Citywalk preference: An expanded measurement for informing data-driven urban planning based on social media analytics

Haoyang He[1], Lexun Wang[2], Jiayu Xu[3], Yuyang Liu[4]
Affiliation[1,2,3,4]: College of Architecture and Urban Planning, Tongji University,
Shanghai 200092, China
Email: 2330091@tongji.edu.cn

Abstract
Citywalk, a popular trend on social media platforms, is transforming people's preference for walking in cities. Despite its growing influence on street and place-making efforts, the relationship between Citywalk preference and built environment remains undisclosed, let alone their concerted efforts for informing more complex urban planning. In this research, a comprehensive evaluation system is proposed based on space analysis of built environment indicators and data of Citywalk from Xiaohongshu, to reconstruct a new urban perception model and recommend route selections for urban renewal practices. We take Shanghai's historical area as the representative example and find the results well correspond to real pedestrian preferences in LBS data, which suggests this measurement is a potential valuable guidance for urban planning in the future.

Key Words
Citywalk, social media data, spatial analysis, route selection

1 Introduction

Citywalk means "roaming around the city" on foot, referring to a popular way of exploring old buildings, browsing boutique shops, sipping a cup of coffee, or indulging in authentic local snacks. (Borucka J., 2019, p.158-163). In China, the emergence of social media platforms in the digital age has led to the evolution of Citywalk into a popular leisure activity among a diverse range of groups, particularly young people. It encourages individuals to engage in outdoor activities, benefiting health and attracting tourist economy to urban areas (Kordi. A.O. et al., 2023, p.34). Besides, Citywalk represents an opportunity for tour guides and travel service providers to offer a more tailored, professional service to meet with ever-changing market demands. As a popular trend in social media, Citywalk is transforming people's behaviour of exploring cities, with popular walking tours furthermore influencing city image and place-making efforts.

Despite many other social media factors preference found in existing researches to influence people's urban perception and behaviour in addition to Citywalk, the relationship between Citywalk indicators and built environment characters remains undisclosed.

In this vein, this study focuses on 'route' as the fundamental premise, proposing a walking tour evaluation system to assess the correlation between Citywalk preference and existing built environment, reconstructing a synthetical urban perception model, and employing new methodologies to measure these public involvement data for informing more effective urban planning practices. The framework of our study is implemented through a practical case study, in response to the emerging Citywalk trend on Xiaohongshu, a popular social media platform, with the aim to provide an expanded measurement for informing data-driven urban planning based on social media analytics.

2 Literature Review

Media data has become a crucial indicator of human activities correlating with the built environment vitality despite traditional methods and basic indicators for urban space evaluation. Data accessible on media platforms were frequently employed to discern urban spatial structures imbued with vitality (Chen et al., 2019, p. 102005), and to reveal a series of potential attractions (Jankowski et al., 2010, p. 833–852). For instance, Zhang et al. (2022) measured virtual visiting preference based on data from TikTok filming locations. Salazar-Miranda et al. (2022) focused on street activity measurement around slow-traffic zones using Twitter comments as indicators. Similarly, existing basic environment indicators are also available to evaluate both physical and social features of streets, such as accessibility, quality, walkability and vitality, influencing subjective feelings (Ma, 2024, p. 104797). Li et al. (2022) and Jiang (2022) have provided insights into the association between street vitality and the built environment, utilising urban data such as Location-Based Services (LBS), road networks, building and street data, street view images and Points of Interest (POI). However they overlook the assessment of a street's latest internet-famous properties. Fortunately, advanced technology has opened up the possibility of quantitative research on social-economic analysis, such as interaction between media data and environmental elements. Improved computation methods could hold practical significance on urban regeneration efforts, including street renewal, urban planning and vitality prediction. For example, an application proposed by Li et al. (2022) and Huang et al. (2021) measured the city image perception through social media analytics and proposed instructions for planning practice. Garay Tamajón and Cànoves Valiente (2017) also utilised various data on social media platforms to explore the formation mechanism of people’s city image.

Furthermore, emerging social media trends are shaping pedestrians’ travel preferences, offering an expanded metric for evaluating urban regeneration practices. Specifically, Ye et al. (2023) observed the alterations in Twitter usage leading to the mobility fluctuations during pandemic era in New York City. Liu et al. (2023) discovered that young women's choice of shopping destinations is more influenced by word-of-mouth marketing on social media platforms. In China, Citywalk has been a pop trend that influences many people’s route preference (Liang, 2023, p.72). To finds motivations for planning and design efforts in specific sites, which helps to address the urban renewal position, Ma et al. (2021) utilised the correlation analysis between visual elements and human perception to support city regeneration decisions. In addition, online opinions from residents about city streets also serve as crucial references for street planning improvements (Wang et al., 2022, p.215–230), such as internet investigations of underground pedestrian networks as valuable guidance for city’s infrastructure construction (Chan et al., 2022, p. 104678).

