Cities 60 (2017) 166–179

Contents lists available at ScienceDirect

Cities

journal homepage: www.elsevier.com/locate/cities

Defining Street-based Local Area and measuring its effect on house price using a hedonic price approach: The case study of Metropolitan London

Stephen Law

Space Syntax Laboratory, Bartlett School of Architecture, University College London, 2nd floor, 140 Hampstead Road, London, NW1 2BX, United Kingdom

ARTICLE INFO

Article history:
Received 30 January 2016
Received in revised form 26 July 2016
Accepted 7 August 2016
Available online 10 September 2016

Keywords:
Community detection
Multi-level regression model
Hedonic price approach
Neighbourhood planning
Space syntax
Dual graph

ABSTRACT

An under-explored topic within the field of planning and housing studies is related to the definition of local area unit. An empirical problem that arises is that different types of local area units can infer different results. This could be in constructing segregation indices, in estimating hedonic price models or in identifying housing submarkets. This research proposes the concept of Street-based Local Area (SLA), in asking to what extent SLA associate with house price. In order to examine this question, this article borrows from network science and space syntax research in defining SLA. This research conjectures that SLA has a significant effect on house price and that this effect is captured more strongly than ad-hoc administrative region-based local area. In order to test this conjecture, this research adopted the multi-level hedonic price approach to estimate local area effects on house prices for the case study of Metropolitan London in the United Kingdom. Results showed significant local area effects on house prices and that SLA is preferred to region-based one. The plausible reasons are firstly, people perceived the local area on a street network. Street-based Local Area is able to capture more precisely subtle perceptual differences in an urban environment than an ad-hoc administrative region. Second, the topology of the street network reinforces the socio-economic similarity/differences overtime. Differences between local areas can become more pronounced as like-minded people bump into each other, cluster together and share information with each other. Third, as people identify these local areas they would make decisions based on it. The local area becomes part of the housing bundle leading to it having an effect on house price. The main contribution of the research is the novel application of community detection techniques on the street-network dual graph to defining SLA. This is important as it links the topology of the street network to how we define and perceive local area and it presents an alternative to ad-hoc administrative geographies that are currently applied in many aspects of neighbourhood planning.

1. Introduction

1.1. Background

Research examining intra-city house price variations often focuses on estimating the implicit price at which buyers and sellers are willing to exchange contracts for structural features, accessibility levels and neighbourhood amenities, using the hedonic price approach (Rosen, 1974; Cheshire & Sheppard, 1998). Applying the hedonic price approach, both geographic and geometric accessibility variables were found to be significant when associating with house prices in London between 1995 and 2011 (Law, Karimi, Penn, & Chiaradia, 2013). The results confirm established relationship between property value with geometric accessibility measures (Chiaradia et al., 2013; Xiao et al., 2015) and geographic accessibility measures (Ahlfeldt, 2011; Shen and Karimi, 2016). However, location differential in house price is argued in this article to not only be captured by accessibility effects but also by local area effect as defined by the street network. This follows from previous spatial configuration research, whereby the topology of a street network relates not only to how we move in space but also how we associate with a place (Dalton, 2006; Yang & Hillier, 2007).

This research will propose the concept of Street-based Local Area (SLA) with the aim to test the extent to which SLA has an effect on house prices. The study employs a multi-level hedonic price approach in estimating the Street-based Local Area effects on house price using the house price dataset of London in 2011. The remainder of the article is organised as follows: Section One introduces previous research on local area units; Section Two introduces the framework for defining Street-based Local Area; Section Three provides details for the multi-level hedonic price empirical method; Section Four introduces the case study of London and the hedonic price model dataset; Section Five reports the estimation results, and Section Six provides a general discussion of the findings and limitations.

1.2. Previous research

An under-explored topic within the field of urban planning and housing studies is the definition of a local area unit. Local area unit here is defined as a geographical unit that is larger than the immediate

E-mail address: Stephen.law@ucl.ac.uk.

http://dx.doi.org/10.1016/j.cities.2016.08.008

home area, but smaller than the city (Kearns & Parkinson, 2001). It is related to the concept of neighbourhood in urban studies which encompass complex overlapping geographical, historical, socio-economic and perceptual constructs (Lebel, Pampalon, & Villeneuve, 2007; Galster, 2001; Kearns & Parkinson, 2001).

Census tracts or ward boundaries are administrative region-based local area units that are commonly used to capture neighbourhood characteristics. Due to convenience, these boundaries were often used in estimating hedonic price models or in defining housing sub-markets (Goodman & Thibodeau, 1998, 2003; Leishman, 2009). However, these local area units are seen as arbitrary as it cut across streets and buildings and researchers recognise these definitions do not necessarily capture the qualities of a neighbourhood (Coulton, Korbin, Chan, & Su, 2001; Ellen and Turner, 1997). Fig. 1 illustrates an area in London known as the Isle of Dogs being overlaid with the Middle Super Output Area (MSOA) UK census boundary. The mapping shows disjointed boundaries of MSOA in red that cuts across the central office areas of Canary Wharf in blue.

One problem of these 'arbitrary' or 'ad-hoc' (Orford, 1999; Goodman, 1978) administrative local area unit is that it creates inconsistent empirical results. Goodman's early studies (1978; 1985) showed traces of this investigation when he found coefficient differences in estimating a hedonic price model comparing between the block level and census tract level for the case study of New Haven. In 1985, Goodman, found segregation indices differed when applied through different levels of aggregation using the case study of Baltimore. Differences can be attributed to the fuzziness of local area geographies. These problems are extended to housing submarket identification as noted by Leishman (2009). For example, Bourassa (Bourassa, Hamelink, Hoesli, & MacGregor, 1999) compared housing submarkets defined using either individual dwellings or census tract level data in both Sydney and Melbourne. He found grouping dwelling data achieved different results than grouping census tract ones. These early research found inconsistencies when calculating segregation indices, when estimating multi-level hedonic price models and when defining housing submarket. Recent research also suggests resident perception maps of neighbourhood could be more meaningful than administrative boundaries (Coulton et al., 2001). It is for these reasons this research will propose the concept of Street-based Local Area (SLA).

2. Conceptual framework

2.1. A framework for Street-based Local Area (SLA)

Street-based Local Area (SLA) is defined as a local area that is; first street-based, second topological/ configurational, third has membership in discrete form and fourth is larger than a home area but smaller than a city. The concept of SLA borrows from two field, network science and space syntax research. It borrows from network science the concept of community structure which is a characteristic found in many social and biological networks (Girvan & Newman, 2002). It also borrows from space syntax research, the use of a spatial network dual graph in representing a city. This research in particular will ask;

Research Question: to what extent do street-based local areas, as defined by the topology of the street network, associate with house price. Secondly, how do street-based local area units compare with ad-hoc administrative region-based local area units in associating with house prices?

Fig. 2 illustrates the comparison between these two types of local area units.

This research conjectures that SLA has significant effect on house price and is preferred to ad-hoc administrative region. The plausible reasons are firstly, people perceived the local area on a street network. The street network is therefore able to capture, more precisely subtle differences in an urban environment and more accurately the perceptual definition of a local area than ad-hoc region. Second, the topology of the street network reinforces the socio-economic similarity overtime. As people identify these local areas, this would have an effect on house price. Further discussions would be presented in the last section. In order to define Street-based Local Area, this research will borrow from network science, community detection techniques and from space syntax, the dual graph representation of the city. We will first describe these methods separately and how combining these two sets of methods can construct Street-based Local Area.

2.2. Community detection method

The objective of community detection is to define a set of subgraphs that maximises internal ties and minimises external ties using strictly the topology of the graph. These techniques found strong association with social, functional and geographical network groupings (Girvan & Newman, 2002; Guimer`a, Mossa, Turtschi, & Amaral, 2005; Caschili, De Montis, Chessa, & Deplano, 2009). A key reason in the use of community detection techniques on defining SLA is the spatial homogeneity within a network cluster could be related to the social-economic or perceptual homogeneity found in neighbourhoods or local areas. Previous research did not apply such techniques on the street network to find locality. Therefore, a key contribution of the research is the application of community detection techniques on the street network dual graph.

