Delft University of Technology

Investigating rural public spaces with cultural significance using morphological, cognitive and behavioural data

Bai, N.; Nourian, Pirouz; Pereira Roders, A.; Bunschoten, Raoul; Huang, Weixin; Wang, Lu

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Article

Investigating rural public spaces with cultural significance using morphological, cognitive and behavioural data

Nan Bai ⓘ**, Pirouz Nourian** ⓘ **and Ana Pereira Roders** ⓘ
TU Delft, the Netherlands

Raoul Bunschoten
TU Berlin, Germany

Weixin Huang ⓘ **and Lu Wang**
Tsinghua University, China

Abstract

During the rural [re]vitalization process in China, national strategies required rural public spaces with cultural significance to be identified before planning decision-making. However, places identified as culturally significant by planners and visitors can differ from the ones mostly used and valued by locals. Even if there is a growing interest in integrating local perspectives and experiences in planning, studies seldom discuss and compare openly the adequacy of spatial configuration, cognition and behaviour to support it. This study took Anyi Historic Village Cluster as a case study to empirically investigate rural public spaces with three distinct, yet related approaches: (1) Morphological: spatial network centralities based on space syntax; (2) Cognitive: Lynchian village images with semi-structured interviews; (3) Behavioural: spatiotemporal occupation patterns using Wi-Fi positioning tracking. Significant places valued and used by locals and non-locals were detected with the multi-source data. Furthermore, multivariant regression models managed to characterize the relationship among different aspects of investigated rural public spaces, which also helped diagnose places of interest to prioritize in planning, demonstrating the advantage of integrating the sources of information in practice instead of studying them apart. Results can also assist rural planning on how to identify what to preserve, what to enhance, and how to develop such spaces, without overlooking the local needs or losing the rural identity.

Keywords
Rural planning, space syntax, cognition, big data, activity space

Corresponding author:
Weixin Huang, School of Architecture, Tsinghua University, S400C, Beijing 100084, China.
Email: huangwx@mail.tsinghua.edu.cn

Introduction

In spatial planning, morphological, cognitive and behavioural data are essential for understanding and representing the spatial configurations and their effect on the vitality of human settlements, as well as identifying the most significant and valuable public spaces empirically (Haq, 2003; Kim, 1999; Kim and Penn, 2004; Penn, 2003). Among them, *morphology* represents how space is topologically and geometrically structured, *cognition* shows how people view, perceive and memorize the space, and *behaviour* demonstrates how people actually occupy and use the space. Methods corresponding to the three sources of information were introduced to spatial analysis in the 1960–1980s as:

1. A family of topological accessibility and visibility analysis methods known as Space Syntax for Morphological Analysis (Hillier and Hanson, 1984; Hillier, 2007);
2. A set of interviewing and sketching techniques studying mental maps of cities through Cognitive Mapping (Lynch, 1960);
3. A group of observation-based strategies for activating human-oriented cities being dubbed as Behavioural Studies (Jacobs, 1993; Gehl, 2011).

Each method has been broadly applied in architecture design, urban planning, tourism and heritage management. Arguably, these three lineages of spatial design methods all focused on the relations between space and human behaviour. Much research has focused on combining two of the three information sources, creating terms such as *space semantics*, *cognitive space* and *behaviour urbanism* (Franz and Wiener, 2008; Hanna, 2020; Haq, 2003; Kim and Penn, 2004; Noyman et al., 2019; Orellana, 2019; Penn, 2003; Stevens, 2006). These studies implied that both perceived and used places are influenced by the spatial configuration and socio-behavioural patterns. Yet, conclusions are not necessarily alike (Franz and Wiener, 2008; Haq, 2003; Ma et al., 2018; Stevens, 2006). Remarkably, 20 years after Kim's dissertation about spatial configuration, cognition and behaviour, empirical studies combining all three sources are still scarce (Kim, 1999).

Moreover, considering that these methods were mostly applied to urban environment, further studies to validate their applications to rural environments are still needed. Yet, it has become urgent in China to investigate the rural spaces and reveal the local identities of villages. In the past decades, rapid urbanization has been challenging the millennial rural-based foundations of the Chinese society (Fei et al., 1992). The two extremes of *non-use* and *over-use* have put the rural identities at risk (Wang, 2000). Fearing the loss of local identities, scholars have realized the urgency of [re] vitalizing rural heritage (Long, 2014; Yang et al., 2016). Consequently, a national policy, *The Five-Year Plan on Rural Vitalization Strategy (2018–2023)*, was released on September 26th, 2018, expressing the strong intention to support [re]vitalizing and preserving rural China in the upcoming years (CPC Central Committee and State Council, 2018).

This national policy calls for '*constructing modernized villages based on the rural society with local identities, avoiding villages mimicking each other, and preventing rural landscape from getting over-urbanized*' (CPC Central Committee and State Council, 2018). The policy requires that before any planning decision-making, local governments and planners need to thoroughly identify the local rural heritage, which may not be listed officially. Some tangible objects (such as a well or a tree) and spaces with intangible attributes of significance (such as the space under a tree providing shade and breeze) in historical villages are not normally recognized as heritage, but contribute significantly to the collective memory and emotions of villagers, which could be recognized and safeguarded as local heritage (Wang, 1999). Scholars have recommended to involve locals as key stakeholders in the [re]vitalization process (Chen and Li, 2008; Wang, 2000). This approach is endorsed by the UNESCO *Recommendation on the Historic Urban Landscape*, which calls for

actively involving various stakeholders in heritage planning and decision-making process, considering their needs and preferences (Bandarin and van Oers, 2012; Pereira Roders, 2019; UNESCO, 2011). Under such circumstances, the aforementioned morphological, cognitive and behavioural analyses have great potential to comprehensively assist the identification of cultural, historical or natural significance conveyed by various stakeholders, and locals in particular, to be included in planning processes and related decision-making (Bai and Wang, 2019a).

This study aims to reveal the places with cultural significance in historic villages to support planning decisions by involving both locals and non-locals to identify the most used and valued significant public spaces. To do this, a methodological framework was proposed following Kim (1999) to link the three spatial information sources, that is, *morphology*, *cognition*, and *behaviour*, and state-of-the-art methods for each source were jointly applied in a Chinese historic village cluster.

Methodology

As is shown in Figure 1, the methodological framework consists of four parts: three parallel substudies and a synthesis sub-study integrating their results. For each information source, a similar workflow was followed:

1. Supporting materials (e.g. maps, interview scripts, Wi-Fi devices) are prepared;
2. A set of public spaces is determined directly or indirectly;
3. Data is collected for each space in the set describing its features;
4. Specific analytical methods are applied to derive key indicators to quantify the information source;
5. The key indicators are inputted to the synthesis sub-study to compare and characterize the relationship among information sources using regression models as dependent and independent variables and
6. Spaces with either high key-indicator scores or obvious mismatch among information sources are identified to prioritize in planning.
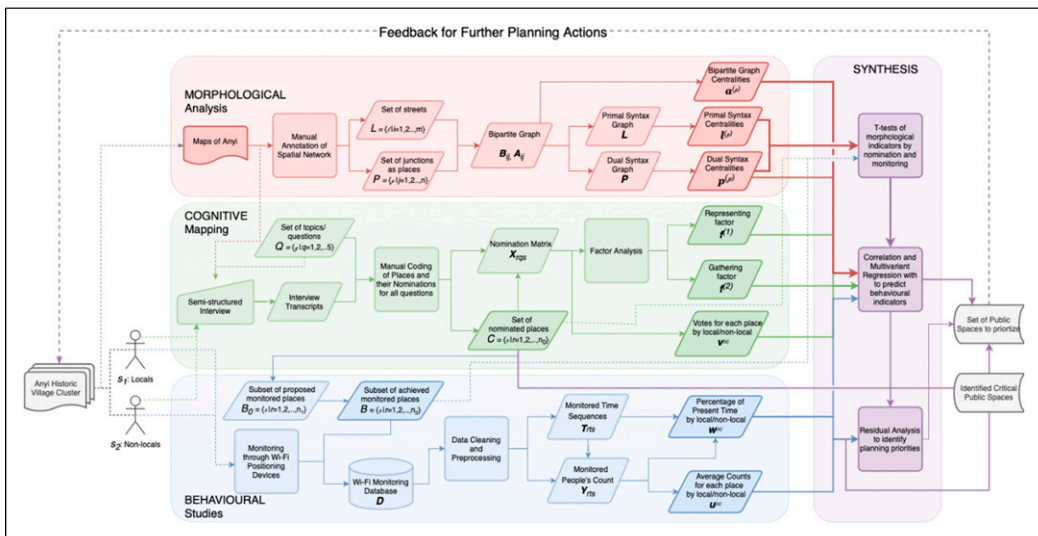

Figure 1. The overall methodological framework to investigate rural public spaces using morphological, cognitive and behavioural sources of information.

Case study

Anyi Historic Village Cluster (AHVC) is composed of three villages, located in central Jiangxi Province, south-eastern China. According to the local municipal archive, AHVC has a history of around 1000 years (Anyi County Archive Committee, 2007). Seven monuments (including residential buildings, ancestral halls and an old opera stage) distributed in AHVC were listed as *Major Historical and Cultural Site Protected* at Provincial Level. Besides, a thousand-year-old camphor tree, the ancient stone-paved trade paths, old wells and stone archways were also considered as traditional elements in the archive though not listed as heritage. Since 2006, the local government has been considering both heritage preservation and exploitation. Within the process of rural [re] vitalization, a few deserted houses were chosen as pilot projects and renovated with modern aesthetics. Consequently, contemporary buildings have been introduced in the historical areas, which can be controversial for the traditional rural context. This research was conducted between November 2018 and March 2019. The entire village cluster was taken as a case study, treating all existing public spaces equally, old or new, listed or unlisted as heritage, without preconception, to investigate their uses and cultural significance. Further information about AHVC could be found in Supplemental Materials 01 and Figure S1.

Morphology – Bipartite-graph space syntax

In the course of Space Syntax theories' development since the 1980s, different approaches have been introduced concerning network centralities and accessibility. Such approaches can be characterized with how they represent spatial relations, such as with axial lines, convex maps, angular segments, visibility graphs, voxel units, spectral methods, etc. (Bafna, 2003; Fisher-Gewirtzman et al., 2013; Haq, 2003; Lu et al., 2019; Penn, 2003; Turner, 2007; Turner et al., 2001; Nourian, 2016; Nourian et al., 2016). These models abstracted the space with different conditions and fitted their own analytical purposes, sometimes causing criticisms and contradictory evidence, especially concerning the use of axial lines (Ericson et al., 2020; Hillier, 1999; Ma et al., 2018; Ratti, 2004; Read, 2011).

