THE UNIVERSITY OF CHICAGO

An Analysis of Neighborhood Attributes in Chicago using Eigenvector Centrality

Benjamin Rothschild

June 2018

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

Faculty Advisor: Prof. Luis Bettencourt
Preceptor: Joshua Mausolf

Abstract

As cities increase in importance in human societies, many scholars have begun to study how they are structured and change to analyze the effects of urbanization, gentrification, and neighborhood change. The difficulty in measuring the structure of cities, though, lies in the fact that there are many independent attributes of neighborhoods as well as complex interactions between them over space and time. One method that has successfully been used to analyze interactions between actors in complex systems is network analysis which is used in a vast array of fields from social network analysis to analyzing the structure of web pages on the internet. In this paper, I use a standard measure of actor importance in network analysis, eigenvector centrality, to analyze how neighborhoods in Chicago are ranked and change over time. Eigenvector centrality is a popular measure of actor importance in networks because it takes into account both the connections of an actor as well as the power of the actors it is connected to. I calculate the eigenvector centrality of neighborhoods in Chicago by using the Longitudinal Employer-Household Dynamics (LEHD) dataset published by the US Census Bureau which is used to create a commuting network of neighborhoods in Chicago. I then explain the meaning of a neighborhood's centrality and use this ranking to analyze attributes of neighborhoods further.

1 Introduction

There are many ways to analyze the effects of neighborhood growth and change that spans academic disciplines from economics to statistics and sociology and employ traditional statistical methods such as regression, causal frameworks, and policy evaluation. While these methods provide a robust theoretical framework to study the effects of neighborhood change, there are fewer methods available to understand how neighborhoods are related to each other and how they make up the overall structure of a city. What makes analyzing neighborhood structure in a city difficult is that there are many independent attributes of neighborhoods and complex interactions between them over both space and time. In this paper, I suggest that network theory, a framework that has been used to study many types of complex systems, can provide a novel way to explore how neighborhoods are structured in a city and how their relationships to each other change over time.

Network theory has been used to study how actors interact in a diverse set of disciplines from social networkS to trading networks and molecular systems. These methods have led to significant new insights in many fields and even serves as the basis of the search algorithm behind Google. I will use a standard measure used to quantify actor importance, eigenvector centrality, to find important neighborhoods for employment and housing in the network of neighborhoods in Chicago. The eigenvector centrality method for computing actor importance will show the relationship between employment and housing centers in Chicago and will allow us to ask additional research questions such as

how wealth and employment are distributed across the city and how connections between neighborhoods influence other local socioeconomic attributes. I hope that this application will provide an additional lens for researchers to analyze the structure of cities and allow them to ask new and different questions.

2 Literature Review

Network theory has been used to study structures in complex systems across a wide array of disciplines. In particular, centrality measures in network analysis are used to model network structures by identifying powerful actors and showing how they relate and interact with each other. Several measures of centrality have been defined to address a central question: *Who is the most important actor in a network*? Centrality measurements differ based on how actors interact and what kind of "importance" researchers want to quantify. In this paper, I will focus on "eigenvector centrality" which is a measure of the influence of an actor in a network and corresponds to the first eigenvector in the connectivity matrix of a network. This measure has been used in a diverse set of applications but is not yet used to analyze the structure of neighborhoods in a city.

Researchers have developed several measures of centrality in networks each of which has its benefits for specific network types and purposes. For example, "degree centrality" is a standard measure of centrality that counts how many connections an actor has. An actor will be ranked higher if it has many neighbors while it will be lower if it has fewer neighbors. "Closeness centrality" measures the degree to which an individual is near all other individuals in the network and is the inverse of the sum of the shortest distances between each actor and every other actor in the network. Researchers have used it to study how long it would take to spread information from one actor to all others sequentially. "Betweenness centrality" quantifies the number of times an actor acts as a bridge along the shortest path between two actors and has been used to quantify the control of information in a social network. In this paper, I will focus on eigenvector centrality which takes into account both the number of connections an actor has as well as the importance of each of the actors it is connected to.

Eigenvector centrality has been studied to explain many of phenomena in the social sciences and lead to new insights where previous measures of centrality failed to describe complex interactions. For example, Cook et al. argued that typical centrality measures like degree centrality fail to predict power distributions in exchange networks (networks where actors bargain or trade with each other).[2] Through theoretical and simulated results, they showed that in negotiations it is advantageous to be connected to those who have few options and being in a central position does not necessarily make an actor more powerful. Thus, they suggest that more general conception of centrality needs to be created that takes into account power dependency as well as closeness. In situa-

tions where degree centrality fails, social scientists have proposed other measures that fuse power-dependency and closeness. Bonacich suggested that eigenvector centrality makes a good centrality measure in these networks because it takes into account the number of connections an actor has as well as the centrality of the actor it is negotiating with.[3] For example, an actor connected to a more powerful actor has a higher eigenvector centrality than if it was connected to a less powerful actor.[**?**].

The first researcher to apply eigenvectors to geography was P.R. Gould. In his paper, *On the Geographical Interpretation of Eigenvalues*, his goal was less motivated by a specific hypothesis but more by a curiosity to determine if this mathematical structure could uncover a pattern in very complex situations.[1] His hope, which many computational social scientists share, was that underlying complex phenomena might be a mathematical idea that could provide a meaningful geographic interpretation. To explore this idea, he mapped the road network of Uganda and created a connectivity matrix of this network on a binary scale, 1 indicating if two cities were connected and 0 if they were not. He calculated the first four eigenvectors of these matrices in 1921 and 1935 and compared the results between the two years. He found that the first eigenvalue, which centered on the city of Kampala, was the most connected town in the country because it had the most direct linkages and was also in the center of Uganda. The city with the next highest value, Entebbe, was also well connected. He then examined a new connectivity matrix of the cities in 1935 and described how several cities have become more important as new roads and cities appeared. He notes that the successive eigenvectors and eigenvalues are a "pull out" of small regional networks within the trading structure of the region. Gould makes an initial attempt, though vague, to describe the meaning of his eigenvector derivation. He explains that vectors representing well-connected towns will not only lie in the middle of a large number of dimensions but will tend to lie close to the principal eigenvalue. On the other hand, towns that are moderately well connected will not lie in the middle of so many dimensions as the well-connected towns and will tend to form small structural clusters on their own. This interpretation has been named the "Gould Index of Accessibility."