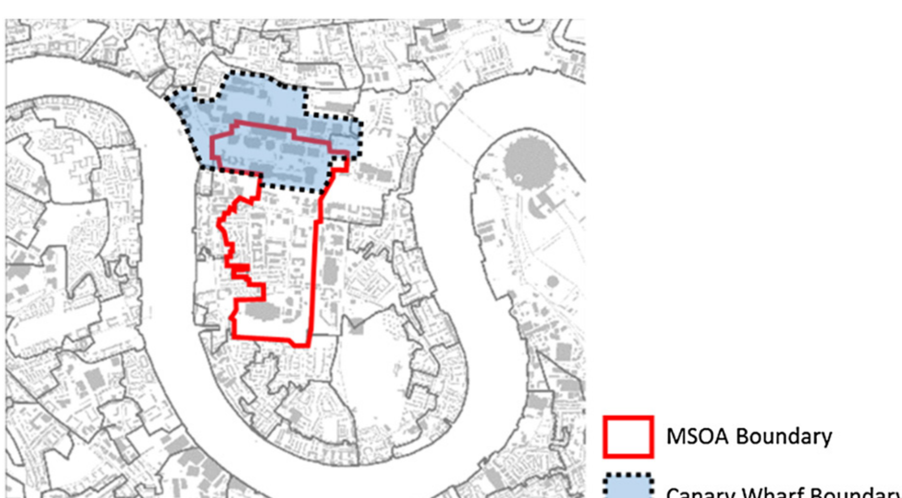


Fig. 1. Canary Wharf Boundary in blue overlaid with MSOA boundary in red.

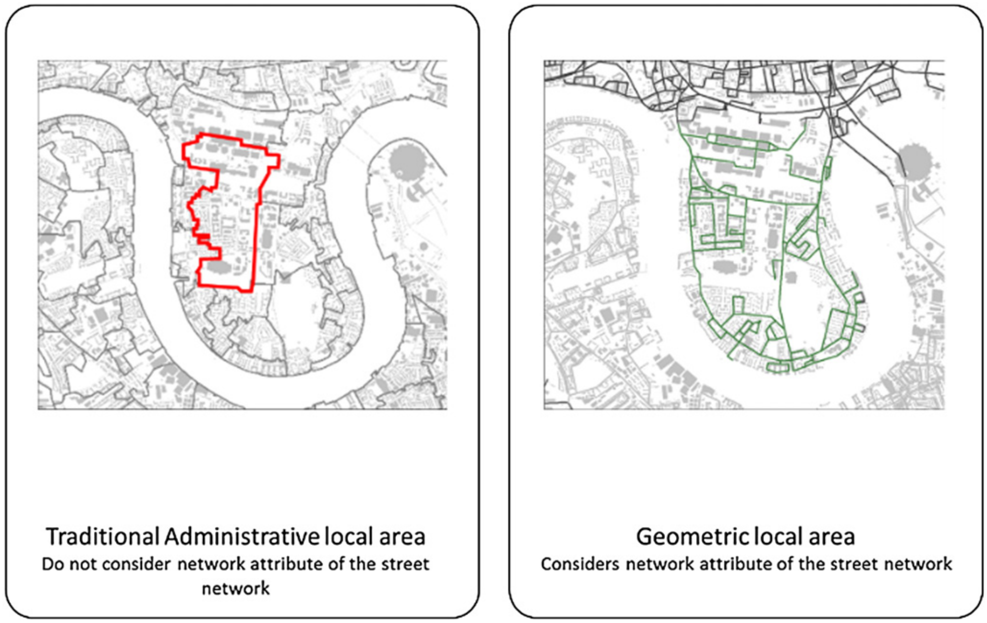


Fig. 2. Traditional administrative local area on the left and Street-based Local Area (SLA) on the right. Ordnance Survey © Crown Copyright. All rights reserved.

2.3. Defining SLA using the street-network dual graph

In graph theory, a spatial street network is a type of planar graph embedded in Euclidean space. Two types of spatial street network graph could be defined: the spatial primal graph (PG), whose vertices are junctions and edges are streets, or the spatial dual graph (DG) whose vertices (u) are streets and edges (e) are junctions (Porta, Crucitti, & Latora, 2006). The ladder had been made popular from space syntax research (Hillier & Hanson, 1984).

$$\begin{aligned}&DG\,(u,e).\\&\text{where}\\&u \text{ is the node (street segments)}\\&e \text{ is the edge (junctions)}\end{aligned} \quad (1)$$

This study will employ community detection technique on the spatial dual graph of the street network in defining SLA (Turner, 2007). More formally, SLA is defined as a discrete subgraph (subset) of the spatial dual graph DG. All vertices (streets) classified within each subgraph shares a membership.

$$\begin{aligned}&SG_k \subseteq DG \text{ where } k = 1, 2, \ldots, K\\&SG \text{ is the subgraph.}\\&DG \text{ is the spatial dual graph.}\\&K \text{ is the number of subgraph.}\end{aligned} \quad (11)$$

A rationale in the use of the dual graph representation is that a property is on a street rather than on a junction. Community detection on a primal graph will pick out clusters of connected junctions rather than clusters of connected streets. The next section will describe the community detection method that identifies the subgraph.

2.4. Multi-level modularity optimisation algorithm on the street-network dual graph

Large numbers of research had been conducted concerning the identification of community structures. Many algorithms were proposed including modularity-based algorithm, the Spinglass Algorithm, the Walktrap algorithm, the Betweenness Cut algorithm and the Vertex Propagation algorithm (Reichardt & Bornholdt, 2004; Raghavan, Albert, & Kumara, 2007; Newman & Girvan, 2004; Pons & Latapy, 2006) This study in particular adopts the Modularity Optimisation algorithm on the street-network dual graph to identify Street-based Local Area (SLA). The technique is one of the most commonly used community detection method that is known for its efficiency and accuracy (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008; Lancichinetti & Fortunato, 2009).[1] The algorithm optimises against a community quality function called Modularity. Modularity (2) calculates the difference between observed number of edges within a subgraph and the expected number of edges. The greater the observed number of edges relative to expected, the higher the modularity. More formally, Modularity (Q) is defined where A is the adjacency matrix, m is the total number of edges in the graph, ki and kj are the degrees for vertex i and vertex j. δ is 1 if i and j are in the same community and zero otherwise.

Eq. (2) Modularity(Q) equation (Girvan & Newman, 2002).

$$\begin{aligned}&Q = \text{Observed number of edges in a community} - \text{Expected number of edges in a community}\\&Q = \frac{1}{2m}\sum (A_{ij} - K_i K_j / 2m)\,\delta(C_i, C_j)\\&\text{Where}\\&A \text{ is the adjacency matrix.}\\&m \text{ is the total number of edges.}\\&Ki \text{ and } Kj \text{ are the degree for the two subgraphs } i,\ j.\\&\delta \text{ is a Kroneckar Delta function which equals 1 when its argument are the same and 0 otherwise.}\end{aligned} \quad (2)$$

Optimisation against the above function is currently impossible to solve for large datasets.[2] As a result, a number of heuristic algorithms had been implemented into finding the optimal sub-graph (Girvan & Newman, 2002). This study will apply specifically the multi-level method (Blondel et al., 2008) in optimising against the modularity function as shown in Fig. 3.

The modularity optimisation algorithm starts where every vertex is a sub-graph. Every vertex would then share sub-graph membership with its neighbour that attains the highest modularity score. This continues for all vertices. After all vertices have been traversed, vertices within the same sub-graph would aggregate into a new super vertex and a new super-graph formed. The super vertices of the new super-graph would again optimise its modularity, sharing sub-graph membership with its neighbours. This aggregation continues until modularity

[1] Initial research suggests other techniques found less satisfactory results. Further comparative research is required.

[2] This is a class NP-hard problem in computation.

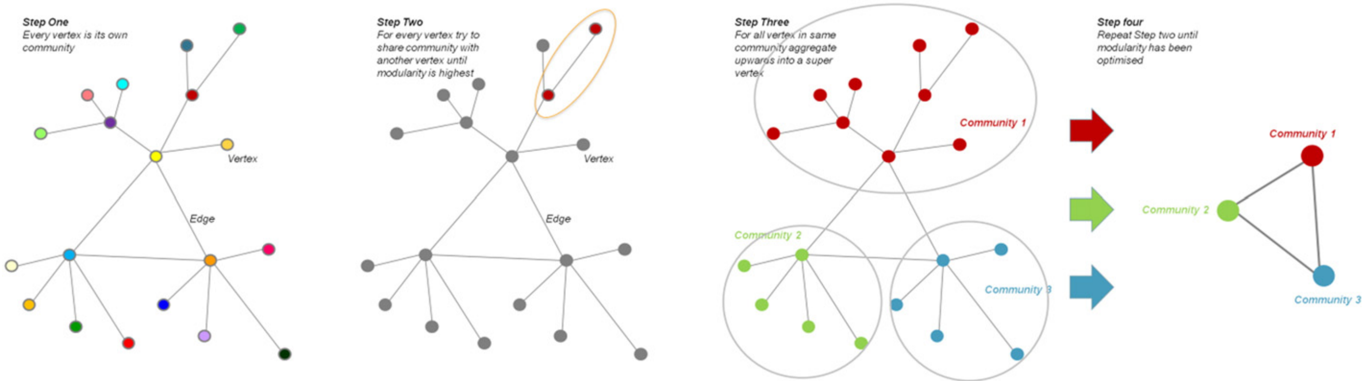


Fig. 3. The Modularity Optimisation method starts where every vertex is a community. Every vertex will then share community membership with its neighbour that attains the highest score. This continues for all vertices. Vertices within the same community will aggregate into a super vertex. These super vertices will again optimise its modularity until modularity can no longer be optimised. Diagram produced by the Author.

could no longer be optimised. The method is multi-level and hierarchical, whereby each subgraph produced is part of a larger super-graph in the next iteration.

