capture of spatial dependence features. Owing to the incompleteness of POI-based features, land-use inference can benefit from aggregating local features of spatial neighborhoods to compensate for the missing local attributes in one spatial unit. However, it is not guaranteed that the GCN mechanism can automatically capture the spatial dependence between the local features of neighboring units, and GCN-based models (e.g. GCN_{dd}) may otherwise perform poorer than local attribute models.

In addition, when holding the same spatial interaction features constant, the HGCN-based models in Group C3—incorporating spatial dependence features—exhibit superior performance over their counterpart HGCN-based models in Group C2, which incorporate local attributes. This is consistent with the finding that the best-performing spatial dependence model (GCN_{adj}) outperforms the best-performing local attribute model (Doc2Vec).

Q2: *Are the proposed HGCN-based models superior to the GCN-based models in terms of capturing spatial interaction features?*

The HGCN-based models in Groups C1, C2, and C3 consistently achieve better performances than their GCN-based counterparts for both the Shenzhen and London datasets. This indicates that HGCN-based models have a superior capability over GCN-based models for extracting spatial interaction features, regardless of whether they are used alone or with the consideration of local attributes or spatial dependence features as the control factor. This is expected because the former captures both demand and supply components through the HGCN mechanism tailored for that purpose, whereas the latter has very limited capability to do so. Considering the same flow data containing spatial interaction information, the simple message-passing mechanism of the GCN cannot explicitly separate the demand and supply components, leaving the true value of the flow data unexploited. Without careful consideration of what should be captured from the spatial interaction data, it is highly possible to deteriorate model performance by simply feeding GCN-based models with spatial interaction data.

5.2. Effects of spatial interaction versus spatial dependence

Table 3 shows that HGCN performs poorer than GCN_{adj}, which may indicate that spatial interaction features alone (HGCN) contribute less than adequately captured spatial dependence features (GCN_{adj}) in land-use inference. This indicates how spatial interaction and spatial dependence interact and how their relationship affects their relative significance in contributing to urban land-use inference.

Notably, the $HGCN_{adj}$ model, the best-performing model in all experiments, has equally divided vector dimensions representing spatial interaction and dependence, that is, 32 dimensions representing spatial interaction and 32 dimensions representing spatial dependence. To disentangle the interrelationship between spatial interaction and spatial dependence and their relative contributions to land-use inference, we varied the number of vector dimensions for each type of feature, indicated by a dimension ratio, while keeping the total number of dimensions for both features constant at 64. Specifically, we began from spatial interaction with 64 dimensions (whereas spatial dependence had zero dimensions) and continued to reduce the dimension used by spatial interaction to 0 (the proportion for spatial dependence increases to 100%, when the model becomes GCN_{adj}).

Figure 9 shows that as the ratio of dimensions of the two features changes from spatial interaction dominance to spatial dependence dominance, the performance curve shows a 'U' shape for Shenzhen data (a similar pattern is shown for London data). The blue vertical line at 32:32 represents the $HGCN_{adj}$ model. The two ends of the curve indicate the models with standalone spatial interaction or dependence features, showing performances of 0.6946 versus 0.5259 on the Shenzhen data and 0.2791 versus 0.2455 on the London data, respectively, in terms of KL divergence. This indicates that the model using spatial dependence features alone has more predictive power than the model using spatial interaction features alone. Once both types of features are adequately used, specifically in a certain range of ratios (8:56–62:2 for Shenzhen data and 22:42–62:2 for London data) indicated by the orange vertical lines, the models generally achieve better performance than those using either type of feature alone. This indicates that with adequate configurations, spatial interaction and dependence features are complementary rather than competitive in urban land-use inference, although they contribute differently to the task.

The next question is based on how they contribute differently to the tasks. Both U curves are highly unbalanced. As the dimension of spatial interaction features decreases and is replaced by increasing the dimensions of spatial dependence features, the model performance continues to improve, indicating that the model performance improvement by adding one dimension of spatial dependence features is larger than the performance loss due to losing one dimension of spatial interaction features. This scenario continues until the ratio of the dimensions reaches a certain point (58:6 for the Shenzhen dataset and 54:10 for the London dataset), where the model performance is the highest. After the optimal point, the model performance improvement from increasing the dimension of the spatial dependence features is smaller than the performance loss from decreasing the dimension of the spatial interaction features; thus, the total model performance begins decreasing.

Two aspects require special attention. First, the optimal points for both datasets indicate that spatial dependence features dominate the combined feature information of spatial interaction and spatial dependence, contributing to the optimal performance of $HGCN_{adj}$. Second, the model performance deteriorates rapidly when the ratio of the dimensions moves to the right side of the optimal point. Although spatial interaction features use only approximately 10% of the combined feature information (six of sixty-four vector dimensions), they play a key role in boosting the model performance to a significant level. This indicates a critical threshold for the number of dimensions accommodating spatial

interaction features, such that there will be a significant loss in model performance if the amount of information for spatial interaction features cannot meet the threshold.

5.3. Generalizability of HGCN-based models

5.3.1. Model generalizability in the same study area

To examine the generalizability of the HGCN-based models, we varied the amount of data for training and evaluated the trained models on the remaining data. Specifically, the size of the training data varied from 10% to 90% of the total data, with 10% internal, and each experiment was repeated for 30 iterations. This also allowed us to assess the generalizability of the proposed model trained from a small sample of randomly distributed locations to other locations in the same study area.

When the size of the training set was varied, the ranking of the performances of all models remained constant for both study areas (Figure 10). Specifically, the performances of the HGCN-based models were consistently better than those of their counterparts. The generalizability of most models increased with the size of the training data (except for GCN_{dd} with the Shenzhen data). Notably, for the Shenzhen data, the L + HGCN model, trained with only 20% (less than 40% for the London data) of the data, outperformed the Doc2Vec model trained with 80–90% (90% for the London data) of the data. In addition, the $HGCN_{adj}$ (GCN_{adj}+HGCN) model, trained with 30% (less than 60% for the London data) of the data, outperformed the GCN_{adj} model trained with 90% (90% for the London data) of the data. This highlights the finding that, in addition to spatial dependence and local attribute features, incorporating spatial interaction features well-captured by the HGCN mechanism significantly enhances the generalizability of models, even on small training samples.

