



The U.S. Supreme Court

A look back on the opinions that have shaped the legal
landscape in the modern political era.

The Problem



The Solution



Goal

Topic Modeling



Navigate the online library of cases.



Acquire opinions relevant to the modern political landscape.



Uncover hidden semantic relationships and themes.

Data

7,741 opinions
[1968 - present]

Preprocessing

1. Removed footnotes and in text citations
2. Fixed Unicode encoding errors
3. Removed '\n', punctuation, numbers and custom stop words
4. Converted all characters to lowercase
5. Applied lemmatization with NLTK

Workflow

PREPROCESS

Pandas, Numpy, Gensim, NLTK

VECTORIZE

Gensim.corpora, sklearn.CountVectorizer, sklearn.TfidfVectorizer

TOPIC MODEL

gensim.LDAMulticore, sklearn.LDA, CorEx

EVALUATE AND INTERPRET

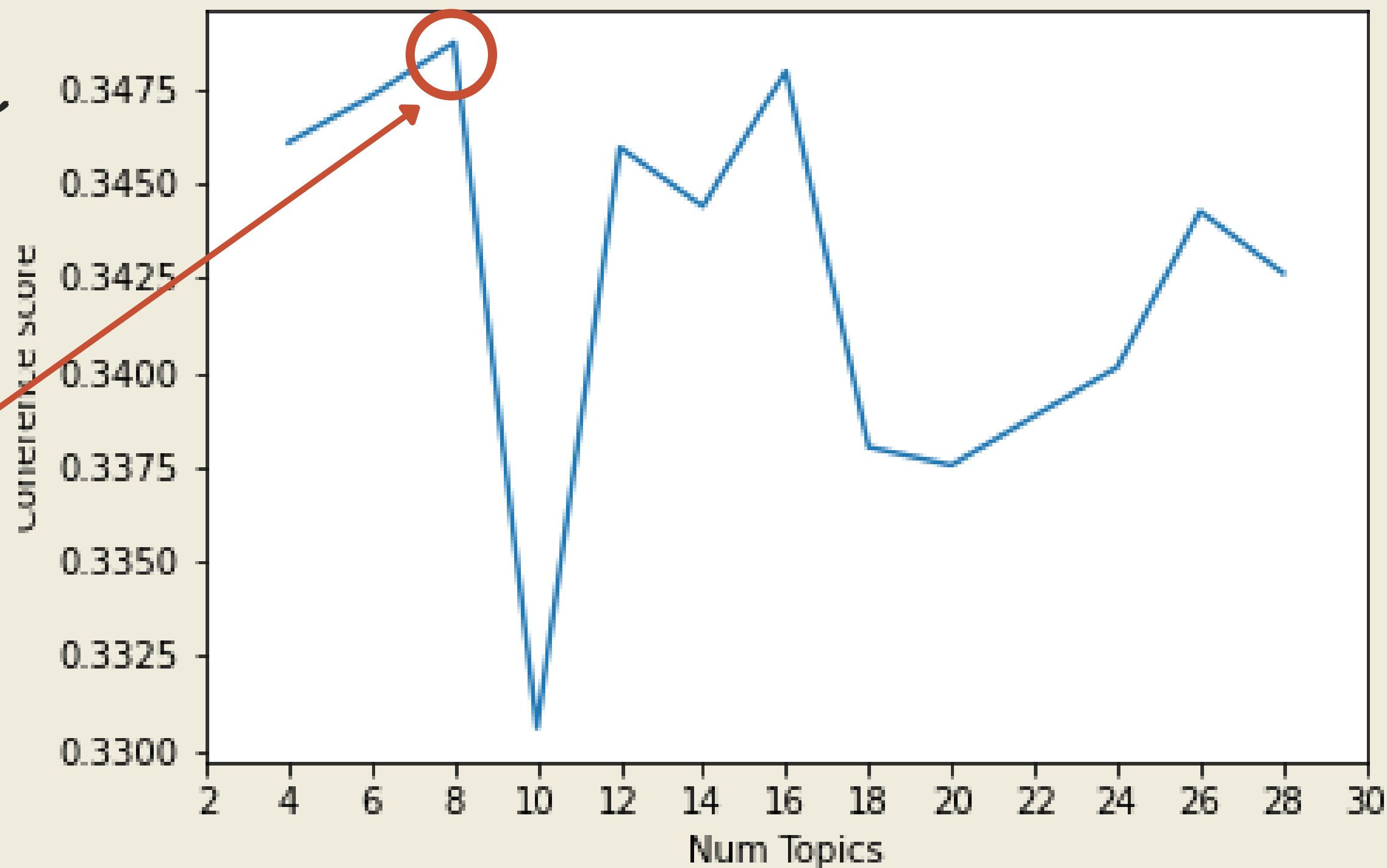
Matplotlib, pyLDavis

Tuning the

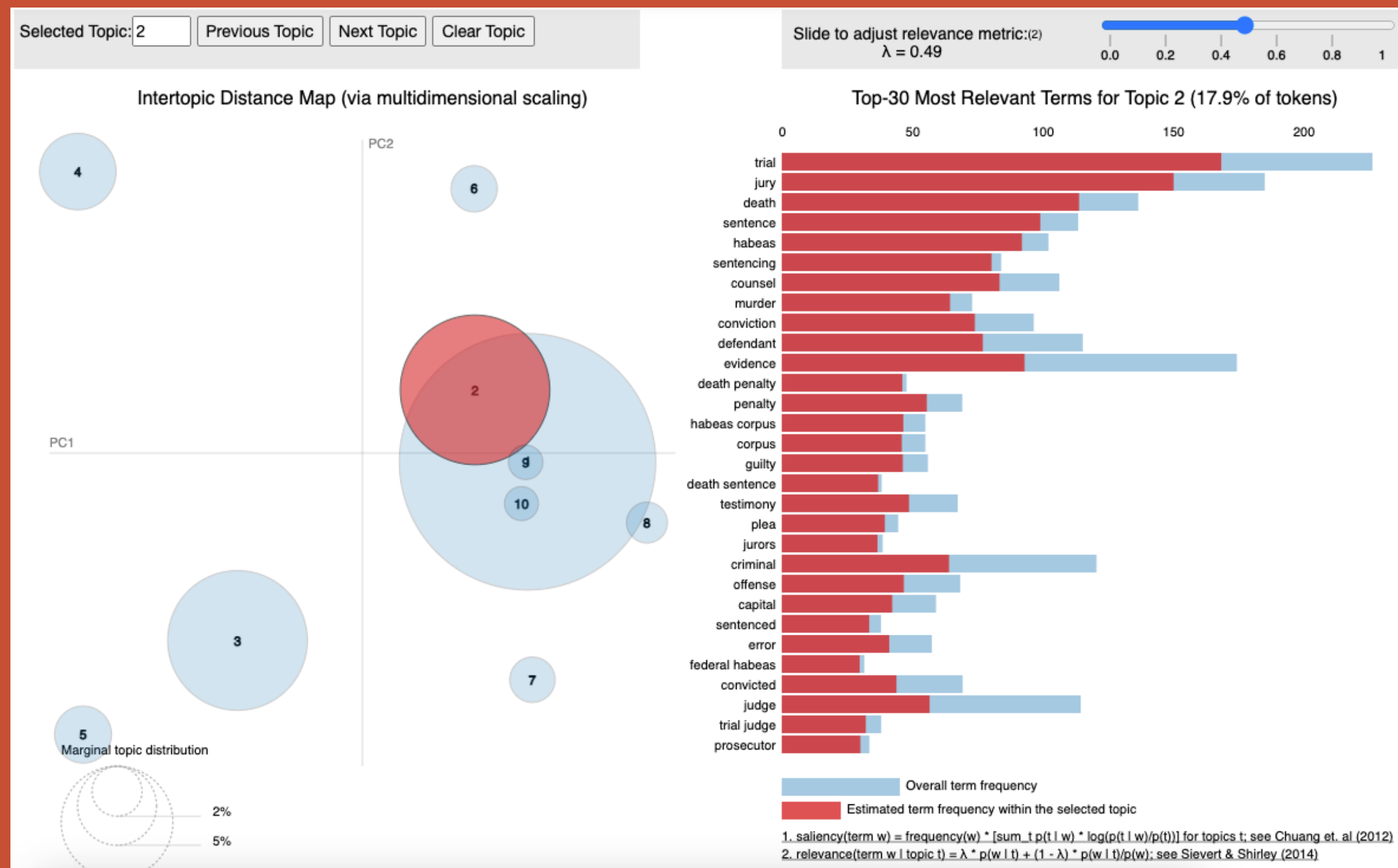
Model

Optimal
Number of
topics = 8

Choosing Optimal Topic Model with Coherence Scores

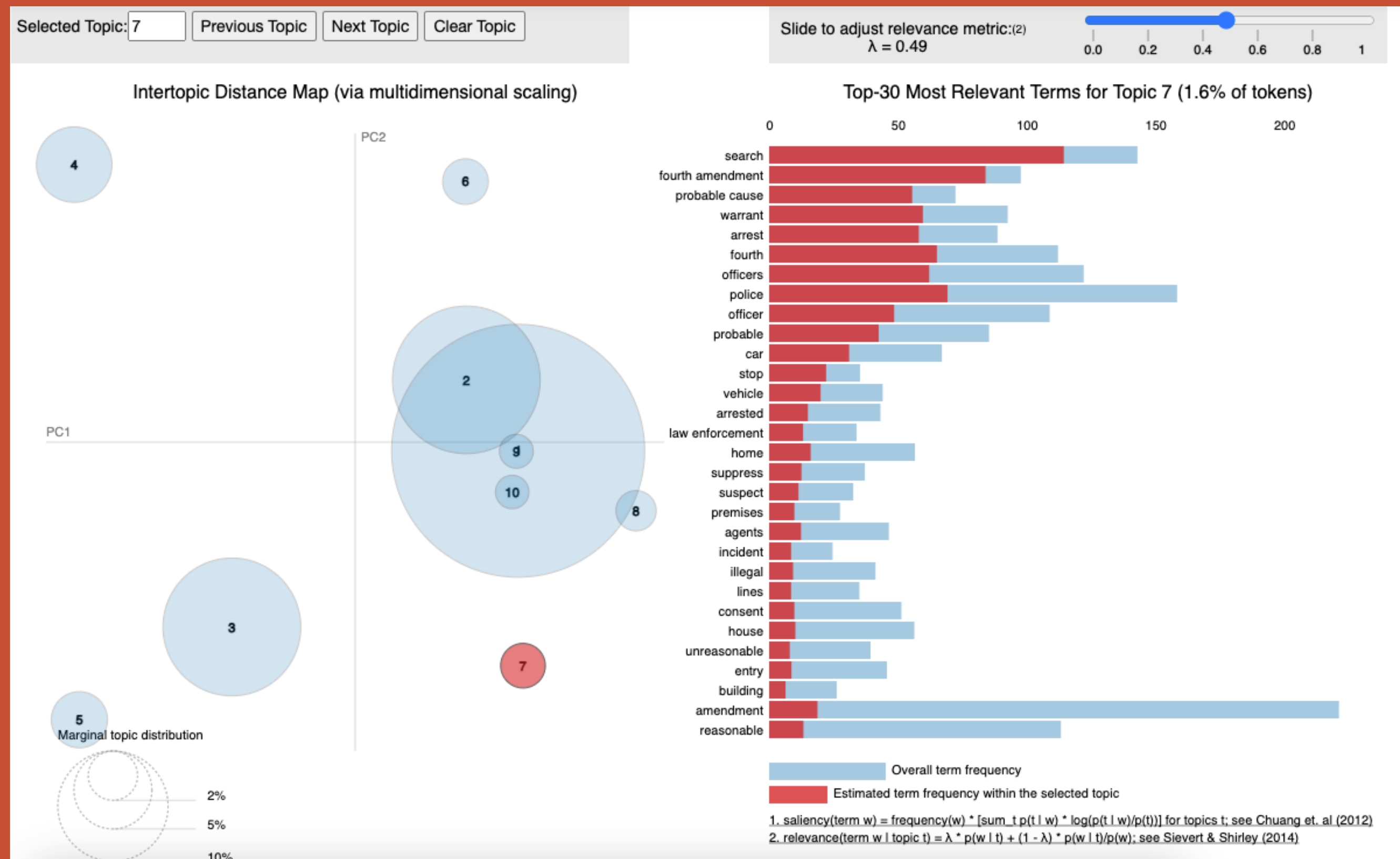


