COVID Risk Is Not Random: A Spatial Network Analysis

Georgia Institute of Technology

Holin Xue   Priscilla Zhang

Georgia Institute of Technology, H. Milton Stewart School of Industrial and Systems Engineering

Georgia Tech College of Engineering
H. Milton Stewart School of Industrial and Systems Engineering

Problem Statement

Research Question: How do socioeconomic characteristics and mobility patterns influence COVID-19 transmission across U.S. counties?

Goal: Attribute socioeconomic and mobility factors to COVID-19 transmission by quantifying their influence on disease spread.

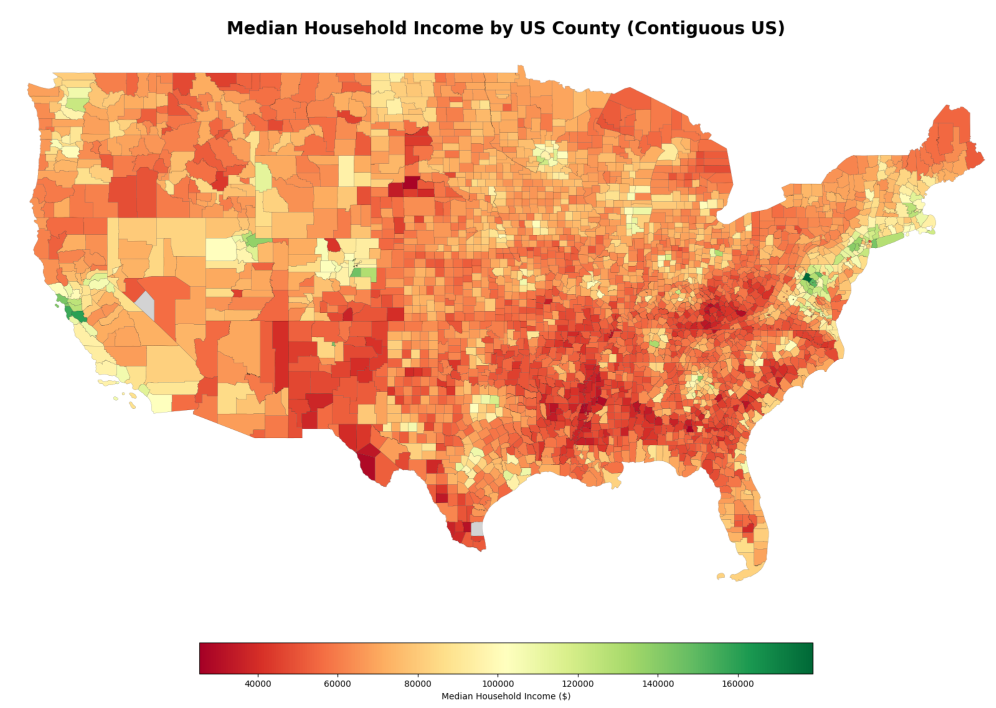


Figure 1. Median Household Income by US County

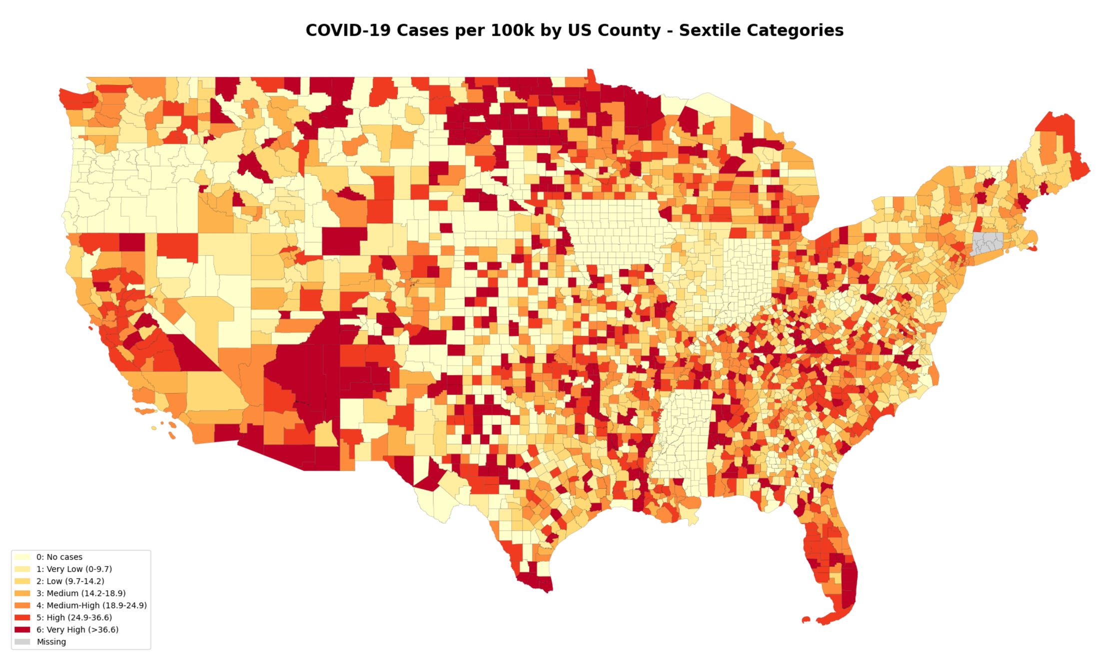


Figure 2. COVID-19 Cases per 100k by US County

Observation: Spatial overlap between low-income regions and high COVID-19 burden.

