Transportation Research Part B 209 (2026) 103464

Contents lists available at ScienceDirect

Transportation Research Part B

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

Graph neural networks for residential location choice: Connection to classical logit models

Zhanhong Cheng [a,b], Lingqian Hu [c], Yuheng Bu [d], Yuqi Zhou [b], Shenhao Wang [b,*]

[a] College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, Zhejiang, China
[b] Department of Urban and Regional Planning, University of Florida, Gainesville, FL, USA
[c] Department of Landscape Architecture and Urban Planning, Texas A&M University, College Station, TX, USA
[d] Department of Computer Science, University of California Santa Barbara, Santa Barbara, CA, USA

ARTICLE INFO

Keywords:
Residential location choice
Spatial choice model
Graph neural network
Discrete choice model
Nested logit model

ABSTRACT

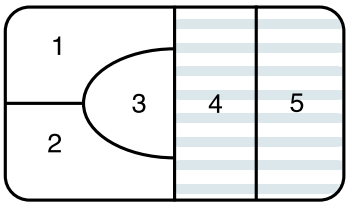
Researchers have adopted deep learning for classical discrete choice analysis as it can capture complex feature relationships and achieve higher predictive performance. However, the existing deep learning approaches cannot explicitly capture the relationship among choice alternatives, which has been a long-lasting focus in classical discrete choice models. To address the gap, this paper introduces Graph Neural Network (GNN) as a novel framework to analyze residential location choice. The GNN-based discrete choice models (GNN-DCMs) offer a structured approach for neural networks to capture dependence among spatial alternatives, while maintaining clear connections to classical random utility theory. Theoretically, we demonstrate that the GNN-DCMs incorporate the nested logit (NL) model and the spatially correlated logit (SCL) model as two specific cases, yielding a novel algorithmic interpretation through message passing among alternatives' utilities. Empirically, the GNN-DCMs outperform benchmark MNL, SCL, and feedforward neural networks in predicting residential location choices among Chicago's 77 community areas. Regarding model interpretation, the GNN-DCMs can capture individual heterogeneity and exhibit spatially-aware substitution patterns. Overall, these results highlight the potential of GNN-DCMs as a unified and expressive framework for synergizing discrete choice modeling and deep learning in the complex spatial choice contexts.

1. Introduction

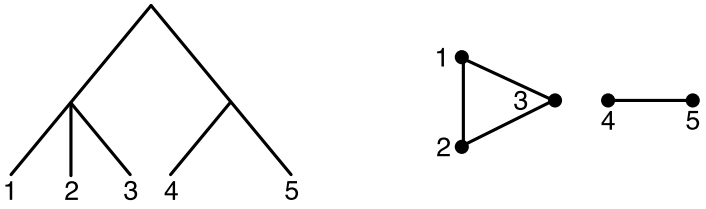
It is essential to understand individual households' residential location choices for transportation planning (Ben-Akiva and Bowman, 1998), urban development (Bhat and Guo, 2007), and economic policy (Bayoh et al., 2006). Residential location choices are typically framed as discrete choice modeling (DCM) tasks. But unlike other widely-studied DCMs, such as travel mode choice, residential location choice is more challenging because it involves a *large* set of *spatially correlated* alternatives (e.g., neighborhoods). Such complexity introduces several challenges for classical DCMs from the 1970s through the 2010s. First, the spatial correlation among alternatives violates the Independence of Irrelevant Alternatives (IIA) assumption inherent in the Multinomial Logit (MNL) model. Second, the large size of the alternative set makes it difficult to specify correlation structures. Third, residential choices often involve complex interactions between individual characteristics and location attributes. In the classical framework, researchers have

* Corresponding author.
E-mail address: shenhaowang@ufl.edu (S. Wang).

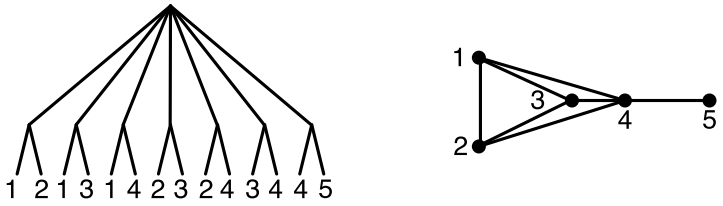
https://doi.org/10.1016/j.trb.2026.103464
Received 31 January 2026; Accepted 30 March 2026
Available online 22 April 2026



(a) Spatial configuration of residential locations.



(b) The nested structure of the NL model (left) and its corresponding alternative graph (right) for the GNN.



(c) The nested structure of the SCL model (left) and its corresponding alternative graph (right) for the GNN.

Fig. 1. Illustration of the alternative graphs in the NL and SCL models for residential location choice.

been seeking to address these challenges using the nested logit (NL) model (McFadden, 1978), spatially correlated logit (SCL) model (Bhat and Guo, 2004), and its extensions (Sener et al., 2011; Perez-Lopez et al., 2022).

In recent years, neural networks (NN) have gained increasing attention because they can capture non-linear relationships and complex interactions between individual and alternative attributes, outperforming traditional DCMs in predictive accuracy (Wang et al., 2024). However, the majority of the existing studies introduce only artificial neural networks (ANNs) and their variants, thus failing to explicitly capture the *dependencies among choice alternatives*. The existing ANN-based discrete choice models assume independent alternatives (Wang et al., 2020a), implicitly encode alternatives' correlations by feeding the attributes of all alternatives (Wang et al., 2020b; Wong and Farooq, 2021), or embed ANNs within traditional DCMs to facilitate utility specification (Sifringer et al., 2020). None of them can explicitly capture alternatives' dependencies while maintaining a clear connection to classical discrete choice theory.

To address such limitations, this paper proposes a Graph Neural Network framework for discrete choice modeling (GNN-DCMs), specifically designed for large sets of spatially correlated alternatives such as residential location choice. Instead of embedding neural networks directly in a Generalized Extreme Value (GEV) model (Sifringer et al., 2020), the GNN-DCMs can capture alternatives' correlation through message passing between utilities of alternatives. Fig. 1 illustrates a novel concept, i.e., alternative graph, which provides a new interpretation for NL and SCL models as message passing among alternative utilities. Combining the alternative graph and GNNs, we demonstrate theoretically that GNN-DCMs incorporate classical NL, SCL, and ASU-DNN (Wang et al., 2020a) as three special cases. Additionally, we derive the relationship between cross-elasticities and the relative positions of two alternatives on the graph, revealing the spatial-aware substitution patterns of the GNN-DCMs. Using the sample of 3,838 households' residential location choices among the 77 community areas in Chicago, our GNN-DCMs outperform MNL, SCL, and ANN benchmarks on predictive performance. This case study also demonstrates the model's ability to capture individual heterogeneity and spatial-aware substitution patterns.

This study seeks to advance the methodological frontier of discrete choice modeling by introducing a GNN framework to capture complex spatial correlation among alternatives, such as housing location choices. The key contributions are as follows:

- Proposing a GNN-DCM framework that handles spatially correlated alternatives through message passing, scales efficiently with large alternative sets, and captures complex non-linear interactions in individual and alternative attributes.
- Establishing theoretical connections between GNN-DCMs and classical DCMs, illustrating that the GNN-DCM framework generalizes both traditional DCM and existing ANN models.
- Demonstrating GNN-DCMs' strong predictive performance, spatially-aware substitution patterns, and the ability to reveal non-linearity and individual heterogeneity in choice behavior.

The remainder of this paper is organized as follows. Section 2 reviews relevant literature on spatially correlated discrete choice models, including both classical and neural network-based approaches. Section 3 presents the GNN-DCM framework, detailing its

structure and theoretical properties. Section 4 applies the framework to a residential location choice problem in Chicago and compares its performance to baseline models. Finally, Section 6 concludes the paper and outlines future research directions.

2. Literature review

2.1. Classical discrete choice models

The Multinomial Logit (MNL) model (McFadden, 1974) has been widely applied in empirical studies to examine the factors influencing residential location choice (Weisbrod et al., 1980; Guevara and Ben-Akiva, 2006; Hu and Wang, 2019). Typically, the MNL model assumes that the random utility components are independent and identically distributed (IID) across alternatives, thus ignoring potential spatial correlations in the choice set. To address this limitation, a range of models have been proposed for modeling location choice, including the Multinomial Probit (MNP) model (Bolduc and Ben-Akiva, 1991), mixed Logit models (Bhat and Guo, 2004), and Generalized Extreme Value (GEV) models (McFadden, 1978). Probit and mixed logit models can capture spatial correlations with the correlated error structure or utility components. However, they often require computationally intensive numerical integration methods for estimation, which can be challenging with large alternative sets. In contrast, GEV models are particularly attractive because many of them offer closed-form choice probabilities and scale well with a large set of correlated alternatives.

GEV represents a broad family of models, which capture the dependency of choice alternatives through a joint generalized extreme value distribution over the unobserved utility components (Train, 2009). The NL model (Ben-Akiva, 1973; McFadden, 1978; Lee and Waddell, 2010) is the very first member in the GEV family, and has been used for residential location choice (illustrated on the left side of Fig. 1b). Building upon the generalized nested model proposed by Wen and Koppelman (2001), Bhat and Guo (2004) introduced the Spatially Correlated Logit (SCL) model for residential location choice, where two alternatives are grouped into the same nest if they are spatially adjacent (as shown in Fig. 1c, left panel). The authors also developed a mixed SCL model to capture individual-level heterogeneity. Further extensions include Sener et al. (2011), who generalized the SCL model by parameterizing the nest allocation parameters using an MNL structure, allowing for more flexible and data-driven correlation patterns. Recently, Perez-Lopez et al. (2022) combined the nested logit model with the SCL model, capturing the spatial correlations for alternatives in the same nest. In short, the GEV model family, such as NL and SCL models, are computationally efficient approaches for capturing spatial dependency among choice alternatives.

Our study focuses on using deep learning to capture the spatial dependence across the discrete alternatives, as tackled by the classical NL and SCL models (McFadden, 1978; Bhat and Guo, 2004; Sener et al., 2011). In this case, spatial dependence manifests as correlation across alternatives, rather than across individuals or independent variables. This dependence of choice alternatives poses a challenge categorically different from that of individuals. For example, prior studies on dependencies among individuals analyze transportation mode choice assuming that individuals residing nearby prefer similar transportation modes (Bhat, 2000; Dugundji and Walker, 2005; Goetzke, 2008) or land use categories (Bina et al., 2006; Wang et al., 2012). In such cases, spatial dependence may arise from spatial spillover effects–where spatial factors (e.g., multiple transit stops) influence one another (LeSage and Pace, 2009)–or from self-selection effects, where individuals with similar preferences (e.g., active travel) or social connections tend to cluster geographically (Bhat and Guo, 2007; Bhat and Eluru, 2009). Nonetheless, our study does not seek to capture the dependence among individuals, as it has been addressed broadly by social network literature.

2.2. Neural network approaches

Researchers have adopted artificial neural networks (ANNs) to tackle discrete choice modeling tasks, outperforming the classical models by capturing complex relationship between individuals and alternatives. In these ANNs, the input layer consists of attributes of both individuals and alternatives, while the output layer produces utility values. These utilities are then transformed into choice probabilities using the softmax function, the same as the MNL formulation. Because of their ability to model complex non-linear relationships between features, ANNs often outperform traditional discrete choice models in predictive accuracy (Wang et al., 2024). ANNs have been applied to discrete choice problems since the 1990s (Kumar et al., 1995; Agrawal and Schorling, 1996). Although early applications were relatively straightforward applications of ANNs, recent research interests are increasingly growing in terms of synergizing ANNs and classical DCMs. For example, the efforts include extracting economic information from ANNs (Wang et al., 2020b), enhancing ANNs' interpretability (Feng et al., 2024; Wang et al., 2020a), or using ANNs to capture individual heterogeneity.

Despite these advances, it remains underexplored how to capture the correlations among alternatives within the deep learning framework. For instance, Wang et al. (2020a) employed alternative-specific utility functions, which only preserve the independence among alternatives without addressing the alternative dependence. Wong and Farooq (2021) proposed ResLogit, where the observed utility is defined as the sum of an alternative-specific MNL component and a cross-alternative ANN component, while the alternative structure is not studied. Sifringer et al. (2020) used ANNs as a part of the utility in a NL model–using the classical GEV model as a backbone to model alternative correlations. These methods for capturing alternative correlations primarily rely on feeding the attributes of all alternatives into the network. Such approaches do not leverage prior knowledge of the alternative correlation, thus exhibiting low inductive bias and increasing the risk of overfitting, especially in large choice sets.

It is possible to analyze the alternative correlation through a network perspective and GNN models (Hamilton et al., 2017; Kipf and Welling, 2017; Veličković et al., 2018), although such a possibility has rarely been explored yet. GNNs can capture network dependence through the graph-based message passing algorithm. They have demonstrated strong performance in a variety of domains

and begin to be applied to DCMs. Tomlinson and Benson (2024) proposed a model with individual-specific coefficients, as an analogy to mixed logit models in the utility function, with a graph regularizer encouraging similarity among socially connected individuals.

