Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Rebalancing Rural Service Supply through Generative Design and Graph-Based Evaluation: Enhancing Accessibility and Diversity for Marginalized Area

Xinlong Li
Universiti Teknologi Malaysia
Zhiwen Cui
Universiti Teknologi Malaysia
Wenwen Shi
20140512@zufedfc.edu.cn

Zhejiang University of Finance and Economics Dongfang College
Qinghua Jia
Universiti Teknologi Malaysia

Article

Keywords: Rural commercial service facilities, Spatial layout, Service accessibility, Service diversity, Spatial relational network analysis

Posted Date: March 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8913227/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The uneven distribution of rural commercial service facilities has led to significant disparities in accessibility and diversity among villages, limiting rural residents' equitable access to essential services and weakening the efficiency of facility allocation. Addressing this challenge, the study aims to establish a generative design–driven evaluation and optimization framework that enhances both equity and efficiency in rural spatial planning. Using Qingshan Town as a case study, the framework integrates k-nearest neighbor graph construction with graph convolutional networks to calculate accessibility and diversity scores, identify facility gaps, and optimize new facility locations. Results show a pronounced improvement, with spatial inequality in both accessibility and diversity being substantially reduced, reflecting notable progress toward greater spatial equity. Efficiency analysis further indicates that population-weighted performance levels increased markedly for both accessibility and diversity, suggesting that facility allocations are not only fairer but also better aligned with actual demand. By demonstrating how generative optimization can rebalance service supply in marginalized rural villages such as Maoping and Baoershan, the study highlights tangible benefits for local residents, particularly those in underserved communities. Beyond its empirical findings, the framework contributes a methodological innovation that integrates generative optimization with spatial equity analysis, offering policymakers and planners a practical tool to support rural.

1. Background

Rural commercial service facilities play a vital role in sustaining residents' daily lives, supporting local economies, and promoting social equity between urban and rural areas. Yet current practice continues to face multiple challenges. First, accessibility varies significantly across villages, with peripheral settlements often experiencing pronounced service gaps. Second, facility diversity remains limited, as provisions for shopping, dining, and everyday services are inadequate in many villages, making it difficult to meet residents' increasingly diverse needs. Third, existing approaches to facility siting and optimization often rely on static standards or expert judgment, lacking data-driven and highly reproducible computational methods. With the rise of data-driven planning and generative design methods, researchers now have the opportunity to integrate accessibility evaluation, diversity analysis, and intelligent optimization, opening new possibilities for more equitable and resilient spatial development. In particular, graph convolutional networks (GCNs) based on knowledge graphs hold promise as a unified and efficient tool for siting commercial service facilities across both urban and rural contexts. To more clearly define the research gaps, a systematic review of the relevant literature is therefore required.

In the study of facility spatial configuration, accessibility and diversity have consistently served as key indicators for evaluating the equity and efficiency of service provision. In existing research on rural commercial service facilities, accessibility is primarily measured using the Two-Step Floating Catchment Area Method (2SFCA) or the Three-Step Floating Catchment Area Method (3SFCA) (Sun et al., 2024; Vaishar & Šťastná, 2021). These methods assess residents' access to facilities through spatial

impedance functions and have been widely applied in urban–rural public service research (Lee et al., 2021; Li et al., 2024; Y. Zhang et al., 2020). Nevertheless, such approaches present two main limitations: first, they emphasize the measurement of “whether access is possible” while overlooking the diversity and balance of service types; second, their optimization strategies often rely solely on static spatial relationships, lacking a systematic framework that integrates multiple indicators. Diversity is typically assessed using the Shannon entropy index to measure the richness and balance of facility types(Z. Gu et al., 2022; Nakase et al., 2021; Xun et al., 2024). This index has been extensively applied in spatial analyses of urban–rural public services and commercial facilities, providing an important tool for uncovering inequalities in facility distribution. Despite their rigorous mathematical foundations, these models are constrained by static assumptions and single-objective optimization. As a result, traditional approaches often fail to fully reflect rural residents’ actual experiences of accessing diverse services and are ill-suited to the complex and dynamic nature of urban–rural environments.

Graph neural networks (GNNs) have gradually become an important tool in facility planning. GNNs are capable of modeling the complex relationships among population, facilities, and road networks within graph structures. Through message passing and neighborhood aggregation, they enable data-driven and integrated optimization (Wu et al., 2020), offering greater adaptability and reproducibility. Among the various GNN models, graph convolutional networks (GCNs) represent a typical branch. By leveraging adjacency matrices and degree normalization mechanisms to aggregate node features(Jiang et al., 2019; S. Zhang et al., 2019), GCNs model graph-structured data and capture neighborhood features, thereby supporting multi-dimensional and multi-objective evaluation and optimization of facility siting. This provides a new technical pathway for urban and rural spatial planning. For instance, Zhao proposed a framework combining knowledge graphs and GNNs (KG-GNN) to optimize the siting of online-to-offline convenience stores in Nanjing, showing that it outperforms traditional methods in capturing multi-source spatial and semantic features (Zhao et al., 2025). Gu integrated GCNs with simulated annealing to optimize the layout of new-energy bus charging stations(T. Gu et al., 2025). Similarly, Xu applied GCNs in healthcare facility design, enhance resource allocation efficiency(Xu, 2025a).

To address the limitations of traditional models, this study proposes an intelligent generative optimization framework that integrates K-nearest neighbor (KNN) graph construction with graph convolutional networks (GCN). By incorporating accessibility and diversity as key indicators, the framework captures the interrelationships among population, facilities, and road networks, thereby enabling a more comprehensive optimization of rural commercial service layouts. This approach not only offers planners and policymakers a practical, evidence-based decision-making tool, but also ensures that residents in underserved peripheral villages gain more equitable and efficient access to essential services.

2. Research materials

The research materials comprise three parts: multi-dimensional data on population, road networks, and POIs; a study area chosen for its representative urban–rural spatial differences; and a research

framework centered on equity and efficiency, covering data preparation, indicator measurement, optimization modeling, and result validation.

2.1 Data Sources

The spatial data required for this study primarily include administrative boundaries, village locations, river distribution, road networks, points of interest (POIs) of commercial service facilities, population data, and exclusion areas. The exclusion areas, administrative boundaries, and village locations were obtained from the Resource and Environment Science Data Platform (RESDC), an official open-source data platform of the Chinese government, which ensures a high level of authority and reliability(*China's administrative village boundary data*, n.d.; *Data on administrative boundaries of Chinese towns and villages in 2015*, n.d.; *Remote sensing monitoring data on China's land use status in 2025*, n.d.). Road network data were collected from *Tianditu*, the official national road surveying platform (*Map World*, n.d.). POI data were obtained from the Amap Open Platform(*Search POI*, n.d.). Population data were sourced from the official website of Jinzhai County Government (*Introduction to Qingshan Town, Jinzhai County*, n.d.), provided at the village level and containing basic demographic information such as population counts.

2.2. Research Area

Qingshan Town is located in the southeastern part of Jinzhai County, Lu'an City, Anhui Province. The town administers six villages and covers an area of 164.4 km², with a total population of 22,164 residents(*Introduction to Qingshan Town, Jinzhai County*, n.d.). The six villages under its jurisdiction are Qingshan, Yaotang, Tangdian, Maoping, Bao'ershan, and Jianghe, with the township government located in Qingshan Village (*Introduction to Qingshan Town, Jinzhai County*, n.d.), As shown in Fig. 1.

Commercial service facilities in the study area are categorized into three types: shopping services, dining services, and life services. This classification is based on relevant literature and the daily needs of local residents, with the specific types of stores included in each category presented in Table 1. According to the POI data, the distribution of commercial facilities in Qingshan Town is highly concentrated in Qingshan Street, which hosts a total of 138 facilities, including 78 shopping services, 40 dining services, and 20 life services—accounting for the vast majority of facilities in the town (see Table 2). By contrast, other villages are significantly underserved. For example, Yaotang, Tangdian, and Maoping each have only two facilities across all categories.

Table 1
Classification of facilities

Service Type	Store Type
Shopping Service	Supermarkets, Convenience stores, Grocery shops, Pharmacies, Daily necessities stores, Clothing stores, Shoe stores, Vegetable shops, and Fruit shops.
Catering Service	Chinese restaurants, Snack bars, Fast food restaurants, Milk tea shops, Burger restaurants, Noodle restaurants.
Daily Life Service	Barbershops, Beauty and nail salons, Express delivery points, Printing and photocopy shops, Motorcycle repair shops.

