
Latent Community Detection for Modeling Legislative Roll Call Votes

Eli Ben-Michael, Runjing Liu, Jake Soloff

Department of Statistics, UC Berkeley

{ebenmichael, runjing_liu, jake_soloff}@berkeley.edu

Abstract

We consider the problem of predicting and modeling legislative roll call votes. Quantitative political scientists have traditionally used ideal point models for this analysis. Based on empirical evidence suggesting the importance of caucus membership in voting behavior, we develop the latent community ideal point model (LC-IPM). This hierarchical Bayesian model extends the ideal point model to incorporate latent community structure. We derive an approximate posterior inference algorithm using variational methods and apply our inference procedure to roll call votes and caucus co-membership data from the U.S. House of Representatives of the 110th Congress. We demonstrate that LC-IPM simultaneously captures community structure and revealed preferences.

1 Introduction

Voting records of legislators are commonly analyzed by political scientists to examine relationships between legislator political leanings, institutional structures, and legislative outcomes [4]. For example, even simple dimensionality reduction techniques on voting data in the US House of Representatives uncover the political characteristics of individual legislators such as party affiliation (Figure 1).

To capture further patterns, quantitative social scientists often use *ideal point models*. In ideal point models, legislators and bills are presumed to lie in a latent “ideological space,” where the probability of a “yea” or “nay” response is a function of the bill’s position and the legislator’s position. The legislator’s position is called an “ideal point” because his or her utility decreases as a bill’s position deviates from this point.

These ideal points enable us to characterize legislators quantitatively. The distribution of ideal points reveals clusters of legislators corresponding, for example, to party lines, region, or caucus membership; furthermore, the distance between two ideal points or two clusters of ideal points can be used as a measure of political division. By visualizing policy preferences along a spectrum, interest groups are able to produce “ratings” of legislators according to their leanings on a certain policy [4]. Ideal point models are also applied to the U.S. Supreme Court to create such ratings [7].

In this paper, we use roll call vote data from the House of Representatives of the 110th Congress (2007-2008) to estimate ideal points and predict voting behavior. In particular, we modify the Bayesian ideal point model proposed in [4]. In the Bayesian ideal point model, the ideal points are assumed to be independent; however, we propose to exploit the interactions among representatives rather than model ideal points as independent.

To take into account these interactions, we posit that representatives in Congress belong to latent communities, and that these latent communities are manifested in two ways: members of the same community tend to share similar caucuses, and members of the same community tend to have similar ideal points. This connection between ideal points and caucus membership is made explicit using a *stochastic block model*.

By incorporating caucus membership data and connecting it to ideal points via latent communities, we hope to improve estimates of the ideal points. In particular, this may result in better estimates of ideal points for junior representatives who have short voting histories. Our approach of incorporating social network or relational data can be applied to broader collaborative filtering problems.

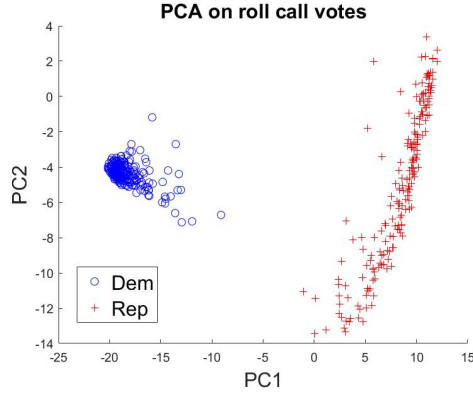


Figure 1: Principal component analysis on roll call vote data in the House of Representatives of the 110th Congress. We computed the eigenvalues and eigenvectors of the $U \times U$ covariance matrix of the $U \times D$ roll call vote matrix ($U = 448$ representatives, $D = 1707$ bills), and projected each representative's voting profile onto the space of the two principal eigenvectors.

1.1 Motivation

We chose to model caucus memberships because initial exploratory data analysis suggested that they are related to a legislator's voting behavior. Figure 2 plots the number of shared caucuses between two representatives against the proportion of bills on which they voted the same way. We see that the more caucuses two members share, the more likely they are to vote the same way.

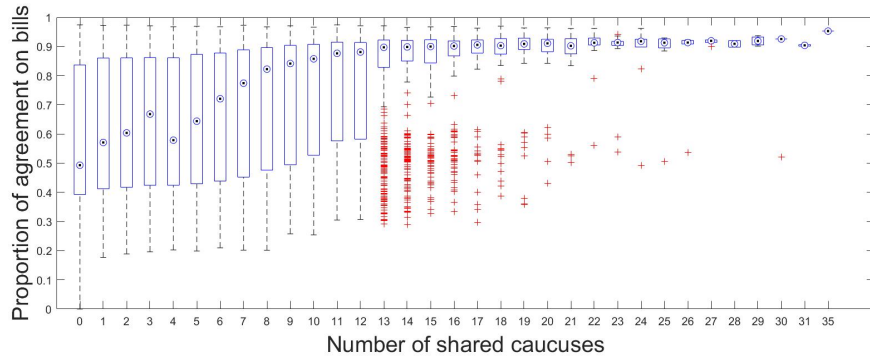


Figure 2: The empirical distribution of agreement on bills versus the number of caucuses two representatives share.

Next we examined the relationship between representatives within several caucuses. We used roll call vote data to infer the graph structure among the representatives in the entire House, assuming pairwise interactions described via an Ising model in which each node denotes a binary variable of a representative voting either yes or no. The edges were inferred using neighborhood selection [6], and in Figure 3 we plot the subgraphs induced by various caucuses. The connectivity (measured by the fraction of total edges present) of the full graph with 448 representatives is 0.064, while the connectivity within the caucus subgraphs is much higher. This suggests that a representative is more likely to be influenced by a member of his or her caucus than another random representative in the House.

