Integrating Graph Neural Networks with Visualizations for Inter-County Food Trade Prediction

Saanvi Sheoran
Canyon Crest Academy, San Diego, CA 92130
Email: saanvisheoran62@gmail.com

Mentor: Dr. Mary Thomas, PhD.
San Diego Supercomputer Center (SDSC), University of California San Diego, La Jolla, CA 92130
Email: mpthomas@ucsd.edu

August 25, 2025

Abstract

The *AI Institute for Intelligent Cyberinfrastructure with Computational Learning in the Environment* (ICICLE) [13] aims to democratize artificial intelligence by building scalable tools to advance AI-driven discovery and integration across society. Within the framework provided in the ICICLE AI Training Catalog [10], I present a graph-based high-performance approach to predict and visualize inter-county food flows across the United States. Using the Expanse supercomputer [16] at SDSC, I implemented ICICLE's Graph Neural Network (GNN) model on Jupyter Notebooks to process large-scale datasets of economic, demographic, and geographic features and generate predictions for food trade flows between Freight Analysis Framework (FAF) zones. To address the sparsity of trade networks, a hurdle modeling pipeline is employed, combining a classification stage to predict trade occurrence with a regression stage to estimate trade volume. Beyond prediction, I integrated static visualizations (Matplotlib [11], NetworkX [12]) and a dynamic, interactive dashboard (Panel [8], GeoViews [6], hvPlot [7]) to enable real-time exploration of trade networks. This combination of machine learning and visualization provides policymakers and researchers with a scalable framework for interpreting complex food trade patterns, supporting data-driven decisions in food security, infrastructure planning, and economic resilience.

1 Introduction

The ICICLE project is an NSF-funded initiative focused on developing cyberinfrastructure to democratize AI through scalable frameworks that integrate diverse data sources and computational methods. By lowering technical barriers and fostering interdisciplinary collaboration, ICICLE seeks to transform AI from a specialized capability into an accessible infrastructure that supports innovation and empowers scientific, educational, and commercial domains [13]. This paper works with ICICLE's sub-project titled "Multiscale Food Flow Prediction Using Graph Neural Networks" [10] and leverages Graph Neural Networks (GNNs) and high-performance computing (HPC) resources to predict and visualize food trade flows across U.S. regions. Post model evaluation, I developed an interactive visualization dashboard on Jupyter Notebooks using Python libraries such as Panel [8], GeoViews [6], and hvPlot [7] to allow users to dynamically explore predictions, apply threshold filters, and interpret large-scale trade networks in real time. This provides a scalable tool for research, policy, and decision-making in food security and infrastructure planning. This report documents work conducted during San Diego Supercomputer Center's 2025 Research Experience for Highschool Students [17] (REHS) program. The following sections outline the motivation, background, methodology, results, and final discussion part of this project.

1.1 Motivation

Understanding and predicting food trade flows is essential for ensuring food security, managing supply chains, and supporting evidence-based policy decisions. We need accurate model predictions to capture the nonlinear, large-scale interactions that govern trade dynamics and to anticipate shifts in supply networks under different conditions. As food systems become increasingly complex and interdependent, visualizing trade dynamics on HPC-systems allows researchers and policy-makers to analyze and interpret high-volume food flow data across the nation.

1.2 Background

Graph structured data is becoming increasingly common across a wide variety of fields and is defined as a network with different entities represented as nodes connected by their relationships (edges). Unlike tabular or sequential data, graph-structured data naturally encodes complex interdependencies, making it particularly useful for modeling systems where relationships are just as important as the individual features of entities. Food trade networks are an example of this, as counties act as nodes and the movement of food between them can be expressed as weighted edges.

There are many variations of GNNs, including Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) [2]. GCNs are models designed to learn from graph-structured data by iteratively aggregating information about a node from nearby nodes. GATs, on the other hand, extend GCNs by applying attention mechanisms onto the network, which allows the model to weight different node-edge relationships higher than others instead of treating them all uniformly. The ICICLE project framework [5] referenced in this project included existing GCN and GAT architectures, though I chose to utilize the GCN in this work because I wanted to prioritize overall structural connections between the different source zones in the network instead of focusing on specific, high-impact trade flows.

