Text-Based Ideal Points

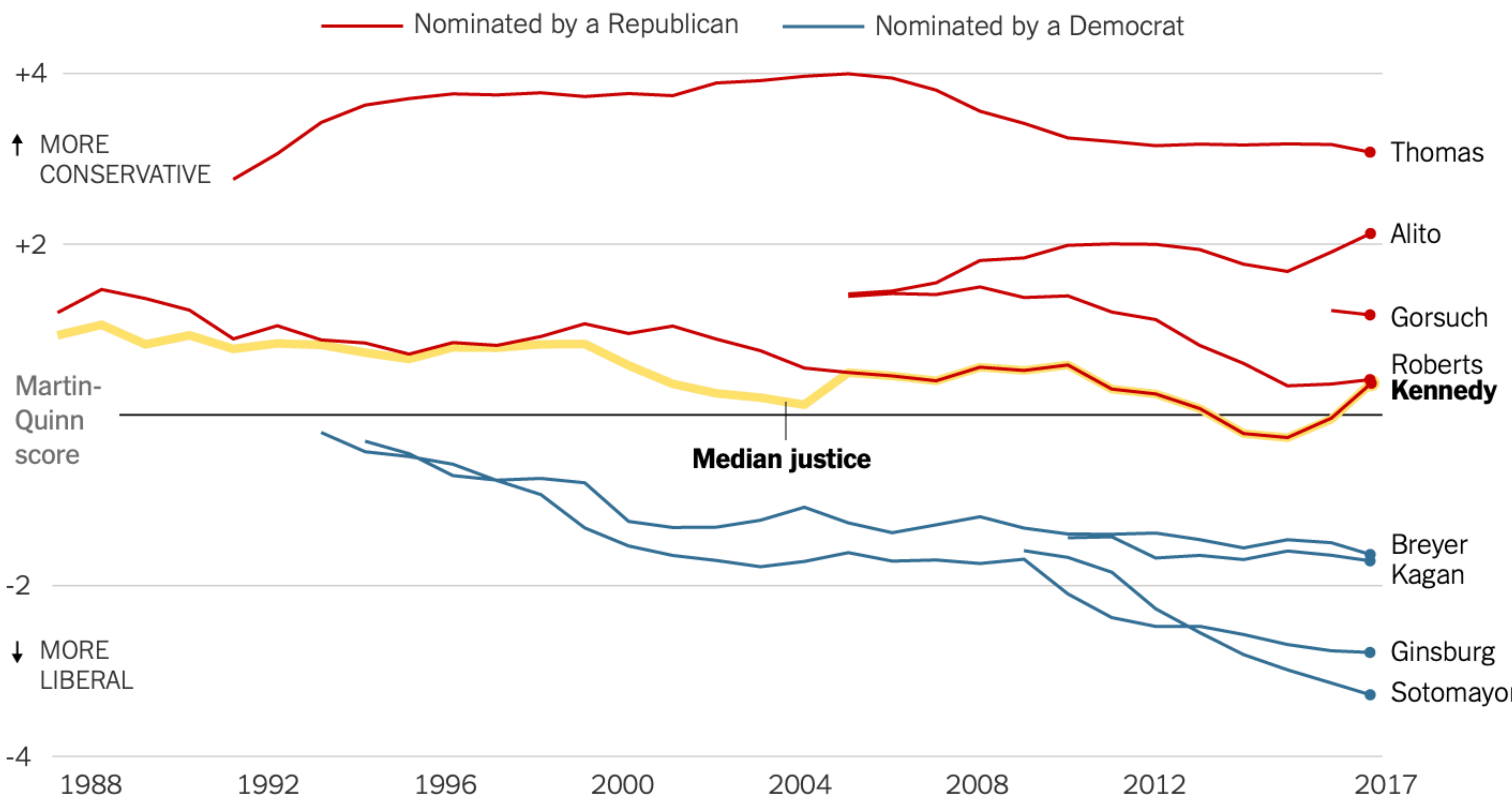
Keyon Vafa
Columbia University

Joint work with:

Suresh Naidu
Columbia University

David Blei
Columbia University

Ideal Points



Source: Lee Epstein, Washington University in St. Louis; and Andrew D. Martin and Kevin Quinn, University of Michigan

Image Source: New York Times

Ideal Points

Bayesian Ideal Points

- Probabilistic method to measure political positions of legislators
- Based solely on voting record

binary vote

legislator ideal point

$$\underbrace{v_{ij}}_{\text{binary vote}} \sim \text{Bern}(\sigma(\underbrace{\beta_j}_{\text{bill popularity}} + \overbrace{x_i}^{\text{legislator ideal point}}\underbrace{\eta_j}_{\text{bill polarity}}))$$

bill popularity

bill polarity

Vote Ideal Points

Analyze **votes** on shared **bills** to infer political positions.