This analysis suggests that caucus memberships may at least partly explain whether two representatives will vote in a similar fashion. Therefore, we proceed in this project by utilizing caucus

membership data and connecting them to ideal points using a stochastic block model; specifically, we hope that caucus memberships will inform a latent community structure among the representatives. By exploiting these interactions, we hope to obtain better estimates of ideal points and hence better predictions of roll call votes.

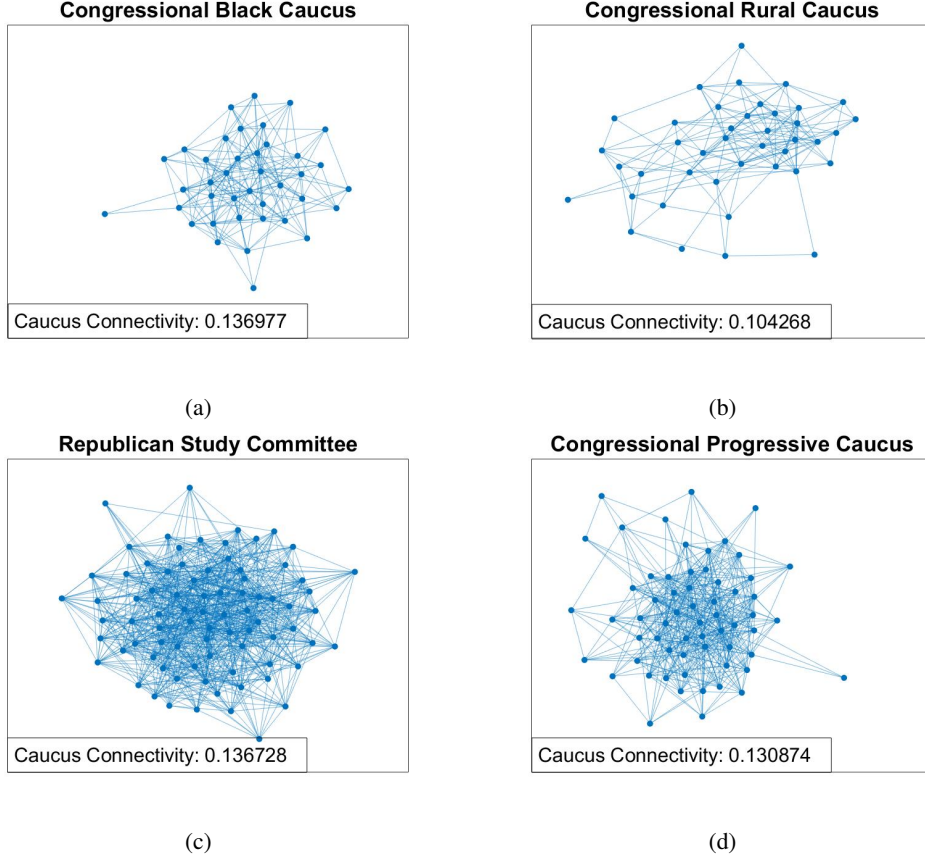


Figure 3: Neighborhood regression on roll call vote data was used to infer an undirected graphical model among all members of the House of Representatives. Shown here are subgraphs induced by various caucuses. The connectivity (fraction of edges present) in the entire House is 0.064

2 Model

2.1 Ideal Point Model

The Bayesian *ideal point model* (IPM) [4] is a basic model for analyzing legislative behavior and revealed preferences with roll call vote data (V_{ud}) for a group of representatives $u \in \{1, \dots, U\}$ voting on a collection of bills $d \in \{1, \dots, D\}$. It assumes that each representative has a latent location $x_u \in \mathbb{R}^S$ called an *ideal point*, and similarly that each bill has two vectors $a_d, b_d \in \mathbb{R}^S$, called the *discrimination* and *difficulty*, respectively. In quantitative political science, the ideal point x_u is often treated as a proxy for one's ideological stance or preference. When $S = 1$, the latent space may be considered a political spectrum [7]. The discrimination a_d quantifies how well votes on this bill separate liberals from conservatives: a bill d is not discriminative ($a_d \approx 0$) if representatives are largely indifferent to it. For positively discriminative bills $a_d > 0$, representatives u such that $x_u > b_d$ are likely to support it. A reasonable likelihood given these quantities is thus

$$V_{ud} \mid x_u, a_d, b_d \sim \text{Bern}(\sigma(a_d \cdot (x_u - b_d))). \quad (1)$$

We place Gaussian priors over each of the latent variables:

$$a_d \sim \mathcal{N}(\eta_a, \sigma_a^2), \quad b_d \sim \mathcal{N}(\eta_b, \sigma_b^2), \quad x_u \sim \mathcal{N}(\nu, \sigma_x^2), \quad (2)$$

where the quantities $\eta_a, \sigma_a^2, \eta_b, \sigma_b^2, \nu, \sigma_x^2$ are fixed hyperparameters.

2.2 Stochastic Block Model

Stochastic block models (SBM) [10] in general seek to model a graph—in our case, over U representatives. Similar to the IPM, the SBM frames our understanding of the structure of our observations using latent variables. The latent variable $M_u \in \{1, \dots, K\}$ associated to representative u designates to which of K communities he or she belongs. The structure of the underlying graph specifies the probability of an interaction R_{uv} between representatives u and v . We assume (R_{uv}) contains counts for the number of interactions between u and v (binary data is typically used). Our version of the SBM assumes the community assignments are i.i.d. draws from a pmf $\pi = (\pi_k)$ on $\{1, \dots, K\}$, and that each pair of communities $k, l \in \{1, \dots, K\}$ has co-expression rate P_{kl} . Our likelihood is

$$R_{uv} \mid P, M_u = k, M_v = l \sim \text{Poisson}(P_{kl}). \quad (3)$$

To achieve conditional conjugacy, we assume $\pi \sim \text{Dir}(\gamma 1_K)$ and $P_{kl} \stackrel{\text{iid}}{\sim} \text{Gamma}(\lambda_0, \lambda_1)$.

