“How can Social Network Analysis and Mapping be leveraged to inform Chicago STEAM Out of School Time (OST) Policy Decisions?”

Sean Liu

Advised by Dr. Nichole Pinkard

Mathematical Methods in the Social Sciences

Northwestern University
TABLE OF CONTENTS:

1. INTRODUCTION
   a. MOTIVATION

2. CONTRIBUTION

3. LITERATURE REVIEW

4. DATA
   a. SOURCE OF ANALYSIS
   b. SCOPE OF ANALYSIS

5. ANALYSIS

6. CONCLUSIONS

7. REFERENCES
ACKNOWLEDGEMENTS:

Thank you to Dr. Nichole Pinkard for always being supportive and available, and to my parents and friends for acting interested when I talked about my thesis work.

MOTIVATION:

With nearly 100,000 after-school programs in the Greater Chicagoland area, there is no shortage of funding or program availability within the city (Cashin, 2022). However, this number of programs fails to guarantee equity in access to programming or diversity in programming type. Segregated housing and redlining policies in the city throughout the 20th century have contributed to the development of markedly different educational opportunities with wide variation for residents of the South Side of Chicago. In more recent years, selective closure of schools has led to further fragmentation and hoarding of opportunity – while some communities are overserved and have programs fighting for participants, other communities lack programs that are critical to career and personal development of Chicago’s youth (Cashin, 2022, Reisner et al., 2007).

Surveys aimed at reconciling and aligning employer offerings with after-school programs found that there is a noticeable lack of continuity between the skills that students are expected to possess and the skills that they are afforded the opportunities to learn, with considerable variance from community to community. By using survey data on community providers, schools, and employers, this paper attempts to use social network analysis and mapping techniques to shine light on areas of inefficiency within the network of Chicago after-school programs, identifying key providers that can be utilized to facilitate interaction between disconnected and siloed programs. Centralizing information on offerings could facilitate a realignment of programs, with
a focus on equipping each community with adequate program diversity and ensuring students are able to access said programs. Finally, this paper will aim to identify policy solutions and incentives that can catalyze and facilitate this movement within the Chicago after-school program network.

**LITERATURE REVIEW:**

The goal of the literature review will be to understand different perspectives on developing educational curriculums. By combining ecological approaches to learning and infrastructural methods of design with existing studies on STEAM Out of School Time (OST) programs, we can aim to approach an understanding of how Chicago can be best served by OST programs.

The first study I looked at was a field-defining book by John Goodlad, “A Place Called School.” The goal of Goodlad’s work was to achieve a nationally representative study on the inner-workings of selected schools. The study looked at over 1,000 schools across seven states, utilizing questionnaire surveys, interviews, and classroom observations across the span of 3 years. A key point in Goodlad’s research is that policymakers misinterpret parent’s expectations for schooling, suggesting a model of schooling within the framework of academic, vocational, personal, and social-civic categories. He notes that you can’t drive change through policy — rather, changes must be driven through a deep understanding of community stakeholder desires and then aligning policy correspondingly (Goodlad, 210). Goodlad goes on to discuss the importance of bringing inter-linked data to the school improvement process with specific goals and data collection methods for each school, as responsibilities for learning are distributed throughout the different parts of the community. While much of Goodlad’s work is done
qualitatively, it provides a strong starting point for developing an ecological understanding of the environment my study will be working within. Being familiar with key drivers of growth within school systems contributes to understanding of how OST programs can fit into them. For my specific study, it reinforces the need to leverage data from specific schools to uplift community voices and provide important insight for policy decisions.

