Modeling Residential Location Choice with Graph Neural Networks

Zhanhong Cheng

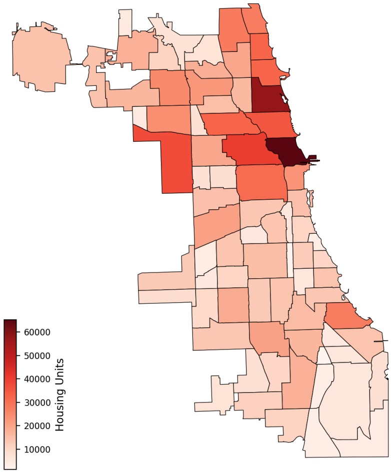
Postdoctoral Associate, University of Florida

zhanhong.cheng@ufl.edu

Co-authors: Shenhao Wang, Lingqian (Ivy) Hu

MIT SRB Seminar on December 11, 2024

Residential location choice – the problem

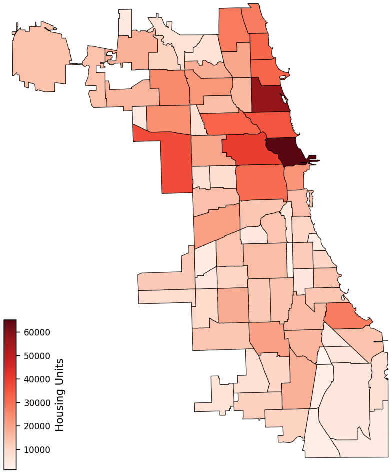


Where to live?

Factors influencing the choice:

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

Residential location choice – literature



Where to live?

Existing models:

- Multinomial Logit (MNL) (McFadden 1972).
- Nested Logit (NL) (McFadden 1978).
- Spatially Correlated Logit (SCL) (Bhat and Guo 2004).
- Other variants of NL and SCL (Sener et al. 2011; Perez-Lopez et al. 2022).

Location choice models: from trees to graphs

Residential location choice – MNL

The utility of an individual n choosing the location i is

$$U_{ni} = V_{ni} + \varepsilon_{ni},$$

where the observed utility V_{ni} is a linear function of features $\boldsymbol{x}_{ni}$ from the alternative and the individual, e.g.,

$$V_{ni} = \boldsymbol{\alpha}^\top \boldsymbol{x}_{ni}.$$

Assume independent random component ε_{ni}. Then, maximizing the utility among alternatives gives a closed-form choice probability:

$$P_{ni} = \frac{\exp(V_{ni})}{\sum_j \exp(V_{nj})},$$

Residential location choice – Nested logit (NL)

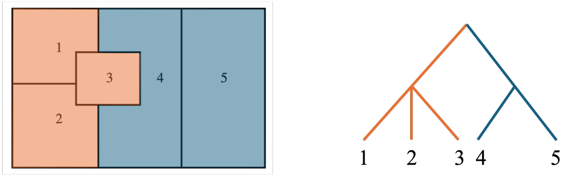


Figure 1: The tree structure in NL.

- Each group of alternatives belongs to a nest.

The choice probability of NL (McFadden 1978):

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

Residential location choice – spatially correlated logit (SCL)

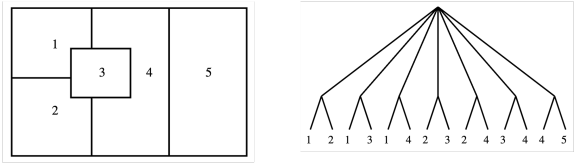


Figure 2: The tree structure in SCL.

- Each pair of neighbors belongs to a nest.

The choice probability of SCL (Bhat and Guo 2004):

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

Residential location choice – GNN

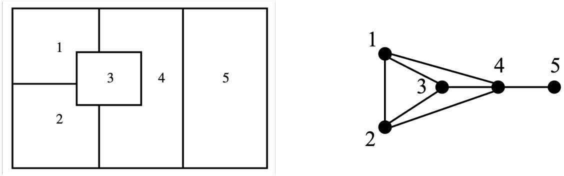


Figure 3: The graph structure in GNN. A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with nodes $\mathcal{V}$ and edges $\mathcal{E}$.

The choice probability of a K-layer GNN model:

$$P_{ni} = \frac{\exp(V_{ni})}{\sum_{j \in \mathcal{V}} \exp(V_{nj})}, V_{ni} = \mathbf{w}^T \mathbf{h}_{ni}^{(K)},$$

$$\mathbf{h}_{ni}^{(k+1)} = \text{Update}^{(k)} \left(\mathbf{h}_{ni}^{(k)}, \text{Aggregate}^{(k)} \left(\left\{ \mathbf{h}_{nj}^{(k)}, \forall j \in \mathcal{N}(i) \right\} \right) \right), \forall k \in \{0, \ldots, K-1\}.$$

- Here $\mathbf{h}_{ni}^{(0)} \equiv \mathbf{x}_{ni}$, and $\mathcal{N}(i)$ are neighbors of node i.
- the Aggregate and Update functions enable message passing among neighbors.

