

# Discovering Issue-Based Voting Groups within the US Senate

Rachel Shorey, Andrew McCallum & Hanna Wallach University of Massachusetts, Amherst

## The United States Senate

- Senate bills are extremely long and complex (The much-discussed health care bill was 2409 pages!)
- Voters often rely on campaign or interest group spin for information about senators
- Policymakers must synthesize content and voting records to find compromisers

## Interest Group Ratings

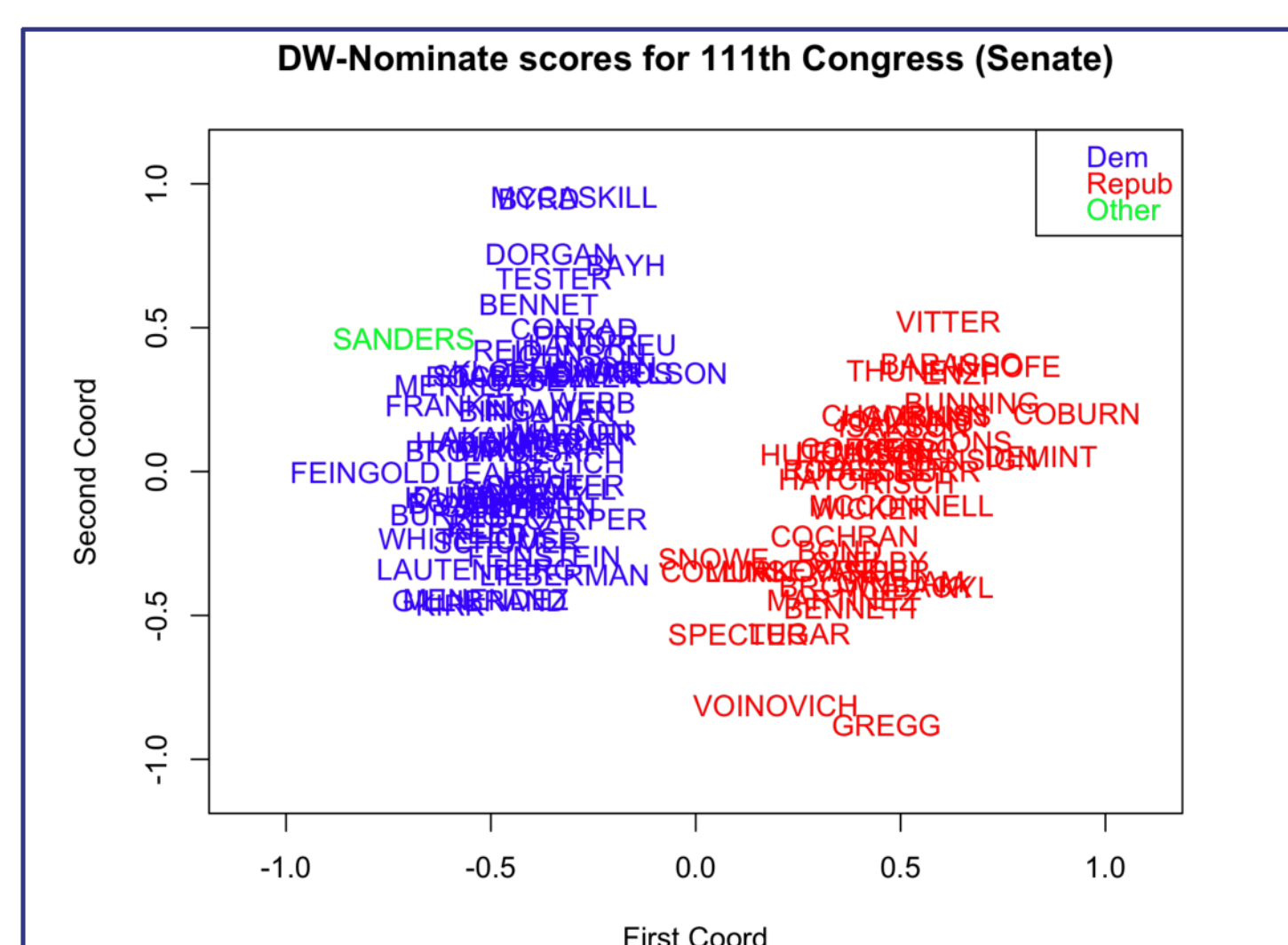
- Interest groups rate senators on a single dimension based on a few votes selected to reflect the group's goals