In this context, based on the established muti-sourced data-driven planning system, we chose Citywalk preference as the fundamental premise, applying a systematic framework of methodologies to comprehensively uncover Citywalk’s internet-famous nature as well as its role within the urban environment, instead of relying on limited manual intuition or isolated data units.

3 Citywalk and Current Situation

Citywalk originated from international media like Twitter, but are currently trending across multiple Chinese social media platforms such as Xiaohongshu and Weibo. This research selected Xiaohongshu, a popular Chinese social media platform, as the channel to obtain relevant data of Citywalk issue in Shanghai. The feasibility demonstrated by previous researches and empirical evidence. Gu et al. (2023) and Zhou et al. (2021) verified that the data crawled from Xiaohongshu notes is correlated with citizens' recreational activities in urban spaces, which affirms its significance in exploring the connection between virtual media space and real physical space. In addition, the data from Xiaohongshu features characteristics of large quantity, time sensitivity, public participation, wide transmissibility, accurate location, etc. Prevailing in its individual convenience for publishing, sharing and recommendation, Xiaohongshu gathers high-quality Citywalk notes, providing representative and operable data for this study.

To improve comprehension of the social media influence on mechanism of route selection, we diagrammed the operation of Citywalk notes on Xiaohongshu platform (refer to **Fig. 1**). Firstly, users browse Citywalk contents using keywords, freely filtering by the hottest, latest, most relevant, most recent, or the notes published by specific users. Secondly, by clicking into any note, users can access routes including their names, mapping locations, itineraries, surrounding internet-famous sites and recommendation reasons, with which travellers can select destinations and travel routes efficiently. Besides, the number of likes, favourites, and comments of the note can additionally help the viewers with insights into the popularity of their travel decisions. Concluded from recommendations on Xiaohongshu, Citywalk primarily takes place on the main streets within the Inner Ring of Shanghai, spanning a comprehensive range of topics including popular commercial activities, public spaces hosting temporary events, and historic districts with pleasant street views.

Fig. 1 The Interface of notes on Xiaohongshu

4 Methodology

4.1 Road as the primary measurement object

This study selects road as the primary measurement object for gathering statistical data on Xiaohongshu and other built environment indicators. When constructing models, we also incorporate roads to analyse correlations and make predictions based on the collected data. The specific methods are as follows.

1) In contrast to activities in huge plots and buildings inside, Citywalk mainly takes place on city streets. However, this does not result in negating the significance of buildings adjacent to or bordering plots, as they will significantly impact the perception of street walking.

2) Besides, considering point data such as Points of Interest (POI), we associate their information with the nearest roads. This is because POIs can significantly influence people's behaviour on the street, particularly when citizens are inclined to follow instructions from social media platforms.

3) The acquisition of street view images provides great help in analysing the built environment of the road itself, including various indicators such as greenery, sky view, pedestrian space, motorisation and diversity.

4) Each road is assigned a unique name, corresponding to the same name used on the social media platform. This allows us to align media data with each road accurately for mapping.

This approach ensures that roads are associated with integrated numerical and spatial data while preserving their spatial connectivity for further applications such as network analysis and route selection.

4.2 Comprehensive evaluation system

To make this comprehensive system more clear, the study divides the paradigm into two main functional modules and four interconnected steps. These parts integrate various data analysis methods based on numerical and spatial aspects, such as Network Analysis and Proximity Analysis within Geographic Information System (GIS), as well as Spearman Correlation and Stepwise regression. Once this synthetic evaluation model is established, it enables quantification of interactions among multiple data sources and facilitates tenable predictions for street utilisation in the future, which is helpful for vitality analysis and urban renewal route selection (see **Fig. 2**).

