Graph Neural Networks for Residential Location Choice

Connection to Classical Logit Models

Zhanhong Cheng

Assistant Professor, Zhejiang University

zhanhong.cheng@zju.edu.cn

26th International Symposium on Transportation and Traffic Theory (ISTTT)

July 21, 2026, Munich, Germany

Co-authors

Lingqian Hu

Texas A&M University

Yuheng Bu

UC Santa Barbara

Yuqi Zhou

University of Florida

Shenhao Wang †

University of Florida

Funding

- U.S. DOE EERE, Vehicle Technologies Program, Award DE-EE0011186.
- UF Research Opportunity Seed Fund 2023.

Outline

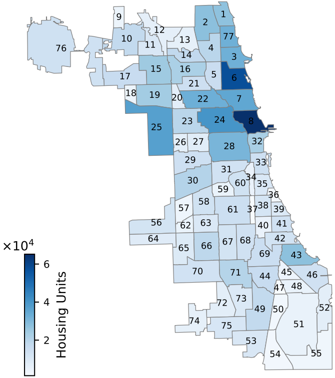
Motivation

What affects where we live?

A discrete choice model (DCM) over location alternatives (Ben-Akiva and Bowman, 1998; Bhat and J. Y. Guo, 2007).

Factors influencing the choice:

- About the location:
  - Housing prices.
  - Travel time to work.
  - School quality.
  - others.
- About the individual:
  - Income.
  - Family size.
  - Age.
  - others.



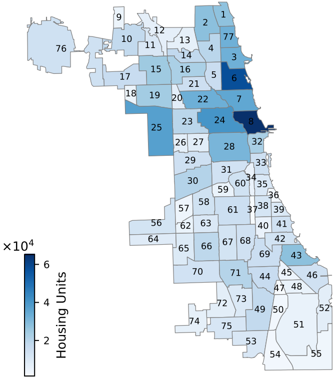
77 communities in Chicago – our case study.

Where we live: important, but hard to model

A discrete choice model (DCM) over location alternatives (Ben-Akiva and Bowman, 1998; Bhat and J. Y. Guo, 2007).

Key challenges:

1. **Spatial correlation** breaks the IIA assumption of the multinomial logit (MNL) (McFadden, 1974).
2. A **large** alternative set makes correlation structures hard to specify.
3. Choices involve complex interactions and non-linearities among **individual** $\times$ **location** attributes.



77 communities in Chicago – our case study.

Location choice – Nested logit (NL)

The nest structure of a two-level NL model.

- Each group of alternatives form a nest.
- The choice probability of the NL (McFadden, 1978):

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

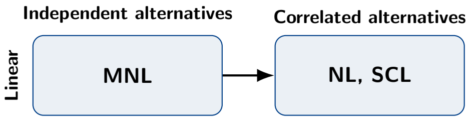
Location choice – spatially correlated logit (SCL)

The nest structure of an SCL model.

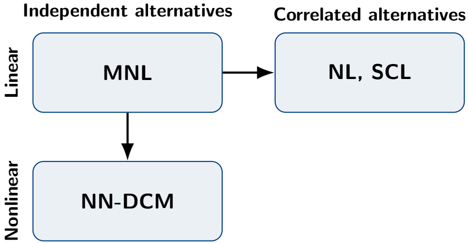
- Each pair of neighbors belongs to a nest.
- SCL (Bhat and J. Guo, 2004) and its extensions (Perez-Lopez et al., 2022; Sener et al., 2011):

$$P_{n,i} = \sum_{j \neq i} P_{n,i|ij} P_{n,ij}.$$

Location choice — the gap

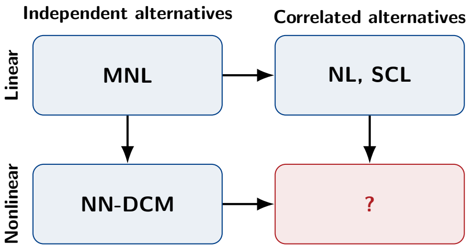


Location choice — the gap



Neural Network-based discrete choice models (NN-DCM) address nonlinearity and attributes interactions, but often **assume independent alternatives**, or mix them only implicitly (Sifringer et al., 2020; Wang et al., 2020; Wong and Farooq, 2021).