Reisner et al.’s “Charting the Benefits of High-Quality After-School Program Experiences” is a comparative study consisting of findings from the Promising Programs study, conducted by researchers at UC Irvine, UW Madison, and Policy Studies Associates, Inc. The study used nearly 3,000 youth in 35 after-school programs that were based either in schools or in neighboring community centers. After vetting the programs to verify quality – based on staff support, available opportunities, and other enrichment activities – the researchers collected longitudinal data on the youth for two years. They used cluster analyses to test benefits associated with the experiences in the OST programs, finding that they improved youths’ conduct and habits over a two-year period (Reisner, 17). The study also noted that current policies focus narrowly on program level growth, and that even when they are part of a larger network, they are still mainly accountable for their own program quality. This results in a lack of emphasis on building programs from a community-wide basis, an area of focus for my thesis. Moreover, they noted that these programs often compete to attract youth with existing structures of support, such as community centers or church organizations. This suggests a lack of efficiency in designing program location, an area that could use expansion.

The next paper I looked at was William Tate’s “Geography of Opportunity,” an investigation into the way in which metropolitan development affects the educational and social environments of surrounding communities. This paper’s significance to my thesis is in
attempting to understand how infrastructural change and development can affect the ability for students to access local resources. The paper looks at two case studies: the Dallas Metroplex case study, which looks at geography of opportunity through the lens of middle schoolers, and a metropolitan St. Louis case study, which uses an ecological approach to analyze geography of opportunity. The first study follows the development of the telecommunications industry in Dallas, growing the economy to a GDP of $254 billion by 2004 (Tate, 24). However, the growth of this economy ousted other hubs, such as a once flourishing cotton and raw materials trade. This led to the development of poorer areas, where the education and living conditions of students were affected by aspects such as the number of liquor stores around their schools. Tate’s paper references studies showing that living in resource-poor neighborhoods negatively influences college aspirations of African American students, and that there is often a lack of functional community support structures for children. The second study discusses large scale industrial disinvestment in metropolitan St. Louis after WWII, and the effect that it had on local communities. It notes the manner in which the introduction of a rich biotech industry displaced thousands of residents and resulted in out-migration of high-income earners from inside the city (Tate, 31). Again, this leaves pockets of underserved and under-resourced communities inside the city, supporting the idea that uneven geographies of opportunity are manifest in metropolitan St. Louis. In my thesis, I would like to take steps to understand how the geographies of opportunity in Chicago affect access to STEAM OST programs.

Next, is Nichole Pinkard’s “Freedom of Movement.” The goal of Pinkard’s paper is to define the framework for a healthy learning ecosystem, in which students are able to move freely and unhindered through learning environments. Past research in the field, the paper says, has focused too much on formal schooling environments, discounting the importance of communal,
casual learning spaces. The paper aims to uncover the connective tissue necessary to facilitate movement through academic spaces. Pinkard’s work references back to her work with the Digital Youth Network, reflecting on how her experiences developing a learning space impacted her understanding of freedom of movement. In this program, she was able to build a walled garden—an ecosystem that has all the necessary infrastructure to mitigate the need for connective tissue but allows students to navigate freely throughout the space (Pinkard, 19). However, when this concept is expanded to larger spaces, Pinkard notes the way in which many aspects of a community are led by different community partners, leading to differences in motivation and frictions between stakeholders. The paper finishes by considering how freedom of movement can be attained via understanding how to link programs to existing transportation routes, leveraging existing spaces, and connecting with community stakeholders. This paper is important to my thesis, as it helps connect the ideas of infrastructure and ecological methods of learning, raising the question as to how communities can be interlinked most efficiently.

On the quantitative side, Nemesure et al.’s “A measure of local uniqueness to identify linchpins in a social network with node attributes” was a large influence in my work. In this paper, Nemesure et al. aim to identify an attribute-aware methodology of measuring a node’s importance to its immediate community. As such, they propose a concept called “linchpin,” which considers a particular attribute of a node—anything from provider type to community—and considers the proportion of first-order connections for which the node in question is the only connection holding that specific attribute (see figure. 5). Using healthcare provider data, Nemesure et al. show the ways in which linchpins can be used to identify areas of a network that would be susceptible to node removal, as well as the power of linchpin in identifying rural specialty providers.
Nemesure et al.’s work is influenced by a pioneering paper on community-aware social network analysis measures, Chen and Hero’s “Node Removal Vulnerability of the Largest Component of a Network.” This paper finds that using a greedy node removal algorithm based on solving a matrix one-norm minimization problem is more efficient in reducing large component sizes of a network with fewer node removals than overall network wide measures such as betweenness and degree. In doing so, their work inspired Nemesure and others to develop measures of analysis to understand which node removals would be most impactful to the overall flow of a network.