3. Methodology.

3.1. Multi-level hedonic price model

In order to answer the research question, this research will adopt the multi-level hedonic price regression model introduced by Orford (1999) and Goldstein (1987) to estimate the Street-based Local Area (SLA) effect on house prices in London. The rationale in the use of multi-level hedonic regression model over a typical OLS hedonic regression model is that it examines hierarchically nested group effects. Simple OLS models simply ignore average variations between groups whereas individual regression between each local area would face sampling problems and poor generalisation. Examples of multi-level hedonic studies include the aforementioned study from Orford (1999; 2001), who provided the evidence to use multi-level models in hedonic price model through the case study of the Cardiff. He found that house price variations from the grand mean can be decomposed into variations across enumeration districts, local communities and individual properties. Orford (2001) also found that primary school quality has greater local effect and parks have greater global effect, resulting in a complex geography of juxtaposing location externalities. Empirically, multi-level methods were also able to account for spatial autocorrelation[3] of the error term otherwise known as neighbourhood effect as properties in local areas are more similar than properties in other areas.

The following section will describe the multi-level hedonic regression model used for this study, specifically in modelling the property effect at level 1 and the local area effect at level 2. Due to the scope and length of the paper, the submarket effect at level 3 will be developed in a separate article. In a typical multi-level hedonic framework, the observed variable is a function of two components, a fixed part and a random part. The fixed part can be the mean or a collection of independent variables and the random part is simply the deviation from the mean. If we want to account for the hierarchical local area effects, we will decompose the fixed part into its mean (u) and fixed level predictor (X_i) and the random part into individual local area effect (u_i) and its error (e_{ijk}).

[3] Spatial autocorrelation refers to spatial association between proximate properties. Please refer to Anselin (1988) for a more informed discussion of the topic.

Eq. (3) multi-level regression model

$$\underbrace{Y}_{Observed} = \underbrace{BX_i + \mu}_{Fixed} + \underbrace{\mu_i + \varepsilon_{ijk}}_{Random}$$

Where
Y is the observed.
B is the coefficient for predictors.
X_i is the predictor.u is the mean.
u_i is the local area random effect.
e_ijk is the error term. (3)

For the empirical study, we estimate first a base grand mean model then four nested multi-level models for the Street-based Local Area (SLA). When local area effects are included, the dimension of the data increases. As a result, we will estimate Model 3 and Model 4 with a narrow set of fixed predictors, namely size and accessibility, and, for Model 5 a, wider set of predictors. The hierarchical multi-level model is illustrated in Fig. 4. The five candidate models are illustrated in Table 1.

The starting point of a multi-level hedonic model is the base model, where there are no explanatory variables specified in the regression model. This is also known as the grand mean model (4).

Eq. (4) Base model 1

$$\mathrm{Log}(HP_{ij}) = \mu + e_{ij} \tag{4}$$
u is the overall mean.
e_ij is the error.

Model 2 is a level two varying intercept model that accounts for the street-based local area effects. Model 2 (5) is as follows, where HP_{ij} is the property price, u is the overall mean, u_j is the local area effect on house prices and e_{ij} is the error term. No explanatory variable is specified for the model.

Eq. (5) Model 2 is the varying intercept model

$$\mathrm{Log}(\mathrm{HP}_{ij}) = \mu + \mu_j = \mathrm{e}_{ij} \tag{5}$$
u is the overall mean.
u_j is the local area effects.
e_ij is the error term.

Model 3 is a level two varying intercept model with fixed predictors. The predictors include space syntax integration (access) and the floor size (Floor) of the property. Model 3 (6) is as follows, where HP_{ij} is the property price, u is the overall mean, u_j is the local area effect on

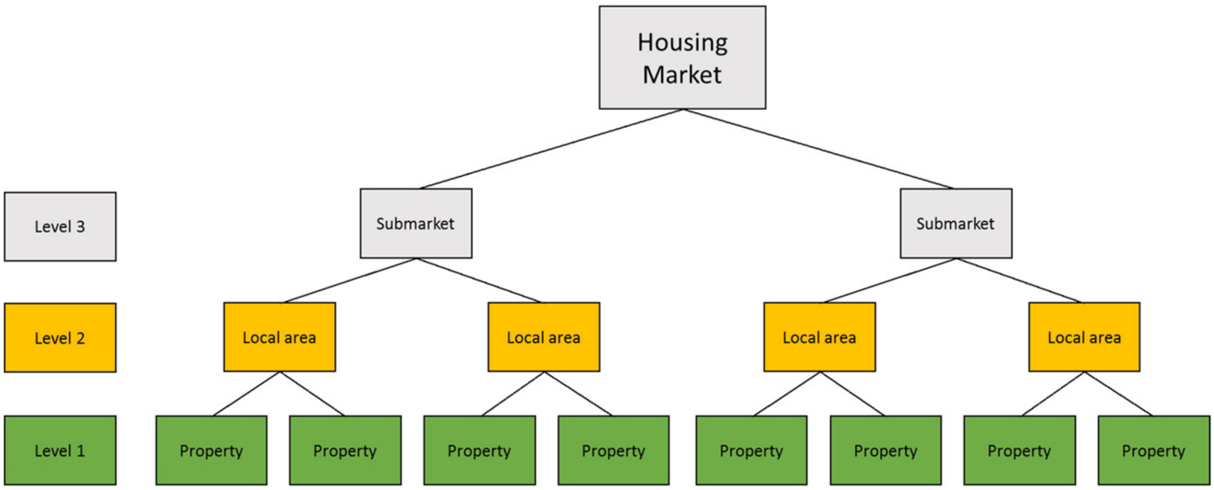


Fig. 4. The hierarchical multi-level model.

house prices and e_{ij} is the error term. B1 is the parameter estimated for space syntax integration and B2 is the parameter estimated for floor area.

Eq. (6) Model 3 varying intercept model with fixed predictors

$$\mathrm{Log}(\mathrm{HP}_{ij}) = \mu + \mathrm{B1}_j * \mathrm{Access}_i + \mathrm{B2} * \mathrm{Floor}_i + \mathrm{u}_j + \mathrm{e}_{ij} \tag{6}$$

u is the overall mean.
B1 is the coefficient for accessibility.
Access is the accessibility variable.
B2 is the coefficient for floor size.
Floor is the floor size variable.
u_j is the local area effects.
e_ij is the error term.

Model 4 is a level two varying intercept and slope model with fixed predictors. The model accounts for local area effect adjusted for fixed effect predictors. This model includes space syntax integration as both a property effect and local area effect. Model 4 (7) is as follows, where HP_{ij} is the property house price, u is the overall mean, u_j is the local area effect on house prices and e_{ij} is the error term. B1 is the parameter for integration and B2 is the parameter estimated for floor area.

Eq. (7) Model 4 varying intercept, varying coefficient model with fixed predictors

$$\mathrm{Log}(\mathrm{HP}_{ij}) = \mu + \mathrm{B1}_j * \mathrm{Access}_i + \mathrm{B2} * \mathrm{Floor}_i + \alpha_j + \mathrm{e}_{ij} \tag{7}$$

u is the overall mean.
B1_j is the coefficient for accessibility.
Access is the accessibility variable.
B2 is the coefficient for floor size.
Floor is the floor size variable.
α_j is the local area effects.
e_ij is the error term.

Table 1
Candidate models for SLA.

Base model	Model 1	Grand mean model **level 1**
Street-based Local Area Model	Model 2	Varying intercept model **level 2**
	Model 3	Fixed predictors and varying intercept model **level 2**
	Model 4	Fixed predictors and varying intercept and slope model **level 2**
	Model 5	Wider set of fixed predictors and varying intercept and slope model **level 2**

Model 5 is a level two varying intercept and slope model with wider set of fixed predictors. This is the same model as the previous one, but with the addition of a wider set of parameters. This includes dwelling type, number of shops in the vicinity[4] and the quality of education as determined by average A-level score.[5] Model 5 (8) is as follows, where HP_{ij} is the property price, u is the overall mean, u_j is the local area effect on house prices, and e_{ij} is the property level error term. $\sum \mathrm{B}_n$ are the parameters estimated for the independent variables.