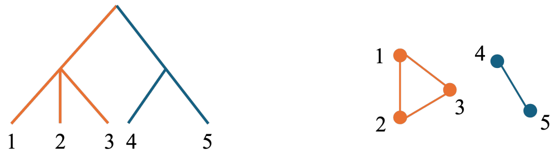


Figure 4: Correspondence between NL and GNN.

Proposition

A two-level nested logit model with each alternative belonging to one nest is a single-layer GNN. Each nest corresponds to a complete subgraph, and there is no edge between nests.

The connection between GNN and NL – proof

Proof.

$$
\begin{aligned}
P_{ni} &= \frac{\exp\left(V_{ni}/\mu_k\right)}{\sum_{j\in B_k}\exp\left(V_{nj}/\mu_k\right)} \times \frac{\left(\sum_{j\in B_k}\exp\left(V_{nj}/\mu_k\right)\right)^{\mu_k}}{\sum_{l=1}^{K}\left(\sum_{j\in B_l}\exp\left(V_{nj}/\mu_l\right)\right)^{\mu_l}} \\
&= \frac{\exp\left(V_{ni}/\mu_k\right)\left(\sum_{j\in B_k}\exp\left(V_{nj}/\mu_k\right)\right)^{\mu_k-1}}{\sum_{l=1}^{K}\left(\sum_{j\in B_l}\exp\left(V_{nj}/\mu_l\right)\right)^{\mu_l}} \\
&= \frac{\exp\left(V_{ni}/\mu_k\right)\left(\sum_{j\in B_k}\exp\left(V_{nj}/\mu_k\right)\right)^{\mu_k-1}}{\sum_{l=1}^{K}\left(\frac{\sum_{m\in B_l}\exp(V_{nm}/\mu_l)}{\sum_{m\in B_l}\exp(V_{nm}/\mu_l)}\left(\sum_{j\in B_l}\exp\left(V_{nj}/\mu_l\right)\right)^{\mu_l}\right)} \\
&= \frac{\exp\left(V_{ni}/\mu_k\right)\left(\sum_{j\in B_k}\exp\left(V_{nj}/\mu_k\right)\right)^{\mu_k-1}}{\sum_{l=1}^{K}\sum_{m\in B_l}\left(\frac{\exp(V_{nm}/\mu_l)}{\sum_{m\in B_l}\exp(V_{nm}/\mu_l)}\left(\sum_{j\in B_l}\exp\left(V_{nj}/\mu_l\right)\right)^{\mu_l}\right)} \\
&= \frac{\exp\left(V_{ni}/\mu_k + (\mu_k - 1)\log\left(\sum_{j\in B_k}\exp\left(V_{nj}/\mu_k\right)\right)\right)}{\sum_{l=1}^{K}\sum_{m\in B_l}\exp\left(\underbrace{V_{nm}/\mu_l + \overbrace{(\mu_l - 1)\log(\sum_{j\in B_l}\exp\left(V_{nj}/\mu_l\right))}^{\text{Aggregate}}}_{\text{Update}}\right)}
\end{aligned}
$$

□

The connection between GNN and SCL

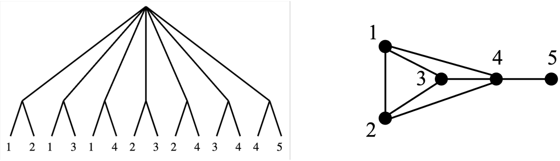


Figure 5: Correspondence between SCL and GNN.

- SCL is also a special case of GNN, where each nest in SCL corresponds to an edge in GNN.
- The corresponding GNN's Update function on vertex i is: $\log\left(\sum_{j\in\mathcal{N}(i)}\left(\alpha_{i,ij}\mathrm{e}^{V_{ni}}\right)^{1/\mu}\left[\left(\alpha_{i,ij}\mathrm{e}^{V_{ni}}\right)^{1/\mu}+\left(\alpha_{j,i}\mathrm{e}^{V_{nj}}\right)^{1/\mu}\right]^{\mu-1}\right)$.

From trees to graphs

- The tree structure has an equivalent form of massage passing in graphs.
- The GNN is a generalization of the NL and SCL.
- The GNN is more flexible in the number of layers, the choice of Aggregate and Update functions, graph structure, etc.
- The GNN naturally integrates with other deep learning models.

