Exploring the Role of Networks in Predicting and Improving Social Mobility

Shricharan Baskaran, Chaundy McKeever, Charumitra Ravi

Introduction

The American dream, that any person from any station in life can achieve success in America because of equal opportunity, is proving to be less and less true. Social mobility, which is one's movement up the socio-economic ladder, has been on the decline in the United States. Research now shows significant evidence supporting generational poverty and recent studies have found that each consecutive generation's likelihood of earning more than their parents and moving into a higher social class is decreasing. This leads to critical questions about what variables promote or hinder social mobility and how to make social mobility more feasible, especially for those from lower socio-economic backgrounds. One possible under-explored factor is social networks.

Several studies discuss relationships between social capital and economic opportunity [1] and income inequality [2]. Here we consider connectedness to those of a higher socio-economic status as a measure of social capital and explore what impact the social capital of an individual's social network has on their evolving economic status. This is supported by the work done by Chetty R. et al. [4], who found that economic connectedness is one of the strongest identified predictors of intergenerational income mobility.

The goal is to develop social networks between zip codes that are analogous to the social networks of individuals within the communities. By analyzing the Socio-Economic Status (SES), connectedness, and social mobility of these communities we can explore the relationships between social networks and social mobility. We hypothesize the following,

- Zip codes with high economic connectedness should be high in terms of economic mobility rate.
- Zip codes with a high pagerank, indicating connectedness, with a high socioeconomic status will have a higher rate of social mobility than comparable communities with lower social capital.

By improving understanding of the factors that contribute to social mobility, policymakers can make better informed decisions that promote economic growth and reduce rates of generational poverty, or individuals may be more proactive in developing their social networks in such a way as to increase their chances of achieving the American dream. Ultimately we hope to use network measures to contribute to the ongoing conversation about the problematic wealth disparity, decreasing rates of intergenerational betterment, generational poverty, and the role social interactions can play in resolving these issues.

Methods

The network focuses on three measures, socio-economic status, social connectedness, and social mobility. The median income in the zip code in the year 2018 [5], determines its socio-economic status. The next critical parameter, Social Connectedness Index (SCI), we obtain from facebook social network using friendship data from users between locations (data is publicly available from [6]).

The data on SCI is only available for the time period around 2018, and there is no data for significantly older periods. Despite this, the geographic structure of networks between regions, which is captured by the SCI, is believed to be stable over time [7]. Therefore, measuring social connectedness today is likely to be a good predictor of interactions in other time periods.

This enables a representation of connections from childhood that make up some of the strongest links in a social network and those relationships will contribute strongly to the choices that significantly impact social mobility, such as secondary education and career interests.

The SCI is obtained from the facebook connectivity data [6] between zip codes, which was originally calculated by the below method for two locations i and j

$$Social\ Connectedness\ Index_{i,j} = \frac{FB\_Connections_{i,j}}{FB\_Users_i * FB\_Users_j}$$

Social connectedness is measured as the proportion of facebook connections that exist between two zip codes and the number of possible connections that could exist, based on the facebook users within those zip codes.

Finally, social mobility is based on the current average income percentile of members of the sample population (around age 35) whose families were in the 25th percentile of the national income distribution during the year they were born [4].

To calculate the Economic connectedness, we use the following technique,

- General Formula for Economic connectedness:
  EC = (number of high-SES friends / total friends) / 0.5
- In our case, since it is a network of weighted edges, with weights as the SCI, and nodes as Zip Codes, we calculate with the following,
  EC = Sum of weights of edges with high SES zip codes / Total sum of weights of all edges

