

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/167012/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Liu, Xiang, Chen, Xiaohong, Orford, Scott , Tian, Mingshu and Zou, Guojian 2024.
Does better accessibility always mean higher house price? Environment and Planning B: Urban Analytics and City Science

Publishers page:

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



This is a pre-copy-editing, author-produced PDF of an article accepted following peer review for publication in ENVIRONMENT AND PLANNING B: URBAN ANALYTICS AND CITY SCIENCE

Does better accessibility always mean higher house price?

Xiang Liu (Corresponding Author)

Urban Mobility Institute, Tongji University
4800 Caoan Road, Shanghai, 201804, China
Email: xliu02@tongji.edu.cn
<https://orcid.org/0000-0001-7890-987X>

Xiaohong Chen

Urban Mobility Institute
Key Laboratory of Road and Traffic Engineering of the Ministry of Education
School of Transportation Engineering, Tongji University
4800 Caoan Road, Shanghai 201804, China
Email: tongjicxh@163.com

Scott Orford

School of Geography and Planning
Cardiff University, Wales, CF10 3WA, UK
Email: orfords@cardiff.ac.uk

Mingshu Tian

Key Laboratory of Road and Traffic Engineering of the Ministry of Education
School of Transportation Engineering, Tongji University
4800 Caoan Road, Shanghai 201804, China
Email: 2133372@tongji.edu.cn

Guojian Zou

Key Laboratory of Road and Traffic Engineering of the Ministry of Education
School of Transportation Engineering, Tongji University
4800 Caoan Road, Shanghai 201804, China
Email: 2010768@tongji.edu.
<https://orcid.org/0000-0003-3597-6874>

For Citation Please use: Liu, X., Chen, X., Orford, S., Tian, M., & Zou, G., (accepted). Does better accessibility always mean higher house prices? *Environment and Planning B: Urban Analytics and City Science*.

Does better accessibility always mean higher house prices?

ABSTRACT

Numerous studies have explored the correlations between house prices and spatial accessibility, but few have delved into the nonlinearities between both. This study uses Cardiff (UK) as a case study and applies interpretable machine learning algorithms, eXtreme Gradient Boosting (XGBoost) and SHapley Additive exPlanations (SHAP), to estimate the nonlinear effects of geometric locational accessibility and street network accessibility on house prices. The findings suggest (1) proximity to the CBD, typically the major determinant of land values in hedonic house price models, does not continuously yield higher prices; (2) street closeness centrality, a network-modelling approach to measuring accessibility, exhibits a more generalised pattern with house prices compared proximity to the CBD regardless of analytical spatial scales. The findings challenge the generalizability of Alonso's bid-rent theory in accurately portraying the relationship between accessibility and house prices in specific urban contexts, highlighting the importance of re-evaluating classical urban theories in different city contexts using new measures and modelling techniques.

Keywords: *street network, accessibility, house prices, nonlinear, machine learning*

Introduction

The economic value of accessibility has been a long-standing area of interest for urban and transport researchers (Webster, 2010; Xiao, 2017). Alonso's bid rent theory, also known as the bid-rent theory, explains the distribution of land uses and variations in land prices throughout an urban area. As a classic theory in urban studies, it posits that different land uses form concentric circles around the Central Business District (CBD), with land uses that people have a higher willingness to pay for in terms of accessibility situated closer to the CBD (Alonso, 1964; Fujita et al., 2001). This insight has influenced housing research, leading to the recognition that accessibility is a significant factor associated with house prices when using the hedonic price approach. This is generally known as the 'access-space' trade-off model, which refers to higher transport costs balancing lower house prices in less accessible areas of the city (Orford, 2002, 2000; Xiao et al., 2016a, 2016b).

In early studies, two types of accessibility measures have been widely used in hedonic house price models. The first one is the conventional planner geometry measure that measures accessibility to specific points of benefits or disbenefits (e.g., proximity to the CBD, parks, hospitals, or transit stations). However, the specification of locational measures has been problematic as these locational advantages are often highly interrelated and are defined somewhat arbitrarily by researchers regarding which ones to measure (Xiao et al., 2016a). By contrast, the second measure of accessibility is rooted in street network centrality. This method is based on Space Syntax and addresses the shortcomings of the conventional planner geometry by directly assessing

general accessibility, serving as a proxy to encapsulate the various dimensions of human interactions inherent in the urban street network (Hillier, 2007; Hillier and Hanson, 1984). This approach provides a more general understanding of accessibility and its impact on housing prices without prior knowledge of significant locational externalities and effective spatial scales (Xiao, 2017).

Although numerous studies highlight that both geometric measures of specific locational externalities and network-accessibility measures are imperative in hedonic house price models (Chakrabarti et al., 2022; Xiao et al., 2016a, 2016b), one critical question has not been thoroughly investigated: to what extent does accessibility influence house prices? In other words, is there a threshold for accessibility to yield a significant effect on house prices? Intuitively, the effects of spatial accessibility on house prices may differ when it falls into different levels of accessibility, and its effect may be saturated or substantially decayed when it reaches a certain distance. Given the complex geography of locational externalities effect and uneven distribution of local amenities in a contemporary urban context (Batty, 2013), the relationship between accessibility and house prices tends to be nonlinear. However, to our knowledge, no studies empirically investigate the possibility of the nonlinear effect of accessibility on housing prices.

This study aims to bridge the existing research gap by investigating the nonlinear effects of accessibility on house prices in the hedonic house price model using interpretable machine learning approaches. It addresses the following research questions: (1) Do accessibility variables have nonlinear effects on housing prices? (2)

Do the effects fluctuate across different analytical spatial scales?

In response, this study first introduces the background of the hedonic house price model and related literature. Then, it describes the methods and analysis results of the study. Finally, it discusses the key findings and concludes the paper with implications and limitations.

Literature review

Hedonic house price model

The hedonic house price model is an econometric method used to estimate the value of a property by examining the different characteristics that influence house prices. This model is based on the principle of hedonic pricing, which suggests that a good or service can be decomposed into the prices of its constituent attributes or features based on the economic principles of utility maximisation (Rosen, 1974). In the context of housing, the hedonic price model considers a wide range of factors that contribute to the overall value of a property (O'Sullivan and Gibb, 2002). These factors can be divided into two main categories: (1) structural attributes: the number of rooms, square footage, age, and overall condition; (2) locational attributes: which refer to a property's spatial characteristics, such as its proximity to amenities, accessibility to the transportation network, neighbourhood quality, and other environmental factors (Can, 1992; Pope and Pope, 2015; Xiao et al., 2016a). The conventional geometric locational measures in hedonic house price models are often used to represent locational attributes of properties, such as their distance or proximity to specific amenities, accessibility measures to key features of the urban area, and local neighbourhood characteristics.

These measures typically include Euclidean distances, travel times, or travel costs to various points of interest, such as CBD, schools, parks, or public transportation stations. Previous studies suggest that the accessibility to the CBD and the city's attractiveness contribute to higher house prices (Ahlfeldt and Maennig, 2010).

