

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

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. 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. The 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.

1.0 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 and 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 – 2011 (Law et al., 2013). The results confirm established relationship between geometric accessibility measures and property values (Desyllas, 1997; Chiaradia et al., 2012; Yang, et al., 2015). 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 et al., 2006; Yang and 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 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 home area, but smaller than the city (Kearns and Parkinson, 2001). It is related to the concept of neighbourhood in urban studies which encompass a more complex historical, socio-economic and perceptual constructs that overlap according to the geographical scale (Lebel et al., 2007; Galster 2001; Kearns and 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 and Thibodeau, 1998; 2002; Leishman, 2009; Orford 2000). 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 et al., 2001; Ellen and Turner 1997). Figure 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.

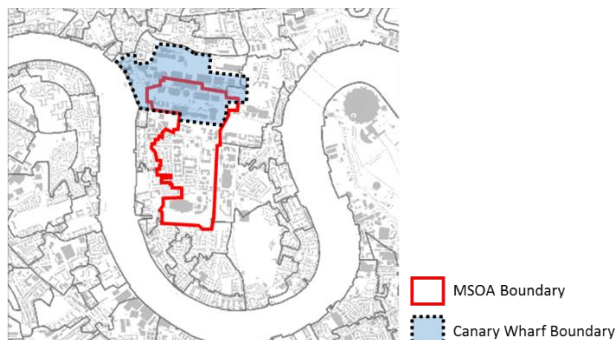


Figure 1 Canary Wharf Boundary in blue overlaid with MSOA boundary in red.

One problem of these 'arbitrary' or 'ad-hoc' (Orford, 1999; Goodman, 1977) administrative local area unit is that it creates inconsistent empirical results. Goodman's early studies (1978;1981) 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 also found segregation indices differed when applied through different levels of aggregation using the case study of Baltimore. Differences between census tracts and the smaller block level aggregation have been attributed to the "fuzziness" and "arbitrariness" of these local geographies. These problems are extended to housing submarket identification as noted by Leishman (2009). For example, Bourassa (Bourassa et al.,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).

2.0 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 and 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?

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

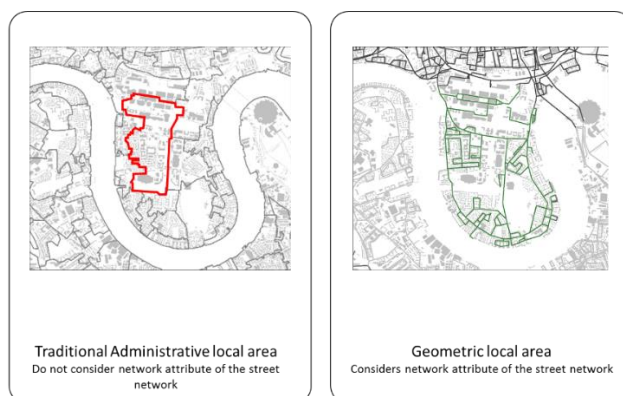


Figure 2 Traditional administrative local area on the left and Street-based Local Area (SLA) on the right. Ordnance Survey © Crown Copyright. All rights reserved.

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 and Newman, 2002; Guimer`a et al., 2005; Caschili et al., 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.

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 et al., 2006). The latter had been made popular from space syntax research (Hillier and Hanson, 1984).

$$DG(u, e)$$

u is the node (street segments)

e is the edge (junctions)

Equation (1)

This study will employ community detection technique on the spatial dual graph of the road centre line 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.

$$SG_k \subseteq DG \text{ where } k = 1, 2, \dots, K$$

SG is the subgraph

DG is the spatial dual graph

K is the number of subgraph

Equation (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 (Reichart and Bornholdt, 2004; Raghavan, et al., 2007; Newman and Girvan 2004; Pons and Latapy, 2006) This study in particular adopts the Multi-level 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 et al., 2008; Lancichinetti and Fortunato, 2009)¹. The algorithm optimises against a community quality function called Modularity. Modularity (Q) 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, k_i and k_j are the degrees for vertex i and vertex j. δ is 1 if i and j are in the same community and zero otherwise.

Q = Observed number of edges in a community – Expected number of edges in a community

$$Q = \frac{1}{2m} \sum (A_{ij} - K_i K_j / 2m) \delta(C_i, C_j)$$

A is the adjacency matrix

m is the total number of edges

K_i and K_j are the degree for the two subgraphs i,j

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

δ is a Kronecker Delta function which equals 1 when its argument are the same and 0 otherwise.

Equation (2) Modularity(Q) equation (Girvan and Newman, 2002)

Optimisation against the above function is currently impossible to solve for large datasets². As a result, a number of heuristic algorithms had been implemented into finding the optimal sub-graph (Girvan and Newman, 2002). This study will apply specifically the multi-level method (Blondel et al., 2008) in optimising against the modularity function as shown in Figure 3.

