Study of the Relationship between the Street Centrality and the Housing Market in Greater London

Haofu Wang

CASA0012 for MSc Dissertation

Supervised by Prof. Michael Batty

Word Count: 9,518

This dissertation is submitted in part

requirement for the degree of

MSc in the Centre for Advanced Spatial Analysis,

Bartlett Faculty of the Build Environment, UCL.

5 September 2022

Abstract

This paper investigates whether there is a relationship between the street centrality and the housing price in the Greater London area. The average closeness centrality index and average betweenness centrality index have been calculated in a distance-weighted driving street network and taken as part of the locational variables in the hedonic pricing model to estimate the influence on the average housing price of Middle Super Output Area (MSOA). The hedonic pricing model is firstly estimated as an OLS regression model. The average closeness centrality index shows a strong influence on the housing price whereas the average betweenness centrality has an opposite result. The spatial autocorrelation has also been studied with the Spatial Lag model. The R-squared value in the OLS regression model and the rho value in the Spatial Lag regression model proves that the models are fitting well. Overall, the research questions have been studied and this research would not only fill the relevant research gap in the Greater London area but help have a more thorough understanding of the urban characteristics of the city. Multiple limitations have been discussed at last and the future direction of the research has been outlooked.

Declaration of Authorship

I, Haofu Wang, hereby declare that this dissertation is all my own original work and that all sources have been acknowledged. It is 9,518 words in length.

Haofu Wang

Haofu Wang

5 September 2022

Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

CBD	Central Business District
MSOA	Middle Layer Super Output Area
OLS	Ordinary Least Squares

Acknowledgements

Firstly, I would like to express my very great appreciation to my supervisor, Prof. Michael Batty for his constant encouragement during the hard times, valuable and constructive suggestions, and selfless teaching of knowledge during the entire project. My grateful thanks are also extended to Prof. Elsa Arcaute for her great help in answering the questions and guiding me for the research topic.

Secondly, I would like to thank all other staff members in the Centre for Advanced Spatial Analysis, Faculty of the Built Environment, University College London, for their extraordinary work and all the academical and social activities they prepared alongside this wonderful year of this MSc program.

Thirdly, great thanks to all the friends and classmates who gave me valuable help in answering the questions and encouraged me throughout the entire project.

Lastly, my thanks would go to my beloved family for their support and great confidence in me all through this year. I would also like to show my heartful gratitude to my girlfriend Yiying for her continuous support and loving considerations.

1. Introduction

1.1. Context

According to the Maslow's Theory, housing, as one of the very basic needs of human beings, has been a popular study topic by the scholars, especially the urban researchers and geographers who had a long history in the exploration of determining the housing price. Since Rosen (1976) utilized the hedonic pricing model over the housing market, a large number of the researchers selected the variables in the field of location, neighbourhood and structure and take them as the characteristics of the housing to estimate the housing prices (Mok et al., 1995, Wen et al., 2005). Among the attributes of the housing price, the locational variables have always played an important role. Due to the monocentric urban theory, it is a common sense that there is an equilibrium between the distance to the city centre and the housing market distribution. Therefore, the locational factors such as the distance to CBD and the accessibility and have been widely studied and proved to have a solid influence on the housing price in some large metropolis areas such as San Francisco Bay Area and Cardiff, Wales. (Maria Kockelman, 1997, Xiao et al., 2016a).

It is believed that the social and economic activities are highly related to the urban street network (Tan et al., 2019). The populated areas and the commercial centres are always located in the streets with important values. The street network is highly structured with the nodes and edges. In this case, nodes would be the junctions of two streets and edges would be the street between two junctions. With the help of the centrality indices, it is able to identify the street network characteristics by locating the influencers (nodes) which have huge impact on the entire network. The type of the street network is decided by the selection of the weights. A distance-weighted street network would be focus on the structure of the network itself, whereas a flow-weighted street network would be connected with the real-life scenario.

In this research, the centrality indices would be taken as one of the locational variables to estimate the housing price in the Greater London area. The detailed research question would be elaborated in the following section.

1.2. Research Questions and Objectives

The research would be studied throughout two questions. The first is that whether the centrality indices of the street network would have any influence on the housing price. Then based on the first question, a progressive second question would be raised that whether there is any further relationship between the centrality indices of the street network and the housing price from the spatial perspective. Therefore, the aim of this research would be estimated the relationship between the centrality indices of the street network and the housing price in the Greater London area. A further spatial relationship would also be involved in the analysis of the study.

In regard to the research questions, a series of study has been spread out. In Section 2, the relative literatures about the hedonic pricing model and the centrality indices would be reviewed. In Section 3, there would be a detailed elaboration of the methodology involved in this research. Section 4 presents the results analysed based on the methodology introduced in Section 3. Section 5 would discuss the results on the basis of the research questions. At last, Section 6 would summarize the entire study, and the limitations and a further outlook of the future work would also be covered in this section.

2. Literature Review

2.1. Hedonic Pricing Model

As one of the most famous econometric models to value the housing price, the hedonic pricing model has been developed for nearly a hundred years.

The approach of the model is using the regression method to estimate the independent characteristics of the property and analyse the importance of these attributes on the value measurement of the property. The function was simply expressed to be:

$P = f(L, S\ and\ N, \gamma)$, (1)

where L is the locational factors, S is the structural factors and N is the neighbourhood factors, meantime γ stands for the coefficients of the attributes of the price which shows the reason why it is called hedonic prices(Abidoye and Chan, 2017, Mok et al., 1995, Maurer and Pitzer, 2004).

Although there is still a controversial issue on who was the first on the hedonic Study, it is commonly believed that the concept of the 'hedonics' was first raised by Haas (1922) in the monograph. (Wen et al., 2005, Colwell and Dilmore, 1999). The idea of the distance to the city centre and the size of the city played an important role in this urban economics paper (Colwell and Dilmore, 1999). However, it was until 1939 that A. T. Court established and used the first hedonic pricing model on the automobiles to explain the relationship between the demands and multiple variables (Colwell and Dilmore, 1999, Wen et al., 2005, Herath and Maier, 2010, Chau and Chin, 2003). Thereafter, Lancaster (1966) and Rosen (1976) built the theoretical foundation which took the hedonic pricing model further (Wen et al., 2005). Lancaster (1966) pointed out that it is the characteristics of the product instead of the product itself that determine the demands. As a result, according to his theory, it is a large group of prices (hedonic prices) that comprises the price of the housing property. To measure the value of the housing property, all the characteristics that the property attributes have to be taken into consideration. Later, Rosen (1976), an American economist, established the equilibrium model between the demand of the products and the characteristics of the products which is the foundation of the modelling to measure the demand of the characteristics (Wen et al., 2005).

On the basis of Rosen's work, a large amount of theoretical and empirical research work using hedonic pricing model emerged, and the modern hedonic pricing model began to be widely used in the field of the housing estate (Abidoye and Chan, 2017).

With the help of the hedonic pricing model, it is convenient for the scholars to explore the inner pattern of the housing estate market in many cities around the world such as Hong Kong (Mok et al., 1995), Seoul (Huh and Kwak, 1997), Hangzhou (Wen et al., 2005), Paris (Maurer and Pitzer, 2004), Perth (Pandit et al., 2013) and Aveiro (Bhattacharjee et al., 2012).

Although there is no requirement for the regression function form of the hedonic pricing model according to the theory (Bhattacharjee et al., 2012), most of the scholars chose among three kinds of the regression model including linear, logarithm and logarithm-linear with the traditional locational, structural and neighbourhood attributes. In this research, the regression model would be conducted in linear specification and logarithm-linear specification.

Among the reviewed literature, Henry M.K. Mok et al. and Raimond Maurer et al. derived from the Box-Cox transformation and estimate the regression model using the attributes with the characteristics. They chose the attributes from the locational, structural and neighbourhood aspects. Distance to CBD, age of the building, story of the building were all included in the analysis. Henry M.K. Mok et al. selected the sea view and school zone as the characteristics of Hong Kong and Raimond Maurer et al. put the additional emphasis on the structural dataset such as number of gardens, elevators, and terrace. The work of Wen's (2005) and the work of Randeniya's (2017) are also involved with OLS methods. Starting with the correlation matrix, it would be able for the researchers to pick out the factors that have larger correlation index with the housing price. As a result, they concluded the hedonic pricing model they estimated was suitable for the estimation of the housing market in Sri Lankan. Wen et al. selected 18 variables which can be divided into the traditional structural, neighbourhood and locational characteristics after combining with the real-life scenario in Hangzhou. In this research, the OLS regression model involving the linear specification and the logarithm-linear specification was selected for the estimation of the hedonic pricing model.

Some of other work put the emphasis on the green infrastructure. For example, Ram Pandit et al. (2013) conduct the research on the relationship between the street trees and the housing price in Perth. When conducting the OLS regression derived from the Box-Cox transformation, they take the number of multiple varieties of trees into account. They also extended the research to the presence of the spatial dependence. Lagrange multiplier and robust Lagrange multiplier were utilized to test for the spatial error and spatial lag. A. Bhattacharjee et al. (2012) studied the spatial interactions as well. They conducted the spatial weights matrix to test the Spatial Lag and Spatial Error dependence.

2.2. Network and Centrality

The definition for the simplest form of the network is that a group of points (nodes) connected together by the lines (edges) (Newman and Oxford University, 2010). From the physics to the sociology, network is a method that can help to find out the inner pattern of a complicated system (Newman and Oxford University, 2010). When it comes to the real-life scenario, a node could represent each person, each station, or each junction between two roads, whilst an edge would be the relationship between two people, the underground line between two stations or the road between two junctions. The capacity or strength that the edge possess would be the weights of the network (Chakrabarti et al., 2022). To analysis a network, centrality would always be one of the most effective and fundamental methods to quantify the main influencer which has been attached to the highest importance in a social network (Latora and Marchiori, 2007). Since Bavelas (1948) firstly applied the idea of the centrality to the communication between humans, multiple varieties of the centrality measures were introduced to analysis the importance of the nodes in the network. In this study, the centrality measures involved are closeness centrality and betweenness centrality due to their correspondence of the research topic of the urban street network. The definitions are as follows.

