UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Areas of physics by complexity



Newton's Mechanics    Electro-Magnetism    Special Relativity    Quantum Mechanics General Relativity    Quantum Field Theory    Complexity Science

# Project 43: Social connectedness index I from Facebook

**Poccianti Gabriele**

**Last update**: September 13, 2024

# Contents

# 1 | Social Connectedness Index I from Facebook
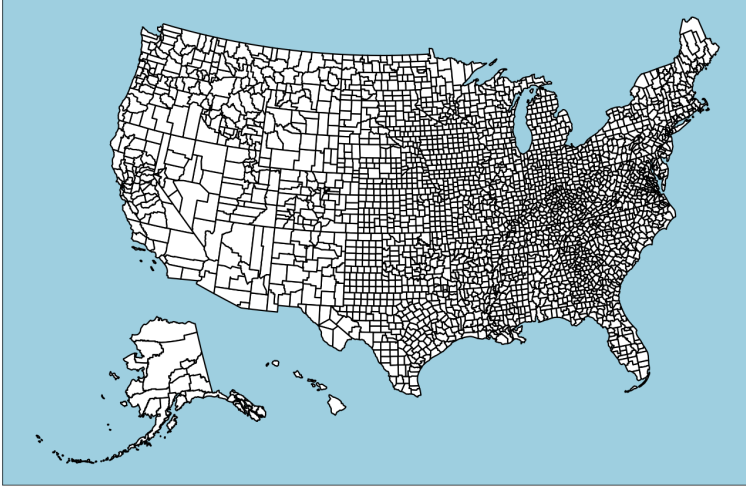
## 1.1 | Task Presentation



Figure 1.1: Map of U.S. counties (Alaska and Hawaii have been displaced for a more compact representation)

The task is to build a network for each U.S. state, where the nodes represent counties, based on a dataset describing social connectivity between different regions [1]. Each edge is characterized by a *Social Connectivity Index (sci)*.

This measure was first proposed and analyzed in [2].

Formally, the *sci* between two locations $i$ and $j$ is defined as:

$$sci_{i,j} = \frac{\text{FB Connections}_{i,j}}{\text{FB Users}_i \times \text{FB Users}_j} \qquad (1.1)$$

where *FB Users*$_i$ and *FB Users*$_j$ are the numbers of Facebook users in locations $i$ and $j$, respectively, and *FB Connections*$_{i,j}$ is the total number of Facebook friendship connections between individuals in the two locations.

To protect individual privacy, the sci has been scaled between 1 and $1,000,000,000$, with a small amount of noise added and the result rounded to the nearest integer.

The sci index can be interpreted as the **relative probability of a Facebook friendship** link between two users located in $i$ and $j$. In other words, if this measure is twice as large, a user in $i$ is twice as likely to be connected with a user in $j$.

The dataset provides edges, characterized by the columns: starting node (**user_loc**), ending node (**fr_loc**), and weight (**scaled_sci**).

We consider the network as simple, ignoring friendships within each county.

## 1.2 | Data Handling

For this analysis, we used the *igraph* package for network manipulation, *ggplot2* for plotting, and *tigris* for retrieving county information. Figure 1.4 was generated using the *usmap* package.

The procedure was as follows: First, we imported a county dataset from *tigris* (from the year 2018 to match the county subdivisions in [2]), modified it (for example, by adding a column for county centroids used later for plotting), and split it by state. Then, we imported the *US counties-US counties sci dataset* from [1] and applied a similar procedure (removing edges between different states). Finally, we generated node and edge files for each state and plotted the corresponding networks.

Here is an example with California:



Figure 1.3: California sci network (node size and color depend on its strength, and edge width is proportional to its weight)
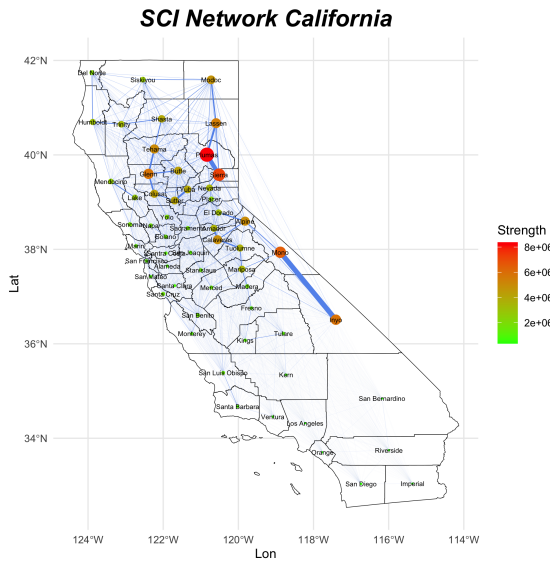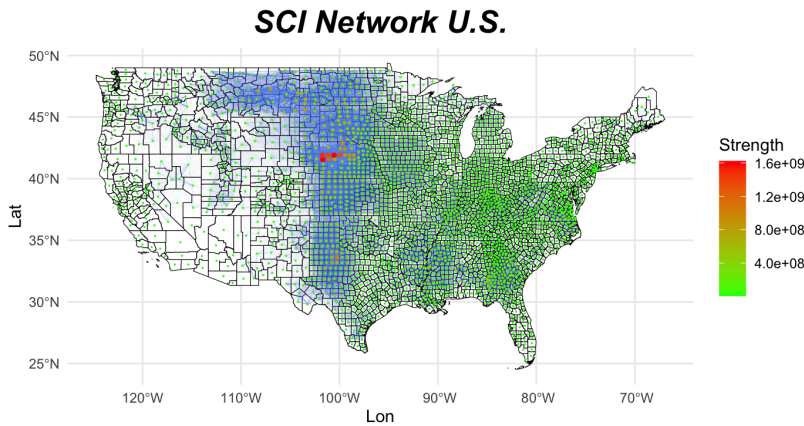
Figure 1.2: California strength heatmap



Figure 1.4: U.S. entire sci network (node size and color depend on its strength, and edge width is proportional to its weight)

## 1.3 | Data Analytics

Regions with only one node were excluded from the analysis. Even with this criterion, network sizes vary significantly, from the *U.S. Virgin Islands* with just 3 nodes, to *Texas* with 254 nodes (the average is around 60 nodes).