Case study on Chicago my daily travel survey data

The dataset

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
Ours	GNN	Chicago	77 communities	3838

- **Community features**: pop density, P_white, P_black, P_single_residential, P_multi_residential, P_office, P_retail, land_mix, transit_a_scaled, median_house_age_scaled, median_value_scaled, h_units_scaled, median_income_scaled, distance_to_work.
- **Household features**: hh_income_scaled_interact, white_interact, black_interact, (household size, whether have children or vehicles).

Study area

- **Area**: Chicago.
- **Zones**: 77 communities.

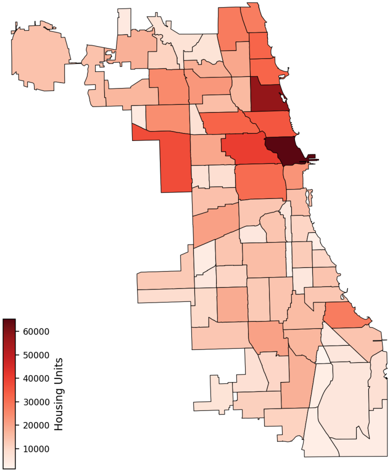


Figure 6: Housing units of 77 Communities.

The graph structure

- **Area**: Chicago.
- **Zones**: 77 communities.
- **Graph**: Assume an edge between two communities with overlapping boundaries.
- **GNN**: 2 layers, Update function uses graph attention network (GAT) (Veličković et al. 2017).

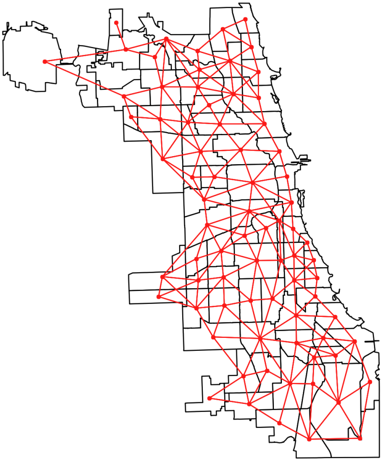


Figure 7: Graph structure of communities.

Model comparison

Table 2: The average performance from ten-fold cross-validation.

	MNL	SCL	NN	GNN
Log-likelihood	-1341.43	-1329.66	-1319.31	**-1310.92**
Accuracy	11.26%	11.98%	12.28%	**13.37%**
Top-5 accuracy	35.90%	36.02 %	38.06%	**39.40%**
F1 score	0.0153	0.0221	0.0467	**0.0514**
Mean reciprocal rank	0.2444	0.2498	0.2598	**0.2667**

- GNN outperforms NN, SCL outperforms MNL, showing the effectiveness of using spatial correlation.
- Deep learning models outperform linear models.

Model interpretation

We can easily interpret the coefficients of a linear model.

Table 3: Results of the Multinomial Logit model.

Variables	Value	t-stats	p-value
Distance to work	-0.09	-22.07	0.00**
Pop density	2.22	15.69	0.00**
Number of dwellings (scaled)	1.26	11.55	0.00**
Med house value (scaled)	-0.06	-0.46	0.64
Transit accessibility (scaled)	-0.18	-1.08	0.28
...	...	...	...

Model interpretation

We can easily interpret the coefficients of a linear model.

Table 3: Results of the Multinomial Logit model.

Variables	Value	t-stats	p-value
Distance to work	-0.09	-22.07	0.00^{**}
Pop density	2.22	15.69	0.00^{**}
Number of dwellings (scaled)	1.26	11.55	0.00^{**}
Med house value (scaled)	-0.06	-0.46	0.64
Transit accessibility (scaled)	-0.18	-1.08	0.28
...	...	...	...

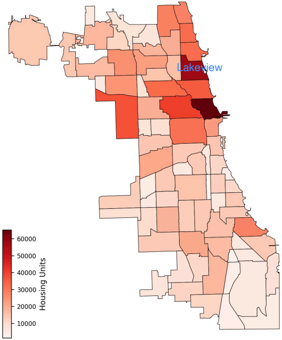
The interpretation of GNNs and other deep learning models:

- Methods: Partial Dependence Plot (PDP), Individual Conditional Expectation (ICE) plots, etc.
- Advantages: instance-level interpretability, feature interactions, etc.

GNN interpretation–housing unit median value at Lakeview

Table 4: Statistics of Lakeview.