The closeness centrality index using the measurement of the average shortest distance to identify the node which is relatively closer to every other nodes in the network and is defined by Sabidussi (1966) and Scott (2000) as (Latora and Marchiori, 2007, Strano et al., 2013):

$$C_i^C = (L_i)^{-1} = \frac{N-1}{\sum_j d_{ij}}, \text{(2)}$$

Where L_i stands for the mean geodesic distance from node i to other nodes, N is the number of the nodes, d_{ij} stands for the minimum number of the edges between i and j. Since the definition of the Closeness Centrality is on the basis of the inverse of the mean geodesic distance, the node with the higher closeness centrality index locates geographically closer with other nodes and would be more influential to others as well in the urban street network.

The betweenness centrality index analyses the shortest path between the nodes, and by counting the number of shortest paths passing through each node, it is able to indicate the level of interaction among the network's nodes. The node with the highest betweenness centrality index typically sits in the centre of two or more node clusters. It is defined by Freeman (1977) and Wasserman and Faust (1994) as (Latora and Marchiori, 2007, Strano et al., 2013):

$$C_i^B = \frac{1}{(N-1)(N-2)} \sum_{j \in G, j \neq i} \sum_{k \in G, k \neq i, k \neq j} \frac{n_{jk}(i)}{n_{jk}}. \text{(3)}$$

Due to its characteristics that has the ability to identify the role of a broker in the system, the betweenness centrality has been proved to be excellent within the research of the urban street network (Kirkley et al., 2018).

The advantage that using centrality measures to study the urban network is quite obvious. For example, Porta et al. (2013) calculated the closeness, straightness and betweenness centrality to compared ten European cities for the urban street networks. Crucitti et al. (2006) studied the centrality measures in the urban street network of different cities around the world. With the clustering analysis of the centrality distribution, they distinguished the difference between the various classes of the cities.

From street perspective (Crucitti et al., 2006b) to the public transportation (Chakrabarti et al., 2022), centrality measures were always taken to be one of the essential tools to understand the structural characteristics of the complex networks (Crucitti et al., 2006a). Literature suggests that since the centrality is one of the decisive factors that influences the locations, it would play very important role in the research of the retails due to the high correlation between the street networks and the economic activities (Porta et al., 2012, Crucitti et al., 2006b) The higher centrality index would indicate the more central neighbourhood where could be the key factor of the placement of the retail stores (Lin et al., 2018). Lin et al. (2018) found that most of the shopping malls' and convenience stores' locations highly depend on the closeness centrality in Guangzhou (Lin et al., 2018).The betweenness and straightness centrality are relatively less correlated with the stores' locations (Lin et al., 2018). However, some of the researchers tend to have different results of the correlation over the centrality measures in different cities. Porta et al. (2009) explored the street centrality and the commercial and service activities in Bologna. They concluded that the areas with better centralities would indicate a more concentrate retail and service activities in Bologna (Porta et al., 2009). They found that the betweenness centrality of the street network would have a higher correlation index than the closeness centrality (Porta et al., 2009). In 2011, Porta et al. did another similar research in Barcelona. Closeness centrality tends to have slightly more impact to the location of the retails than betweenness centrality and straightness centrality (Porta et al., 2012). At the end of the research, the researchers found that the economic activities would not be influenced by the three centrality measures separately, but in a mutual way (Porta et al., 2012). In conclusion, the centrality of the street network has a large impact with the locations of the economic activities, but it would also vary among the different cities. The reason of this phenomenon could be the regional characteristics related to the cities' size and the history background (Lin et al., 2018).

2.3. *Centrality Indices in Hedonic Pricing Model*

Owing to the high correlation between the street centrality and the locations of the economic activities, the scholars started to have the research on the relationship between the street centrality and the price of the housing estate property. To fit the hedonic pricing model, they would interpret the centrality as one of the explanations of the accessibility (Xiao et al., 2016b). The theory summarized by Batty (2009) is that one of the accessibilities is defined on the basis of the shortest path between the components in the network which according to the definition, is the centrality measure.

The studies between the accessibility and the housing price are plentiful. Xiao et al. (2016) conducted the research over the use of the network accessibilities and the hedonic pricing model in Cardiff, Wales. To reach for the best accuracy, the researchers calculated the centrality to measure the network accessibilities. With the advantage of the reduction in heteroscedasticity, spatial autocorrelation, and multicollinearity, they concluded the measurement of the accessibility in the urban network can make the improvement of the model's performance (Xiao et al., 2016a). Later that year, Xiao et al. connected the attributes of the housing price with the accessibilities of the street network and tried to find the spatial relationship in Nanjing, China. They concluded that the positive hypothesis of the correlate relationship between the accessibility and the housing price is valid in Nanjing, a densely populated Asian city (Xiao et al., 2016b). Kang et al. (2019) combined the housing price, street network, and the land use density to measure how much the citizens valued the urban accessibility and centrality to the neighbourhood, commercial centre, office building. They concluded that the property price in Seoul would increase with the higher accessibility and centrality to the residential and commercial centre (Kang, 2019). When it comes to the office building and industrial area, it would have less or negative effect (Kang, 2019).

Across the studies, it is widely acknowledged by the scholars that the relationship between the housing price and the accessibility or centrality would vary among the cities. In Cardiff (Xiao et al., 2016a), it was found that the closeness centrality would have a positive relationship with the housing price, but the betweenness centrality had the opposite result. The same conclusion was also made in Nanjing (Xiao et al., 2016b). However, in Seoul, the betweenness centrality would had the positive relationship (Kang, 2019).

2.4. Research Gap

Throughout the literature review, it is found that most of the case studies involving the relationship between the centrality indices and the housing price are focused on the traditional hedonic pricing model (OLS regression model). Few of them pay attention to the spatial dependences and compare it with the traditional one.

Since the inspiration of the research is to not only cover the existing body of knowledge about this topic but extend to the new perspective, the relationship between the centrality indices of the street network and the housing price would be firstly estimated by means of the traditional hedonic pricing model. Then, the spatial dependences of the housing price would be checked and one of the spatial models would be selected to deal with the spatial autocorrelation.

The study areas involved in the past literature are also limited. There are few scholars analysing the relationship between the street network centrality and the housing price in the Greater London area. As a result, extending the study area to the Greater London area of has also been one of the inspirations of this research.

3. Methodology

3.1. Ethical Considerations

All the datasets involved in the research can be found in the public databases. The details of the data sources would be listed in the sections below. As a result, there is no ethical problems raised in the research of this project.

3.2. Study Area

The study area of this research is the Greater London, generally known as London, one of the administrative areas in England. The Greater London consists of 32 London boroughs and City of London with 9,002,488 of the population according to the Office for National Statistics (ONS). (Estimates of the population for the UK, England and Wales, Scotland, and Northern Ireland - Office for National Statistics, 2022)

Due to the limitations of the data acquirements, statistical geographies selected for the research are the Middle Layer Super Output Areas (MSOAs). The splitting limits for an MSOA are between 5,000 to 15,000 people or 2,000 to 6,000 households (2011 Census geographies - Office for National Statistics, 2022). The average population of an MSOA in London in 2010 was 8,346 as for the reference (Greater London Authority, 2014). The map of the study area is presented in Fig. 3.1.

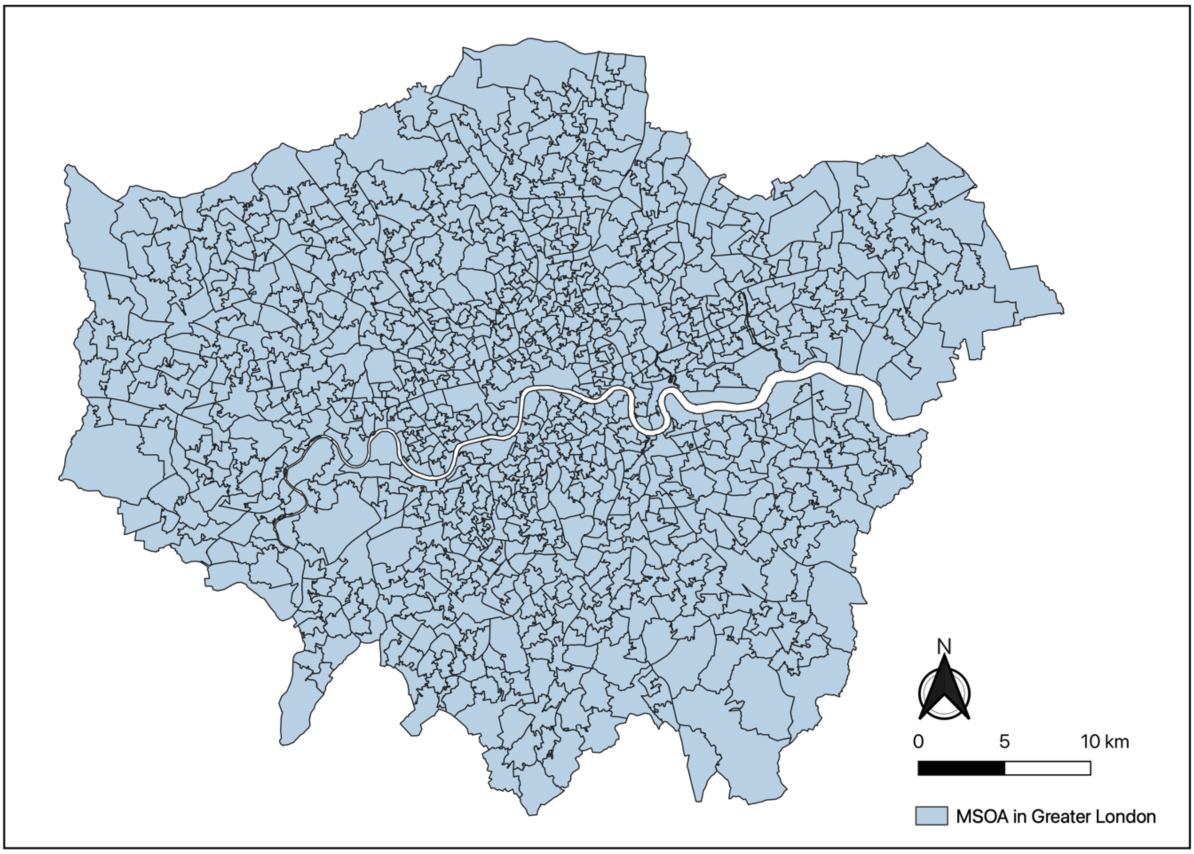

Fig. 3.1. The map of MSOA in the Greater London area.