In this study, a more generic representation of the space – the bipartite graph approach – is proposed following previous studies (Batty, 2004; 2013; Nourian et al., 2016; Porta et al., 2006). A set of street-centrelines $\mathcal{L} = \{\ell_i | i \in [0,m) \subset \mathbb{N}\}$ and a set of street junctions $\mathcal{P} = \{\boldsymbol{p}_j | j \in [0,n) \subset \mathbb{N}\}$ were annotated from the base-map of AHVC. The topological interrelation of the two sets was annotated to construct a bipartite graph with the adjacency matrix $\boldsymbol{A} := [A_{i,j}]_{(m+n)\times(m+n)}$ and bi-adjacency matrix $\boldsymbol{B} := [B_{i,j}]_{m\times n}$ such that

$$B_{i,j} = \begin{cases} 1, & if \quad \boldsymbol{p}_j \in \ell_i \\ 0, & otherwise \end{cases}, \text{and } \boldsymbol{A} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{B} \\ \boldsymbol{B}^T & \boldsymbol{0} \end{bmatrix}$$

The bi-adjacency matrix $\boldsymbol{B}$ of Line-to-Point connections was projected to unipartite Line-to-Line (primal-syntax) graph $\boldsymbol{L} = \boldsymbol{B}\boldsymbol{B}^T$ and unipartite Point-to-Point (dual-syntax) graph $\boldsymbol{P} = \boldsymbol{B}^T\boldsymbol{B}$, following Batty (2013) and Nourian et al. (2016). All three graphs (two unipartite and one bipartite) were constructed in topological distance based on the definition of matrix $\boldsymbol{B}$.

Degree Centrality (whether a place directly connects to many places), *Closeness Centrality* (whether a place is closely connected to many other places), *Betweenness Centrality* (whether a place appears in many shortest paths connecting the network as bridge) and *PageRank [Centrality]* (whether a place connects to many 'circles' of 'central' places, and thus demonstrating the steady state of a random walk) of the graphs were computed using *NetworkX* Python Library (Hagberg et al., 2008). The computation of the four centralities relies on the concepts of shortest paths and

random walks, respectively, which is believed to correspond with the space discovery modes with different levels of familiarity (Nourian, 2016). All four centralities are common in social network analysis and space syntax, capturing related yet distinct characteristics of the nodes as either street-lines or junctions (Agryzkov et al., 2019; Agryzkov et al., 2019; Boulmakoul et al., 2017; Curado et al., 2020; Nourian, 2016; Nourian et al., 2016; Page et al., 1998; Porta et al., 2006; Taheri et al., 2017; Wasserman and Faust, 1994). The centrality computation yields three groups of vectors

$$l^{(\gamma)} := \left[\sigma i^{(\gamma, L)} \middle| \, i \in \mathcal{L}\right]_{m \times 1},$$

$$p^{(\gamma)} := \left[\sigma i^{(\gamma, P)} \middle| \, i \in \mathcal{P}\right]_{n \times 1},$$

$$\text{and} \quad a^{(\gamma)} := \left[\sigma i^{(\gamma, A)} \middle| \, k \in \mathcal{L} \cup \mathcal{P}\right]_{(m+n) \times 1},$$

where $\sigma_v^{(\gamma, G)}$ denotes the centrality index γ (one of the four indices) of node v (either a street-centreline or a street junction) computed in graph G (one of the three graphs). The vectors characterizing the bipartite and unipartite graphs were applied to describe the spatial morphological structure and to identify the most central places in AHVC.

Cognition – semi-structured interview

Along with the 'five elements' of cities, Lynch also proposed a methodology with semi-structured interviews and cognitive map sketching to understand how people sense the city they live in (Lynch, 1960). Since then, his theories have been broadly referenced, adapted and sometimes critiqued in urban planning, tourism and environmental psychology (Ford, 1999; Filomena et al., 2019; Hall and Page, 2014; Morello and Ratti, 2009; Stevens, 2006). Although his purpose was originally about wayfinding and the structure of city imageability, Lynch did interview about the significance of places for citizens and the *symbolic* and *emotional* values in particular (Lynch, 1960), allowing for amendments by adding space usage, playability and cultural significance into the framework (Bai & Wang, 2019b; Peng et al., 2020; Stevens, 2006; Van der Aa, 2005).

A semi-structured interview was designed under the Lynchian paradigm, while dimensions of cultural significance and space usage were added by inquiring *representative*, *meaningful*, *crowded* and *attractive* places for participants. Twelve participants sampled from key stakeholders joined the interview, including six local villagers and six non-locals (architects, government staff and managers). Maps of AHVC were presented during the interview to help participants recall the places. Participants were asked to mark the places where they refer to, or orally describe them if they could not read maps (Bai & Wang, 2019b). The interviews were recorded and transcribed, being consented to by all participants.

The transcription was coded inductively on five main topics (denoted as q_q) in *NVivo 11 Plus* (Richards, 1999; Bazeley, 2007). All places mentioned by participants associating with one or more topics were coded as r_r, forming a set $\mathcal{C} = \{r_r | r \in [0, n_0)\}$ of cognitively significant places, where $\mathcal{C} \subset (\mathcal{L} \cup \mathcal{P})$. The two stakeholder groups (locals s_0 and non-locals s_1) were coded separately. The number of mentioning $x_{r,q,s}$ in the transcript of each place r_r for each topic q_q by each stakeholder group s_s formed a matrix $X_{r,q,s} := [x_{r,q,s} \in \mathbb{N}]_{n_0 \times q_0 \times s_0}$.

Factor analysis with Varimax rotation was conducted along the r axis of $X_{r,q,s}$ to get the first two factors $f^{(1)} = [f_r^{(1)}]_{n_o \times 1}$ and $f^{(2)} = [f_r^{(2)}]_{n_o \times 1}$ of each place s_r. The two factors explained 76.1% of the variance, suggesting that they could represent the cognitive dataset. The first factor $f^{(1)}$ was mainly comprised of answers indicating that places were representative, meaningful and symbolic, forming a '**Representing Axis**' of the data; while the second factor $f^{(2)}$ was mainly about places that were crowded and attractive for people, forming a '**Gathering Axis**'. Details of the interviews,

mentioned places and factor analysis are, respectively, shown in Supplemental Materials 03, 04, 09, Figure S2 and Tables S2-3.

Furthermore, the total number that each place $\boldsymbol{r}_r$ was mentioned by both groups were computed to get the votes for the place by stakeholders, similar to Kim (1999)

$$\boldsymbol{v}^{(\boldsymbol{s}_s)} := \left[v_r^{(\boldsymbol{s}_s)} = \sum_{q=1}^{|q\_0|} x_{r,q,s} \right]_{n_0 \times 1}, \quad s = \{0, 1\},$$

where $\boldsymbol{v}^{(\boldsymbol{s}_0)}$ stands for the vote of locals, and $\boldsymbol{v}^{(\boldsymbol{s}_1)}$ stands for that of non-locals.

Behaviour – Wi-Fi position tracking

Classical space syntax is a good model to predict traffic flow as it focuses on primal line syntax (Major et al., 1999; Major et al., 2000; Ma et al., 2018; Penn, 2003). Switching to dual point syntax provides possibilities to regard public spaces as locations to stay at in instead of merely to pass through (Batty, 2017; Gehl, 2011; Zhang et al., 2020). The staying activities provide meaningful information of the public spaces in traditional villages like AHVC, which can be either where locals gather daily for chatting, or where tourists visit occasionally for sightseeing. To capture the crowd staying behaviour in open public spaces, Wi-Fi Position Tracking has proved effective for its balance between applicability and accuracy (Huang et al., 2018; 2019; Lin and Huang, 2018; Mashhadi et al., 2016; Sapiezynski et al., 2015).

In this sub-study, a Wi-Fi positioning system was set up in AHVC under the IEEE 802.11 protocol, following previous studies (Huang et al., 2017 2017; Huang et al., 2018; Lin and Huang, 2017, 2018). A set of Wi-Fi Access Points (APs) were installed at pre-selected locations to record the presence of Wi-Fi-enabled mobile devices (as seen in Supplemental Figure S3). The record dataset $\boldsymbol{D}$ updated every 5 min, capturing Access Control Address (MAC) near the APs at least once during the time window. All MAC-addresses were strictly anonymized and desensitized as data privacy protection measures. A subset $\mathcal{B}_0 = \{\boldsymbol{r}|r \in [0, n_1)\}$, $\mathcal{B}_0 \subset (\mathcal{C} \cap \mathcal{P})$ of perceived places was chosen as locations to install the AP network, where monitoring devices were all anchored deliberately at street junctions. The system monitored AHVC for 56 days from Jan 28th to Mar 24th, 2019. Five out of 32 devices failed to capture data due to power issues, leaving the monitored places to a smaller subset $\mathcal{B} = \{\boldsymbol{r}|r \in [0, n_2)\}$, $\mathcal{B} \subset \mathcal{B}_0$.

Data pre-processing (as described in Supplemental Material 06 and Figures S4-5) was conducted to distinguish three groups of people (non-locals, locals and others) based on the MAC-addresses' presence frequency. Following Lin and Huang (2018), the arrival-departure pattern can be plotted for each group as 2D heatmaps in Figure 2. Clear patterns can be observed: (1) points along the diagonal represented the moving behaviour when people passed by; (2) points parallel to the main diagonal demonstrated the sightseeing behaviour when people stayed for a short time and (3) points distributed in the top left part of the plots showed the staying behaviour when people came and stayed for a longer period. As the patterns of locals and non-locals are clearer and fit better to the research question, further analyses will only consider these two groups.