In social networks, Tinkler described the eigenvector centrality of a person in the context of a rumor spreading through a social network. A social network can be modeled by a square matrix where an entry $E_{i,j} = 1$ if person i know person j and 0 if they do not. If a rumor starts in the network, as time progresses, the rumor will spread throughout the network according to the connections between people in the social network–if someone knows the rumor, they tell it as many times as they heard it to all the people they are connected to. As this process repeats the distribution of the rumor will also be given by the principal eigenvector. In other words, the eigenvector is the chance that the rumor has spread to a specific actor in the network.[4]

Another interpretation of eigenvector centrality was given by J.W. Moon who

described eigenvector centrality in the context of player rankings after an iterative round-robin competition. In his example, there is a tournament between n players and the win-loss outcomes create a square matrix where an entry $E_{i,j}$ is the win percentage of player i against player j. A player's ranking increases by beating another player, however, if they win against a stronger opponent, they will get a higher rating boost than if they beat a weaker player. After the tournament has elapsed and rankings reach an equilibrium, the player's rankings will correspond to the principal eigenvector of the win-loss matrix.[5]

One last application of eigenvector centrality I will explain was applied by Sergey Brin and Larry Page. They created a database with the hyperlink network of over 24 million web pages and used eigenvector centrality to rank the importance of web pages in response to a query. Web pages are arranged in a network based on their hyperlinks to other pages, and a PageRank algorithm ordered results of a query by counting the number of pages that linked to each web page. The PageRank of a webpage was calculated using an iterative algorithm that corresponded to retrieving the principal eigenvector of the normalized matrix of hyperlinks. They give a few intuitive justifications why this ranking works. One was imagining a "random surfer" who is given a web page at random and keeps clicking links. The PageRank of the web page is the probability the random surfer will land on a page. This method for ranking web pages was the basis of the original version of Google's search engine.[6]

A related area of research is studying the second eigenvector in the connectivity matrix. While the first eigenvector reflects volumes and strengths of connections among the actors, the second eigenvector can extract separate groups within the network who behave in similar but distinct manners. These subnetworks can be informative in analyzing the overall structure of the network. In Gould's analysis, the second eigenvector was able to pick out significant geographic subsystems in the transportation network of Uganda. Iaocobucci et al. also demonstrate that the extraction of only the first eigenvector can be insufficient in gaining a comprehensive understanding of the network.[7] The example they give is from a communication network between researchers. While the first eigenvector retrieves the principal structure, it is often similar to standard measures of centrality like degree centrality. By extracting the second and third eigenvector, several classes of network structures and actor attributes were pulled out and interpreted. These eigenvectors are interesting because by necessity they are uncorrelated with the principle eigenvectors and therefore uncorrelated with the traditional degree of centrality. Another example is given by Bonacich in analyzing cliques in a social network.[8] Consider a social network that is made up of many cliques where each clique has zero communalities between another clique and all individuals within the clique are connected to each other. In this case, each clique will be represented as an eigenvalue with the most significant clique being the principal (largest eigenvalue) and the magnitude of the eigenvalue will be a measure of how well the eigenvector is at summarizing the relationships within the clique. The eigenvector will be the

popularity score of individuals within the clique. In my analysis, I will also try to explain the interpretation of the first and second eigenvectors.

Eigenvector centrality is used as a general ranking measure of both power and influence in many independent fields and contexts where interactions between actors are modeled as a network. It is a popular measure of centrality because it takes in to account both how many connections an actor has and the importance of each actor it is connected to. Eigenvector centrality also has benefits over other measures of centrality as it can be used with signed graphs, adjacency graphs, or value-based graphs. For example, networks graphs with negative connections include dating and friendship networks where reciprocation is not necessary or trade where one actor sells a product to another. While there have been several explanations of eigenvector centrality in networks across many fields, it is often difficult to interpret the meaning of exactly what an eigenvector centrality score means in the real world, and I will also attempt to understand the meaning of the principal eigenvector by comparing its value to other neighborhood attributes over time.

3 Data

The main dataset in my analysis is the Longitudinal Employer-Household Dynamics Dataset (LEHD) that is published by the United States Census Bureau. This is a synthetic dataset that joins firm employment data and census demographic data on the census block level and provides a fine-grained view of the connections between where people live and work. Synthetic data represents an innovative way for a government to create and release data as it provides researchers data at a low cost since it leverages existing datasets and there is no additional burden on respondents such filling out additional surveys. The datasets that are used to produce the LEHD dataset include Unemployment Insurance wage records, the Quarterly Census of Employment and Wages, and the Statistical Administrative Records System. Some of the data sources that are used to produce the LEHD dataset are confidential and not themselves made public. The current coverage of employment data is limited to jobs covered by the Unemployment Insurance Program which is approximately 95% of jobs in the United States.

Jobs are broken down among job categories, employee age brackets, and employee monthly salary as follows:

1. Job Category:
   (a) Goods Producing
   (b) Trade, Transportation, and Utilities
   (c) Other

2. Age:
   (a) 29 and younger
   (b) 30 - 45
   (c) 55 and older
3. Monthly Salary:
   (a) under $1,250
   (b) $1,251 - $3,333
   (c) over $3,333

The data is published every year from 2002 to 2015 and shows the number of people who live and work between each census block in the United States.[1] Census blocks are currently the smallest geographic units used in the US Census Bureau statistics. The number of census blocks in the 2010 Census was 11,155,486 with an average size of 27 people so the resultant dataset provides a very fine-grained view of the relationship between places of work and employment. Since the data is highly specified, the Census Bureau employs a few techniques to protect confidentiality of citizens such as noise infusion and synthetic data creation using probabilistic differential privacy. [2].

In my analysis I focus on data within the Chicago Metropolitan Statistical Area and a summary of the employment data for the 2002 is below.

	Count	Percent of Total
Total Jobs	3,924,152	100%
Age: 29 or younger	1,027,445	26.1%
Age: 30 to 54	2,328,093	59.3%
Age: 55 or older	568,614	14.4%
Earnings: $1250month or less	1,090,632	27.7%
Earnings: $1251/month to $3333/month	1,467,733	37.3%
Earnings: greater than $3333/month	1,365,787	34.7%
Goods Producing	708,324	18.0%
Trade, Transportation, and Utilities	819,502	20.8%
Other	2,396,326	61.6%

Much of my analysis is based off of the eigenvector centrality of the live-work commuting matrix between census tracts in Chicago broken down by different job and demographic characteristics. Though the LEHD dataset provides census tract level statistics I found that interpretability improves at a slightly larger

[1] Data is omitted for 9 state-year combinations where the Census Bureau notes there are data integrity issues. Illinois was not noted on this list, so my study is unaffected by missing data.

[2] More information about noise infusion and confidentially protection can be found on the census website

area so decided to use census blocks and neighborhoods of which there are 2,215 and 190 respectively in the Chicago metropolitan statistical area. To visualize the dataset, I first created a connectivity graph between all census tracts with more than 25 people commuting between them shown below.

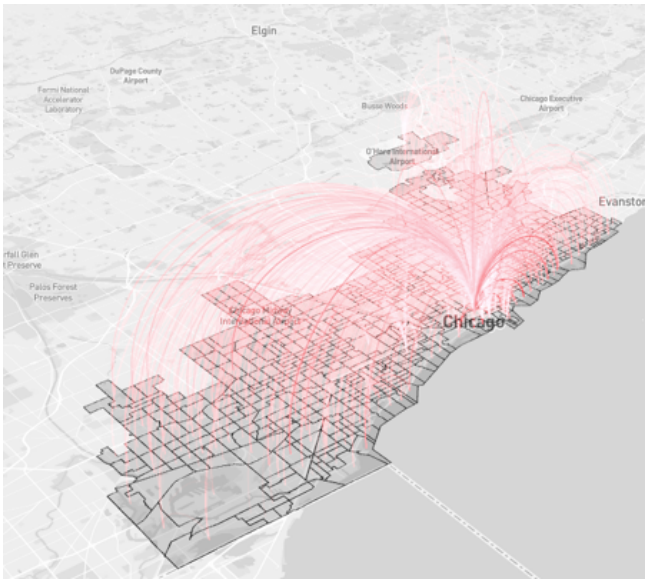


Figure 1: This map shows all the connections between where people live and work in Chicago. Note the dominance of The Loop as the main center of employment.

