Computers, Environment and Urban Systems 117 (2025) 102246

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

From points to patterns: An explorative POI network study on urban functional distribution

Xuhui Lin [a], Tao Yang [b,*], Stephen Law [c,*]

[a] The Bartlett School of Sustainable Construction, University College London, London, UK
[b] School of Architecture, Tsinghua University, Beijing, China
[c] Department of Geography, University College London, London, UK

ARTICLE INFO

Keywords:
Urban analytics
Topic modelling
Urban function characterisation
Spatial network science
Spatial colocation mining

ABSTRACT

In the context of rapid urbanization, urban spaces not only accommodate a growing population but also produces complex socio-economic activities and cultural exchanges. Cities are complex systems, and conventional Points of Interest (POI) analysis methods, which usually assess the density and diversity of POIs in various neighbourhoods, often fails to capture this complexity. To address these limitations, this study introduces a novel approach by transforming POI sequences into words along streets and applying Latent Dirichlet Allocation (LDA) model to identify urban functional regions. Unlike traditional approaches that rely on subjective delineation of administrative boundaries, Voronoi cells or regular grids, our approach identifies street level functional areas that align more closely with human experience. Based on these functional topics, a multi-layered Poi-Topic network is then constructed to help better understand the roles specific POI plays within urban functional regions. This approach effectively distills the spatial distributional patterns of urban functions and provides a micro-level foundations for analysing the contextual interrelationships between POIs, thereby offering a more nuanced understanding of urban spaces. The effectiveness of the approach is demonstrated through the London case study. The results show that the proposed approach can effectively identify and delineate urban functional areas based on the co-occurrence patterns and network structure of POI vocabularies. The network centrality analysis further reveals the structural properties and interaction patterns, providing valuable insights into the roles and positions of different POI types in the functional organization of urban space. This method of using POI sequences and network analysis offers a new tool for urban planners, geospatial scientists, and policymakers, enabling them to understand and plan urban spaces with greater precision.

1. Introduction

As urbanization accelerates, cities will increasingly serve as a main stage for human social life (Jacobs, 1961).Urban spaces not only accommodate a growing population but also mirror complex socio-economic activities and cultural exchanges (Balland et al., 2020). In this context, Points of Interest (*POI*) data emerges as a critical geographic information resource, offering rich details for understanding urban places that support activities like jobs, amenities and entertainment (Long & Shen, 2015; Psyllidis et al., 2022; Yuan, Zheng, & Xie, 2012; Zhong, Huang, Arisona, Schmitt, & Batty, 2014). *POI* data reveals the location information of facilities such as shops, restaurants, schools, and hospitals, playing a key role in analysing urban functional layouts, optimizing urban resource allocation, enhancing urban management levels and sustaining an important part of our daily lives (Psyllidis et al., 2022). However, despite the recognized value of *POI* data, existing analysis methods exhibit significant limitations. Traditional *POI* analyses often rely on statistical approaches, characterizing urban spaces with the quantity, type, density and diversity of *POIs* within specific areas. Though intuitive, these methods typically fail to capture the interactions and contextual relationships between *POIs*, overlooking the complexity of urban spaces (Batty, 2013). For instance, an area densely linked by cafes, bookstores, and theatres may signify a culturally vibrant community, yet mere quantitative statistics fall short of revealing such urban functional connections. While spatial colocation mining (Chen, Chen, Liu, & Li, 2020; Yu, 2016) aims to capture the interactions between land use and amenities, it is often conducted at a single geographical scale, such as the city level, which can result in missing the

* Corresponding authors.
E-mail addresses: yangtao128@tsinghua.edu.cn (T. Yang), stephen.law@ucl.ac.uk (S. Law).

https://doi.org/10.1016/j.compenvurbsys.2024.102246
Received 17 July 2024; Received in revised form 12 December 2024; Accepted 20 December 2024
Available online 3 January 2025
0198-9715/

nuances within different urban functional regions of large metropolitan area.

To overcome these limitations, this study proposes a novel approach: transforming *POI* data into a textual format and applying Natural Language Processing (*NLP*) techniques and complex network science methods to mine the co-occurrences between *POIs* and the functional Topics of urban spaces. At the heart of this method is the conceptualization of urban spaces as a form of language, where *POIs* act as words that come together to form meaningful sentences or paragraphs, reflecting the functional and structural characteristics of urban spaces. Leveraging Latent Dirichlet allocation (*LDA*), latent Topics that represent different functional areas and activity patterns within the city can be identified from a vast amount of textualized *POI* data. This analysis of textualized *POI* data enables researchers to uncover details of urban functional areas, identifying commercial hotspots, residential quarters, industrial sectors, etc., and understanding how these areas are formed through the interactions and combinations of *POIs*. Furthermore, this method allows for the exploration of contextual relationships between *POIs*, such as cafe situated next to a library may reflect a community environment centred around learning and leisure. This in-depth analysis provides urban planners with insights, enabling them to comprehend more accurately how urban spaces are utilized thereby making more scientific and effective planning decisions. Furthermore, this research takes a step forward by integrating the *LDA* model with network science techniques (Hidalgo, Castaner, & Sevtsuk, 2020). By constructing multilayer networks based on the identified functional topics and their associated *POIs*, we can uncover the complex relationships and interaction patterns within different urban functional areas. The proposed approach not only reveals the semantic structure of urban space but also sheds light on the underlying organizational logic of urban functions.

The main contributions of this research are threefold. First, it introduces a novel perspective of textualizing POI data by incorporating *NLP* analysis at the street-level. Second, it integrates topic modelling with network analysis techniques to uncover the multi-dimensional relationships between POIs and urban functional areas, offering a more comprehensive and nuanced understanding of its spatial structure. Third, it conducts a detailed empirical analysis based on the Greater London in the UK, demonstrating the effectiveness of the proposed approach in deconstructing complex urban spatial structures for a large metropolitan region.

This paper is structured as follows: Section 2 provides a comprehensive review of previous research and practice relevant to the topic of this study. Section 3 presents the research framework. Section 4 presents a case study based on London, demonstrating the effectiveness of the proposed framework. Finally, the discussion and conclusions of this study are presented in Sections 5 and 6.

2. Literature review

2.1. Urban function characterisation

The study of urban functional characteristics is crucial for understanding and planning urban spaces (Crooks et al., 2015; Zhong et al., 2014). These characteristics reveal the types of activities and purposes within urban areas, providing essential information for urban planning, management, and development. Urban functional characteristics not only assist in identifying different areas within cities, such as commercial centres, residential areas, mixed use quarters, industrial zones, and recreational facilities, but are also closely related to urban sustainability, residents' quality of life, economic vitality, and social interactions (Liu & Long, 2016; Yao et al., 2017). Thus, the research on these characteristics is significant for optimizing urban resource allocation, improving residents' living standards, fostering economic growth, and enhancing urban social cohesion. Traditionally, the identification and analysis of urban functional characteristics have relied on macro-statistical data such as land use information, census information, and records of economic activities. These methods outline the city's basic functional layout by analysing land use patterns, population distribution, and types of economic activities. For example, land use data can show whether specific areas are designated for commercial, residential, or industrial purposes, while census data can provide information on population density, socio-economic characteristics of residential areas (Zhong, Zeng, Tu, & Yoshida, 2018) and mapping a city's geodemographic profiles (Singleton, Alexiou, & Savani, 2020). By integrating these data, planners and researchers can identify the city's main functional areas, serving as a basis for urban planning and policymaking. However, these traditional methods have limitations. These methods generally focus on macro-level analysis, potentially overlooking the complex interactions and functional diversity at the micro-level within urban spaces. To uncover deep structure and dynamic changes of urban functional characteristics, especially the subtle differences at the street or neighbourhood level, more fine-grained data and analytical methods are often required (Gao, Janowicz, & Couclelis, 2017; Liu & Long, 2016). With the development of big data and geographic information technology (Kitchin, 2014), new types of geographic data sources such as POI data (Liu & Long, 2016), social media data (Lansley & Longley, 2016), and mobile sensing data (Tao, Wang, Zhuo, & Li, 2019) offer new perspectives and methods for studying urban functional characteristics. These data sources provide richer and more real-time information on urban space usage, helping researchers capture the dynamic changes in urban activities at the micro-level (Gao et al., 2017). Through the analysis of these data, researchers can more precisely identify urban functional areas, understand how urban spaces are used, and how cities respond to different socio-economic activities' demands. However, effectively utilizing these new geographic data sources for urban functional characteristics analysis faces challenges in data processing and analysis methods. Traditional statistical and spatial analysis methods may not be suitable for handling large-scale, unstructured geographic data.

2.2. Emerging use of natural language processing in urban function characterisation

With recent advances in machine learning, there has been an increasing use of *NLP* on POI or social media data, employing supervised and unsupervised text mining methods to retrieve urban functions. For instance, a classifier was trained from Twitter data (based on TF-IDF) (Zhou & Zhang, 2016) to extract grid-level urban function density. Another study applied a variant of LDA (Blei, Ng, & Jordan, 2003), a widely accepted topic modelling method, on POI and mobile data in the Beijing area to discover large areas with different functions (Yuan et al., 2012). Similarly, LDA and cluster analysis on Twitter data were used to find large aggregated functional areas in London (Zhong et al., 2018), revealing the city's spatial structure at the administrative regional level. A framework was proposed to discover urban functional areas using LDA and K-Means clustering on *POI* and mobile data (Gao et al., 2017). Neural embedding methods have also been utilized to derive land-use function types. For example, CBOW-Word2Vec (Mikolov, Chen, Corrado, & Dean, 2013) was applied to project words within a transportation zone into an embedding space, which was then clustered (K-Means) to identify land-use types, later validated within designated land-use types in the Pearl River Delta. Liu et al.'s (Liu, Yin, Lu, & Mou, 2020) involved visualizing and exploring the semantic space of POI type embeddings through dimensionality reduction (T-SNE). Also, Yao et al. (Yao et al., 2017) integrated POIs with the Google Word2Vec model to analyse urban land use distribution, demonstrating high accuracy in land use classification. Xu et al. (Morioka, Okabe, Kwan, & McLafferty, 2022) combined POIs' spatial context with graph convolutional neural networks, improving urban land use classification by capturing complex spatial relationships among POIs. Niu & Silva (Xu et al., 2022) employed Doc2Vec to vectorize POIs and urban areas in London, enhancing the classification of urban functional areas through spatial context learning.

All these previous studies used arbitrary census tracts, transportation zones, regular grids, or plots/city blocks as units to identify urban functional areas.

2.3. Application of street perspective in urban functional representation

In urban planning and geographic information science, characterizing urban functions from a street-level perspective is gaining increasing attention (Niu & Silva, 2021; Yu, 2016). The underlying assumption of this geographic representation posits that streets are not merely physical components of urban spaces but are vibrant stages of urban life (Jacobs, 1961). Streets, or the linear spaces between buildings, form the fundamental units of urban structure and geography, serving both as a context and a foreground for human social activities, linking different places and people together (Gehl, 2010; Hillier & Hanson, 1984). Despite carrying the majority of daily activities, such as shopping, working, studying and entertaining, they are often marginalized in urban studies and planning, merely considered as containers for buildings or a context of the city, rather than as the key geographical units for understanding urban structure and functions.

The functional combination and spatial layout of streets significantly impact urban residents' quality of life and the city's overall image. Urban function characterisation at the street level analyses usage patterns and social activities in urban spaces in detail. Points of Interest (POI) data along streets, which include commercial facilities and public spaces, are crucial for understanding these functional layouts. Additionally, social media and mobile signal data provide rich insights into human activities, emotions, and movement patterns, offering a real-time view of how urban spaces are used from a street-level perspective.