After pre-processing, the number of both groups of people $\boldsymbol{s}_s$ at each AP position $\boldsymbol{r}_r$ at each time window τ_t in database $\boldsymbol{D}$ was aggregated to construct the matrix $\boldsymbol{Y}_{r,t,s}$

$$\boldsymbol{Y}_{r,t,s} := \left[y_{r,t,s} \in \mathbb{N} \right]_{n_2 \times \tau_m \times s_0}, \quad \text{where } \tau_m = 288 \times 56 = 16,128$$

As not all devices managed to capture data all the time, another matrix $\boldsymbol{T}_{r,t,s} := [t_{r,t,s}]_{n_2 \times \tau_m \times s_0}$ was constructed to mask the influence of missing data
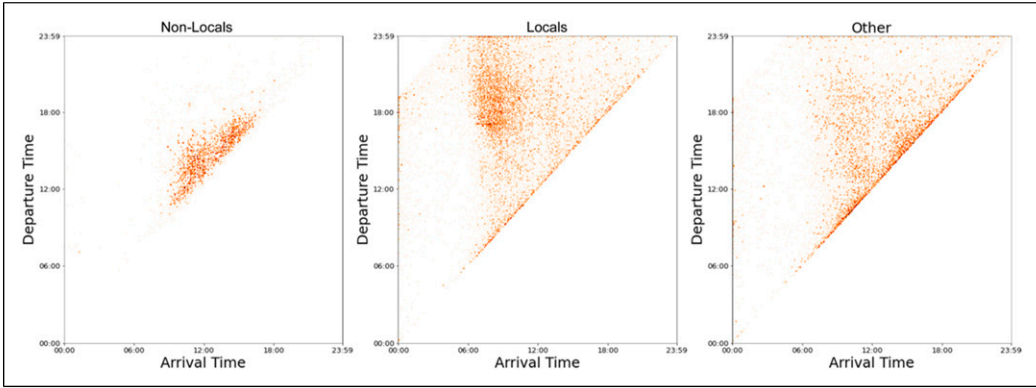
Figure 2. The arrival-departure time plot of the three labelled groups of people: left for non-locals, middle for locals, right for others.

$$t_{r,t,s} = \begin{cases} 0, & if \quad y_{r,t,s} = 0 \\ 1, & otherwise \end{cases}.$$

Analysis of Variance (ANOVA) was conducted on $Y_{r,t,s}$ to distinguish the places that generally attracted more people during the monitoring period. The matrices $Y_{r,t,s}$ and $T_{r,t,s}$ were then aggregated into four vectors $u^{(s_0)}, u^{(s_1)}, w^{(s_0)}, w^{(s_1)}$:

$$u^{(s_s)} := \left[u_r^{(s_s)} = \frac{\sum_{t=1}^{\tau_m} y_{r,t,s}}{\sum_{t=1}^{\tau_m} t_{r,t,s}} \right]_{n_2 \times 1}, \; s = \{0,1\}$$

$$w^{(s_s)} := \left[w_r^{(s_s)} = \frac{\sum_{t=1}^{\tau_m} t_{r,t,s}}{\sum_{t=1}^{\tau_m} V_{s=1}^2 t_{r,t,s}} \right]_{n_2 \times 1}, \; s = \{0,1\}$$

where $u^{(s_s)}$ refers to the average count of people appeared at each place in a time window, and $w^{(s_s)}$ refers to the percentage of the time windows a certain group of people appeared at each place over the total number of time windows this place managed to captured data at all.

Synthesis of information sources

All the previously listed variables from each source of information could be referred to in Supplemental Table S1. The following key indicators were selected to investigate their relations:

1. For morphological analysis, four dual-syntax centralities $p^{(\gamma)}$ and bipartite-graph PageRank $a^{(\text{PgRk})}$ were selected, since: (1) the AP devices were all located at street junctions and this study focused on space occupation rather than movement; (2) bipartite graph containing both streets and junctions as a superset yielded chances to compare the centralities of points and lines at once and (3) bipartite-graph PageRank was the only centrality index that had a non-trivial correlation with dual-syntax indices (as shown in Supplemental Figure S10), since the random walk setting was different.
2. For cognitive mapping, the factors $f^{(1)}, f^{(2)}$ and the votes of both groups $v^{(s_s)}$ were included.
3. For behavioural studies, the average counts $u^{(s_s)}$ and appearance percentages $w^{(s_s)}$ were considered.

As hypothesis, the behavioural indicators could be characterized and predicted with indicators from the other two sources. A regression model for the local count $\boldsymbol{u}^{(s_0)}$, for example, could be formulated as

$$\boldsymbol{u}^{(s_0)} = \beta_0 + \beta_1 \boldsymbol{p}_z^{(\text{deg})} + \beta_2 \boldsymbol{p}_z^{(\text{cls})} + \beta_3 \boldsymbol{p}_z^{(\text{betwn})} + \beta_4 \boldsymbol{p}_z^{(\text{PgRk})}$$
$$+ \beta_5 \boldsymbol{a}_z^{(\text{PgRk})} + \beta_6 \boldsymbol{f}_z^{(1)} + \beta_7 \boldsymbol{f}_z^{(2)} + \beta_8 \boldsymbol{v}_z^{(s_0)} + \beta_9 \boldsymbol{v}_z^{(s_1)} + \epsilon$$

All independent variables were standardized into z-scores (subtracted with means and divided by SDs within set $\mathcal{B}$) to make the coefficients directly comparable.

Furthermore, as the dual-syntax centralities $\boldsymbol{p}^{(\gamma)}$ and the bipartite-graph PageRank $\boldsymbol{a}^{(\text{PgRk})}$ might have complex nonlinear effects since the latter also contains structural information of streets, interaction terms (Wooldridge, 2013) as Hadamard product of vectors were added in alternative models, such that

$$\boldsymbol{u}^{(s_0)} = \beta_0 + \ldots + \beta_{1\&5} \boldsymbol{p}_z^{(\text{deg})} \dot{\mathbf{c}} \boldsymbol{a}_z^{(\text{PgRk})} + \beta_{2\&5} \boldsymbol{p}_z^{(\text{cls})} \dot{\mathbf{c}} \boldsymbol{a}_z^{(\text{PgRk})}$$
$$+ \beta_{3\&5} \boldsymbol{p}_z^{(\text{betwn})} \dot{\mathbf{c}} \boldsymbol{a}_z^{(\text{PgRk})} + \beta_{4\&5} \boldsymbol{p}_z^{(\text{PgRk})} \dot{\mathbf{c}} \boldsymbol{a}_z^{(\text{PgRk})} + \epsilon$$

The indicators were first compared pair-wise with Pearson Correlation using SciPy python library before regression models were constructed, where instances were omitted when at least one variable contained empty values. Ordinary Least Squares (OLS) was used to estimate the regression models with **Statsmodel** Python library. The independent variables can have multicollinearity due to correlations, yielding redundancy in regression models. Therefore, backward elimination method with Bonferroni correction was applied to select optimal models with significant variables, compensating the multiple tests (Abdi, 2007; Green et al., 2001; Mundfrom et al., 2006). The statistical details of the independent and dependent variables are listed in Supplemental Table S7.

Making behavioural data dependent variables in regression is a natural choice since the ultimate goal of rural [re]vitalization and sustainable heritage planning is arguably to stimulate the human activities in public spaces, to engage stakeholders through active usage, and therefore to enhance human-centred heritage (Janssen et al., 2017; Ginzarly, Pereira Roders and Teller, 2019). It is valuable to both investigate the interpretability of relevant indicators for predicting behavioural patterns, and compare predictions with observations (Wooldridge, 2013).

Since the correlation and regression analyses only investigated places in set $\mathcal{C}$ and/or $\mathcal{B}$, additional t-tests were conducted on centrality indices $\boldsymbol{l}^{(\gamma)}$, $\boldsymbol{p}^{(\gamma)}$ and $\boldsymbol{a}^{(\text{PgRk})}$ between places in $\mathcal{C}$ and $(\mathcal{P} \cup \mathcal{L})\backslash\mathcal{C}$ (please check Supplemental Figure S6 for a Venn Diagram showing the relations among the sets of places), to check the general configurational features of the cognitively significant places in AHVC, compared to the ones not mentioned.

Results

Evidence distinguishing locals and non-locals

Cognitive mapping enables the distinction of perceptions between locals and non-locals. The behavioural tracking, on the other hand, can distinguish the behavioural patterns of different groups by observing the raw occupation data. However, although intended by introducing centrality indices within different radii and depths, such distinctions cannot be directly derived from morphological analyses.

K-means clustering on the two cognitive factors $\boldsymbol{f}^{(1)}$ and $\boldsymbol{f}^{(2)}$ (as shown in Figure 3 left) found three clusters in the set C of nominated places: **Central Places** (7 cases) that were more representative and symbolic than others but not able to gather people for a long time; **Gathering Places**
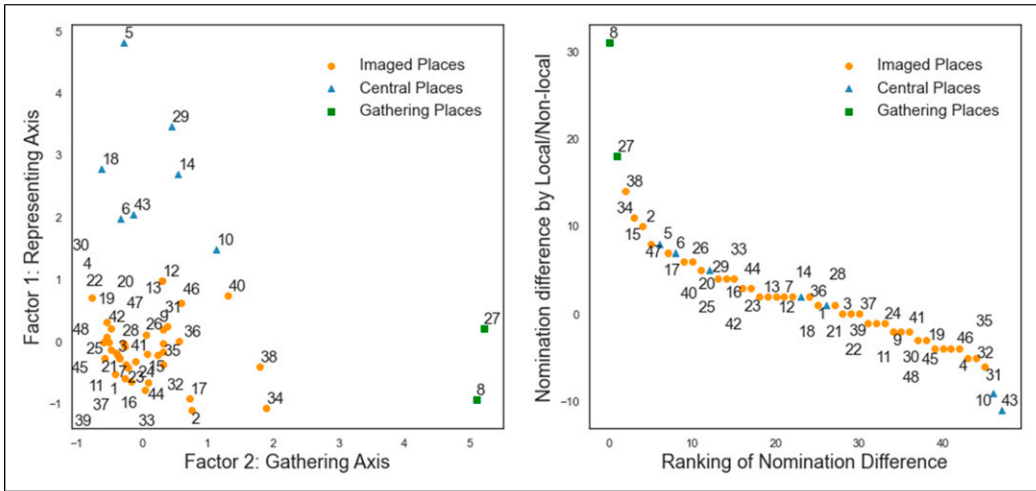
Figure 3. Left: the two main factors $f^{(1)}$ and $f^{(2)}$ of all three clusters in mentioned places; Right: the rank-size of the nomination difference $v^{(s_0)} - v^{(s_1)}$ between locals and non-locals. The points were clustered and labelled based on K-means clustering.

