Redrawing the Map of the United States from Commuter Data

DAVID OGUTU

Massachusetts Institute of Technology, Cambridge

ogutu@mit.edu

Abstract

This paper offers an alternative way of mapping the United States based on commuter data. I argue that this method is a better indicator of interaction than two SMS and call-based maps made by Sensable labs at MIT because it covers all of the United States mainland and is not biased by carrier (or mode of transportation for our case).

Figure 1: US Map redrawn from Commuter Data

I. INTRODUCTION

Mainland United States is made up of 48 contiguous states whose boundaries were based mainly on political lines at the time of division. A lot has also changed from the time the borders were drawn and it is unclear whether they are still representative of how US citizens are split in this age. This paper seeks to redefine state borders with respect to how people interact across space. Commuter data based on inter- and intra-county trips was used to form regional clusters across the United States. Given a geographical area and some measure of the strength of links between its inhabitants, the area is partitioned into smaller, non-overlapping regions. The most surprising detail was that some of the states have very strong inter-county dealings amongst themselves and almost none with neighboring counties across the state border.
The economic consequences of inference from such a map are enormous. Businesses can use the insights to know where to strategically position their regional offices across the country. Some states clearly have multiple non-defined clusters while clusters merged into one across a few states in our analysis – a good indicator of inter-state trade. The social inferences are also insightful. The people of the state of Colorado, for instance, associate more with Texas and almost never with any of their neighboring states. Although the two states do not border each other, travel occurs across one county in Oklahoma into the two states. If the state borders were to be redefined, both states would merge into one. Such interactions tend to define cultural identities and may greatly influence future political ideologies and affiliations due to increased interaction.

There are, surprisingly, a lot of clusters that defy spatial restrictions. These regions engage in a lot of air traffic and residents associate more with each other than with neighboring states perhaps due to a similar interest in commerce. Florida, Washington and Oregon are a good example. They are in the same cluster but Florida is as far away as one can get from both Washington and Oregon.

II. Data and Findings

Commuter data collected from all counties in the United States was used to make the map. Nodes were defined as different counties with assigned geo-locations and edges were the trips between counties. No regard was given to the mode of transportation. Intra-county trips were filtered out as they were found to be roughly proportional to the population size and told us nothing about connectivity between counties.

The distribution of degrees was found to obey a power law and was fitted to $P_k = 192.39(k^{-2.3})$. This was then compared to a scale free network and results were closely co-related but slightly offset. This suggests that if the United States was unbounded and people were still allowed to travel freely, the degrees will continue to grow following a power law $P(k) \sim k^{-\gamma}$. In light of the fact that some states are landlocked and that the diversity of business specializations across counties and states, these results are very shocking. It doesn’t really matter whether you live in the country or city.

The degree of nearest neighbors did not offer a lot of insight beside the fact that both the data points and the scale free network both cluster around the $10^2$ range for both degree and the degree of nearest neighbors. The $k_{nn}$ values also become more focused as $k$ increases. This differs slightly from a strictly scale free distribution where the the number of nearest neighbors increases with increasing degree. Spatial data may offer a better insight in this regard and may be a subject for further investigation.
Commuter data is a good indicator of physical connections across counties. A map would be a good indicator of social interactions and economic activity as most people travel either to see their family, friends or business partners. I sought to cluster regions with a lot of commuter interaction.

The map was made as follows: I started the analysis with commuter data across all the counties of the United States combining all modes of transportation and ran a randomized clustering algorithm that used edge weights and had a resolution of 1 to come up with 29 modularity classes. The county location co-ordinates were then laid out on a Mercator map and colored according to their respective clusters. A political map of the United States was overlaid on the map to make it more insightful.

A question naturally arises: Was the grouping done correctly? What is the best way to group these pixels into larger regions? A similar question has been the focus of network research over the past decade where one seeks the best way to partition a network into separate, non-overlapping communities and the leading approach is based on optimizing the network’s “modularity” – the method employed in this paper. High modularity values occur when the network is subdivided such that there are many links within communities and few between them, as compared to a randomly generated network with otherwise similar characteristics. (Ratti, 2010)

I was, however, not the first to do this. Researchers, especially at the MIT Sensable lab have sought to draw maps of the United States and the United Kingdom based solely on social connection. They used text messaging data to divide the country into modular classes and come up with a more social map. The problem with the map they drew for the United States was that it did not cover the whole country (perhaps due to technical or legal reasons). Skeptics have also cited that the data is skewed toward city residents as they are more likely to use AT&T compared to country dwellers. (Sensible labs, CSA website)
This new map based on commuter data overcomes some of the challenges that were faced by the teams at the Sensable labs. Data is not biased as all modes of transportation were considered. Although different modes of transport were considered, preference was not given to one over another. Also, in human interactions, calling normally takes more involvement of both parties than texting but physically meeting trumps them both. It can be viewed as a stronger indicator of engagement or connection among persons. I believe that this map will not only help economists and sociologists make better models but will also help the public to better understand the fabric of relationships and deep seated social biases in regular interaction and might help explain communal biases, stereotypes, rivalries and affiliations. CEOs can also have a standard for choosing branch locating across the country. Obviously this last suggestion is to be weighed against many other company goals and customer locations.

Figure 6: Comparison of Commuter and Sensable lab’s SMS maps

REFERENCES

- Article Source: Redrawing the Map of Great Britain from a Network of Human Interactions