Finally, area-defining work on social network analysis in afterschool programs was conducted by Martha Russell and Marc Smith in “Network Analysis of a Regional Ecosystem of Afterschool Programs,” where they investigated the Dallas, Texas after-school program infrastructure. The paper aimed to identify strengths and vulnerabilities within the network, as well as provide insights for resource development and program advocacy. Using data on 525 afterschool programs and 25 support organizations, Russell and Smith found that many programs worked in isolation and relied on only 1 or 2 sources of funding. Additionally, many afterschool programs were unaware of other afterschool programs in the area. The network, as in Chicago, was a scale-free network, meaning that while some nodes had a lot of edges, the majority of the nodes had very few edges. Scale-free networks are resistant to accidental disruptions, as most edges are inconsequential to the health of the overall network. However, coordinated attacks on the network’s main hubs can be heavily consequential – an example of this would be a disruption of funding towards a large afterschool provider.
DATA:

I. Data Sources

The data used in the survey derives from two sources — a survey administered by the Digital Youth Network (DYN) in conjunction with Northwestern’s Office of Community Education Partnerships, and program data from City of Learning and My Chi, My Future (MCMF), a youth-focused initiative from former Mayors Lori Lightfoot and Rahm Emanuel focused on connecting Chicago youth to meaningful out of school experiences and organizing Chicago’s opportunity ecosystem.

The survey from the Digital Youth Network is categorized into four categories: School, Employer, Workforce Development, and Community Organizations. In each category, the survey recipient is assessed on factors such as topic offerings, program structure, scheduling and seasonality, and partnership with other organizations. The survey received 128 responses, culminating in a directed network with 128 nodes and 143 edges.

Data from MCMF was accessed courtesy of the City of Chicago and the office of Mayor Lori Lightfoot. My Chi, My Future consists of an online database and app that allow community providers to post opportunities and advertise positions. The program also provides subsidies and grant funding to programs in 15 neighborhoods across Chicago, providing up to $175,000 per program with a total budget of 2.6 million. In exchange for their support and coordination, MCMF collects information on key indicators of programs, as well as collaborators and partners.

Aggregating the past 10 years of MCMF data produces a network of 2,172 nodes and 2,579 edges consisting of local providers and partners – a network model of the composite data using Force Atlas 2 can be seen below:

1 A full copy of the survey can be located in the appendix
Data collected from MCMF is divided into community providers and locations within an undirected network. For example, Chicago Park District may work with a number of different locations throughout the city, and data is collected on collaboration between the provider and location. In addition to the 10-year aggregation, I will be conducting analysis on the data reported thus far in 2023, which consists of 124 nodes and 303 edges, as well as a dataset consisting of providers and location by neighborhood from 2017-2021. In doing so, I aim to identify ways in which the construction of the after-school network has evolved and adapted over the last decade.

II. DATA SCOPE:

The primary focus of the data will be on the DYN survey and MCMF data. Both datasets are limited to the Chicago area – MCMF data is specific to programs targeting 15–24-year-olds, while the DYN survey refers to K-12 education programs and related offerings. Findings from this paper hope to be applicable to programs looking to offer any after-school programming for
grade-school to high-school age students in the Chicago area, as well as policy-makers and funding sources aiming to determine resource allocations to specific organizations.

**ANALYSIS:**

Analysis of the data will follow a multi-pronged approach to data-visualization and representation. By identifying general trends using measures such as modularity, individual statistics such as betweenness centrality, and tools specifically built to identify uniquely positioned organizations, social network analysis will be used to identify which organizations should be targeted in conversation and incentive structures. Metrics and terms used in analysis include the following:

- **Degree:** the number of edges that a node has.
- **Betweenness:** the number of times that a node is dissected by the shortest distance between two other nodes.
- **Scale-Free:** a network with a small number of high-degree nodes, and a high number of low-degree nodes.
- **Modularity:** A measurement of connection density between sets of nodes.