Eq. (8) Model 5 varying intercept, varying coefficient model with fixed predictors

$$\mathrm{Log}(\mathrm{HP}_{ij}) = \mu + \mathrm{B1}_j * \mathrm{Access}_i + \mathrm{B2} * \mathrm{Floor}_i + \mathrm{B3} * \mathrm{Dwelling1}_i + \mathrm{B4} * \mathrm{Dwelling2}_i + \mathrm{B5} * \mathrm{Shop}_i + \mathrm{B6} * \mathrm{School}_i + u_j + \varepsilon_{ij} \tag{8}$$

u is the overall mean.
B is the coefficient for predictors.
Access is the accessibility variable.
Floor is the floor size variable.
Dwelling1 represents flats.
Dwelling2 represents terraces.
Shop represents the number of shops within 800 m.
School is the Average A-level score within 800 m.
u_j is the local area effects.
e_ij is the error term.

Multilevel models are commonly estimated using a maximum likelihood estimator (MLE).[6] Standard statistics for Multilevel models would be reported. This includes the Likelihood Ratio (LR) and Intra-class Correlation Coefficient (ICC). The LR is a test statistic that compares how well each candidate model fits with its respective null model. The test statistic is chi-square distributed and would be calculated to test the significance of the local area effect on house prices.[7] The null model is rejected in favour of the multilevel model if the P-Value > 0.05. The

[4] Active use is classified under the retail category in the Valuation Office Agency's business rates data. Data provided by the Valuation Office Agency contains public sector information licensed under the Open Government Licence v1.0.

[5] A-Level scores (General Certificate of Education Advanced Level) is an academic qualification offered by educational institutions in England, Wales and Northern Ireland to students completing secondary or pre-university education

[6] MLE have been estimated using the Stata software which uses the Newton-Raphson gradient-based method.

[7] Log likelihood ratio is a common statistical test for MLE to compare fit between the null model and alternate model. The test statistic has an approximate chi-squared distribution with the degree of freedom equal to the df of alternative model – df of null model. It is calculated as follows. $\begin{array}{l}\mathrm{LR} = -2 * [\ln(\mathrm{LL}_{null})] + 2 * [\ln(\mathrm{LL}_{multilevel})] \\ LL = loglikelihood\end{array}$ (12)

Table 2
Candidate models for all local areas.

	SLA	LSOA	MSOA	Ward
Grand mean model **L1**	Model 1	Model 6	Model 11	Model 16
Varying intercept model **L2**	Model 2	Model 7	Model 12	Model 17
Fixed predictors and varying intercept model **L2**	Model 3	Model 8	Model 13	Model 18
Fixed predictors and varying intercept and slope model **L2**	Model 4	Model 9	Model 14	Model 19
Wider set of fixed predictors and varying intercept and slope model **L2**	Model 5	Model 10	Model 15	Model 20

null model in each case is the same as the Ordinary Least Square (OLS) model without the local area effect. This allows the isolation of local area effect for each multi-level model. The ICC, on the other hand, would also be calculated for each SLA multilevel model to measure the amount of variation the local area effect captures, in proportion to the overall house price variance.[8]

In order to compare across the five candidate models, the Akaike Information Criterion (AIC) will be computed. The AIC[9] is a common metric of good fit (in terms of loglikelihood), adjusted for the number of parameters. AIC is calculated for all the candidate models and compared where the lower the criterion, the better the quality of the model. It is a more robust measure of good fit than the loglikelihood.

3.2. SLA and administrative local area model comparison

This section compares the extent to which Street-based Local Area (SLA) differs from other administrative units. The same multi-level hedonic price approach specified in Section 3.1 is applied to three commonly used administrative units in the UK, namely, statistical ward, lower super output area and medium super output area. Similar to the last section, the candidate models are compared through the AIC goodness of fit. In total, twenty models are being estimated. See Table 2.

4. Datasets and case study

4.1. The Greater London Area

The Greater London Area in the UK is used as the case study. The extent of the study area is presented in Fig. 5, where the black line indicates the study boundary, the red line indicates the 33 administrative borough boundaries of Greater London (Ordnance Survey, 2015), and the grey line indicates the meridian line street network.

4.2. Residential sold price

This study uses the house price dataset from the Nationwide Building Society.[10] House price in this research is defined as the exchange value between the buyer and seller. A total of 5344 observations from 2011 are used. Fig. 6 shows house prices in London for 2011 mapped

Fig. 5. Study area boundary.

in GIS, whereby red indicates a higher house price and blue indicates a lower house price. The thematic distribution in GIS is calculated using the natural break method for 8 bands.

4.3. London street network

The London pedestrian street network is used to compute the accessibility measure and to construct street-based local areas (SLA) for the empirical study. The basis of the London street network is the Ordnance Survey Meridian 2 network.[11] (Ordnance Survey, 2014). The street network dataset has a total of 113,555 street segments as illustrated in Fig. 7.

4.4. Descriptive statistics

In the hedonic price approach, structural features, accessibility levels and neighbourhood amenities were often included in the empirical model (Rosen, 1974; Cheshire & Sheppard, 1998). Below are a set of variables included for the study. This includes structural features such as property size, dwelling type (flat, house, terrace), location accessibility such as street network closeness centrality (Law et al., 2013; Xiao, Webster, & Orford, 2015)[12] and neighbourhood amenities such as the number of retail units within 800 m (Des Rosiers et al., 1996), the secondary school average score within 800 m (Black, 1999; Gibbons and Machin, 2003; 2008). Table 3 describes the basic statistics for the London house price dataset in 2011. The mean house price is approximately 350,000GBP with a mean floor size of 99sqm, a mean bedroom of 2.6 and a mean property age of 85 years old (See Table 3).

4.5. London street-based local area

Applying the modularity optimisation algorithm described in 2.4 on the OS Meridian line network, a total of 207 Street-based Local Area

---

8 $\mathrm{ICC} = \frac{(\mathrm{Var}_{\mathrm{Level}\ 2})^2}{[(\mathrm{Var}_{\mathrm{Level}\ 2})^2 + (\mathrm{Var}_{\mathrm{Level}\ 1})^2]}$ (13)
$\mathrm{Var} = \text{variance}$

9 $\mathrm{AIC} = -2 * \mathrm{LL} + 2 * \mathrm{k}$
$\mathrm{LL} = \text{loglikelihood}$ (14)
$\mathrm{k} = \text{number of parameters}$

10 The data was provided by the Nationwide through a licensing agreement with London School of Economics. The Nationwide dataset is a subset of the Land Registry dataset. The origins of all data on sold house prices in United Kingdom is owned by Land Registry/Registers of Scotland © Crown copyright 2013.

11 Ordnance Survey Open Data Meridian 2 Dataset. © Crown Copyright {2014}

12 Spatial network closeness centrality or integration in the spatial configuration literature measures the reciprocal sum of the shortest path between every origins (i) to every destinations (j). (Hillier and Iida, 2005) Spatial Network Closeness centrality were found to have significant positive association with house price suggesting places that are more central achieve higher house price (Xiao et al., 2015; Law et al., 2013).

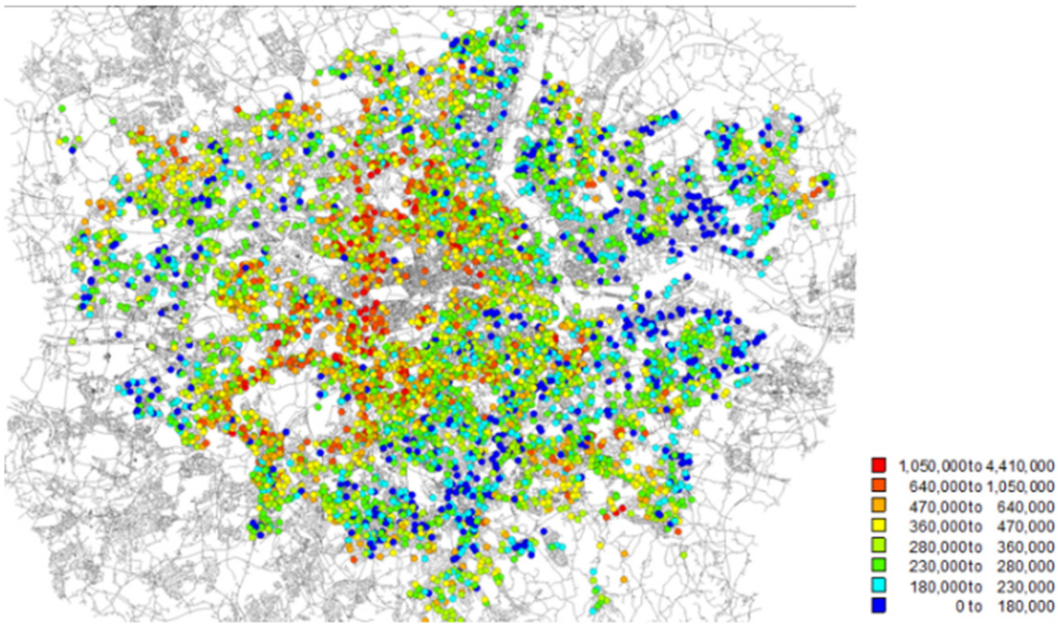


Fig. 6. Visualisation of London House Price in 2011 from red indicating high house price to blue indicating low house price.