3.3. Data

Since the hedonic pricing model is on the basis of the regression model specification, the dependent variable would be the average housing price. As for the independent variables in the model, there are three fields for the attributes that make up the hedonic pricing model as mentioned above in Section 2. Across the three fields, locations, structures and neighbourhoods, there are eight attributes that have been selected as listed in Table 3.1. The descriptive statistics of the variables involving in the hedonic pricing model are listed in Table 3.2. The details of each variable would be introduced in the following paragraphs.

Table 3.1. Details of Housing Attributes.

Fields	Variables	Definition
Locational	Closeness_w	Closeness Centrality index calculated based on the distance-weighted street network
	Betweenness_w	Betweenness Centrality index calculated based on the distance-weighted street network
	Distance_to_City	Distance from other MSOAs' centroids to the centroid of City of London
Neighbourhood	NOx_Index	The index of Nitrogen Oxide (NOx) emissions
	Num_Schools_MSOA	Number of Schools in each MSOA
	Crime_Rate	Crime rate
	Green_Blue_Rate	The rate of parks' and water's area in each MSOA
Structural	Housing_Age	The age of the housing since built (Year, the age of housing built in 2021 is 1)

Table 3.2. Descriptive Statistics.

Variables	N	Mean	Standard Deviation	Min	Max
Average Housing Price	946	595,209.360	381,420.187	225,405	4,348,686.500
Closeness_w	946	4.812	0.869	2.699	6.447
Betweenness_w	946	37,235,174.382	35,028,637.920	302,024.375	295,932,693.100
Distance_to_City	946	12,486.223	5,905.813	0	28,214.673
NOx_Index	946	99.920	22.619	57	203
Num_Schools_MSOA	946	3.432	2.238	0	22.000
Crime_Rate	946	85.824	98.489	19	1,934
Green_Blue_Rate	946	43.519	15.124	5.742	89.330
Housing_Age	946	79.960	16.418	22.978	118.886

3.3.1. Average Housing Price

As mentioned above, all the dataset was statistically split into MSOAs. The dataset of the average housing price by MSOA was collected from Land Registry in London Datastore. According to the London Datastore, this dataset contained annual mean and median property prices was published by the HM Land Registry and calculated by the Greater London Authority in 2017 (Land Registry, 2017). Python was used during the data processing. Since the distribution of the housing price is almost symmetrical and there are no obvious outliers, the mean property prices were selected for the research instead of the median price for the accuracy of the research. There are 983 MSOAs that have the available housing prices. The dataset was spatial-joined with the MSOA shapefile obtained from data.gov.uk in R programming language and plotted as a distribution map in QGIS as presented in Fig. 3.2.

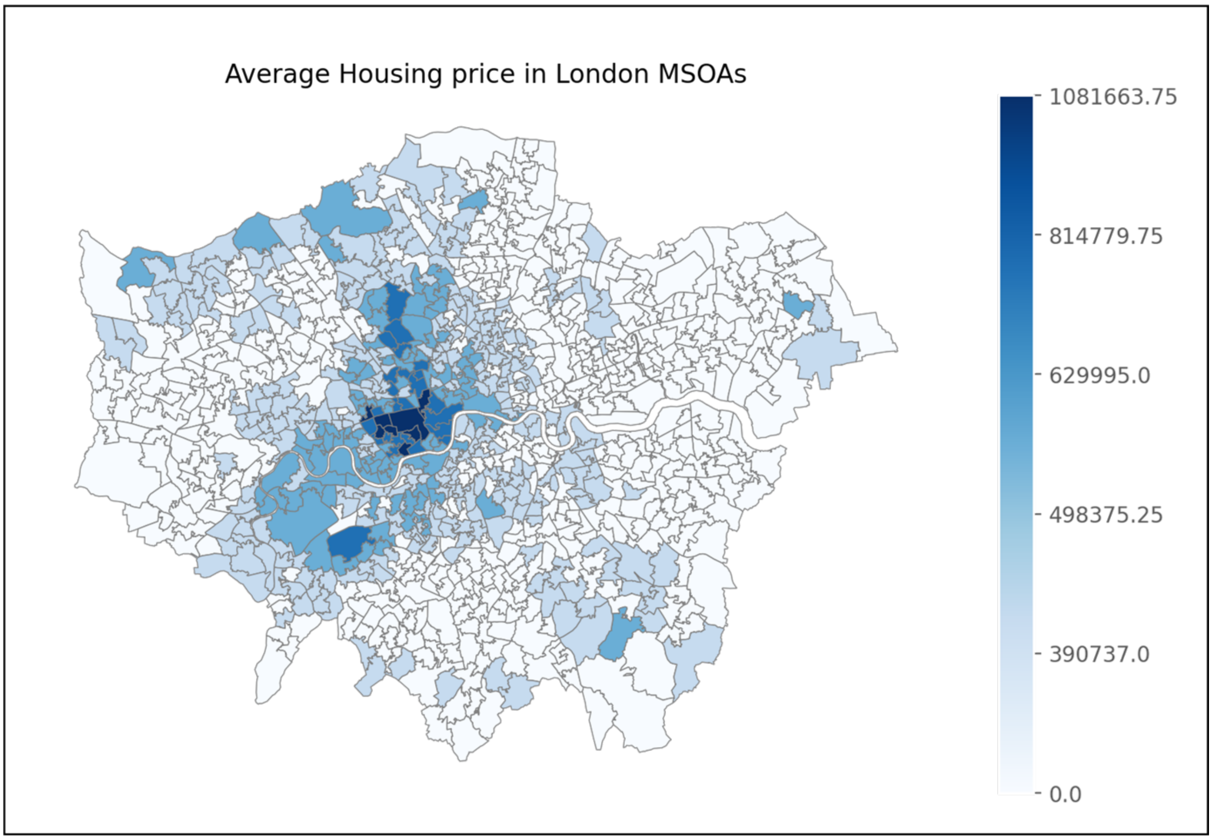

Fig. 3.2 The distribution map of the average MSOA housing price in the Greater London area.

3.3.2. Locational Variables

3.3.2.1. Centrality Indices

The street network of Greater London was obtained from the OpenStreetMap. With the help of the Python's package 'OSMnx', the street network can be constructed to be multidigraphs (containing 162405 edges and 127307 nodes) and exported as a GraphML file (Boeing, 2017). In the street network, the node is the junction of the two streets, and the edge is the street between the two nearest junctions. The type of the transportation type is selected to be 'drive'. Both one-way and two-way streets are included by filtering the edges' details. However, what needs to be mentioned is that the driving street network obtained excludes the entire City of London due to the highly complicated roads inside.

The type of network that constructed in the research is the weighted network. The length of the edges that derived from the network make up the weights in the weighted network. The street network within the study area was plotted by the coordinates of the nodes and presented in Fig. 3.3.

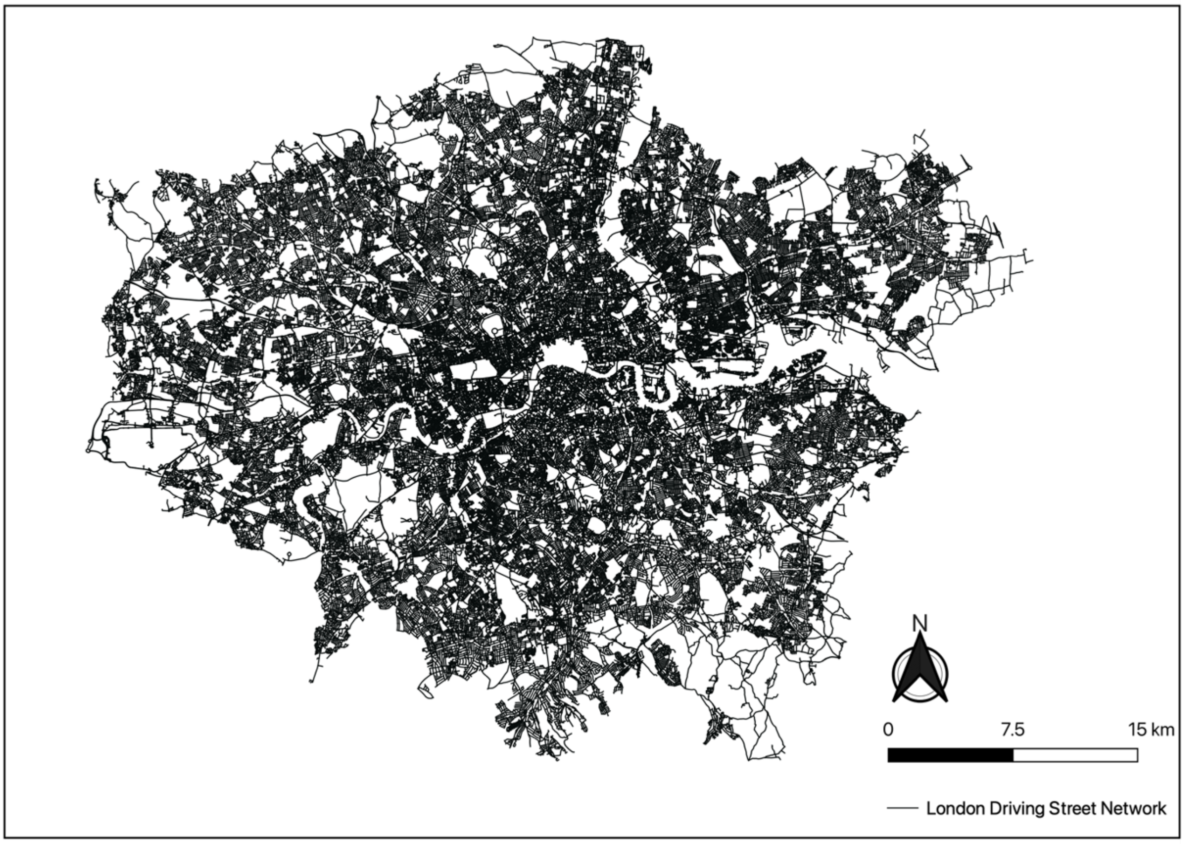

Fig. 3.3 The map of the street network of the Greater London area in types of driving.