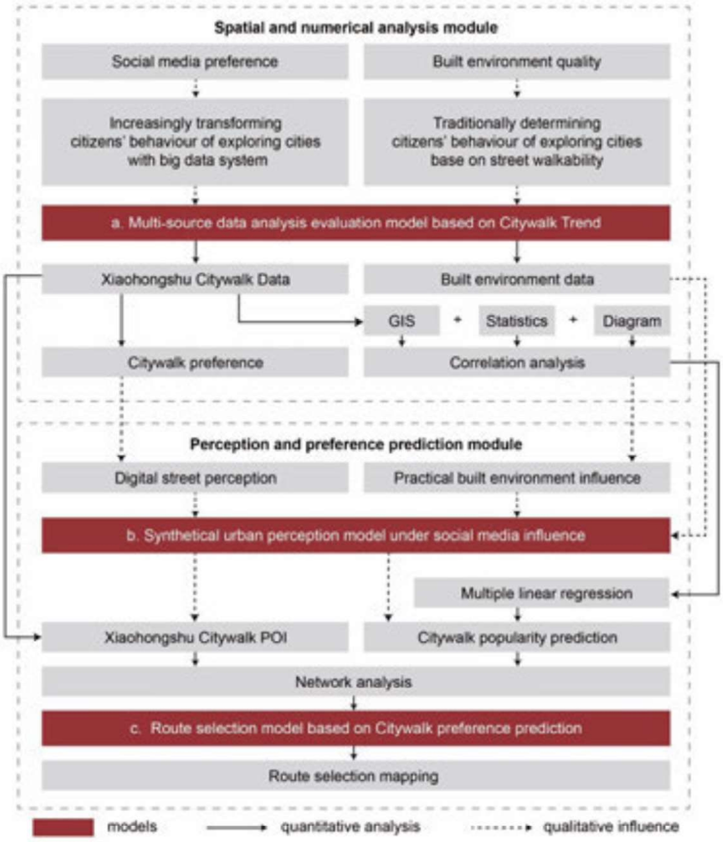


Fig. 2 Research framework and comprehensive evaluation system

4.3 Study area

To effectively establish the data analysis model, this study selects top-typical roads from the crawled data on Xiaohongshu platform as research objects (specific details are in analytical framework *step 1*). We mapped these roads onto the road network, revealing that social media platforms tend to recommend areas with numerous attractions for pedestrians. For instance, roads proximate to tourist attractions and historical districts are more apt to be suggested compared to those situated within residential communities. Therefore, the inner ring of Shanghai (refer to **Fig. 3**) is more suitable for conducting correlation analysis and predicting Citywalk patterns, following the selected Hengfu historical area in introduction.

Fig. 3 Most popular Citywalk roads are circled in the Shanghai Inner Ring.

4.4 Analytical framework

Step 1 Data acquisition and preparation

To thoroughly validate the quantifiable relationship between Citywalk preference and built environment, this study adapts three categories of data based on key indicators from existing research. Firstly, popularity data related to Citywalk destinations on social media platforms is employed to inform Citywalk preferences. Secondly, built environment data including transportation stops, street accessibility, Points of Interest (POIs), adjacent Floor-Area Ratio (FAR), and building density is integrated for route preference correlation analysis in different Citywalk themes among 'shopping districts', 'tourist attractions' and 'historical streets'. Lastly, street view image data containing information on greenery, sky view, pedestrian space, motorisation diversity and other factors is also included as street quality measurements, which can influence Citywalk preference.

The quantifiable indicators and quantification methods are presented in **Table. 1** below. Given that roads are the final measurement units, we adopt the proximity principle to attach the data to each road segment. This involves directly adding average numerical data to road attributes and converting spatial data into area density for calculation.

Data Categories	Indicators	Quantification method

Citywalk preference	Citywalk point vitality Citywalk road vitality	Points | average value Numerical attributes | average value
Built environment	Living facilities Tourist attractions Shopping places Restaurants Metro Stations Bus stops	Points | density (numbers/m²)
	FAR Building density	Polygons | average value
Street view image	motorisation ratio sidewalk ratio Building ratio Greenery Sky view Urban furniture Colour diversity Façade permeability	Numerical attributes | average value(0~100%)
Accessibility	500m_accessibility 300m_accessibility Full_accessibility	Numerical attributes | value
Vitality	Location Based Services (LBS)	Grids | qualitative mapping

Table. 1 Citywalk preference data and built environment data.

For data acquisition, we utilise all publicly accessible open data online, which is also readily available in other regions worldwide if using the similar technique framework of this study. The following are the resources we utilised in Shanghai for example, along with the detailed data processing methods.

1) To quantify Citywalk preference, the data obtained from the Xiaohongshu platform (a widely used social media app, particularly among individual users in China) is classified into three categories: texts, number of likes, and release time (refer to **Table. 2**).

Texts(i)	Number of likes	Release time
从[**武康路**]步行到[**徐家汇**]，今天成功打卡[**武康大楼**]……	6667	2023.11.1
…	…	…

Table. 2 Each text includes many places of interest with preference data.

Considering Citywalk is a time-sensitive activity and may change over time, this study only takes data from November 2023 to December 2023 as an example， And we convert it into a quantifiable indicator, as shown in the following formula:

$$Activity_{j-place} = \sum_{i=1}^{i} \frac{likes_{i,j}}{time_{i,j}} \tag{1}$$

Here, $Activity_j$ represents the vitality of each place with its unique name code such as [武康路]. $likes_{i,j}$ and $time_{i,j}$ represent the *likes* number and *duration time* of each *i* text containing each j place in this stratified framework.

