



Department of Computer Science
UNIVERSITY OF COLORADO **BOULDER**



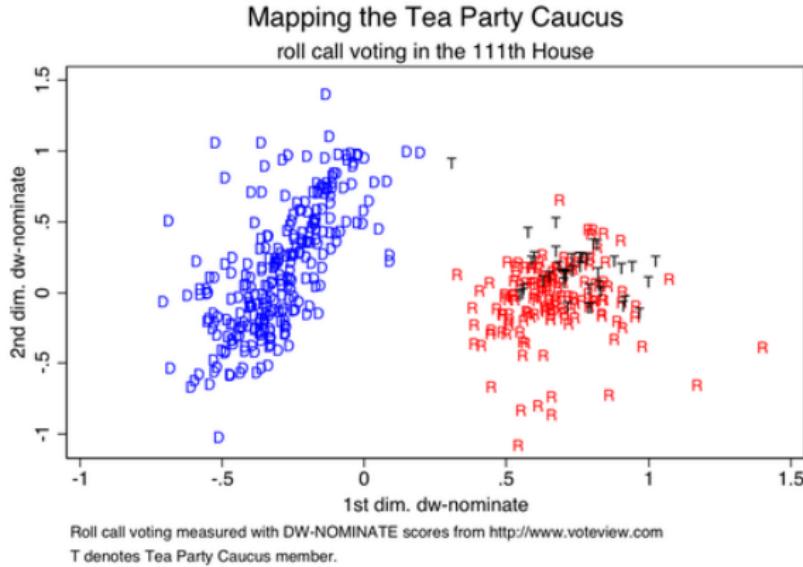
Supervised Topic Models

Advanced Machine Learning for NLP

Jordan Boyd-Graber

OVERVIEW

Motivation: Representing Elected Officials with Ideal Points



An essential tool in political science: distinguish trends and characterize subgroups

Evaluation: Tea Party in the House

The Tea Party

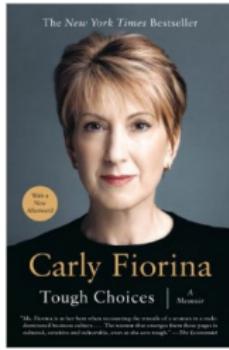
- American political movement for freedom, small government, lower tax
- Disrupting Republican Party and recent elections
- Organizations:
 - Institutional: Tea Party Caucus
 - Other: Tea Party Express, Tea Party Patriots, Freedom Works
- **“Conventional views of ideology as a single-dimensional, left–right spectrum experience great difficulty in understanding or explaining the Tea Party.”**

(?, ARPS)

Goal

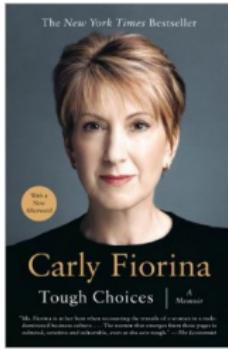
- Explain Tea Partiers in terms of issues and votes
- Identify Tea Partiers from their rhetoric

Not everyone has a voting record



- Ideal points estimated based on voting record
- Not all candidates have a voting record
 - Governors
 - Entertainers
 - CEOs

Not everyone has a voting record



- Ideal points estimated based on voting record
- Not all candidates have a voting record
 - Governors
 - Entertainers
 - CEOs
- But all politicians—by definition—talk

Let's use whatever data we have

 Dr. Ben Carson @RealBenCarson · May 7

I'm pleased the Senate just passed the Corker-Menendez bill requiring Congressional review of the administration's proposed treaty with Iran

  333  662  ...

 Dr. Ben Carson @RealBenCarson · May 7

Met with some Pastors & community leaders from the inner city #OneBaltimore

A single model that uses:

- Bill text
- Votes
- Commentary

to map political actors to the same continuous space.

Let's use whatever data we have

 Dr. Ben Carson @RealBenCarson · May 7

I'm pleased the Senate just passed the Corker-Menendez bill requiring Congressional review of the administration's proposed treaty with Iran

  333  662  ...

 Dr. Ben Carson @RealBenCarson · May 7

Met with some Pastors & community leaders from the inner city #OneBaltimore

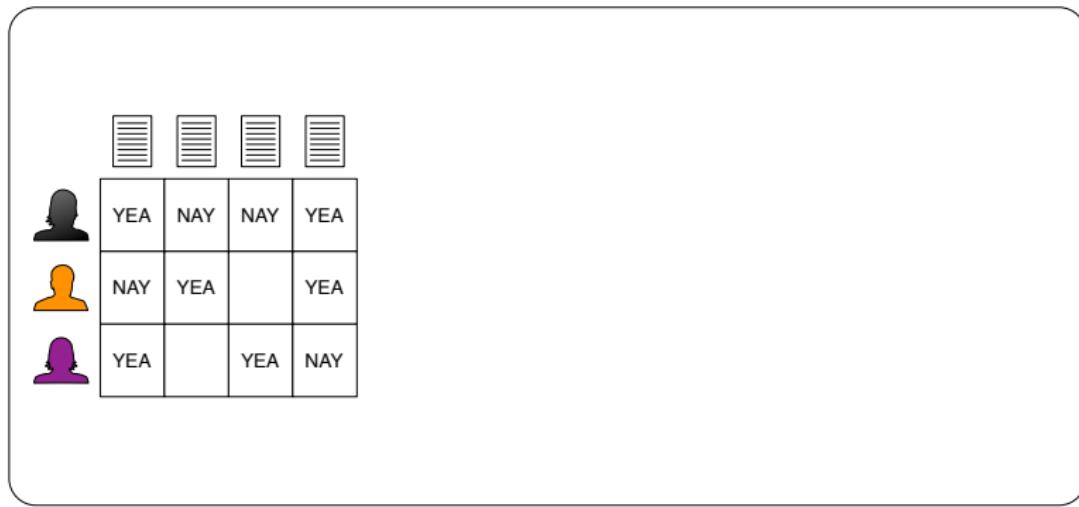
A single model that uses:

- Bill text
- Votes
- Commentary

to map political actors to the same continuous space. This work: congressional floor speeches

Outline

One-dimensional Ideal Point using Votes



| | | | | |
|--|-----|-----|-----|-----|
| | | | | |
| | YEA | NAY | NAY | YEA |
| | NAY | YEA | | YEA |
| | YEA | | YEA | NAY |

?

One-dimensional Ideal Point using Votes

Legislator a votes 'Yea' on bill b with probability



| | | | | |
|--|-----|-----|-----|-----|
| | YEA | NAY | NAY | YEA |
| | NAY | YEA | | YEA |
| | YEA | | YEA | NAY |

$$p(v_{a,b} = \text{Yea}) = \Phi(u_a x_b + y_b)$$

$$\Phi(\alpha) = \frac{\exp(\alpha)}{\exp(\alpha) + 1}$$

?

One-dimensional Ideal Point using Votes

Legislator a votes 'Yea' on bill b with probability

| | | | | |
|--------|-----|-----|-----|-----|
| | YEA | NAY | NAY | YEA |
| YEA | | YEA | | YEA |
| NAY | | | | |
| PURPLE | YEA | | YEA | NAY |

One-dimensional ideal point
of legislator a

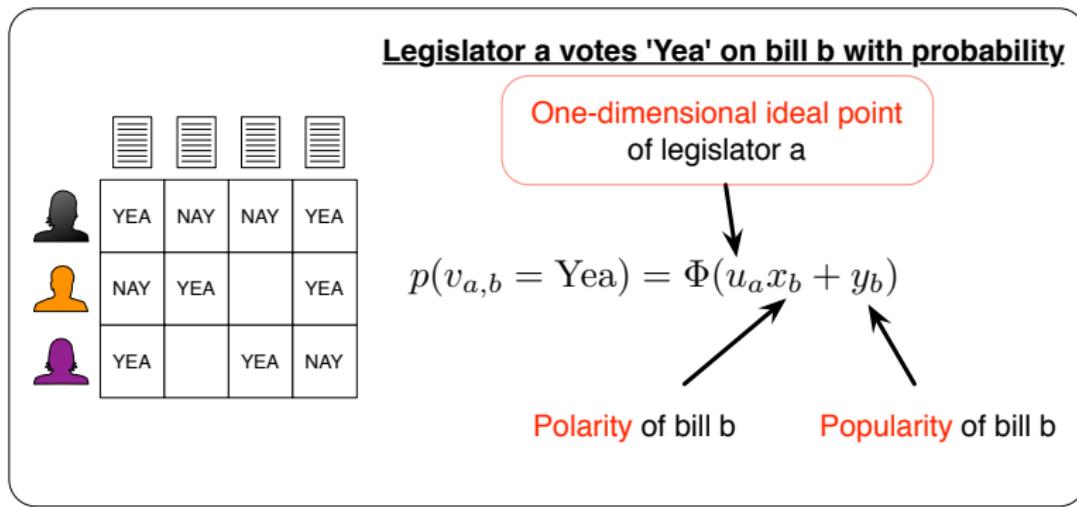
$p(v_{a,b} = \text{Yea}) = \Phi(u_a x_b + y_b)$

Polarity of bill b Popularity of bill b

The diagram illustrates the formula for the probability of a legislator voting 'Yea' on a specific bill. It features three icons representing different individuals above a 4x5 grid of YEA/NAY votes. To the right, a red arrow points down from the formula to the term y_b , labeled 'Popularity of bill b'. Another red arrow points diagonally up and to the left from the formula to the term $u_a x_b$, labeled 'Polarity of bill b'.