Limitations:

- Cannot compare groups who do not vote together (e.g. judges on different courts).

- Votes on decisions must be available (e.g. cannot extend to presidential candidates).

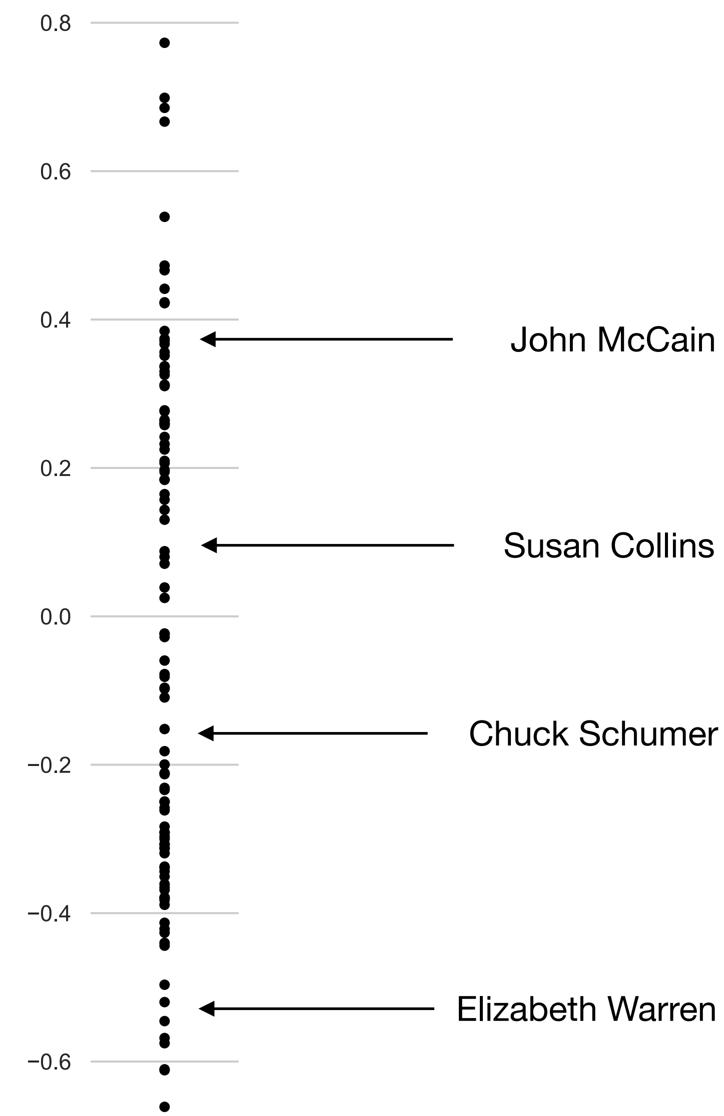
Solution: Text-based ideal points!

- Analyze language of speeches to infer political preferences.