Since the locations mentioned in the texts include both roads and points of interest, we compute the average value of the vitality-related data surrounding the measurement roads, using GIS buffer with the following formula:

$$Redu_{z-road} = \sum_{j=1}^{n} \frac{Activity_{j-place,z}}{n} \tag{2}$$

As roads are linear features, this study utilizes a spatial buffer consisting of circles at both ends of each road segment and a band area along its length to gather road-related data. Here, $Activity_{j-place,z}$ represents the places within this buffer, and n represents the total number of places nearby (refer to **Fig. 4**). Therefore, each road is able to obtain an average activity $Redu_{z-road}$ from places around. Considering pedestrian's walking habit, we apply the distance of 500 metres (5-7min walking) from *15-Min Pedestrian Neighbourhood* theory in China's urban planning.

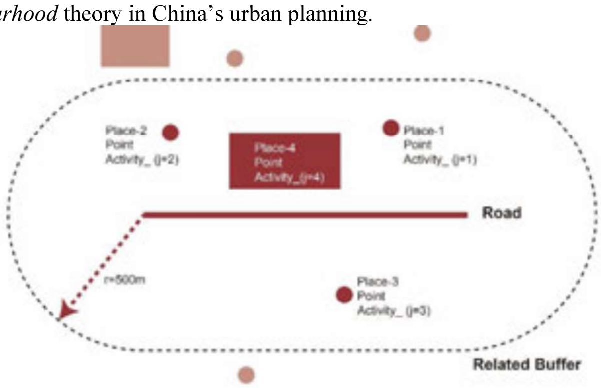
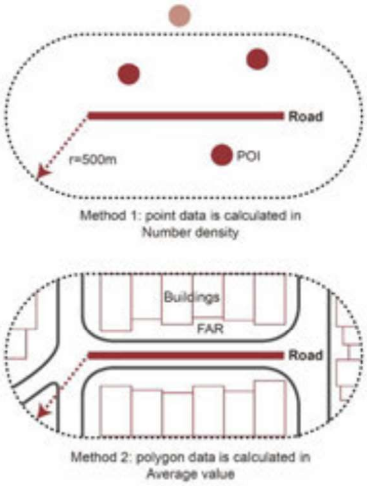


Fig. 4 Buffer is employed to collect Citywalk preference data for each road.

2) To pithily measure the various built environment indicators, we mainly acquire POI data and basic urban 3D spatial information from Amap (a popular mapping APP in China), utilising GIS for subsequent processing. Similarly, the data of each indicator is linked within different buffers of each road, whose radius vary according to the service distance stipulation in planning policy documents and pedestrian perception of the built environment. The two different computing methods and distance radius are shown below (**Fig. 5, Table .3**).



Indicators	Radius(metres)
Living facilities	200
Tourist attractions	600
Shopping places	300
Restaurants	300
Metro Stations	500
Bus stops	300
FAR	50
Building density	50

Fig. 5 Two different computing method;
Table. 3 Different buffer radius according to planning stipulation and urban perception

These two methods mainly focus on numerical computation while the spatial influential factors are embodied in distance radius specified in **Table. 3**. The detailed calculation formulas are as follows:

$$Density_{i-point,\ z} = \frac{N_{i,z}}{Area_z};\ AveValue_{i-area,\ z} = \frac{\sum_i (Value_i \times Area_i)}{Area_z} \qquad (3), (4)$$

Here, $Area_z$ represents the buffer area of road z. In method 1, $N_{i,z}$ is the total number of each kind of point data within this buffer. In method 2, $Value_i$ represents either the FAR of including plots or the total floor area of intersecting buildings, while $Area_i$ only represents to the part of buildings and plots in the buffer area. Despite that the average value may be lower than the practical data due to the indicator area being smaller than that of the buffer, it still provides a quantifiable measurement of the FAR and building density along the road.

3) To incorporate factors directly from the street itself, this study supplements street view image to provide indicators of human-scale perception. These measurements primarily concentrate on the visual proportion of different categories of graphic elements. The specific diagram and indicators are shown in **Fig. 6** and **Table. 4**. Given the linear features of street, average computing method is similarly employed for each road.

Fig. 6 Semantic segmentation

Indicators	Practical meaning
motorisation ratio	Proportion of motor vehicle lanes, reflecting motorisation
sidewalk ratio	Proportion of sidewalks, reflecting walking environment
Building ratio	Proportion of building objects, reflecting building density
Greenery	Street greening degree, reflecting environment quality
Sky view	Sky visibility, reflecting street openness or width
Urban furniture	Urban Infrastructure construction degree
Colour diversity	Colour richness, which reflects visual richness
Façade permeability	Proportion of transparent doors and windows on façade

Table. 4 Each text includes many places of interest with preference data.