Fig. 7. London OS Meridian 2 spatial street network.

(SLA) were identified for the Greater London Area. Each SLA has on average 549 segments, with a standard deviation of 257 segments (See Table 4).

Fig. 8 below shows the SLA of London. The figure shows distinct SLA mapped in GIS where the different colour corresponds to different membership. Visually the results show clear topologic distinction for SLA separated by the River Thames such as the Isle of Dogs and local areas separated by the Lea Valley and railway tracks. This discrete representation is also a limitation as neighbourhood boundaries are often described with varying level of sharpness/fuzziness.

4.6. London administrative local area units

Three different administrative local area units will be compared with the street-based local area in the following empirical study namely Lower Super Output Area (LSOA), Medium Super Output Area and Statistical Ward[13] (See Table 5).

[13] Electoral wards/divisions is the key local area unit for UK administrative geography. There are a total of 9456 wards in the UK with an average population of 5,500 person per ward. (ONS, 2015).

Table 3
Descriptive statistics for house price and attributes.

Variables	Description	(1) Mean	(2) sd
Price	Transaction price	356,481	213,846
Bedrooms	Number of bedrooms	2.604	1.006
Floorsz	Floor size	99.03	40.73
Age	Age of property	85.05	36.35
CC	Closeness centrality	8721	2719
BC	Betweenness centrality	2.643e + 07	1.087e + 08
Shops	No. Shops within 800 m	354.3	407.7
Park	Distance to royal parks	10,355	5272
School	Average A-level score in 800 m	366.2	390.6
New_build_dum1	More than five years old	0.986	0.117
New_build_dum2	New-build	0.0139	0.117
Tenure_dum1	Freehold	0.592	0.492
Tenure_dum2	Leasehold	0.408	0.492
Type_dum1	Terrace	0.314	0.464
Type_dum2	Flat	0.405	0.491
Type_dum3	House	0.281	0.449

Table 4
Street-based Local Area statistics.

	N
SLA	207
Total street segments	113,555

	Segments per local areas
Average	549
Std Dev	257
Min	73
Max	1243

Fig. 9 illustrates the Isle of Dogs area overlaid with the Lower Super Output Area, Medium Super Output Area and Electoral Wards boundaries. The cyan line denotes the LSOA boundaries, the blue line denotes the MSOA boundaries and the red line denotes the ward boundaries and the dark grey regions represent the built form. As seen, the three boundaries largely follow the separation created by the River Thames. However, the divisions are clearly more arbitrary in the central area as boundary lines cut across streets and buildings. In contrast, the SLA level illustrated in the previous section largely follows both the spatial separation caused by the River Thames as well as the morphology of the local area.

Table 5
Administrative Local area statistics.

	N
LSOA	4765
MSOA	983
Ward	629

5. Empirical results

The following section illustrates the empirical results for this article. We will first test the significance of SLA on house prices using the multilevel hedonic approach as specified in Section 3.1. This will be followed by a comparison with the different administrative units as specified in Section 3.2.

5.1. Street-based local area - multilevel regression results

The first part in the analysis is to study the extent to which SLA effects are evident in associating with house price. Table 6 illustrates the regression results for the five candidate models (See Table 6).

Model 1 is the null model where the grand mean of Log house price is 12.660 and the residual illustrates the total variance away from the mean is 0.220. Model 2 is a level two varying intercept model where the between-SLA (level 2) variance in house price is 0.085 and the between property (level 1) variance is 0.151. Model 3 is a level two varying intercept model with fixed predictors where the between-SLA (Level 2) variance in house price is 0.078 and the between property variance is 0.043. The reduction of property variance is to be expected due to predictor inclusion. Model 4 is a level two varying intercept and slope model with fixed predictors where the between-SLA (Level 2) variance in house price is 0.070, the between-SLA integration (Level 2) variance is 0.016 and the between-property (level 1) variance in house price is 0.041. Model 5 is a level 2 varying intercept and slope model with wider sets of fixed predictors where the between-SLA (Level 2) variance in house price is 0.058, the between-SLA integration (Level 2) variance is 0.014, and the within-property (level 1) variance in house price is 0.034. The reduction in property variance is again to be expected due to a wider set of fixed predictors.

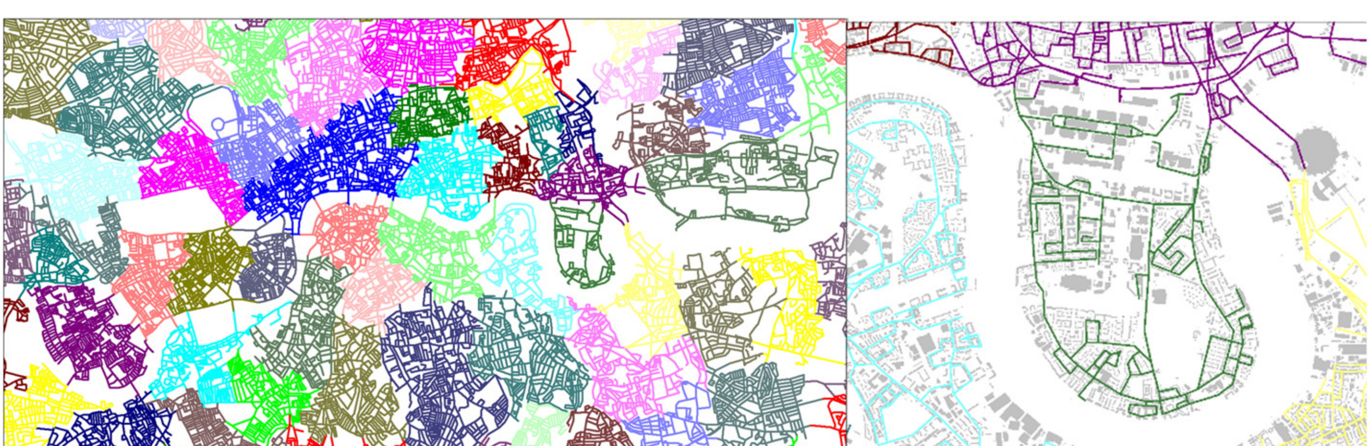


Fig. 8. Visualisation of Street-based Local Area (SLA) membership for the Greater London Area.

Fig. 9. Isle of Dogs area as denoted by LSOA, MSOA and Ward boundaries.

Local area effect remains relatively stable, with a small reduction due to the inclusion of fixed effect predictors in Model 3 and Model 5. This shows the relative stability of local area effect on house price. The overall loglikelihood ratio test (Prob > chi2 = 0***) shows significance for all the candidate models. This shows robust evidence that street-based local area effect is significant.

Table 7 summarises the goodness of fit as measured by AIC between the five candidate SLA models. The reduction in AIC shows clear improvement in statistical significance, as we allow for progressive inclusion of local area effect and fixed predictor effect. Both local area effect in model 2, fixed predictor in model 3 and model 5 have significant improvements in statistical significance. These improvements show that evidently the housing market is hierarchically nested in at least two levels, namely property level and local area level (See Table 7).

Table 6
Hedonic price regression model results.

SLA	Model 1	Model 2	Model 3	Model 4	Model 5
Lnprice					
Integration			0.046	0.058	0.034
			0.008	0.013	0.012
Floor size			0.342	0.340	0.277
			0.003	0.003	0.003
Age					0.030
					0.003
Park					0.127
					0.013
Shops					0.034
					0.005
Terrace					0.093
					0.004
Flat					0.030
					0.003
School					0.010
					0.003
_cons	12.660	12.620	12.620	12.610	12.630
	0.006	0.022	0.020	0.020	0.018
_e	0.220	0.151	0.043	0.041	0.034
	0.004	0.003	0.001	0.001	0.001
Local area effect		0.085	0.078	0.070	0.058
		0.010	0.009	0.009	0.007
Local area integration				0.016	0.014
				0.00297	0.00245
N	5334	5334	5334	5334	5334
LR test (Prob > 0.05)	0***	0***	0***	0***	0***

Fig. 10 below summarises the intra-class correlation coefficient for each model. The ICC measures the amount of variation captured by local area effect and property effect, in proportion to the overall house price variance. Blue denotes property variance and orange denotes local area variance. Property variance dropped in model 3 to approximately 30% due to the inclusion of predictors. Local area variance is constantly above 30% and remained relatively stable after the inclusion of fixed predictors.