Vote-Based Ideal Points

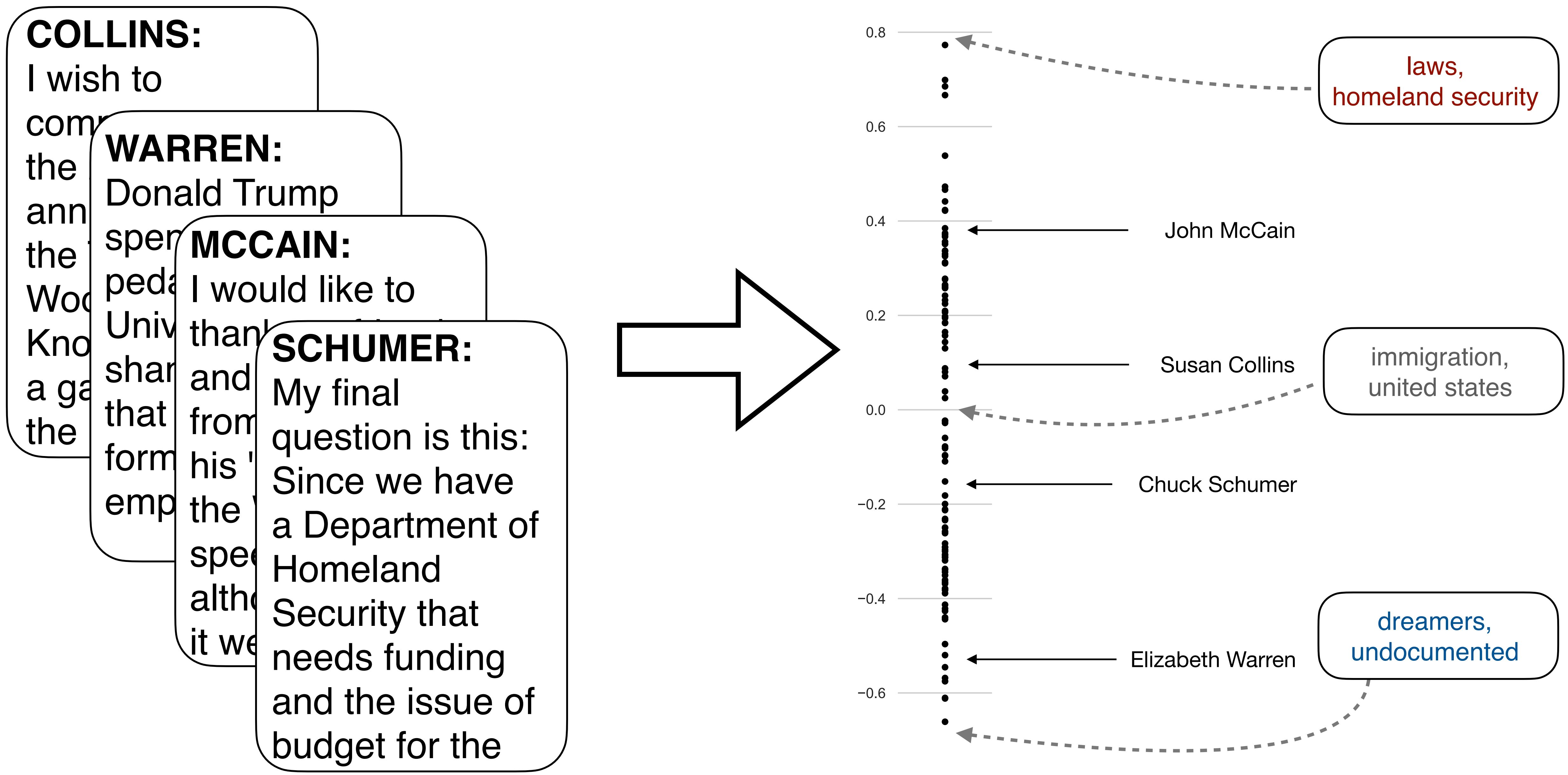
	1	2	3	4	5
Susan Collins	Y	N	Y	Y	Y
Elizabeth Warren	N	Y	N	Y	N
John McCain	Y	Y	Y	N	Y
…					
Chuck Schumer	N	Y	N	Y	N

IN: Voting Record



OUT: Ideal Points

Text-Based Ideal Points



IN: Speeches

OUT: Ideal Points + Ideological Topics

Existing Methods

Existing methods for inferring political positions from text either:

- Use party labels
- Combine text with voting records
- Use hand-labeled political text
- Require grouping of texts into single issues

Text-Based Ideal Points

The Text-Based Ideal Point Model (TBIP) is **completely unsupervised**:

- Does not require party labels, voting records, hand-labeled political text, or grouping of text into single issues

Advantages of being unsupervised:

- Applicable to unlabeled political discourse
- Does not force hard membership into binary groups
- Does not depend on subjectivity of coders

Political Framing

Entman's definition of framing (Entman, 1993):

> *"[Selecting] some aspects of a perceived reality and [making] them more salient in a communicating text, in such a way as to promote problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described."*

Political framing: When discussing a topic, word choice is affected by political message.

Frames for abortion (Boydstun et al., 2014; Johnson et al., 2017):

- "life" and "unborn" invoke morality and religion
- "choice" and "freedom" invoke constitutionality and personal liberty

Text-Based Ideal Points

Vote-based ideal points:

- Inferred by **vote** differences on shared **bills**.