Furthermore, the functional characteristics and activity levels can vary significantly between different sides of urban blocks. For example, one side of a street block in Central London might be bustling with shops and restaurants (as shown in Fig. 1), while the other side could be a quiet residential street. This heterogeneity highlights the importance to understand and plan urban functions from a street-level perspective. We argue in this research, that traditional geographical administrative units such as census tracts, plots, Voronoi cells or grids fails to address the usage of urban spaces and capture the perceptual experiences of residents (Law, 2017). Therefore, this research will map and analyse urban functional characteristics at the street level to better understand usage patterns, residents' behaviour, and the dynamic changes in urban functions.

2.4. Extending spatial co-location mining to urban functional clusters

Despite advancements in using machine learning for urban functional characterisation, there is still limited research on the interrelationship within each urban functional cluster. A related topic in geographic information systems is spatial co-location mining *SCM* of urban poi and amenities (Huang, Shekhar, & Xiong, 2004; Morioka et al., 2022). The goal of *SCM* is to identify co-located or co-occurred pairs or triplets (size-k) of amenities in a neighbourhood that are within some street network or Euclidean distances away (Yu, 2016).

Some studies cast these co-location pairs as an interaction graph. For example, (Chen et al., 2020) proposed a co-location mining method using a Voronoi graph as a neighbourhood to identify and compare.

POI co-location graphs across 25 Chinese cities. Another example is (Morioka et al., 2022), who proposed a street-based statistical methods to define a collocate-and-core pulse graph for a neighbourhood in Tokyo. In economic geography, based on the principles of relatedness, (Hidalgo et al., 2020) proposed an amenity co-location graph by analysing the pairwise correlations within an accessible neighbourhood for cities in the United States.

This research employs similar methods but focuses on examining the co-location subgraphs for each urban functional cluster, rather than studying the overall co-location network of amenities at the city-level. This approach will enable us to understand not only the general co-location pattern at the city-level but a more nuanced co-location mechanism at the functional area level. Additionally, we will extend this line of work by applying network science centralities on the co-location subgraphs of these different urban functional clusters to identify the most central and connected POI.

3. Methodology

This study proposes a novel analytical framework for uncovering urban functional patterns from POI data. As illustrated in Fig. 2, the framework consists of four main modules: Data Representation, Topic Modelling, Network Construction, and Network Evaluation. Module 1, Data Representation, focuses on constructing POI sequences based on street networks. By integrating road network data and raw POI data, this module transforms the spatial distribution of *POIs* into a structured sequence format, which serves as the foundation for subsequent analyses. Module 2, Topic Modelling, applies the Latent Dirichlet Allocation (LDA) model to the POI sequences to discover latent functional topics. The LDA model treats each POI as a word and each POI sequence as a document and uncovers the underlying semantic structure by identifying co-occurrence patterns of *POIs*. The output of this module is a set of interpretable functional topics and their associations with individual *POIs*. Module 3, Network Construction, builds a spatial-functional network based on the results of topic modelling. Nodes in the network

Fig. 1. Figure illustrating the heterogeneity for different sides of urban streets block in Central London, with Point of Interests (POIs) represented as black dots. The top right image shows a streetview filled with shops, whereas the bottom right image, just one step away, shows a residential street view. ©2024 Google Inc. Google and the Google logo are registered trademarks of Google Inc.

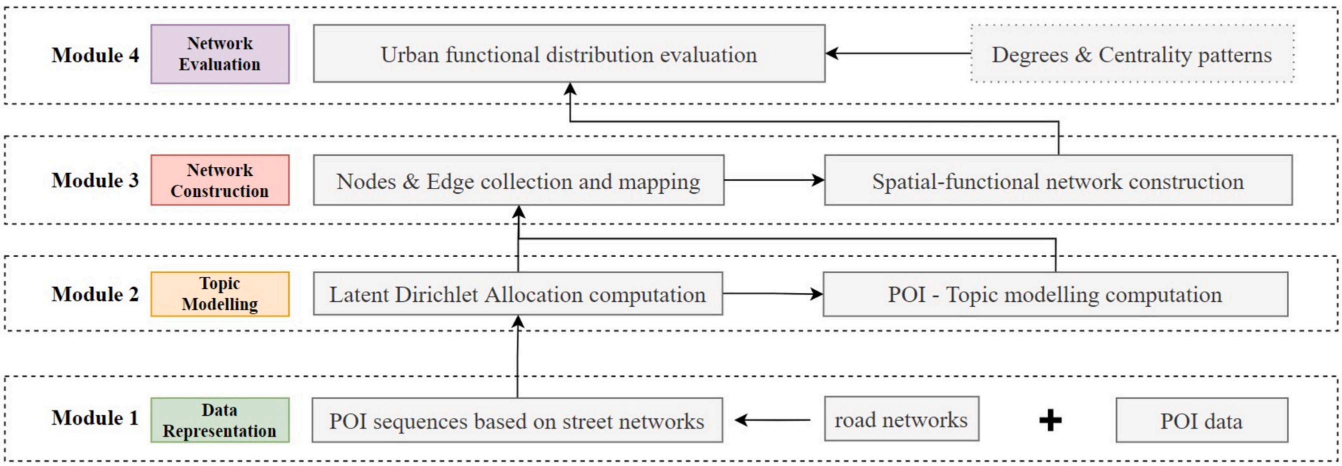


Fig. 2. A framework for analysing urban functional distribution through street-level POI sequences. (a) Module 1: Integrates Road network data with POI data to construct POI sequences along street networks, forming the foundation for analysis. (b) Module 2: Applies Latent Dirichlet Allocation to analyse POI sequences, identifying latent functional topics within the urban environment. (c) Module 3: Creates a spatial-functional network by collecting nodes and edges, mapping the relationships between POIs and their identified topics. (d) Module 4 (Network Evaluation): Assesses the urban functional distribution using network analysis metrics, particularly focusing on degrees and centrality patterns. The arrows indicate the flow of information and dependencies between modules, showing how street-level POI data is progressively transformed into a comprehensive understanding of urban functional distribution.

represent individual *POIs* and latent topics, while edges represent the topic assignment relationships and spatial proximity between *POIs*. This module transforms the abstract topic modelling results into a tangible network structure, enabling further analysis of the spatial and functional interactions among *POIs*. Module 4, Network Evaluation, assesses the topological properties and spatial patterns of the constructed network using various network metrics, such as degrees and centrality. By analysing these metrics, this module reveals the underlying structure and organization principles of urban functions, such as the existence of functional centres, the spatial clustering of related functions, and the diversity and balance of different functions. The proposed framework integrates techniques from data representation, topic modelling, and network science, providing a comprehensive and flexible tool for mining urban functional patterns from POI data. The modular design allows each component to be improved or replaced independently, making the framework adaptable to different data contexts and research questions. The following subsections will elaborate on the technical details and implementation procedures of each module.

3.1. Module 1: POI data representation

3.1.1. Why chose POI sequence to express street?

The utilization of point-of-interest (*POI*) sequences along city streets to characterize urban spatial distributions is founded upon a fundamental geographical and urban planning concept: the structure and morphology of a city can be discerned and comprehended through the arrangement and interplay of its constituent elements. In this context, each street is a microcosm of the urban fabric, with its amalgamation of *POIs* unveiling the city's functional layouts and patterns of activities. Theoretically, the distribution of *POIs* within a cityscape is not arbitrary but is determined by an amalgam of factors encompassing economic, social, cultural, and planning dimensions. For instance, commercial districts are often typified by a concentration of restaurants, shopping centres, and office complexes, whereas residential areas may predominantly feature homes, educational institutions, and small-scale retailers. These clusters of *POIs* correspond to different functional zones within urban space. Analysing these *POIs* enables the beginning of an understanding of how the city's structure and functional regions are spatially arranged. A key advantage of employing POI sequences to express urban distributions lies in the capacity to capture the interdependencies and spatial continuities among *POIs*. This perspective emphasizes the interrelationships between *POIs* and the overall structure that they collectively form. For example, the proximity of a coffee shop to a library may reflect a pattern of cultural and social interaction, a pattern that may vary across different urban localities. Through such analyses, it becomes possible to identify'Spatial-Functional Patterns' within the city, illustrating how various commercial and living functions interact and are distributed within specific geographic locales.

3.1.2. Constructing POI sequences based on street networks

We introduce a new method of embedding *POIs* along urban streets sequentially. Fig. 3 shows the construction process.

(a) We first depicts an overhead map view with numerous scattered *POI* points. (b) We then create a buffer based on its road type shown in pink that captures a subset of *POIs* that falls within its bounds. (c) We then project *POIs* to the street by casting a ray from the *POIs* to the road's network. Each point within the buffer is labelled with a unique identifying number, illustrating their respective order based on their position along the street.

Selecting POI sequences as fundamental geographical units, we subsequently introduce a novel methodology to represent these sequences within the urban streets. Initially, Fig. 3(a) depicts an overhead map view with numerous points scattered across. These points presumably represent different *POIs*. There is a prominent black road visible amidst these points. A buffer analysis of the road network is conducted, creating specific distance buffers for each road, contingent upon its distinct type and tier. Subsequent application of spatial join techniques allows for the association of each *POI* with its respective road buffer, ensuring the spatial distribution of the *POIs* aligns closely with the structure of the road network. As shown in Fig. 3(b), an elongated pink rectangular buffer overlays this section, capturing a subset of the *POIs* that fall within its bounds.

Secondly, further computations then determine the proportional position of each POI on its corresponding road. This determination is facilitated by a projection algorithm founded on shortest distance principles: a ray is cast from the POI to the road's centreline, and the intersection of this ray with the road's centreline elucidates the POI's position in relation to the road's total length. Armed with these position details, a sorting process arranges the *POIs* based on their proportional locations on their respective roads. Fig. 3(c) further magnifies the section from Fig. 3(b), detailing the exact positioning of each POI inside the buffer. Each point within the buffer is labelled with a unique identifying

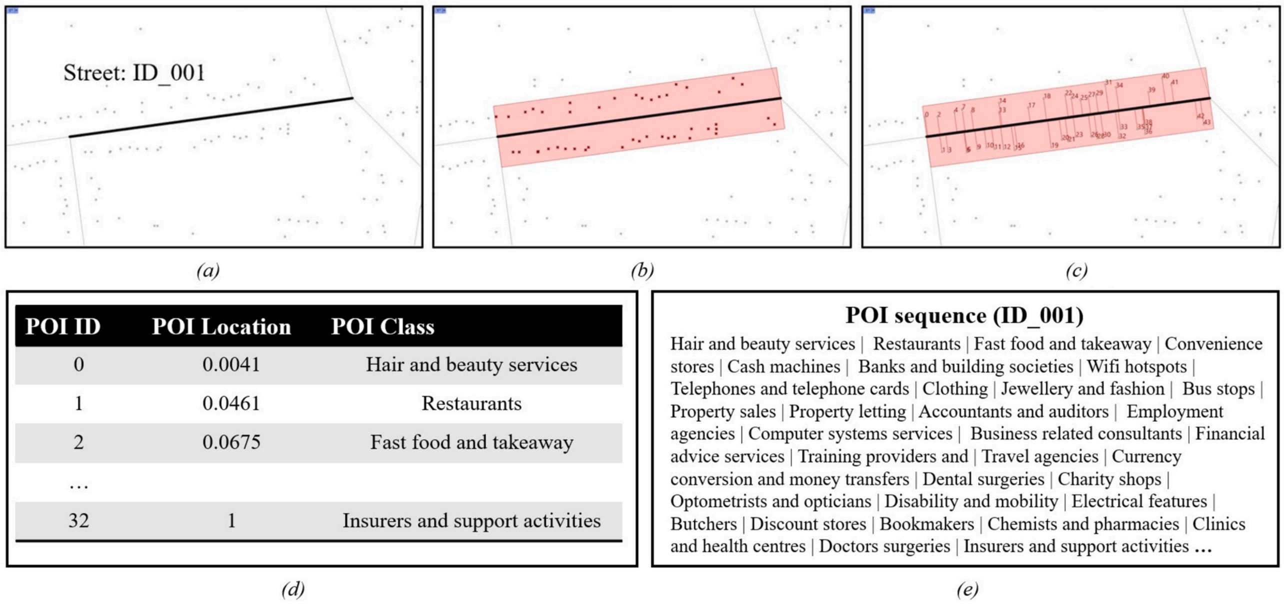


Fig. 3. Construction process of POI sequences based on street networks. The process of transforming street-level POIs into sequential data: (a) Original street segment with POIs distributed around it; (b) Creation of a buffer zone (shown in pink) to capture relevant POIs for the street segment; (c) POIs are projected onto the street segment, maintaining their relative positions; (d) Detailed POI information including unique identifier, relative location along the street, and functional class; (e) Final POI sequence showing the complete functional composition of the street segment, where POIs are arranged in order of their appearance along the street. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

number, illustrating their respective order based on their position along the street. Thirdly, post-sorting, these *POIs* undergo grouping by their associated road IDs and a subsequent merger with the road's attribute data, class name. Consequently, each road not only retains its original attributes but is also appended with an ordered sequence of associated *POIs*.