The centrality indices were calculated based on the 127307 nodes in the weighted street networks. Closeness centrality and betweenness centrality would be taken into the research with considering of the length of the edges as the weights. The definitions of the closeness and betweenness centrality have been specified in the Literature Review. To sum up, closeness centrality identifies the shortest distance, basically how close of a node to every other node in the network, whilst betweenness centrality indicates the number of times a node is located on the shortest path between other nodes (Golbeck, 2013). For the consistency of the research, the centrality indices would be firstly classified with the MSOA boundaries using the coordinates of the nodes by means of 'Count Points In Polygon' in QGIS and then calculated for the average centrality of each MSOA. Considering the results of the closeness centrality are all under the scale of 10^(-5), hence all the results have been multiplied by 10^5 to avoid any rounding up of the decimal points in the programming language. The distribution of average closeness centrality index and betweenness centrality index across the MSOAs within the weighted network that plotted by means of QGIS and RStudio has been depicted in Fig. 3.4 and Fig. 3.5 separately.

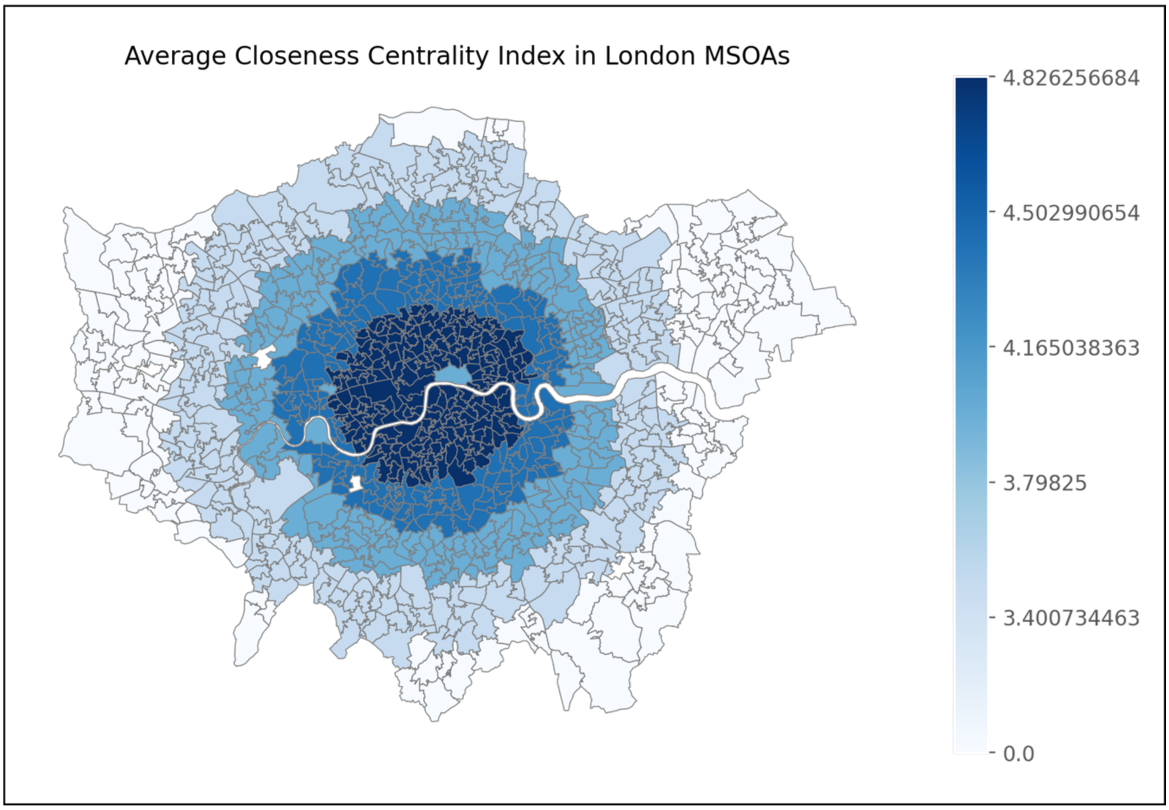

Fig. 3.4 The distribution map of the average MSOA closeness centrality index in the Greater London area.

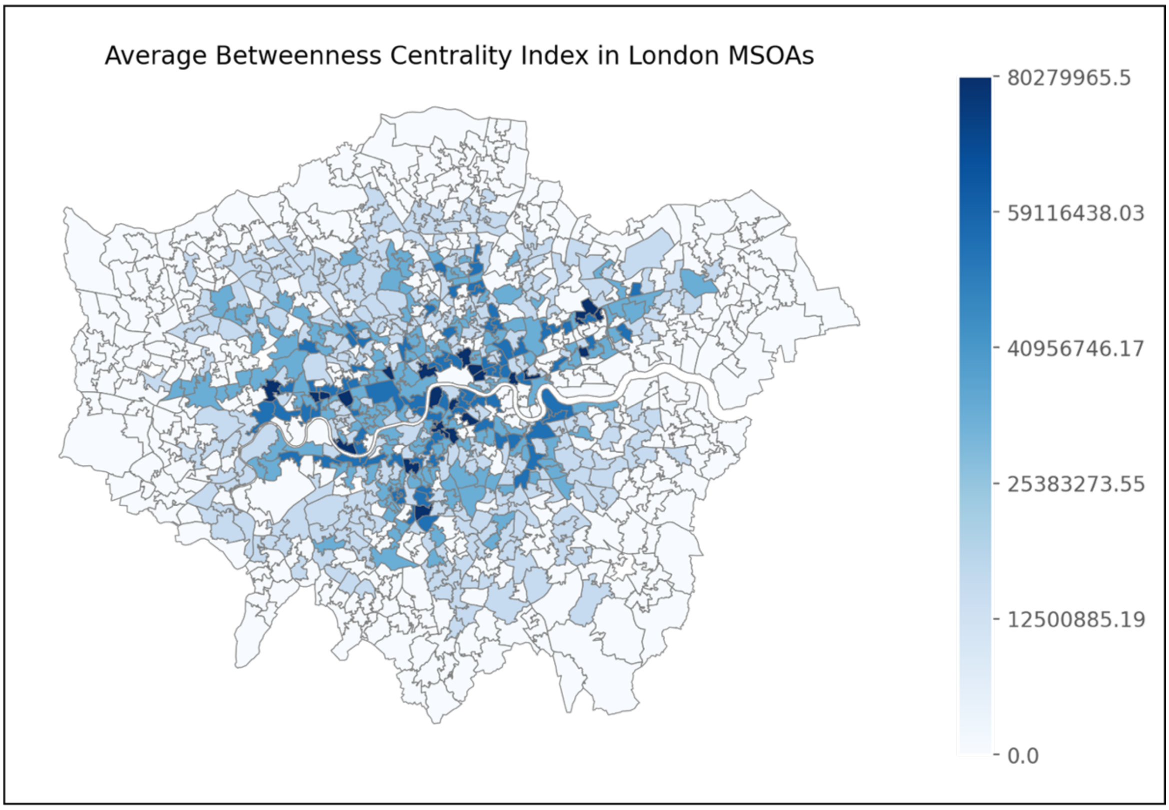

Fig. 3.5 The distribution map of the average MSOA betweenness centrality index in the Greater London area

As presented in Fig. 3.5, the average closeness centrality index in the distance-weighted network shows a distribution of the concentric circles. The less distance to the centre of the city, the higher the closeness centrality index would be. When it comes to the betweenness centrality index, as shown in Fig. 3.5, the distribution across the MSOAs tends to be relatively disperse in the city but still in the pattern that decreases due to the further distance to the centre of the city. Normally, the relatively higher closeness centrality index would imply the travel distance between one given node to all other nodes is relatively shorter (Chakrabarti et al., 2022). Thus, it is relatively more convenient for those nodes with higher closeness centrality index to travel to all other nodes. Considering the definition of the distance-weighted network, in can be inferred that the areas with higher average closeness centrality index could locate in a more centred location which also could be a more popular residential neighbourhood. As a result, it can also lead to a potentially higher price in the housing property market.

Since the definition of the betweenness centrality is of the frequency, the higher betweenness centrality index would be inferred to the higher probability of the occurrence that the distance between two nodes happen to be the shortest path (Chakrabarti et al., 2022). When it comes to the context of the research, the higher betweenness centrality index could lead to the higher traffic flow around that node (junction) between two streets. Since the weight of the network was chosen to be the street length between the nodes (junctions), the betweenness centrality would simply indicate the importance of the nodes without considering the traffic capacity (Chakrabarti et al., 2022). However, it is imprecise to simply deduce the relationship between the importance of the nodes inferred from the betweenness centrality index and the average housing price around that area. The further study of the relationship would be elaborated in the next section.