Text-based ideal points:

- Inferred by **word choice** differences on shared **topics**.

Model

The TBIP is based on **Poisson factorization**:

$$y_{dv} \sim \text{Pois}\left(\sum_k \theta_{dk}\beta_{kv}\right)$$

word counts

document intensities

topics

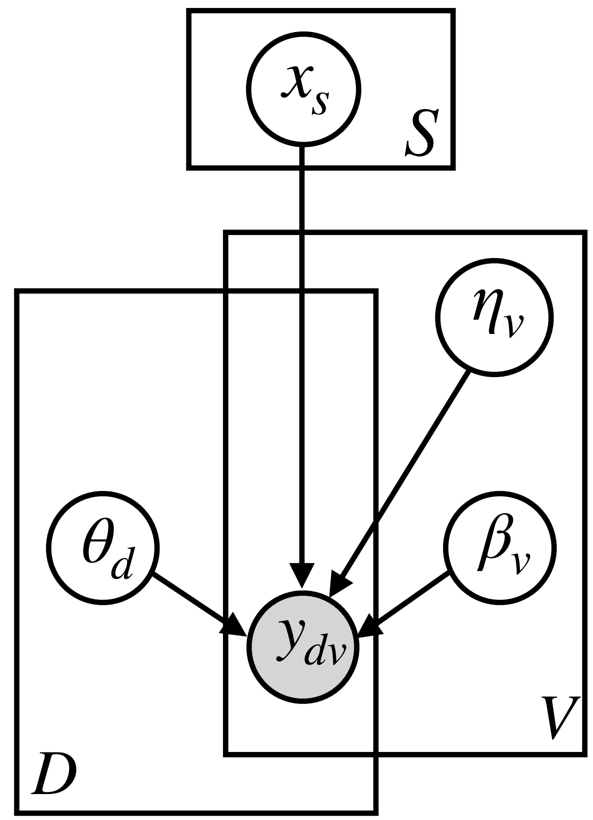
We add two terms to the Poisson factorization log-likelihood:

"ideological" topics

$$y_{dv} \sim \text{Pois}\left(\sum_k \theta_{dk}\beta_{kv}\exp\{x_{a_d}\eta_{kv}\}\right)$$

ideal point for author
of document *d*

Inference



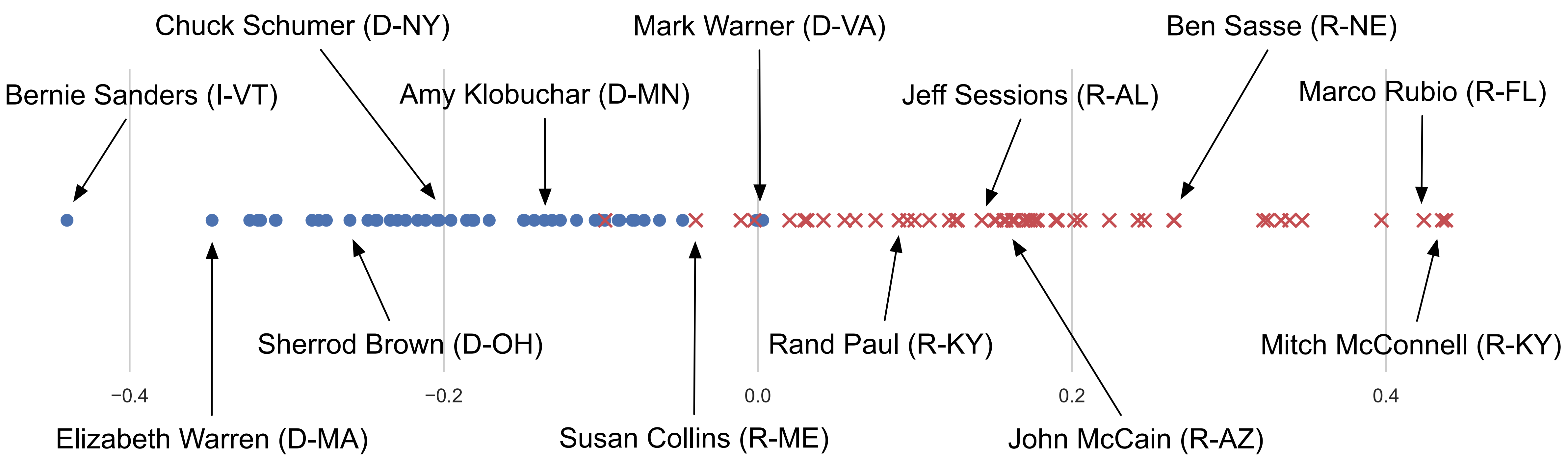
Posterior distribution for latent parameters (θ, β, η, x) is approximated with variational inference.

