International Journal of
Geo-Information

Article

Differences in Urban Development in China from the Perspective of Point of Interest Spatial Co-Occurrence Patterns

Guangsheng Dong [1,2,3], Rui Li [1,2,3,*], Fa Li [4], Zhaohui Liu [1,2,3], Huayi Wu [1,2,3], Longgang Xiang [1,2,3], Wensen Yu [5], Jie Jiang [6,7], Hongping Zhang [7,8] and Fangning Li [9]

1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; guangshengdong@whu.edu.cn (G.D.); liuzhaohui@whu.edu.cn (Z.L.); wuhuayi@whu.edu.cn (H.W.); geoxlg@whu.edu.cn (L.X.)
2 Hubei Luojia Laboratory, Wuhan 430079, China
3 Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China
4 Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA; fali2@lbl.gov
5 Fujian Key Laboratory of Big Data Application and Intellectualization for Tea Industry, Wuyi University, Wuyishan 354300, China; npyws@wuyiu.edu.cn
6 Key Laboratory of Urban Spatial Information, Ministry of Natural Resources, Beijing 100044, China; jiangjie@bucea.edu.cn
7 School of Geomatics and Urban Spatial Information, Beijing University of Civil Engineering and Architecture, Beijing 100044, China; zhanghongping@ngcc.cn
8 National Geomatics Center of China, Beijing 100830, China
9 Beijing Urban Construction Exploration & Surveying Design Research Institute Co., Ltd., Beijing 100101, China; hr@cki.com.cn
* Correspondence: ruili@whu.edu.cn; Tel.: +86-136-0716-8286

Abstract: An imbalance in urban development in China has become a contradiction. Points of Interest (POIs) serve as representations of the spatial distribution of urban functions. Analyzing POI spatial co-occurrence patterns can reveal the agglomeration patterns of urban functions across cities at different levels, providing insights into imbalances in urban development. Using POI data from 297 cities in China, the Word2vec model was employed to model the POI spatial co-occurrence patterns, allowing for the quantification of fine-granular urban functionality. Subsequently, the cities were clustered into five tiers representing different levels of development. An urban hierarchical disparity index and graph were introduced to examine variations in urban functions across different tiers. A significant correlation between POI spatial co-occurrence patterns and the GDP of cities at different levels was demonstrated. This study revealed a notable polarization trend characterized by the development of top-tier cities and lagging tail-end cities. Top-tier cities exhibit advantages in terms of their commercial environments, such as international banks, companies, and transportation facilities. Conversely, tail-end cities face deficiencies in urban infrastructure. It is crucial to coordinate resource allocation and establish sustainable development strategies that foster mutual support between the top-tier and tail-end cities.

Keywords: point of interest (POI); POI semantic space; spatial co-occurrence; urban function; urban development

Citation: Dong, G.; Li, R.; Li, F.; Liu, Z.; Wu, H.; Xiang, L.; Yu, W.; Jiang, J.; Zhang, H.; Li, F. Differences in Urban Development in China from the Perspective of Point of Interest Spatial Co-Occurrence Patterns. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 24. https://doi.org/10.3390/ijgi13010024

Academic Editor: Wolfgang Kainz

Received: 23 October 2023
Revised: 22 December 2023
Accepted: 8 January 2024
Published: 10 January 2024

1. Introduction

China has experienced rapid urbanization in the past four decades [1], and the Chinese government has emphasized the contradiction between unbalanced and inadequate development and the ever-growing need for a better life. This contradiction is particularly evident in the unbalanced development of cities at different tiers [2]. Cities are vast and complex systems that serve as a foundation for economic development. However, the unbalanced development of cities has resulted in a range of unsustainable socioeconomic issues, such as industrial deterioration, cultural heritage destruction, widening income disparities, inefficient use of resources, environmental degradation, and heightened social unrest [3,4]. Consequently, a comprehensive understanding of disparities in urban development is crucial for formulating sustainable urban development strategies [5].

The unbalanced development of cities in China is evident in various aspects, such as population migration, talent flow, economic indicators, the environment, and transportation facilities. Xu studied the population migration during a Spring Festival in 293 prefecture-level cities and four municipalities to reveal the unbalanced development of cities [4]. Zhou analyzed the imbalance in educational resources from the perspective of high-level talent flow [6]. Zünd examined disparities in cities based on Gross Domestic Product (GDP) and population [7]. Manoli analyzed the urban heat island effect in different cities [8], whereas Chen analyzed the imbalance in the development of high-grade highways [9]. In recent years, Point of Interest (POI) data have emerged as a valuable resource for inferring urban functions, offering a novel perspective for elucidating the unbalanced development of cities [10].

Urban functions involve the use of urban spaces where human activities occur and represent the fundamental attributes of cities [11–13]. Different sorts of urban functions are spatially intertwined and co-exist with each other, with both common foundational urban functions and distinctive regional variations in urban functions [10]. Developed cities outperform underdeveloped cities in urban functions such as education, healthcare, environment, and transportation [14]. The analysis of urban functions can provide insight into the disparities between cities with different development levels, thus aiding government planning and resource allocation efforts aimed at fostering economic development [15].

POIs serve as a proxy for sensing fine-granular urban functions [16–18]. POIs provide a clear representation of the types of urban activities and the specific locations where these activities occur [10]. They serve as carriers for urban industries and provide functional services [19]. For instance, the educational industry consists of kindergartens, middle schools, high schools, and universities, which collectively support the educational function of a city. POI data in urban areas play a crucial role in representing urban functions. As a result, the utilization of POI data to infer functional types in different regions of a city has become a popular research topic in recent years [20–23].

POI spatial co-occurrence refers to the spatial clustering phenomenon of diverse POI types [24]. This phenomenon is grounded in the First Law of Geography, which posits that "everything is related to everything else, but near things are more related than distant things" [25]. Therefore, the spatial co-occurrence patterns of POIs signify their inherent correlations in proximity. Chen demonstrated that the concept of POI spatial co-occurrence is rooted in location theories that explain how geographic locations are related to different economic activities and processes [10]. POIs associated with different industries tend to aggregate based on similar or complementary functional relationships, which is a common phenomenon between the industries with upstream and downstream industry associations, thereby facilitating industry-specific services, reducing costs, and enhancing operational efficiency [26]. Liu employed spatial co-occurrence patterns of POIs to identify the geographic concentrations of related businesses, known as industrial clusters, in Dongguan, China [19]. Thus, POI spatial co-occurrence patterns serve as representations of the spatial distribution patterns of urban functions. Moreover, compared with previous studies that adopted a coarse granular definition of urban functions, the fine granular perspective offers a more comprehensive and detailed depiction [10].

Due to the unbalanced development of cities, researchers classify them into different levels. For instance, Xu divided Chinese cities into five tiers [4]. Different levels of urban development exhibit significant differences in the functional composition and industrial distribution, which are important manifestations of the unbalanced development of cities [14]. Rodrigo compared European second- and first-tier cities concerning the presence of urban functions and how these are spread over their urban regions, revealing an association between urban development levels and urban functions [27]. However, the current research has focused on coarse granular variations in urban functions in developed cities, overlooking the fine granular differences in urban functions within widely dispersed underdeveloped city clusters. This constitutes a crucial aspect of understanding urban development disparities, which can be addressed by exploring POI spatial co-occurrence patterns [10].

The objective of this study is to show the imbalances in urban development among cities with different tiers from the perspective of POI spatial co-occurrence. We used a dataset of 65.27 million POIs covering 297 prefecture-level cities in China. The Word2vec model was employed to model the POI spatial co-occurrence patterns of each city, representing the fine granular spatial distribution of urban functionality. Based on the POI spatial co-occurrence patterns, the cities were clustered into five tiers, encompassing developed first- and second-tier cities, as well as undeveloped third-, fourth-, and fifth-tier cities. The association between the tiers and GDP was also analyzed. We introduced an urban hierarchical disparity index and graph to evaluate the fine granular functional differences among the tiers and disclose their relationship with urban development.

2. Study Area and Data Source

This study focused on an extensive area comprising 297 prefecture-level cities in mainland China. The dataset for each city comprised two primary components: POI and GDP data. POI data collected from Amap in 2018 comprise a comprehensive collection of 65.27 million distinct POIs. These POIs are classified into three hierarchical levels: first, second, and third. At the first level, there are 24 distinct types, followed by 268 at the second level and 899 at the third level. A subset of these POI types was selected, as listed in Table 1 (https://lbs.amap.com/api/webservice/download, accessed on 15 October 2023).

Table 1. Some examples of the POI types used in the research.

First Level (Level 1)	Second Level (Level 2)	Third Level (Level 3)
Auto Dealers	Toyota Franchised Sales	Toyota Sales
Auto Repair	Chrysler Franchised Repair	Jeep Repair
Motorcycle Service	Motorcycle Sales	BMW Motorcycle Sales
Food and Beverages	Chinese Food Restaurant	Shanghai Food
Shopping	Supermarket	Wal-Mart
Daily Life Service	Information Centre	Enquire of Hotel
Sports and Recreation	Sports Stadium	Gym Center
Medical Service	Special Hospital	Special Hospital
Accommodation Service	Hotel	Five-star Hotel
Tourist Attraction	Park and Square	Park
Commercial House	Building	Commercial-residential Building
Governmental Organization and Social Group	Governmental Organization	State Level Organization and Institution
Science/Culture and Education Service	Media Organization	TV Station
Culture and Education	School	Facilities within the School
Transportation Service	Parking Lot	Public Parking Lot
Finance and Insurance Service	ATM	Bank of China ATM
Enterprises	Company	Network Science and Technology
Road Furniture	Warning Sign	Camera
Place Name and Address	Normal Place Name	Country Name
Public Facility	Public Toilet	Public Toilet
Incidents and Events	Public Event	Conference
Indoor facilities	Indoor facilities	Indoor facilities
Pass Facilities	Gate of Buildings	Main Gate of Buildings

While many studies have selectively screened specific POI types, leading to a limited representation of urban functions [28,29], our research adopted a comprehensive approach by incorporating all available POI types in the dataset. This enabled us to present a detailed analysis of fine granular urban functions, with particular emphasis on POI types at level 3.

The distribution of POI data normalized by population in each city is depicted in Figure 1. The findings reveal that cities with sparse populations in the western regions demonstrate a higher per capita POI count than their counterparts in densely populated cities in the eastern regions. Additionally, provincial capital cities and municipalities exhibit a higher per capita POI count compared to ordinary cities, such as Shanghai, Beijing, Wuhan, Changsha, Nanjing, Guiyang, Kunming, and Guangzhou.

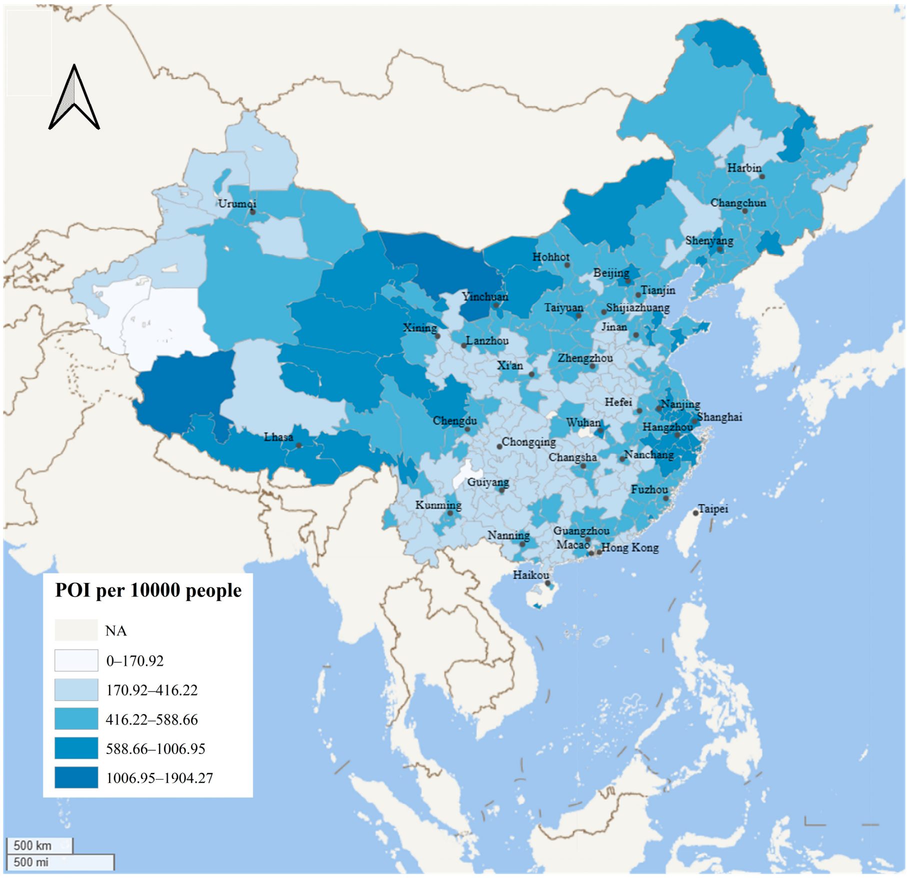

Figure 1. The distribution of POI data normalized by population in Chinese cities.