Table 2
Number of village facilities

Village name	Catering Service	Shopping Service	Daily Life Service	Total
Qingshan	40	78	20	138
Yaotang	0	2	0	2
Tangdian	1	1	0	2
Maoping	0	1	1	2
Baoershan	1	4	1	6
Jianghe	0	9	1	10

2.3 Research Framework

The study constructs an integrated framework consisting of six components, as illustrated in Fig. 2: data collection and processing, accessibility calculation, diversity calculation, gap identification (GCN), optimization generation, and performance evaluation. First, administrative boundaries, population data, road networks, commercial service facility POIs, and exclusion areas were obtained through government open platforms and map service APIs. Next, accessibility for each village was measured using the Two-Step Floating Catchment Area Method (2SFCA), while the diversity of facility types was evaluated using Shannon entropy. Together, these measures reveal disparities in service access and structural balance across different areas.

On this basis, a graph convolutional network (GCN) was introduced to identify facility gaps. This model is capable of simultaneously capturing the interrelationships among population distribution, facility layout, and road network structure, thereby overcoming the limitations of traditional methods that rely on static assumptions and single-objective evaluation. Furthermore, the study integrates accessibility, diversity, and GCN-based gap indicators into a multidimensional framework to generate optimized facility layouts. Under the constraint of exclusion areas, this process identifies new siting options for commercial facilities. Finally, performance evaluation was conducted to verify the effectiveness of the optimized results in improving both accessibility and diversity.

3. Method

This study adopts a comprehensive methodological framework to assess and optimize rural commercial service facilities. Accessibility was measured using the Two-Step Floating Catchment Area Method (2SFCA), and diversity with the Shannon index. A KNN-based graph and graph convolutional network (GCN) were then applied to identify service gaps and generate optimized facility layouts. The recalculated accessibility and diversity indicators enabled comparison before and after optimization, while the Gini coefficient and population-weighted means evaluated outcomes from equity and efficiency perspectives.

3.1 Accessibility of Commercial

Accessibility is a key indicator for assessing the convenience of residents in obtaining commercial service facilities, reflecting both the spatial equity and efficiency of facility distribution(Pirie, 1979). Existing studies commonly adopt the Two-Step Floating Catchment Area (2SFCA) method to evaluate spatial accessibility of facilities (M. Gu et al., 2025; Pérez-Fernández & Michel, 2025; Y. Zhang et al., 2024). This approach comprehensively accounts for the spatial matching relationship between supply and demand as well as travel impedance, making it suitable for the evaluation needs of this study.

At the operational level, this study designates village committees as demand points, as they are typically located near the center of settlements and can reasonably represent the overall service needs of each village, thereby serving as a reliable proxy for the overall service demand of each village. Points of interest (POIs) of commercial service facilities were set as supply points and categorized into three groups: shopping, dining, and life services. The accessibility of facilities was calculated using the following formula(Liu et al., 2023).

Step 1: Calculation of the Supply–Demand Ratio

Taking facility j as the center, all demand points k within its service range are identified. The supply–demand ratio of facility j is expressed as:

$$R_j^{(t)} = \frac{S_j^{(t)}}{\sum k \in \left\{ d_{ij} \leq d_0^{(t)} \right\} D_k}$$

Here, t denotes the facility service type (shopping, catering, or daily life services). $S_j^{(t)}$ represents the service capacity of facility j of type t (set to 1 in this study). dij refers to the network distance between demand point iii and facility j; the smaller the value of dij, the more easily residents can access commercial services within a reasonable travel range. $d_0^{(t)}$ is the service threshold radius for type t facilities (10 km for catering, 5 km for shopping, and 7 km for daily life services). $\sum D_k$ denotes the

total population of all demand points within the service radius. This step reflects the service capacity of an individual facility relative to the population it covers.

Step 2: Calculation of Demand-Point Accessibility

Taking demand point i (i.e., the village committee) as the center, all facilities j within the threshold distance $d_0^{(t)}$ are identified. The supply–demand ratios Rj of these facilities are then summed to obtain the accessibility of village (demand point) iii to facility type t, denoted as Ai(t). The formula is expressed as:

$$A_i^{(t)} = \sum_{j \in \left\{ d_{ij} \leq \ d_0^{(t)} \right\}} R_j^{(t)}$$

Step 3: Calculation of Comprehensive Village Accessibility

In the third step, different accessibility weights were assigned to each type of service facility according to both the frequency of residents' use and the degree of influence on their daily lives. Based on these weights, the final comprehensive accessibility score for each village was calculated using the following formula:

$$A_i^{total} = \sum_{t=1}^{3} w_t \bullet \ A_i^{(t)}$$

Here, A_i^{total} represents the comprehensive accessibility score of village i. $A_i^{(t)}$ denotes the accessibility score of villages i for service category t; and w_t is the weight assigned to service category t, reflecting its relative importance in residents' daily lives. In this study, the weights are set as follows: catering services = 25%, shopping services = 40%, and daily life services = 35%.

3.2 Diversity of Commercial

Diversity is an important indicator for assessing whether residents can access a wide and balanced range of service categories within a given spatial scope(Shannon, 1948). It reflects the spatial distribution of different facility types: if a village contains only a single type of service facility, its diversity is considered low; conversely, if multiple facility types coexist and are relatively evenly distributed, diversity is higher(Lin et al., 2008).

In this study, villages are used as the basic analytical units. Three categories of commercial service facilities—shopping, catering, and life services—were extracted from POI data. The Shannon Diversity Index was then applied to calculate the diversity of commercial service facilities for each village (Gómez-Varo et al., 2024). The Shannon entropy formula is given as follows:

$$H = -\sum_{i=1}^{n} p_i \ln\left(p_i\right)$$

Here, n denotes the total number of facility categories (three in this study), and pi represents the proportion of facility type i within a given village. The implication of the formula is that the greater the

number of facility categories within a village, and the more balanced their proportions, the higher the value of H, indicating greater diversity. Conversely, if one type of facility is either overly concentrated or absent, the value of H decreases, reflecting lower diversity. This method enables a quantitative comparison of the commercial service facility configurations across villages and provides a basis for subsequent facility gap identification and optimization.

3.3 Facility Gap Identification

After completing the measurements of accessibility and diversity, this study further adopts a gap identification method based on graph convolutional networks (GCNs) to systematically evaluate deficiencies in the configuration of commercial service facilities across villages. The core idea is to transform villages, facilities, and their spatial relationships into a graph structure, within which multiple attributes—including population, road networks, exclusion areas, accessibility, and diversity—are integrated. Through model training, the GCN identifies facility gaps at both the overall and category-specific levels, thereby providing quantitative evidence for subsequent optimization.

3.3.1 KNN Graph Construction

Taking villages as nodes, the node feature matrix X incorporates attributes such as population size, existing facility categories, construction constraints, accessibility scores, and diversity scores, thereby reflecting service disparities among villages within a spatial context. K-nearest neighbor (KNN) graph is then constructed based on geographic coordinates to capture the spatial proximity relationships between villages(Cover & Hart, 1967; Regression, 1992).

$$d_{ij} = \sqrt{\left(X_i + X_j\right)^2 + \left(Y_i + Y_j\right)^2}$$

$$A_{ij} = \begin{cases} 1, if\ j \in\ KNN\left(i\right) \\ 0,\ \ otherwise \end{cases}$$

The Euclidean distance between village iii and village j is denoted as d_{ij}, where Xi and Yi represent the geographic coordinates of village i. The adjacency matrix element Aij indicates whether there exists an edge between villages iii and j. Specifically, KNN(i) denotes the set of the k-nearest neighbors of village i.

3.3.2 GCN Model and Gap Score

On the constructed graph structure, this study applies a two-layer graph convolutional network (GCN) to predict the gap intensity of each village. The core formulation of the GCN is as follows(Kipf, 2016; Wu et al., 2020).

$$H^{(l+1)} = \sigma\left(\bar{D}^{-\frac{1}{2}}\ \bar{A}\ \bar{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right)$$

$$Need_i \in\ [0,1]$$

Here, $H^{(l)}$ represents the node feature matrix at layer l. The input layer is defined as $H^{(0)} = X$, which is the input feature matrix consisting of attributes such as population size, existing facility categories, construction constraints, accessibility scores, and diversity scores. $W^{(l)}$ denotes the learnable weight matrix at layer l, and σ(·) is the activation function. $\bar{A} = A + I$ represents the adjacency matrix with self-loops added, while $\bar{D}$ is the corresponding degree matrix of $\bar{A}$.The final output of the model is the overall gap score for each village, denoted as $Need_i$. The score ranges from 0 to 1, where higher values indicate greater deficiencies in the provision of commercial service facilities.