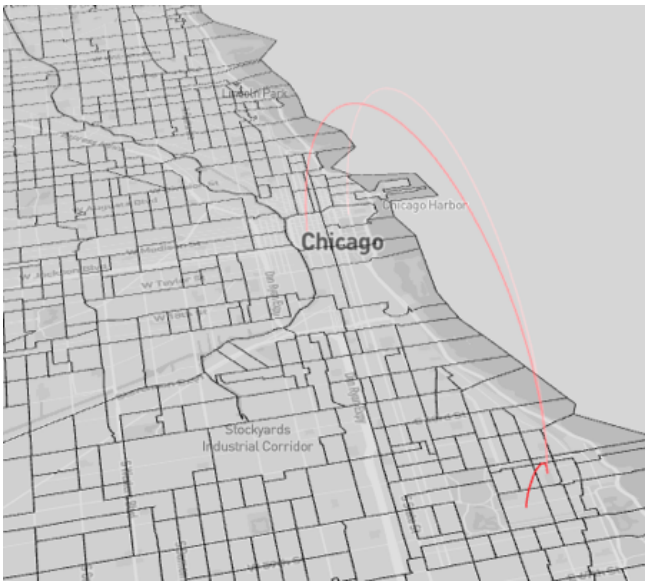


Figure 2: This map shows the connections between a people living Hyde Park who either commute to the University of Chicago or the The Loop. Note: this map only shows combinations on tracts with 25 more people commuting between them

In order to correlate the LEHD calculated centrality measures to other datasets I make the following considerations. When comparing eigenvector centrality to home values, I calculate centrality on the neighborhood level as this is the most fine-grained data level that is available through Zillow. When comparing centrality with data that is only available from the City of Chicago such as Business Licenses and Crime, I calculate centrality on the census block level for the entire Chicago MSA which includes many tracts outside of the City of Chicago Boundary and counties in Wisconsin and Indiana but restrict my analysis to census blocks within the City of Chicago boundary.

Real estate listing data is provided by Zillow, an online real estate database company that tracks detailed listing data of home sales throughout the US starting in 1994. They publish data of home sales for the top 50 markets on the neighborhood level. In my paper I use the following statistics:

1. Median List Price Per Square Foot ($)
2. Monthly Home Sales (Number, Seasonably Adjusted)
3. Zillow Home Value Index: All Homes (SFR, Condo, Co-Op)[3]
4. Zillow Rent Index: All Homes (Multifamily, SFR, Condo, Co-Op)footnotehttps://www.zillow.com/researc rent-index-methodology-2393/

[3]https://www.zillow.com/research/zhvi-methodology-6032/

Since I am comparing real estate prices to neighborhood rankings, I make the following adjustments to be able to compare the data. First I normalize each neighborhood to its initial value to get an index of the statistic over time. Next I divide by the median value for all neighborhoods in Chicago for the given time period. The result is a statistic that shows how the neighborhood compares to other neighborhoods in Chicago over time. If the value is less than 1, then it is below average whereas if it is greater than 1 it is above average.
For a complete overview of how data was collected, cleaned, visualized, and calculated, I provide the scripts and programs I used on github at `https://github.com/b-nroths/chi-data`.

4 Methods

In order to find the employment centers of a city I will represent the city's commuting network as a matrix and calculate the important neighborhoods using eigenvector centrality, a method that has been used in studying networks such as trade routes, social networks, and webpages. The goal of the method is to take the network matrix and to output the most influential actors in the network. The definition of influential varies depending on the context and will be explore later in this paper. The matrix representing the connections between actors can be represented a number of way. In social networks, edges represent a connection, in which case the matrix is called an adjacency matrix and is filled with 1 and 0's depending on if two nodes are connected to each other. In a round-robin tournament the matrix could be filled in with 1's and 0's depending on if a team beat the other or by a percentage which would represent a team's win percentage against another team. In a trading network there are multiple ways to describe the network matrix. One simple way is to to use an adjacency matrix (as Gould did) of 1's and 0's if an actor trades with another. Another way would be to represent the size of the actor compared to the country they are trading with or the percent of their total trade an actor makes with a specific partner. Further it could also be represented as a distance of a trading route between two actors. In this case it is common to normalize the distances to the sum of each column is equal to one.

In this paper I will call the network matrix a "commuting matrix". The actor is a neighborhood and the value in the matrix represents the percent of people who live in one neighborhood who commute to the other. This will have the benefit of producing an easy to interpret eigenvector and eigenvalue. Before I demonstrate how the network matrix is built in the context of this paper, it is important to understand why we are guaranteed a positive eigenvalue from the following theorem proven by Oskar Perron and Georg Forbenius.

Theorem 1 (Perron-Frobenius Theorem). Let $C \in \mathbb{R}^{nxn}$ represent a nonnegative primitive matrix (i.e. C: $C_{i,j} > 0$). There exists a positive real number λ_{max}, such that:

1. $\lambda_{max} > 0$
2. λ_{max} has a unique (up to a constant) eigenvector v which has all positive entries
3. $\lambda_{max} > |\lambda|$ for any eigvenvalue $\lambda \neq \lambda_{max}$

Theorem 2. A column stochastic matrix will always have an eigenvalue 1. All other eigenvalues are in absolute value smaller or equal to 1.

To illustrate the network I will explore in my paper, consider a simplified example of a city of 110 people, 100 of whom live Downtown and 10 of whom live in Hyde Park. Of the 100 people who live downtown, 90 works downtown and 10 works in Hyde Park while of the 10 people, 5 work downtown and 5 work in Hyde Park. This network can be represented by the following matrices.

$$C = \begin{bmatrix} hydepark -> hydepark & hydepark -> downtown \\ downtown -> hydepark & downtown -> downtown \end{bmatrix} = \begin{bmatrix} 5 & 10 \\ 5 & 90 \end{bmatrix} \quad (1)$$

$$l = \begin{bmatrix} hydepark \\ downtown \end{bmatrix} = \begin{bmatrix} 10 \\ 100 \end{bmatrix} \quad (2)$$

$$w = \begin{bmatrix} hydepark \\ downtown \end{bmatrix} = \begin{bmatrix} 15 \\ 95 \end{bmatrix} \quad (3)$$

From this information we can create a commuting matrix which normalizes the flow between regions of the city and transforms the "live" matrix (2) into the "work" matrix (3). This transforms the above matrices into the following equation.

$$Cw = l \quad (4)$$

$$\begin{bmatrix} 5/15 & 5/95 \\ 10/15 & 90/95 \end{bmatrix} \begin{bmatrix} 15 \\ 95 \end{bmatrix} = \begin{bmatrix} 10 \\ 100 \end{bmatrix} \quad (5)$$

We can also calculate write out the commuting flow from work to home as

$$\mathrm{C}\ l = w$$

$$\begin{bmatrix} 5/10 & 10/100 \\ 5/10 & 90/100 \end{bmatrix} \begin{bmatrix} 10 \\ 100 \end{bmatrix} = \begin{bmatrix} 15 \\ 95 \end{bmatrix} \quad (6)$$

The last connectivity matrix I will study looks at the flow of money between regions modeled by the salary of workers that commute between regions. The first model will be the total flow of money between regions represented by the sum of the salaries of all the workers who commute between regions. For example, if 10 people commute from Hyde Park to Downtown and they each make an average of \$2,500 per month I will consider the money flow between Hyde Park and downtown to be \$25,000. In my dataset, salaries are broken down into three ranges less than \$1,250, \$1,250-\$3,333 and over \$3,333. To simplify the

problem, I will represent the buckets as \$1,250, \$2,500, and \$5,000. Consider the same commuting flows as above but now with income added according to the following breakout.

