

Graph Theory Techniques for Analyzing City Structure Regarding Socioeconomic Factors, Real Estate Market, and Climate Change

Yunzhong Ding 

College of Art & Science, University of North Carolina, Chapel Hill, USA
Email: dingpropro@gmail.com, dingding@unc.edu

How to cite this paper: Ding, Y. (2025) Graph Theory Techniques for Analyzing City Structure Regarding Socioeconomic Factors, Real Estate Market, and Climate Change. *Journal of Data Analysis and Information Processing*, 13, 487-503.
<https://doi.org/10.4236/jdaip.2025.134028>

Received: September 12, 2025
Accepted: November 1, 2025
Published: November 4, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The primary purpose of this study is to present mathematical modeling methods inspired by graph theory operations and logic as a tool to structurally analyze the socio-economic composition of a city based on the geographical location of the investigated areas. We've incorporated graph theory concepts like connectivity, subgraph, degree, tree, complete graph, and dual graph as our model's main components. We applied these methods to study the geographical distribution of food hardship in New York City, as well as housing prices in Boise, Idaho, and Miami, Florida. We conducted a structural analysis of our models and concluded several notable properties within the model results. We've also included the ocean's current direction and location for the New York City model to speculate on the mechanism behind our results further. Graphs and quantitative data regarding each one of these factors are simulated and created through Gephi and R Studio, where the combination of these factors will be concluded and presented as the result of the study. In this way, the result of our model brings a step-by-step demonstration of how the graph theory and analysis techniques we've developed can be applied to any city with proper quantitative or qualitative data. The report in our prototype model focuses on population poverty and socio-economic conditions, emphasized and indicated through food hardship and the housing market within the area. We also discussed several plausible applications of our methods, including topics like climate change and the real estate market. Considering that our model is a skeletal position-based map exhibiting the functionality of the analysis techniques we've developed, the graph is a prototype for environmental science and mathematics researchers to examine, where they can further improve and optimize it for more accurate and informative results.

Keywords

Graph Theory, Modeling Techniques, Mathematical Modeling, Location-Based Data, City Structure Analysis, Socioeconomic Analysis

1. Introduction

1.1. What Is Graph Theory

Graph theory studies graphs consisting of elements like vertices, edges, faces, colors, cycles, and paths. With simple yet straightforward edges connecting the vertices of a graph, the relationship between objects can be presented in a skeletal manner, and the model can provide a direct relationship among each set of vertices. A graph's connectivity refers to the number of edges that must be removed for a connected graph to become separated [1].

1.2. What Is Climate Change

According to the definition provided by the Intergovernmental Panel on Climate Change (IPCC, 2024), climate change is defined as “a change in the state of the climate that can be identified by changes in the mean and/or the variability of its properties and that persists for an extended period, typically decades or longer”. Such a change can be induced by natural internal processes or external factors related to solar cycles or volcanic activities. Furthermore, the Framework Convention on Climate Change has indicated that human activity also, directly and indirectly, contributes to climate change by altering the composition of the global atmosphere, leading to natural-cause-induced climate variability such as ocean acidification and global warming [2].

Considering the devastating yet somewhat irreversible effects of climate change, the COP, Conference of the Parties, has established policies discussed by the participating nations around the globe to promote effective environmental strategies combating its negative impact. The primary purpose of the COP is to review policies proposed by the Parties. The COP meets yearly to promote effective communications, seeking the most promising Convention or legal instrument to implement [3].

1.3. Possible Impacts on Coastal Real Estate

In the United States, the Environmental Protection Agency is responsible for keeping the environment safe [4]. Socially, environmental justice advocates have argued that individuals, regardless of their race, income, or origin, should enjoy the same rights to access proper ecological protections and benefits [5]. Therefore, it is paramount to investigate the correlation between the possible impacts of climate change on the coastal real estate market and the socio-economic and environmental mechanisms behind the market dynamics. Researchers from the Natural Resources Defense Council have suggested that people of color and poorer

individuals were the predominant groups forced to leave the coastal area due to the lack of environmental protection available in their community, thus stripping their right to reside in the coastal area without huge financial burdens and excessive repercussions regarding their living condition due to the deteriorating level of clean air and clean water. Note that a similar conclusion can be drawn about the elderly portion of society.

Access to clean water and clean air is one of many issues preventing individuals from rightfully and properly residing in the coastal area. Hurricanes have caused major environmental and economic disruptions [6]. Summing the damage caused by Hurricane Harvey in Texas, Florence in North Carolina, Michael in Florida, and Ida in Louisiana, researchers have calculated that these hurricanes have caused \$300 million in damage and a death toll of 287 in the corresponding coastal areas. Such a phenomenon indicated that the frequency and intensity of possible hurricanes must be considered when evaluating the real estate market of the selected coastal area. The same devastating impact can be said regarding the rising sea level, swallowing the adjacent land [7]. A study published by the Washington Post has stated that about 25,000 properties in coastal Louisiana could find themselves below the tidal boundary lines by the year 2050, sinking these houses under the sea. In this way, the longevity of the properties would pose significant concerns for their potential residents, and the market would be perceived as skeptical by those buyers, thus hindering the real estate market in those areas.