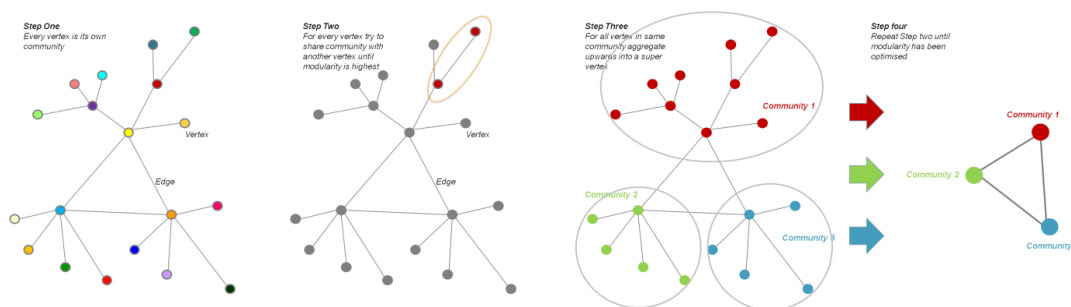


Figure 3 The Multi-level Modularity Optimisation method starts where every vertex is a community. Every vertex will then share community membership with its neighbour that attains the highest score. This continues for all vertices. Vertices within the same community will aggregate into a super vertex.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

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

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

3.0 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 (2001), who provided the evidence to use multi-level models in capturing the hierarchical effects 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 autocorrelation³ 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

³ Spatial autocorrelation refers to spatial association between proximate properties. Please refer to Anselin (1998) for a more informed discussion of the topic.

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 (μ) and fixed level predictor (X_i) and the random part into individual local area effect (u_i) and its error (ϵ_{ijk}).

$$Y = \beta X_i + \mu + \mu_i + \epsilon_{ijk}$$

$\underbrace{\hspace{1.5cm}}_{\text{Observed}} \quad \underbrace{\hspace{1.5cm}}_{\text{Fixed}} \quad \underbrace{\hspace{1.5cm}}_{\text{Random}}$

Y is the observed
B is the coefficient for predictors
X_i is the predictor
μ is the mean
u_i is the local area random effect
ε_{ijk} is the error term

Equation (3) multi-level regression model

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 figure 4.

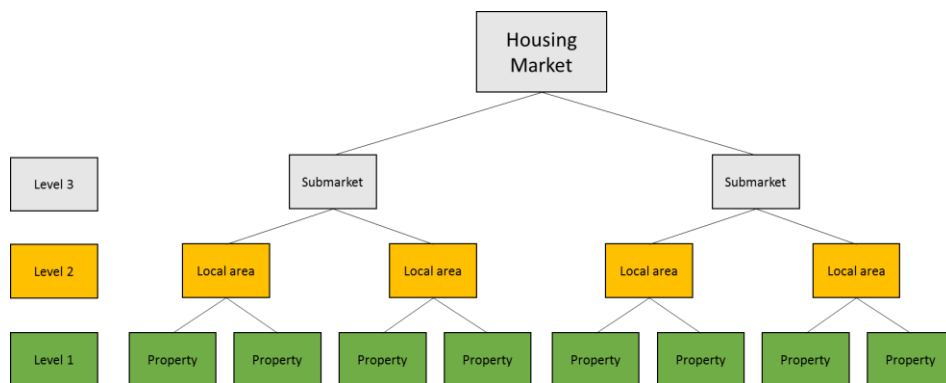


Figure 4 the hierarchical multi-level model

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

Table 1 Candidate models

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).

$$\text{Log}(HP_{ij}) = \mu + \varepsilon_{ij}$$

μ is the overall mean

ε_{ij} is the error

Equation (4) Base model 1

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, μ is the overall mean, u_j is the local area effect on house prices and ε_{ij} is the error term. No explanatory variable is specified for the model.

$$\text{Log}(HP_{ij}) = \mu + u_j + \varepsilon_{ij}$$

μ is the overall mean

u_j is the local area effects

ε_{ij} is the error term

Equation (5) Model 2 is the varying intercept model

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, μ is the overall mean, u_j is the local area effect on house prices and ε_{ij} is the error term. B_1 is the parameter estimated for space syntax integration and B_2 is the parameter estimated for floor area.

$$\text{Log}(HP_{ij}) = \mu + \beta 1_j * \text{Access}_i + \beta 2 * \text{Floor}_i + u_j + \varepsilon_{ij}$$

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

Equation (6) Model 3 varying intercept model with fixed predictors

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 ε_{ij} is the error term. B1 is the parameter for integration and B2 is the parameter estimated for floor area.