The supervisory signal is defined as the ratio of the number of missing facility categories to the total number of categories. In this way, the more facility types a village lacks, the higher its label value. After training, the GCN is able to predict the overall deficiency in service provision for each village by integrating both its own characteristics and those of neighboring villages under the constraints of spatial adjacency.

3.3.3 Category-Specific Gap Decomposition

After obtaining the overall gap score, the study further decomposes the deficiencies by facility category. First, within the neighborhood of each village, the numbers of catering, shopping, and daily life service facilities are counted, and the Shannon entropy index is calculated to measure the level of diversity. Next, villages with high diversity are selected as the reference set, from which the target proportion $p*$ of facility categories is derived. Based on this benchmark, each village is evaluated by comparison, and a gap vector is computed to indicate the specific types of services that are underrepresented(Talen & Anselin, 1998).

$$Gap_{i,c} = \max\left(0, round\left(T_i \bullet p_*^{(c)} - Count_{i,c}\right)\right)$$

The category-specific gap for village i in service type ccc (catering, shopping, or daily life services) is denoted as $Gap_{i,c}$. Here, Ti represents the target total number of facilities in the neighborhood of village i. $p_*^{(c)}$ is the target proportion of category c, derived from the high-diversity reference set. $Count_{i,c}$ denotes the existing number of facilities of category c within the neighborhood of village i. The operator $\mathrm{Max}(\bullet)$ ensures that the gap value is non-negative, while $round(\bullet)$ rounds the result to the nearest integer, thereby guaranteeing integer gap values.

For example, if the gap vector is [2,1,0], this indicates that the village has significant shortages in catering and shopping facilities, while daily life services are relatively sufficient. Once the hard deficits are satisfied but additional capacity remains, the model supplements facilities by prioritizing the categories with the largest deviation between the current proportions and the target proportions, thereby gradually improving overall diversity.

3.4 Optimization Strategy

After completing the measurements of accessibility and diversity as well as the identification of facility gaps, this study further proposes a multi-indicator optimization approach designed to generate rational candidate facility layouts. The method takes the GCN-derived gap score as its core input, while simultaneously incorporating diversity constraints, population weighting, and spatial neighborhood constraints. In this way, the newly added facilities are ensured not only to enhance overall service coverage but also to improve category structure and spatial balance(Fotheringham et al., 2000; T. Gu et al., 2025; Xu, 2025b).The optimization strategy can be divided into three levels: slot allocation, type and location selection, and secondary population-weighted optimization.

3.4.1 Slot Allocation

First, to determine the number of additional facilities that can be allocated to each village, this study integrates the GCN-predicted gap score ($need_{pred}$) with the degree of diversity deficiency (1 − H) through a power-based fusion approach.

$$Score_u \propto \left(need_{pred}{}^{\propto} + (1-H)^{\propto} \right)$$

Here, $Score_u$ represents the comprehensive gap intensity of village u, $\propto$ is the adjustment coefficient, and H denotes the Shannon diversity index. Based on this comprehensive score, and in combination with the base slot parameter (SLOT_BASE) and the proportional scaling factor (SLOT_SCALE), an initial number of slots is allocated to each village.

$$slot_{count(u)} = SLOT_{BASE} + SLOT_{SCALE} \times score_u$$

To ensure equity, each village is assigned at least TARGET_S_MIN slots. The generated slot locations are distributed as random scatter points around the village center, with a minimum distance constraint applied to prevent excessive clustering of facilities.

3.4.2 Facility Type and Location Selection

After determining the number of slots, the study further specifies both the types and precise locations of the new facilities. Type selection is driven by diversity considerations. In this process, priority is given to meeting the hard gap vector target add[i] identified during the gap analysis stage, ensuring that the most severely underrepresented categories are addressed first. Once these hard deficits are satisfied, the allocation proceeds by reducing proportional deviations (gap = p∗−q), so that the facility category structure gradually approaches the target distribution p∗ (Luo & Wang, 2003).

Location selection, by contrast, is guided primarily by accessibility. Within each allocated slot, accessibility is taken as the foremost criterion and distance as a secondary factor when ranking placement priorities. For a facility of type t, a service radius rt is defined (10 km for catering, 5 km for shopping, and 7 km for daily life services), and the covered population is then calculated. The population covered by facility j of type t is expressed as(Wang, 2012).

$$Acc_{i,t} = \sum_{j \in N(i,r_t)} Pop_j$$

Here, $Acc_{i,t}$ represents the population coverage of facility type t at location i; Pop_j denotes the population of village j within the service radius; and $N(i, r_t)$ refers to the set of neighboring villages located within the service radius r_t. A higher value of this indicator implies stronger service coverage, and thus a higher priority for the corresponding location to be selected.

3.4.3 Secondary Weighted Optimization

After the preliminary allocation is completed, this study introduces a secondary weighting mechanism based on population and gap scores to refine the distribution results. The weight of newly added facilities is defined by the following formula.

$$w_i = Pop_i \times need_{pred_i}$$

Here, Pop_i represents the population size of village i, and $need_{pred_i}$ denotes the predicted gap score of that village. Based on the weight w_i, an additional proportion of facilities—accounting for approximately POP_EXTRA_PCT of the total new allocation (set to 25% by default)—is redistributed, so that areas with both high population concentration and severe service deficits receive further supplementation.

The proposed optimization approach emphasizes the integration of multiple indicators with spatial constraints in its overall design. On the one hand, it comprehensively incorporates accessibility, diversity, and gap scores; on the other hand, it embeds exclusion-area constraints, neighborhood-radius constraints, and diversity safeguards into the layout generation process. This method not only ensures the rationality of the number and spatial distribution of newly added facilities but also prevents excessive clustering and category imbalance. As a result, it forms a reproducible and generalizable data-driven optimization framework that advances the application of AI in rural commercial service facility planning.

3.5 Method Validation

To ensure the scientific soundness and effectiveness of the proposed optimization framework, validation is conducted from two perspectives: (1) comparison with the current situation and traditional methods, and (2) evaluation using quantitative indicators. The first perspective examines the degree of improvement achieved by the optimization results relative to the existing configuration and conventional models, while the second quantitatively assesses overall performance from the dimensions of equity and efficiency.

In terms of comparison with the current situation and traditional methods, the study adopts the two-step floating catchment area method (2SFCA) and the Shannon diversity index as benchmark measures to evaluate results before and after optimization. Accessibility comparison reveals the extent of service improvements for both the overall study area and peripheral villages, whereas diversity comparison

examines the effectiveness of newly added facilities in enhancing the balance and richness of service types. This dual comparison not only demonstrates whether optimization improves service availability but also reveals whether the structural configuration of facilities has been improved.

With respect to quantitative evaluation, the study incorporates both accessibility and diversity indicators under the dual dimensions of equity and efficiency. For equity, the Gini coefficient is employed to measure distributional disparities in accessibility and diversity among villages(Karam & Ryerson, 2023; Wang, 2012). The formula is expressed as follows:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n^2\,\bar{\mu}}$$

Here, xi represents the accessibility score or diversity index of village i, n is the total number of villages, and $\bar{\mu}$ denotes the mean value of the indicator. A smaller value of G indicates a more balanced distribution of the indicator across villages, thereby reflecting the spatial equity of the optimization scheme.

In terms of efficiency, this study calculates the population-weighted mean to examine the effectiveness of the optimization results in enhancing overall performance and improving residents' actual service acquisition(Tahmasbi & Haghshenas, 2019). The formula is as follows:

$$E_{pop} = \frac{\sum_{i=1}^{n} p_i x_i}{\sum_{i=1}^{n} p_i}$$

where n denotes the number of villages, xi represents the accessibility score or diversity index of village i, and pi is the population size of village iii. E_{pop} thus reflects the extent of improvement in residents' actual experience.

The methodological approach of this study aims to reveal the spatial supply–demand patterns of rural commercial service facilities and their improvement pathways through quantitative assessment and optimization. Accessibility was measured using the Two-Step Floating Catchment Area Method (2SFCA), and diversity with the Shannon index. Service gaps were then identified through a KNN-based graph and a graph convolutional network (GCN), which also generated optimization strategies. The newly proposed or reallocated sites were reintroduced into the baseline models, and accessibility and diversity were recalculated for comparison. Finally, the Gini coefficient and population-weighted means were applied to evaluate the effectiveness of optimization in improving service levels and reducing spatial disparities.