Consequently, each road not only retains its original attributes but is also appended with an ordered sequence of associated *POIs*. As shown in Fig. 3(d), it provides:

- POI ID: A unique identifier for each POI.
- POI Location: A decimal number, possibly representing the POI's relative position on the street.
- POI Class: A description of the type of establishment or service the POI offers (e.g., Hair and beauty services, Restaurants, and so on).

Lastly, all the POI class could be aggregated as a complete POI sequence in representing a street, Take Fig. 3(e) for instance, it provides a cumulative sequence of POIs as they appear on the street, translating the visual information from (c) into long textual list.

3.2. Module 2: LDA-based POI topic modelling

In the domain of *NLP*, topic modelling has become an indispensable tool, widely recognized for its efficacy in uncovering implicit semantic relationships among words in extensive corpora. Topic modelling refers to a suite of algorithms designed to discern latent semantic themes and topics within large-scale textual corpora by analysing lexical co-occurrence patterns. The extracted topics, defined as clusters of co-occurring words, represents predominant themes for the underlying documents.

Recognizing its potential, we adopt this technique in the current study, specifically aiming to mine associations of urban spatial functionality from city roads with POI sequences, effectively discerning spatial-functional patterns. The rationale behind this endeavour is simple: while individual documents or POI sequences may encapsulate multiple topics or functionalities, each topic manifests its unique lexical frequency distribution. Among various topic models, LDA stands prominent, primarily due to its versatility and computational efficiency. In this context, the application of LDA to POI sequences is selected.

3.2.1. Latent Dirichlet allocation: principles and assumptions

Topic modelling comprises a set of algorithms designed to identify latent semantic themes within large-scale textual corpora. This method analyses the co-occurrence patterns of words to uncover clusters that represent the predominant themes of the documents. Topic modelling moves beyond basic clustering techniques like k-means, which limit documents to a single cluster, and instead uses a mixed-membership approach. This allows documents to be associated with multiple topics, reflecting the nuanced and contextual nature of language. Bayesian inference techniques are applied to iteratively refine and optimize topic assignments, enhancing the accuracy of the semantic analysis. This method not only captures the context-dependent nature of word meanings but also provides a comprehensive framework for extracting meaningful insights from extensive textual data. For our study, we have selected the Latent Dirichlet Allocation (LDA) model, a well-established method in topic modelling, known for its effectiveness in handling large datasets and revealing intricate topic structures.

Latent Dirichlet Allocation (LDA) is an unsupervised Bayesian topic modelling approach that assumes documents are mixtures of topics and topics are characterized by a distribution of words. In essence with LDA, documents that share similar topics are likely to contain similar words.

For the inference of model parameters, we employed the Online Variational Bayes (Online VB) method. This method is particularly advantageous for large datasets as it facilitates faster convergence and scalability, the calculation process is depicted in Fig. 4.

3.2.1.1. Model parameters. In the model parameters, α pertains to the Dirichlet prior associated with the document's topic distribution. θ represents the distribution of topics within document *d*, essentially reflecting the probability of each topic in that document. β is the distribution of words associated with each topic, which is a crucial aspect of the model, outlining the probability of words under each topic.

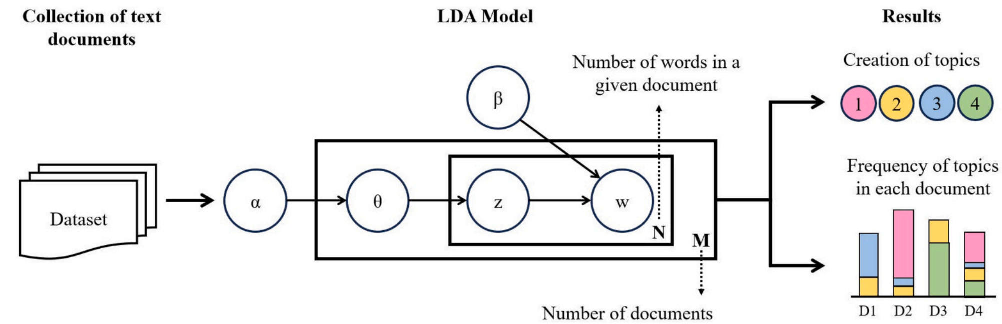


Fig. 4. The calculation process of LDA. Illustration of the LDA process showing three main components: left - input data collection, where each street's POI sequence is treated as a document; centre - the LDA model structure with parameters α (document-topic distribution), β (word distribution), θ (topic mixture), z (topic assignment), and w (observed words); right - model outputs including topic creation and topic frequency distribution across documents (D1-D4).

3.2.1.2. Generative process. For each word w in document d, a topic z is initially chosen from the topic distribution θ of the document. Subsequently, a word w is selected from the word distribution pertinent to topic z.

3.2.1.3. Topic representation. As shown in the model's outputs, one can see the various topics inferred from LDA. These topics manifest as probabilistic distributions based on co-occurrence patterns of words within documents.

3.2.1.4. Distribution of topics. The bar charts delineate the probabilistic distribution of individual topics across different documents, as exemplified by $D1$, $D2$, $D3$, and $D4$.

Owing to its probabilistic foundations, LDA allows for discovering inherent semantic structures and extracting explanatory topics from large swaths of unlabelled text in an unsupervised manner. These characteristics render LDA ideally suited for our exploratory study of uncovering and examining latent topical patterns amidst urban POI sequences. By coupling LDA with spatial analysis, our methodology can uncover functional regions mixtures, and probe the spatial logic behind urban context.

3.2.2. Applying LDA to POI sequences: treating POIs as words and sequences as documents

To uncover the latent spatial-functional patterns from massive POI sequences, we propose a novel approach that combines LDA topic model with the representation of POI sequences as documents. As illustrated in Fig. 5(a), each POI sequence along a street is treated as a 'document' (denoted as d), and each unique POI type is considered as a 'word'. The POI sequence (document d) is represented as a bag-of-words vector, containing various POI types such as Hair and beauty services, Restaurants, Fast food and takeaway, Convenience stores, etc.

The combination of LDA with POI sequences involves the following key steps. First, the LDA model is trained on the corpus of POI sequence documents to learn the latent topic distributions and the topic-POI type distributions. As exemplified in Fig. 5(b), each topic T_k is represented by a distribution ϕ_k over the POI types, indicating the functional composition of that topic. Second, for each POI sequence document d, the trained LDA model is used to infer its topic proportions θ_d, which represent the functional mixture of the corresponding street segment. Fig. 5(c) shows an example of the topic proportions for a specific street ($ID_0$01), where the proportions of different topics (e.g., 0.5793, 0.1762, 0.2445) characterize the street's functional composition.

By applying this LDA-based approach to massive POI sequences across the entire urban road network, we can uncover the latent functional topics and their spatial distributions in the urban space. Streets

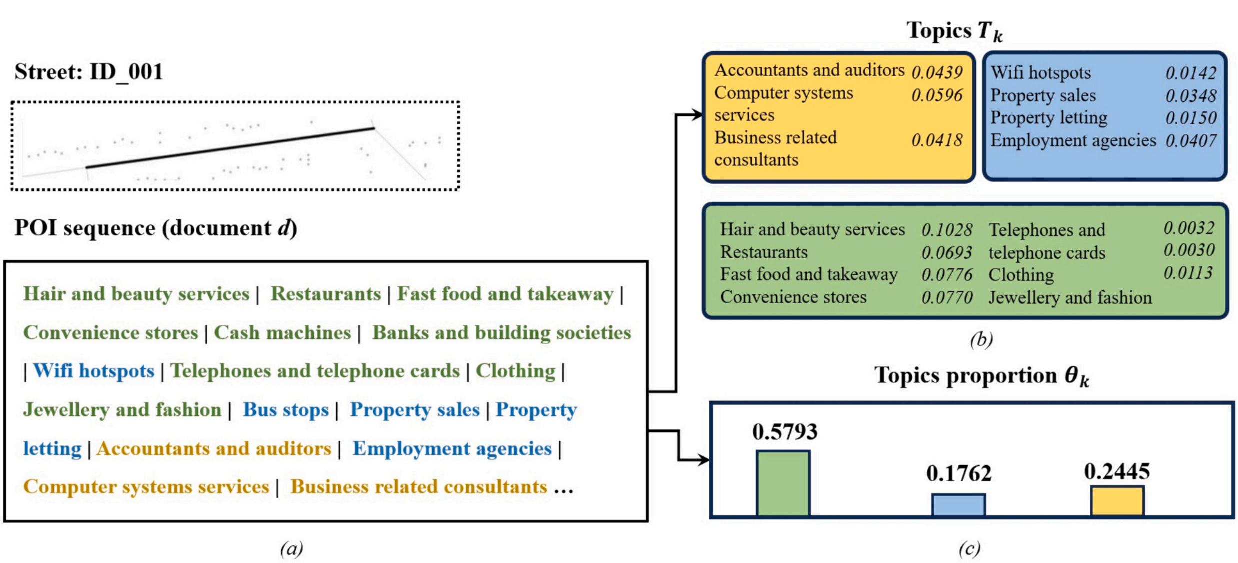


Fig. 5. Topic modelling results for POI sequences based on LDA analysis. Demonstration of the topic modelling process and its results: (a) Input street segment with its corresponding POI sequence treated as a document; (b) Extracted topics showing the probability distribution of POIs within each topic, where different colours represent different functional groups; (c) Topic proportion for the given street segment, indicating the relative strength of each topic's presence. The colours in the POI sequence correspond to their associated topics.

with similar topic proportions can be grouped together to form coherent urban functional zones, revealing the inherent spatial structure of urban functions. This combination of topic modelling with POI sequences provides a data-driven and interpretable way to characterize urban functional zones, enabling a deeper understanding of the complex interactions among POIs and urban functions. Moreover, the scalability of LDA allows this approach to be applied to massive POI datasets, making it suitable for large-scale urban analysis and modelling.

3.3. Module 3: spatial-functional network construction

3.3.1. Defining network nodes: POIs and topics

In the process of constructing a spatial-functional network, we define two types of nodes: large nodes and small nodes. As shown in Fig. 6(c), large nodes (represented by T_{k1}, T_{k2}, T_{k3}) represent the latent functional topics discovered by applying the LDA model to the POI sequences, as shown in Fig. 6(a). Each topic T_k, where $k = 1,2, \ldots, K$, corresponds to a distinct urban function or a mix of related functions. On the other hand, small nodes (represented by the colored circles in Fig. 6(c)) represent individual POIs within the city, categorized according to their functional attributes. Each POI i, where $i = 1,2, \ldots, N$ and N denotes the total number of POIs, is classified as a small node N_i. Through this node definition, the spatial-functional network captures the hierarchical structure of urban functions, with large nodes representing high-level functional topics and small nodes representing specific POIs. (See Fig. 7.)