$$\text{Log}(HP_{ij}) = \mu + \beta 1_j * \text{Access}_i + \beta 2 * \text{Floor}_i + \alpha_j + \varepsilon_{ij}$$

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

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

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 vicinity⁴ and the quality of education as determined by average A-level score⁵. Model 5 (8) is

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

as follows, where HP_{ij} is the property price, μ is the overall mean, u_j is the local area effect on house prices, and ε_{ij} is the property level error term. β_n are the parameters estimated for the independent variables.

$$\text{Log}(HP_{ij}) = \mu + \beta_1 * \text{Access}_i + \beta_2 * \text{Floor}_i + \beta_3 * \text{Dwelling1}_i + \beta_4 * \text{Dwelling2}_i + \beta_5 * \text{Shop}_i + \beta_6 * \text{School}_i + u_j + \varepsilon_{ij}$$

μ is the overall mean

β is the coefficient for predictors

Access is the accessibility variable

Floor is the floor size variable

Dwelling1 represents flats

Dwelling 2 represents terraces

Shop represents the number of shops within 800m

School is the Average A-level score within 800m

u_j is the local area effects

ε_{ij} is the error term

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

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 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

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.

$$LR = -2 * [\ln(LL_{null})] + 2 * [\ln(LL_{multilevel})]$$

LL = loglikelihood

Equation (12)

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 AIC⁹ 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 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 as described in the chart below.

8

$$ICC = \frac{(Var_{Level 2})^2}{[(Var_{Level 2})^2 + (Var_{Level 1})^2]}$$

Var = variance

Equation (13)

9

$$AIC = -2 * LL + 2 * k$$

LL = loglikelihood

k = number of parameters

Equation (14)

	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

Table 2 Candidate models for all local areas

4.0 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 Figure 5, where the black line indicates the study boundary, the red line indicates the 33 administrative borough boundaries of Greater London (ONS, 2014), and the grey line indicates the meridian line street network.



Figure 5 Study area boundary

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. Figure 6 shows house prices in London for 2011 mapped in GIS, whereby red indicates a higher house

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.

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

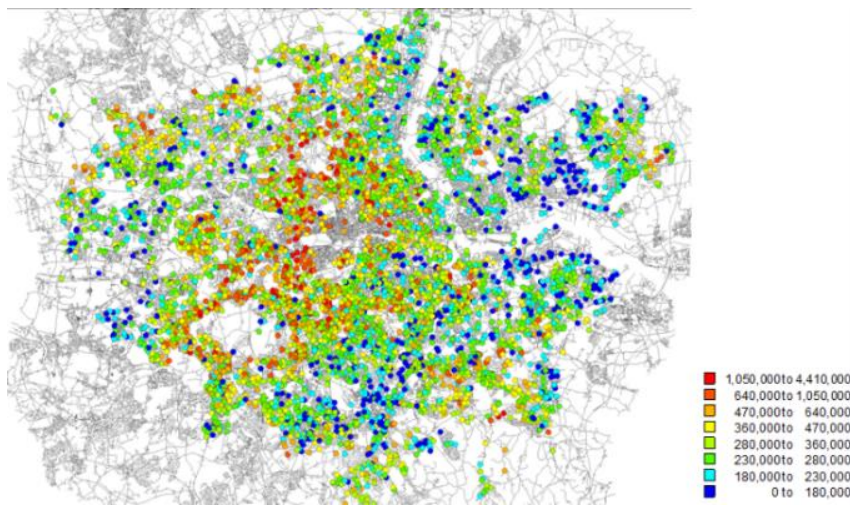


Figure 6 Visualisation of London House Price in 2011 from red indicating high house price to blue indicating low house price.

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 spatial network dataset has a total of 113,555 street segments as illustrated in Figure 7.

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



Figure 7 London OS Meridian 2 spatial street network

4.4 Descriptive Statistics

In the hedonic approach, structural features, accessibility levels and neighbourhood amenities were often included in the empirical model (Rosen, 1974; Cheshire and 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 et al., 2015)¹² and neighbourhood amenities such as the number of retail units within 800m (Des Rosier et al., 1996) or the secondary school average score within 800m (Black, 1999; Gibbons and Machin, 2003; 2008). The table below 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.

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). (Sabidussi, 2005 as mentioned in Iida and Hillier, 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., 2014; Law et al., 2013).

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	8,721	2,719
BC	Betweenness centrality	2.643e+07	1.087e+08
Shops	No. Shops within 800m	354.3	407.7
Park	Distance to royal parks	10,355	5,272
School	Average A-level score in 800m	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 3. descriptive statistics for house price and attributes

4.5 London Street-based Local Area

Applying the multi-level modularity algorithm described in 2.4 on the OS Meridian line network, a total of 207 spatial network local areas were identified for the Greater London area. Each street-based (SLA) local area has on average 549 segments, with a standard deviation of 257 segments. The chart below summarises the summary statistics.

