Latent Community Detection for Predicting Legislative Roll Call Votes

Elijahu Ben-Michael, Runjing Liu, Jake Soloff

Department of Statistics, UC Berkeley

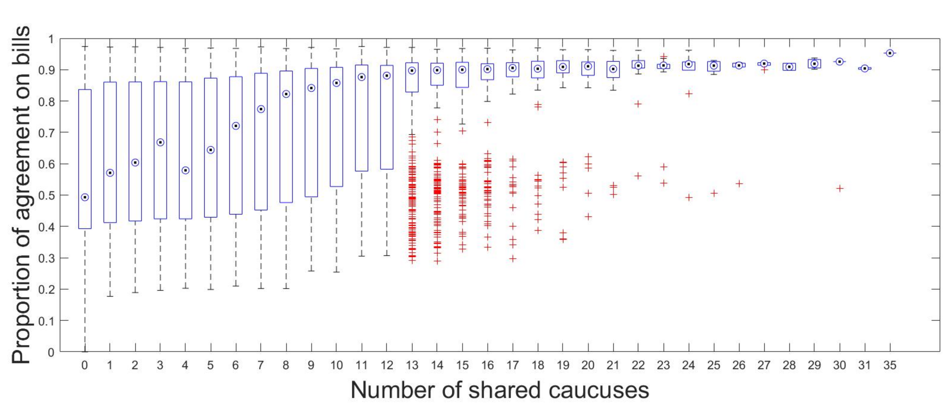
Objectives

We analyze voting data in the House of Representatives in the 110th Congress (2007-2009). We extend a traditional ideal point model to incorporate caucus membership data via a stochastic block model. In doing so, we aim to

- Use caucus membership data to infer latent communities among members
- Exploit this community structure to inform estimates for each representative's ideal point
- Predict a representative's voting behavior

Motivation

We chose to model caucus membership because we found that caucus memberships influence legislative behavior. For example, the more caucuses two legislators share, the more likely they are to vote the same way on a bill.



Moreover, representing interactions among members of the House using an undirected graphical model, we found that subgraphs corresponding to caucuses were denser than the graph of the whole House.

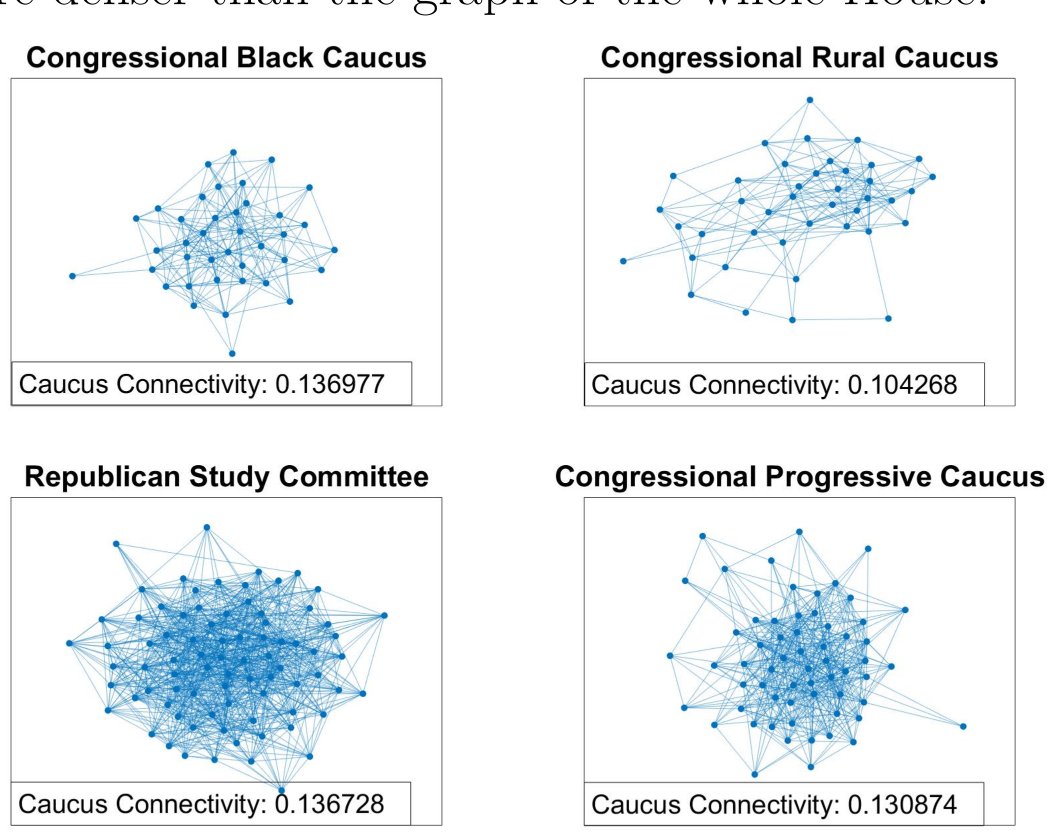


Figure: The connectivity is measured by the fraction of total edges present; connectivity of the whole House is 0.064