?

One-dimensional Ideal Point using Votes



?



One-dimensional Ideal Point using Votes

Legislator a votes 'Yea' on bill b with probability

| | | | |
|-----|-----|-----|-----|
| YEA | NAY | NAY | YEA |
| NAY | YEA | | YEA |
| YEA | | YEA | NAY |

One-dimensional ideal point
of legislator a

$$p(v_{a,b} = \text{Yea}) = \Phi(u_a x_b + y_b)$$

Polarity of bill b Popularity of bill b



Multi-dimensional Ideal Point using Votes

Legislator a votes 'Yea' on bill b with probability

| | | | | |
|--|-----|-----|-----|-----|
| | | | | |
| | YEA | NAY | NAY | YEA |
| | NAY | YEA | | YEA |
| | YEA | | YEA | NAY |

$$p(v_{a,b} = \text{Yea}) = \Phi \left(\sum_{k=1}^K u_{a,k} x_{b,k} + y_b \right)$$

???

Multi-dimensional Ideal Point using Votes

Legislator a votes 'Yea' on bill b with probability

| | | | | |
|-----|-----|-----|-----|-----|
| | YEA | NAY | NAY | YEA |
| YEA | | YEA | | YEA |
| NAY | | | | |
| YEA | | YEA | YEA | NAY |

Multi-dimensional ideal point
of legislator a

$$p(v_{a,b} = \text{Yea}) = \Phi \left(\sum_{k=1}^K u_{a,k} x_{b,k} + y_b \right)$$

??? K dimensions

Multi-dimensional Ideal Point using Votes

Legislator a votes 'Yea' on bill b with probability

| | | | | |
|-----|-----|-----|-----|-----|
| | YEA | NAY | NAY | YEA |
| YEA | | YEA | | |
| NAY | | | | YEA |
| | YEA | | YEA | NAY |

Multi-dimensional ideal point
of legislator a

$$p(v_{a,b} = \text{Yea}) = \Phi \left(\sum_{k=1}^K u_{a,k} x_{b,k} + y_b \right)$$

K dimensions

???



Multi-dimensional Ideal Point using Votes

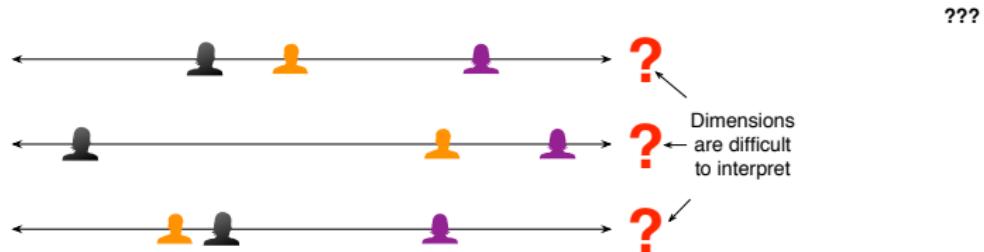
Legislator a votes 'Yea' on bill b with probability

| | | | | |
|-----|-----|-----|-----|-----|
| | YEA | NAY | NAY | YEA |
| YEA | | YEA | | |
| NAY | | | | YEA |
| | YEA | | YEA | NAY |

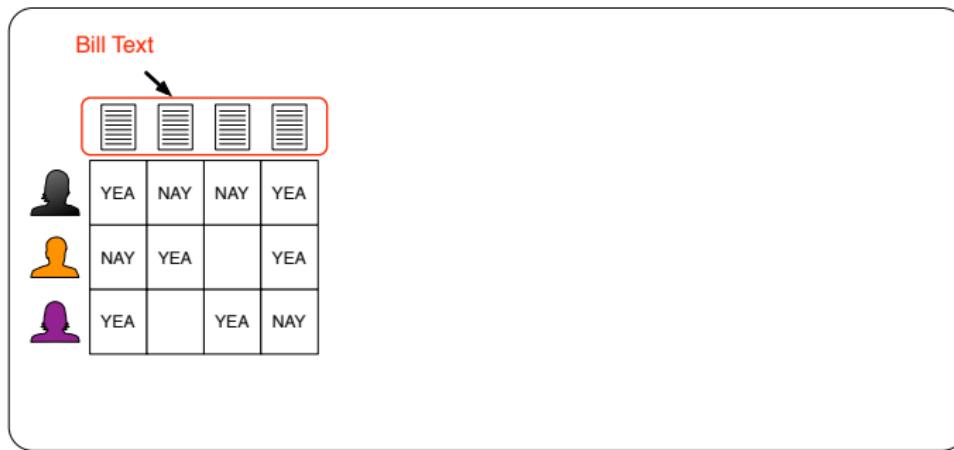
Multi-dimensional ideal point
of legislator a

$$p(v_{a,b} = \text{Yea}) = \Phi \left(\sum_{k=1}^K u_{a,k} x_{b,k} + y_b \right)$$

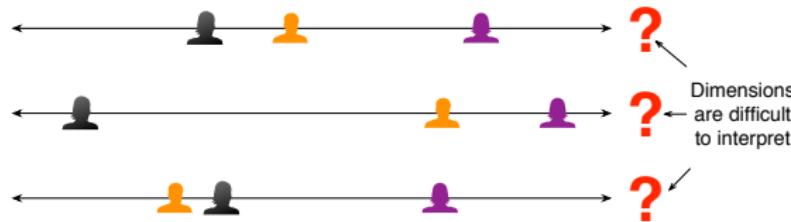
K dimensions



Multi-dimensional Ideal Point using Votes & Text



????



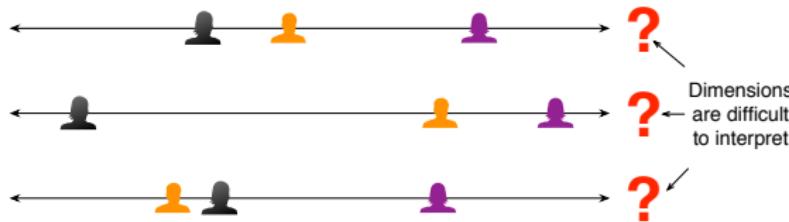
Multi-dimensional Ideal Point using Votes & Text

Legislator a votes 'Yea' on bill b with probability

| | | | | |
|--|-----|-----|-----|-----|
| | YEA | NAY | NAY | YEA |
| | NAY | YEA | | YEA |
| | YEA | | YEA | NAY |

$$p(v_{a,b} = \text{Yea}) = \Phi \left(x_b \sum_{k=1}^K u_{a,k} \vartheta_{b,k} + y_b \right)$$

????



Multi-dimensional Ideal Point using Votes & Text

Legislator a votes 'Yea' on bill b with probability

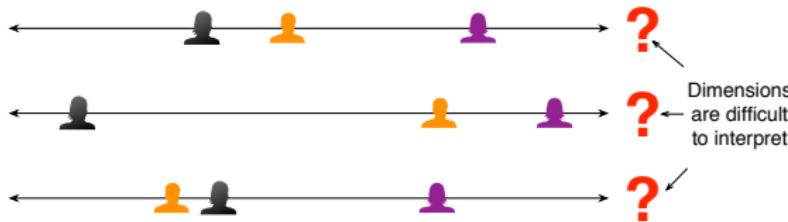
| | | | | |
|-----|-----|-----|-----|-----|
| | YEA | NAY | NAY | YEA |
| YEA | NAY | YEA | | YEA |
| NAY | YEA | | | |

Multi-dimensional ideal point
of legislator a

$$p(v_{a,b} = \text{Yea}) = \Phi \left(x_b \sum_{k=1}^K u_{a,k} v_{b,k} + y_b \right)$$

Topic proportion of bill b
estimated from its text

????