Number of segments
113,555
Number of Local areas
207

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

Table 4. 207 Street-based Local Areas detected in Greater London

Figure 8 below shows the Street-based Local Areas (SLA) obtained from applying the multi-level modularity optimisation method on the London Meridian Line map. 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 is also a limitation as some areas might be considered fuzzy and some less. fuzzier and some areas less. This would be discussed in the discussion.

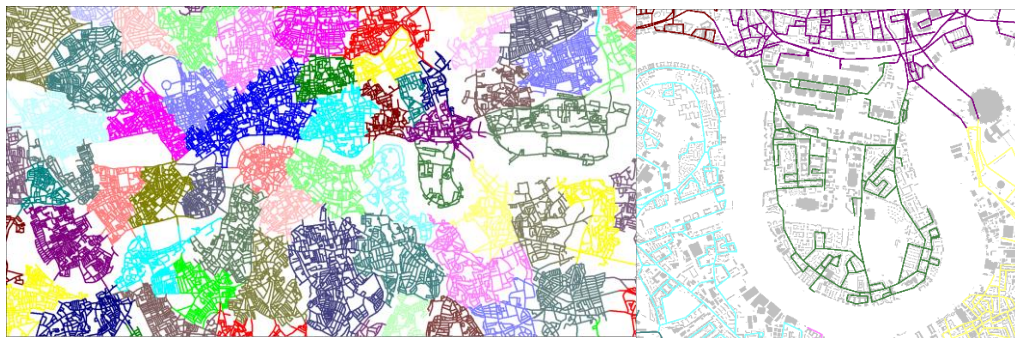


Figure 8 Visualisation of Street-based Local Area (SLA) membership for the Greater London Area.

4.6 London Administrative Local Area Units

Below is a table describing the three administrative local area units to be compared with the street-based local area in the following empirical study. The smallest are LSOA level followed by the MSOA level and Ward level¹³.

	N
LSOA	4765
MSOA	983
Ward	629

Table 5. Local area statistic

¹³ Electoral wards/divisions is the key local area unit for UK administrative geography. There are a total of 9,456 wards in the UK with an average population of 5,500 person per ward. (ONS 2015)

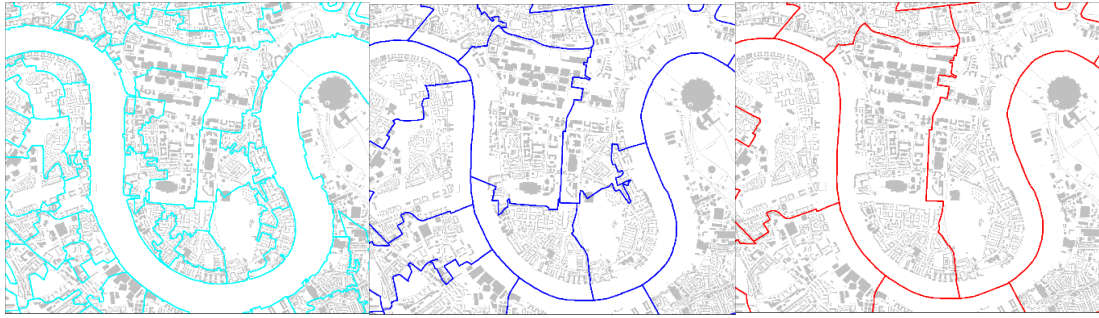


Figure 9 Isle of Dogs area as denoted by LSOA, MSOA and Ward boundaries

Figure 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.

5.0 Empirical Results

The following section illustrates the empirical results for this article. We will first test the significance for house prices using the multi-level hedonic approach as specified in section 3.1. This will be followed by a comparison between the SLA housing submarket as specified in section 3.2 and a comparison with other administrative units as specified in section 3.3.

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 variations as specified in section 3.1. The table below illustrates the regression results for the five candidate models.

SLA	Model 1	Model 2	Model 3	Model 4	Model 5
Lnprice					
Integration			0.046	0.058	0.034
Floor size			0.008	0.013	0.012
Age			0.342	0.340	0.277
Park			0.003	0.003	0.003
Shops					0.030
Terrace					0.003
Flat					0.127
School					0.013
_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
<i>N</i>	5334	5334	5334	5334	5334
LR test (Prob>0.05)	0***	0***	0***	0***	0***

Table 6. Multi-level hedonic regression results

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.