Most networks are complete, except for *American Samoa*, *Hawaii*, and the *Northern Mariana Islands*—each with 5 or fewer nodes—making degree analysis uninformative. Since the networks are weighted, we focused on node strength, as shown in figure **??**, and used SCI in the descriptors.

Interestingly, counties with the highest strength often have smaller populations. This may be because intra-county connections were ignored, and people in less populated areas tend to have more long-range friendships compared to those in denser regions.

We also generated histograms to analyze the strength distribution per state and across the U.S. While no clear patterns emerged from the frequency histograms, focusing on states with more than 100 counties (where statistical analysis is more reliable) reveals a consistent shape: an initial peak followed by a sharp decrease in frequency for higher strengths.
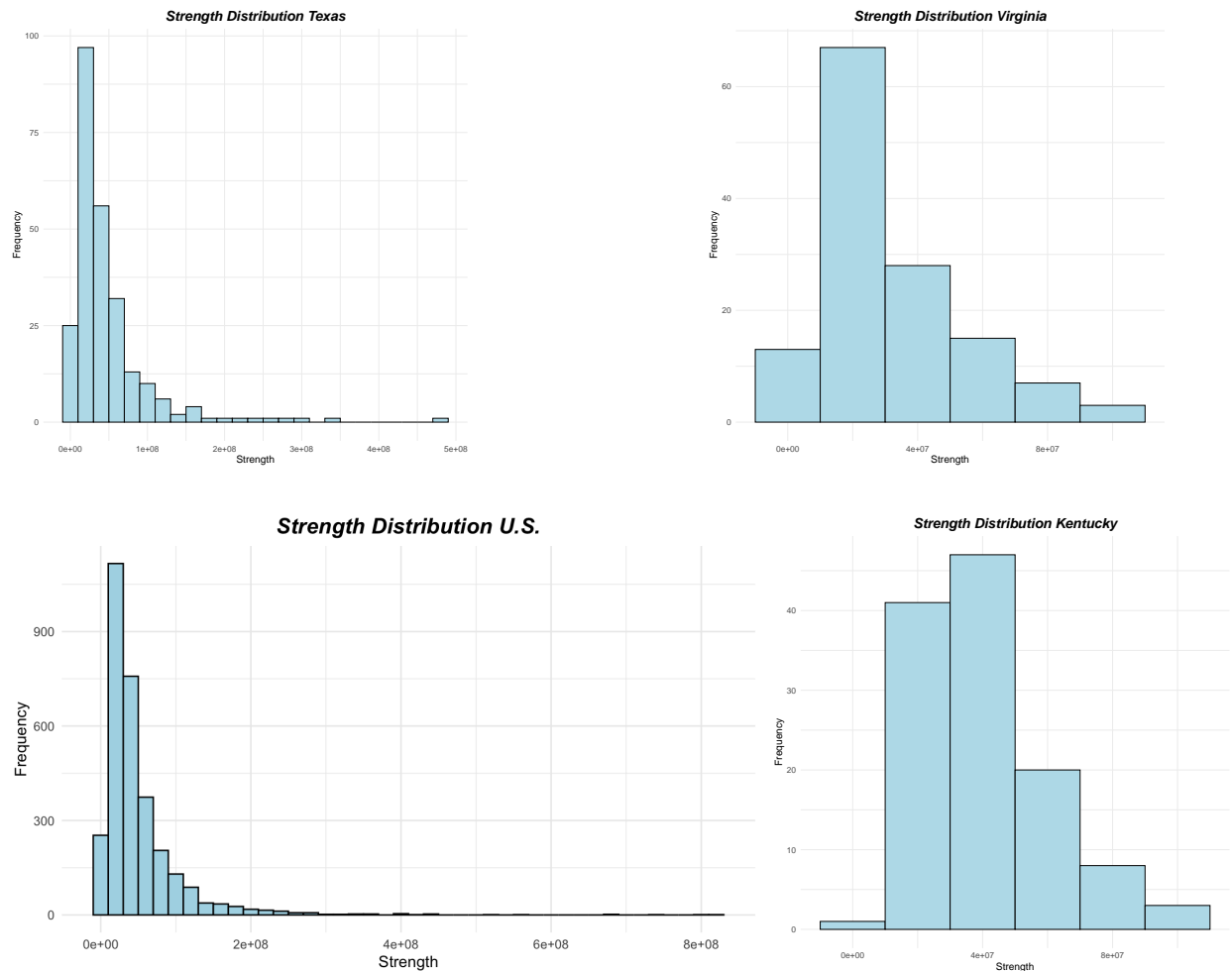
# Remarks

Part of the code was generated with the assistance of ChatGPT (v4), though never used blindly. For example, it suggested using the *tigris* package, helped build functions for degrees-DMS conversions, and provided a base layout for network and state map plotting. Additionally, this tool was helpful in checking grammar and making minor adjustments to improve the readability of this text.

# 2 | Bibliography

[1] Fb social connectedness index. https://data.humdata.org/dataset/social-connectedness-index. [Modified: 13 October 2021].

[2] Michael Bailey, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. Social connectdness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3):259–80, 2018.