Using house sales transactions data to identify potentially gentrifying neighbourhoods

Comber A*[1], Usher M[1], Wang Y[1], Minh Le Kieu[2], Bui Thanh Quang[3], Hang Nguyen Thi Thuy[4], Phe Hoang Huu[5] and Maleeson N[1]

[1]School of Geography, University of Leeds, UK
[2]Department of Civil and Environmental Engineering, University of Auckland, New Zealand
[3]Faculty of Geography, VNU University of Science, Hanoi, Vietnam
[4]VNU Vietnam Japan University, Hanoi, Vietnam
[5]R&D Consultants, Hanoi, Vietnam

GISRUK 2025

Summary

This paper describes initial seeking to identify locales of neighbourhood gentrification. It summarises house sales transactions over MSOAs, compares average annual house price with neighbouring areas over a 10 year period, and uses a time series analysis to identify neighbourhoods with high rates of change of these as potential gentrifying neighbourhoods. The next steps are to link to analyses of social media data to characterise the nature of the observed neighbourhood change. A number of areas of further work are described and a number of critical considerations are discussed.

KEYWORDS: Hedonic house price models, Social media data, Time series analysis.

1 Introduction

There is now long standing research interest in and use of using social media data (SMD) as a means to capture information about people's preferences and behaviours, including within the spatial and geographic sciences (Hess et al., 2017; Matthews, 2015; Johnston and Pattie, 2011; Redi et al., 2018). This is because a small proportion of social media posts and content have some form of location attached either through the use of geo-tags, which directly indicate user or longitude and latitude at the time of posting, or through location inferred from social media content, for example through mention of placenemes or landmarks, which themselves can be located.

Some research has sought to leverage social media content to to answer specific questions and to understand particular spatial processes including. However, SMD is commonly been subject to some form of sentiment analysis in order to indicate place and neighbourhood related perceptions.

*a.comber@leeds.ac.uk

The recently funded INTEGRATE project (`https://urban-analytics.github.io/INTEGRATE/intro.html`) extends the analysis of SMD in order to capture neighbourhood dynamics and how they change over time, in order to identify locales of gentrification. It does this by linking analyses of to traditional hedonic house price (HHP) data and models. The project seeks to explore the use of SMD and its ability to capture dynamic place-based perceptions and thus intangible elements of neighbourhood attractiveness (Huu Phe and Wakely, 2000), within the construct of Gentrification.

This paper describes work analysing traditional housing sales data, neighbourhood level measures of housing affordability and socio-economic deprivation, and Twitter data in order to identify locales of potential gentrification in the UK. In so doing it identifies and starts to address a number of methodological and philosophical challenges: 1) Can signals of neighbourhood change be derived from analysis of SMD and HHP data? 2) Over what timescales do they operate and to what degree do they lag or coincide? 3) How separable is Gentrification from other more general neighbourhood changes? 4) Is the concept of Gentrification still relevant in the UK?

2 Background

HHP models typically seek to link house price to property characteristics (such as age, number of bedrooms, floor area, etc) (Rosen, 1974; Holmes et al., 2017; Follain and Jimenez, 1985), location utility (such as access to amenities, distance to transportation, workplace) (Osland, 2010; Poudyal et al., 2009) and to location socio-economic characteristics (such as crime, ethnic composition, environmental quality, etc) (Lynch and Rasmussen, 2001; Hui et al., 2007).

Gentrification is a significant driver of house price dynamics. It is widely studied but contentious (Lees et al., 2010) and takes different forms in different contexts. Examples include super-gentrification (Lees, 2003), green gentrification (Gould and Lewis, 2016), rural gentrification (Phillips et al., 2021), phenomenon-specific gentrification types such as student driven (studentification) (Smith, 2005) and development driven (new-build gentrification) (Davidson, 2018). Across these different manifestations, gentrification is typically associated with the displacement of residents by people from a higher social class (Lees et al., 2010). The displaced populations are often working-class (Paton, 2016) and / or ethnic minorities (Huse, 2018). Gentrification has been closely associated with enhanced SMD use (Gibbons et al., 2018; Bronsvoort and Uitermark, 2022).