Multi-dimensional Ideal Point using Votes & Text

Legislator a votes 'Yea' on bill b with probability

| | | | | |
|-----|-----|-----|-----|-----|
| | YEA | NAY | NAY | YEA |
| YEA | | | | |
| NAY | YEA | | | YEA |
| | YEA | | YEA | NAY |

Multi-dimensional ideal point
of legislator a

$$p(v_{a,b} = \text{Yea}) = \Phi \left(x_b \sum_{k=1}^K u_{a,k} \theta_{b,k} + y_b \right)$$

Topic proportion of bill b
estimated from its text

????



Outline

Hierarchical Ideal Point Topic Model: Intuition

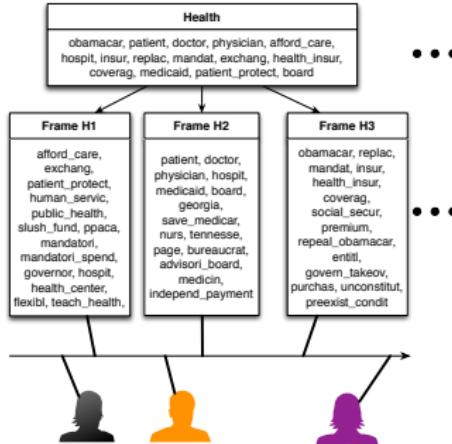
What are your thoughts on the issue of **immigration**?



Hierarchical Ideal Point Topic Model: Overview

Using both votes and text to learn

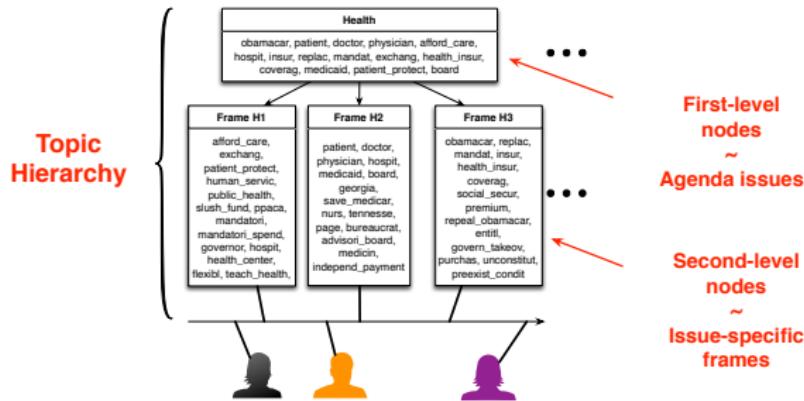
- Two-level topic hierarchy:
- Ideal points in multiple interpretable dimensions



Hierarchical Ideal Point Topic Model: Overview

Using both votes and text to learn

- Two-level topic hierarchy:
 - First-level nodes map to agenda issues
 - Second-level nodes map to issue-specific frames
- Ideal points in multiple interpretable dimensions

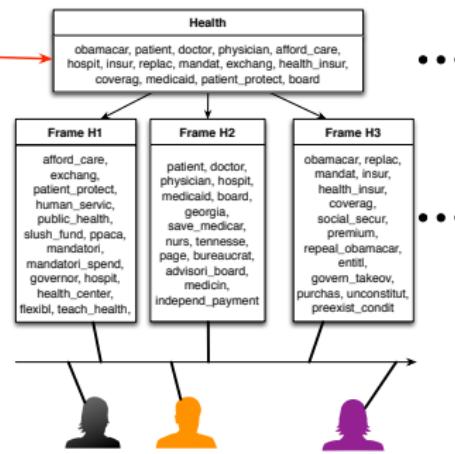


Hierarchical Ideal Point Topic Model: Overview

Using both votes and text to learn

- Two-level topic hierarchy: Use existing labeled data to learn priors for interpretable issues
- Ideal points in multiple interpretable dimensions

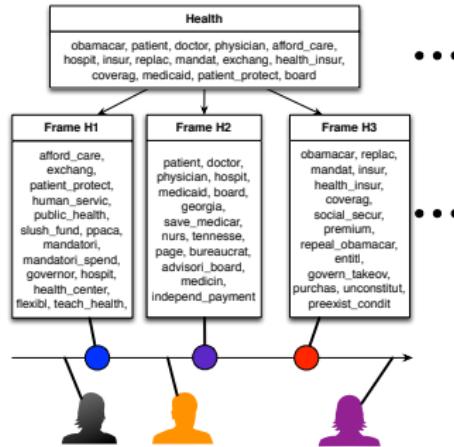
Use prior to learn
interpretable
issue topics



Hierarchical Ideal Point Topic Model: Overview

Using both votes and text to learn

- Two-level topic hierarchy: Ideal points for frames for predictions using text only
- Ideal points in multiple interpretable dimensions

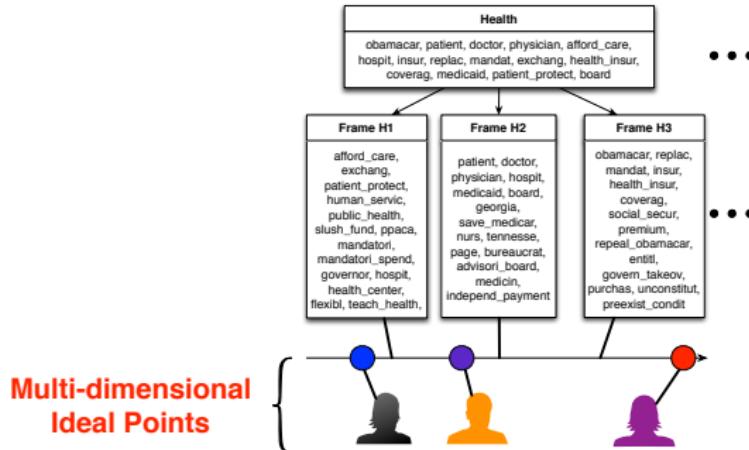


Learn ideal point
for each frame

Hierarchical Ideal Point Topic Model: Overview

Using both votes and text to learn

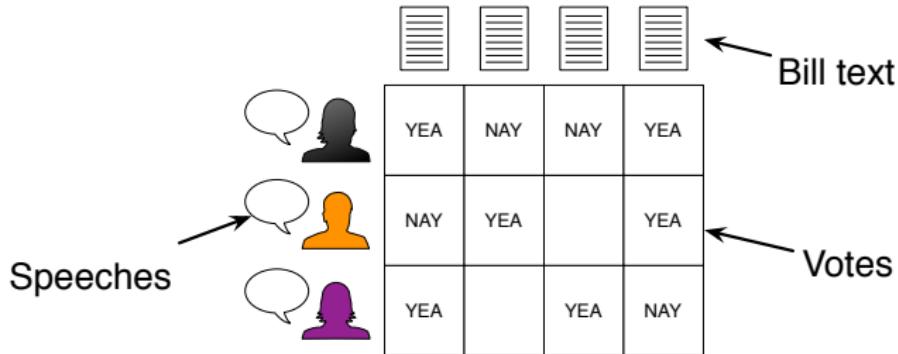
- Two-level topic hierarchy:
- Ideal points in multiple interpretable dimensions



Hierarchical Ideal Point Topic Model: Overview

Hierarchical Ideal Point Topic Model: Inputs

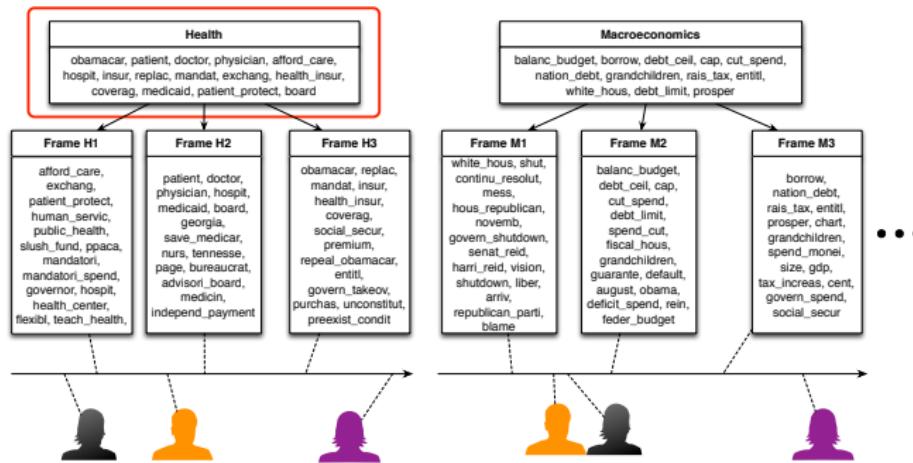
- A collection of votes $\{v_{a,b}\}$
- A collection of D speeches $\{\mathbf{w}_d\}$, each of which is given by legislator a_d
- A collection of B bill text $\{\mathbf{w}'_b\}$