The GDP is a prominent indicator of urban development. The GDP data used in this study were collected in 2019. Figure 2 provides a visual depiction of the GDP distribution normalized by population across each city. The per capita GDP distribution highlights a more pronounced advantage in provincial capital cities and municipalities compared to ordinary cities, and this advantage is more prominent compared to Figure 1. The unbalanced distribution of per capita POI count and per capita GDP among cities is apparent in both Figures 1 and 2, with provincial capital cities and municipalities displaying a significant advantage.

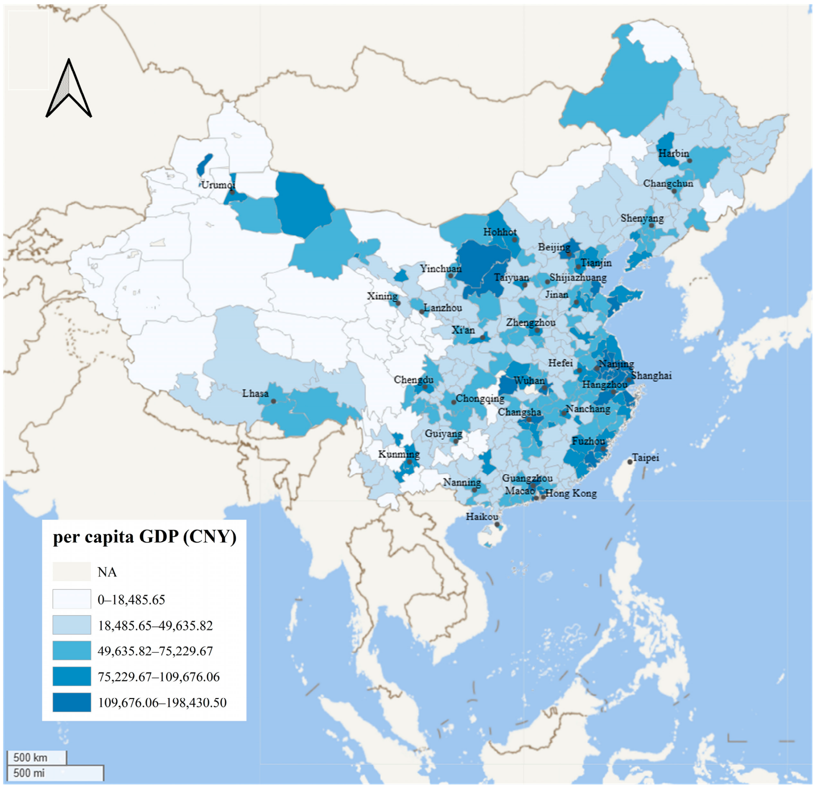

Figure 2. The distribution of GDP data normalized by population in Chinese cities.

3. Methods

We employed POI types to represent urban functions and developed a framework in Figure 3 that encompasses both quantitative and qualitative perspectives to reveal differences in urban functionality, thereby analyzing the imbalance in urban development.

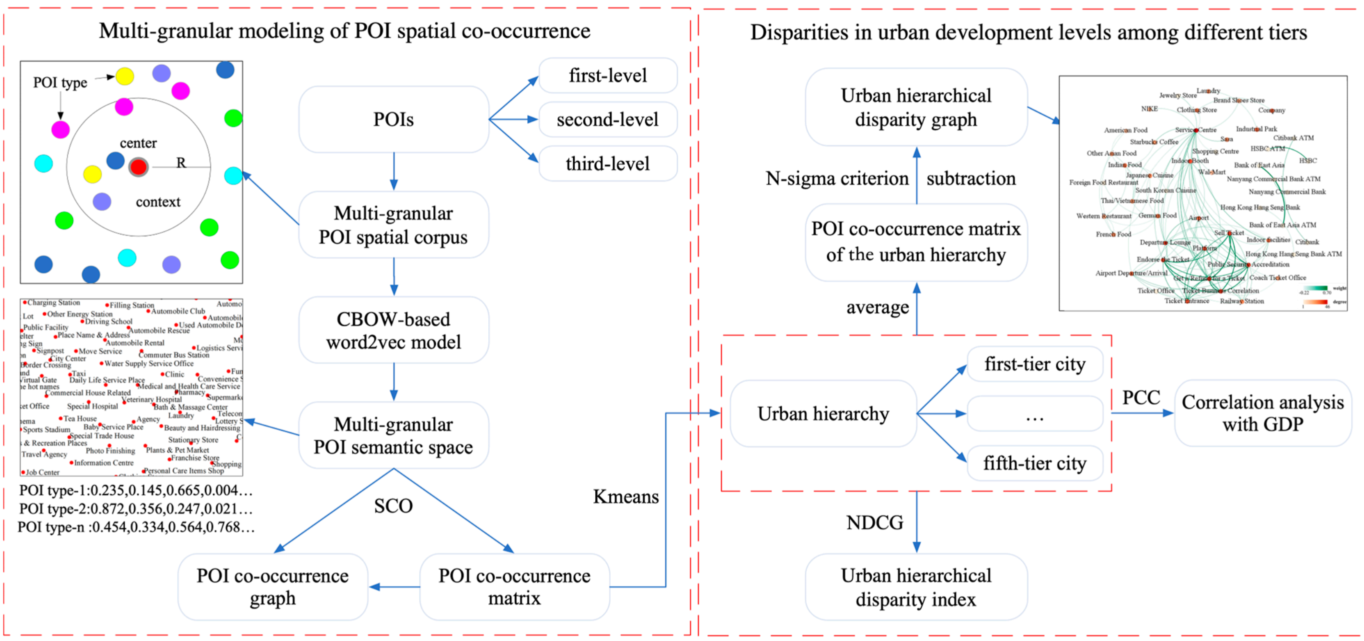

Figure 3. A framework to reveal the imbalance in urban development from the perspective of POI spatial co-occurrence patterns.

We constructed a three-level POI spatial corpus for each city in China, in which POIs were treated as words and POI contexts as sentences. Utilizing the Word2vec model, we developed multigranular POI semantic spaces based on spatial co-occurrence modeling. Each POI type is converted into a corresponding POI vector, with smaller vector distances indicating a higher likelihood of spatial co-occurrence. Subsequently, the multigranular POI semantic spaces are transformed into multigranular POI co-occurrence matrices, which serve as representations of urban functions.

The cities were classified into five tiers using the POI co-occurrence matrix, facilitating the establishment of an urban hierarchy in which cities within the same tier exhibited similar urban development levels. We then calculated the average of the POI co-occurrence matrices for cities within each tier to extract the commonalities of the urban functions. A correlation analysis between POI spatial co-occurrence and GDP was employed to assess differences in urban development levels among the different tiers. We introduced an urban hierarchical disparity index and graph to quantify the extent of variation in urban functions across different tiers and visualized the nature of these differences, which could reveal the imbalance in urban development.

3.1. Multi-Granular Modeling of POI Spatial Co-Occurrence

3.1.1. Construction of a Multi-Granular POI Spatial Corpus

We constructed a POI spatial corpus by drawing inspiration from the corpus construction process in natural language processing. A corpus typically encompasses a vast and structured collection of texts such as documents, sentences, and words [30]. We regarded POIs as words, whereas spatially adjacent POIs formed sentences referred to as POI contexts. The complete set of POI types constituted a dictionary. Consequently, the spatially distributed POIs were transformed into a POI spatial corpus aligned with the principles of the first law of geography [31].

The process of constructing the POI spatial corpus is illustrated in Figure 4. Each POI document is composed of (P_X, P_Y), where P_X indicates the center POI and $P_Y = \{P_1, P_2, \cdots, P_k\}$, denoted as k nearest POIs retrieved with a buffer zone with a radius of R. $\mathrm{dis}(P_X, P_{\varphi-1}) < \mathrm{dis}(P_X, P_\varphi), 1 < \varphi \leq k$, where $\mathrm{dis}(P_X, P_\varphi)$ indicates the Euclidean distance between P_X and P_φ. To evaluate the impact of POI granularity, we establish the multi-granular $P_Y^L, L = 1,2,3$, where L denotes the level of POI types. The dictionary sizes of $P_Y^L, L = 1,2,3$ were 24, 268, and 899, respectively. In accordance with the parameter settings described in a previous study [32], we set the maximum retrieval radius to 1000 m to ensure the inclusion of nearby POIs within the surrounding area. However, in urban areas with a high density of POIs, the sentence length in the POI spatial corpus exceeds the average sentence length found in typical text corpora, which is approximately 30. Therefore, we imposed a maximum limit of 30 neighbors [33].

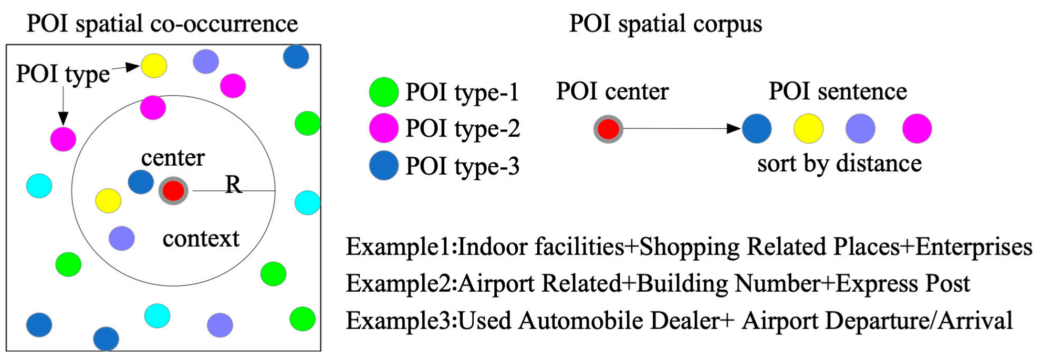

Figure 4. Construction of a POI spatial corpus, where circles represent POIs, and their colors signify distinct POI types.

3.1.2. Learning the POI Spatial Co-Occurrence Based on the Multi-Granular POI Semantic Spaces

The Word2vec model is widely used for POI spatial co-occurrence modeling because of its stability and efficiency [24]. We employed the Word2vec model to construct POI semantic spaces by employing high-dimensional vector representations to capture spatial co-occurrence patterns. We assumed that the size of the POI spatial corpus at level L is H, the sampling window of the context of P_h is c, and the maximum likelihood estimation of the Word2vec model can be expressed as:

$$log(\theta) = \frac{1}{H}\sum_{h=1}^{H} log\ \rho\left(P_h \middle| P_{h-c}^{h+c}\right), \tag{1}$$

where P_{h-c}^{h+c} represents using P_h as the center and c as the sampling window to construct the POI context. $\rho\left(P_h \middle| P_{h-c}^{h+c}\right)$ is defined as:

$$\rho\left(P_h \middle| P_{h-c}^{h+c}\right) = \frac{exp\left(-E\left(P_h, P_{h-c}^{h+c}\right)\right)}{\sum_j^T exp\left(-E\left(P_h, P_{h-c}^{h+c}\right)\right)}, \tag{2}$$

where E is an energy function and $E(P_a, P_b) = -(P_a \cdot P_b)$ [34]. We construct multigranular POI semantic spaces for each city using a multigranular POI spatial corpus. We employed the Gensim package to implement the Word2vec model, utilizing a continuous bag-of-words training method with a window size of five [32]. The dimension of the vectors in the semantic space of NLP trained by Word2vec typically ranges from 50 to 1000 and is determined by the size of the corpus [32]. Given that our POI spatial corpus comprised 65.27 million POIs, which is the largest known spatial corpus, we set the dimensions of the POI vectors to 400.

