



# Analyzing Presidency Similarity Using NLP

*Jie Liang*

# Similarity in Terms of Focus Areas

In a 200 topic space:

- For each president, get the percentage of his documents on every topic, forming a 200D vector.
- The smaller the angle between two vectors, the more similar these presidents are



# NOT Similarity in Terms of Views!

- Two presidents are similar if they focus on similar issues
- **Even if they have completely opposite views!**



# Data

## The American presidency project



- 120K presidential documents from 1789 to 2018
- Including papers, speeches, addresses, economic reports, etc.

# Data Cleaning

Use Spacy for speedy tokenization, lowercasing, removing stop words and punctuation

Example:

Thank **you all for** coming. Welcome **to the** South Lawn. Laura **and I are** thrilled **you're here. This is called "Rodeo on the** South Lawn.

thank come welcome south lawn laura thrilled rodeo south lawn



# Data Transformation

- Convert documents to word document matrix
- Derive 200 topics using topic modelers.

## Example topics:

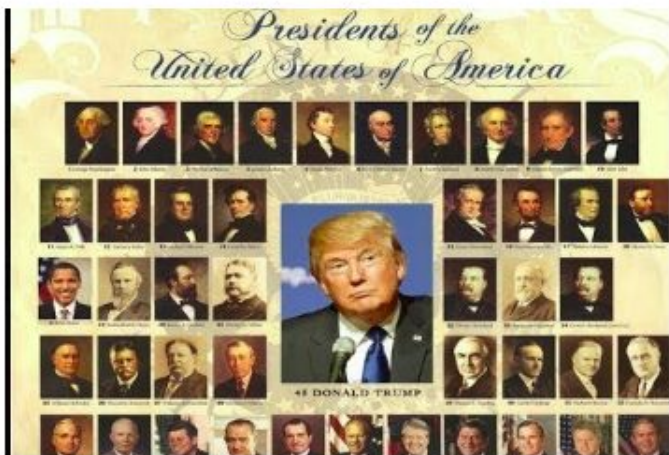
job, worker, economy, create, create job, unemployment, new job, new, million, manufacturing

terrorist, attack, terror, terrorism, enemy, threat, september, terrorist attack, al, war  
terror





# Finding Nearest Neighbors



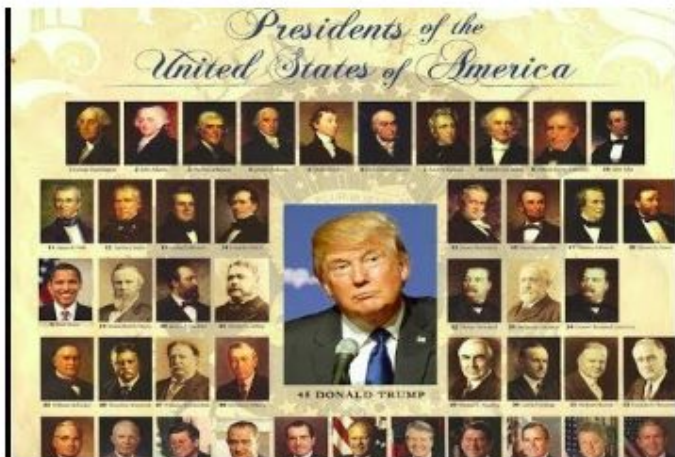
## Similarities Among US Presidents and Presidential Candidates

Find presidents most similar to

Most similar presidents are:

Barack Obama ,William J. Clinton ,George W. Bush

# Finding Nearest Neighbors



## Similaritis Among US Presidents and Presidential Candidates

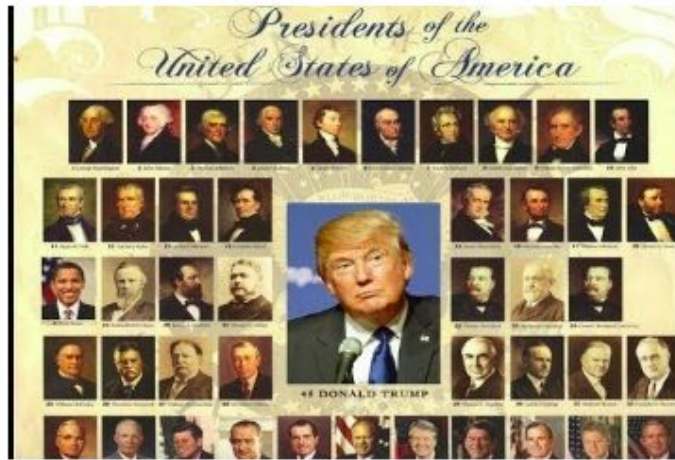
Find presidents most similar to

Most similar presidents are:

