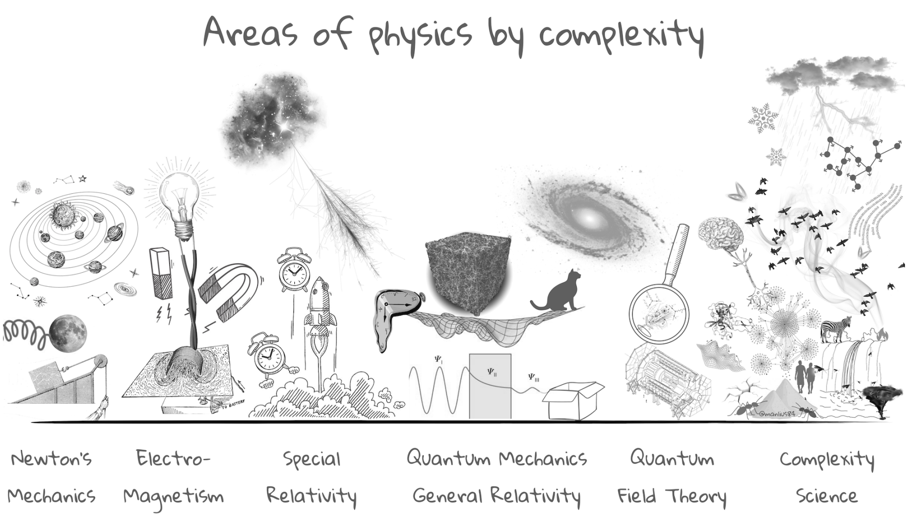
Final Report

Physics of Complex Networks: Structure and Dynamics

Università degli Studi di Padova



Project # 43: Social connectedness index I from Facebook

Bedin Veronica

Last update: July 12, 2025

Contents

1 | Social Connectedness Index

1.1 | Introduction

The Social Connectedness Index (SCI) is a metric used to evaluate the strength of social connections between locations. To calculate this index, we rely on a snapshot of Facebook user networks and their interactions with friends. Locations are assigned to users based on information from their Facebook profiles, including self-reported locations as well as data derived from their device and network connections. The SCI for two locations, i and j, is defined as [1]:

$$SocialConnectednessIndex_{i,j} = \frac{FB\_Connections_{i,j}}{FB\_Users_i * FB\_Users_j}$$

Where:

- FB_Users_i and FB_Users_j represent the number of Facebook users in positions i and j, respectively.
- $FB\_Connections_{i,j}$ denotes the number of connections between positions i and j.

The Social Connectedness Index (SCI) is a measure of the relative probability that a Facebook user in location i is friends with a user in location j. Higher values indicate a stronger likelihood of connections between two locations, reflecting underlying social, economic, and demographic ties.

One key question in analyzing social connectedness is how the strength of these connections varies with geographic distance. Intuitively, friendships tend to be more frequent between closer locations, but the exact nature of this decline can provide insights into regional cohesion, migration patterns, and economic integration. [2]. A useful concept for quantifying this relationship is *elasticity*, defined as:

$$Elasticity = \frac{\%ChangeinFriendshipLinks}{\%ChangeinDistanceElasticity}$$

This captures how quickly social connectedness decays with distance.

Beyond elasticity, I will also explore key network properties that characterize social connectedness. In particular, I will analyze the *strength distribution*, as well as *strength centrality*, which identifies the most socially integrated counties based on their total connection weight. Additionally, I will examine the *modularity* of the network to assess the presence of distinct clusters of highly connected regions. These structural metrics will provide a broader perspective on the patterns of social connectivity, complementing the elasticity analysis by highlighting variations in the strength and organization of social ties.

1.2 Data analysis

During the initial data preparation, the county-level SCI dataset from The Humanitarian Data Exchange [4] was enhanced by adding county and state names as well as latitude and longitude coordinates. These modifications facilitated a more detailed spatial analysis of SCI flows.

To visualize these flows, I generated maps where each county is represented as a node (colored dot) labeled with its FIPS code, and edges connecting counties are scaled by the SCI weight. For instance, in Figure 1.1, the SCI flows in North Dakota are depicted; here, edge thickness is normalized to the highest SCI value to ensure even the weakest connections remain visible.