Hierarchical Ideal Point Topic Model

Modeling bill text

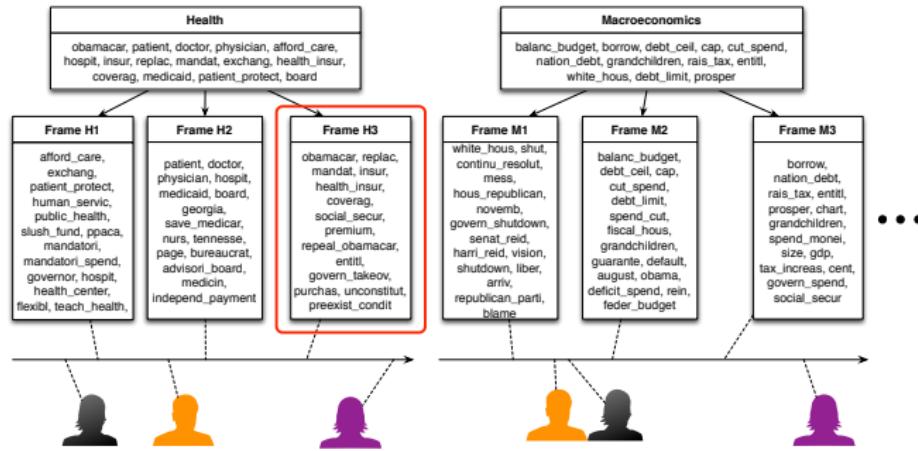
- Each bill text b is a mixture over K issues ϑ_b
- Each bill token generated from topic at **first-level issue node**



Hierarchical Ideal Point Topic Model

Hierarchical Ideal Point Topic Model: Generative Process

- Each speech d also has a distribution θ_d over K issues
- Each issue k , each speech d has distribution over frames $\psi_{d,k}$
- Each speech token from topic at **second-level frame node**

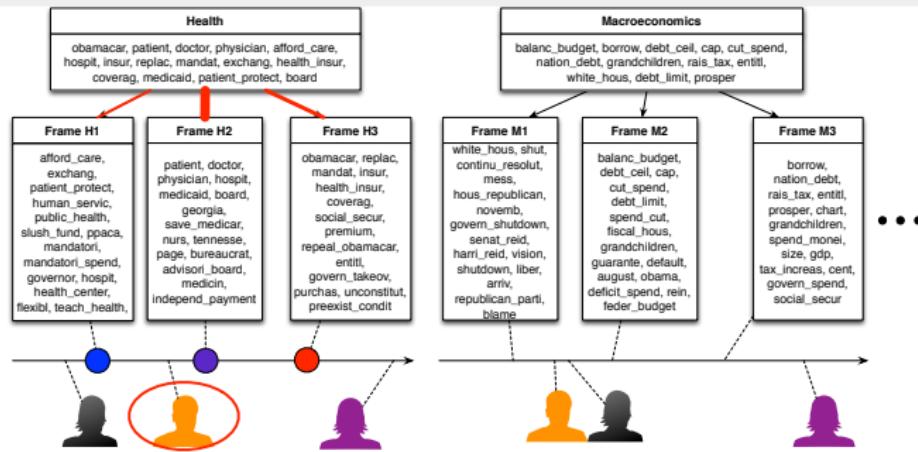


Hierarchical Ideal Point Topic Model

Hierarchical Ideal Point Topic Model: Modeling votes

- Legislator a votes ‘Yea’ on bill b with probability

$$p(v_{a,b} = \text{Yea}) = \Phi(x_b \sum_{k=1}^K \vartheta_{b,k} u_{a,k} + y_b)$$
- Ideal point $u_{a,k} \sim \mathcal{N}(\sum_{j=1}^{J_k} \eta_{k,j} \psi_{a,k,j}, \rho)$

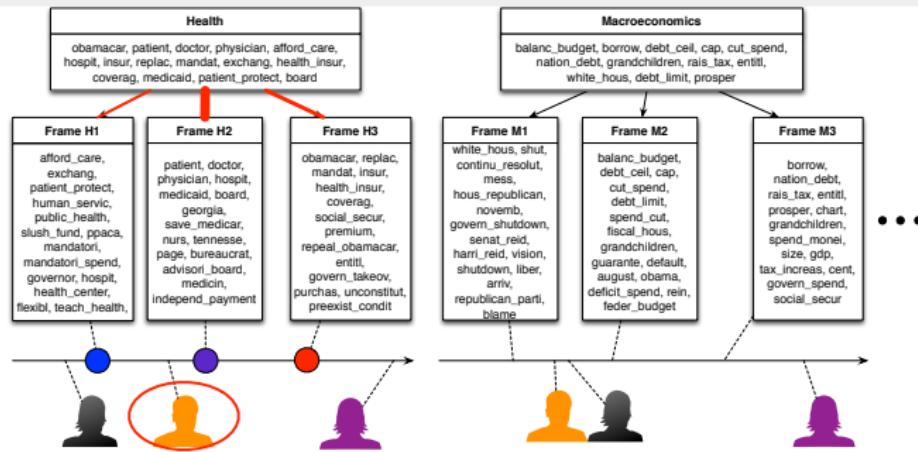


Hierarchical Ideal Point Topic Model

Hierarchical Ideal Point Topic Model: Modeling votes

- Legislator a votes ‘Yea’ on bill b with probability

$$p(v_{a,b} = \text{Yea}) = \Phi(x_b \sum_{k=1}^K \vartheta_{b,k} u_{a,k} + y_b)$$
- Ideal point $u_{a,k} \sim \mathcal{N}(\sum_{j=1}^{J_k} \eta_{k,j} \psi_{a,k,j}, \rho)$

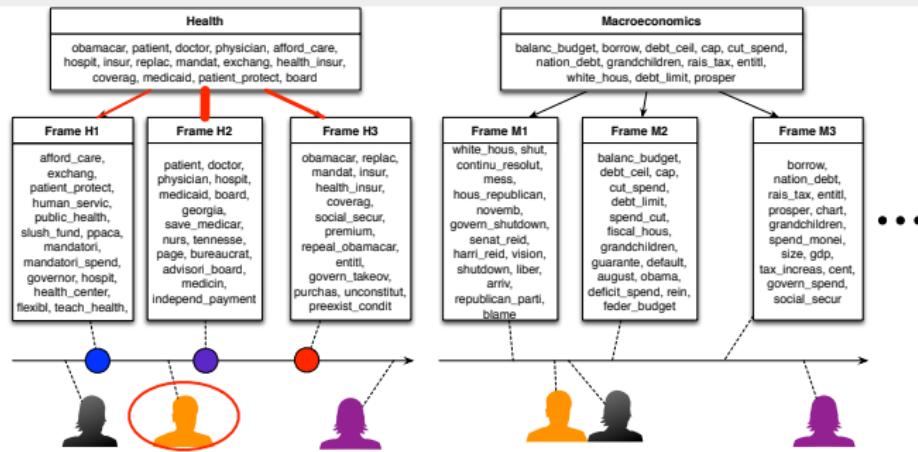


Hierarchical Ideal Point Topic Model

Hierarchical Ideal Point Topic Model: Modeling votes

- Legislator a votes ‘Yea’ on bill b with probability

$$p(v_{a,b} = \text{Yea}) = \Phi(x_b \sum_{k=1}^K \vartheta_{b,k} u_{a,k} + y_b)$$
- Ideal point $\mathbf{u}_{a,k} \sim \mathcal{N}(\sum_{j=1}^{J_k} \eta_{k,j} \psi_{a,k,j}, \rho)$