However, there are several limitations to conventional geometric locational measures. Firstly, the geometric locational measures in the hedonic house price model are insufficient to capture all the effects of location attributes on house prices (Xiao, 2017). Homebuyers have diverse preferences when it comes to locational attributes. Different individuals and households may prioritise and value various aspects of a location, such as proximity to amenities, access to public transportation, school quality, and environmental factors (Rodríguez and Mojica, 2009). This heterogeneity in preferences can make it difficult to accurately measure and understand the effects of locational attributes on house prices (Xiao et al., 2016b). Secondly, euclidean distances do not necessarily represent the actual travel distances or travel times between locations. They fail to account for the complex street network, topography, and potential barriers, such as rivers or highways, which can affect travel paths (Xiao, 2017). Thirdly, conventional geometric locational measures may not capture the spatial autocorrelation or the dependency between nearby properties. This can lead to biased estimates in hedonic price models, as spatially correlated errors can violate the model assumptions (Orford, 2002).

Street network and house prices

Inspired by Space Syntax, numerous studies have tried to use street network

accessibility to understand accessibility-house relationships (Chakrabarti et al., 2022; Xiao et al., 2016a, 2016b). In contrast to conventional geometric locational measures to capture the locational externality effect, Space Syntax focuses on the configuration and connectivity of urban spaces and street networks as a proxy for the multiple dimensions of human interaction that are contained in the urban street network (Hillier and Hanson, 1984). Theoretically, the value of accessibility is itself an aggregation and averaging of many separate sources of negative and positive externalities (Webster, 2010), this suggests that street layout, a product of urban and social development, may serve as a more holistic proxy for spatial accessibility in the context of multi-scaled and multi-centric urban areas (Hillier and Hanson, 1984).

Specifically, betweenness and closeness, as graph-theoretic measures of street network accessibility, are commonly used metrics in investigating the effects of street network accessibility on house prices. Betweenness measures the extent to which a particular street segment or node acts as a connector within a network. It is calculated by counting the number of shortest paths passing through a specific segment or node. A high betweenness score indicates that the segment or node is crucial for connecting different network parts, thereby serving as a critical point for movement and interaction. On the other hand, closeness refers to how easily a location can be reached from all other locations within the network. It is determined by the average length of the shortest paths from a given space to all other spaces in the network. A high closeness score suggests that space is easily accessible from other spaces, indicating a central location within the network. In other words, properties with high closeness values might be more

attractive due to their proximity to various amenities and reduced travel times (Xiao, 2017). These two metrics allow for a more comprehensive representation of how people navigate and interact with their environment, accounting for street connectivity, accessibility, and pedestrian or vehicular movement patterns (Hillier, 2007).

A need and opportunity for nonlinear analysis

Since prior studies highlight the complementary role of street network accessibility in correlating the effects of accessibility with house prices, one critical question emerges: to what extent do accessibility factors influence house prices? In other words, are there thresholds for the effect of accessibility on house prices? Previous studies often assume that the accessibility factor has a linear or log-linear relationship with house prices. However, the impact of accessibility tends to be more complex than what is typically pre-defined, with some studies acknowledging the potential for a nonlinear relationship between public transit accessibility and house prices (Jin et al., 2022). For example, an incremental increase in public transit accessibility may not invariably result in a corresponding incremental increase (or decrease) in house prices. Such complexities underscore the need for exploring the potential nonlinear relationship between accessibility and housing prices, given the increasingly complex urban form and uneven distribution of local amenities with respect to multiscaled, multi-centric urban contexts (Batty, 2013).

XGBoost (eXtreme Gradient Boosting), one of the tree-based boosting machine learning algorithms, has gained popularity in recent years for addressing regression problems in various fields, including urban and transport planning (Liu et al., 2023a,

2023b; Zou et al., 2022). In contrast to conventional linear regressions, the XGBoost model can learn complex patterns and relationships from data, making them well-suited for handling nonlinearities and high-dimensional data (Hastie et al., 2009). The XGBoost model offers several advantages over conventional linear regression models. First, machine learning algorithms are known for their excellent predictive accuracy, while the flexible structure of machine learning algorithms allows them to model complex, nonlinear relationships without imposing strict assumptions, leading to more accurate and reliable results in hedonic house price models (Zou et al., 2022). Second, by drawing SHAP (SHapley Additive exPlanations) dependence plots, researchers can identify the most influential intervals in the model, helping to identify the thresholds that have the most significant effects on the target variable (Liu et al., 2023a). Third, the SHAP-explained XGBoost model can estimate spatial effects similar to those in the spatial lag and geographically weighted regression models (Li, 2022). In other words, it is an alternative to spatial statistical models and performs better when considering unknown nonlinear effects. However, to our knowledge, no studies have specifically adopted these techniques to hedonic house price research that considers geometric locational and street network accessibility measures. This study, therefore, aims to fill these research gaps by adopting the XGBoost and SHAP models to explore the nonlinear effects of accessibility variables on house prices in Cardiff, UK.

Methods

Study area and variables

This study focuses on Cardiff, the capital city of Wales, UK, chosen for its data

availability and our familiarity with the city's structure and housing market. As of 2021, Cardiff has a population of approximately 362,310, sharing similar characteristics in size and urban form with other UK cities. This makes Cardiff an apt representation of a British medium-sized town, characterized by a classic concentric urban structure and housing stock. The findings from this study would be indicative of general trends in British cities.

The dataset for this study was sourced from the HM Land Registry for England and Wales, and enhanced with the addition of floor area measures, which have been utilised in prior research (Orford, 2002; Xiao et al., 2016b). The dataset contains the following attributes: property price, floor area, sale year, property type (detached, semi-detached, terraced, flat, or maisonette), new build status, tenure status (freehold or leasehold), and Area Classification for Output Areas (OAC). A total of 16,297 properties sold within the study area between 2001 and 2007 were included in the dataset. The properties selected for this study encompass a representative sample of the Cardiff housing market, including Victorian and Edwardian terraces in the inner city and interwar and postwar semi-detached and detached houses in suburban areas. Due to privacy considerations, the map illustrating the properties under investigation has been consolidated to the Output Area level (as shown in Fig. 1 and Fig. 2).

Two categories of accessibility variables were included in this study. The first category refers to conventional geometric measures of accessibility, this was measured using the network distance from each property to the CBD, park, and hospital. The second category refers to street network accessibility metrics measured using Spatial

Design Network Analysis (sDNA) based on the street centerline data from Ordnance Survey Mastermap ITN. Similar to Space Syntax, sDNA is a set of advanced spatial analysis tools designed for studying street network in urban research using various graph-theoretic measures such as closeness and betweenness (Cooper and Chiaradia, 2020). In our study, we employ angular distance to assess street network accessibility. Empirical studies have demonstrated that angular segment analysis is efficient and robust in measuring pedestrian movement (Cooper and Chiaradia, 2015; Hillier and Iida, 2005). Closeness and betweenness were calculated by the network quantity penalised by distance in radius angular (NQPDA) as Equation (1):

$$NQPDA(i) = \sum_{j \in R_i} \frac{p(j)}{d(i,j)} \quad (1)$$

where $NQPDA(i)$ represents the closeness of link i ; $p(j)$ is the weight of link j within the search radius R ; $d(i,j)$ denotes the shortest topological distance from link i to link j ; R_i is the link set within the search radius R of link i . Betweenness is calculated by two-phase betweenness angular (TPBtA) as defined in Equation (2):

$$TPBtA(i) = \sum_{j \in N} \sum_{k \in R_j} OD(j, z, i) \frac{P(z)}{Links(j)} \quad (2)$$

where $TPBtA(i)$ denotes the betweenness of link i . In this context, $OD(j, z, i)$ signifies the shortest angular route between link j and z that passes through link i within the search radius, denoted as R . The weight of node z within the search radius R is expressed as $P(z)$; the link set within the search radius R of link j is represented by R_j , while N stands for the set of links in the global spatial system.