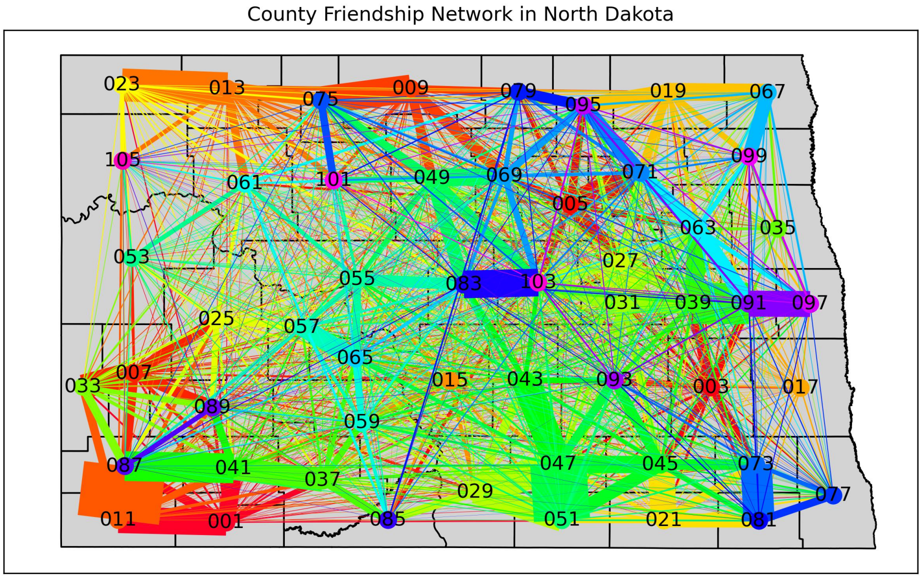


Figure 1.1: County-level SCI flows in North Dakota. Nodes (colored dots) represent counties (labeled with FIPS codes) and edge thickness is proportional to the SCI weight, normalized to the maximum value.

Subsequent analysis examined how both geographical features and migration patterns affect social connectedness. In areas like Philadelphia and surrounding states, counties within the Appalachian Mountains tend to exhibit higher SCI values, indicating stronger local connections. In contrast, in Florida, SCI maps reveal lower local connectivity, likely reflecting migration patterns in which residents maintain ties with more distant northern regions [5].

Table 1.1 compares my elasticity estimates with those reported by Bailey et al. The results indicate that elasticity is more negative for shorter distances—implying a sharper decline in social ties with increasing separation—and becomes less negative as the distance grows. This suggests that while geographic distance is a significant factor in determining friendship links, its impact diminishes over larger scales [2].

Distance Range	My Results	Bailey et al.
Full dataset	-1.63 ± 0.03	–
<200 miles	-1.92 ± 0.07	-1.99
≥ 200 miles	-1.11 ± 0.27	-1.16

Table 1.1: Comparison of elasticity estimates. A more negative elasticity indicates a steeper decline in social connectedness with increasing distance.

Finally, community detection was performed using the Louvain algorithm. When these communities were superimposed on topographical maps, notable spatial patterns emerged. In Montana, for example, the Louvain method identified five distinct modules. Particularly striking were two modules: one encompassing counties north of the Missouri River, and another covering regions within the Rocky Mountains. This spatial differentiation highlights the role of natural features —such as rivers and mountain ranges— in segmenting social connectivity and structuring local networks.

These preliminary findings point toward a promising direction for further research. Incorporating additional layers of data—such as migration trends, trade flows, and socioeconomic indicators—could enrich our understanding of how social connectedness is influenced by both natural and socio-economic factors, as demonstrated in the Bailey paper.

Additionally, the correlation analysis between total population and strength centrality of SCI across different states suggests that population size alone is not a strong determinant of centrality, with correlation coefficients mostly ranging between -0.1 and -0.35. This weak negative relationship implies that other structural factors, such as geographic distribution, connectivity, and economic interactions, may play a more significant role in shaping social ties than sheer population numbers.

