Dynamic Dirichlet process mixture model for identifying voting coalitions in the United Nations General Assembly human rights roll call votes

Qiushi Yu

View supplementary material

Published online: 20 May 2021.

Submit your article to this journal

Article views: 131

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

Check for updates

Dynamic Dirichlet process mixture model for identifying voting coalitions in the United Nations General Assembly human rights roll call votes

Qiushi Yu

University of Michigan, Ann Arbor, MI, USA

ABSTRACT

Scholars have been interested in the politicization of humans rights within the United Nations for some time. However, previous research typically looks at simple associations between voting coalitions and observable variables, such as geographic location or membership in international organizations. Our study is the first attempt at estimating the latent coalition structure based on the voting data. We propose a Bayesian Dynamic Dirichlet Process Mixture (DDPM) model to identify voting coalitions based on roll call vote data across multiple time periods. We also propose post-processing methods for analyzing the outputs of the DDPM model. We apply these methods to the United Nations General Assembly (UNGA) human rights roll call vote data from 1992 to 2017. We identify human rights voting coalitions in the UNGA after the Cold War, and polarizing resolutions that divide countries into different coalitions.

1. Introduction

In his 2018 address to the United Nations General Assembly (UNGA), President Trump stated that the United Nations' body overseeing human rights issues 'had become a grave embarrassment to this institution, shielding egregious human rights abusers while bashing America and its many friends [49].' Indeed, others have lodged similar critiques of the United Nations' handling of human rights. Many critiques point out that human rights are highly politicized, making it difficult for the United Nations to fulfill its original mission of defending human rights [16,22,25,40].

One important avenue for countries to make their voices heard on human rights issues is to vote on the human rights resolutions in the UNGA. These human rights resolutions concern both general human rights principles, such as the right to development, the use of mercenaries, children's and women's rights, and human rights violation reports on a specific country. Researchers have analyzed countries' human rights votes from two main perspectives. First, researchers believe that countries' votes on human rights resolutions reflect their individual preferences in terms of human rights protection, negligence, or

---

violation [10]. Researchers found that countries' domestic political regimes and domestic human rights records are predictive of their votes on human rights resolutions [26]. This approach is in line with many studies that apply multidimensional scaling models or Bayesian Item Response Theory (IRT) models to the UNGA voting data to estimate countries' individual latent preferences in terms of global governance and world affairs [4,5,50,51].

Second, other researchers focus on the 'peer group' effect. Studies of the United Nations have shown evidence for 'voting blocs' in the UNGA vote data in the early days of the institution [24]. Namely, there are many observations of clustering structures in the UNGA data [23,29,39]. However, most of these studies take a qualitative approach, and explain this phenomena in terms of the historical and political backgrounds. Specifically, researchers of human rights voting in the UNGA and the Human Rights Council of the UN also contend that 'peer groups' of countries affect individual countries' votes on human rights resolutions [10,25,26]. The 'peer groups' may arise from geographic closeness, such as regional country groups, or cultural connections, such as countries sharing the same religion or language. Specifically, researchers contend that the European Union (EU) member countries tend to coalesce in human rights voting [30]. Other researchers find that member countries of the Organization of Islamic Cooperation (OIC) are inclined to vote together on human rights issues [7].

Our study on the human rights votes in the UNGA engages with the 'peer group' argument in the literature. The existing analyzes rely solely on observable variables to group countries, such as geographic region or international organization membership. To the best of our knowledge, there has been no attempt to learn the latent clustering structure in the UNGA human rights votes data.[1] To better understand the politicization and the challenges of dealing with human rights issues in the UN, we propose a new statistical clustering model to analyze the UNGA human rights vote data. The goal of this study is to inductively identify the voting coalitions in the UNGA human rights vote data without making assumptions about countries' political, historical, or cultural backgrounds. We also identify the most polarizing resolutions that divide countries into different voting coalitions. The latent voting coalition structures and the polarizing resolutions advance our understanding of the lines of conflict on human rights and the driving force behind the politicization of human rights issues in the UNGA. In addition, our study also engages and contributes to the broad literature on model-based clustering methods for identifying voting blocs [20,21].