3.1.3. Urban Functional Representation Based on POI Co-Occurrence Matrix

We constructed POI semantic spaces by modeling the spatial co-occurrence patterns of POIs. Each POI type is represented as a vector, with shorter distances between vectors indicating a higher likelihood of spatial co-occurrence. To establish independent POI semantic spaces for each city, we trained individual Word2vec models. However, this approach introduced random factors, rendering the comparison of POI semantic spaces across different cities challenging. Previous studies have attempted to overcome this limitation by jointly training models for all cities assuming a homogeneous POI context. Nevertheless, the accuracy of these approaches was found to be inferior to that of independent modeling [35]. Consequently, we introduced the concept of POI co-occurrence matrices, which facilitates the comparison of POI semantic spaces across diverse cities.

POI vectors represent absolute positions within their respective POI semantic spaces, with their similarity indicating relative positions. Consequently, the relative relationships between the POI vectors across different cities exhibited similarity and stability. We constructed a POI co-occurrence matrix by leveraging the similarity of the POI vectors. Denoting POI vectors P_b and P_b as a POI pair in city M, we define the spatial co-occurrence (SCO) of a POI pair using cosine similarity as:

$$S(P_a, P_b) = \frac{\sum_{i=1}^{D} P_{a,i} P_{b,i}}{\sqrt{\sum_{i=1}^{D} {P_{b,i}}^2}\sqrt{\sum_{i=1}^{D} {P_{b,i}}^2}}, \tag{3}$$

where S represents the cosine similarity. D denotes the dimensions of the POI vectors. We then calculated the similarity of all the POIs in city M to construct the POI co-occurrence matrix as follows:

$$Mat(M) = \begin{bmatrix} S(P_1, P_1) & \cdots & S(P_1, P_n) \\ \cdots & \cdots & \cdots \\ S(P_n, P_1) & \cdots & S(P_n, P_n) \end{bmatrix} \tag{4}$$

where $Mat(M)$ is the POI co-occurrence matrix, which serves as an alternative representation of the POI semantic space. The comparability of POI co-occurrence matrices across different cities arises from the construction approach, which relies on the similarity of POI vectors [35].

$Mat(M)$ is a symmetric matrix that can be interpreted as the adjacency matrix of an undirected graph. Consequently, we transformed the POI co-occurrence matrix into the POI co-occurrence graph $POIGraph$ as follows:

$$POIGraph = (Vertex, Edge, Weight), \tag{5}$$

$$Edge_{ab} = < P_a, P_b, Weight_{ab} >, \tag{6}$$

$$Weight_{ab} = S(P_a, P_b), \tag{7}$$

where $Vertex$ represents the nodes, which are POI types. $Edge$ denotes the spatial co-occurrence between nodes, and $Weight_{ab}$ indicates the SCO of a POI pair. $POIGraph$ is a completely connected, undirected graph for visualizing the spatial co-occurrence patterns of urban functions.

3.2. Analysis of Disparities in Urban Development Levels among Different Tiers

POI spatial co-occurrence modeling was employed to establish POI semantic spaces, which served as a representation of the spatial distribution of urban functionality. However, cities at different levels of development exhibit variations in their functional distribution, leading to differences in the spatial co-occurrence patterns of POIs. In this section, we explore variations in urban functionality by analyzing differences in POI spatial co-occurrence patterns among cities at different levels of development.

3.2.1. Correlation Analysis between POI Spatial Co-Occurrence and GDPs

We adopted the GDP as a comprehensive indicator to measure the urban development level. The GDP is a monetary metric that quantifies the market value of all final goods and services produced by a city within a specific period. In China, the GDP encompasses various sectors, including farming, forestry, animal husbandry, fishing, industry, construction, wholesale and retail, transport, storage, post, hotels and catering, finance, and real estate. Given its broad coverage, the GDP serves as a comprehensive indicator for assessing the level of urban development [36].

The POI co-occurrence matrix encompasses both the frequency of POIs and their spatial co-occurrence patterns, which reflect the clustering patterns of the functional types represented by POIs during urban development. Hence, we conducted a correlation analysis between the POI co-occurrence matrix and GDPs to investigate the influence of POI spatial co-occurrence patterns on the urban development level.

The POI co-occurrence matrices were transformed into one-dimensional vectors, where each value represents the SCOs of the POI pairs. The dimensions of the vectors corresponding to the POI types of Levels 1, 2, and 3 were 253, 34,453, and 372,816, respectively, within a dataset comprising 297 cities. However, the dimensions of these vectors exceed the number of cities, rendering the regression model susceptible to overfitting. Consequently, we opted to utilize the Pearson Correlation Coefficient (PCC) [32] to evaluate the correlation between SCO and GDP as follows:

$$PCC\left(G, S^{a,b}\right) = \frac{\sum_{i=1}^{N} \left(G_i - \overline{G}\right)\left(S_i^{a,b} - \overline{S^{a,b}}\right)}{\sqrt{\sum_{i=1}^{N} \left(G_i - \overline{G}\right)^2} \sqrt{\sum_{i=1}^{N} \left(S_i^{a,b} - \overline{S^{a,b}}\right)^2}}, \tag{8}$$

where G represents the GDP, $S^{a,b}$ denotes the SCO between P_b and P_b, and N indicates the number of cities.

3.2.2. Constructing the Urban Hierarchy Based on the POI Co-Occurrence Matrix

China has over 300 cities, each exhibiting a variety of geographical locations and cultural disparities. Hence, conducting a comparative analysis of individual cities without considering their commonalities may yield inadequate results. Thus, previous studies have employed classifications to group cities with similar developmental conditions [4,37]. Beijing, Shanghai, Shenzhen, and Guangzhou are widely recognized as first-tier cities in China.

We constructed an urban hierarchy using k-means clustering. Compared to other clustering algorithms, K-means is widely recognized and utilized in urban clustering scenarios owing to its validity, simplicity, and reliability [20,23]. Although determining the number of clusters is a primary challenge in k-means clustering [28], we employed a predetermined number of clusters, five, to represent the five-tier city classification. This choice was based on empirical values established in previous studies [4]. Consequently, we divided the cities into five tiers, representing the first through fifth tiers, by considering the patterns of POI spatial co-occurrence.

3.2.3. Urban Hierarchical Disparity Index and Graph

In the previous section, cities were classified into five tiers. Analyzing the variations in fine granular urban functions among cities with different tiers could reveal an imbalance in urban development. Therefore, we introduced an urban hierarchical disparity index/graph to assess and represent disparities in urban hierarchies from quantitative and qualitative perspectives.

To capture the commonness of POI spatial co-occurrences within a specific tier, we computed the average of the POI co-occurrence matrices from all cities within that tier as follows:

$$Mat_t = \frac{\sum_{i=1}^{n} Mat(i)}{n}, \tag{9}$$

where Mat_t represents the POI co-occurrence matrix of the urban hierarchy, and $Mat(i)$ signifies the POI co-occurrence matrix of a city within tier t. Mat_t can enhance the prominence of common characteristics exhibited in spatial co-occurrence patterns while concurrently attenuating distinctive characteristics.

We introduced the Normalized Discounted Cumulative Gain (NDCG) as the urban hierarchical disparity index, which allows us to measure the similarity between two matrices [38]. First, the POI co-occurrence matrix is flattened into a one-dimensional vector, which is then sorted based on the SCOs, thereby transforming the matrix into a sequence of POI co-occurrences. A smaller difference between the two sequences indicates a greater similarity in the POI co-occurrence matrices and a smaller discrepancy in urban functionality between the two tiers. NDCG is defined as follows:

$$NDCG_p = \frac{DCG_p}{IDCG_p}, \tag{10}$$

$$DCG_p = \sum_{i=1}^{p} \frac{rel(i)}{log_2(i+1)}, \tag{11}$$

where $rel(i)$ represents the correlation score with the 0–1 strategy, where 0 denotes that the two sequences do not have the same element. $IDCG$ represents the maximum DCG. The range of $NDCG$ is [0, 1]. When NDCG is larger, the urban hierarchical disparity is smaller.

Mat_t indicates the commonness of spatial co-occurrence patterns within a specific tier. To examine the disparities in urban functions between the two tiers, we performed a subtraction operation between the POI co-occurrence matrices Mat_r and Mat_z, as follows:

$$Mat_{r-z} = Mat_r - Mat_z, \tag{12}$$

where Mat_{r-z} indicates the urban hierarchical disparity matrix. Given that both Mat_r and Mat_z are symmetric, Mat_{r-z} also possesses symmetry. Consequently, Mat_{r-z} can be converted into a POI co-occurrence graph, that is urban hierarchical disparity graph, facilitating the visualization of urban functional disparities.

4. Experiments and Analysis

4.1. A Case Study of Urban Function with POI Spatial Co-Occurrence

The dimensionality of POI vectors in a semantic space is prohibitively high, making direct visualization impractical. Therefore, we employed a t-distributed Stochastic Neighbor Embedding (t-SNE) model to map the POI semantic space onto a lower-dimensional space for dimension-reduction visualization [39].

Figure 5 presents a visualization of the second-level POI semantic space in Beijing. The figure excludes POI types associated with auto services to enhance clarity in the presentation. The proximity of the POIs in the visualization indicates a higher likelihood of spatial co-occurrence and similarity in spatial semantics. This implies the existence of spatial clustering patterns related to urban functions, which are characterized by relationships based on functional similarity and functional complementarity. Notable examples include pairs such as <School, Library>, <Personal Care Item Shop, Clothing Store, Sports Store>, <Airport Related, Taxi>, <Food & Beverage Related, Chinese Food Restaurant, Fast Food Restaurant, Bakery>, and <Foreign Food Restaurant, Dessert House, Coffee House, and Icecream Shop>. Consequently, the POI semantic space effectively captures and quantifies the spatial co-occurrence of POIs.

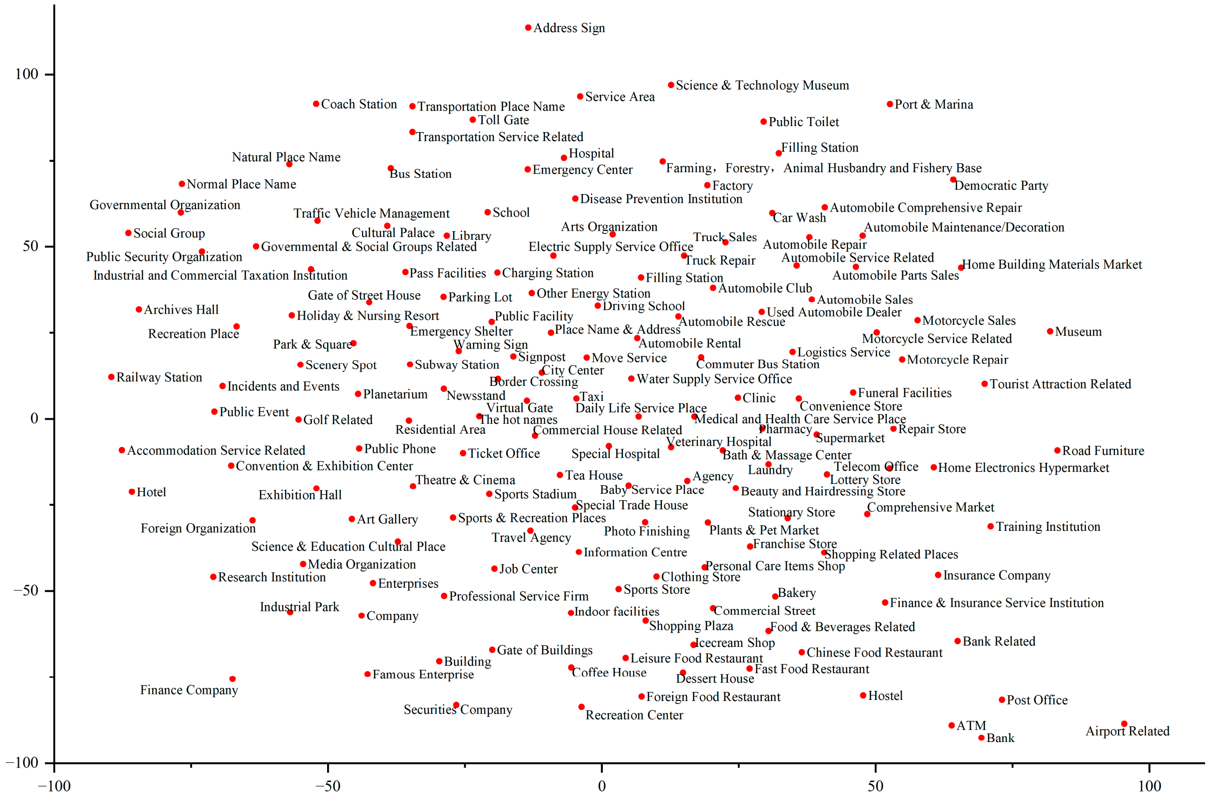

Figure 5. Visualization of the second-level POI semantic space in Beijing.