4. Result

4.1 Results Before and After Optimization

Before proceeding with the analysis, this study first employed the Two-Step Floating Catchment Area method (2SFCA) and the Shannon Diversity Index to calculate the accessibility and diversity scores of the six villages in Qingshan Town. The accessibility comparison reveals the extent of service improvement for both the overall region and peripheral villages after optimization, while the diversity comparison evaluates the effectiveness of the newly added facilities in enhancing the balance and richness of service types. Following the identification of facility gaps and the generation of additional facility locations using the KNN + GCN method, the accessibility and diversity scores of the six villages were recalculated. This section compares the changes in accessibility and diversity before and after optimization. To verify the effectiveness of the proposed methodological framework, this study compared the accessibility and diversity results of commercial service facilities in Qingshan Town before and after optimization.

As shown in Fig. 3, significant spatial disparities exist in the accessibility of commercial service facilities across the six villages of Qingshan Town. Prior to optimization, core villages such as Qingshan (0.00804) and Yaotang (0.00802) maintained relatively high accessibility, whereas peripheral villages such as Maoping (0.00060) and Baoershan (0.00048) were almost in a “service vacuum.” Tangdian (0.00299) and Jianghe (0.00207) demonstrated moderate levels of accessibility, yet these remained insufficient to meet residents’ daily needs. Overall, the results highlight a highly uneven “strong core–weak periphery” pattern of facility distribution before optimization.

With the introduction of the KNN + GCN method and the addition of new facility points, the overall level of accessibility across the region was substantially improved. Tangdian Village and Jianghe Village saw their accessibility scores rise from 0.00299 and 0.00207 to 0.00418 and 0.00296, respectively, effectively narrowing the gap with the core villages. Peripheral villages such as Maoping (0.00158) and Baoershan (0.00059) also recorded improvements. Although the magnitude of these increases was relatively modest, they nonetheless broke through the previous condition of “extremely low accessibility,” thereby alleviating the problem of service insufficiency. Meanwhile, core villages such as Qingshan (0.00928) and Yaotang (0.00889) further strengthened their already strong positions, demonstrating that the optimization process did not compromise the service advantages of central locations.

Overall, the post-optimization accessibility pattern demonstrates a trend of “general uplift and narrowing disparities”. Before optimization, the gap between the highest and lowest scores exceeded sixteen fold. after optimization, this disparity was significantly reduced, and peripheral villages were no longer entirely “left behind”. Furthermore, with regard to different service types, the optimization strategy effectively addressed deficiencies in categories such as Daily Life Services in several villages, thereby further enhancing the composite accessibility scores. These results indicate that the KNN + GCN method not only raised the overall level of accessibility but also improved intra-regional equity.

In terms of diversity, the results presented in Fig. 4 reveal a markedly uneven distribution of commercial service facilities across the six villages of Qingshan Town. The core village, Qingshan, possessed 138 facility points and achieved a diversity score of 0.961, far exceeding the other villages. By contrast,

Yaotang Village lacked both catering and daily life service facilities, resulting in a diversity score of 0 and placing it in a state of complete service deficiency. Although Tangdian and Maoping recorded diversity scores of 0.693, each village contained only two facilities in total, which is insufficient to meet residents' daily needs. Comparatively, Baoershan (0.868) and Jianghe (0.325) demonstrated slightly greater service variety but remained severely inadequate overall.

After the introduction of the KNN + GCN optimization method (Fig. 4), the diversity levels of all villages were significantly improved. Qingshan Village's diversity score rose from 0.961 to 0.978, further enhancing service balance while maintaining its core advantage. Improvements were particularly notable in peripheral villages. For instance, Yaotang's score increased from 0 to 1.099, Tangdian's from 0.693 to 1.095, and Maoping's from 0.693 to 1.099—all representing a leap from "low-level" to "high-level" diversity. Jianghe Village also experienced a substantial improvement, with its score rising from 0.325 to 1.005, indicating a strong level of completeness in multi-service coverage. The only village showing limited improvement was Baoershan, where diversity decreased slightly from 0.868 to 0.956; however, its score still remained within a relatively high range, suggesting that the optimization had only a minor effect on its structural balance.

Overall, the optimization measures effectively reduced inter-village disparities. Before optimization, the gap between the highest and lowest diversity scores was striking—0.961 in Qingshan versus 0 in Yaotang—nearly a twofold difference. After optimization, the diversity scores of all six villages converged within a narrow range of 0.95–1.10, eliminating extreme disparities and producing a highly balanced spatial distribution of service diversity. This transformation not only reflects a marked improvement in the equity of facility allocation but also implies enhanced efficiency, as more residents across villages can now access multiple types of services—catering, shopping, and daily life—thereby unlocking the overall potential of regional service utilization.

4.2 Equity Analysis

In terms of equity analysis, Table 3 shows that the initial Gini coefficient for accessibility was 0.45800, indicating a relatively high level of inequality. This suggests that significant disparities existed among villages in their opportunities to access commercial service facilities. Such disparities were most evident between core villages and peripheral ones: core villages such as Qingshan and Yaotang exhibited substantially higher accessibility compared with peripheral villages like Maoping and Baoershan, resulting in an overall spatial distribution characterized by a "strong center–weak periphery" pattern.

After the introduction of the KNN + GCN optimization method, the accessibility Gini coefficient decreased to 0.40402, reflecting a notable reduction in intra-regional inequality. Although core villages continued to maintain relatively high accessibility levels, peripheral villages experienced significant improvements, narrowing the gap with central settlements. This indicates that the optimization not only enhanced overall service coverage, thereby expanding opportunities for residents in disadvantaged areas to access essential services.

Table 3
Gini coefficients of accessibility and diversity before and after optimization

	Accessibility Gini coefficient	Diversity Gini coefficient
Before Optimization	0.45800	0.31567
After Optimization	0.40402	0.03072

In terms of diversity, the initial Gini coefficient was 0.31567, indicating substantial disparities among villages in the coverage and distribution of service types. Some villages, such as Qingshan, offered relatively rich service categories, while others, such as Yaotang and Jianghe, exhibited extremely low diversity, approaching a “single-function” state that reflected pronounced inequity.

After optimization, the diversity Gini coefficient dropped sharply to 0.03072, suggesting that service diversity had become far more balanced across villages.

4.3 Efficiency Analysis

In terms of efficiency, the overall mean accessibility score increased from 0.0037 before optimization to 0.00458 afterward, representing an improvement of approximately 23.8% (Table 4). This result indicates that the KNN + GCN optimization method can effectively raise the overall accessibility of service facilities without undermining the advantages of existing core villages. More importantly, the population-weighted mean accessibility rose from 0.00404 to 0.00493, an increase of 22.1%.

Table 4
Efficiency Analysis of Accessibility

Time and Type	Original accessibility average score	Population-weighted average accessibility score
Before Optimization	0.0037	0.00404
After Optimization	0.00458	0.00493

When combined with population data, the results highlight a more efficient resource allocation. Qingshan, Jianghe, and Baoershan villages—each with relatively large populations of 5,506, 4,320, and 3,950 respectively—were previously in a state of “population concentration but insufficient service coverage.” Following optimization, the population-weighted accessibility of these villages improved significantly, ensuring that larger settlements received resource allocations more proportional to their population size. At the same time, smaller villages such as Maoping and Tangdian, which had suffered from service deficits before optimization, also experienced marked improvements in accessibility through the addition of new facilities.

Furthermore, the spatial distribution of accessibility before optimization displayed a distinct “center–periphery” pattern, where several villages with larger populations experienced efficiency constraints due

to facility shortages. After optimization, however, the configuration of facilities became more closely aligned with the spatial distribution of the population. This not only raised the overall level of regional accessibility but also enhanced efficiency when weighted by population. These findings indicate that the KNN + GCN optimization method achieved a dual improvement in efficiency: on the one hand, the increase in the overall mean reflects a universal enhancement; on the other hand, the rise in the population-weighted mean highlights a more precise response to demographic demand.

In addition, the optimized layout significantly improved the efficiency of service diversity across the study area, as shown in Table 5. Prior to optimization, the average diversity score was only 0.59005, revealing that most villages suffered from functional deficiencies. Peripheral villages such as Yaotang and Maoping were particularly disadvantaged, existing in a “single-function” or even “no-service” state. This condition not only constrained residents’ access to multiple types of services but also limited the ability of larger settlements to meet their populations’ daily and diversified needs. Following optimization, the average diversity score rose to 1.03856, nearly doubling the pre-optimization level.