(2 cases) that were popular destinations for people to gather and spend their days, and the remaining **Imaged (Perceived) Places** (39 cases) which were mentioned but not prominent in both axes. ANOVA with *F*-tests showed that the three clusters were significantly different in both factors $[F_1(2, 45) = 64.278, p < .001 , \eta^2 = .741, F_2(2, 45) = 66.274, p < .001, \eta^2 = .747]$ (check Supplemental Tables S4-5 for *post-hoc* analyses). Furthermore, the rank-size distribution of the nomination difference of each place $v^{(s_0)} - v^{(s_1)}$ is shown in Figure 3 right. The places on both tails of the distribution had different significance for locals and non-locals. For example, the two Gathering Places r_8 (The Lotus Pond) and r_{27} (The Front Street) were mentioned far more by local villagers; and the two Central Places r_{43} (Xilu Academy) and r_{10} (The Back Street) were valued more by non-local planners. This result indicates that the locals and non-locals have different perceptions on the places.

Inspecting the behavioural patterns at the monitored places, clear distinctions between locals and non-locals could also be observed. The average occupation patterns of both groups on a typical day for a subset of exemplary APs are shown in Figure 4 (the complete temporal patterns of all APs could be found in Supplemental Figures S11-12). For some APs (for example, LT04 located at r_8 the Lotus Pond), locals always dominantly occupied the place. Yet for many other places, locals and non-locals shared the space equally. This pattern also resembled the rank difference of nominations in Figure 3 right. Furthermore, for non-locals (probably visitors), a single peak would usually appear at noon or in early afternoon (11 a.m. to 3 p.m., varying from APs), reflecting the visitation mode and suggesting that there seemed to be a preferred touring sequence. And for the locals, normally two or three peaks would appear in the morning (11 a.m.), late afternoon (4 p.m.) and early evening (8 p.m.), reflecting the local rural habit to wander, chat and play cards before/after lunch and after dinner.

Ability to identify places of cultural significance

As the nature of three sources of information differs, 'significant' places have different meanings. For morphological analysis, places with higher centrality values reflect the connectivity of the
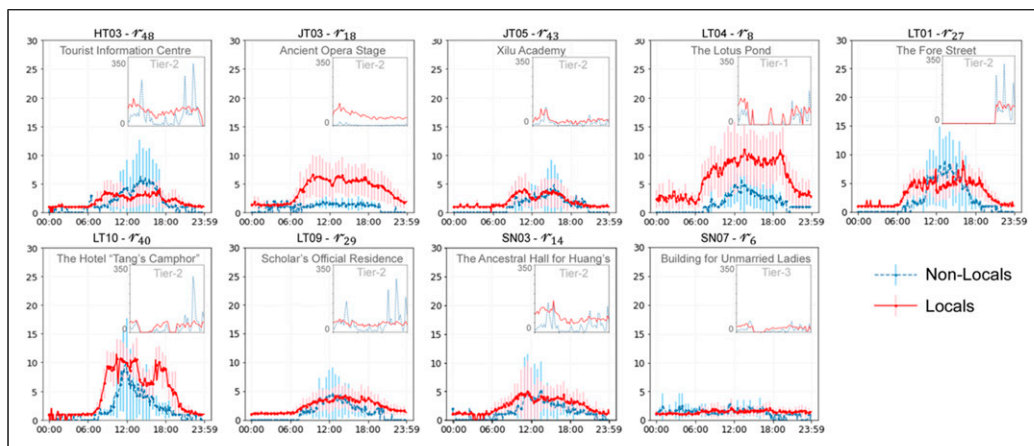
Figure 4. The average number of people present at each AP for every time window in a 'typical' day. The solid curves showed the average numbers, while the shades showed the standard deviations. The daily numbers of captured people from Jan 28 to Mar 24 were plotted in the subgraphs, showing the peaks of non-locals during weekends and the relatively steady curves of locals. For brevity, only a typical subset of $\mathcal{B}$ (9 spots) was demonstrated here.

spatial structure in different senses; for cognitive mapping, places mentioned more frequently by people indicate stronger cognitive associations; and for behavioural studies, places that attract more people for a longer time imply that they are more popular and actively used. However, the term 'significant' (not to confuse with the statistical term) used here was only a common denominator and did not intend to generalize and/or deviate from the real meanings in each context.

Figure 5(a)-(c) mapped some centrality indices that proved to have influence in regression models, while Figure 5(d) showed the cognitive mapping of all participants in the research. The places with higher centrality values included the main streets, squares and entrances to each village in AHVC, overlapping with the places nominated by people in the interview (the complete mapping of centrality indices in all three types of graphs could be found in Supplemental Figures S7-10). The *t*-tests between the nominated places and other places in Table 1 showed significant differences (with moderate or large Cohen's d_s effect sizes) in all junction-based dual-syntax centrality indices and most street-based primal-syntax indices, suggesting that places perceived, remembered and 'imaged' by people in AHVC were generally also located at the spatially central places. However, this does not promise the other way around.

Furthermore, the average number of people captured by each AP for each time window in a typical day formed three tiers of crowdedness (see Supplemental Table S2). ANOVA with *F*-test showed significant difference among the tiers $[F(2, 6336) = 1075.584, p < .001, \eta^2 = .254]$ (check Supplemental Table S6 for *post-hoc* analyses), suggesting that the tier-1 place r_8 (The Lotus Pond) was generally more actively used, supporting the gathering factor $f^{(2)}$ from cognitive data.

Comparative analyses

The correlation analysis among the indicators showed strong associations between all three information sources, as shown in Figure 6. Intra-group correlations between variables from the same source were generally higher than inter-group ones across the sources. The correlation patterns among behavioural indicators and the other two sources supported the construction of regression models.

Table 2 displayed the estimated regression models for the four behavioural indicators as dependent variables with OLS. Models (1) and (3) only used one independent variable, and,
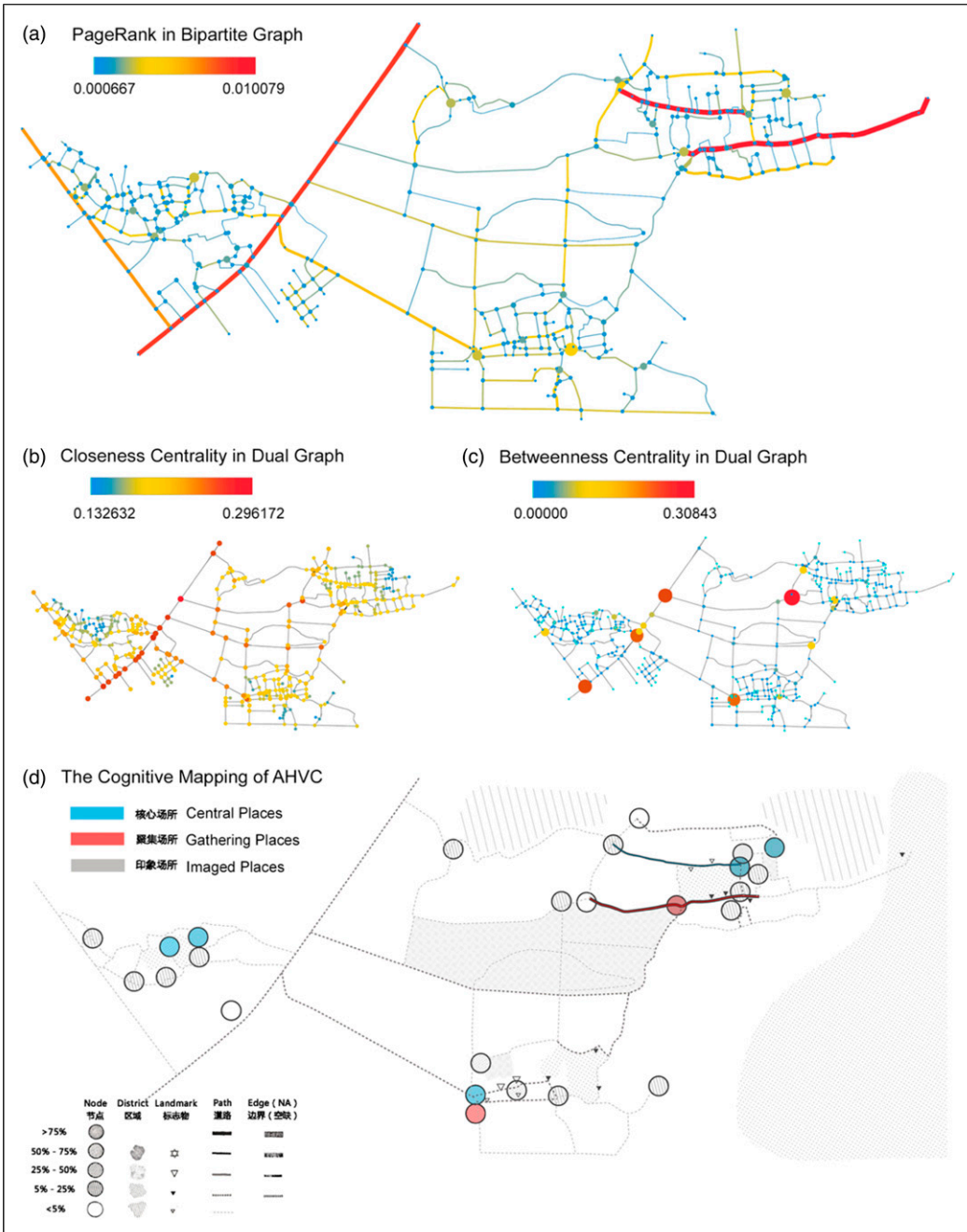
Figure 5. The spatial centralities and cognitive mapping of AHVC. a) PageRank in the bipartite graph; b) closeness centrality in the dual point syntax; c) betweenness centrality in the dual point syntax; d) the cognitive mapping of all participants. For space syntax graphs, the size of the points, the width of the lines, and colours showed the centrality indices. For cognitive mapping, the spatial elements were plotted based on Lynch's rules.

Table 1. The result of Welch's t-tests on the centrality indices between the nominated places and the places not mentioned. Welch's t-tests were used since the samples have unequal variance based on Levene test. The significance levels were adjusted based on Bonferroni correction to compensate multiple comparison.