I aim to use a combination of these methods of analysis to propose tactics that ensure different communities across Chicago have equity in access to programs.

In Figure 2, seen below, we see a representation of MCMF data in the context of a modularity map. Modularity, a measure of network detection, measures the strength of division within a network. Nodes within the same modularity class are revealed to have many connections within their communities but few facing outwards. In the context of the social
network analysis, the larger, central nodes within the modularity map represent strong ties to both their own modularity classes and others, allowing for areas of high collaboration.

Figure 2: MCMF Modularity SNA – Node size correlates to # of Edges

In the context of the social network analysis, the larger, central nodes within the modularity map represent strong ties to both their own modularity classes and others, allowing for areas of high collaboration. Here, we can see the importance of programs such as After School Matters, the Chicago Park District, Digital Youth Network, and the Chicago Public Library system.

In Figure 3, we observe the same network after applying Fruchterman-Reingold analysis to untangle and group the nodes via attraction-repulsion, and then Force Atlas, a strong-force algorithm used to disperse groups and give space around larger nodes. The graph is then color coded by modularity score, with the node size corresponding to betweenness centrality, a measure of node influence over information flow.
The more times that a node is dissected by the shortest distance between two other nodes, the higher the betweenness centrality of that node. Applied to the data, we can identify nodes that have the capability to both expedite and slow-down data transfer and communication through the system. Research into social networks has shown that central actors may create bottlenecks within large networks – these factors are worth factoring in when considering the redesign of a network for greater efficiency (Daly, 2015).

Figures 4 and 5 show the same analysis done on the DYN Survey Data. Here, we can see the manner in which Bowa Group, Bank of America, UIUC, and Kenwood Academy occupy unique spaces within modularity classes, serving as key connections between them. Similarly, analyzing betweenness centrality amongst the survey data points redirects attention towards organizations such as NextWave Stem, which occupies a strong position within the information flow of the network with its higher betweenness centrality.
Within social networks, attention is often given to the nodes in the network with higher degree, meaning that low degree nodes can be easily overlooked. This is due to the fact that network-wide measures, such as degree and betweenness, are calibrated to identify the highest influence actors within the network. However, it is important to consider the positions that lower degree nodes in a network occupy as well, as they contribute to their local communities. Only by focusing on both inter-community and intra-community strength can the overall network be strengthened (Goodlad, 2005).
Using linchpin analysis, we can identify how unique the contribution of a node is to its surrounding communities, based on the attributes that it holds. In the case of our data, attributes include program offering type and neighborhood. If the removal of a node from a network affects the surrounding nodes’ ability to access that attribute, it receives a higher linchpin score. This allows our analysis to identify nodes that may not have many connections but have connections of higher value. In figure 6, we can observe a diagram depicting linchpin score calculation. Because linchpin scoring is community and attribute-aware, it reveals information about a network that is distinct from the aforementioned network-wide measures such as betweenness and modularity. In figure 7, we can see a measure of the manner in which linchpin scoring doesn’t correlate with those measures. In figure 8, we can see further evidence of this holding true specifically in our own MCMF datasets, with varying levels of betweenness centrality not having any effect on the linchpin scores of a specific node.