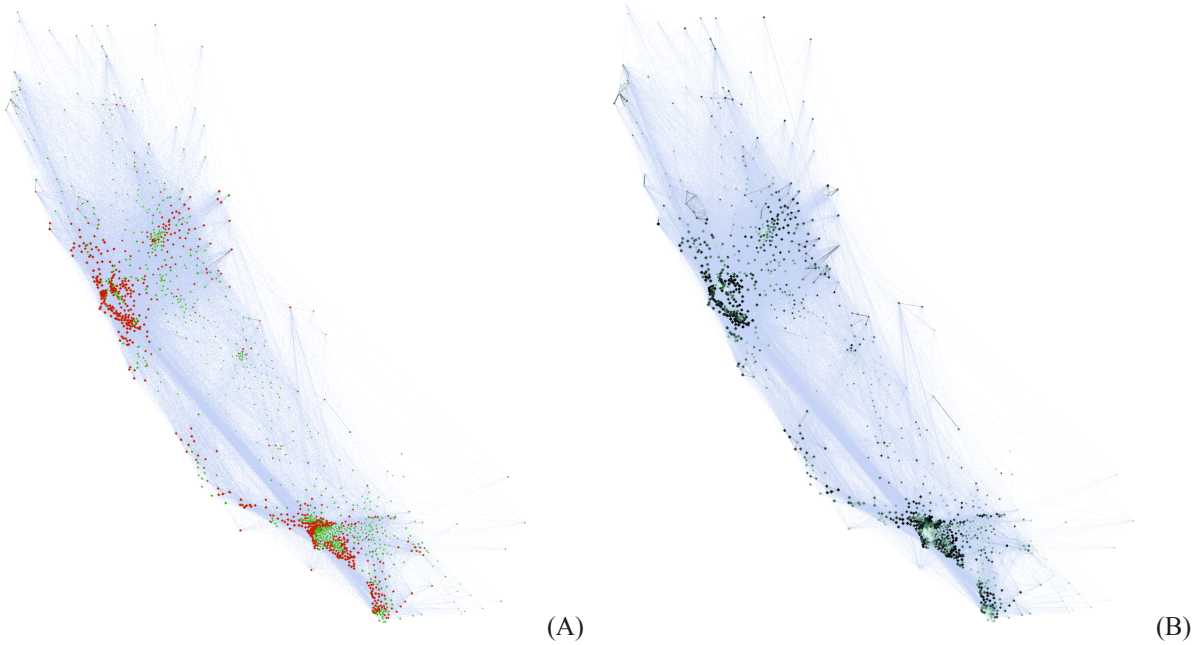
We obtain the mean incomes for each zip code from US Census [5]. For the calculation of economic connectedness to be implemented, we need to have data to decide if a zip code belongs to the High SES or Low SES category. To find that, we select the zip codes we would like to do the analysis (in this case, we did analysis on New York and New Jersey States and on the state of California) and find the median income of the mean incomes of zip codes that are being analyzed. If a particular zip code's mean income is less than the median of mean incomes of zip codes that are being analyzed, we assign it as low SES and if not as high SES.

We had over 110 million edges, since the network is almost a fully connected one. To limit the size of the Network and make the analysis feasible, the analysis was restricted to three regions in the United States. We did analyze New York and New Jersey zip codes together as one and the state of California zip codes as the second. In both cases, nodes without available data on the three measures were removed from the set. To make it easier for comparison, we calculate the mobility rate percentile of each zip and update it as an attribute to the node.

To test our second hypothesis, which is checking for any correlation between the page rank of a node (zip) and the mobility performance of the zip, we remove the low SES zip code to low SES zip code edges. This is done particularly to limit the transfer of page rank score from a low SES node to another low SES node, since we are particularly hypothesizing that the connectedness with a high SES node makes a better chance for better social mobility rate. The page rank should consider the weights of the edges in calculating the page rank scores.
Gephi was used to create visualizations of, and analyze, the data.

Results

Analysis on California:



(A) (B)

Fig 1 - California zip codes plotted as nodes whose sizes are proportional to their economic connectedness with weighted edges of connectivity between zip codes. (A) High SES, those above the 50th percentile, are colored red and low SES nodes are colored green. (B) Node darkness corresponds to mobility, where high mobility nodes are darkest.

Fig 1 shows a pattern between high SES zip codes and high mobility zipcodes.There is also clear grouping amongst nodes of similar mobility and SES, this is to be expected in the regions with dense population where it is customary to see segregation of richer and poorer areas. In the larger bottom cluster you can see some blending in the color of the low SES nodes that may indicate that they are benefitting from proximity to high SES nodes in terms of mobility. In Fig 2 the page rank of each node is graphed versus its mobility. There is a slight but dubious trend to the data, in that higher page rank corresponds to higher mobility. This is what we hypothesized but it is not well supported by this data. Most notably the lower page rank values become sparser at higher mobility percentiles.