The POI co-occurrence matrix represents the spatial co-occurrence patterns of the urban functions, where the matrix values reflect the SCOs between POI pairs. Figure 6 presents the POI co-occurrence matrices for Beijing at level 1, offering a succinct depiction of their spatial relationships. Due to the abundance of POI types at level 2, the POI co-occurrence matrices cannot be clearly presented. Therefore, we focus on displaying the POI co-occurrence matrices at level 1.

In Figure 6, the horizontal and vertical axes represent the POI types, while the color gradient illustrates the SCOs of POI pairs. The deeper the color between any two POI types, the larger the SCO, indicating a closer proximity in the POI semantic space. The notable POI spatial co-occurrence patterns with larger SCO values include <Auto Service, Auto Dealers>, <Pass Facilities, Transportation Service>, <Accommodation Service, Sports & Recreation>, <Indoor Facilities, Public Facility>, and <Daily Life Service, Medical Service>, among others. These POI spatial co-occurrence patterns align with common expectations.

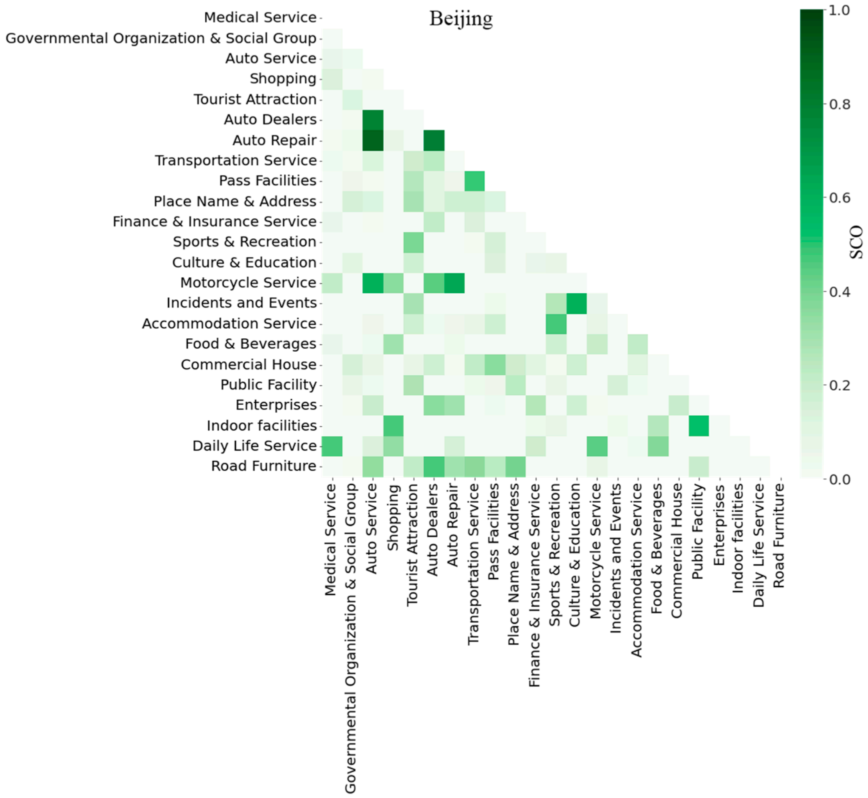

Figure 6. POI co-occurrence matrices in Beijing at level 1.

4.2. Significant Association between POI Spatial Co-Occurrence and Urban Development Level

POI spatial co-occurrence patterns were correlated with GDPs. We computed the PCCs between the POI co-occurrence matrices and GDPs for 297 cities. Figure 7 displays the histograms of the PCCs and p-values at levels 1, 2, and 3. The distributions of PCCs at level 1 and level 2 approximate a normal distribution, with PCCs ranging from -0.6 to 0.6 and -0.8 to 0.8, respectively. However, the PCCs at level 3 followed a positively skewed distribution, indicating a predominantly positive correlation between POI pairs in the co-occurrence matrix and GDPs. We also examined the significance of PCCs and found that PCCs with p-values less than 0.05 accounted for 61.02% (level 1), 59.79% (level 2), and 40.87% (level 3). Additionally, PCCs with p-values less than 0.1 accounted for 67.32% (level 1), 64.75% (level 2), and 45.63% (level 3), respectively. These results suggest that PCC is a reliable indicator. The POI co-occurrence graph at level 3 is presented in Figure 8. To ensure readability while including a sufficient number of POI types, we set the PCC filtering threshold to 0.69, based on empirical considerations.

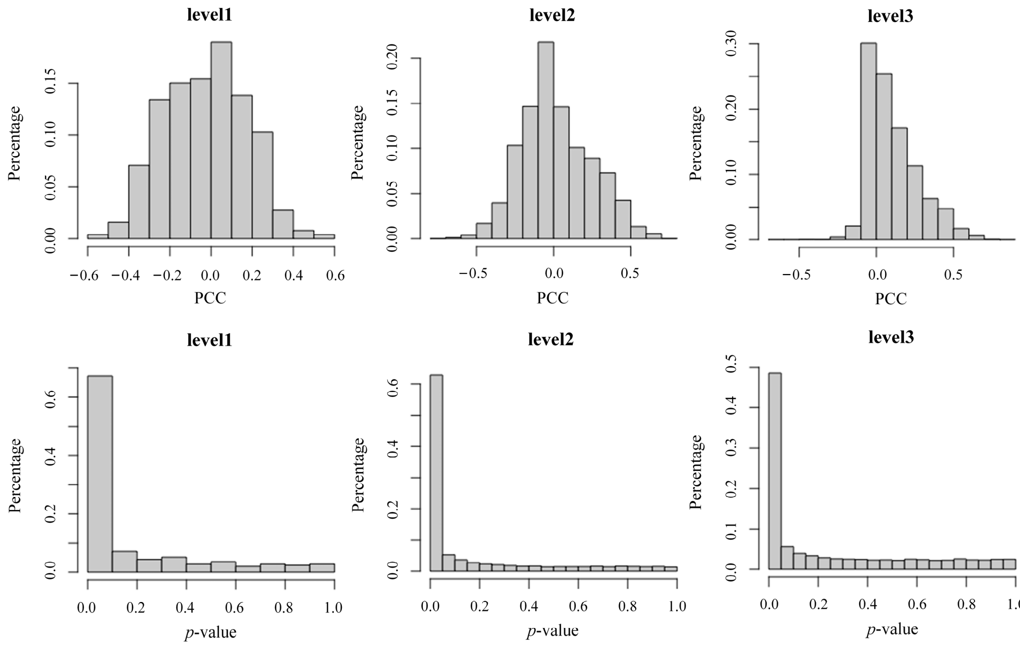

Figure 7. Histograms of PCCs and p-values between the POI co-occurrence matrix and GDP at level 1, level 2, and level 3.

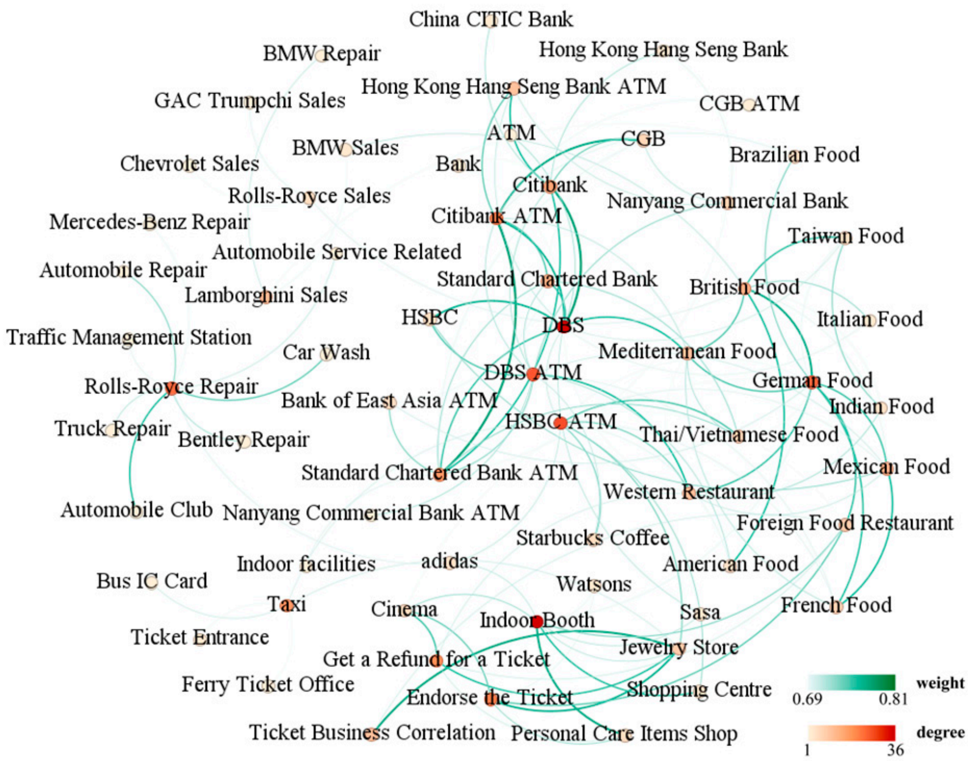

Figure 8. The POI co-occurrence graph at level 3 consists of POI pairs that are correlated with GDP.

In Figure 8, the nodes represent POI types, whereas the edges represent the PCCs between POI pairs. The colors of the nodes and edges correspond to the node degree and edge weight, respectively. Nodes with a higher degree indicate a stronger relevance of the corresponding POI types to the GDPs. These nodes encompass various establishments, such as international banks (Nanyang Commercial Bank, HSBC, Bank of East Asia, Standard Chartered Bank, etc.), foreign restaurants (German Food, British Food, American

Food, etc.), luxury car services (Rolls-Royce Sales, Lamborghini Sales, Bentley Repair, etc.), transportation services (Taxi, Endorse the Ticket, Ticket Business Correlation, etc.), and shopping services (Indoor Booth, Starbucks Coffee, Watsons, Jewelry Store, etc.). The majority of these POI types are prevalent in developed cities. Notably, urban functions tend to aggregate into communities, comprising POI types that exhibit strong correlations with GDPs.

We employed the k-means algorithm to cluster cities based on their POI co-occurrence matrices, which allowed us to establish an urban hierarchy consisting of five tiers: Clu1, Clu2, Clu3, Clu4, and Clu5. Table 2 provides exemplar cities from each tier at level 3. Figure 9 displays the distribution of the GDPs across the different tiers.

Table 2. Some examples of urban hierarchy at level 3.

Clu1	Clu2	Clu3	Clu4	Clu5
Shanghai	Taiyuan	Fuyang	Datong	Tianshui
Beijing	Nanchang	Hengshui	Tongliao	Yan'an
Chengdu	Jinhua	Luan	Xining	Baoshan
Tianjin	Quanzhou	Bengbu	Jilin	Bazhong
Hefei	Guiyang	Dezhou	Anshan	Panzhihua
Jinan	Yantai	Zaozhuang	Daqing	Baise
Qingdao	Nantong	Tai'an	Ordos	Hami
Dongguan	Xuzhou	Liaocheng	Hebi	Zhangye
Foshan	Luoyang	Heze	Panjin	Wuwei
Guangzhou	Taizhou	Shantou	Yingkou	Baiyin
Shenzhen	Jiaxing	Nanyang	Huludao	Jiuquan
Nanjing	Shaoxing	Shangqiu	Yangquan	Lhasa
Wuxi	Baotou	Anyang	Hulunbuir	Anshun
Suzhou	Changchun	Jiaozuo	Bayannaoer	Bijie
Zhengzhou	Lanzhou	Xuchang	Siping	Tongren
Ningbo	Linyi	Zhangzhou	Songyuan	Qingyang
Hangzhou	Jining	Leshan	Chaoyang	Sanya
Wuhan	Baoding	Deyang	Jinzhou	Zhongwei
Changsha	Tangshan	Anqing	Fuxin	Ankang
Xiamen	Langfang	Linfen	Jiamusi	Lincang
Fuzhou	Zhangjiakou	Changzhi	Mudanjiang	Lijiang
Dalian	Cangzhou	Qingyuan	Suihua	Zhaotong
Shenyang	Qinhuangdao	Zhanjiang	Qiqihar	Pu'er
Chongqing	Xingtai	Shangrao	Fushun	Ya'an
Xi'an	Handan	Jiujiang	Ulanqab	Wuzhong
Harbin	Hohhot	Ganzhou	Dandong	Congzuo
Shijiazhuang	Yinchuan	Xinyang	Shizuishan	Laibin

Figure 9 illustrates a decreasing trend in the average GDP values from Clu1 to Clu5. ANOVA tests conducted at Levels 1, 2, and 3 confirmed the significance ($p < 0.05$) of the GDP differences among the tiers. These findings provide evidence of the unbalanced nature of urban development across different tiers.