90th ACLU		AMERICAN CIVIL LIBERTIES UNION		90 YEARS OF PROTECTING YOUR LIBERTY															
Senator	Party	State	Score	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
Akaka, Daniel	Democrat	HI	92%	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
Alexander, Lamar	Republican	TN	8%	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓
Barrasso, John	Republican	WY	8%	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓
Baucus, Max	Democrat	MT	83%	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
Bayh, Evan	Democrat	IN	75%	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
Bejich, Mark	Democrat	AK	92%	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
Bennet, Michael	Democrat	CO	91%	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
Bennett, Robert	Republican	UT	9%	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓
Bingaman, Jeff	Democrat	NM	92%	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
Bond, Christopher	Republican	MO	8%	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓
Boxer, Barbara	Democrat	CA	92%	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
Brown, Scott	Republican	MA	33%	*	✓	*	✓	*	✓	*	✓	*	✓	*	✓	*	✓	*	✓
Brown, Sherrod	Democrat	OH	92%	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
Brownback, Sam	Republican	KS	8%	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓
Bunning, Jim	Republican	KY	8%	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓
Burr, Richard	Republican	NC	8%	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓
Burr, Roland	Democrat	IL	92%	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
Byrd, Robert	Democrat	WV	83%	-	*	✓	*	-	*	✓	*	-	*	✓	*	-	*	✓	*
Cantwell, Maria	Democrat	WA	83%	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
Cardin, Benjamin	Democrat	MD	92%	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗

Copyright 2010 American Civil Liberties Union  
Reprinted with permission of the American Civil Liberties Union <http://www.aclu.org>

## Political Science Analysis

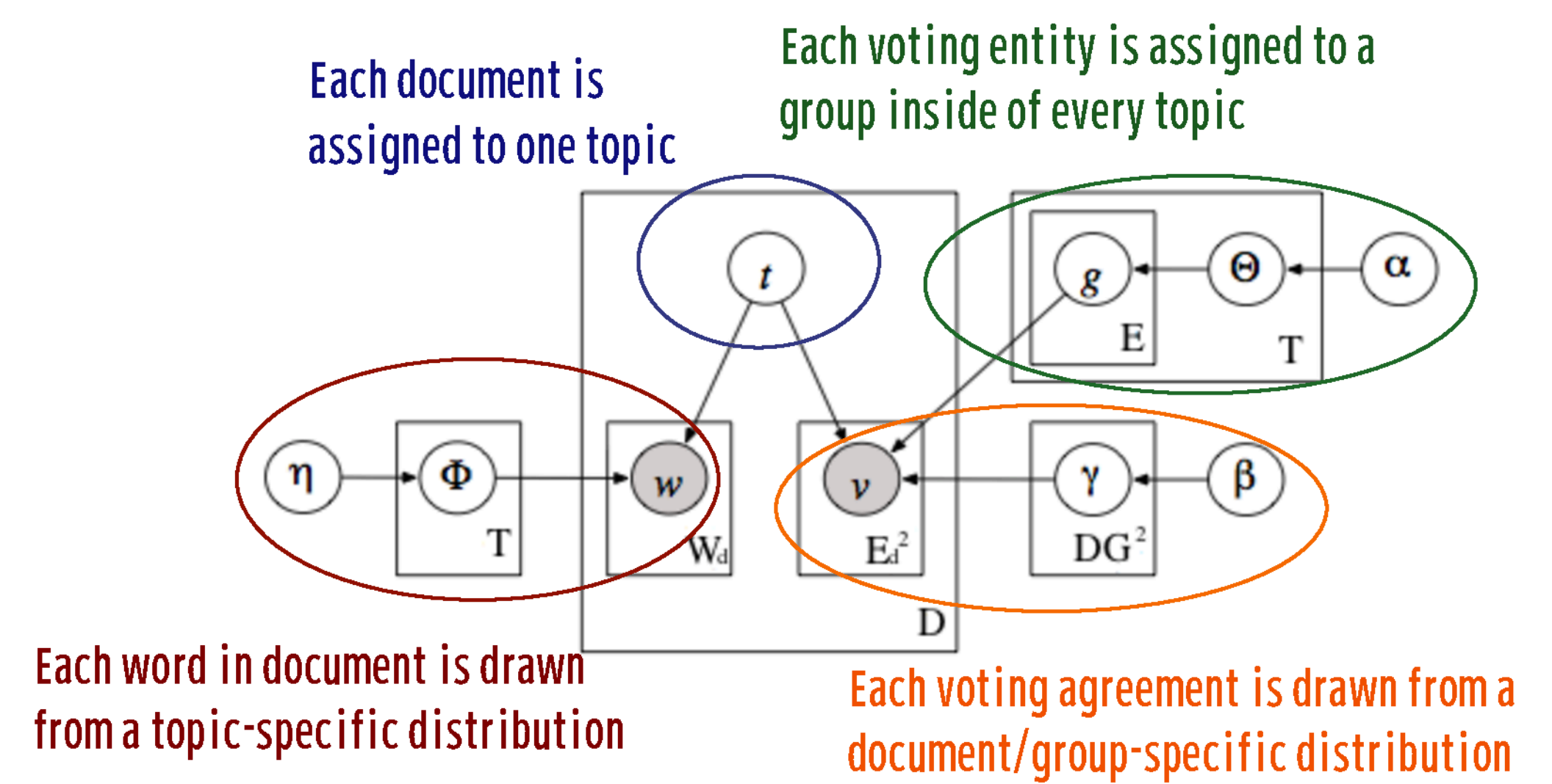
- Political Scientists have created tools for multi-dimensional analysis that rely on voting records but ignore context