In this paper, we propose a Dynamic Dirichlet Process Mixture (DDPM) model to identify voting coalitions based on roll call vote data across multiple time periods. We also propose post-processing methods for analyzing the outputs of the DDPM model. The strength of the Dirichlet Process Mixture (DPM) model lies in its weak assumptions and flexibility. In this study, the DPM model models an individual voter's vote on a specific bill as a Bernoulli trial.[2] We make less data generation assumptions than other common models applied to roll call vote data, such as multidimensional scaling models and Item Response Theory models. In addition, as a Bayesian Non-parametrics clustering algorithm, the DPM model does not require users to pre-specify the number of clusters for data analyzes. Theoretically, the method can identify as many clusters as there are individual voters. With proper priors, the DPM model lets the data decide how many clusters there should be. Moreover, as a strength of Bayesian modeling in general, the DPM model

allows researchers to add hierarchical priors on appropriate parameters to model dynamic dependence.

Many existing roll call vote datasets cover decades of voting records in one institution. The dynamic nature of the datasets provide important information on the varying numbers of voting coalitions across time. Intentionally modeling the dynamic dependence of the numbers of voting coalitions allows the DDPM model to borrow information across time periods, resulting in more accurately identifying voting coalitions in each period. Other than the purpose of borrowing information across time, the variation and trend of the numbers of voting coalitions across time could be the direct quantity of interest. For example, the number of voting coalitions in the UNGA human rights vote data bears on many important topics in international relations, such as polarity and politicization of human rights issues. To model the dynamic dependence among roll call votes across periods, we propose a Dynamic Linear Model (DLM) to model the precision parameters for each period's DPM. The precision parameter in a period directly reflects a researcher's prior belief about how many voting coalitions there should be in this period. The DLM on the precision parameters adds a dynamic dependence structure on modeling the numbers of voting coalitions across time.

Due to the label switching problem, interpreting the MCMC output from the DDPM model can be challenging. To aid applied researchers in analyzing the outputs of the DDPM models, we propose a Maximum A Posteriori estimation method to extract fixed cluster labels for each voter in a period based on the posterior samples. Moreover, we propose a visualization method to present the coalition structure in each time period, based on the posterior probability for any voter pair to be in the same cluster. We implement the DDPM model and post-processing methods in the `MCMCddpmbb` R package, and make it freely available to the public.

The rest of the paper proceeds as follows: First, we illustrate the model setup for a single period in a roll call vote setting. Second, we specify the DLM structure to model time dependence in the data. Third, we outline the sampling scheme for the DDPM model. Fourth, we propose post-processing methods for analyzing the posterior samples of the DDPM model. Last, we use the DDPM model to analyze the UNGA human rights roll call vote data after the Cold War.

2. Model setup for a single period

We illustrate the proposed model for a single period in the setting of identifying voting coalitions (clusters) based voters' roll call votes in a parliament-like institution across multiple periods. The number of time periods is T, i.e. $t = 1, \ldots, T$. The number of voters and bills can both vary across time. For any given time period t, there are I_t voters, i.e. $i = 1, \ldots, I_t$, and there are J_t bills (items), i.e. $j = 1, \ldots, J_t$. Let v_{ijt} be an indicator variable representing the observed vote of voter i on bill j at time t.

$$
v_{ijt} = \begin{cases} 1 & \text{if voting yea} \\ 0 & \text{if voting nay} \\ NA & \text{if missing} \end{cases}
$$

In each time period, a roll call vote matrix $\boldsymbol{V}_t$ stores all the data. Column i of $\boldsymbol{V}_t$ stores voter i's voting record vector $\boldsymbol{v}_{it}$ that consists of 1, 0 or NA. We use J_t independent Bernoulli

distributions to model each $\boldsymbol{v}_{it}$. For the voters belonging to the same cluster l, they share the same J_t Bernoulli parameters, which are stored in $\boldsymbol{\theta}_{lt}$. If voter i is in cluster l, then we know $v_{ijt} \sim \text{Bernoulli}(\theta_{ljt})$, for $j = 1, 2, \ldots, J_t$, and we express the likelihood of voter i's voting record, $\boldsymbol{v}_{it}$, as follows:

$$F(\boldsymbol{v}_{it} \mid \boldsymbol{\theta}_{lt}) = \prod_{j=1}^{J_t} \theta_{ljt}^{\mathbf{I}(v_{ijt}=1)} (1 - \theta_{ljt})^{\mathbf{I}(v_{ijt}=0)}$$

where $\mathbf{I}()$ is an indicator function.

