Narrative

Sonja Castañeda

11/24/2019

Brief substantive background / goal

This project began with the aim of exploring the commonly held belief that pan-indigenous cooperation that supports the revitalization of dead and dying indigenous languages in the United States was largely made possible by oppressive federal and state English-only legislation that accounted for the decline of these languages in the first place. However, the lack of reliable/accessible data on language status and policy implementation at the tribal level revealed this to be a difficult undertaking for a term-length project. I therefore narrowed my substantive focus, but I maintained the goal of mapping key indigenous social networks in the United States in the late "Termination" period (late 1950s and early 1960s), with particular attention to the American Indian Chicago Conference, which I suggest should yet allow me to critically evaluate cultural "tipping point" theories that pertain to coalition building among Native Americans.

Collecting data

In June of 1961, approximately 700 indigenous people, representing sixty-four tribes in the United State, met at the University of Chicago, to formulate a joint political agenda and to set forth a declaration of shared principles. I am in the process of building a database that contains information related to pan-indigenous network building before, during, and after the 1961 Conference. I focus on primary and secondary archival sources, for example conference registration and committee lists, organization-related correspondence, personal letters, newsletters and reports, as well as online tribal records, obituaries, and newspapers to locate data on interactions between key figures in the emerging pan-indigenous movement. I am also gathering information on the contexts from which these individuals emerged in terms of their tribal and reservation status, their kin, their education, and their individual and tribal relationships with state and Federal government structures.

Cleaning / pre-processing data

I encountered problems in R when trying to read warped images of text , using photos from archives. I first tried converting images to pdf, and tried to use simple route to converting pdf to text: library(pdftools) trial_pdf_txt <- pdf_text("Sol.pdf") trial_pdf_tx • This didn't return an error, but it also didn't return text or anything else, just [1] "" One route to reading these files could be using Adobe fixes, but it seemed like discussions online pointed to better solutions (perhaps because some people who chat about data online prefer obscure solutions to popular ones?). Adobe does provide some trouble-shooting steps that could help. For example, one can "embed fonts" to help Adobe more easily recognize text. I considered this, but my documents used several different fonts, most of which likely came from 1920s to 1950s typewriters with some from printing presses. Olde tyme type all kind of looks the same when I call it up from memory, but in actuality it varied quite a bit. Adding to the difficulty of loading fonts to aid in extracting text from my warped documents was the fact that, as far as I could learn, fonts weren't generally named prior to the mid 1950s, when one of the larger manufacturers introduced switchable type on a home typewriter, so I would have to try to find similar fonts to the ones in my documents and hope these would help. Also, a font can be embedded (in Adobe) only if it contains a setting by the font vendor that permits it to be embedded. I decided to try another, hopefully more efficient, routes first. So on this project I got to learn all sorts of cool stuff I never meant to pick up.

Adobe references https://helpx.adobe.com/acrobat/kb/missing-or-garbled-text-converting.html https://helpx.adobe.com/acrobat/using/pdf-fonts.html

I next tried the Tesseract OCR engine, including deskewing my document photos. I found some improvement in the output, but it still wasn't nearly accurate enough to prefer it to hand-entering the data for the size of my database (even though my pain database is 457 rows long, and getting quite wide)

I tried a few ways to consolidate duplicated edges in both directions, but I didn't find a solution yet (I could get it to sum edges, then I could ignore edges, but I found this a tricky proposition). I will return to this shortly a workable solution surely exists, and it will be important for efficiency in future work.

try 1 committee_links_Dupless <- edge.duplicates(committee_links_NoDup)

try 2 df_dupless <- edges df_dupless

try 3 dfnew <- data.frame(matrix(unlist(committee_links_Dupless)),stringsAsFactors=FALSE)

try 4 install.packages("linkcomm") library(linkcomm) Eliminate duplicate edges and create new dataframe committee_links_NoDup <- committee_links

I will return to this soon, just had to move on for the moment. In the meantime, I cleaned up and assembles edges via csv.

Analysis and visualization

The "story" that I associate with my data emerged gradually over the course of data collection, then more suddenly when reading through correspondences to/from a key conference organizer (a non-indigenous anthropology professor at the University of Chicago). In short, I find a number of subtle-at-first cleavages that divide committee members (and cliques) and appear to result in a minority position being featured prominently in the "Statement of Shared Sovereignty." In particular, a few key committee members who were very young (early twenties, at oldest) appear to drive the consensus toward more inclusive definitions of "Indian," such that those who had lost federal recognition, rejected it, or never achieved it, are brought into the movement on equal ground to "Reservation Indians" and "Tribed Indians." I believe these individuals (and especially one, Clyde Warrior) were partly motivated in this direction by their own social/political/tribal position, but I also believe they were impressed by a very vocal "Eastern Indian" (from the East of the United States), William Rickard, who refused to participate as a committee member, but was very active and antagonistic to what he saw as "colonial sympathies" among the older committee "tribed" members. Centrality measures support this story, but much work remains to be done to draw any causal conclusions.