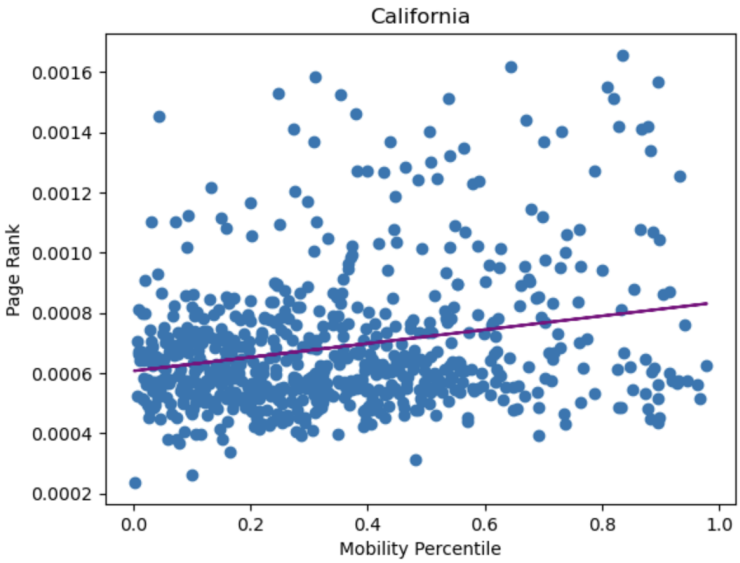


Fig 2 - Plot of mobility percentile versus page rank for California zipcodes, slight upward trendline

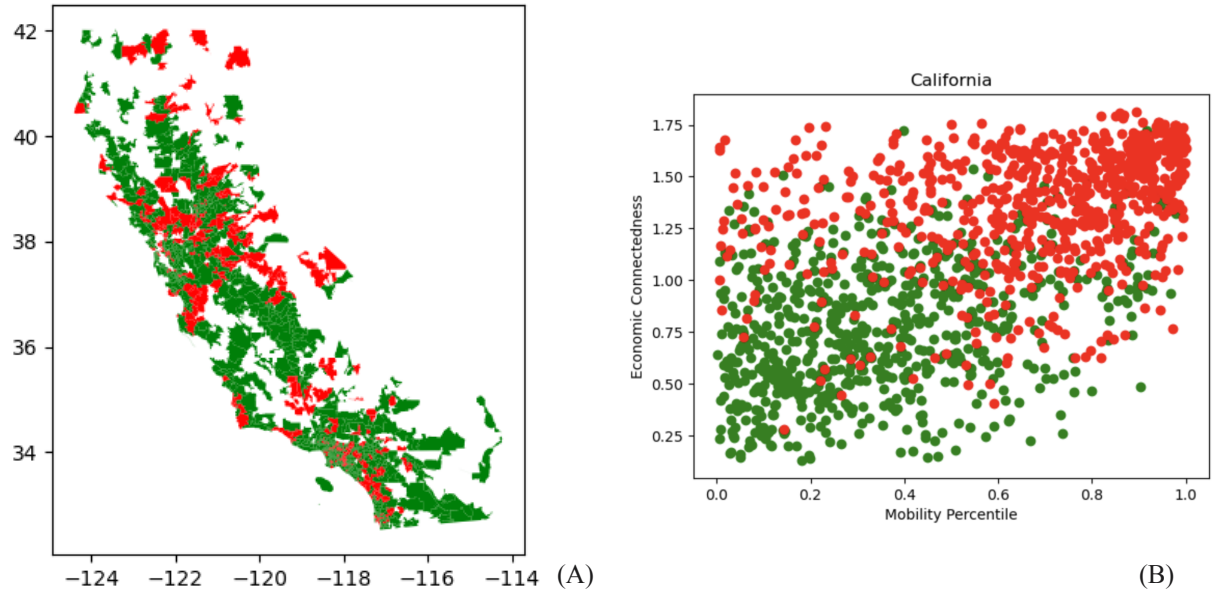


Fig 3 - (A) Geographic plot of zip codes where the area is colored green if the zip codes EC is proportional to its mobility percentile and red if not. (B) Plot of mobility percentile versus EC, high SES nodes are colored red and low SES nodes are colored green

For a large number of zip codes in California the economic connectedness is found to be proportional with the mobility percentile. Fig 3(A) shows this and when compared to Fig 1(A) there is no hard pattern between proportionality and SES of the zip. In Fig 3(B) it can be seen

that there are clear areas where the two SES groups lie and a correlation between economic connectedness and mobility percentile. It is as expected that high SES nodes have high economic connectedness

