

Social Connectedness in America

Analysis of Facebook's Social Connectedness Index

Bridger Gordon

12/4/2019

Abstract

Today, Facebook is one of the most popular websites in the world and connects nearly two and a half billion people. It has become the digital representation of people's social circles; the average person's Facebook friends include their family, friends, coworkers, former classmates, and more. This digital representation of people's natural social networks gives a unique opportunity to analyze the way that people interact with other people in society and who those people are, as the composition of your social circle can greatly influence the experiences you have, the ideas you encounter, the opportunities you have, and more. Why are some people so deeply woven into the social fabric as others are marginalized, excluded, and isolated from it? To look at this question, I analyzed Facebook's Social Connectedness Index, a metric of how connected people in a state or county are with other states or counties. Overall, the most socially connected states are western states like Montana, Colorado, and Wyoming and the least connected states are Southern and Rust Belt states like Alabama, Mississippi, Wisconsin, and Michigan. The most important factor in determining social connectedness is geographic closeness, which leads to social bubbles based on the demographic similarities of geographically close states (by race, politics, class, etc).

Social Connectedness in America

Today, Facebook is one of the most popular websites in the world and connects nearly two and a half billion people. It has become the digital representation of people's social circles; the average person's Facebook friends include their family, friends, coworkers, former classmates, and more. This digital representation of people's natural social networks gives a unique opportunity to analyze the way that people interact with other people in society and who those people are, as the composition of your social circle can greatly influence the experiences you have, the ideas you encounter, the opportunities you have, and more.

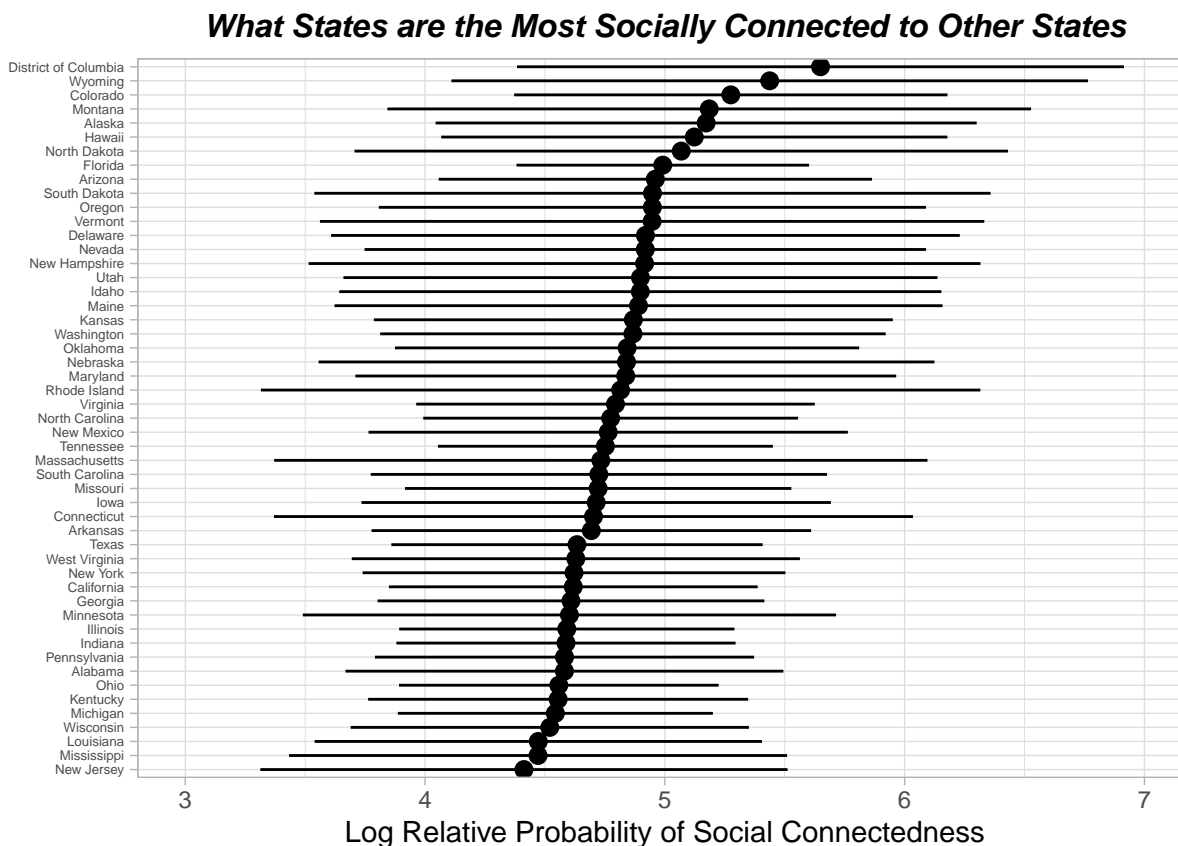
Understanding how social circles form in our society and the relative social connectedness of peoples is important because it helps us understand the lives and divides in the ways life is experienced in our society. ***Why are some people so deeply woven into the social fabric as others are marginalized, excluded, and isolated from it?*** Are differences in social connectedness based in politics? Socioeconomic status? Race? Geography? How do the nodes of the American people connect?