Closeness and betweenness variables were calculated at the following four radii:

600 m, 3,000 m, 6,000 m, and 10,000 m. The choice of four radii as distances corresponding to different human movement scales is based on the idea that people have other preferences and travel behaviour patterns at various spatial scales. These distances are chosen to provide a range of spatial scales that capture the varying degrees of human movement and interaction with the urban environment (Xiao et al., 2016a). By considering these different scales, researchers can better understand how location-specific factors and street network accessibility influence house prices.

The descriptive statistics of all variables are presented in Table 1. The logarithmic transformation of the dependent variable (house prices) was used in the models, which helps to achieve a normal distribution of the data and is a common approach in hedonic house prices (Xiao et al., 2016a). Following the multicollinearity check, the variables “DU_TER”, “DU_CL”, “Dist_Roath”, and “Dist_Health” were excluded from the models due to Variance Inflation Factor (VIF) scores being greater than 5.

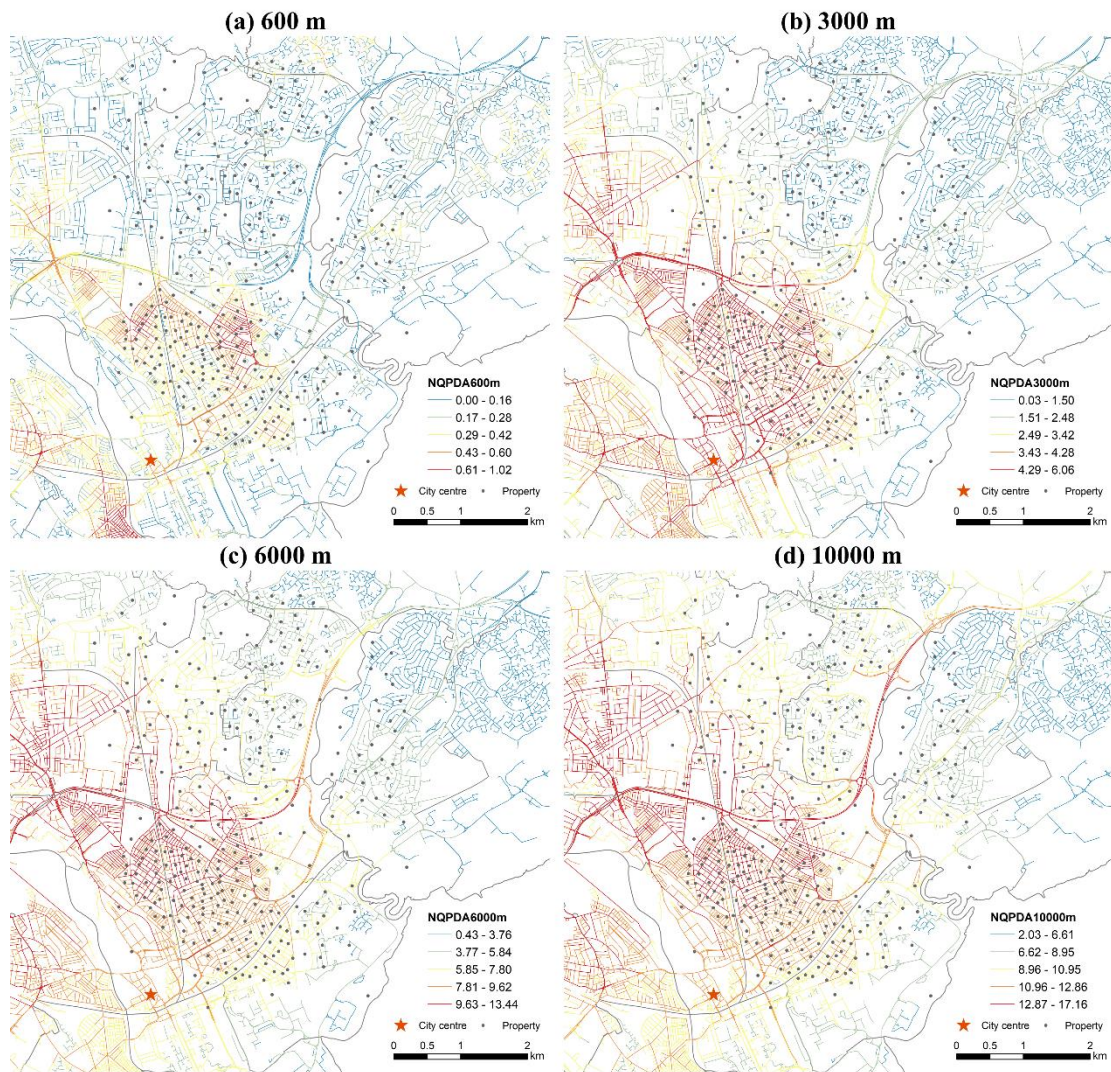


Fig. 1. The closeness of the street network at four different spatial scales

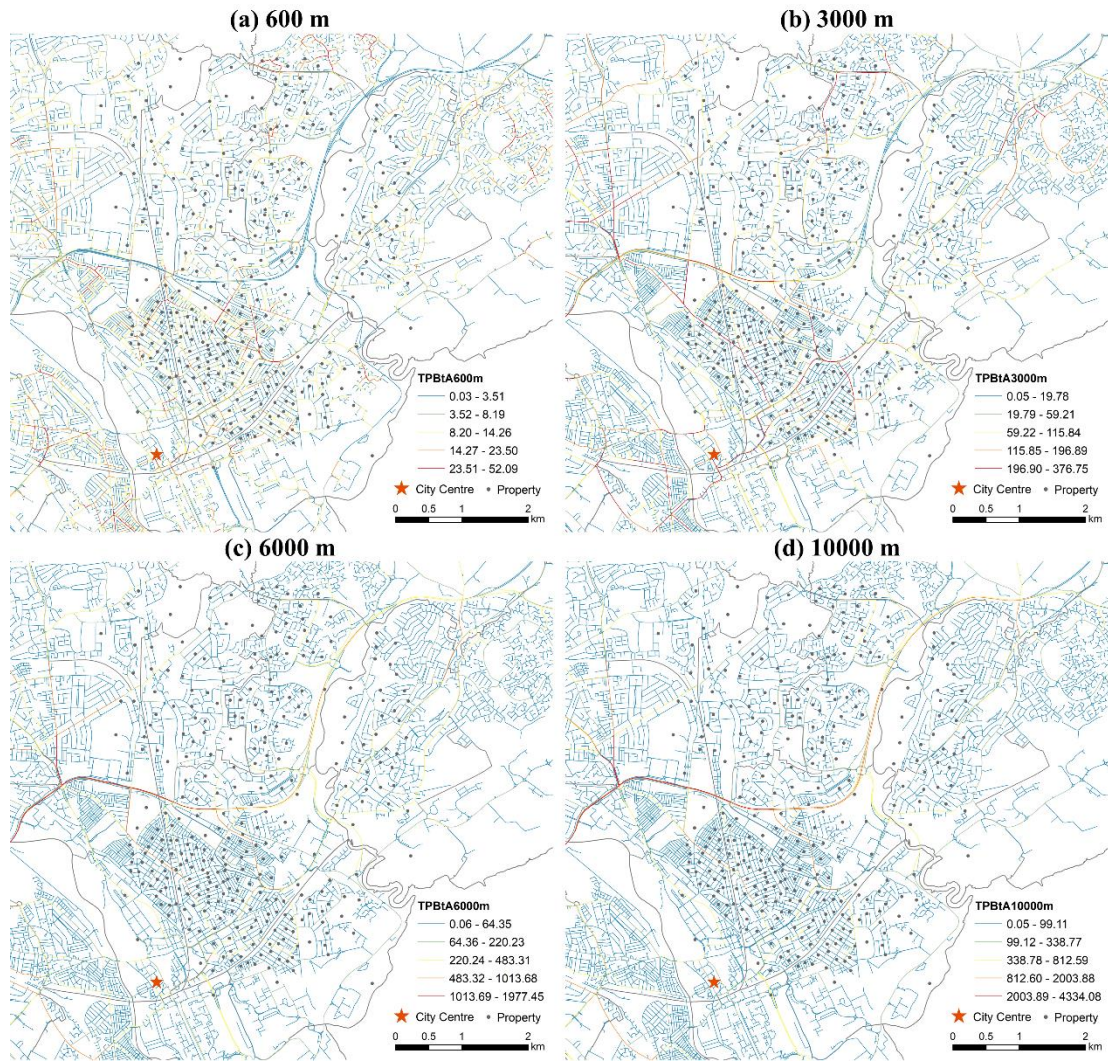


Fig. 2. The betweenness of street network at four different spatial scales

Table 1. Descriptive statistics for all variables.

Variable	Description	Mean	St.dev.
Dependent variable			
LN_PRICE	Natural log of house prices	11.67	0.57
Hedonic variables			
LN_FLOOR	Natural log of floor area	4.90	1.07
DU_NEW	New build (yes = 1, no = 0)	0.07	0.26
DU_DET	Detached house (yes = 1, no = 0)	0.09	0.28
DU_SEMI	Semidetached house (yes = 1, no = 0)	0.21	0.41
DU_TER	Terrace house (yes = 1, no = 0)	0.54	0.50
DU_FLAT	Flat (yes = 1, no = 0)	0.15	0.36

DU_TEN	Tenure (freehold = 1, leasehold = 0)	0.78	0.41
DU_BC	OAC blue-collar communities (yes = 1, no = 0)	0.11	0.31
DU_CL	OAC living in the city (yes = 1, no = 0)	0.27	0.45
DU_PS	OAC prosperous suburbs (yes = 1, no = 0)	0.15	0.35
DU_CC	OAC constrained by circumstances (yes = 1, no = 0)	0.05	0.21
DU_TT	OAC typical traits (yes = 1, no = 0)	0.28	0.45
DU_MU	OAC multicultural (yes = 1, no = 0)	0.15	0.35
TR_YRAR	Transactions year	2003.47	2.29
Dist_CBD	Network distance to the CBD (km)	2.57	1.44
Dist_Roath	Network distance to Roath Park (km)	2.27	0.87
Dist_Health	Network distance to Health hospital (km)	2.78	0.95
Street network variables			
NQPDA600	NQPDA at a radius of 600 m	0.27	0.16
TPBtA600	TPBtA at a radius of 600 m	4.17	4.42
NQPDA3000	NQPDA at a radius of 3,000 m	3.17	1.42
TPBtA3000	TPBtA at a radius of 3,000 m	19.08	38.36
NQPDA6000	NQPDA at a radius of 6,000 m	7.48	2.18
TPBtA6000	TPBtA at a radius of 6,000 m	30.92	79.39
NQPDA10000	NQPDA at a radius of 10,000 m	10.92	2.16
TPBtA10000	TPBtA at a radius of 10,000 m	41.33	134.14

Methodology

XGBoost was introduced by Chen and Guestrin (2016), the algorithm was designed to optimise the predictive performance, speed, and scalability of decision tree ensembles, making it suitable for various applications, including regression, classification, and ranking problems. This study compared the preliminary XGBoost model results with those obtained with the random forest and gradient-boosting decision tree models. The results confirmed that the XGBoost models outperformed the random forest and GBDT models regarding predictive accuracy and model fit.