Table 5
Efficiency Analysis of Accessibility

Time and Type	Original Diversity Average Score	Population-weighted average diversity score
Before Optimization	0.59005	0.32508
After Optimization	1.03856	1.02646

From the population-weighted perspective, the efficiency gains brought by optimization become even more evident. Before optimization, the population-weighted average diversity score was only 0.32508. This indicates that although some core villages exhibited relatively high levels of diversity, a large share of the population resided in areas with insufficient service variety, thereby constraining the overall service experience of residents. After optimization, this indicator rose to 1.02646, surpassing the critical threshold of 1.0. This outcome demonstrates that the optimization strategy successfully distributed newly added facilities in villages with higher population concentrations, enabling a greater number of residents to access multiple categories of services within their daily travel range.

The optimized layout achieved a “dual enhancement” of diversity and efficiency. Service types became more balanced, raising overall diversity, while higher population-weighted scores showed that supply expansion was effectively aligned with population distribution, maximizing resource use. This not only improved residents’ daily convenience but also offered a practical pathway for rational resource allocation. The KNN and GCN method simultaneously advanced equity and efficiency: the pre-optimization “strong center–weak periphery” pattern was eased, peripheral villages gained in accessibility and diversity, and Gini coefficients declined, reflecting reduced disparities. At the same time, higher mean and population-weighted values indicated better demographic matching, allowing more residents access to a wider range of services.

5. Discussion

From an equity perspective, this study underscores the critical role of facility layout optimization in reducing spatial disparities. Before optimization, Qingshan Town exhibited a clear “core–periphery” pattern: core villages such as Qingshan and Yaotang had accessibility scores near 0.008, while peripheral villages like Maoping and Baoershan were almost in a “service vacuum” at about 0.0005. Diversity showed similar imbalance, with Qingshan Village scoring 0.961 and Yaotang registering zero, reflecting both uneven resource allocation and unmet daily needs in education, shopping, and healthcare. After applying the KNN and GCN optimization, disparities narrowed significantly: the Gini coefficient for accessibility dropped from 0.458 to 0.404, and for diversity from 0.316 to 0.031. Most strikingly, Yaotang Village, once at zero diversity, rose to 1.099, among the highest in the region. These results demonstrate that optimization not only improved overall service provision but also delivered compensatory benefits to long-marginalized villages, thereby advancing spatial equity.

On the efficiency side, the optimization delivered notable gains. The overall mean accessibility score rose from 0.0037 to 0.00458, and the population-weighted mean from 0.00404 to 0.00493, indicating improved provision and closer alignment with population distribution. Larger villages such as Qingshan and Jianghe, once constrained by low accessibility relative to their size, gained more equitable service access, reflecting a better match between spatial supply and demographic demand. Diversity improvements were even more striking: the average score increased from 0.59005 to 1.03856, and the population-weighted mean from 0.32508 to 1.02646. This shows that new facilities not only broadened service types but also enriched provision in larger settlements like Jianghe and Baoershan, substantially enhancing residents’ service experience. Overall, the results highlight a dual achievement of expanding supply and strengthening population matching in the efficiency dimension.

Taken together, the findings indicate that facility optimization is not a zero-sum trade-off but a “win–win mechanism.” While equity and efficiency are often viewed as conflicting—where improving disadvantaged areas might weaken core regions—this study shows the opposite: by applying a scientifically grounded siting method, optimization reduced disparities, enhanced services in peripheral villages, and preserved or even strengthened advantages in core settlements, demonstrating real synergies between equity and efficiency. For rural public service systems, equity means ensuring that peripheral areas are not overlooked, while efficiency requires aligning limited resources with population demand—both are indispensable. The empirical evidence confirms that data-driven, model-based approaches can meet these dual challenges, providing strong support for rural revitalization and balanced service provision. Future planning should therefore integrate equity and efficiency explicitly, using optimization models to achieve “more with less” and promote sustainable, balanced spatial development across urban and rural contexts.

6. Conclusion

This study, taking Qingshan Town in Anhui Province as a case, developed a generative design–driven framework for optimizing rural commercial service facilities. This integrative approach moves beyond static models or subjective judgment, offering a reproducible, data-driven pathway for optimizing public service facilities. Results revealed pronounced “core–periphery” disparities before optimization and showed that the proposed strategy not only raised overall service levels but also mitigated deficiencies in peripheral villages, achieving simultaneous improvements in accessibility and diversity. These findings provide data-driven evidence for rural revitalization and balanced urban–rural service provision, enhancing residents’ sense of access and satisfaction.

Some limitations remain. This study focused mainly on spatial distribution and physical accessibility, without considering facility quality, service capacity, or residents’ perceptions, which constrains comprehensiveness. Future work could integrate dynamic data such as mobile signaling or real-time POI updates to improve adaptability, and link facility optimization with socio-economic variables like income levels and aging demographics for a more nuanced equity analysis. Extending the framework to larger-scale urban–rural or regional contexts would also test its generalizability.

Declarations

Acknowledgements

Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors.

Disclosure statement

The authors report there are no competing interests to declare.

Fund

This research received no funding or support.

Author Contributions

Conceptualization, W.S.; methodology, X.L.; software, W.S.; validation, X.L.; formal analysis, Z.C.; resources, Z.C.; data curation, Y.L.; data analysis, Z.C.; writing – original draft preparation, X.L.; writing and editing, X.L.; visualization, W.S.; supervision, X.L.; project administration, Q.J. All authors have read and agreed to the published version of the manuscript.

ORCID

Xinlong Li, https://orcid.org/0009-0005-2329-5248

Zhiwen Cui, https://orcid.org/0009-0004-5072-8099

Wenwen Shi, https://orcid.org/0009-0006-0356-1702

Qinghua Jia, https://orcid.org/0009-0001-5718-0829

Data availability statement

All data generated or analyzed during this study are included in this published article.

References

1. *China's administrative village boundary data*. (n.d.). Resource and Environment Science Data Platform. Retrieved September 24, from (2025). https://www.resdc.cn/data.aspx?DATAID=333
2. Cover, T. & Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory*. **13** (1), 21–27 (1967).
3. *Data on administrative boundaries of Chinese towns and villages in 2015*. (n.d.). Resource and Environment Science Data Platform. Retrieved September 24, from (2025). https://www.resdc.cn/data.aspx?DATAID=360
4. Fotheringham, A. S., Charlton, M. & Brunsdon, C. *Quantitative geography: Perspectives on spatial data analysis*. (2000). https://www.torrossa.com/gs/resourceProxy?an=4913049&publisher=FZ7200
5. Gómez-Varo, I., Delclòs-Alió, X., Miralles-Guasch, C. & Marquet, O. Unravelling retail diversity through its spatial relationship with social vulnerability and tourism pressures in Barcelona. *J. Urban Des.* **29** (3), 280–300. https://doi.org/10.1080/13574809.2023.2245761 (2024).
6. Gu, M., Li, S. & Zheng, G. Optimizing spatial accessibility and equity of hierarchical older adult care facilities using a multi-modal two-step floating catchment area method: A case study of lin'an district, hangzhou. *Frontiers in Public Health*, *13*. (2025). https://doi.org/10.3389/fpubh.2025.1559463
7. Gu, T., Zhu, M., Jiao, J. & Chung, H. Regional optimization of new-energy bus charging stations using a hybrid model integrating graph convolutional networks and simulated annealing. *J. Transp. Geogr.* **128**, 104382. https://doi.org/10.1016/j.jtrangeo.2025.104382 (2025).
8. Gu, Z. et al. Density, Diversity, and Design: Evaluating the Equity of the Elderly Communities in Three Measures of the Built Environment. *Land*, *11*(11), 1976. (2022). https://doi.org/10.3390/land11111976
9. *Introduction to Qingshan Town, Jinzhai County*. (n.d.). Jinzhai County People's Government. Retrieved September 24, from (2025). https://www.ahjinzhai.gov.cn/public/6596691/37737037.html
10. Jiang, B., Zhang, Z., Lin, D., Tang, J. & Luo, B. Semi-supervised learning with graph learning-convolutional networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11313–11320. (2019). http://openaccess.thecvf.com/content_CVPR_2019/html/Jiang_Semi-Supervised_Learning_With_Graph_Learning-Convolutional_Networks_CVPR_2019_paper.html

