CIPHER: Interpretable Price Inference by Fusing Mobility Networks and ZIP-Level Socioeconomics

Shenghuan Wang

College of Letters and Science, University of California Davis, 95616 Davis CA, USA

dvswang@ucdavis.edu

Abstract. Urban housing prices arise out of the nuanced interplay of neighbourhood socioeconomic, land-use, and mobility. This paper presents the Construction–Integration–Prediction–Hypothesis–Explanation–Reproducibility (CIPHER) framework—an interpretable cross-city model combining different types of mobility graphs (parking, metro, and bike) and ZIP-level features within ZIP Code Tabulation Areas (ZCTAs) including parks, income, education, and schools, through a Heterogeneous Graph Transformer (HGT). Here, CIPHER denotes: Construction of a type-aware urban graph; Integration of ZIP context at the node level for early fusion; Prediction with a type-specific multi-task head centered on median price; Hypothesis-driven probes separating structure from parameters via hard edge deletions and value-only scaling; Explanation via Input×Gradient (IXG) and Integrated Gradients (IG) attributions aggregated from nodes to ZCTAs; and Reproducibility through schemas and manifests across cities. Comparisons of San Francisco (SF) and New York City (NY) show CIPHER achieving price prediction accuracy with R^2 between 0.77 and 0.81, with stronger performance for metro in NY and for parking in SF. The ancillary population task has poor fit. Discussion of edge removal points out neighbourhood structure dependence, and value-only scaling captures parameter robustness. Attributions at the ZIP-level provide policy-relevant maps distinguishing between corridor-driven and community-driven areas.

Keywords: Heterogeneous Graph Transformer (HGT), Typed Mobility Graphs, Housing Price Inference, ZIP-Level Socioeconomics, Explainable Graph Learning.

1. Introduction

Urban computing is increasingly employing graph neural networks (GNNs) to model non-Euclidean spatial structure and spatiotemporal dynamics. This shift has driven advances in traffic analysis, land-use classification, and neighborhood inference [1]. Meanwhile, research on land-use has moved from single-modality classification to multimodal integration, involving mobility data, points of interest (POI), and imagery, together with more advanced experimental evaluations and reproducible pipelines for heterogeneous signal integration [2, 3].

Two streams of work are especially relevant to housing and land-use analytics. First, neighborhood-embedding studies learn representations by jointly encoding street-view imagery and POI data or by embedding geospatial networks, which learn semantically rich place representations. For instance, Urban2Vec fuses street-view imagery with POI distributions to derive multimodal neighborhood embeddings; meanwhile, geospatial network embedding exploits property–amenity topology modeling, also increases the precision of price forecasting [4, 5]. These findings motivate fusing operational mobility signals with ZIP-level indicators in a single graph-learning framework. Second, a complementary line develops heterogeneous multi-network graphs that connect infrastructure layers: integrating street and waterway systems reveals latent urban structures that single-layer graphs may miss, and Transformer-based visual fusion combined with GNN backbones for multimodal traffic forecasting demonstrates the practical value of cross-source fusion in smart-city prediction [6, 7].

At the intersection of urban analytics and real-estate appraisal, graph-based models are increasingly outperforming tabular baselines. The MugRep framework, which couples a dynamic transaction graph with a community-level heterogeneous graph, shows that co-modeling asynchronous spatiotemporal dependencies and inter-community relations improves appraisal accuracy over traditional Automated Valuation Models (AVMs) [8]. Even more recently,

Transformer-GNNs (T-GNNs) report gains over strong baselines and include attention-based comparable retrieval, reminding that graph formalisms are able to encapsulate the time–space organization underlying prices [9]. Simultaneously, applications of explainable methods to correlated statistics or knowledge graphs make it clear that valuation models must be interpretable and accurate [10, 11]. However, experiments with benchmarks disclose that GNN advantages are non-generalizable across datasets and definitions of tasks, eliciting more robust ablations and clearer problem definitions [12].

Explainability is more than a desirable property; it is an urban science lever. In environmental perception, explainable learning elucidates intricate, multi-faceted factors influencing the green-view index [13]. Interpretable machine learning quantifies how built-environment attributes influence outcomes with policy-relevant transparency [14, 15]. In mobility, neural relational inference (NRI) links prediction to mechanism by learning the graph structure itself and inspecting the inferred edges, highlighting both the potential and the limits of graph-level interpretation [16]. Taken together, these strands suggest that post-hoc attribution (such as IXG/IG) and structural probes (such as edge ablations and learned-adjacency inspection) should be integral components of urban graph modeling.

Gaps. (i) Most price-oriented GNNs operate on parcel/transaction graphs and treat mobility only implicitly (via distance or amenity features), leaving open how operational networks (bike, parking, metro) and their flows relate to economic outcomes such as housing prices; (ii) Even with multi-source fusion, many models lack type-aware heterogeneous message passing (for example, bike-station ↔parking-meter) or omit area-level socioeconomic context from node features; while multi-network embeddings show the value of type awareness, they are seldom coupled directly to price inference at neighborhood/ZCTA scales [7]; (iii) Although explanations are common, task-aligned evaluation — target-specific attributions, counterfactuals, and self-versus-neighbor contribution analyses — remains underdeveloped [10, 16].

This paper combines a heterogeneous mobility graph with ZCTA-level socioeconomic indicators for interpretable housing-price inference. The study extends an open-source Heterogeneous Graph Transformer (HGT) land-use pipeline with: (i) a multi-task regression head that augments outputs with median housing price; (ii) node-level enrichment that maps each mobility node to a ZIP Code (ZIP) and appends ZIP-derived attributes (median income, educational attainment, park coverage, student–teacher ratio); (iii) a post-hoc explanation suite (IXG, counterfactuals, residual maps, and self-versus-neighbor ablations). The framework is instantiated and evaluated in San Francisco using mobility datasets manually collected from SFMTA and Bay Wheels, and ZIP-level tables built from Zillow ZHVI, U.S. Census ACS, and OpenStreetMap. Experiments follow best-practice evaluation, including solid tabular baselines, ablations, and explicit explainability tests—with competitive accuracy and interpretable explanations. This pipeline is constructed and designed for cross-city replication, such as NY.