Key Findings

Model Performance:

- Baseline SIR: $R^2 = -0.028$ (fails to capture heterogeneity)
- Extended SIR: $R^2 = -4.609$ (linear parameterization fails)
- **GNN:** $R^2 = 0.441$ (only model with positive explanatory power)

Key Findings

Attribution Results:

- **Top predictors:** Median household income, residential mobility
- Structural factors (income, education, housing) and behavioral mobility jointly drive transmission
- **Main insight:** Effects are highly non-linear—requiring flexible models (GNN) to capture

Methods

Dual-Model Framework:

1. Mechanistic SIR Models:

- Baseline: Constant β, spatial diffusion
- Extended: $\beta_i = \exp(\alpha_0 + \sum_{j=1}^{18} \alpha_j \cdot X_{ij})$ from 18 SE features
- **Attribution:** Causal pathway (SE/mobility → transmission rate → disease burden)

2. Graph Neural Network (GNN):

- Counties as nodes, adjacency as edges, 23 features (18 SE + 5 mobility)
- Non-linear learning with spatial message passing
- **Attribution:** Identifies which factors matter most (predictive)

Why Both? SIR reveals *how* factors influence transmission; GNN identifies *which* factors matter most.

Data and Experimental Design

Dataset: 3,222 U.S. counties

- COVID-19 outcomes: CDC Community Levels
- 18 socioeconomic indicators: Census Bureau
- 6 mobility metrics: Google Mobility Reports
- Spatial structure: County adjacency network

Spatial SIR Framework:

$$\frac{dI_i}{dt} = \beta_i S_i \sum_{j \in \mathcal{N}(i)} w_{ij} \frac{I_j}{N_j} - \gamma I_i + m \sum_{j \in \mathcal{N}(i)} w_{ij} \left(\frac{I_j}{N_j} - \frac{I_i}{N_i} \right)$$

Baseline: $\beta_i = \beta$ (constant)

Data and Experimental Design

Extended: $\beta_i = \exp(\alpha_0 + \sum_{j=1}^{18} \alpha_j \cdot X_{ij})$

GNN: Counties as graph nodes, message passing for spatial learning, non-linear feature interactions.

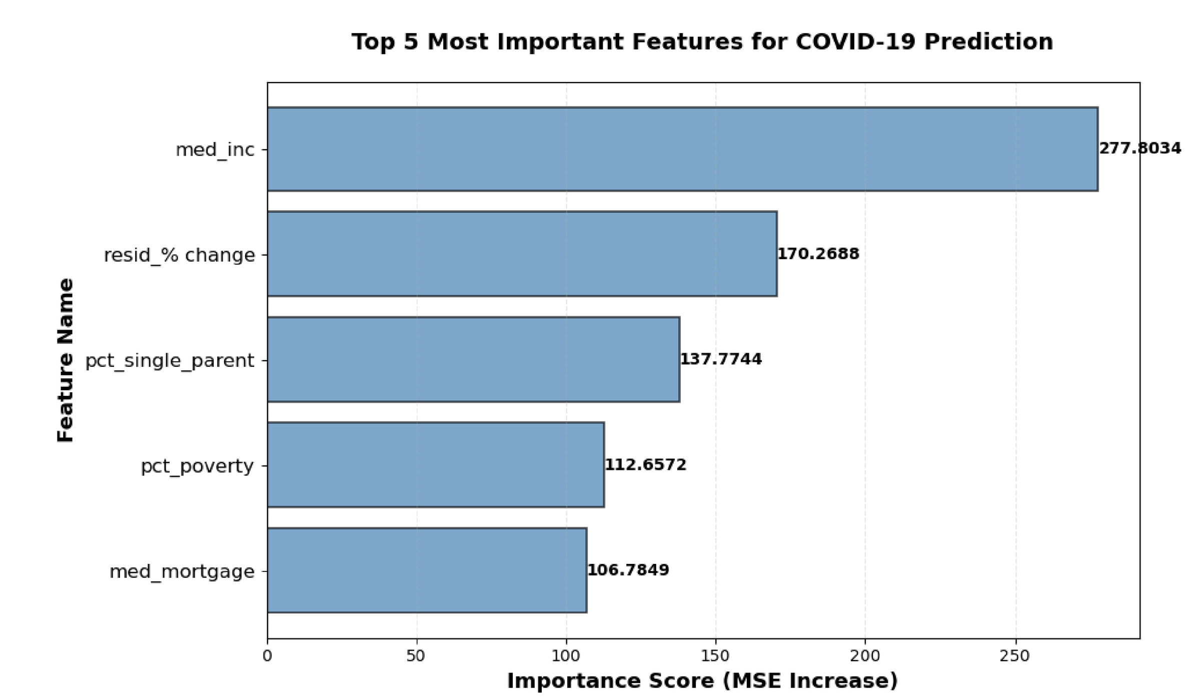
Results and Attribution

Model	MAE	RMSE	R^2
Baseline SIR	1.26	7.65	-0.028
SE-Mobility SIR	13.53	17.88	-4.609
GNN	27.48	39.59	0.441

Table 1. Predictive performance (cases per 100k). GNN explains 44.1% of variance.

Key Findings:

- **SIR models fail:** Linear parameterization cannot capture complex non-linear interactions
- **GNN succeeds:** Only model with positive R^2; learns spatial patterns and feature interactions



Conclusion: Structural socioeconomic factors (income, education, housing) and behavioral mobility patterns jointly drive COVID-19 transmission. Effects are highly non-linear, requiring flexible models for accurate attribution.

References

Khanijahani, A., Iezadi, S., Gholipour, K., Azami-Aghdash, S., and Naghibi, D. (2021). A systematic review of racial/ethnic and socioeconomic disparities in covid-19. *International Journal for Equity in Health*, 20(1):248.