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 ($\text{Prob} > \chi^2 = 0^{***}$) shows significance for all the candidate models. This shows robust evidence that street-based local area effect is significant.

The chart below summarises the goodness of fit as measured by AIC between the five candidate SLA multi-level 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.

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

Table 7. AIC comparison

Figure 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.

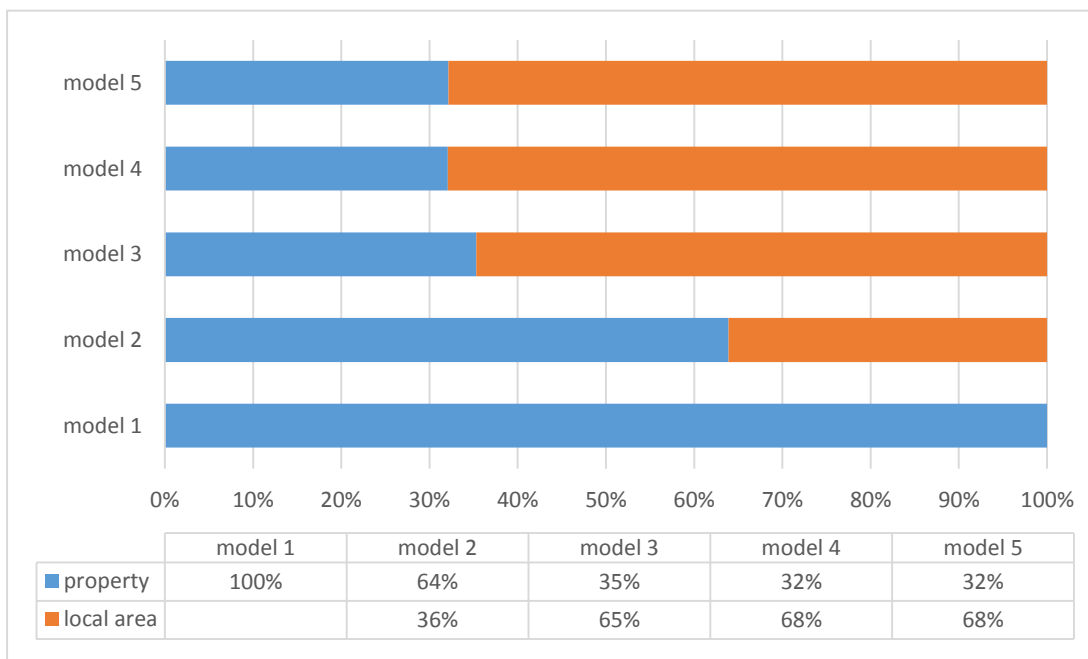


Figure 10. Intra-class correlation coefficient comparison.

For empirical reasons, spatial autocorrelation effects were checked. Moran's I were calculated with a minimum radius of 2400m 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. Figure 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.