Donald J. Trump ,John F. Kennedy ,Lyndon B. Johnson



# Finding Nearest Neighbors



## Similaritis Among US Presidents and Presidential Candidates

Find presidents most similar to

Most similar presidents are:

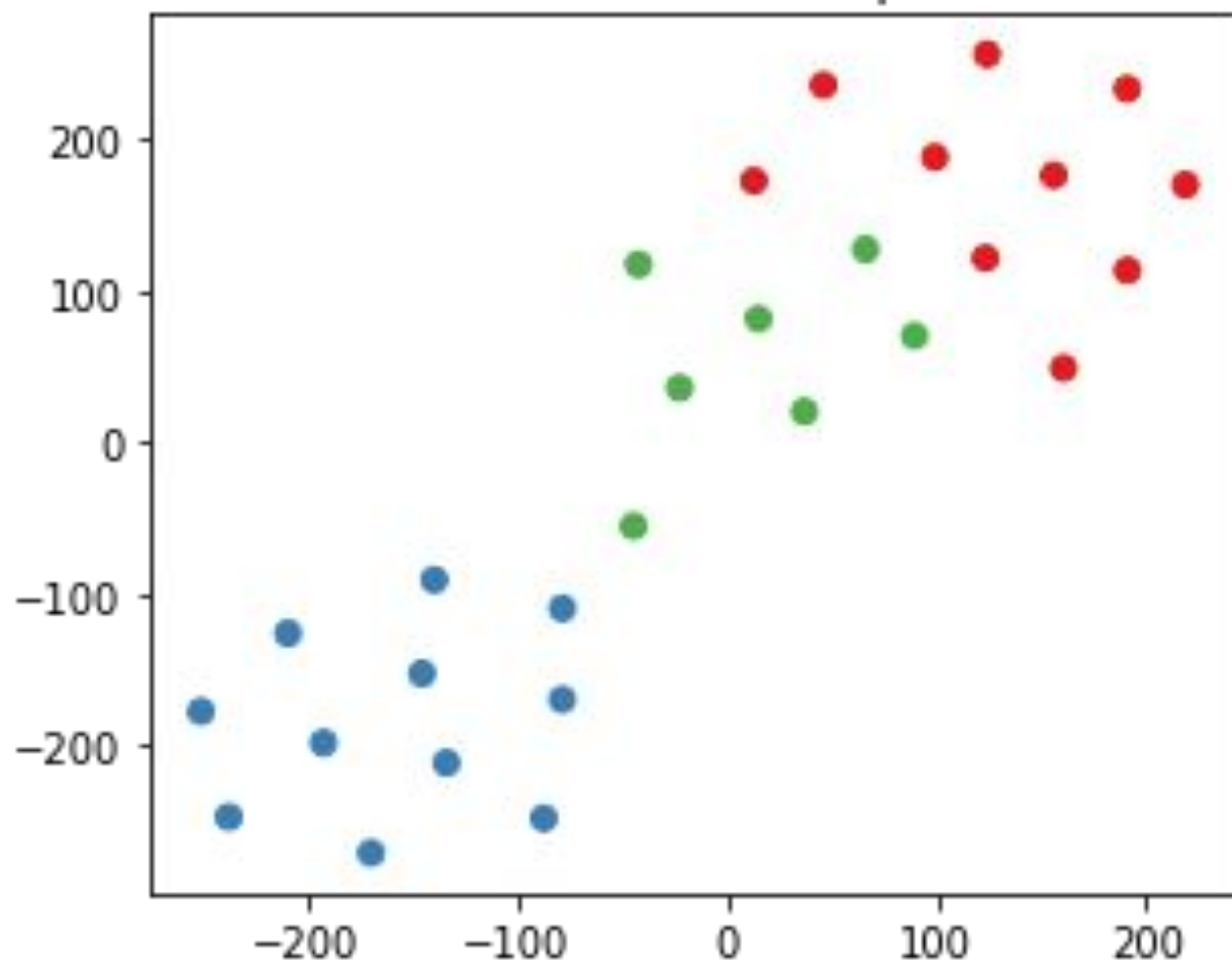
William J. Clinton ,George W. Bush ,Gerald R. Ford