TensorFlow and PyTorch implementations are available at:

`github.com/keyonvafa/tbip`

U.S. Senate Speeches

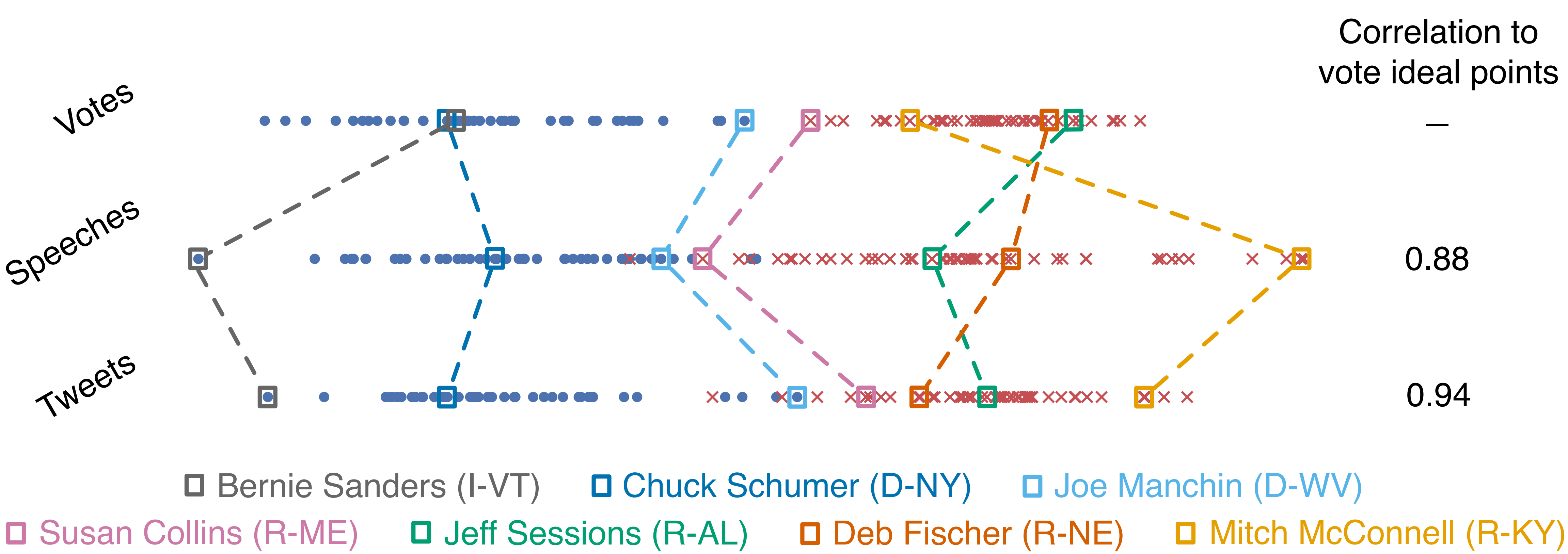
Ideal Points



U.S. Senator Tweets

209,779 tweets from senators between 2015-2017

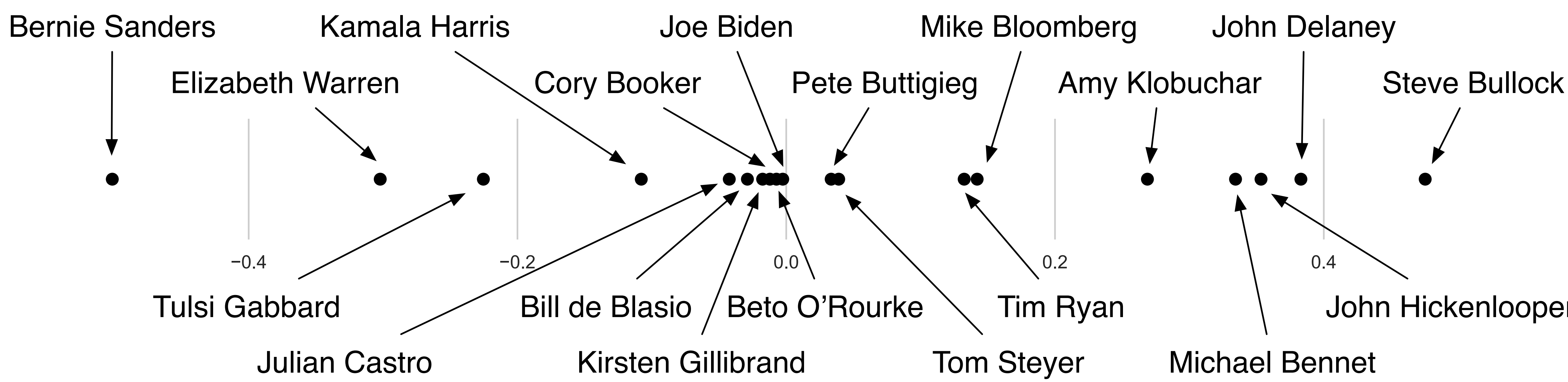
Votes vs Speeches vs Tweets



2020 Democratic Presidential Candidate Tweets

45,927 tweets from 19 candidates between 2019-2020

2020 Democratic Candidates



2020 Democratic Candidates

more progressive

#medicareforall, insurance companies, profit, health care

health care, plan, medicare, americans, care, access

more moderate

healthcare, universal healthcare, public option, plan

more progressive

green new deal, fossil fuel industry, fossil fuel, planet, pass

climate change, climate, climate crises, plan, planet, crisis

more moderate

solutions, technology, carbon tax, climate change, challenges

Comparisons

Other methods: Wordfish (Slapin and Proksch, 2008) and
Wordshoal (Lauderdale and Herzog, 2016)

Evaluate each ideal point method by measuring correlation and rank correlation to vote ideal points.

	Speeches 111		**Speeches 112**		**Speeches 113**		**Tweets 114**	
	Corr.	SRC	Corr.	SRC	Corr.	SRC	Corr.	SRC
WORDFISH	0.47	0.45	0.52	0.53	0.69	0.64	0.87	0.80
WORDSHOAL	0.61	0.64	0.60	0.56	0.45	0.44	—	—
TBIP	**0.79**	**0.73**	**0.86**	**0.85**	**0.87**	**0.84**	**0.94**	**0.84**

Recap

We develop an unsupervised model to learn ideal points and ideological topics solely from text.

Text-based ideal points can be used to learn political preferences for non-voting entities (e.g. presidential candidates).