Model

All ideal points a_d, b_d, x_u live in $\mathbb{R}^S$, where S is a free parameter and we assume K latent communities

- Sample community proportions $\pi \sim \text{Dir}(\gamma 1_K)$ and each community's ideal point $\nu_k \sim \mathcal{N}(\varpi, \sigma_\nu^2)$.
- Draw representative u's community $M_u \overset{\text{iid}}{\sim} \text{Cat}(\pi)$ and ideal point $x_u \mid M_u = k, \nu \sim \mathcal{N}(\nu_k, \sigma_x^2)$.
- Draw coexpression rates $P_{kl} \overset{\text{iid}}{\sim} \text{Gamma}(\lambda_0, \lambda_1)$.
- Observe the number of common caucuses $R_{uv} \mid P, M_u = k, M_v = l \sim \text{Poisson}(P_{kl})$.
- Draw a discrimination $a_d \sim \mathcal{N}(\eta_a, \sigma_d^2)$ and a difficulty $b_d \sim \mathcal{N}(\eta_b, \sigma_d^2)$ for each bill d.
- Observe the votes $V_{ud} \mid x_u, a_d, b_d \sim \text{Bern}(\sigma(a_d \cdot (x_u - b_d)))$.

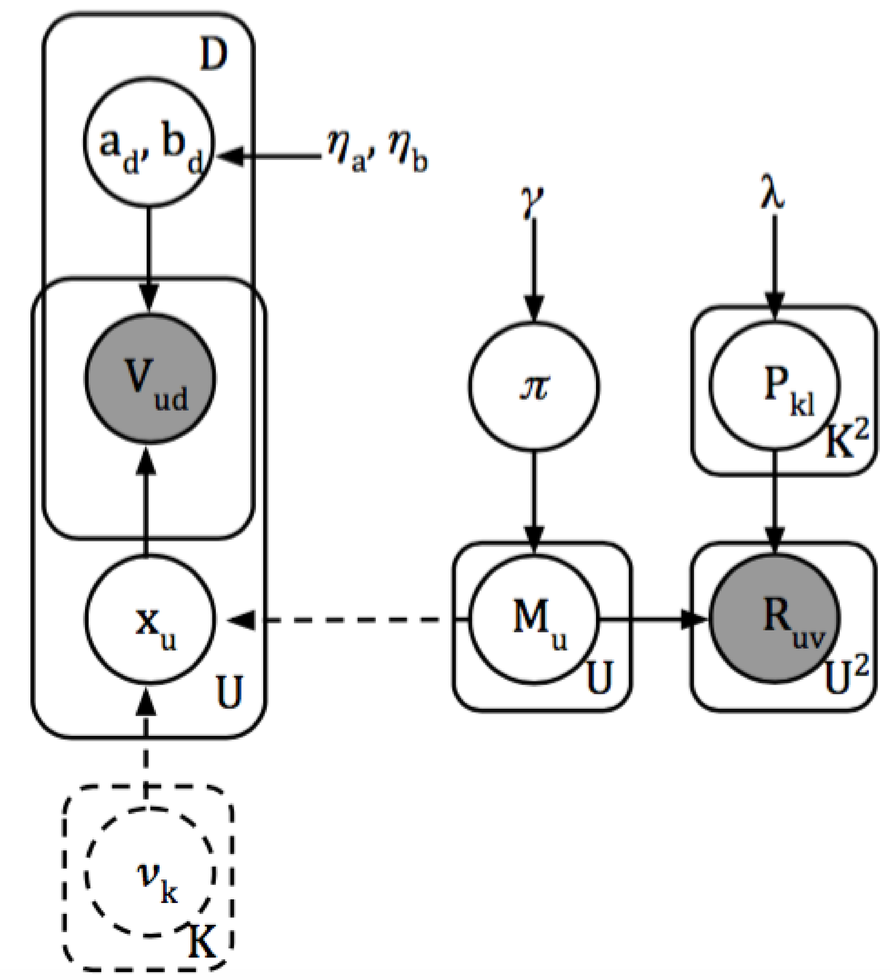
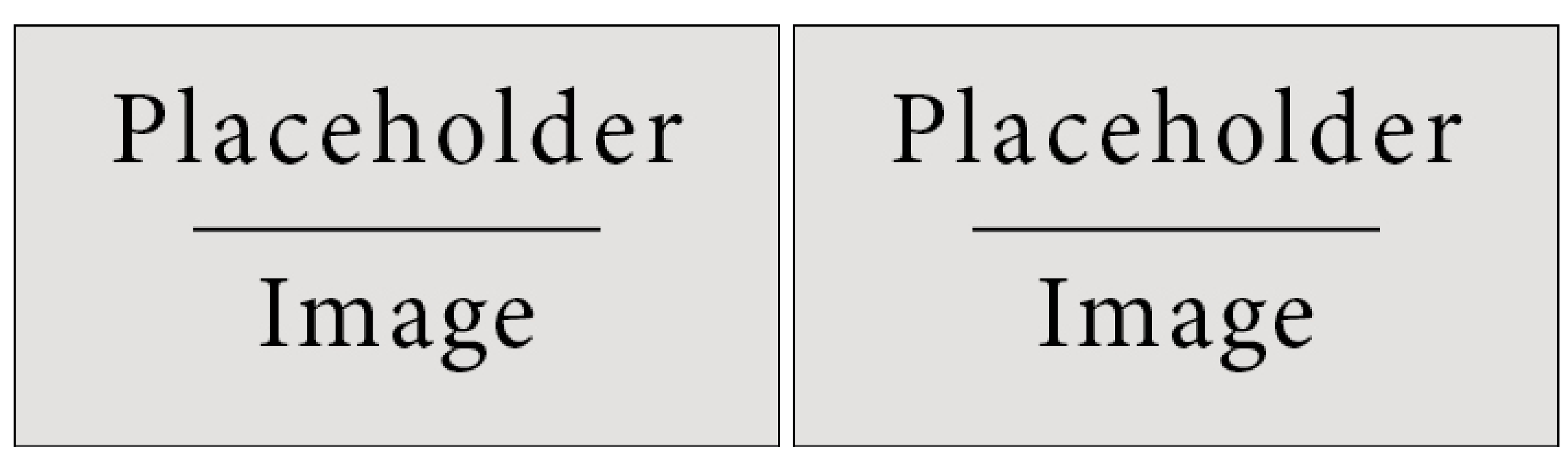