3.3.2. Constructing the topic-POI network: POI-topic edges and POI-POI edges

To uncover the functional distribution and interrelations of POIs within urban spaces, we propose a novel network construction method that integrates the results of LDA topic modelling with the spatial proximity information derived from POI sequences. As illustrated in Fig. 6(b), the LDA model is applied to the POI sequences (represented as documents d) to identify a set of latent functional topics $T = \{T_1, T_2, \ldots, T_K\}$. Each POI p_i is then assigned to one or more topics based on its posterior topic distribution, forming a POI-topic mapping M: $P \rightarrow T$.

We establish two types of edges in the spatial-functional network. First, for each POI p_i belonging to topic T_k, an edge $e_{pt} = (p_i, T_k)$ is created to connect the POI node N_i with the corresponding topic node T_k. The set of all POI-topic edges is denoted as.

$$E_{pt} = \{(p_i, T_k) | p_i \in P, T_k \in T, M(p_i) = T_k\}$$

Second, to capture the spatial proximity between POIs, we create edges between adjacent POIs in each POI sequence $S = \{p_1, p_2, \ldots, p_m\}$. Specifically, for each pair of adjacent POIs (p_i, p_{i+1}) in sequence S, an edge $e_{pp} = (p_i, p_{i+1})$ is established. The set of all POI-POI edges is denoted as.

$$Epp = \{(pi, pi+1) | pi, pi+1 \in S\}$$

The resulting spatial-functional network can be represented as a graph $G = (V, E)$, where V is the set of all nodes (including both POIs and topics) and E is the set of all edges, consisting of POI-topic edges E_{pt} and POI-POI edges E_{pp}. As shown in Fig. 6(c), this network structure captures the complex relationships between POIs and their associated functional topics, as well as the spatial proximity between POIs along the street network. By analysing the topological properties and connectivity patterns of this network, we can gain valuable insights into the functional organization and spatial distribution of urban activities.

3.4. Module 4: network analysis and interpretation

After constructing the Topic-POI functional network, we employ advanced network analysis methods to uncover the intricate structure and relationships within urban functionalities. This multi-layered network approach provides crucial insights into the organization and dynamics of urban functions, offering robust support for evidence-based urban planning and development decisions. To explore the roles of Points of Interest (POIs) and functional Topics in the urban functional network, we focus on three key centrality measures, each adapted for our unique two-layer network structure.

3.4.1. Degree centrality

Degree centrality quantifies the connectivity of each node within the multi-layer Topic-POI network. It is calculated differently for Topic nodes and POI nodes:

- For Topic nodes: Degree centrality DC_T is defined as the number of POIs directly associated with that Topic:

  $DC_T = |\text{POIs connected to Topic } T|$

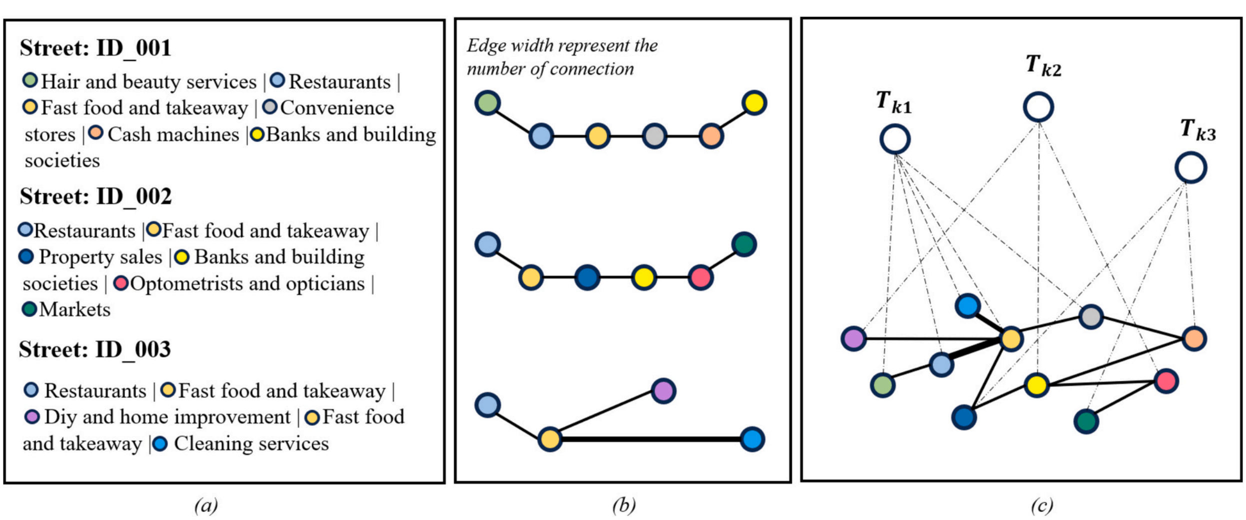
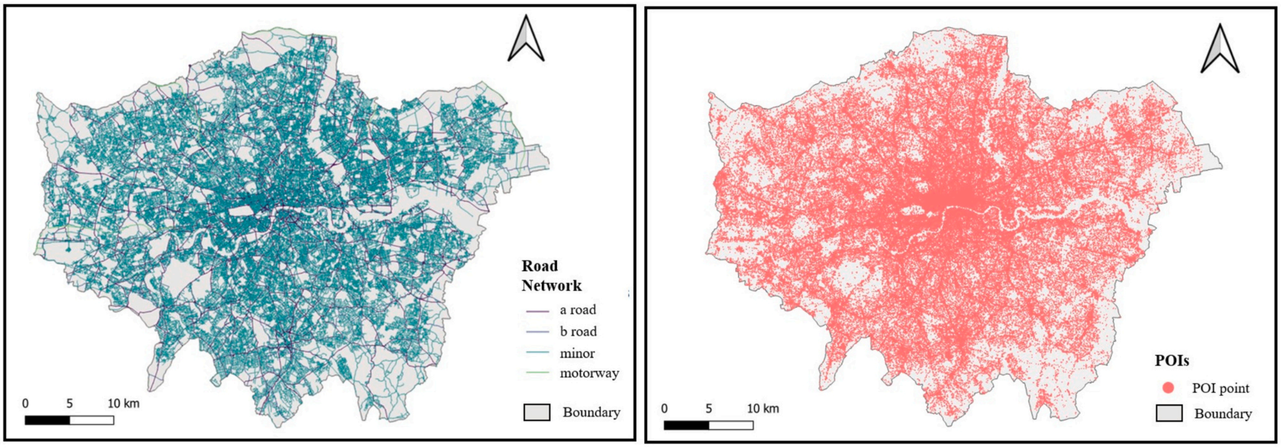


Fig. 6. The multi-layered topic-POI network construction process. Visualization of the network construction process: (a) POI sequences for three example streets, with colored circles representing different types of POIs; (b) Network representation of the POI sequences, where edge width indicates connection strength between POIs; (c) Final multi-layered network structure showing both POI-POI connections (solid lines) and POI-Topic associations (dashed lines), where T_{k1}, T_{k2}, and T_{k3} represent different functional topics.



a. Road networks in London

b. POIs in London

Fig. 7. The street network and POI distribution in London. Spatial distribution of the studied data in London: (a) Road network hierarchy showing A roads, B roads, minor roads, and motorways, with different line styles representing different road types; (b) Distribution of POIs across London, where each red dot represents a POI location. Both maps include the administrative boundary of Greater London for reference. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

- For POI nodes: Degree centrality DC_p includes both the connections to other POIs within the same Topic and the connection to the Topic node itself:

$$DC_P = |\text{POIs connected to POI } P \text{ within the same Topic}| + 1 \text{ (for the Topic connection)}$$

A high degree centrality for a Topic node indicates a diverse and multifaceted urban function, while a high degree centrality for a POI node suggests a multifunctional urban element that plays roles in various aspects of that urban function. A high degree for a Topic node indicates a diverse and multifaceted urban function, while a high degree for a POI node suggests a multifunctional urban element that plays roles in various aspects of that urban function.

3.4.2. Betweenness centrality

This measure identifies nodes that act as bridges or connectors within the network. Our calculation considers paths that can traverse both Topic and POI nodes, reflecting the multi-layer structure of our network. For both Topic and POI nodes, betweenness centrality BC is calculated as:

$$BC(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where:

- σ_{st} is the total number of shortest paths from node s to node t.
- $\sigma_{st}(v)$ is the number of those paths that pass-through node v.

High betweenness centrality for a Topic node indicates a function that bridges different urban activities, while for a POI node, it suggests a location that connects various aspects of urban functionality.

3.4.3. Closeness centrality

Closeness centrality measures how central a node is based on its average distance to all other nodes in the network. This calculation considers the shortest paths between all pairs of nodes, including paths from Topic-to-Topic, Topic-to-POI, and POI-to-POI.

The formula for closeness centrality (CC) for both Topic and POI nodes is:

$$CC(v) = \frac{N-1}{\sum_{u \neq v} d(v,u)}$$

where:

- N is the total number of nodes in the network.
- $d(v,u)$ is the shortest path distance between nodes v and u.

A high closeness centrality indicates that the node (either a Topic or POI) is well-integrated within the overall urban functional fabric, facilitating efficient interaction or service distribution across different urban functions.

These metrics are calculated using custom algorithms that account for the multi-layered nature of our network. By analysing these metrics across both Topic and POI nodes, we can identify key functional hubs (high-degree Topics), multifunctional urban elements (high-degree POIs), bridge functions or locations (high-betweenness nodes), and core urban functions or central locations (high-closeness nodes). This multi-layered network analysis approach provides a nuanced understanding of urban spatial structure, capturing both the thematic relationships between urban functions and the spatial connections between POIs. It allows us to identify not only the important individual elements (POIs or Topics) but also the critical connections and transitions between different urban functions. Through this in-depth network analysis and visualization, the spatial-functional network serves as a powerful analytical tool for understanding the complex organization and usage patterns of urban spaces. It provides empirical support for urban planning, management, and policy formulation by revealing the intricate interactions within and between urban spatial functionalities.

4. Case study

4.1. Data source

In this study, all data were sourced from secondary databases provided by Edina-digiMAP in the United Kingdom. These databases have been extensively used in both academia and practice, thus ensuring the reliability and validity of the research findings. The key datasets extracted from Edina-digiMAP include the Ordnance Survey transportation road network and Ordnance Survey POIs in September 2019.

- Road Network (7(a)): The dataset comprises 114,582 road segments (data type: Linestring), categorized into A Roads (major roads), B Roads (minor roads), and Minor Roads. Hierarchical buffer zones were applied with A Roads utilizing a 60-m buffer, B Roads a 40-m buffer, and Minor Roads a 30-m buffer. This classification allows for the analysis of road network influences on urban functionality at varying levels of thoroughfare significance.
- London Points of Interest (POI): Encompassing 329,058 points (data type: Point), the London POI dataset is hierarchically structured into three levels of detail: groups, categories, and classes. This dataset encompasses a comprehensive taxonomy of 11 groups, 62 categories, and 736 classes. For the purposes of this investigation, the class level was selected for its fine granularity, offering in-depth insights into urban characteristics and facilitating a detailed analysis of the urban landscape. This data offers a nuanced view of London's urban stratification, capturing the diverse array of commercial and community hubs within the city.

All the datasets utilize the British National Grid reference system (coordinate system: EPSG 27700), ensuring geographic accuracy and compatibility.