3.3.2.2. Distance_to_City

The distance to the Central Business District (CDB) has always been treated as one of the most classic locational factors that used in the hedonic pricing model (Haider and Miller, 2000). Multiple studies have revealed that the distance to the CBD has a highly correlation relationship with the housing price in San Francisco Bay Area (Maria Kockelman, 1997), Cleveland, Ohio (Simons et al., 1998), and Greater Toronto Area (Haider and Miller, 2000). However, there are also some scholars claiming that the relationship between the distance to CBD and the housing price would be influenced by the distribution of the employment in the city and the size of the city since the large modern cities tends to be not monocentric but polycentric (Hui et al., 2007, Levine, 1992). For the completeness of the research, the distance to CBD would be selected. Since London is the largest city in the UK, there are the historical area, City of London, also known as the Square Mile, and the modern area, Canary Wharf (Authority, 2008). Considering from the global recognition and geographical level, City of London, one of the MSOA with the id 'E02000001' located in central London, is selected to be the CBD in the research. To calculate the distance to City of London, the centroids of the MSOAs were calculated with tools of QGIS on the basis London MSOA boundary map. The distance to the City of London was derived from the distance of all other MSOA centroids to the centroid of City of London by means of Distance Matrix in QGIS.

3.3.3. Neighbourhood Variables

Goodman (1989) remarked that even though it is difficult to measure the value of the neighbourhood variables to the housing market, by means of the hedonic pricing model, it is still able to differentiate the housing price with the varying attributes (Goodman, 1989). According to the classification that Chin and Chau (2003) summarized, the neighbourhood variables could be divided into social factors such as crime rate, and environmental factors such as air quality.

3.3.3.1. Crime_Rate

Lots of research conducted by various scholars were about the impact of the crime on the housing price (Gibbons, 2004, Naroff et al., 1980, Tita et al., 2006). Unlike other factors, the effect of the crime to housing price is cumulative which is the reason why it is complicated to have a complete examination on the impact. The increase of the crime would potentially lead to a movement to another neighbourhood with higher safety level (Morenoff et al., 2001) and change the spatial distribution of the employment in the city which could also have the effect on the housing price (Tita et al., 2006).

As the result, crime data would always be one of the key factors in the hedonic pricing model. The research of Tita et al. (2006) involved the statistics of the burglaries and assaults in the suburb areas of Perth, Australia and the result of the OLS regression clearly present a negative relationship with the housing price (Pandit et al., 2013).

The dataset of crime in this research was obtained from the MSOA Atlas in London Datastore uploaded by Greater London Authority. The average crime rate of MSOA is selected to be one of the structural attributes involved in the hedonic pricing model of this research.

3.3.3.2. NOx_Index

Across the research of over 30 years, it has been found that there is a weak correlation index between the air quality and housing price (Chay and Greenstone, 2005). Chay and Greenstone (2005) pointed out that the result could depend on the model the scholars worked with to modify the relationship and the preference of the local individuals to the air quality. For example, in the research of Wang and Lee (2022), it is shown that the Chinese residents tend to attach importance to the air quality and this behaviour brings about the negative influence on the housing markets.

London, as a large city with more than 2.6million cars registered, has always troubled by the air quality problem. As a result, this research selects the Nitrogen Oxides (NOx) index, of which the main emission source is the pollution automobiles, trucks, and various non-road vehicles, to be the factor of the air pollution. The dataset was collected from the MSOA Atlas in London Datastore uploaded by Greater London Authority, which is the same as the crime index.

3.3.3.3. Green_Blue_Rate

A lot of research has shown the evidence that the green and blue places (parks and waters) in the city not only improve the urban environment but enrich the happiness of the residences and benefit the health (Noor et al., 2015, Pandit et al., 2013, Biao et al., 2012). The relationship between the green and blue rate and the housing price has been studied by a few scholars. Anderson and Cordell (1988) found that the local trees could contribute more than 100,000 dollars which would has large influence on the housing markets in Athens, Georgia (U.S.A.). The similar conclusion was also drawn by Kong et al. (2007) in Jinan City, China and Donovan and Butry (2010) in Portland, Oregon.

London is one of the greenest cities around the globe of which roughly 47% is ‘green’ including open-space natural habitats and private, domestic garden land, and 2.5% is blue space including rivers, lakes, and reservoirs (London National Park City, 2022). The dataset of green and blue covering rate of each ward in London was collected from the London Datastore uploaded by

Greater London Authority. With the help of gov.uk, it is able to convert the Ward (2015) to Middle Layer Super Output Area (2011) so that the geographical boundary of the research would be consistent.

3.3.3.4. Num_Schools_MSOA

Since 1956, the urban economists have revealed that the public schools would have strong effect on the housing price (Tiebout, 1956). Most of studies focused on the quality of schools (Li and Brown, 1980, Ge and Du, 2007, Uju and Iyanda, 2012). However, due to the limitation of the dataset, the educational factor in the research would be the number of schools in an area. By means of 'Count Points In Polygon' in QGIS, it would be able to convert the dataset 'location of the schools in Greater London area' to the counts of schools per MSOA.

3.3.4. Structural Variables

3.3.4.1. Housing_Age

The housing price is closely related to the structural variables (Chau and Chin, 2003) and the building age of the house was selected because it is suitable for the building age to calculate for the average value within an area. It is widely acknowledged by the scholars that the building age should have a negative influence on the housing price because of the more maintenance & repair fee, and out-of-date electrical & plumbing systems (Clapp and Giaccotto, 1998). For example, Chang et al. (1999) conducted the research of the hedonic pricing model for the housing market in Taipei, China. It is shown in the OLS regression model that the houses' ages would indicate the falling of the housing price. However, Li and Brown (1980) made the conclusion in an opposite way. They found that the historical value of the house would have a positive effect on the housing price.

In the research, the dataset of the average building age of the houses at MSOA resolution was obtained from Consumer Data Research Centre (CDRC) who integrated from HM Land Registry.

3.4. Model Specification and Estimation

3.4.1. OLS Regression Model

Since the objective of the hedonic pricing model is to measure the linear relationship between the housing price and the centrality of the street network. With the other 6 attributes, the specification of the OLS model is presented in the following form:

$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon_i, i \in \{1, 2, 3, \dots, n\}$. (4)

In the equation, i stands for each MSOA and in this case, n would be 946. Meantime, y is the independent variable (housing price), x is the principle independent variable (centrality indices), z is the other independent variables (attributes to the housing price), β stands for the coefficients and γ would be the random error in the function. In order to avoid the overfitting problem of the model, the multicollinearity among the independent variables would be checked by means of Variance Inflation Factor (VIF). Since the key characteristics of the regression model are randomness and unpredictability, the plot of residuals versus fits should be randomly distributed without showing any pattern and the distribution of the residuals should be normal distribution.

3.4.2. Spatial Regression Model

However, the OLS regression model still excludes the consideration of the spatial interactions. Therefore, the residuals of the regression model need to be mapped to check whether there is obvious spatial pattern. As it depicted in Fig. 3.2, it is clearly to see that the blue areas and orange areas are gathering next to the similar colours which indicates that there is spatial autocorrelation in the regression model. The result of Moran's I index based on the Spatial Weighted Matrix W shown in Table 3.3. also confirms the existence of the spatial autocorrelation. The Spatial Weighted Matrix W is generated by means of the k-nearest neighbours algorithm.

Table 3.3. The results of the Moran's I test on the OLS regression's residuals.

Moran's I	p-value
0.562	5.67e-146

Toward this problem, most of the scholars introduced the Spatial Lag regression model (Kawamura and Mahajan, 2005) and Spatial Error regression model (Chakrabarti et al., 2022, Bhattacharjee et al., 2012). Bhattacharjee et al. (2012) pointed out that when certain hedonic features were applied, inferences were generally similar so that both the spatial error model and the spatial lag model would be suitable for the data.

Hence, the Lagrange Multiple Tests would be utilized to select the spatial regression model. As shown in Table 3.4, the results of the Lagrange Multiple Tests includes the Spatial Error model test (LMerr), the Spatial Lag model test (LMlag), variants of one model's robust to the presence of the other model (RLMerr, RLMlag), and a portmanteau test (SARMA). Since both of the LMerr and LMlag are insignificant ($p < 2.2e-16$), it is necessary to compare the p-values of the robust forms. In this case, it suggests that Spatial Lag regression model would be the choice ($p < 2.2e-16$ compared to p-value = 0.052).

Table 3.4. The results of the Lagrange Multiple Tests.

Tests	Results	p-value
LMerr	654.21	< 2.2e-16
LMlag	773.74	< 2.2e-16
RLMerr	3.7805	0.05185
RLMlag	123.31	< 2.2e-16
SARMA	777.52	< 2.2e-16

In this context, the equation of the Spatial Lag regression model would be specified as (Chakrabarti et al., 2022):

$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + u_i$. (5)

Where

$u_i = \rho \sum_{j=1}^{n} W_{ij} \cdot u_j + \varepsilon_i,\ i, j \in \{1,2,3, \dots, n\}$ (6)

In the equation, ρ is the autoregressive parameter and W_{ij} is the Spatial Weighted Model generated from above containing the attributes' information of all the neighbouring MSOAs j ($j \in \{1,2,3, \dots, n\}$) of MSOA i that could influence the residual u_i. All the process during the

Spatial Lag model was finished by means of R programming language. With the help of '*spatialreg*' package, it is able to estimate the Spatial Lag regression model on the basis of the OLS regression model.

4. Results

In this section, the results of the analysis would be presented. At first, the association between the average housing price of MSOAs and the attributes including the average centrality index would require a preliminary estimation. Therefore, as shown in Fig. 4.1, the correlation matrix was utilized for all the dataset involving in the research. Note that both of the average closeness centrality index and average betweenness centrality index have been calculated based on the street network taking distance as weights.