We use a Beta-Binomial DPM model to cluster voters, which is equivalent to clustering all the $\boldsymbol{v}_{it}$ vectors. We put a Dirichlet Process prior on cluster l's Bernoulli parameter vector $\boldsymbol{\theta}_{lt}$:

$$\boldsymbol{\theta}_{lt} \mid G_t \sim G_t$$
$$G_t \mid \alpha_t, \boldsymbol{\lambda}_t \sim DP(\alpha_t G_{0t}(\cdot \mid \boldsymbol{\lambda}_t))$$

α_t is the precision parameter in period t. G_t does not have a specific parametric form but is assumed to be generated from a Dirichlet process, whose center distribution is G_{0t}. G_{0t} is the product of J_t independent Beta distributions, acting as the priors for an arbitrary cluster l's Bernoulli parameter vector $\boldsymbol{\theta}_{lt}$. $\boldsymbol{\lambda}_t$ is a J_t by 2 matrix, the j'th row of which stores the two parameters for the Beta prior distribution corresponding to bill j in time t. For an arbitrary new cluster l, the prior for its Bernoulli parameter vector, $G_{0t}(\boldsymbol{\theta}_{lt} \mid \boldsymbol{\lambda}_t)$, is expressed as follows:

$$G_{0t}(\boldsymbol{\theta}_{lt} \mid \boldsymbol{\lambda}_t) = \prod_{j=1}^{J_t} \left\{ \frac{\Gamma(\lambda_{1jt} + \lambda_{0jt})}{\Gamma(\lambda_{1jt})\Gamma(\lambda_{0jt})} \theta_{ljt}^{\lambda_{1jt}-1} (1 - \theta_{ljt})^{\lambda_{0jt}-1} \right\}$$

We introduce the parameter of direct interest, c_{it}, indicating the cluster affiliation of voter i in time t. Given $\boldsymbol{\lambda}_t$ and α_t, c_{it}'s are drawn separately across different time periods. When considering the distribution for c_{it}, we take the current cluster affiliations of all the other voters, $\boldsymbol{c}_{-i,t}$, as given. Assuming that there are currently L_t existing unique clusters in period t among all the voters except for voter i, then the cluster affiliation of voter i is determined by a categorical distribution. When drawing a new value of c_{it}, we know that the drawn value is either one of the existing cluster's label, or a new label representing a completely new cluster ($\boldsymbol{v}_{it}$ will be the first member of this new cluster). Therefore, the categorical distribution on c_{it} is defined on all the existing clusters' labels and a potentially newly generated cluster's label. We can express the probability for c_{it} as follows:

$$p(c_{it} = l, \text{ for } l \in \boldsymbol{c}_{-i,t} \mid \boldsymbol{c}_{-i,t}, \boldsymbol{V}_t) = b \frac{n_{-i,lt}}{I_t - 1 + \alpha_t} \int F(\boldsymbol{v}_{it} \mid \boldsymbol{\theta}_{lt}) \, \mathrm{d}H_{-i,l}(\boldsymbol{\theta}_{lt})$$

$$p(c_{it} \neq c_{i't}, \forall \, i' \neq i \mid \boldsymbol{c}_{-i,t}, \boldsymbol{V}_t) = b \frac{\alpha_t}{I_t - 1 + \alpha_t} \int F(\boldsymbol{v}_{it} \mid \boldsymbol{\theta}_{0t}) \, \mathrm{d}G_{0t}(\boldsymbol{\theta}_{0t})$$

$\boldsymbol{c}_{-i,t}$ represents all voters' cluster affiliations except for voter i's, c_{it}. b is an appropriate normalizing constant. $n_{-i,lt}$ is the size of cluster l, not considering voter i. $F(\boldsymbol{v}_{it} \mid \boldsymbol{\theta}_{lt})$ is the likelihood above, and $H_{-i,l}$ is the updated prior density of $\boldsymbol{\theta}_{lt}$ based on the Dirichlet Process

prior G_{0t} and all the vote vectors belonging to cluster l, except for voter i's. Note that we analytically integrate out $\boldsymbol{\theta}_{lt}$ and $\boldsymbol{\theta}_{0t}$ vectors by relying on the conjugate Beta prior, so we never sample $\boldsymbol{\theta}$ parameters in the MCMC sampler. The intuition to use the product of cluster l's size and the likelihood of voter i's vote vector is to induce sparsity of cluster assignment. Without considering cluster sizes, assigning each voter to her own cluster is guaranteed to achieve the largest likelihood for the data. However, identifying only single-member clusters provides no additional insight than the raw data already do.