Future research could incorporate additional regional characteristics to better capture the complexities of social connectivity and its underlying determinants.

Structure as[1]*:*

- *A short (max 1 page) explanation of the task, including references. Include mathematical concepts.*
- *Max 2 pages for the whole task (including figures)*
- *It is possible to use appendices for supplementary material, at the end of the report. Max 5 pages per task*

A total of 3 pages + 5 supplementary pages per task

[1] Remove this part from the report

2 | Supplementary material

2.1 | SCI index

The Social Connectedness Index (SCI) is based on aggregated, anonymized data from Facebook friendship links as of April 2016. It is built from the connections between Facebook users, reflecting the extent of social ties within and between U.S. counties and foreign countries. As of September 2014, over 58% of U.S. adults and 71% of U.S. internet users were on Facebook. Usage was fairly consistent across income, education, and racial groups but declined with age, from 87% among 18-29 year-olds to 56% for those over 65.

In the U.S., Facebook serves primarily as a platform for real-world friends to interact, with individuals typically connecting only with people they know personally. Friendship links on Facebook require mutual consent, and the total number of friends a person can have is capped at $5,000$. This makes Facebook an effective tool for representing U.S. friendship networks on a large scale. The SCI is constructed by mapping Facebook users to their county and country locations, based on profile data and activity, and measuring the total number of friendship links between these regions. Only friendship links among Facebook users who have interacted with Facebook over the 30 days prior to the April 2016 snapshot are considered. The SCI is normalized to a maximum value of 1,000,000, with Los Angeles County having the highest value due to its dense network of local connections.

2.2 | Data preparation and integration

Initially, I merged two datasets, which I will refer to as *Dataset 1* [3] and *Dataset 2* [7], to create a comprehensive table of county-level data. *Dataset 1* contains the names of the counties and their associated states, while *Dataset 2* provides the positions (latitude and longitude) of each county. After merging the datasets, I manually added the missing values from either of the two to ensure all relevant information was available. These datasets are crucial for ensuring the completeness of the analysis.

Next, I compared the merged dataset with the county_county dataset, which was obtained from *The Humanitarian Data* Exchange site [4]. To further refine the data, I identified and added counties that were missing from the first two datasets.

The final dataset was manipulated to focus on county-level SCI flows. First, I kept only the rows where the origin and destination countries matched, then I excluded same-county SCI flows. Then, I created a graph for each country.

To evaluate the elasticity, I had to introduce another dataset, called *Population*, obtained from the *American Community Survey (ACS)*, with population data of American counties from 2017 to 2021 [8]. Some countries were missing from this dataset but they were composed of $2-3$ nodes so they were not included in the analysis.

2.3 SCI considerations

The topographical map of Pennsylvania 2.1 alongside the corresponding Social Connectedness Index (SCI) network 2.2 illustrate the high SCI values observed in counties within the Appalachian Mountains.

Figure 2.1: Topographical Map of Pennsylvania

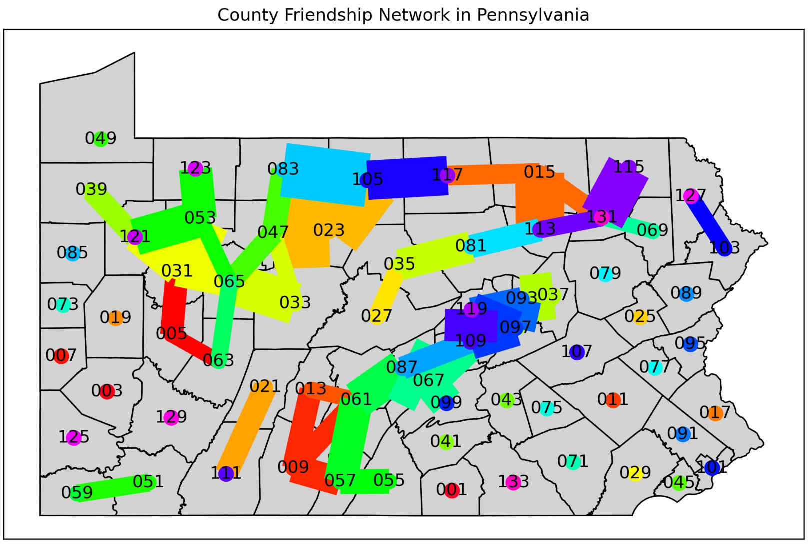


Figure 2.2: Pennsylvania SCI map where edges with SCI smaller than 20% of the maximum SCI value in the country have been removed, to enhance clarity

In 2.3, on the other hand, the SCI map of Florida where migration patterns caused strong connections to non-adjacent northern regions but low SCI values inside the country.