The categories in Table 2 generally align with established city classifications in China (https://www.chinacheckup.com/places-in-china#groupings, accessed on 18 December 2023). Particularly, Clu1 predominantly consists of well-developed cities in China, mainly provincial capital cities. Nevertheless, it also encompasses Tier 1+ cities—Beijing, Shanghai, Guangzhou, and Shenzhen. While an association exists between POI spatial co-occurrence patterns and GDP, it does not imply complete consistency in patterns among cities. POI spatial co-occurrence patterns consider the spatial distribution features of POIs rather than their quantity. Despite Tier 1+ cities leading in POI data volume and GDP, they share analogous spatial co-occurrence structures with other cities in Clu1.

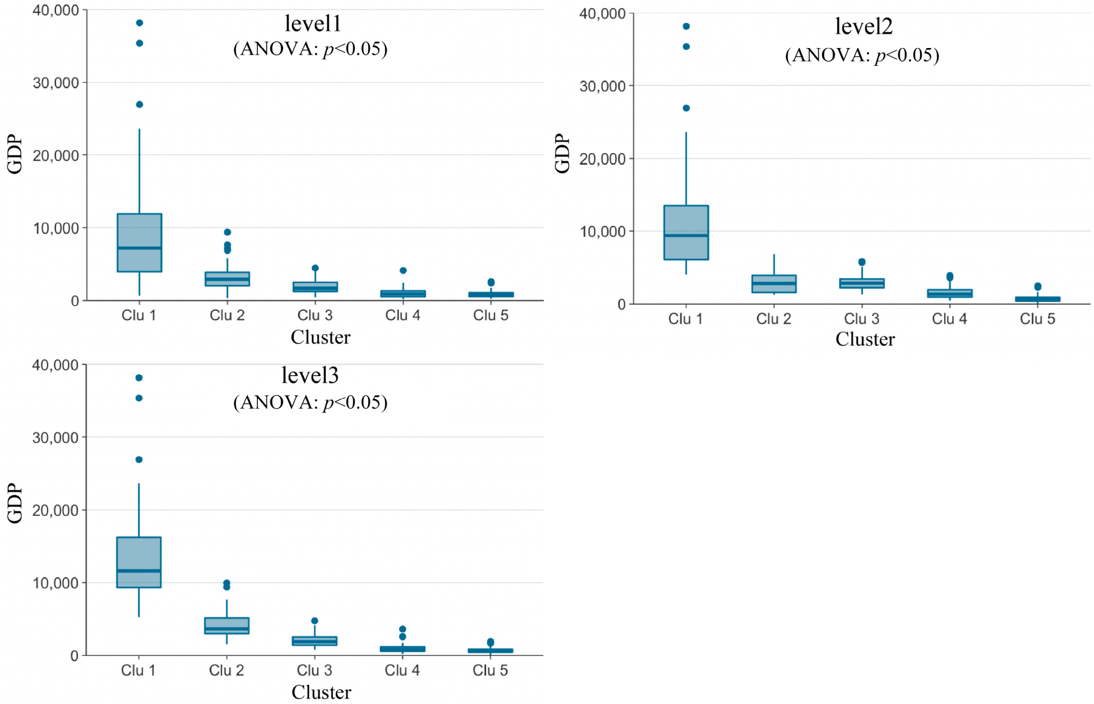

Figure 9. The GDP distribution across the urban hierarchy, ranging from Clu1 to Clu5, demonstrates a declining trend.

4.3. Analysis of Urban Hierarchical Disparity

In the preceding section, we examined the correlation between the POI spatial co-occurrence and the level of urban development. In this section, we use the NDCG as the urban hierarchical disparity index to quantify the variations in urban functions between adjacent tiers and analyze the nature of these differences based on the urban hierarchical disparity graph.

4.3.1. Urban Hierarchical Disparity Index

We clustered the cities into five tiers, ranging from Clu1 to Clu5. To assess the disparity in the urban hierarchy, we introduced NDCG as a measure to quantify the differences between POI co-occurrence matrices. Figure 10 illustrates the NDCGs at level 1, level 2, and level 3. The vertical axis represents NDCG, and the horizontal axis represents the SCO threshold. By selecting elements with SCO values greater than 0.2 in the POI co-occurrence matrix, NDCG was computed, focusing on contributions from significant SCOs. By weighting higher SCO values more prominently, the SCO threshold effectively assigns weights to different elements.

As shown in Figure 10, NDCG exhibited a decreasing trend as the SCO threshold increased at levels 1, 2, and 3. As the SCO threshold increased, the number of elements included in the calculation decreased, resulting in a greater emphasis on elements with higher SCOs and, subsequently, leading to smaller NDCG values. This indicates a large disparity between adjacent tiers in terms of urban functions.

The stability of the NDCG values between the two tiers highlights the superiority of fine granular urban functions. When calculating NDCGs using POI types at levels 1 and 2, the order of NDCG values was not consistent. For example, at level 1, when the threshold was set to 0.2, the NDCG values followed the order: "Clu1–Clu2", "Clu1–Clu3", "Clu2–Clu3", "Clu4–Clu5", "Clu3–Clu4", "Clu1–Clu5", and "Clu1–Clu4". However, as the threshold increased to 0.4, the order changed to "Clu2–Clu3", "Clu1–Clu2", "Clu1–Clu3", "Clu1–Clu5", "Clu4–Clu5", "Clu3–Clu4", and "Clu1–Clu4". In contrast, at level 3, the order of NDCG values remained consistent across different SCO thresholds, indicating that utilizing POI types at level 3 to represent urban functions provided greater stability than POI types at levels 1 and 2.

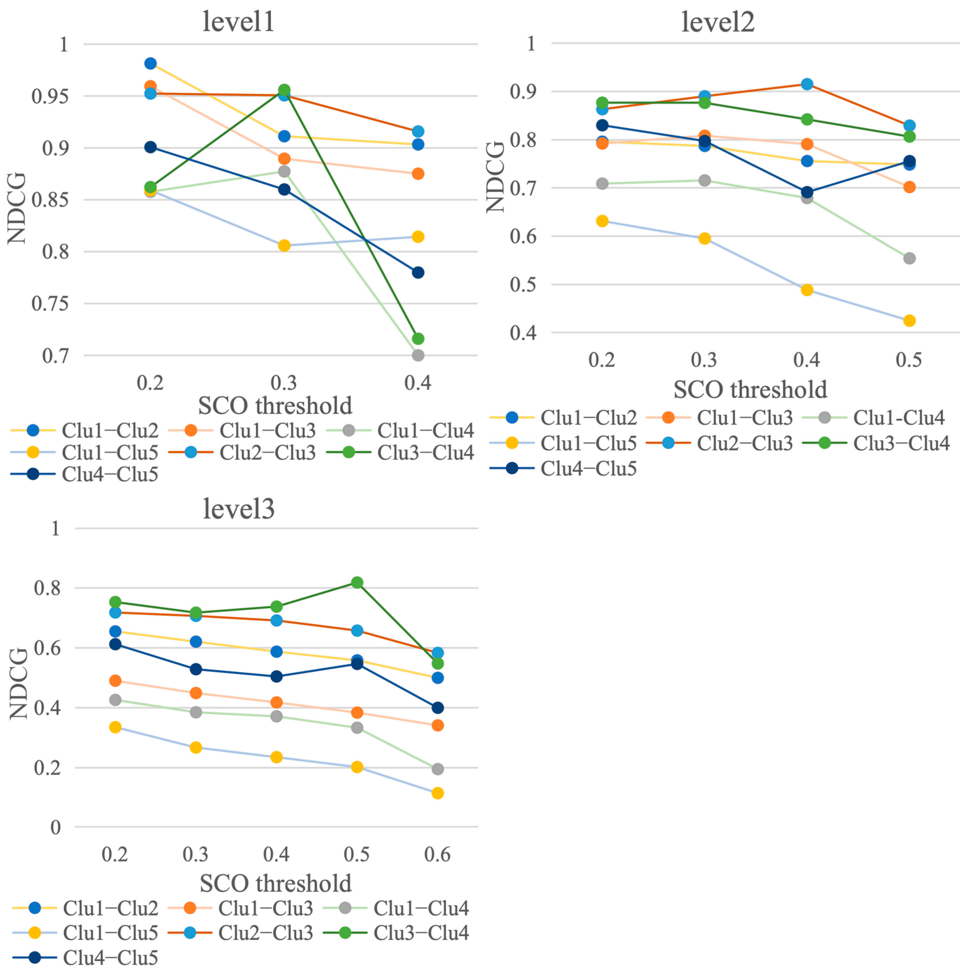

Figure 10. NDCG was used as the urban hierarchical disparity index at level 1, level 2, and level 3.

The NDCG values exhibited a decreasing trend from level 1 to level 3. For example, the NDCG ranges at levels 1, 2, and 3 were [0.7, 1], [0.4, 0.95], and [0.1, 0.85], respectively. These results indicate that the types of POIs at level 1 tend to weaken the differences between cities, whereas the types of POIs at level 3 have the potential to strengthen these differences. Level 3 represents a detailed subdivision of levels 1 and 2, leading to the largest dimension of the POI co-occurrence matrix at level 3. Therefore, level 3 achieves a fine granular expression of urban functions, allowing for a detailed analysis of variations among cities.

The differences between adjacent tiers were smaller than those between non-adjacent tiers. Taking Clu1 as the baseline at level 3, the average NDCG value for "Clu1–Clu2" was 28.77%, 41.5%, and 60.59% greater than that for "Clu1–Clu3", "Clu1–Clu4", and "Clu1–Clu5", respectively. Consequently, the most significant difference was observed in "Clu1–Clu5". These results align with expectations and provide support for the credibility of the urban hierarchical disparity index.

The ranking of urban differences within level 3 is as follows: "Clu3–Clu4" < "Clu2–Clu3" < "Clu1–Clu2" < "Clu4–Clu5" < "Clu1–Clu3" < "Clu1–Clu4" < "Clu1–Clu5". This indicates a trend of polarization in urban differences, particularly evident in the ranking of adjacent level cities: "Clu3–Clu4" < "Clu2–Clu3" < "Clu1–Clu2" < "Clu4–Clu5". Taking "Clu3–Clu4" as the benchmark, the average NDCG values increased by 6.08%, 18.3%, and 27.53% for "Clu2–Clu3", "Clu1–Clu2", and "Clu4–Clu5", respectively. The largest differences were observed between fourth- and fifth-tier cities, followed by first- and second-tier cities, whereas the smallest differences existed among second-, third-, and fourth-tier cities. The results indicate substantial imbalances in urban development as revealed by POI spatial co-occurrence patterns. To mitigate these disparities, concerted efforts should be focused on ameliorating the development gaps in tail-end cities (“Clu4–Clu5”) and head-end cities (“Clu1–Clu2”).

4.3.2. Rationality Analysis of Urban Hierarchical Disparity Matrix

Histograms are used to illustrate the distribution of elements within this matrix to assess the rationality and validity of the urban hierarchical disparity matrix. Figure 11a shows histograms, where the horizontal axis represents the SCOs and the vertical axis denotes the frequencies. The blue and green histograms correspond to Clu1, Clu2, and “Clu1–Clu2”.