Data from DW-NOMINATE Scores with Bootstrapped Standard Error, Carroll et al <http://voteview.com/dwnominate.aspransitair>

## Topic Models

- Unsupervised methods to distill document collections into interpretable topics
- Topic models can be easily extended to account for votes as well as text



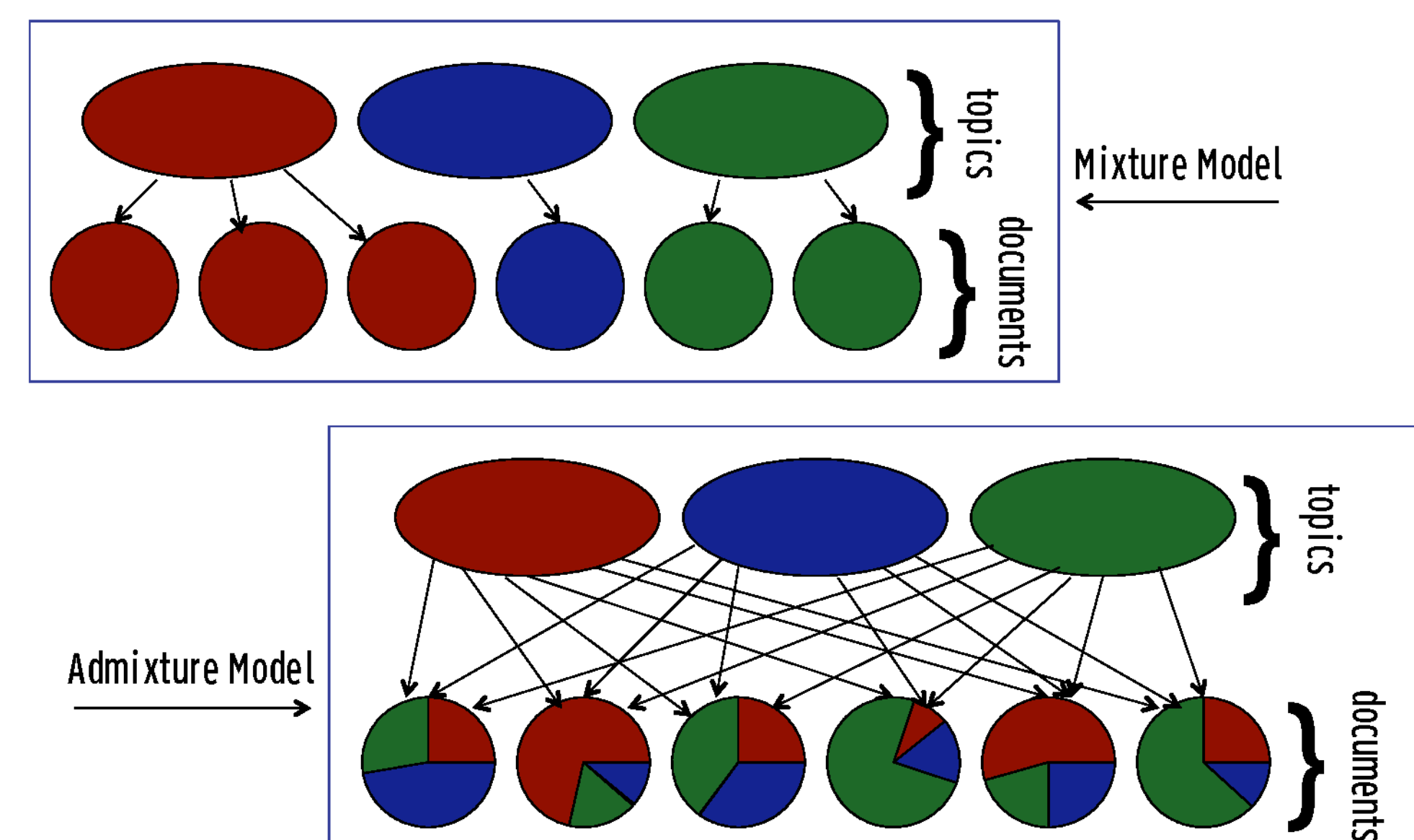
t=topic indicator  
w=word  
g=group indicator  
v=do entities e and f vote the same way on document d?  
T=number of topics  
W=number of words ( $W_d$ =number of words in doc d)  
G=number of groups  
E=number of voting entities words ( $E_d$ =number of entities voting on doc d)  
D=number of documents

