

## WHY THE WARREN COURT?

The Supreme Court acts as the highest court in all the United States. It acts as the supreme power for the judicial branch of government. As such, it has immense power to shape our country and arguably other countries. Due to the lifetime appointments of the justices of the Supreme Court, the court changes slowly and is separated into eras based on who was Chief Justice. The Warren Court (1953-1969) was arguably one of the most significant courts in the history of the Supreme Court. It made a series of landmark decisions. They decided in *Brown v. Board of Education* (1954), which reversed precedent from *Plessy v. Ferguson* (1896), declared segregation of schools unconstitutional. In addition, the Warren Court incorporated much of the Bill of Rights to ensure that individual states could not violate those rights. In *Engel v. Vitale* (1962), they halted mandatory Bible readings in public schools. In *Mapp v. Ohio* (1941), the court asserted that evidence obtained illegally could not be used in criminal courts. In *Gideon v. Wainwright* (1963), the court established that states were required to provide attorneys. *Miranda v. Arizona* (1966) gave us the famous Miranda rights in which suspects must be made aware of their rights. *Griswold v. Connecticut* (1965) established the right to privacy as something to be constitutionally protected which would later be precedent cited in *Roe v. Wade* (1973).

## THE DATA SET

The data consists of the texts of the Supreme Court opinions and related documents from the year 1953 to 1969. (Language: Legal English)

## OBTAINING THE DATA

Source:

FindLaw provides numerous resources regarding law and court cases. They keep a record of all Supreme Court opinions going back to 1760. Its organization and ease of navigation made it an excellent candidate for obtaining the texts of opinions and related documents.

Tools Used:

### Python

The programming language used for all fetching and processing of documents for this project was Python 3.8 in an Anaconda Environment.

### Jupyter Notebooks

For organizing, coding, and some visualization, Jupyter Notebooks was used.

### Beautiful Soup

Beautiful Soup 4 was the package used to parse the web pages and scrape the necessary data.

### Regex

To clean and unify the data, regex was used.

### Pandas

Pandas dataframes were used for storage and conversion between files and Python data structures.

### Google Drive

After the data was processed, Google Drive was used for storage and sharing. Colaboratory was used for further processing.