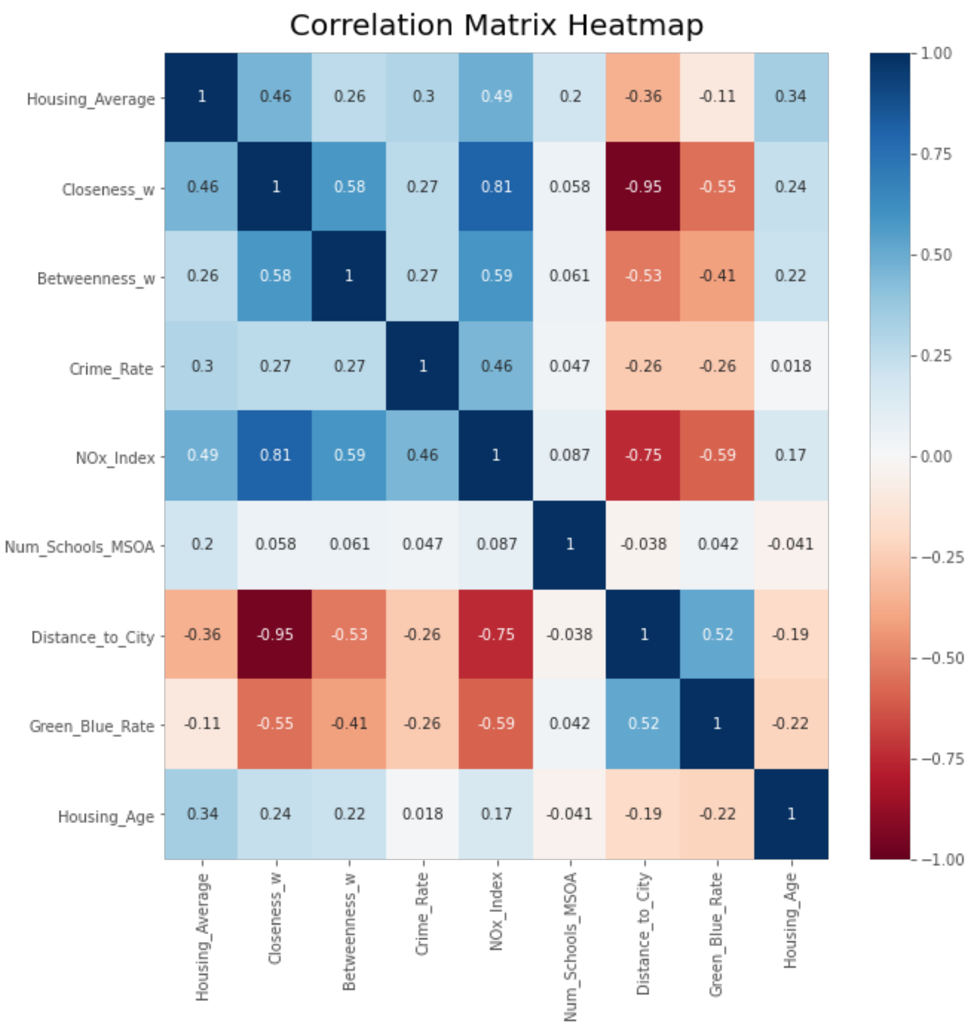

Fig. 4.1. The heatmap of the correlation matrix based on all the variables involved in the research.

After the initial estimation, the OLS regression model was introduced to have a detailed measurement of the relationship between the average closeness centrality index and average betweenness centrality index and average housing price of each MSOA. All other attributes to housing price were also included in the OLS regression model. The Variance inflation factor (VIF) result utilized to check the multicollinearity among the variables is presented in Table 4.1. It is clearly shown in Table 4.1. that both *Closeness_w* and *Distance_to_City* have relatively higher scores. Since the closeness centrality is one of the main research targets and as mentioned above, the implication of the closeness centrality in the distance-weighted network is similar with the distance to city centre, *Distance_to_Cit*y would be removed from the variables in the OLS regression model.

Table 4.1. VIF Results.

Variables	VIF
Closeness_w	17.659
Betweenness_w	1.661
Crime_Rate	1.375
NOx_Index	4.351
Num_Schools_MSOA	1.063
Distance_to_City	12.608
Green_Blue_Rate	1.818
Housing_Age	1.266

Note: Variance Inflation Factor (VIF) approach is used. The details of variables could be checked in Section 3.

Table 4.2. presents the OLS regression results of two different model specification to the hedonic pricing model as shown in Equation (4). Both average closeness centrality index and betweenness centrality index are included in the estimation. Note that all the variables involving in the model are statistically significant to the average housing price. The threshold for p-value to be statistically significant was set to be 0.05 throughout the research. In the Model A, the average housing price and other variables are taken to be fit in the linear specification. After the logarithm on average housing price in Model B, the equation turns into the log-linear specification (Mok et al., 1995). In Model B, the R-squared value has an increase

comparing with in Model A. Among the variables in Model A and Model B, *Betweenness_w* is the only factor that owns the negative coefficient in the regression function, whereas all others have a positive influence on the regression function. What needs to be mentioned is that since the equations of the models are different, it is not accurate to directly put the coefficients of the variables in two models into comparison. However, it is comparable for each variable's significance in their own model. The numbers of observations are 946 for both Model A and B.

Table 4.2. OLS Regression Models before and after Logarithm on Average Housing Price.

Variables	Model A (Before Logarithm)	Model B (After Logarithm)
Closeness_w	1.096e+05***	0.2112***
Betweenness_w	-0.0012***	-1.118e-09**
Crime_Rate	569.2699***	0.0003*
NOx_Index	7085.2029***	0.0073***
Num_Schools_MSOA	2.606e+04***	0.0260***
Green_Blue_Rate	8382.2224***	0.0112***
Housing_Age	7331.7935***	0.0091***
Constant	-1.686e+06***	10.1435***
R-Squared	0.430	0.507
Adj. R-Squared	0.426	0.504
N	946	946

Note: OLS regression model is based on Equation (4). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. See Section 3 for the details of the variables.

Next, as mentioned in Section 3, the Spatial Lag regression model (Model C) would be used to address the spatial autocorrelation that could influence the estimation of the relationship among average housing prices and attributes in the OLS regression model. Table 4.3 shows the summary of the results conducted from the Spatial Lag regression model.

Table 4.3. The summary of Spatial Lag regression model (Model C).

Variables	Coefficients
Closeness_w	7.007733e-02***
Betweenness_w	-3.075073e-08*
Crime_Rate	2.092552e-04**
NOx_Index	1.332579e-03*
Num_Schools_MSOA	1.210655e-02***
Green_Blue_Rate	4.946422e-03***
Housing_Age	4.750531e-03***
Constant	1.946101e+00***
rho	0.768

Note: Spatial Lag regression model is based on Equation (5). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. See Section 3 for the details of the variables.

In Table 4.3, the coefficients of the variables present the specification to the function of the Spatial Lag regression model as shown in Equation (5) and (6). All the variables are statistically significant since their p-values are below the threshold 0.05. The coefficients of the Model C are not comparable to the Model A and B directly, but it could be used to measure the importance of each variable in Model C. With the combination of all three models, Model A, B and C, the importance of each variable could be estimated thoroughly, which further leads to a detailed understanding of the relationship between the average housing price, centrality indices and other attributes. The rho value in Model C stands for the Spatial Lag parameter which measures the spatial dependence of the average housing price on the centrality indices and other attributes in the surrounding spatial areas as the definition of the Spatial Weighted Matrix in Section 3.

5. Discussion

Throughout the analysis, the first research question that whether the centrality indices of the street network could determine the housing price was studied. There are two kinds of the centrality indices involved in the research, closeness centrality and betweenness centrality. With the estimation of the hedonic pricing model, the results of the analysis have indicated that there is a relationship between the centrality indices and the housing price. However, it is found that the different selection of the centrality indices would have different results. The closeness centrality index and betweenness centrality index would have contrasting scales of the influence on the housing price. The details of the results will be elaborated in the following paragraphs.

As depicted in Fig. 4.2, the result of the correlation matrix based on all the variables suggests that the centrality indices of the street network in the Greater London area is positively contributed to the average housing price. The result of the OLS regression models and the Spatial Lag regression model presented in Table 4.2. and Table 4.3. indicates a positive relationship between the centrality indices and the average housing price as well. Among the three regression models (OLS regression model before logarithm, OLS regression model after logarithm and Spatial Lag regression Model), the betweenness centrality index tends to have the relatively smallest absolute value of the coefficients in the functions, which suggests that the average betweenness centrality index is relatively insignificant to the average housing price in Greater London area. Note that the statistical significances in the models confirm that the betweenness index fits the models well. The betweenness centrality index in the models is calculated as an average value of the MSOA. As defined above in Section 3, the average betweenness centrality index involved in this research indicates the scale of the traffic importance in that MSOA. Therefore, it can be deduced that in the Greater London area, the importance of one junction between two streets has relatively little impact to the housing price around that area comparing to other attributes. The reason could be that the junctions with the higher importance level could appeal to the commercial developers who would take the shopping centres as the land use (Chakrabarti et al., 2022). As for the negative sign of the coefficients of *Betweenness_w* shown in Model A, B and C, this could be explained with the noise and the air pollution caused by traffic. This can be justified by the highly positive correlation index between *Betweenness_w* and *NOx_Index* shown in Fig. 4.1. Many scholars have also concluded the similar explanation (Xiao et al., 2016b).