5.3.2. Model transferability in a different spatial context

Transferability is a hot topic in geospatial artificial intelligence (Skobalski *et al*. 2024, Wang *et al*. 2024). The transferability of models to different spatial contexts is one of the most crucial indicators of model generalizability. However, there have been few successful efforts in this field, and most studies have dedicated their models to one study area and have not been able to transfer their models because of the spatial heterogeneity between different study areas. This section aims to assess the transferability of the best-performing HGCN-based model from previous experiments, that is, $HGCN_{adj}$ trained on Shenzhen data, to the study area of London. As the two study areas have vastly different POI data, spatial units, and flow data types, zero-shot learning is unreasonable (Garcia and Bruna 2017). Therefore, we transferred the pre-trained model on the Shenzhen dataset to London and adopted few-shot fine-tuning on the pre-trained model using training sets of varying sizes (from 10 to 90%) from the London dataset, which is the transferred model shown in Table 8. As a baseline, the same type of model trained entirely on the London dataset was used, which is the original model listed in Table 8.

Table 8 shows that the transferred model outperformed the original model in all experiments, while the performance of both models increased with the size of the training data. Notably, the transferred model fine-tuned on 40% of the London dataset

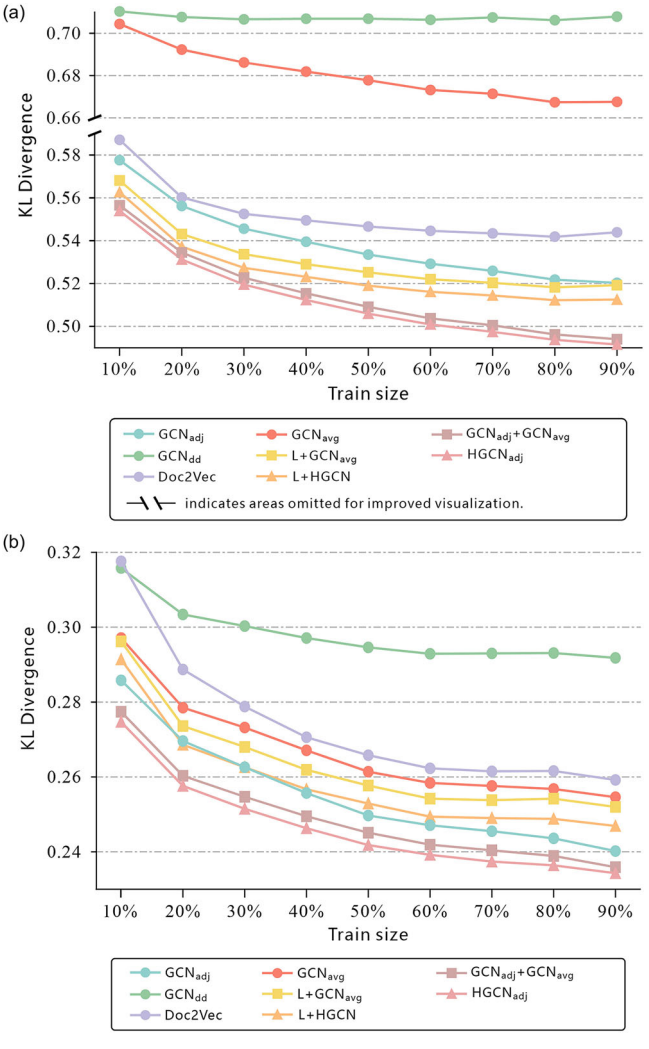


Figure 10. Model generalizability in the study areas: Shenzhen (a) and London (b).

Table 8. Model transferability from Shenzhen to London.

	KL[Loss]↓	
Train size in London	Transferred model	Original model
10%	0.2714	0.2747
20%	0.2544	0.2576
30%	0.2416	0.2515
40%	0.2332	0.2463
50%	0.2298	0.2418
60%	0.2238	0.2392
70%	0.2135	0.2374
80%	0.2143	0.2364
90%	0.2114	0.2343

performed even better than the original model trained on 90% of the same dataset, which reveals the high transferability of the proposed model pre-trained on the Shenzhen dataset. This finding is reasonable, as the Shenzhen dataset covering larger areas consists of more diversified features, such as spatial interaction flows in Shenzhen including all types of travel modes rather than those including only a single mode, biking, in London. The high transferability of models has a significant practical implication, which is that transferable models are highly preferable for regions where data are scarce, enabling land-use planning applications that are otherwise impossible.

6. Conclusions and future research

In the context of urban functions/land uses, this study conceptually differentiates spatial interaction and spatial dependence and argues that spatial interaction features derived from spatial interaction data must be recognized, contributing differently from local attributes and spatial dependence features in the task of urban land-use inference. To address the deficiencies of existing approaches in modeling spatial interactions for land-use inference, this study proposes a novel framework, HGCN, which explicitly accounts for the spatial demand and supply components embedded in spatial interaction data. The proposed framework was tested through a series of experiments involving datasets from Shenzhen and London.

First, the results revealed that spatial interaction features play a distinctly crucial role in contributing to urban land-use inference compared to that of local attributes and spatial dependence features. Therefore, in addition to incorporating the latter two types of features in urban function modeling, it is imperative to consider the spatial interaction characteristics conceptualized within the scope of this study. Second, HGCN-based models exhibit superior performance over GCN-based models in extracting spatial interaction features, regardless of whether they are used alone or with the consideration of local attributes or spatial dependence features as control factors. Specifically, heterogeneous representations were learned to capture the multimodal nature of places (as nodes) as demanders or suppliers of urban functions and the asymmetry of spatial interactions (as meta-paths) between places from a flow-production or flow-attraction perspective. This is attributable to an innovative accessibility-based operator informed by spatial interaction theory, which aids the explanatory power of the model. In summary, the HGCN framework can distinguish between functional complementarity and functional similarity, which are the underlying mechanisms of spatial interaction and spatial dependence, respectively. In addition, it has the capacity to incorporate both features for urban land-use inference.