	\$1,250	\$2,500	\$5,000	Total People	Total Salaries
$HydePark -> HydePark$	1	2	2	5	\$16,250
$HydePark -> Downtown$	2	3	5	10	\$35,000
$Downtown -> HydePark$	0	1	4	5	\$22,500
$Downtown -> Downtown$	10	20	60	90	\$362,500

This is represented by the following matrix.

$$C = \begin{bmatrix} Salaries_{HP->HP} & Salaries_{HP->D} \\ Salaries_{D->HP} & Salaries_{D->D} \end{bmatrix} = \begin{bmatrix} \$16,250 & \$35,000 \\ \$22,500 & \$362,500 \end{bmatrix} = \begin{bmatrix} .419 & .088 \\ .580 & .912 \end{bmatrix} \quad (7)$$

Next, I take the C matrix in (5), (6), & (7) and computing the corresponding eigenvectors and eigenvalues.

Eigenvalue and Eigenvector for (5)

	$\lambda_1 = 1.0$	$\lambda_2 = 0.28$
$Downtown$	0.99689815	0.70710678
$HydePark$	0.07870249	-0.70710678

Eigenvalue and Eigenvector for (6)

	$\lambda_1 = 1.0$	$\lambda_2 = 0.4$
$Downtown$	0.98058068	0.70710678
$HydePark$	0.19611614	-0.70710678

Eigenvalue and Eigenvector for (7)

	$\lambda_1 = 1.0$	$\lambda_2 = 0.33$
$Downtown$	0.98869689	0.70710678
$HydePark$	0.14992818	-0.70710678

These results demonstrate a few things. First, note that the principal eigenvalue is 1.0 as expected from the theorem above. The vector that corresponds to the principal eigenvalue is the Gould Index of Accessibility in the network and represents the relative strength of each node in the principal network. In the three examples above you can see that the Downtown area dominates this vector. The second eigenvalue represents a "pull out" of the principal network. It is important to remember that the second eigenvector will be orthogonal to the principal eigenvector and represents a completely different sub-network

or clique. In the system in equation 5 and 7 the second eigenvalue is lower which indicates the dominance of the first eigenvalue compared to the network examined in the 6 equation. These results have a similar interpretation to the previous example where the principal eigenvector show the most dominant housing neighborhoods in the network. The eigenvector that corresponds to the principal eigenvalue are the relative ranking of neighborhoods in this setting.

Also, consider a more balanced commuting flow with its corresponding eigenvalues.

$$\mathrm{C}\ l = w$$

$$\begin{bmatrix} 0.60 & 0.55 \\ 0.40 & 0.45 \end{bmatrix} \begin{bmatrix} 50 \\ 50 \end{bmatrix} = \begin{bmatrix} 57.5 \\ 42.5 \end{bmatrix} \tag{8}$$

Eigenvalue and Eigenvector for (8)

	$\lambda_1 = 1.0$	$\lambda_2 = 0.05$
Downtown	0.8087	0.70710678
HydePark	0.5882	-0.70710678

Here not only is the primary eigenvector more equal across neighborhoods with values of 0.8087 and 0.5882, the value of the second eigenvalue is also very small compared to the first $\lambda_2 = 0.05$. This means that the majority of the network is explained by its primary eigenvalue.

Next I will use this same set up to analyze the neighborhoods of Chicago using the LEHD dataset.

5 Results

5.1 Centrality for Employment

First, I will study what neighborhoods are the most important for different job sectors, income ranges, and age profiles. I will use the commuting matrix described in equation 6 to rank the neighborhoods according to the eigenvector that corresponds to the principal eigenvalue. Answering the question *How important is a neighborhood to jobs in a specific industry, age group or salary range?* To do this, I create a connectivity matrix over all neighborhoods in the Chicago MSA and rank them from the years 2002 - 2015. Below I list the top 10 neighborhoods ranked according to the eigenvalue that corresponds to the principal eigenvector. In the last column I note the percentage change from 2002 to 2015 of the neighborhood.

Table 1: Employment Neighborhood Rank

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	%
The Loop	.958	.952	.954	.939	.947	.955	.947	.960	.961	.962	.962	.964	.962	.964	.63%
Streeterville	.162	.176	.180	.203	.206	.168	.174	.157	.146	.144	.134	.131	.115	.129	-2.37%
River North	.074	.070	.076	.084	.083	.082	.085	.079	.097	.091	.098	.094	.099	.116	56.76%
West Loop Gate	.073	.084	.083	.097	.085	.087	.087	.085	.083	.094	.089	.094	.099	.105	43.84%
O'Hare	.121	.084	.079	.078	.057	.079	.120	.063	.077	.070	.090	.074	.095	.082	-32.23%
Near North	.074	.098	.085	.101	.098	.091	.088	.088	.085	.086	.073	.070	.073	.074	.00%
South Loop	.102	.124	.113	.125	.115	.110	.114	.109	.075	.081	.086	.089	.090	.065	-36.27%
Hyde Park	.049	.043	.063	.092	.077	.080	.100	.067	.096	.087	.094	.088	.088	.053	8.16%
West Town	.039	.052	.049	.052	.054	.048	.049	.049	.041	.045	.046	.042	.047	.053	35.90%
Near West Side	.033	.040	.033	.039	.016	.017	.040	.036	.036	.038	.042	.042	.042	.044	33.33%

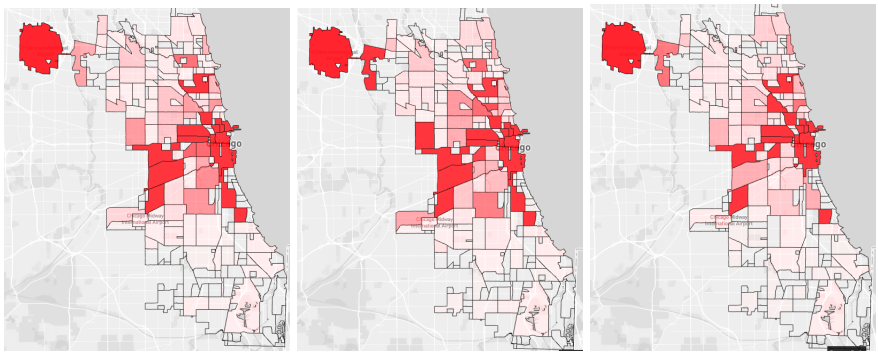


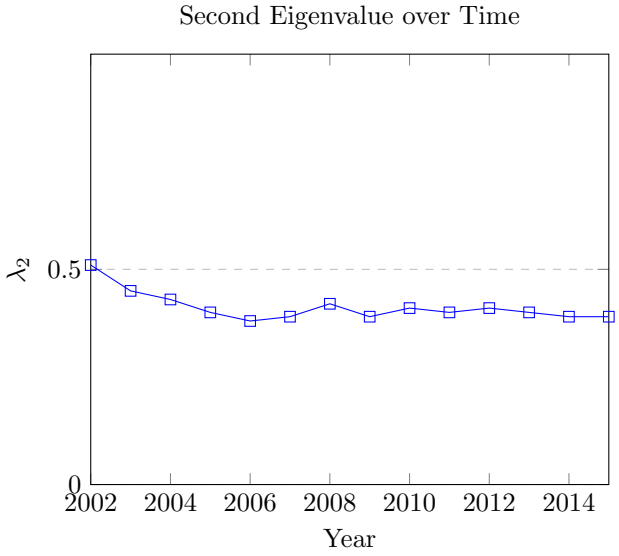
Figure 3: This map shows the eigenvector corresponding to the principal eigenvalue for 2002, 2008, and 2015