2. Related work

2.1. Urban representation learning and multimodal fusion

Surveys from the recent past indicate a movement from grid-based to graph-based forms that more effectively capture accessibility, movement dynamics, and community interaction [1]. With such a change, neighborhood embeddings like Urban2Vec which are trained with street-view images and POI data extract semantically augmented place vectors have been prominent. Geospatial network embeddings that include explicit property–amenity topology modeling also increase the precision of price forecasting [4, 5].

Two major design intuitions are deducible: (i) operational signals and contextual semantics are complementary; (ii) integration at the representation stage may yield improved outcomes over late concatenation. Following these statements, the current research integrates ZIP-level features so that node features are combined before the first message-passing layer, as opposed to being merely added in the final output phase.

2.2. Heterogeneous and multi-network graphs

Urban structure is characterized by layered infrastructural systems. Combinations of street and waterway networks reveal structure obscured in single-layer perception [7]. For mobility forecasting, hybrids of Transformers and GNNs of multi-source visual features also corroborate the promise of type-aware architectures [6]. This framework builds on this theme: typed mobility graphs formally differentiate bike, parking, and metro nodes, and relation-specific parameters in HGT disambiguate same-type and cross-type messages separately. This contrasts with isotropic or homogeneous GNNs that compressing edge semantics and disambiguating based on features alone. For cross-city replication, cross-type connectivity is also regulated through a geodesic radius limit to avoid spurious cross-river links in New York City.

2.3. Graph-based appraisal and domain-specific gains

In real estate appraisal, graph-based models have delivered competitive results over AVMs. MugRep combines transaction dynamics with community-level heterogeneous graphs and reports sizable improvements over strong AVM baselines [8]. T-GNN attributes gains to attention-based comparable and Transformer-style context windows [9]. These studies indicate that relational structure and temporal context jointly matter. This study is complementary: instead of parcel graphs (nodes as properties/parcels with proximity-based edges) or transaction graphs (nodes as sales events linked via spatiotemporal comparable), the focus is on operational nodes (bike docks, parking meters, metro stops) and proximity-based relations, augmented with ZIP-level socioeconomic attributes. Shifting the focus from parcels to mobility nodes provides a distinct perspective on how infrastructure relates to prices.

2.4. Explainable graph learning in urban science

Interpretable GNNs applied to related statistical data emphasize the need for models of valuation to disclose predictive power sources [10, 11]. For urban vitality and environmental perception, interpretable ML quantifies nonlinear built environment effects [13-15]. For transport forecasting, NRI learns edges and provides a basis for structural explanation [16]. Following these ideas, the existing framework combines hard edge deletions with value-only soft perturbations that scale relation-specific value vectors while keeping topology and attention weights fixed. These two probes allow the model to disentangle structural dependency from parameter-level stability. To make explanations policy-relevant, node-level attribution summaries are rolled up to ZIP-level and encapsulated through a single balancing ratio of mobility vs. socioeconomic signals.

2.5. Benchmarking and the case for stronger ablations

Geerts et al. warn that the hypothesized benefits of GNNs may fail to generalize; outcomes may strongly depend on design decisions, data leakage, and weakly formatted tasks [12]. These issues are alleviated with the CIPHER framework's reproducibility processes, comprising schema manifests, type-specific standardization, and cross-city alignment tests that quickly surface mismatches. Additionally, ablation experiments isolate self from neighbor channels, and metrics are provided in real units instead of normalized proxies—two potential fallacies in urban deep learning.

2.6. How prior art shapes the design

From multimodal embeddings comes the principle of early fusion, hence, ZIP Code attributes participate in message passing from the first layer [4, 5]. From heterogeneous and multi-network GNNs come type-aware message passing with relation-specific parameters and type-specific decoders [6, 7]. From appraisal-centric models come multi-task learning and architectures that admit price-oriented heads [8, 9]. From explainable GNNs and NRI come post hoc gradient attributions and structural probes; value-only soft perturbations are formalized here for an HGT encoder to reduce

confounds relative to edge deletions [10, 11, 16]. Following benchmarking critiques, diagnostic ablations and alignment-first engineering are prioritized, as codified in CIPHER [12].

3. Method

This section details how CIPHER turns operational mobility data and ZIP attributes into an interpretable, cross-city price inference system.

3.1. C — Construction: typed urban graph

Each operational station (bike dock, parking meter, metro stop) is a node v. This subsection summarizes hourly inflow/outflow as fixed-length statistics (peaks, means, dispersions, diurnal ratios) and concatenates light station metadata, forming $x_v \in \mathrm{R}^F$ (in the experiments F=52).

City schemas differ (SF: bike/parking; NY: bike/metro), so this part records node and edge schemas at build time to avoid hard-coded types. Edges include same-type KNN (bike↔bike; parking↔parking or metro↔metro) to capture local connectivity and cross-type KNN (bike↔parking/metro) to expose inter-layer coupling such as last-mile effects. In NY this study caps cross-type edges by a geodesic radius to avoid cross-river artifacts. Self-loops are retained for interpretability tests.

3.2. I — Integration: ZIP-level attributes

Stations are spatially joined to ZIP polygons (ZCTAs). This study appends median income, education (bachelor+), park coverage (area ratio, parks per capita), school count, and student–teacher ratio. These are harmonized to ZIP-level and merged by a consistent key, enabling early fusion: ZIP context participates in message passing from the first layer.