2.3 Latent Community Ideal Point Model

We relate the ideal points x_u of IPM to the relational data (R_{uv}) by stitching together the two models described above. We call this the *latent community ideal point model* (LC-IPM). We introduce K community ideal points $\nu_k \stackrel{\text{iid}}{\sim} \mathcal{N}(\varpi, \sigma_\nu^2)$. The ideal point x_u now follows a normal distribution centered around its community mean ν_{M_u} . LC-IPM simultaneously models the voting data (V_{ud}) and the relational data (R_{uv}) in a hierarchical manner, which has the advantage of being able to model ideal points for representatives with little voting information. The graphical model is provided below.

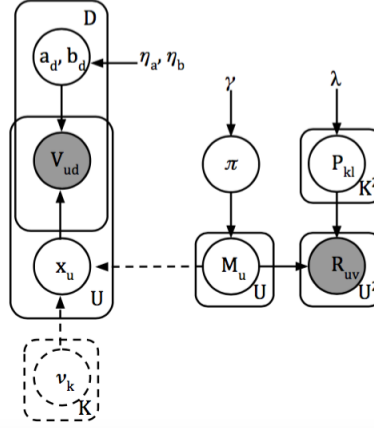


Figure 4: Graphical model for LC-IPM. Left: ideal point model. Right: stochastic block model.

2.4 Related Work

Bayesian methods for ideal point modeling have become an important tool in analyzing revealed preferences in government [4]. Previously these methods have been extended to consider the discrimination a_d and difficulty b_d as response variables with covariates related to the topical content of the bill using *supervised latent Dirichlet allocation* [5]. Relating the posterior of a_d and b_d to the text of the bill (in addition to the votes) addresses the *cold start problem*, i.e. the ability to predict votes on new bills. We adopt a similar approach where the ideal points x_u for representatives with limited voting data can lean on the community mean ν_{M_u} . Both methods are more broadly applicable to collaborative filtering for behavior data when related data sources are present. Previous analyses have performed SVD on roll call voting data to reveal correlations in political structure and latent community structure [8].

3 Posterior Inference

Upon observing votes $V = (V_{ud})$ and caucus co-membership counts $R = (R_{uv})$, computing the posterior distribution of the latent variables given the observations

$$p(x, a, b, \pi, P, M \mid V, R)$$

is intractable. We employ *naïve mean field variational inference*, finding the distribution q closest in KL divergence to the posterior among all fully factorized distributions. Minimizing the KL divergence is equivalent to maximizing the evidence lower bound (ELBO) [2]:

$$\begin{aligned} \mathcal{L}(q) &= \mathbb{E}_q \left[\log \frac{p(\nu, x, a, b, \pi, P, M, V, R)}{q(\nu, x, a, b, \pi, P, M)} \right] \\ &= \mathbb{E}_q \left[\log \frac{p(a, b)p(V \mid a, b, x)p(x \mid M, \nu)p(\nu)}{q(\nu)q(x)q(a)q(b)} \right] + \mathbb{E}_q \left[\log \frac{p(\pi, P, M, R)}{q(\pi)q(P)q(M)} \right]. \end{aligned}$$

Each factor of q follows a simple parametric form, described in appendix A. The simplest approach to variational inference maximizes the ELBO $\mathcal{L}(q)$ via coordinate-ascent, choosing the best value of a variational parameter with all others fixed. Iteratively applying these updates, the variational approximation q improves at every step toward some local optimum. Conditional conjugacy yields closed form updates for the variational parameters corresponding to π and P [2]. Furthermore, due to the decomposition of the ELBO $\mathcal{L}(q)$ above, the mean field factorization yields the same variational updates for π , P , a_d , and b_d as in SBM and IPM, and we exploited this modularity. We devised our implementation of coordinate-ascent variational inference (CAVI) for IPM, SBM, and LC-IPM to have as little repeat code as possible by defining LC-IPM as a class which extends from both of its component models.

4 Empirical Results

To evaluate the performance of LC-IPM we attempted to recover known model parameters from simulated data; we also evaluated the model’s predictive performance on data from the 110th congress. For the following results we implemented CAVI for LC-IPM in Python ¹

4.1 Simulated Data

To ensure that CAVI produces reasonable posterior inferences, we sampled vote interactions and caucus co-membership counts from the generative model and compared the true model parameters to the naïve mean field approximation. We generated 50 representatives, 500 documents, and two communities, using one dimensional ideal points and ran the CAVI routine for 35 iterations. To evaluate convergence we output the ELBO at each iteration. To evaluate the approximate posterior means we compared the variational ideal point mean $\hat{\tau}_u = \mathbb{E}_q[x_u]$ and the posterior responsibility of the first community $\hat{r}_{u1} = \mathbb{E}_q[1_{M_u=1}]$ (See A.1 and A.2 for notation) to the true ideal point \tilde{x}_u and the true community assignment \tilde{M}_u for each representative. We score according to Mean Square Error (MSE) and Brier Score (Brier), given by:

$$\text{MSE}(\tilde{x}, \hat{\tau}) = \frac{1}{U} \sum_{u=1}^U (\tilde{x}_u - \hat{\tau}_u)^2 \quad \text{Brier}(\hat{r}_1, \tilde{M}) = \frac{1}{U} \sum_{u=1}^U (\tilde{M}_u - \hat{r}_{u1})^2$$

Because the model is invariant to permutations of the communities and reflections of the ideal points, we chose the permutation and reflection which minimized the scores. We ran this analysis on 5 different data sets generated according to the model; the results of this experiment are in Figure 5. We see that CAVI very quickly converges to the correct ideal points while taking longer to converge to the correct community assignments. Overall the mean square error is close to zero, and the naïve mean field approximation recovers the correct parameters.