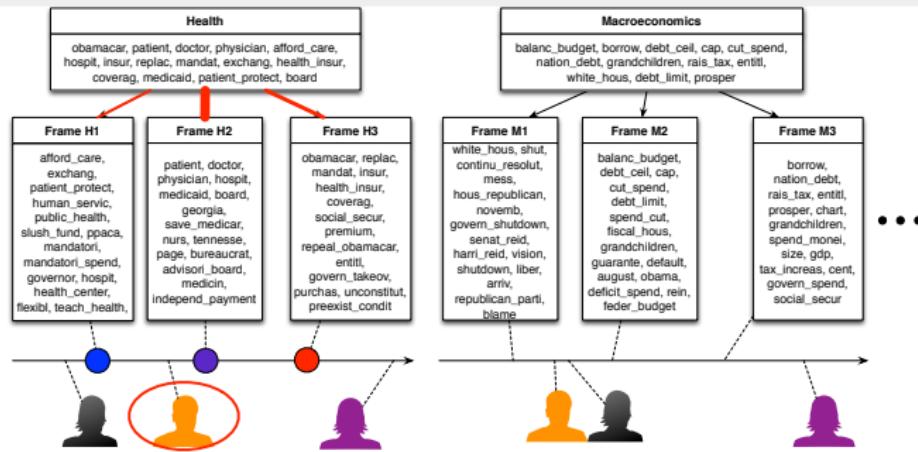


Hierarchical Ideal Point Topic Model

Hierarchical Ideal Point Topic Model: Modeling votes

- Legislator a votes ‘Yea’ on bill b with probability

$$p(v_{a,b} = \text{Yea}) = \Phi(x_b \sum_{k=1}^K \vartheta_{b,k} u_{a,k} + y_b)$$
- Ideal point $u_{a,k} \sim \mathcal{N}(\sum_{j=1}^{J_k} \eta_{k,j} \psi_{a,k,j}, \rho)$

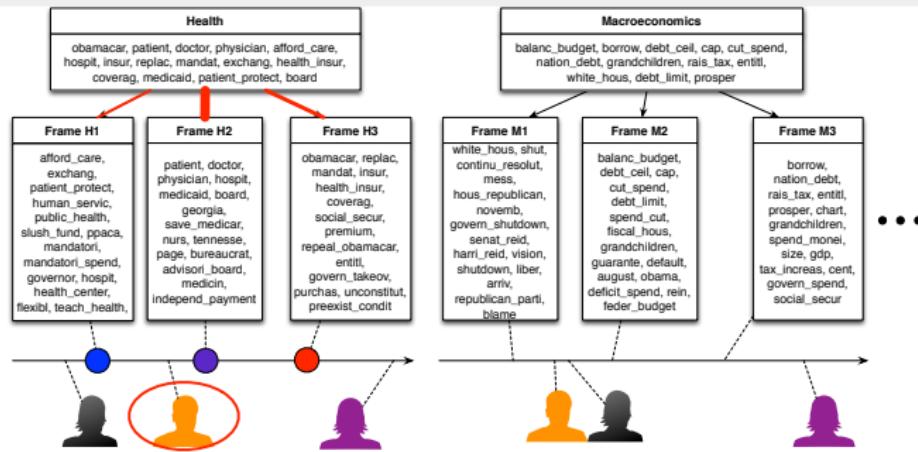


Hierarchical Ideal Point Topic Model

Hierarchical Ideal Point Topic Model: Modeling votes

- Legislator a votes ‘Yea’ on bill b with probability

$$p(v_{a,b} = \text{Yea}) = \Phi(x_b \sum_{k=1}^K \vartheta_{b,k} u_{a,k} + y_b)$$
- Ideal point $u_{a,k} \sim \mathcal{N}(\sum_{j=1}^{J_k} \eta_{k,j} \psi_{a,k,j}, \rho)$



Evaluation: Tea Party in the House

The Tea Party

- American political movement for freedom, small government, lower tax
- Disrupting Republican Party and recent elections

Evaluation: Tea Party in the House

The Tea Party

- American political movement for freedom, small government, lower tax
- Disrupting Republican Party and recent elections

Data

- 240 Republican Representatives in the 112th U.S. House
- 60 are members of the Tea Party Caucus (self-identified)
- 60 key votes selected by Freedom Works (2011-2012)
- Speeches, bill text and voting records from the Library of Congress

Outline

Tea Party Caucus Membership Prediction

Experiment setup

- Task: Binary classification of whether a legislator is a member of the Tea Party Caucus
- Evaluation metric: AUC-ROC
- Classifier: SVM^{light}
- Five-fold stratified cross-validation

Tea Party Caucus Membership Prediction

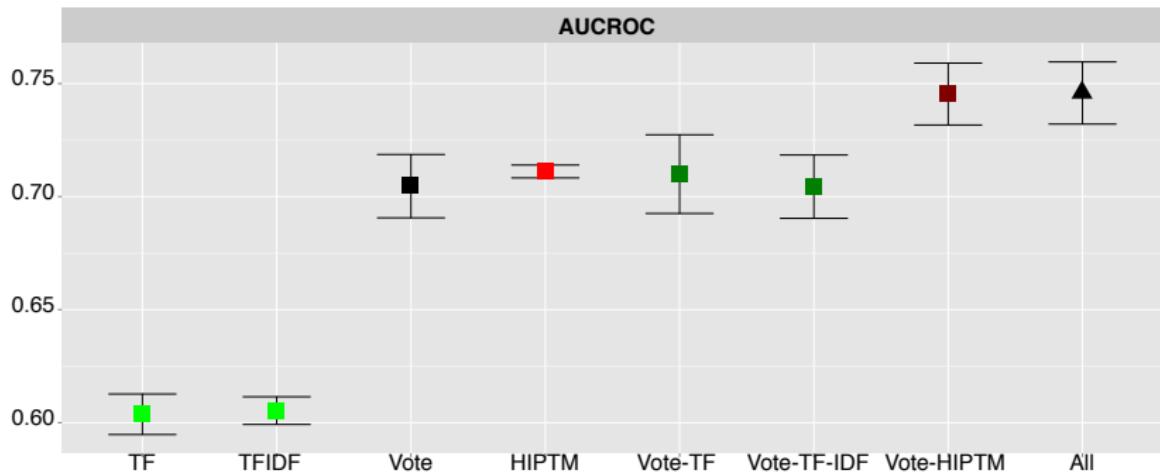
Experiment setup

- Task: Binary classification of whether a legislator is a member of the Tea Party Caucus
- Evaluation metric: AUC-ROC
- Classifier: SVM^{light}
- Five-fold stratified cross-validation

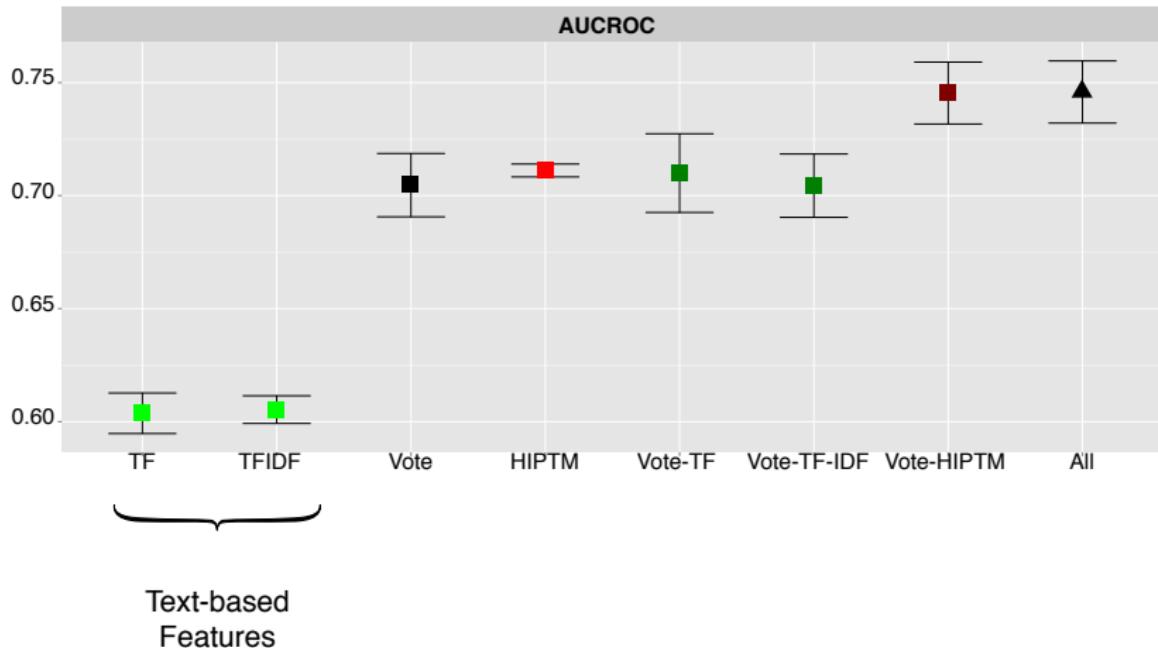
Features

- Text-based features: normalized term frequency (**TF**) and **TF-IDF**
- **Vote**: binary features
- **HIPTM**: features extracted from our model including
 - K -dim ideal point $u_{a,k}$ estimated from both votes and text
 - K -dim ideal point estimated from text only $\eta_k^T \hat{\psi}_{a,k}$
 - B probabilities estimating a 's votes $\Phi(x_b \sum_{k=1}^K \vartheta_{b,k} u_{a,k} + y_b)$