In a single period, we draw c_{it}'s with the Chinese Restaurant Process [1,9,43]. The Chinese Restaurant Process induces large clusters to grow larger and small clusters to disappear. The Chinese Restaurant Process is not a very strong assumption in itself, because we can theoretically tune the algorithm with the precision parameter, α_t, to achieve a vast possibility of cluster configurations. α_t acts as the imaginary cluster size for a cluster yet to be generated. If $\alpha_t \to 0$, then we put all voters in one cluster, which is equivalent to a completely pooled model. If $\alpha_t \to +\infty$, then we put each voter in a single-member cluster, which is equivalent to separately modeling each voter's vote vector without borrowing information. We can add proper priors on α_t's to let the data tell how many clusters there should be, and this modeling strategy reflects the flexibility of DPM.

We assume that the values of $\boldsymbol{\lambda}_t$ are known a priori. As pointed out in Spirling & Quinn (2010), roll call vote data usually do not provide enough information for sampling $\boldsymbol{\lambda}_t$ parameters, so models that sample values of $\boldsymbol{\lambda}_t$ are not ideal [48]. This finding is also confirmed by our exploration with models that sample $\boldsymbol{\lambda}_t$ with simulated data. Sampling $\boldsymbol{\lambda}_t$ prevents the posterior samples of cluster labels from converging to the true values. For this paper, we assume that all values in each matrix $\boldsymbol{\lambda}_t$ are 1. This is equivalent to assuming a uniform$(0,1)$ prior distribution for all θ_{ljt}'s. However, it's reasonable to assume other values for elements in $\boldsymbol{\lambda}_t$ based on alternative theories or prior beliefs.

We assume that a voter's voting record on a bill is missing completely at random (MCAR) [46]. The implication of the MCAR assumption is that a missing voting record does not affect estimating any voter's cluster affiliation. For a single period, if voter i's voting record on bill j is missing ($v_{ijt} = NA$), then the likelihood of v_{ijt}, $F(v_{ijt} \mid \theta_{ljt})$, is a constant in voter i's cluster affiliation, c_{it}.[3] Therefore, v_{ijt} does not affect computing the weight parameters for the categorical distribution for drawing voter i's cluster affiliation, c_{it}, and only observed voting records for voter i have an impact. Similarly, suppose voter i' belongs to cluster l for the current sampling iteration, and we are about to draw the cluster affiliation parameter for voter i, c_{it}. Then voter i''s missing voting record on bill j does not affect computing $H_{-i,l}(\boldsymbol{\theta}_{lt})$, the updated prior density of $\boldsymbol{\theta}_{lt}$.[4]

3. Modeling dynamic dependence

For each period t, the precision parameter α_t is a very important parameter. The value of α_t reflects a researcher's prior belief about the number of clusters in period t. Let L_t be the number of clusters in period t. The prior distribution for L_t depends critically on α_t, and is stochastically increasing with α_t. As shown in the DPM model literature, for I_t, the number of voters, moderately large, $E(L_t \mid \alpha_t, I_t) \approx \alpha_t \ln(1 + \frac{I_t}{\alpha_t})$ [14]. α_t is always positive, so for a single time period t, we put a log-normal prior on α_t.

$$\ln(\alpha_t) \sim N(\gamma_t, V)$$

We assume an underlining Bayesian Dynamic Linear Model (DLM) on $z_t = \log(\alpha_t)$ for the purpose of smoothing the values of α_t's by borrowing information from adjacent time periods. The reason for adding this dynamic smoothing structure on α_t's is that adjacent time periods should show similar degrees of division or cooperation in the voting records. Therefore, the prior belief of the number of clusters for one period should be correlated with those in adjacent periods. The sampler will be able to borrow information from draws of α_t in neighboring periods for sampling α_t for a specific period.

