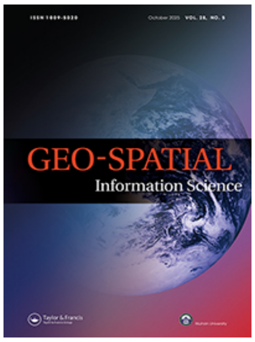

Taylor & Francis
Taylor & Francis Group

Geo-spatial Information Science

ISSN: 1009-5020 (Print) 1993-5153 (Online) Journal homepage: www.tandfonline.com/journals/tgsi20

Semantic co-location model: a novel approach to explore the urban spatial structure with a social phenomenon

Shi Zeng, Xiaoliang Meng, Chen Zhong, Yao Shen, Yichun Xie & Michael Batty

To cite this article: Shi Zeng, Xiaoliang Meng, Chen Zhong, Yao Shen, Yichun Xie & Michael Batty (23 Dec 2025): Semantic co-location model: a novel approach to explore the urban spatial structure with a social phenomenon, Geo-spatial Information Science, DOI: 10.1080/10095020.2025.2601314

To link to this article: https://doi.org/10.1080/10095020.2025.2601314

Published online: 23 Dec 2025.

GEO-SPATIAL INFORMATION SCIENCE
https://doi.org/10.1080/10095020.2025.2601314

Semantic co-location model: a novel approach to explore the urban spatial structure with a social phenomenon

Shi Zeng [a,b], Xiaoliang Meng [a], Chen Zhong [b], Yao Shen [b,c], Yichun Xie [d] and Michael Batty [b]

[a]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China; [b]Centre for Advanced Spatial Analysis, University College London, London, UK; [c]College of Architecture and Urban Planning (CAUP), Tongji University, Shanghai, China; [d]Department of Geography and Geology, Eastern Michigan University, Ypsilanti, MI, USA

ABSTRACT

This study explores the intersection of spatial co-location patterns and social phenomena through an innovative analysis of Twitter data, addressing a gap in existing spatial co-location research that predominantly focuses on geographical phenomena. Spatial co-location pattern analysis is fundamental to understanding spatial data and enhancing geographic context-awareness in applications. While traditional studies have concentrated on identifying spatial proximity of physical features to discern interactions among geographical phenomena, this research integrates social phenomena, acknowledging the intrinsic relationship between geographic and social dynamics. Through the analysis of georeferenced Twitter data, this study identifies spatial features associated with social interactions and activities, providing a comprehensive understanding of social-spatial interplay. The research introduces an innovative Semantic Co-Location (SCL) model to analyze spatial co-location patterns from individual tweets at aggregated spatial levels. This includes developing spatial co-location mining techniques, analyzing topical categories of spatial co-location based on contextual information, and uncovering previously unknown patterns that expand current research boundaries. The findings advance our understanding of urban discourse and illuminate the relationship between place and people, specifically within spatial and social networks.

ARTICLE HISTORY
Received 21 January 2025
Accepted 1 December 2025

KEYWORDS
Spatial analysis; co-location; spatial co-location; semantic co-location; urban studies

1. Introduction

Cities are constantly changing, with urban elements interacting through both physical configurations and spatial forms, as well as through digital interactions, all of which contribute to creating complex systems (Batty 2008, 2018). Urban spaces are not merely physical constructs, but they are unique vibrant ecosystems where social interactions, economic transactions, and cultural expressions converge (Batty 2024). In the evolving landscape of urban studies, the concept of co-location – the phenomenon where activities or social interactions occur in close spatial proximity, regardless of the temporal dimension, has emerged as a vital area of inquiry.

Most prior research has concentrated on identifying spatial co-location patterns, characterized by subsets of features whose occurrences are frequently found in close spatial proximity (Liu et al. 2022; Morioka et al. 2022; Shekhar and Huang 2001). However, those studies have focused more on the co-location of objects in space, whereas this study is primarily motivated by the research gap in understanding the co-location of both spatial and social phenomena, and will more extensively use the definition of co-location to explore this relationship.

People gathering in urban spaces is not merely a matter of physical proximity; it also reflects underlying social issues and the quality of life in cities. As Giddens (1984) argues, co-presence is essential for the maintenance of even the most complex social structures, as it facilitates interaction and cohesion within societies. This foundational concept, however, can vary in definition depending on the field, methodology, and spatial scale in question. While terms such as "co-existence" and "co-occurrence" may not traditionally be linked with urban studies, their application in urban analysis reveals a profound connection to co-presence and co-location, illustrating how these concepts evolve into broader social phenomena. Here, the difference between co-presence and co-location is that co-presence refers to the simultaneous presence of individuals or entities in the same space at the same time, often implying a more real-time or interaction-based relationship. On the other hand, co-location focuses more on the spatial proximity of entities regardless of the temporal dimension. As cities evolve, the spaces where people choose to come together reveal significant insights into the social fabric and the effectiveness of urban design in meeting people's needs (Silver and Clark 2016). Researches such as

CONTACT Xiaoliang Meng xmeng@whu.edu.cn

analyzing the spatial co-occurrence of Points of Interest (POIs) and co-location analysis for better search results in Public Map Service Platforms (Dong et al. 2024; Zhao, Zhu, and Qin 2025), as well as measuring spatial similarity among COVID-19 epicenters, each offer a unique perspective for understanding the dynamics of urban issues (Kaffash et al. 2023).

In this context, co-location can be viewed as a subset of "co-existence" and 'co-occurrence." However, the concept of co-location varies widely across different disciplines, methodologies, and spatial scales. While co-location in public urban spaces does not always involve intentional gatherings, factors such as friendship ties and small group behaviour are critical elements in co-location studies (Xu et al. 2017). Conversely, in telecommunications and network infrastructure planning, 'co-location' refers to the practice of housing multiple tenants" servers or equipment within a single data center facility. Here, the focus is on physical security, power/cooling efficiency, and network interconnectivity at the building or rack level, with little consideration for the social interactions central to urban studies. Traditionally, much of the research has focused on physical, "real-world" co-location, and the importance of virtual co-location (defined as technology-mediated shared presence across geographically dispersed settings (e.g. video calls, shared digital spaces)) is increasingly recognized. Technological improvements have enabled more distant methods of communication and collaboration. They encompass both the condition of face-to-face interactions in tangible spaces and behaviors shared through virtual communications, like "tagging" on social media platforms, tele- and video-conferencing, e-mail, and chats (Bahrehdar and Purves 2018; Carmody et al. 2022; Yan, Schultz, and Zipf 2019)

Existing explorations of co-location have laid a foundation, but they often fail to fully capture the interplay between people's engagement in physical and digital realms and the inherent social logic of urban space. This limitation means that our understanding of how proximity and social interaction shape urban life remains incomplete – especially regarding how spatial contexts and gathering patterns mutually influence each other. Expanding the concepts of co-location thus becomes necessary to address this gap, as it can enrich the comprehension of these dynamics.

To contribute to filling this gap, this study primarily focuses on two aspects of these transformations: a) space (location) itself, which we treat not as a neutral backdrop but as a carrier of social logic; and b) the way people gather within such spaces. Building on scholars' insights that location semantics can be examined from spatial, temporal, and thematic dimensions (Hu 2018; Zhu et al. 2019), we aim to advance this line of inquiry. By using Twitter data and POI data, this paper proposes a novel method for studying co-location between individuals – the Semantic Co-location (SCL) model – with the goal of establishing a bridge between the social and spatial dimensions that existing research has yet to fully connect.

The rest of this paper is organized as follows. Section 2 reviews the related work and proposes a new strategy. Section 3 first defines and explains the concepts and scientific issues addressed in this study and then introduces the innovative SCL model. Section 4 presents the experimental evaluation and a case study on our Twitter and POI dataset. Finally, Section 5 concludes the paper and outlines future work.

2. Related works

Mining co-location patterns is a foundational task in spatial data mining, aimed at uncovering nonrandom associations among spatial features. A co-location pattern is defined as a subset of spatial features that frequently occur in geographic proximity, typically due to positive spatial interactions (Shekhar and Huang 2001). These patterns have been applied across domains including ecology (Keddy 2007), epidemiology, criminology, and urban studies – reflecting their versatility and importance.

For traditional co-location methods, classic approaches focused primarily on static spatial proximity, where co-location was detected using geometric thresholds and frequency-based pattern mining. For instance, studies have examined the spatial clustering of crimes (Li et al. 2022; Shiode, Shiode, and Inoue 2022), the distribution of public health risks (Iyer et al. 2023), and POI-based retail patterns. These methods typically rely on rule-based or distance-based algorithms to quantify spatial co-occurrence but are limited by fixed proximity thresholds and lack contextual interpretation.

Recent methods have introduced topological and network constraints to better capture co-location along urban infrastructures. Morioka et al. (2022) proposed a graph-theoretic approach that analyzes bilateral (mutual proximity where both entities benefit/influence each other, e.g. complementary businesses) and unilateral (one-sided proximity, e.g. a dependent facility near a hub) co-location patterns among store types on street networks, overcoming limitations of traditional cross-K functions. Similarly, Liu et al. (2022) developed adaptive, Monte Carlo-based methods to detect regional co-location patterns constrained by urban road networks, enhancing applicability in complex city structures.

Temporal dynamics have been incorporated through flow-based co-location models. Cai and Kwan (2022) analyzed commuting patterns to uncover occupational co-location biases, while Shen et al.

(2019) proposed the People-Space-Time (PST) model to measure face-to-face interaction probabilities based on temporal availability and spatial accessibility. These methods emphasize the importance of movement and interaction over time, but they often remain limited to structured mobility datasets.

A growing body of work now leverages crowd-sourced and user-generated content, such as social media, to reveal patterns of urban activity (Gong et al. 2021; Jing et al. 2023; Yan et al. 2020; Zheng, Han, and Sun 2018). Shwartz-Asher et al. (2020) and Carmody et al. (2022) discuss the role of digital interactions as proxies for co-location, expanding the concept to include virtual proximity and shared engagement. These data sources provide a rich, real-time layer of information about human presence and behavior, enabling more responsive interpretations of urban space.

However, the methods discussed above for mining co-location patterns face three challenges: first, establishing a suitable proximity threshold to recognize nearby instances in unfamiliar areas is challenging, and second, these approaches often overlook the influence of distance values between instances and the impact of instances at longer distances on the significance of the patterns. Finally, as we discussed in Section 1, although foundational works (Huang and Zhang 2006; Shekhar and Huang 2001; Yu 2016) formalize theoretical concepts of spatial proximity, and while traditional research methods are abundant, the application of new types of data and new research approaches still leaves much room for development (Li et al. 2017; Cai and Kwan 2022). This study mainly addresses the third gap directly by proposing a semantic spatial co-location framework that Leverages Topic Modeling (LDA) to infer thematic meanings from textual data (e.g. tweets) and combines them with POI distributions. This approach enables the detection of not just where entities co-occur, but what kind of social or functional meaning emerges from those co-locations. While this work also makes pragmatic contributions to the first two challenges, semantic integration forms its primary methodological contribution.

By shifting from fixed-feature datasets to crowd-sourced semantic content, this study introduces a method that is capable of uncovering latent social functions in urban space. This approach helps understand urban dynamics through the lens of community participation and engagement. While traditional spatial data provides foundational insights into fixed geographic features, crowdsourced data, particularly social media data, offers real-time information on human activities and interactions across urban spaces (Li et al. 2025). This shift allows for a deeper exploration of how people engage with different areas, complementing static spatial analysis with responsive, human-centered insights, and finally helps us to shift from spatial features to social meanings.

3. Methodology

This study looks at the bigger picture, focusing not just on lone individuals but on groups of people coming together. This framework reconceptualizes co-location, moving beyond the simple notion of spatial-temporal coincidence to encompass the shared purpose and activities of individuals. By shifting the analytical focus from individual entities to groups, this approach provides a more nuanced understanding of human clustering and spatial interaction patterns. Employing this methodology enables researchers to delve into the complex social dynamics that emerge when individuals congregate in urban settings. It transcends mere movement tracking by investigating how collectives perceive, engage with, and attribute meaning to shared spaces. Through this analytical lens, the intricate relationship between group behavior and urban geography is revealed, shedding light on how collective experiences shape the semantic landscape of cities. This framework underscores the multifaceted nature of place-making, where locations derive significance from the intersection of collective practices, shared intentions, and communal experiences.

This research unfolds in four distinct phases, beginning with the essential process of acquiring and refining data from Twitter interactions, and advancing toward the discovery of urban co-location as reflected in tweet patterns. Here is how it breaks down:

(a) **Translating tweets to topics**: The research initially involves the collection of Twitter data via the Twitter Streaming API from July 2015 to June 2017, focusing on the Greater London Authority (GLA). The Twitter API uses the JSON format to communicate with third-party apps. This study used geo-tweets only to extract spatial and temporal structures. Geo-tweets refer to those with valid coordinates, while this research excludes those with only a location tag. Other useful information includes user ID, tweet ID, creation time, and tweet content. The data has 3,043,753 records from 239,882 different users. The collected dataset undergoes a rigorous pre-processing stage, aimed at filtering out irrelevant, invalid, or automated content, thereby ensuring the integrity of the data for human-centric analysis.

(b) **Topics to functions**: The second phase is dedicated to the development of a novel analytical approach for topic modeling, specifically tailored to address geo-spatial questions. This involves the creation of a metric for determining the optimal number of topics, which in turn facilitates a more precise classification of entangled activity types and urban functions.