Figure: Graphical model depiction of Latent Community Ideal Point Model (LC-IPM)

Results

Model	K	S	Acc	AUC
Logistic Reg	-	-	78.340295	85.10542
IPM	-	1	94.768623	98.418436
IPM	-	2	95.451139	98.997592
LC-IPM	1	1	94.759033	98.439207
LC-IPM	1	2	95.463571	99.011618
LC-IPM	2	1	94.749798	98.420599
LC-IPM	2	2	95.441016	99.002679
LC-IPM	10	1	94.764538	98.421727
LC-IPM	10	2	95.443324	99.018565



(a) (b)

Figure: figure caption

Nunc tempus venenatis facilisis. Curabitur suscipit consequat eros non porttitor. Sed a massa dolor, id ornare enim:

Variational Inference

Upon observing vote behavior $V = (V_{ud})$ and caucuses $R = (R_{uv})$, computing the posterior distribution of the latent variables given the observations is intractable. We employ *mean field variational inference*, finding the distribution q which factorizes over the latent variables closest in KL divergence to the posterior. Since the graphical model for LC-IPM joins SBM and IPM via one edge, this composabiity and the mean field factorization implies that variational updates for π, P, a_d, and b_d do not change from their respective updates in SBM and IPM, and we exploited this modularity in our implementation.

Conclusion

Nunc tempus venenatis facilisis. **Curabitur suscipit** consequat eros non porttitor. Sed a massa dolor, id ornare enim. Fusce quis massa dictum tortor **tincidunt mattis**. Donec quam est, lobortis quis pretium at, laoreet scelerisque lacus. Nam quis odio enim, in molestie libero. Vivamus cursus mi at *nulla elementum sollicitudin*.

Future Directions

By incorporating caucus data, we were able to place more informative priors on the representatives ideal points. On the other hand, we may also wish to place similarly informed priors on a bill's difficulty and discrimination. Following the example of Gerrish and Blei 2011, one may apply *supervised topic modeling* and infer the latent topics in a bill from a bill's text; these latent topics then generate a bill's difficulty and discrimination.

References

[1] Wainwright, M. J. & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning.*

[2] Gerrish, S. M. & Blei, D. M. (2011). Predicting legislative roll calls from text. *Proceedings of the 28th International Conference on Machine Learning.*

[3] Blei, D. M., Kucukelbir, A. and McAuliffe, J. D. (2016). Variational inference: a review for statisticians. *arXiv:1601.00670.*