Last but not least, saltwater intrusion has contaminated the clean water source for those coastal households and significantly disrupted the soil and the coastal ecosystems [8]. Migrating the marsh plants to further inland areas, thus shrinking the size of viable agricultural soil and damaging the agricultural industry run by the nearby residents. Under such notions, the environmental stability and the economic potential of the coastal real estate market would be significantly sabotaged. To prevent such disasters in the coastal real estate market, we would create a model evaluating each of these factors and concluding possible solutions for each of these phenomena, thus revitalizing the coastal real estate market, especially in a time where the demand for such properties is on an up-rise considering that more people wish to live at the coast in this day of age [9].

1.4. Plan of the Paper

The paper's plan is as follows: In Section 2, a model illustrating New York City's population distribution regarding food hardship will be presented, along with two other graphs exhibiting Miami's and Boise's housing price distribution. The presentation and the model's result will be given in Section 3. Then, in Section 4, the results and model will be discussed, along with their implications.

2. Graph Theory Model of City Structures

2.1. Beginning Model—Single City

To begin the progression of our model, we will consider only one coastal city, in

this case, New York City, in the continental United States (see **Figure 1**). From the map provided by Columbia University, we are able to gather data regarding the geographical location of areas with various levels of food hardship in New York City (see **Figure 2**). We will utilize the open-source software Gephi to simplify the map and create a conceptual graph illustrating the relationship between the rate of food hardship, divided by city regions of New York City, and ocean currents [10]. Gephi is an open-source network visualization platform. Created to be the Photoshop of network visualization, it combines a rich set of built-in functionalities and a friendly user interface aggregated around the visualization window.



Figure 1. The three cities of study, including Boise (ID), New York City (NY), and Miami (FL) (Britannica).

In this study, we utilized data collected from the Robin Hood Poverty Tracker, a longitudinal survey launched in 2012 in partnership with Columbia University. The survey follows approximately 4000 New York City households, surveyed quarterly over several years. The report draws on the most recent four years of data available through 2019, with respondents classified by their residence into the city’s 59 Community Districts, which serve as the spatial units of analysis. Each district includes an average of about 213 respondents, providing representative neighborhood-level estimates of food hardship across New York City [11].

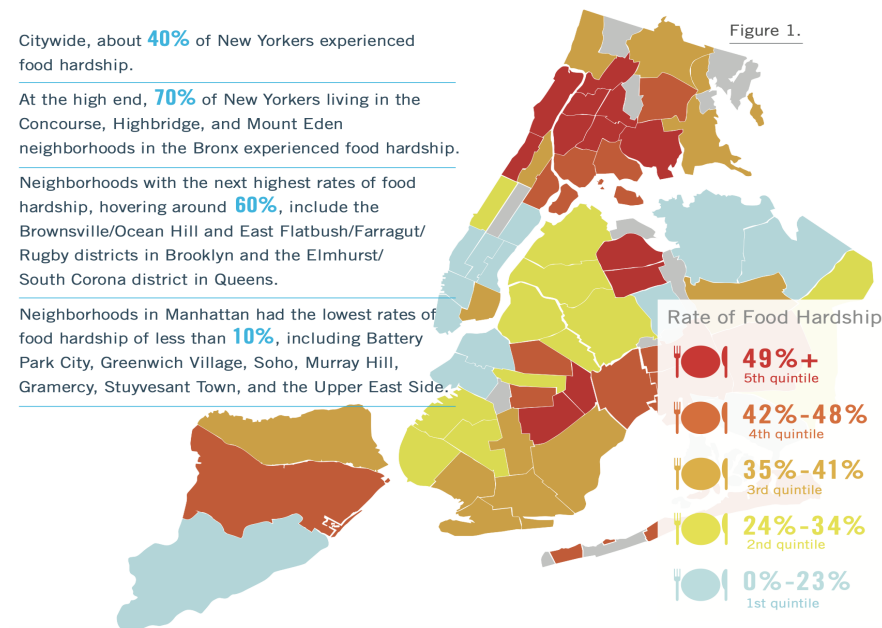


Figure 2. New York City map of Food Hardship Rate (Columbia University, 2019).

To compare the geographical pattern regarding poverty conditions in New York City, we have created Figure 3, simplifying the areas into vertices and denoting their adjacencies through edges to skeletally illustrate the geographical adjacencies of these investigated areas, thus analyzing the spatial relations and patterns regarding these vertices.