	$I^{(deg)}$	$I^{(cls)}$	$I^{(betwn)}$	$I^{(PgRk)}$	$a^{(PgRk)}$	$p^{(deg)}$	$p^{(cls)}$	$p^{(betwn)}$	$p^{(PgRk)}$	$a^{(PgRk)}$
t-statistics	3.565*	2.011	2.120	3.522*	3.040*	9.641***	5.457***	3.920***	8.128***	3.308**
Cohen's d_s	1.214	0.684	0.722	1.200	1.035	1.023	0.579	0.416	0.863	0.351
Counts	Places in $\mathcal{L} \cap \mathcal{C}$ ($n = 9$) versus Places in $\mathcal{L} \setminus \mathcal{C}$ ($n = 209$)					Places in $\mathcal{P} \cap \mathcal{C}$ ($n = 123$) versus Places in $\mathcal{P} \setminus \mathcal{C}$ ($n = 319$)				

* $p < \frac{.05}{5} = .01$; ** $p < \frac{.01}{5} = .002$; *** $p < \frac{.001}{5} = .0002$
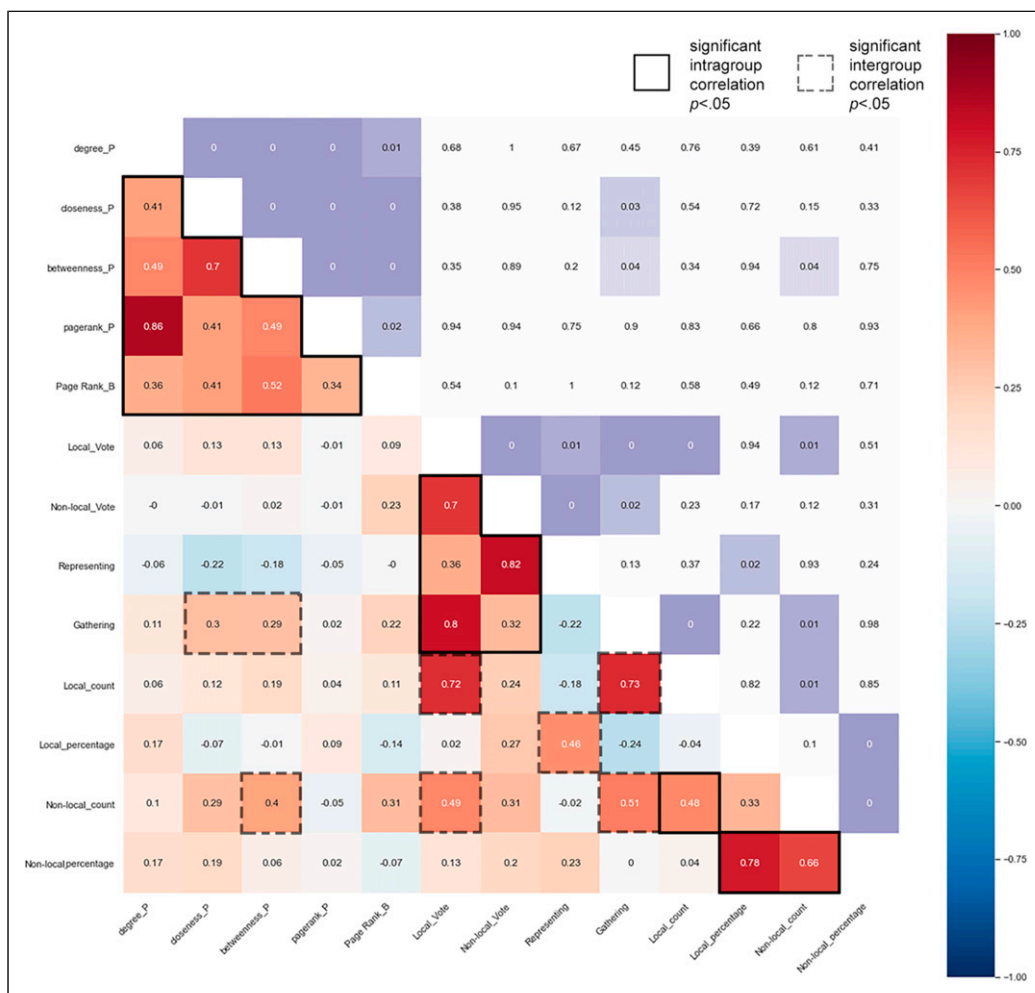

Figure 6. The result of correlation analysis within key indicators from all three sources. The lower triangle showed the r-values, while the upper triangle showed the p-values. The significant correlations were highlighted.

Table 2. OLS Results for coefficients' estimations. Only independent variables that managed to enter regression models at least once were presented. While interaction terms were significant, the associated [insignificant] main variables were also kept. The standard errors for coefficients' estimations were shown in parenthesis. For all models, 25 observations were used. The significance levels for coefficients and F-statistic were adjusted with Bonferroni correction (with, respectively, 13 and 9 variables for models with and without interaction terms).

Dependent variables	$\widehat{u}^{(s_0)}$ –local count		$\widehat{u}^{(s_1)}$ –non-local count		$\widehat{w}^{(s_0)}$ –local percentage	$\widehat{w}^{(s_1)}$ –non-local percentage
Independent variables	(1)	(2)	(3)	(4)	(5)	(6)
$p_z^{(deg)}$ – degree in **P**					0.11 (0.05) †	0.09 (0.03) *
$p_z^{(cls)}$ – closeness in **P**		−0.72 (0.24) *			−0.03 (0.05)	0.02 (0.01) †
$p_z^{(betwn)}$ – betweenness in **P**				0.20 (0.26)		
$p_z^{(PgRk)}$ – PageRank in **P**		0.65 (0.23) †		0.18 (0.24)		−0.13 (0.05) †
$a_z^{(PgRk)}$ – PageRank in **A**		−0.59 (0.22) †		−0.18 (0.20)	−0.07 (0.05)	0.04 (0.03)
$f_z^{(1)}$ – representing factor			0.67 (0.20) **			0.02 (0.01) †
$f_z^{(2)}$ – gathering factor					−0.25 (0.08) *	
$v_z^{(s_0)}$ – local votes	1.11 (0.26) ***	1.61 (0.24) ***		0.79 (0.21) **	0.27 (0.08) **	
$p_z^{(cls)} \odot a_z^{(PgRk)}$				1.27 (0.33) **	0.12 (0.05) †	0.06 (0.01) ***
$p_z^{(betwn)} \odot a_z^{(PgRk)}$				−0.94 (0.30) *		
Constant	3.07 (0.25) ***	3.07 (0.20) ***	2.43 (0.20) ***	2.50 (0.17) ***	0.37 (0.04) ***	0.09 (0.01) ***
R^2	.449	.695	.323	.666	.519	.598
Adjusted R^2	.425	.634	.294	.555	.358	.464
AIC	84.11	75.34	72.69	65.01	−5.82	−75.60
F- statistic	18.730***	11.380***	10.990**	5.993**	3.234 †	4.459*
P-value for F-statistic	.00025	.00001	.00301	.00138	.02460	.00619

† $p < .05$, * $p < \frac{.05}{9} = .006$, ** $p < \frac{.05}{13} = .004$, *** $p < \frac{.01}{13} = .0008$.

respectively, reached moderate Adjusted R^2 for both predicants, showing that cognitive indicators local votes $v^{(s_0)}$ and gathering factor $f^{(2)}$ were good predictors for average local and non-local counts. However, adding morphological indicators significantly improved the predictive power in models (2) and (4) with higher Adjusted R^2, larger than .55. Moreover, an additional sensitivity check with alternative regression models is given in Supplemental Material 14 and Table S8, which proves to be similar to the results presented in this section.

For local counts, other than the most influential variable local votes with the highest estimated standardized coefficient ($\widehat{\beta}_8^{(2)} = 1.61$, CI $= [1.11, 2.12]$), bipartite-graph PageRank index $a^{(\text{PgRk})}$ also contributed positively, while closeness centrality $p^{(\text{cls})}$ and representing factor $f^{(1)}$ had negative effects. This suggests that the closeness of spots to other places and the representativeness of spots were less influential for locals given that the spots have a crucial status in the street-junction bipartite network: that is, since the villages were relatively small and familiar for locals, they did not have to stay at somewhere specifically close or symbolic, but rather at natural random-walk destinations.

For non-local counts, however, the inclusion of interaction terms and local votes $v^{(s_0)}$ made the coefficient of gathering factor $f^{(2)}$ insignificant, as a result of the high correlation between the latter two ($r = .73$, $p < .001$). The interaction of dual-syntax closeness and bipartite-graph PageRank contributed most significantly for the prediction ($\widehat{\beta}_{2\&5}^{(4)} = 1.27$, CI $= [.57, 1.97]$), suggesting that the spots being both natural destinations and close to other spots would gain more visitors. Similarly, the negative interaction effect of dual-syntax betweenness and bipartite-graph PageRank ($\widehat{\beta}_{3\&5}^{(4)} = -0.94$, CI $= [-1.58, -0.31]$) suggested that visitors were likely to stay at places that were natural destinations but not 'bridge connections' among the villages.

Models (5) and (6), predicting the percentage of time windows when people were present at the spots, were both [marginally] significant based on Bonferroni-corrected F-statistic. Similar to average counts, local votes $v^{(s_0)}$ was most influential for local percentage ($\widehat{\beta}_8^{(5)} = 0.27$, CI $= [0.11, 0.44]$), and the interaction of dual-syntax closeness and bipartite-graph PageRank for non-local percentage ($\widehat{\beta}_{2\&5}^{(6)} = 0.05$, CI $= [0.03, 0.09]$). Furthermore, dual-syntax degree $p^{(\text{deg})}$ contributed positively to both percentages, and the representing factor $f^{(1)}$ had a weak effect on non-local percentage. This suggests that even though the symbolic values did not necessarily attract more non-locals/visitors for any specific moment as in model (4), it did attract them more frequently during the monitored periods. The constants $\widehat{\beta}_0^{(5)}$ is much higher than $\widehat{\beta}_0^{(6)}$, meaning that the locals appeared at the spots more frequently than the non-locals, consistent with the visiting pattern of visitors during weekends, as demonstrated in Figure 4 and Supplemental(s)11.

Interestingly, the non-local votes $v^{(s_1)}$ entered none of the models. This was partially due to the collinearity shown in Figure 6. Yet it also suggests that the locals, rather than the non-locals, revealed to know the villages better, and their perception could be a better predictor for the actual usage of public spaces.