McCallum et al [2007] Group-Topic model

- “Group-Topic” model is a mixture model: each document is assigned to just one topic

## Mixtures vs. Admixtures

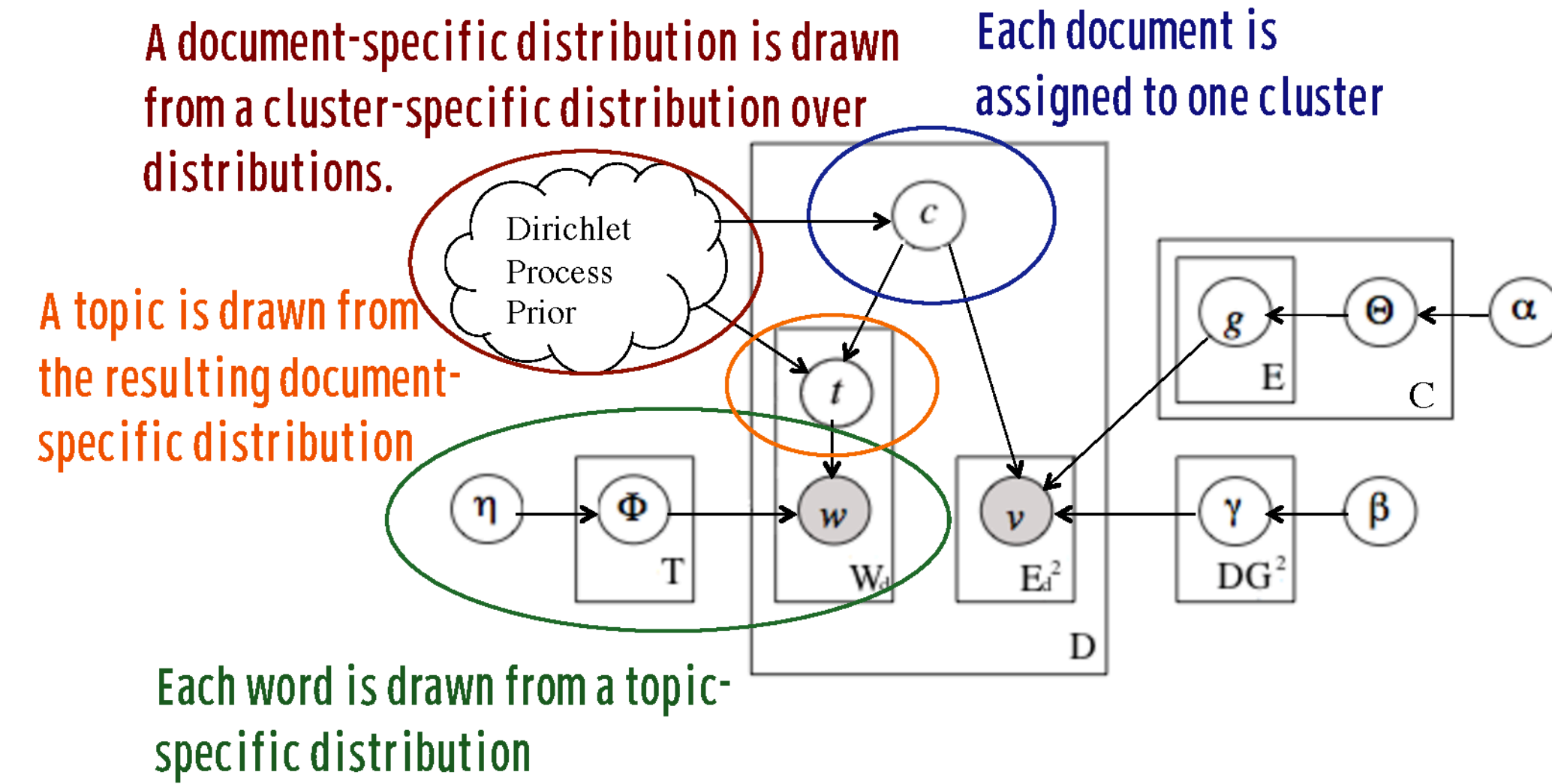
- Most topic models are admixture models, meaning that multiple topics are responsible for the words in each document



- Admixtures can give a more nuanced accounting of the documents in a collection if documents deal with multiple subjects

## Admixture Group Topic

- We propose an admixture version of the Group-Topic model



- Like McCallum et al, voting groups are inferred based on entity vote agreements

- Documents are placed in clusters based on the vote actions and the semantic content

- Each word in the document is assigned to a topic based on a document-and-cluster-specific topic distribution

## Results

- As in McCallum et al, topics are semantically coherent and groups make intuitive sense