Med house value	Med income	White prop	Black prop	Transit accessibility
398288	87330	85.8%	3.5%	0.97

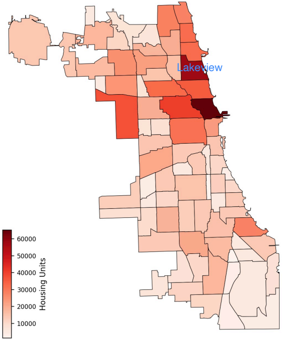


Lakeview location.

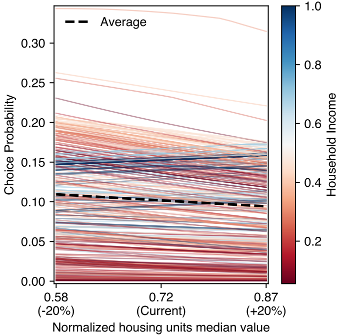
GNN interpretation–housing unit median value at Lakeview

Table 4: Statistics of Lakeview.

Med house value	Med income	White prop	Black prop	Transit accessibility
398288	87330	85.8%	3.5%	0.97



Lakeview location.

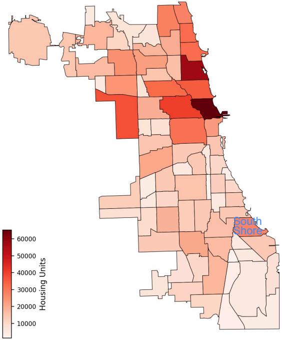


ICE plots for housing unit median value at Lakeview.

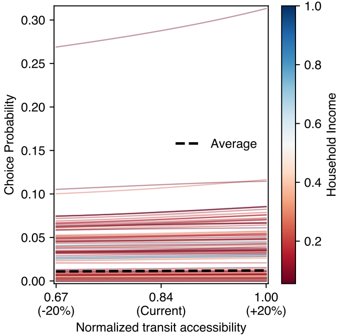
GNN interpretation–transit accessibility at South Shore

Table 5: Statistics of South Shore.

Med house value	Med income	White prop	Black prop	Transit accessibility
184142	24814	3.2%	94.5%	0.84



South Shore location.



ICE plots for transit accessibility at South Shore.

Visualizing the attention weights

- Weights are not equal.
- Conner communities have larger weights (because they have fewer neighbors).
- Stronger connections along lakeshore communities.

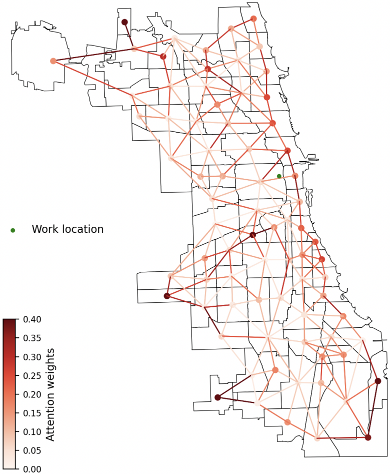


Figure 8: The attention weights of the GNN.

Summary

Table 6: The relationship between models.

	Independent alternatives	→	Correlated alternatives
Linear	Multinomial Logit	→	SCL (Nested logit)
↓	↓		↓
Nonlinear	Neural Networks	→	GNNs

Summary

Table 6: The relationship between models.

	Independent alternatives	$\rightarrow$	Correlated alternatives
Linear	Multinomial Logit	$\rightarrow$	SCL (Nested logit)
$\downarrow$	$\downarrow$		$\downarrow$
Nonlinear	Neural Networks	$\rightarrow$	GNNs

- GNN is a generalization of traditional choice models.
- Nonlinearity and alternatives' correlation are important.
- Applications: social networks, spatiotemporal correlations in mode choice, etc.

Summary

Table 6: The relationship between models.

	Independent alternatives	→	Correlated alternatives
Linear	Multinomial Logit	→	SCL (Nested logit)
↓	↓		↓
Nonlinear	Neural Networks	→	GNNs

- GNN is a generalization of traditional choice models.
- Nonlinearity and alternatives' correlation are important.
- Applications: social networks, spatiotemporal correlations in mode choice, etc.

Thank you! `zhanhong.cheng@ufl.edu`

Questions? https://chengzhanhong.github.io/

References

- 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.
- McFadden, Daniel (1972). **"Conditional logit analysis of qualitative choice behavior".** In.
- — (1978). **"Modeling the Choice of Residential Location".** en. In: *Transportation Research Record* 673.

References ii

- Perez-Lopez, Jose-Benito et al. (2022). **"Spatially correlated nested logit model for spatial location choice".** In: *Transportation Research Part B: Methodological* 161, pp. 1–12.
- Sener, Ipek N et al. (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.
- Veličković, Petar et al. (2017). **"Graph attention networks".** In: *arXiv preprint arXiv:1710.10903*.