¹Implementation and experiments can be found at www.github.com/ebenmichael/congress-bills

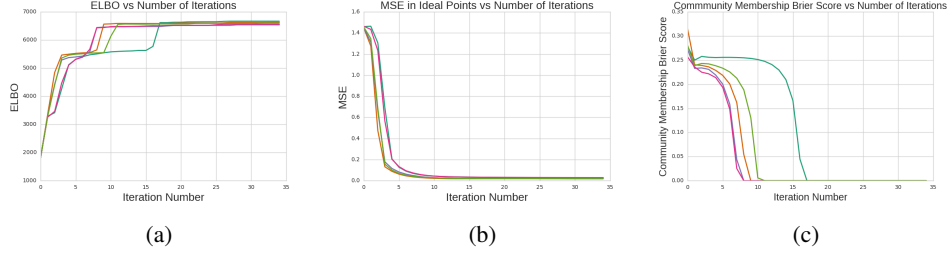


Figure 5: (a) ELBO (b) MSE between variational ideal point means and true ideal points and (c) Brier score between posterior responsibility of the first community and true assignments vs number of CAVI iterations on simulated data for 5 random data sets generated from the model.

4.2 The 110th Congress

4.2.1 Data

We obtained roll call vote records for the House of Representatives of the 110th Congress from GovTrack.us, an independent open government data website which publishes and stores data related to the United States Congress. Our data set contains $D = 1707$ documents (those categorized as bills, amendments, and motions) and $U = 448$ representatives. We only consider those roll call votes where a representative voted *yea* or *nay*; there are 703,830 such interactions. We gathered caucus membership data from [12].

4.2.2 Evaluation

To evaluate our model we assess predictive accuracy. We compared a $K = 2$ community and $K = 10$ community LC-IPM with dimension $S = 2$ to (a) predicting *yea* each time, (b) L^2 regularized Logistic Regression, and (c) the standard Ideal Point Model with dimension $S = 1$ and $S = 2$. We first chose the regularization parameter for Logistic Regression using a 5 fold cross validation procedure. We performed another 5 fold cross validation procedure where we trained each model on 4 folds, tested on the 5th validation fold, and averaged the results (Table 1). We found that LC-IPM and IPM both perform very well and similarly.

Model	Accuracy	AUC
<i>yea</i>	67.123	50.000
Logistic Regression	78.340	85.105
IPM ($S = 1$)	94.769	98.418
IPM ($S = 2$)	95.451	98.998
LC-IPM ($S = 2, K = 2$)	95.405	99.013
LC-IPM ($S = 2, K = 10$)	95.415	99.015

Table 1: Averaged held out test set prediction metrics

4.2.3 Inferred Ideal Points and Communities

In this section we examine inferred ideal points and communities from the House of Representatives of the 110th Congress.

We first ran CAVI on an LC-IPM with dimension $S = 2$ and $K = 2$ communities. With this configuration the ideal points are well separated by party (Figure 6a) but the inferred communities straddle party lines. We sorted the caucus co-membership counts by inferred community and found that there is one community where members share a large number of caucuses, and another where the members do not share very many caucuses (Figure 7a). Running CAVI on an LC-IPM model with dimension $S = 2$ and $K = 10$ communities retains the clean separation by party and overall shape of the ideal points up to rotation (Figure 6b). However, the community structure is markedly different (Figure 7b). The inferred communities simultaneously capture fine block structure in the caucus co-membership counts and inter/intra-party spatial structure of the ideal points.

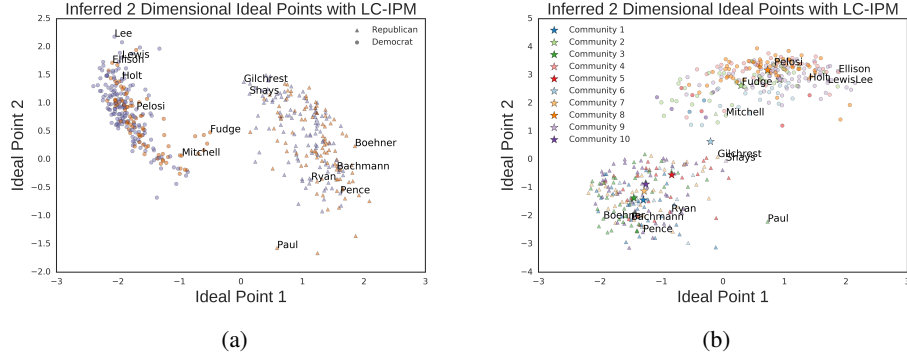


Figure 6: Inferred 2 dimensional ideal points for an LC-IPM model with (a) 2 and (b) 10 communities. Shape denotes party affiliation; color denotes community; stars in (b) indicate inferred community means. Both models separate Republicans and Democrats with noted centrists in between.

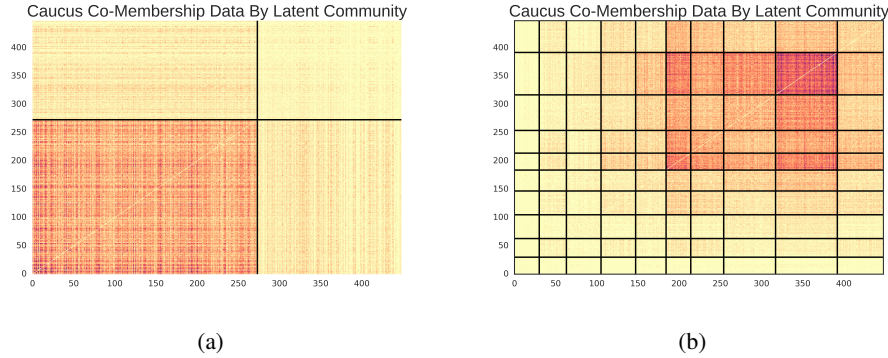


Figure 7: Caucus co-membership counts grouped by inferred community for (a) 2 and (b) 10 communities. Each entry denotes the number of caucuses shared by a pair of representatives. Red denotes many (up to thirty) common caucuses; yellow denotes close to no common caucuses. The 10 community model captures finer block structure.