Notably, the HGCN model, combining both spatial dependence and interaction features, achieved the highest performance among all models, and spatial interaction features play a key role in boosting model performance to a significant level, even if it uses only approximately 10% of the combined feature information. It was also highlighted that the HGCN model significantly enhanced the generalizability of models, even for small training samples in the same study area. Moreover, a remarkable finding revealed the high transferability of the HGCN model pre-trained on the Shenzhen dataset to the London dataset; the transferred model fine-tuned on 40% of the London

dataset performed even better than the original model trained on 90% of the London dataset. This spatial generalizability indicates that the proposed model framework has a potential to become an approach to building urban function foundation models. Besides, model transferability has substantial practical implications for regions where land use data are scarce. It can be compensated by land uses inferred by the transferable models, enabling land-use planning applications that are otherwise impossible.

There is scope for further improvement of the proposed models. Possible future studies are as follows: The effects of spatial interaction versus spatial dependence in Section 5.2 indicate a critical threshold for the dimension of spatial interaction features, such that there is a great loss in model performance if the amount of information for spatial interaction features cannot meet the threshold. The investigation of this threshold will help optimize the two sets of features in terms of balancing spatial interactions and spatial dependencies in modeling urban functions. In addition, it is crucial to incorporate the temporal dynamics of spatial interactions by extending the proposed framework.

Acknowledgments

The authors would like to express sincere thanks to the Editor-in-Chief, Prof. May Yuan and all anonymous reviewers for their insightful comments and suggestions, which have helped significantly enhance the quality of this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Shenzhen Science and Technology Program [KQTD20221101093604016]; the Guangdong Basic and Applied Basic Research Foundation [2024A1515011924]; the Shenzhen Science and Technology Program [JCYJ20220818100810024]; the National Natural Science Foundation of China [41925003, 42130402]; and the Shenzhen Science and Technology Program [RCBS20221008093330064].

Notes on contributors

Zhaoya Gong is an assistant professor at the School of Urban Planning and Design, Peking University Shenzhen Graduate School, Shenzhen, China. His research interests include geospatial artificial intelligence, urban data science and big data analytics. He contributed to the conceptualization, resources, methodology, formal analysis, investigation, validation, funding acquisition, projection administration, supervision, writing-original draft, writing-review & editing.

Chenglong Wang is a master student at the School of Urban Planning and Design, Peking University Shenzhen Graduate School, Shenzhen, China. His research interests include geospatial artificial intelligence, urban computing, and cartography. He contributed to the conceptualization, methodology, software, formal analysis, investigation, validation, visualization, writing-original draft, writing-review & editing.

Yuting Chen is a postdoctor at the School of Urban Planning and Design, Peking University Shenzhen Graduate School, Shenzhen, China. Her research interests include geospatial artificial

intelligence and human mobility. She contributed to the conceptualization, data curation, resources, investigation, writing-review & editing.

Bin Liu is a master student at the School of Urban Planning and Design, Peking University Shenzhen Graduate School, Shenzhen, China. Her research interests include geospatial artificial intelligence and human mobility. She contributed to the investigation, validation, visualization, writing-original draft, writing-review & editing.

Pengjun Zhao is a full professor and the dean of School of Urban Planning and Design of Peking University. His research interests include spatial planning, and sustainable transportation. He contributed to the conceptualization, resources, funding acquisition, projection administration, supervision, writing-review & editing.

Zhengzi Zhou was a research assistant at the School of Urban Planning and Design, Peking University Shenzhen Graduate School, Shenzhen, China. His research interests include geospatial artificial intelligence and human mobility. He contributed to the data curation, resources, formal analysis, visualization, writing-review & editing.

Data and codes availability statement

The data, codes, and instructions that support the findings of this study are available with the identifier (https://github.com/SupdGeoAI/HGCN). The mobile phone data purchased from China UniCom and the official land-use survey data for Shenzhen cannot be made publicly available, because the former is subject to the commercial restriction and the latter is subject to the legal restriction in China.

References

Alonso, W., 1960. *A model of the urban land market: location and densities of dwellings and businesses*. Philadelphia: University of Pennsylvania.

Anselin, L., 2013. *Spatial econometrics: methods and models* (Vol. 4). Dorddrecht, The Netherlands: Kluwer Academic Publishers.

Batty, M., 2008. The size, scale, and shape of cities. *Science (New York, N.Y.)*, 319 (5864), 769–771.

Cha, S.H., 2007. Comprehensive survey on distance/similarity measures between probability density functions. *City*, 1 (2), 1.

Chen, B., Xu, B., and Gong, P., 2021. Mapping essential urban land use categories (EULUC) using geospatial big data: Progress, challenges, and opportunities. *Big Earth Data*, 5 (3), 410–441.

Cheng, J., *et al.*, 2006. Urban land administration and planning in China: Opportunities and constraints of spatial data models. *Land Use Policy*, 23 (4), 604–616.

Crooks, A., *et al.*, 2015. Crowdsourcing urban form and function. *International Journal of Geographical Information Science*, 29 (5), 720–741.

Deng, Z., *et al.*, 2022. Identification of urban functional zones based on the spatial specificity of online car-hailing traffic cycle. *ISPRS International Journal of Geo-Information*, 11 (8), 435.

Fischer, M.M., and Gopal, S., 1994. Artificial neural networks: a new approach to modeling interregional telecommunication flows. *Journal of Regional Science*, 34 (4), 503–527.

Fresco, L. O., Stroosnijder, L., Bouma, J., and Keulen, H. V. (Eds.). 1994. *The future of the land: mobilising and integrating knowledge for land use options*. Chichester: Wiley, xviii+–409.

Gao, S., Janowicz, K., and Couclelis, H., 2017. Extracting urban functional regions from points of interest and human activities on location-based social networks. *Transactions in GIS*, 21 (3), 446–467.

Garcia, V., and Bruna, J., 2017. Few-shot learning with graph neural networks. arXiv preprint arXiv:1711.04043.

Geng, X., 2016. Label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 28 (7), 1734–1748.