The most relevant work to our approach is by Villarraga and Daziano (2025), who used Graph Convolutional Networks (GCNs) (Kipf and Welling, 2017) to capture social network effects. Their approach seeks to incorporate individual correlations, rather than the alternative correlation as in our study. This fundamental distinction leads to several differences in model specification and interpretation. In our case, the graph structure on the alternative set is fixed, and the entire graph is used directly during training. By contrast, social network usually involves a much larger graph than the choice set, and the GNN training often relies on subgraph sampling or semi-supervised learning to handle large-scale networks. Interpretation of the learned parameters also differs substantially. The learned weights or attention coefficients in our GNN encode substitution patterns and dependencies among alternatives, whereas in social-network GNNs, these weights on the graph reflect the influence or similarity between individuals. To the best of our knowledge, our work is the first to apply GNNs to the DCM tasks by structurally analyzing a large number of spatially correlated alternatives, while retaining the economic interpretation by analyzing the substitution patterns and elasticities.

3. Methodology: GNN-DCM framework

This section proposes a GNN-DCM framework for residential location choice with spatial correlations. To capture dependence among alternatives, we introduce the concept of an alternative graph, which enables the integration of GNNs into discrete choice modeling through a network approach. We demonstrate that this GNN-DCM framework contains certain GEV and ANN models as specific cases. Finally, we derive the elasticities of GNN-DCMs, which help interpret the model's predictions and provide insights into spatially-aware substitution patterns.

3.1. Alternative graph

Alternative graph enables us to capture the alternative dependence through GNNs. Specifically, this alternative graph is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents the set of choice alternatives, (e.g., residential location), and $\mathcal{E}$ denotes the set of edges that define the relationships among these alternatives. If two alternatives are independent, no edge exists between them; otherwise, they are connected by an edge. The edge set $\mathcal{E}$ is represented by an adjacency matrix $\mathbf{A}_{|\mathcal{V}|\times|\mathcal{V}|}$, where $\mathbf{A}[i,j] = 1$ if $(i,j) \in \mathcal{E}$ and $\mathbf{A}[i,j] = 0$ otherwise. For residential location choice, edges can be defined based on geographical adjacency. For example, residential location alternatives in Fig. 1a can be represented as an alternative graph with a structure shown on the right side of Fig. 1c. Currently, the alternative graph incorporates only undirected edges to represent these connections, which ensures a symmetric adjacency matrix.

Let $\mathbf{x}_{ni}$ denote the vector of explanatory variables of alternative i and individual n. In the context of residential location choice, $\mathbf{x}_{ni}$ encompasses a set of individual- and alternative-specific attributes, such as household income, housing value, and distance to work. Consequently, when an individual n makes the choice, each node in the alternative graph is associated with an attribute vector specific to that individual.

3.2. Utility specification using GNN and alternative graphs

Recall the random utility maximization (RUM) framework, where the utility of alternative i for individual n contains a deterministic component V_{ni} and a random component ε_{ni}, i.e., $U_{ni} = V_{ni} + \varepsilon_{ni}$. Under the assumption that individuals choose the alternative with the highest utility, the probability that individual n selects alternative i is given by

$$P_{ni} = P(U_{ni} > U_{nj}, \forall j \neq i).$$

When assuming the random components ε_{ni} follow independently and identically distributed (i.i.d.) Type I Extreme Value (Gumbel) distribution. This assumption yields a closed-form expression for the choice probability:

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j\in\mathcal{V}} e^{V_{nj}}}, \tag{1}$$

which is equivalent to applying the softmax function to the deterministic utilities. In the MNL and many ANN-based choice models (e.g. Wang et al., 2020a) that rely on the i.i.d. error assumption, the utility term[1] is specified as a function only based on the alternative itself, i.e., $V_{ni} = f_i(\mathbf{x}_{ni})$, without considering the dependence between alternatives. However, in spatial choice problems, the deterministic component V_{ni} may exhibit spatial dependencies across alternatives.

Hence we propose GNNs to specify the utility function, where the utility of each alternative is determined not only by its own attributes but also by those of its neighbors in a predefined alternative graph $\mathcal{G}$. This approach absorbs spatial dependencies into the deterministic part of the utility function, analogous to spatial econometric models (LeSage and Pace, 2009). Specifically, in a GNN model with K_g layers, the utility of an alternative i is computed as:

$$V_{ni} = \mathbf{w}^\top \mathbf{h}_{ni}^{(K_g)}, \tag{2}$$

[1] Unless otherwise specified, the term "utility" refers to the deterministic component V_{ni}.

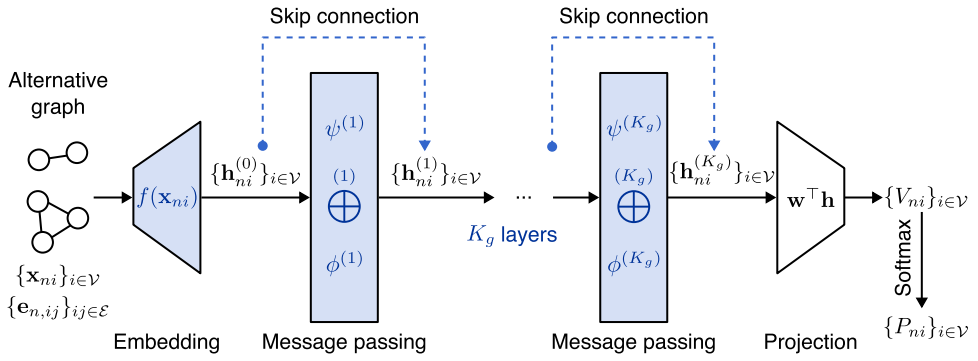


Fig. 2. Illustration of the GNN-DCM framework. Customizable design components are colored in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\mathbf{h}_{ni}^{(k)} = \phi^{(k)}\left(\mathbf{h}_{ni}^{(k-1)}, \bigoplus_{j\in\mathcal{N}(i)}^{(k)} \psi^{(k)}\left(\mathbf{h}_{ni}^{(k-1)}, \mathbf{h}_{nj}^{(k-1)}, \mathbf{e}_{n,ij}\right)\right) \forall k \in \{1,\dots,K_g\}, \tag{3}$$

where $\mathbf{h}_{ni}^{(k)} \in \mathbb{R}^{h_k}$ is a vector for the hidden representation of node i in the kth layer of the GNN, with initial representation $\mathbf{h}_{ni}^{(0)} \equiv \mathbf{x}_{ni}$ or using an input embedding function $\mathbf{h}_{ni}^{(0)} = f(\mathbf{x}_{ni})$. The set $\mathcal{N}(i)$ denotes the direct neighbors of node i in the alternative graph $\mathcal{G}$. Each layer of the GNN contain a "message-passing" process, where the hidden representation of each node is updated based on its previous representation and the aggregated messages from its neighbors: First, the *message* function ψ encodes the information of node i and its neighbor j, along with possible edge features $\mathbf{e}_{n,ij}$ that may represent the weight between nodes i and j. Next, the *aggregation* function $\bigoplus$ is a permutation-invariant operation over the neighbor set $\mathcal{N}(i)$, such as element-wise summation, averaging, or maximum function, which aggregates the messages received from all neighboring nodes of alternative i. Finally, the *update* function ϕ updates the node representation based on the aggregated message. Several common GNN update functions are provided in Appendix A.2. After reaching the last GNN layer, the utility is obtained by projecting node representation $\mathbf{h}_{ni}^{(K_g)}$ onto a scalar via a learnable weight vector $\mathbf{w}$ as shown in Eq. (2). This utility is then used in Eq. (1) to compute the choice probability.

Fig. 2 illustrates the GNN-DCM framework. Given the extensive research on GNNs, this framework is highly flexible and can adapt to a wide range of specifications. Such flexibility arises from choices regarding the embedding function, number of layers, message, aggregation, and update functions. The customizable design components in Fig. 2 are colored in blue. For instance, the aggregation function can be a simple sum, mean, or more complicated attention-based sum (Veličković et al., 2018), or any other permutation-invariant operation. In addition, Fig. 2 highlights the use of optional skip connections, which are not part of the standard message-passing formulation in Eq. (3) but commonly employed in deep GNNs to alleviate the over-smoothing problem (Hamilton et al., 2017). These skip connections allow information to bypass intermediate layers, facilitating more effective gradient flow from earlier to later layers. The implementation details of the skip connection are elaborated in Appendix A.3.

For additional techniques in GNN design, such as regularization strategies and multi-relational graphs, we refer readers to You et al. (2020), Hamilton (2020). In the next subsection, we demonstrate that the GNN-DCM utility functions generalize classical DCMs and various ANN-based DCMs.

3.3. GNN-DCM generalizes classical logit models

The NL model by McFadden (1978) is the most common member in the GEV family, which captures the correlation between alternatives assuming random components following a GEV distribution. The SCL model by Bhat and Guo (2004) is a variant of the NL model tailored for residential location choice, where the nests are formed based on spatial adjacency of alternatives. We examine how NL and SCL models are special cases in the GNN-DCM framework, showing the connections between the nested structure and the alternative graph.

3.3.1. NL is a specific case in GNN-DCMs

Proposition 1. *A two-level nested logit model, where each alternative belongs to a single nest, is a single-layer GNN. In this representation, each nest corresponds to a complete subgraph with a self-loop at each node, and there is no edge between nests.*

Proof. In a NL model, denoted by K_n as the total number of nests, and μ_k as nest-specific independence parameter for nest B_k, the choice probability of an alternative i in nest B_k is typically given by:

$$\begin{aligned} P_{ni} &= P_{ni|B_k} P_{nB_k} \\ &= \frac{\exp\left(V_{ni}/\mu_k\right)}{\sum_{j\in B_k}\exp\left(V_{nj}/\mu_k\right)} \times \frac{\left(\sum_{j\in B_k}\exp\left(V_{nj}/\mu_k\right)\right)^{\mu_k}}{\sum_{l=1}^{K_n}\left(\sum_{j\in B_l}\exp\left(V_{nj}/\mu_l\right)\right)^{\mu_l}} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\exp\left(V_{ni}/\mu_k\right)\left(\sum_{j\in B_k}\exp\left(V_{nj}/\mu_k\right)\right)^{\mu_k-1}}{\sum_{l=1}^{K_n}\left(\sum_{j\in B_l}\exp\left(V_{nj}/\mu_l\right)\right)^{\mu_l}} \\
&= \frac{\exp\left(V_{ni}/\mu_k\right)\left(\sum_{j\in B_k}\exp\left(V_{nj}/\mu_k\right)\right)^{\mu_k-1}}{\sum_{l=1}^{K_n}\left(\frac{\sum_{m\in B_l}\exp\left(V_{nm}/\mu_l\right)}{\sum_{m\in B_l}\exp\left(V_{nm}/\mu_l\right)}\left(\sum_{j\in B_l}\exp\left(V_{nj}/\mu_l\right)\right)^{\mu_l}\right)} \\
&= \frac{\exp\left(V_{ni}/\mu_k\right)\left(\sum_{j\in B_k}\exp\left(V_{nj}/\mu_k\right)\right)^{\mu_k-1}}{\sum_{l=1}^{K_n}\sum_{m\in B_l}\left(\frac{\exp\left(V_{nm}/\mu_l\right)}{\sum_{m\in B_l}\exp\left(V_{nm}/\mu_l\right)}\left(\sum_{j\in B_l}\exp\left(V_{nj}/\mu_l\right)\right)^{\mu_l}\right)} \\
&= \frac{\exp\left(V_{ni}/\mu_k\right)\left(\sum_{j\in B_k}\exp\left(V_{nj}/\mu_k\right)\right)^{\mu_k-1}}{\sum_{l=1}^{K_n}\sum_{m\in B_l}\left(\exp\left(V_{nm}/\mu_l\right)\left(\sum_{j\in B_l}\exp\left(V_{nj}/\mu_l\right)\right)^{\mu_l-1}\right)} \\
&= \frac{\exp\left(V_{ni}/\mu_k+(\mu_k-1)\log\left(\sum_{j\in B_k}\exp\left(V_{nj}/\mu_k\right)\right)\right)}{\sum_{l=1}^{K_n}\sum_{m\in B_l}\exp\left(V_{nm}/\mu_l+(\mu_l-1)\log\left(\sum_{j\in B_l}\exp\left(V_{nj}/\mu_l\right)\right)\right)}.
\end{aligned}
\tag{4}
$$

Eq. (4) follows the MNL form given in Eq. (1), and the exponential terms in both the numerator and the denominator have the same functional form. In comparison with Eq. (3), we can find that the exponential terms follow the GNN message passing scheme:

$$
V_{ni}^{(1)} = \underbrace{V_{ni}^{(0)}/\mu_k + (\mu_k-1)\overbrace{\log\left(\sum_{j\in\mathcal{N}(i)\cup\{i\}}\exp\left(V_{nj}^{(0)}/\mu_k\right)\right)}^{\text{Aggregate }\bigoplus}}_{\text{Update }\phi}. \tag{5}
$$

Eq. (5) corresponds to a single-layer GNN message passing ($K_g = 1$) with scalar node representations, where the hidden representation of node i is given by a scalar $\mathbf{h}_{ni}^{(k)} = V_{ni}^{(k)}$, the message function is a scaled utility $\psi = V_{nj}^{(0)}/\mu_k$, the aggregate function $\bigoplus$ takes a Log-Sum-Exponential (LSE) form, and the update function ϕ computes a linear combination of the node's own utility and the aggregated message, with weights determined by the nest-specific parameter μ_k.