# The Warren Court

Jack Stehn - 2020/03/30

Language Models, Text Analysis - Adam Anderson

## WORKS CITED

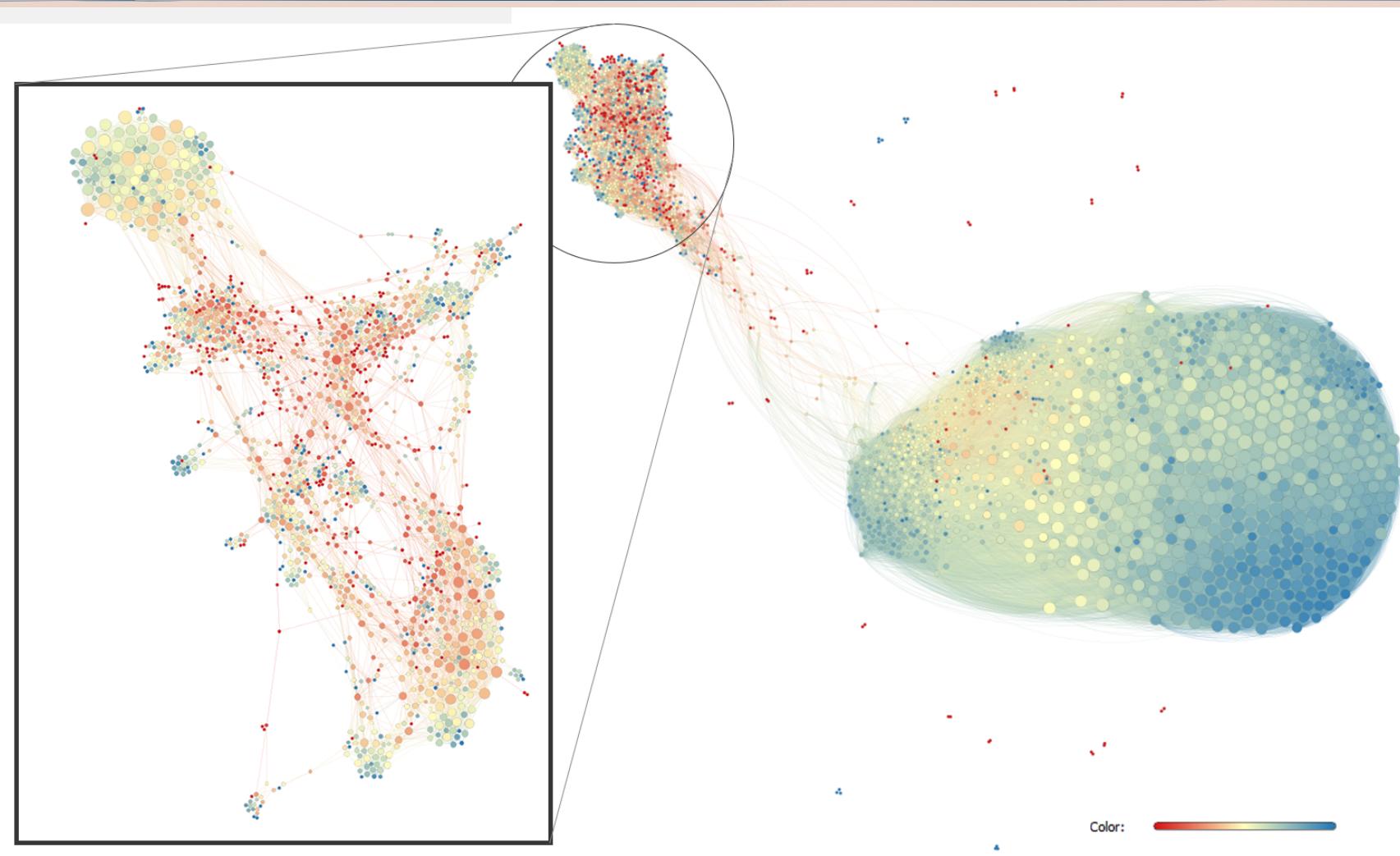
- B. S., Texas A&M University, and Facebook. "The Warren Court: Its Impact and Importance." ThoughtCo. Accessed March 27, 2020. <https://www.thoughtco.com/the-warren-court-4706521>.  
"Beautiful Soup Documentation – Beautiful Soup 4.4.0 Documentation." Accessed March 27, 2020. <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>.  
Findlaw. "FindLaw's United States Supreme Court Cases and Opinions." Accessed March 27, 2020. <https://caselaw.findlaw.com/court/us-supreme-court>.  
Ganesan, Kavita. "Gensim Word2Vec Tutorial - Full Working Example." Kavita Ganesan (blog). January 30, 2018. <https://kavita-ganesan.com/gensim-word2vec-tutorial-starter-code/>.  
Gephi. "The Open Graph Viz Platform." Accessed March 27, 2020. <https://gephi.org/>.  
Kenan Jiang. "Doc2Vec Jupyter Notebook." Accessed March 27, 2020. <https://color.research.google.com/drive/1nCtJQuMKf6UWq75agUZ01Fq0A9iH0FF>.  
"Natural Language Toolkit – NLTK 3.5.1 Documentation." Accessed March 27, 2020. <https://www.nltk.org/>.  
pythonspot. "NLTK - Stemming." Accessed March 27, 2020. <https://pythonspot.com/nltk-stemming/>.  
"Pandas Documentation – Pandas 0.3 Documentation." Accessed March 27, 2020. <https://pandas.pydata.org/pandas-docs/stable/>.

## AREAS FOR ANALYSIS

### SIMILARITY AND CENTRALITY

The Supreme Court chooses the cases that it will take. With such a historic court, I wondered if there are some cases that act centrally for the duration of the court.

Which cases are most central in the dataset? Which ones don't compare to the corpus?



3054 NODES / 173,675 EDGES NODE COLOR = CLUSTERING COEFFICIENT NODE SIZE = DEGREE

Each pair of documents is assigned a cosine similarity score as produced by doc2vec. In short, this was a score from 0 to 1 that judged how similar two documents are. That similarity score, if over .9, then is saved as the weight of an edge between those two documents as nodes in an undirected edge graph. This data was generated by a Jupyter Notebook written by Kenan Jiang. Graph produced in Gephi. What we can see here is that the cases are largely like each other with one distinct massive hub and a smaller hub on the outside. While it may not look it, the two hubs have the same number of nodes. The larger hub is connected much more heavily. When looking at a sample of that massive hub, I noted that every case seemed to have very little text. Either the case was not accepted by the Supreme Court and dismissed, or judgments vacated. In the smaller hub, the connections are less organized, but 2 sub-hubs appear. The top left appears to be mostly lawsuits against corporations while the bottom right appears to largely be related to criminal rights. The remaining cases appear to have lower degree. There seems to be much smaller clusters of 2 or 3 cases separate from the two large groups. In reading them, there seemed to be no discernible pattern (as to be expected with them being not largely disconnected from others). Interestingly, no documents seemed to be particularly influential or central in this corpus.

## TOPICS

Each case that the Supreme Court takes up addresses important questions related to some (often fundamental) aspect of our country. Some topics come up over and over again. Some amendments or sections of the constitution are hardly ever considered at all.

Which issues are brought up most often? Which part of the constitution is most debated? Which one is least debated? Which parts have almost no affect at all?