# Finding Neighborhoods(Clustering)

- Group presidents by similarity in topic space
- Visualization in tSNE space



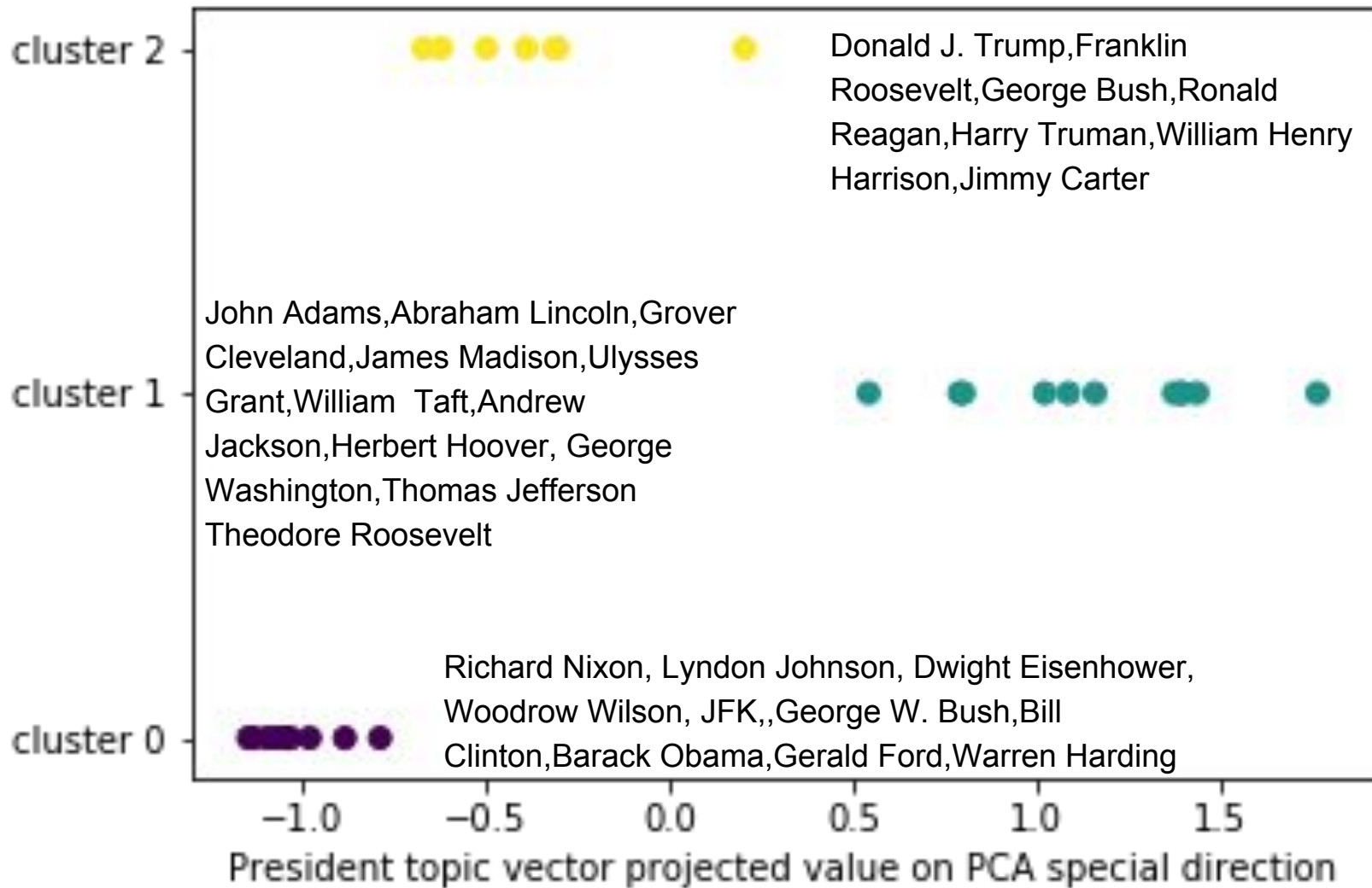
3 Clusters in tSNE space



# Finding Neighborhoods(Dimension Reduction)

- 200 D topic vectors : curse of dimensionality!
- PCA to the rescue: 200 D to 1 D while preserving 85% importance





# Main Takeaway

By measuring angular distance between topic vectors representing each president, we found

- nearest neighbors: most similar presidents in terms of topics frequencies
- neighborhoods: bigger groups each president belongs to





# Next Steps

- Segment presidents by eras and explore period specific topics like cold war and terrorist attacks.
- Analyze documents by categories such as campaign speeches and executive orders over time to compare different writing styles.