To avoid leakage and preserve cross-city comparability, this study fits scalers per city and node type:

$$\tilde{x}_{i,f}^{(c)} = \frac{x_{i,f}^{(c)} - \mu_f^{(c)}}{\sigma_f^{(c)} + \varepsilon} \tag{1}$$

In Equation 1: c: city (SF or NY); i: node index; f: feature index; $x_{i,f}^{(c)}$: raw feature; $\tilde{x}_{i,f}^{(c)}$: standardized feature; $\mu_f^{(c)}/\sigma_f^{(c)}$: mean and standard deviation of feature f computed on the training split of city c; $\epsilon > 0$: a small constant for numerical stability (constant).

Targets $y_v \in \mathbb{R}^T$ combine six land-use–derived signals, median price, and optionally population (T=7 or T=8). To stabilize training while keeping outputs in real units, the study standardizes the price target on the training split and inverts exactly at evaluation:

$$\tilde{y}_{\text{price}}^{(c)} = \frac{y_{\text{price}}^{(c)} - \mu_{\text{price}}^{(c)}}{\sigma_{\text{price}}^{(c)} + \varepsilon}, \quad (\hat{y})_{\text{price}} = \left(\hat{\tilde{y}}\right)_{\text{price}} \sigma_{\text{price}}^{(c)} + \mu_{\text{price}}^{(c)} \tag{2}$$

In Equation 2: $y_{\text{price}}^{(c)}$: ground-truth median price (real units, USD) for city c; $\tilde{y}_{\text{price}}^{(c)}$: standardized price target used in training; $\mu_{\text{price}}^{(c)}, \sigma_{\text{price}}^{(c)}$: mean and standard deviation of price computed on the training split of city c; $\left(\hat{\tilde{y}}\right)_{\text{price}}$: model output for standardized price; $(\hat{y})_{\text{price}}$: de-standardized predicted price (real units). Other targets (land-use, optional population) are scaled per city/task and inverted before reporting.

3.3. P — Prediction: HGT encoder–decoder and multi-task head

For each layer and node type τ, type-specific projections produce $\mathbf{K}, \mathbf{Q}, \mathbf{V}$:

$$\left[K_u, Q_v, V_u\right] = \text{split} W_\tau^{\mathbf{kqv}} h, \; h \in \mathrm{R}^{H_{\text{hid}}} \tag{3}$$

In Equation 3: τ: node type (bike, parking, metro); u: source node; v: destination node; $h \in R^{H_{hid}}$: input hidden state for the current layer; H_{hid}: hidden size; W_{τ}^{kqv} : type-specific linear maps producing $\mathbf{K}, \mathbf{Q}, \mathbf{V}$; $[K_u, Q_v, V_u]$: key of source, query of destination, value of source after the type-specific projection and head reshaping; $split(\cdot)$: splits one linear output into $\mathbf{K}, \mathbf{Q}, \mathbf{V}$ blocks (and heads internally).

Relation-specific transforms map them to $\widetilde{K}_e$, $\widetilde{Q}_e$, $\widetilde{V}_e$:

$$\widetilde{K}_e = W_{\rho(e)}^{(K)} K_u,\ \widetilde{Q}_e = W_{\rho(e)}^{(Q)} Q_v,\ \widetilde{V}_e = W_{\rho(e)}^{(V)} V_u \tag{4}$$

In Equation 4: e: an edge; $\rho(e)$: relation type of edge e; $W_{\rho(e)}^{(K)}$, $W_{\rho(e)}^{(Q)}$, $W_{\rho(e)}^{(V)}$: relation-specific linear transforms for $\mathbf{K}, \mathbf{Q}, \mathbf{V}$; $\widetilde{K}_e$, $\widetilde{Q}_e$, $\widetilde{V}_e$: relation-colored key, query, and value associated with edge e.

Then this study computes multi-head attention with a relation bias:

$$\alpha_{e,h} = \mathrm{softmax}_{e \in N(v)} \left(\frac{\langle (\widetilde{Q})_{e,h}, (\widetilde{K})_{e,h} \rangle}{\sqrt{d}} + b_{\rho(e),h} \right) \tag{5}$$

In Equation 5: $\alpha_{e,h}$: attention weight on edge e for head h; $N(v)$: set of incoming edges to node v (destination); $\langle \cdot, \cdot \rangle$: inner product; $\sqrt{d}$: (d) dimension per head (so $\sqrt{d}$ scales scores); $b_{\rho(e),h}$: relation and head-specific bias term; $\mathrm{softmax}_{e \in N(v)}$: normalizes over all incoming edges to v.

After that, values are aggregated, and a type-specific residual is added:

$$z_v = W_{\tau(v)}^{out} \left(\oplus_{h=1}^{H} \sum_{e \in N(v)} \alpha_{e,h} \left(\widetilde{V} \right)_{e,h} \right) + \gamma_{\tau(v)} h_v \tag{6}$$

In Equation 6: z_v: post-aggregation representation for node v; $\left(\widetilde{V} \right)_{e,h}$: value vector for edge e, head h after relation transform; H: number of heads; $\oplus$: concatenation across heads; $W_{\tau(v)}^{out}$: type-specific output projection for node type $\tau(v)$; $\gamma_{\tau(v)}$: type-specific residual scaling coefficient (scalar); h_v: skip-connected hidden state from the layer input at node v.

Next comes stacking and decoding: This study stacks L HGT layers with nonlinearity $\varphi(\cdot)$ and applies a type-specific decoder $W_{\tau(v)}^{dec}$ to the final state $h_v^{(L)}$ to produce a T-dimensional prediction $\hat{y}_{\tau(v)}$.

Training objective: This study optimizes a task-weighted loss over nodes, using either MAE or MSE per task. Let $\lambda_t > 0$ denote task weights ($\lambda_{price} > 1$, $\lambda_{pop} < 1$). All metrics are reported in real units after inverting any standardization (Eq. 2).

3.4. H — Hypothesis-Driven perturbations

CIPHER's last three letters—H, E, R—summarize how this paper diagnosis's structure and features and keeps cross-city runs sound. This section quantifies reliance on self vs. neighborhood channels via hard deletions and soft value-only perturbations.