In tuning hyperparameters, 90 percent of the dataset was randomly chosen for training the XGBoost model, while the remaining 10 percent was reserved for testing purposes. In addition, the study evaluated the XGBoost models with varying values for tree complexity (4, 8, 12, 16, 20, 22) and shrinkage (0.1, 0.05, 0.01) using a fivefold cross-validation procedure. The number of trees was examined by setting the indicator from 5,000 to 50,000 at intervals of 5,000. Ultimately, the XGBoost models gained the best hyperparameters with 50,000 trees, a learning rate of 0.1, and a maximum depth of 22 to prevent overfitting. As the results, the R^2 , MAE, and RMSE values for the XGBoost model predictions are 0.74-0.76, 0.18-0.19, and 0.28-0.30, respectively.

SHapley Additive exPlanations (SHAP) is one of the methods for interpreting complex machine learning models. SHAP values provide instance-level explanations, meaning they can help explain the output of a model for a specific input. SHAP can offer more local and granular insights into how features impact individual predictions. SHAP values also provide a global measure of feature importance, making it easier to compare the contributions of different features to the model's predictions (Lundberg and Lee, 2017). Therefore, this study used the XGBoost and SHAP models to investigate the nonlinear effects of street network accessibility on house prices.

Results

Relative importance

Table 2 demonstrates the relative significance and ranking of independent variables in predicting house prices. Across the four spatial scales, the ranking of most independent variables is generally consistent. The prosperous suburbs (DU_PS) stand

out as the most crucial factor in house prices, with the highest relative importance (RI) of 18.6% to 23.1%. The second most important variable is transaction year (TR_YEAR), followed by household tenure (DU_TEN), flat house (DU_FLAT) or detached house (DU_DET). Regarding the locational variables, the importance of distance to the CBD was similar to closeness accessibility, with a relative contribution of 2.9%-3.4% for distance to the CBD and 2.7%-3.9% for closeness (NQPDA). The importance of closeness accessibility becomes more critical as the spatial scale expands. This result aligns with the findings of a previous study (Xiao et al., 2016a), suggesting that the distance to the CBD is not a consistent factor influencing house prices. In contrast, closeness accessibility exhibits a more significant effect on house prices as analytical spatial scales increase. This could be attributed to the fact that people place a higher value on the convenience of being close to amenities such as schools, shops, public transportation, and recreational facilities, rather than simply being near the CBD, especially as the analytical spatial enlarges, closeness encapsulates a broader coverage of amenities. Moreover, the importance of betweenness accessibility (TPBtA) in predicting house prices is significantly lower than closeness accessibility across the four different spatial scales, accounting for only 2.0%-2.5% and ranking 12-13 among 15 variables. This confirms that street betweenness accessibility has a relatively weak relationship with house prices, capturing less potential locational externality effects by the metric.

Table 2. Estimation results of the XGBoost models.