As can be seen in the map and table employment is Chicago is highly concentrated in the downtown The Loop neighborhood which consistently is ranked above .95 and dominating the rank of all other neighborhoods. Other consistently important neighborhoods include include Streeterville which is just north of The Loop as well as O'Hare Airport. Looking at the change in rankings overtime can give us an interesting view of how neighborhoods are becoming more or less important centers of employment over time. For example, West Loop Gate which is the neighborhood just west of the Chicago River has become much more important relative to other neighborhoods. There have been a number of high tech companies that have opened offices in the West Loop since 2014 such Uber and Google to name a few.[4]. Hyde Park has also grown in importance driven by the census blocks that cover the University of Chicago and Harper

[4]https://www.builtinchicago.org/2014/09/18/meet-neighbors-4-tech-companies-sign-huge-leases-west-loop

Court.[5] [6]

Another interesting view of the data is to look at the value and location of the second eigenvector of the network over time. The second eigenvector will depict a completely orthogonal network compared to the primary eigenvector and can give an idea of how more or less important the primary network is than the next network. One way to think of the second eigenvector would be a completely different "clique" that is somehow separate from the primary "clique" in the city. For example, it is common for people to live between these regions with less access to the primary region. In the graph below I show the value of the second eigenvalue over time which decreases from 0.51 to 0.41 from 2002 to 2015. This means that compared to the primary network the influence of the secondary network is decreasing. There could be a few reasons for this, for example the primary network could be increasing in size, people from the secondary network are joining the primary network, or people are leaving the secondary network.



Next I will take a look at how the different neighborhoods specialize in different job industries. The eigenvector centrality is able to pull out the different networks of employment in Chicago such as how dispersed the Goods Producing sector is compared to the Trade, Transportation, and Utilities sector which is dominated by O'Hare International Airport. The Other category is dominated by The Loop area where most of the jobs in the city are located.

[5] https://fiftythird.uchicago.edu/category/tags/harper-court-partners

[6] this feels anecdotal, bring in the other LEHD employment data, how is this conclusion different from just analyzing the counts?

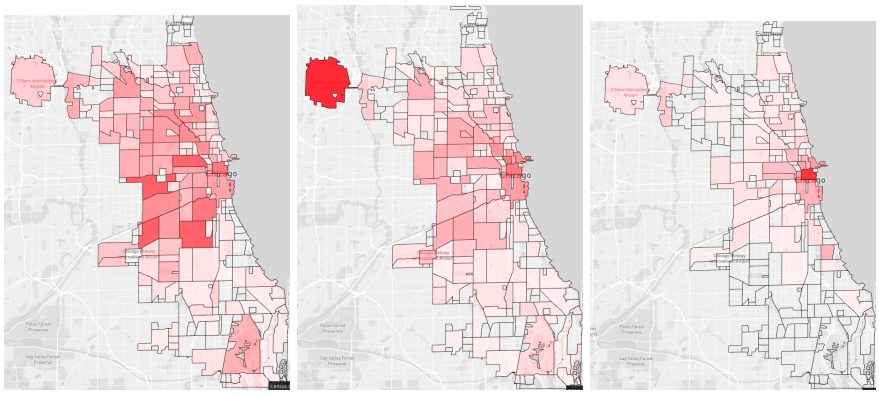


Figure 4: 2015 eigenvectors for the three different industry categories: Goods Producing (left), Trade, Transportation, Utilities (center), Other (right)

Lastly I will look at the eigenvectors across different income levels. What is interesting about these graphs is how similar the principal eigenvectors are be by income levels. Even though parts of Chicago are segregated by income level, the importance of each neighborhood for jobs is very similar.

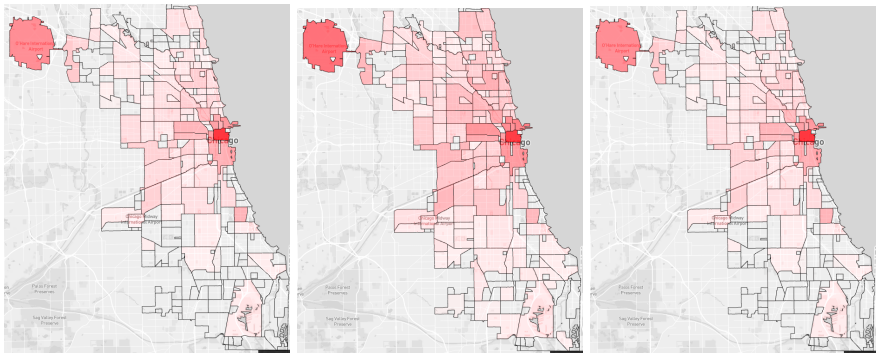


Figure 5: 2015 eigenvectors for the three different salary levels: Under $1,250 left, $1,250 - $3,333 center, Over $3,333 right for 2015.

Above I used the theory of eigenvector centrality to rank the most important neighborhoods of employment for neighborhoods in Chicago. This perspective gives us a rich view of the composition of a neighborhood as well as how it changes over time in comparison to other neighborhoods in the city.

5.2 Centrality for Housing

Similar to how employment centrality was calculated, I calculate housing centrality, answering the question *What neighborhoods are the most important for housing*? I also break down the neighborhood centrality by most important

for High and Low income earners and Old age and Young age based off of the LEHD Dataset. Below I map the principal eigenvector across the whole population in 2015. From this map it is apparent the network structure for housing is a lot different than for employment. The distributions of the rankings are a lot more even across neighborhoods compared to the employment dataset which was dominated by The Loop neighborhood. The top 10 neighborhoods based on their centrality from 2015 are listed below. The rankings are much more stable over time than they were for the employment dataset as most neighborhoods only increasing or decreasing in the single digits over the 14 year timespan.[7]

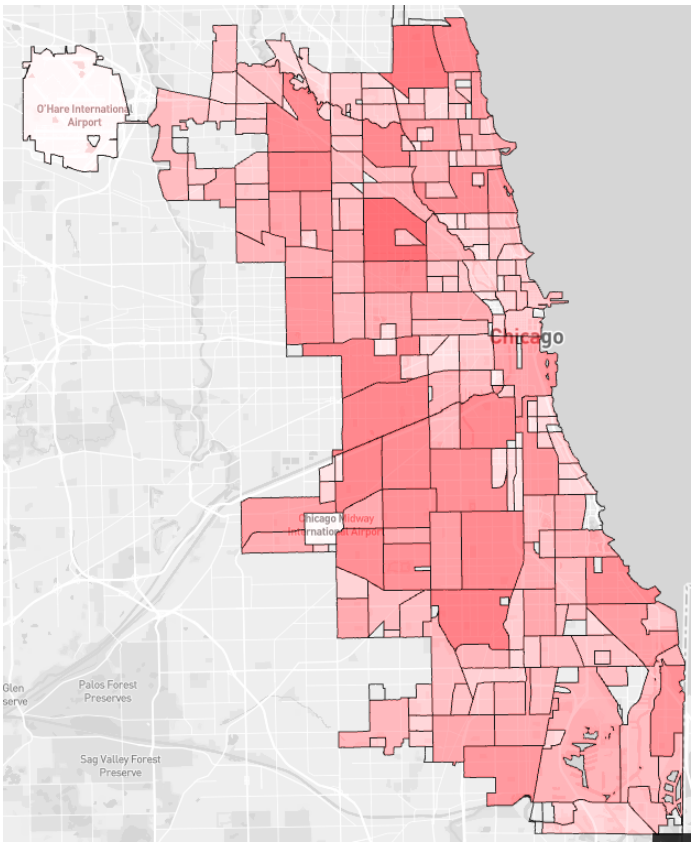


Figure 6: This map shows the eigenvector that corresponds to the principal eigenvalue in 2015.

fAfter the centrality measurements are created we can start to ask interesting questions such as how neighborhoods are changing overtime in comparison to each other. The main dimensions I will look at are how the age and income

[7]Graphs for this section are also included in the Appendix.