(c) **Functions to social phenomena**: By utilizing the pre-processed tweets, the third phase is that the study introduces an innovative measure of co-location, designed to capture both the geographical patterns of social gatherings and the cognitive perceptions of individuals associated with different urban locations. This is achieved by analyzing the content of a tweet's category to infer the hierarchical organization of urban spaces, thereby providing insights into the spatial distribution of various urban activities.

(d) **Social phenomena to structure insights**: In the final phase, the study also uses the POI dataset, which is taken from the Ordnance Survey (https://www.ordnancesurvey.co.uk/products/points-of-interest) to further delineate urban structure. After pre-processing, there are more than 850,000 records in the study area. Each record contains various information such as name, postcode, geographic county, and, most importantly, georeferencing. In addition, each record is classified into three-level categories. There are nine groups at the first level, which further expand into 52 subcategories and over 600 classes at the third level (the technical information can refer to their official website). This research focuses on the first level. This dual analysis enables a comprehensive exploration of spatiotemporal patterns in urban space usage, linking social phenomena with the physical geography of the city.

Through this approach, this research goes beyond distant observation of the city, employing detailed analysis of tweet content to understand the dynamics of urban spaces. The discovery uncovers the hidden patterns of how people share and experience our cities. Figure 1 illustrates the structure of this research framework.

The figure straightforwardly shows that the framework integrates thematic information (from topic modeling) with spatial distributions to identify activity types from pre-processed geotagged tweets. It has different topic categories for all geotagged tweets. Then, this study utilizes the categorized Twitter data to propose an innovative method to find out semantic co-location patterns and combines it with POI data to interpret the spatial structure.

The framework for co-location analysis in this study is designed to uncover patterns of human activity and urban functions by identifying where people are likely to gather with similar purposes. The primary aim is to understand the spatial organization of activities and explore the relationships between locations and the purposes or functions they serve. This approach provides insights into the dual questions of where certain activities are concentrated and what types of activities are most likely to occur in specific areas.

Co-location of different activity types is used as a way to identify purpose-driven spatial clusters in the city. By analyzing social media data and POI

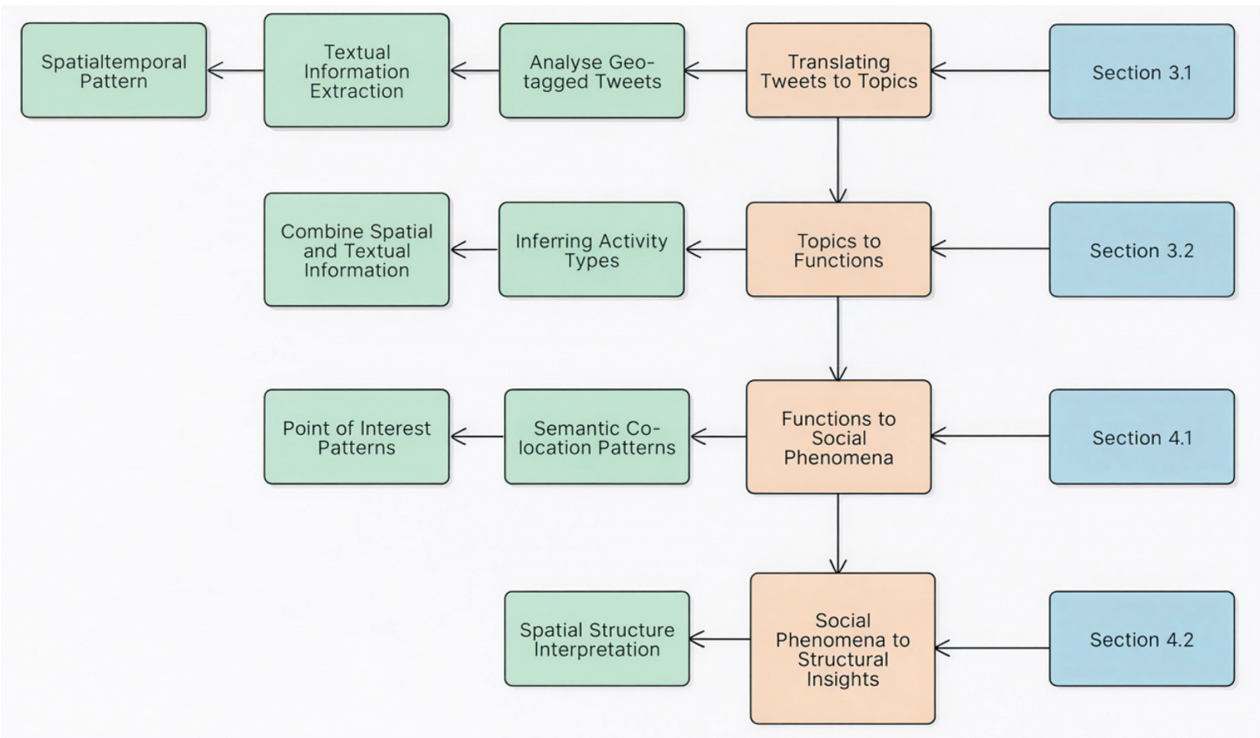

Figure 1. The general research framework for our semantic co-location model.

distributions, the study revealed areas where people converge with shared intentions, such as shopping, leisure, or commuting. For example, entertainment districts may show co-location patterns between topics related to nightlife and dining, highlighting their multifunctional nature. Beyond identifying gathering spots, the framework explores the types of activities that are more likely to occur in particular areas. This involves integrating semantic themes from social media with spatial clustering to infer implied activity types. For instance, co-location patterns might indicate that areas near parks are associated with recreational activities, while areas near transportation hubs support a mix of commuting and retail.

3.1. Indicators for selecting the number of topics

To extract the content within the Twitter dataset, the study employed the Latent Dirichlet Allocation (LDA), which is one of the most widely used topic modeling approaches (Blei et al. 2003). Briefly summarizing, LDA assumes that each document in a corpus contains a mixture of latent topics, and each word belongs to one of these topics. Words are treated as vectors, resulting in each topic having a unique word probability distribution. This facilitates the grouping of similar semantic information through the application of underlying mathematical techniques. A universally acknowledged challenge in topic modeling is the interpretation of these topics, a factor critical to selecting the most suitable topic model. The selection of the optimal topic model often relies on the coherence value, a metric evaluating the quality of the topics identified (O'Callaghan et al. 2015). Generally, a higher coherence value indicates superior topic quality. This study implemented a grid search to identify the best topic models and reached the same conclusion as that discussed by Fang et al. (2016).

While increasing the number of topics (k) generally leads to higher coherence values, a higher value does not necessarily guarantee a better model (Fang et al. 2016). This indicates that the spatial distribution of topics became increasingly random, complicating topic classification and interpretation while considering their associated spatial patterns. Introducing an excessive number of topics hinders in-depth analysis and obscures the identification of distinct urban functions and spatial patterns. To balance these considerations, the study adopts a global spatial autocorrelation measure in addition to the coherence value (Zhong et al. 2018). The selection of the optimal topic model involves balancing semantic coherence and spatial relevance. This process evaluates models across different numbers of topics (k) using two key metrics: coherence scores to assess the thematic consistency of topics and Moran's I (Moran 1950) to measure the spatial clustering of thematic patterns. Moran's I is a widely used statistic for spatial autocorrelation, and it is defined as:

$$I = \frac{n\sum_i \sum_j w_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{\left(\sum_{i\neq j} w_{ij}\right)\sum_i (Y_i - \bar{Y})^2} \tag{1}$$

where n is the number of regions; Y_i is the observed value of the variable of interest in region i; $\bar{Y}$ is the mean of all values; w_{ij} are the elements of a matrix of spatial weights between regions i and j, with $w_{ii} = 0$ and $i, j = 1, \ldots, n$. The definition of the spatial weights depends on the variable of study and the specific setting.

Then, Figure 2 shows the method of how to select the better topic number k. Specifically, if 10 topics are used, 10 autocorrelation values will be calculated, and their average is taken. A trade-off is observed, as increasing k often improves coherence but can fragment spatial relevance. By identifying the k-value where these metrics are balanced, the optimal model ensures that the topics are both interpretable and aligned with meaningful spatial distributions, forming the foundation for the semantic co-location analysis presented in this study.

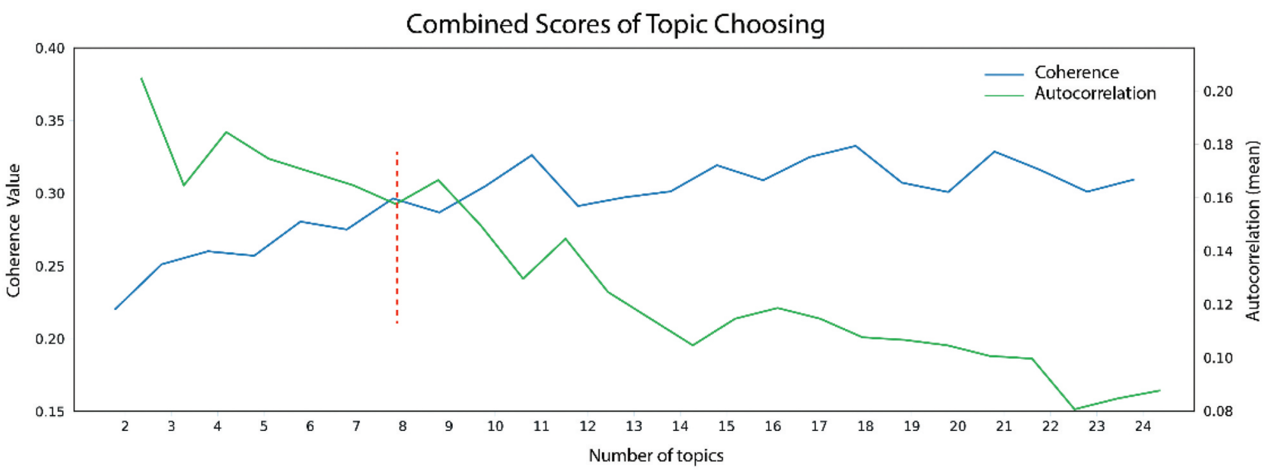

Figure 2. Method for the selection of the topic number.

The trade-off between the two becomes critical. First, areas with high Moran's I but low coherence scores might represent spatial clustering of activities or interactions that are not thematically related. For example, there might be a high concentration of tweets around a transportation hub, but if the topics being discussed are unrelated (e.g. food, entertainment), the coherence would be low. High coherence scores with low Moran's I suggest that the topics are thematically cohesive but do not exhibit strong spatial clustering. This could indicate that certain social or cultural activities are important across the entire city, but they do not concentrate in any specific geographic area.

The most ideally informative co-location patterns in urban dynamics would be those where both spatial clustering and thematic coherence are high. While this method lacks a strict mathematical proof, it provides a logical and potentially valuable approach for researchers using LDA or similar topic models across various fields, offering a new direction for balancing model quality and the interpretability of spatially embedded data. Understanding the trade-off is crucial for urban planners and policymakers, as it helps them recognize which areas of the city are not only socially active but also thematically focused. The reason the trade-off between Moran's I and the coherence score is so important is that it provides a holistic and innovative view of how urban spaces foster social interactions.

The study, therefore, selected the eight-topic model to balance topic coherence and spatial autocorrelation. This decision point is indicated by the red dashed line in Figure 2, which we have included specifically to guide the reader in identifying the closest topic number for each intersection point. The plot that overlaps Moran's I and coherence score visually demonstrates the balance between where co-location occurs (spatial clustering) and what the co-located activities represent (semantic coherence). Table 1 shows the outlines of the clusters of eight topics generated by LDA, showcasing the representative words within each topic, and delineates the mapping from words to topics alongside their correlated urban functions.

This is followed by a wordcloud (Figure 3) to visualize textual data and allow researchers to quickly identify key themes, concepts, and the relative importance of words within a given dataset. In research, especially in areas like content analysis or topic modeling, word clouds help to provide an intuitive, at-a-glance understanding of large volumes of text, facilitating the discovery of patterns and insights that might otherwise be overlooked. This makes them particularly useful for summarizing qualitative data and enhancing the accessibility of complex findings.

These results revealed that each tweet document is characterized by a probability distribution that suggests its association with multiple topics. Given the brevity of most tweets, it is assumed for analytical simplicity that each tweet is primarily influenced by a single dominant topic, the one with the highest probability. Based on this approach, tweets corresponding to the top eight identified topic groups are assigned specific labels reflecting these topics, while tweets that did not align clearly with these groups are categorized under a generic "other topics" label. Subsequently, each tweet is tagged with a single label that best represented the activity or theme most relevant to its content. These labels, referred to as activity labels in this study, play a crucial role in categorizing tweet content within the framework of the research.

As a form of verification, the coherence of detected keywords aligns with findings from previous studies (Lansley and Longley 2016; Steiger, Resch, and Zipf 2016), guiding the topic labeling process based on previously defined labels. Urban function classification

Table 1. The interpretation of words in topics and the corresponding urban functions.