For empirical reasons, spatial autocorrelation effects were checked. Moran's I were calculated with a minimum radius of 2400 m for Model 1 and Model 5. This radius was used to ensure there was a significant sample for each data-point to calculate the statistic. Global spatial autocorrelation reduced from 0.27 for Model 1 to 0.004 for Model 5. The P-value shows weak significance at the Prob > 0.01 level. This confirms previous research on the use of multi-level hedonic models in reducing spatial autocorrelation (Orford, 1999). For details please see Appendix A.

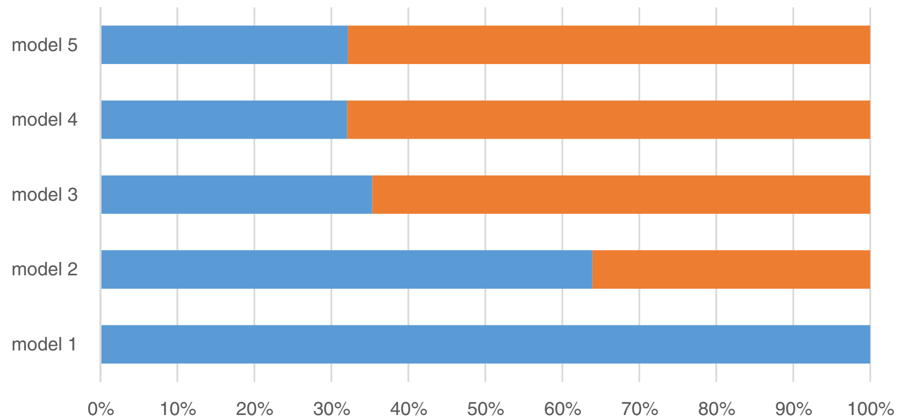
5.2. SLA and administrative local area comparison

This section compares the five candidate models using different local area units. Similarly, the LR tests for all five candidate models are significant at the prob. > 0.01 level. Fig. 11 below shows a goodness of fit comparison between the five local area units using AIC. SLA is denoted in light blue, ward is denoted in orange, MSOA is denoted in grey and LSOA is denoted in yellow.

The downward trend of the AIC shows the joint effect of property characteristics and local area effect on house prices progressively across local area units. This confirms Orford's research (1999) on the hierarchical nature of the housing submarket, where the London housing market is nested in at least two levels. The result also showed clear differences in results across different administrative units confirming previous research. The result also showed SLA is consistently preferred to all the other administrative units, including electoral wards, MSOA and LSOA. Together, this evidence confirmed street-based local area effect on house prices.

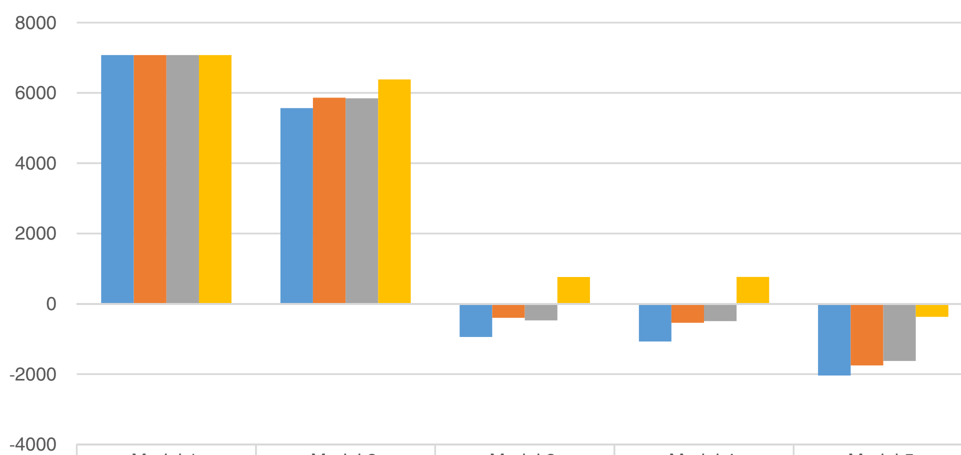
Table 7
Candidate model AIC comparison.

SLA	Model 1	Model 2	Model 3	Model 4	Model 5
AIC	7075.495	5567.136	−944.593	−1072.618	−2040.18



	model 1	model 2	model 3	model 4	model 5
property	100%	64%	35%	32%	32%
local area		36%	65%	68%	68%

Fig. 10. Intra-class correlation coefficient comparison.



	Model 1	Model 2	Model 3	Model 4	Model 5
GLA	7075.495	5567.136	-944.5928	-1072.618	-2040.18
Ward	7075.495	5863.705	-400.0294	-540.859	-1751.261
MSOA	7075.495	5845.4	-472.6239	-492.6706	-1626.684
LSOA	7075.495	6382.405	762.4071	764.4071	-370.5573

Fig. 11. Goodness of fit comparison (AIC) across different local area units.

6. Discussion

The main contribution of the research is the novel application of community detection techniques on the street network dual graph to defining Street-based Local Areas (SLA) in London. The results showed that local areas have a significant effect on house price variations. The results also showed that SLA is able to capture socio-economic similarity more accurately than region-based local area units. The plausible reasons are threefold; firstly, people perceived the local area on a street network. The street network is therefore able to capture, more precisely subtle differences in an urban environment and more accurately the perceptual definition of a local area than ad-hoc administrative region. To explain this concept, we will go to Fig. 12, which illustrates two distinct local areas connected via a bridge (one in orange and one in green).

If we pick any orange node randomly in the network, the chances of ending up in another orange node is much greater than a green node. Using this analogy, the probability of walking within the same subgraph or identifying the highly connected subgraph as a local area is much greater than in another subgraph. On aggregate, the topology of the street network could capture more accurately the perceptual definition of a local area. The result could also provide linkage between spatial network clusters and collective perception of neighbourhoods. To verify this, future empirical research would be needed where individual perception maps are compared to street-based local area units (Lynch, 1960; Coulton et al., 2001).

Secondly, the topology of the street network reinforces socio-economic similarity within the local area and overtime reinforces the perception of the local area. To illustrate this, we go back to the conceptual diagram in Fig. 12 where we run a conceptual simulations of an agent walking in this network in Fig. 13.[14]

[14] This notional simulation takes inspiration from the Walktrap algorithm (Pons & Latapy, 2006).

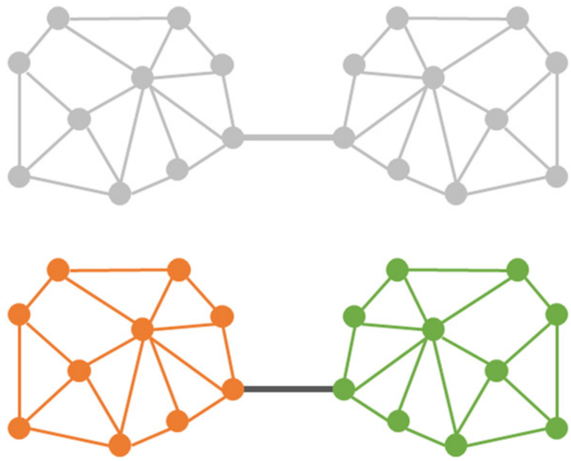


Fig. 12. The top image illustrates a graph. The bottom image illustrates an obvious division between two subgraphs connected via a bridge. The subgraph to the left is coloured orange and the subgraph to the right is coloured green and the bridge is coloured grey.

The simulation starts by having an agent that starts from a random orange node then randomly walks to a connected node. The number of random steps required to reach the green subgraph would then be recorded. The first simulation in the figure shows an agent took 9 steps to reach a green node. The second simulation in the figure shows an agent took 8 steps to reach a green node. A plausible future is that, over time, differences between areas will become more pronounced as like-minded people cluster together and bump into each other. This thus reinforces socio-economic similarity within a SLA and the boundaries between SLA. Plausible processes allowing this to happen include crowd behaviour and bounded rationality where information is constrained within the local area. To verify this, a key question to ask in the future is to what extent do social constructs, perceptual clusters and topologic clusters overlap with SLA across space and time.

Thirdly, as people identify these local areas they would make decisions based on it. The local area becomes part of the housing bundle leading to it having an effect on house price. For example, when we purchase a property in Kensington, we are also buying a Kensington local area premium as part of the housing bundle. Therefore, given the exact same house, a buyer would value a house more similarly to one within the same local area than to one in another local area. From the geographical science perspective, this could also be interpreted through Tobler's first law, where properties that are closer to each other are likely to be more socio-economically similar than properties that are further away (Tobler, 1970). Overtime, local areas would become more socio-economically homogeneous reinforcing further the effect on house price.