Table 2: Overall Housing Rank

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	— %
Logan Square	.303	.307	.307	.296	.287	.207	.296	.201	.313	.316	.323	.333	.340	.322	6.27%
West Rogers Park	.220	.221	.221	.224	.230	.201	.221	.209	.246	.260	.261	.254	.245	.239	8.64%
Gresham	.238	.239	.239	.233	.231	.245	.242	.254	.202	.205	.213	.193	.201	.215	-9.66%
Portage Park	.188	.185	.185	.181	.182	.214	.191	.220	.201	.203	.190	.198	.188	.189	.53%
Archer Heights	.181	.178	.178	.186	.182	.187	.173	.185	.198	.173	.175	.184	.186	.188	3.87%
Little Village	.208	.200	.200	.196	.193	.201	.176	.201	.178	.171	.183	.187	.197	.188	-9.62%
Albany Park	.189	.195	.195	.179	.172	.144	.181	.133	.186	.198	.201	.193	.183	.179	-5.29%
Englewood	.247	.244	.244	.236	.239	.258	.223	.252	.197	.185	.178	.172	.177	.179	-27.53%
Bridgeport	.139	.147	.147	.146	.157	.171	.165	.181	.189	.171	.176	.175	.171	.174	25.18%
Rogers Park	.154	.151	.151	.150	.159	.133	.155	.127	.158	.172	.175	.177	.175	.174	12.99%

are affecting neighborhoods in Chicago. Below are tables of the top 10 neighborhoods by centrality ranking for Young/Old workers and Low-Income/High Income workers.

Table 3: Housing Rank - Age < 29

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	— %
Logan Square	.355	.364	.341	.353	.338	.262	.368	.235	.388	.385	.398	.408	.391	.385	8.45%
Little Village	.248	.247	.244	.253	.234	.235	.200	.228	.207	.202	.224	.224	.230	.217	-12.50%
Gresham	.194	.201	.215	.197	.190	.209	.219	.227	.164	.173	.174	.158	.187	.209	7.73%
Englewood	.244	.257	.253	.239	.234	.262	.226	.254	.183	.157	.165	.162	.175	.205	-15.98%
Archer Heights	.197	.184	.191	.199	.191	.198	.174	.205	.209	.166	.186	.195	.198	.189	-4.06%
Lake View	.155	.171	.191	.185	.212	.186	.214	.156	.240	.215	.208	.191	.177	.187	2.65%
Brighton Park	.187	.189	.182	.186	.189	.145	.158	.145	.156	.164	.159	.183	.170	.179	-4.28%
West Rogers Park	.169	.165	.168	.176	.169	.169	.178	.185	.192	.206	.208	.191	.175	.178	5.33%
Rogers Park	.142	.139	.140	.144	.153	.126	.145	.135	.154	.166	.154	.154	.158	.174	22.54%
Marquette Park	.140	.144	.150	.176	.180	.164	.151	.160	.131	.132	.135	.124	.142	.167	19.29%

Above we can notice a few interesting trends, for example, some neighborhoods becoming a lot more influential for young workers such as Lake View, Rogers Park, and Marquette Park while others becoming more influential for older workers such as West Rogers Park, Portage Park, Bridgeport, and Albany Park.

There has also have been some larger shifts in where low income workers are moving such as moving into Gresham and Bridgeport whereas High Income earners have been flocking to Logan Square, Albany Park, and especially the South Loop whose importance has risen 66% since 2002.

Another interesting statistic we can calculate from using the eigenvector cen-

Table 4: Housing Rank - Age over 55

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	— %
West Rogers Park	.229	.248	.234	.220	.225	.186	.247	.183	.273	.293	.301	.302	.273	.281	22.71%
Logan Square	.231	.224	.223	.201	.211	.129	.196	.167	.234	.225	.240	.246	.265	.239	3.46%
Portage Park	.183	.202	.213	.204	.189	.220	.219	.239	.216	.240	.225	.215	.242	.235	28.42%
Gresham	.314	.292	.319	.312	.301	.280	.261	.283	.238	.238	.221	.220	.228	.229	-27.07%
Jefferson Park	.193	.200	.185	.194	.184	.189	.225	.194	.231	.208	.204	.214	.235	.208	7.77%
Bridgeport	.119	.139	.148	.134	.151	.189	.170	.190	.200	.185	.191	.209	.197	.194	63.03%
Albany Park	.155	.164	.159	.164	.150	.118	.176	.121	.174	.179	.192	.182	.175	.186	2.00%
Archer Heights	.159	.156	.155	.181	.173	.178	.168	.167	.194	.160	.163	.172	.158	.168	5.66%
Little Village	.150	.154	.164	.162	.169	.181	.150	.196	.155	.145	.167	.163	.171	.161	7.33%
West Pullman	.185	.205	.189	.177	.241	.172	.175	.179	.164	.153	.142	.138	.142	.161	-12.97%

Table 5: Housing Rank - Income < $1,250/month

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	— %
Gresham	.254	.259	.259	.256	.263	.247	.278	.287	.237	.234	.240	.220	.255	.280	1.24%
Englewood	.310	.296	.296	.290	.314	.325	.292	.299	.245	.223	.227	.222	.257	.271	-12.58%
Logan Square	.272	.283	.283	.262	.247	.183	.255	.194	.293	.286	.284	.304	.280	.246	-9.56%
West Rogers Park	.201	.224	.224	.232	.222	.190	.220	.195	.249	.272	.240	.248	.220	.218	8.46%
Archer Heights	.189	.204	.204	.195	.182	.191	.187	.181	.198	.190	.192	.207	.204	.211	11.64%
Bridgeport	.158	.185	.185	.182	.195	.196	.191	.206	.217	.196	.207	.197	.186	.190	2.25%
Little Village	.212	.208	.208	.199	.200	.201	.180	.184	.172	.164	.192	.206	.199	.183	-13.68%
West Pullman	.196	.180	.180	.153	.174	.142	.170	.171	.172	.137	.173	.151	.173	.177	-9.69%
South Shore	.172	.153	.153	.166	.174	.184	.190	.203	.144	.161	.155	.152	.158	.176	2.33%
Back of the Yards	.183	.172	.172	.185	.188	.164	.182	.164	.194	.171	.169	.171	.166	.172	-6.01%

trality scores is how neighborhood composition is changing. For example, while previously I just calculated ranking along one demographic indicator I can combine rankings to describe neighborhoods in new ways. One way to do this is by subtracting the neighborhood ranking of young and old workers or high income and low income workers. This could help detect neighborhood changes such as gentrification or income changes and potentially produce different rankings than if neighborhoods were scored according to a single demographic indicator. For example, it could be that a neighborhood is becoming a higher rank for both young and old workers. This would most likely be a different kind of neighborhood change than a neighborhood that becomes a higher rank for young workers and a lower rank for old workers. Below I calculate this metric by subtracting the trend lines of rankings between these two eigenvectors.