Using this quantifiable measurement with Citywalk preference and built environment data, we can establish the statistical table with each selected road as the fundamental data preparation for next computation.

Step 2 Correlation analysis

In the first analysis module, we primarily address the qualitative correlation between the Citywalk preference and miscellaneous built environment indicators. Since there is no direct computation on practical meaning of the data, normalisation is employed to simplify small or large values into range [0,1] for improved calculation. The normalisation is indicated in the following fumula:

$$Value_{indicator-normalised} = \frac{Value_i - \min(Value_i)}{\max(Value_i) - \min(Value_i)} \tag{5}$$

Meanwhile, to enhance precision, we excluded abnormal high $Redu_{z-road}$ values which may be more influenced by recommendation algorithm of Xiaohongshu platform, together with those low values that include only one or two Citywalk places.

Step 3 Synthetical urban perception model

To convincingly establish prediction and route selection models from human-scale perspective, we employ a new synthetical urban perception model to address digital influence like Citywalk preference or internet-famous phenomenon on social platforms. Extended from the research framework on spatial perception based on psychological processes, this urban perception model integrates digital street perception from media platforms and practical built environment feelings (refer to **Fig. 7**), reflecting on the various indicators selected above. In addition, the potential causation from objective feelings to subjective behaviour generates a more complex issue, for which we utilise another relatively quantifiable regression method in step 4 to explain.

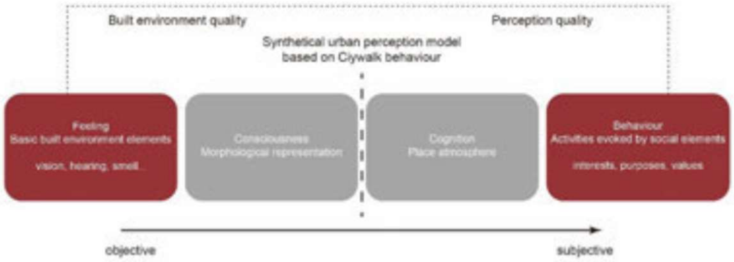


Fig. 7 Synthetical urban perception model based on Citywalk

Step 4 Regression and prediction model

To quantitatively address the relationship between built environment and Citywalk preference, this study apply multiple regression analysis, constructing an objective expression to estimate the popularity of future Citywalk routes upon the built environment data (refer to **Fig. 8**). Through this regression, other roads without current Citywalk preference may realise their potential to become lively in the next social media trend.

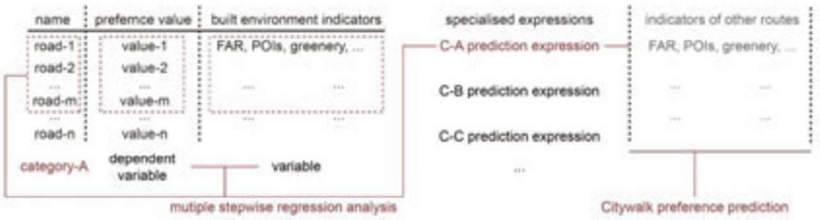


Fig. 8 Regression analysis and preference prediction of different route categories

To improve the precision of regression, we classified the selected road into three main groups through text analysis in module 1 and subsequently implemented three pertinent prediction from 'shopping districts' to 'tourist attractions' to 'historical streets', which allows for more nuanced insights into the relationship between the built environment and different Citywalk preference.

Step 5 Route selection

Route selection model is established based on regression analysis, integrating popular Citywalk destinations identified during data preparation as penetration points. Additionally, subjective manual adjustments are supplemented to connect road segments into cohesive routes in the final part.

Considering the diverse characteristics of Shanghai's political districts, we chose four central areas, with built environments similar to those of the popular Citywalk sample roads selected in data preparation, ensuring that the selected roads are not excessively repeated (refer to **Fig. 9**). Subsequently, network analysis in GIS is applied with popular Citywalk places as penetration points and three categories of Citywalk preference predictions to ascertain road throughput (shown in **Fig. 10**).

Fig. 9 Four prediction areas in contrast to popular Citywalk sample roads.

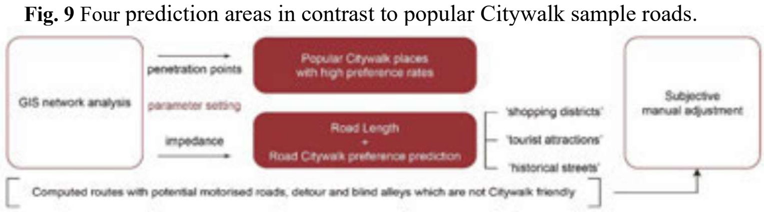


Fig. 10 Route selection framework using GIS and manual adjustment.