4.2. Topic number selection

Topic coherence is a metric that measures the frequency with which words belonging to the same topic co-occur within documents. Define $D(v)$ as the document frequency of the term v (i.e., the number of documents containing at least one instance of v, and $D(v, v^{'})$ as the co-document frequency of terms v and $v^{'}$ (i.e., the number of documents containing at least one instance of both v and $v^{'}$). The coherence of topic t is defined by the following formula:

$$\text{Coherence}(t) = \sum_{i<j} log \frac{D(v_i(t), v_j(t)) + 1}{D(v_i(t))}$$

where $(v_1(t), \ldots, v_M(t))$ represents the list of terms most likely to appear under the topic, and the smoothing term 1 is added to avoid taking the logarithm of zero. For K topics, the overall coherence is calculated as:

$$C(K) = \frac{1}{K} \sum_{t=1}^{K} \nabla \text{coherence } (t.)$$

A higher value indicates a higher quality of the topics. To select an appropriate number of topics for study, we ran the LDA model on the POI sequence dataset with varying numbers of topics from 3 to 10 (Fig. 8), incrementing by 1, and each configuration was run 20 times to calculate the corresponding topic coherence. As illustrated in Fig. 1, overall coherence tends to decrease with an increasing number of topics, but peaks at 3 and 6 topics before gradually declining. Based on the relationship between topic coherence scores and the number of topics as shown in Fig. 1, the selection of 6 as the optimal number of topics is justified for several reasons: First, the topic coherence score reaches a local peak when the number of topics is set to 6. This indicates that the LDA model generates topics with high internal consistency at this number of topics, where words belonging to the same topic frequently co-occur in documents, signifying good topic quality. Second, although the coherence score is highest at 3 topics, considering topic granularity, 3 topics may be too broad and insufficient to fully reveal the semantic patterns contained in POI data. Choosing 6 topics achieves a better balance between topic quality and granularity. Furthermore, when the number of topics increases from 6 to 7 or more, there is a clear downward trend in coherence scores, suggesting that further increasing the number of topics may lead to overfitting, generating overly detailed topics that reduce its interpretability.

Therefore, selecting 6 as the optimal number of topics achieves a balanced compromise between topic quality, granularity, and interpretability, providing a suitable framework for uncovering the hidden semantic patterns in POI data.

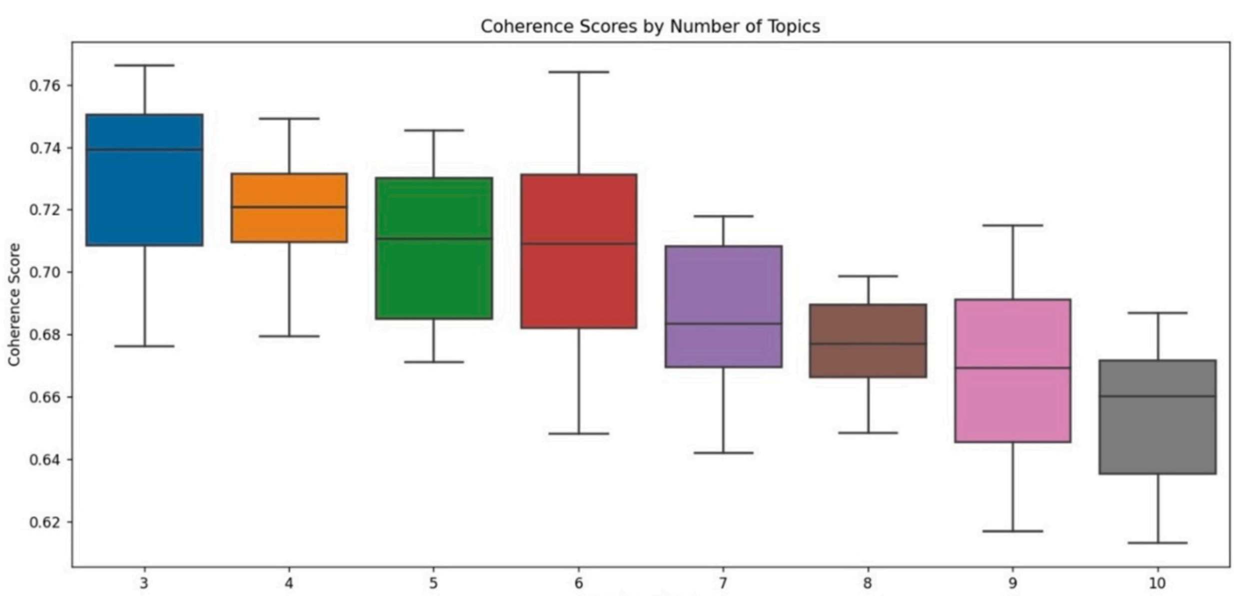


Fig. 8. The box plot displaying coherence scores for different number of topics.

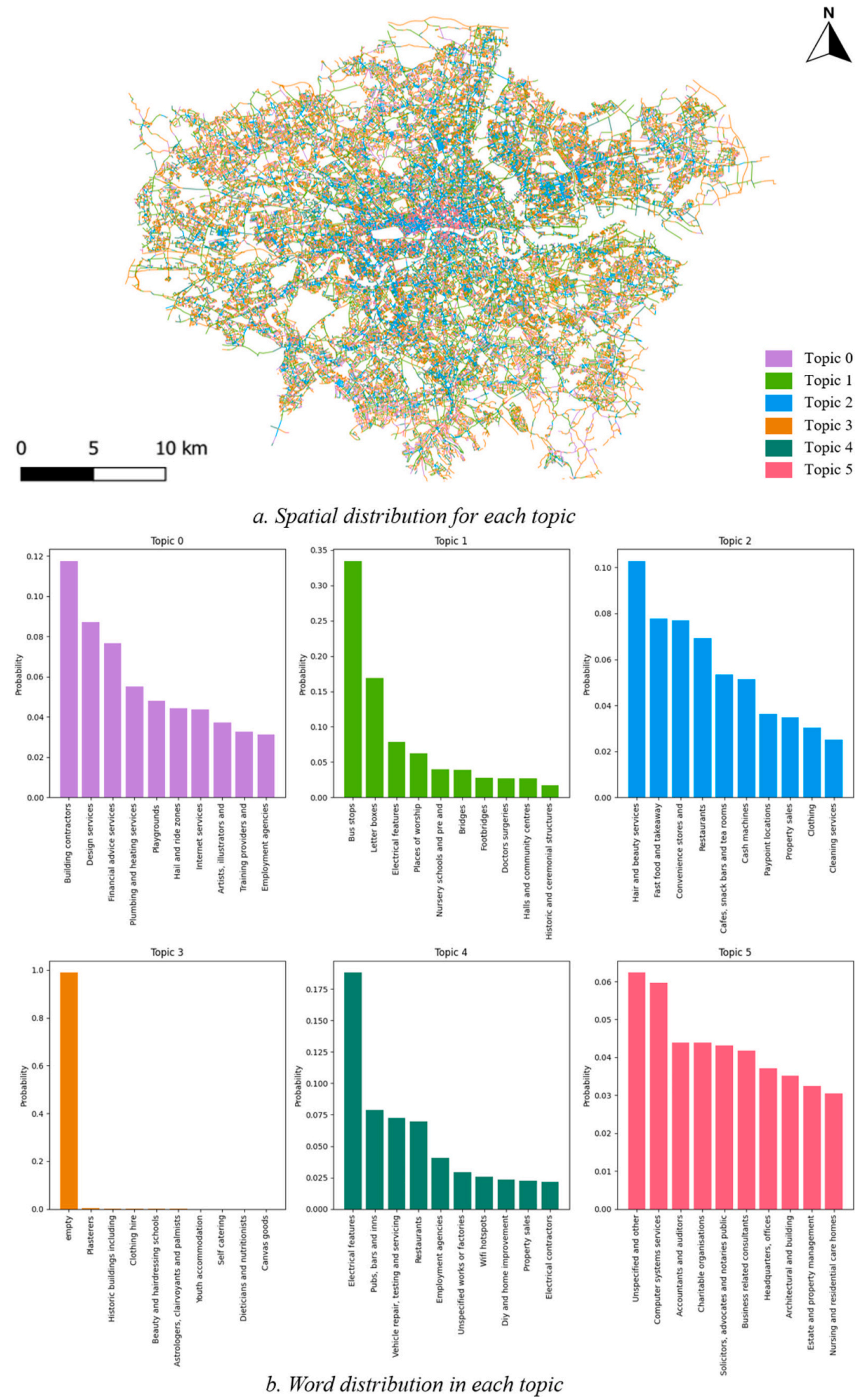


Fig. 9. Spatial distribution and POI composition of the six identified urban functional topics in London. Analysis of the six urban functional topics identified through LDA: (a) Spatial distribution map showing the geographical pattern of each topic across London, with different colours representing different topics; (b) POI composition within each topic shown through probability distributions, where each bar chart represents the most frequent POI types within that topic and their relative probabilities of occurrence. The colours in the bar charts correspond to their respective topics in the spatial distribution map.

4.3. Result analysis

4.3.1. LDA result

Fig. 9 illustrates the spatial distribution characteristics of six topics identified using the LDA model on Points of Interest (POI) data within London's urban area. Each Topic comprises a group of highly associated POI types, reflecting the latent semantic structure of urban functional zones. Observing the spatial distribution of Topics reveals distinct differences and clustering patterns. 1) Topic 1 predominantly occupies the periphery of London, especially in the eastern and southern regions, and is sparser in the city centre. This Topic's leading POI types include bus stations, mailboxes, electrical facilities, and religious places, indicating it represents peripheral urban areas primarily characterized by infrastructure and public services. 2) In contrast, Topics 2 and 4 exhibit pronounced central clustering, predominantly located in London's core area. Topic 2 is leading POI types include beauty services, fast food takeaways, convenience stores, restaurants, and cafes, reflecting the commercial services and consumption functions of the metropolitan core; 3) Topic 4 is dominated by bars, car repairs, restaurants, and human resource services, representing the economic and social activities of the city centre. 4) Topics 0 and 5 are relatively dispersed but tend to cluster in the western and northern parts of London, with leading POI types involving architectural design, financial consulting, plumbing maintenance, amusement parks, and high-end producer services like business offices, accounting auditing, and legal consulting, symbolizing the city's production, innovation, and business functions. 5) Topic 3 is evenly distributed across districts but at the lowest proportion, mainly comprising specific POI types like decoration, relics, and clothing rental. The distribution of POI types across different Topics shows significant disparities.

Overall, by conducting LDA topic modelling on London's POI data, this study identified six semantically meaningful and spatially distinct latent Topics. These Topics each represent different functional areas of the city, including peripheral zones dominated by public services, core areas characterized by commercial consumption, and innovation zones led by productive services. The composition of POI types within different Topics is distinctive, reflecting the intrinsic logic of corresponding urban functions. These results demonstrate the effectiveness of the LDA model in mining urban POI semantic patterns but also offer a new perspective for understanding the internal structure and functional zoning of cities.

4.3.2. Comparison with traditional clustering methods

To demonstrate the advantages of our street-level Topic-POI network approach, we conducted a comparative analysis using three different methods on the same London POI dataset: (a) direct POI category classification, (b) Gaussian Mixture Model (GMM) clustering, and (c) K-means clustering based on POI type proportions.

The traditional methods reveal several limitations. 1) Direct POI Category Classification (Fig. 10(a)) simply categorizes streets based on predefined functional categories (e.g., Accommodation, Commercial services, public infrastructure). While straightforward, this approach oversimplifies urban functionality by forcing each street into a single category, ignoring the complex mix of functions that typically exist. Both GMM and K-means clustering (Fig. 10(b), (c)) show significant limitations. They produce highly imbalanced clusters, as evidenced by Cluster 2 in both methods containing only single-function POIs with 100 % concentration. Several clusters are dominated by single POI types, such as Sewage services 100 % in GMM Cluster 1 and Sculptors 100 % in K-means Cluster 2. The larger clusters, such as Cluster 0 in both methods, show similar POI compositions without clear functional differentiation. Moreover, these methods fail to capture the spatial continuity and functional relationships between different POI types.