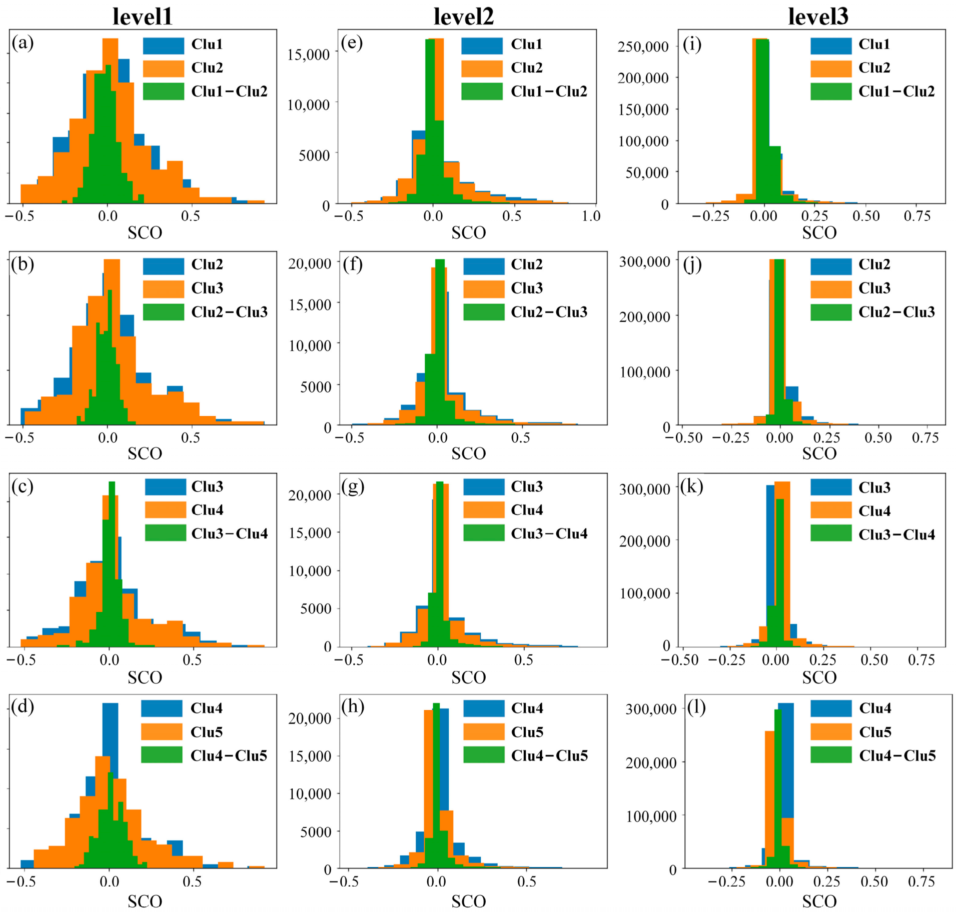

Figure 11. Histograms of POI spatial co-occurrence matrices and urban hierarchical disparity matrices at level 1 (**a**–**d**), level 2 (**e**–**h**), and level 3 (**i**–**l**).

According to the findings presented in Section 4.3.1, the differences observed between adjacent tier “Clu1–Clu2” are 28.77%, 41.5%, 60.59% smaller than those between non-adjacent tiers at level 3. Consequently, the urban hierarchical disparity matrix proved capable of eliminating commonalities. In Figure 11a, the green histogram representing “Clu1–Clu2” exhibits a more concentrated distribution around zero, indicating that the mean and standard deviation of “Clu1–Clu2” were smaller than those of Clu1 and Clu2. Similar patterns can be observed in other figures, further supporting the effectiveness of the urban hierarchical disparity matrix in capturing distinctions within the urban hierarchy.

Figure 12 shows the mean and standard deviations (SD) of the urban hierarchical disparity matrix. The horizontal axis represents the matrices, and the vertical axis represents the mean and SD values. At levels 1, 2, and 3, the mean and SD of the original matrices exceeded those of the urban hierarchical disparity matrix. For instance, the SD of Clu1 and Clu2 at level 1 was three times larger than that of "Clu1–Clu2". Furthermore, the mean value of "Clu3–Clu4" was closer to zero than those of Clu3 and Clu4, indicating that the original POI co-occurrence matrices contained similar elements that were effectively eliminated in the urban hierarchical disparity matrix.

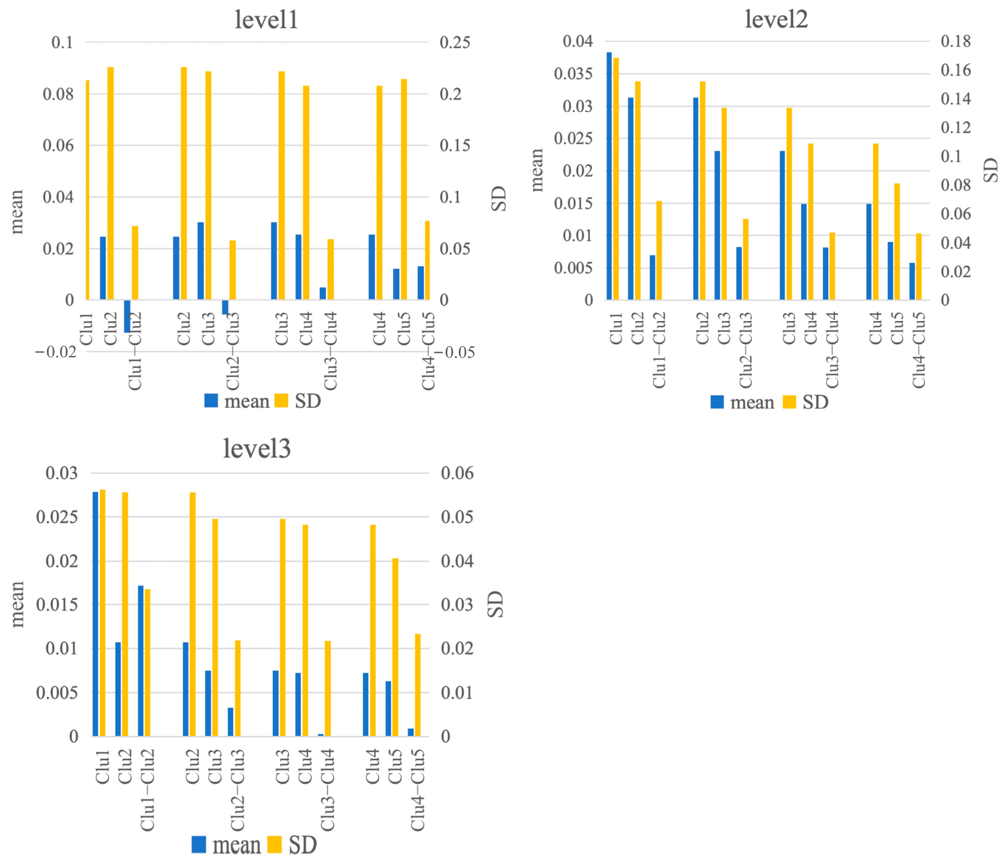

Figure 12. Mean and standard deviation of the urban hierarchical disparity matrix.

4.3.3. Urban Hierarchical Disparity Graph

To comprehensively comprehend the implications of the urban hierarchical disparity matrix and explore the nature of urban hierarchy differences, we transformed the urban hierarchical disparity matrix into an urban hierarchical disparity graph representing the urban hierarchy. To optimize the clarity and interpretability, we employed the N-sigma criterion to eliminate edges with negligible weights. For this experiment, we specifically chose N = 6 to include the maximal number of elements while preserving visual coherence. The POI types at level 3 were utilized to provide a detailed depiction of fine-granular urban functions.

Figure 13 illustrates the disparities in urban functions between the first- and second-tier cities. The nodes in the graph represent specific POI types, and the edges represent variations in urban functions. The colors of the nodes and edges correspond to their respective degrees and weights, where higher values indicate greater differences between the two tiers.

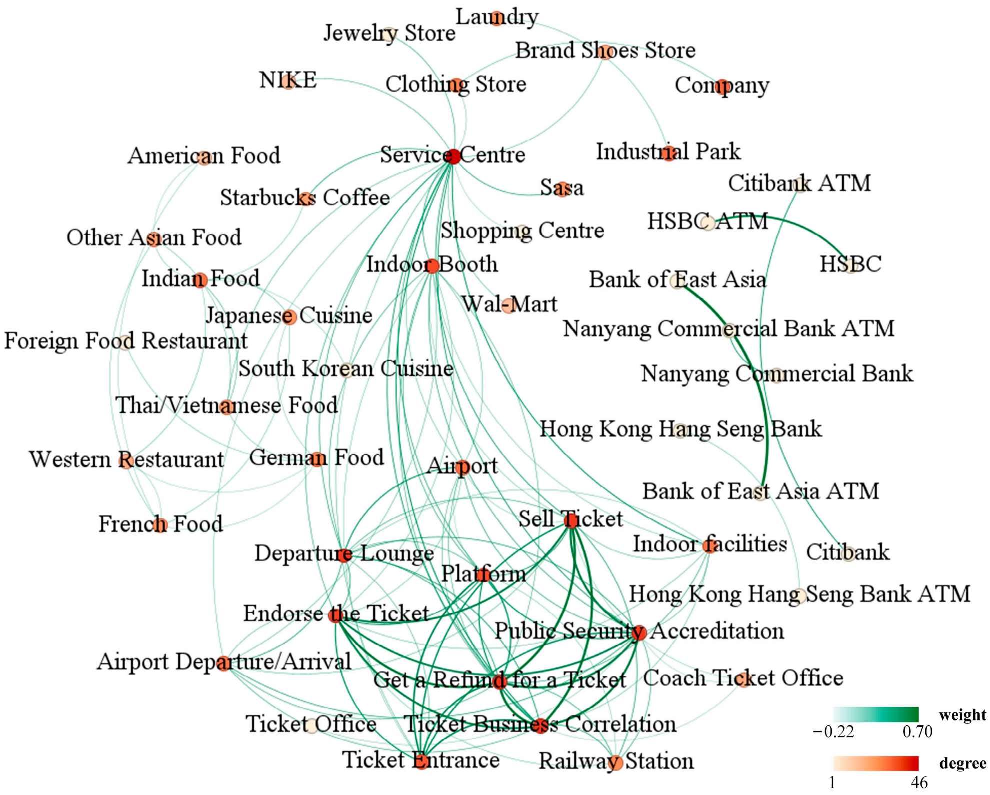

Figure 13. Urban hierarchical disparity graph for Clu1–Clu2 at level 3 indicates the disparities in urban functions between first-tier cities and second-tier cities.

Among the nodes, the one with the highest degree is labeled as "Service Center". This node encompasses co-occurring POI types, such as Indoor Booth, Shopping Center, Jewelry Store, Wal-Mart, Starbucks Coffee, Nike, and Sasa. The presence of Service Centers within large shopping malls signifies that first-tier cities possess more comprehensive business services than second-tier cities. We could identify the community of Foreign Food Restaurants, which contains Japanese Cuisine, American Food, French Food, German Food, South Korean Cuisine, Other Asian Food, Indian Food, and Thai/Vietnamese Food. Foreign restaurants fall under high-end catering consumption, and first-tier cities exhibit a more widespread presence of high-end catering compared to second-tier cities. Additionally, there is a clustering of transportation communities, including Airport, Railway Station, Ticket Business Correlation, and Departure Lounge, implying that first-tier cities boast more advanced transportation facilities than second-tier cities. The figure also highlights the appearance of nodes associated with Industrial Park, Company, and Banks (such as HSBC, Nanyang Commercial Bank, and Citibank), indicating the thriving commercial environment prevalent in first-tier cities.

Figure 14 depicts the disparities in urban functions between second- and third-tier cities. The graph reveals several distinct communities, including hospitals (AAA Class Hospital, Special Hospital, Chest Hospital, Brain Hospital), banks (Industrial Bank, China Minsheng Bank, Huaxia Bank, China CITIC Bank), and Noble Sports (Golf Course, Horse Riding Club, Bowling Hall). In comparison to Figure 12, the spatial co-occurrence patterns of foreign restaurants decreased. There is an emergence of Chinese local dining, such as Chinese Food Restaurant—Anhui Food, Local Special Food, Sichuan Food, Shandong Food, Islamic Food, Beijing Food, Yunan and Guizhou Food. The spatial co-occurrence patterns of Shopping Mall (Wal-Mart, Watsons, Brand Shoes Store, Sports Store, Nike, KFC, Pizza Hut, Bakery, Starbucks Coffee) no longer revolve around the Service Center. These communities illustrate the superiority of second-tier over third-tier cities in terms of healthcare facilities, banking services, sports amenities, and catering.

