ISSN: 2710-0170

Research on the Spatial Heterogeneity of Urban Vitality Driven by Social Big Data: A Case Study of Xi'an

Hao Zhang

Henan Polytechnic University, China

Abstract

Based on multi-scale geographically weighted regression (MGWR) and spatial autocorrelation analysis, this paper explores the impact of different urban functional elements on urban vitality and their spatial heterogeneity. The results show that urban vitality exhibits significant spatial clustering within the study area, with both high and low values showing a clustered state. Scenic spots, business residences, and science, education, and culture facilities have a significant negative effect on urban vitality, while the positive effects of public facilities and life services are not significant. This suggests that the spatial layout and complexity of functional land use play an important role in urban vitality. The MGWR model significantly outperforms the traditional OLS model in terms of goodness of fit and explanatory power, indicating that considering spatial heterogeneity is crucial for revealing the formation mechanism of urban vitality. This study provides empirical reference and methodological support for urban spatial planning and functional layout optimization.

Keywords

Urban Vitality, MGWR, Spatial Autocorrelation, Urban Spatial Planning.

1. Introduction

The concept of urban vitality can be traced back to Jane Jacobs's theory on the subject. She proposed that urban vitality originates from urban diversity, stating that "vitality is a concentrated manifestation of urban diversity." This perspective underscores the importance of human activities and spatial diversity for a city's liveliness[1].

In contrast, Kevin Lynch defined urban vitality as the ability of an urban system to sustain its internal survival, growth, and development. His view highlights the quality of urban form and urban space as the key determinants. Specifically, scholars have noted that "the measurement of urban vitality reflects the interactions among social, economic, and environmental factors," and that the work of Jane Jacobs and others laid the theoretical foundation for the concept.

Consequently, the notion of urban vitality is based on a comprehensive understanding of urban diversity, spatial form, and human activities, and has received widespread attention in the fields of urban planning and sociology[2].

The core essence of urban vitality primarily originates from the four key "generators" proposed by Jane Jacobs in her 1961 book, The Death and Life of Great American Cities: diversity of land use, small block size, diversity of economic activities, and a high concentration of people[3].

Together, these factors foster continuous pedestrian activity within urban spaces, creating the dynamism of city life [4]. The diversity of land use, which includes a mix of residential, office, industrial, and cultural uses, attracts diverse groups of people for various activities at different times, thereby enhancing the safety and appeal of a space. Small blocks and short streets increase the number of intersections, which promotes social interaction and natural surveillance and helps reduce crime. The diversity of economic activities and high population density, in turn, inject a continuous flow of people and a dynamic energy into a neighborhood.

ISSN: 2710-0170

Recent studies have also empirically validated the effectiveness of these indicators using big data and remote sensing technologies, providing a scientific basis for urban planning and vitality assessment[5].

Existing research reveals the primary factors influencing urban vitality through a variety of data sources. In top-tier international journals, common data sources include population distribution, mobile signaling, and socioeconomic statistics. For example, one study noted that "the most important data directly reflecting urban vitality are population distribution and human activity data, such as census and mobile signaling data; however, these are difficult to use widely due to data privacy and limited acquisition frequency." Additionally, some studies use social media check-in data, taxi GPS data, and street view imagery to capture urban vitality from different angles[6].

However, these data sources have their own issues, including high acquisition costs, strict privacy requirements, and a lack of temporal and spatial resolution. For instance, census data has a long cycle, while mobile signaling data is limited by specific carriers. These issues constrain the dynamic and fine-grained evaluation of urban vitality[7].

In contrast, the use of POI (Point of Interest) data has emerged as an effective alternative. POI data, which covers a city's various functional facilities and activity locations, can reflect its functional diversity and dynamism[8]. Research indicates that "POI data represents urban functional facilities, reflecting the diversity and density of urban vitality, making it a crucial indicator for its assessment." Furthermore, POI data offers advantages like open access, high spatial accuracy, and rapid updates, effectively compensating for the shortcomings of traditional data[9].

Empirical studies have shown that POI density significantly influences community or urban vitality. Research has also found that POI mixedness has a significant threshold effect, with urban vitality increasing notably once a certain level of mixedness is achieved[9].

In summary, while traditional population and mobile data face issues with accessibility and dynamic assessment, POI data, due to its advantages in timeliness and diversity, has become a vital alternative. By reflecting urban functional density and diversity, it effectively reveals the key factors influencing urban vitality, providing a reliable theoretical basis and practical guidance for related research[10].