The aim of this research is to identify neighbourhoods that have gentrified by integrating HHP data and SMD. In this gentrification is conceptualised as resulting in, among other things, an increase in house price relative to other neighbourhoods in the locale, an influx of higher socio-economic status people displacing others of lower socio-economic status. Increases in house price, relative to those in the nearby areas, are not on their own sufficient to indicate gentrification processes. However, analysis of SMD offers opportunities to capture neighbourhood socio-economic characteristics [(Lansley and Longley, 2016) and dynamics (Poorthuis et al., 2022; Nguyen et al., 2016). SMD has proved useful in revealing emergent gentrification (Bronsvoort and Uitermark, 2022; Gibbons et al., 2018; Walters and Smith, 2024) as new residents seek to reinforce and solidify the new neighbourhood identity status (Bronsvoort and Uitermark, 2022), particularly among gentrifiers

(Friesenecker and Lagendijk, 2021).

3 Methods

One of the key methodological questions was how to approach the analysis: from the SMD or the HHP data? Analysis of SMD could reveal gentrifying sentiment. Initial work analysed the spatio-temporal distributions of sentiment and gentrification sentiment in Twitter data using a basic gentrification lexicon (Comber et al. 2024). This has been extended to explore different approaches for analysing Tweet content including classification, natural language processing and more recently large language models (LLMs) applied to a variety of SMD. Malleson et al. at this conference (*Using Large Language Models to Predict Neighbourhood Change*) used LLM analysis of SMD to identify potentially gentrifying neighbourhoods. However, in both cases results were not sufficiently spatially resolved (specific), with many, for example, city centre areas indicated that are just experiencing uplift and development. An alternative approach is to use the HHP data to identify neighbourhood where house prices have increased over time relative to those in their local area. This paper describes work undertaking this. The approach and data used was as follows:

1. WhenFresh/Zoopla property sales data was obtained from the CDRC (`https://data.cdrc.ac.uk`). This describes nearly 9 million individual transactions for 2014 to 2023.
2. Property transactions were located via their postcode using look-up tables from ONS (`https://geoportal.statistics.gov.uk/datasets/9ac0331178b0435e839f62f41cc61c16/about`).
3. Annual summaries of the median house price for each Medium Super Output Area (MSOA) in England and Wales were calculated. The MSOAs form the neighbourhoods in this analysis and contain around 7,500 people (2,500 households).
4. For each MSOA, a measure of local median house price was generated using a k-nearest neighbours approach, with $k = 50$. An example of this is shown in Figure 1.
5. From this, a ratio of neighbourhood house price to nearby house prices was constructed for each MSO for each year.
6. The ratio used to construct a regression of house price ratio against time for each MSOA, to indicate the rate of change over time (β_{time}).
7. The (β_{time}) values were used to identify MSOA neighbourhoods experiencing much greater increases in house prices, relative to other neighbourhoods in their local area.

4 Results

The mapped results are for England and Wales are shown in Figure 2, along with some local detail for the the MSOAs in the Leeds and Bradford area. This shows the spatial distribution of the coefficients of the ratio of MSOA house price to local nearby house prices over a 10 year period (2014 to 2023). There are some evident *high* (positive) and *low* (negative) areas indicating MSOAs that exhibit very different house price trends over time to those of nearby MSOAs.

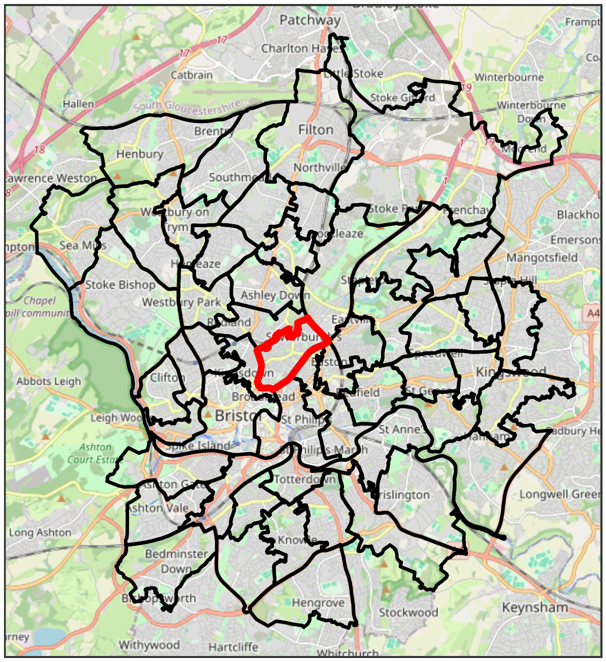

Figure 1: An example of an MSOA and its 50 neighbouring areas.