Tea Party Caucus Membership Prediction: Votes & Text

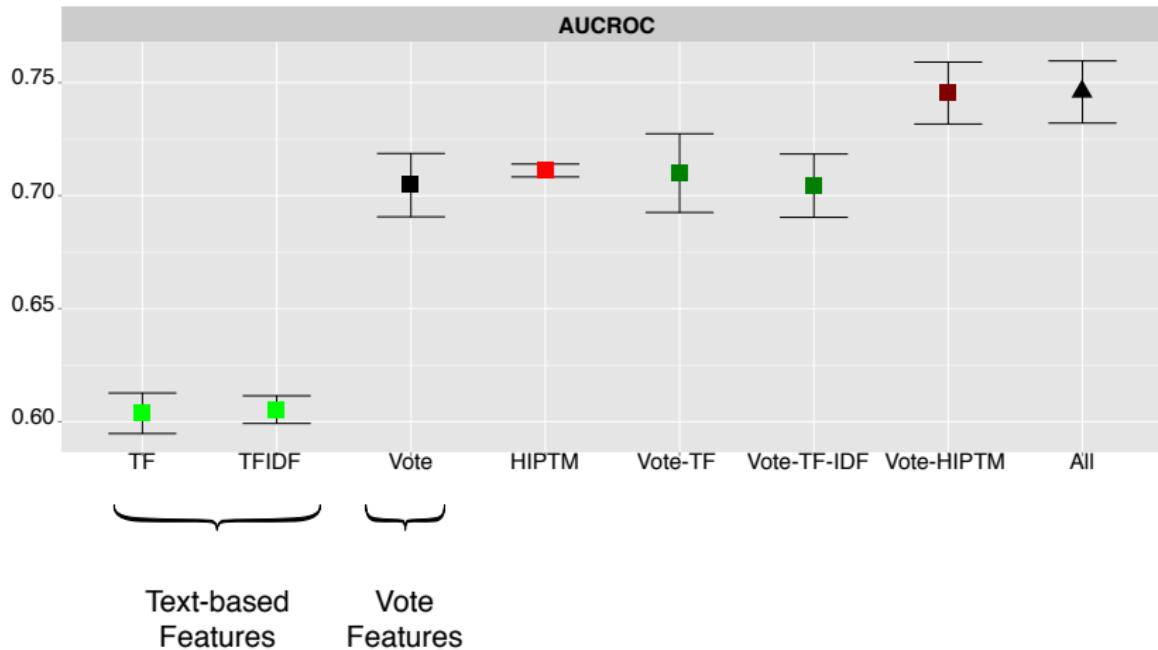


Tea Party Caucus Membership Prediction: Votes & Text

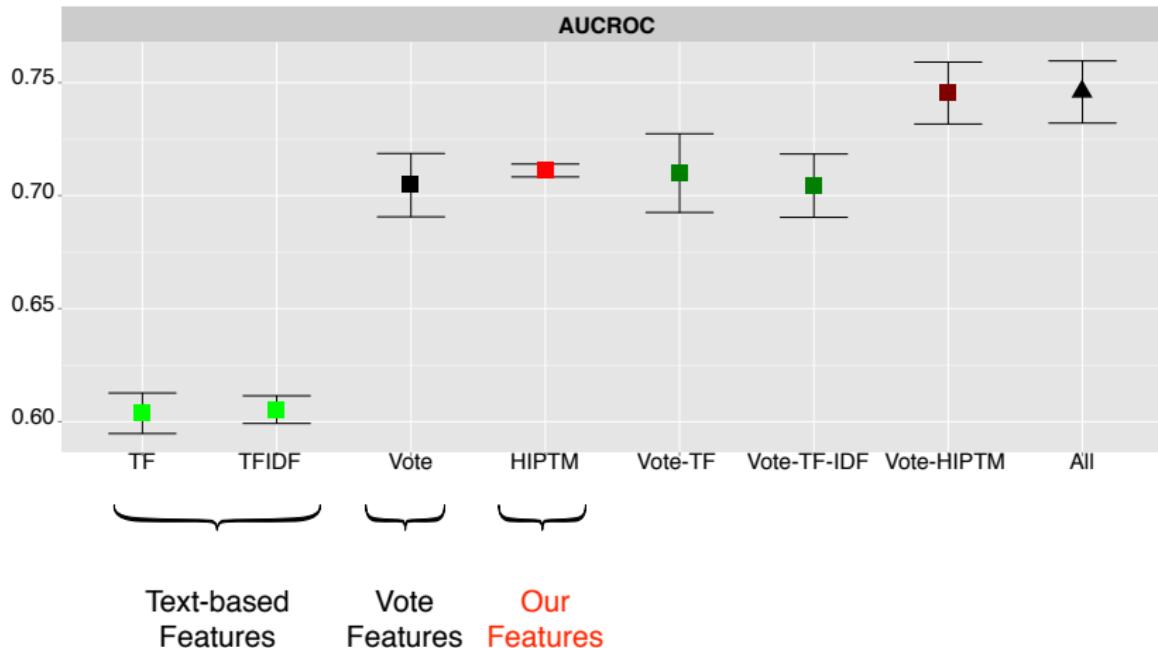


Text-based
Features

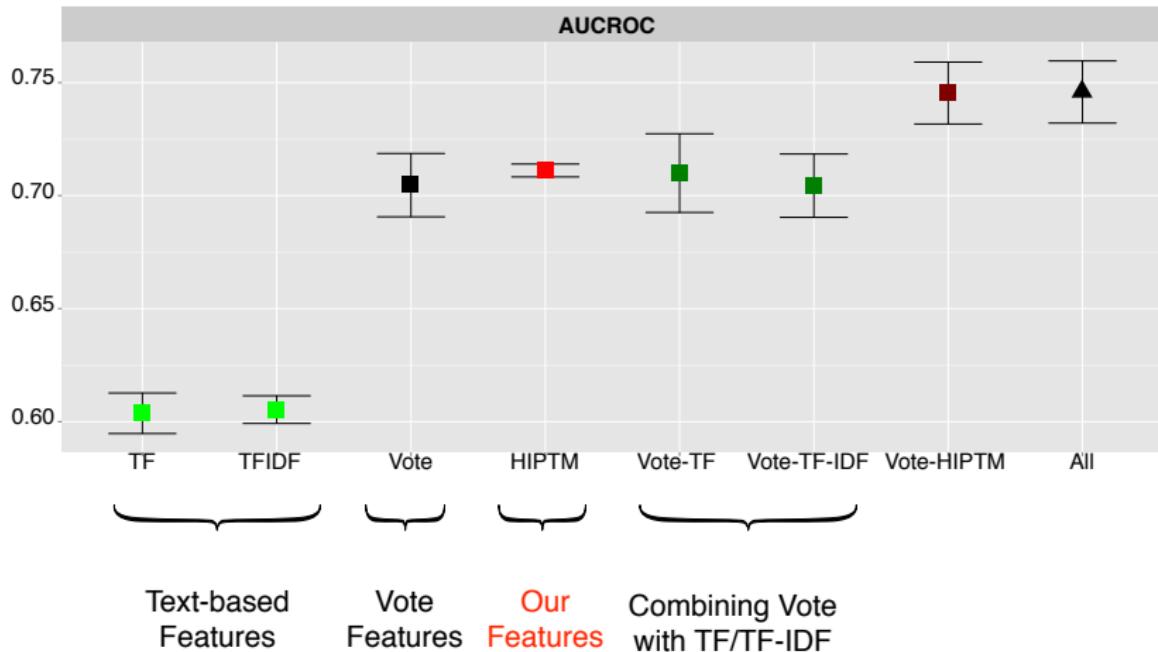
Tea Party Caucus Membership Prediction: Votes & Text