Location choice — the gap

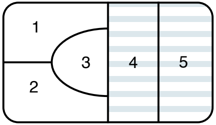


Neural Network-based discrete choice models (NN-DCM) address nonlinearity and attributes interactions, but often **assume independent alternatives**, or mix them only implicitly (Sifringer et al., 2020; Wang et al., 2020; Wong and Farooq, 2021).

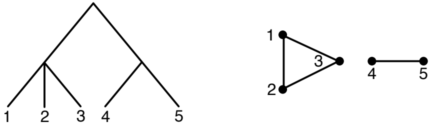
The gap

How can NN-DCM explicitly capture **alternatives' correlations**, while keeping a clear link to discrete choice theory?

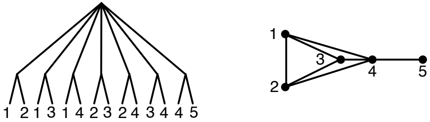
Graph Neural Network (GNN) — a natural bridge



Spatial configuration of residential locations.



The nested structure of an NL model and its corresponding alternative graph.



The nested structure of an SCL model and its corresponding alternative graph.

Key idea

GNN is a natural way to encode alternative correlations for NN-based choice models, while keeping a clear link to classical DCMs.

The GNN-DCM framework

GNN-DCM: The alternative graph

An **alternative graph** $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

- $\mathcal{V}$ **nodes**: alternatives (locations);
- $\mathcal{E}$ **edges**: dependence (e.g. spatial adjacency);
- independent $\Rightarrow$ no edge.

Each node carries the attributes $\mathbf{x}_{ni}$ that individual n sees for alternative i.

GNN-DCM: The alternative graph

An **alternative graph** $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

- $\mathcal{V}$ **nodes**: alternatives (locations);
- $\mathcal{E}$ **edges**: dependence (e.g. spatial adjacency);
- independent $\Rightarrow$ no edge.

Each node carries the attributes $\mathbf{x}_{ni}$ that individual n sees for alternative i.

Note

Different from social-network GNNs for choice (Tomlinson and Benson, 2024; Villarraga and Daziano, 2025), which graph the *individuals*. Here the graph is on the *choice set*.

GNN-DCM: Utility by message passing on the graph

Random utility $U_{ni} = V_{ni} + \varepsilon_{ni}$; i.i.d. Gumbel $\Rightarrow P_{ni} = e^{V_{ni}} / \sum_j e^{V_{nj}}$.

Now let V_{ni} depend on a node *and its neighbors* via a K_g-layer GNN:

$$\mathbf{h}_{ni}^{(k)} = \phi^{(k)}\Big(\mathbf{h}_{ni}^{(k-1)},\ \bigoplus_{j\in\mathcal{N}(i)} \psi^{(k)}\big(\mathbf{h}_{ni}^{(k-1)}, \mathbf{h}_{nj}^{(k-1)}\big)\Big), \qquad V_{ni} = \mathbf{w}^\top \mathbf{h}_{ni}^{(K_g)},$$

where

- $\mathbf{h}_{ni}^{(k)}$ is the k-th layer embedding, and $\mathbf{h}_{ni}^{(0)} = \mathbf{x}_{ni}$ or using an input function $\mathbf{h}_{ni}^{(0)} = f(\mathbf{x}_{ni})$;
- $\mathcal{N}(i)$ is the neighborhood set of alternative i;
- $\psi^{(k)}$ is a message function;
- $\bigoplus$ is an aggregation operator (e.g. sum, mean, max, attention-based sum);
- $\phi^{(k)}$ is an update function.