5 Results

In this present study, after cleaning the relevant data from Xiaohongshu platform, a total of 822 outstanding Citywalk road segments were included in the research. Following text analysis of all preferred Citywalk places, we categorized them into three groups with different themes - 'shopping districts', 'tourist attractions' and 'historical streets'. Category 'shopping districts' comprised 362 roads, featuring texts related to special shops, exquisite stores and restaurants. Category 'tourist attractions' included 350 roads, containing more texts of tourist destinations and scenic spots. Category 'historical streets' consist of 361 roads, representing streets with texts focusing on historical buildings and sites. All categories of road segments are attached with the computed built environment data, including both direct and normalised values (details refer to **Appendix Table. 1**).

5.1 Indicators Correlation analysis

Centred on the calculated Citywalk preference value as the dependent variable, we separately listed the Spearman's rank correlation for three categories of road datasets (refer to **Table. 5**). Despite the key computational indexes R^2 remaining around 0.3~0.4, which are tenable to reveal socially interactive relationships, the outcomes potentially indicate the comparative advantage of certain indicators such as Tourist attractions, Shopping places, Restaurants and Metro Stations.

Indicators (normalised)	C-shopping districts Preference R²	C- tourist attractions Preference R²	C- historical streets Preference R²
Living facilities	0.142(0.007***)	0.165(0.002***)	0.153(0.004***)
Tourist attractions	0.351(0.000***)	0.313(0.000***)	0.404(0.000***)
Shopping places	0.262(0.000***)	0.259(0.000***)	0.284(0.000***)
Restaurants	0.235(0.000***)	0.209(0.000***)	0.275(0.000***)
Metro Stations	0.302(0.000***)	0.292(0.000***)	0.283(0.000***)
Bus stops	0.118(0.024**)	0.113(0.035**)	0.191(0.000***)
FAR	-0.084(0.112)	-0.079(0.141)	-0.053(0.316)
Building density	0.045(0.389)	0.088(0.101)	0.041(0.433)
motorisation ratio	-0.098(0.062*)	-0.03(0.577)	-0.042(0.425)
sidewalk ratio	0.22(0.000***)	0.292(0.000***)	0.256(0.000***)
Building ratio	-0.07(0.183)	-0.108(0.044**)	-0.126(0.017**)
Greenery	0.148(0.005***)	0.172(0.001***)	0.169(0.001***)
Sky view	-0.265(0.000***)	-0.255(0.000***)	-0.285(0.000***)
Urban furniture	-0.086(0.103)	-0.158(0.003***)	-0.147(0.005***)
Colour diversity	-0.063(0.233)	-0.074(0.171)	-0.085(0.109)
Façade permeability	0.256(0.000***)	0.318(0.000***)	0.32(0.000***)
500m_accessibility	-0.032(0.539)	-0.05(0.355)	-0.038(0.478)
300m_accessibility	-0.037(0.482)	-0.048(0.369)	-0.051(0.334)
Full_accessibility	-0.018(0.732)	-0.017(0.748)	-0.01(0.857)

Table. 5 Spearman analysis of Citywalk preference and built environment indicators.

Conclusively based on the results, Citywalk preference is partly correlated with street elements. Individuals are more likely to walk on streets with greater amenities, wider sidewalks, and more permeable façades. This further validates our findings regarding the complexity of citizens' urban perception and Citywalk preference, which are synthetically influenced by physical feelings of built environment indicators in addition to recommendations from social media platforms such as Xiaohongshu.

5.2 Regression analysis

To make predictions of Citywalk preference under different themes for the unselected roads, we employed three distinctive stepwise regression expressions, extracting the original data computed in module 1, which is imbued with practical meanings. The final prediction results should be absolute values, thus it is also essential for other indicators to be realistic values throughout the calculation process. After eliminating the irrelevant indicators through stepwise multiple regression analysis, we obtained three expressions with indexes R^2 around 0.4~0.5 (refer to **Table. 6,** the number is computed without normalisation). In the context of social-economic analysis, it attains a convincing level, particularly for predictions of potential Citywalk roads derived from computed outcomes rather than relying solely on planners' empirical intuition.