6.1. Benefits and limitations

There are a number of benefits in defining Street-based Local Area (SLA). Firstly, by using the street network as the geographic unit, it can reduce the modifiable areal unit problem of using region-based geographies. Second, SLA can capture more accurately subtle differences in urban environments than ad-hoc administrative regions. Third, as the street network is clearly the most permanent of all morphological elements, SLA can be considered as a slow dynamic. The slowness allows the data to be consistently compared across time but at the same time dynamic enough to reflect the changes of the street network and morphology. To demonstrate the benefits, we run the simulation as described in Fig. 13 multiple times. Fig. 14 illustrates the average number of random steps an agent at an orange subgraph would take to reach the green subgraph 500 times and four different configurations.

With one bridge, it takes on average 40 random steps, with two bridges it takes on average 30 steps, with three bridges it takes on average 15 steps and with four bridges it takes on average 10 steps. One can clearly see that the more bridges there are between the two subgraphs the lower the number of average random steps any agent will need to reach the adjacent subgraph. The simulation shows if we add one

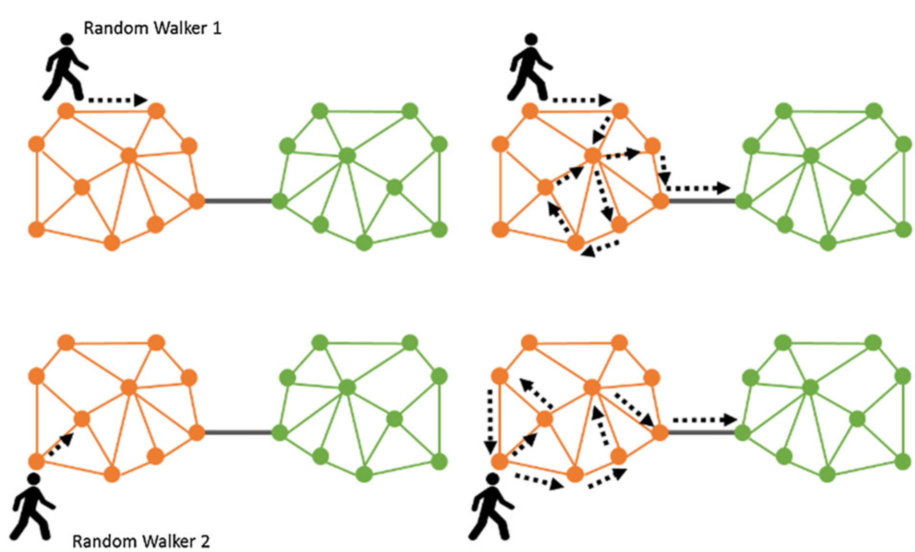


Fig. 13. This figure illustrates two simulations of an agent starting from a different orange node who walks randomly around the graph. The first simulation shows the walker took 9 steps to reach the bridge. The second simulation shows the walker took 8 steps to reach the bridge. This illustrates that a random walker is likely to stay longer within a local area purely by chance when there is greater intra-cluster connectivity.

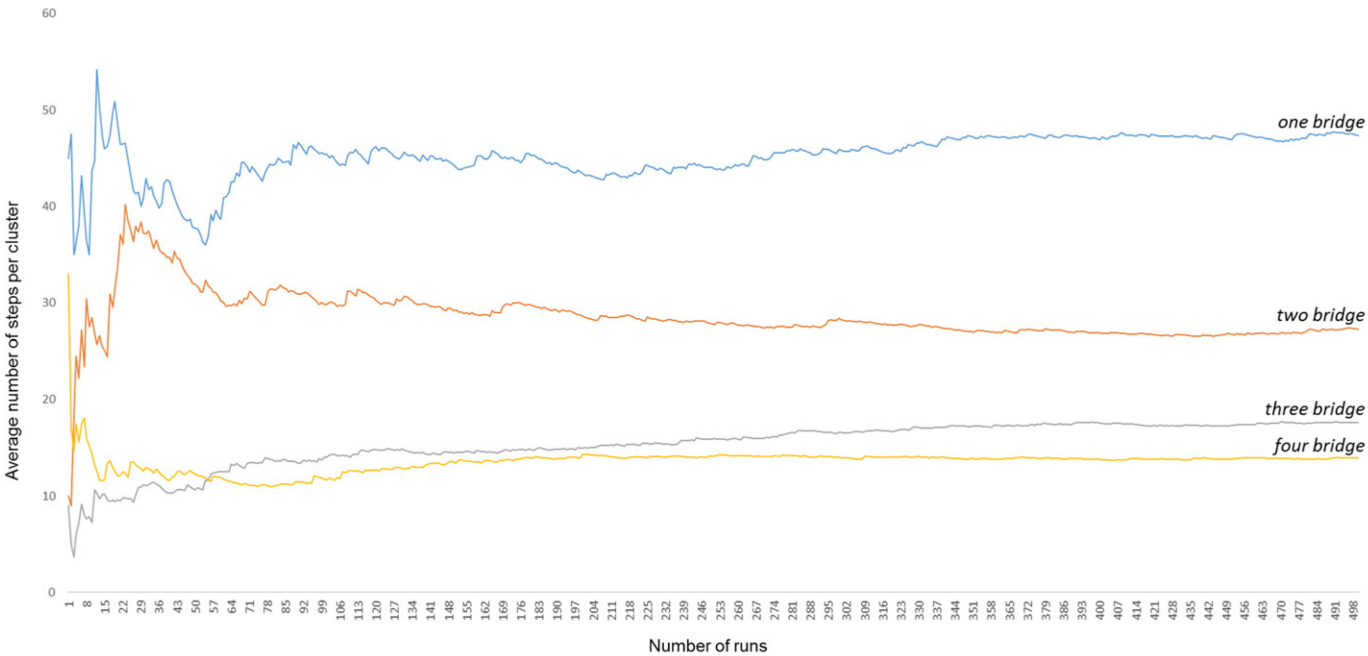


Fig. 14. Average number of steps required to jump between clusters.

more bridge across the two Street-based Local Areas (SLA), the probability of ending up in a green node would increase substantially. The result shows how subtle differences in an urban grid could influenced socio-economic spillover.

The definition of SLA is not without its concerns. First, this research suggests on aggregate, SLA is able to capture subtle differences of an urban environment more accurately than region based methods. However, at an individual-level more research is required to understand and confirm how this happened and what are the processes that influences the construction of individual cognitive boundaries (Tolman, 1948). Second, considering the street network provides an entirely singular approach to defining local area. When a grid is highly uniform and connected, street network connectivity might not be adequate in defining local area. For example, in central London or in many American CBD, the grid is too uniform to be separated by the grid. Instead, these areas might be more defined by other dimensions such as morphological, sociological, economical and historical characteristics. This constraint the feasibility of the method to be used in spatial planning. Future research are recommended to focus on joining these inner constructs in defining a more comprehensive definition of local area or neighbourhood for planning.

Third, the use of the modularity optimisation method defines sharp local area boundaries which contradicts to previous research in describing neighbourhoods as overlapping (Alexander, 1965). To overcome this limitation, one approach is to apply fuzzy-logic memberships in community detection. Lastly, research is needed to examine how Street-base Local Areas (SLA) can improve existing housing research such as the definition of housing submarkets. This will be discussed in a future article in applying street network attributes in forming housing submarket.