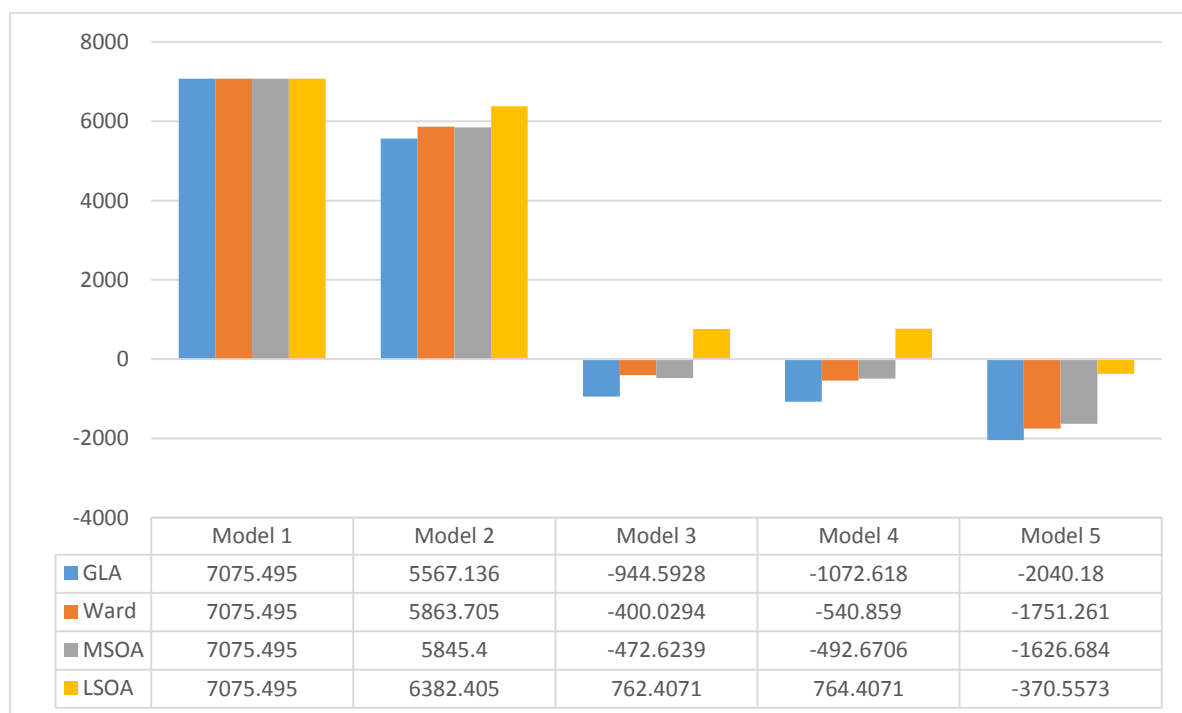


Figure 11. Goodness of fit comparison (AIC) across different local area units

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) and Goodman’s research (1998) 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 (Goodman, 1978). 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.

6.0 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 figure 12, which illustrates two distinct local areas connected via a bridge (one in orange and one in green).

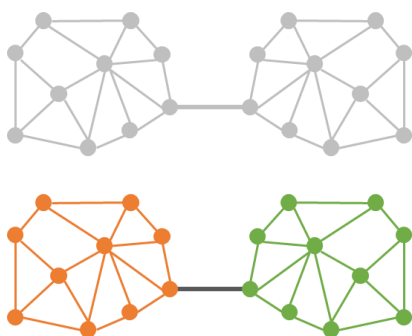


Figure 12 The top image illustrates a graph. The bottom image illustrates an obvious division between two subgraphs connected via a bridge. The subgraph to the left is coloured orange and the subgraph to the right is coloured green and the bridge is coloured grey.

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 (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 figure 12 where we run a conceptual simulations of an agent walking in this network in figure 13.¹⁴

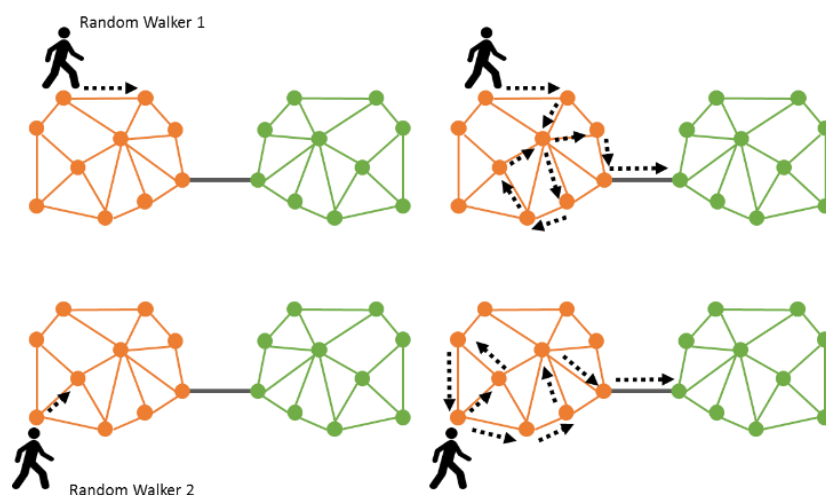


Figure 13 This figure illustrates two simulations of an agent starting from a different orange node who walks randomly around the graph. The first simulation shows the walker took 9 steps to reach the bridge. The second simulation shows the walker took 8 steps to reach the bridge. This illustrates that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

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

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 herding behaviour and bounded rationality where information is constrained within the local area (Benerjee, 1992; Simon, 1957). 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.

6.2 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 figure 13 multiple times. Figure 14 illustrates the average number of random steps an agent at an orange subgraph would take to reach the green subgraph over 500 times and four different configurations.