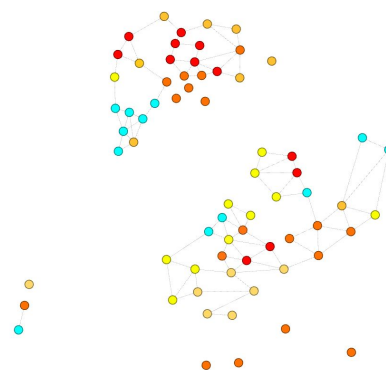


Figure 3. New York City graph of Food Hardship Rate (Created through Gephi).

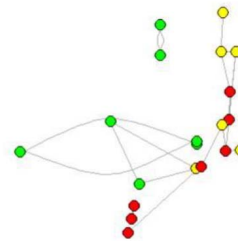
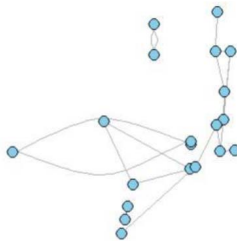
To create this graph, we considered each confined region in the map as a separate vertex and connected these vertices with edges. The adjacencies of these vertices were determined by whether the confined areas were territorially and contiguously connected or not. Individual vertices were labeled based on the data provided by the map regarding that area's quintile for food hardship rate. The gray areas were ignored due to the lack of data provided by the map, where the corresponding adjacencies were also excluded from the graph.

To further mathematically clarify the creation of the graph through graph theory, if we consider each confined surveyed region as a face, their barriers as edges, and each of the faces' endpoints or turning points of the barriers' sides as a vertex, we will simply create the dual graph G_d of such a graph G . Such a transformation can be operated by taking the faces of the original graph G , and make these faces the new vertices of G_d connected through edges determined from the shared edges in G . For the purpose of notation, we will denote the dual graph we created as G instead of G_d .

2.2. Beginning Model—Other Single Cities

It's critical to note that Gephi is not the only medium for utilizing our graph theory modeling method. To achieve a similar style of graph, we can use R Studio to construct the investigated neighborhoods based on the city's main road structure and the neighborhoods' latitude and longitude. Such a technique becomes more effective for more even-sized neighborhoods, considering that each edge represents a bridge, highway, or other forms of significant connection instead of a territorial connection or shared sides.

Graph of Miami Neighborhoods Graph of Miami Neighborhoods with Listing Prices per SqFt



Graph of Boise Neighborhoods Graph of Boise Neighborhoods with Listing Prices per SqFt

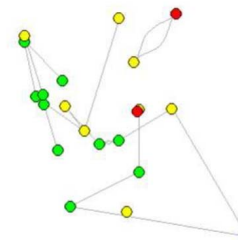
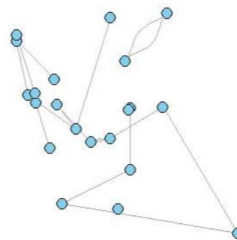


Figure 4. Miami and Boise graph of Housing Price (Created through R Studio).

Data collected by Trading Economics was used to determine the color label of

each vertex to compare and visualize the geographic distribution of housing prices in Boise and Miami (see **Figure 4**). In graph theory, vertex coloring is typically used in a concept where no two adjacent vertices share the same color. We use vertex coloring to label the quantitative value of the study's variable associated with the given vertex. To categorize each of the neighborhoods into three separate colors, we sorted each neighborhood's average listing price per square foot into a list and concluded three quintiles defined by the top 33%, the middle 34%, and the bottom 33% from the most expensive to the least, where each corresponded to red, yellow, and green. The adjacencies of such vertices were determined by the local highway and major roads in the area [12] [13]. Although the actual location of the vertices doesn't have any significant meaning or value mathematically, having the vertices placed based on the neighborhood's location can help visualize the sheer structure in a map-like fashion. Therefore, we will quantitatively cast the location data of these neighborhoods into longitude and latitude and connect the highways accordingly.

2.3. Annotated Model—New York City

To recognize sub-structures within the graph, we can manually annotate prominent Graph Theory structures among homogeneous vertices in terms of the quintile of food hardship that a neighborhood falls into (see **Figure 5**). Based on the sub-structures that we've labeled, such as path, cycle, and complete subgraph, we are able to analyze the given area of the city structurally.