The above equivalence indicates that the following properties are needed to turn a GNN into a NL model. (1) Self-loops: Message for each nest is aggregated from the set $i \in \mathcal{N}(i)\cup\{i\}$, meaning that each node not only receives messages from its neighbors but also from itself via a self-loop. (2) Complete subgraphs within nests: All alternatives within a nest share the same message with each other, forming a complete subgraph for each nest. (3) No inter-nest edges: Eq. (4) shows that the message for each alternative is only aggregated from its own nest. Thus, there are no edges between alternatives in different nests. (4) Nest-specific parameter: The parameter μ_k in the nested logit model quantifies the degree of independence among alternatives within a nest. □

An example in Fig. 3 illustrates the equivalence between an NL model and a single-layer GNN. The toy choice set contains five alternatives grouped into two nests: $B_1 = \{1,2,3\}$ and $B_2 = \{4,5\}$ with nest-specific independence parameters $\mu_1 = 0.6$ and $\mu_2 = 0.5$. In the equivalent GNN representation (right panel), every pair of alternatives within the same nest is linked. Each alternative is annotated with a housing price, and the observed utility is defined as $\text{Utility}_i = -\text{Price}_i$. When using the UPDATE rule from Eq. (5), the GNN produces exactly the same choice probability for alternative 1, $P_1 = 0.6959$, as the nested logit model. The identical probabilities apply for all other alternatives, thereby confirming Proposition 1.

3.3.2. SCL is a specific case in GNN-DCMs

SCL model (Bhat and Guo, 2004) is a type of NL model tailored for residential location choice, where each nest is a pair of adjacent alternatives (see its structure from the left panel of Fig. 1c). The choice probability of the SCL model is:

$$
P_{ni} = \frac{\sum_{j\neq i}\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}\left[\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}+\left(\alpha_{j,ij}e^{V_{nj}}\right)^{1/\mu}\right]^{\mu-1}}{\sum_{k=1}^{|\mathcal{V}|-1}\sum_{l=k+1}^{|\mathcal{V}|}\left[\left(\alpha_{k,kl}e^{V_{nk}}\right)^{1/\mu}+\left(\alpha_{l,kl}e^{V_{nl}}\right)^{1/\mu}\right]^{\mu}}, \tag{6}
$$

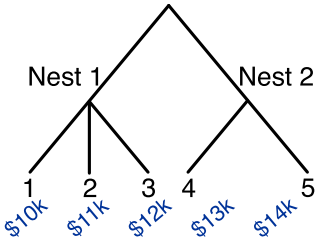
where $\alpha_{i,ij}$ is a parameter representing the allocation of alternative i to the paired nest with alternatives i and j, there is $0 < \alpha_{i,ij} < 1$ and $\sum_j \alpha_{i,ij} = 1$; μ is the dissimilarity parameter with $0 < \mu \leq 1$, when $\mu = 1$, the SCL model reduces to the MNL model.

Proposition 2. *A SCL model is a single-layer GNN, where each nest with a pair of alternatives is an edge in the graph. The message passing scheme takes the following form:*

$$
V_{ni}^{(1)} = \log\left(\sum_{j\in\mathcal{N}(i)}\exp\left(V_{ni}^{(0)}+(\mu-1)\log\left[\exp(V_{ni}^{(0)})+\exp(V_{nj}^{(0)})\right]\right)\right), \tag{7}
$$

where $V_{ni}^{(0)} = f(\mathbf{x}_{ni}) = \frac{\mathbf{b}^\top\mathbf{x}_{ni}}{\mu} + \frac{1}{\mu}\log\alpha_{i,ij}$ *when using a linear utility function.*

The nest structure of alternatives:



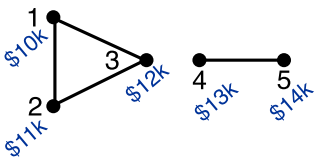
The nested logit view to calculate P_1:

$$P_{B_1} = \frac{\left(e^{\frac{-13}{0.5}} + e^{\frac{-14}{0.5}}\right)^{0.5}}{\left(e^{\frac{-10}{0.6}} + e^{\frac{-11}{0.6}} + e^{\frac{-12}{0.6}}\right)^{0.6} + \left(e^{\frac{-13}{0.5}} + e^{\frac{-14}{0.5}}\right)^{0.5}} = 0.9523$$

$$P_{1|B_1} = \frac{e^{\frac{-13}{0.5}}}{e^{\frac{-13}{0.5}} + e^{\frac{-14}{0.5}}} = 0.7307$$

$$P_1 = P_{1|B_1} P_{B_1} = 0.7307 \times 0.9523 = 0.6959$$

The equivalent alternative graph:



The GNN view to calculate P_1:

$$V_1 = \frac{-10}{0.6} + (0.6-1)\log(e^{\frac{-10}{0.6}} + e^{\frac{-11}{0.6}} + e^{\frac{-12}{0.6}}) = -10.063$$
$$V_2 = \frac{-11}{0.6} + (0.6-1)\log(e^{\frac{-10}{0.6}} + e^{\frac{-11}{0.6}} + e^{\frac{-12}{0.6}}) = -11.313$$
$$V_3 = \frac{-12}{0.6} + (0.6-1)\log(e^{\frac{-10}{0.6}} + e^{\frac{-11}{0.6}} + e^{\frac{-12}{0.6}}) = -12.563$$
$$V_4 = \frac{-13}{0.5} + (0.5-1)\log(e^{\frac{-13}{0.5}} + e^{\frac{-14}{0.5}}) = -13.028$$
$$V_5 = \frac{-14}{0.5} + (0.5-1)\log(e^{\frac{-13}{0.5}} + e^{\frac{-14}{0.5}}) = -14.140$$
$$P_1 = \frac{e^{V_1}}{e^{V_1} + e^{V_2} + e^{V_3} + e^{V_4} + e^{V_5}} = 0.6959$$

Fig. 3. Using GNN as the NL model to calculate the choice probability for alternative 1, where the observed utility is defined as $\text{Utility}_i = -\text{Price}_i$; $\mu_1 = 0.6$, $\mu_2 = 0.5$. The two approaches yield the same result.

The proof of Proposition 2 is shown in Appendix A.1. Eq. (7) presents a single-layer GNN, which can be decomposed into a three-step message passing algorithm:

1. **Node embedding.** The node utility value is initialized as $V_{ni}^{(0)} = f(\mathbf{x}_{ni}) = \frac{\mathbf{b}^\top \mathbf{x}_{ni}}{\mu} + \frac{1}{\mu}\log \alpha_{i,ij}$, where $\mathbf{b}$ is a vector of coefficients.
2. **Message calculation.** The message of a neighboring node j to node i is computed by $M_{i,j} = \psi(V_{ni}^{(0)}, V_{nj}^{(0)}) = V_{ni}^{(0)} + (\mu - 1)\log[\exp(V_{ni}^{(0)}) + \exp(V_{nj}^{(0)})]$.
3. **Node aggregation.** Messages from all neighbors of a node i are aggregated by a LSE function: $V_{ni}^{(1)} = \bigoplus(\{M_{i,j}, j \in \mathcal{N}(i)\}) = \log \sum_{j\in\mathcal{N}(i)} \exp(M_{i,j})$.

While here we demonstrate a GNN perspective into SCL models, the message passing algorithm also highlights the unique characteristics of the SCL model, as opposed to the standard GNN family. (1) The passed message in SCL models is a one-dimensional utility value with mostly a linear form. While the initialization function $f(\cdot)$ in standard GNN is often a neural network with a high-dimensional mapping. (2) The LSE aggregation is rarely used in standard GNNs, while it is the standard form for SCL and NL models. (3) The SCL model corresponds to a one-layer GNN since it aggregates the information from the one-hop neighbors of the targeting node i. This one-layer GNN is also a unique choice in SCL models because a typical GNN model would adopt around two to three layers to incorporate the information from multi-hop neighborhoods. (4) Lastly, the SCL model shares the parameter μ across its whole algorithm, as opposed to the standard nested logit model, where each nest has its own μ_k.

3.4. GNN-DCMs generalize ANN-based choice models

Beyond the classical NL family, the GNN-DCM framework also generalizes a broad class of ANN-based discrete choice models. This generalization depends on how the attributes of different alternatives are used to construct the utility function of the target alternative. For example, the GNN-DCM framework incorporates the alternative-specific utility functions (ASU-DNN) (Wang et al., 2020a) as a specific case:

Proposition 3. *The GNN-DCM reduces to the ASU-DNN model (Wang et al., 2020a) when the embedding function $f(\cdot)$ is a multilayer perceptron (MLP) as in a feedforward neural network, and without message passing, i.e., $\phi(\cdot) = \emptyset, \psi(\cdot) = \emptyset$.*

Essentially, the ASU-DNN model corresponds to a zero-layer GNN, where no information is exchanged between alternatives through message passing over the spatial network. Similarly, a wide range of ANN-based DCMs can be interpreted as special cases of the GNN-DCM framework without message passing, e.g., the TasteNet by Han et al. (2022).

The ANN-based DCMs typically capture the alternative interactions in two ways: (1) incorporating the attributes of all alternatives as joint inputs into the ANN-based utility function (e.g., Wang et al., 2020b; Wong and Farooq, 2021); (2) embedding neural networks within the utility function structure of classical logit models (e.g., Sifringer et al., 2020). These two approaches can both be interpreted within the GNN-DCM framework. The first corresponds to a fully connected GNN, where each alternative is connected to all others, and the message passing functions are implemented using neural networks. The second corresponds to a GNN with neural networks

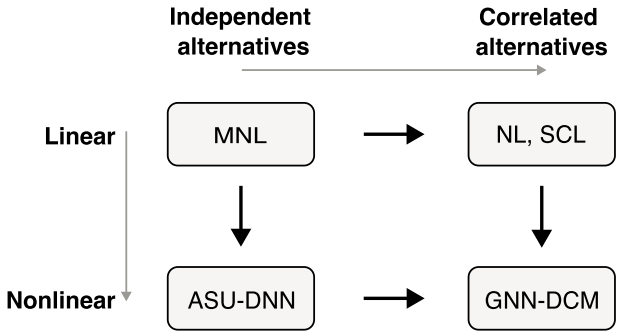


Fig. 4. The relationship between classical and NN-based discrete choice models.

as the input embedding function followed by a message passing scheme defined by Eq. (5). Therefore, the GNN-DCM framework unifies both modeling strategies under a single, graph-structured formulation.

Fig. 4 summarizes the key relationships between the GNN-DCM, existing DCMs, and ANNs. The second row of the figure represents the early development of DCMs, from the MNL model with independent alternatives to the NL model for hierarchically correlated alternatives, and both models use linear utility functions. The second column illustrates recent advances in deep learning-based models that introduce nonlinearity into utility specification. The GNN-DCM framework bridges these two dimensions: it generalizes ANN-based models by enabling flexible alternatives' correlation through message passing, and it extends classical models like the NL and SCL models by allowing richer forms of utility specification via vectorized representations, more expressive message passing scheme, and flexible network designs.

3.5. *Interpreting GNNs for travel demand modeling*

Previous studies have demonstrated that deep neural networks can provide full economic information for discrete choice models (Wang et al., 2020b). The GNN-DCM framework is no exception. Through numerical methods, we can compute the elasticity, substitution patterns, and visualize the choice probability functions for GNNs. Particularly, substitution patterns describe how the change in an alternative's attributes (e.g., the housing price) affects the choice probabilities of other alternatives. The MNL model is well known for its proportionate substitution pattern: a change in the utility of an alternative proportionally affects the choice probabilities of other alternatives. This behavior is a direct consequence of the IIA property, meaning that the relative choice probabilities between two alternatives are unaffected by the presence or attributes of other alternatives. The IIA property has been criticized for being unrealistic in some applications. For example, in residential location choice models, an increase in housing prices in one community is expected to impact neighboring communities more than distant ones – the MNL model fails to capture this. We will see that GNN-DCMs mitigate this constraint imposed by the IIA.

Substitution patterns of a choice model can be measured by elasticities $E_{i\,z_{nj}}$, which is defined by the percentage change in the choice probability of individual n on alternative i in response to a percentage change in the attribute of alternative j:

$$E_{i\,z_{nj}} = \frac{\partial P_{ni}}{\partial z_{nj}} \frac{z_{nj}}{P_{ni}},$$

where z_{nj} is the attribute of alternative j that is being changed. When $i = j$, it is called direct elasticity; when $i \neq j$, it is called cross-elasticity. The cross-elasticities of the MNL model are the same for all $i \neq j$.