In contrast, our Topic-POI network approach (as shown in Fig. 9)

Fig. 10. Comparison of different POI clustering methods in London. Three clustering approaches applied to London's POI data, showing different perspectives on urban functional distribution: (a) Direct POI category classification based on dominant functional types, offering a simplistic view through predefined categories such as accommodation, retail, and transport; (b) GMM clustering results showing POI composition within each cluster, revealing highly imbalanced distributions with some clusters dominated by single functions; (c) K-means clustering results demonstrating similar limitations with extreme concentrations in certain clusters. The percentages shown for clusters in (b) and (c) represent the proportion of different POI types within each cluster. All three maps cover the same Greater London area with a 10 km scale bar, demonstrating how different analytical approaches can yield varying interpretations of the same urban space.

offers several advantages. It captures mixed functionality through topic proportions rather than strict categorization, reveals more balanced and interpretable functional patterns, and preserves street-level spatial relationships. The method identifies functional interactions through network metrics and shows clear differentiation between different urban functions while maintaining their interconnections. For example, while traditional clustering methods struggle with areas of mixed use, our Topic 2 clearly identifies urban lifestyle and consumption patterns with a balanced distribution of retail, food, and service establishments. Similarly, our Topic 1 reveals coherent public infrastructure patterns that are fragmented across multiple clusters in traditional approaches. This comparison demonstrates that our method provides a more nuanced and sophisticated understanding of urban functional distribution by considering both the co-occurrence patterns of POIs and their spatial-functional relationships, rather than simply grouping POIs based on their proportional distribution or predefined categories.

4.3.3. POI network construction result

To further explore the semantic and structural characteristics of these Topics, we constructed a multi-layer network of Topics and

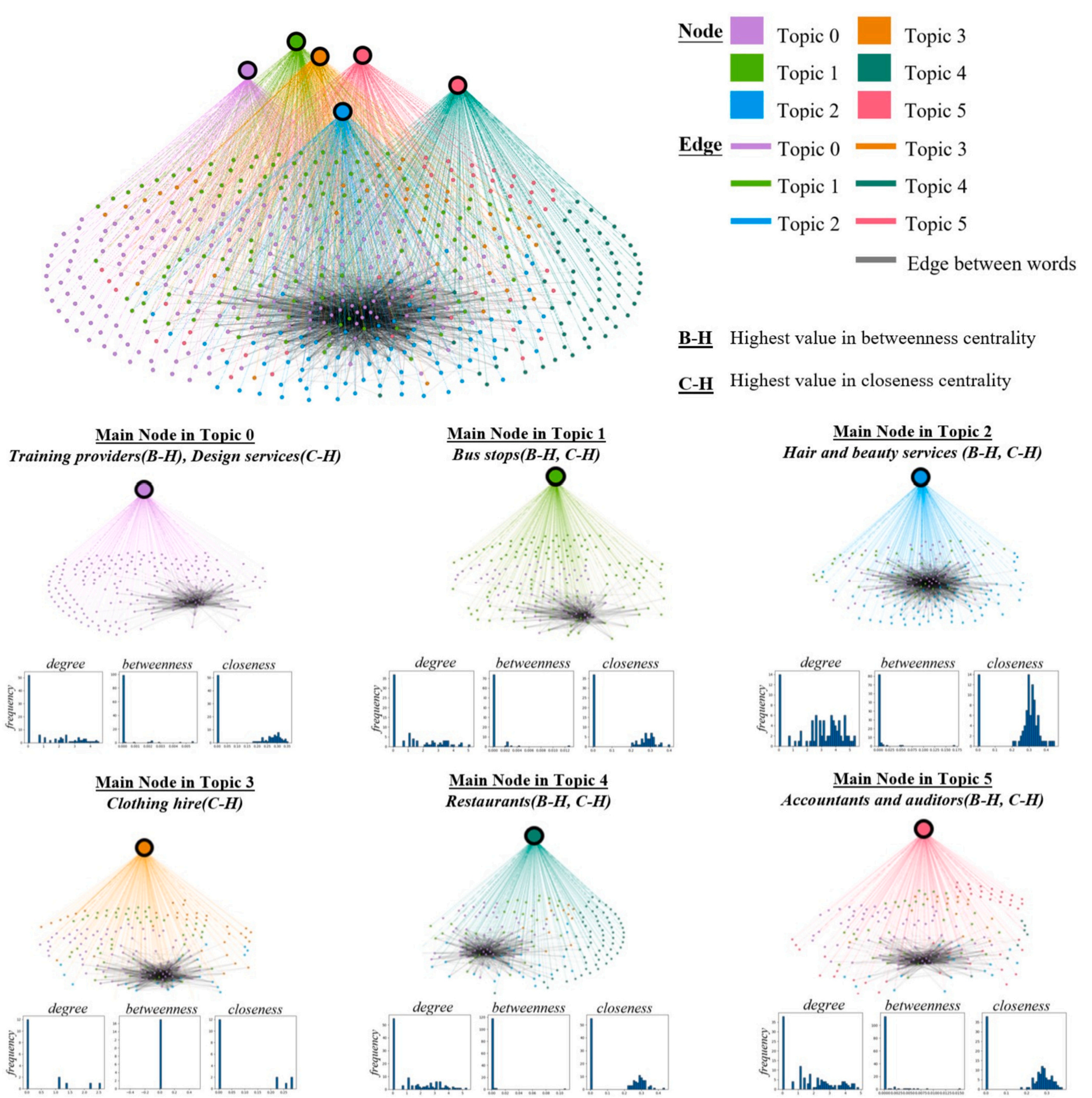


Fig. 11. The multi-layered POI-network structure and centrality analysis of main nodes within each topic. Visualization of the network structure and centrality analysis: Top - complete network showing all topics and their interconnections, with different colours representing different topics and grey edges indicating connections between POIs; Bottom - individual topic subnetworks centred around their main nodes, each accompanied by corresponding centrality measurements (degree, betweenness, and closeness). B—H indicates highest betweenness centrality, and C—H indicates highest closeness centrality. The main nodes identified in each topic (e.g., Training providers in Topic 0, Bus stops in Topic 1) represent key functional elements within their respective urban areas.

vocabulary and analysed its topological structure and network metrics. Fig. 11 depicts the overall topology of the Topic-vocabulary multi-layer network. The nodes in the network represent the six Topics (Topic 0–5), and the edges between nodes signify the associations between Topics at the POI vocabulary level. The greater the weight of an edge, the more POI types the two Topics share, indicating a higher semantic similarity. The results reveal significant differences in the strength of associations between different Topics. The edges between Topics 0, 2, and 4 are thicker, suggesting they have more overlap in POI composition, possibly representing interrelated urban functional zones, whereas Topics 1, 3, and 5 are relatively independent, with weaker connections to other Topics, reflecting their semantic uniqueness. Fig. 11 showcases the internal POI vocabulary network of each Topic. Key network metrics of these sub-networks, including degree, betweenness, and closeness, were calculated and visualized.

By comparing the aforementioned network metrics, we can assess the differences in POI composition characteristics and centrality within different Topics.

Topic 0's POI network **(Professional Services Cluster)**: This cluster is primarily characterized by professional services and construction-related industries, reflecting a highly specialized business district. The main POI types include: Building contractors (11.76 %), Design services (8.72 %), Financial advice services (7.66 %) and so on. Network analysis reveals a pronounced centralization in the POI network of this topic. The degree distribution histogram is highly right skewed, indicating the presence of a few core nodes with degrees far exceeding others. This structure reflects the functional concentration within the professional services area, with a small number of key service types playing central roles in the network. The betweenness centrality distribution also shows a peak feature, further confirming the crucial role of these key nodes in controlling the network's information flow. This suggests that certain professional services (such as training providers and financial advisory services) act as important bridges connecting different functional groups. In contrast, most nodes have lower closeness centrality, indicating structural holes between them and the core nodes, suggesting room for improvement in overall network connectivity. This may reflect the specialized division of labor between different professional services, but also hints at potential opportunities for enhancing collaboration among various service types. The most influential POI types in this theme are:

1. Training providers: Highest betweenness centrality (0.0056), indicating its crucial bridging role connecting different functional groups.
2. Financial advice services: Second-highest betweenness centrality (0.0055), signifying its key connecting position in the network.
3. Design services: Highest closeness centrality (0.3545), suggesting its central position with quick access to other nodes.
4. Building contractors: Second-highest closeness centrality (0.3443), indicating its core position in the network.

This network structure reflects a highly concentrated professional services district, where key POI types such as construction, design, finance, and training play central functional roles, while other types of services form supportive functional clusters around these core services. This structure facilitates the agglomeration and synergy of knowledge-intensive services, but may also risk over-dependence on core services.

Topic 1's POI network **(Public Infrastructure Cluster)**: This cluster is characterized by public infrastructure and community services, reflecting a zone dedicated to essential public amenities. The main POI types include: Bus stops (33.45 %), Letter boxes (16.91 %), Electrical features (7.89 %). Network analysis reveals an unbalanced topology in the POI network of this topic, albeit less pronounced than in Topic 0. The degree distribution histogram shows a lower degree of right-skew, indicating a more even distribution of connections among nodes. This structure suggests a more distributed pattern of public infrastructure, with fewer dominant hubs compared to the Professional Services District. The betweenness centrality distribution exhibits a peak, but the overall distribution is more uniform compared to Topic 0. This indicates the presence of several secondary bridging nodes in addition to a few key nodes, suggesting that multiple types of public infrastructure play significant roles in connecting different parts of the network. Closeness centrality of nodes within Topic 1 is generally higher, indicating better overall connectivity. However, a gap still exists between head nodes and others, reflecting the network's hierarchical characteristics. This suggests that while public infrastructure is more evenly distributed, there are still key facilities that serve as central points in the network. The most influential POI types in this theme are:

1. Bus stops: Highest betweenness centrality (0.0128) and highest closeness centrality (0.4030), indicating its crucial role as both a connector and a central point in the public infrastructure network.
2. Places of worship: Second-highest betweenness centrality (0.0042), suggesting its significant role in bridging different parts of the community.
3. Dental surgeries: Second-highest closeness centrality (0.3563), indicating its importanceas a centrally located public health service.

This network structure reflects a public infrastructure zone with a more distributed yet still hierarchical arrangement of services. Key POI types such as bus stops play central roles in both connecting and anchoring the network, while other types of public amenities form a more evenly distributed support structure. This arrangement facilitates widespread access to essential services while maintaining efficient connections between different types of public infrastructure.

Topic 2's POI network **(Urban Lifestyle and Consumption Cluster)**: This cluster is characterized by a diverse mix of retail and personal services, reflecting a vibrant urban lifestyle and consumption center. The main POI types include:Hair and beauty services (10.28 %), Fast food and takeaway (7.76 %), Convenience stores and retail (7.70 %). Network analysis reveals a markedly different structure compared to the previous two topics, presenting a more dispersed and balanced characteristic. The degree distribution histogram approaches a normal distribution, with most nodes' degree values concentrated at a medium level. This indicates a lack of dominant hubs or isolated peripheries, suggesting a more evenly distributed and interconnected urban commercial landscape. The betweenness centrality distribution is more gradual, with several peaks corresponding to different functional communities. This pattern suggests that multiple types of businesses and services play important roles in connecting various parts of the network, contributing to a diverse and well-integrated urban environment. Closeness centrality values among nodes are relatively similar, indicating a higher efficiency in information and resource flow across different POI types. This homogeneous structural feature facilitates collaborative development within the region, promoting a synergistic relationship among various urban lifestyle and consumption services.

The most influential POI types in this theme are:

1. Hair and beauty services: Highest betweenness centrality (0.1702) and highest closeness centrality (0.4493), indicating its central role in both connecting different parts of the network and being easily accessible within the urban fabric.
2. Property sales: Second-highest betweenness centrality (0.0548), suggesting its significant role in bridging different functional communities within the urban consumption landscape.
3. Fast food and takeaway: Second-highest closeness centrality (0.4339), reflecting its importance as a widely accessible and central service in the urban lifestyle network.