The concurrent relational and spatial models allow for interesting conclusions. For instance, we can see that the 9th community has a high rate of caucusing together and is comprised of Democrats with ideal points far from Republicans and centrists, leading to an interpretation that the 9th community is a strongly interconnected group of staunch liberals. The four depicted members of this community (Rush Holt, Keith Ellison, John Lewis, and Barbara Lee) are regarded as true-blue liberal representatives. Interestingly, Chris Shays—criticized by hard-line conservatives as a RINO (Republican in name only) and the only Republican representative from New England—is also in this community. We can also see representatives such as centrist Wayne Gilchrest, who broke ranks with the Republican party more often than any other representative in 2007, and libertarian Ron Paul, who challenges Republican orthodoxy, both of whom are in communities which nonetheless are deeply conservative.

5 Conclusions and future directions

While LC-IPM offers similar predictive performance to the standard ideal point model, we gain the ability to detect more subtle clusters within and between parties. This ability to simultaneously analyze the social network and voting behavior of Congress is of particular interest to social scientists. Our approach applies to any collaborative filtering problem where there is associated relational data. For instance, recommendation systems could benefit from modeling social network information. In future work we hope to expand the relational component of our model using Bayesian non-parametric methods and mixed membership models [1].

Finally, a key advantage to our generative model is its modularity, and we are interested in generalizing this notion of composability of hierarchical models for multiple data sources. While this project incorporated caucus data to better inform estimates for the representatives’ ideal points, another direction to explore would be to utilize data on bills to better estimate their difficulty and discrimination. Following the example of [5], we can combine LC-IPM with supervised topic modeling and infer the latent topics in a bill from a bill’s text; these latent topics then inform a bill’s difficulty and discrimination. On the other hand, we may also explore incorporating legislators’ speech transcripts in deliberating the bill [9, 11].

Acknowledgments

We would like to thank Professor Wainwright, Arturo, Jeffrey, and Peter for their dedication to teaching an excellent course.

References

- [1] Airoldi, E. M., Blei, D. M., Fienberg, S. E. & Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*.
- [2] Blei, D. M., Kucukelbir, A. & McAuliffe, J. D. (2016). Variational inference: a review for statisticians. *arXiv:1601.00670*.
- [3] Braun, M. & McAuliffe, J. (2010). Variational inference for large-scale models of discrete choice. *Journal of the American Statistical Association*. 105(489): 324-334.
- [4] Clinton, J., Jackman, S. & Rivers D. (2004). The statistical analysis of roll call data. *American Political Science Review*. 98(2): 353-370.
- [5] Gerrish, S.M. & Blei, D.M. (2011) Predicting legislative roll calls from text. *Proceedings of the 28th International Conference on Machine Learning*.
- [6] Hastie, T. J., Tibshirani, R. & Wainwright, M. J. (2015). Statistical learning with sparsity: the Lasso and generalizations. *CRC Press*.
- [7] Martin, A. D. & Quinn, K. M. (2002). Dynamic ideal point estimation via Markov chain monte carlo for the U.S. Supreme Court, 1953-1999. *Political Analysis*. 10: 134-153.
- [8] Porter, M. A., Mucha, P. J., Newman, M. E. J. & Friend, A. J. (2007). Community structure in the United States House of Representatives. *Physica A*.
- [9] Quinn, K M., Monroe, B L., Colaresi, M., Crespin, M H. & Radev, D R. (2006). An automated method of topic-coding legislative speech over time with application to the 105th- 108th u.s. senate. *In In Midwest Political Science Association Meeting*
- [10] Snijders, T. A. B. & Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*.
- [11] Thomas, M., Pang, B. & Lee, L. (2006). Get out the vote: Determining support or opposition from congressional floor-debate transcripts. *Proceedings of EMNLP*. 327–335.
- [12] Victor, J. N. (2013). “membership_110.csv.” and “membership_111.csv.” Data available for scholarly use from <http://mason.gmu.edu/~jvictor3/Data/> 11/15/2016.
- [13] Wainwright, M. J. & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*.

A Variational updates

A.1 Stochastic Block Model (SBM)

After observing the symmetric matrix $R = (R_{uv})$, where R_{uv} is the number of caucuses that representatives u and v have in common, we see to find a distribution q over the latent community assignments $M = (M_u)$, the community coexpression rates $P = (P_{kl})$, and the community proportions $\pi = (\pi_k)$ which is close in relative entropy to the true posterior and lies in the factorized family $q(M)q(P)q(\pi)$. Each factor has free parameters described below and denoted with hats. The approximation q is equivalently scored by the ELBO objective \mathcal{L} , which we break down as:

$$\mathcal{L}(q) = \underbrace{\mathbb{E}_q \left[\log p(R \mid M, P) + \log \frac{p(P)}{q(P)} \right]}_{\mathcal{L}_{\text{data}}} + \underbrace{\mathbb{E}_q [-\log q(M)]}_{\mathcal{L}_{\text{ent}}} + \underbrace{\mathbb{E}_q [\log p(M \mid \pi)]}_{\mathcal{L}_{\text{local}}} + \underbrace{\mathbb{E}_q \left[\log \frac{p(\pi)}{q(\pi)} \right]}_{\mathcal{L}_{\text{global}}}$$

Variational Factors. To each u we associate variational parameters $\hat{r}_u = (\hat{r}_{uk})_{k=1}^K$, so

$$q(M) = \prod_{u=1}^U q(M_u \mid \hat{r}_u) = \prod_{u=1}^U \prod_{k=1}^K \hat{r}_{uk}^{\delta_k(M_u)}. \quad (4)$$

We define $q(\pi) \triangleq \text{Dir}(\hat{\gamma}_1, \dots, \hat{\gamma}_K)$ and $q(P) \triangleq \prod_{kl} \text{Gamma}(\hat{\lambda}_{0kl}, \hat{\lambda}_{1kl})$.