Since the models reached high Adjusted R^2, they could be assumed to have good fit and fair prediction powers. The predicted values $\widehat{u}^{(s_s)}$ and $\widehat{w}^{(s_s)}$ could characterize the theoretical popularity of the spots (please check Supplemental Figure S13 for the predicted values), while the residuals would suggest the mismatch between estimation and observation, as shown in Figure 7. The configuration of spots' residuals of both locals and non-locals could reveal 'over-estimated/under-used' spots located in the bottom-left quadrants that needed further improvement, and 'under-estimated/well-used' spots located in the top-right quadrants being more attractive than estimation. For example, r_{43} (the Xilu Academy) and r_6 (Building for Unmarried Ladies) were public places that both occupied central locations and raised awareness among locals, yet attracted significantly less people than they should, suggesting that the local government should strengthen these spots in
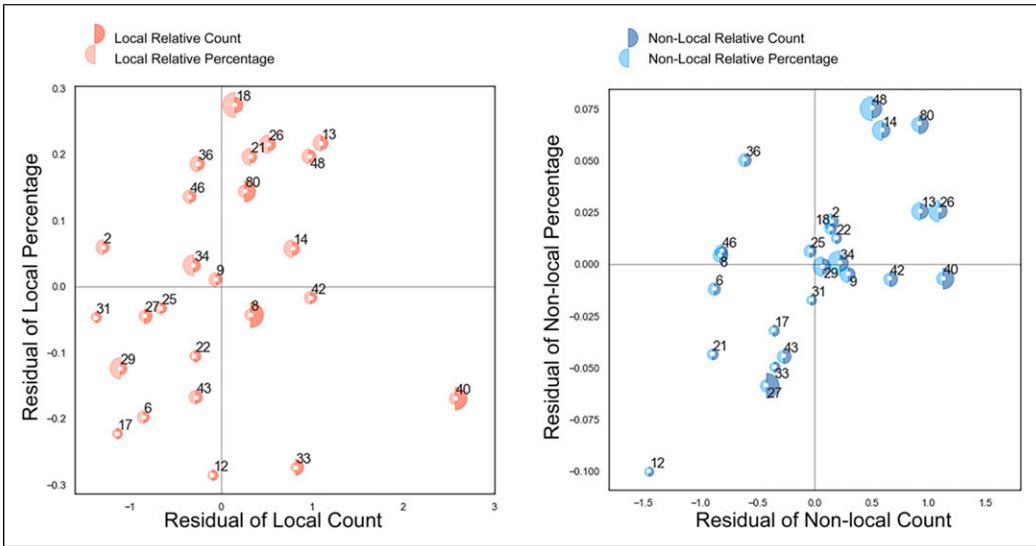
Figure 7. Residual Analysis of the regression models as scatter plots. X-axes showed $u^{(S_s)} - \widehat{u}^{(S_s)}$ for both local (left) and non-local (right) counts, and Y-axes showed $w^{(S_s)} - \widehat{w}^{(S_s)}$ for local and non-local percentages, while the sizes of the indexed points showed the relative popularity of that spot $\left(\frac{u^{(S_s)}}{max(u^{(S_s)})} \text{ or } \frac{w^{(S_s)}}{max(w^{(S_s)})}\right)$

[re]vitalization plans. On the other hand, r_{80} (the Lotus Pond) and r_{14} (The Ancestral Hall for Huang's) attracted more people than estimation, suggesting that they might be well-designed public spaces to learn from. However, the residuals should be interpreted together with the actual numbers of people: although r_{27} (The Fore Street) and r_{29} (Scholar's Official Residence) also had high residuals, they were less worrisome since they already attracted many people (indicated with the point size in Figure 7). Moreover, planners are suggested to conduct site-specific investigations on the spatiotemporal patterns to give detailed explanations on high residuals.

Discussions

Planning support recommendations

Following the presented methodological framework, the outcomes could be utilized in support of strategic spatial planning in similar historic villages with cultural significance:

1. The different needs of locals and non-locals should be identified, respected and balanced, especially when constant changes are on-going. During the past decades, many rural public spaces have been implanted with fancy attractions, disregarding the essential needs of local residents to use public spaces to maintain their lifestyle, for example, to chat, dance and/or play cards. A successful development plan should consider both visitors and residents, allowing for a symbiosis of stakeholders without disturbing or destroying local needs. Places that are popular tourist destinations do not necessarily attract local people, and places where locals spend most of their time do not always draw planners' and visitors' attention. Both types can be identified as public places with cultural significance, and should be handled with different planning strategies. This study provides empirical tools to distinguish the 'central' and 'gathering' places used and

valued differently by both stakeholders, following a systematic analytical workflow that could also be applied elsewhere, possibly armoured with more effective technologies in the future.

2. Multiple sources of information focus on divergent aspects of 'significant' public spaces and they all ought to be recognized by planners. In this study, systematic methods have been proposed to capture three sorts of 'significant' public spaces: places with high centralities in different morphological configurative contexts, 'central' and 'gathering' places perceived cognitively by people, and frequently-used 'tier-1' places standing out in behavioural monitoring. The comparative synthesis showed that the sources are related yet divergent. This refers back to earlier philosophical discussions on the triad of 'conceived/perceived/ lived' spaces (Lefebvre, 1991). Places such as Lotus Pond could be valued as vivid living spaces though it might not be considered as valuable in conventional heritage perspectives. The finding that cognitively imaged places also appeared at spatially central spots suggests that initial morphological analyses (especially on dual-syntax and bipartite graphs) in future studies could effectively provide sets of potentially significant spaces, to be followed by cognitive mapping and behavioural tracking for a comprehensive understanding.

3. The evaluation of behavioural occupation patterns could become a diagnosis tool for public spaces. Through occupying and using the spaces, emotional connections could be built and/or strengthened, helping locals and visitors better convey the cultural significance within the spaces, beneficial for both heritage preservation and planning development (Janssen et al., 2017). By combining and comparing the observational popularity of places with theoretical estimations, one may suggest priorities in development through investigating the detected 'under-used/well-used places' one by one. Rural [re]vitalization plans could use such diagnosis tools and seek for suitable strategies to help determine what to preserve, renovate and demolish in traditional villages, while the spatial structure and human perception are non-trivial to change directly in a planning process.

Towards a general theory

Although a totally different set of indicators (variables) were selected from the three information sources, measured in case-studies from various cultural contexts, this study confirms most conclusions from Kim (1999), especially that cognitive variables are better predictors for behaviour
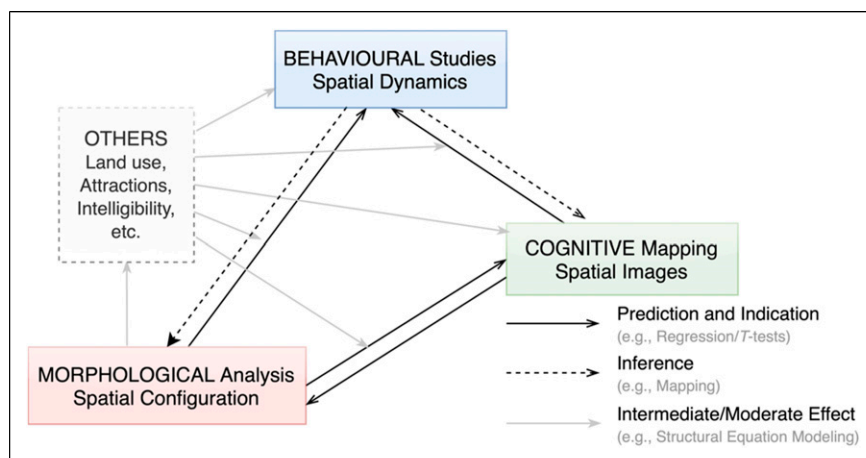

Figure 8. The theoretical framework characterising the interrelation among morphology, cognition and behaviour. The diagram is adapted and updated from Kim (1999).

than spatial configuration. This is to suggest that there might exist a general theory to explain the relationship between human behavioural patterns, complex morphological structure of space and its cognitive representation, as demonstrated in Figure 8. This study mainly investigated the solid arrows in the diagram, while the remainders could also be scientifically interesting:

1. Would other factors such as demographics, attractions, land use, spatial intelligibility, etc., have intermediate and/or moderate effects on the complex interrelation (Penn, 2003)?
2. Given a behavioural pattern distribution effectively collected with digital technologies, could the underlying spatial/social structure and human preference be inferred, as a reversed process to the conventional prediction and modelling of human behaviour (Noyman et al., 2019)?

To approach a general theory, more cases are needed following the methodological framework proposed by this pilot study, probably using different time periods and economic status as control variables.

Limitation and future steps

Although we meant to be neutral without preconception during the research, especially during the interview, biases to value certain typology could have been introduced unintentionally. However, the accordance of the chronologically first-conducted interview with other analyses added reliability to the proposed methodology. The models presented might suffer from over-fitting to the collected dataset, while the sensitivity check in Supplemental Table S8 suggests that they are relatively robust. With more cases, machine learning techniques such as cross-validation, Lasso-regression and Elastic-Net could be applied to perform variable selection while avoiding over-fitting, providing more concrete proofs on the generalizability of the models. Furthermore, specific techniques have to be adopted if a concrete causality inference is desired in future studies, in order to suggest changes in planning actions.

Moreover, the following aspects can be improved: (1) network centrality indices adjusted by other distance measures may contribute to a more precise evaluation of spatial morphology; (2) a wider range of stakeholders could be involved; (3) Wi-Fi devices could be supplemented with other spatiotemporal data collected by GPS and/or wearable devices; (4) horizontal and vertical comparative studies for rural [re]vitalization projects would help monitor and assess the changing dynamics and (5) time series as dynamic panel data instead of aggregated values could be used to characterize the behavioural observations in higher resolution, benefitting more from the collected big data.

Conclusion

This study proposed new methods for processing three sources of information to draw a comprehensive picture of the rural public spaces with cultural significance in Anyi Historic Village Cluster (AHVC), suggesting that the spaces are valued and used differently for various stakeholders. The three approaches – morphology, cognition and behaviour – proposed and jointly evaluated were strongly interrelated and provided useful information for spatial planning, consistent with the rich urban study cases in the literature. The complex relationship of the spatial structure, the valuation and usage of spaces in AHVC was characterized with regression models, providing tools to identify places to prioritize in planning through comparing estimation and observation. This could only be achieved by integrating all three sources of information, such as the methodological framework proposed here.