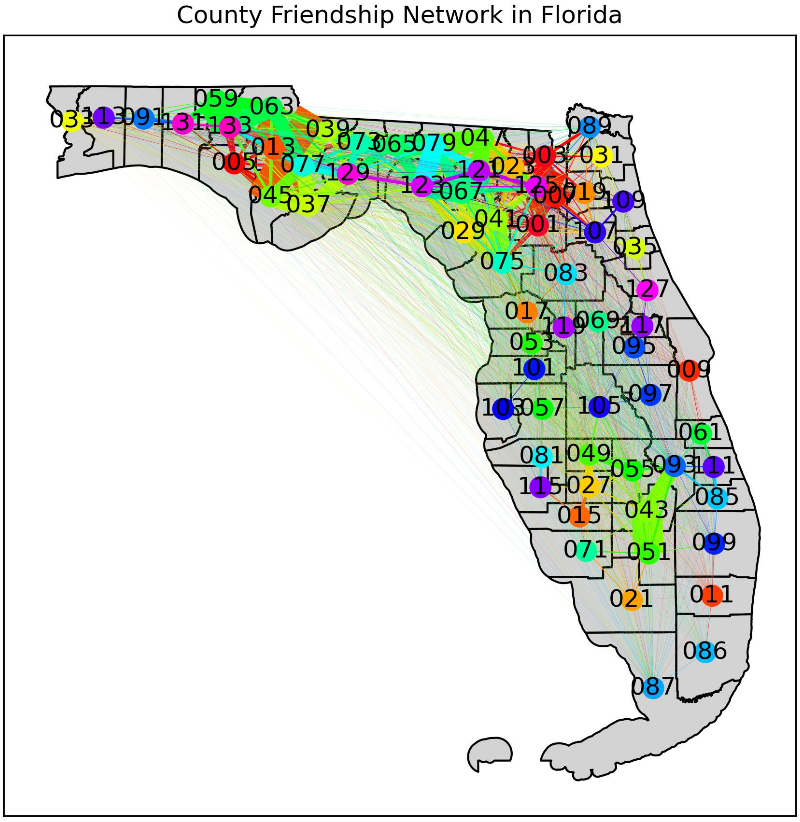


Figure 2.3: Florida SCI map

2.4 Elasticity

Table 2.1 shows the regression coefficients for states with more than 50 counties, ensuring a more reliable fit for the elasticity model. The coefficients are reported for three models: β_1^{all}, β_1^{lt} (short distances, less than 200 *miles*), and β_1^{ge} (long distances, more than 200 *miles*).

The results align with Bailey et al.'s findings, showing that elasticity is more negative for short distances, indicating a steeper decline in social connectedness as distance increases. For example, Alabama shows a sharp decrease in short-distance elasticity (-2.57), which then flattens at longer distances (-1.19).

The variation across states may reflect differences in migration, local culture, or infrastructure. In Florida, for instance, the elasticity for short distances is quite negative (-2.47), which could suggest that local ties are weakened by long-distance connections, as already mentioned. This idea could be further explored by examining migration patterns in more detail.