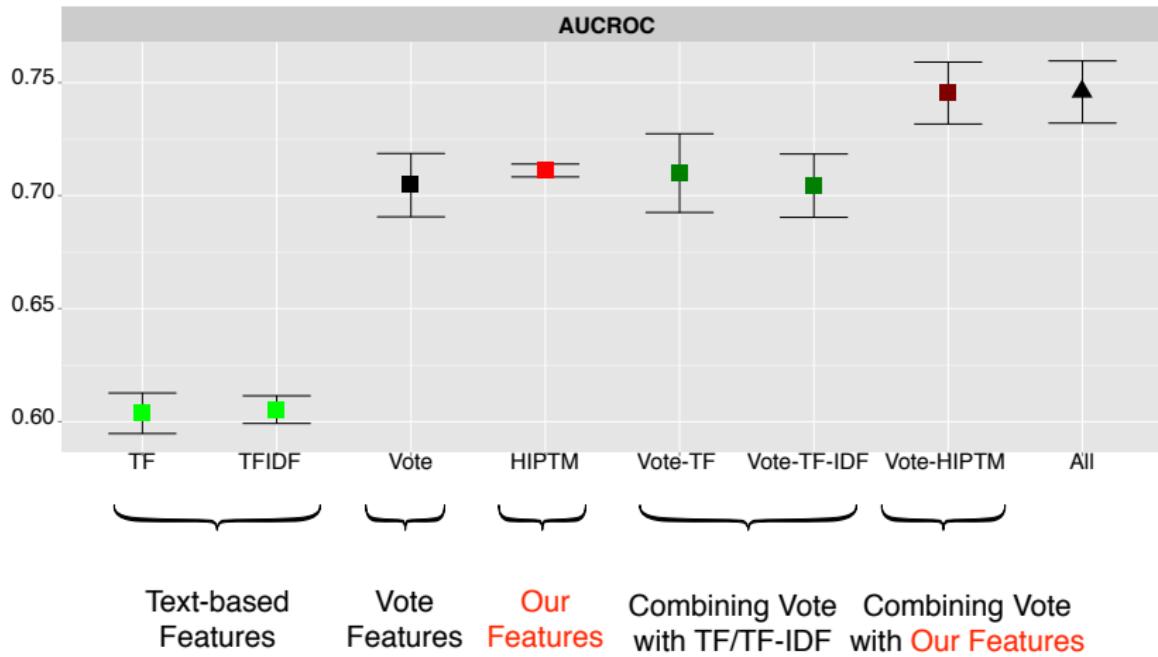
Tea Party Caucus Membership Prediction: Votes & Text



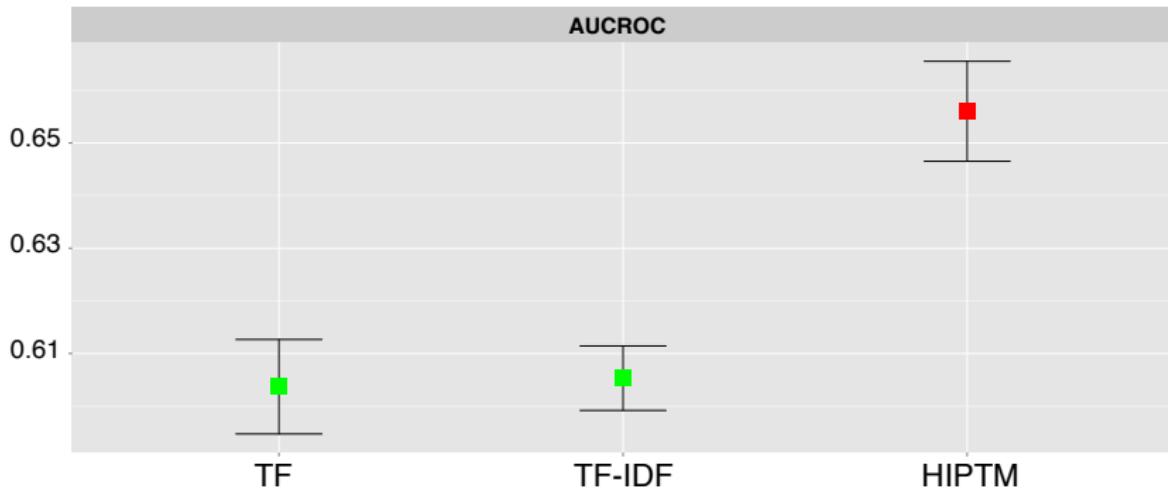
Tea Party Caucus Membership Prediction: Votes & Text



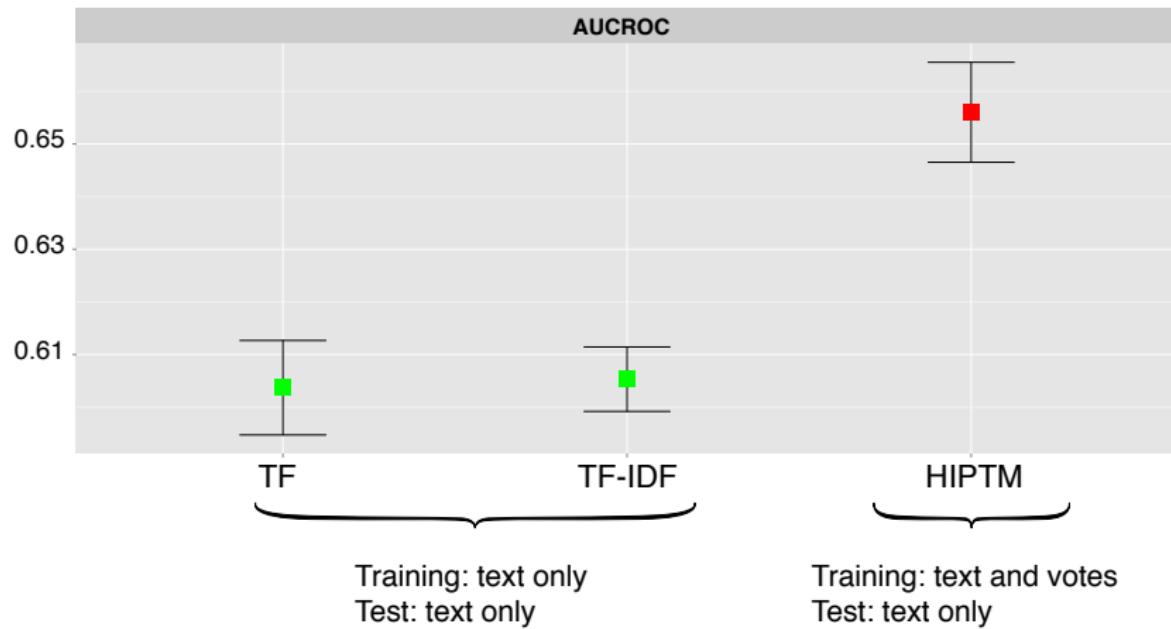
Tea Party Caucus Membership Prediction: Votes & Text



Tea Party Caucus Membership Prediction: Text Only



Tea Party Caucus Membership Prediction: Text Only



Outline

One-dimensional Ideal Points



One-dimensional Ideal Points



One-dimensional Ideal Points



- **Alexander and Crenshaw's votes** only agree with Freedom Works 48% and 50% respectively
- Both voted for raising the debt ceiling and are listed as “traitor”

John T. Reed on Headline News

points and perspectives not offered elsewhere

| House Tea Party Caucus members | how they voted on debt ceiling increase |
|---------------------------------------|---|
| Sandy Adams, Florida | traitor |
| Robert Aderholt, Alabama | traitor |
| Todd Akin, Missouri | no |
| Rodney Alexander, Louisiana | traitor |
| Michele Bachmann, Minnesota, Chairman | no |
| Rob Bishop, Utah | no |
| Ander Crenshaw, Florida | traitor |
| Michael C. Burgess, Texas | traitor |