ID	Topic	Words in the topic	Urban functions/land use type	Moran's I	Ratio (%)
1	Socialising	Good, great, love, day, night, amazing, dance, event, gallery, visit, show, school	Recreation and leisure/ community services	0.071	11.74
2	Events	Time, think, come, year, today, start, ready, people, tomorrow, way, feel, leave, excited	Recreation and leisure/retail	0.094	10.31
3	Daily routines	work, back, morning, home, run, nice, week, session, walk, sunday, long, class, train, office, sun, hour, finish, gym, Monday, training, start, break, hard, early, follow	Residential/industry and business/ transport	0.167	18.34
4	Sight and view	Photo, house, bar, place, park, street, view, road, stop, gym, Tesco, TFL, check, road, free	Recreation and leisure/retail	0.164	12.14
5	Lifestyle	Big, watch, live, play, man, club, boy, head, baby, moment, video, dream, heart, picture, win, listen, rock	Recreation and leisure/retail	0.067	8.52
6	City hub	station, bridge, victoria_station, central, royal, tower, wembley, hospital, square, town, Camden, Greenwich, Kensington, road	Recreation and leisure/retails/ transport	0.458	19.17
7	Food and drink	Drink, lunch, food, coffee, restaurant, dinner, beer, eat, ale, pale, arm, pub, brewery, cider	Retail	0.102	10.23
8	Sports and games	Finish, run, park, walk, cycle, miles, game, football, club, arsenal, win, workout, season, stadium, team	Residential/ recreation and leisure	0.169	9.55

Note: The words in the topic are simply the most representative ones ranked by probability – they do not represent all the words in the topic.

Figure 3. A wordcloud of Twitter topics.

draws from the UK's National Land Use Classification (https://assets.publishing.service.gov.uk/media/5a78b635ed915d07d35b1d52/144275.pdf), which specifies 14 principal land uses and their further subdivisions into sub-types. The effort to directly map these activities to urban functions revealed the complexity of achieving a one-to-one correspondence; activities identified could span multiple urban functional areas. However, an exception is found with topic 7 – food and drink – which straightforwardly correlates with the retail function. Conversely, other topics demonstrate a one-to-many (1-to-N) relationship. For example, topic 3, labeled "routine activities," amalgamates work, education, residential activities, and more identifiable through keyword analysis. Thus, the urban functions corresponding to these activities could encompass residential areas, offices, educational institutions, and transit routes to various destinations, indicating multifunctional zones characterized by high people flow. Keywords also spotlight significant transport hubs and popular tourist sites, underscoring the multifaceted nature of urban functionality.

3.2. The SCL model

This research delineates a novel integrated framework, termed the semantic co-location model, devised to elucidate the intricacies of interactions among individuals discussing similar subjects on social platforms. This model has four pivotal dimensions: spatial (location), temporal (time), social (people), and thematic (content of tweets), each offering a distinct perspective on social dynamics.

Spatial dimension: This first dimension quantifies the physical location of individuals, positing that the geographical proximity significantly influences the propensity for social connections. It underscores the role of physical space in facilitating or hindering social interactions.

Temporal dimension: It captures temporal variance in human activities, reflecting how differences in time-specific behaviors (e.g. travel patterns) can impact social engagements. This dimension acknowledges that synchronicity in time is crucial for enabling interactions.

Social dimension: Represented through the "people" aspect, this dimension addresses the diversity among individuals, typically measured by demographic characteristics. It highlights how personal attributes can mediate the formation of social ties.

Thematic dimension: This focuses on the intent or purpose of the interaction. This is the most important dimension as it allows the approach to distinguish from most existing work on co-location. It suggests that the nature of the content shared by individuals can be a significant predictor of their spatial and temporal convergence.

The interplay between spatial and temporal proximity is especially emphasized, as it substantially affects social interactions. Spatial proximity delineates the necessity for individuals to traverse physical distances to engage, whereas temporal proximity necessitates their presence within the same timeframe for interaction. The model posits that the likelihood of face-to-face encounters escalates as individuals exhibit closer spatial and temporal proximity, a hypothesis supported by extensive research in the field.

This study adopts an innovative approach to discern co-location patterns, opting for a grid-based analysis. After experimenting with various grid sizes ranging from $100\,\mathrm{m}\times100\,\mathrm{m}$ to $500\,\mathrm{m}\times500\,\mathrm{m}$, this research selected the $300\,\mathrm{m}\times300\,\mathrm{m}$ grid. The decision to use a $300\,\mathrm{m}\times300\,\mathrm{m}$ grid size for the co-location analysis is based on a series of preliminary tests evaluating different grid resolutions, ranging from $100\,\mathrm{m}\times100\,\mathrm{m}$ to $500\,\mathrm{m}\times500\,\mathrm{m}$. The $300\,\mathrm{m}\times300\,\mathrm{m}$ grid is chosen because it strikes a balance between capturing detailed spatial behaviors, managing the computational and analytical limitations of the data, and ensuring the readability of the visualizations. Smaller grids (e.g. $100\,\mathrm{m}\times100\,\mathrm{m}$) could provide greater detail but risk fragmenting meaningful patterns, especially for activities distributed across multiple small areas. Larger grids (e.g. $500\,\mathrm{m}\times500\,\mathrm{m}$) aggregate more data but reduce the resolution of the detected patterns, potentially obscuring localized co-location phenomena. The potential for co-location clusters to be split at the grid boundaries is acknowledged as an inherent limitation of grid-based analyses. Future enhancements could use overlapping grids or adaptive grid sizes in areas with high activity density to further address edge effects.

In this approach, each tweet of an individual user can be considered as a list of tuples $\{(l_1,t_1),(l_2,t_2)\ldots(l_n,t_n)\}$, where l indicates the user's location at time t. The study first defines the probability of a user x at location l at time slot Δt as follows:

$$P_x^{(l,\Delta t)} = m_x^{(l,\Delta t)} \div n_x^{\Delta t} \tag{2}$$

where $m_x^{(l,\Delta t)}$ is calculated as the cumulative times of all appearance occurring to location l during a certain time-period Δt. The relationship between m and n can be described as follows:

$$n_x^{\Delta t} = \textstyle\sum_{\Delta \mathrm{l}\in \mathrm{Loc}(T)} m_x^{(l,\Delta t)} \tag{3}$$

where $\mathrm{Loc}(T)$ is the set of all locations visited by user x during time-period Δt.

To better explain these equations, this could be represented by all the tweets posted by User A in Chinatown between 6 and 7 pm. $n_x^{\Delta t}$ indicates the total number of appearances at any location during time-period Δt of a user. For the same example, it shows all tweets posted by User A at any location in London (the study area) between 6 and 7 pm. $P_x^{(l,\Delta t)}$ therefore indicates the probability that User A is in Chinatown between 6 and 7 pm based on all his tweets. However, this study divided the study area into standard grid cells instead of irregular areas like Chinatown.

After assigning subjects to groups, the study next measures the co-location of users. Noting that the focus of this model is not to measure physical co-location (face-to-face encounter). Instead, a different approach to co-location is proposed. This research first defines the equation of co-location, as follows:

$$\mathrm{Col}_{x,y}^{l} = P_x^{(l,\Delta t)} \times P_y^{(l,\Delta t)} \tag{4}$$

where it does not measure the time of each day but instead considers the timeslot at an aggregate level. In other words, this approach assumes that if people are in the same grid during the same period (regardless of date), they exhibit co-location behavior. In this study, each day represents a parallel event in time, meaning that people do not need to physically encounter each other. For instance, $\mathrm{Col}_{x,y}^{l}$, the parameter can indicate the probability that both User A ($P_x^{(l,\Delta t)}$) and User B ($P_y^{(l,\Delta t)}$) are present in Chinatown (l) between 6 and 7 pm (Δt) in the evening. However, this definition is not entirely adequate because it is arguable that people are at those locations randomly. User A went to Chinatown for food and posted a tweet about it and User B otherwise posted a tweet about how crowdy this place is. In this case, User A and User B went to Chinatown between 6 and 7 pm, and it is crucial to remember that the dates they visited Chinatown are very likely to be different, which suggests that their purposes for visiting Chinatown are not the same. Therefore, in this study, they are not considered to be co-located. In order to improve the detection methods, the study further developed the approach as follows:

If two users ("people") are close to each other geographically ("location") during certain time-period ("time") and they mention similar information ("content C_i"), then the study defines this scenario as SCL:

$$\mathrm{SCL}_{x,y}^{l,t} = \mathrm{Col}_{x,y}^{l,t} \cdot 1_{C_x=C_y} \tag{5}$$

where users and locations:

- x, y : Two users.
- l_x, l_y: Locations of users x and y.

Time:

- t_x, t_y: Times at which users x and y are active.

Topics:

- C_x, C_y: Content associated with user x and y.
- $C_x, C_y \in \{1, 2, \ldots, 8\}$: content is defined by eight different topics generated by LDA.
- $1_{C_x=C_y}$ or C_i : indicator function that equals 1 if $C_x = C_y$ (users share the same topic), and 0 otherwise.

Using the same example, the $\mathrm{SCL}_{x,y}^{l,t}$ only records when User A and User B went to Chinatown between 6 and 7 pm, and they both tweeted about food (or other similar topics).

Finally, this research calculates the accumulative SCL for group of users $U = \{u_1, u_2 \ldots u_n\}$: in grid cells and then divides this by total number of unique pairs $\binom{N}{2}$ in the grid to get the SCL intensity (SCL_i) for each location, as shown below:

$$\mathrm{SCL}_i^{l,t} = \frac{\sum_{x=1}^{N} \sum_{y=x+1}^{N} \mathrm{SCL}_{x,y}^{l,t}}{\binom{N}{2}} \tag{6}$$

The inclusion of the SCL transforms the co-location analysis from a purely spatial exercise to a comprehensive examination of spatial, temporal, social, and thematic dimensions. This innovation enhances the model's capability to link urban dynamics with shared human behaviors, providing insights that are not only academically valuable but also practically relevant for urban design and policy-making. In the following sections, we will use the value of SCL_i to explore which areas in London people are more likely to visit for the same reasons and are therefore more likely to be co-located.

4. Results and analysis

In this exploration, the concept of co-location is used as a pivotal tool to unravel the intricate ties binding place to tweet content, positing co-location not merely as a factor but as a linchpin in the web of social connectivity. Co-location emerges as a significant social construct, one that underpins the establishment and reinforcement of social bonds among individuals. This section explores the social dynamics of London's urban spaces through two complementary aspects. First, we examine the temporal and spatial patterns of semantic co-location intensity across different times of day, revealing how social interactions evolve throughout the urban landscape. By analyzing these patterns at key temporal intervals (3:00–4:00 am, 8:00–9:00 am, 6:00–7:00 pm, and 11:00 pm–12:00 am), we uncover the rhythms of urban social life and identify distinct behavioral signatures in different areas of the city. We then extend this analysis by comparing these patterns with the distribution of POIs across London, offering insights into the relationship between physical urban infrastructure and patterns of social interaction. This dual approach allows us to not only map where and when social interactions occur but also to understand how these patterns relate to the built environment. Through this analysis, we aim to reveal both expected alignments and surprising disconnects between urban form and social function, providing valuable insights for urban planning and policy-making.

Drawing inspiration from Aristotle's assertion that "Man is by nature a social animal; an individual who is unsocial naturally and not accidentally is either beneath our notice or more than human," this investigation delves into the profound implications of social differences on human interactions. These differences reveal the diverse tapestry of individual variances, which, in turn, shape their motives for movement and social engagement. It is the pursuit of fulfilling specific needs that propels individuals into motion, with the utility of a location acting as a pivotal draw. Therefore, a symbiotic relationship unfolds between the demand, manifested in people's motives and movements, and the supply, represented by the functionalities of various locations. This dynamic interplay between social behavior and spatial configuration serves as a testament to the complex dialog between social dynamics and urban spaces.

4.1. Semantic co-location intensity in London

According to the last section, semantic co-location intensity ($\mathrm{SCL}_i^{(l,t)}$) shows the accumulative density of co-location behavior in each grid cell. It is necessary to emphasize that the heatmap here does not represent the activity level of Twitter data. In fact, according to the calculation formula of $\mathrm{SCL}_i^{(l,t)}$, a high $\mathrm{SCL}_i^{(l,t)}$ indicates that the topics people discuss in this location are relatively similar, or the categories of what they discuss are closer or more related. Conversely, a low $\mathrm{SCL}_i^{(l,t)}$ suggests that the topics discussed at this location are diverse, objectively indicating that this place may have buildings or facilities with more varied functions. It could also mean the volume of Twitter usage is low, as most suburban areas reveal in the following figure.

Figure 4 shows the spatiotemporal evolution of semantic co-location intensity ($\mathrm{SCL}_i^{(l,t)}$), capturing the dynamic interplay of social interactions across distinct time periods within a day, and aligning with the exploration of urban form and function through the lens of social media data. This figure highlights four pivotal time periods that epitomize the diurnal pulse of Twitter activity and its implications for understanding urban dynamics.

The time between 3:00 am and 4:00 am is usually the quietest hour for Twitter usage, reflecting a general decrease in social media activity that aligns with the sleep patterns of most urban residents. In contrast, the 8:00 am to 9:00 am period represents the morning rush hour, a time of increased activity that provides insights into the temporal and spatial aspects of urban life, especially regarding commuting patterns and the city's social awakening. After 9:00 am, the global city center becomes the focus, reasserting its activity and influence throughout the day, reflecting the flow of urban life. This continuous activity highlights the importance of urban cores as centers of social and economic interaction, which is also evident in the clustering of semantic co-locations within these areas. The time from 6:00 pm to 7:00 pm is one of the most popular periods for Twitter, indicating a significant increase in social media engagement. This period likely represents the end of daily activities, including the evening commute, leisure activities, and the beginning of nighttime socialization, providing rich data for analyzing urban social patterns. Finally, the time from 11:00 pm to 12:00 am is chosen to investigate nighttime behaviors, revealing the nocturnal aspect of urban life and the continuation of social interactions beyond typical hours. This examination of nighttime activity patterns supplements the daytime analysis, offering a comprehensive understanding of the rhythm of urban life as reflected in social media usage.