GNN-DCM: The framework

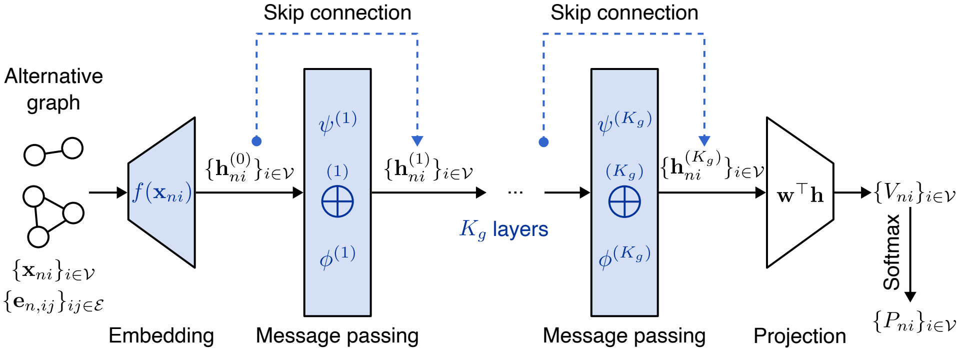


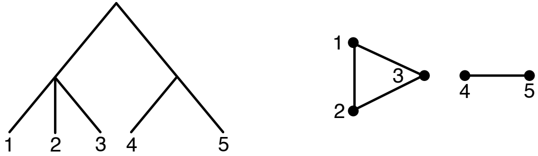
Illustration of the GNN-DCM framework. Customizable design components are colored in blue.

Connection between GNN-DCM and classical DCMs

Nested logit is a single-layer GNN-DCM

Proposition

A two-level NL model is a single-layer GNN-DCM: each nest is a **complete subgraph**; there are no edges between nests.



The nested structure of the NL and its corresponding alternative graph for the GNN

GNN message passing scheme for alternative i:

$$P_{ni} = \text{softmax}_i(\widetilde{V}_n), \quad \widetilde{V}_{ni} = \underbrace{\frac{V_{ni}}{\mu_k} + (\mu_k - 1)\overbrace{\log \sum_{j \in \mathcal{N}(i) \cup \{i\}} e^{V_{nj}/\mu_k}}^{\text{aggregate } \bigoplus}}_{\text{update } \phi}$$

Proof: nested logit = single-layer GNN-DCM

The choice probability of an alternative i in the nest B_k of a nested logit model is:

$$
\begin{aligned}
P_i &= \frac{\exp\left(V_i/\mu_k\right)}{\sum_{j\in B_k}\exp\left(V_j/\mu_k\right)} \times \frac{\left(\sum_{j\in B_k}\exp\left(V_j/\mu_k\right)\right)^{\mu_k}}{\sum_{l=1}^{K}\left(\sum_{j\in B_l}\exp\left(V_j/\mu_l\right)\right)^{\mu_l}} \\
&= \frac{\exp\left(V_i/\mu_k\right)\left(\sum_{j\in B_k}\exp\left(V_j/\mu_k\right)\right)^{\mu_k-1}}{\sum_{l=1}^{K}\left(\sum_{j\in B_l}\exp\left(V_j/\mu_l\right)\right)^{\mu_l}} \\
&= \frac{\exp\left(V_i/\mu_k\right)\left(\sum_{j\in B_k}\exp\left(V_j/\mu_k\right)\right)^{\mu_k-1}}{\sum_{l=1}^{K}\left(\frac{\sum_{m\in B_l}\exp(V_m/\mu_l)}{\sum_{m\in B_l}\exp(V_m/\mu_l)}\left(\sum_{j\in B_l}\exp\left(V_j/\mu_l\right)\right)^{\mu_l}\right)} \\
&= \frac{\exp\left(V_i/\mu_k\right)\left(\sum_{j\in B_k}\exp\left(V_j/\mu_k\right)\right)^{\mu_k-1}}{\sum_{l=1}^{K}\sum_{m\in B_l}\left(\frac{\exp(V_m/\mu_l)}{\sum_{m\in B_l}\exp(V_m/\mu_l)}\left(\sum_{j\in B_l}\exp\left(V_j/\mu_l\right)\right)^{\mu_l}\right)} \\
&= \frac{\exp\left(V_i/\mu_k+(\mu_k-1)\log\left(\sum_{j\in B_k}\exp\left(V_j/\mu_k\right)\right)\right)}{\sum_{l=1}^{K}\sum_{m\in B_l}\exp\left(V_m/\mu_l+(\mu_l-1)\log\left(\sum_{j\in B_l}\exp\left(V_j/\mu_l\right)\right)\right)},
\end{aligned}
$$

which can be viewed as the MNL (softmax) form with the **transformed utility**:

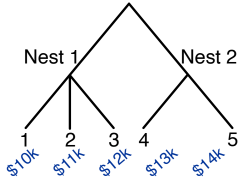
$$P_i = \mathrm{softmax}_i(\widetilde{V}_n)$$

$$\widetilde{V}_{ni} = \underbrace{\frac{V_{ni}}{\mu_k} + (\mu_k-1)\overbrace{\log\sum_{j\in\mathcal{N}(i)\cup\{i\}} e^{V_{nj}/\mu_k}}^{\text{aggregate }\bigoplus}}_{\text{update }\phi}.$$

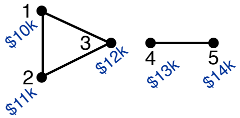
This is exactly a single-layer message passing with **complete subgraphs** per nest. ■

It checks out: same probability, two views

Nested-logit view of P_1



GNN view of P_1



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

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

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

$$V_1 = \frac{-10}{0.6} + (0.6-1)\log(e^{\frac{-10}{0.6}} + e^{\frac{-11}{0.6}} + e^{\frac{-12}{0.6}}) = -10.063$$

$$V_2 = \frac{-11}{0.6} + (0.6-1)\log(e^{\frac{-10}{0.6}} + e^{\frac{-11}{0.6}} + e^{\frac{-12}{0.6}}) = -11.313$$

$$V_3 = \frac{-12}{0.6} + (0.6-1)\log(e^{\frac{-10}{0.6}} + e^{\frac{-11}{0.6}} + e^{\frac{-12}{0.6}}) = -12.563$$

$$V_4 = \frac{-13}{0.5} + (0.5-1)\log(e^{\frac{-13}{0.5}} + e^{\frac{-14}{0.5}}) = -13.028$$

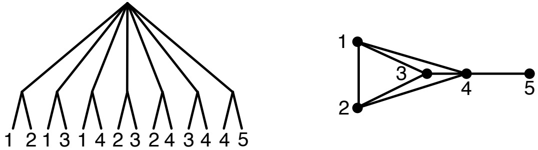
$$V_5 = \frac{-14}{0.5} + (0.5-1)\log(e^{\frac{-13}{0.5}} + e^{\frac{-14}{0.5}}) = -14.140$$

$$P_1 = \frac{e^{V_1}}{e^{V_1} + e^{V_2} + e^{V_3} + e^{V_4} + e^{V_5}} = 0.6959$$

$\text{Utility}_i = -\text{Price}_i$; $\mu_1 = 0.6$, $\mu_2 = 0.5$. Both views give the same $P_1 = 0.6959$.

Proposition

The SCL model is a single-layer GNN-DCM, where each nest with a pair of alternatives is an edge in the graph.



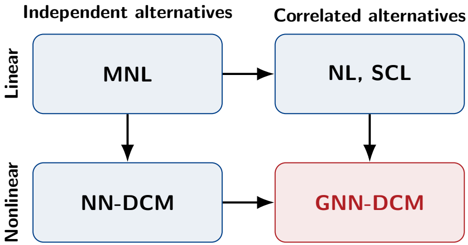
The nested structure of the SCL and its corresponding alternative graph for the GNN

GNN message passing scheme for alternative i in nest B_k:

$$P_{ni} = \text{softmax}_i(\widetilde{V}_n), \quad \widetilde{V}_{ni} = \log\left(\sum_{j \in \mathcal{N}(i)} \exp\left(V_{ni} + (\mu - 1)\log\left[\exp(V_{ni}) + \exp(V_{nj})\right]\right)\right),$$

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

GNN-DCM: The big picture



The relationship between classical and NN-based DCMs.

- GNN-DCM provides a new interpretation for NL and SCL models as message passing among alternative utilities.
- GNN-DCM offers a structured approach for neural networks to capture dependency among alternatives.