Recent research has demonstrated the applicability of GNNs across diverse applications. An example of this is investigating GNN applications for Bio-informatics. In 2023, Zhang et al. used GNNs to model protein–protein interactions, gene regulatory networks, and molecular structures for tasks such as function prediction and disease association studies. In this setting, biological entities such as proteins, genes, or molecules are represented as nodes, while their relationships—such as interactions, regulatory links, or chemical bonds—are represented as edges, allowing GNNs to capture the underlying topology of complex biological systems [18]. Additionally, in 2023 Yaniv et al. applied the node-edge architecture employed by GNNs towards power flow estimation and solving power optimization problems for electrical distribution networks [1]. These works, among others, showcase the versatility of GNNs beyond biological and social domains.

2 Methodology

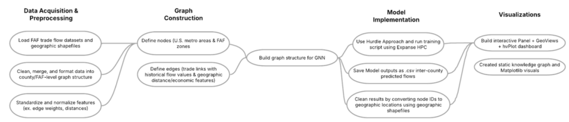

Figure 1: Methodology Pipeline

I began by downloading the GNN Food Flow Github repository [5] provided by the ICICLE team and uploading it to the Expanse [16] supercomputer. Next, I created a new Conda environment so that I could run the Jupyter Notebooks provided from ICICLE and installed the necessary dependencies to run my code (Pytorch [4], Pandas [15], Numpy [14], MatPlotLib [11], GeoViews [6], hvPlot [7], Panel [8], NetworkX [12]).

I first had to reorganize the folder structures and gather some missing shapefiles from the US Census Bureau [3] in order to run the training.ipynb notebook. Once this was done, I ran the preprocessing code [5] on Expanse [16], implemented a Graph Convolutional Network (GCN), and saved the model predictions in a .csv file.

A GCN was chosen for this project because food trade networks are inherently graph-structured, with counties represented as nodes and trade flows as weighted edges. GCNs are uniquely effective at capturing spatial and relational dependencies, allowing the model to compute information across neighboring regions and better reflect the interdependent nature of food systems.

While Graph Attention Networks (GATs) offer enhanced node-specific weighting through attention mechanisms, I chose to use a Graph Convolutional Network (GCN) due to its computational efficiency and proven performance on large-scale, homogeneous graph data [2]. I felt this made it well-suited for modeling nationwide food trade networks where it is important to consider all the receiving zones around the source zone.

The GCN implemented in this paper followed a two-step hurdle model approach [9]: the model first classifies whether trade occurs or not, and if trade occurs, then predicts how much trade occurs. A pipeline explanation of this process is displayed below in Figure 2.

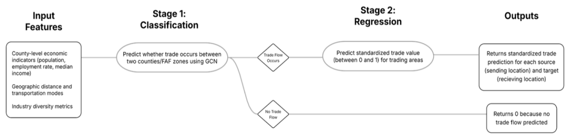

Figure 2: Hurdle Model Pipeline

Initially, I developed a knowledge graph using NetworkX [12] to show the node and edge connections illustrated by the GCN, though this visualization lacked geographic context and interactivity, which limited its effectiveness in conveying real-world trade networks. Figure 3 featured in the Results section shows the knowledge graph.

Then, I wanted to transform the nodes and edges into geographic locations to better show the relationship between food source zones and food target zones. I coded dictionaries using Python to map the node IDs created by the GCN back to the original source names and latitude/longitude data from the US Census Bureau shapefiles [3]. I integrated MatPlotLib [11] and Panel [8] together to build a dashboard where users can select a source zone from a dropdown and visualize a map of the US with food networks moving from sources to targets. The lines are colour-coordinated based on prediction strength, with the darker lines indicating stronger predictions. Though this visualization was more useful than the knowledge graph, the maps also had limited interactivity and numerous network lines felt visually overwhelming. This knowledge graph is shown in Figure 4 in the Results section.