To analyze these networks of social connectivity, it would be extremely helpful to have a metric for the connectedness of people across the US. And for better or for worse, Facebook has been tracking exactly this as they've accumulated the digital representation of our social networks through our Facebook friends. One specific metric, called the Social Connectedness Index, gives the relative probability that a person in County A is socially connected (i.e. Facebook friends) with a person in County B.

In September, I became aware of the Social Connectedness Index from an article in the New York Times about Facebook's data collection. I emailed Facebook's Data Research Department asking if there was any way I could study / research the data for my Gov1005: Data final project. And to my surprise, they responded the next day saying that I'd need to sign a waiver but then would be able to have access to the Social Connectedness Index dataset. The data was initially broken down by county, and I also aggregated it up another level to look at associations by state. The data was originally in CSV and Stata files, and I analyzed the data using R.

So which states are the most socially connected? Well it turns out it might not be the states you expect. In first place is Washington DC. After that the top of the list is overwhelmingly Western states (like Wyoming, Montana, Colorado, and the Dakotas). This means that the average American is (as a whole) relatively more likely to be Facebook friends with someone from these states. Also with high Social Connectedness Index scores: Arizona and Florida. One possible explanation for this is the gradual migration of elderly Americans from their home states to these warmer-weather states. At the bottom of the list are overwhelmingly Rust Belt (Ohio, Michigan, Pennsylvania) and Southern states (Mississippi, Louisiana, Alabama).