Computing the ELBO. Now we can write out the component terms of the ELBO more explicitly:

$$\begin{aligned} \mathcal{L}_{\text{data}} &= \mathbb{E}_q \left[\log p(R \mid M, P) + \log \frac{p(P)}{q(P)} \right] = \sum_{kl} \mathbb{E}_q \left[\sum_{u,v} \delta_k(M_u) \delta_l(M_v) \log p(R_{uv} \mid P_{kl}) + \log \frac{p(P_{kl})}{q(P_{kl})} \right] \\ &= - \sum_{u,v} \log R_{uv}! + \sum_{k,l} \left(\lambda_0 \log \lambda_1 - \hat{\lambda}_{0kl} \log \hat{\lambda}_{1kl} - \log \frac{\Gamma(\lambda_0)}{\Gamma(\hat{\lambda}_{0kl})} \right) + \sum_{k,l} \mathcal{L}_{kl}(R) \\ \mathcal{L}_{\text{ent}} &= \mathbb{E}_q [-\log q(M)] = - \sum_{u,k} \mathbb{E}_q [\delta_k(M_u) \log \hat{r}_{uk}] = - \sum_{u,k} \hat{r}_{uk} \log \hat{r}_{uk} \\ \mathcal{L}_{\text{local}} &= \mathbb{E}_q [\log p(M \mid \pi)] = \sum_{u,k} \mathbb{E}_q [\delta_k(M_u) \log \pi_k] = \sum_k N_k \mathbb{E}_q [\log \pi_k] \\ \mathcal{L}_{\text{global}} &= \mathbb{E}_q \left[\log \frac{p(\pi)}{q(\pi)} \right] = \log \Gamma(C\gamma) - C \log \Gamma(\gamma) - \log \Gamma \left(\sum_k \hat{\gamma}_k \right) + \sum_k \{ \log \Gamma(\hat{\gamma}_k) + (\gamma - \hat{\gamma}_k) \mathbb{E}_q [\log \pi_k] \} \end{aligned}$$

where $N_k = \sum_u \hat{r}_{uk}$, $N_{kl} = \sum_{uv} \hat{r}_{uk} \hat{r}_{vl}$, $S_{kl} = \sum_{uv} \hat{r}_{uk} \hat{r}_{vl} R_{uv}$, and

$$\mathcal{L}_{kl}(R) = (S_{kl} + \lambda_0 - \hat{\lambda}_{0kl}) \mathbb{E}_q [\log P_{kl}] - (N_{kl} + \lambda_1 - \hat{\lambda}_{1kl}) \mathbb{E}_q [P_{kl}],$$

and the posterior expectations can also be computed explicitly as

$$\mathbb{E}_q [P_{kl}] = \frac{\hat{\lambda}_{0kl}}{\hat{\lambda}_{1kl}}, \quad \mathbb{E}_q [\log P_{kl}] = \psi(\hat{\lambda}_{0kl}) - \log \hat{\lambda}_{1kl}, \quad \mathbb{E}_q [\log \pi_k] = \psi(\hat{\gamma}_k) - \psi \left(\sum_l \hat{\gamma}_l \right)$$

CAVI Updates. The simplest approach to variational inference maximizes the ELBO \mathcal{L} via coordinate-ascent, i.e. choosing the best value of a variational parameter with all others fixed. Iteratively applying these updates, the variational approximation q improves at every step toward some local optimum. Conditional conjugacy yields closed form updates for $\hat{\gamma}_k$ and $\hat{\lambda}_{kl}$:

- **Global Update to $q(\pi)$.** We have $\hat{\gamma}_k = \gamma + N_k$.
- **Global Update to $q(P)$.** We have $\hat{\lambda}_{0kl} = \lambda_0 + S_{kl}$ and $\hat{\lambda}_{1kl} = \lambda_1 + N_{kl}$.
- **Local Update to $q(M)$.** Differentiating the ELBO with respect to \hat{r}_{uk} ,

$$0 = \frac{\partial \mathcal{L}}{\partial \hat{r}_{uk}} = -\log \hat{r}_{uk} - 1 + \mathbb{E}_q [\log \pi_k] + \sum_{v \neq u} \sum_l \hat{r}_{vl} (R_{uv} \mathbb{E}_q [\log P_{kl}] - \mathbb{E}_q [P_{kl}]).$$

Thus, we take

$$\hat{r}_{uk} \propto_k \exp \left(\mathbb{E}_q [\log \pi_k] + \sum_{v \neq u} \sum_l \hat{r}_{vl} (R_{uv} \mathbb{E}_q [\log P_{kl}] - \mathbb{E}_q [P_{kl}]) \right).$$

A.2 Ideal Point Model (IPM)

We observe the votes matrix $V = (V_{ud})$ where V_{ud} is the vote of congressperson u on bill d . We have the ideal point for congressperson u , $x_u \in \mathbb{R}^s$, and the discrimination and difficulty for bill d , $a_d, b_d \in \mathbb{R}^s$. The variational distribution is fully factorized $\prod_{u=1}^U \prod_{d=1}^D q(x_u)q(a_d)q(b_d)$ where $q(x_u) \triangleq \text{Normal}(\hat{\tau}_u, \hat{\sigma}_\tau^2 I_S)$, $q(a_d) \triangleq \text{Normal}(\hat{\kappa}_{ad}, \hat{\sigma}_{\kappa_a}^2 I_S)$, and $q(b_d) \triangleq \text{Normal}(\hat{\kappa}_{bd}, \hat{\sigma}_{\kappa_b}^2 I_S)$.