Gjoka, M., *et al.*, 2011. Multigraph sampling of online social networks. *IEEE Journal on Selected Areas in Communications*, 29 (9), 1893–1905.
Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69 (4), 211–221.
Guan, X., *et al.*, 2023. HGAT-VCA: integrating high-order graph attention network with vector cellular automata for urban growth simulation. *Computers, Environment and Urban Systems*, 99, 101900.
Hansen, W.G., 1959. How accessibility shapes land use. *Journal of the American Institute of Planners*, 25 (2), 73–76.
Huang, W., *et al.*, 2022. Estimating urban functional distributions with semantics preserved POI embedding. *International Journal of Geographical Information Science*, 36 (10), 1905–1930.
Huang, W., *et al.*, 2023. Learning urban region representations with POIs and hierarchical graph infomax. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196, 134–145.
Hunt, J.D., Kriger, D.S., and Miller, E.J., 2005. Current operational urban land-use–transport modelling frameworks: a review. *Transport Reviews*, 25 (3), 329–376.
Jiang, S., *et al.*, 2015. Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems*, 53, 36–46.
Ju, J., Kolaczyk, E.D., and Gopal, S., 2003. Gaussian mixture discriminant analysis and sub-pixel land cover characterization in remote sensing. *Remote Sensing of Environment*, 84 (4), 550–560.
Kipf, T.N., and Welling, M., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.
Krizhevsky, A., Sutskever, I., and Hinton, G., 2012. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
Lambin, E. F., and Geist, H. J. (Eds.). 2008. *Land-use and land-cover change: local processes and global impacts*. Berlin, Heidelberg: Springer Science & Business Media.
Le, Q., and Mikolov, T., 2014. Distributed representations of sentences and documents. International conference on machine learning. PMLR, 1188–1196.
Liu, X., *et al.*, 2016. Incorporating spatial interaction patterns in classifying and understanding urban land use. *International Journal of Geographical Information Science*, 30 (2), 334–350.
Ma, L., *et al.*, 2019. Deep learning in remote sensing applications: a meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166–177.
Mikolov, T., *et al.*, 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
Niu, H., and Silva, E.A., 2021. Delineating urban functional use from points of interest data with neural network embedding: a case study in Greater London. *Computers, Environment and Urban Systems*, 88, 101651.
Parker, D.C., *et al.*, 2003. Multi-agent systems for the simulation of land-use and land-cover change: a review. *Annals of the Association of American Geographers*, 93 (2), 314–337.
Prieto Curiel, R., *et al.*, 2021. The heartbeat of the city. *PLoS One*, 16 (2), e0246714.
Reades, J., Souza, D. J., and Hubbard, P., 2019. Understanding urban gentrification through machine learning. *Urban Studies*, 56 (5), 922–942.
Seto, K.C., and Fragkias, M., 2005. Quantifying spatiotemporal patterns of urban land-use change in four cities of China with time series landscape metrics. *Landscape Ecology*, 20 (7), 871–888.
Skobalski, J., *et al.*, 2024. Bridging the gap between crop breeding and GeoAI: soybean yield prediction from multispectral UAV images with transfer learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 210, 260–281.
Sun, Y., *et al.*, 2011. Pathsim: meta path-based top-k similarity search in heterogeneous information networks. *Proceedings of the VLDB Endowment*, 4 (11), 992–1003.
Ullman, E.L., 1954. Amenities as a factor in regional growth. *Geographical Review*, 44 (1), 119–132.
Vaswani, A., *et al.*, 2017. Attention is all you need. *In*: *Advances in neural information processing systems*. Long Beach, CA: Long Beach Convention & Entertainment Center, 5998–6008.

Viegas, J.M., Martinez, L.M., and Silva, E.A., 2009. Effects of the modifiable areal unit problem on the delineation of traffic analysis zones. *Environment and Planning B: Planning and Design*, 36 (4), 625–643.

Von Thünen, J.H., 2018. The isolated state. *In: The economics of population*. Milton Park, Abingdon, Oxfordshire: Routledge, 211–216.

Wang, C., *et al.*, 2024. TransMI: a transfer-learning method for generalized map information evaluation. *Cartography and Geographic Information Science*, 1–17. doi: 10.1080/15230406.2024.2306827.

Wang, F., 2021. From 2SFCA to i2SFCA: integration, derivation and validation. *International Journal of Geographical Information Science: IJGIS*, 35 (3), 628–638.

Wang, J., 2016. Economic geography: spatial interaction. *In: International encyclopedia of geography*. Hoboken, NJ: John Wiley & Sons, Ltd., 1–4.

Xu, Y., *et al.*, 2022. Application of a graph convolutional network with visual and semantic features to classify urban scenes. *International Journal of Geographical Information Science*, 36 (10), 2009–2034.

Yan, B., *et al.*, 2017. November). From itdl to place2vec: reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. Proceedings of the 25th ACM SIGSPATIAL international conference on advances in geographic information systems,1–10.

Yao, Y., *et al.*, 2017. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *International Journal of Geographical Information Science*, 31 (4), 825–848.

Yuan, J., Zheng, Y., and Xie, X., 2012. Discovering regions of different functions in a city using human mobility and POIs. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 186–194.

Zhai, W., *et al.*, 2019. Beyond Word2vec: an approach for urban functional region extraction and identification by combining Place2vec and POIs. *Computers, Environment and Urban Systems*, 74, 1–12.

Zhang, C., *et al.*, 2018. An object-based convolutional neural network (OCNN) for urban land use classification. *Remote Sensing of Environment*, 216, 57–70.

Zhong, M., *et al.*, 2022. Advances in integrated land use transport modeling. *In: Advances in transport policy and planning* (Vol. 9). Amsterdam, Netherlands: Academic Press, 201–230.

Zhu, D., *et al.*, 2020. Understanding place characteristics in geographic contexts through graph convolutional neural networks. *Annals of the American Association of Geographers*, 110 (2), 408–420.

Appendix A. Delineation of spatial units in the study area of London

To delineate the spatial units, the selection of a suitable geographical scale is the first challenge for quantitative spatial analysis. Different spatial analysis units are usually chosen for different research objects. Studies on regional objects are commonly based on statistical data, so they use census units as spatial units. For example, Reades *et al.* (2019) used lower layer super output area (LSOA) as the spatial unit that reflects spatial variations and is statistically robust to analyze neighborhood change in London. Niu and Silva (2021) measured the richness and diversity of POI classes for three census units (including OA, LSOA, and MSOA) and used LSOA as the optimal spatial unit to analyze urban functional use. On the other hand, the dividing lines between different communities in a city are often main roads and transport stations that are also crucial nodes of urban mobility. In studies focusing on spatial mobility, census or grid cells tend to divide areas with the same spatial interaction characteristics into different observation units, so the layout of public transport systems is often used to divide cities into tiles (Prieto Curiel *et al.* 2021).