Hard partition of edges:

$$\varepsilon_{self} = \{(u,v) \in \varepsilon : \tau(u) = \tau(v) \wedge u = v\},\ \varepsilon_{neigh} = \varepsilon \setminus \varepsilon_{self} \tag{7}$$

In Equation 7: ε: set of directed edges; ε_{self}: self-loop edges (source equals destination within the same type); ε_{neigh}: non-self-edges (neighborhood channel): $\tau(\cdot)$: node-type function; u, v: node identifiers. The analysis runs forward, passes with self-only vs. neighbor-only graphs and compares output distributions.

Soft value-only scaling keeps topology and attention intact while scaling relation-specific values:

$$\widetilde{V}_{e,h}^{(soft)} = s_{\rho(e)} \widetilde{V}_{e,h},\ s_{\rho} \in [0, 1] \tag{8}$$

In Equation 8: $\widetilde{V}_{e,\mathrm{h}}^{(\mathrm{soft})}$: scaled value for edge e, head h; $s_\rho \in [0, 1]$: scaling factor applied uniformly to all edges of relation ρ; Topology and attention ($\alpha_{e,\mathrm{h}}$) are held fixed; only values are scaled.

Distributional and effect-size summaries. This study reports (i) the Kolmogorov–Smirnov distance between the two empirical CDFs of outputs (self-only vs. Neighbor-only), defined as the maximum vertical gap between the CDFs, and (ii) the standardized mean difference (Cohen's d) below.

Cohen's d with pooled variance.

$$d=\frac{\mu_{\mathrm{self}}-\mu_{\mathrm{neigh}}}{s_p},\ s_p=\sqrt{\frac{(n_s-1)\sigma_s^2+(n_n-1)\sigma_n^2}{n_s+n_n-2}} \tag{9}$$

In Equation 9: μ_{self}, μ_{neigh}: sample means of outputs under self-only and neighbor-only interventions; σ_s, σ_n: sample standard deviations for the two settings; n_s, n_n: sample sizes; s_p: pooled standard deviation; d: Cohen's effect size (standardized mean difference).

3.5. E — Explanation at node and ZIP scales

Node-level Input×Gradient (IXG).

$$\mathbf{s}_{\mathbf{v}}^{(\mathbf{t})}=\left|\mathbf{x}_{\mathbf{v}}\odot\nabla_{\mathbf{x}_{\mathbf{v}}}(\hat{y})_{v,t}\right| \tag{10}$$

In Equation 10: $\mathbf{s}_{\mathbf{v}}^{(\mathbf{t})}\in\mathrm{R}$: per-feature importance vector for node v and task t; $\mathbf{x}_{\mathbf{v}}$: input feature vector at node v; $\nabla_{\mathbf{x}_{\mathbf{v}}}(\hat{\mathbf{y}})_{\mathbf{v},\mathbf{t}}$: gradient of prediction with respect to the input features; $\odot$: element-wise multiplication; $|\cdot|$: element-wise absolute value.

ZIP aggregation: For each ZIP z, this study computes the mean importance of feature f over nodes in $V(z)$, such as the arithmetic average of $s_{v,f}^{(t)}$ across all v joined to z.

ZIP-share ratio: This study reports the proportion of ZIP-group feature importance within the total (mobility + ZIP), defined as the sum of ZIP-feature importances divided by the sum of mobility-feature plus ZIP-feature importances (with a small $\varepsilon > 0$ to avoid division by zero). Values near 1 indicate socioeconomic dominance; values near 0 indicate mobility dominance.

3.6. R — Reproducibility and alignment

The proposed system enforces schema manifests, covering node/edge types, feature order, scalers, split masks, and target definitions. Cross-city checkpoints load only when source and destination schemas are identical, avoiding silent misalignment. Metrics are reported in real units (MAE/RMSE/R^2) at both type-specific and pooled levels. Any cross-city mismatch in the manifest (feature order/length; X/Y shapes; per-type scalers; target order and indices; node-type sets and split masks; equality of edge-type sets (modulo parking↔metro swap); decoder out-dim; and the hyperparameter profile.) raises a clear error.

4. Experiment

4.1. General accuracy in real units (price as the primary objective)

The model is tailored to predict house prices. This study includes population as an ancillary task to test whether the same representation captures coarse-grained demographic variation, but parity is not expected. Many of the predictors—mobility intensity, station/parking context, and neighborhood socioeconomic proxies—are closer to the economic and accessibility mechanisms that set prices than to the slow-moving counts that define resident population.

Table 1. Real-unit accuracy by city and node type on the test split (Price)

City	Node	MAE	RMSE	R^2	N
SF	Bike	$100,860	$149,044	0.761	86
SF	Parking	$138,455	$176,349	0.807	2077
SF	ALL	$136,960	$175,345	0.806	2163
NY	Bike	$173,584	$229,481	0.765	338
NY	Metro	$128,142	$172,156	0.807	64
NY	ALL	$166,349	$221,351	0.772	402

Table 1 (price) confirms this asymmetry. Price is accurately predicted in both cities: SF (ALL) reaches MAE $136,960; RMSE $175,345; R^2=0.806, and NY (ALL) reaches MAE $166,349; RMSE $221,351; R^2=0.772. Node-type comparisons reinforce the structural storyline: NY metro R^2=0.807 exceeds NY bike R^2=0.765 despite a smaller test set (N=64 vs 338), consistent with the corridor-like predictability of subway lines. In SF, parking R^2=0.807 also exceeds bike R^2=0.761, suggesting that static facilities internalize stronger neighborhood and socioeconomic signals than bike nodes. These patterns are typical of price formation: accessibility, amenity capital, and corridor organization are proximate determinants, and each is captured by the graph and features.