Measures of this type are particularly important because our data represents a scale-free network, meaning that many of the connections are concentrated within a select few nodes, while most of the nodes have very few connections. This is due to the nature of key central providers connecting to a variety of different locations and areas within the city. We can see this represented in figure 3, where the Chicago Park District and After School Matters nodes dominate the network based on betweenness centrality. Due to the fact that linchpin score only considers the interaction of a node with its first order connections, it is able to bring attention to individual providers within the larger network, allowing insight into micro-interactions on the Chicago afterschool program community level.
Figure 6 (LEFT): Linchpin Scoring Representation (Nemesure, 2021)

Figure 7 (RIGHT): Correlation Heatmap of SNA measures on a Providence, RI physician network (Nemesure, 2021)

Figure 8 (LEFT): MCMF data plotted with respect to linchpin score and betweenness centrality

Figure 9 (RIGHT): MCMF Network – node size represents linchpin score, color represents modularity class

In figure 9 above, we can see a representation of the same MCMF as figure 2, but with node size corresponding to linchpin score. Here, linchpin score is representative of modularity uniqueness – for any given node, the proportion of its connections that aren’t connected to a different node from the same modularity determines the size of the node. Because modularity is determined by closeness of connection with other nodes in the same modularity, each modularity serves as a
proxy neighborhood within the greater community. A node with a high linchpin score stands as the only connection of that modularity for a majority of its connections. As such, if that node were to dissolve or move, it would remove these lines of connection between modularities, disrupting important cross-community connections.

By moving the area of focus away from the high-degree nodes and towards identifying nodes that are important to their individual networks, linchpin analysis helps us to identify potential areas of weakness within the after-school program infrastructure. In Krackhardt and Stern’s “Informational Networks and Organizational Crises” (1988), they study the manner in which networks respond to interruptions and crises within a network. Through their research, they found that organizations with high density of strong connections between subunits responded significantly better as a whole than organizations where the strong connections were mostly contained within the subunit. Using modularity as a proxy for neighborhood, we can identify cross-neighborhood connections that are important to support and foster in order to sustain the health of the community at large. This informal construction of neighborhoods is useful in scenarios where we don’t have strict community data, but also in situations where that data is present.

Nemesure et al. observe that within a healthcare network, distributing specialties randomly into different communities resulted in higher average linchpin scores than when calculating the scores based on observed network, implying that the emergence of patient-sharing patterns that make networks less vulnerable. Working with a subset of the MCMF data where we have access to neighborhood information, we are able to run a similar comparison between arbitrarily assigned modularities and established neighborhood networks. Taking the top 11 most populated neighborhoods from the data, I ran a linchpin analysis based on
neighborhood, and compared the average linchpin score to the same set of data when assigned into modularity classes – these modularity classes don’t consider geographic location, neighborhood type, or program offering, rather focusing strictly on the density of connections towards other nodes in a particular modularity. In figures 10 and 11, below, you can find a depiction of the networks described. In figures 12-15, we can see a depiction of the distribution and density of linchpin scores for the modularity-based network and the neighborhood-based networks. Importantly, we can see that the modularity-based network has significantly lower average linchpin scores and lower densities of high linchpin scores, indicating that individual subsets of the community are less siloed and disconnected from others. Lower linchpin scores by neighborhood could also point towards geographical and infrastructural limitations, such as limited public transportation into particular neighborhoods, or improperly distributed resources across different communities.

Note that this deviates from the findings of the Nemesure study – in this case, the subsets constructed without attentiveness to individual attributes of the community were less vulnerable, not more vulnerable. However, while the attributes in the Nemesure study were based on provider type, the attributes in our MCMF data were geographic, suggesting that neighborhoods themselves may not allocate enough attention to connecting with providers in other neighborhoods, focusing on the strength of their individual subset’s network rather than the network at large. These network subsets may not be aware of their lack of connectedness with other networks and can be susceptible to opportunity hoarding when working as siloed communities (Cashin, 2022). While we don’t currently have access to information regarding provider type, future work would be interested in seeing if Nemesure et al.’s work could be replicated using provider type data.
Figure 10 (LEFT): 2017-2021 MCMF Network – node size represents linchpin score, color represents neighborhood
Figure 11 (RIGHT): 2017-2021 MCMF Network – node size represents linchpin score, color represents modularity

Figure 12 (TOP): 2017-2021 MCMF Network – Average linchpin score per modularity class
Figure 13 (BOTTOM): 2017-2021 MCMF Network — Average linchpin score per neighborhood
Figure 14 (TOP): 2017-2021 MCMF Network – Linchpin score density by modularity class
Figure 15 (BOTTOM): 2017-2021 MCMF Network — Linchpin score density by neighborhood
CONCLUSIONS:

While there has been limited applications of social network analysis to after-school program data, the majority of the work consists of network-wide measures or focuses on student experience rather than program to program or program to location connections. Combining network wide analysis with a focus on individual communities allows us to not only derive stronger micro-level conclusions regarding provider interactions, but also allows us to provide key information to the providers and communities that they would not typically have access to.