Two strategies are evident for research that, as in this study, target both spatial flows and regional objects. One is region-based, where flow data are aggregated into census units, i.e. all

flows within a census unit are aggregated. The other is flow-based, i.e. aggregating regional features around the origin and destination points according to where the flow occurs. Both approaches are theoretically feasible, and the actual choice depends on the resolution of the data. High-precision data can be aggregated into lower-resolution units. For example, when we have flow data with GPS coordinates and statistics data based on census units, it is more appropriate to aggregate flows to census units. This case study uses POI and spatial flow data from bicycle systems. Since the resolution of the flow data is relatively low, we choose to use the flow-based strategy of determining spatial units.

Existing flow-based strategies for determining spatial units generally construct Voronoi diagrams based on the collection of public transport stations, thus partitioning the space into units (Prieto Curiel *et al*. 2021). It is worth noting that the shape of the spatial unit affects the result of aggregating spatial phenomena into regions, i.e. the modifiable areal unit problem (MAUP) will exist not only because of the size (scale effect) but also because of the shape (zoning effect) of the spatial unit (Viegas *et al*. 2009; Batty 2008). However, in traditional Voronoi diagrams,

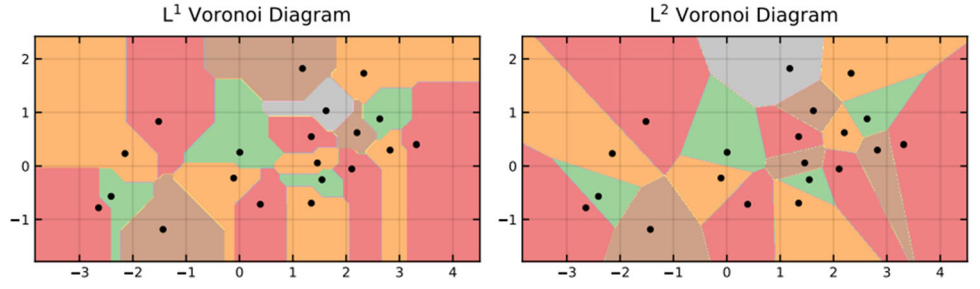


Figure A1. A demonstration of Voronoi diagrams using Euclidean distance (Right) and Manhattan distance (Left).

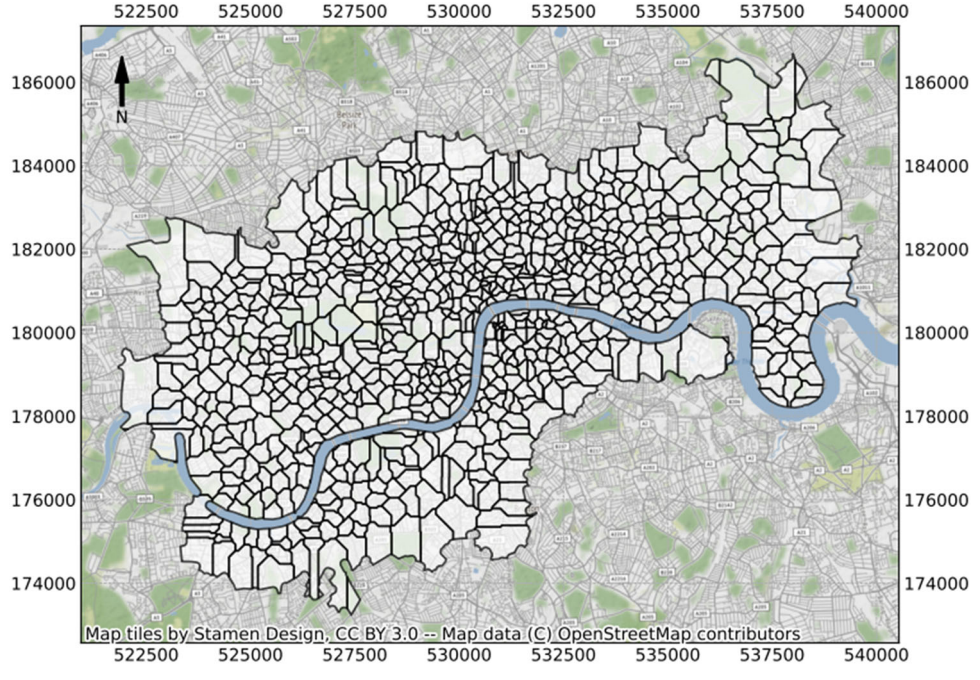


Figure A2. Spatial unit of London's study area.

Table A1. Model performance with London data.

GP	Models	Evaluation Measures			Rank
		KL[Loss]↓	L1↓	Cosine↑	
A	Doc2Vec (L)	$0.2615_{\pm 0.0154}$	$0.0597_{\pm 0.0018}$	$0.8989_{\pm 0.0070}$	14
B	GCN_{dd}	$0.2930_{\pm 0.0163}$	$0.0634_{\pm 0.0016}$	$0.8888_{\pm 0.0070}$	15
	GCN_{adj}	$0.2455_{\pm 0.0150}$	$0.0568_{\pm 0.0018}$	$0.9074_{\pm 0.0067}$	5
C1	GCN_{avg}	$0.2576_{\pm 0.0152}$	$0.0601_{\pm 0.0017}$	$0.8999_{\pm 0.0067}$	11
	GCN_{in}	$0.2613_{\pm 0.0154}$	$0.0606_{\pm 0.0018}$	$0.8980_{\pm 0.0069}$	13
	GCN_{out}	$0.2580_{\pm 0.0145}$	$0.0603_{\pm 0.0016}$	$0.8995_{\pm 0.0066}$	12
	HGCN	$0.2566_{\pm 0.0127}$	$0.0601_{\pm 0.0017}$	$0.9007_{\pm 0.0053}$	10
C2	$L + GCN_{avg}$	$0.2539_{\pm 0.0166}$	$0.0585_{\pm 0.0019}$	$0.9033_{\pm 0.0075}$	8
	$L + GCN_{in}$	$0.2543_{\pm 0.0167}$	$0.0586_{\pm 0.0019}$	$0.9030_{\pm 0.0075}$	9
	$L + GCN_{out}$	$0.2538_{\pm 0.0166}$	$0.0585_{\pm 0.0019}$	$0.9033_{\pm 0.0074}$	7
	L + HGCN	$0.2490_{\pm 0.0130}$	$0.0581_{\pm 0.0017}$	$0.9054_{\pm 0.0059}$	6
C3	$GCN_{adj}+GCN_{avg}$	$0.2407_{\pm 0.0147}$	$0.0562_{\pm 0.0019}$	$0.9103_{\pm 0.0069}$	3
	$GCN_{adj}+GCN_{in}$	$0.2412_{\pm 0.0147}$	$0.0563_{\pm 0.0019}$	$0.9102_{\pm 0.0069}$	4
	$GCN_{adj}+GCN_{out}$	$0.2404_{\pm 0.0147}$	$0.0562_{\pm 0.0018}$	$0.9104_{\pm 0.0068}$	2
	$HGCN_{adj}$	$0.2374_{\pm 0.0131}$	$0.0558_{\pm 0.0017}$	$0.9120_{\pm 0.0060}$	1