Table 2. Real-unit accuracy by city and node type on the test split (Population)

City	Node	MAE	RMSE	R^2	N
SF	Bike	8,410	11,835	0.194	86
SF	Parking	14,614	18,185	-0.072	2077
SF	All	14,367	17,976	-0.066	2163
NY	Bike	15,689	19,936	0.357	338
NY	Metro	16,508	19,131	0.402	64
NY	All	15,819	19,810	0.369	402

By contrast, Table 2 (population) exhibits a clear mismatch with the present representation. NY (ALL) attains moderate explanatory power (R^2=0.369; MAE 15,819 persons), whereas SF (ALL) is weak (R^2=−0.066; MAE 14,367). This is not a failure of the modeling class so much as a target–feature misalignment. Population shows significant responsiveness to temporal variations, such as seasonality, migration, and short-term residency compared to long-term residency. The representation emphasizes the movement of nodes and accessibility, which are good evidentiary indicators for demand pressure impacting price setting; however, these are only weakly related to officially registered residents. Briefly, the representation is optimized for the price signal, and the learned representation does not directly carry over to population without time-resolved covariates (such as event calendars, terms of schooling, tourist cycles) and more refined spatial reconciliation. The divergence therefore justifies the selected design: the model's strength lies exactly in the spot designated: price, whereas the additional task highlights directions for future feature augmentation rather than contradicting the approach.

Two operational nuances contextualize Table 1 (price). First, the error spread scales with market heterogeneity: broader price dispersion inflates RMSE even at similar R^2. Second, sample-size asymmetry (for example, NY-metro N=64) cautions against over-interpreting small type-level gaps; the ordering with metro higher than bike is nonetheless robust and consistent with network theory.

4.2. Structural dependence: hard deletions

To test whether price inference indeed rests on neighborhood structure, this analysis reruns the model under hard ablations that retain self-only or neighbor-only channels.

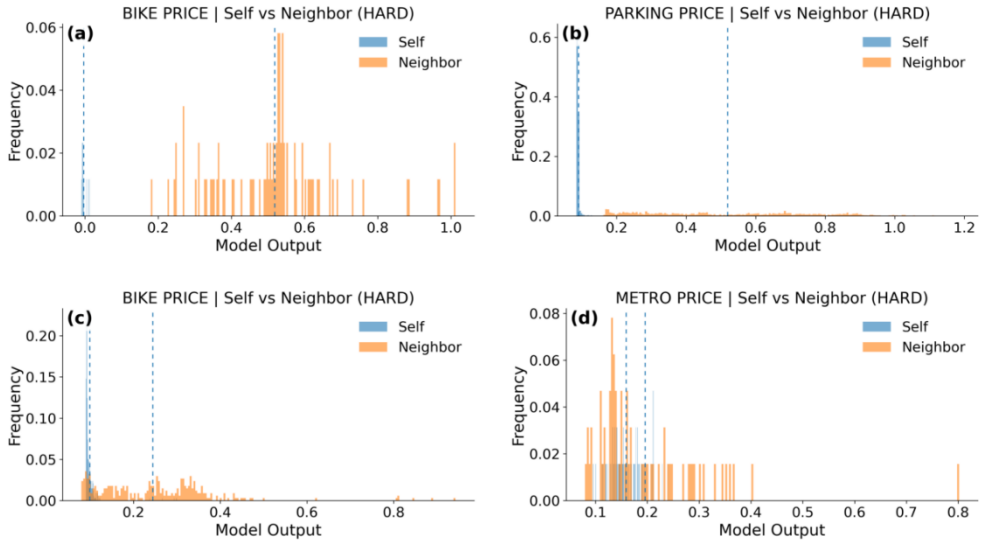

Figure 1. Hard ablation (self-only vs neighbor-only) histograms for price

Figure 1: (a) SF bike, (b) SF parking, (c) NY bike, (d) NY metro.

Removing neighborhood edges induces large distributional shifts, confirming neighborhood dependence.

Table 3: Report KS distance (KS D), approximate p-value, Cohen's d and mean outputs under hard ablation (self-only vs neighbor-only graphs).

Table 3. Hard ablation statistics for price

City	Node	KS D	p (approx)	Cohen's d	Mean (self-only)	Mean (neigh-only)
NY	Bike	0.781	<1e-6	-1.54	0.099	0.244
NY	Metro	0.281	0.012659	-0.44	0.159	0.196
SF	Bike	1.000	<1e-6	-4.28	-0.003	0.518
SF	Parking	0.994	<1e-6	-2.45	0.091	0.519

Figure 1 shows stark distributional shifts; Table 3 quantifies them. For SF bike, KS D = 1.000, $p<10^{-6}$, and Cohen's d=-4.28; the mean moves from −0.003 (self-only) to 0.518 (neighbor-only). SF parking behaves similarly (KS D = 0.994, $p<10^{-6}$, d=−2.45; 0.091→ 0.519). NY bike remains strongly structure-dependent (KS D = 0.781, $p<10^{-6}$, d=−1.54; 0.099→ 0.244). NY metro is significant but milder (KS D = 0.281, p≈0.0127, d=−0.44; 0.159 → 0.196), indicating that station-level self-features retain explanatory value even without neighbor aggregation.

These findings reveal structural anchoring: without neighborhood edges, the price signal cannot be reconstructed; with edges preserved, aggregation amplifies economically meaningful context. The near-maximal KS in SF bike/parking indicates not merely a mean shift but a reshaping of the distribution—a signature of graph-borne information rather than node-local regressors.

4.3. Parametric robustness: value-only scaling

The analysis then asks whether performance is sensitive to parameter rescaling when structure is held fixed. In the soft ablation, the study rescales the value while keeping attention and topology unchanged. Figure 2 shows that self-only and neighbor-only histograms largely overlap; Table 4 reports KS D collapsing from the hard-ablation regime to near-zero across all types (NY bike: 0.781 → 0.006; SF bike: 1.000 → 0.058; NY metro: 0.281 → 0.047; SF parking: 0.994 → 0.002), with p-values near 1 and $d \approx 0$.