Rural [re]vitalization is ongoing in China, continuously changing the space quality in villages. This study provides planners with a methodological framework to systematically investigate important rural public spaces, both as perceived and used places that need to prioritize in planning. This approach can assist planners to make more comprehensive village development plans, while respecting the local needs and rural identities.

Acknowledgements

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

Author's Note

The advice about statistical methods and reporting from Professor Pei Sun, the Department of Psychology, Tsinghua University, the language support from Sören Johnson, Centre for Languages and Academic Skills, TU Delft, and the help on the initial data processing algorithm for Wi-Fi positioning data from Yuming Lin and Yu Zhang, Tsinghua University, are gratefully appreciated. The field work in AHVC could not be conducted without the support from Yihui Zheng, the local management team from Anyi, and Rural Vitalization Station of Tsinghua University. This paper is based on Nan Bai's master thesis in Chinese at Tsinghua University, Beijing, China, and Technische Universität Berlin, Berlin, Germany. A preliminary version of the cognitive mapping sub-section was presented (unpublished) at the Leiden-Delft-Erasmus (LDE) joint Conference "*Heritage and the Sustainable Development Goals*", November 26-28, 2019, Delft, the Netherlands.

ORCID iDs

Nan Bai 🔘 https://orcid.org/0000-0001-7637-3629
Pirouz Nourian 🔘 https://orcid.org/0000-0002-3817-7931
Weixin Huang 🔘 https://orcid.org/0000-0002-3278-4490
Ana Pereira Roders 🔘 https://orcid.org/0000-0003-2571-9882

Supplemental Material

Supplemental material for this article is available online.

References

Abdi H (2007) *'The Bonferonni and Šidák Corrections for Multiple Comparisons'*. Thousand Oaks, CA, Encyclopedia of Measurement and Statistics, 103–107. DOI: 10.4135/9781412952644.

Agryzkov T, Curado M, Pedroche F, et al. (2019) 'Extending the adapted PageRank algorithm centrality to multiplex networks with data using the PageRank two-layer approach', *Symmetry*, 11(2): 1–17. DOI: 10. 3390/SYM11020284.

Agryzkov T, Tortosa L, Vicent F José, et al. (2019) 'A centrality measure for urban networks based on the eigenvector centrality concept', *Environment and Planning B: Urban Analytics and City Science*, 46(4): 668–689. DOI: 10.1177/2399808317724444.

Anyi County Archive Committee (2007) *County Archive for Anyi 1986-2000 (in Chinese)*. Xi'an, China: Sanqin Publication.

Bafna S (2003) 'Space syntax: A brief introduction to its logic and analytical techniques', *Environment and Behavior*, 35(1): 17–29. DOI: 10.1177/0013916502238863.

Bai N and Wang L (2019a) 'Village heritage conservation in the new data age'. In: Intelligent and Informed-Proceedings of CAADRIA 2019, Wellington, New Zealand, 15–18 April 2019.

Bai N and Wang L (2019b) 'Village Heritage Preservation in the New Data Age: Rereading and Reinterpreting the Image of the Village'. In: International Conference on Built Heritage Studies.

Bandarin F and Van Oers R (2012) *The Historic Urban Landscape, The Historic Urban Landscape*. Chinchester, NY: WILEY-BLACKWELL. DOI: 10.1002/9781119968115.

Batty M (2004) *A New Theory of Space Syntax*. Working Paper No. 75. London, UK: Center for Advanced Spatial Analysis.

Batty M (2013) *The New Science of Cities*. Cambridge, MA: MIT Press.

Batty M (2017) *Space Syntax and Spatial Interaction: Comparisons, integrations, Applic*. Working Paper No. 208. London, UK: Center for Advanced Spatial Analysis.

Bazeley P (2007) *Qualitative Data Analysis with NVivo*. London, UK: SAGE Publications Ltd.

Boulmakoul B, Boulmakoul A, Besri Z, et al. (2017) Galois's algebraic structure and bipartite graph spatio-structural analytics for urban public transportation system assessment, *Procedia Computer Science*. Elsevier B.V., 109: 172–179. DOI: 10.1016/j.procs.2017.05.314.

Chen Z and Li Q (2008) *Vernacular Architecture Heritage Preservation (in Chinese)*. Huangshan, China: Huangshan Publishing House.

CPC Central Committee and State Council (2018) *The Five-Year Plan on Rural Vitalization Strategy (in Chinese)*. Beijing, China: Xinhua News Agency. Available at: http://www.xinhuanet.com/politics/2018-09/26/c_1123487123.htm (Accessed: 20 September 2020).

Curado M, Tortosa L, Yeghikyan G, et al. (2020) Analysis and comparison of centrality measures applied to urban networks with data. *Journal of Computational Science*. Elsevier B. V., 43: 101127. DOI: 10.1016/j.jocs.2020.101127.

Ericson JD, Chrastil ER and Warren WH (2020) Space syntax visibility graph analysis is not robust to changes in spatial and temporal resolution. *Environment and Planning B: Urban Analytics and City Science* 48(6): 1–17. DOI: 10.1177/2399808319897624.

Fei H, Hamilton GG and Wang Z (1992) *From the Soil: The Foundataion of Chinese Society*. California, CA: University of California Press.

Filomena G, Verstegen JA and Manley E (2019) A computational approach to "The Image of the City". *Cities*. Elsevier, 89: 14–25. DOI: 10.1016/j.cities.2019.01.006.

Fisher-Gewirtzman D, Shashkov A and Doytsher Y (2013) Voxel based volumetric visibility analysis of urban environments, *Survey Review*, 45(333): 451–461. DOI: 10.1179/1752270613Y.0000000059.

Ford LR (1999) Lynch revisited. New urbanism and theories of good city form, *Cities*, 16(4): 247–257. DOI: 10.1016/S0264-2751(99)00023-2.

Franz G and Wiener JM (2008) From space syntax to space semantics: a behaviorally and perceptually oriented methodology for the efficient description of the geometry and topology of environments. *Environment and Planning B: Planning and Design*, 35(4): 574–592. DOI: 10.1068/b33050.

Gehl J (2011) *Life Between Buildings: Using Public Space*. Washington, DC: Island Press. DOI: 10.3368/lj.8.1.54.

Ginzarly M, Pereira Roders A and Teller J (2019) Mapping historic urban landscape values through social media, *Journal of Cultural Heritage*. Elsevier Masson SAS, 36: 1–11. DOI: 10.1016/j.culher.2018.10.002.

Green SB, Thompson MS and Poirer J (2001) An adjusted bonferroni method for elimination of parameters in specification addition searches. *Structural Equation Modeling* 8(1): 18–39. DOI: 10.1207/S15328007SEM0801_2.

Hagberg AA, Schult DA and Swart PJ (2008) Exploring network structure, dynamics, and function using NetworkX. In: Varoquaux G, Vaught T and Millman J (eds) *Proceedings of the 7th Python in Science Conference (SciPy2008)*. Pasadena, CA: SciPy, 11–15.

Hall C and Page S (2014) *The Geography of Tourism and Recreation: Environment, Place and Space, Annals of Tourism Research*. Milton Park, UK: Routledge. DOI: 10.1016/0160-7383(79)90100-2.

Hanna S (2020) Random walks in urban graphs: a minimal model of movement. *Environment and Planning B: Urban Analytics and City Science*, 48(6): 1697–1711. DOI: 10.1177/2399808320946766.

Haq S (2003) Investigating the syntax line: configurational properties and cognitive correlates', *Environment and Planning B: Planning and Design*, 30(6): 841–863. DOI: 10.1068/b2960.

Hillier B (1999) The hidden geometry of deformed grids: or, why space syntax works, when it looks as though it shouldn't, *Environment and Planning B: Planning and Design*, 26(2): 169–191. DOI: 10.1068/b4125.

Hillier B (2007) *Space is the Machine: A Configurational Theory of Architecture*. London, UK: Space Syntax. DOI: 10.1016/s0142-694x(97)89854-7.

Hillier B and Hanson J (1984) *The Social Logic of Space*. Cambridge, MA: Cambridge University Press. DOI: 10.1017/cbo9780511597237.

Huang W et al. (2018) Tourist behavioral analysis of smart national park based on wifi positioning data: case study on Huangshan National Park (in Chinese). *Chinese Landscape Architecture* 34(3): 5–11.

Huang W, Yuming L, Borong L, et al. (2019) Modeling and predicting the occupancy in a China hub airport terminal using Wi-Fi data. *Energy and Buildings* Elsevier B.V., 203. DOI: 10.1016/j.enbuild.2019.109439..et al.

Huang W, Lin Y and Wu M (2017) Spatial-temporal behavior analysis using big data acquired by wi-fi indoor positioning system. In: Proceedings of CAADRIA, Suzhou, China.

Jacobs J (1993) *The Death and Life of Great American Cities*. New York, NY: Vintage. DOI: 10.4324/9781912282661.

Janssen J, Luiten E, Renes R, et al. (2017) Heritage as sector, factor and vector: conceptualizing the shifting relationship between heritage management and spatial planning. *European Planning Studies. Taylor & Francis* 25(9): 1654–1672. DOI: 10.1080/09654313.2017.1329410.

Kim YO (1999) *Spatial Configuration , Spatial Cognition and Spatial Behaviour: The Role of Architecture Intelligibility in Shaping Spatial Experience*. London, UK:University College.

Kim YO and Penn A (2004) Linking the spatial syntax of cognitive maps to the spatial syntax of the environment. *Environment and Behavior* 36(4): 483–504. DOI: 10.1177/0013916503261384.

Lefebvre H (1991) *The Production of Space*. Oxford, UK: Blackwell Publishing Ltd. DOI: 10.2307/378107.

Lin Y and Huang W (2017) Behavior analysis and individual labeling using data from Wi-Fi IPS. In: Disciplines and Disruption-Proceedings Catalog of the ACADIA, Boston, MA.

Lin Y and Huang W (2018) A new perspective of environmental behaviour study: A brief introduction of Wi-Fi indoor positioning system. *International Journal of Sustainable Society* 10(4): 282–299. DOI: 10.1504/IJSSOC.2018.099013.

Long H (2014) Land consolidation: An indispensable way of spatial restructuring in rural China. *Journal of Geographical Sciences* 24(2): 211–225. DOI: 10.1007/s11442-014-1083-5.