	600 m		3,000 m		6,000 m		10,000 m	
	RI	Rankin g	RI	Rankin g	RI	Rankin g	RI	Rankin g
Hedonic variables								
LN_FLOOR	4.4%	8	4.2%	7	4.5%	6	4.7%	6
DU_NEW	1.5%	14	1.5%	14	1.6%	14	1.7%	14
DU_DET	7.1%	5	6.9%	5	7.2%	5	7.7%	4
DU_SEMI	2.5%	11	2.4%	11	2.8%	12	4.6%	7
DU_FLAT	9.2%	4	9.8%	4	8.5%	4	7.1%	5
DU_TEN	10.7%	3	11.3%	3	10.2%	3	13.6%	3
DU_BC	5.7%	6	6.8%	6	4.4%	7	3.9%	10
DU_PS	22.8%	1	22.2%	1	23.1%	1	18.6%	2
DU_CC	1.4%	15	1.0%	15	1.1%	15	1.0%	15
DU_TT	2.0%	12	2.3%	12	3.0%	11	2.1%	13
DU_MU	5.1%	7	4.0%	8	3.2%	10	4.5%	8
TR_YEAR	19.6%	2	19.1%	2	20.8%	2	21.1%	1
Dist_CBD	3.4%	9	3.1%	10	3.3%	9	2.9%	11
Street network variables								
NQPDA	2.7%	10	3.4%	9	3.9%	8	4.1%	9
TPBtA	2.0%	13	2.1%	13	2.4%	13	2.5%	12
Model performance								
R ²	0.76		0.76		0.76		0.74	
MAE	0.18		0.18		0.18		0.19	
RMSE	0.28		0.28		0.28		0.30	

RI: Relative Importance

Nonlinear relationship

Given the focus of this study, we only discuss the SHAP value of the variables of distance to the CBD, closeness accessibility, and betweenness accessibility. Fig. 3 depicts the marginal effects of distance to the CBD on predicting house prices across four different spatial scales in the models. All the plots display a nonlinear relationship

between the distance to the CBD and house prices, regardless of the spatial scales of analysis. Specifically, SHAP values of house prices decrease slightly as the distance to the CBD increases from 0 to 1.2 km. However, house prices increase within 1.2 km to 2.6 km as the distance to the CBD grows. Then, house prices slightly decrease and are sparse when the distance to the CBD exceeds 2.6 km. This indicates that the effects of distance to the CBD on house prices are not always consistent.

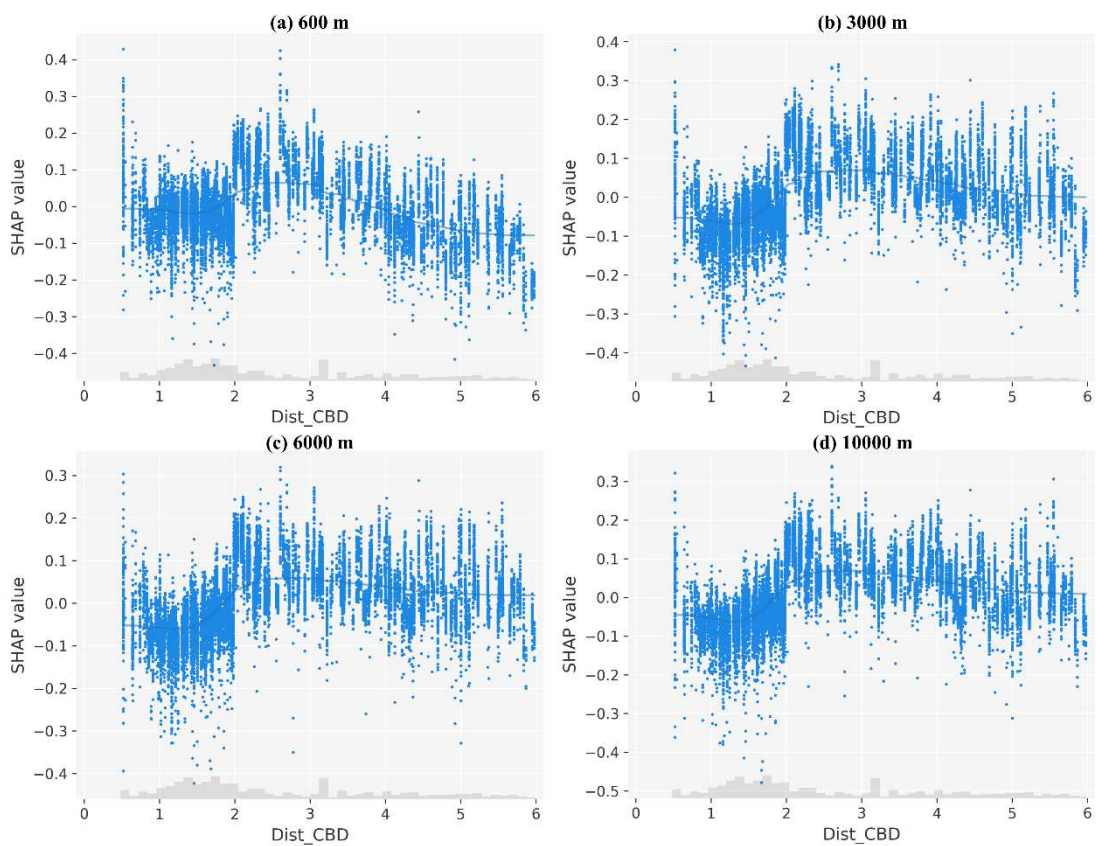


Fig. 3. The effects of distance to the CBD on house prices.

Fig. 4 displays the effects of closeness accessibility on house prices across four different spatial scales. All plots indicate a positive correlation between closeness accessibility and house prices for all spatial scales of analysis. This pattern aligns with our expectations and is consistent with Xiao et al. (2016a). Among the four different

scales, closeness accessibility within 600 m radii exhibits a more fluctuating pattern compared to the other three models. The results reveal a significant positive relationship between closeness accessibility and house prices from 0 to 0.2. Between 0.2 and 0.5, the curve remains constant. Once it exceeds 0.5, the relationship turns positive again. The other three models share a similar pattern: a generally upward curve with slight fluctuations. Compared with the curves of distance to the CBD, the results suggest that closeness accessibility is a more stable locational attribute factor for capturing locational externalities compared to the distance to the CBD.