The methodology of this paper is primarily based on spatial big data analysis and spatial econometric models.First, Sina Weibo check-in data is used to construct a proxy indicator for urban vitality, which characterizes the dynamic spatio-temporal activity patterns of residents. Second, urban land use structure and street network characteristics are extracted from multi-source data to create explanatory variables from the two dimensions of function and form. Finally, a Multi-scale Geographically Weighted Regression (MGWR) model is employed to estimate the relationship between different spatial elements and urban vitality. This approach can reveal multi-scale differences and spatial heterogeneity in the strength of variable effects across geographic space, thus providing a more precise explanation of how urban spatial form influences the distribution mechanism of urban vitality.

2. Material and Methods

2.1. Study Area

The case study area for this research is Xi'an, the capital of Shaanxi Province. Located in the central Guanzhong Plain in Northwest China, Xi'an is a core node city for the "Belt and Road" Initiative, a major national scientific research and high-tech industrial base, and the largest rail, air, and road hub in the Northwest region.

ISSN: 2710-0170

As of the end of 2022, the city governs 11 districts and 2 counties, with a total administrative area of 10,096.81 km^2 and a concentrated built-up area of approximately 1,295 km^2. Its permanent population reached 12.996 million.

As a megacity with a history of over 3,100 years, Xi'an has experienced rapid expansion and renewal in recent years under its designation as a "National Central City." Over the past decade, its permanent population has grown by 52.8%, and its subway system has expanded to over 300 km. The coupling relationship between the city's socioeconomic advancement and its spatial restructuring—particularly the optimization of land use allocation and the redistribution of public services and industrial resources—has become a key focus for academics, policymakers, and local planning departments.

To meticulously characterize this coupling relationship, this study uses a 1 km × 1 km grid as the statistical analysis unit[11]. This ensures a high-resolution evaluation of urban vitality spatial heterogeneity while guaranteeing that each grid cell contains at least one sampling point. As shown in Figure 1, after excluding the Qinling Ecological Protection Zone, the Weihe River wetland buffer, and areas without signal coverage, the study area was divided into 936 valid grid cells. These cells cover the main urban area of Xi'an and maintain spatial continuity. The data was sourced from Weibo check-in data.

2.2. Data Collection and Processing

Our study's dataset integrates multiple heterogeneous sources, which not only broadens the scope of our research but also fundamentally ensures the comprehensiveness and reliability of our findings. We utilized two key data types: first, 2024 Gaode Maps POI (Point of Interest) data, which, with its high coverage and standardized classification system, accurately depicts the static spatial distribution of urban functional facilities such as commerce, living services, and public infrastructure. Second, we used Sina Weibo check-in data from October 2023 to October 2024. This dynamic dataset captures residents' actual spatial locations at different times, effectively reflecting their spatio-temporal behavior patterns and the dynamic shifts of urban hotspots. The complementary nature of these two datasets allows our study to be analyzed from the dual perspectives of "what the city provides" and "how residents use it," jointly forming a multi-dimensional data foundation.

To ensure the accuracy of our analysis, we performed a rigorous three-stage data preprocessing on the Weibo check-in data. First, we implemented deduplication based on unique identifiers, using the distinct ID of each Weibo record for precise matching and deletion, which efficiently eliminated entries that were completely duplicated due to multiple crawls or system errors. Second, we addressed content redundancy through filtering and deduplication based on content similarity. We employed locality-sensitive hashing techniques, such as MinHash, to convert each Weibo's text content into a digital fingerprint. We then set a strict similarity threshold to identify and remove highly similar records, effectively reducing noise from user reposts or copy-pasted content. Finally, we conducted a crucial comprehensive filtering process based on time, user, and content. This involved removing records with abnormal timestamps and excluding low-quality data from potential marketing or bot accounts by analyzing features like follower-to-following ratios and posting frequency. We also filtered out Weibo posts that were too short or lacked valid geographic information to ensure that every retained record held genuine analytical value. Through this series of meticulous data processing steps, we ultimately obtained a high-quality, reliable dataset that provides a solid foundation for subsequent urban spatial analysis and behavioral pattern research.

2.3. Measurement of Urban Vitality

This study evaluates urban vitality based on an intensity dimension, which refers to the spatial concentration of human activities and reflects the vitality level of different areas. The specific

ISSN: 2710-0170

method involves aggregating Weibo check-in data from October 2023 to October 2024 on an hourly (24-hour) basis for each day. This approach refines the sample at a temporal scale, enhancing its representativeness. The methodology is supported by the large user base of the Sina platform and previous research experience in Xi'an.