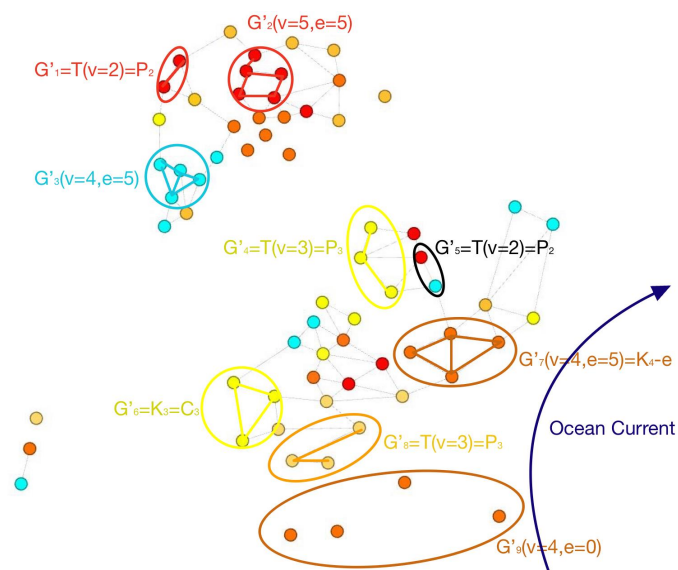


Figure 5. Annotated New York City Graph of Food Hardship Rate (Created through Gephi, Adjusted for Ocean Current Circulation).

2.4. Analytical Model—New York City

The Poverty Tracker Report from Columbia University also provided detailed quantitative data regarding each neighborhood in all five of the New York City

boroughs. An example of such data can be seen in the chart below, where data of each investigated neighborhood in the Bronx was recorded (see [Table 1](#)).

Table 1. Food hardship rate data table of the Bronx (Columbia University).

Community District Number	Community District Name	Food Hardship Rate	Margin of Error	Quintile	Sample Size
1	Melrose, Mott Haven, Port Morris	45%	+/- 11%	4	243
2	Hunts Point, Longwood	45%	+/- 11%	4	243
3	Claremont, Crotona Park East, Melrose, Morrisania	49%	+/- 9%	5	331
4	Concourse, Highbridge & Mount Eden	70%	+/- 7%	5	284
5	Morris Heights, Fordham South & Mount Hope	53%	+/- 12%	5	216
6	Belmont, Bathgate & East Tremont	49%	+/- 9%	5	331
7	Bedford Park, Fordham North & Norwood	57%	+/- 11%	5	239
8	Riverdale, Fieldston & Kingsbridge	39%	+/- 10%	3	213
9	Castle Hill, Clason Point & Parkchester	52%	+/- 7%	5	334
10	Co-Op City, Pelham Bay & Schuylerville	39%	+/- 9%	3	218
11	Pelham Parkway, Morris Park & Laconia	47%	+/- 13%	4	186
12	Wakefield, Williamsbridge & Woodlawn	35%	+/- 9%	3	184

Compiling such data for all five of the boroughs, we will be able to acquire a more comprehensive list of the food hardship rate of all community districts in New York City. Then, I constructed a new metric for this table, labeled “degree”, for each of these community districts based on their adjacencies with surrounding neighborhoods. The term degree is extracted from Graph Theory, where the number of neighborhoods that are adjacent to a given vertex is recorded as the degree of that vertex. However, simply knowing the amount of the neighborhood that is territorially connected with a district would not contribute much in terms of predicting the food hardship rate of the given district—the level of food hardship in the neighboring community district matters. Therefore, I adjusted the metric “degree” to “weighted degree”, where a neighborhood with a food hardship rate that

falls into the first quintile will receive an index score of “+2” for any community district neighboring that area. Similarly, the second quintile districts would be assigned a “+1”, the third quintile districts would be labeled as a “0”, the fourth quintile districts would be considered as a “-1”, and the fifth quintile districts would subsequently be a “-2”. The metrics used to categorize which quintile each community district falls into are listed below in **Figure 6** and are consistent across all boroughs.

Here is a formal equation for our newly constructed “weighted degree”. For each neighbor j of district i , define binary indicators:

$$s_{j,q} = \begin{cases} 1 & \text{if neighbor } j \text{ is in quintile } q, \\ 0 & \text{otherwise.} \end{cases}$$

Then the weighted degree of the district i is:

$$WD_i = \sum_{j \in N(i)} (2s_{j,1} + 1 \cdot s_{j,2} + 0 \cdot s_{j,3} - 1 \cdot s_{j,4} - 2s_{j,5})$$

Note that $N(i)$ denotes the set of community districts adjacent to i .

Rate of Food Hardship



Figure 6. Food Hardship Rate quintile categorization (Columbia University).

With the newly constructed variable “weighted degree” added to the data table (see **Table 2**), we can use different statistical models to quantitatively analyze the correlation between the weighted degree of a district and its food hardship rate, thus constructing a model that can be useful to predict food hardship rate, as well as other socio-economic factors if given enough relevant data, based on the variables constructed through Graph Theory. Note that these variables, inspired by Graph Theory, are not limited to quantitative data like weighted degree or weighted edge, but can also be presented as categorical binary variables, like the presence of a cycle of 3 that includes a given vertex, or whether or not a vertex has a shortest

path under a particular value to a selected vertex based on the Dijkstra’s algorithm.