Computing the ELBO. We can write the ELBO as

$$\begin{aligned}\mathcal{L}(q) &= H(q) + \sum_u \mathbb{E}_q[\log p(x_u)] + \sum_d \mathbb{E}_q[\log p(a_d)] + \sum_d \mathbb{E}_q[\log p(b_d)] + \mathbb{E}_q[\log p(V|x, a, b)] \\ H(q) &= (US \log 2\pi e \hat{\sigma}_\tau^2 + DS \log 2\pi e \hat{\sigma}_{\kappa_a}^2 + DS \log 2\pi e \hat{\sigma}_{\kappa_b}^2) / 2 \\ \mathbb{E}_q[\log p(x_u)] &= \mathbb{E}_q \left[-\frac{S}{2} \log 2\pi \sigma_x^2 - \frac{1}{2\sigma_x^2} \|x_u - \nu\|_2^2 \right] = -\frac{S}{2} - \frac{1}{2\sigma_x^2} \hat{\sigma}_\tau^2 S + \|\hat{\tau}_u - \nu\|_2^2 \\ \mathbb{E}_q[\log p(a_d)] &= \mathbb{E}_q \left[-\frac{S}{2} \log 2\pi \sigma_a^2 - \frac{1}{2\sigma_a^2} \|a_d - \eta_a\|_2^2 \right] = -\frac{S}{2} - \frac{1}{2\sigma_a^2} \hat{\sigma}_{\kappa_a}^2 S + \|\hat{\kappa}_{ad} - \eta_a\|_2^2 \\ \mathbb{E}_q[\log p(b_d)] &= \mathbb{E}_q \left[-\frac{S}{2} \log 2\pi \sigma_b^2 - \frac{1}{2\sigma_b^2} \|b_d - \eta_b\|_2^2 \right] = -\frac{S}{2} - \frac{1}{2\sigma_b^2} \hat{\sigma}_{\kappa_b}^2 S + \|\hat{\kappa}_{bd} - \eta_b\|_2^2\end{aligned}$$

We can deal with the last expectation by using the 2nd order delta method (Braun McAullife 2008) which takes

$$\mathbb{E}[f(V)] \approx f(\mathbb{E}[V]) + \frac{1}{2} \text{trace}(\nabla^2 \mathbb{E}[V] \text{Cov}(V)).$$

Letting $u(i), d(i)$ be the users and documents for data point i , and applying this gives the approximation to the ELBO contribution from the likelihood

$$\begin{aligned}\mathbb{E}_q[\log p(V|x, a, b)] &= \sum_{i=1}^n \mathbb{E}_q[V_i(a_{d(i)} \cdot (X_{u(i)} - b_{d(i)}))] + \mathbb{E}_q[\log(1 - \sigma(a_{d(i)} \cdot (X_{u(i)} - b_{d(i)})))] \\ &\approx \sum_{i=1}^n V_i(\hat{\kappa}_{ad(i)} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)})) - \log(1 + \exp(\hat{\kappa}_{ad(i)} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)}))) \\ &\quad - \frac{1}{2} \sigma''(\kappa_{ad(i)} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)})) (\hat{\sigma}_{\kappa_a}^2 \|\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)}\|_2^2 + (\hat{\sigma}_\tau^2 + \hat{\sigma}_{\kappa_b}^2) \|\hat{\kappa}_{ad(i)}\|^2)\end{aligned}$$

CAVI Updates. There are no closed form updates for $\hat{\tau}_u, \hat{\kappa}_{ad}$, and $\hat{\kappa}_{bd}$, so we maximize \mathcal{L} numerically when updating these parameters. Let $V(u)$ be the set of votes for user u , and similarly let $V(d)$ be the set of votes on bill d . Also let $\rho_{ud} = \sigma(\kappa_{ad(i)} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)}))$. The gradients are

$$\begin{aligned}\nabla_{\hat{\tau}_u} \mathcal{L} &= -\frac{1}{\sigma_x^2} (\hat{\tau}_u - \nu) + \sum_{i \in V(u)} (V_i - \rho_{ud(i)}) \hat{\kappa}_{ad(i)} - \sigma'(\hat{\kappa}_{ad(i)} \cdot (\hat{\tau}_u - \hat{\kappa}_{bd(i)})) \hat{\sigma}_{\kappa_a}^2 (\hat{\tau}_u - \hat{\kappa}_{bd(i)}) \\ &\quad - \frac{1}{2} \sigma''(\hat{\kappa}_{ad(i)} \cdot (\hat{\tau}_u - \hat{\kappa}_{bd(i)})) (\hat{\sigma}_{\kappa_a}^2 \|\hat{\tau}_u - \hat{\kappa}_{bd(i)}\|_2^2 + (\hat{\sigma}_\tau^2 + \hat{\sigma}_{\kappa_b}^2) \|\hat{\kappa}_{ad(i)}\|^2) \hat{\kappa}_{ad(i)} \\ \nabla_{\hat{\kappa}_{ad}} \mathcal{L} &= -\frac{1}{\sigma_a^2} (\hat{\kappa}_{ad} - \eta_a) + \sum_{i \in V(d)} (V_i - \rho_{u(i)d}) (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd}) - \sigma'(\hat{\kappa}_{ad} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd})) (\hat{\sigma}_\tau^2 + \hat{\sigma}_{\kappa_b}^2) \hat{\kappa}_{ad} \\ &\quad - \frac{1}{2} \sigma''(\hat{\kappa}_{ad} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd})) (\hat{\sigma}_{\kappa_a}^2 \|\hat{\tau}_{u(i)} - \hat{\kappa}_{bd}\|_2^2 + (\hat{\sigma}_\tau^2 + \hat{\sigma}_{\kappa_b}^2) \|\hat{\kappa}_{ad}\|^2) (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd}) \\ \nabla_{\hat{\kappa}_{bd}} \mathcal{L} &= \frac{1}{\sigma_b^2} (\hat{\kappa}_{bd} - \eta_b) - \sum_{i \in V(d)} (V_i - \rho_{u(i)d}) \hat{\kappa}_{ad} + \sigma'(\hat{\kappa}_{ad} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd})) \hat{\sigma}_{\kappa_a}^2 (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd}) \\ &\quad + \frac{1}{2} \sigma''(\hat{\kappa}_{ad} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd})) (\hat{\sigma}_{\kappa_a}^2 \|\hat{\tau}_{u(i)} - \hat{\kappa}_{bd}\|_2^2 + (\hat{\sigma}_\tau^2 + \hat{\sigma}_{\kappa_b}^2) \|\hat{\kappa}_{ad}\|^2) \hat{\kappa}_{ad}\end{aligned}$$