Analysis on New York:

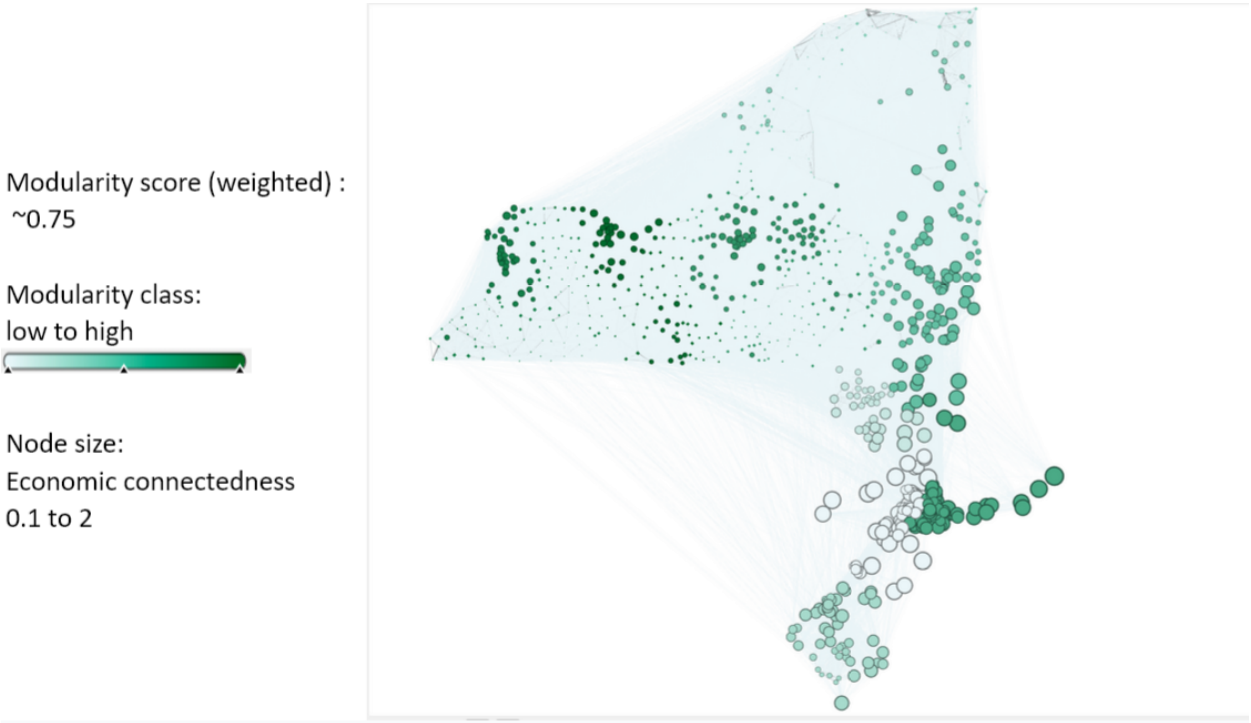


Fig 4 - New York and New Jersey Zip codes visualized based on their Modularity Class Value color shaded and Node size based on economic connectedness

From Fig 4, it can be seen that the similar communities (similar colored nodes) are regionally localised,, showing that mobility and economic connectedness are somewhat similar in a community of nodes in the same region . Also, in the NJ central area, we can see a cluster of red nodes, which shows the presence of a large number of high ses zip codes. Within those nodes, we are also able to observe large green nodes(low ses nodes with high economic connectedness).