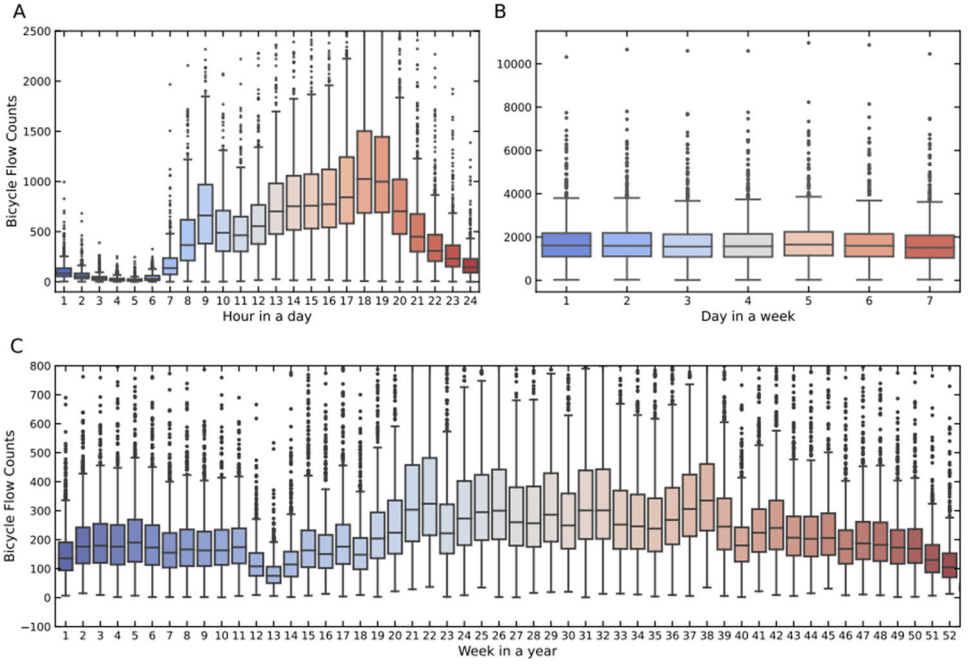


Figure A3. Aggregated flow counts in the hour of the day (a), the day of the week (b), and the week of the year (c).

narrow acute triangular units are often produced due to the use of Euclidean metrics as distances, and this type of partitioning provides a rigid segmentation of the urban spaces. In this paper, we introduce a Manhattan distance-based Voronoi partitioning approach that takes into account the urban morphology while ensuring the regularity of the shape of the spatial units. The comparison of space partition based on L1 metrics (Manhattan distance) and L2 metrics (Euclidean distance) is demonstrated in Figure A1.

Combining the above considerations, our procedure for constructing the spatial analysis unit is as follows. First, we generate spatial units by performing L1 Voronoi diagrams for bicycle parking locations. Then the name of each parking location is used to label the unit. Finally, the

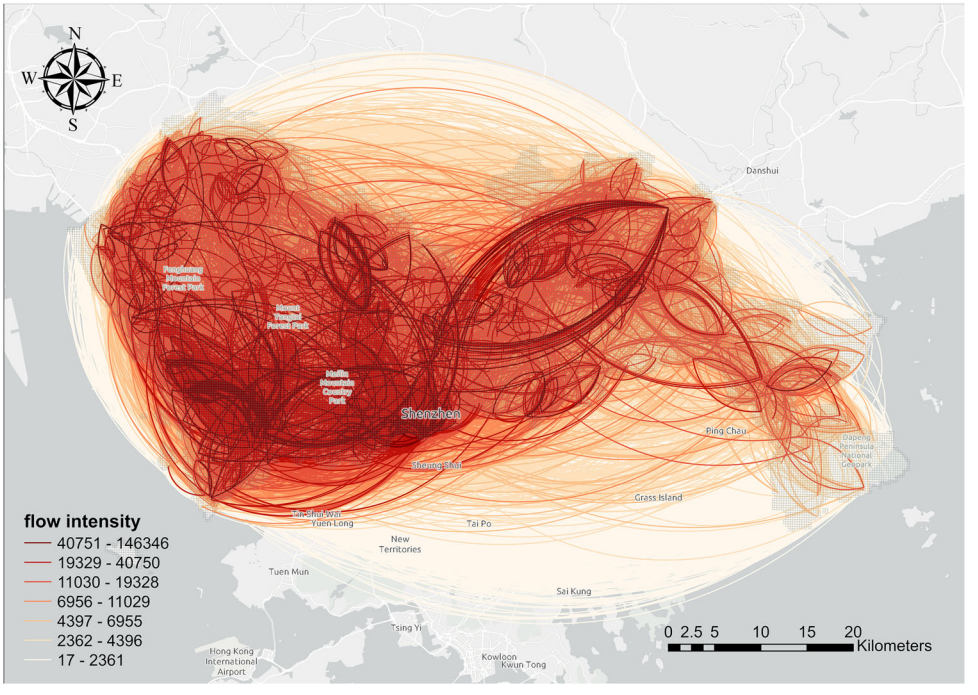


Figure A4. An origin-destination (OD) flow visualization for Shenzhen.