The 1-h interval strikes a balance between capturing meaningful temporal variations and maintaining sufficient data density within each bin. Shorter intervals might fragment data, leading to sparsity, particularly during periods of lower social media activity (e.g. early morning hours). Longer intervals could mask rapid changes in co-location dynamics, such as transitions

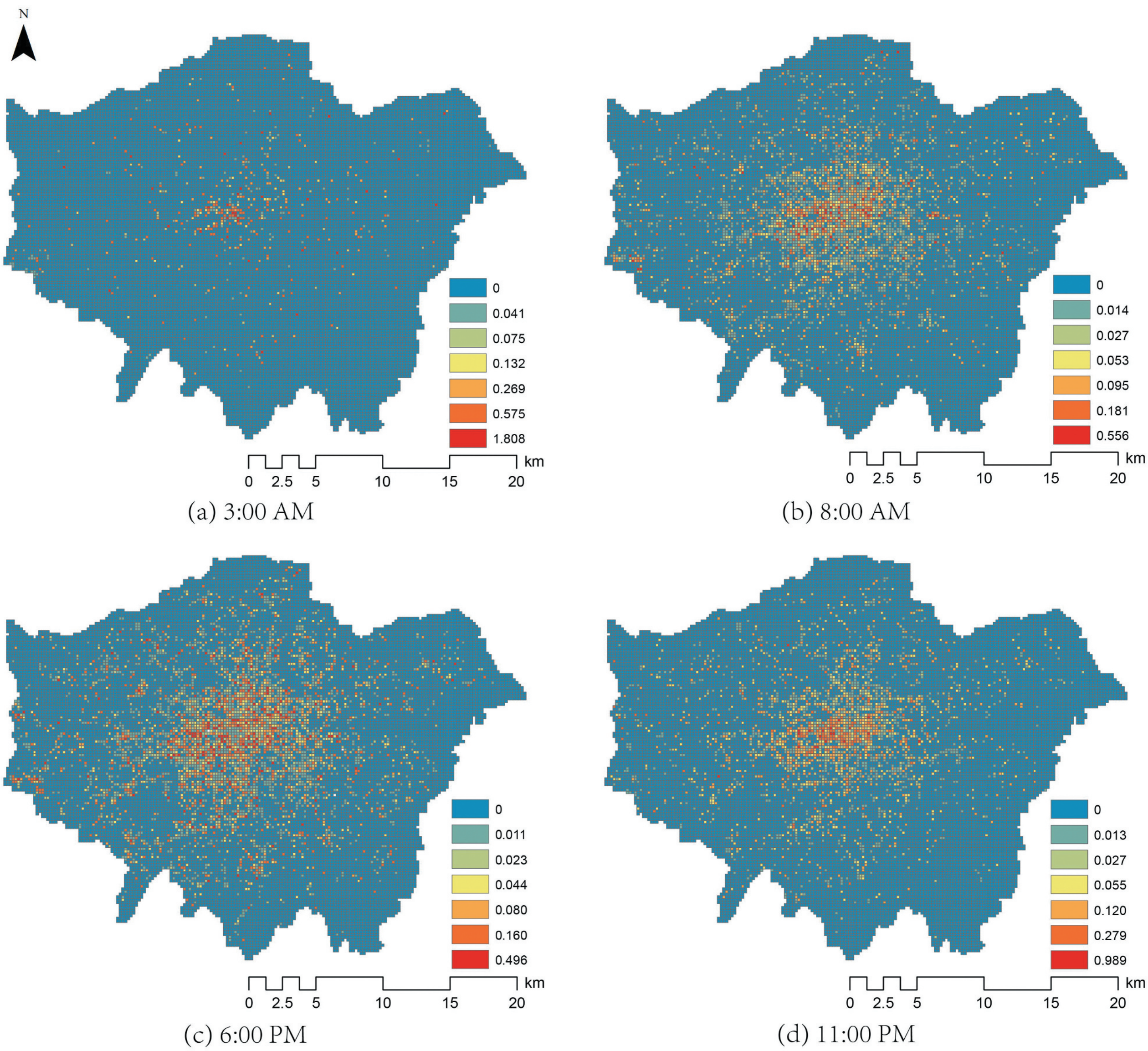

Figure 4. Temporal signatures of semantic co-location intensity ($SCL_i^{(l,t)}$).

between commuting and leisure activities. It is worth mentioning that edge effects arise when activities or tweets occurring near the boundary of a bin (e.g. 6:59 pm vs. 7:00 pm) are assigned to separate intervals, potentially fragmenting clusters that span multiple bins. Future iterations could incorporate overlapping time bins or smoothing techniques to account for continuity across bin boundaries. This would reduce the impact of sharp temporal cutoffs and ensure that co-location clusters are not artificially divided. Preliminary explorations might include testing longer intervals (e.g. 2-h or 4-h bins) to evaluate the trade-off between capturing temporal detail and aggregating sufficient data for robust pattern detection. Through the temporal lens of these selected periods, this study leverages the semantic co-location model to unravel the complex relationship between urban spaces and the social dynamics that animate them, offering nuanced insights into how digital footprints on social media platforms like Twitter can illuminate the multifaceted nature of urban living.

The analysis also reveals the dynamic spatiotemporal patterns of urban space, illustrating a shift from a centric to a more polycentric structure of co-location behavior throughout the day. In the early morning, the Soho area emerges as a focal point of activity, reflecting its vibrant nightlife. In contrast, daytime populous areas like Bloomsbury, Holborn, and the City of London exhibit a diminished co-location intensity, with the majority of grid cells showing minimal to no co-location patterns. As commuting hours commence, the distribution of co-location intensity becomes spatially more homogeneous, mirroring the widespread dispersal of people transitioning between residential areas and workplaces. The analysis further identifies a regeneration of activity in the global urban center post-morning and evening peaks, maintaining a central trend of dominance, especially during midday. This pattern is illustrated in Figure 4(c), where central dominance peaks alongside numerous dispersed clusters extending from the city center to the suburbs. This dispersion aligns with the evening commuting peak at 6:00 pm, marking a period of outward movement from the city center across the city. As the evening progresses, the number of hotspots diminishes. The central area experiences a notable decrease in both areas of high $SCL_i^{(l,t)}$ popularity and the extent of moderate and low $SCL_i^{(l,t)}$ areas in surrounding suburbs by 11:00 pm. Additionally, the presence of tweets within scattered hotspots decreases as well.

Next, when taking an example of the 6:00 pm interval as depicted in Figure 5, an intriguing pattern emerges within the Soho district. This area is represented in a grass green hue, indicating a diminished intensity of co-location activity spanning from 2:00 pm to 10:00 pm. Despite Soho's notable popularity, as evidenced by a significant volume of tweets, this apparent "hollow" in the figure shows its low semantic co-location intensity. The immediate assumption of Soho's "unpopularity" is quickly dispelled by the robust social media engagement, prompting a pivot in the hypothesis revealing a phenomenon of diminished "loyalty" among Soho's patrons. The "loyalty" here indirectly indicates whether people prefer to stay in one place (more "loyal") to tweet about related topics or tweet these topics from various locations (less "loyal"). This suggests that the individuals within Soho during the specified time may exhibit a propensity for higher mobility, possibly engaging in multiple locales concurrently within the hour, thus diluting the co-location intensity.

In contrast, Heathrow Airport presents a paradigm of "loyalty," with a sustained pattern of co-location that begins to taper off from 10:00 pm to 4:00 am, before witnessing a resurgence in intensity. This fluctuation not only mirrors the operational peak times of the airport but also lends itself to dual interpretations: first, the functional clarity of the airport environment fosters a uniformity in tweet categorization; second, the behavior of individuals frequenting Heathrow suggests a limited radius of movement, consistent with the typical behaviors of airport patrons who are likely confined to the airport premises for the duration of their stay.

While the analysis of other areas within London might offer additional perspectives, the distinct behaviors observed in Soho and Heathrow provide a foundational basis for further exploration. These insights underscore the need for a more granular understanding of urban mobility and social interaction patterns, warranting deeper investigation in the subsequent phases of this research. Readers who are interested in further exploration can find all figures depicting semantic co-location intensity by hour in Appendix A (Figures A1–A4).

The analysis of semantic co-location intensity across London reveals complex spatiotemporal dynamics in urban social behavior. Our findings demonstrate a distinct evolution from a centric to polycentric urban structure throughout the day, with early morning activity concentrated in nightlife areas like Soho, followed by a more homogeneous distribution during commuting hours. The global urban center experiences a notable regeneration of activity post-morning and evening peaks, maintaining central dominance while spawning numerous dispersed clusters extending from the city center to the suburbs. This pattern is particularly evident during the evening commuting peak at 6:00 pm, marking significant outward movement across the city. These insights, uniquely

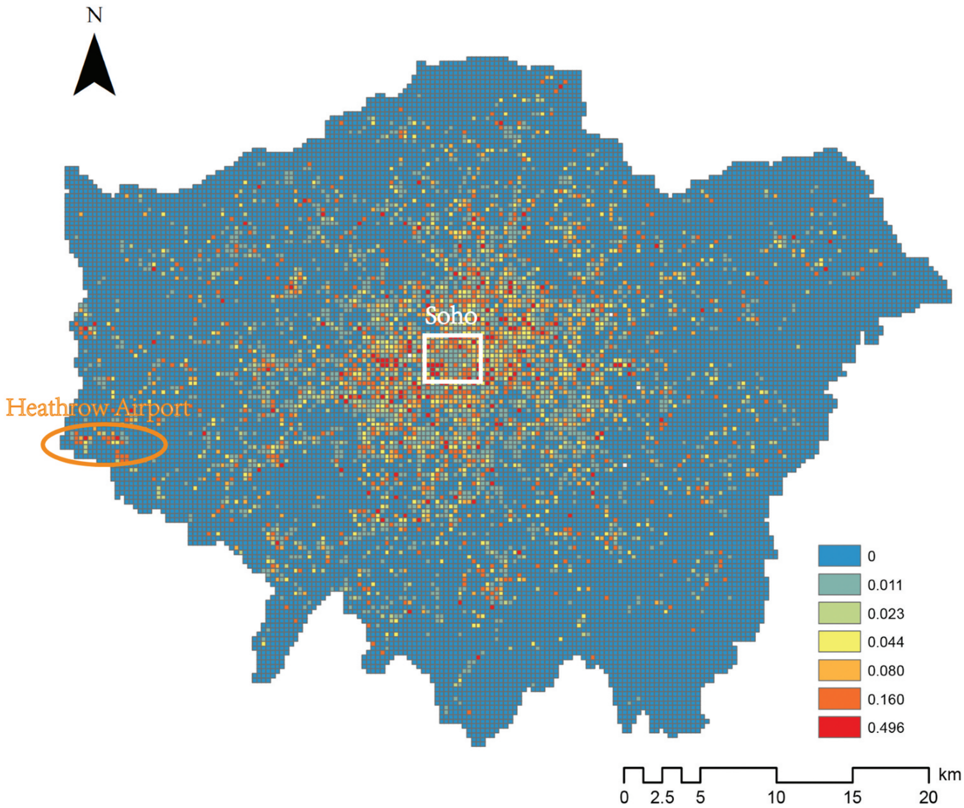

Figure 5. Semantic co-location intensity at 6:00 pm.

captured through semantic co-location analysis of social media content, go beyond traditional spatial metrics to reveal how different urban spaces function not just as physical locations, but as contexts for specific types of social interaction and behavior. The temporal signatures of semantic co-location can further enrich our understanding of urban dynamics by revealing not just when people gather, but how the nature and cohesion of their social interactions vary across different urban contexts throughout the day.

4.2. Comparison between SCL intensity and POI distribution

While the temporal analysis of semantic co-location intensity reveals distinct patterns of urban activity, these patterns cannot be fully understood without considering the underlying urban infrastructure that shapes them. To deepen our understanding of these spatial dynamics, we now turn to examine how the distribution of POIs, physical anchors of urban activity, relates to the observed patterns of semantic co-location. This comparison allows us to explore whether areas of high social media engagement align with concentrations of urban amenities, potentially revealing mismatches between physical infrastructure and actual patterns of social interaction. In this study, a heatmap is created to visualize the distribution of POI density across the GLA, as shown in Figure 6(b). The map clearly demonstrated a significant concentration of POIs within the central area, particularly retail establishments in Soho and Chinatown. Additionally, transportation-related POIs are widely distributed across the GLA, with Inner London containing a high proportion.