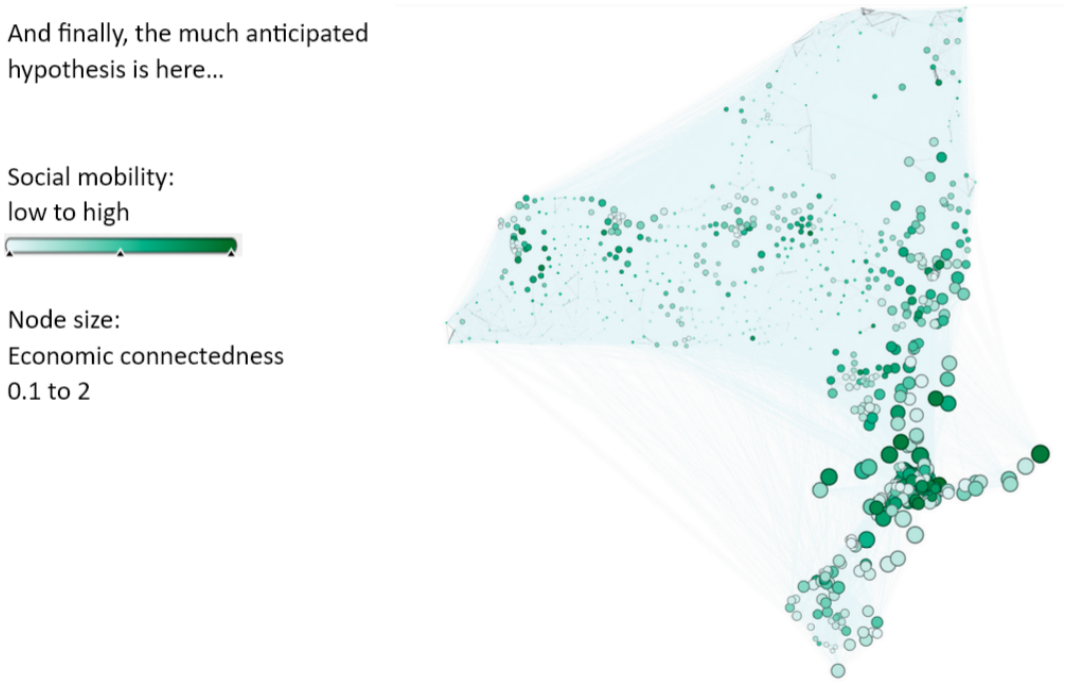


Fig 5 - New York and New Jersey Zip codes visualized based on their Mobility percentile color shaded and Node size based on economic connectedness

From Fig 5, it can be seen that, just like the earlier visualization, we are now focusing on the social mobility as the color of the nodes. Proving our hypothesis, we can see that larger nodes are mostly darker colored, meaning stronger ec = higher mobility, centered around the busy jersey area. As you can see, most of the interesting activity is around the manhattan brooklyn area, where we can see a variety of lighter and darker nodes