Topic 2: Eighth Amendment: Death Sentence

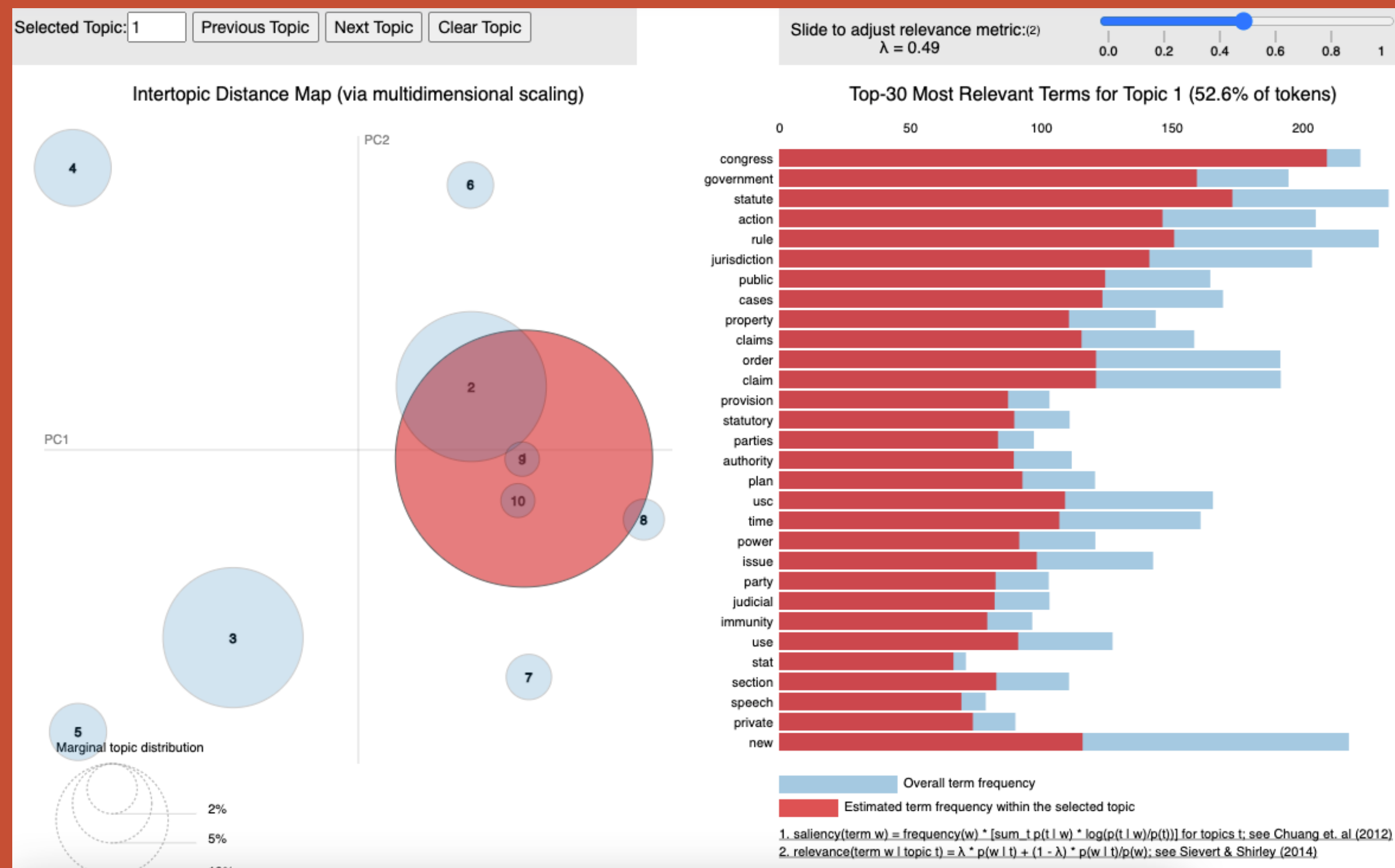


Topic 7:

Unlawful Search and Seizure



Topic 1:
Shows need
for model
improvement.



Future Work

Tuning and Clustering

Perform additional preprocessing to eliminate common legal terms.

Tune LDA model further.

Analyze data through clustering techniques.

Thank you!