This network structure reflects an Urban Lifestyle and Consumption Hub with a well-balanced and integrated arrangement of services. The even distribution of connections and the presence of multiple bridging nodes suggest a diverse and resilient urban environment. This structure promotes easy access to a variety of services and facilitates the flow of people and resources, contributing to a dynamic and cohesive urban experience. The balanced nature of this network also suggests that the area is adaptable to changing consumer preferences and lifestyle trends.

Topic 3's POI network **(Urban Regeneration Cluster)**: This cluster is characterized by a high concentration of empty spaces and a sparse distribution of specific services, suggesting an urban zone potentially undergoing regeneration or redevelopment. The main POI types include Empty spaces (98.88 %), Plasterers (0.36 %), Historic buildings (0.24 %). It's important to note that these 'unclassified or vacant spaces' are not a single POI category, but rather represent locations in the Ordnance Survey POI database that are recorded but either temporarily lack specific functional attributes or are marked as vacant. Our cross-reference with satellite imagery reveals that these spaces often correspond to vacant lots, buildings temporarily without active use, or areas undergoing redevelopment. We chose to retain these spaces in our analysis as they represent actual elements of the urban landscape, and their inclusion is crucial for maintaining the spatial continuity of our street-level analysis.

Network analysis in Topic 3's POI network reveals a significantly different structure compared to the previous topics, with a smaller scale and exceptionally simple topological structure. The degree distribution histogram shows a high concentration in the low-value range, indicating generally weak connectivity among nodes. This suggests a fragmented urban landscape with limited interaction between the few existing services. Notably, the betweenness centrality for all nodes is 0. This is a crucial observation that indicates there are no nodes acting as bridges or connectors within the network. In other words, there is no flow of information or resources between different parts of the network, as each node (if connected at all) is only connected to its immediate neighbours. This extreme lack of bridging nodes is consistent with an area undergoing significant regeneration, where established functional relationships are minimal or non-existent. Closeness centrality values are generally high but given the network's small scale and the lack of betweenness centrality, this likely reflects mere geographical proximity between the few existing nodes rather than meaningful functional connections. This suggests that while the limited services in the area are physically close, they do not form a cohesive functional network. The most influential POI types in this theme, based solely on closeness centrality, are:

1. Clothing hire: Highest closeness centrality (0.2843)
2. Beauty and hairdressing schools: highest closeness centrality (0.2843)

However, given the overall network characteristics, their influence appears extremely limited. This network structure reflects an area that lacks the complex interconnections typical of established urban zones. The high proportion of POIs with missing information presents a challenge in fully understanding the area's characteristics, but the available data suggests a poorly integrated urban fabric. The absence of any betweenness centrality indicates that this cluster has not yet developed any functional relationships typical of established urban zones. This presents both significant challenges and unique opportunities for urban planners and developers.

Topic 4's POI network **(Mixed-Use Entertainment Cluster)**: This cluster is characterized by a diverse mix of entertainment, hospitality, and urban services, reflecting a vibrant, multifunctional urban zone. The main POI types include: Electrical features (18.81 %), Pubs, bars and inns (7.90 %), Vehicle repair, testing and servicing (7.23 %), Restaurants (6.93 %). Network analysis reveals a structure that falls between Topics 2 and 0, featuring both dispersion and centrality. The degree distribution histogram shows a gentle slope in the low-to-middle value range but peaks in the high-value range, corresponding to a few key nodes. This indicates a mixed structure where many POIs have moderate connectivity, but a few plays central roles in the network. The betweenness centrality distribution displays a clear stepped feature, with different levels of nodes playing varied roles in information dissemination. This suggests a hierarchical structure where different types of services contribute to the network's connectivity at various levels, reflecting the diverse nature of a mixed-use district. Closeness centrality is generally high but shows a significant head effect, indicating that while many nodes are well-connected, the network's overall connectivity relies heavily on a few core nodes. This core-periphery structure is characteristic of mixed-function urban zones, where key services act as anchors for a diverse range of supporting functions. The most influential POI types in this theme are:

1. Restaurants: Highest betweenness centrality (0.1090) and highest closeness centrality (0.4365), indicating their crucial role as both connectors and central points in the mixed-use entertainment network.
2. Property letting: Second-highest betweenness centrality (0.0064), suggesting its significant role in bridging different parts of the district, possibly reflecting the area's diverse residential and commercial mix.
3. Wi-Fi hotspots: Second-highest closeness centrality (0.3769), highlighting the importance of digital connectivity in this modern urban environment.

This network structure reflects a Mixed-Use Entertainment District with a balanced blend of centralized and distributed services. Key POI types like restaurants play central roles in both connecting and anchoring the network, while a diverse range of other services contribute to the area's multifunctional character. This arrangement facilitates a dynamic urban environment where entertainment, dining, residential, and other urban services coexist and interact closely.

Topic 5's POI network **(Business and Charitable Services Cluster)**: This cluster is characterized by a diverse mix of professional business services and charitable organizations, reflecting a zone dedicated to knowledge-intensive and socially responsible activities. The main POI types include: Unspecified and other services (6.23 %), Computer systems services (5.96 %), Accountants and auditors (4.39 %), Charitable organizations (4.39 %). Network analysis reveals a structure similar to Topic 2, exhibiting a dispersed and balanced characteristic. The degree distribution histogram is relatively flat, without marked peaks or troughs, indicating a more even distribution of connections among nodes. This suggests a well-integrated business environment where various services have similar levels of connectivity. The betweenness centrality distribution is also relatively uniform, without significant peaks. This pattern implies that many different types of business and charitable services play important roles in connecting various parts of the network, contributing to a diverse and well-integrated professional ecosystem. Closeness centrality values among nodes are generally high and close to each other, indicating tight connections among POI types within this region. This suggests a functionally homogeneous and highly integrated network structure, where different services can easily interact and collaborate. The most influential POI types in this theme are:

1. Accountants and auditors: Highest betweenness centrality (0.0155) and highest closeness centrality (0.3762), indicating their central role in both connecting different parts of the network and being easily accessible within the business services landscape.
2. Architectural and building services: Second-highest betweenness centrality (0.0088), suggesting their significant role in bridging different functional communities within the professional services network.
3. Computer systems services: Second-highest closeness centrality (0.3581), reflecting their importance as a widely accessible and central service in this knowledge-intensive cluster.

This network structure reflects a Business and Charitable Services Cluster with a well-balanced and highly integrated arrangement of services. The even distribution of connections and the presence of multiple types of services with similar network characteristics suggest a diverse and resilient professional environment.

Comparing the network features of the six Topics reveals significant differences in the organization of POIs across urban functional areas, shedding light on the complex structure of modern cities. These differences reflect diverse spatial layouts and interaction patterns among economic activities and social services, revealing underlying socio-economic dynamics. Topics 0 (Professional Services) and 1 (Public Infrastructure) show the highest degree of centralization. In professional services, this mirrors the clustering of knowledge-intensive industries to promote innovation, suggesting high daytime activity but limited after-hours vitality. For public infrastructure, centralization indicates strategic service placement to maximize accessibility, though it raises questions about equitable access across urban areas. In contrast, Topics 2 (Urban Lifestyle and Consumption) and 5 (Business and Charitable Services) exhibit the most homogeneous network structures. This suggests vibrant, mixed-use areas for lifestyle and consumption that support walkability and community interaction, while business and charitable services reflect a well-distributed support system contributing to social resilience. Topic 4 (Mixed-Use Entertainment Cluster) combines characteristics of centralized and homogeneous networks, representing areas that balance concentrated entertainment hubs with a variety of supporting services. These areas serve as focal points for local communities and city-wide attractions, shaping urban identity and fostering social interactions. The structured variation across different urban functions highlights the interplay between historical development, planning policies, and evolving needs, emphasizing the value of area-specific approaches. For example, centralized professional service areas could benefit from policies promoting mixed-use development to enhance after-hours activity, while homogeneous consumption areas might focus on preserving local character to prevent gentrification.

The network characteristics across the six topics, as shown in Table 1, reveal diverse urban structures and functional relationships within London, providing insight into the city's socio-economic landscape. Topic 2 (Urban Lifestyle and Consumption) stands out with the highest network density (0.2016), average degree (18.3478), clustering coefficient (0.613), and transitivity (0.5784), indicating a highly interconnected and cohesive environment. This suggests vibrant, mixed-use areas that serve as key community hubs, offering diverse services and adapting well to changing consumer demands. In contrast, Topic 3 (Urban Regeneration) has no connections between its 17 nodes, reflecting areas undergoing significant transition, such as redevelopment zones. The lack of connectivity indicates a temporary 'functional vacuum,' presenting both challenges and opportunities for urban renewal, particularly in addressing affordable housing and new public spaces. Topics 0 (Professional Services), 1 (Public Infrastructure), 4 (Mixed-Use Entertainment), and 5 (Business and Charitable Services) represent varying levels of urban functionality. Topic 5, with the most nodes (129) and moderate connectivity, reflects London's polycentric business landscape, supporting economic resilience through distributed activity hubs. Topic 1, despite fewer nodes, shows higher density and clustering, indicating localized planning efforts to enhance public service delivery through 'community hubs.' Topic 4's moderate connectivity aligns with mixed-use entertainment areas like the West End, combining cultural venues and recreational spaces. Topic 0, with the lowest density and clustering, suggests a specialized, hierarchical structure typical of financial districts. Fragmentation analysis reveals that Topic 2's high integration (17 components) suggests cohesive, community-oriented areas, while greater fragmentation in Topics 0, 1, 4, and 5 reflects London's compartmentalized urban landscape. This diversity allows for specialized districts but can hinder city-wide integration and equitable access. In conclusion, London's urban fabric ranges from highly integrated lifestyle hubs to more specialized, fragmented areas, underscoring its character as a global city with distinct but interconnected zones. For urban planners, these insights emphasize the need for tailored strategies that enhance each area's unique character while promoting integration and accessibility across the city. Future development should focus on bridging fragmented areas, revitalizing transitional zones, and ensuring the benefits of well-connected districts reach all residents.

5. Discussion

This study introduces a novel approach to identify and characterize urban functional areas by integrating topic modelling (LDA) and network science methods, offering new insights into the spatial distribution and intrinsic organization logic of POIs in London. The results demonstrate the effectiveness of this approach in uncovering the latent semantic structure and spatial interaction patterns of urban functions from POI data.

5.1. Methodological innovation

The main innovation of this research lies in the combination of topic modelling and network analysis techniques to investigate street-level urban functional areas from both semantic and structural perspectives. Unlike traditional methods that rely on subjective delineation of administrative boundaries, Voronoi cells or regular grids, our approach identifies street-level functional areas based on the co-occurrence patterns and network characteristics of POI vocabularies, providing a more objective and data-driven framework for urban spatial analysis. The LDA model enables us to discover latent themes in POI data, while network analysis reveals the complex interrelationships and organizational structure among different types of POIs. This integration of methods offers a comprehensive and nuanced understanding of urban functional areas that goes beyond simple statistical aggregation, spatial clustering or colocation mining.

5.2. Necessity of the network perspective

Cities are complex systems where different functional areas are not isolated but closely interconnected. The network analysis method is particularly suitable for capturing these complex, non-linear relationships among POIs, as it focuses on the interconnectedness and interaction patterns of urban elements. By constructing multi-layer networks of Topics and POIs, we can uncover the hidden associations and similarities between different functional areas, which are difficult to capture using

Table 1
Multi-layered POI-network statistics for each topic.

Topic	Num nodes	Num edges	Density	Average degree	Clustering coefficient	Transitivity	Num connected components	Largest component size
Topic 0	110	97	0.0162	1.7636	0.0654	0.2600	69	42
Topic 1	83	102	0.0300	2.4578	0.1679	0.3401	49	35
Topic 2	92	844	0.2016	18.3478	0.6130	0.5784	17	76
Topic 3	17	0	0.0000	0.0000	0.0000	0.0000	17	1
Topic 4	123	162	0.0216	2.6341	0.1817	0.2894	71	53
Topic 5	129	298	0.0361	4.6202	0.2358	0.3875	59	71

traditional statistical methods. The network perspective thus provides a more realistic and holistic representation of urban spatial structure, taking into account the inherent interdependencies and complexities of urban systems.