The average closeness centrality index shows a statistically significant influence on the average housing price. As shown in Table 4.2, the coefficient value of *Closeness_w* is the highest with a positive sign among the attributes in both Model A and Model B. Recall that the closeness centrality index in the distance-weighted street network measures how close in the distance it would be when travelling from one node (junction) to the other node (junction), which means it would be more convenient for one to travel in the area with higher closeness centrality index. It can be said that the high coefficient value indicates that the convenience of travelling in that MSOA would have a huge impact on the housing price around. The importance of the centrality in the housing price is well demonstrate in this finding. This also suggests that those neighbourhoods with a better-connected street network in itself and with the others tend to be more popular and have a higher housing value in the Greater London area. What need to be mentioned is that since the closeness centrality in distance-weighted network is distributed in the Greater London area as the concentric circles as shown in Fig. 3.4, the potential influence of distance to the city centre on the housing price could also lead to such high coefficient value of *Closeness_w*.

Other possible reason for the relationship between the centrality indices and the housing price derived from this analysis could be explained from other scholars. There is very few research on the Greater London's street network centrality. However, it can still be explained from other cities' study case. Xiao et al. concluded in their study of Cardiff that the relationship between centrality and housing price fits the theoretical expectation because the betweenness centrality stands for the likelihood of traffic congestion and the closeness centrality stands for the convenient access to opportunities (Xiao et al., 2016b). The impact is extremely evident in Cardiff, which is a relatively single CBD city clearly with an identifiable city centre. When it comes to the Greater London area, where there are multiple CBDs inside the city centre and the centre of commercial activities are relatively vague, the housing price can be less influenced by some accessibility measurements such as the betweenness centrality.

The second research question that whether there is any further relationship between the centrality indices of the street network and the housing price was studied in the analysis as well. The result presented in Table 4.3. has clearly shows the fitting of the Spatial Lag regression model is well. The *rho* value indicates a strong spatial dependence in the model. The reason could be since the streets are connecting multiple neighbourhoods, the housing price would not be defined by the single centrality index in one MSOA in multiple MSOA around the area with the consideration of Spatial Lag effect.

The high coefficient value of *Closeness_w* in Table 4.3. indicates that there is a solid spatial relationship between the average closeness centrality index. When it comes to the betweenness centrality index, just as in Model A and B, *Betweenness_w* still shows a relatively weak influence on the average housing price determined from the coefficient value. The p-values have proved that the estimation has the robustness and reliability. It can be concluded that the Spatial Lag regression model is suitable for the Greater London area to identify the relationship between the centrality indices of the street network and the housing price. As for the reason that leads to the different reaction of the closeness centrality index and the betweenness centrality index to the housing price, it can be deducted from the distribution maps shown in Fig. 3.4. and Fig. 3.5. Fig. 3.4. presents a distribution similar to the concentric circles which proves that the average closeness centrality index is highly spatial autocorrelated since it directly reveals the distance to the city centre. And what Fig. 3.5 presents is a relatively random distribution which indicate that the average betweenness centrality is less spatial dependent. Therefore, when the centrality indices were applied in the Spatial Lag regression model, the closeness centrality index would have much more spatial influence on the housing price than the betweenness centrality index.

This analysis presents three different hedonic pricing models to estimate the relationship between the average centrality indices and the average housing price. The first two, Model A and Model B, conducted with the OLS regression in the linear specification and the log-linear specification individually. The increasement of the R-squared value (0.430 to 0.507) shown from Model A to Model B in Table 4.2. indicates that the logarithm on the average housing price improves the fitting of the attributes to the regression model. However, the disadvantage of the OLS regression model is also obvious since the strength of the relationship between the Model A and Model B and the dependent variable is only around 40%-50%. The possible reasons would be explained in the following paragraph.

Firstly, since the focus of the research is on the basis of the MSOA level rather than each single neighbourhood, the housing prices and the centrality indices were calculated in average of each MSOA. During this process, the characteristics of the single neighbourhood would be ignored, and the variables would be analysed from a more global perspective. Therefore, if the access to a more detailed dataset becomes available, it could lead to a better estimated model. Then, as mentioned above, the spatial autocorrelation is occurring in the OLS regression model. The spatial dependence of the housing price in the Greater London area needs to be considered. As a result, the Spatial Lag regression model was conducted to solve such problem.

What needs to be discussed as well is the other 5 variables (*Crime_Rate, NOx_Index, Num_Schools_MSOA, Green_Blue_Rate, Housing_Age*). Among all three models, the high coefficient value of *Num_Schools_MSOA* has indicated that the number of the schools around the area plays an important and positive role on the housing market in the Greater London area. The medium level of the *Green_Blue_Rate*'s and *Housing_Age*'s coefficients shows that the housing price in Greater London area is not highly related to the green and blue covering rate and the age of the housings. However, it is surprising to see that *Crime_Rate* and *NOx_Index* both have a positive coefficient in the models. Reviewed from the literature, the crime rate and the pollution rate should have a negative effect on the housing price (Pandit et al., 2013, Wang and Lee, 2022). An assumption of this case is that in the Greater London area, those with the higher crime rate and the pollution rate tends to be a highly populated area with more attraction. This could lead to a relatively higher price in the housing market.

6. Conclusion

In summary, the research analysed the relationship between the average centrality indices in the distance-weighted driving street network and the average housing price from the MSOA perspective in the Greater London area based on the two research questions. From the traditional hedonic pricing model to the spatial hedonic model, the OLS regression model and Spatial Lag regression model were utilized to estimate the average housing price with the selected locational, neighbourhood and structural variables. The average closeness centrality index and average betweenness centrality index were taken as part of the locational variables.

The results of the analysis have presented that there is a strong impact of the average closeness centrality index on the average housing price in both of the OLS regression models (OLS regression model before logarithm and OLS regression model after. Whereas the average betweenness centrality index has relatively weak and even negative influence on the housing price. The research has also been extended to the spatial perspective. The spatial autocorrelation has been detected and the Spatial Lag regression model was utilized to estimate the spatial dependence of the housing price. The result also indicates that the average closeness centrality index has a solid spatial influence on the average housing price while the average betweenness centrality index has an opposite reaction. The potential reasons that lead to these cases have been elaborated.

This study is expected to make contributions to the literature. Firstly, this research covers the current body of knowledge about this topic that studies the housing price and the street network while also provides new insights including new study area, the Greater London area. The findings of the research enhance the existing literature in the fields of the urban street network planning and the housing market and makes new contributions in the fields such as the application of the centrality. This study includes the closeness centrality index and the betweenness centrality index, both of which are calculated in a distance-weighted driving street network. When it comes to the distance-weighted street network itself, the advantage of this application is obvious and useful. The distance-weighted street network neglects the traffic flow which would differ depending on the peak hours and off-peak hours. It is able to put the emphasis on the structure of the network itself rather than considering other weights. Secondly, this research on the spatial model also provides new perspective to the study of the influence of the street network centrality indices on the housing market. By means of Moran's I test on the residuals and the Lagrange Multiple Tests, the Spatial Lag regression model is selected on a scientific basis. Thirdly, with the positive relationship between the closeness centrality index and the housing price, it helps identify the area with a higher housing price from the locational perspective, which could be very useful for the urban planners have a better explanation facing the case such as the prediction of the housing price. Lastly, this research also includes multiple factors in the locational, neighbourhood and structural perspective which enhances the understanding of the characteristics of the Greater London area. By means of the hedonic pricing model, there variables help to have a comprehensive study on the local housing market as well.

There are also some limitations in this study. One limitation is that the analysis is constrained by the availability of the datasets. Since it is not able to obtain the detailed neighbourhood-level information, the estimation of the models has to be based on the calculation of the average value of the centrality index in the MSOA level rather than an area with smaller radius. This partly leads to a less ideal fitted regression model as mentioned above in Section 5. The other limitation that results from the dataset is that as mentioned above in Section 3, the dataset of the street network excludes the entire City of London area, which would lead to a critical problem when calculating the centrality indices. Both closeness centrality and betweenness centrality would be influenced by the removal of the nodes in City of London. This could bring about the imprecise estimation of the models. The next limitation is that this research focused on the driving street network. To conduct the complete estimation of the street network and the housing price, it would be more accurate with the transportation perspective such as the bus route network. What's more, if the bus route network takes the passenger flows as the weight and is combined with the traffic flow-weighted street network, the housing price would be better estimated due to the complement of the urban transportation accessibility.

From the model perspective, the analysis lacks the comparison between the closeness centrality and betweenness centrality individually affecting the housing price, with which it is able to tell which one is more effective, the induvial influence or the mutual influence. From the spatial perspective, the last limitation of this study is lack of the further examination of the direct and indirect effect of the Spatial Lag regression model. With this examination, it is able to identify the detailed influence of the centrality indices in the neighbourhood area on the housing price in the local area.

As a result, the future work of this study would firstly expand to the more detailed study area based on the availability of the datasets. The urban transportation network would be taken into consideration to enrich the analysis. To estimate the complete relationship, the closeness centrality and the betweenness centrality would be considered to be put separately in the model so that a comparison can be made. A further detailed analysis in the spatial perspective would also be involved in the future work.

Bibliography

ABIDOYE, R. B. & CHAN, A. P. 2017. Critical review of hedonic pricing model application in property price appraisal: A case of Nigeria. *International Journal of Sustainable Built Environment,* 6**,** 250-259.

ANDERSON, L. M. & CORDELL, H. K. 1988. Influence of trees on residential property values in Athens, Georgia (USA): A survey based on actual sales prices. *Landscape and urban planning,* 15**,** 153-164.

AUTHORITY, G. L. 2008. London's central business district: Its global importance. *London, England: GLA Economics*.

BAVELAS, A. 1948. A mathematical model for group structures. *Human organization,* 7**,** 16-30.

BHATTACHARJEE, A., CASTRO, E. & MARQUES, J. 2012. Spatial interactions in hedonic pricing models: the urban housing market of Aveiro, Portugal. *Spatial Economic Analysis,* 7**,** 133-167.

BIAO, Z., GAODI, X., BIN, X. & CANQIANG, Z. 2012. The effects of public green spaces on residential property value in Beijing. *Journal of Resources and Ecology,* 3**,** 243-252.