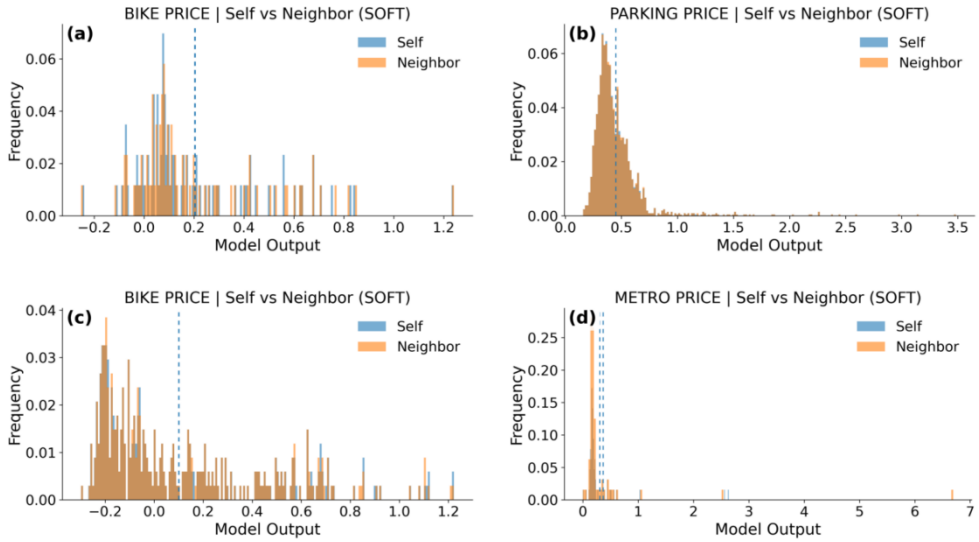

Figure 2. Soft value-only scaling ablations for price (attention/topology fixed)

Figure 2: (a) SF bike, (b) SF parking, (c) NY bike, (d) NY metro. Distributional gaps are substantially smaller than in Figure 1, evidencing parametric robustness.

Table 4. Soft ablation statistics for price

City	Node	KS D	p (approx.)	Cohen's d	Mean (self-only)	Mean (neigh-only)
NY	bike	0.006	1	0.00	0.101	0.100
NY	metro	0.047	1	-0.09	0.304	0.366
SF	bike	0.058	1	0.01	0.204	0.203
SF	parking	0.002	1	-0.00	0.450	0.450

Table 4: Report KS distance (KS D), approximate p-value, Cohen's d and mean outputs under soft ablation (value-only; attention/topology fixed).

The encoder–decoder thus absorbs value-scale perturbations, consistent with multi-head aggregation and type-specific residuals acting as internal normalizers. A few long-tail outliers remain for NY metro, but their rarity and negligible impact on means indicate robustness, not brittleness.

Taken together, the hard setting has a much larger effect than the soft setting, which cleanly separates structure from parameters and helps explain why price generalizes across node types whereas population does not.

4.4. Node-level attributions (IXG)

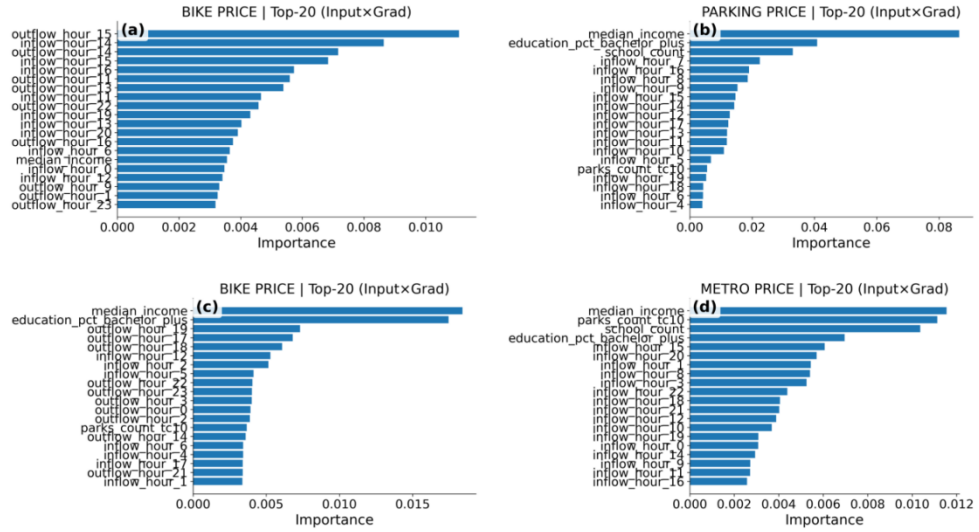

Figure 3. Lists the Top-20 Input×Gradient features for each city/type

Figure 3: (a) SF bike, (b) SF parking, (c) NY bike, (d) NY metro. Both mobility and ZIP-features contribute, with relative weighting varying by city and node type.

SF bike is mobility-dominant: the highest ranks are dominated by hourly mobility (in/out) flow features—most prominently the 14–16 h block and adjacent late-morning/early-evening hours (11–13, 17)—with ZIP covariates significantly lagging behind. This emphasizes that hour-of-day flow intensity and local accessibility around bike stations underpin bike-area prices.

SF Parking is ZIP-dominant: ZIP-level socioeconomic variables —*median_income*, *education_pct_bachelor_plus*, *school_count*—predominate, with only a weak tail of hourly mobility (7–9, 15–17), suggesting that neighborhood institutions and capital matter more than transient traffic near parking facilities.

NY bike is mixed: *median_income* and *education_pct_bachelor_plus* anchor the list, while evening-peak mobility (in/out) flow features (17–19, plus midday inflows) contribute meaningfully—suggesting a socioeconomic core modulated by peak-time flows.