### Coherent Topics:

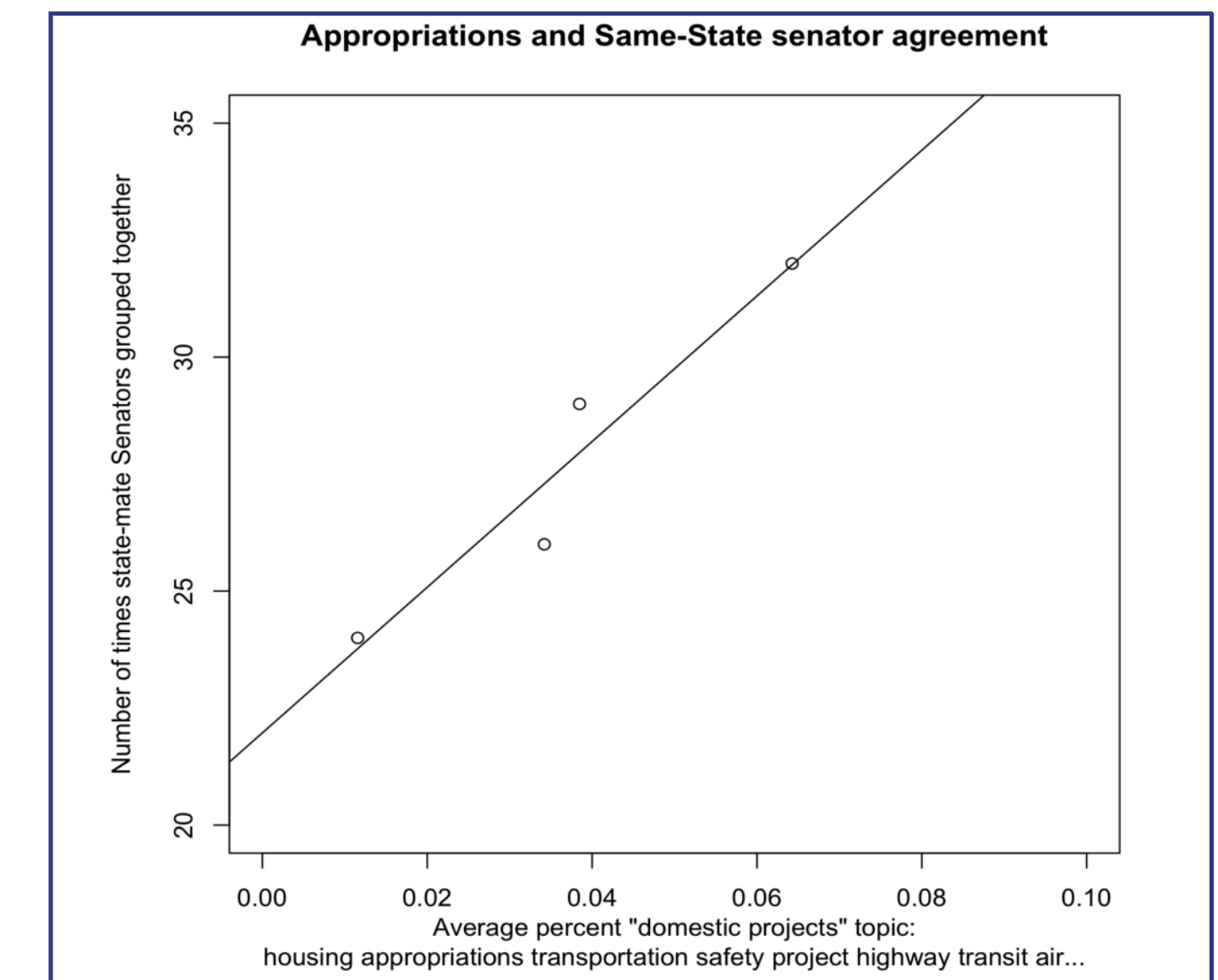
Land Management	Tobacco Regulation	Appropriations/Budgeting
land water wilderness management project system river	marketing tobacco pay risk submit practices financial	project billion summary reductions community improvement transportation

Words commonly found in selected topics, 111th Senate

Intuitive Groups: For a cluster with high percentages of tobacco and regulation topics, most senators are in one group. The exceptions: anti-regulation senators (e.g., Bennett (UT), Coburn (OK)), and tobacco-staters (e.g., Graham (SC), Hagan (NC))

### Results Continued

- What do admixtures do?  
•Since documents draw their content from multiple topics, we can investigate correlations between topic percentage and group behavior. For example:



Clusters and topics from 111th congress, 5 clusters, 6 groups, 30 topics, cluster containing only 1 bill excluded.

## Future Directions

- Automatically infer numbers of groups and topics using Dirichlet Process priors  
•Take the human out of the equation!  
•What issues are the most divisive?  
•Internationally, which legislatures are the most fractured?  
•Predict votes on unseen data  
•Useful for model validation and predicting positions on new bills  
•Include data linking US Senate to House or state legislatures to predict the votes of a newly elected US Senator

## References

McCallum, A., Wang, X., and Mohanty, N., *Joint group and topic discovery from relations and text*. Statistical Network Analysis: Models, Issues and New Directions, Lecture Notes in Computer Science 4503, pp. 28-44, 2007.

Carroll, R., Lewis, J., Lo, J., McCarty, N., Poole, K., and Rosenthal, H. *DW-NOMINATE Scores With Bootstrapped Standard Errors*. <http://voteview.com/downloads.asp>. Accessed 11/25/10.

This work was supported in part by the Center for Intelligent Information Retrieval and in part by The CIA, the NSA and NSF under NSF grant #115-0326249. Any opinions, findings and conclusions or recommendations expressed in this material are the authors' and do not necessarily reflect those of the sponsor.