When comparing this map to the semantic co-location intensity at 6:00 pm (Figure 6(a)), an interesting pattern emerges. Areas like Soho and Liverpool Street, identified as “hollows” with lower semantic co-location intensity, actually show a high density of POIs. This suggests that the “hollows” may reflect the temporary presence of individuals in these areas. They tend to spread their tweets over a wider area within a given timeframe, likely due to the diverse functions offered by the many POIs. Additionally, the detection of adjacent cold spots intimates that active Twitter users might also exhibit propensities to engage in proximities to these “hollow” regions. Given the pragmatic constraints on travel distances within an hour, particularly during peak traffic periods, it stands to reason that the majority of movement is likely confined to these “hollow” vicinities. This suggests that Twitter users present in these locales at 6:00 pm exhibit elevated mobility. Besides Soho and Liverpool Street, many

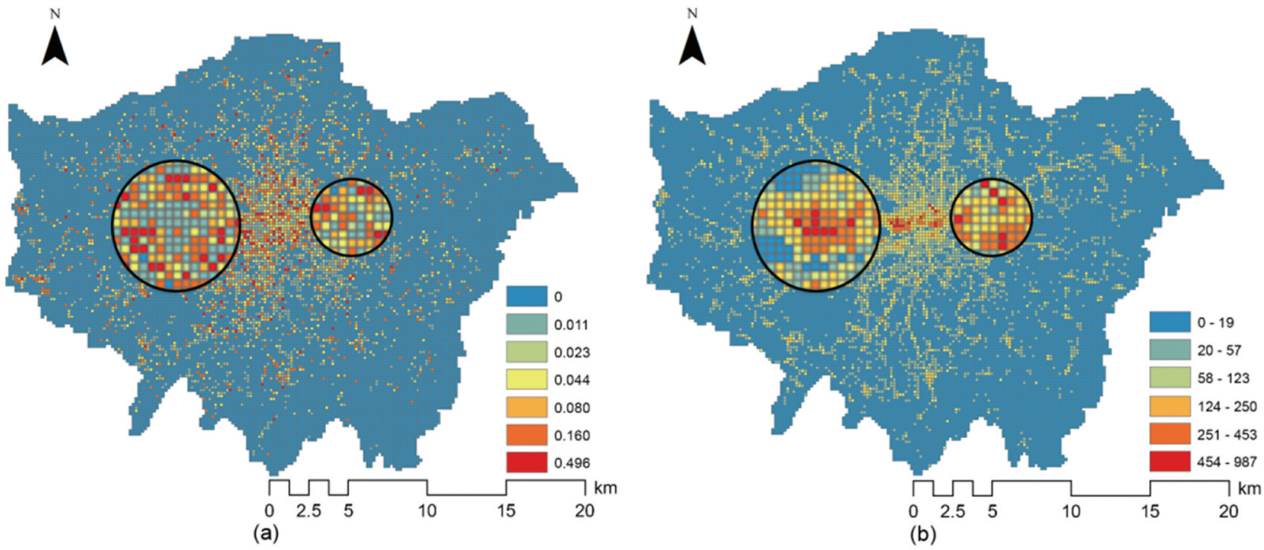

Figure 6. Comparison of semantic co-location intensity at 6:00 pm (a) and distributed density of POIs (b).

other popular areas also exhibit similar characteristics, that is, places with many POIs have low $\mathrm{SCL}_i^{(l,t)}$ value. Many suburban areas with fewer POIs have high $\mathrm{SCL}_i^{(l,t)}$ value, which also confirms the previous conclusion.

To further investigate the correlation between SCL_i and POI data, a choropleth map is created to visualize the spatial relationship between the density of POIs and semantic co-location intensity across London, as shown in Figure 7. Each grid cell is assigned to one of five categories: high POI density with low semantic co-location intensity (orange), low POI density with high semantic co-location intensity (blue), high values in both measures (red), low values in both measures (light green), and else values (light yellow). For balancing the visualization and data sparsity, the normalized thresholds for high and low values are set to above 80% and below 30%.

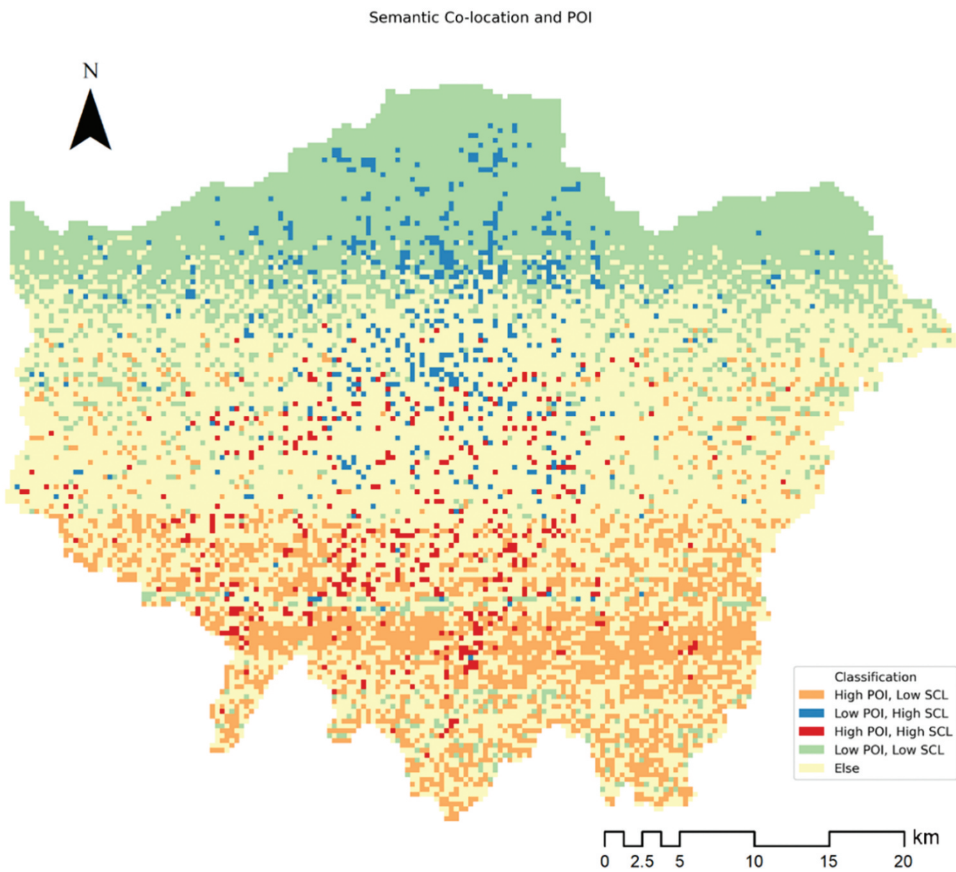

Figure 7. The choropleth map for SCL and POIs.

The spatial pattern aligns closely with London's known urban structure but also with some surprising results because of the small study unit. The red clusters (high in both POI and gathering intensity) appear in locations like Soho, Covent Garden, and South Bank, Heathrow Airport-areas famous for entertainment, dining, cultural activities, and traveling that naturally attract large crowds. However, looking at the red clusters in southwest and south London, there are several interesting explanations for these high POI and high gathering intensities in what are traditionally considered suburban areas. These patterns likely correspond to what urban planners call "suburban town centers" or historical village cores that have been absorbed into Greater London (Vaughan et al. 2010). Places like Kingston upon Thames, Richmond, Croydon, and Bromley are originally independent market towns before becoming part of London's metropolitan area. These areas have maintained their historical role as local centers while developing into significant commercial and social hubs, which explains the high POI density and gathering intensity. Taking Croydon as an example – despite being considered suburban, it has evolved into London's second-largest commercial center with significant retail presence (Whitgift Center, Centrale Shopping Center), and transportation connectivity (East and West Croydon stations). Similarly, Kingston upon Thames combines historical significance with modern retail (Bentall Center), university presence (Kingston University), and riverside attractions, creating natural gathering points.

The blue clusters (Low POI, High SCL) in north London also show a distinctive pattern. These blue clusters appear to be dispersed rather than concentrated, forming several small groupings across the northern part of the map. What is particularly interesting is that many of these blue clusters are surrounded by light green areas (low in both measures), suggesting that these are isolated pockets of high gathering activity in otherwise quiet areas. This suggests that these might be local community focal points that have emerged organically rather than being planned commercial or institutional centers. The scattered nature of these blue clusters, combined with their location in predominantly low-density areas, indicates that they might represent community-level gathering spaces that serve local populations rather than being major destination points.

The most striking feature is the substantial concentration of orange cells forming a broad band across the southern portion of the map. This indicates areas with high POI density but surprisingly low semantic co-location intensity. What is particularly interesting about this pattern is: first, it forms a more continuous and denser pattern compared to other color classifications on the map; second, the orange zones appear to be most concentrated at the southern edges of the mapped area, and third, these areas occasionally have red spots (High POI, High SCL) interspersed within them, suggesting some locations within these POI-dense areas do manage to attract gatherings. The pattern suggests a mismatch between the presence of amenities and actual gathering behavior. Some possible explanations could be that these areas might have numerous POIs that serve functional rather than social purposes – places people visit briefly but do not linger to gather

The resulting classification helps identify areas where the presence of POIs either aligns with or diverges from semantic co-location patterns, offering insights into the relationship between physical urban infrastructure and social-semantic dynamics. In addition, the high density of POIs might be spread out rather than clustered in ways that would encourage social gathering. Last but not least, there might be some hidden aspects not captured in the map – these areas might have high POI density, but the types of POIs might not encourage sustained gathering behavior.

The choropleth map reveals fascinating insights about the relationship between physical urban amenities (POIs) and human gathering behaviors (semantic co-location) across London's geography. The comparison between semantic co-location patterns and POI distribution reveals insights that challenge traditional urban planning assumptions. Our analysis shows that high POI density does not necessarily correlate with sustained social engagement, as evidenced by the "orange zones" in southern London where abundant urban amenities coexist with low semantic co-location intensity. The model's ability to identify these mismatches between physical amenities and social behavior patterns, along with the detection of successful suburban town centers like Croydon and Kingston upon Thames, suggests that effective urban spaces depend not merely on the quantity of amenities but on how these spaces resonate with communities. Our approach provides urban planners with a more nuanced understanding of how urban spaces function as social environments, informing strategies that better align physical infrastructure with actual patterns of social behavior. It shows that successful urban spaces are not just about the quantity of amenities, but rather how these spaces resonate with and are adopted by communities for social interaction.

5. Discussion

The analysis of semantic co-location patterns across London, from their temporal evolution to their relationship with physical urban infrastructure, has revealed intricate layers of urban social dynamics. By examining both the spatiotemporal rhythms of social media engagement and their alignment with POI distribution, our semantic co-location model has uncovered patterns ranging from the "hollow" phenomenon in high-activity areas to the emergence of suburban social hubs, demonstrating the complex relationship between urban form and social function. These findings, derived from the semantic analysis of social media content rather than mere activity patterns, open up broader considerations about the role of such analytical approaches in understanding and shaping urban spaces.

5.1. Potential impact

The potential impact of this study is multifaceted. First, it demonstrates the utility of social media data in elucidating urban spatial structures and human interactions, thereby enriching the discourse on human mobility. Second, by leveraging topic modeling, this research unveils the underlying urban structure and associated social phenomena, extending beyond the individual social media user as the sole unit of analysis. Inspired by the significant social phenomenon of co-location, this study posits that the content of discussions can serve as a lens through which to analyze urban spatiotemporal structures. It posits that spatial–functional interactions across areas drive the variations in spatiotemporal dynamics and topic discussions across different locales. Third, the identified spatial patterns offer insights into the specific locales people discuss, highlighting the inherent data bias toward social media users. Nonetheless, the spatial attributes of locations influence where users go and what they discuss. Despite such biases, at an aggregated level, valuable spatiotemporal structures can be elucidated, informing the spatial planning of urban resources. This includes the strategic development of economic clusters within cities and the incorporation of community heterogeneity into urban analysis, showcasing the utility of co-location analysis for comprehensive urban planning.

5.2. Limitations and future studies

Despite the contributions of this study, several limitations warrant further investigation. The utilization of the LDA model in this research represents a pivotal approach to categorizing Twitter data into distinct topics. However, one of the primary concerns revolves around the inherent limitations of LDA in processing short texts, such as tweets, which are characteristic of the Twitter platform. Earlier studies focus primarily on using and exploiting external knowledge to improve the performance of topic inference for short texts. Phan, Nguyen, and Horiguchi (2008) present a general framework for building short-text classifiers; their model aids in finding most of the hidden topics from Wikipedia and Medline data. Jin et al. (2011) provide a similar approach by clustering short text messages via transfer learning from auxiliary long texts. To enhance modeling of short texts, Yan et al. (2013) propose a novel semantic model known as the Biterm Topic Model (BTM). The primary distinction between BTM and traditional document-generation-based topic models lies in BTM's direct modeling of word co-occurrence patterns (i.e. biterms) across the entire corpus. A more recent topic modeling technique BERTopic (Grootendorst 2022) uses Bidirectional Encoder Representations From Transformers (BERT) embeddings and clustering algorithms to create cohesive topics based on semantic similarity in the data (Devlin 2018), and this technology has been applied to the geospatial field (Xu et al. 2025). In parallel with BERTopic's development, Top2Vec is another advanced method that generates topics by embedding both documents and words into a shared vector space, allowing for topics to emerge organically based on semantic similarity (Angelov 2020). There is also a deep learning model specifically for spatial relation extraction in text (Wu et al. 2023), which indicates a potential for future information integration, allowing for more accurate and contextually aware analyses of relationships in various domains.