We investigate the substitution patterns of GNN-DCMs by deriving their elasticities. Eqs. (2) and (3) show that V_j is a function of the attributes of alternative j and its k_g-hop neighbors, $\mathcal{N}_{k_g}(j)$, but not of other alternatives. This leads to the following results:

- For an alternative i that is not a k_g-hop neighbors of j, i.e. $i \notin \mathcal{N}_{k_g}(j) \cup \{j\}$, the cross-elasticity in response to the change in the attribute z_{nj} of alternative j is:

$$\begin{aligned}
E_{i\,z_{nj}} &= \frac{\partial\left(e^{V_{ni}} / \sum_{k\in\mathcal{V}} e^{V_{nk}}\right)}{\partial z_{nj}} \frac{z_{nj}}{P_{ni}} \\
&= \frac{-e^{V_{ni}}\left(\sum_{k\in\mathcal{N}_{k_g}(j)\cup j} e^{V_{nk}} \frac{\partial V_{nk}}{\partial z_{nj}}\right)}{\left(\sum_{k\in\mathcal{V}} e^{V_{nk}}\right)^2} \frac{z_{nj}}{P_{ni}} \\
&= -P_{ni} \sum_{k\in\mathcal{N}_{k_g}(j)\cup j} P_{nk} \frac{\partial V_{nk}}{\partial z_{nj}} \frac{z_{nj}}{P_{ni}} \\
&= -z_{nj} \sum_{k\in\mathcal{N}_{k_g}(j)\cup j} P_{nk} \frac{\partial V_{nk}}{\partial z_{nj}},
\end{aligned}$$

Table 1
Dataset in residential location choice studies.

Paper	Methods	City	# Zones	# Households
Bhat and Guo (2004)	SCL	Dallas	98	236
Sener et al. (2011)	GSCL	San Francisco	115	702
Perez-Lopez et al. (2022)	SCNL	Santander (Spain)	26	534
This work	GNN	Chicago	77	3838

which is irrelevant to the alternative i. This means that the cross-elasticities are the same for all alternatives outside of the k_g-hop neighbors of j, the same as the proportionate substitution pattern of the MNL model.

- For an alternative j belongs to the k_g-hop neighbors of j, i.e., $i \in \mathcal{N}_{k_g}(j) \cup \{j\}$, the (cross-)elasticity takes the form

$$
\begin{aligned}
E_{i\,z_{nj}} &= \frac{e^{V_{ni}} \frac{\partial e^{V_{ni}}}{\partial z_{nj}} \sum_{k\in\mathcal{V}} e^{V_{nk}} - e^{V_{ni}} \left(\sum_{k\in\mathcal{N}_{k_g}(j)\cup j} e^{V_{nk}} \frac{\partial V_{nk}}{\partial z_{nj}} \right)}{\left(\sum_{k\in\mathcal{V}} e^{V_{nk}} \right)^2} \frac{z_{nj}}{P_{ni}} \\
&= \left(P_{ni} \frac{\partial e^{V_{ni}}}{\partial z_{nj}} - P_{ni} \sum_{k\in\mathcal{N}_{k_g}(j)\cup j} P_{nk} \frac{\partial V_{nk}}{\partial z_{nj}} \right) \frac{z_{nj}}{P_{ni}} \\
&= \left(\frac{\partial e^{V_{ni}}}{\partial z_{nj}} - \sum_{k\in\mathcal{N}_{k_g}(j)\cup j} P_{nk} \frac{\partial V_{nk}}{\partial z_{nj}} \right) z_{nj},
\end{aligned}
$$

which varies across alternatives i. This means that the cross-elasticities of alternatives within the k_g-hop neighbors of j are alternative-specific and depend on the attributes of i, j and its k_g-hop neighbors. This is a key difference from the MNL model.

These results demonstrate that GNN-DCMs allow for more flexible substitution patterns, where a change in an alternative's attributes has dedicated localized effects on k_g-hop neighbors, rather than globally proportionate as in the MNL model. Since NL and SCL models are specific one-layer GNN-DCMs, deeper GNN models also present more flexible elasticities than NL and SCL models. By choosing the number of layers (i.e., k_g) and constructing different graph topologies, we can control the behavior of substitution patterns between alternatives. We name it spatially-aware substitution patterns, since the substitution patterns depend on the spatial relationships defined by the graph structure. An illustration of this behavior is shown in Section 5.4.

4. Experiment setup

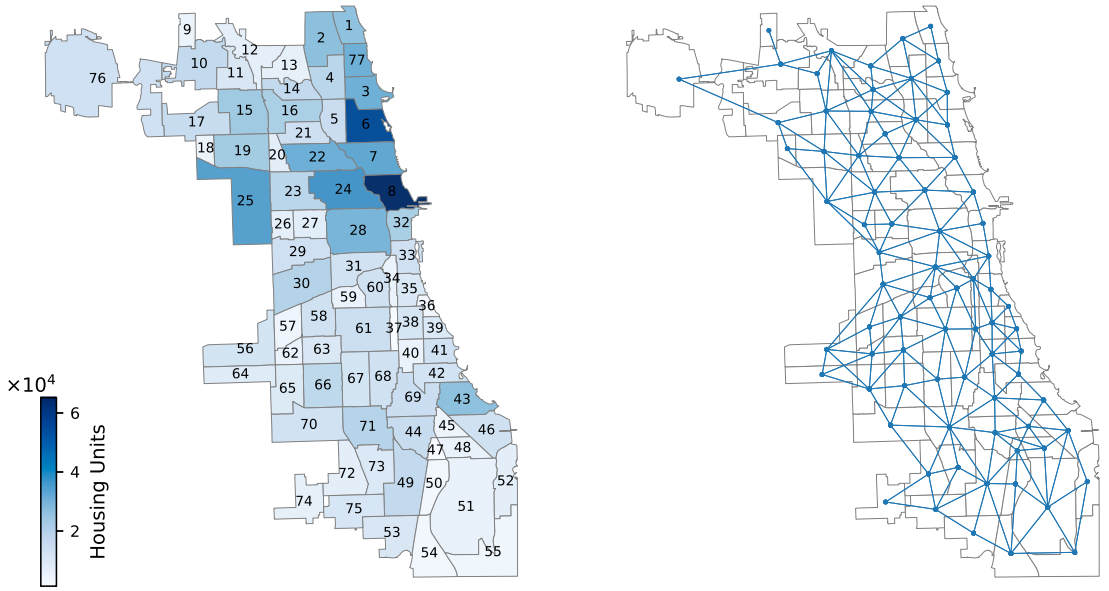
4.1. Dataset

We perform a case study on residential location choice across 77 communities in Chicago. As of the 2020 census, the study area has a population of approximately 2.7 million. The locations of these communities and the number of housing units in each are illustrated in Fig. 5a. The 77 community areas were originally defined in the 1920s by the University of Chicago Social Science Research Committee and were intended to represent moderately coherent social characteristics across urban space at a generalized geographical scale. Although the geography and population of these communities have evolved over time, their boundaries have remained relatively stable. These community areas continue to provide a consistent framework for urban planning, statistical analysis, and policy-making (Chicago Historical Society, 2004). The Chicago Metropolitan Agency for Planning (CMAP) and other agencies have adopted these community areas as a standard for reporting demographic and socioeconomic data.

The goal of this study is to investigate how households select their residential locations within these communities. The social and demographic characteristics of the communities are derived from census data obtained from the American Community Survey 5-year data (2013–2017)[2] and land use data (2017) from the CMAP Data Hub.[3] Household-level information, such as home location and household income, is sourced from the My Daily Travel Survey data, conducted by CMAP (released in 2019 with data collected in 2017). The travel survey captures the travel patterns of over 12,000 households in the seven-county region of northeastern Illinois. For this study, we extracted a subset of 3838 households residing in these 77 communities who reported their work locations. Although the sample size is relatively small compared to the total population of these communities, it serves as a convenient dataset for our analysis and is sufficiently large when compared to similar studies in the literature, as indicated in Table 1.

After reviewing the literature (Bhat and Guo, 2004; Sener et al., 2011; Schirmer et al., 2014), considering data availability, and conducting significance tests in the MNL model, we selected thirteen attributes for modeling household residential choice behavior, as detailed in Table 2. These variables are categorized into four groups: housing, land use, transportation, and demographics. All attributes are scaled to ensure comparable magnitudes across variables. Below are additional explanations for some attributes.

[2] https://www.nhgis.org/
[3] https://datahub.cmap.illinois.gov/



(a) The 77 communities in Chicago. (b) The graph structure.

Fig. 5. The study area and the alternative graph.

The land use mixture of a community is calculated using the entropy index: $-\sum_k p_k \log(p_k)/\log(K)$, where p_k is the proportion of land area allocated to land use type k, and K is the total number of land use types considered. We consider 14 land use types according to the CMAP land use dataset. A higher entropy value indicates a more diverse land use mix within the community.

It is common to use distance or travel time to work as an input attribute in residential location choice models (Bhat and Guo, 2004; Sener et al., 2011; Perez-Lopez et al., 2022). However, this approach of using work attributes as inputs may introduce endogeneity bias because the work location itself may be influenced by the chosen residential location. Nonetheless, for simplicity, we treat work locations as exogenous to focus on discussing the prediction, interpretation, and other theoretical potentials of GNN-DCM.

Household-level attributes, such as race and income, do not directly enter the utility function because adding household-level constants to all alternatives does not change the relative utility differences and the choice probabilities in the MNL framework. To incorporate these attributes, we interact household-level attributes with community-level attributes, creating attributes "black interact", "white interact", and "income interact", as shown in Table 2. This approach aligns with common practices in the literature (Bhat and Guo, 2004; Sener et al., 2011; Perez-Lopez et al., 2022).

It is also noteworthy that GNNs can directly utilize household-level attributes as node features, as the projection matrices in Eqs. (2) and (3) inherently interact household-level with community-level attributes. This eliminates the need for manually constructing interactive terms and allows for the inclusion of additional household-level attributes–such as vehicle ownership or the presence of children–that are challenging to incorporate into the MNL model. However, to maintain a fair comparison, we retain the same set of attributes for the GNNs as those used in the MNL model. This also enables us to focus on differences in model structures rather than variations in input features.

Finally, we use spatial adjacency–the most straightforward approach–to construct the graph structure for the 77 communities. Each community is represented as a node in the graph, with edges connecting adjacent communities. The resulting graph structure is illustrated in Fig. 5b.

4.2. Experiment design

The experiments are designed to address the following research questions:

- **RQ1:** How does the performance of GNN-DCM compare to the MNL, SCL, and ASU-DNN models?
- **RQ2:** How do different components of GNN design affect the performance of GNN-DCM?
- **RQ3:** How do GNN-DCMs compare with traditional models in terms of interpretability and elasticities?

To address **RQ1**, we compare the GNN-DCM with the standard MNL model and SCL model (Bhat and Guo, 2004). The MNL model assumes independence of irrelevant alternatives, and the SCL model can be viewed as a single-layer GNN with a specific update function defined in Eq. (7). We also include a comparison with the ASU-DNN model (Wang et al., 2020b), which serves as a benchmark representing GNNs without any message passing between alternatives. Comparative results are reported in Section 5.1.

Table 2
Summary of variables.

Categories	Attributes	Short name
Housing	Number of housing units in the community[a]	# Units
	Median housing value in the community[a]	House value
	Median house age in the community[a]	House age
Land use	Land use mixture measured by entropy	Land mixture
	Percentage of land allocated to single-family housing	% Single house
	Percentage of land allocated to multi-family housing	% Multi house
	Percentage of land allocated to office	% Office
Transportation	Transit accessibility in the community (released by CMAP)[a]	Transit access
	Logarithm of distance to work	Work distance
Demographics	Population density of the community[a]	Pop density
	Percentage of the black population in the community multiplied with the black household dummy variable[b]	Black interact
	Percentage of the white population in the community multiplied with the white household dummy variable[b]	White interact
	The difference between the household income with the community median household income[a]	Income interact

[a] Scaled to have comparable magnitude across attributes.
[b] The black (white) household dummy variable equals 1 if the household identifies as black (white) and 0 otherwise.

To address **RQ2**, we evaluate three representative GNN update methods: Message Passing Neural Network (MPNN), Graph Convolutional Networks (GCN) (Kipf and Welling, 2017), and Graph Attention Network (GAT) (Veličković et al., 2018). For MPNN, we also explore different aggregation functions, including sum, max, mean, and log-sum-exp (LSE), the latter being inspired by the SCL model. We use the same number of hidden dimensions for all layers. Our base configuration is a two-layer GNN with 64 hidden dimensions and skip connections. We then systematically vary the number of layers, hidden dimensions, and the use of skip connections. Details of these GNN update methods are provided in Appendix A.2, details of the skip connection can be found in Appendix A.3, and the experimental results are presented in Section 5.2.

To address **RQ3**, we assess the interpretability of the GNN-DCM by analyzing how predicted choice probabilities respond to changes in alternative attributes. We also compute the cross-elasticities of the GNN-DCM and compare them with those from the MNL and SCL models. The results are presented in Sections 5.3 and 5.4.

4.3. Implementation details

Model parameters are estimated by minimizing the negative log-likelihood (NLL) of the training data:

$$\mathcal{L}(\theta) = -\sum_{n=1}^{N} \log P_{ni},$$

where P_{ni} denotes the predicted choice probability of household n selecting alternative i, and θ represents the model parameters. The MNL and SCL models are trained using the NLL computed over the entire training dataset, where N is the total number of households. In contrast, DNN and GNN models are trained using mini-batch gradient descent, where the NLL is calculated for each batch and used to update parameters iteratively. We use a mini-batch size of 32, a common choice in the deep learning literature (Goodfellow et al., 2016).