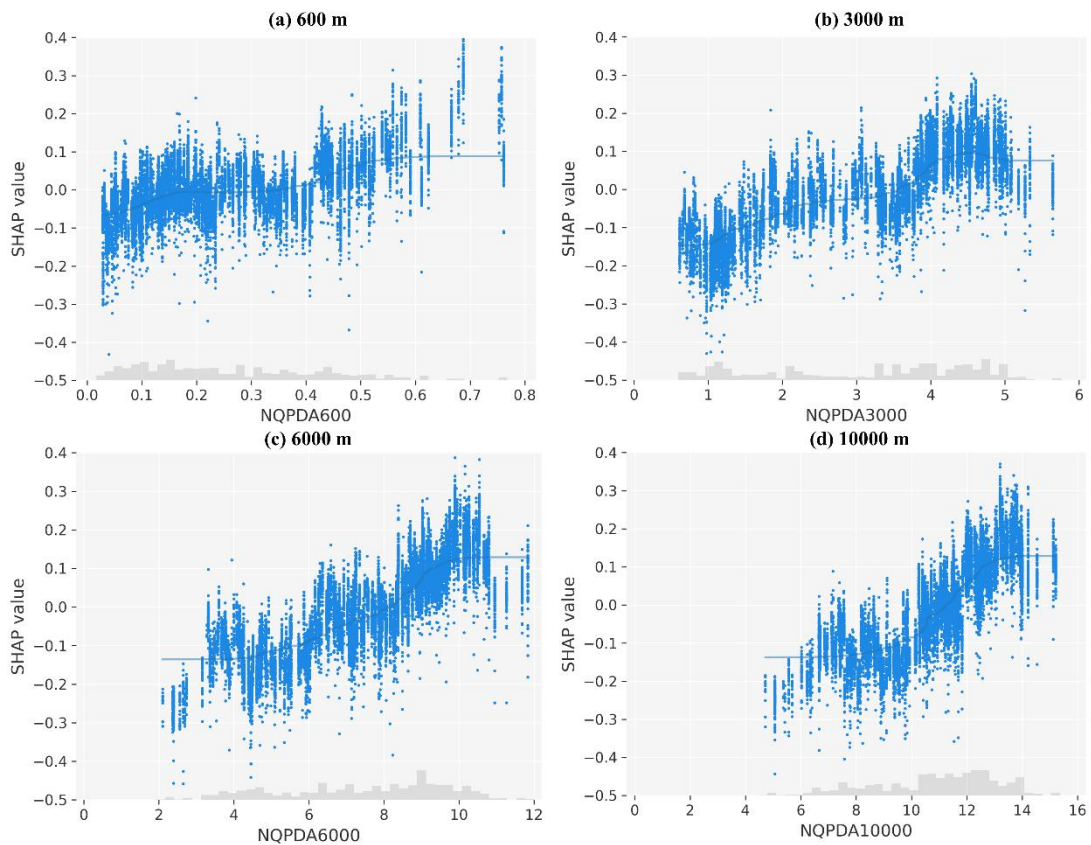


Fig. 4. The effects of NQPDA (closeness) on house prices.

The SHAP plots for betweenness accessibility are shown in Fig. 5. The relationship between betweenness accessibility and house prices, as opposed to closeness

accessibility, demonstrates a scattered curve with a mild downward tendency across all four different spatial scales. This may explain why betweenness accessibility holds less importance in predicting house prices in Table 2, because the metric tends to indicate the frequency with which a particular street route is likely to be utilised. It is worth noting, however, that the histogram on the horizontal axes reveals many properties with very low betweenness accessibility values, while only a small number of streets are prominent in terms of betweenness accessibility. This indicates significant variation in betweenness accessibility among properties.

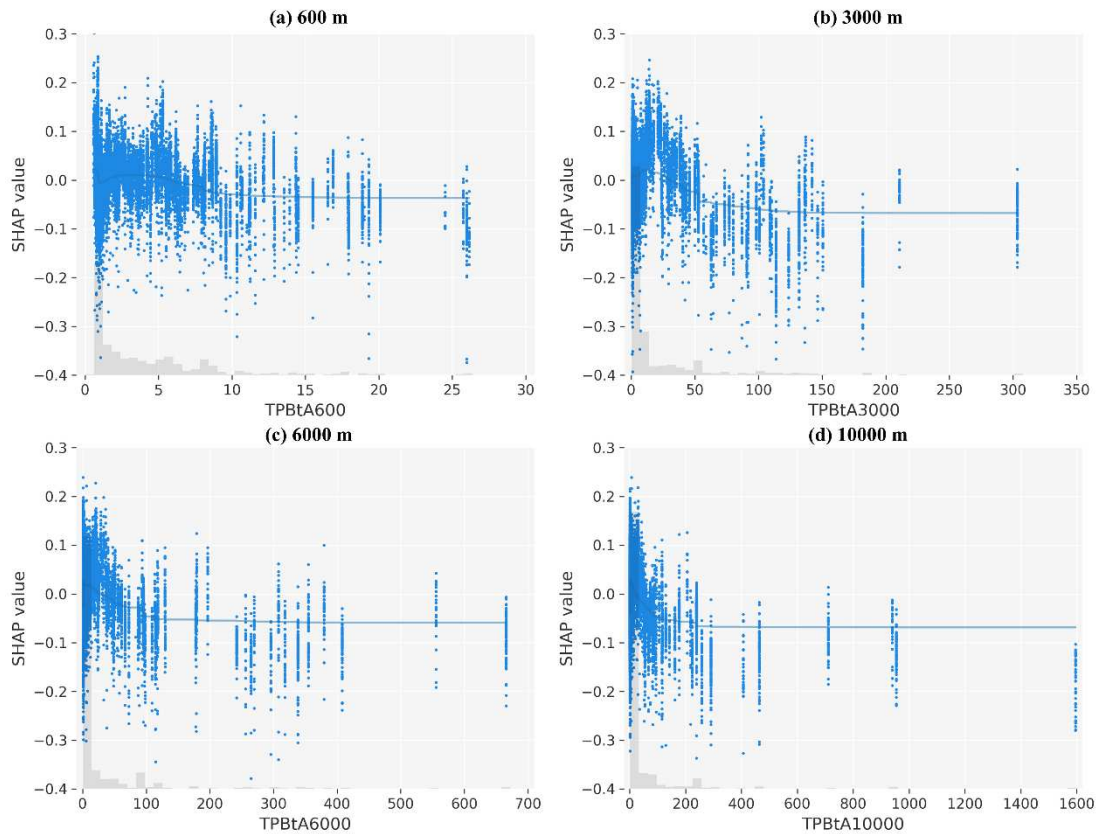


Fig. 5. The effects of TPBtA (betweenness) on house prices.

Discussion

Theoretically, without external influences such as natural environments and

government policies, land value is significantly dominated by accessibility. Land value increases for locations where travel costs for access have been reduced (Alonso, 1964; McDonald and Osuji, 1995). This theory has been continuously applied to housing research by incorporating accessibility as a substitute for hedonic factors in predicting house prices. Most literature uses specifications of locational attributes (e.g. distance to the CBD) as a proxy for accessibility in hedonic house price models, yet this measure can be problematic as (1) it is difficult to conceptualise and measure; (2) the effect of this measure tends to be nonlinear.

This study overcomes the limitations by using street-network accessibility measures and machine learning approaches to disentangle the complex relationships between accessibility and house prices. We found that the impact of distance to the CBD on house prices does not always follow a simple linear pattern (see Fig. 6), which is somewhat counterintuitive to our expectations and existing theory. Alonso's bid theory posits a negative correlation between the distance to the CBD and house prices, consistent with the access-space theory of land value. This theory has been empirically found in cities such as Shanghai (Zou et al., 2022), Nanjing (Xiao et al., 2016b), and Seoul (Kang, 2019).

A possible explanation may be heterogeneous preferences. In developed countries like the United States and Europe, residents prefer living in suburban areas due to the better environment and more spacious living spaces. Suburban areas tend to feature larger homes with more expansive yards than their urban counterparts and often boast abundant green spaces, parks, and outdoor recreational opportunities. This abundance

of space and nature, usually lacking in city centres, can contribute to higher housing prices in suburban and rural areas. Thus, the appeal of suburban living may often extend beyond the confines of the home itself, encompassing the surrounding environment and lifestyle opportunities (Clark and Huang, 2003).

Contrary to this, individuals and households residing in China and other East Asian countries typically prefer shorter commuting distances, easy access to public transportation, and proximity to amenities for daily needs (Li et al., 2019; Wang and Lin, 2014). This preference persists despite the negative externalities that high-density living may bring, such as crowded conditions, noise pollution, and less appealing natural environments. These conditions have shaped the built environment and urban landscape they inhabit today and have cultivated a greater tolerance for high-density living than their counterparts in Western countries (Wu, 2015).