GNN-DCM substitution patterns

Elasticities $E_{i\,z_{nj}}$ measure the change in the choice probability P_{ni} on alternative i in response to a change in the attribute z_{nj} of alternative j:

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

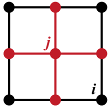
GNN-DCM substitution patterns

Elasticities $E_{i\,z_{nj}}$ measure the change in the choice probability P_{ni} on alternative i in response to a change in the attribute z_{nj} of alternative j:

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

For a GNN-DCM with K_g layers,

- if i is **outside** the k_g-hop neighborhood of j: the elasticity is the **same for all such** i — like MNL/IIA (proportional);



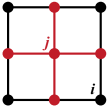
GNN-DCM substitution patterns

Elasticities $E_{i\, z_{nj}}$ measure the change in the choice probability P_{ni} on alternative i in response to a change in the attribute z_{nj} of alternative j:

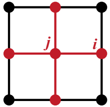
$$E_{i,z_{nj}} = \frac{\partial P_{ni}}{\partial z_{nj}} \frac{z_{nj}}{P_{ni}}.$$

For a GNN-DCM with K_g layers,

- if i is **outside** the k_g-hop neighborhood of j: the elasticity is the **same for all such** i — like MNL/IIA (proportional);



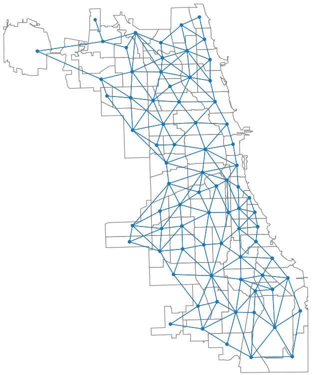
- if i is **within** k_g hops of j: the elasticity is alternative-specific, and depends on its k_g-hop neighborhood — like NL/SCL (non-proportional).



Case study: residential location choice in Chicago

Case study: 77 communities in Chicago

- **77** community areas; **3,838** households (CMAP *My Daily Travel* survey, 2017).
- **13** attributes: housing, land use, transportation, demographics.
- Construct edges based on **spatial adjacency** of communities.
- Compare MNL, SCL, ASU-DNN (Wang et al., 2020), and GNN-DCM with 1–3 layers.
- We use Graph Attention (GAT) (Veličković et al., 2018) for the message passing in GNN-DCM.



Alternative graph of 77 communities in Chicago

Predictive performance

	MNL	SCL	ASU-DNN	GNN-1	GNN-2	GNN-3
Log-likelihood	-1342.1	-1342.1	-1319.0	-1308.5	-1306.7	-1309.4
Accuracy	12.1%	12.1%	12.7%	12.6%	13.0%	**13.3%**
Top-5 accuracy	36.4%	36.4%	37.5%	37.8%	**38.8%**	38.2%
Avg. distance [km]	7.38	7.38	7.43	7.01	**6.97**	7.01
Mean recip. rank	0.250	0.250	0.258	0.279	**0.285**	0.284

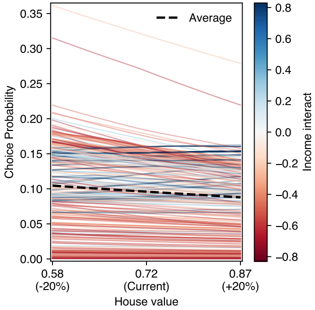
- Neural models $>$ classical: nonlinearity $+$ interactions (Wang et al., 2024).
- GNN $>$ ASU-DNN: message passing yields **spatially coherent** predictions (smaller avg. distance).
- SCL $\approx$ MNL here ($\mu \to 1$) (Sener et al., 2011).

Interpretability: MNL coefficients

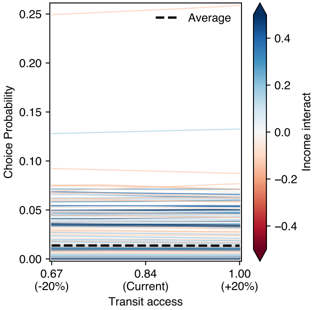
Attribute	Param.	t	Attribute	Param.	t
# Units	1.972	18.85	Transit access	−1.473	−8.09
House value	−0.322	−2.90	Work distance	−0.908	−21.63
House age	0.782	7.23	Pop density	2.341	17.12
Land mixture	−0.944	−8.96	Black interact	4.001	20.02
% Single house	−0.802	−5.65	White interact	2.812	12.34
% Multi house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

Signs match intuition: households prefer more units, lower house value, higher density, shorter commute, and same-race communities.