Lu Y, Gou Z, Ye Y, et al. (2019) Three-dimensional visibility graph analysis and its application. *Environment and Planning B: Urban Analytics and City Science* 46(5): 948–962. DOI: 10.1177/2399808317739893.

Lynch K (1960) *The Image of the City*. Cambridge, MA: MIT Press. DOI: 10.1002/9780470775707.ch10.

Ma D, Omer I, Osaragi T, et al. (2018) Why topology matters in predicting human activities. *Environment and Planning B: Urban Analytics and City Science* 46(7): 1297–1313. DOI: 10.1177/2399808318792268.

Major MD, Penn M and Spiliopoulou G (1999) In with the right crowd: crowd movement and space use in Trafalgar Square during the new year's eve celebrations. In: Space Syntax Second International Symposium, Brasília, Brazil.

Major MD, Penn A, Fong P, et al. (2000) Following the crowd: spatial layout and crowd behaviour. *Architectural Research Quarterly* 4(3): 257–264. DOI: 10.1017/S1359135500000282.

Mashhadi A, Acer UG, Boran A, et al. (2016) Exploring space syntax on entrepreneurial opportunities with Wi-Fi analytics. In: Luckowicz P (ed) *UbiComp 2016-Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. New York, NY: Association for Computing Machinery; 2016, 658–669. DOI: 10.1145/2971648.2971745.

Morello E and Ratti C (2009) A digital image of the city: 3D isovists in Lynch's urban analysis. *Environment and Planning B: Planning and Design* 36(5): 837–853. DOI: 10.1068/b34144t.

Mundfrom DJ, Perrett J, Schaffer J, et al. (2006) Bonferroni adjustments in tests for regression coefficients. *Multiple Linear Regression Viewpoints* 32(1): 1–6.

Nourian P (2016) *Configraphics: Graph Theoretical Methods for Design and Analysis of Spatial Configurations, A+BE Architecture and the Built Environment*. Delft, Netherlands: Delft Architectural Studies on Housing. DOI: 10.7480/abe.2016.14.

Nourian P, Rezvani S, Sariyildiz S, et al. (2016) Spectral modelling for spatial network analysis. In: Proceedings of the Symposium on Simulation for Architecture and Urban Design (simAUD 2016), London, UK, 11 July 2016.

Noyman A, Doorley R, Xiong Z, et al. (2019) Reversed urbanism: inferring urban performance through behavioral patterns in temporal telecom data. *Environment and Planning B: Urban Analytics and City Science* 46(8): 1480–1498. DOI: 10.1177/2399808319840668.

Orellana D (2019) Exploring the influence of road network structure on the spatial behaviour of cyclists using crowdsourced data. *Environment and Planning B: Urban Analytics and City Science* 46(7): 1314–1330. DOI 10.1177/2399808319863810.

Page L, Motwani S, Winograd R, et al. (1998) *The PageRank Citation Ranking: Bringing Order to the Web January*. Technical Report. Standford, CA: Stanford InfoLab. DOI: 10.1109/IISWC.2012.6402911.

Peng X, Bao Y and Huang Z (2020) Perceiving Beijing's "city Image" across different groups based on geotagged social media data *IEEE Access* 8: 93868–93881. DOI: 10.1109/ACCESS.2020.2995066.

Penn A (2003) Space syntax and spatial cognition: or why the axial line? *Environment and Behavior* 35(1): 30–65. DOI: 10.1177/0013916502238864.

Pereira Roders A (2019) The Historic Urban Landscape Approach in Action: Eight Years Later. In: Roders AP and Bandarin F (eds) *Reshaping Urban Conservation*. Berlin, Germany: Springer, 21–54. DOI: 10.1007/978-981-10-8887-2_2.

Porta S, Crucitti P and Latora V (2006) The network analysis of urban streets: a primal approach. *Environment and Planning B: Planning and Design* 33(5): 705–725. DOI: 10.1068/b32045.

Ratti C (2004) Space syntax: some inconsistencies. *Environment and Planning B: Planning and Design* 31(4): 487–499. DOI: 10.1068/b3019.

Read S (2011) History, structure and technique: a reply to batty and another challenge to space syntax. *The Journal of Space Syntax* 2(1): 120–124.

Richards L (1999) *Using NVivo in Qualitative Research*. London, UK: SAGE Publications Ltd.

Sapiezynski P, Stopczynski A, Gatej R, et al. (2015) Tracking human mobility using WiFi signals. *PLoS ONE* 10(7): 1–11. DOI: 10.1371/journal.pone.0130824.

Stevens Q (2006) The shape of urban experience: a reevaluation of Lynch's five elements. *Environment and Planning B: Planning and Design* 33(6): 803–823. DOI: 10.1068/b32043.

Taheri SM, Mahyar H, Firouzi M, et al. (2017) HellRank: a Hellinger-based centrality measure for bipartite social networks, social network analysis and mining *Springer Vienna* 7(1): 1–16. DOI: 10.1007/s13278-017-0440-7.

Turner A (2007) From axial to road-centre lines: a new representation for space syntax and a new model of route choice for transport network analysis. *Environment and Planning B: Planning and Design* 34(3): 539–555. DOI: 10.1068/b32067.

Turner A, Doxa M, Penn A, et al. (2001) From isovists to visibility graphs: a methodology for the analysis of architectural space. *Environment and Planning B: Planning and Design* 28(1): 103–121. DOI: 10.1068/b2684.

UNESCO (2011) *RECOMMENDATION ON the Historic Urban Landscape*. Paris, France: UNESCO.

Van der Aa BJM (2005) *Preserving the Heritage of Humanity ? Obtaining World Heritage Status and the Impacts of Listing*. Groningen, Netherlands: Rijksuniversiteit Groningen.

Wang L (1999) Preservation and revitalization for rural architectures in traditional villages: revelation from rural revitalization planning in Germany (in Chinese). *Architectural Journal* 11: 16–21.

Wang L (2000) The future image for villages: experiences from traditional villages and contemporary settlement planning (in Chinese). *Architectural Journal* 11: 16–22. DOI: 10.1017/CBO9781107415324.004.

Wasserman S and Faust K (1994) *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge University Press.

Wooldridge JM (2013) *Introductory Econometrics: A Modern Approach*. 5th edition. Mason: South-Western Cengage Learning. DOI: 10.4324/9781351140768-8.

Yang R, Xu Q and Long H (2016) Spatial distribution characteristics and optimized reconstruction analysis of China's rural settlements during the process of rapid urbanization. *Journal of Rural Studies* Elsevier Ltd, 47: 413–424. DOI: 10.1016/j.jrurstud.2016.05.013.

Zhang A, Li W, Wu J, et al. (2020) How can the urban landscape affect urban vitality at the street block level? A case study of 15 metropolises in China. *Environment and Planning B: Urban Analytics and City Science* 48(5): 1245–1262. DOI: 10.1177/2399808320924425.

Nan Bai is currently a PhD candidate in the UNESCO chair in Heritage and Values: Heritage and the Reshaping of Urban Conservation for Sustainability, Department of Architectural Engineering and Technology at the TU Delft. He completed his BA in Architecture and BSc in Psychology from Tsinghua University, Beijing in 2016, and Master of Architecture and MSc in Architecture from the dual degree program with Tsinghua University and TU Berlin, Berlin in 2019. He also participates in the HERILAND project in European Union as a Marie Sklodowska-Curie Early-Stage Researcher. His research interests mainly involve computational social sciences, big data analysis, machine/deep learning applications, and cultural heritage planning.

Pirouz Nourian is an assistant professor of Design Informatics, Department of Architectural Engineering and Technology at TU Delft. He completed his BSc in Control Engineering at KN Toosi University of Technology, Tehran in 2004, MSc in Architecture at University of Art, Tehran in 2009, and PhD in Design Informatics at TU Delft in 2016. He is the founder and director of Genesis Lab: Laboratory of Generative Systems and Sciences. He has the expertise and interest in computational generative design, specifically by means of mathematical modelling, geometrical, topological, or graph-theoretical computing. He has developed several computational tool-suites, most well-known among which are Space Syntax and Cheetah for Grasshopper 3D.

Weixin Huang is an Associate Professor in the School of Architecture, Tsinghua University, the Associate Director of the Digital Architectural Technology Education Committee of China, former committee member of Computer-Aided Architectural Design Research in Asia (CAADRIA), and one of the founders of Digital Architectural Design Association (DADA) of the Architectural Society of China (ASC). He completed his BE in Structural Engineering at Tsinghua University, Beijing in 1999, ME in Underground Architectural Space at Tsinghua University in 2002, and PhD in Architectural Planning at Kyoto University, Japan in 2007. With his multi-disciplinary

background in architecture, structural engineering and computation, his research focuses on three different but related areas, including big data architectural environmental behaviour analysis, intelligent algorithm supported spatial design and cognition, and architecture & structure integrated digital construction system.

Lu Wang is Professor at the School of Architecture at Tsinghua University. He completed his Bachelor and Master of Architecture in Tsinghua University, Beijing in 1987, and PhD in Hannover University, Germany in 1997. He was chief editor of World Architecture Magazine from 2000 to 2012. He is the founder and leader of the architectural office "in+of architecture". His main research and practice interests include architecture design and theory, regional architecture and vernacular settlement planning, and evolving rural architecture.

Raoul Bunschoten is Professor of Sustainable Urban Planning and Urban Design at TU Berlin. He completed his Diplom in Architecture at ETH, Zurich, and further studied at the Cooper Union School of Art and Architecture, New York, and at Cranbrook Academy of Art in Michigan, USA. He has initiated the Conscious City and has founded the Conscious City Lab in TU Berlin. He is the founder and director of CHORA, an architectural design and urban planning group, including CHORA Berlin and the UK-based CHORA research. He is a specialist in Climate Crisis Impact planning and Urban Intelligence planning.

Ana Pereira Roders is Professor in the UNESCO chair in Heritage and Values: Heritage and the Reshaping of Urban Conservation for Sustainability, Department of Architectural Engineering and Technology at the TU Delft, former Head of the Department of Architectural Engineering and Technology at TU Delft. She completed her MSc in Architecture at Universidade Lusiada de Lisboa, Lisbon in 2002, and PhD in Building Technology, TU Eindhoven in 2007. She is the co-founder and was the editor of the Journal Cultural Heritage Management and Sustainable Development between 2010–2020. Her research and scholarship address historic urban landscapes and the role of cultural heritage in their sustainable transformation, with particular interest in heritage (impact) assessments and related decision-making.