One-dimensional Ideal Points

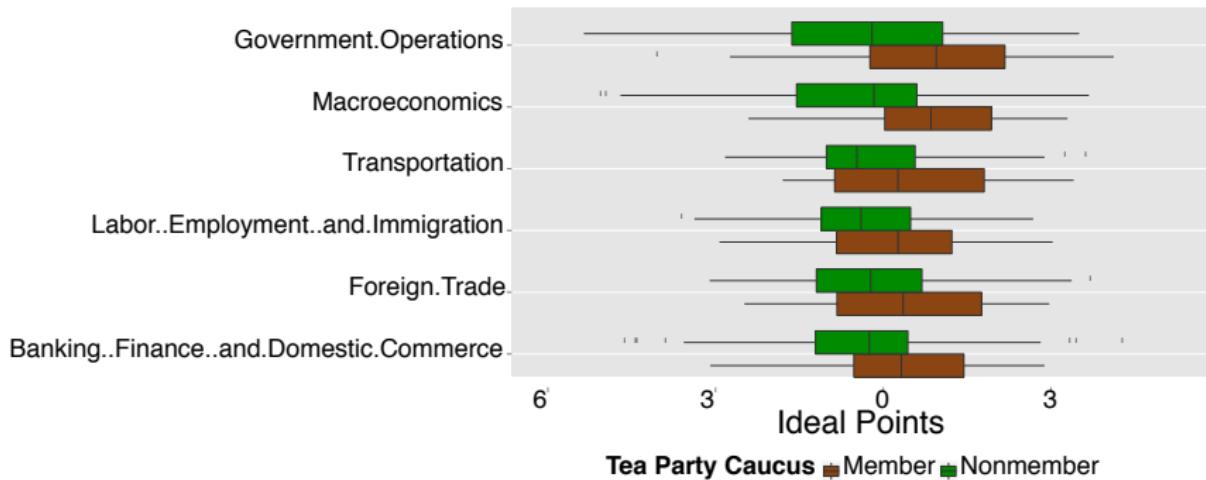
- **Flake** and **Amash** didn't self-identify as members of the Tea Party Caucus but have been endorsed by other Tea Party organizations

NEW REPUBLIC

"Some 46 House members and six senators had been [Tea Party] . . . In addition, there were about 18 other House members like Trey Gowdy, Mark Meadows, and **Justin Amash**, and several senators, including **Jeff Flake** and Pat Toomey, who owed their election to support from the Tea Party and its Washington allies."



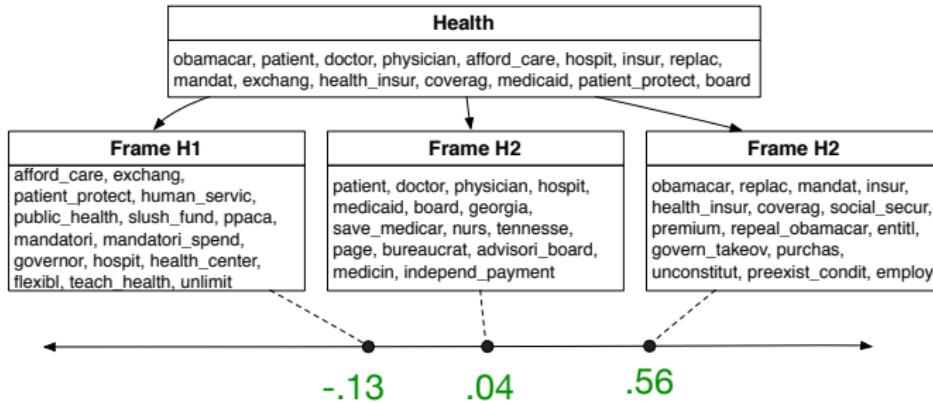
Multi-dimensional Ideal Points



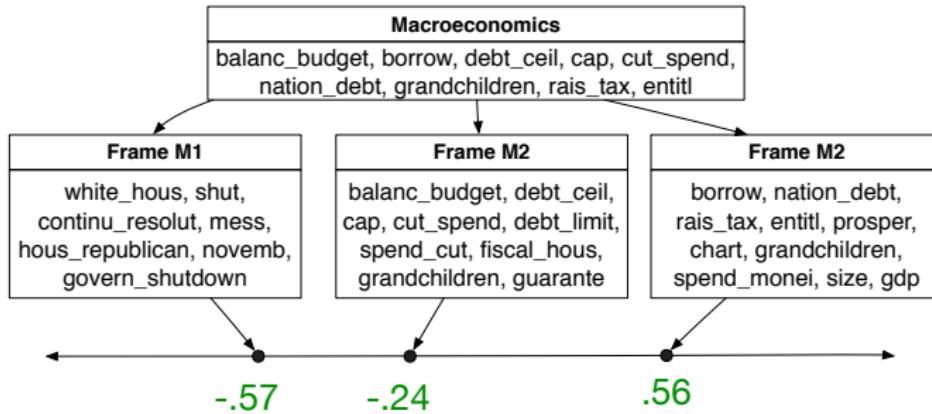
Freedom Works' key votes on most highly polarized dimensions are about government spending

Outline

Framing Healthcare

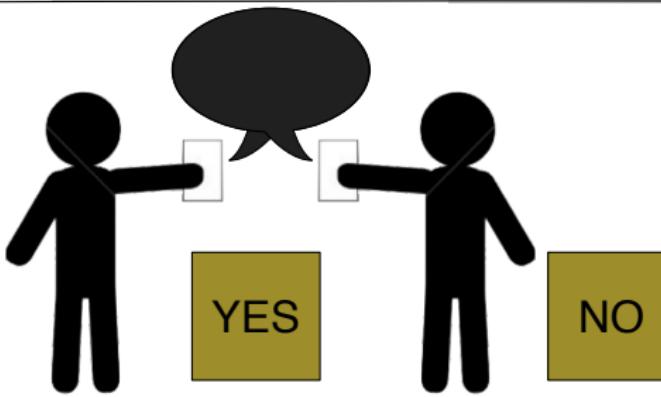


Framing Macroeconomics



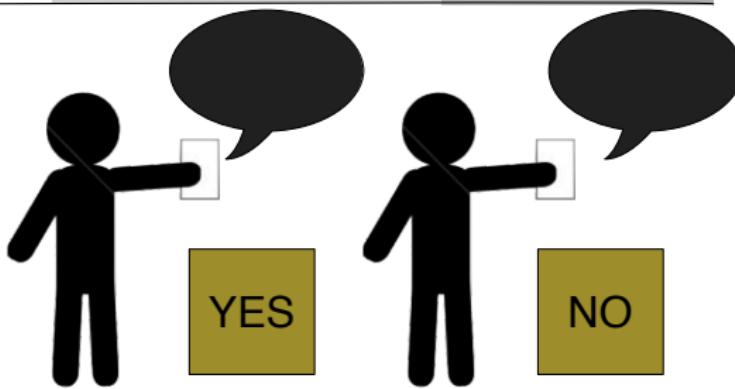
Polarization

| Ideal Point Distributions | | Not | Polarized |
|------------------------------|-----------|--|---------------------------------------|
| Distribution of Issue Frames | Not | Civil Rights, Minority Issues, Civil Liberties | Banking and Finance; Transportation |
| | Polarized | Health; Public Lands and Water Management | Macroeconomics; Government Operations |



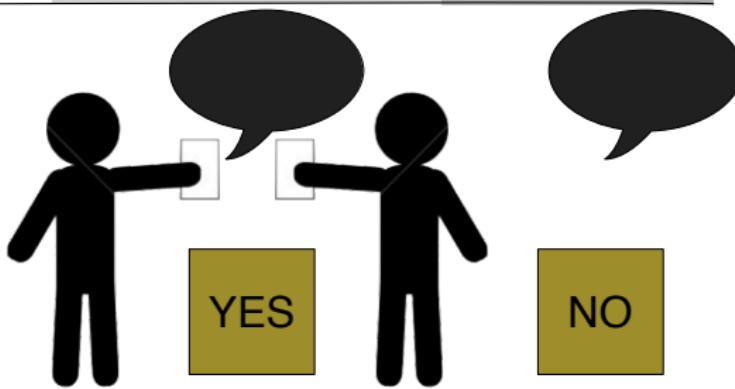
Polarization

| Ideal Point Distributions | | Not | Polarized |
|------------------------------|-----------|--|--|
| Distribution of Issue Frames | Not | Civil Rights, Minority Issues, Civil Liberties | Banking and Finance Transportation |
| | Polarized | Health; Public Lands and Water Management | Macroeconomics; Government Operations |



Polarization

| Ideal Point Distributions | | Not | Polarized |
|------------------------------|-----------|--|--|
| Distribution of Issue Frames | Not | Civil Rights, Minority Issues, Civil Liberties | Banking and Finance Transportation |
| | Polarized | Health; Public Lands and Water Management | Macroeconomics; Government Operations |



Polarization

| Ideal Point Distributions | | Not | Polarized |
|------------------------------|-----------|--|--|
| Distribution of Issue Frames | Not | Civil Rights, Minority Issues, Civil Liberties | Banking and Finance Transportation Macroeconomics, Government Operations |
| | Polarized | Health; Public Lands and Water Management | |