Finally, I built an interactive dashboard using Panel [8], GeoViews [6], and hvPlot [7] to allow users to dynamically analyze and read the food flow predictions. The GeoViews library was used to create an interactive geographic map, hvPlot was used to efficiently plot food flow networks using the prediction data, and Panel was used to join all the features together into one platform. Keeping some functionality from the Panel+Matplotlib approach, this visualization allows users to select a source zone from a dropdown menu and then visualize all the trade flow lines coming from that source. Hovering on a specific flow will visualize the prediction strength, source location, and target location. Users can interact with the geographic map and zoom in on different locations. I also added a prediction threshold bar where viewers can filter the predictions based on strength. Figures 5 and 6 in the Results section show an overview of the dashboard as well as a zoomed in view.

3 Results

Below is the initial knowledge graph visualization as described in the Methodology Section.

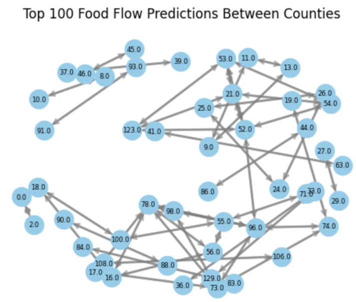

Figure 3: Initial Knowledge Graph using NetworkX

Below is the MatPlotLib visualization as described in the Methodology Section.

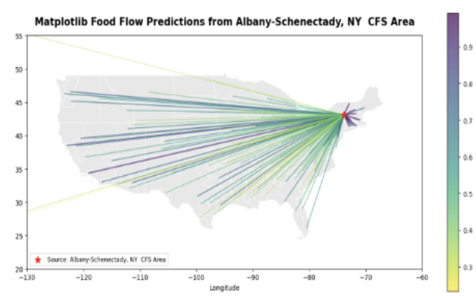

Figure 4: MatPlotLib Static Visualization

Below is the final interactive dashboard using Panel, GeoViews, and hvPlot.

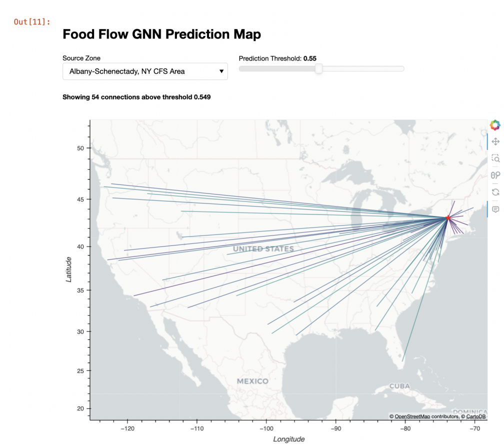

Figure 5: Overview of Interactive Dashboard

Below is a zoomed in display of the final interactive dashboard.

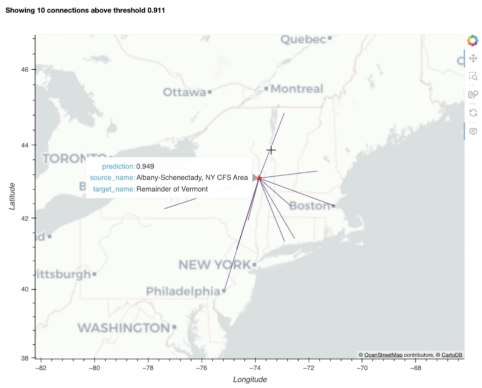

Figure 6: Zoomed in Display of Dashboard

4 Discussion

The results of this project demonstrate the usability of combining Graph Neural Networks (GNNs) with interactive visualization tools to interpret large-scale food trade predictions. The application of a Graph Convolutional Network (GCN) successfully captured spatial and relational dependencies across U.S. counties, allowing for structured modeling of complex food flows. The hurdle model [9] pipeline effectively addressed potential sparsity issues within the trade networks by first classifying whether trade occurred and then estimating the volume of trade flow that occurred.

While the model produced meaningful outputs, several challenges were encountered throughout the project. Missing files and shapefiles in the provided ICICLE GitHub repository required external sourcing and manual corrections. Additionally, a code mismatch in the original training pipeline led to implementation issues, which were eventually resolved through communication with the ICICLE team, who pushed a fix.