State	β_1^{all}	β_1^{lt}	β_1^{ge}
Alabama	-2.15 ± 0.02	-2.573 ± 0.03	-1.19 ± 0.06
Arkansas	-2.03 ± 0.01	-2.31 ± 0.03	-1.59 ± 0.05
California	-1.17 ± 0.02	-1.59 ± 0.07	-1.12 ± 0.03
Colorado	-1.29 ± 0.03	-1.74 ± 0.07	-0.70 ± 0.06
Florida	-1.46 ± 0.02	-2.47 ± 0.05	-0.93 ± 0.03
Georgia	-1.89 ± 0.01	-2.26 ± 0.02	-1.34 ± 0.02
Illinois	-2.22 ± 0.01	-2.64 ± 0.02	-1.75 ± 0.03
Indiana	-2.05 ± 0.01	-2.37 ± 0.01	-1.65 ± 0.04
Iowa	-1.99 ± 0.01	-2.25 ± 0.01	-1.91 ± 0.03
Kansas	-1.61 ± 0.01	-1.98 ± 0.04	-1.40 ± 0.03
Kentucky	-2.01 ± 0.01	-2.35 ± 0.01	-1.40 ± 0.02
Louisiana	-1.96 ± 0.02	-2.22 ± 0.03	-1.47 ± 0.04
Michigan	-1.30 ± 0.01	-1.98 ± 0.03	-0.85 ± 0.03
Minnesota	-1.47 ± 0.01	-1.92 ± 0.03	-1.05 ± 0.03
Mississippi	-2.09 ± 0.01	-2.38 ± 0.02	-1.80 ± 0.04
Missouri	-2.14 ± 0.01	-2.50 ± 0.02	-1.87 ± 0.02
Montana	-1.33 ± 0.02	-1.6 ± 0.1	-1.18 ± 0.04
Nebraska	-1.45 ± 0.02	-1.71 ± 0.04	-1.51 ± 0.04
New York	-1.90 ± 0.10	-1.77 ± 0.04	-2.13 ± 0.07
North Carolina	-1.65 ± 0.05	-2.26 ± 0.02	-1.05 ± 0.02
North Dakota	-1.67 ± 0.10	-2.09 ± 0.04	-1.38 ± 0.05
Ohio	-2.17 ± 0.05	-2.41 ± 0.02	-1.80 ± 0.04
Oklahoma	-1.57 ± 0.08	-2.09 ± 0.03	-0.90 ± 0.03
Pennsylvania	-2.14 ± 0.07	-2.25 ± 0.03	-2.10 ± 0.04
South Dakota	-1.5 ± 0.1	-1.68 ± 0.04	-1.41 ± 0.05
Tennessee	-1.95 ± 0.05	-2.48 ± 0.02	-1.40 ± 0.02
Texas	-1.74 ± 0.04	-2.17 ± 0.03	-1.62 ± 0.01
Virginia	-1.73 ± 0.04	-1.94 ± 0.02	-1.40 ± 0.02
West Virginia	-2.11 ± 0.07	-2.22 ± 0.02	-2.22 ± 0.09
Wisconsin	-1.66 ± 0.07	-1.97 ± 0.03	-1.28 ± 0.04

Table 2.1: Regression Coefficients for States with more than 50 counties

2.5 Modularity

Building on our previous discussion, intriguing spatial patterns emerge when examining community modules in relation to geographic features. For example, in Montana 2.4 the map reveals distinct clusters: green dots indicate counties associated with the Rocky Mountain regions, while blue dots represent counties in the northern part of the river.

A similar pattern is observed in other states along the Rocky Mountains, such as Colorado, as well as in the Appalachian region, including Georgia, suggesting that natural geographic barriers play a key role in shaping social connectivity.

Beyond geography, modularity patterns may also reflect urban-rural divisions, economic dependencies, and historical migration trends, which could be explored in