Though I have shown that much can be learned about a neighborhood by its eigenvector centrality. The neighborhood rankings produced are potentially hard to interpret and analyze. For example, while we know that a neighborhood ranks high for an age group and income level, what else does this ranking tell us

Table 6: Housing Rank - Income > $3,333/month

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	— %
Logan Square	.235	.261	.249	.240	.262	.198	.282	.221	.295	.305	.300	.331	.350	.335	42.55%
Lake View	.331	.342	.307	.332	.331	.263	.319	.289	.345	.323	.359	.330	.312	.321	-3.02%
West Rogers Park	.291	.273	.293	.272	.290	.262	.268	.302	.284	.281	.270	.265	.266	.264	-9.28%
Lake View East	.241	.257	.243	.278	.279	.202	.262	.185	.233	.239	.241	.237	.242	.243	.83%
Jefferson Park	.248	.257	.267	.246	.235	.251	.259	.237	.207	.214	.228	.237	.247	.226	-8.87%
Portage Park	.239	.216	.203	.196	.195	.226	.194	.212	.186	.190	.189	.191	.183	.202	-15.48%
Albany Park	.159	.161	.166	.171	.164	.159	.186	.138	.178	.204	.205	.204	.190	.191	2.13%
Rogers Park	.166	.174	.167	.180	.186	.165	.181	.157	.191	.190	.191	.188	.175	.187	12.65%
Ravenswood	.169	.171	.164	.157	.166	.145	.175	.150	.187	.190	.174	.170	.172	.187	1.65%

Table 7: Housing Rank Income Changes

Table 8: Low to High

Neighborhood	%
Bridgeport	0.456
Portage Park	0.386
West Rogers Park	0.384
Brighton Park	0.356
West Lawn	0.329
Albany Park	0.325
Little Village	0.306
Gage Park	0.261
Back of the Yards	0.219
Jefferson Park	0.208

Table 9: High to Low

Neighborhood	%
Gresham	0.640
Calumet Heights	0.437
Fernwood	0.409
Roseland	0.390
West Pullman	0.334
West Englewood	0.327
Chatham	0.324
Morgan Park	0.298
Rosemoor	0.294
Hegewisch	0.271

Table 10: Housing Rank Income Changes

Table 11: Low to High

Neighborhood	Rank
Marquette Park	0.423
Jefferson Park	0.420
Portage Park	0.405
Morgan Park	0.365
Beverly	0.343
West Lawn	0.328
West Rogers Park	0.327
East Side	0.292
Wrightwood	0.261
Garfield Ridge	0.254

Table 12: High to Low

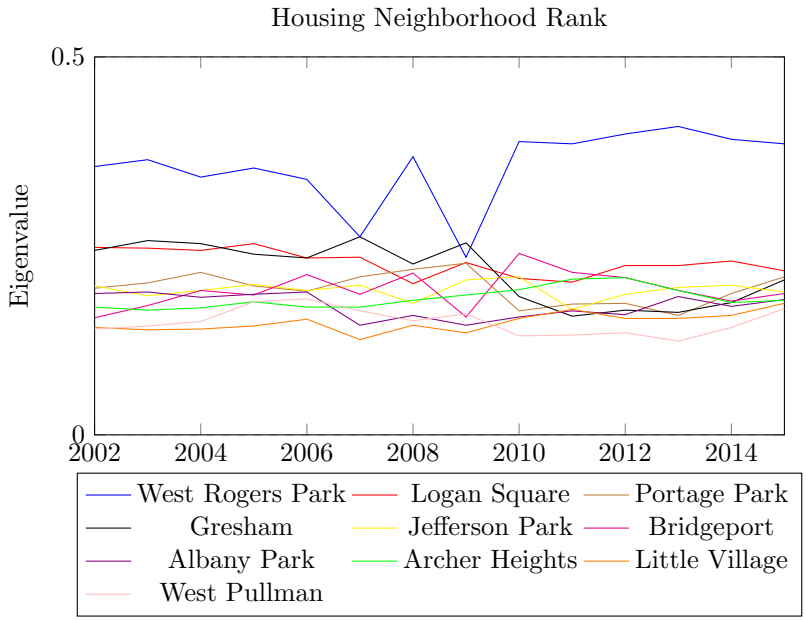
Neighborhood	Rank
Logan Square	0.757
Albany Park	0.492
South Loop	0.329
Humboldt Park	0.311
Ravenswood	0.227
Englewood	0.196
West Town	0.177
Heart of Chicago	0.176
The Loop	0.170
East Ukrainian Village	0.157

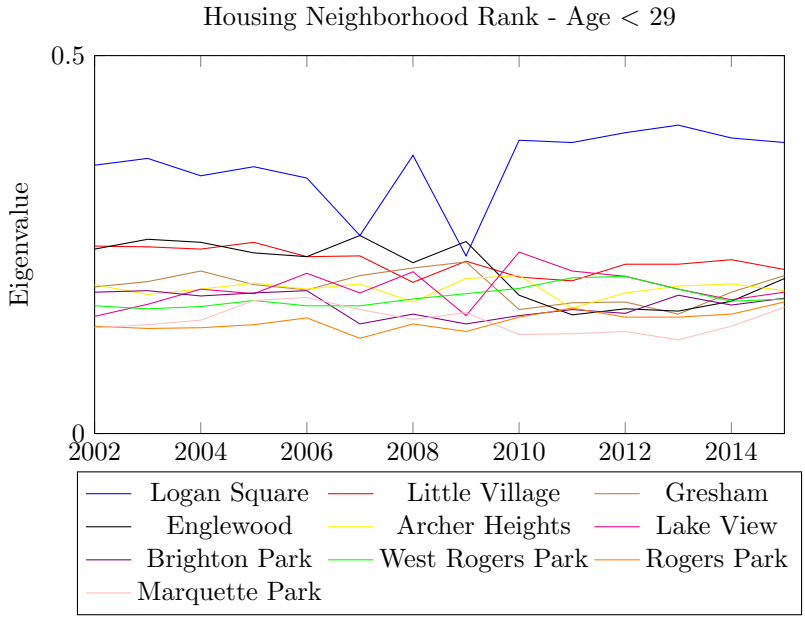
about a city. To analyze this I will look how centrality rankings correspond to other attributes about a neighborhood first by looking at correlation to housing stock and value.