11. Karam, S. & Ryerson, M. S. Operating at the individual level: A review of literature and a research agenda to support needs-forward models of transport resource allocation. *Transp. Res. Interdisciplinary Perspect.* **21**, 100887 (2023).
12. Kipf, T. N. Semi-supervised classification with graph convolutional networks. *arXiv Preprint arXiv:1609.02907*. (2016). https://bibbase.org/service/mendeley/bfbbf840-4c42-3914-a463-19024f50b30c/file/25dbdd06-4704-a33f-23d9-c626b08adc1e/160902907.pdf.pdf
13. Lee, J. H., Goh, S., Lee, K. & Choi, M. Y. Spatiotemporal distributions of population in Seoul: Joint influence of ridership and accessibility of the subway system. *EPJ Data Sci.* **10** (1), 41. https://doi.org/10.1140/epjds/s13688-021-00298-3 (2021).
14. Li, H. et al. Rural residential choices: Unravelling the nexus of accessibility and preferences. *J. Asian Archit. Building Eng.* 1–13. https://doi.org/10.1080/13467581.2024.2407580 (2024).
15. Lin, H. T., Sun, C. Y. & Hung, C. T. A study in the relationship between greenery of urban parks and bird diversity in tainan city, taiwan. *WIT Trans. Ecol. Environ.* **117**, 193–202 (2008).
16. Liu, L. et al. Spatial accessibility and equity evaluation of medical facilities based on improved 2SFCA: A case study in xi'an, China. *International Journal of Environmental Research and Public Health*, *20*(3), 2076. (2023).
17. Luo, W. & Wang, F. Measures of spatial accessibility to health care in a GIS environment: Synthesis and a case study in the chicago region. *Environ. Plan.* **30** (6), 865–884. https://doi.org/10.1068/b29120 (2003).
18. *Map World*. (n.d.). Map World. Retrieved September 24, from (2025). https://map.tianditu.gov.cn/
19. Nakase, R., Chou, C. C., Aoki, Y., Yoh, K. & Doi, K. Evaluating hierarchical diversity and sustainability of public transport: From metropolis to a weak transport demand area in western japan. *Front. Sustainable Cities*. **3**, 667711 (2021).
20. Pérez-Fernández, O. & Michel, G. R. Spatial accessibility to healthcare facilities: GIS-based public–private comparative analysis using floating catchment methods. *ISPRS Int. J. Geo-Information*. **14** (7), 253 (2025).
21. Pirie, G. H. Measuring accessibility: A review and proposal. *Environ. Plann. A: Econ. Space*. **11** (3), 299–312. https://doi.org/10.1068/a110299 (1979).
22. Regression, N. An introduction to kernel and nearest-neighbor. *Am. Stat.* **46** (3), 175–185 (1992).
23. *Remote sensing monitoring data on China's land use status in 2025*. (n.d.). Resource and Environment Science Data Platform. Retrieved September 24, from (2025). https://www.resdc.cn/data.aspx?DATAID=360
24. Search POI. (n.d.). Amap API. Retrieved September 24. from (2025). https://amap.apifox.cn/api-14636359
25. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27** (3), 379–423 (1948).
26. Sun, S., Sun, Q., Zhang, F. & Ma, J. A Spatial Accessibility Study of Public Hospitals: A Multi-Mode Gravity-Based Two-Step Floating Catchment Area Method. *Appl. Sci.* **14** (17), 7713.

https://doi.org/10.3390/app14177713 (2024).

27. Tahmasbi, B. & Haghshenas, H. Public transport accessibility measure based on weighted door to door travel time. *Comput. Environ. Urban Syst.* **76**, 163–177 (2019).
28. Talen, E. & Anselin, L. Assessing spatial equity: An evaluation of measures of accessibility to public playgrounds. *Environ. Plann. A: Econ. Space*. **30** (4), 595–613. https://doi.org/10.1068/a300595 (1998).
29. Vaishar, A. & Šťastná, M. Accessibility of Services in Rural Areas: Southern Moravia Case Study. *Sustainability* **13** (16), 9103. https://doi.org/10.3390/su13169103 (2021).
30. Wang, F. Measurement, optimization, and impact of health care accessibility: A methodological review. *Ann. Assoc. Am. Geogr.* **102** (5), 1104–1112. https://doi.org/10.1080/00045608.2012.657146 (2012).
31. Wu, Z. et al. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Networks Learn. Syst.* **32** (1), 4–24 (2020).
32. Xu, H. *Sustainability enhancement in healthcare facility design: Structural and functional optimization based on GCN*. Engineering. (2025a). https://doi.org/10.20944/preprints202503.0805.v1
33. Xu, H. Sustainability enhancement in healthcare facility design: Structural and functional optimization based on GCN. *Proceedings of the 2025 International Conference on Health Big Data*, 227–233. (2025b). https://doi.org/10.1145/3733006.3733043
34. Xun, J., Zhang, M., Xu, G. & Guo, X. Diversity and Influencing Factors of Public Service Facilities in Urban (Suburban) Railway Life Circle—Evidence from Beijing Subway Line S1, China. *Land* **13** (8), 1286. https://doi.org/10.3390/land13081286 (2024).
35. Zhang, S., Tong, H., Xu, J. & Maciejewski, R. Graph convolutional networks: A comprehensive review. *Comput. Social Networks*. **6** (1), 11. https://doi.org/10.1186/s40649-019-0069-y (2019).
36. Zhang, Y., Fu, X., Yu, Z. & Luo, S. How does multi-modal travel enhance tourist attraction accessibility? A refined two-step floating catchment area method using multi-source data. *Trans. GIS*. **28** (2), 278–302. https://doi.org/10.1111/tgis.13136 (2024).
37. Zhang, Y., Li, W., Deng, H. & Li, Y. Evaluation of Public Transport-Based Accessibility to Health Facilities considering Spatial Heterogeneity. *Journal of Advanced Transportation*, *2020*, 1–10. (2020). https://doi.org/10.1155/2020/7645153
38. Zhao, Y., Xi, G., Zhang, H. & Zhen, F. Optimization site selection for online-to-offline convenience facilities using knowledge graph and graph neural network: A case study of nanjing, China. *Appl. Spat. Anal. Policy*. **18** (2), 57. https://doi.org/10.1007/s12061-025-09662-6 (2025).