Table 2. Full data of Food Hardship Rate in NYC with weighted degrees.

COMMUNITY_DISTRICT_NUMBER	COMMUNITY_DISTRICT_NAME	FOOD_HARDSHIP_RATE	MARGIN_OF_ERROR	QUINTILE	SAMPLE_SIZE	DEGREES	WEIGHTED_DEGREES	BOROUGH
1	MELROSE, MOTT HAVEN, PORT MORRIS	0.45	0.11	4	243	3	-5	Bronx
2	HUNTS POINT, LONGWOOD CLAREMONT,	0.45	0.11	4	243	3	-5	Bronx
3	CROTONA PARK EAST, MELROSE, MORRISANIA CONCOURSE,	0.49	0.09	5	331	6	-10	Bronx
4	HIGHBRIDGE & MOUNT EDEN	0.7	0.07	5	284	4	-7	Bronx
5	MORRIS HEIGHTS, FORDHAM SOUTH & MOUNT HOPE	0.53	0.12	5	216	4	-8	Bronx
6	BELMONT, BATHGATE & EAST TREMONT	0.49	0.09	5	331	6	-11	Bronx
7	BEDFORD PARK, FORDHAM NORTH & NORWOOD	0.57	0.11	5	239	4	-4	Bronx
8	RIVERDALE, FIELDSTON & KINGSBRIDGE	0.39	0.1	3	213	1	-2	Bronx
9	CASTLE HILL, CLASON POINT & PARKCHESTER	0.52	0.07	5	334	5	-6	Bronx
10	CO-OP CITY, PELHAM BAY & SCHUYLERVILLE	0.39	0.09	3	218	3	-3	Bronx
11	PELHAM PARKWAY, MORRIS PARK & LACONIA	0.47	0.13	4	186	4	-4	Bronx
12	WAKEFIELD, WILLIAMSBRIDGE & WOODLAWN	0.35	0.09	3	184	3	-3	Bronx
1	GREENPOINT & WILLIAMSBURG	0.33	0.09	2	186	3	2	Brooklyn
2	BROOKLYN HEIGHTS & FORT GREENE	0.22	0.09	1	213	4	3	Brooklyn
3	BEDFORD-STUYVESANT	0.42	0.1	4	219	5	3	Brooklyn
4	BUSHWICK	0.32	0.1	2	183	4	-3	Brooklyn
5	EAST NEW YORK & STARRETT CITY	0.47	0.11	4	212	3	-1	Brooklyn

Continued

6	PARK SLOPE, CARROLL GARDENS & RED HOOK	0.2	0.08	1	224	3	4	Brooklyn
7	SUNSET PARK & WINDSOR TERRACE	0.31	0.15	2	91	4	4	Brooklyn
8	CROWN HEIGHTS NORTH & PROSPECT HEIGHTS	0.26	0.07	2	214	6	-2	Brooklyn
9	CROWN HEIGHTS SO., PROSPECT LEFFERTS & WINGATE	0.47	0.11	4	156	3	-1	Brooklyn
10	BAY RIDGE & DYKER HEIGHTS	0.24	0.09	2	163	3	2	Brooklyn
11	BENSONHURST & BATH BEACH	0.36	0.09	3	171	4	2	Brooklyn
12	BOROUGH PARK, KENSINGTON & OCEAN PARKWAY	0.3	0.15	2	103	5	2	Brooklyn
13	BRIGHTON BEACH & CONEY ISLAND	0.39	0.13	3	105	2	0	Brooklyn
14	FLATBUSH & MIDWOOD	0.41	0.1	3	195	6	-2	Brooklyn
15	SHEEPSHEAD BAY, GERRITSEN BEACH & HOMECREST	0.36	0.11	3	124	5	1	Brooklyn
16	BROWNSVILLE & OCEAN HILL	0.66	0.08	5	278	6	-3	Brooklyn
17	EAST FLATBUSH, FARRAGUT & RUGBY	0.58	0.11	5	150	5	-4	Brooklyn
18	CANARSIE & FLATLANDS	0.37	0.1	3	230	5	-5	Brooklyn
1	BATTERY PARK CITY, CIVIC CENTER	0.05	0.03	1	215	2	2	Manhattan
2	GREENWICH VILLAGE & SOHO	0.05	0.03	1	215	5	8	Manhattan
3	CHINATOWN & LOWER EAST SIDE	0.36	0.12	3	248	3	6	Manhattan
4	CHELSEA, CLINTON & HUDSON YARDS	0.21	0.06	1	259	3	5	Manhattan
5	FLATIRON, MIDTOWN BUSINESS DISTRICT	0.21	0.06	1	259	5	9	Manhattan
6	MURRAY HILL, GRAMERCY & STUYVESANT TOWN	0.07	0.04	1	196	4	6	Manhattan
7	UPPER WEST SIDE & WEST SIDE	0.24	0.06	2	379	4	2	Manhattan
8	UPPER EAST SIDE HAMILTON HTS,	0.07	0.05	1	262	3	3	Manhattan
9	MANHATTANVILLE & WEST HARLEM	0.49	0.12	5	242	3	-1	Manhattan
10	CENTRAL HARLEM	0.39	0.08	3	359	4	-5	Manhattan
11	EAST HARLEM	0.48	0.09	4	334	2	2	Manhattan