BOEING, G. 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems,* 65**,** 126-139.

CHAKRABARTI, S., KUSHARI, T. & MAZUMDER, T. 2022. Does transportation network centrality determine housing price? *Journal of Transport Geography,* 103**,** 103397.

CHANG, H. J. & LEE, Y. H. 1999. Specification of the hedonic price model for Taipei housing market. *International journal of information and management sciences,* 10**,** 1-14.

CHAU, K. W. & CHIN, T. 2003. A critical review of literature on the hedonic price model. *International Journal for Housing Science and its applications,* 27**,** 145-165.

CHAY, K. Y. & GREENSTONE, M. 2005. Does air quality matter? Evidence from the housing market. *Journal of political Economy,* 113**,** 376-424.

CLAPP, J. M. & GIACCOTTO, C. 1998. Residential hedonic models: A rational expectations approach to age effects. *Journal of Urban Economics,* 44**,** 415-437.

COLWELL, P. F. & DILMORE, G. 1999. Who was first? An examination of an early hedonic study. *Land Economics***,** 620-626.

CRUCITTI, P., LATORA, V. & PORTA, S. 2006a. Centrality in networks of urban streets. *Chaos: an interdisciplinary journal of nonlinear science,* 16**,** 015113.

CRUCITTI, P., LATORA, V. & PORTA, S. 2006b. Centrality measures in spatial networks of urban streets. *Physical Review E,* 73**,** 036125.

DONOVAN, G. H. & BUTRY, D. T. 2010. Trees in the city: Valuing street trees in Portland, Oregon. *Landscape and urban planning,* 94**,** 77-83.

FREEMAN, L. C. 1977. A set of measures of centrality based on betweenness. *Sociometry***,** 35-41.

GE, X. J. & DU, Y. Main variables influencing residential property values using the entropy method–the case of Auckland. Proceedings of the 5th International Structural Engineering and Construction Conference. Shunan, Japan, 2007.

GIBBONS, S. 2004. The costs of urban property crime. *The Economic Journal,* 114**,** F441-F463.

GOLBECK, J. 2013. *Analyzing the social web*, Newnes.

GOODMAN, A. C. 1989. Topics in empirical urban housing research. *The economics of housing markets***,** 49-143.

GREATER LONDON AUTHORITY, 2014. MSOA Atlas – London Datastore. [online] DATA.LONDON.GOV.UK. Available at: <https://data.london.gov.uk/dataset/msoa-atlas> [Accessed 14 August 2022].

HAAS, G. C. 1922. A statistical analysis of farm sales in blue earth county, Minnesota, as a basis for farm land appraisal.

HAIDER, M. & MILLER, E. J. 2000. Effects of transportation infrastructure and location on residential real estate values: application of spatial autoregressive techniques. *Transportation Research Record,* 1722**,** 1-8.

HERATH, S. & MAIER, G. 2010. The hedonic price method in real estate and housing market research: a review of the literature.

HUH, S. & KWAK, S.-J. 1997. The choice of functional form and variables in the hedonic price model in Seoul. *Urban Studies,* 34**,** 989-998.

HUI, E. C., CHAU, C. K., PUN, L. & LAW, M. 2007. Measuring the neighboring and environmental effects on residential property value: Using spatial weighting matrix. *Building and environment,* 42**,** 2333-2343.

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.

KAWAMURA, K. & MAHAJAN, S. 2005. Hedonic analysis of impacts of traffic volumes on property values. *Transportation Research Record,* 1924**,** 69-75.

KIRKLEY, A., BARBOSA, H., BARTHELEMY, M. & GHOSHAL, G. 2018. From the betweenness centrality in street networks to structural invariants in random planar graphs. *Nature communications,* 9**,** 1-12.

LANCASTER, K. J. 1966. A new approach to consumer theory. *Journal of political economy,* 74**,** 132-157.

LAND REGISTRY, 2017. Average House Prices by Borough, Ward, MSOA & LSOA – London Datastore. [online] Data.london.gov.uk. Available at: <https://data.london.gov.uk/dataset/average-house-prices> [Accessed 16 August 2022].

LATORA, V. & MARCHIORI, M. 2007. A measure of centrality based on network efficiency. *New Journal of Physics,* 9**,** 188.

LEVINE, J. C. 1992. *Decentralization of jobs and emerging suburban commute*.

LI, M. M. & BROWN, H. J. 1980. Micro-neighborhood externalities and hedonic housing prices. *Land economics,* 56**,** 125-141.

LIN, G., CHEN, X. & LIANG, Y. 2018. The location of retail stores and street centrality in Guangzhou, China. *Applied geography,* 100**,** 12-20.

MARIA KOCKELMAN, K. 1997. Effects of location elements on home purchase prices and rents in San Francisco Bay Area. *Transportation Research Record,* 1606**,** 40-50.

MAURER, R. & PITZER, M. 2004. Hedonic price indices for the Paris housing market. *Allgemeines Statistisches Archiv,* 88**,** 303-326.

MOK, H. M., CHAN, P. P. & CHO, Y.-S. 1995. A hedonic price model for private properties in Hong Kong. *The Journal of Real Estate Finance and Economics,* 10**,** 37-48.

MORENOFF, J. D., SAMPSON, R. J. & RAUDENBUSH, S. W. 2001. Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology,* 39**,** 517-558.

NAROFF, J. L., HELLMAN, D. & SKINNER, D. 1980. Estimates of the impact of crime on property values. *Growth and Change,* 11**,** 24-30.

NATIONALPARKCITY.LONDON. 2022. London National Park City. [online] Available at: <https://www.nationalparkcity.london/> [Accessed 28 August 2022].

NEWMAN, M. E. J. & OXFORD UNIVERSITY, P. 2010. *Networks : an introduction,* Oxford, Oxford University Press.

NOOR, N. M., ASMAWI, M. Z. & ABDULLAH, A. 2015. Sustainable Urban Regeneration: GIS and Hedonic Pricing Method in determining the value of green space in housing area. *Procedia-Social and Behavioral Sciences,* 170**,** 669-679.

ONS.GOV.UK. 2022. Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland - Office for National Statistics. [online] Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland> [Accessed 14 August 2022].

ONS.GOV.UK. 2022. 2011 Census geographies - Office for National Statistics. [online] Available at: <https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeographies/2011censusgeographies> [Accessed 30 August 2022].

PANDIT, R., POLYAKOV, M., TAPSUWAN, S. & MORAN, T. 2013. The effect of street trees on property value in Perth, Western Australia. *Landscape and Urban Planning,* 110**,** 134-142.

PORTA, S., LATORA, V., WANG, F., RUEDA, S., STRANO, E., SCELLATO, S., CARDILLO, A., BELLI, E., CARDENAS, F. & CORMENZANA, B. 2012. Street centrality and the location of economic activities in Barcelona. *Urban studies,* 49**,** 1471-1488.

PORTA, S., STRANO, E., IACOVIELLO, V., MESSORA, R., LATORA, V., CARDILLO, A., WANG, F. & SCELLATO, S. 2009. Street centrality and densities of retail and services in Bologna, Italy. *Environment and Planning B: Planning and design,* 36**,** 450-465.

PORTA, S., STRANO, E., VENERANDI, A., ADAM, R., ROMICE, O., PASINO, P. & BIANCHI, G. 2013. Multiple centrality assessment: understanding and designing mixed use streets in professional masterplanning. *Urban Design***,** 12-14.

RANDENIYA, T., RANASINGHE, G. & AMARAWICKRAMA, S. 2017. A model to estimate the implicit values of housing attributes by applying the hedonic pricing method. *International Journal of Built Environment and Sustainability,* 4.

ROSEN, S. 1974. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy,* 82**,** 34-55.

SABIDUSSI, G. 1966. The centrality index of a graph. *Psychometrika,* 31**,** 581-603.

SCOTT, J. 2000. Social network analysis: A handbook. 2nd edn sage Governmentalations.

SIMONS, R., QUERCIA, R. & LEVIN, I. 1998. The value impact of new residential construction and neighborhood disinvestment on residential sales price. *Journal of Real Estate Research,* 15**,** 147-161.

STRANO, E., VIANA, M., DA FONTOURA COSTA, L., CARDILLO, A., PORTA, S. & LATORA, V. 2013. Urban street networks, a comparative analysis of ten European cities. *Environment and Planning B: Planning and Design,* 40**,** 1071-1086.

TAN, R., ZHOU, K. & XU, H. 2019. Effects of urban road centrality on property values: Spatial hedonic analysis of the housing market in Wuhan, China. *Journal of Urban Planning and Development,* 145**,** 05019005.

TIEBOUT, C. M. 1956. A pure theory of local expenditures. *Journal of political economy,* 64**,** 416-424.

TITA, G. E., PETRAS, T. L. & GREENBAUM, R. T. 2006. Crime and residential choice: a neighborhood level analysis of the impact of crime on housing prices. *Journal of quantitative criminology,* 22**,** 299-317.

UJU, I. V. & IYANDA, S. T. 2012. Comparative analysis of the determinants of residential land values. *Chinese Business Review,* 11.

WANG, J. & LEE, C. L. 2022. The value of air quality in housing markets: A comparative study of housing sale and rental markets in China. *Energy Policy,* 160**,** 112601.

WASSERMAN, S. & FAUST, K. 1994. Social network analysis: Methods and applications.

WEN, H.-Z., SHENG-HUA, J. & XIAO-YU, G. 2005. Hedonic price analysis of urban housing: an empirical research on Hangzhou, China. *Journal of Zhejiang University-Science A,* 6**,** 907-914.

XIAO, Y., ORFORD, S. & WEBSTER, C. J. 2016a. Urban configuration, accessibility, and property prices: A case study of Cardiff, Wales. *Environment and Planning B: Planning and Design,* 43**,** 108-129.

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.