Figures

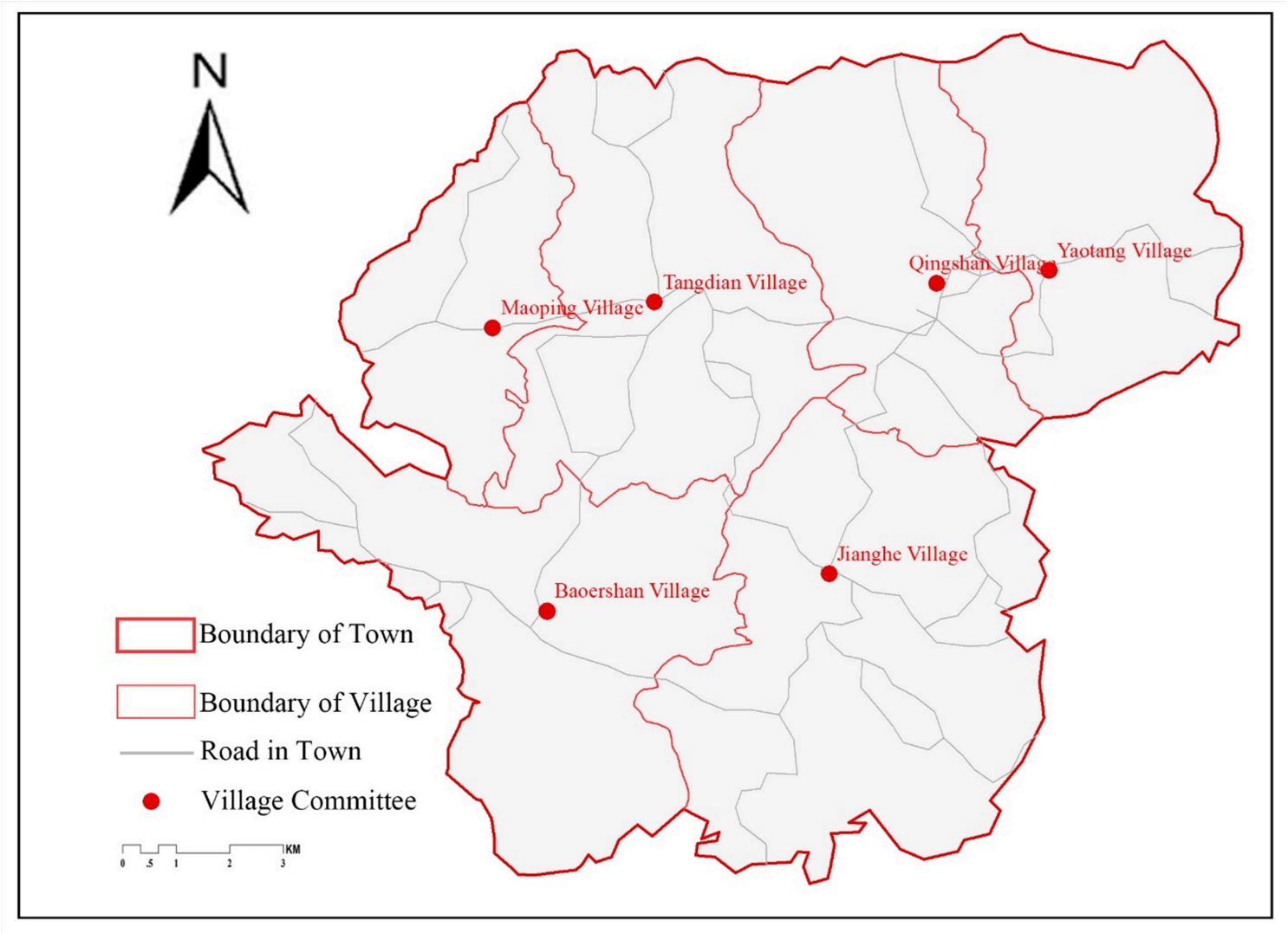


Figure 1

Distribution of villages in Qingshan Town

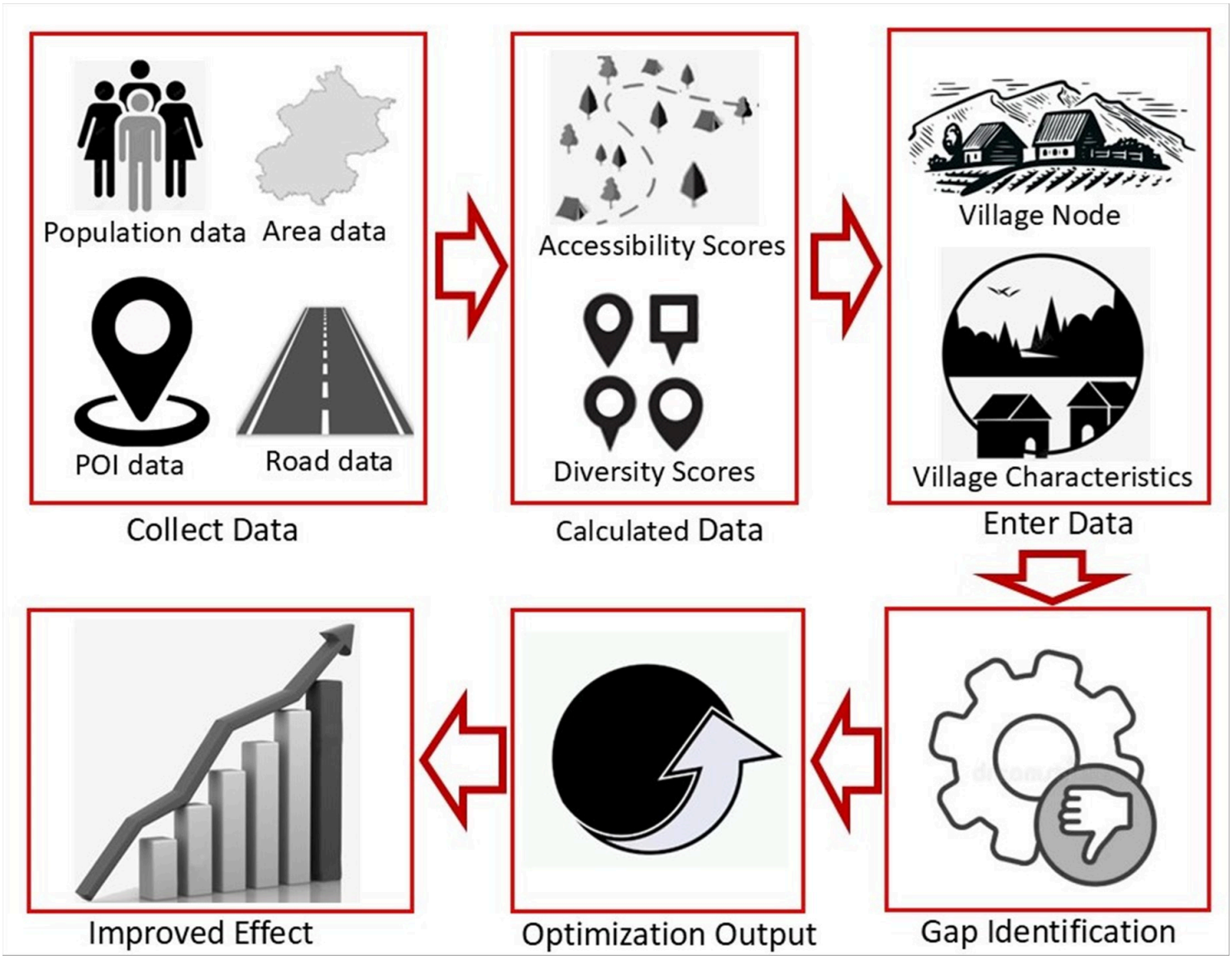


Figure 2

Research framework

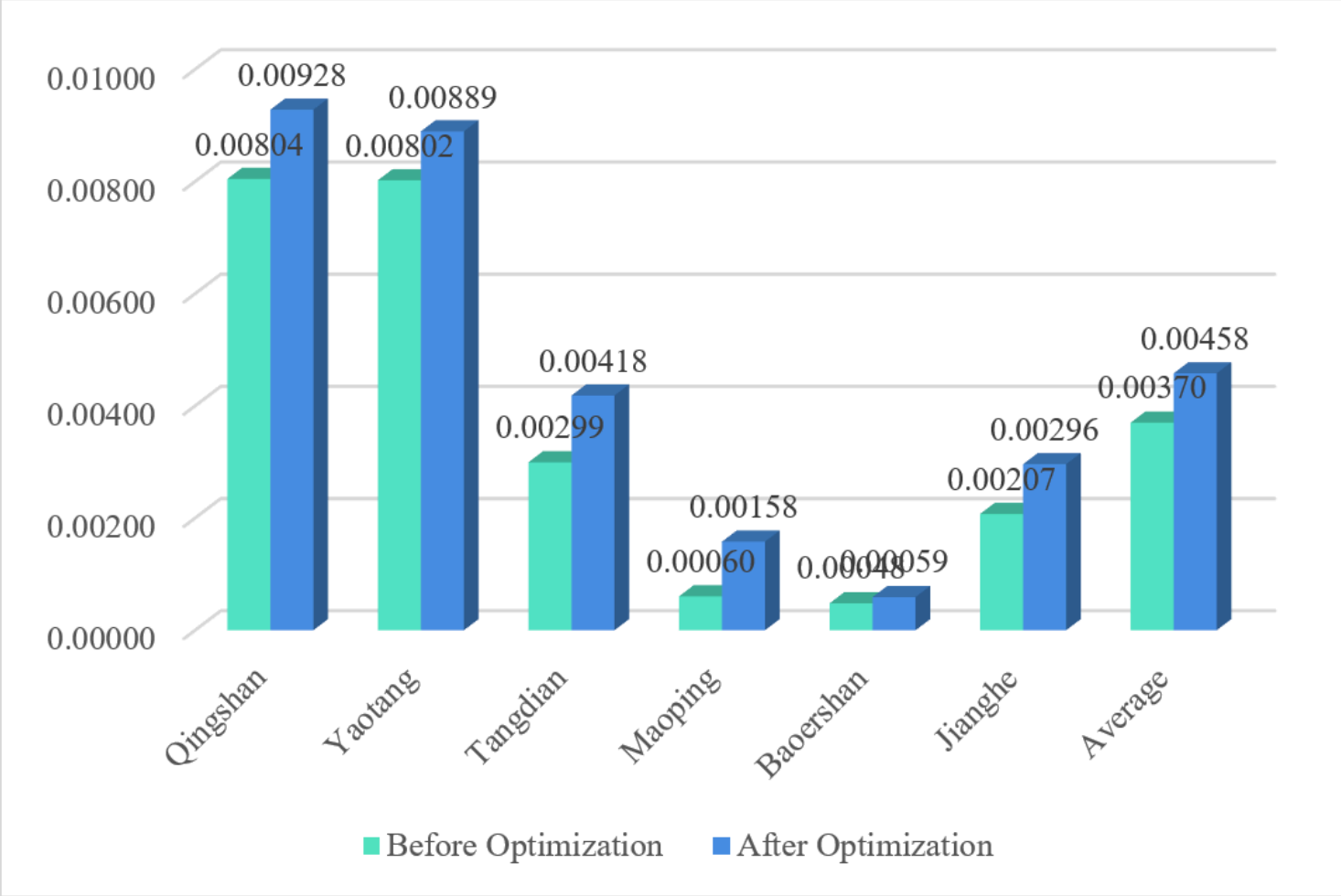


Figure 3