Specifically, we put a random walk DLM prior on z_t. The random walk DLM prior specification is extensively used in various Bayesian dynamic IRT models for modeling latent variables in political science research, such as the Supreme Court Justices' ideal points [3,34] and countries' human rights scores [47]. The popularity of the random walk DLM prior is due to its flexibility and the efficient 'Forward Filtering Backward Sampling' Gibbs sampler developed for it [11,17,53]. The random walk DLM prior on z_t is specified as follows:

$$z_t \sim N(\gamma_t, V)$$
$$\gamma_t \sim N(\gamma_{t-1}, W)$$
$$\gamma_0 \sim N(m_0, H_0)$$
$$V \sim \text{Inverse} - \text{Gamma}\left(\frac{r_0}{2}, \frac{s_0}{2}\right)$$
$$W \sim \text{Gamma}(r_1, s_1)$$

V is the variance of the Gaussian disturbance, e_2, in the observation equation, $z_t = \gamma_t + e_2, e_2 \sim N(0, V)$. V has a conjugate Inverse-Gamma prior with user-supplied parameters r_0 and s_0. W is the variance of the Gaussian disturbance, e_1, in the evolution equation, $\gamma_t = \gamma_{t-1} + e_1, e_1 \sim N(0, W)$. The smaller W is, the more smoothing of z_t is achieved across time, whereas large W value allows the value of z_t to change more drastically across time. W has a Gamma prior with user-supplied shape parameter r_1 and rate parameter s_1. We choose a Gamma prior for W over a conjugate Inverse-Gamma prior commonly used in the DLM literature [42]. The Gamma prior is a better alternative, because of its potentially desirable shrinkage property [8]. If a user believes that cluster numbers are very stable across time, then she can pull the estimated W toward a small value, by tuning r_1 and s_1 to reduce both the prior mean $\frac{r_1}{s_1}$ and the prior variance $\frac{r_1}{s_1^2}$. Conversely, if a user does not have a strong belief about the stability of cluster numbers across time, then she can use a more diffused Gamma prior on W, and little shrinkage effect will take place in this case. m_0 and H_0 are the mean and variance for Gaussian prior on γ_0 for the imaginary period before period 1. Except for the Gamma prior on W, the above specification is the standard setup of a unidimensional Gaussian random walk DLM [53].

4. Sampling scheme

The sampler for the entire model is a mixture of Gibbs sampler and Metropolis-Hastings sampler. Due to conjugate prior specifications, the DPM part of the model only relies on a Gibbs sampler for sampling c_{it}'s. The conjugate property breaks down for sampling α_t, so we use a random walk Metropolis-Hastings step to sample α_t in each period separately.

Given the last draws of α_t's, the DLM part of the model mostly relies on a Gibbs sampler, except for the sampling step on W, which employs a random walk Metropolis-Hastings sampler. We lay out the sampling scheme in the following three parts:

(1) For each period t, we sample cluster labels, c_{it}'s, given the current value of α_t, vote matrix, $\boldsymbol{V}_t$, and user-supplied values of $\boldsymbol{\lambda}_t$.
(2) For each period t, we draw the precision parameter, α_t, given the current value of γ_t, the current number of clusters, L_t, and the constant number of voters, I_t.
(3) Given the current values of α_t's and the user-supplied values of $m_0, H_0, r_0, s_0, r_1, s_1$, we sample the latent means, γ_t's, the observation equation variance, V, and the innovation equation variance, W, in the DLM model.

4.1. Sample $c_{it} \mid \boldsymbol{c}_{-i,t}, \alpha_t, \boldsymbol{V}_t, \boldsymbol{\lambda}_t$

Cluster affiliation parameters, c_{it}, are sampled separately in each time period given $\boldsymbol{V}_t$ and $\boldsymbol{\lambda}_t$, so we drop subscript t for simplification in this subsection. Every cluster affiliation corresponds to either a unique existing cluster, or to a newly generated cluster. $\boldsymbol{c}_{-i}$ represents all voters' cluster affiliations in period t except for voter i's, c_i. We use the Chinese Restaurant Process to draw c_i given $\boldsymbol{c}_{-i}$. Voter i is assigned to a new cluster or an existing cluster with a categorical distribution. Due to the conjugacy of the Beta-Binomial prior-likelihood specification, we can integrate out $\boldsymbol{\theta}_0$ or $\boldsymbol{\theta}_l$ when computing the weights for a new cluster or the existing clusters [38]. Therefore, the sampler never directly samples or stores the values of $\boldsymbol{\theta}_0$ or $\boldsymbol{\theta}_l$.