To mitigate randomness during training, we conduct ten-fold cross-validation and report the average performance across all folds. Each fold the dataset is partitioned into a training set (90%) and a held-out test set (10%). All models are implemented in PyTorch (Ansel et al., 2024). The GNN models are built using the PyTorch Geometric library (Fey and Lenssen, 2019), which provides efficient implementations of various GNN layers. The MNL and SCL models are optimized using the L-BFGS optimizer (Liu and Nocedal, 1989), while the GNN and ASU-DNN models are trained with the Adam optimizer (Kingma and Ba, 2014) using a learning rate of 0.01. Dropout regularization with a rate of 0.05 is applied after each GNN layer and after each fully connected layer to prevent overfitting. We train all neural network-based models for 20 epochs, which we find sufficient for convergence. Early stopping is not used, as mini-batch training and dropout provide effective regularization against overfitting. It takes around 1 min for training a GNN-DCM at a MacBook Pro-with an M3 Pro-chip and 18GB of RAM. Additional implementation details are available in the code repository.

The GNN-DCM introduces a larger number of parameters than the MNL and SCL models and therefore requires a large number of data for reliable estimation. Nevertheless, its parameterization remains efficient because the parameters in each message-passing layer are shared across all alternatives and locations; consequently, model complexity does not scale with the number of spatial alternatives. In practice, reliable estimation of the GNN-DCM requires (1) sufficient variation in explanatory variables, (2) adequate representation of each alternative in the training data, and (3) appropriate regularization to mitigate overfitting. In this study, overfitting is addressed through mini-batch training and dropout. Empirically, we follow the standard process of training and testing GNN-DCM using separate data points, and a high testing performance indicates effective regularization and generalizability. When sample sizes are limited, practitioners may prefer reduced-capacity GNN architectures (e.g., fewer layers or hidden units). More generally, we recommend validation-based model selection to tune architectural complexity and to assess generalization performance prior to deployment.

Table 3
The average performance of ten-fold cross-validation.

	MNL	SCL	ASU-DNN	GNN (1 layer)	GNN (2 layers)	GNN (3 layers)
Log-likelihood	−1342.05	−1342.05	−1319.00	−1308.47	*−1306.68*	**−1309.41**
Accuracy	12.14%	12.14%	12.71%	12.64%	*13.03%*	**13.26%**
Top-5 accuracy	36.37%	36.37%	37.52%	37.78%	**38.75%**	*38.15%*
Avg. distance [km]	7.38	7.38	7.43	7.01	**6.97**	*7.01*
F1 score	0.033	0.033	0.045	0.065	*0.066*	**0.068**
Mean reciprocal rank	0.250	0.250	0.258	0.279	**0.285**	*0.284*

* In each row, **the best** number is highlighted in bold, and *the second-best* number is underlined and italicized.

Conditioning on a specific data set, it is challenging to examine how sample size influences model performance, but the authors' past work provides some practical guidance regarding the relationship between sample size, effective regularization, and out-of-sample performance (Wang et al., 2021b).

5. Results

We compare the GNN-DCMs with a set of baseline models using a case study of residential location choice in Chicago. This section showcases the performance, interpretability, elasticities, and attention weights of the GNN model. The code and data for this study are available at https://github.com/chengzhanhong/GNN_residential_choice.

5.1. Comparing GNN-DCMs to benchmark models

We first compare the GNN-DCM model with the highest performance to the benchmark models, including MNL, SCL, and ASU-DNN. Based on the results in Section 5.2, we set the base configuration of the GNN-DCM to a two-layer GAT with 64 latent dimensions and skip connections. For a fair comparison, the ASU-DNN uses the same specification, except that it does not have the message passing component. We also present the performance of GNN with one, two, and three layers, thus revealing the impacts of GNN depth. The average performance of the ten-fold cross-validation of the models is summarized in Table 3. Evaluation metrics include log-likelihood, prediction accuracy, top-5 accuracy, the average distance between the centroids of the predicted and actual communities (Avg. distance), F1 score, and mean reciprocal rank (MRR). The best performance is highlighted in bold, and the second-best performance is underlined and italicized.

Both neural network-based models (ASU-DNN and GNNs) outperform the classical MNL and SCL models in terms of log-likelihood, accuracy, top-5 accuracy, F1 score, and MRR, highlighting the enhanced capability of using non-linear utility functions and attribute interactions. This finding validates the past research that overall the machine learning and deep learning models outperform the classical econometrics models. However, the ASU-DNN exhibits the largest average distance between predicted and actual communities, even larger than that of the MNL model. This counterintuitive result can be explained by the fact that ASU-DNN is trained solely to maximize the log-likelihood without incorporating any spatial structure. As a result, it may overfit to the training data and assign high probability to similar-attribute but geographically distant communities, resulting in less spatially coherent top predictions despite stronger likelihood-based performance. In contrast, the GNN-based models effectively leverage spatial dependencies through message passing, leading to smaller average distances. Nonetheless, ASU-DNN still outperforms MNL and SCL models in terms of log-likelihood and accuracy - the most common metrics in the evaluation of DCMs - and this result aligns with the findings from other studies.

The GNN model consistently outperforms the ASU-DNN, MNL, and SCL models across all evaluation metrics, with the one exception of accuracy in the one-layer GNN configuration. This demonstrates the effectiveness of incorporating spatial dependencies in modeling residential location choices. Among the GNNs, the two-layer and three-layer models perform better than the one-layer model, indicating that deeper architectures can capture more complex interactions among alternatives. However, the three-layer GNN does not substantially outperform the two-layer version and performs slightly worse in top-5 accuracy and mean reciprocal rank, suggesting that adding more layers may not always lead to better performance in this context. This finding could be caused by the oversmoothing issue in GNN. While deeper GNN architecture can pass more messages to the targeting alternative node, deeper architecture also diminish the model's learning capability of differentiating across nodes.

Note that the SCL and the MNL models have the same performance because the estimation value of the dissimilarity parameter approaches one, making the SCL model equivalent to the MNL model. This observation is consistent with findings reported by Sener et al. (2011). A plausible explanation for this behavior is the SCL model's use of equally distributed allocation parameters across all neighbors and a single dissimilarity parameter for all alternatives. Such an approach may fail to capture the heterogeneous dependencies in some data sets.

5.2. Impact of GNN-DCM configurations

We then investigate which configuration in the GNN-DCMs contributes the most to their predictive performance. Table 4 presents the average accuracy from ten-fold cross-validation across various GNN configurations, including different update and aggregation

Table 4
The average accuracy of ten-fold cross-validation for various GNN-DCM designs.

Models		$h = 64$			$K_g = 2$				$h = 64, K_g = 2$
Name	$\oplus$	$K_g = 1$	$K_g = 2$	$K_g = 3$	$h = 1$	$h = 16$	$h = 32$	$h = 128$	W/o skip
MPNN	Sum	12.38%	**12.95%**	*12.72%*	10.49%	12.38%	12.82%	12.59%	11.33%
	Max	**13.03%**	12.87%	12.38%	10.58%	12.82%	*13.03%*	12.74%	12.07%
	Mean	**12.87%**	12.51%	12.79%	10.48%	12.69%	*12.85%*	12.61%	11.49%
	LSE	12.66%	**13.05%**	12.79%	10.50%	12.64%	*12.92%*	12.74%	11.70%
GCN	Sum	12.66%	*12.69%*	**12.92%**	8.96%	12.46%	12.61%	12.63%	11.13%
GAT	Sum	12.64%	*13.03%*	**13.26%**	11.02%	12.64%	12.87%	12.96%	12.48%

* In each row, **the best** number is highlighted in bold, and *the second-best* number is underlined and italicized.

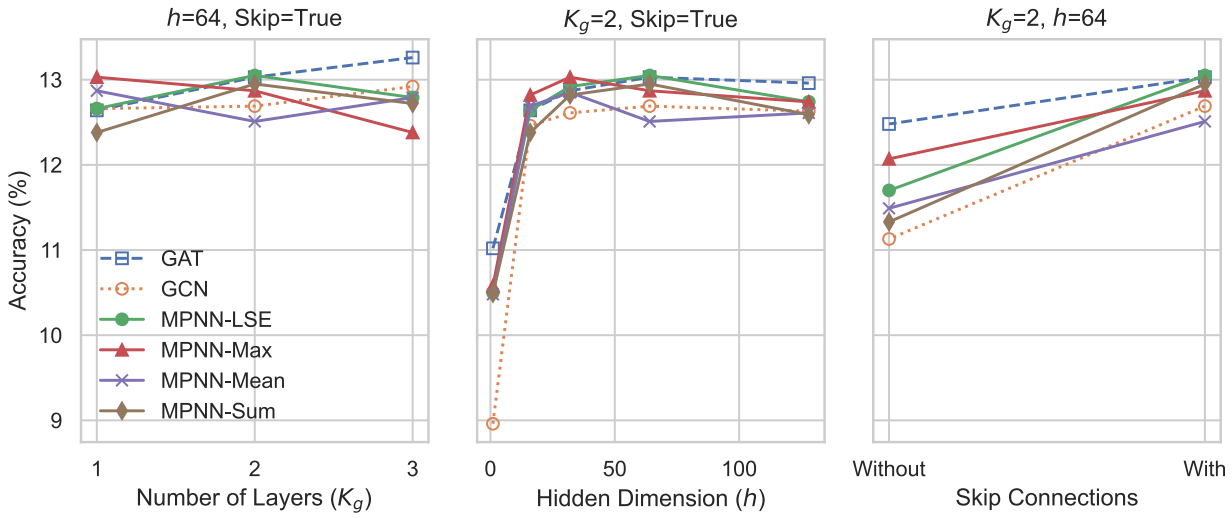


Fig. 6. Visualization of forecast accuracy of ten-fold cross-validation for different GNN designs.

functions, hidden dimensions (h), number of graph convolution layers (K_g), and the use of skip connections. Fig. 6 further visualizes these results.

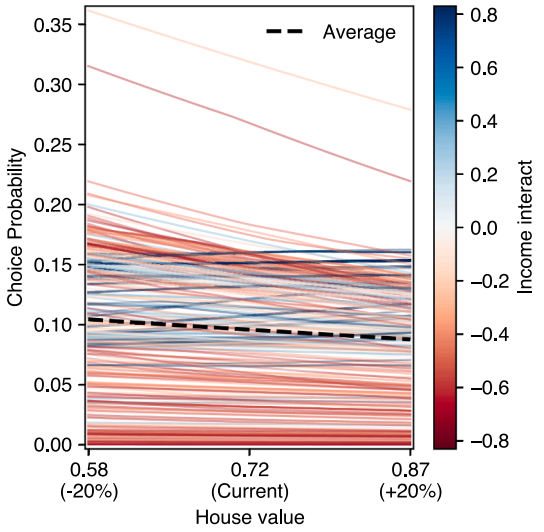
Overall, the results suggest that model performance is relatively insensitive to the number of layers (K_g), but highly sensitive to the hidden dimension. Notably, all the best-performing configurations across models (Message Passing, GCN, and GAT) share a hidden dimension of $h = 64$. The middle panel of Fig. 6 illustrates this finding, showing that the performance of all models improves significantly when the hidden dimension increases from 1 to 64, but remains relatively stable for larger dimensions. This indicates that a hidden dimension of 64 is sufficient to capture the complexity of the residential location choice problem, while larger dimensions do not yield substantial performance gains. These results highlight the importance of utilizing high-dimensional representations to capture the complexity of residential location choices.

We further compared the LSE aggregation to other aggregation methods because the LSE aggregation is the most common practice in the classical choice modeling. It also has an appealing utility interpretation in NL models because the log-sum form indicates the utility of the alternatives in a nest. However, the predictive differences across various aggregation functions appear minor. As suggested by Fig. 6, the LSE aggregation in a message passing network achieves comparable performance to other aggregation approaches, including max, mean, and sum. This result suggests that other aggregation methods are as effective as the LSE aggregation in capturing spatial dependencies, as least for predicting residential location choices.

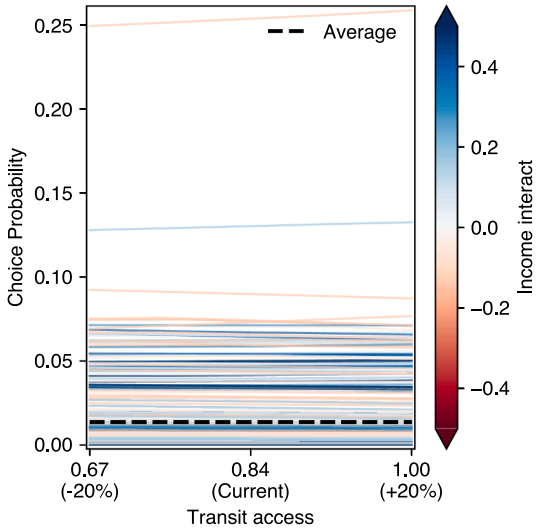
Finally, skip connections consistently lead to higher accuracy across all configurations. When skip connections are removed ("W/o skip" column), the performance of each model drops substantially. This is because skip connections maintain the alternative's original representation after each layer, preventing the node's information from being replaced by the aggregated information from its neighbors. This finding resonates with the past results from both the machine learning and the choice modeling communities. For example, Wang et al. (2021a) and Wong and Farooq (2021) both designed a structure similar to ResNet by integrating the MNL outputs with neural networks, indicating the importance of skipping the deeper architecture and retaining the linear components in models.