The MSOAs with high slopes are potential locales of gentrification, those with negative local coefficients of house price change over time potentially indicate areas in decline. The distribution of the coefficient estimates is shown in Figure 3 and indicates a normal distribution, with then the potential use standard deviations to determine potentially gentrifying MSOAs. Finally, MSOAs with high rates of house price change over time (greater than 2 standard deviations) relative to their neighbouring areas, are highlighted in Figure 4.

The next steps will be to investigate the content of SMD in these areas, and the degree to which gentrifying sentiment can be identified, through text analysis and LLMs.

5 Discussion

It is possible to characterise neighbourhood change through analyses of time series data (Reades et al., 2019, 2023; Gray et al., 2021, 2023). This paper describes exploratory work investigating changes in house price as a first step in characterising neighbourhood gentrification. It summarised house sale transactions over MSOAs, generating a median annual price for a 10 year period, 2014 to 2023. For each MOSOA, and for each year these were compared with nearby values ($n = 50$) to generate a ratio of MSOA price to nearby MSOA price. In turn, for each MSOA a simple OLS regression was used to estimate a local coefficient for time as a predictor of the ratio. The coefficient estimates β_{time} greater than 2 standard deviations from the mean coefficient estimate identified MSOAs that were candidates for further investigation (Figure 4). This conceptualisation thus considers large increases in house price relative to house prices in nearby, as a potential indicator of

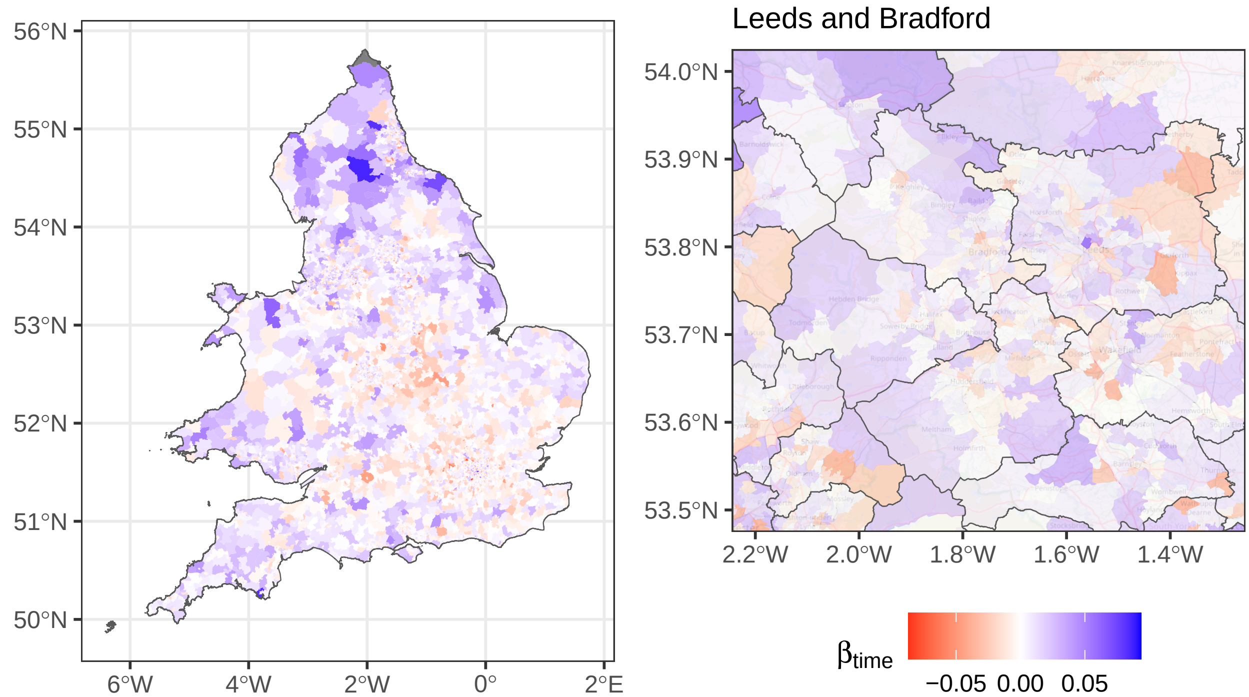

Figure 2: The trajectories (2014-2023) of the ratio of MSOA house price to prices in nearby areas for England and Wales, with some local detail.