2.4. Factors Influencing Urban Vitality

Based on a synthesis of existing research, the characteristics of Xi'an's central urban area, and data availability, we constructed indicators for urban form diversity and destination accessibility using POI (Point of Interest) data, initially selecting four built environment indicators. Subsequently, a multicollinearity test was performed, and any indicators with a Variance Inflation Factor (VIF) greater than 7.5 were removed to ensure the robustness and scientific validity of the indicator system.

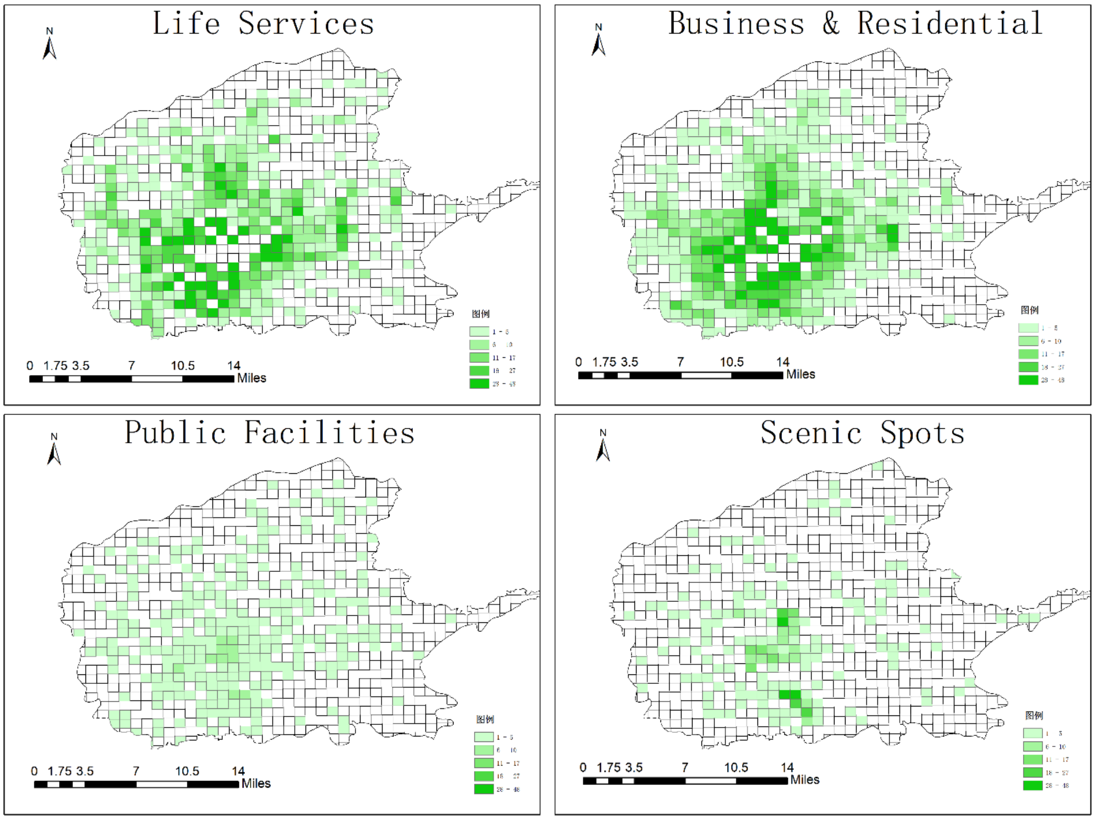

Figure 1Distribution of POIs in Xi'an.

Urban living services in this region are highly concentrated within and between the Second and Third Ring Roads, forming a "chessboard-plus-radial" pattern with density decreasing from the city center outwards. Beyond the Third Ring Road, medium density is only maintained along subway and major arterial lines, with other areas being notably sparse. The commercial and residential landscape is "monocentric with multiple poles," where the old city center (Zhonglou-Xiaozhai) serves as the absolute core with the highest density. Three secondary hotspots have emerged in the periphery: the northern administrative center-high-speed rail station, the eastern Chanba World Expo area, and the southern High-tech Zone-Qujiang area, with these hotspots extending in a "beaded" pattern along subway lines. Public facilities exhibit a "ring-plus-axis" distribution, with the highest density forming a distinct ring between the First and Second Ring Roads. Outside the Second Ring Road, facilities form radial axes along subway lines 2, 3, 4, and 5, with density diminishing with distance. In remote suburban areas, isolated high-value points only appear at district and county government seats. Finally, tourist attractions are "south-heavy and north-light," with a continuous high-density area in the south (including the Big Wild Goose Pagoda, Qujiang, Datang Everbright City, and the northern

ISSN: 2710-0170

foothills of the Qinling Mountains). Medium-density scattered points are found in the east (Chanba World Expo), west (Han Chang'an City Ruins), and north (Daming Palace), while other areas have a sparse distribution of tourist POIs.