For each parameter we solve this optimization problem using L-BFGS. Finally, there are closed form updates for the variational variance parameters by taking the derivative and setting to zero

$$\begin{aligned}\hat{\sigma}_\tau^2 &= \frac{US}{\frac{US}{\sigma_x^2} + \sum_{i=1}^n \sigma'(\kappa_{ad(i)} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)}))(S\hat{\sigma}_{\kappa_a}^2 + \|\hat{\kappa}_{ad(i)}\|_2^2)} \\ \hat{\sigma}_{\kappa_a}^2 &= \frac{DS}{\frac{DS}{\sigma_a^2} + \sum_{i=1}^n \sigma'(\kappa_{ad(i)} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)}))(S(\hat{\sigma}_\tau^2 + \hat{\sigma}_{\kappa_b}^2) + \|\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)}\|_2^2)} \\ \hat{\sigma}_\tau^2 &= \frac{DS}{\frac{DS}{\sigma_b^2} + \sum_{i=1}^n \sigma'(\kappa_{ad(i)} \cdot (\hat{\tau}_{u(i)} - \hat{\kappa}_{bd(i)}))(S\hat{\sigma}_{\kappa_a}^2 + \|\hat{\kappa}_{ad(i)}\|_2^2)}\end{aligned}$$

A.3 Latent Community Ideal Point Model (LC-IPM)

The factorization for q is the same, with one more factor $q(\nu) = \prod_k q(\nu_k)$ where $q(\nu_k) \triangleq \mathcal{N}(\hat{\mu}_k, \hat{\sigma}_\mu^2)$. Due to the factorization in the LC-IPM generative model, the only contribution to the ELBO from IPM which changes is that corresponding to (x_u) . This becomes

$$\begin{aligned}\mathcal{L}_x &= \mathbb{E}_q \left[\log \frac{p(x|\nu, M)}{q(x)} \right] = \mathbb{E}_q \left[\log \prod_{uk} \phi(x_u|\nu_k)^{\delta_k(M_u)} \right] + H(q) \\ &= \sum_{uk} \hat{r}_{uk} \mathbb{E}_q [\log \phi(x_u|\nu_k)] + H(q)\end{aligned}$$

In particular, the gradient of the ELBO w.r.t. the responsibility \hat{r}_{uk} is

$$\begin{aligned}0 &= \frac{\partial \mathcal{L}_{\text{SBM}}}{\partial \hat{r}_{uk}} + \frac{\partial \mathcal{L}_x}{\partial \hat{r}_{uk}} = -\log \hat{r}_{uk} - 1 + \mathbb{E}_q [\log \pi_k] + \mathbb{E}_q [\log \phi(x_u|\nu_k)] \\ &\quad + \sum_{v \neq u} \sum_l \hat{r}_{vl} (R_{uv} \mathbb{E}_q [\log P_{kl}] - \mathbb{E}_q [P_{kl}])\end{aligned}$$

so the update is

$$\hat{r}_{uk} \propto_k \exp \left(\mathbb{E}_q [\log \pi_k] + \sum_{v \neq u} \sum_l \hat{r}_{vl} (R_{uv} \mathbb{E}_q [\log P_{kl}] - \mathbb{E}_q [P_{kl}]) + \mathbb{E}_q [\log \phi(x_u|\nu_k)] \right).$$

To determine the updates for the variational mean $\hat{\mu}_k$ corresponding to ν_k , we need the ELBO term

$$\mathcal{L}_\nu = \mathbb{E}_q \left[\log \frac{p(\nu)}{q(\nu)} \right] = \sum_k \mathbb{E}_q [\log p(\nu_k)] + KH(q(\nu_1)) = -\frac{1}{2\sigma_\nu^2} \sum_k \|\hat{\mu}_k - \varpi\|^2 + \frac{KS}{2} \log(2\pi e \hat{\sigma}_\mu^2) + \text{const.}$$

Setting the gradient of the ELBO w.r.t. $\hat{\mu}_k$ equal to zero, we obtain

$$0 = \frac{\partial (\mathcal{L}_\nu + \mathcal{L}_x)}{\partial \hat{\mu}_k} = \frac{1}{\sigma_x^2} \sum_u \hat{r}_{uk} (\hat{\tau}_u - \hat{\mu}_k) - \frac{1}{\sigma_\nu^2} (\hat{\mu}_k - \varpi) = \frac{1}{\sigma_x^2} \sum_u \hat{r}_{uk} \hat{\tau}_u - \left(\frac{N_k}{\sigma_x^2} + \frac{1}{\sigma_\nu^2} \right) \hat{\mu}_k + \frac{1}{\sigma_\nu^2} \varpi$$

and thus

$$\hat{\mu}_k = \left(\frac{\sum_u \hat{r}_{uk} \hat{\tau}_u}{\sigma_x^2} + \frac{\varpi}{\sigma_\nu^2} \right) \hat{\sigma}_{\hat{\mu}_k}^2; \text{ where } \hat{\sigma}_{\hat{\mu}_k}^2 = \left(\frac{N_k}{\sigma_x^2} + \frac{1}{\sigma_\nu^2} \right)^{-1}.$$