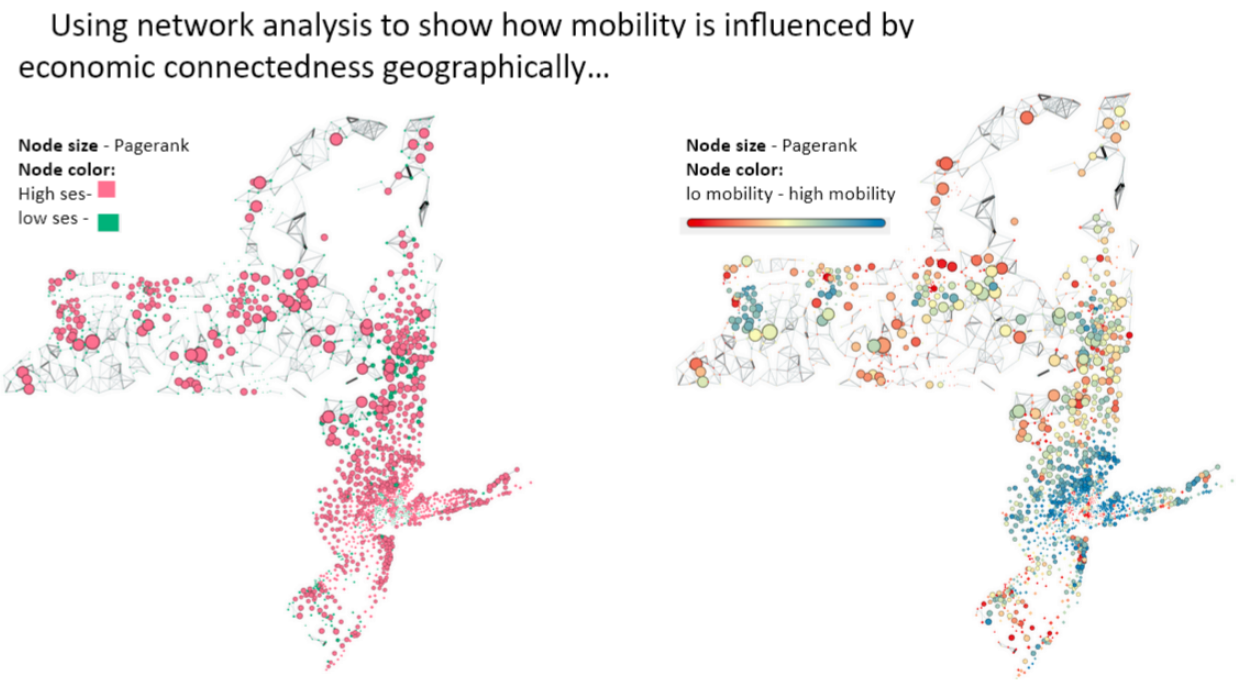


Fig 6: The two figures show the relationship between different measures of connectivity and socioeconomic status (SES) in a geographic network of zip codes in the New York-New Jersey region. The first figure shows node size as the PageRank measure of connectivity and node color as the SES measure, where green represents low SES and red represents high SES. The second figure shows node size as the PageRank measure of connectivity and node color as the mobility measure, where the color range goes from light to dark blue for low to high mobility percentile values. Both figures suggest that zip codes with higher PageRank tend to have higher SES and mobility, indicating that they are more connected to other regions and likely more prosperous.

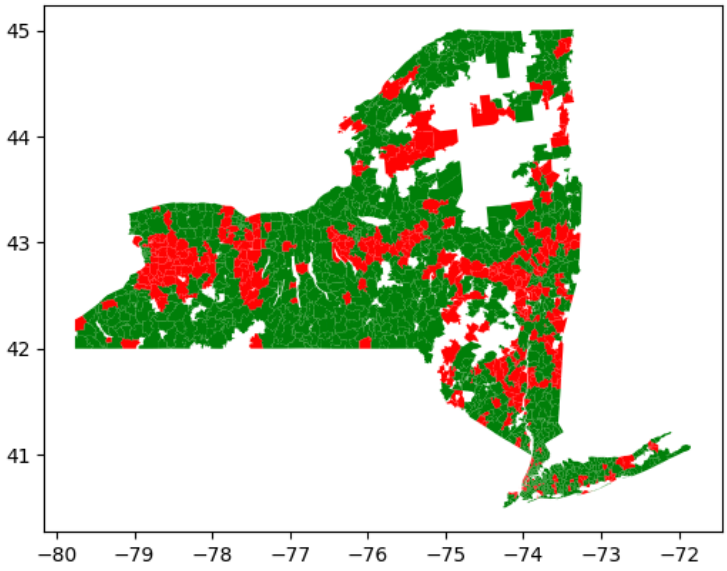


Fig 7 - New York and New Jersey Zip codes Geo Plot showing the zip codes where hypothesis was held (green) and where it wasn't held (red).

The Fig 7 indicates that with New York and New Jersey state zip codes, we observe that with a good number of zip codes, we see that the hypothesis of Economic Connectedness being proportional to Social mobility performance to be right.

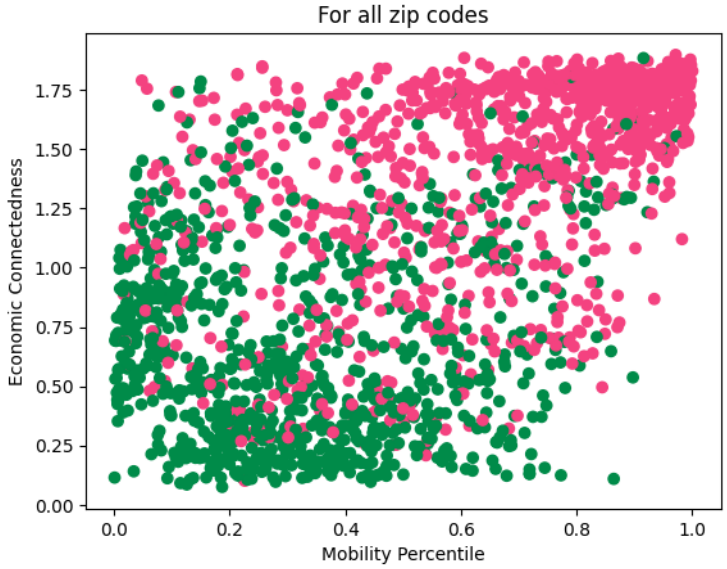


Fig 8 - New York and New Jersey Zip codes' scatter plot of Economic connectedness against Mobility Percentile

In the Fig 8, we plotted for the New York and New Jersey Zip Codes of Economic Connectedness against Mobility Percentile by faceting with the SES type through the color of the points (Green - Low SES, Red - High SES). It can be observed that with high SES zip codes, the performance is significantly better than low SES. Also the correlation of Economic connectedness and Mobility percentile is more pronounced with the high SES zip codes.