NY metro is ZIP-first, mobility-follow: *median_income*, *parks_count_t10*, *school_count*, and *education_pct_bachelor_plus* dominate, with secondary inflow hourly (8, 10–12, 15, 20–22), consistent with the metro's role as a high-coverage accessibility backbone. Attributions are not causal; they indicate where the model finds evidence and mirror the hard-ablation findings.

Evaluation: ZIP structure and Mobility attributes both allow for price inference, but weights differ across cities/types: SF bike is most subject to hour-based mobility, while SF parking, NY metro, and NY bike are more subject to ZIP structure (income, education, schools, parks). This cross-city variation conforms with Section 4.2's structural dependence, and Section 4.3's parametric robustness, with the corollary here being that edges/topology are an unavoidable and salient determinant of the model, while parameter rescaling has negligible effect in these experiments. The attribution profiles also comply with domain evidence—socioeconomic capital and institutional density correlate with high prices, and peak-time flows capture reachability and demand—thus increasing interpretability.

4.5. From nodes to ZIPs: aggregated importance & heterogeneity

Aggregating IXG by ZIP vs mobility clarifies the global balance.

Table 5. Node-level group shares (ZIP vs mobility) for price

City	Node	ZIP share (%)
NY	bike	23.8
NY	metro	36.6
SF	bike	5.9
SF	parking	46.3

Table 5: Report the ZIP share= ZIP/ (ZIP + Mobility) aggregated over test nodes.

Table 5 shows SF bike at 5.9% ZIP share (near pure mobility), SF parking at 46.3% (near parity), NY bike at 23.8%, and NY metro at 36.6%. Metro's higher ZIP share relative to bike implies that station-area attributes matter, yet both remain below 50%, so mobility is the majority driver in NY. This reconciles the IXG Top-20 with the ablations: price depends on whom nodes connect to (flows) but is modulated by who they are (ZIP endowments), with weights shifting by node type and urban form.

Table 6: Surfaces ZIP-level heterogeneity that is decision-useful.

From Table 6: In NY, 11109/11215/10006/11208/11201 populate the high ZIP share tail (≈ 46–72 %), pointing to community-driven areas where investments in education, parks, and amenities are most likely to shift prices. 10036/10009/10001/11216/11205 occupy the low tail (≈ 4–7 %), indicating mobility-driven locales where transfer efficiency, corridor capacity, and first/last-mile improvements dominate the payoff. SF presents the same dichotomy (94123/94131/94134/94114/94112 high at 58–66 %; 95126/94612/95125/94601/94609 very low at ~1–1.5 %). Because these shares arise from the trained model rather than ex-ante heuristics, they provide a lightweight diagnostic for policy triage.

Table 6. ZIP-level heterogeneity: Top-5 and Bottom-5 ZIPs

SF Top-5 by zip_share		SF Bottom-5 by zip_share	
ZIP	**zip_share**	**ZIP**	**zip_share**
94123	66.2%	95126	1.0%
94131	65.2%	94612	1.3%
94134	63.7%	95125	1.4%
94114	63.6%	94601	1.5%
94112	58.4%	94609	1.5%
NY Top-5 by zip_share		NY Bottom-5 by zip_share	
ZIP	**zip_share**	**ZIP**	**zip_share**
11109	71.6%	10036	4.0%
11215	57.3%	10009	5.8%
10006	55.6%	10001	6.1%
11208	47.0%	11216	6.7%
11201	46.4%	11205	7.4%

4.6. Cross-city synthesis and implications

The image built by this work emerges gradually into focus. The results achieve high precision for price prediction and disclose enhancements with certain urban structural configurations, such as metro and parking. However, population-related predictors indicate limitations of the existing set of features: lacking time-varying and boundary-constrained covariates, a price-optimized model cannot capture resident-population information. Structural information is the primary cause of this limitation, as seen from cliff-like drops under hard ablations and stability under soft ones. Attribution analysis connects the mechanistic chain: flows, in other words, human mobility signals, transmit through nodes in the network, while ZIP-level resource endowments regulate expression strength node-wise. Resulting heterogeneity dictates potential intervention entry points—transportation vs. community resources—in ways consistent with existing governance regimes. The ensuing modeling benefits fall overwhelmingly in graph architecture: origin–destination (OD) weighted cross-ZIP edges; peak and off-peak relation types; attention masks aligned with scheduled service; and time-varying edge weights corresponding to seasonality or congestion behavior. For planning, the tables yield an operational switchboard: areas with high ZIP resource share are amenable to amenity-oriented interventions, while areas with low ZIP share are amenable to mobility-oriented policy measures. Population should be considered an independent prediction target with behavioral attributes of its own, rather than an incidental product of price modeling here. This makes the signal clearer and disentangles different concepts of "demand."

5. Conclusion

This work introduces CIPHER, an interpretable and typed graph-learning framework that infers median house prices at the ZIP code scale by combining mobility-network data with ZIP-level socioeconomic features. The theoretical and empirical contributions of CIPHER lie in two aspects: (i) In principle, experiments show that early ZIP fusion (injecting area features into node features before the first message-passing layer) has the potential to benefit cross-city portability and interpretability to an enormous degree. (ii) Empirically, a diagnostic set of criteria distinguishes structural effects from the effects of parameters: hard deletions (self-only vs. neighbor-only) establish dependence on neighborhood context, while value-only soft scaling reveals encoder robustness with fixed topology and attention scheme. In New York and San Francisco, CIPHER achieves high precision on prices and provides interpretable spatial explanations, with type-level leads for parking (SF) and metro (NY). Addressing population as an ancillary task is educational, emphasizing the fact that when the target is residency dynamics rather than market-related socioeconomic data, covariates must be time-sensitive and boundary-aware. ZIP-aggregated attributions also transform node-level