Table 5
Estimation results of the MNL model.

Attribute name	Parameter	t-statistic	Attribute name	Parameter	t-statistic
# Units	1.972	18.85	Transit access	−1.473	−8.09
House value	−0.322	−2.90	Work distance	−0.908	−21.63
House age	0.782	7.23	Pop density	2.341	17.12
Land mixture	−0.944	−8.96	Black interact	4.001	20.02
% Single house	−0.802	−5.65	White interact	2.812	12.34
% Multi house	0.581	7.63	Income interact	−0.875	−5.70
% Office	0.107	5.48			



(a) ICE plot of choosing Lake View (community 6) by housing units median value.



(b) ICE plot of choosing South Shore (community 43) by transit accessibility.

Fig. 7. The impact of housing unit value and transit accessibility on the choice probability.

5.3. Understanding the behavior of residential location choice

It is also essential to understand how various factors influence residential location choices, thus enhancing model interpretation and informing decision-making for transportation policy and planning. Classical MNL and NL models use linear utility functions, where the coefficients directly reflect an attribute's global effect on choice probabilities. This characteristic enhances the interpretability of these models, making them valuable tools in analysis and practice. Table 5 shows the estimated coefficients of the MNL model from one of the ten-fold cross-validation experiments. The results generally align with intuitive expectations: households are more likely to choose communities with more housing units, lower house values, and higher population densities. The negative coefficient for distance to work indicates that households prefer to reside closer to their workplaces. Additionally, the positive coefficients for the black and white interaction terms indicate that households are more likely to select communities with a higher percentage of residents matching their own race. Our results show a negative coefficient of the transit accessibility, which has also been observed in previous studies (Hu and Wang, 2019); although the effect of accessibility in household residential location choice may have different answers depending on contexts (Zondag and Pieters, 2005; Chen et al., 2008).

Neural network-based choice models capture the interactions between attributes in a non-linear way. As a result, the influence of an attribute (e.g., house value) varies with households' attributes, reflecting individuals heterogeneity in their choice behavior. While the deep learning approach improves predictive accuracy, the absence of single interpretable coefficients poses challenges for model interpretation. To overcome this limitation, we employ Individual Conditional Expectation (ICE) plots (Goldstein et al., 2015) to understand how the GNNs capture choice behavior. Since the ICE plots are alternative- and attribute-specific, we focus on two communities–Lake View (community 6) and South Shore (community 43)–to examine the effects of housing units' median value and transit accessibility on choice probabilities, as shown in Fig. 7. These communities were selected due to their high respondent counts, representing the entire region and the southern area, respectively.

Each ICE plot displays one line per individual from the test set to show how the choice probability of an individual changes as an attribute changes. To highlight interactions between attributes, we color the lines in Fig. 7 based on the income interaction attribute. Fig. 7a reveals that the probability of choosing Lake View generally decreases as the median value of housing units rises, consistent with the MNL model's findings. Interestingly, this effect varies across households with different income levels: high-income

Table 6
Elasticities of a selected household on community 6 (Lake View).

Attribute name	MNL			GNN (2 layers)		
	Direct elasticity	W.r.t. a neighbor	W.r.t. a non-neighbor	Direct elasticity	W.r.t. a neighbor	W.r.t. a non-neighbor
# Units	1.114	−0.630	−0.630	0.609	−0.392	−0.244
House value	−0.149	0.084	0.084	−0.648	0.333	0.321
House age	0.611	−0.346	−0.346	0.856	−0.450	−0.389
Land mixture	−0.403	0.228	0.228	−0.146	−0.017	0.121
% Single house	−0.057	0.032	0.032	0.002	−0.004	0.000
% Multi house	0.117	−0.066	−0.066	0.250	−0.142	−0.110
% Office	0.000	0.000	0.000	−0.004	0.001	0.003
Transit access	−0.910	0.514	0.514	0.098	−0.130	−0.011
Work distance	−0.191	0.108	0.108	−0.290	0.275	0.088
Pop density	1.859	−1.051	−1.051	0.411	−0.074	−0.210
Black pop	0.000	0.000	0.000	0.000	0.000	0.000
White pop	1.542	−0.872	−0.872	0.742	−0.189	−0.413
HH income	0.162	−0.091	−0.091	0.138	−0.011	−0.094

households (blue curves) are less sensitive to house value and even show a slight increase in choice probability with higher house value. This suggests that high-income households may prioritize community amenities, social status, or potential capital gains over housing costs. Fig. 7 shows that most households' choices of living in South Shore remain largely unaffected by transit accessibility, with only a few showing increased probability as accessibility improves. Combined with the MNL results, this implies that transit accessibility is not a primary consideration for most households when selecting South Shore as a residence.

In summary, the GNN models, while lacking single coefficients, can reflect nuanced attribute interactions and household heterogeneity, yielding a more detailed understanding of how factors influence specific alternatives and households. We acknowledge that the MNL model provides straightforward interpretability through its global attribute effects, though its results may be confounded by attribute correlations or non-linear relationships. Results based on GNN offer valuable insights into how the preference of a certain group of households shifts in response to changes in specific factors.

5.4. Analysis of elasticities

We also investigate how choice probabilities respond to changes in community attributes and how these responses differ between the MNL and the GNN-DCM. To do so, we randomly select a household and calculate the direct and cross elasticities of its choice probability w.r.t. changes in the attributes of Lake View (community 6). The elasticities are computed for both the MNL and GNN (2 layers) models, as shown in Table 6.

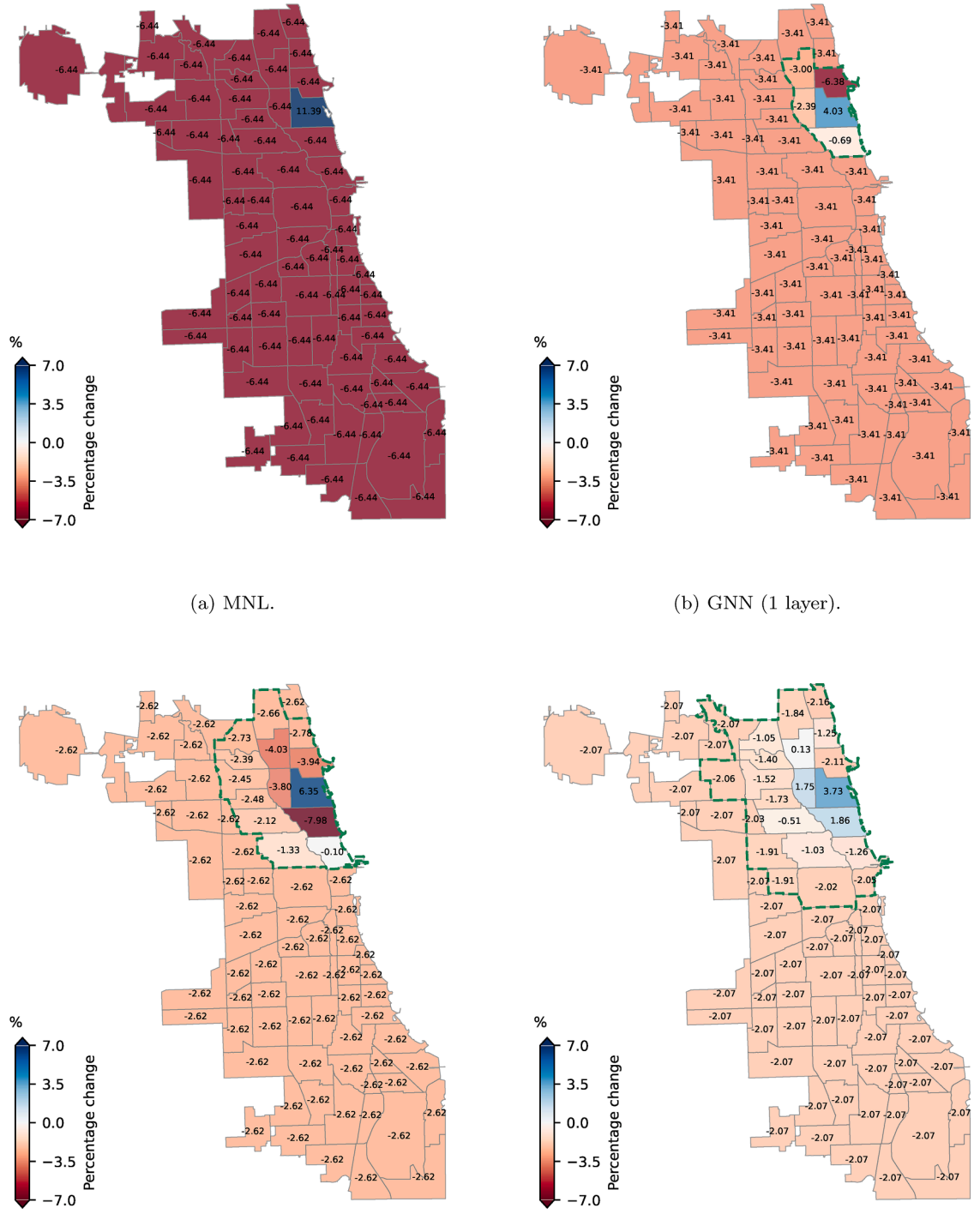
The signs of elasticities are generally consistent across the two models, suggesting that they capture similar choice behaviors for the selected household. For example, both models have positive direct elasticities for the number of housing units, indicating that an increase in the number of housing units in Lake View will lead to a higher probability of choosing Lake View itself. Conversely, the negative cross-elasticities w.r.t. neighboring and non-neighboring communities imply that such an increase reduces the likelihood of choosing other communities. The MNL model, because of the proportionate substitution patterns, the cross-elasticities w.r.t. a neighbor and a non-neighbor are the same for each attribute. In contrast, the GNN model exhibits varying cross-elasticities for neighbors versus non-neighbors, reflecting heterogeneous substitution patterns that account for spatial dependencies.

We further demonstrate the spatial patterns of the cross-elasticities, particularly focusing on how they evolve for GNNs with different layers. To avoid small numbers and for better visualization, Fig. 8 presents the percentage change w.r.t a 10% increase in the number of housing units in the Lake View community (community 6), which approximates ten times the elasticities. The boundaries of k_g-hop neighbors of GNNs are indicated by dashed lines, which helps to visualize how the influence of neighboring communities varies across different GNN architectures.

As depicted in Fig. 8, cross-elasticities of GNNs vary among communities within the k_g-hop neighborhood but remain constant for communities outside the k_g-hop neighbors, consistent with the theoretical analysis in Section 3.5. The MNL model can be viewed as a zero-layer GNN and thus exhibits a constant cross-elasticity across all communities. Additionally, the absolute values of the elasticities appear to decrease as the number of GNN layers increases. This trend arises from the GNN's message-passing mechanism, which introduces a smoothing effect: the change in the utility of a community is "averaged out" by its neighbors. Note that the results in Fig. 8 and Table 6 are derived from a single household, and specific elasticity values may vary across households. Nevertheless, the observed patterns remain consistent across multiple households. Overall, the elasticity analysis highlights the flexibility of the GNN-DCMs. While the classical MNL is limited by the proportional substitution across the whole map (Fig. 8a), the GNNs can mitigate this constraint by increasing the depth and expanding its reach through the message passing algorithm.

5.5. Examining graph attention weights

The best-performing GNN-DCM in this study adopts a GAT as the underlying GNN. The update function of the GAT is the weighted sum of the message from a node and its neighbors, where the weights (referred to as attention weights) are outputs of learned functions



(a) MNL. (b) GNN (1 layer).

(c) GNN (2 layer). (d) GNN (3 layer).

Fig. 8. Percentage change in the residential choice probability for a selected household following a 10% increase in housing units in Lake View (Community 6). The boundaries of k_g-hop neighbors are indicated by dashed lines.

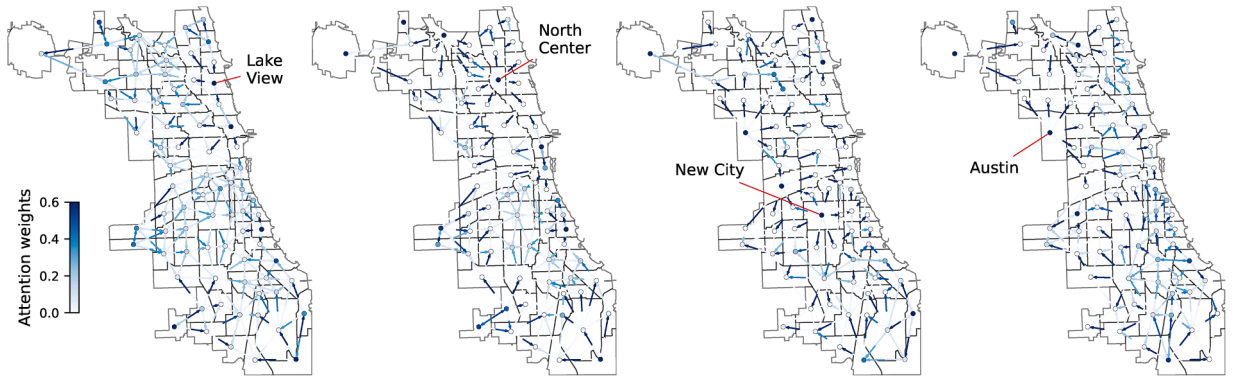


Fig. 9. Visualization of graph attention weights from the first layer of a learned GAT model. From left to right are head 1 to head 4 of the GAT. Arrows indicate the direction of attention weights from one node to another. Node colors represent self-attention weights.

of the corresponding node representations. These attention weights are directional, allowing the model to capture heterogeneous and context-dependent dependencies among alternatives. The technical details of the GAT architecture are provided in Appendix A.2. Here, we visualize the learned attention weights to gain insights into how the GAT captures spatial dependencies among alternatives.