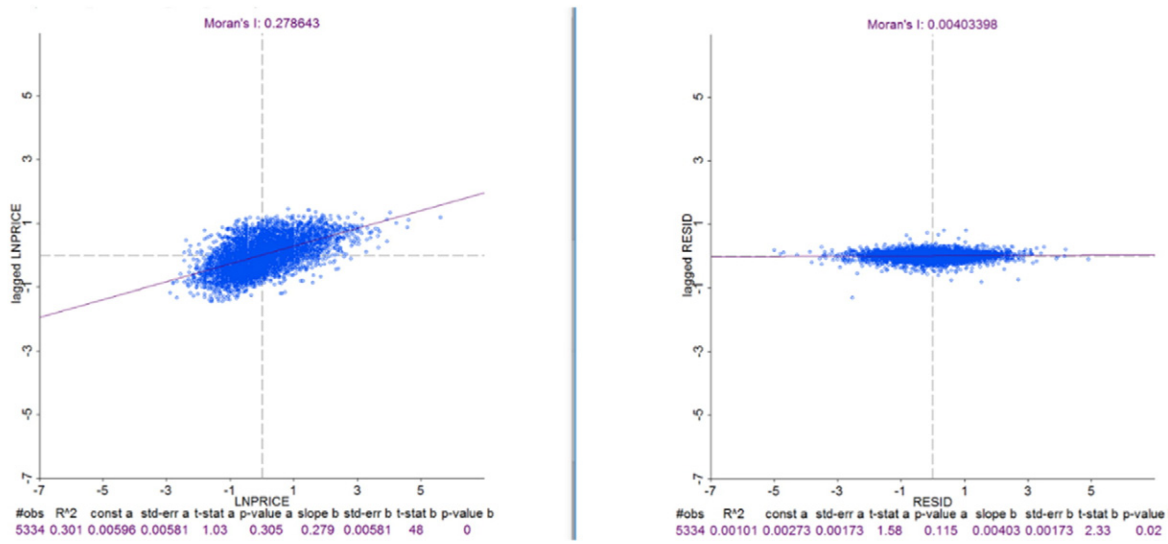
Despite the limitations of the approach, the definition of Street-based Local Area (SLA) is important as, it links the geometry of the street network to how we perceive local area. For real estate economists, this research highlights local area effects on house price which is important in house price prediction models. For urban planners, this research reveals considerable evidence or a belief that neighbourhoods are not only defined by socio-economic or historic dimensions but perhaps also through their spatial network topology/configuration. This is important, as administrative census tract had been used in many aspects of spatial planning. street-based methods could therefore provide an alternative to ad-hoc administrative local geographies in many aspects of neighbourhood planning.

Acknowledgement

A special acknowledgement to Professor Alan Penn, Dr. Kayvan Karimi and Dr. Jorge Gil at University College London and Professor Stephen Gibbons at London School of Economics for their help and discussions. I would also like to thank the reviewers for their comments and the editors for their editorial help and patience.

Appendix A. Global Moran's I

Left SLA Model 1.**Right** SLA Model 5.



References

Ahlfeldt, G. (2011). If Alonso was right: modeling accessibility and explaining the residential land gradient. *Journal of Regional Science*, *51*(2), 318–338.
Alexander, C. (April 1965). A City is Not a Tree. *Architectural Forum*, *Vol 122*(No 1), 58–62.
Anselin, L. (1988). *Spatial econometrics: Methods and models.* Dordrecht: Kluwer Academic Publishers.
Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics*, *114*(2), 577–599.
Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech.* (2008:P10008, 2008).
Bourassa, S. C., Hamelink, F., Hoesli, M., & MacGregor, B. D. (1999). Defining housing submarkets. *Journal of Housing Economics*, *8*, 160–183.
Caschili, S., De Montis, A., Chessa, A., & Deplano, G. (2009). Weighted networks and community detection: planning productive districts in Sardinia. In G., & M. (Eds.), *Planning, complexity and new ICT* (pp. 27–36) (Alinea Editrice s.r.l., 2009).
Cheshire, P., & Sheppard, S. (1998). Estimating the demand for housing, land, and neighbourhood characteristics. *Oxford Bulletin of Economics and Statistics*, *60*, 357–382.
Chiaradia, A. J., Hillier, B., Barnes, Y., & Schwander, C. (2013). Residential property value and urban form patterns in Greater London. *Urban Design and Planning, Issue DP3*, *166*(3), 176–199.
Coulton, C. J., Korbin, J., Chan, T., & Su, M. (2001). Mapping residents' perceptions of neighborhood boundaries: A methodological note. *American Journal of Community Psychology*, *29*, 371–383.
Dalton, N. S. C. (2006). Configuration and neighborhood: Is place measurable. *Space Syntax and Spatial Cognition Workshop of the Spatial Cognition.*
Des Rosiers, F., Lagana, A., Theriault, M., & Beaudoin, M. (1996). Shopping centres and house values : an empirical investigation. *Journal of Property Valuation & Investment*, *14*(4), 41–62.
Ellen, I. G., & Turner, M. A. (1997). Does neighborhood matter? Assessing recent evidence. *Housing Policy Debate*, *8*(4), 833–866.
Galster, G. (2001). *On the nature of neighbourhood. Urban studies. 12.* (pp. 2111–2124), 2111–2124 (Publisher Full Text OpenURL).
Gibbons, S., & Machin, S. (2003). Valuing English Primary Schools. *Journal of Urban Economics*, *53*, 197–219.
Gibbons, S., & Machin, S. (2008). Valuing school quality, better transport, and lower crime: evidence from house prices. *Oxford review of economic policy*, *24*(1), 99–119.
Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(12), 7821.
Goldstein, H. (1987). *Multi-level models in educational and social research.* London: Charles Griffin.
Goodman, A. C. (1978). Hedonic price, price indices and housing markets. *Journal of Urban Economics*, *5*, 471–484.
Goodman, A. C. (1985). A note on neighbourhood size and the measurement of segregation indices. *Journal of Regional Science*, *25*(3), 1985.
Goodman, A. C., & Thibodeau, T. G. (1998). Housing market segmentation. *Journal of Housing Economics*, *7*(2), 121–143.
Goodman, A.C.,and Thibodeau, T.G. (2003) "Housing market segmentation and hedonic prediction accuracy" Journal of Housing Economics, Elsevier, Vol. 12(3), Pages 181–201, (September).
Guimer`a, R., Mossa, S., Turtschi, A., & Amaral, L. A. N. (2005). The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 7794–7799 (2005).
Hillier, B., & Hanson, J. (1984). *The social logic of space.* Cambridge: Cambridge University Press.
Hillier, B., & Iida, S. (2005). Network and psychological effects in urban movement. In A. G. Cohn, & D. M. Mark (Eds.), *Proceedings of Spatial Information Theory: International Conference, COSIT 2005, Ellicottsville, N.Y., U.S.A., September 14-18, 2005. Lecture Notes in Computer Science*, *Vol. 3693*. (pp. 475–490). Berlin, Germany: Springer-Verlag.
Kearns, A., & Parkinson, M. (2001). The significance of neighbourhood. *Urban Studies*, *2001*(38), 2103–2110.
Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Physical Review E*, *80*(5), 056117.
Law, S., Karimi, K., Penn, A., and Chiaradia, A. J. (2013) Measuring the influence of spatial configuration on the housing market in metropolitan London. In: Kim, Y.O., Park, H. T. and Seo, K. W. (eds.), Proceedings of the Ninth International Space Syntax Symposium, Seoul: Sejong University, Article 121.
Lebel, A., Pampalon, R., & Villeneuve, P. Y. (2007). A multi-perspective approach for defining neighbourhood units in the context of a study on health inequalities in the Quebec City region. *International Journal of Health Geographics*, *6*, 27. http://dx.doi.org/10.1186/1476-072X-6-27.
Leishman, C. (2009). Spatial change and the structure of urban housing sub-markets. *Housing Studies*, *24*(5), 563–585.
Lynch, K. (1960). *The image of the city.* the MIT Press.
Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*(026113), 2004.
Ordnance Survey (2014). *Ordnance survey open data meridian 2 dataset. Open source.* (Access on 01/08/2014 from the world wide web http://www.ordnancesurvey.co.uk/business-and-government/products/opendata-products.html. © Crown Copyright {2014}).
Ordnance Survey, & Survey, O. (2015). *Open data boundary line dataset. Open source.* (Access on 18/03/2015 from the world wide web http://www.ordnancesurvey.co.uk/business-and-government/products/opendata-products.html. © Crown Copyright {2015}).
Orford, S. (1999). *Valuing the built environment: GIS and house price analysis.* Aldershot: Ashgate.
Orford, S. (2001). Modelling spatial structures in local housing market dynamics: A multilevel perspective. *Urban Studies*, *37*(9), 1643–1671.
Pons, P., & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, *10-2*, 191–218.
Porta, S., Crucitti, P., & Latora, V. (2006). The network analysis of urban streets: A dual approach. *Physica A: Statistical Mechanics and its Applications*, *369*(2), 853–866.

Raghavan, U. N., Albert, R., & Kumara, S. (2007). *Physical Review E*, *76*(3), 036106.

Reichardt, J., & Bornholdt, S. (2004). *Physical Review Letters*, *93*(21), 218701.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55.

Shen, Y., & Karimi, K. (2016). Urban function connectivity: Characterisation of functional urban streets with social media check-in data. *Cities*, *55*, 9–21.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*(2), 234–240.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208.

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

Xiao, Y., Webster, C., & Orford, S. (2015). Urban configuration, accessibility, and property prices: A case study of Cardiff, Wales. *Environment and Planning B: Planning and Design*, *42*, 1–22 2015.

Yang, T., & Hillier, B. (2007). The fuzzy boundary: the spatial definition of urban areas. In A. S., Ö., & Y. I. (Eds.), *Proceedings, 6th international space syntax symposium. Istanbul, Turkey*. Cenkler, Istanbul: Istanbul Technical University 091.01-22.