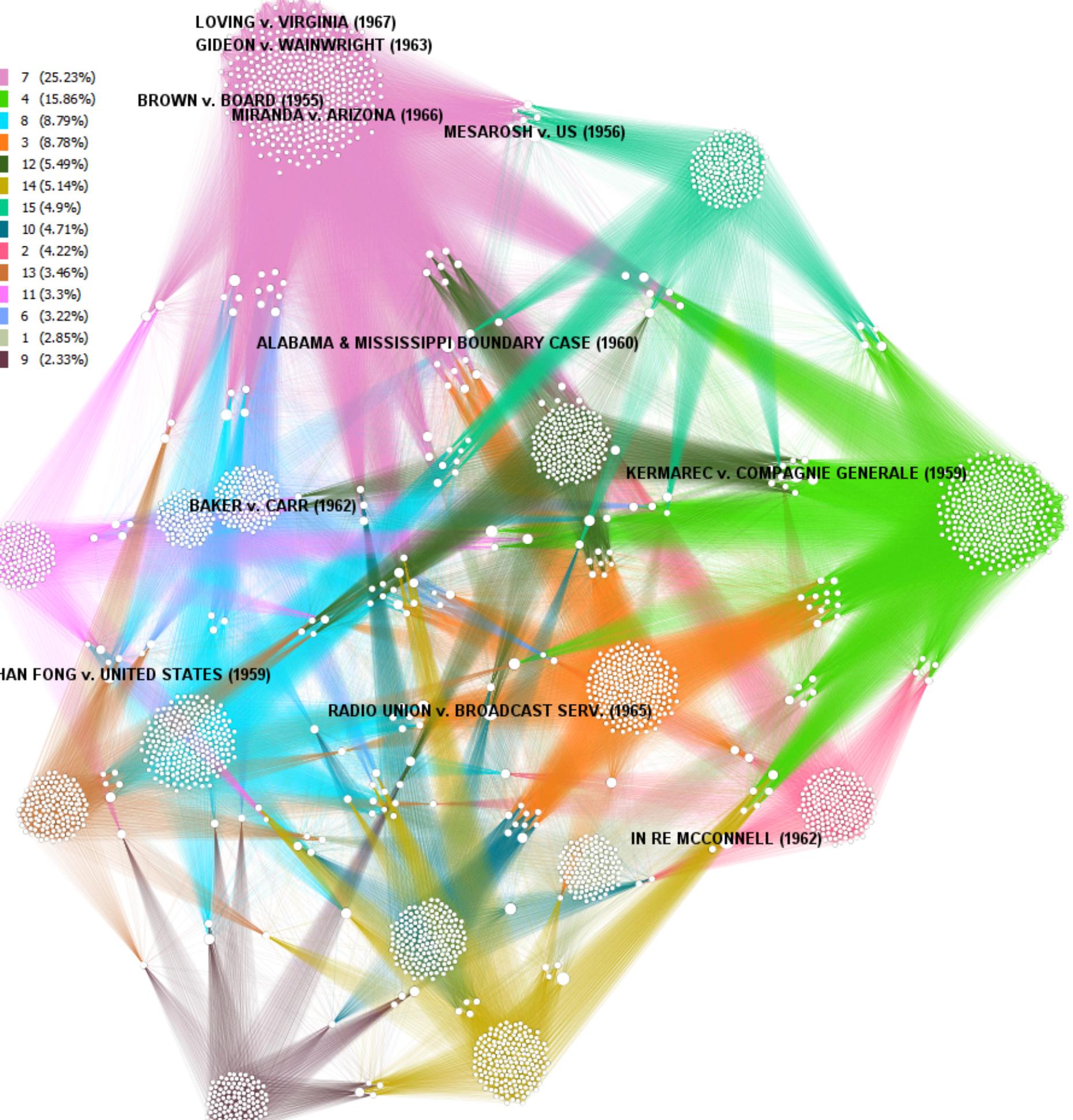
### 3321 NODES / 434,881 EDGES

### EDGE COLOR = TOPIC

### NODE SIZE = BETWEENNESS CENTRALITY

The LDA results are averaged and assigned as the weights for directed edges. Each source node is the highest scoring document for a topic and connects to the subsequent highest scoring documents in the same topic. 15 topics were generated and the threshold for edge creation was set at 0.4. The processing was done in a Jupyter Notebook that was primarily programmed by Kenan Jiang with slight modifications to work with this specific dataset.

It shows that a large body of the cases had to do with the rights of individual citizens, as the four largest topics focus on ideas of freedoms. I found that many of the cases with high betweenness were largely disputes about regulation and trade such as (*Mesaros v. US*, *Radio Union v. Broadcast Serv.*, *In re McConnell*) or matters of determining jurisdiction (*Kermarec v. Compagnie Generale*, *Tak Shan Fong v. US*). This leads to questions about whether issues of jurisdiction and trade may be somehow more interdisciplinary in law than matters of freedoms and rights.



## TOPIC KEYWORDS

- Topic 7 - confession, income, school, insurance, habeas, lien, confessions, corpus, accused, arrest  
Topic 4 - patent, religious, juvenile, election, religion, ship, vessel, sunday, negligence, owner  
Topic 8 - communist, conspiracy, death, collective, bargaining, venue, income, indictment, coal, organizations  
Topic 3 - lands, income, wife, estate, corpus, waters, arbitration, death, water, indian