Continued

12	WASHINGTON HEIGHTS, INWOOD & MARBLE HILL	0.55	0.08	5	445	2	-2	Manhattan
1	ASTORIA & LONG ISLAND CITY	0.32	0.08	2	236	2	-1	Queens
2	SUNNYSIDE & WOODSIDE	0.28	0.11	2	158	4	-2	Queens
3	JACKSON HEIGHTS & NORTH CORONA	0.49	0.1	5	221	3	0	Queens
4	ELMHURST & SOUTH CORONA	0.57	0.12	5	125	4	2	Queens
5	RIDGEWOOD, GLENDALE & MIDDLE VILLAGE	0.3	0.1	2	143	3	1	Queens
6	FOREST HILLS & REGO PARK	0.17	0.09	1	128	4	-2	Queens
7	FLUSHING, MURRAY HILL & WHITESTONE	0.16	0.08	1	172	2	2	Queens
8	BRIARWOOD, FRESH MEADOWS & HILLCREST	0.39	0.14	3	126	5	3	Queens
9	RICHMOND HILL & WOODHAVEN	0.43	0.13	4	117	4	0	Queens
10	HOWARD BEACH & OZONE PARK	0.46	0.13	4	124	2	-2	Queens
11	BAYSIDE, DOUGLASTON & LITTLE NECK	0.12	0.07	1	144	3	3	Queens
12	JAMAICA, HOLLIS & ST. ALBANS	0.46	0.1	4	264	4	-1	Queens
13	QUEENS VILLAGE, CAMBRIA HEIGHTS & ROSEDALE	0.26	0.07	2	233	3	1	Queens
14	FAR ROCKAWAY, BREEZY POINT & BROAD CHANNEL	0.45	0.09	4	229	0	0	Queens
1	PORT RICHMOND, STAPLETON & MARINER'S HARBOR	0.35	0.07	3	317	1	0	Staten Island
2	NEW SPRINGVILLE & SOUTH BEACH	0.41	0.11	4	134	2	2	Staten Island
3	TOTTENVILLE, GREAT KILLS & ANNADALE	0.17	0.07	1	185	1	0	Staten Island

3. Results from the Model

3.1. Beginning Model Structural Analysis

Analyzing the graph (see [Figure 7](#)), it's straightforward to observe that vertices with the same quintile for hardship rate are highly clustered, forming multiple 1-vertex-connected neighborhoods (like G_2 , as well as all other denoted subgraphs

besides G_9), indicating that there exist clustered areas in New York City with the same level of food hardship being inner-connected through at least one surveyed area. In other words, the graph contained several connected subgraphs with vertices of the same color label. Within these 1-vertex-connected subgraphs, some exhibited a higher degree of connection in terms of edge-connectivity. Indeed, the graph consists of multiple 2-edge-connected neighborhoods with vertices within the same quintile (like G_6 and G_7), illustrating several clusters of regions with similar food hardship, having each individual surveyed area being inter-continently-connected with at least two other regions in the vertex-neighborhood with the same quintile label, thus forming geographically triangle-like or diamond-like areas within New York City. Much like the research triangle located in North Carolina, residents in these survey neighborhoods within such triangle-like or diamond-like areas often share a similar socio-economic status, displaying a relatively comparable level of food hardship. Mathematically, such a phenomenon makes sense, considering that such geographical structures are shown as complete subgraphs (like $G_6 = K_3 = C_3$) and complete subgraphs with a missing edge (like $G_7 = K_4 - e$) in G , and a complete graph maximizes vertex-connectivity, edge-connectivity, and the number of edges, giving a fixed amount of vertices. In this way, a high connectivity level brings neighboring areas a similar level of living location-wise, in this case, the level of food hardship. Additionally, a triangle-like area in New York City not only fulfills a complete graph of 3 vertices in our graph but also completes a cycle of 3 vertices, displaying properties like total degree maximization (2-regular), graph symmetry, vertex-transitivity, edge-transitivity, as well as strong regularity. All of these properties further cemented the geographic significance behind the G_6 structure's socio-economic connectedness. Note that continentally and territorially connected indicates that the two confined regions share at least some degree of side, where the two regions aren't connected through a single point, in this case, not being bridged through a very small number of roads. Such a rule can also be illustrated by defining a dual graph, where each edge in G_d requires a shared edge in G .