Despite the rise of advanced context extraction models, LDA remains a valuable tool in the modern NLP toolkit due to its unique blend of interpretability, efficiency, and probabilistic foundations. While newer models may excel in capturing contextual relationships, LDA's ability to generate clear and interpretable topic distributions without the need for substantial computational resources makes it particularly well suited for practical applications. Its statistical framework not only offers theoretical guarantees but also supports principled extensions, making it especially valuable in cases where understanding the underlying topic structure is more important than achieving marginal improvements in performance metrics.

Furthermore, the determination of an optimal number of topics within the LDA framework emerges as a contentious issue. This aspect of the model is highly sensitive to changes in model variables, leading to variability in the optimal topic number across different analytical scenarios. This variability necessitates a degree of subjective judgment on the part of the researcher, who must apply their expertise and experience to identify the most appropriate topic number for their specific study context.

In addition, the demographic profile of Twitter users, who contribute to this dataset, is not fully representative of the broader population. This user base tends to be skewed to younger persons, with a particular affinity for digital and mobile technology. One of the inaugural studies on social media data bias was conducted by Mislove et al. (2011). They discovered an overrepresentation of populous US counties and an underrepresentation in the US Midwest within Twitter data. Furthermore, the unbalanced distribution of minorities highlighted an undersampling issue.

Zickuhr (2013) investigated the demographics of location service users through a smartphone survey, noting that 12% of respondents activated their geolocation services. Hecht and Stephens (2014) explored Twitter biases using US Census data, employing a method to amplify spatial dependencies by minimizing the effective sample size. Their findings indicated that geotagged tweets from urban areas are 5.3 times more prevalent than those from rural areas. Longley, Adnan, and Lansley (2015) examined bias by utilizing forenames and surnames to ascertain the gender, age, and ethnicity of Twitter users within the GLA, comparing their findings to the 2011 UK Census data. Their research revealed an overrepresentation of young male and white British users, alongside an underrepresentation of middle-aged and older females, as well as minority groups like South Asian, West Indian, and Chinese users.

While acknowledging the inherent biases embedded within crowdsourced data, particularly from social media platforms, it is crucial to recognize the transformative potential it offers for researchers across various fields. Therefore, future research must prioritize the development of robust methods to identify, quantify, and address the biases present in crowdsourced social media data. With the easier accessibility of high-quality data and developing techniques, this will ensure that the insights derived from such data accurately reflect the complexities of urban environments and the diverse lived experiences of people.

Finally, this study, based on data from 2015 to 2017, may raise concerns about its timeliness given the rapid changes in urban dynamics and social behaviors. However, the primary strength of this research lies in the straightforward and accessible nature of its approach, which does not require complex methodologies. Once recent data becomes available, the approach can be readily applied, making it both a practical tool and a valuable asset for future studies. While the reliance on older data presents a limitation, it simultaneously offers opportunities for further refinement and broader applicability, particularly with the integration of more current data.

Similarly, the methodologies devised in this study are not limited to the context of London or simply to Twitter-type social media. Their universal design enables application across varied urban landscapes and media assuming the availability of comprehensive and dependable data sets. While this study analyzes phenomena in a single city, drawing broader conclusions requires a more comprehensive approach. Gathering stronger evidence by applying this framework and advanced techniques to social media and other data in other cities, making this work a valuable reference. While traditional theories explore the nature and impacts of these phenomena, they have not investigated, measured, or quantified such social behaviors using new data from social media and new methods from data science and statistics (Che et al. 2025; Li et al. 2025). In fact, in recent years, over the past few years, a growing body of literature has concentrated on semantic information, spanning from vehicle trajectory analysis (Zhang et al. 2025) to building extraction based on deep learning techniques (Wang et al. 2025). Such versatility enables comparative studies of urban environments, facilitating the examination of co-location trends in different cities to discern distinct or common urban attributes. Future research will also explore the regularity and variability of urban activity patterns using other social phenomena like co-location.

6. Conclusions

In conclusion, this research has advanced urban spatial analysis through its innovative examination of semantic co-location patterns in Greater London using geotagged tweets. The study's primary contributions lie in two methodological innovations: (1) the integration of coherence with spatial autocorrelation for optimal topic model selection, and (2) the development of semantic co-location as a novel metric for understanding urban spatial structure. Through the segmentation of analysis into distinct phases – translating tweets to topics, topics to functions, functions to social phenomena, and social phenomena to structural insights – a granular understanding was achieved of urban dynamics at various levels of aggregation.

Our findings demonstrate that urban social patterns are more nuanced than simple activity concentrations would suggest. For example, the identification of "hollow" phenomena in high-activity areas like Soho contrasted with sustained "loyalty" patterns in functional spaces like Heathrow Airport, reveals how semantic co-location analysis can uncover distinct patterns of urban usage. Furthermore, the comparison with POI distribution has challenged conventional assumptions about the relationship between physical infrastructure and social dynamics, particularly in identifying mismatches between amenity density and actual social engagement patterns. Due to length constraints, many interesting patterns revealed by our semantic co-location analysis await future exploration.

The insights derived from this research offer significant implications for urban planning. By identifying areas of high thematic and spatial co-location, planners can better allocate resources to meet localized demands. For instance, areas with strong co-location patterns of residential and commercial functions can inform the design of mixed-use neighborhoods, fostering walkability and reducing reliance on transportation networks. Additionally, the co-location methodology provides a scalable framework that can adapt to new data sources, such as sensor networks, other social media platforms, or mobility data, ensuring its relevance in the context of evolving urban technologies and data modalities. This approach equips urban planners with innovative tools to design cities that are more responsive, sustainable, and inclusive. With the rapid advancement of artificial intelligence, particularly large language models like GPT and their enhanced capabilities in semantic analysis, there is substantial potential to further refine and expand this research (Liang et al. 2025). These AI tools could enhance topic modeling accuracy, enable more nuanced semantic similarity measurements, and potentially reveal even more subtle patterns in urban social dynamics. The integration of such advanced language models with spatial analysis opens up exciting new possibilities for understanding the complexity of urban spaces through the lens of social media content.

Author contributions

CRediT: **Shi Zeng:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing; **Xiaoliang Meng:** Funding acquisition, Resources, Supervision, Validation, Writing – review & editing; **Chen Zhong:** Conceptualization, Data curation, Formal analysis, Methodology, Resources; **Yao Shen:** Conceptualization, Data curation, Funding acquisition, Methodology, Resources; **Yichun Xie:** Supervision, Validation, Visualization, Writing – review & editing; **Michael Batty:** Conceptualization, Supervision, Validation, Writing – review & editing.

Disclosure statement

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

Funding

This research was funded by National Natural Science Foundation of China [Grant number 41971352], National Natural Science Foundation of China [Grant number 52278074], and the Shanghai Rising-Star Program [Grant number 22QB1404800].

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Notes on contributors

Shi Zeng received his PhD from Centre for Advanced Spatial Analysis, University College London. His research interests are spatial analysis, social and spatial connections, urban studies, and spatial temporal intelligence.

Xiaoliang Meng received a PhD degree from Wuhan University in 2009. He was a Visiting Scholar and a Post-Doctoral Scientist in the USA for three years and participated in the NASA ICCaRS Project. He is currently a Distinguished Professor with Hubei Luojia Laboratory, Wuhan University. His main research interest is intelligent geospatial sensing. Dr. Meng has received the "Best Young Authors Award" from the International Society for Photogrammetry and Remote Sensing (ISPRS). He is currently the Deputy Director of the Education and Science Popularization Committee, China Association of Geographic Information Industry.

Chen Zhong is currently an Associate Professor in Urban Analytics at Centre for Advanced Spatial Analysis, University College London. She received her PhD from ETH Zurich. Her research interests include location data mining, machine learning, spatial network analysis, and the use of such techniques for city planning.

Yao Shen received his PhD from University College London. He is currently an Associate Professor at the College of Architecture and Urban Planning (CAUP), Tongji University. He is also the Director of Centre for Urban Science and Planning, CAUP, Tongji University, and Honorary Fellow of Centre for Advanced Spatial Analysis, University College London. His research involves the development of urban science and its applications across multiple scales for planning and design, and relevant topics include urban modelling, spatial analysis, urban geometry, spatial econometrics, data visualisation, urban resilience, transport and land use, data augmented design, etc.

Yichun Xie received his PhD from State University of New York at Buffalo. He is currently a professor of geographic information science and environmental geography at the Department of Geography and Geology, Eastern Michigan University. He is also the Founding Director of the Institute of Geospatial Research and Education (IGRE, from 1998), Eastern Michigan University. His research includes spatiotemporal modelling of urban growth, grassland ecosystem, coupled impacts of human dynamics and environmental changes on resource management and ecosystem recovery, and land-use and land-cover changes; and he is a pioneer in the development of urban dynamic evolution theory based on cellular automata, which is widely regarded as a start of the second generation of dynamic modelling of urban complex systems.

Michael Batty received his PhD from the University of Wales. He is Bartlett Professor of Planning at University College London where he is Chair of the Centre for Advanced Spatial Analysis. He is a Fellow of the British Academy (FBA), the Academy of Social Sciences (FAcSS), the Royal Society (FRS), and Foreign Academician of Chinese Academy of Sciences. His research work involves the development of computer models of cities and regions, and he has published many books and articles in this area. His book *Cities and Complexity* (MIT Press, Cambridge, MA, 2005) won the Alonso Prize of the Regional Science Association in 2010. His most recent books are *The New Science of Cities* (MIT Press, Cambridge, MA, 2013) and the edited volumes *Virtual Geographic Environments* (ESRI Press, Redlands, CA, 2011) and *Agent Based Models of Geographical Systems* (Springer, Berlin, 2012). He is editor of the journal Environment and Planning B: Planning and Design.

ORCID

Shi Zeng http://orcid.org/0000-0002-7801-2517

Xiaoliang Meng http://orcid.org/0000-0002-3271-9314
Chen Zhong http://orcid.org/0000-0003-3582-1266
Yao Shen http://orcid.org/0000-0002-7261-4234
Yichun Xie http://orcid.org/0000-0002-2045-6406
Michael Batty http://orcid.org/0000-0002-9931-1305

References

Angelov, D. 2020. "Top2vec: Distributed Representations of Topics." *arXiv Preprint arXiv: 2008.09470*. https://doi.org/10.48550/arXiv.2008.09470.

Bahrehdar, A. R., and R. S. Purves. 2018. "Description and Characterization of Place Properties Using Topic Modeling on Georeferenced Tags." *Geo-Spatial Information Science* 21 (3): 173–184. https://doi.org/10.1080/10095020.2018.1493238.

Batty, M. 2008. "The Size, Scale, and Shape of Cities." *Science* 319 (5864): 769–771. https://doi.org/10.1126/science.1151419.

Batty, M. 2018. "Digital Twins." *Environment and Planning B: Urban Analytics and City Science* 45 (5): 817–820. https://doi.org/10.1177/2399808318796416.

Batty, M. 2024. "Digital Twins in City Planning." *Nature Computational Science* 4 (3): 192–199. https://doi.org/10.1038/s43588-024-00606-7.

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

Cai, J., and M. P. Kwan. 2022. "Discovering Co-Location Patterns in Multivariate Spatial Flow Data." *International Journal of Geographical Information Science* 36 (4): 720–748. https://doi.org/10.1080/13658816.2021.1980217.

Carmody, D., M. Mazzarello, P. Santi, T. Harris, S. Lehmann, T. Abbiasov, R. Dunbar, and C. Ratti. 2022. "The Effect of Co-Location on Human Communication Networks." *Nature Computational Science* 2 (8): 494–503. https://doi.org/10.1038/s43588-022-00296-z.

Che, M., C. Wang, Y. Nian, and P. Rao. 2025. "Using Social Media for Environmental Insight: A Multi-Model Deep Learning Framework Approach." *Geo-Spatial Information Science*: 1–16. https://doi.org/10.1080/10095020.2025.2541877.

Devlin, J. 2018. "Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *arXiv Preprint arXiv: 1810.04805*. https://doi.org/10.48550/arXiv.1810.04805.

Dong, G., R. Li, H. Wu, H. Wei, H. Zhang, V. Tao, and Q. Liu. 2024. "Learning the Spatial Co-Occurrence for Browsing Interests Extraction of Domain Users on Public Map Service Platforms." *Geo-Spatial Information Science* 27 (2): 455–474. https://doi.org/10.1080/10095020.2022.2140078.

Fang, A., C. Macdonald, I. Ounis, and P. Habel. 2016. "Examining the Coherence of the Top Ranked Tweet Topics." *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Italy: Presented at the Pisa. https://doi.org/10.1145/2911451.2914731.

Giddens, A. 1984. *The Constitution of Society: Outline of the Theory of Structuration*. Oakland, California: University of California Press.

Gong, J., S. Li, X. Ye, Q. Peng, and S. Kudva. 2021. "Modelling Impacts of High-Speed Rail on Urban Interaction with Social Media in China's Mainland." *Geo-Spatial Information Science* 24 (4): 638–653. https://doi.org/10.1080/10095020.2021.1972771.

Grootendorst, M. 2022. "BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure." *arXiv: 2203.05794. Preprint, arXiv*. https://doi.org/10.48550/arXiv.2203.05794.

Hecht, B., and M. Stephens. 2014. "A Tale of Cities: Urban Biases in Volunteered Geographic Information." *Proceedings of the International AAAI Conference on Web and Social Media* 8 (1): 197–205. https://doi.org/10.1609/icwsm.v8i1.14554.