Interpretability: ICE plots for GNN-DCM



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

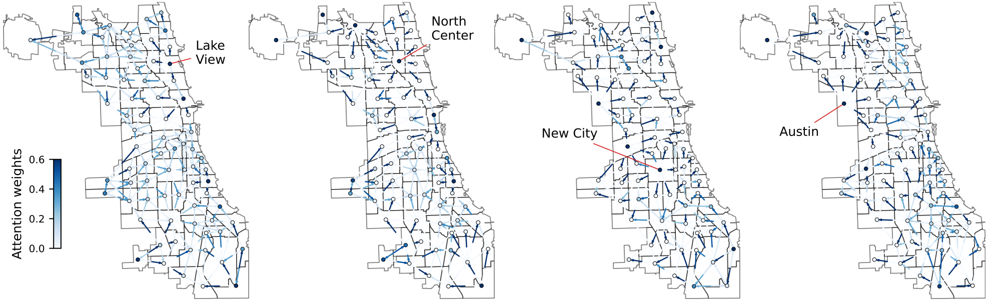


(b) ICE plot of choosing South Shore by transit accessibility.

The Individual Conditional Expectation (ICE) plots show:

- High-income households are far less sensitive to house value.
- Transit access barely moves most choices of South Shore.

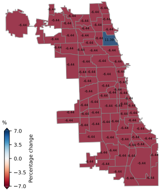
What the attention learned



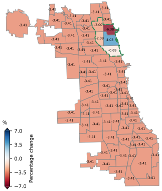
Attention weights, layer 1 of the GAT (Veličković et al., 2018) (heads 1–4); arrows show direction, node color is self-attention.

- Learned dependence is **directional and asymmetric** ($i \to j \neq j \to i$).
- A few communities strongly influence neighbors (Lake View, North Center, New City, Austin).
- Goes **beyond** the symmetric correlation NL/SCL can express.

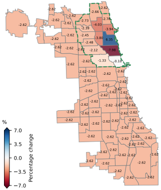
Spatially-aware substitution



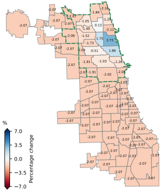
MNL ($k_g = 0$)



GNN-1



GNN-2



GNN-3

% change in choice probability after a **10% rise in housing units in Lake View**; dashed line = k_g-hop boundary.

- MNL: **same everywhere** (IIA).
- GNN: cross-elasticities vary among communities within k_g hops of Lake View.

Conclusion & outlook

Takeaways & outlook

What to remember

- GNN is a natural bridge for NN-DCMs to capture alternatives' dependencies;
- NL and SCL are single-layer GNN-DCMs;
- Substitution patterns are spatially-aware, depending on the k_g-hop neighborhood.

Code & data: https://github.com/chengzhanhong/GNN_residential_choice

Where this goes next:

- Richer, learned, or multi-relational alternative graphs (beyond adjacency).
- Correlation among *individuals* (Villarraga and Daziano, 2025) *and* alternatives, jointly.
- Joint residence–mode models; embedding GNN-DCM in the broader random utility / GEV theory.

Thank you!

Questions & discussion welcome.

zhanhong.cheng@zju.edu.cn · chengzhanhong.github.io

References i

Ben-Akiva, Moshe and John L Bowman (1998). **"Integration of an activity-based model system and a residential location model".** In: *Urban Studies* 35.7, pp. 1131–1153.

Bhat, Chandra R and Jessica Guo (2004). **"A mixed spatially correlated logit model: formulation and application to residential choice modeling".** In: *Transportation Research Part B: Methodological* 38.2, pp. 147–168.