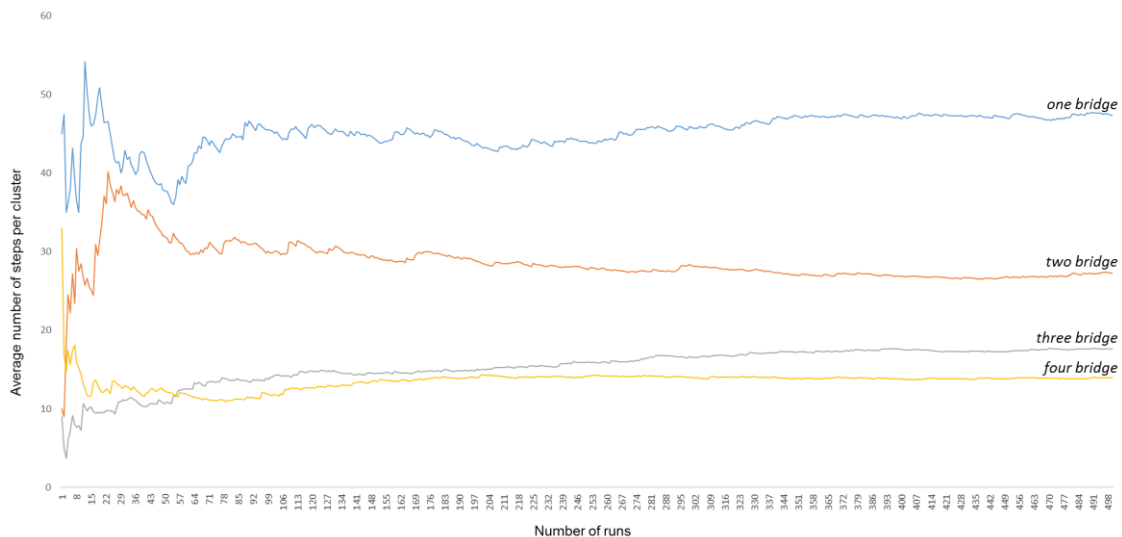


Figure 14. Average number of steps required to jump between clusters

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 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 the network approach can identify subtle differences in an urban environment but also how local area boundary and socio-economic spillover could be influenced by these subtle differences of the grid.

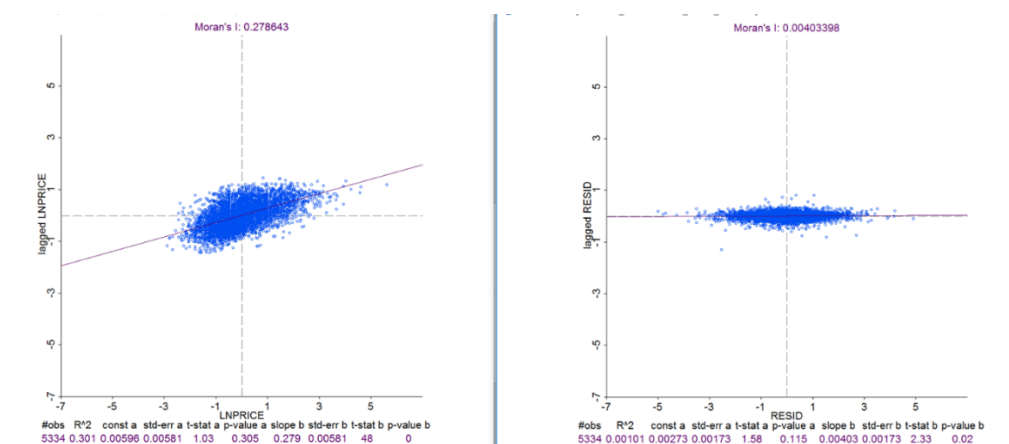
The definition of SLA is not without its concerns. First, this research suggests on aggregate, SLA is able to capture subtle differences of an urban environment more accurately than region based methods. However, at an individual-level more research is required to understand and confirm how this happened 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 multi-level modularity optimisation method defines sharp local area boundaries which contradicts to previous research in describing neighbourhoods as fuzzy and overlapping (Alexander 1964). 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 housing submarket formation.

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 for neighbourhood planning and policies.

Appendix A



Global Moran's I
Left SLA Model 1
Right SLA Model 5

References

- 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., and Lefebvre. E. (2008)** Fast unfolding of communities in large networks. *J. Stat. Mech.*, 2008:P10008, 2008.
- Bourassa, S. C., Hamelink, F., Hoesli, M., and MacGregor, B. D. (1999)** "Defining Housing Submarkets," *J. Housing Econ.* 8, 160-183.
- Caschili, S., De Montis, A., Chessa, A., and Deplano. G. (2009)** Weighted networks and community detection: planning productive districts in Sardinia. In: G. Rabino and M. Caglioni, editors, *Planning, complexity and new ICT*, pages 27 – 36. Alinea Editrice s.r.l., 2009.
- Cheshire, P., and Sheppard, S. (1998)** Estimating the demand for housing, land, and neighbourhood characteristics. *Oxford Bulletin of Economics and Statistics* 60: 357–382.
- Coulton, C. J., Korbin, J., Chan, T. and 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. In: *Space Syntax and Spatial Cognition Workshop of the Spatial Cognition*.
- Fortunato, S. (2010)** "Community detection in graphs". *Phys. Rep.* 486 (3-5): 75–174. doi:10.1016/j.physrep.2009.11.002.
- Galster, G. (2001)** On the Nature of Neighbourhood. *Urban studies* 12:2111-2124. Publisher Full Text OpenURL.
- Girvan, M., and 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), p.7821.
- Goldstein, H. (1987)** *Multi-level Models in Educational and Social Research*. London: Charles Griffin.
- Goodman, A.C., and Thibodeau, T.G. (2007)** "The Spatial Proximity of Metropolitan Area Housing Submarkets," *Real Estate Economics*, American Real Estate and Urban Economics Association, vol. 35(2), pages 209-232, 06.
- 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*, Vol.25, No.3, 1985.

- Goodman, A.C., and 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.
- Greater London Authority. (2011)** <http://data.london.gov.uk/dataset/ward-profiles-and-atlas> UK Open Government Licence (OGL).
- Grisby, W., Baratz, M., Galster, G., and Maclennan, N.D. (1987)** *The Dynamics of Neighbourhood Change and Decline*. Oxford : Pergamon.
- Guimer`a, R., Mossa, S., Turtschi, A., and Amaral. L.A.N. (2005)** The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci. USA*, 102:7794–7799, 2005.
- Helbich, et al. (2013)** Data-Driven Regionalization of Housing Markets, *Annals of the Association of American Geographers* 103, 871-889.
- Hillier, B., and Hanson, J. (1984)** *The Social Logic of Space*, Cambridge University Press: Cambridge.
- Hillier, B., Burdett, R., Peponis, J., and Penn, A. (1987)** “Creating Life: Or, Does Architecture Determine Anything?”, *Architecture & Comportment/ Architecture & Behaviour*, 3 (3). pp. 233-250.
- Jenks, M. and Dempsey, N. (2007)** ‘Defining the Neighbourhood: challenges for empirical research’ *Town Planning Review*, 78(2), pp.153-177.
- Jones, K. (1991)** *Multi-level Models for Geographical Research*. Norwich: Geo Books.
- Jones, K., and Bullen, N. (1993)** A multilevel analysis of the variations in domestic property prices: southern England 1980–1987, *Urban Studies*, 30, pp. 1409–1426.
- Kearns, A., and Parkinson, M. (2001)** The significance of neighbourhood. *Urban studies* 2001, 38:2103-2110.
- Lancichinetti, A., and Fortunato, S. (2009)** Community detection algorithms: a comparative analysis, *Physical Review E* 80 (5), 056117.
- Lancichinetti, A., and Fortunato, S. (2011)** Limits of modularity maximization in community detection, *Physical Review E* 84 (6), 066122.
- 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., and 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* 2007, 6:27 doi:10.1186/1476-072X-6-27.

Leishman, C. (2009) Spatial Change and the structure of urban housing sub-markets. *Housing Studies*. Vol.24. Issue 5. Pp 563-585.

Lynch, K. 1960. *The image of the city*, the MIT Press.

Maclennan, D., and Tu, Y. (1996) Economic perspectives on the structure of local housing systems, *Housing Studies*, Vol.11, Issue 3, pp 387-406.

Newman, M.E.J., and Girvan, M. (2004) Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113, 2004.

ONS (2015) Office for National Statistics website. Access on 18/03/2015 from the world wide web <http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/administrative/england/electoral-wards-divisions/index.html>

Openshaw, S. (1983) *The modifiable areal unit problem*. Norwick: Geo Books. ISBN 0860941345. OCLC 12052482.

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. (2015) Ordnance Survey 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), pp. 1643–1671.

Peponis, J. (1988) Social structure of space and relations of community in six Greek cities, Secretariat of Research and Technology, Athens, Ministry of Energy Industry, Research and technology.

Pons, P., and Latapy, M. (2006) Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications* Vol.10-2. P.191-218.

Porta, S., Crucitti, P., and 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., and Kumara, S. (2007) *Phys. Rev. E* 76(3), 036106.

Read, S. (1999) Space syntax and the Dutch city. *Environment and Planning B*, 26, pp.251–264.

Reichardt, J., and Bornholdt, S. (2004) *Phys. Rev. Lett.* 93(21), 218701.

- Rosen, S. (1974)** Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34-55.7.
- Rossum, G. (2007)** Python programming language. In: USENIX Annual Technical Conference.
- 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., and Orford, S. (2015)** Urban Configuration, Accessibility, and Property Prices: A Case study of Cardiff, Wales. *Environment and Planning B: Planning and Design*. 2015. Vol.42, Pages 1-22.
- Yang, T., and Hillier, B. (2007)** The fuzzy boundary: the spatial definition of urban areas. In: *Proceedings, 6th international space syntax symposium*. Istanbul, Turkey, ed. A.S.Kubat, Ö Ertekin, and Y.Io. Güney, 091.01-22. Cenkler, Istanbul: Istanbul Technical University.