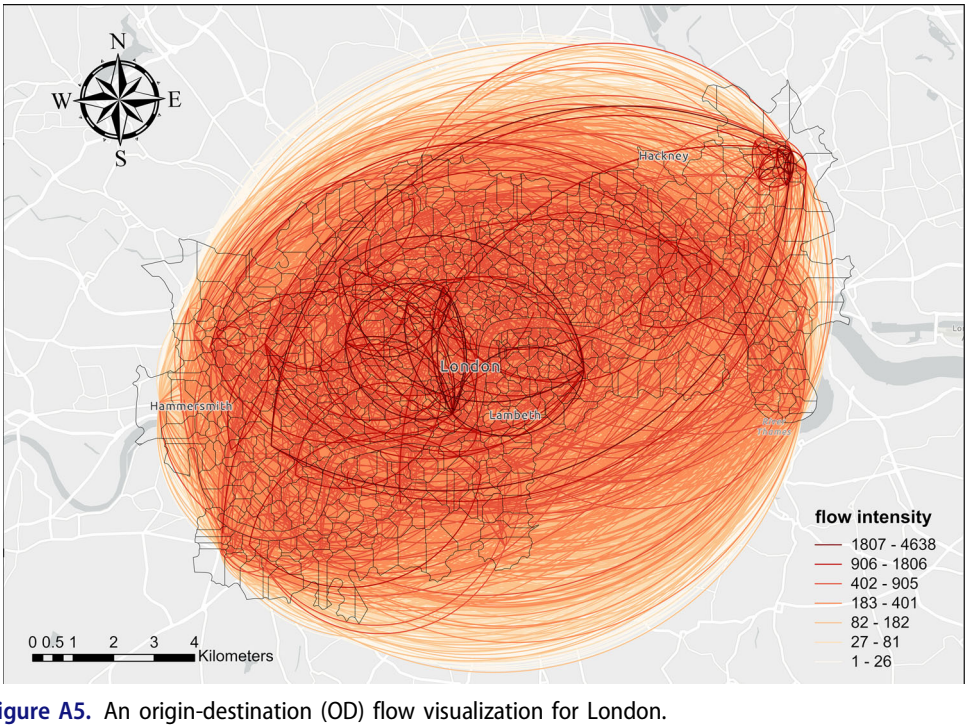


Figure A5. An origin-destination (OD) flow visualization for London.

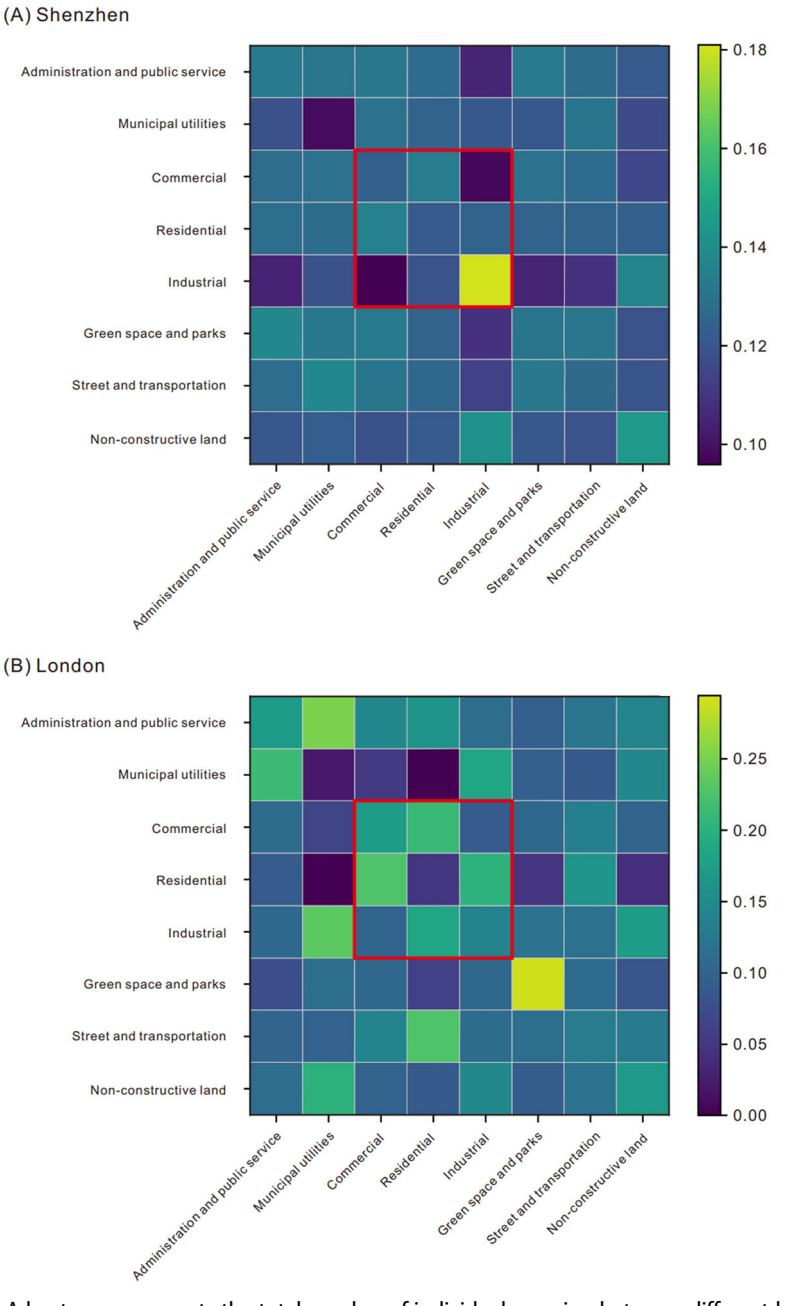


Figure A6. A heatmap represents the total number of individuals moving between different land uses.

land functional use data and POI data are divided into each spatial unit. For the land-use data, we calculate the proportion of each type of functional land use in each spatial unit. The study area and spatial units we finally determined are shown in Figure A2.