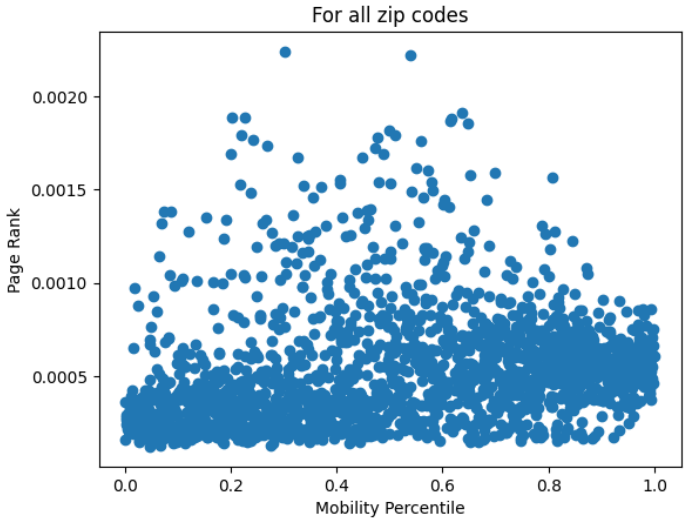


Fig 9 - New York and New Jersey Zip codes' Page Rank score against Mobility Percentile

From the Fig 9, we cannot establish any concrete observation to support the hypothesis of high page rank could correlate with high mobility percentile.

Conclusion

Based on the correlation plot and the Geo Plot, it is difficult to make a concrete conclusion regarding the sole role of Economic Connectedness by testing with zip code connectivity alone. However, we did observe that High SES zip Codes tended to perform well in mobility, especially with high Economic connectedness. In contrast, with Low SES zip codes, it appears to be more random. Also the second hypothesis of a node with higher page rank to correlate with higher social mobility doesn't look to hold true with respect to both the New York, New Jersey and the California data set.

To further investigate the role of Economic Connectedness in mobility, future work can be conducted on an individual user level instead of at the zip code level. Additionally, the analysis could be expanded beyond just two metropolitan states to better represent the trend of the entire nation. By conducting more comprehensive analyses, we can gain a better understanding of how Economic Connectedness affects mobility, which can help inform policy decisions to improve social connectedness and mobility in the United States.

References

1. "The Wealth of Relations. Expanding Opportunity by Strengthening Families, Communities, and Civil Society". SCP Report No. 3-19 (Social Capital Project, 2019).
2. Putnam, R. "Our Kids: The American Dream in Crisis" (Simon and Schuster, 2016).

3. U.S. Census Bureau (2023). *Zipcode Tabulation Areas* – Retrieved from https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2022&layergroup=ZIP+Code+Tabulation+Areas

4. Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenberg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barber´a, Monica Bhole, and Nils Wernerfelt (2022a). "Social Capital I: Measurement and Associations with Economic Mobility." Nature, 608(7921), 108−121.

5. U.S. Census Bureau (2023). Income in the past 12 months – Retrieved from https://data.census.gov/table?q=median+income&g=010XX00US$8600000&tid=ACSST5Y2021.S1901.

6. Bailey, M., Cao, R., Kuchler, T., Stroebel, J., & Wong, A. (2018). Social Connectedness: Measurement, Determinants, and Effects. Journal of Economic Perspectives, 32(3), 259-80. Data retrieved from https://data.humdata.org/dataset/social-connectedness-index?fbclid=IwAR1a2gwDSNQ9_PUhZgn6fzmzgQG9DzlBUihtAIPqO7o2jYSKhnaDksiUgeI

7. Bailey, M., Gupta, A., Hillenbrand, S., Kuchler, T., Richmond, R. J., & Stroebel, J. (2021). International Trade and Social Connectedness. Journal of International Economics, 129, 103418