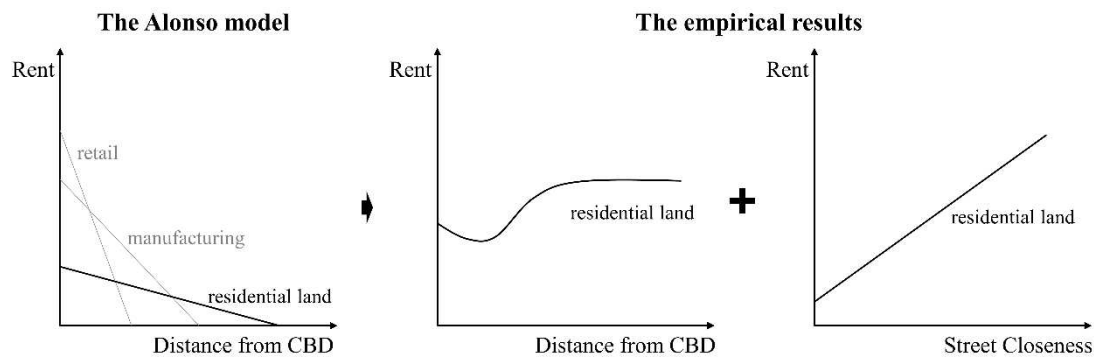


Fig. 6. The Alonso model and the empirical results.

Additionally, this study finds that street closeness accessibility consistently contributes to increased house prices, demonstrating a stable upward curve. This result substantiates the advantage of using measures of street closeness accessibility as more robust predictors of locational valuations compared to proximity to the CBD (Ozuduru

et al., 2021; Pan et al., 2018; Webster, 2010). There are two possible explanations for this outperformance. Firstly, street closeness accessibility often indicates how readily and widely residents can access services and opportunities, including jobs, schools, hospitals, shops, and recreational facilities. As a result, increased accessibility to these amenities enhances a property's desirability, leading to higher housing prices. According to Space Syntax theory, more integrated spaces tend to be more frequently used, which could attract more services and amenities and ultimately lead to higher property values (Hillier and Hanson, 1984). In comparison, the integration of locational attributes into hedonic house price models can be unstable, considering their wide range and the inconsistency in their identification and measurement (Xiao et al., 2016a).

Secondly, as one of the foundational theories in neo-classical microeconomics, Alonso's bid-rent theory evolved from von Thünen's model to emphasise an 'access-space' trade-off. This model describes a balance between transportation costs and land rents, developed under the presumption of a monocentric city situated on an isotropic plane with a perfectly competitive housing market. However, this assumption does not fully capture the complexities of cities. The predetermined specification for locations may not adequately represent 'access' in the context of urban systems. In contrast, using a closeness street network for accessibility measurements, employing aggregated algorithms to calculate potential opportunities or attractiveness, offers a more appropriate approach. Because this method considers both the ease of interaction and the availability of opportunities (Hansen, 1959: 73), which been intricately woven into the fabric of the urban street network (Hillier and Hanson, 1984).

To validate our main findings, we further utilised direct line distance to measure proximity to the CBD and geometric measures for street network accessibility in our robustness analysis. The analysis revealed consistent nonlinear relationships between proximity to the CBD and house prices, and an **upward** pattern between street closeness centrality and house prices. This robustness analysis confirms the main findings.

However, while street network accessibility is a useful metric for measuring accessibility, it is crucial to recognise that these metrics may encounter issues related to distance, scale, and the selection of transport networks (Law et al., 2012; Ratti, 2004), which has long been discussed these issues in Space Syntax research (Pafka et al., 2020). Our study attempted to mitigate potential biases using various scales and distance measures. Nonetheless, recognising and addressing these concerns is still crucial, as it greatly determines the effectiveness of this tool in diverse urban planning and analysis contexts.

Conclusions

By utilising the XGBoost and SHAP techniques and taking Cardiff as a case study, this study investigated the nonlinear effects of street network accessibility on house prices across different spatial scales. The study offers two novel contributions to the literature and insights into house policies. First, this study relaxes the assumption of generalised linearity and confirms the nonlinear effects of accessibility metrics on house prices. The results show that the effects of the distance to the CBD on house prices demonstrate a nonlinear relationship regardless of analytical scales. This finding potentially challenges the classic Alonso's bid rent theory, which suggests a downward-

sloping land value curve with proximity to the CBD. Our findings suggest that homebuyers and developers should not solely focus on proximity to the CBD as the primary determinant of house price evaluation.

Second, this study reveals a more stable pattern in the relationship between house prices and street closeness accessibility compared to proximity to the CBD. The effects of street closeness accessibility on house prices demonstrate a generally upward curve with slight fluctuations regardless of analytical spatial scales. In contrast, the effect of the street betweenness accessibility metric on house prices only shows a scattered curve with a mild downward tendency across all four different spatial scales. Webster (2010) emphasized the concept of "capitalized" accessibility within the street network. Building on this foundation, our study delves deeper, exploring the specific aspects of street network accessibility and the extent to which accessibility contributes to the capitalized value of the street network.

From a policy perspective, the study highlights that enhancing street closeness accessibility in areas inhabited by low-income households could be an effective strategy to address spatial inequality (Jin et al., 2022), local governments could incorporate street network accessibility into their affordable housing initiatives. For example, legislative mechanisms could be employed to encourage developers and investors to construct affordable housing, such as a prescribed percentage of homes or apartments, within areas of high street network accessibility. Another innovative approach could be the introduction of an infrastructure levy (Shanahan et al., 2019), which allows local authorities to impose a tax on development projects that could fund the enhancement

of the accessibility of the surrounding street network. This method alleviates the government's responsibility for funding public projects and provides economic benefits for developers through enhanced accessibility.

Several questions warrant further investigation. First, this study was conducted in Cardiff, a city characterised as monocentric and post-industrialised, the conclusions and findings may primarily apply to cities with similar urban structures, population levels, and household living preferences. To broaden the applicability of our research, future studies are encouraged to utilise the same framework and techniques in larger cities with more intricate urban configurations. Second, this study only considers the relationship between street network accessibility and house prices. It would be interesting to investigate how commercial, industrial, and other economic activities have been affected by street network accessibility, helping update Alonso's bid-rent theory and other classic urban theories in contemporary cities.

Acknowledgments

The authors also like to thank the reviewers for their invaluable comments and the editors for their editorial help and timely response.

Declaration of competing interest

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was sponsored by the National Natural Science Foundation of China (No.71734004) and the International Exchange Program for Graduate Students, Tongji University (No. 2023020023).

References

- Ahlfeldt, G.M., Maennig, W., 2010. Substitutability and Complementarity of Urban Amenities: External Effects of Built Heritage in Berlin: Substitutability and Complementarity of Urban Amenities. *Real Estate Economics* 38, 285–323. <https://doi.org/10/dqfngf>
- Alonso, W., 1964. *Location and Land Use: Toward a General Theory of Land Rent*. Harvard University Press. <https://doi.org/10.4159/harvard.9780674730854>
- Batty, M., 2013. *The New Science of Cities*. The MIT Press. <https://doi.org/10.7551/mitpress/9399.001.0001>
- Can, A., 1992. Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics* 22, 453–474. <https://doi.org/10/dhvqvqf>
- Chakrabarti, S., Kushari, T., Mazumder, T., 2022. Does transportation network centrality determine housing price? *Journal of Transport Geography* 103, 103397. <https://doi.org/10/g3xmk>
- Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Presented at the KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA, pp. 785–794. <https://doi.org/10/gdp84q>
- Clark, W.A.V., Huang, Y., 2003. The Life Course and Residential Mobility in British Housing Markets. *Environ Plan A* 35, 323–339. <https://doi.org/10/bf2hhg>
- Cooper, C., Chiaradia, A.J., 2015. sDNA: how and why we reinvented Spatial Network Analysis for health, economics and active modes of transport, in: Malleon, Nick Ed. *GIS Research UK (GISRUK) 2015 Proceedings*. Presented at the GISRUK, Leeds, pp. 122–127.
- Cooper, C.H.V., Chiaradia, A.J.F., 2020. sDNA: 3-d spatial network analysis for GIS, CAD, Command Line & Python. *SoftwareX* 12, 100525. <https://doi.org/10/gg6ssv>
- Fujita, M., Krugman, P.R., Venables, A., 2001. *The spatial economy: cities, regions, and international trade*, 1. paperback ed. ed. MIT Press, Cambridge, Mass.
- Hansen, W.G., 1959. How Accessibility Shapes Land Use. *Journal of the American Institute of Planners* 25, 73–76. <https://doi.org/10/cb2mht>
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning*, Springer Series in Statistics. Springer New York, New York, NY. <https://doi.org/10.1007/978-0-387-84858-7>
- Hillier, B., 2007. *Space is the machine: a configurational theory of architecture*. Space Syntax.
- Hillier, B., Hanson, J., 1984. *The Social Logic of Space*, 1st ed. Cambridge University Press. <https://doi.org/10.1017/CBO9780511597237>