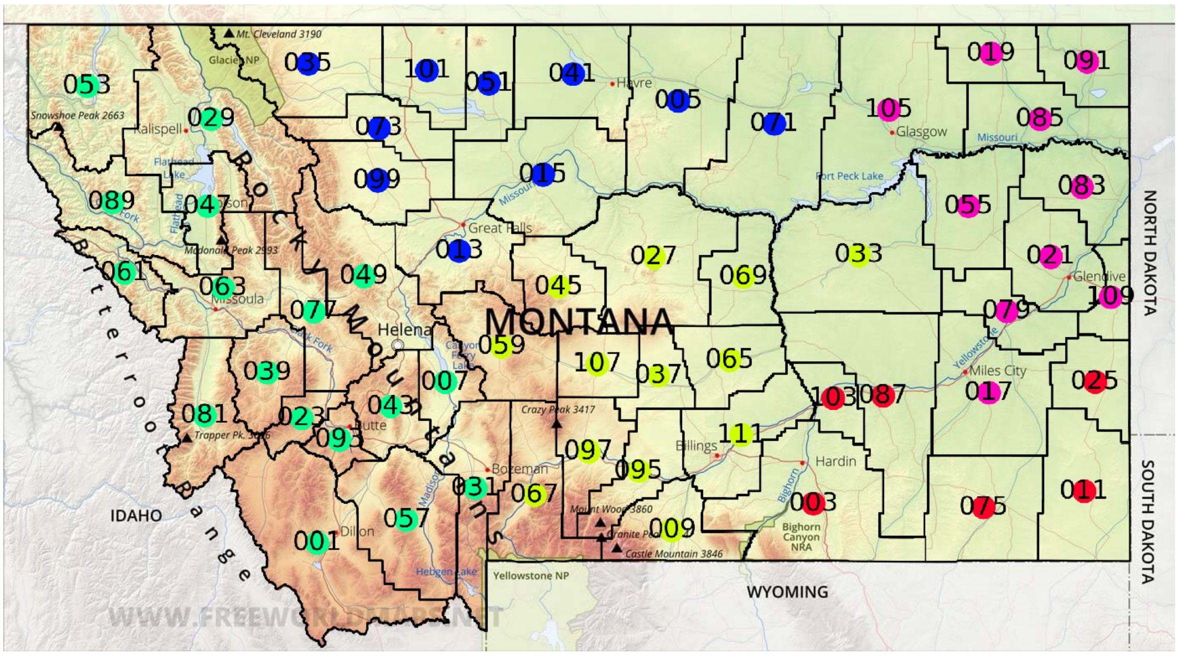


Figure 2.4: Topographical map of Montana and modules found via Louvain algorithm

future research.

2.6 | Use of ChatGPT

In this project, ChatGPT [6] primarily assisted with the superimposition plot of the network and U.S. counties, including the automatic generation of the plot and the implementation of saving functions for the obtained results. It also contributed to locating some datasets, such as U.S. FIPS codes and population data, though further evaluation was required on my part to identify the most complete and reliable sources. Additionally, ChatGPT played a role in the final revision of the review, refining my initial draft to enhance clarity, professionalism, and appropriateness of language.

3 | Bibliography

[1] Social Connectedness Index — dataforgood.facebook.com. https://dataforgood.facebook.com/dfg/docs/methodology-social-connectedness-index. [Accessed 13-Feb-2025].

[2] R. Bailey, T. Cao, J. Kuchler, J. Stroebel, and A. Wong. Social connectedness: Measurements, determinants, and effects. *Journal of Economic Perspectives*, 32(3): 259–280, 2018.

[3] Chuck Connell. Fips2county dataset. https://github.com/ChuckConnell/articles/blob/master/fips2county.tsv. [Accessed 1 February 2025].

[4] Humanitarian Data Exchange. Social connectedness index - county dataset. https://data.humdata.org/dataset/social-connectedness-index#. [Accessed 1 February 2025].

[5] Shruthy Nair and Clio Andris. The relative probability of facebook friendship in the united states. *EPB: Urban Analytics and City Science*, 2024. doi: 10.1177/23998083241272094. URL https://journals.sagepub.com/doi/10.1177/23998083241272094.

[6] OpenAI. Chatgpt (feb 15, 2025 version), 2025. URL https://openai.com/chatgpt. Accessed: 2025-02-15.

[7] Russell Samora. Gist: 12be4f9f574e92413ea3f92ce1bc58e6. https://gist.github.com/russellsamora/12be4f9f574e92413ea3f92ce1bc58e6. [Accessed 1 February 2025].

[8] U.S. Census Bureau. Population estimates - acs 2017-2021 (counties), 2021. URL https://mce-data-uscensus.hub.arcgis.com/datasets/USCensus::population-acs-2017-2021-counties/about. Accessed: February 13, 2025.