Fig. 9 visualizes the attention weights from the first layer of a trained GAT model. Darker colors indicate higher attention weights. Arrows indicate the direction of attention weights from one node to another. Node colors represent self-attention weights. For each node, the attention weights from its neighbors and itself sum to one. The four subfigures correspond to the four attention heads in the GAT, each of which can attend to different aspects of the data. The outputs of these heads are concatenated to form the final node representations.

The different heads exhibit distinct patterns of spatial dependencies. Overall, the learned attention weights are highly asymmetric: the attention from node i to node j generally differs from that from node j to node i. This asymmetry implies that the influence among alternatives is directional. In particular, several communities exhibit relatively high self-attention weights and strong outgoing attention to neighboring communities. For example, Lake View, North Center, New City, and Austin, highlighted in the figure, indicate a relatively strong influence to the representations of nearby communities. Conversely, these communities receive weaker attention from their neighbors, showing an asymmetric pattern of dependence.

In contrast to classical NL and SCL models, which impose symmetric correlation structures among alternatives through shared unobserved components, the GAT allows for directional and context-dependent interactions. While the attention weights should not be interpreted directly as error correlations, this visualization illustrates the flexibility of the GAT in representing complex and heterogeneous spatial dependencies among alternatives.

6. Conclusion and discussion

This paper presents a GNN-DCM framework for modeling residential location choice with a large set of spatially correlated alternatives. The GNN-DCM framework provides a deep learning approach for integrating spatial dependencies into discrete choice models. In addition, we demonstrate that the GNN-DCM framework generalizes key classical models, including the NL and SCL models, offering a novel interpretation for DCMs from the perspective of message passing among utilities. The case study of residential location choices in Chicago demonstrates the advantage of the GNN-DCM in predictive performance, design flexibility, capturing individual-level heterogeneity, and spatially-aware substitution patterns.

There are limitations in this paper. This study focuses primarily on developing the GNN-DCM framework and its connections to classical and ANN-based DCMs. As a result, the empirical analysis is intentionally simplified to demonstrate the properties of GNN-DCMs, and thus does not incorporate the full range of household and community-level attributes commonly used in spatial choice modeling (e.g., Sivakumar and Bhat, 2007; Bhat and Guo, 2007; Hu and Wang, 2019). A second limitation is the potential endogeneity introduced by using the distance to work as an explanatory attribute, as residential and work location choices are often determined jointly (Paleti et al., 2013). Future work should integrate the GNN-DCM framework with a joint modeling structure, such as a two-stage model or an instrumental variables approach within the GNN framework, to address this endogeneity issue. Finally, the GNN-DCMs in the case study are designed to enable a fair comparison with baseline models and to highlight the model's spatially-aware substitution patterns. Consequently, the underlying graph is also simplified using only spatial adjacency to construct the adjacency matrix. The full potential of alternative graphs and deep learning techniques, such as using more sophisticated non-adjacency-based graph structure strategies, is not yet fully explored in this study.

This work creates new research opportunities by bridging the perspectives in graph, deep learning, and discrete choice models, enabling researchers to further synergize classical DCMs with the powerful AI toolbox. From the modeling perspective, future work could investigate other GNN variants and more advanced graph construction techniques, including the use of spatial distance, community similarity, land use patterns, or transportation connectivity to define the graph. There is also potential to explore automatic or learned graph structures, as well as multi-relational graphs that capture multiple types of spatial or functional connections between

alternatives. GNN-DCMs address the correlations among alternatives, but it is equally important to use GNNs for capturing correlations among individuals (Brock and Durlauf, 2001; Villarraga and Daziano, 2025) and their connection to classical models. From the application perspective, future research can extend the GNN-DCM framework to other choice modeling tasks. For example, GNN-DCM can be applied to travel mode choice by representing motorized and non-motorized alternatives as separate subgraphs (Zhou et al., 2025), analogous to the nesting structure in the NL framework. GNN-DCMs can also serve as key modules in more complex joint choice models, such as integrated models of residential location and travel mode choice (Anas, 1982; Ben-Akiva and Bowman, 1998; Bhat and Guo, 2007; Bhat et al., 2010, e.g.,). Lastly, while this paper builds the connection between GNN and DCM in the spatial choice context, potential research opportunities exist in answering deeper theoretical questions, such as how to discover new graph theory in the DCM context or how to embed GNN-DCMs into the more general random utility theoretical frameworks.

CRediT authorship contribution statement

Zhanhong Cheng: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Lingqian Hu:** Funding acquisition, Investigation, Project administration, Resources, Writing – review & editing; **Yuheng Bu:** Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing; **Yuqi Zhou:** Data curation, Formal analysis, Investigation, Writing – review & editing; **Shenhao Wang:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Validation, Writing – original draft, Writing – review & editing.

Data availability

Research Link Provided

Acknowledgements

The authors acknowledge the support from the Research Opportunity Seed Fund 2023 at the University of Florida and the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Vehicle Technology Program Award Number DE-EE0011186. The views expressed herein do not necessarily represent the views of the U.S. Department of Energy or the United States Government. The authors also acknowledge the early discussions with Dr. Kara Kockelman, Dr. Joan Walker, and Dr. Jinhua Zhao in the research seminars at UT Austin, UC Berkeley, and MIT, along with the early contributions from Dr. Rui Yao (Technion) and Mr. Siqi Feng (Cornell).

Appendix A. Appendix

A.1. Proof of Proposition 2

Proof. The choice probability of the SCL model is:

$$P_{ni} = \frac{\sum_{j\neq i}\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}\left[\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}+\left(\alpha_{j,ij}e^{V_{nj}}\right)^{1/\mu}\right]^{\mu-1}}{\sum_{k=1}^{|\mathcal{V}|-1}\sum_{l=k+1}^{|\mathcal{V}|}\left[\left(\alpha_{k,kl}e^{V_{nk}}\right)^{1/\mu}+\left(\alpha_{l,kl}e^{V_{nl}}\right)^{1/\mu}\right]^{\mu}}.$$

Since $\alpha_{i,ij}=0, \forall j\notin\mathcal{N}(i)$, the choice probability can be rewritten as:

$$\begin{aligned}
P_{ni} &= \frac{\sum_{j\in\mathcal{N}(i)}\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}\left[\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}+\left(\alpha_{j,ij}e^{V_{nj}}\right)^{1/\mu}\right]^{\mu-1}}{\sum_{kl\in\mathcal{E}}\left[\left(\alpha_{k,kl}e^{V_{nk}}\right)^{1/\mu}+\left(\alpha_{l,kl}e^{V_{nl}}\right)^{1/\mu}\right]^{\mu}} \\
&= \frac{\sum_{j\in\mathcal{N}(i)}\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}\left[\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}+\left(\alpha_{j,ij}e^{V_{nj}}\right)^{1/\mu}\right]^{\mu-1}}{\sum_{kl\in\mathcal{E}}\left[\frac{\left(\alpha_{k,kl}e^{V_{nk}}\right)^{1/\mu}+\left(\alpha_{l,kl}e^{V_{nl}}\right)^{1/\mu}}{\left(\alpha_{k,kl}e^{V_{nk}}\right)^{1/\mu}+\left(\alpha_{l,kl}e^{V_{nl}}\right)^{1/\mu}}\right]\left[\left(\alpha_{k,kl}e^{V_{nk}}\right)^{1/\mu}+\left(\alpha_{l,kl}e^{V_{nl}}\right)^{1/\mu}\right]^{\mu}} \\
&= \frac{\sum_{j\in\mathcal{N}(i)}\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}\left[\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}+\left(\alpha_{j,ij}e^{V_{nj}}\right)^{1/\mu}\right]^{\mu-1}}{\sum_{kl\in\mathcal{E}}\sum_{m\in\{k,l\}}\left\{\left(\alpha_{m,kl}e^{V_{nl}}\right)^{1/\mu}\left[\left(\alpha_{k,kl}e^{V_{nk}}\right)^{1/\mu}+\left(\alpha_{l,kl}e^{V_{nl}}\right)^{1/\mu}\right]^{\mu-1}\right\}} \\
&= \frac{\sum_{j\in\mathcal{N}(i)}\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}\left[\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}+\left(\alpha_{j,ij}e^{V_{nj}}\right)^{1/\mu}\right]^{\mu-1}}{\sum_{k\in\mathcal{V}}\sum_{l\in\mathcal{N}(k)}\left\{\left(\alpha_{k,kl}e^{V_{nl}}\right)^{1/\mu}\left[\left(\alpha_{k,kl}e^{V_{nk}}\right)^{1/\mu}+\left(\alpha_{l,kl}e^{V_{nl}}\right)^{1/\mu}\right]^{\mu-1}\right\}}
\end{aligned}$$

$$= \frac{\exp\left\{\log\left(\sum_{j\in\mathcal{N}(i)}\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}\left[\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}+\left(\alpha_{j,ij}e^{V_{nj}}\right)^{1/\mu}\right]^{\mu-1}\right)\right\}}{\sum_{k\in\mathcal{V}}\exp\left\{\log\left(\sum_{l\in\mathcal{N}(k)}\left(\alpha_{k,kl}e^{V_{nl}}\right)^{1/\mu}\left[\left(\alpha_{k,kl}e^{V_{nk}}\right)^{1/\mu}+\left(\alpha_{l,kl}e^{V_{nl}}\right)^{1/\mu}\right]^{\mu-1}\right)\right\}}. \tag{A.1}$$

Eq. (A.1) shows that the SCL model can be reformulated into the MNL form given by Eq. (1), and the exponential terms in both the numerator and the denominator have the same functional form. Continuing to simplify the exponential terms, we have:

$$\log\left(\sum_{j\in\mathcal{N}(i)}\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}\left[\left(\alpha_{i,ij}e^{V_{ni}}\right)^{1/\mu}+\left(\alpha_{j,ij}e^{V_{nj}}\right)^{1/\mu}\right]^{\mu-1}\right)$$
$$=\log\left(\sum_{j\in\mathcal{N}(i)}e^{V_{ni}/\mu+\log(\alpha_{i,ij})/\mu}\left[e^{V_{ni}/\mu+\log(\alpha_{i,ij})/\mu}+e^{V_{nj}/\mu+\log(\alpha_{j,ij})/\mu}\right]^{\mu-1}\right).$$

Letting $V_{ni}^{(0)} = V_{ni}/\mu + \log(\alpha_{i,ij})/\mu$, the above expression can be rewritten as a one-layer GNN update function:

$$V_{ni}^{(1)} = \log\left(\sum_{j\in\mathcal{N}(i)}e^{V_{ni}^{(0)}}\left[e^{V_{ni}^{(0)}}+e^{V_{nj}^{(0)}}\right]^{\mu-1}\right)$$
$$=\log\left(\sum_{j\in\mathcal{N}(i)}\exp\left(V_{ni}^{(0)}+(\mu-1)\log\left[e^{V_{ni}^{(0)}}+e^{V_{nj}^{(0)}}\right]\right)\right),$$

where $V_{ni}^{(0)} = \frac{V_{ni}}{\mu} + \frac{\log(\alpha_{i,ij})}{\mu} = \frac{\mathbf{b}^\top\mathbf{x}_{ni}}{\mu} + \frac{1}{\mu}\log\alpha_{i,ij}$ serves as the initial node representation of the GNN. □

Note that the same proof also holds for generalized spatially correlated logit (GSCL) models (Sener et al., 2011), which takes the same form as the SCL but parameterizes the allocation parameters, $\{\alpha_{i,ij}, \forall j \in \mathcal{N}(i)\}$, by an additional MNL model.

A.2. Details of GNN update functions

The details of the GNN update functions used in the experiments are provided in this section.