- Hillier, B., Iida, S., 2005. Network and Psychological Effects in Urban Movement, in: Cohn, A.G., Mark, D.M. (Eds.), *Spatial Information Theory, Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 475–490. https://doi.org/10.1007/11556114_30
- Jin, T., Cheng, L., Liu, Z., Cao, J., Huang, H., Witlox, F., 2022. Nonlinear public transit accessibility effects on housing prices: Heterogeneity across price segments. *Transport Policy* 117, 48–59. <https://doi.org/10/gr8dzz>
- Kang, C.-D., 2019. Effects of spatial access to neighborhood land-use density on housing prices: Evidence from a multilevel hedonic analysis in Seoul, South Korea. *Environment and Planning B: Urban Analytics and City Science* 46, 603–625. <https://doi.org/10/gddtg7>
- Law, S., Chiaradia, A., Schwander, C., 2012. Towards a multimodal space syntax analysis: A case study of the London street and underground network, in: *Published in Proceedings of the 8th International Space Syntax Symposium*. Santiago de Chile, Chile.
- Li, H., Wei, Y.D., Wu, Y., Tian, G., 2019. Analyzing housing prices in Shanghai with open data: Amenity, accessibility and urban structure. *Cities* 91, 165–179. <https://doi.org/10/ghtmdn>
- Li, Z., 2022. Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. *Computers, Environment and Urban Systems* 96, 101845. <https://doi.org/10/gqh3tt>
- Liu, X., Chen, X., Potoglou, D., Tian, M., Fu, Y., 2023a. Travel impedance, the built environment, and customized-bus ridership: A stop-to-stop level analysis. *Transportation Research Part D: Transport and Environment* 122, 103889. <https://doi.org/10.1016/j.trd.2023.103889>
- Liu, X., Chen, X., Tian, M., De Vos, J., 2023b. Effects of buffer size on associations between the built environment and metro ridership: A machine learning-based sensitive analysis. *Journal of Transport Geography* 113, 103730. <https://doi.org/10.1016/j.jtrangeo.2023.103730>
- Lundberg, S., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. <https://doi.org/10/gp6hf4>
- McDonald, J.F., Osuji, C.I., 1995. The effect of anticipated transportation improvement on residential land values. *Regional Science and Urban Economics* 25, 261–278. <https://doi.org/10/btk3r>
- Orford, S., 2002. Valuing Locational Externalities: A GIS and Multilevel Modelling Approach. *Environ Plann B Plann Des* 29, 105–127. <https://doi.org/10/fqtg55>
- Orford, S., 2000. Modelling Spatial Structures in Local Housing Market Dynamics: A Multilevel Perspective. *Urban Studies* 37, 1643–1671. <https://doi.org/10/fqcr8v>
- O’Sullivan, T., Gibb, K. (Eds.), 2002. *Housing Economics and Public Policy*, 1st ed. Wiley. <https://doi.org/10.1002/9780470690680>

- Ozuduru, B.H., Webster, C.J., Chiaradia, A.J.F., Yucesoy, E., 2021. Associating street-network centrality with spontaneous and planned subcentres. *Urban Studies* 58, 2059–2078. <https://doi.org/10/ghwr2g>
- Pafka, E., Dovey, K., Aschwanden, G.D., 2020. Limits of space syntax for urban design: Axiality, scale and sinuosity. *Environment and Planning B: Urban Analytics and City Science* 47, 508–522. <https://doi.org/10/gdzzrt>
- Pan, H., Deal, B., Chen, Y., Hewings, G., 2018. A Reassessment of urban structure and land-use patterns: distance to CBD or network-based? — Evidence from Chicago. *Regional Science and Urban Economics* 70, 215–228. <https://doi.org/10/gds3h8>
- Pope, D.G., Pope, J.C., 2015. When Walmart comes to town: Always low housing prices? Always? *Journal of Urban Economics* 87, 1–13. <https://doi.org/10/f7ch8d>
- Ratti, C., 2004. Space Syntax: Some Inconsistencies. *Environ Plann B Plann Des* 31, 487–499. <https://doi.org/10/bqz6jf>
- Rodríguez, D.A., Mojica, C.H., 2009. Capitalization of BRT network expansions effects into prices of non-expansion areas. *Transportation Research Part A: Policy and Practice* 43, 560–571. <https://doi.org/10/fv8zvn>
- Rosen, S., 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy* 82, 34–55. <https://doi.org/10/dghdwb>
- Shanahan, D., Astell–Burt, T., Barber, E., Brymer, E., Cox, D., Dean, J., Depledge, M., Fuller, R., Hartig, T., Irvine, K., Jones, A., Kikillus, H., Lovell, R., Mitchell, R., Niemelä, J., Nieuwenhuijsen, M., Pretty, J., Townsend, M., Van Heezik, Y., Warber, S., Gaston, K., 2019. Nature–Based Interventions for Improving Health and Wellbeing: The Purpose, the People and the Outcomes. *Sports* 7, 141. <https://doi.org/10/dppr>
- Wang, D., Lin, T., 2014. Residential self-selection, built environment, and travel behavior in the Chinese context. *JTLU* 7, 5–14. <https://doi.org/10/gtdvzq>
- Webster, C., 2010. Pricing accessibility: Urban morphology, design and missing markets. *Progress in Planning* 73, 77–111. <https://doi.org/10/bmmn6m>
- Wu, F., 2015. *Planning for growth: urban and regional planning in China*, The RTPI library series. Routledge, New York London.
- Xiao, Y., 2017. *Urban Morphology and Housing Market*, Springer Geography. Springer Singapore, Singapore. <https://doi.org/10.1007/978-981-10-2762-8>
- Xiao, Y., Orford, S., Webster, C.J., 2016a. Urban configuration, accessibility, and property prices: a case study of Cardiff, Wales. *Environ Plann B Plann Des* 43, 108–129. <https://doi.org/10/f77g7b>
- Xiao, Y., Webster, C., Orford, S., 2016b. Identifying house price effects of changes in urban street configuration: An empirical study in Nanjing, China. *Urban Studies* 53, 112–131. <https://doi.org/10/gtdvzr>

Zou, G., Lai, Z., Li, Y., Liu, X., Li, W., 2022. Exploring the nonlinear impact of air pollution on housing prices: A machine learning approach. *Economics of Transportation* 31, 100272. <https://doi.org/10/gtdvzs>