2.5. Variable Standardization and Model Setting

This study systematically examines the linear impact of multi-source environmental factors on urban vitality by integrating both OLS and MGWR models. To eliminate dimensional differences and ensure comparability, all explanatory variables are standardized using min-max normalization, mapping them to the [0, 1] range. The normalization formula is:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

Subsequently, a multicollinearity diagnostic is performed on all variables using the Variance Inflation Factor (VIF) to ensure the robustness of the estimations.

In the model implementation, an OLS model is first established as a global baseline to estimate the average marginal effects of each variable on urban vitality. Significant variables are then incorporated into the subsequent MGWR framework. MGWR extends GWR by relaxing the "single spatial bandwidth" assumption, adaptively selecting an optimal bandwidth for each explanatory variable to characterize its heterogeneous effect at different spatial scales. Unlike the global estimations of OLS, MGWR re-estimates parameters at each observation location (u,v). The model's expression is simplified as:

$$y(u,v) = \beta_0(u,v) + \sum_k \beta_k(u,v)x_k + \varepsilon(u,v) \tag{2}$$

This approach allows the study to reveal local differences and spatial non-stationarity in how various factors influence urban vitality within the city.

3. Results

3.1. Multiple Linear Regression

We used stepwise multiple linear regression to evaluate the relationship between urban vitality and its related factors. The dependent variable was the Vitality Index (VI), and the independent variables included built environment elements as detailed in Section 2.4.The regression analysis results satisfied the tests for normality of residuals and homoscedasticity. The Variance Inflation Factor (VIF) values for all independent variables were below 10, indicating no severe multicollinearity issues among the variables. The regression parameters are detailed in a table.

Table 1OLS Model Results

Variable	Coefficient[a]	Standard Error	t-Statistic	VIF[c]
Intercept	-1.141583	0.136874	-8.340368	—
Scenic Spots	-0.047365	0.014949	-3.168459	1.610028
Public Facilities	0.040144	0.039021	1.028778	2.069024
Business & Residential	-0.027031	0.0068	-3.975174	5.80071
Life Services	0.005233	0.005031	1.040133	4.925128

3.2. Moran's I

At a 0.001 confidence level, with a z-score greater than 4, all variables were found to be statistically significant. The Moran's I results show that all variables have significant positive spatial autocorrelation (P = 0.000), indicating a clear clustered distribution in urban space, where both high-value and low-value areas are clustered together.Among these, Business &

Residential has the highest Moran's I (0.819), followed by Life Services (0.719). The indices for Science, Education & Culture, Urban Vitality, Public Facilities, and Scenic Spots decrease in order (0.422–0.577), but all variables maintain a medium to high degree of clustering.

Table 2The Moran's I index results for Xi'an city.

Variable Name	Moran’s I	Z-score	P-Value
Urban Vitality (VI)	0.544923	28.836642	0.000000
Scenic Spots	0.421941	22.58269	0.000000
Public Facilities	0.449444	23.21355	0.000000
Business & Residential	0.81878	41.89312	0.000000
Life Services	0.718936	36.8861	0.000000

3.3. Comparison of MGWR and OLS Results

Compared to OLS, the goodness of fit of MGWR is significantly improved: R^2 increases from 0.487 to 0.741, and Adjusted R^2 rises from 0.478 to 0.706. Meanwhile, AICc decreases from 1628.404 to 1293.035, and the residual sum of squares (RSS) drops from 377.735 to 190.729. This indicates that MGWR is clearly superior to OLS in explaining the spatial differences in urban vitality.