Let $n_{-i,l}$ be the size for an existing cluster l, not considering voter i. For each existing cluster l (not considering voter i), we define the count of yea vote on bill j as $n_{-i,lj1}$, and the count of nay vote on bill j as $n_{-i,lj0}$, based on all the affiliating voters' observed voting records. A yet-to-be-generated cluster has α size and zero counts of yea or nay on any bill. Let b be an appropriate normalizing constant, and I be the number of voters. We express the conditional posterior probability of c_i as below.[5] We can thus directly draw c_i with a closed-form Gibbs sampler.

$$p(c_i = l, \text{ for } l \in \boldsymbol{c}_{-i} \mid \boldsymbol{c}_{-i}, \boldsymbol{V}) = b \frac{n_{-i,l}}{I - 1 + \alpha} \prod_{j=1}^{J} \left(\left(\frac{n_{-i,lj1} + \lambda_{1j}}{n_{-i,lj1} + n_{-i,lj0} + \lambda_{0j} + \lambda_{1j}} \right)^{\mathbf{I}(v_{ij}=1)} \right.$$

$$\left. \times \left(\frac{n_{-i,lj0} + \lambda_{0j}}{n_{-i,lj1} + n_{-i,lj0} + \lambda_{0j} + \lambda_{1j}} \right)^{\mathbf{I}(v_{ij}=0)} \right)$$

$$p(c_i \neq c_{i'}, \forall\, i' \neq i \mid \boldsymbol{c}_{-i}, \boldsymbol{V}) = b \frac{\alpha}{I - 1 + \alpha} \prod_{j=1}^{J} \left(\left(\frac{\lambda_{1j}}{\lambda_{0j} + \lambda_{1j}} \right)^{\mathbf{I}(v_{ij}=1)} \left(\frac{\lambda_{0j}}{\lambda_{0j} + \lambda_{1j}} \right)^{\mathbf{I}(v_{ij}=0)} \right)$$

Researchers find that pure Gibbs samplers for DPM models have poor mixing, so it's necessary to attempt at merging and splitting existing clusters to improve mixing [27]. Therefore, in addition to the Gibbs sampler for c_i, we also use the Sequentially-Allocated Merge-Split (SAMS) sampler for c_i. We implement the above Gibbs sampler and SAMS sampler in alternative turns. Interested readers should refer to [13] for the details of SAMS sampler.

4.2. Sample $\alpha_t(z_t) \mid \gamma_t, V, L_t, I_t$

For period t, the conditional likelihood of α_t, given the number of clusters L_t and the number of voters I_t, is $p(\alpha_t \mid L_t, I_t) \propto \alpha_t^{L_t} \frac{\Gamma(\alpha_t)}{\Gamma(\alpha_t + I_t)}$ [2,14,15]. We have a log-normal prior on α_t, $\ln(\alpha_t) \sim N(\gamma_t, V)$. We can express the conditional posterior of α_t as follows:

$$p(\alpha_t \mid \gamma_t, V, L_t, I_t) \propto \underbrace{\frac{1}{\alpha_t} \exp\left(-\frac{(\ln \alpha_t - \gamma_t)^2}{2V}\right)}_{\text{Log-Normal prior kernel}} \underbrace{\alpha_t^{L_t} \frac{\Gamma(\alpha_t)}{\Gamma(\alpha_t + I_t)}}_{\text{conditional likelihood}}$$

The conditional posterior of α_t does not have a kernel of any recognizable distribution. We use a unidimensional random walk Metropolis-Hastings sampler to sample α_t for each period separately, and users supply the positive value τ to generate the random walk step, δ, from $\delta \sim \text{uniform}(-\tau, \tau)$ for updating α_t. We accept the proposed new $\alpha_{t,\text{new}} = \alpha_t + \delta$ value, with the following acceptance rate.

$$\min\left\{1, \frac{p(\alpha_{t,\text{new}} \mid \gamma_t, V, L_t, I_t)}{p(\alpha_t \mid \gamma_t, V, L_t, I_t)}\right\}$$

Users can monitor the acceptance rate for α_t and tune the τ parameter to achieve the desirable acceptance rate around 0.45 [12,45]. After sampling α_t, we can use an logarithm transformation to directly get z_t by $z_t = \ln(\alpha_t)$.