Hu, Y. 2018. "Geospatial Semantics." *Comprehensive Geographic Information Systems* 1:80–4. https://doi.org/10.1016/B978-0-12-409548-9.09597-X.

Huang, Y., and P. Zhang. 2006. "On the Relationships Between Clustering and Spatial Co-Location Pattern Mining." *200618th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06)*, 513–522. https://doi.org/10.1109/ICTAI.2006.91.

Iyer, S., B. Karrer, D. T. Citron, F. Kooti, P. Maas, Z. Wang, E. Giraudy, A. Medhat, P. A. Dow, and A. Pompe. 2023. "Large-Scale Measurement of Aggregate Human Colocation Patterns for Epidemiological Modeling." *Epidemics* 42:100663. https://doi.org/10.1016/j.epidem.2022.100663.

Jin, O., N. N. Liu, K. Zhao, Y. Yu, and Q. Yang. 2011. "Transferring Topical Knowledge from Auxiliary Long Texts for Short Text Clustering." *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, 775–784. Presented at the Glasgow, Scotland, UK.

Jing, F., Z. Li, S. Qiao, J. Zhang, B. Olatosi, and X. Li. 2023. "Using Geospatial Social Media Data for Infectious Disease Studies: A Systematic Review." *International Journal of Digital Earth* 16 (1): 130–157. https://doi.org/10.1080/17538947.2022.2161652.

Kaffash, C. N., A. Sadeghi-Niaraki, S.-M. Choi, and T. Abuhmed. 2023. "An Approach for Measuring Spatial Similarity Among COVID-19 Epicenters." *Geo-Spatial Information Science* 26 (3): 496–513. https://doi.org/10.1080/10095020.2022.2088303.

Keddy, P. 2007. *Plants and Vegetation: Origins, Processes, Consequences*. Cambridge: Cambridge University Press.

Lansley, G., and P. A. Longley. 2016. "The Geography of Twitter Topics in London." *Computers, Environment and Urban Systems* 58:85–96. https://doi.org/10.1016/j.compenvurbsys.2016.04.002.

Li, B., D. Hutchinson, S. Lim, and C. R. MacIntyre. 2025. "An Attention-Based Hybrid Model for Spatial and Temporal Sentiment Analysis of COVID-19 Related Tweets in the Contiguous United States." *Geo-Spatial Information Science* 28 (4): 1846–1865. https://doi.org/10.1080/10095020.2024.2408343.

Li, F., Z. Feng, P. Li, and Z. You. 2017. "Measuring Directional Urban Spatial Interaction in China: A Migration Perspective." *PLOS ONE* 12 (1): e0171107. https://doi.org/10.1371/journal.pone.0171107.

Li, L., J. Cheng, J. Bannister, and X. Mai. 2022. "Geographically and Temporally Weighted Co-Location Quotient: An Analysis of Spatiotemporal Crime Patterns in Greater Manchester." *International Journal of Geographical Information Science* 36 (5): 918–942. https://doi.org/10.1080/13658816.2022.2029454.

Liang, J., S. Hou, A. Zhao, Q. Xu, L. Xiang, R. Li, and H. Wu. 2025. "Design and Application of a Semantic-Driven Geospatial Modeling Knowledge Graph Based on Large Language Models." *Geo-Spatial Information Science* 1–20. https://doi.org/10.1080/10095020.2025.2483884.

Liu, W., Q. Liu, M. Deng, J. Cai, and J. Yang. 2022. "Discovery of Statistically Significant Regional Co-Location Patterns on Urban Road Networks."

International Journal of Geographical Information Science 36 (4): 749–772. https://doi.org/10.1080/13658816.2021.1981335.

Longley, P. A., M. Adnan, and G. Lansley. 2015. "The Geotemporal Demographics of Twitter Usage." *Environment and Planning A* 47 (2): 465–484. https://doi.org/10.1068/a130122p.

Mislove, A., S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, and J. Rosenquist. 2011. "Understanding the Demographics of Twitter Users." *Proceedings of the International AAAI Conference on Web and Social Media* 5 (1): 554–557. https://doi.org/10.1609/icwsm.v5i1.14168 .

Moran, P. A. 1950. "Notes on Continuous Stochastic Phenomena." *Biometrika* 37 (1–2): 17–23. https://doi.org/10.1093/biomet/37.1-2.17.

Morioka, W., A. Okabe, M.-P. Kwan, and S. L. McLafferty. 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. https://doi.org/10.1080/13658816.2021.1976409.

O'Callaghan, D., D. Greene, M. Conway, J. Carthy, and P. Cunningham. 2015. "Down the (white) Rabbit Hole: The Extreme Right and Online Recommender Systems." *Social Science Computer Review* 33 (4): 459–478. https://doi.org/10.1177/0894439314555329.

Phan, X.-H., L.-M. Nguyen, and S. Horiguchi. 2008. "Learning to Classify Short and Sparse Text & Web with Hidden Topics from Large-Scale Data Collections." In *17th International Conference on World Wide Web*, 91–100. Presented at the Beijing, China. https://doi.org/10.1145/1367497.1367510.

Shekhar, S., and Y. Huang. 2001. "Discovering Spatial Co-Location Patterns: A Summary of Results." In *Advances in Spatial and Temporal Databases. SSTD 2001. Lecture Notes in Computer Science*, edited by C.S. Jensen, M. Schneider, B. Seeger, and V.J. Tsotras, Vol. 2121. Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-47724-1_13.

Shen, Y., K. Karimi, S. Law, and C. Zhong. 2019. "Physical Co-Presence Intensity: Measuring Dynamic Face-To-Face Interaction Potential in Public Space Using Social Media Check-In Records." *PLOS ONE* 14 (2): e0212004. https://doi.org/10.1371/journal.pone.0212004.

Shiode, N., S. Shiode, and R. Inoue. 2022. "Measuring the Colocation of Crime Hotspots." *GeoJournal* 88 (3): 3307–3322. https://doi.org/10.1007/s10708-022-10806-1.

Shwartz-Asher, D., S. Chun, N. R. Adam, and K. L. Snider. 2020. "Knowledge Sharing Behaviors in Social Media." *Technology in Society* 63:101426. https://doi.org/10.1016/j.techsoc.2020.101426.

Silver, D. A., and T. N. Clark. 2016. *Scenescapes: How Qualities of Place Shape Social Life*. Chicago: University of Chicago Press.

Steiger, E., B. Resch, and A. Zipf. 2016. "Exploration of Spatiotemporal and Semantic Clusters of Twitter Data Using Unsupervised Neural Networks." *International Journal of Geographical Information Science* 30 (9): 1694–1716. https://doi.org/10.1080/13658816.2015.1099658.

Vaughan, L., C. E. Jones, S. Griffiths, and M. Haklay. 2010. "The Spatial Signature of Suburban Town Centres." *The Journal of Space Syntax* 1 (1): 77–91.

Wang, W., P. Yu, M. Li, X. Zhong, Y. He, H. Su, and Y. Zhou. 2025. "TDFNet: Twice Decoding V-Mamba-CNN Fusion Features for Building Extraction." *Geo-Spatial Information Science* 1–20. https://doi.org/10.1080/10095020.2025.2514812.

Wu, K., X. Zhang, Y. Dang, and P. Ye. 2023. "Deep Learning Models for Spatial Relation Extraction in Text." *Geo-Spatial Information Science* 26 (1): 58–70. https://doi.org/10.1080/10095020.2022.2076619.

Xu, L., J. Zhang, C. Zhang, X. Zheng, Z. Du, and X. Xue. 2025. "Beyond Extraction Accuracy: Addressing the Quality of Geographical Named Entity Through Advanced Recognition and Correction Models Using a Modified BERT Framework." *Geo-Spatial Information Science* 28 (3): 1195–1213. https://doi.org/10.1080/10095020.2024.2354229.

Xu, Y., A. Belyi, I. Bojic, and C. Ratti. 2017. "How Friends Share Urban Space: An Exploratory Spatiotemporal Analysis Using Mobile Phone Data." *Transactions in GIS* 21 (3): 468–487. https://doi.org/10.1111/tgis.12285.

Yan, X., J. Guo, Y. Lan, and X. Cheng. 2013. "A Biterm Topic Model for Short Texts." In *Proceedings of the 22nd International Conference on World Wide Web*, 1445–1456. Brazil: Presented at the Rio de Janeiro. https://doi.org/10.1145/2488388.2488514.

Yan, Y., C.-C. Feng, W. Huang, H. Fan, Y. Wang, and A. Zipf. 2020. "Volunteered Geographic Information Research in the First Decade: A Narrative Review of Selected Journal Articles in GIScience." *International Journal of Geographical Information Science* 34 (9): 1765–1791. https://doi.org/10.1080/13658816.2020.1730848.

Yan, Y., M. Schultz, and A. Zipf. 2019. "An Exploratory Analysis of Usability of Flickr Tags for Land use/Land Cover Attribution." *Geo-Spatial Information Science* 22 (1): 12–22. https://doi.org/10.1080/10095020.2018.1560044.

Yu, W. 2016. "Spatial Co-Location Pattern Mining for Location-Based Services in Road Networks." *Expert Systems with Applications* 46 (March): 324–335. https://doi.org/10.1016/j.eswa.2015.10.010.

Zhang, Y., Y. Xie, C. Shi, Q. Li, B. Yang, and W. Hao. 2025. "Identifying Vehicle Types from Trajectory Data Based on Spatial-Semantic Information." *Geo-Spatial Information Science* 28 (4): 1757–1773. https://doi.org/10.1080/10095020.2024.2430294.

Zhao, F., A.-X. Zhu, and C. Qin. 2025. "Spatial Distribution Pattern Analysis Using Variograms Over Geographic and Feature Space." *Geo-Spatial Information Science* 28 (4): 1774–1788. https://doi.org/10.1080/10095020.2024.2439371.

Zheng, X., J. Han, and A. Sun. 2018. "A Survey of Location Prediction on Twitter." *IEEE Transactions on Knowledge and Data Engineering* 30 (9): 1652–1671. https://doi.org/10.1109/TKDE.2018.2807840.

Zhong, C., S. Zeng, W. Tu, and M. Yoshida. 2018. "Profiling the Spatial Structure of London: From Individual Tweets to Aggregated Functional Zones." *ISPRS International Journal of Geo-Information* 7 (10): 386. https://doi.org/10.3390/ijgi7100386.

Zhu, X., T. Hu, X. Ye, W. Guo, L. Huang, H. Xiong, H. Ai, B. She, Q. Xiong, and L. Duan. 2019. "Development and Implementation of a Dynamic and 4D GIS Based on Semantic Location Model." *Geo-Spatial Information Science* 22 (3): 193–213. https://doi.org/10.1080/10095020.2019.1649192.

Zickuhr, K. 2013. "Location-Based Services." *Pew Research* 679:695. https://www.pewresearch.org/internet/2013/09/12/location-based-services/.

Appendix

Appendix A

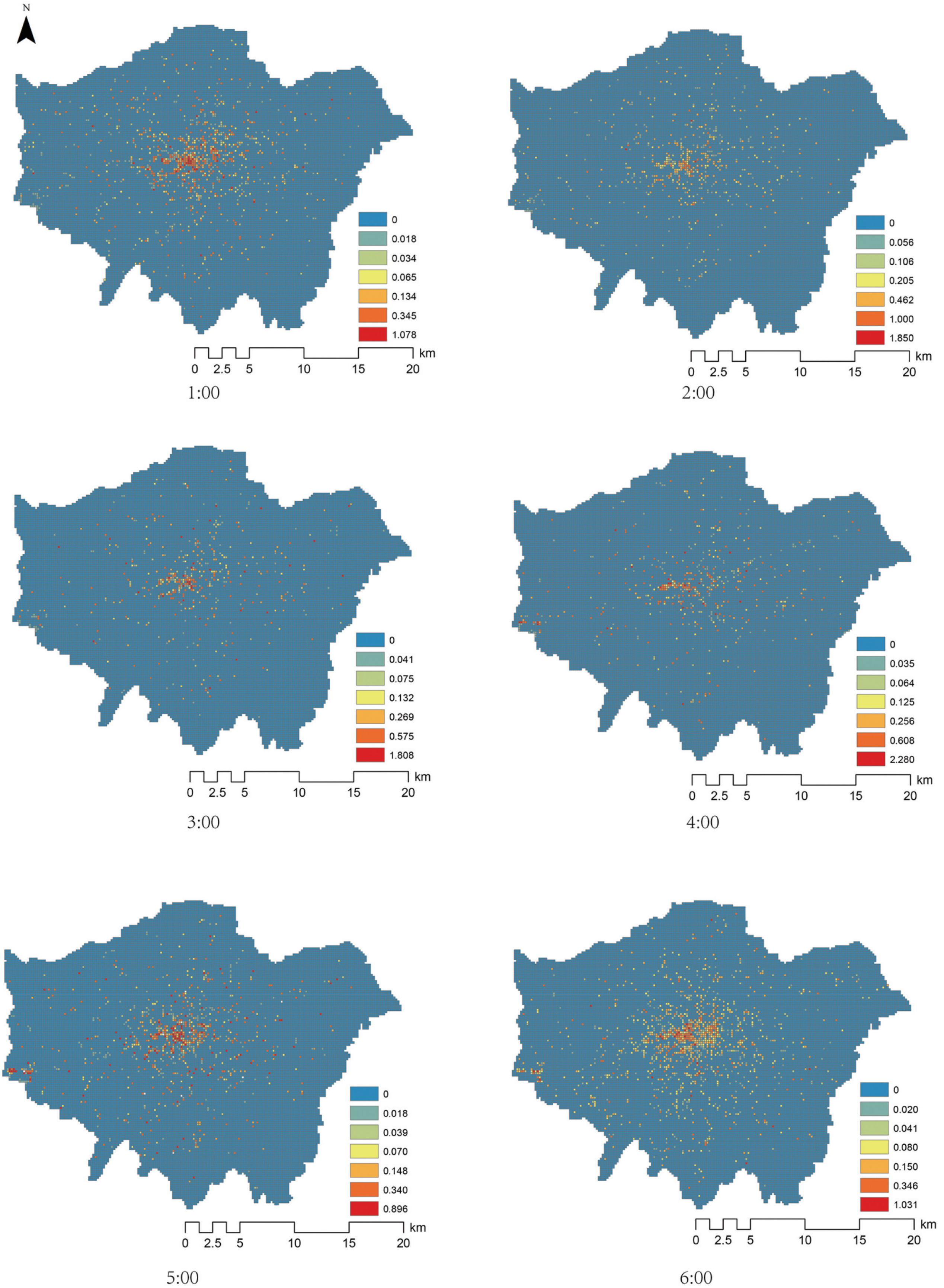

Figure A1. Visualizations of semantic co-location intensity from 1 am to 6 am. (Each hour indicates the time period of 1 h. For example, 1:00 means 1:00 to 1:59. The same applies to the following figures).