- **Message passing neural network (MPNN)**: The update function for MPNN is defined as:

$$\mathbf{h}_i^{(k)} = \sigma\left(\mathbf{W}\mathbf{h}_i^{(k-1)} + \bigoplus_{j\in\mathcal{N}(i)}\mathbf{W}\mathbf{h}_j^{(k-1)}\right),$$

  where σ is an element-wise non-linear activation function (we use ReLU Nair and Hinton, 2010 for all activations in this study), $\mathbf{W}$ is a learnable weight matrix, and the second term aggregates the representations of neighboring nodes. We test sum, mean, max, and log-sum-exp (LSE) as the aggregation functions $\bigoplus$ in the MPNN update step.
- **Graph convolutional network (GCN)**: The update function for GCN is defined as:

$$\mathbf{h}_i^{(k)} = \sigma\left(\sum_{j\in\mathcal{N}(i)\cup\{i\}}\frac{1}{\sqrt{|\mathcal{N}(i)|}\sqrt{|\mathcal{N}(j)|}}\mathbf{W}\mathbf{h}_j^{(k-1)}\right),$$

  where σ is the activation function, $\mathbf{W}$ is a learnable weight matrix, and the normalization term ensures that the updates are invariant to the node degrees.
- **Graph attention network (GAT)**: The update function at each layer computes a weighted sum of a node's neighbor representations, defined as:

$$\mathbf{h}_i^{(k)} = \sigma\left(\sum_{j\in\mathcal{N}(i)\cup\{i\}}\alpha_{ij}\mathbf{W}\mathbf{h}_j^{(k-1)}\right), \tag{A.2}$$

  where σ is the activation function, $\mathbf{W}$ is a learnable projection matrix, and α_{ij} represents the attention weight indicating the importance of node j's representation to node i. The attention weights are calculated by a softmax function over the nodes' representations:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top\left[\mathbf{W}\mathbf{h}_i||\mathbf{W}\mathbf{h}_j\right]\right)\right)}{\sum_{k\in\mathcal{N}(i)\cup\{i\}}\exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top\left[\mathbf{W}\mathbf{h}_i||\mathbf{W}\mathbf{h}_k\right]\right)\right)},$$

  where $||$ denotes concatenation, and $\mathbf{a}$ is a learnable coefficient vector. It is common to use separate Eq. (A.2) multiple times (i.e., stacking the outputs of multiple functions), referring to the multi-head attentions. We choose the number of heads to be 1 when the hidden dimension $h = 1$ and 4 otherwise.

A.3. Details of skip connections

Skip connection was first popularized by the ResNet architecture (He et al., 2016). It connects the input of a layer to the output of a deeper layer, enabling the training of very deep neural networks by mitigating the vanishing gradient problem. In GNNs, skip connections are implemented in various ways, such as through summation (He et al., 2016) or concatenation (Huang et al., 2017). These approaches serve the same purpose–preserving the original node representation–and typically have a similar performance in GNNs (You et al., 2020).

In this paper, we adopt the gated skip connection via summation (Srivastava et al., 2015). For example, the update function of a GNN with skip connections for MPNN is:

$$\mathbf{h}_i^{(k)} = \sigma\left((1-c)\mathbf{h}_i^{(k-1)} + c\left(\mathbf{W}\mathbf{h}_i^{(k-1)} + \bigoplus_{j\in\mathcal{N}(i)} \mathbf{W}\mathbf{h}_j^{(k-1)}\right)\right),$$

where $c = \text{sigmoid}(\mathbf{W}_c\mathbf{h}_i^{(k-1)} + \mathbf{b}_c) \in (0,1)$ is a learnable gate parameter, which allows the model to adaptively balance between retaining the original node representation and incorporating the output from the GNN message passing. This layer reverts to which of the ASU-DNN when $c \to 0^+$. In addition, it keeps the same dimensionality across hidden layers. We use the same skip connection mechanism for all GNNs used in this paper, including the MPNN, GCN, and GAT.

References

Agrawal, D., Schorling, C., 1996. Market share forecasting: an empirical comparison of artificial neural networks and multinomial logit model. J. Retailing 72 (4), 383–407.

Anas, A., 1982. Residential Location Markets and Urban Transportation. Economic Theory, Econometrics and Policy Analysis with Discrete Choice Models. Monograph.

Ansel, J., Yang, E., He, H., Gimelshein, N., Jain, A., Voznesensky, M., Bao, B., Bell, P., Berard, D., Burovski, E., Chauhan, G., Chourdia, A., Constable, W., Desmaison, A., DeVito, Z., Ellison, E., Feng, W., Gong, J., Gschwind, M., Hirsh, B., Huang, S., Kalambarkar, K., Kirsch, L., Lazos, M., Lezcano, M., Liang, Y., Liang, J., Lu, Y., Luk, C.K., Maher, B., Pan, Y., Puhrsch, C., Reso, M., Saroufim, M., Siraichi, M.Y., Suk, H., Suo, M., Tillet, P., Wang, E., Wang, X., Wen, W., Zhang, S., Zhao, X., Zhou, K., Zou, R., Mathews, A., Chanan, G., Wu, P., Chintala, S., 2024. PyTorch 2: faster machine learning through dynamic python bytecode transformation and graph compilation. In: 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24). ACM. https://doi.org/10.1145/3620665.3640366.

Bayoh, I., Irwin, E.G., Haab, T., 2006. Determinants of residential location choice: how important are local public goods in attracting homeowners to central city locations? J. Reg. Sci. 46 (1), 97–120.

Ben-Akiva, M., Bowman, J.L., 1998. Integration of an activity-based model system and a residential location model. Urban Stud. 35 (7), 1131–1153.

Ben-Akiva, M.E., 1973. Structure of Passenger Travel Demand Models. Ph.D. thesis. Massachusetts Institute of Technology.

Bhat, C.R., 2000. A multi-level cross-classified model for discrete response variables. Transp. Res. Part B: Methodol. 34 (7), 567–582.

Bhat, C.R., Eluru, N., 2009. A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. Transp. Res. Part B: Methodol. 43 (7), 749–765.

Bhat, C.R., Guo, J., 2004. A mixed spatially correlated logit model: formulation and application to residential choice modeling. Transp. Res. Part B: Methodol. 38 (2), 147–168.

Bhat, C.R., Guo, J.Y., 2007. A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. Transp. Res. Part B: Methodol. 41 (5), 506–526.

Bhat, C.R., Sener, I.N., Eluru, N., 2010. A flexible spatially dependent discrete choice model: formulation and application to teenagers' weekday recreational activity participation. Transp. Res. Part B: Methodol. 44 (8-9), 903–921.

Bina, M., Warburg, V., Kockelman, K.M., 2006. Location choice vis-à-vis transportation: apartment dwellers. Transp. Res. Rec. 1977 (1), 93–102.

Bolduc, D., Ben-Akiva, M., 1991. A multinomial probit formulation for large choice sets. In: Les Methodes D'analyse DES Comportements DE Deplacements Pour LES Annees 1990-6E Conference Internationale Sur Les Comportements De Deplacements, Volume 2.

Brock, W.A., Durlauf, S.N., 2001. Discrete choice with social interactions. Rev. Econ. Stud. 68 (2), 235–260.

Chen, J., Chen, C., Timmermans, H. J.P., 2008. Accessibility trade-offs in household residential location decisions. Transp. Res. Rec. 2077 (1), 71–79.

Chicago Historical Society, 2004. Community areas. Encyclopedia of Chicago. Accessed Mar 23, 2025. http://www.encyclopedia.chicagohistory.org/pages/319.html.

Dugundji, E.R., Walker, J.L., 2005. Discrete choice with social and spatial network interdependencies: an empirical example using mixed generalized extreme value models with field and panel effects. Transp. Res. Rec. 1921 (1), 70–78.

Feng, S., Yao, R., Hess, S., Daziano, R.A., Brathwaite, T., Walker, J., Wang, S., 2024. Deep neural networks for choice analysis: enhancing behavioral regularity with gradient regularization. Transp. Res. Part C: Emerging Technol. 166, 104767.

Fey, M., Lenssen, J.E., 2019. Fast graph representation learning with PyTorch geometric. In: ICLR Workshop on Representation Learning on Graphs and Manifolds.

Goetzke, F., 2008. Network effects in public transit use: evidence from a spatially autoregressive mode choice model for new york. Urban Stud. 45 (2), 407–417.

Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. J. Comput. Graphical Stat. 24 (1), 44–65.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press. http://www.deeplearningbook.org.

Guevara, C.A., Ben-Akiva, M., 2006. Endogeneity in residential location choice models. Transp. Res. Rec. 1977 (1), 60–66.

Hamilton, W., Ying, Z., Leskovec, J., 2017. Inductive representation learning on large graphs. Adv. Neural Inf. Process. Syst. 30, 1025–1035.

Hamilton, W.L., 2020. Graph Representation Learning. Morgan & Claypool Publishers.

Han, Y., Pereira, F.C., Ben-Akiva, M., Zegras, C., 2022. A neural-embedded discrete choice model: learning taste representation with strengthened interpretability. Transp. Res. Part B: Methodol. 163, 166–186.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Hu, L., Wang, L., 2019. Housing location choices of the poor: does access to jobs matter? Hous. Stud. 34 (10), 1721–1745.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.

Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kipf, T.N., Welling, M., 2017. Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations .

Kumar, A., Rao, V.R., Soni, H., 1995. An empirical comparison of neural network and logistic regression models. Mark. Lett. 6, 251–263.

Lee, B. H.Y., Waddell, P., 2010. Residential mobility and location choice: a nested logit model with sampling of alternatives. Transportation 37, 587–601.

LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics. Chapman and Hall/CRC.

Liu, D.C., Nocedal, J., 1989. On the limited memory BFGS method for large scale optimization. Math. Program. 45 (1), 503–528.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Fontiers in Econometrics. Academic Press, pp. 105–142.

McFadden, D., 1978. Modelling the choice of residential location. Transp. Res. Rec. 673, 72–77.
Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 807–814.
Paleti, R., Bhat, C.R., Pendyala, R.M., 2013. Integrated model of residential location, work location, vehicle ownership, and commute tour characteristics. Transp. Res. Rec. 2382 (1), 162–172.
Perez-Lopez, J.-B., Novales, M., Orro, A., 2022. Spatially correlated nested logit model for spatial location choice. Transp. Res. Part B: Methodol. 161, 1–12.
Schirmer, P.M., Van Eggermond, M. A.B., Axhausen, K.W., 2014. The role of location in residential location choice models: a review of literature. J. Transp. Land Use 7 (2), 3–21.
Sener, I.N., Pendyala, R.M., Bhat, C.R., 2011. Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modeling residential location choice behavior. J. Transp. Geogr. 19 (2), 294–303.
Sifringer, B., Lurkin, V., Alahi, A., 2020. Enhancing discrete choice models with representation learning. Transp. Res. Part B: Methodol. 140, 236–261.
Sivakumar, A., Bhat, C.R., 2007. Comprehensive, unified framework for analyzing spatial location choice. Transp. Res. Rec. 2003 (1), 103–111.
Srivastava, R.K., Greff, K., Schmidhuber, J., 2015. Highway networks. arXiv preprint arXiv:1505.00387.
Tomlinson, K., Benson, A.R., 2024. Graph-based methods for discrete choice. Netw. Sci. 12 (1), 21–40.
Train, K.E., 2009. Discrete Choice Methods with Simulation. Cambridge university press.
Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y., 2018. Graph attention networks. In: International Conference on Learning Representations.
Villarraga, D.F., Daziano, R.A., 2025. Designing graph convolutional neural networks for discrete choice with network effects. arXiv preprint arXiv:2503.09786.
Wang, S., Mo, B., Zhao, J., 2020a. Deep neural networks for choice analysis: architecture design with alternative-specific utility functions. Transp. Res. Part C: Emerging Technol. 112, 234–251.
Wang, S., Mo, B., Zhao, J., 2021a. Theory-based residual neural networks: a synergy of discrete choice models and deep neural networks. Transp. Res. Part B: Methodol. 146, 333–358. https://doi.org/10.1016/j.trb.2021.03.002.
Wang, S., Mo, B., Zheng, Y., Hess, S., Zhao, J., 2024. Comparing hundreds of machine learning and discrete choice models for travel demand modeling: an empirical benchmark. Transp. Res. Part B: Methodol. 190, 103061.
Wang, S., Wang, Q., Bailey, N., Zhao, J., 2021b. Deep neural networks for choice analysis: a statistical learning theory perspective. Transp. Res. Part B: Methodol. 148, 60–81.
Wang, S., Wang, Q., Zhao, J., 2020b. Deep neural networks for choice analysis: extracting complete economic information for interpretation. Transp. Res. Part C: Emerging Technol. 118, 102701.
Wang, X.C., Kockelman, K.M., Lemp, J.D., 2012. The dynamic spatial multinomial probit model: analysis of land use change using parcel-level data. J. Transp. Geogr. 24, 77–88.
Weisbrod, G., Lerman, S.R., Ben-Akiva, M., 1980. Tradeoffs in residential location decisions: transportation versus other factors. Transp. Policy Decis. Making 1 (1), 13–26.
Wen, C.-H., Koppelman, F.S., 2001. The generalized nested logit model. Transp. Res. Part B: Methodol. 35 (7), 627–641.
Wong, M., Farooq, B., 2021. ResLogit: a residual neural network logit model for data-driven choice modelling. Transp. Res. Part C: Emerging Technol. 126, 103050.
You, J., Ying, Z., Leskovec, J., 2020. Design space for graph neural networks. Adv. Neural Inf. Process. Syst. 33, 17009–17021.
Zhou, Y., Cheng, Z., Hu, L., Bu, Y., Wang, S., 2025. NestGNN: a graph neural network framework generalizing the nested logit model for travel mode choice. arXiv preprint arXiv:2509.07123.
Zondag, B., Pieters, M., 2005. Influence of accessibility on residential location choice. Transp. Res. Rec. 1902 (1), 63–70.