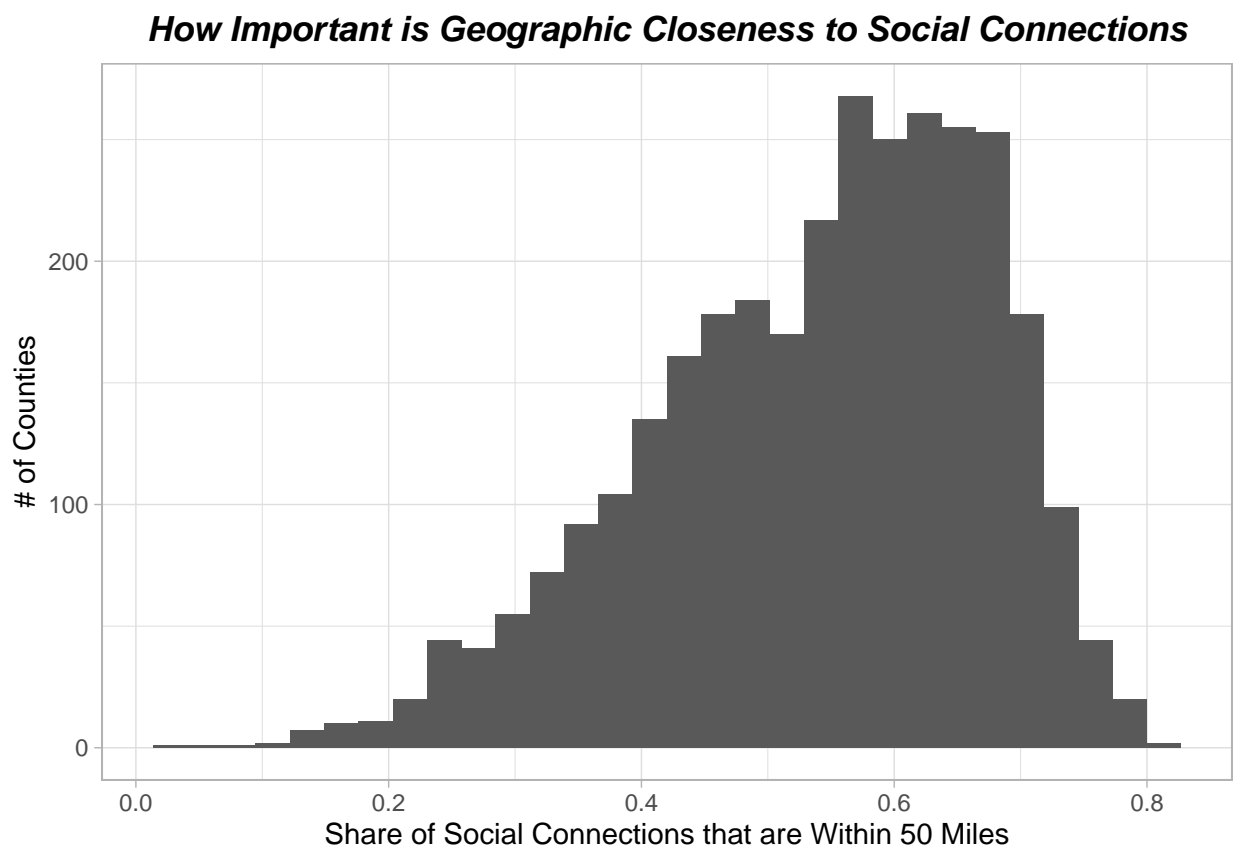
Figure 1: Mean Social Connectedness with 68% confidence interval (\pm one s.dev.)



I combined the Social Connectedness Index with demographic data from the American Economic Association, including racial composition, median income, population, education, and partisan lean. This would give me the opportunity to analyze those questions of how social connectedness and demographic divides are related. We can fit a linear model based on this demographic data to see how well it predicts the social connectedness between two counties. Overall, the demographic data does not explain particularly well the level of social connectedness – meaning that social connectedness is overall much more determined by something other than demographic similarity.

Perhaps one of the most explanatory variables is just the geographic closeness of two places. Counties and states that are near each other are much more likely to be socially connected. With this comes the demographic similarity, so it may be that since counties and states next to each other are demographically similar they tend to be more likely to be connected to demographically similar people but not all people across the nation who are demographically similar. This suggests that social connectedness is similar to the concept of a gravity model in economics (where nations that are near each other are more likely to trade goods). The histogram shows just how important geographic closeness is by showing the average share of connections that are within 50 miles (roughly 60+%).

Figure 2: Histogram of Geographic Effect on SCI



This means that many of the social bubbles that exist may not be due to intentional self segregating but rather because states or counties next to each other tend to be demographically similar. America is not a well-mixed random demographic sample and that leads to people having unrepresentative social circles.

This initial exploration of the data is just a start at looking into the trends of the Social Connectedness Index, and future studies of the data could do more to understand the relationships between how social connections between locations form (besides geography and demographics) and how these social connections impact our lives (are more socially connected counties healthier? wealthier? etc.). I look forward to further analyzing this dataset in future projects.