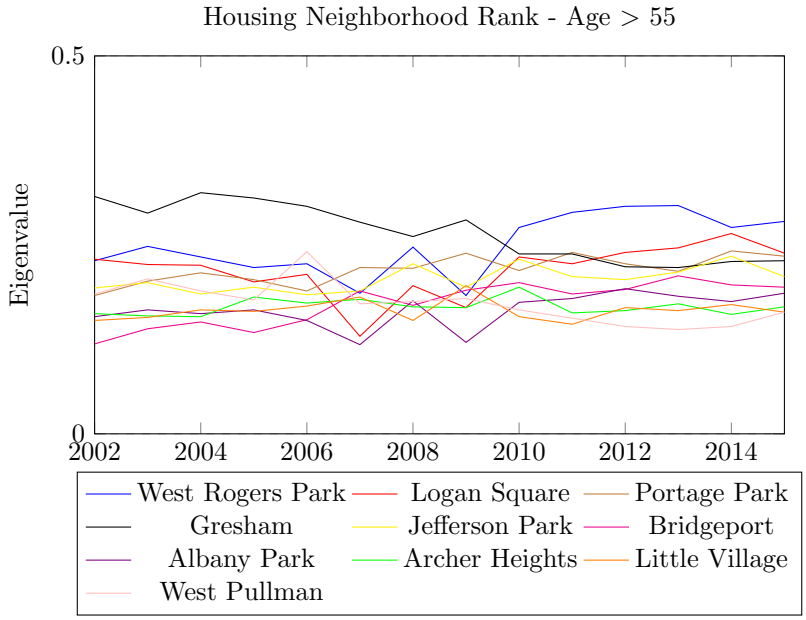
6 Conclusion

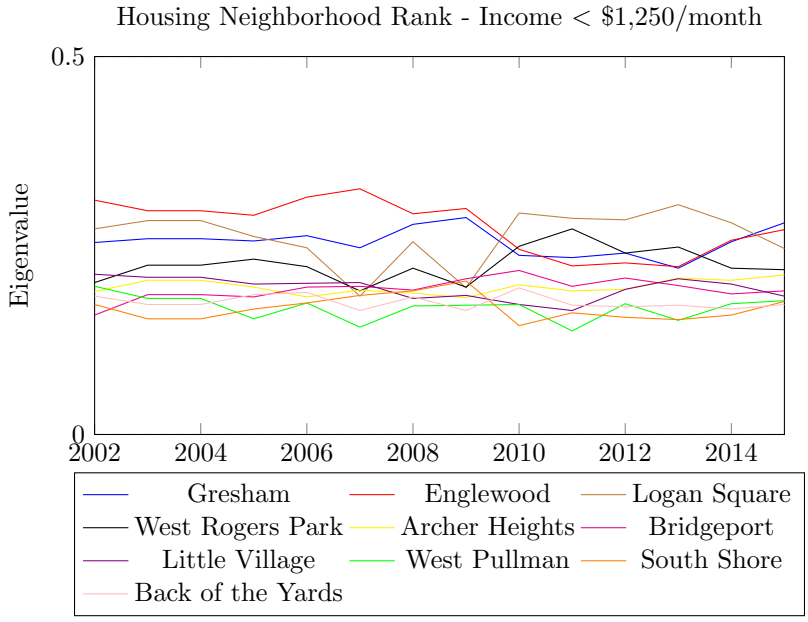
In this paper I presented a novel application of eigenvector centrality to dataset of Chicago's neighborhoods. After exploring the data through visualizations, I derived the math necessary for this application and outlined three examples where eigenvector centrality can help explore attributes of Chicago neighborhoods. First, I looked at how jobs were distributed across the city. In this analysis I found how employment was centralized in the Loop. While jobs in some industries such as Goods Producing are also located a west of downtown, The Loop neighborhood ranked at .95 while the next closest neighborhood ranked .17. Even low-income jobs were concentrated in The Loop which was surprising. While The Loop was a mixing pot of employment across industries, incomes, and ages, Housing centrality was much more segregated and dispersed throughout the city. For example, I found that

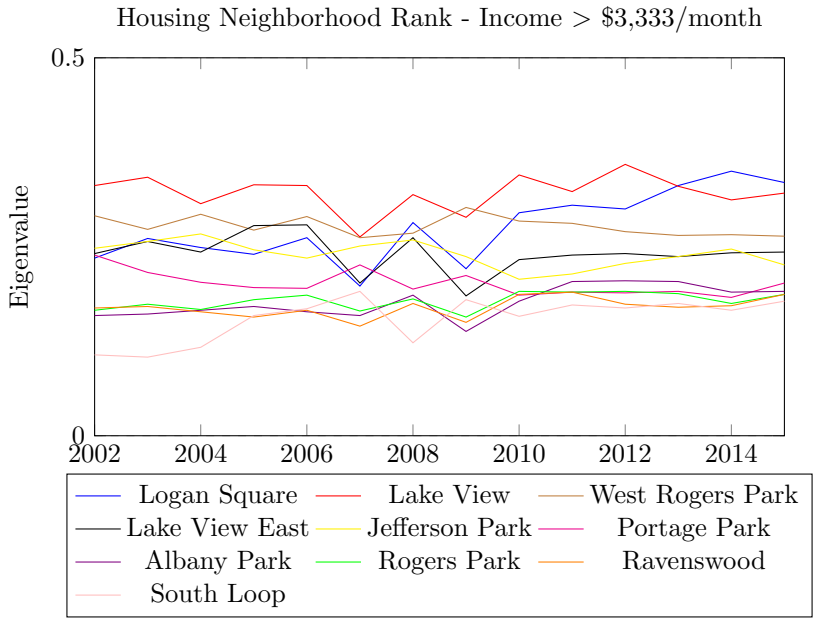
7 Appendix




Housing Neighborhood Rank - Age < 29
Eigenvalue
0.5
0
2002
2004
2006
2008
2010
2012
2014
Logan Square
Little Village
Gresham
Englewood
Archer Heights
Lake View
Brighton Park
West Rogers Park
Rogers Park
Marquette Park


Housing Neighborhood Rank - Age > 55
Eigenvalue
0.5
0
2002
2004
2006
2008
2010
2012
2014
West Rogers Park
Logan Square
Portage Park
Gresham
Jefferson Park
Bridgeport
Albany Park
Archer Heights
Little Village
West Pullman


Housing Neighborhood Rank - Income < $1,250/month
Eigenvalue
0.5
0
2002
2004
2006
2008
2010
2012
2014
Gresham
Englewood
Logan Square
West Rogers Park
Archer Heights
Bridgeport
Little Village
West Pullman
South Shore
Back of the Yards


Housing Neighborhood Rank - Income > $3,333/month
Eigenvalue
0.5
0
2002
2004
2006
2008
2010
2012
2014
Logan Square
Lake View
West Rogers Park
Lake View East
Jefferson Park
Portage Park
Albany Park
Rogers Park
Ravenswood
South Loop

References

[1] Peter R Gould. On the geographical interpretation of eigenvalues. *Transactions of the Institute of British Geographers*, pages 53–86, 1967.

[2] Karen S Cook, Richard M Emerson, Mary R Gillmore, and Toshio Yamagishi. The distribution of power in exchange networks: Theory and experimental results. *American journal of sociology*, 89(2):275–305, 1983.

[3] Phillip Bonacich. Power and centrality: A family of measures. *American journal of sociology*, 92(5):1170–1182, 1987.

[4] Keith J Tinkler. The physical interpretation of eigenfunctions of dichotomous matrices. *Transactions of the Institute of British Geographers*, pages 17–46, 1972.

[5] JW Moon and NJ Pullman. On generalized tournament matrices. *SIAM Review*, 12(3):384–399, 1970.

[6] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[7] Dawn Iacobucci, Rebecca S McBride, and Deidre Popovich. Eigenvector centrality: Illustrations supporting the utility of extracting more than one eigenvector to obtain additional insights into networks and interdependent structures. 2017.

[8] Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of mathematical sociology*, 2(1):113–120, 1972.