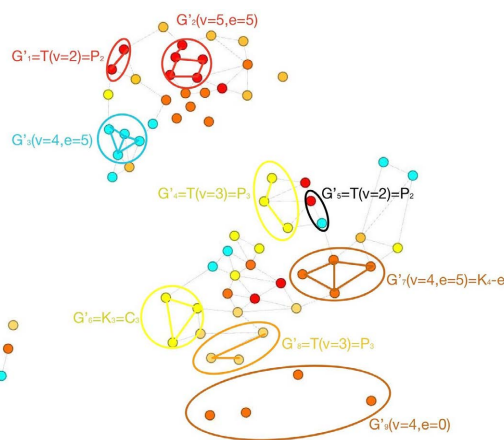


Figure 7. Annotated New York City graph of Food Hardship Rate (Created through Gephi).

Outside of these primary observations, tree-structured subgraphs were also frequently seen throughout the map. By definition, a tree is a connected graph where any two vertices are connected by exactly one path, meaning that no cycles or faces are allowed within a tree. In the graph that we've created, paths are the most common form of trees in New York City. Subgraph G_1 .

The graph also displayed a significant degree of socio-economic separation among different quintiles of food hardship levels. Among the graphs, only one edge reflected the adjacency between two vertices with more than three quintile differences.

Additionally, the existence of a cut vertex within a subgraph can signify the geographical and socio-economic significance of that particular district. In other words, in a given connected subgraph, by removing the cut vertex, the subgraph will no longer be connected. An example of such a structure can be represented through G_2 , where the cut vertex, "Morris Heights, Fordham South & Mount Hope", was recorded as having a food hardship rate upwards of 75%, one of the highest in the area, further emphasizing the district's regional significance.

3.2. Analytical Model Results

Using the given data, we constructed a linear regression model to illustrate the relationship between the weighted degree of a district and its food hardship rate (see [Figure 8](#)).

```
Call:
lm(formula = FOOD_HARDSHIP_RATE ~ WEIGHTED_DEGREES, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.256813 -0.076451 -0.006813  0.093368  0.263187

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.353359   0.015454   22.86 < 2e-16 ***
WEIGHTED_DEGREES -0.023273   0.003853   -6.04 1.24e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.118 on 57 degrees of freedom
Multiple R-squared:  0.3902,    Adjusted R-squared:  0.3795
F-statistic: 36.48 on 1 and 57 DF,  p-value: 1.238e-07
```

Figure 8. Linear regression model result (Created Through R Studio).

Although having a relatively low R^2 , with the correlation between weighted degree and food hardship level not being particularly strong, the p-value of the model is very low, revealing that the significance level of the model is extremely high, and the relationship between weighted degrees and food hardship rate is significant. The result stated that a higher weighted degree of a community district, which signifies the adjacencies of neighborhoods with a relatively lower food hardship rate, will likely have a low food hardship rate as well.

Despite the relatively low R^2 , ordinary least squares were chosen because it provides an interpretable baseline and allows us to establish whether any statistically significant association exists between weighted degree and food hardship rates.

While more complex functional forms were explored, including quadratic, logarithmic, exponential, and interaction specifications, none yielded statistically significant results. Spatial-lag models were not tested in this study; therefore, ordinary least squares was the most appropriate and transparent choice for the analysis.

I also attempted to use the variable “borough” as a factor for the model. However, due to the limitation of data variability and the relatively low specification of neighborhoods within each borough, the result was not significant at a 95% level.

It is worth noting that when providing more variables regarding the neighborhoods’ income, population, education level, and other socio-economic factors to construct the model, we can utilize stepwise regression to determine which particular variables contribute the most to the model’s predictive power. It is also vital to avoid multicollinearity, as many of these socio-economic factors are highly interconnected; keeping two highly correlated factors would vastly diminish the predictive power of our model. When given large enough data to train and cross-validate our model, we should also avoid overfitting, since some factors, like “borough” in New York City, can’t be easily applied to smaller cities like Cleveland or Milwaukee. By balancing a proper level of model complexity and reducing the predictive error of both our training sample and our testing sample to an optimal state, we will be able to accurately predict different socio-economic factors of each neighborhood when given data from new cities.