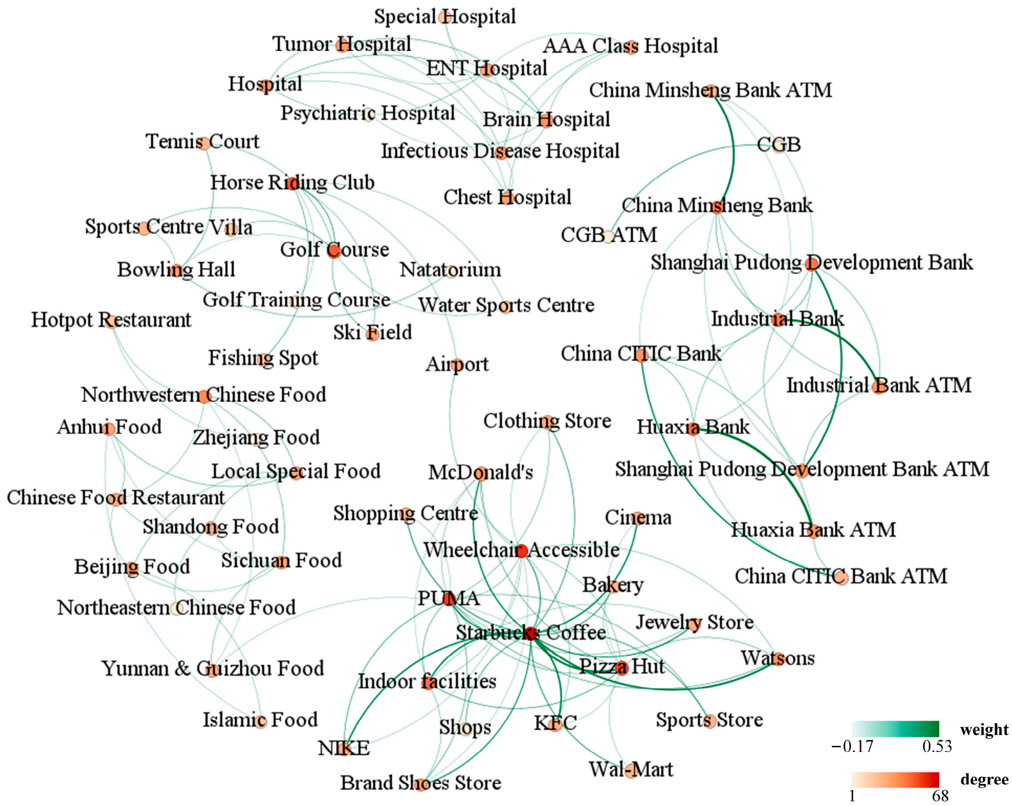

Figure 14. Urban hierarchical disparity graph for Clu2-Clu3 at level 3 indicates the disparities in urban functions between second-tier cities and third-tier cities.

Figure 15 illustrates the disparities in urban functions between third- and fourth-tier cities. Notably, the node with the highest degree is the Ski Field; however, the weights of the corresponding connecting edges are uniformly negative. This observation indicates that Ski Fields are more prevalent in fourth-tier cities than in third-tier ones. This discrepancy can be attributed to the climatic variations between Northern and Southern China. Colder winters in the northern regions enable the establishment of Ski Fields, whereas insufficient snowfall in the southern regions limits their presence. Consequently, the distribution of Ski Fields reflects the economic disparities between Northern and Southern China, with a greater number of fourth-tier cities located in the northern regions.

The distinctions between third- and fourth-tier cities manifest in various sectors, including catering, shopping, and agriculture. Catering services include establishments such as Steak House, KFC, Hotpot Restaurant, Local Special Food, Icecream Shop, and Bakery. Shopping services encompass Brand Clothing Store, Nike, Li-Ning, Clothing Store, Deco Cloth Store, Supermarket, and Convenience Store. The agricultural domain includes Fruit Cultivation Base, Farm, Fishing Farm, Forest Farm, and Flower Nursery Base. Thus, the disparities between third- and fourth-tier cities predominantly revolve around fundamental service offerings.

Figure 16 shows the disparities in urban functions between fourth- and fifth-tier cities. This aligns with Figure 15, in which the community primarily comprises fundamental services related to catering and shopping. Figure 16 shows the abundance of POI types associated with natural features, including Lake, Mountain, Bridge, Garden, and Park, and their co-occurrence with Cemetery and Funeral House. These POI types are less prevalent in developed cities but exhibit a higher likelihood of co-occurrence in underdeveloped cities.

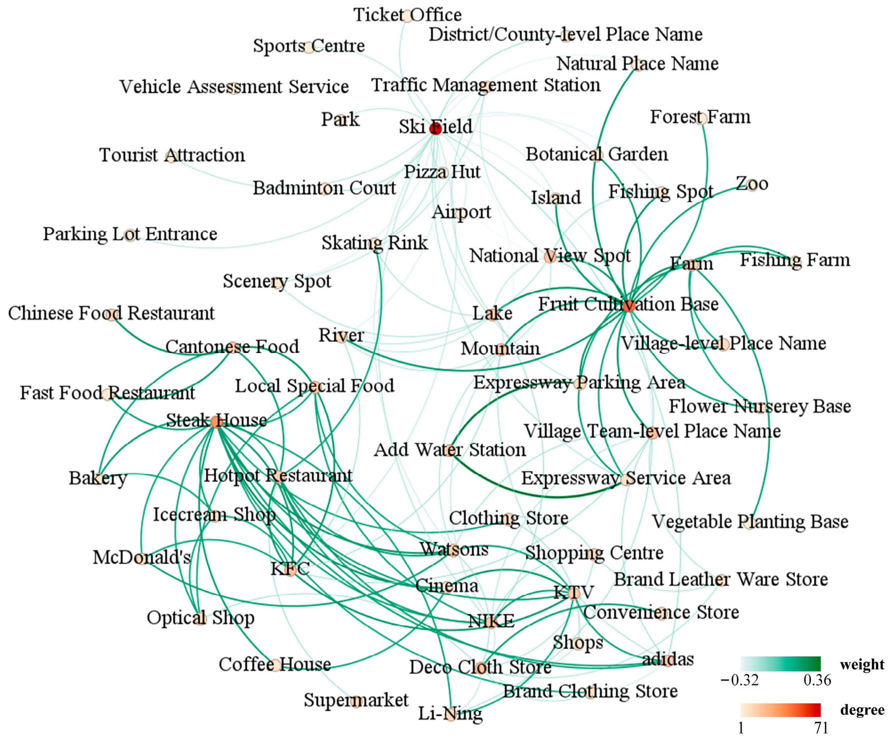

Figure 15. Urban hierarchical disparity graph for Clu3-Clu4 at level 3 indicates the disparities in urban functions between third-tier cities and fourth-tier cities.

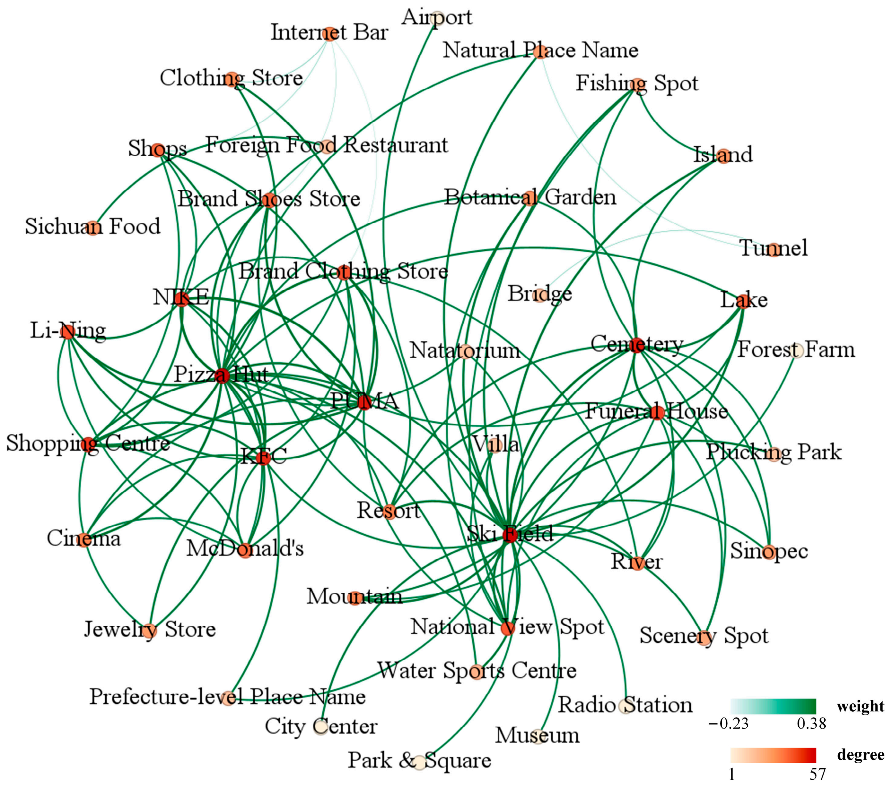

Figure 16. Urban hierarchical disparity graph for Clu4-Clu5 at level 3 indicates the disparities in urban functions between fourth-tier cities and fifth-tier cities.

5. Discussion

Unbalanced development among cities at different levels has become a consensus and is considered a prominent social contradiction in China [2]. Urban functions play a crucial role in reflecting the spatial distribution of industries and are closely associated with imbalances in urban development [19]. This study aims to reveal urban development imbalances from a fine-granular perspective of urban functions, with a specific emphasis on utilizing POI spatial co-occurrence modeling. We confirmed that defining fine-granular urban functions based on third-level POIs yields more precise and robust outcomes. Additionally, previous research on urban functions has predominantly focused on developed first- and second-tier cities [10,27], failing to explore variations in urban functions among cities at different levels of development. This study employed POI spatial co-occurrence modeling to extract urban functions for 297 cities in China and classified them into five tiers. We established a significant correlation between POI spatial co-occurrence patterns and GDPs, enabling the further analysis of functional disparities among cities at different tiers. These findings provide insights into urban development imbalances and offer valuable references for urban planning and policymaking.

Notably, this study substantiates the validity of employing POI spatial co-occurrence for elucidating the disparities in urban development across Chinese cities. Initially, we converted POI co-occurrence patterns into vectors and classified them into five categories, as delineated in Table 2. These categories align with the established city classifications in China. Additionally, an exponential decrease in GDP is observed from Clu1 to Clu5, as depicted in Figure 9. In Figure 8, we delineate the POI types exhibiting correlation with GDP, encompassing international banks, foreign restaurants, luxury car services, transportation services, and shopping services. These POI types are typically linked with elevated consumption, and their prevalence augments with urban development. The evidence implies that POI spatial co-occurrence patterns function as discernible indicators of the urban development.

The variations in POI spatial co-occurrence among cities highlight imbalances in urban development. A quantitative analysis revealed the following ranking of fine-granular functional differences across city levels: "Clu3–Clu4" < "Clu2–Clu3" < "Clu1–Clu2" < "Clu4–Clu5". Taking "Clu3–Clu4" as the baseline, the average NDCG values increased by 6.08%, 18.3%, and 27.53% for "Clu2–Clu3", "Clu1–Clu2," and "Clu4–Clu5", respectively. The most significant disparities are evident between fourth- and fifth-tier cities. Their POI spatial co-occurrence patterns exemplify urban fundamental service, including catering, shopping, and farming, with POIs associated with natural features and rural areas. The fifth-tier cities are underdeveloped urban areas that exhibit deficiencies in their basic functions and economic progress. Unfortunately, these cities often go unnoticed and do not receive public or government attention. Therefore, it is imperative for government development strategies to prioritize the advancement of fifth-tier cities to effectively address urban imbalances. Notably, first-tier cities differed significantly from second-tier cities. First-tier cities typically serve as provincial capitals, consolidating the advantageous resources of their respective provinces and augmenting their dominant positions. First-tier cities exhibit greater levels of internationalization and commercialization than second-tier cities, hosting a larger number of international banks, companies, and upscale shopping malls, indicating a more sophisticated business environment. Enhancing the business environment has emerged as a critical developmental focus in second-tier cities.

Past urban development strategies in China have facilitated rapid economic growth; however, they have also significantly widened the gap between leading and lagging cities. In contrast, the differences among the second-, third-, and fourth-tier cities were relatively smaller. Coordinating resource allocation and establishing sustainable development strategies, where leading cities support the development of lagging cities, are essential for addressing the issue of imbalanced urban development.