We use an efficient variational inference algorithm to apply the model to large datasets.

All code (including Tensorflow and PyTorch implementations) available at:

`www.github.com/keyonvafa/tbip`

Thank you!

References

- Boydstun, A. E., Card, D., Gross, J., Resnick, P., and Smith, N. A. (2014). Tracking the development of media frames within and across policy issues.

- Lewis, J. B., Poole, K. T., Rosenthal, H., Boche, A., Rudkin, A. and Sonnet, L. (2020). Voteview: Congressional roll-call votes database.

- Entman, R. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication.*

- Gentzkow, M., Shapiro, J. M. and Taddy, M. (2016). Congressional record for the 43rd-114th Congresses: Parsed speeches and phrase counts. *Stanford Libraries* [distributor], https://data.stanford.edu/congress_text

- Gopalan, P., Hofman, J.M. and Blei, D. M. (2013). Scalable recommendation with Poisson factorization. *Proceedings of UAI.*

- Johnson, K., Lee, I. T., and Goldwasser, D. (2017). Ideological phrase indicators for classification of political discourse framing on Twitter. In *Proc. of the Workshop on NLP and Computational Social Science collocated with ACL*.

- Lauderdale, B. E. and Herzog, A. (2016). Measuring political positions from legislative speech. *Political Analysis*.

- Poole, K. T. and Rosenthal, H. (2000). Congress: A political-economic history of roll call voting. *Oxford University Press on Demand*.

- Slapin, J. B. and Proksch, S.-O. (2008). A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*.

- VoxGovFEDERAL (2020). U.S. senators tweets from the 114th Congress.