Consider the fact that a program is not always aware of its second-degree partners – the partners of their partners. As such, programs that have high linchpin scores, meaning that the network would be left vulnerable if the program were to dissolve or leave, are not aware of this fact. Communities can leverage linchpin data to provide additional resources that shore up vulnerable spots for their overall community. Furthermore, a program’s individual knowledge that they have a high linchpin score gives them agency to request additional funding on the premise that they hold a particularly important role within their local community. On the contrary, the realization that a program has a low linchpin score – their neighborhood is infrastructurally capable of handling their departure – may provide them with the freedom to move to new neighborhoods and regions that may not share the same luxury. At a city-wide level, the recognition that certain neighborhoods have higher linchpin scores, meaning that they are more siloed and closed off, can be leveraged to foster cross-neighborhood initiatives.
APPENDIX:

Survey Questions for Digital Youth Network:

<table>
<thead>
<tr>
<th>School</th>
<th>Employer</th>
<th>Workforce Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Instructional spaces (computer lab, digital media studios, recording studios, maker spaces, science or engineering labs, etc.)</td>
<td>• Industry/Category</td>
<td>• STEM-related workforce development offered</td>
</tr>
<tr>
<td>• STEM topic areas</td>
<td>• STEM Topics related to jobs</td>
<td>• Frequency of STEM-focused opportunities</td>
</tr>
<tr>
<td>• Partnerships with organizations</td>
<td>• Top skills they look for</td>
<td>• Availability of virtual programming</td>
</tr>
<tr>
<td>• Frequency of Staff professional development</td>
<td>• Partnerships with Chicago schools in finding employees</td>
<td>• Schedule of opportunities</td>
</tr>
<tr>
<td>• Structure of programming</td>
<td>• Education levels of current employees</td>
<td>• Age range of current participants</td>
</tr>
<tr>
<td>• Source of STEM programming at</td>
<td>• Local workforce development opportunities offered</td>
<td>• Advertising methods of programs</td>
</tr>
<tr>
<td>• Cost to families</td>
<td>• Remote opportunities beyond COVID</td>
<td>• Primary service area</td>
</tr>
<tr>
<td>• # of students reached</td>
<td>• Workforce needs</td>
<td>• Partnerships with organizations</td>
</tr>
<tr>
<td>• STEM needs and/or opportunities</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Community Organizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Availability of publicly accessible computers/wifi</td>
</tr>
<tr>
<td>• STEM topic areas</td>
</tr>
<tr>
<td>• Ages of young people served</td>
</tr>
<tr>
<td>• Formal leadership opportunities for ages 16-24</td>
</tr>
<tr>
<td>• Presence of unstructured “hang out” spots</td>
</tr>
<tr>
<td>• Dedicated instructional spaces</td>
</tr>
<tr>
<td>• For example: computer labs, art studios, digital media studio, maker spaces, science or engineering lab, auditorium, kitchen, or gardens.</td>
</tr>
<tr>
<td>• Structure and timing of STEM opportunities</td>
</tr>
<tr>
<td>• Primary goals of STEM opportunities</td>
</tr>
<tr>
<td>• # of young people reached</td>
</tr>
<tr>
<td>• Transportation support</td>
</tr>
<tr>
<td>• Advertisement methods of programs</td>
</tr>
<tr>
<td>• Partnerships with organizations</td>
</tr>
<tr>
<td>• Challenges and opportunities in STEM learning and development</td>
</tr>
</tbody>
</table>
Bibliography:


