Final Project Report – Analyzing the performance of Chicago Public Schools with a Bipartite graph model and Community Detection

Srivasa Pranathy Mutcherla - A20555875

Kiran Velamati - A20525555

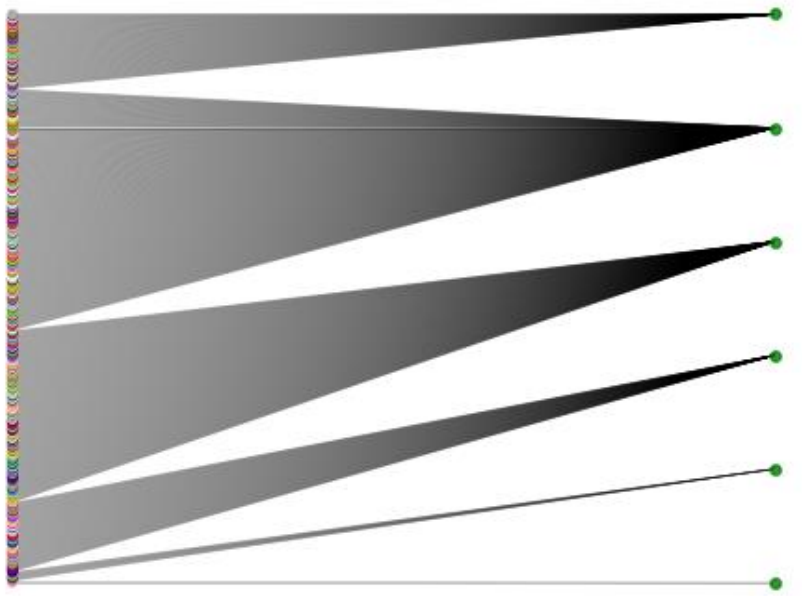
When we were assigned this project, we were told we had the absolute freedom to go with a piece of work that inspired us. That flexibility encouraged us to look into the Chicago Data Portal, as we wanted to take up a problem statement with meaning and impact. Our first finding was the energy dataset, which covered the Chicago buildings and their energy usage. We planned to derive insights that could help go greener.

Since the network model was not readily designable, we moved on to the education department, where we found a dataset that was a progress report of the Chicago Public Schools that spanned over a year. The city had 566 schools, with 79 metrics associated with it.

We intended to identify patterns for the school's performance, and location was an attribute. Owing to the reasonable number of rows, we did not have to sample them. This decision remained the same as we mentioned in our project proposal. While iterating through the columns, we worked on determining the most relevant columns that could help us determine the school's overall performance. The extensive scoring on a school's holistic growth helped us. Once we identified the columns, we could discretize them after analyzing the scoring patterns. This helped define a cumulative column that gave each school an overall score that we normalized. Once our data was ready, we needed to finalize a network model to capture the data and derive the insights.

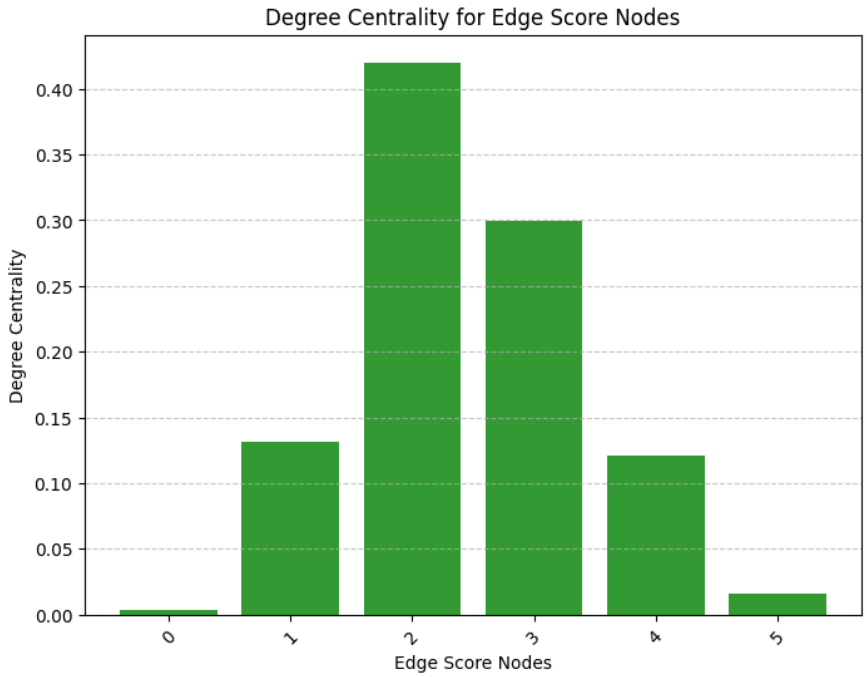
With Prof. Cynthia Hood's help, we chose a bipartite model that mapped school nodes to their performance scores. Our next hurdle involved representing the graph, as there were 566 schools and only scores ranging from 0-6. With that, the spacing proved to be a little complicated.

Bipartite Graph of Schools and Edge Scores with Community Detection

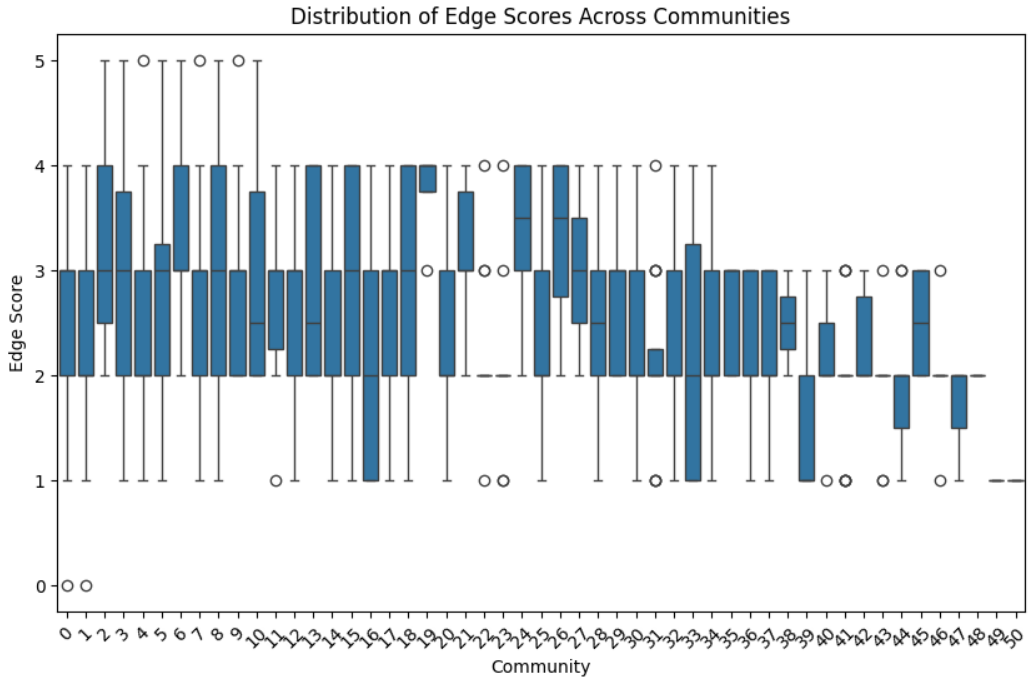


Once satisfied with the network model, we wanted to work on community detection to see the relationship between a school's location and performance. We used NetworkX documentation and ChatGPT to help with the code and insight derivation.

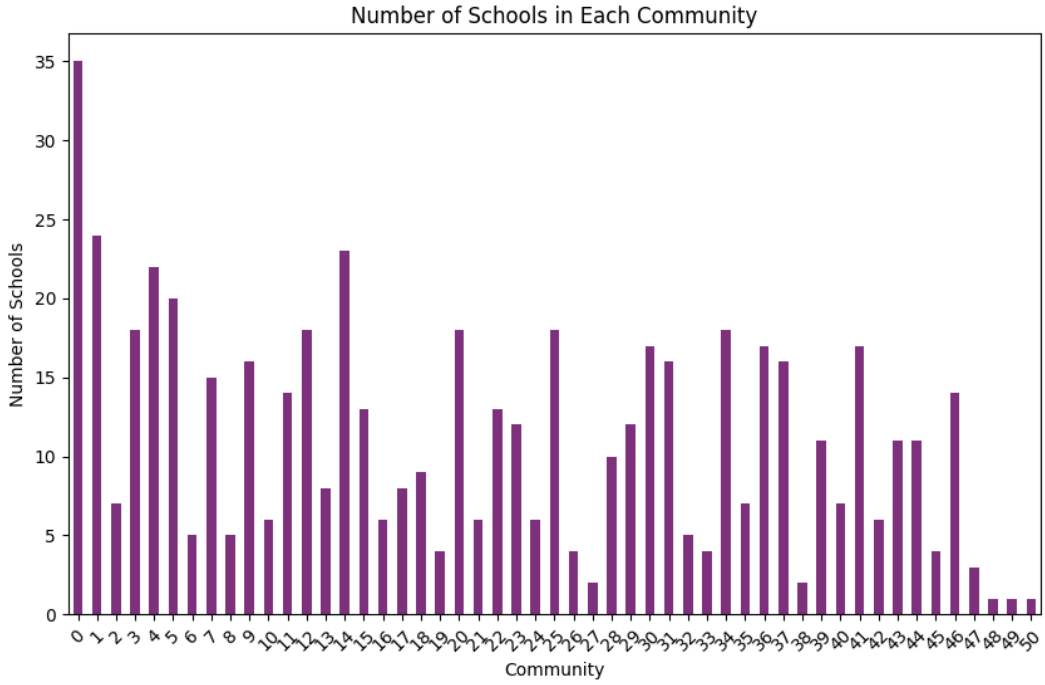
Community detection could be done with various algorithms, some of which we tried, including the Louvain, Girvan-Newman, and zip-code-based grouping algorithms. The most relevant one was to map schools in the same code to a community. This is also visible in our graph, where schools in the same community have the same color. We calculated the degree centrality of the score nodes, which is high for the score the schools around the city most achieve.



We also estimated the distribution of the scores over the geographic locations, and while there seems to be some pattern, we haven't pinpointed it yet. There wasn't any linear proportionality that could readily identify patterns. This plot helped us mark the maximum, minimum, and mean scores of performances in the communities, and we could see some communities perform relatively better than the others.



The way we see it, with a little more time, we can find other variables that could have a pattern. Like the network managers, to see if they contribute to the performance and identify any pain points that need to be improved. The number of schools in a particular zip code was 35, with the least number of schools in a community being 1 or 2. This raises the question of what could be causing an uneven distribution of schools throughout the city.



We could further examine if any socioeconomic factors come into play with these numbers. Even the scoring mechanism we went with could be improved to allow for more spread-out values that could help with better pattern identification. Having an open-ended question pushed our boundaries to help us analyze the data and think about what makes sense to be included and what does not, what insights could be derived, and what would be helpful. It was an exciting experience.

References:

https://chatgpt.com/

https://networkx.org/documentation/stable/