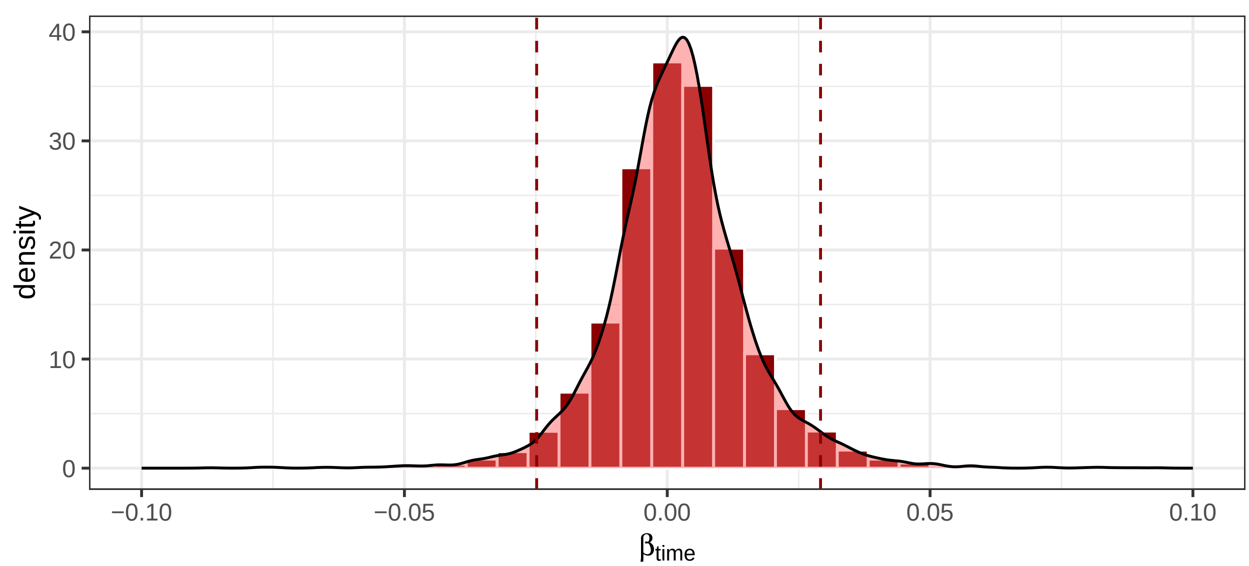

Figure 3: A density histogram of the distribution of the coefficients of the ratio of MSOA house price to local nearby house prices over a 10 year period (2014 to 2023), with 2 standard deviations from the mean indicated.