Bhat, Chandra R and Jessica Y Guo (2007). **"A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels".** In: *Transportation Research Part B: Methodological* 41.5, pp. 506–526.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2016). **"Deep residual learning for image recognition".** In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.

Kipf, Thomas N and Max Welling (2017). **"Semi-supervised classification with graph convolutional networks".** In: *International Conference on Learning Representations*.

McFadden, Daniel (1974). **"Conditional logit analysis of qualitative choice behavior".** In: *Frontiers in Econometrics*. Ed. by Paul Zarembka. Academic Press, pp. 105–142.

McFadden, Daniel (1978). **"Modelling the choice of residential location".** In: *Transportation Research Record* 673, pp. 72–77.

Perez-Lopez, Jose-Benito, Margarita Novales, and Alfonso Orro (2022). **"Spatially correlated nested logit model for spatial location choice".** In: *Transportation Research Part B: Methodological* 161, pp. 1–12.

Sener, Ipek N, Ram M Pendyala, and Chandra R Bhat (2011). **"Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modeling residential location choice behavior".** In: *Journal of Transport Geography* 19.2, pp. 294–303.

Sifringer, Brian, Virginie Lurkin, and Alexandre Alahi (2020). **"Enhancing discrete choice models with representation learning".** In: *Transportation Research Part B: Methodological* 140, pp. 236–261.

Tomlinson, Kiran and Austin R Benson (2024). **"Graph-based methods for discrete choice".** In: *Network Science* 12.1, pp. 21–40.

Veličković, Petar et al. (2018). **"Graph Attention Networks".** In: *International Conference on Learning Representations*.

Villarraga, Daniel F and Ricardo A Daziano (2025). **"Designing Graph Convolutional Neural Networks for Discrete Choice with Network Effects".** In: *arXiv preprint arXiv:2503.09786*.

Wang, Shenhao, Baichuan Mo, and Jinhua Zhao (2020). **"Deep neural networks for choice analysis: Architecture design with alternative-specific utility functions".** In: *Transportation Research Part C: Emerging Technologies* 112, pp. 234–251.

Wang, Shenhao et al. (2024). **"Comparing hundreds of machine learning and discrete choice models for travel demand modeling: An empirical benchmark".** In: *Transportation Research Part B: Methodological* 190, p. 103061.

Wong, Melvin and Bilal Farooq (2021). **"ResLogit: A residual neural network logit model for data-driven choice modelling".** In: *Transportation Research Part C: Emerging Technologies* 126, p. 103050.

Appendix: which design choices matter?

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

- **Hidden dimension** matters most (≈ 64 suffices).
- **Depth** (K_g): far less sensitive.
- **Skip connections** consistently help (He et al., 2016; Wong and Farooq, 2021).
- Aggregation (LSE vs. sum/mean/max): minor.

Appendix: GNN update functions

Message passing NN (MPNN). $\mathbf{h}_i^{(k)} = \sigma\big(\mathbf{W}\mathbf{h}_i^{(k-1)} + \bigoplus_{j\in\mathcal{N}(i)} \mathbf{W}\mathbf{h}_j^{(k-1)}\big)$, $\bigoplus \in \{\text{sum, mean, max, LSE}\}$.

Graph convolution (GCN) (Kipf and Welling, 2017). $\mathbf{h}_i^{(k)} = \sigma\big(\sum_{j\in\mathcal{N}(i)\cup\{i\}} \frac{1}{\sqrt{|\mathcal{N}(i)||\mathcal{N}(j)|}}\mathbf{W}\mathbf{h}_j^{(k-1)}\big)$.

Graph attention (GAT) (Veličković et al., 2018). $\mathbf{h}_i^{(k)} = \sigma\big(\sum_{j\in\mathcal{N}(i)\cup\{i\}} \alpha_{ij}\mathbf{W}\mathbf{h}_j^{(k-1)}\big)$, with learned attention weights α_{ij}.

Gated skip connection (He et al., 2016). $\mathbf{h}_i^{(k)} = \sigma\big((1-c)\mathbf{h}_i^{(k-1)} + c\,(\text{GNN update})\big)$, $c \in (0,1)$ learnable.