1 Model

Stochastic Block Model (SBM)

- Choose the community proportions $\pi \sim \text{Dir}(\gamma)$, where $\pi \in \mathbb{R}^C$ with C latent communities
- For each representative u
  1. Choose a community membership assignment $M_u \overset{\text{iid}}{\sim} \text{Cat}(\pi)$
- For each pair of communities $k, l \in \{1, \dots, C\}$, draw coexpression rate $P_{kl} \overset{\text{iid}}{\sim} \text{Gamma}(\lambda_0, \lambda_1)$
- For each pair of representatives $u, v \in \{1, \dots, U\}$, draw $R_{uv} \mid P, M_u = k, M_v = l \sim \text{Poisson}(P_{kl})$

Ideal Point Model (IPM)

- For each document d
  1. Choose a discrimination $a_d \sim \mathcal{N}(\eta_a, \sigma_d^2)$
  2. Choose a difficulty $b_d \sim \mathcal{N}(\eta_b, \sigma_d^2)$
- For each representative u
  1. Choose a position $x_u \mid \nu \sim \mathcal{N}(\nu, \sigma_x^2)$
- Draw representative u's vote on document d as $V_{ud} \mid x_u, a_d, b_d \sim \text{Bern}(\sigma(a_d \cdot (x_u - b_d)))$

Latent Dirichlet Allocation (LDA)

- Draw a topic $\varphi_k \overset{\text{iid}}{\sim} \text{Dir}(\beta), \varphi_k \in \mathbb{R}^V$ as a distribution over words, for each $k \in \{1, \dots, K\}$
- For each document, draw the topic proportions $\theta_d \overset{\text{iid}}{\sim} \text{Dir}(\alpha)$, where $\theta_d \in \mathbb{R}^K$
- For each document $d \in \{1, \dots, D\}$ and each word $n \in \{1, \dots, N_d\}$ in the document
  1. Choose a topic $z_{dn} \mid \theta_d \overset{\text{ind}}{\sim} \text{Mult}(\theta_d)$
  2. Choose a word $W_{dn} \mid z_{dn} = k, \varphi_k \overset{\text{ind}}{\sim} \text{Mult}(\varphi_k)$

1.1 Frankenstein Model

The ideal point model (IPM) is useful to us as a baseline model for the roll call voting data (V_{ud}) for a couple of reasons. For one, using it alone we can attempt to predict missing votes, a problem of interest in political science. Another problem of more qualitative interest is analyzing and interpreting the factors a_d, b_d specific to a document and those x_u specific to the representative. All are assumed to reside in some latent space $\mathbb{R}^S$ and so depending on how we set up the model, we might be able to interpret quantities like x_u as u's *political stance* or *ideological position* or $x_u - b_d$ as representative u's propensity for the bill/document d. There are a number of problems we cannot address in IPM. A major problem is predicting on heldout documents (the 'cold start'), which is a potentially useful performance measure. Similarly if we have relatively junior representatives, they may not have had enough votes for the inferred x_u to represent something (1) meaningful / interpretable or (2) reliable. We want to incorporate more information to inform the choices of a_d, b_d and x_u.

Ideal Point Allocator (IPA)

- Run the generative processes for SBM and LDA as described above. Then,
- For each document d
  1. Calculate the empirical topic proportions $\overline{z}_d = \frac{1}{N_d}\sum_{i=1}^{N_d} z_d$ (a $K \times 1$ vector)
  2. Generate $S \times K$ matrices η_a, η_b with iid normal entries
  3. Choose a discrimination $a_d \sim \mathcal{N}(\eta_a' \overline{z}_d, \sigma_d^2)$
  4. Choose a difficulty $b_d \sim \mathcal{N}(\eta_b' \overline{z}_d, \sigma_d^2)$
- For each representative u
  1. Generate the community means $\nu_k \sim \mathcal{N}(\tau, \sigma_x^2)$
  2. Choose a position $x_u \mid M_u = k, \nu \sim \mathcal{N}(\nu_k, \sigma_x^2)$
- Draw representative u's vote on document d as $V_{ud} \mid x_u, a_d, b_d \sim \text{Bern}(\sigma(a_d \cdot (x_u - b_d)))$

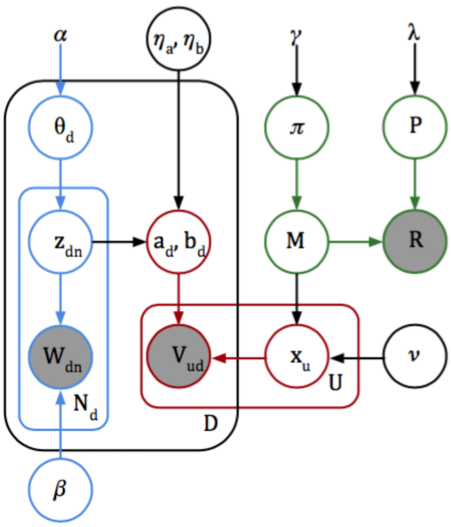

Figure 1: IPA graphical model