Understanding the Impact of Retail Composition on Residential House Prices in London: A Graph Attention Network Approach

Jihun Jung
University College London, Green Street

Background and Motivation

Retail environments shape how residents value and experience their neighbourhoods. This study applies Graph Neural Networks (GNNs) to examine how surrounding retail composition influences residential property prices across Greater London between 2015 and 2024.

Data and Methods

GNNs are used to model the relationship between residential property prices and surrounding retail environments. A cleaned dataset of ~1.1 million property transactions across Greater London is linked to floor area data from Energy Performance Certificates (EPCs) and enriched with retail POI data from Green Street's *Retail Analytics Pro* platform. Retail data is reconstructed annually, focusing on everyday categories with labelled brand tiers.

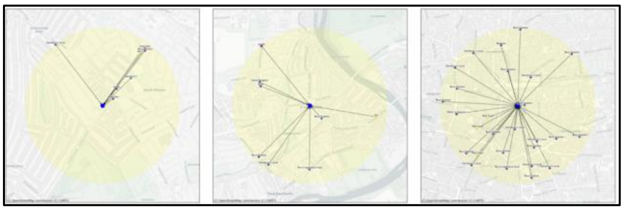
Bipartite graphs are constructed for each year linking homes to retail locations within 1 km, with edges weighted by an exponential distance-decay function. Graphs are trained using a Graph Attention Network (GAT) to predict price per square metre for new-build homes.

Key Findings

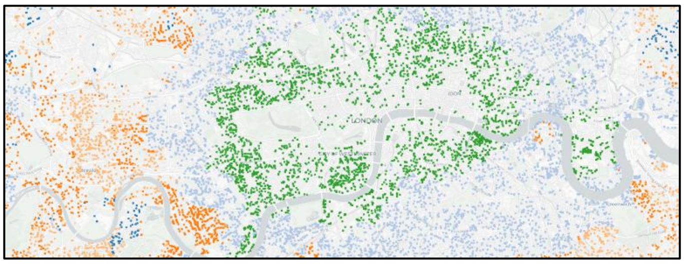
Metric	Log-Scale	Original Scale (Price per m²)
RMSE	0.2382	2,037.81
MAE	0.1770	1,422.97
R^2	0.5883	0.5850

GAT achieves strong predictive performance on new-build transactions, with MAE of £1,423/m² and R^2 of 0.585. Predicted prices align closely with observed values, particularly around the modal range (£6,000–£8,000/m²). High-value homes are slightly under-predicted due to unmodelled structural factors. These results show that the surrounding retail context, captured through graph structure and attention mechanisms, can systematically explain variation in residential property values.

Attention analysis reveals that supermarkets receive the strongest model focus throughout, underscoring their consistent importance in price prediction. Premium chains like Waitrose exert an especially strong influence on new-build, low-density housing, with the highest attention scores observed for detached properties. In contrast, flats show more diffuse attention, due to greater baseline access to nearby retail.



Graph visualisations demonstrate how the model prioritises retail connections. In low-value areas, attention concentrates on value anchors like Aldi, even when they are not the nearest. In mid-value areas, Lidl receives the highest weight, outperforming closer but less distinctive options. In high-value areas, attention is more diffuse but consistently favours premium chains like M&S Food. These examples show that the model captures not just proximity, but the differentiated importance of retail types across the spectrum.



Clustering of GNN embeddings identifies five distinct residential–retail ecologies across London, revealing an affordability–amenity gradient. Lower-value clusters are dominated by convenience/value retail, while mid-range zones mix value and premium brands. The highest-value areas show a concentration of premium chains like M&S and Waitrose. These patterns emerge without encoding explicit socioeconomic variables, showing how the model captures neighbourhood structure.

Value of the Research

This research demonstrates the value of GNNs in providing interpretable links between retail amenities and residential property values. For developers and local authorities, the findings offer insights into how different retail types and brand tiers shape neighbourhood desirability, supporting more informed approaches to urban planning, regeneration, and retail zoning. The model also uncovers affordability-amenity gradients without explicit socioeconomic data, revealing retail access disparities as markers of neighbourhood structure. Finally, the framework is scalable and transferable, with applications in other cities and contexts.