The initial static visualizations—such as the knowledge graph and Matplotlib map—were useful for validating output but limited in interactivity and scalability. These limitations motivated me to build a Panel-based interactive dashboard, which allowed for real-time filtering and geographic interpretation. This enhanced visualization experience offered users a more intuitive understanding of the model's predictions, particularly when zooming into specific source-target trade relationships or applying prediction strength filters.

Despite these successes, there are opportunities for improvement. The current GCN treats all neighboring node contributions equally, which may limit model precision in regions where certain trade routes are disproportionately influential. Incorporating attention mechanisms, as enabled by Graph Attention Networks (GATs), may offer improved performance by weighting edges according to their relative importance. Additionally, integrating temporal data or more granular socioeconomic variables could further enhance model sensitivity to real-world trade dynamics.

5 Conclusion

As climate change and disruptions of the supply chain challenge food security, understanding how food moves between regions has become increasingly important. This project leverages Graph Neural Networks trained and executed on high-performance computing systems to model and predict complex intercounty trade flows across the US. By integrating ICICLE's large-scale GNN models with Expanse at SDSC [16], we are able to process massive datasets of economic and geographic features and scale predictions to the national level. Coupled with interactive visualization tools in Jupyter Notebooks, these predictions allow researchers to dynamically filter, explore, and interpret food-trade networks in real time. This combination of HPC-driven machine learning and geographic visualization provides a scalable framework for analyzing trade patterns and supports advanced research into food security and economic resilience.

References

[1] Yuval Beck Arbel Yaniv, Parteek Kumar. Towards adoption of gnns for power flow applications in distribution systems. *Computers and Electronics in Agriculture*, 2022. Discussion of neural methods for power flow.

[2] Joyce Birkins. Comprehensive guide to gnn, gat, and gcn: A beginner's introduction to graph neural networks after reading 11 gnn papers. *Medium*, 2024. Accessed: 2025-08-20.

[3] United States Census Bureau. Cartographic boundary files, 2025. Accessed: 2025-08-20.

[4] PyTorch Contributors. Pytorch documentation, 2025. Accessed: 2025-08-20.

[5] GeoDS. Gnnfoodflow. https://github.com/GeoDS/GNNFoodFlow, 2025.

[6] HoloViz Team. Geoviews user guide. https://geoviews.org/user_guide/, 2025.

[7] HoloViz Team. hvplot documentation. https://hvplot.holoviz.org/en/docs/latest/, 2025.

[8] HoloViz Team. Panel documentation. https://panel.holoviz.org/, 2025.

[9] ICICLE AI Institute. How-to implement hurdle model for trade prediction: Multi-scale food flow prediction using graph neural networks, 2025. Accessed: 2025-08-20.

[10] ICICLE AI Institute. Multi-scale food flow prediction using graph neural networks. https://icicle-ai.github.io/training-catalog/docs/category/multi-scale-food-flow-prediction-using-graph-neural-networks-2, 2025.

[11] Matplotlib Development Team. Matplotlib documentation. https://matplotlib.org/stable/index.html, 2025.

[12] NetworkX Developers. Networkx reference manual. https://networkx.org/documentation/stable/reference/index.html, 2025.

[13] NSF ICICLE Consortium. Intelligent cyberinfrastructure with computational learning in the environment (icicle). https://icicle.osu.edu/, 2025. NSF-funded AI-CI initiative led by The Ohio State University to democratize AI through cyberinfrastructure.

[14] NumPy Developers. Numpy documentation, 2025. Accessed: 2025-08-20.

[15] pandas Development Team. pandas documentation, 2025. Accessed: 2025-08-20.

[16] San Diego Supercomputer Center. Expanse. Service. University of California San Diego, 2025.

[17] San Diego Supercomputer Center StudentTech. Ucsd research experience for high school students (rehs). https://education.sdsc.edu/studenttech/rehs/, 2025. Eight-week research internship providing computational science exposure, mentoring, and outreach at UCSD's SDSC.

[18] Xiao-Meng Zhang, Li Liang, Lin Liu, and Ming-Jing Tang. Graph neural networks and their current applications in bioinformatics. *Frontiers in Genetics*, 12:690049, 2021.