6. Conclusions

Urban functional agglomerations play a significant role in shaping the widespread spatial co-occurrence patterns of POIs. In this study, we aimed to uncover the imbalances in urban development among cities at different levels by examining POI spatial co-occurrence patterns. Our analysis focused on 297 cities in China, from which we extracted and analyzed patterns to gain insights into comprehensive and fine-granular differences in urban functionality. The results confirm that POI spatial co-occurrence patterns serve as crucial indicators of imbalances in urban development. The fine-granular POI analysis yields optimal results in characterizing urban functionality, albeit at the cost of reduced interpretability. The findings underscore a notable polarization in the development of Chinese cities, with highly advanced top-tier cities and significantly underdeveloped tail-end cities. Top-tier cities exhibit disparities in the commercial environment, while cities at the lower end of the hierarchy manifest discrepancies in terms of urban infrastructure.

This study contributes to a comprehensive understanding of the functional disparities among cities at distinct levels, thereby providing valuable insights into avenues for mitigating urban imbalances in China. Similarities in POI spatial co-occurrence patterns are observed across cities at each tier. Consequently, governmental efforts to enhance the GDP of a specific city can draw insights from the spatial co-occurrence patterns of higher-level cities as reference models for urban planning. For example, if the objective is to augment the GDP of a third-tier city, planners may consult the spatial co-occurrence patterns of second-tier cities for strategic urban functional layout planning. This study serves as a valuable reference for urban development and planning strategies.

This study has some limitations that should be acknowledged. Employing fine granular urban functions based on third-level POIs proves to be optimal for characterizing urban functionality. However, addressing the challenge of enhancing clarity in presentation remains essential. The classification of POIs in this study was built using Amap, which may have issues of imbalance in category distribution, such as the overrepresentation of the automobile category. The impact of the classification method on the research results has not been thoroughly assessed. Furthermore, this study focused solely on cities in China, and its applicability to cities in other countries remains to be evaluated. These limitations suggest directions for future research.

Author Contributions: Conceptualization, Rui Li and Huayi Wu; methodology, Guangsheng Dong and Fa Li; software, Guangsheng Dong; validation, Zhaohui Liu; investigation, Wensen Yu and Fangning Li; resources, Jie Jiang and Hongping Zhang; writing—original draft preparation, Guangsheng Dong; writing—review and editing, Rui Li and Longgang Xiang; visualization, Zhaohui Liu; project administration, Rui Li; funding acquisition, Rui Li and Guangsheng Dong. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by these fundings: National Natural Science Foundation of China [grant numbers: U20A2091; 41771426; 42301485]; China Postdoctoral Science Foundation [Certificate Number: 2023M732682]; Hubei Provincial Natural Science Foundation of China [Grant Number: 2023AFB184]; Open Project Program of Fujian Key Laboratory of Big Data Application and Intellectualization for Tea Industry, Wuyi University; Open Fund of Key Laboratory of Urban Spatial Information, Ministry of Natural resources.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chen, M.; Liu, W.; Tao, X. Evolution and Assessment on China's Urbanization 1960–2010: Under-Urbanization or over-Urbanization? *Habitat Int.* **2013**, *38*, 25–33. [CrossRef]
2. Xi, J. Secure a Decisive Victory in Building a Moderately Prosperous Society in All Respects and Strive for the Great Success of Socialism with Chinese Characteristics for a New Era. In Proceedings of the 19th National Congress of the Communist Party of China, Beijing, China, 18–24 October 2017; Volume 18, pp. 2011–2017.

3. Cao, S.; Lv, Y.; Zheng, H.; Wang, X. Challenges Facing China's Unbalanced Urbanization Strategy. *Land Use Policy* **2014**, *39*, 412–415. [CrossRef]
4. Xu, J.; Li, A.; Li, D.; Liu, Y.; Du, Y.; Pei, T.; Ma, T.; Zhou, C. Difference of Urban Development in China from the Perspective of Passenger Transport around Spring Festival. *Appl. Geogr.* **2017**, *87*, 85–96. [CrossRef]
5. Tan, Y.; Xu, H.; Zhang, X. Sustainable Urbanization in China: A Comprehensive Literature Review. *Cities* **2016**, *55*, 82–93. [CrossRef]
6. Zhou, Y.; Guo, Y.; Liu, Y. High-Level Talent Flow and Its Influence on Regional Unbalanced Development in China. *Appl. Geogr.* **2018**, *91*, 89–98. [CrossRef]
7. Zünd, D.; Bettencourt, L.M.A. Growth and Development in Prefecture-Level Cities in China. *PLoS ONE* **2019**, *14*, e0221017. [CrossRef]
8. Manoli, G.; Fatichi, S.; Schläpfer, M.; Yu, K.; Crowther, T.W.; Meili, N.; Burlando, P.; Katul, G.G.; Bou-Zeid, E. Magnitude of Urban Heat Islands Largely Explained by Climate and Population. *Nature* **2019**, *573*, 55–60. [CrossRef]
9. Chen, J.; Chen, J.; Miao, Y.; Song, M.; Fan, Y. Unbalanced Development of Inter-Provincial High-Grade Highway in China: Decomposing the Gini Coefficient. *Transp. Res. Part Transp. Environ.* **2016**, *48*, 499–510. [CrossRef]
10. Chen, Y.; Chen, X.; Liu, Z.; Li, X. Understanding the Spatial Organization of Urban Functions Based on Co-Location Patterns Mining: A Comparative Analysis for 25 Chinese Cities. *Cities* **2020**, *97*, 102563. [CrossRef]
11. Bettencourt, L.M.A. The Origins of Scaling in Cities. *Science* **2013**, *340*, 1438–1441. [CrossRef]
12. Li, D.; Shan, J.; Shao, Z.; Zhou, X.; Yao, Y. Geomatics for Smart Cities-Concept, Key Techniques, and Applications. *Geo-Spat. Inf. Sci.* **2013**, *16*, 13–24. [CrossRef]
13. Qian, J.; Liu, Z.; Du, Y.; Liang, F.; Yi, J.; Ma, T.; Pei, T. Quantify City-Level Dynamic Functions across China Using Social Media and POIs Data. *Comput. Environ. Urban Syst.* **2021**, *85*, 101552. [CrossRef]
14. Liu, Z.; Liu, H.; Lang, W.; Fang, S.; Chu, C.; He, F. Scaling Law Reveals Unbalanced Urban Development in China. *Sustain. Cities Soc.* **2022**, *87*, 104157. [CrossRef]
15. Guan, Q.; Zhou, J.; Wang, R.; Yao, Y.; Qian, C.; Zhai, Y.; Ren, S. Understanding China's Urban Functional Patterns at the County Scale by Using Time-Series Social Media Data. *J. Spat. Sci.* **2022**, *68*, 631–649. [CrossRef]
16. Huang, W.; Cui, L.; Chen, M.; Zhang, D.; Yao, Y. Estimating Urban Functional Distributions with Semantics Preserved POI Embedding. *Int. J. Geogr. Inf. Sci.* **2022**, *36*, 1905–1930. [CrossRef]
17. Mei, Y.; Gui, Z.; Wu, J.; Peng, D.; Li, R.; Wu, H.; Wei, Z. Population Spatialization with Pixel-Level Attribute Grading by Considering Scale Mismatch Issue in Regression Modeling. *Geo-Spat. Inf. Sci.* **2022**, *25*, 365–382. [CrossRef]
18. Aslam, N.S.; Ibrahim, M.R.; Cheng, T.; Chen, H.; Zhang, Y. ActivityNET: Neural Networks to Predict Public Transport Trip Purposes from Individual Smart Card Data and POIs. *Geo-Spat. Inf. Sci.* **2021**, *24*, 711–721. [CrossRef]
19. Liu, Z.; Chen, X.; Xu, W.; Chen, Y.; Li, X. Detecting Industry Clusters from the Bottom up Based on Co-Location Patterns Mining: A Case Study in Dongguan, China. *Environ. Plan. B Urban Anal. City Sci.* **2021**, *48*, 2827–2841. [CrossRef]
20. Gao, S.; Janowicz, K.; Couclelis, H. Extracting Urban Functional Regions from Points of Interest and Human Activities on Location-based Social Networks. *Trans. GIS* **2017**, *21*, 446–467. [CrossRef]
21. Li, T.; Zheng, X.; Zhang, C.; Wang, R.; Liu, J. Mining Spatial Correlation Patterns of the Urban Functional Areas in Urban Agglomeration: A Case Study of Four Typical Urban Agglomerations in China. *Land* **2022**, *11*, 870. [CrossRef]
22. Ma, S.; Long, Y. Functional Urban Area Delineations of Cities on the Chinese Mainland Using Massive Didi Ride-Hailing Records. *Cities* **2020**, *97*, 102532. [CrossRef]
23. Zhai, W.; Bai, X.; Shi, Y.; Han, Y.; Peng, Z.-R.; Gu, C. Beyond Word2vec: An Approach for Urban Functional Region Extraction and Identification by Combining Place2vec and POIs. *Comput. Environ. Urban Syst.* **2019**, *74*, 1–12. [CrossRef]
24. Dong, G.; Li, R.; Wu, H.; Huang, W.; Zhang, H.; Tao, V.; Liu, Q. Learning the Spatial Co-Occurrence for Browsing Interests Extraction of Domain Users on Public Map Service Platforms. *Geo-Spat. Inf. Sci.* **2022**, 1–20. [CrossRef]
25. Goodchild, M.F. Commentary: General Principles and Analytical Frameworks in Geography and GIScience. *Ann. GIS* **2022**, *28*, 85–87. [CrossRef]
26. Xue, B.; Xiao, X.; Li, J. Identification Method and Empirical Study of Urban Industrial Spatial Relationship Based on POI Big Data: A Case of Shenyang City, China. *Geogr. Sustain.* **2020**, *1*, 152–162. [CrossRef]
27. Cardoso, R.V.; Meijers, E.J. Contrasts between First-Tier and Second-Tier Cities in Europe: A Functional Perspective. *Eur. Plan. Stud.* **2016**, *24*, 996–1015. [CrossRef]
28. Hu, S.; He, Z.; Wu, L.; Yin, L.; Xu, Y.; Cui, H. A Framework for Extracting Urban Functional Regions Based on Multiprototype Word Embeddings Using Points-of-Interest Data. *Comput. Environ. Urban Syst.* **2020**, *80*, 101442. [CrossRef]
29. Zhang, X.; Li, W.; Zhang, F.; Liu, R.; Du, Z. Identifying Urban Functional Zones Using Public Bicycle Rental Records and Point-of-Interest Data. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 459. [CrossRef]
30. Zeroual, I.; Lakhouaja, A. MulTed: A Multilingual Aligned and Tagged Parallel Corpus. *Appl. Comput. Inform.* **2020**, *18*, 61–73. [CrossRef]
31. Tobler, W.R. A Computer Movie Simulating Urban Growth in the Detroit Region. *Econ. Geogr.* **1970**, *46*, 234–240. [CrossRef]
32. Liu, K.; Yin, L.; Lu, F.; Mou, N. Visualizing and Exploring POI Configurations of Urban Regions on POI-Type Semantic Space. *Cities* **2020**, *99*, 102610. [CrossRef]

33. Chen, J.; Chen, Q.; Liu, X.; Yang, H.; Lu, D.; Tang, B. The Bq Corpus: A Large-Scale Domain-Specific Chinese Corpus for Sentence Semantic Equivalence Identification. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; Association for Computational Linguistics: Kerrville, TX, USA, 2018; pp. 4946–4951.
34. Yao, Y.; Li, X.; Liu, X.; Liu, P.; Liang, Z.; Zhang, J.; Mai, K. Sensing Spatial Distribution of Urban Land Use by Integrating Points-of-Interest and Google Word2Vec Model. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 825–848. [CrossRef]
35. Liu, X.; Andris, C.; Rahimi, S. Place Niche and Its Regional Variability: Measuring Spatial Context Patterns for Points of Interest with Representation Learning. *Comput. Environ. Urban Syst.* **2019**, *75*, 146–160. [CrossRef]
36. Hu, Z.; Wu, G.; Han, Y.; Niu, Y. Unraveling the Dynamic Changes of High-Speed Rail Network with Urban Development: Evidence from China. *Socioecon. Plann. Sci.* **2023**, *85*, 101380. [CrossRef]
37. Jin, X.; Long, Y.; Sun, W.; Lu, Y.; Yang, X.; Tang, J. Evaluating Cities' Vitality and Identifying Ghost Cities in China with Emerging Geographical Data. *Cities* **2017**, *63*, 98–109. [CrossRef]
38. McKenzie, G.; Romm, D. Measuring Urban Regional Similarity through Mobility Signatures. *Comput. Environ. Urban Syst.* **2021**, *89*, 101684. [CrossRef]
39. Van der Maaten, L.; Hinton, G. Visualizing Data Using T-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.