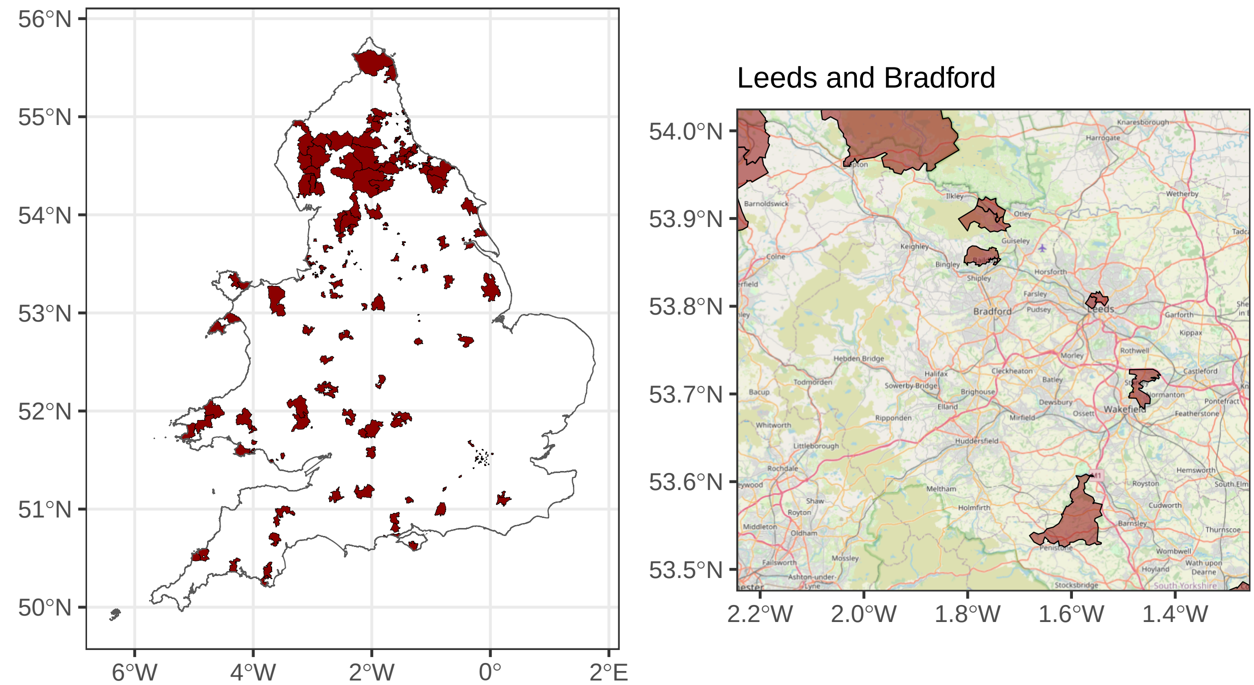

Figure 4: The MSOAs with high rates of house price increase (2014-2023) relative to nearby areas, with some local detail.

gentrification. The next steps in this investigation are to link analysis of social media data (Twitter), and to determine the degree to which this can provide confirmatory evidence of gentrification.

However, a number of important assumptions were made that will also be investigated in future work. First, the analysis uses MSOAs as neighbourhoods. These contain around 7,500 people, are perhaps too spatially coarse in that they may not contain homogeneous trends and behaviours. Exploratory work undertaking the same analysis with LSOAs (1,500 people) in many cases resulted in too few annual house sale transactions. However, there may be advantages to analysing SMD over LSOAs *within* MSOAs indicated as potentially gentrifying areas, to overcome such granularity issues. Second, median house price for each MSOA was calculated. Different areas often have very different types of housing stock, and perhaps *median price per bedroom* would provide a more representative and sensitive measure. Third, the 10 years of MSOA median house price were compared to those for the 50 nearest MSOAs. This number was somewhat arbitrarily chosen, but does represent the population of a typical city, within which people may chose to relocate. The sensitivity of the results to this definition of *nearby* will be investigated, along with other definitions of nearby based on distance and adjacency (1st, 2nd... nth order, etc).

A final, broader set of philosophical considerations have also emerged during this work. First, practically, earlier work as (Comber et al., 2024) sought to identify potentially gentrifying areas using a simple text analysis of SMD over time and LSOAs. However, the signal of any potential gentrification was lost in the noise in the morass non-gentrifying sentiment. So the question is where to start: from the identification gentrifying sentiment in SMD? Or from (relative) house price changes? Parallel work (see Malleson et al. at this conference) is exploring and extending previous text analysis methods - from gentrification sentiment lexicons to the use of large language models - to examine the extent to which SMD analysis can highlight potentially gentrifying areas. Second, conceptually, is Gentrification still a relevant concept in the UK? An increasing number of people simply do not have an any choice about *where* they buy house: increasingly they can either afford to buy a house or they cannot, and if they can it may only be one neighbourhood - the cheapest. This is definitely the case now in London, where the concept of gentrification is largely irrelevant - all of London has gentrified.

6 Acknowledgements

This research is supported by UKRI (ESRC) funding ES/Y006259/1 under the Digital Footprints scheme. The data for this research have been provided by the Consumer Data Research Centre, an ESRC Data Investment, under project ID CDRC, ES/L011840/1; ES/L011891/1.

References

Bronsvoort, I. & Uitermark, J. L. (2022). Seeing the street through instagram. digital platforms and the amplification of gentrification. *Urban Studies*, *59*(14), 2857–2874.

Comber, A., Kieu, M., Bui, Q.-T., & Malleson, N. (2024). Using social media data to identify neighbourhood change. *AGILE: GIScience Series*, *5*, 20.

Davidson, M. (2018). New-build gentrification. In *Handbook of gentrification studies* (pp. 247–261). Edward Elgar Publishing.

Follain, J. R. & Jimenez, E. (1985). Estimating the demand for housing characteristics: a survey and critique. *Regional science and urban economics*, *15*(1), 77–107.

Friesenecker, M. & Lagendijk, A. (2021). Commercial gentrification in arnhem and vienna: The local enactment of, and resistance towards ‘globally’circulating creative neighbourhood policies. *City*, *25*(5-6), 698–719.