evidence into policy advice, marking out socioeconomic driver dominance regions and mobility driver dominance regions, thus guiding planners towards transport–corridor–oriented or amenity-oriented plan levers. CIPHER also institutionalizes reproducibility: schema manifests, per-type feature scaling, cross-city alignment checks, and real-unit reporting reduce common pitfalls like data-type mismatches in urban deep learning and assist robust replication. Future extensions include graph design (origin–destination weighting, temporal relation types, scheduled-service masks), cross-city transfer (domain adaptation of encoders with calibrated heads), and uncertainty (intervals, calibration curves). Incorporating time explicitly through spatiotemporal Heterogeneous Graph Transformers (HGTs) (such as, weekly cycles and seasonal regimes) will also help ancillary targets like population and short-term analysis. Briefly, CIPHER extends a land-use Heterogeneous Graph Transformer (HGT) into a cross-city, interpretable price-inference mechanism accurate on the core signal (price), explains causes and circumstances on which its operation depends, and opens systematic comparison across and beyond urbanized areas. Such tools, especially the structural probes and ZIP-level interpretations, must be an exemplar of urban forecasting programs which should also have predictive capability as well as explanatory processes.

References

[1] G. Jin, Y. Liang, Y. Fang, Z. Shao, J. Huang, J. Zhang, Y. Zheng, Spatio-temporal graph neural networks for predictive learning in urban computing: A survey. arXiv: 2303.14483 (2023). https://doi.org/10.48550/arXiv.2303.14483.

[2] Z. Li, B. Chen, S. Wu, M. Su, J. Chen, B. Xu, Deep learning for urban land use category classification: A review and experimental assessment. Remote Sens. Environ. 311, 114290 (2024). https://doi.org/10.1016/j.rse.2024.114290.

[3] M. Z. U. Rehman, S. M. S. Islam, D. Blake, A. Ulhaq, N. Janjua, Deep learning for land use classification: A systematic review of hyperspectral–LiDAR data fusion. Artif. Intell. Rev. 58, 272 (2025). https://doi.org/10.1007/s10462-025-11265-z.

[4] S. S. S. Das, M. E. Ali, Y.-F. Li, Y.-B. Kang, T. Sellis, Boosting house price predictions using geo-spatial network embedding. Data Min. Knowl. Discov. 35, 2221–2250 (2021). https://doi.org/10.1007/s10618-021-00789-x.

[5] Z. Wang, H. Li, and R. Rajagopal, Urban2Vec: Incorporating street-view imagery and POIs for multi-modal urban neighborhood embedding, in Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI-20), New York, NY, USA, February 7–12 (2020), 1013–1020. https://doi.org/10.1609/aaai.v34i01.5450.

[6] S. Dhanasekaran, D. Gopal, J. Logeshwaran, N. Ramya, A. O. Salau, Multi-model traffic forecasting in smart cities using graph neural networks and transformer-based multi-source visual fusion for intelligent transportation management. Int. J. Intell. Transp. Syst. Res. 22, 518–541 (2024). https://doi.org/10.1007/s13177-024-00413-4.

[7] P. Liu, Y. Wang, S. De Sabbata, B. Lei, F. Biljecki, J. Tang, R. Stouffs, Living upon networks: A heterogeneous graph neural embedding integrating waterway and street systems for urban form understanding. Environ. Plan. B Urban Anal. City Sci. 0(0) (2025). https://doi.org/10.1177/23998083251358527.

[8] W. Zhang, H. Liu, L. Zha, H. Zhu, J. Liu, D. Dou, H. Xiong, MugRep: A multi-task hierarchical graph representation learning framework for real estate appraisal, in Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Singapore, August 14–18 (2021), 3937–3947. https://doi.org/10.1145/3447548.3467187.

[9] F. Moghimi, R. A. Johnson, A. Krause, Rethinking real estate pricing with transformer graph neural networks (T-GNN), in Proceedings of 2023 IEEE International Conference on Machine Learning and Applications, Jacksonville, FL, USA, December 15–17 (2023), 1405–1411. https://doi.org/10.1109/ICMLA58977.2023.00212.

[10] P. Brimos, A. Karamanou, E. Kalampokis, M. E. Mamalis, K. Tarabanis, Explainable graph neural networks on linked statistical data for predicting Scottish house prices, in Proceedings of the 27th Pan-

Hellenic Conference on Informatics, Lamia, Greece, November 24–26 (2023), 36–41. https://doi.org/10.1145/3635059.3635065.

[11] A. Karamanou, P. Brimos, E. Kalampokis, K. Tarabanis, Explainable graph neural networks: An application to open statistics knowledge graphs for estimating house prices. Technologies. 12, 128 (2024). https://doi.org/10.3390/technologies12080128.

[12] M. Geerts, S. vanden Broucke, J. De Weerdt, Graph neural networks for house price prediction: Do or don't? Int. J. Data Sci. Anal. 20, 3563–3593 (2025). https://doi.org/10.1007/s41060-024-00682-y.

[13] C. Chen, J. Wang, D. Li, X. Sun, J. Zhang, C. Yang, B. Zhang, Unraveling nonlinear effects of environment features on green view index using multiple data sources and explainable machine learning. Sci. Rep. 14, 30189 (2024). https://doi.org/10.1038/s41598-024-81451-6.

[14] Y. Xie, J. Zhang, Y. Li, Z. Zhu, J. Deng, Z. Li, Integrating multi-source urban data with interpretable machine learning for uncovering the multidimensional drivers of urban vitality. Land. 13, 2028 (2024). https://doi.org/10.3390/land13122028.

[15] W. Liu, Z. Yang, C. Gui, G. Li, H. Xu, Investigating the nonlinear relationship between the built environment and urban vitality based on multi-source data and interpretable machine learning. Buildings. 15, 1414 (2025). https://doi.org/10.3390/buildings15091414.

[16] M. N. Tygesen, F. C. Pereira, F. Rodrigues, Unboxing the graph: Towards interpretable graph neural networks for transport prediction through neural relational inference. Transp. Res. Part C Emerg. Technol. 146, 103946 (2023). https://doi.org/10.1016/j.trc.2022.103946.