Village Accessibility Scores Before and After Optimization

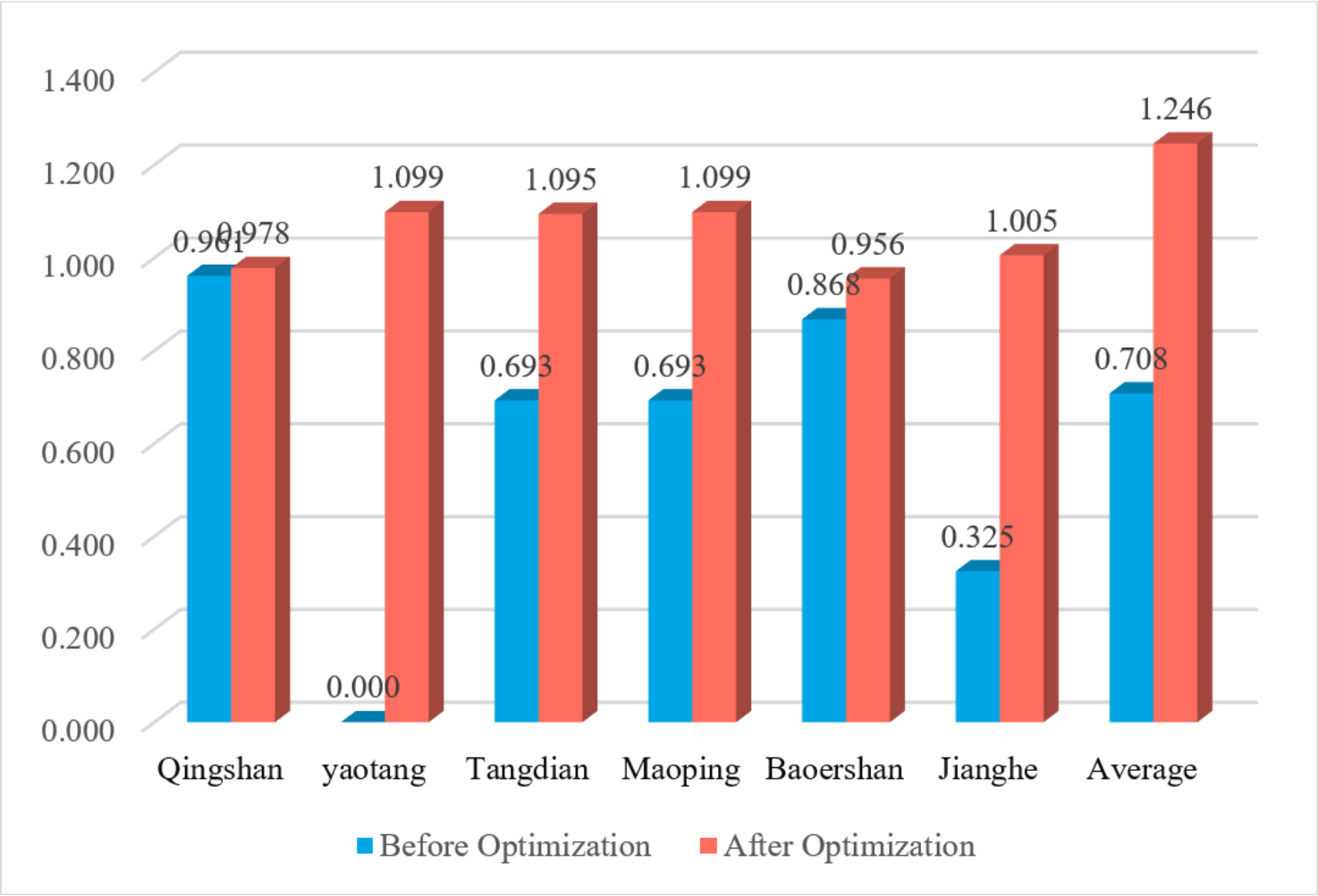


Figure 4

Village Diversity scores before and after optimization

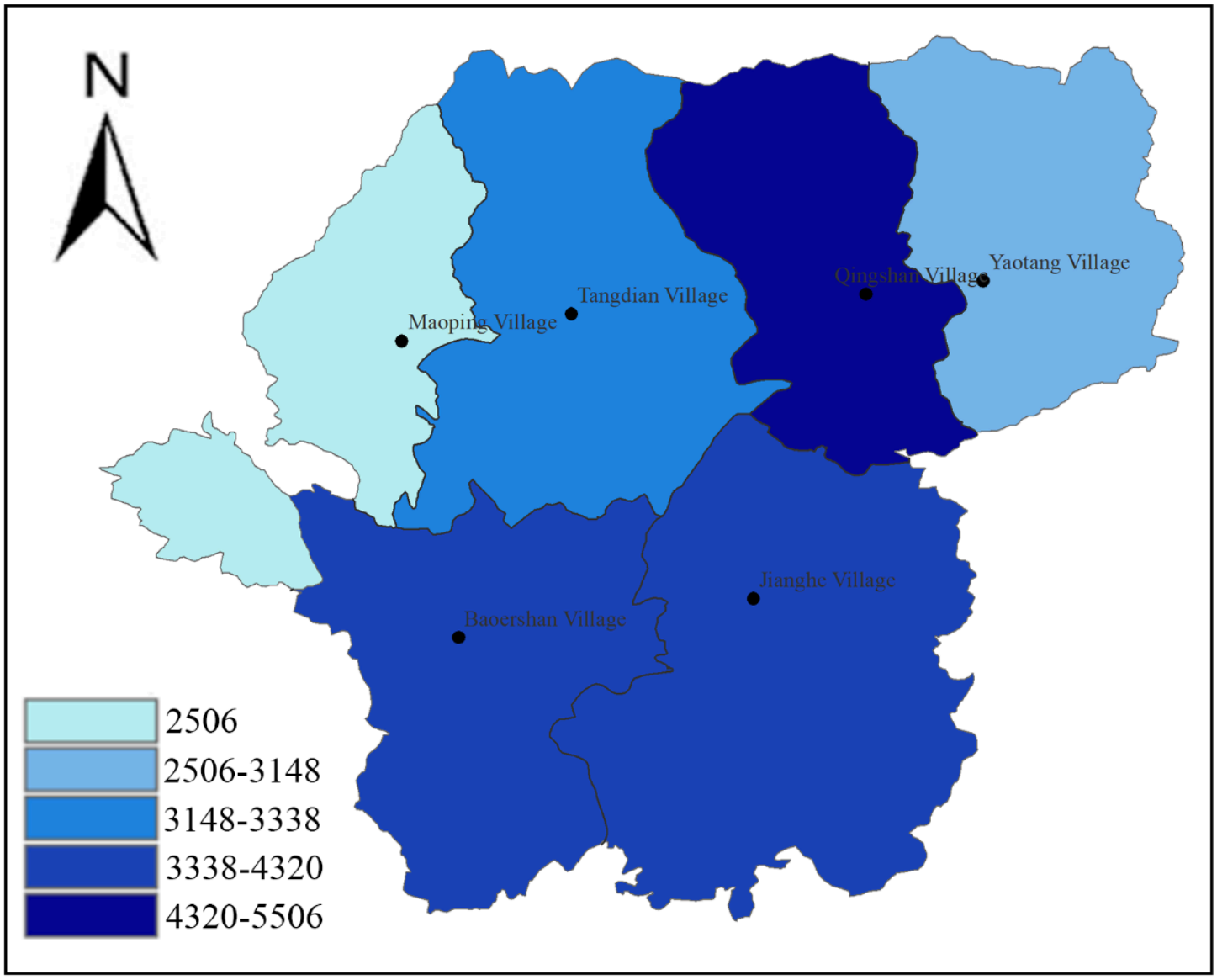


Figure 5

Village Population in Qingshan Town