Table 3Comparison of MGWR and OLS Results

Metric	OLS	MGWR
R^2	0.487	0.741
Adjusted R^2	0.478	0.706
AICc	1628.404	1293.035
RSS	377.735	190.729

4. Conclusion

From the results of the MGWR model, it is clear that different types of urban functional elements have varying impacts on urban vitality. Overall, the model's intercept is negative and significant, indicating a low baseline level of urban vitality after controlling for various functional factors. Among the explanatory variables, Scenic Spots, Business & Residential, and Science, Education & Culture facilities all show a significant negative effect on urban vitality. This implies that while scenic spots attract people, they primarily represent tourism and leisure functions; their spatial clustering does not drive daily urban vitality and may even compete with residential or commercial activities. The clustering of Business & Residential land also reduces urban vitality, likely due to its singular function and lack of diverse services, which limits interaction and activity among a variety of people. Meanwhile, the spatial layout of Science, Education & Culture facilities is relatively closed, with a functional orientation often linked to teaching and research, leading to a spatial dislocation from vitality hotspots like commercial and leisure areas. In contrast, while Public Facilities and Life Services show a positive relationship in their coefficients, they are not statistically significant, suggesting a limited or only locally effective contribution to urban vitality.

The spatial autocorrelation analysis further corroborates these differences. The Moran's I results show that urban vitality exhibits significant positive spatial autocorrelation within the study area, with both high and low vitality values displaying a clear spatial clustering state. At the same time, various functional elements also demonstrate clustering characteristics, with Business & Residential and Life Services showing the highest degree of clustering. This indicates that these two types of functions are often co-located, whereas Scenic Spots, Public

ISSN: 2710-0170

Facilities, and Science, Education & Culture exhibit a medium level of clustering. This pattern suggests that the relationship between urban vitality and functional elements is not random but highly dependent on spatial aggregation effects.

Further comparison of the OLS and MGWR model results clearly shows that MGWR significantly outperforms OLS in both goodness of fit and explanatory power. As a global regression model, OLS can only provide an average relationship, ignoring spatial differences, with an R^2 of only 0.487. By incorporating spatial heterogeneity, MGWR's R^2 increases to 0.741, and its Adjusted R^2 also rises to 0.706, far exceeding the OLS level. Simultaneously, MGWR's AICc value significantly decreases to 1293.035 from 1628.404, and its residual sum of squares (RSS) drops by nearly half from 377.735 to 190.729. This indicates that MGWR not only improves the model's explanatory power but also more effectively captures the spatial interaction mechanisms among variables. Therefore, when explaining the formation mechanism of urban vitality, considering spatial differences is crucial, and MGWR is able to more comprehensively reveal the differentiated effects of various functional elements at different spatial scales.

References

[1] Row A T, Jacobs J. The death and life of great american cities[J]. Yale Law Journal, 1962, 71(8): 1597.

[2] Cai Z, Li D, Ji B, et al. Exploring the influence mechanisms and spatial heterogeneity of urban vitality recovery in the university fringe areas of nanjing[J]. Sustainability, 2024, 17(1): 223.

[3] Built environment, urban vitality and social cohesion: do vibrant neighborhoods foster strong communities?[J]. Landscape and Urban Planning, 2020, 204: 103951.

[4] Gómez-Varo I, Delclòs-Alió X, Miralles-Guasch C. Jane jacobs reloaded: a contemporary operationalization of urban vitality in a district in barcelona[J]. Cities, 2022, 123: 103565.

[5] Chen Y, Yu B, Shu B, et al. Exploring the spatiotemporal patterns and correlates of urban vitality: temporal and spatial heterogeneity[J]. Sustainable Cities and Society, 2023, 91: 104440.

[6] Jiang Y, Zhang Y, Liu Y, et al. A review of urban vitality research in the chinese world[J]. Transactions in Urban Data, Science, and Technology, 2023, 2(2-3): 81-99.

[7] Liu K, Yin L, Lu F, et al. Visualizing and exploring POI configurations of urban regions on POI-type semantic space[J]. Cities, 2020, 99: 102610.

[8] Ouyang J, Wang L, Zhu D, et al. Revealing urban vibrancy stability based on human activity time-series[J]. Sustainable Cities and Society, 2022, 85: 104053.

[9] Yue Y, Zhuang Y, Yeh A G O, et al. Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy[J]. International Journal of Geographical Information Science, 2017, 31(4): 658-675.

[10] Taecharungroj V, Ntounis N. Categorising neighbourhoods using OpenStreetMap POIs: affinity propagation clustering of 7,213 subdistricts in Thailand[J]. Journal of Urban Management, 2024.

[11] Chen Z, Huang B. Achieving urban vibrancy through effective city planning: a spatial and temporal perspective[J]. Cities, 2024, 152: 105230.