Visualizations

I created a couple of maps in preparation for creating network maps, which will be one of the next steps in my project. I think others have done more elaborate mapping projects, so I will only mention that I found it useful to use a less busy map for my purposes, which are mostly to show that my sample (or those who attended the conference) were somewhat geographically disbursed. I can show networks on this map, but because even my smaller network is quite large (450+ rows), and because individuals and committees are disbursed (something I can also calculate from my data), they just crisscross everywhere and don't provide much information other than letting us know that there is no highly obvious geographic clustering. I encountered one issue I couldn't solve:

#combine us_map with chicago_data "address" (lon, lat) and color nodes to indicate committee status (e.g. drafting committee) ggmap(us_map)+ geom_point(aes(lon, lat, color = STEERING_COMM+ DRAFTING_COMM+ CHAIRMEN+ RULES_COMM+ CREDENTIAL_COM+ NYIM), data = chicago_data)

This didn't return an Error, but the output is just some of the code, not a map. It mapped fine with 2 vectors, but not more. I tried to trouble shoot this but did not find a solution. It's not a crucial map at this point in the project, as I can calculate distance and other things to confirm diversity of committees, so I moved on for the moment.

Networks/igraph

I found the igraph tutorial in datacamp useful for learning to use an already constructed graph object, but the tutorial I used did not cover constructing objects. This is simple if all you want to do is analyze an n-by-n matrix (a square matrix). Adding a single attribute to this, for example a weight, is also very easy, since you can assign a number to the intersection of two vertices. But if you have a wide matrix, like mine, and/or one that includes several committees, finding instructions on how to create the graph object(s), and executing these instructions on for your particular dataframe(s), can be much more complicated. I found the R igraph manual somewhat helpful with this, but, as with most of the references to graph object information I could find easily, it was most useful for very small networks, and all of the examples centered around constructing a tiny dataset from scratch and operating on that.

My biggest regret, in terms of making my way through datacamp, is that my data wasn't ready when I was doing so, and so I wasn't trying out my own code alongside doing the exercises that involved igraph's data. I think this is especially important if you are working with a fairly large dataset, because the igraph tutorial tends to use very small data. It seems like this is often the case in other examples throughout the R world: the data is entered as part of the example. This often works fine in general, but when you have a large network, the visualization is often vastly different than with small data, and things like organization and specifying parameters can become much more important. I would suggest having 4 types of data on-hand when you start the igraph tutorial, datasets with which you are familiar. Two small datasets Two large datasets, one of each undirected and directed. Undirected means the edges/links between the nodes/vertices do not run in one direction. a -> b or a<-b vs a<->b, which is just written as a-b If you have these four sets, I think you will be in great shape if you work through the data for each at each stage. That might sound terrible, and it might be slow at first, but even if it doubles your time working through igraph, you will have an incredibly useful resource if you plan to visualize networks using R in the future.

Sources: igraph manual https://igraph.org/r/doc/graph_from_data_frame.html Author: Gabor Csardi csardi.gabor@gmail.com

For the size of my dataset, the number of vertices, attributes, and "relations" (in my case, whether or not 73 individuals sat on any of 6 committees together) a lot of the igraph examples I found were impractical, so in the absence of examples that used a wide dataframe with many vertices, I needed to find a series of shortcuts to replicating the steps used in constructing small graph objects. Other resources on the web were helpful with this part: For example, a particular stack overflow thread & its responses were helpful in creating my edgelist:

https://stackoverflow.com/questions/34670145/generating-an-edge-list-from-id-and-grouping-vectors

This website/course was also helpful: Network Analysis and Visualization with R and igraph Katherine Ognyanova, www.kateto.net NetSciX 2016

School of Code Workshop, Wroclaw, Poland https://kateto.net/netscix2016.html I found this especially useful once I got my object constructed as it provides a lot of information on identifying cliques, subgroups, etc.

Future work

My network visualizations remain quite messy. I look forward to improving my skill in assessing how to build and display networks in ways that best fit the data. I hope to continue with this project and to expand my use of legal documents to evaluate the story. Doing so would be much easier if I were able to scan in documents easily and use regex to manage them. I will continue to look for better ways of doing