-	**C-shopping districts**		**C- tourist attractions**		**C- historical streets**	
Indicators (without normalisation)	Unstandardized Coefficients	R^2	Unstandardized Coefficients	R^2	Unstandardized Coefficients	R^2
Living facilities	-(abolished)		-		-	
Tourist attractions	37276519.308		35079820.375		35233670.141	
Shopping places	292515.051		421365.147		284811.133	
Restaurants	-3314557.03		-5176136.127		-2749375.194	
Metro Stations	62395195.844		46120586.525		63003256.789	
Bus stops	-13282497.56		-		-10715763.545	
FAR	-		-		-	
Building density	-		-		-	
motorisation ratio	-		-		-	
sidewalk ratio	-	**0.387**	6294.353	**0.456**	12687.666	**0.388**
Building ratio	-		-3345.4		-	
Greenery	1444.488		-		1260.088	
Sky view	-		-		-	
Urban furniture	-14363.228		-15051.24		-14187.162	
Colour diversity	-		-		-	
Façade permeability	-		16336.327		-	
500m_accessibility	-		-		-	
300m_accessibility	-		3804.477		3617.647	
Full_accessibility	-		-		-	
Constant term	1253.475		1738.654		722.856	

Table. 6 Stepwise multiple regression analysis results from three categories of roads

5.3 Route selection

According to preference predictions computed by expressions with the given built environment indicators of each roads in the four selected regions, we subsequently employed network analysis in GIS to generate complete potential Citywalk routes.

By utilising high score Citywalk places as penetration points (shown in **Fig. 11** as red points) and Citywalk preference values of roads, we delineated suggested routes in each prediction area following a few manual adjustments (refer to **Fig. 11**), which anticipated prospective Citywalk behaviour in next social media trends. To further validate the potential vitality of the final selected routes, we compared the data with the layer of Shanghai's daytime average Location-Based Services (LBS) data (depicted by points in **Fig. 12**). The strongly correlated results indicate that the roads outlined in red from the diagrams could indeed provide quantitative recommendations of urban renewal streets for urban planning endeavours in the future.

Fig. 11 Citywalk route selection in four Prediction areas

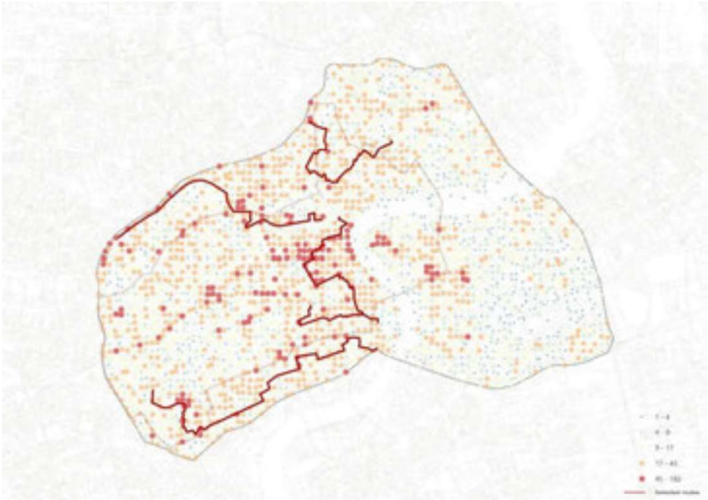


Fig. 12 Citywalk routes compared with LBS vitality data

6 Conclusion and Disscusion

This study focuses on Citywalk, a spontaneous urban exploration behaviour significantly influenced by social media, and constructs a comprehensive framework for route research, encompassing the entire process from exploration to application.

By examining particular behaviour pattern, this research responds to the changing patterns of urban experiences and interaction modes in social media era. Additionally, the quantitative relationship between Citywalk preference and built environment elements is unveiled in our experiment, which is often neglected in previous studies. As a supplement, this study develops a walking targeted comprehensive urban perception model under the influence of social media. To validate our findings in practice, we ultimately employ a prediction model to identify potential Citywalk routes for future urban renewal endeavours.

In terms of theoretical contributions, the study extends the boundaries of existing theories of urban perception and behavioural analysis. The findings deepened understanding of the relationship between urban behaviour and built environment elements, and enriches the understanding of urban walking behaviour. From a practical perspective, this study incorporates multi-source data into the analysis, providing a new data-driven paradigm for public participation in urban planning. Meanwhile, the prediction model offers decision-making reference for urban pedestrian system renewal, providing new possibilities into digitally-driven urban planning practices.

From a technical perspective, taking Shanghai as a significant exploration case allows for the development of an analysis paradigm that can be broadly applicable to other cities globally, by collecting the local data and making route selections tailored to each city's unique characteristics. However, we claim the limitations of the insufficiency of the quantitative indicators of online social media data, since incorporating more city samples and diversified social media platform resources may raise the coherent influence of superfluous varieties of data on the analysis and modelling, resulting in lower correlation index R^2. Additionally, extended comprehensive analysis of geographic covariance and more precise extraction of human perceptions and evaluations from social media information that we didn't pay enough attention to may enhance the accuracy of the model as well.