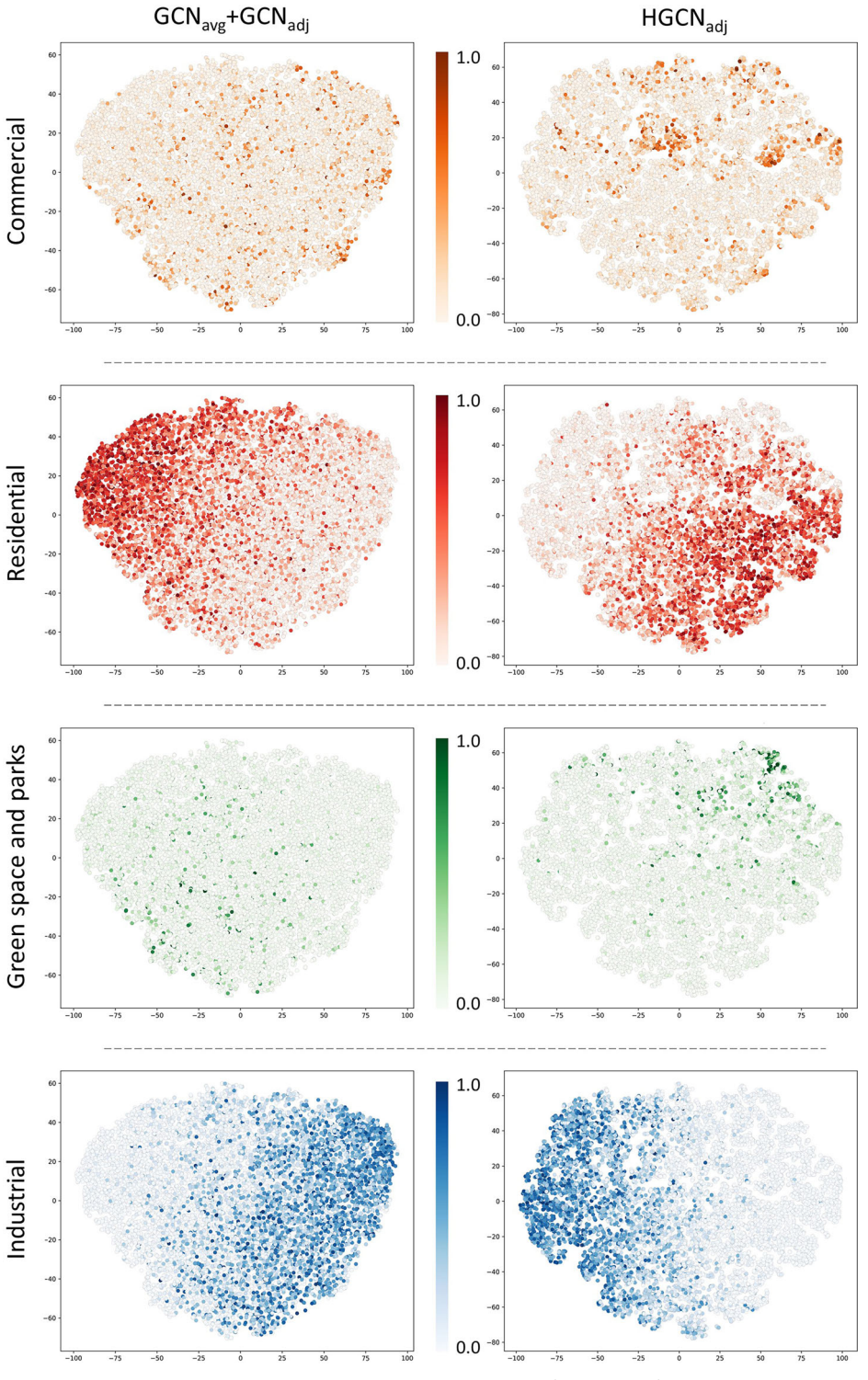


Figure A7. Dimensionality reduction and visualization on the features before input into the classifier for both $HGCN_{adj}$ and GCN_{adj}+GCN_{avg}.

Appendix B. The validity of the bike flows dataset in London

To test the data's validity, we examined this dataset's trend characteristics. The data were aggregated to the week of the year, the day of the week, and the hour of the day to show the flows' weekly, daily, and hourly patterns. As shown in Figure A3(a), we see that this dataset has an evident hourly pattern, with a pellucid morning and evening peak; the daily pattern of the data (Figure A3(b)) is not significant, with no notable differences in flows each day of the week; As seen in the weekly pattern (Figure A3(c)), week 12 and 39 have two clear downward trends, which correspond to the lockdown in London. The seasonal component can be seen as a validation of mobility patterns, which are expected to be stable and remain constant over time.

Appendix C. Detailed elaboration on the raw spatial interaction data

The visualization in Figures A4–A5 illustrates the Origin-Destination (OD) flow between two cities, showcasing an uneven spatial distribution characterized by long-distance connections. This data type exhibits heterogeneity, diverging from mere spatial proximity. Moreover, as depicted in Figure A6, we designate the land-use type with the highest proportion within each spatial unit as its dominant land-use type. Subsequently, we aggregate all flows based on these dominant land-use types into an 8×8 matrix. Here, each (i,j) entry represents the total number of individuals moving from land-use type i to land-use type j. Following normalization, we present the results in the form of a heatmap. Several noteworthy phenomena emerge: (1) There are strong connections between commercial and residential areas, surpassing connections between areas of the same type (commercial-commercial or residential-residential). This highlights the complementary functions of these areas; (2) Interactions between commercial and residential zones prevail over those between commercial and industrial zones. This suggests a tendency for residents to return home after shopping, rather than commuting to work from commercial areas; (3) Residential zones exhibit strong connections with both commercial and industrial areas, indicating multifaceted interactions. These underlying patterns of spatial interactions align closely with the assumptions of our modeling approach.

Appendix D. Visualization of the features before input into the classifier

To further illustrate the superior feature extraction capabilities of $HGCN_{adj}$, we applied t-SNE for dimensionality reduction and visualized the features before input into the classifier for both $HGCN_{adj}$ and $GCN_{adj}+GCN_{avg}$. In Figure A7, each dot represents a spatial unit, with color intensity indicating the proportion of commercial, residential, green space and parks, and industrial areas. The visualization reveals that the features extracted by $HGCN_{adj}$ are more clustered, highlighting the model's enhanced feature extraction abilities.