4. Discussion

Our analysis provides preliminary evidence of measurable correlations between city district location and food hardship rates, as well as how these variables can be incorporated into a graph-based framework. However, the results are modest in explanatory power and should be interpreted cautiously. Importantly, the study does not establish causal relationships, and further work—potentially integrating spatial-lag models or richer datasets—would be required to capture the full complexity of climate impacts on pricing dynamics.

A key limitation of this study lies in several gray-area omissions, including unobserved socioeconomic or infrastructural variables that may shape both adjacency patterns and food hardship rates but were not incorporated into the model. In addition, potential spatial autocorrelation—the tendency of nearby districts to share similar characteristics beyond measured adjacencies—was not formally tested, leaving open the possibility that residual dependence biases the results. Finally, the analysis relied on a single socioeconomic indicator, the food hardship rate, as the outcome measure; while meaningful, it captures only one dimension of disadvantage and may not fully reflect the broader dynamics of poverty or well-being across neighborhoods.

In the future, our analysis techniques can also be utilized to help investigate other location-based phenomena, such as the correlation between climate change and the coastal real estate market. This provides a foundation for understanding

how pricing might fluctuate in response to shifts in climate factors, and offers relevant data, including temperature, precipitation, and humidity. To reflect and specify the length of certain roads and connections between two vertices, we can also utilize the concept of weighted edges, where each edge is assigned a numerical value as its weight, signifying the distance between two vertices.

Under the current rapid shifting state of climate change, when constructing our analytical models, we should also consider the environmental factors of the area. Factors such as proximity to the ocean, precipitation levels, and the directional relationship with nearby mountains are much more dynamic when describing the “condition” of a neighborhood for its residents. For instance, Southern Miami vertices adjacent to the sea would face a decrease in housing prices due to sea level rise. Since sea level rise is a continuous and often gradual process that occurs over a long period, one can also utilize time series data when constructing and training the model to further enhance its predictive power.

Acknowledgements

I would like to thank Dr. David M. Holland, as well as everyone from the Capstone program, for their scholastic, technical, and resource support throughout this long and intensive project, and for allowing me to initiate it in the first place. Dr. David M. Holland’s extensive knowledge in statistics and climate change was crucial during the middle stage of my research, when I was struggling to find the proper direction.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Wilson, R.J. (1972) Introduction to Graph Theory. 4th Edition, Oliver & Boyd. <https://www.maths.ed.ac.uk/~v1ranick/papers/wilsongraph.pdf>
- [2] UNFCCC (2024) United Nations Climate Change. UNFCCC. <https://unfccc.int/>
- [3] COP (2024) United Nations—Climate Change. Conference of the Parties. <https://unfccc.int/process/bodies/supreme-bodies/conference-of-the-parties-cop>
- [4] EPA (2023) Climate Change Impacts on Coasts. United States Environmental Protection Agency. <https://www.epa.gov/climateimpacts/climate-change-impacts-coasts>
- [5] Miller, V. and Skelton, R. (2023) The Environmental Justice Movement. NRDC. <https://www.nrdc.org/stories/environmental-justice-movement>
- [6] Frank, T. and E&E News (2024) Hurricanes Caused Lost Income among at Least Half of Local Residents. Wikipedia. <https://www.scientificamerican.com/article/hurricanes-caused-lost-income-among-at-least-half-of-local-residents/>
- [7] Dennis, B. (2022) Rising Seas Could Swallow 650,000 Privately-Owned Properties by 2050. Washington Post. <https://www.washingtonpost.com/climate-environment/2022/09/08/sea-level-rise-climate-central/>
- [8] Miller, J. (2021) Saltwater Intrusion: A Growing Threat to Coastal Agriculture. USDA

- Climate Hubs.
<https://www.climatehubs.usda.gov/hubs/northeast/topic/saltwater-intrusion-growing-threat-coastal-agriculture>
- [9] Creel, L. (20035) Ripple Effects: Population and Coastal Regions. PRB. Population Reference Bureau.
<https://www.prb.org/resources/ripple-effects-population-and-coastal-regions/>
- [10] Gephi (2024) Gephi—The Open Graph Viz Platform. <https://gephi.org/>
- [11] Columbia University (2019) Can Mapping Hunger in New York City Help People in Poverty? Columbia Giving.
<https://giving.columbia.edu/can-mapping-hunger-new-york-city-help-people-poverty>
- [12] Trading Economics (2024) All-Transactions House Price Index for Boise County, ID. Trading Economics.
<https://tradingeconomics.com/united-states/all-transactions-house-price-index-for-boise-county-id-fed-data.html>
- [13] Trading Economics (2024) All-Transactions House Price Index for Miami-Dade County, FL. Trading Economics.
<https://tradingeconomics.com/united-states/all-transactions-house-price-index-for-miami-dade-county-fl-fed-data.html>