7 References

Chan, H.-Y., Ip, L.-C., Mansoor, U., & Chen, A. (2022). Pedestrian route choice with respect to new lift-only entrances to underground space: Case study of a metro station area in hilly terrain in Hong Kong. Tunnelling and Underground Space Technology, 129, 104678.

Chen, T., Hui, E. C. M., Wu, J., Lang, W., & Li, X. (2019). Identifying urban spatial structure and urban vibrancy in highly dense cities using georeferenced social media data. Habitat International, 89, 102005.

Garay Tamajón, L., & Cànoves Valiente, G. (2017). Barcelona seen through the eyes of TripAdvisor: Actors, typologies and components of destination image in social media platforms. Current Issues in Tourism, 20(1), 33–37.

Gu Jinyuan, Yang Dongfeng. (2023). New Features of Urban Leisure Space Distribution under the Influence of Social Media: A Comparison between Xiaohongshu Check-ins and POI. Journal of Geo-information Science, 02, 1–20.

Huang, J., Obracht-Prondzynska, H., Kamrowska-Zaluska, D., Sun, Y., & Li, L. (2021). The image of the City on social media: A comparative study using "Big Data" and "Small Data" methods in the Tri-City Region in Poland. Landscape and Urban Planning, 206, 103977.

Jankowski, P., Andrienko, N., Andrienko, G., & Kisilevich, S. (2010). Discovering Landmark Preferences and Movement Patterns from Photo Postings. Transactions in GIS, 14(6), 833–852.

Jiang, Y., Han, Y., Liu, M., & Ye, Y. (2022). Street vitality and built environment features: A data-informed approach from fourteen Chinese cities. Sustainable Cities and Society, 79, 103724.

Justyna Borucka. (2019). City walk: a didactic innovative experiment in architectural education. World Transactions on Engineering and Technology Education, 17(2), 158-163.

Kordi, A. O., & Galal Ahmed, K. (2023). Towards a socially vibrant city: Exploring urban typologies and morphologies of the emerging "CityWalks" in Dubai. City, Territory and Architecture, 10(1), 34.

Liang Xiaobi. (2023). The Citywalk tourism boom. Insight China, 30, 72.

Liu, J., Wang, C., Zhang, T. (Christina), Qiao, H., (2023). Delineating the Effects of Social Media Marketing Activities on Generation Z Travel Behaviors. Journal of Travel Research, 62, 1140–1158.

Li, Q., Cui, C., Liu, F., Wu, Q., Run, Y., & Han, Z. (2022). Multidimensional Urban Vitality on Streets: Spatial Patterns and Influence Factor Identification Using Multisource Urban Data. ISPRS International Journal of Geo-Information, 11(1), Article 1.

Li, Y., Yabuki, N., & Fukuda, T. (2022). Exploring the association between street built environment and street vitality using deep learning methods. Sustainable Cities and Society, 79, 103656.

Ma, S., Wang, B., Liu, W., Zhou, H., Wang, Y., & Li, S. (2024). Assessment of street space quality and subjective well-being mismatch and its impact, using multi-source big data. Cities, 147, 104797.

Ma, X., Ma, C., Wu, C., Xi, Y., Yang, R., Peng, N., Zhang, C., & Ren, F. (2021). Measuring human perceptions of streetscapes to better inform urban renewal: A perspective of scene semantic parsing. Cities, 110, 103086.

Salazar-Miranda, A., Heine, C., Duarte, F., Schechtner, K., & Ratti, C. (2022). Measuring the impact of slow zones on street life using social media. Cities, 131, 104010.

Wang, L., Han, X., He, J., & Jung, T. (2022). Measuring residents' perceptions of city streets to inform better street planning through deep learning and space syntax. ISPRS Journal of Photogrammetry and Remote Sensing, 190, 215–230.

Ye, Q., Ozbay, K., Zuo, F., Chen, X., 2023. Impact of Social Media on Travel Behaviors during the COVID-19 Pandemic: Evidence from New York City. Transportation Research Record: Journal of the Transportation Research Board 2677, 219–238.

Zhang, Y., Li, Y., Zhang, E., & Long, Y. (2022). Revealing virtual visiting preference: Differentiating virtual and physical space with massive TikTok records in Beijing. Cities, 130, 103983.

Zhou Kai, Zhang Haitao, Xia Yining, Liu Chong. (2021). New Urban Consumption Space Shaped by SociaI Media： A Case Study of the Geo-Tagging PIaces of Changsha on Xiaohongshu. Modern Urban Research, 09, 20-27.

Appendices

Pictures and detailed excel data included in:

https://drive.google.com/drive/folders/1PQLpmnvqjwk11EzIT4xHS7uERfGludHr?usp=drive_link

This study is supported by **National Natural Science Foundation of China (52078343, 72361137008)**