Gibbons, J., Nara, A., & Appleyard, B. (2018). Exploring the imprint of social media networks on neighborhood community through the lens of gentrification. *Environment and Planning B: Urban Analytics and City Science*, *45*(3), 470–488.

Gould, K. & Lewis, T. (2016). *Green gentrification: Urban sustainability and the struggle for environmental justice*. Routledge.

Gray, J., Buckner, L., & Comber, A. (2021). Extending geodemographics using data primitives: A review and a methodological proposal. *ISPRS International Journal of Geo-Information*, *10*(6), 386.

Gray, J., Buckner, L., & Comber, A. (2023). Identifying neighbourhood change using a data primitive approach: the example of gentrification. *Applied Spatial Analysis and Policy*, *16*(2), 897–921.

Hess, D. B., Iacobucci, E., & Väiko, A. (2017). Network connections and neighbourhood perception: using social media postings to capture attitudes among twitter users in estonia. *Architecture and Urban Planning*, *13*(1), 67–78.

Holmes, T. P., Adamowicz, W. L., Carlsson, F., Champ, P. A., Boyle, K. J., & Brown, T. C. (2017). A primer on nonmarket valuation: The economics of nonmarket goods and resources.

Hui, E. C., Chau, C. K., Pun, L., & Law, M. (2007). Measuring the neighboring and environmental effects on residential property value: Using spatial weighting matrix. *Building and environment*, *42*(6), 2333–2343.

Huse, T. (2018). Gentrification and ethnicity. In *Handbook of gentrification studies* (pp. 186–204). Edward Elgar Publishing.

Huu Phe, H. & Wakely, P. (2000). Status, quality and the other trade-off: Towards a new theory of urban residential location. *Urban studies*, *37*(1), 7–35.

Johnston, R. & Pattie, C. (2011). Social networks, geography and neighbourhood effects. *The SAGE handbook of social network analysis*, (pp. 301–311).

Lansley, G. & Longley, P. A. (2016). The geography of twitter topics in london. *Computers, Environment and Urban Systems*, *58*, 85–96.

Lees, L. (2003). Super-gentrification: The case of brooklyn heights, new york city. *Urban studies*, *40*(12), 2487–2509.

Lees, L., Slater, T., & Wyly, E. K. (2010). *The gentrification reader*, volume 1. Routledge London.

Lynch, A. K. & Rasmussen, D. W. (2001). Measuring the impact of crime on house prices. *Applied Economics*, *33*(15), 1981–1989.

Matthews, P. (2015). Neighbourhood belonging, social class and social media—providing ladders to the cloud. *Housing Studies*, *30*(1), 22–39.

Nguyen, Q. C., et al. (2016). Leveraging geotagged twitter data to examine neighborhood happiness, diet, and physical activity. *Applied Geography*, *73*, 77–88.

Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research*, *32*(3), 289–320.

Paton, K. (2016). *Gentrification: A working-class perspective.* Routledge.

Phillips, M., Smith, D., Brooking, H., & Duer, M. (2021). Re-placing displacement in gentrification studies: Temporality and multi-dimensionality in rural gentrification displacement. *Geoforum*, *118*, 66–82.

Poorthuis, A., Shelton, T., & Zook, M. (2022). Changing neighborhoods, shifting connections: mapping relational geographies of gentrification using social media data. *Urban Geography*, *43*(7), 960–983.

Poudyal, N. C., Hodges, D. G., & Merrett, C. D. (2009). A hedonic analysis of the demand for and benefits of urban recreation parks. *Land Use Policy*, *26*(4), 975–983.

Reades, J., De Souza, J., & Hubbard, P. (2019). Understanding urban gentrification through machine learning. *Urban Studies*, *56*(5), 922–942.

Reades, J., Lees, L., Hubbard, P., & Lansley, G. (2023). Quantifying state-led gentrification in london: Using linked consumer and administrative records to trace displacement from council estates. *Environment and Planning A: Economy and Space*, *55*(4), 810–827.

Redi, M., Aiello, L. M., Schifanella, R., & Quercia, D. (2018). The spirit of the city: Using social media to capture neighborhood ambiance. *Proceedings of the ACM on human-computer interaction*, *2*(CSCW), 1–18.

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

Smith, D. (2005). Patterns and processes of'studentification'in leeds. *The Regional Review*, *12*, 14–16.

Walters, P. & Smith, N. (2024). It's so ridiculously soulless: Geolocative media, place and third wave gentrification. *Space and Culture*, *27*(1), 94–109.