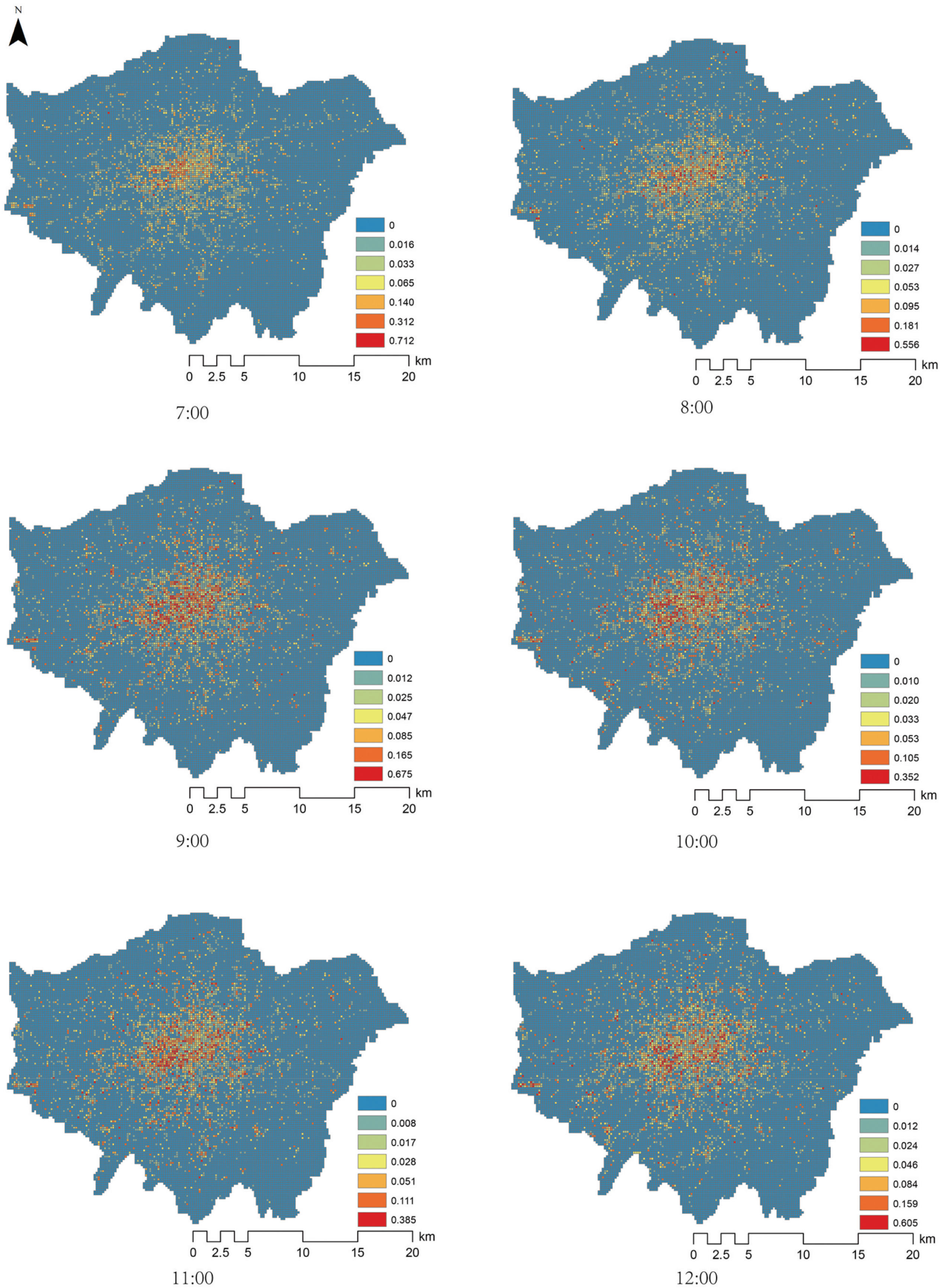

Figure A2. Visualizations of semantic co-location intensity from 7 am to 12 am.

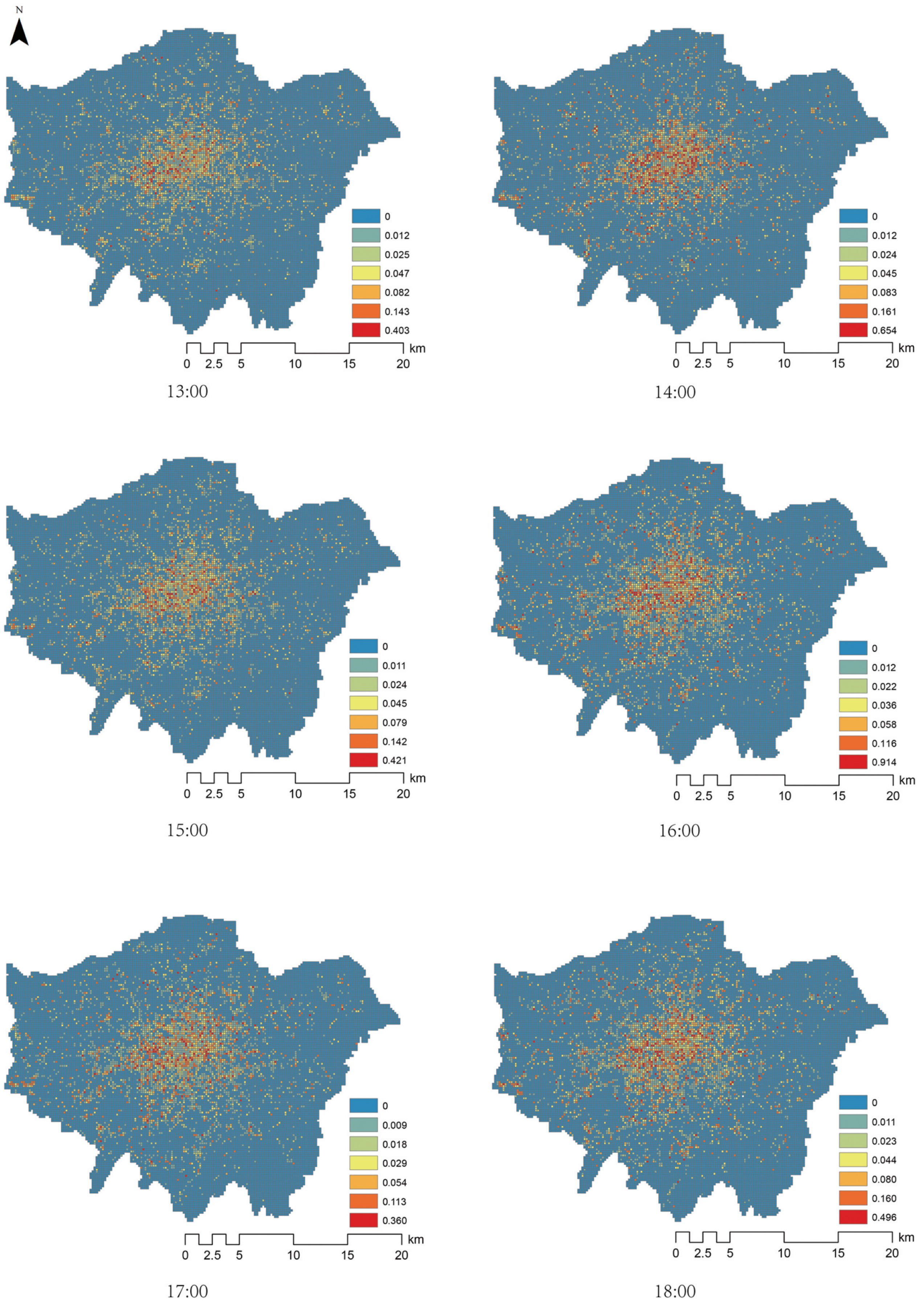

Figure A3. Visualizations of semantic co-location intensity from 1 pm to 6 pm.

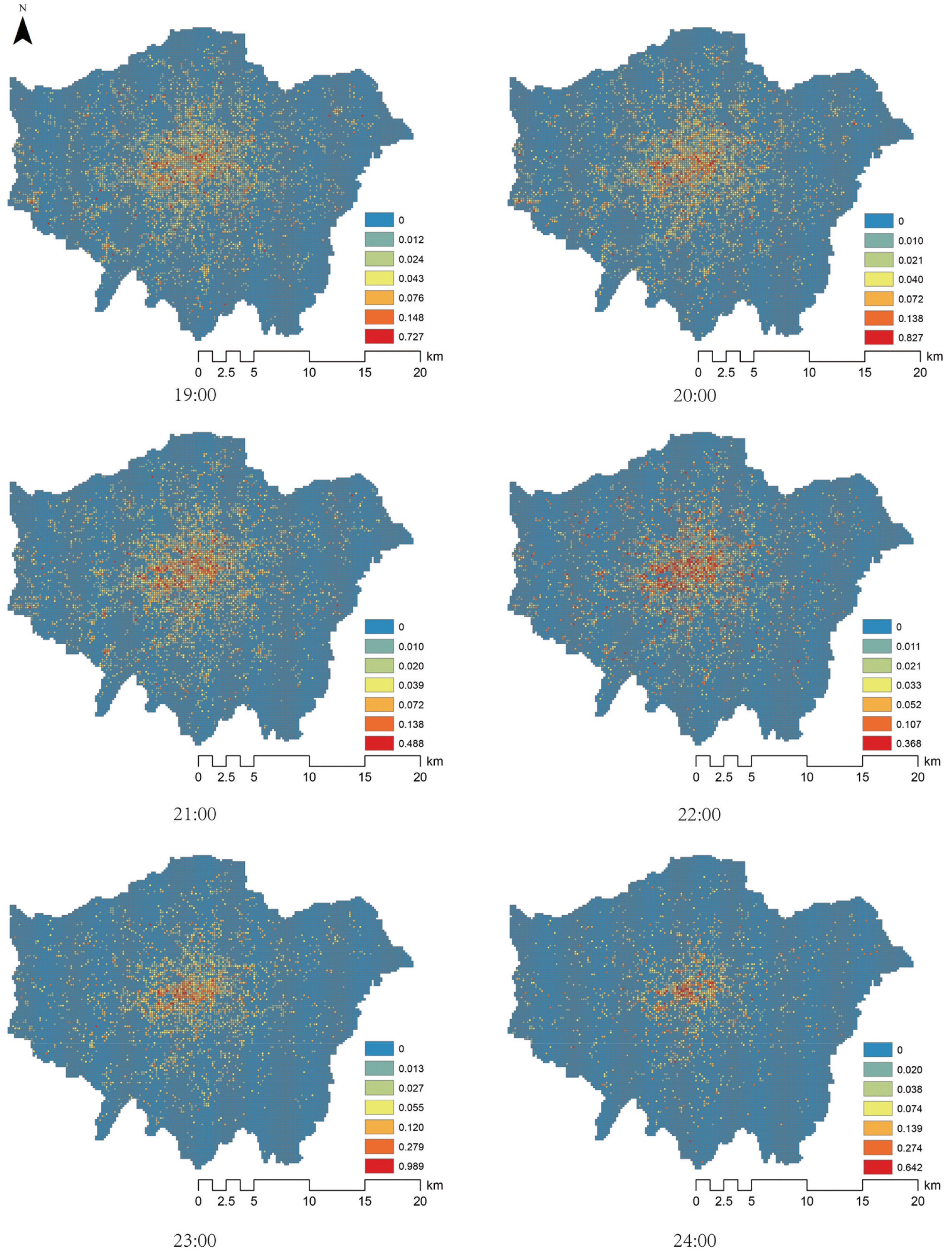

Figure A4. Visualizations of semantic co-location intensity from 7 pm to Midnight.