5.3. Interpretation of network characteristics

The network metrics, such as degree, betweenness centrality, and closeness centrality, offer valuable insights into the roles and positions of different types of POIs in the functional organization of urban space. By comparing the network characteristics across different Topics, we can identify the key nodes and bridging elements that play crucial roles in shaping the structure and dynamics of urban functional areas. For example, the high centralization of POI networks in Topics 0 and 1 suggests the presence of a few dominant POI types with highest centrality, that control the flow and distribution of resources and activities in these areas, while the more homogeneous and balanced structure of POI networks in Topics 2 and 5 indicates a higher level of functional integration and collaboration. These findings provide a deeper understanding of the organizational logic and optimization strategies for different functional areas, informing urban planning and management practices.

5.4. From street-level interactions to city-wide functional patterns: an emergent perspective

Our study demonstrates a novel approach to understanding urban functional relationships by bridging micro-level street interactions with macro-level city patterns. This multi-scale perspective offers unique insights into the complex organization of urban functions, complementing the network characteristics discussed in the previous section. The street-level POI sequences that form the foundation of our analysis capture the fine-grained functional fabric of the city. As we aggregate these sequences into our multi-layered Topic-POI network, we observe the emergence of larger functional structures. This emergence is a key concept in our approach, illustrating how micro-interactions among POIs contribute to the generation of multi-scaled centrality in terms of functional relationships. For instance, the high centrality nodes in our network, as identified in Topics 0 and 1, often correspond to streets or areas where diverse POIs coexist and interact intensively. These locations emerge as functional hotspots that influence the broader urban fabric. The betweenness centrality of certain POIs or topics in our network reveals how some urban functions act as bridges between different functional areas, a pattern that only becomes apparent when viewing the city as an interconnected system. This bottom-up perspective provides a more nuanced understanding of how local, street-level dynamics shape the overall functional organization of the city. It reveals that city-wide functional patterns are not merely the sum of individual POIs, but rather the result of complex interactions and relationships at multiple scales, as evidenced by the varying network structures observed across different Topics.

5.5. Implications for urban planning and design

The insights gained from POI network analysis have important implications for urban planning and design. By identifying the optimal combination patterns and spatial configurations of different types of POIs, planners can develop more fine-grained and evidence-based strategies for land use zoning, spatial layout optimization, and public amenity facility allocation. For example, the network centrality measures can help prioritize the locations for key public service facilities to maximize their accessibility and service coverage. Moreover, the network-based approach enables planners to assess the functional diversity and resilience of urban areas, informing the design of more liveable, sustainable, and adaptive urban environments

5.6. Limitations and future directions

Despite the promising results, this study has several limitations that need to be acknowledged. First, the representativeness and quality of POI data may affect the reliability of the findings. Future research should test and incorporate other POI datasets, as well as validate the results with field surveys and local knowledge. Second, the classification system of POIs may influence the interpretation of functional areas. More refined and context-specific classification schemes should be developed to better capture the nuances of urban functions. Third, the study focuses on a single case of London, limiting the generalizability of the findings. Comparative studies across different cities and regions are needed to uncover the common patterns and variations in urban functional organization.

Moreover, future research should extend the analysis to incorporate dynamic network methods to investigate the evolution and transformation of urban functional areas over time (Jin, Wang, Ge, & Yan, 2023). Integration with other urban datasets, such as population, socioeconomic, social media and mobility data, can provide a more comprehensive understanding of the drivers and impacts of urban functional changes. Advanced *NLP* techniques, such as deep learning-based word embeddings (Grootendorst, 2022), can also be applied to improve the accuracy and granularity of POI semantic mining and possibly the use of language models for next POI prediction (Brown et al., 2020).

In conclusion, this study demonstrates the great potential of integrating topic modelling and network analysis methods for street-level urban functional area identification and characterisation using POI data. The proposed approach offers a novel and powerful tool for urban planners, geographers, and policymakers to uncover the underlying patterns and mechanisms of urban spatial organization, supporting evidence-based and human-centric urban planning and management practices. Future research should continue to explore the methodological innovations, theoretical implications, and practical applications of this approach in various urban contexts, contributing to the understanding and the development of more sustainable, resilient, and liveable cities.

6. Conclusion

This study presents a novel approach to identifying and characterizing urban functional areas by integrating topic modelling and network analysis methods, using Points of Interest (POI) data in Greater London. Crucially, this research emphasizes the importance of street-level analysis, providing a more nuanced and human-centric understanding of urban functions compared to traditional methods based on administrative boundaries or grid cells. The proposed framework offers a data-driven and multi-dimensional perspective to uncover the latent semantic structure and spatial organization logic of urban functions at the street level. This street-based approach aligns more closely with how people experience and interact with urban spaces in their daily lives, offering insights that are both more granular and more relevant to real-world urban dynamics. The main contributions of this research include: 1) A methodological innovation that combines topic modelling and network analysis techniques to investigate urban functional areas from both semantic and structural perspectives, with a focus on street-level patterns; 2) The construction of multi-layer networks to bridge the gap between thematic, functional, and spatial dimensions of urban space, reflecting the complexity of street-level urban interactions; 3) A case study of London demonstrating the effectiveness of the proposed approach in revealing the spatial distribution patterns and network characteristics of different functional Topics at the street level.

The results show that the proposed approach can effectively identify and delineate urban functional areas based on the co-occurrence patterns and network structure of POI "vocabularies" along streets. This street-level analysis captures the fine-grained functional mix and spatial relationships that characterize urban neighbourhoods, providing a more accurate representation of urban functionality than broader-scale approaches. The network analysis further reveals the structural properties and interaction patterns of different Topics, providing valuable insights into the roles and positions of different POI types in the functional organization of urban space. This street-based network perspective highlights the complex interactions and interdependencies among POIs that occur at the human scale of urban experience. This study underscores the necessity and value of both the network perspective and street-level analysis in understanding urban functional areas. By focusing on street-level patterns, the research captures the complex interactions and interdependencies among POIs that form the fabric of urban life, offering insights that are more directly applicable to urban design and community planning. However, the limitations of data quality, classification systems, and network construction should be acknowledged and addressed in future research. Additionally, further work could explore how this street-level approach can be integrated with other scales of urban analysis to provide a comprehensive understanding of urban functionality.

Code availability

The code used for analysis in this study is available from the corresponding author upon reasonable request.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Funding

This research was funded by Outstanding Young Scientists Program for Beijing Universities, grant number JJWZYJH01201910003010.

CRediT authorship contribution statement

Xuhui Lin: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Tao Yang:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision, Project administration. **Stephen Law:** Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization.

Declaration of competing interest

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

Data availability

The data that support the findings of this study are available from Edina-digiMAP in the United Kingdom and are publicly available.

References

Balland, P.-A., Jara-Figueroa, C., Petralia, S. G., Steijn, M. P., Rigby, D. L., & Hidalgo, C. A. (2020). Complex economic activities concentrate in large cities. *Nature Human Behaviour, 4*(3), 248–254.

Batty, M. (2013). *The new science of cities*. MIT press.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993–1022.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., , … Askell, A., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901.

Chen, Y., Chen, X., Liu, Z., & Li, X. (2020). Understanding the spatial organization of urban functions based on co-location patterns mining: A comparative analysis for 25 chinese cities. *Cities, 97*, Article 102563.

Crooks, A., Pfoser, D., Jenkins, A., Croitoru, A., Stefanidis, A., Smith, D., Karagiorgou, S., Efentakis, A., & Lamprianidis, G. (2015). Crowdsourcing urban form and function. *International Journal of Geographical Information Science, 29*(5), 720–741.

Gao, S., Janowicz, K., & 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.

Gehl, J. (2010). *Cities for people*. Island Press.

Grootendorst, M. (2022). *Bertopic: Neural topic modeling with a class-based tf-idf procedure*. arXiv preprint. arXiv:2203.05794.

Hidalgo, C. A., Castaner, E., & Sevtsuk, A. (2020). The amenity mix of urban neighborhoods. *Habitat International, 106*, Article 102205.

Hillier, B., & Hanson, J. (1984). *The social logic of space*. Cambridge University Press.

Huang, Y., Shekhar, S., & Xiong, H. (2004). Discovering colocation patterns from spatial data sets: A general approach. *IEEE Transactions on Knowledge and Data Engineering, 16*(12), 1472–1485.

Jacobs, J. (1961). *The death and life of great American cities*. Random House Inc.

Jin, M., Wang, L., Ge, F., & Yan, J. (2023). Detecting the interaction between urban elements evolution with population dynamics model. *Scientific Reports, 13*(1), Article 12367.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal, 79*, 1–14.

Lansley, G., & Longley, P. A. (2016). The geography of twitter topics in London. *Computers, Environment and Urban Systems, 58*, 85–96.

Law, S. (2017). Defining street-based local area and measuring its effect on house price using a hedonic price approach: The case study of metropolitan London. *Cities, 60*, 166–179.

Liu, K., Yin, L., Lu, F., & Mou, N. (2020). Visualizing and exploring poi configurations of urban regions on poi-type semantic space. *Cities, 99*, Article 102610.

Liu, X., & Long, Y. (2016). Automated identification and characterization of parcels with openstreetmap and points of interest. *Environment and Planning. B, Planning & Design, 43*(2), 341–360.

Long, Y., & Shen, Z. (2015). Discovering functional zones using bus smart card data and points of interest in Beijing. In *Geospatial analysis to support urban planning in Beijing* (pp. 193–217).

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. arXiv preprint. arXiv:1301.3781.

Morioka, W., Okabe, A., Kwan, M.-P., & McLafferty, S. L. (2022). An exact statistical method for analyzing co-location on a street network and its computational implementation. *International Journal of Geographical Information Science, 36*(4), 773–798.

Niu, H., & 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*, Article 101651. https://doi.org/10.1016/j.compenvurbsys.2021.101651

Psyllidis, A., Gao, S., Hu, Y., Kim, E.-K., McKenzie, G., Purves, R., Yuan, M., & Andris, C. (2022). Points of interest (poi): A commentary on the state of the art, challenges, and prospects for the future. *Computational Urban Science, 2*(1), 20.

Singleton, A., Alexiou, A., & Savani, R. (2020). Mapping the geodemographics of digital inequality in great britain: An integration of machine learning into small area estimation. *Computers, Environment and Urban Systems, 82*, Article 101486.

Tao, H., Wang, K., Zhuo, L., & Li, X. (2019). Re-examining urban region and inferring regional function based on spatial– Temporal interaction. *International Journal of Digital Earth, 12*(3), 293–310.

Xu, Y., Zhou, B., Jin, S., Xie, X., Chen, Z., Hu, S., & He, N. (2022). A framework for urban land use classification by integrating the spatial context of points of interest and graph convolutional neural network method. *Computers, Environment and Urban Systems, 95*, Article 101807.

Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., & Mai, K. (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.

Yu, W. (2016). Spatial co-location pattern mining for location-based services in road networks. *Expert Systems with Applications, 46*, 324–335.

Yuan, J., Zheng, Y., & Xie, X. (2012). Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 186–194).

Zhong, C., Huang, X., Arisona, S. M., Schmitt, G., & Batty, M. (2014). Inferring building functions from a probabilistic model using public transportation data. *Computers, Environment and Urban Systems, 48*, 124–137.

Zhong, C., Zeng, S., Tu, W., & Yoshida, M. (2018). Profiling the spatial structure of London: From individual tweets to aggregated functional zones. *ISPRS International Journal of Geo-Information, 7*(10), 386.

Zhou, X., & Zhang, L. (2016). Crowdsourcing functions of the living city from twitter and foursquare data. *Cartography and Geographic Information Science, 43*(5), 393–404.