4.3. Sample $\{\gamma_t\}_{t=0}^T, V, W \mid z_t, m_0, H_0, r_0, s_0$

Based on the random walk DLM prior imposed on z_t's, we use a Gibbs sampler to draw $\{\gamma_t\}_{t=0}^T$, V and W, given the latest draws of z_t. We use the efficient Forward Filtering Backward Sampling (FFBS) algorithm for this Gibbs sampler. Interested readers can refer to [11,17] or [53] for the detailed proof for the FFBS algorithm. We lay out the sampling steps below.

Let D_t denote all the information available up to period t, and we know $D_t = \{D_{t-1}, z_t\}$. For period 0, we know $D_0 = \{m_0, H_0, V, W\}$. We exploit the factorization of the joint probability of $\{\gamma_t\}_{t=1}^T$:

$$p(\{\gamma_t\}_{t=0}^T \mid \{D_t\}_{t=0}^T) = p(\gamma_T \mid D_T) p(\gamma_{T-1} \mid \gamma_T, D_{T-1}) \cdots p(\gamma_1 \mid \gamma_2, D_1) p(\gamma_0 \mid \gamma_1, D_0)$$

We can thus draw a joint sample of $\{\gamma_t\}_{t=0}^T$ by sampling $p(\gamma_T \mid D_T)$ first, then we can subsequently draw γ_t by $p(\gamma_t \mid \gamma_{t+1}, D_t)$ for $t = T - 1, T - 2, \ldots, 1, 0$.

Given the current values of $\{z_t\}_{t=1}^T$, V and W, and the fixed values of m_0 and H_0, we start the forward filtering procedure. For $t = 1, 2, \ldots, T$, we compute the following quantities in sequence. First, we compute, a_t, the prior mean of γ_t, given D_{t-1} by $a_t = E[\gamma_t \mid D_{t-1}] = E[\gamma_{t-1} + e_1 \mid D_{t-1}] = m_{t-1}$. Second, we compute, R_t, the prior variance of γ_t, given D_{t-1} by $R_t = Var[\gamma_t \mid D_{t-1}] = Var[\gamma_{t-1} + e_1 \mid D_{t-1}] = H_{t-1} + W$. Third, we compute, f_t, the prior mean of the one-step forecast of z_t, given D_{t-1} by $f_t = E[z_t \mid D_{t-1}] = E[\gamma_t + e_2 \mid D_{t-1}] = a_t$. Fourth, we compute, Q_t, the prior variance of the one-step forecast of z_t, given D_{t-1} by $Q_t = Var[z_t \mid D_{t-1}] = Var[\gamma_t + e_2 \mid D_{t-1}] = R_t + V$. Fifth, we

compute, $m_t = E[\gamma_t \mid D_t]$, the posterior mean of γ_t given D_t as follows:

$$m_t = E[\gamma_t \mid D_{t-1}] + \frac{Var[\gamma_t \mid D_t]}{Var[z_t \mid D_{t-1}]}(z_t - E[z_t \mid D_{t-1}]) = a_t + \frac{R_t}{Q_t}(z_t - f_t)$$

Sixth, we compute, $H_t = Var[\gamma_t \mid D_t]$, the posterior variance of γ_t given D_t as follows:

$$H_t = Var[\gamma_t \mid D_{t-1}] - \left(\frac{Var[\gamma_t \mid D_{t-1}]}{Var[z_t \mid D_{t-1}]}\right)^2 Var[z_t \mid D_{t-1}] = R_t - \left(\frac{R_t}{Q_t}\right)^2 Q_t$$

After the above forward filtering procedure, we store the following values: $\{a_t\}_{t=1}^T$, $\{R_t\}_{t=1}^T$, $\{f_t\}_{t=1}^T$, $\{Q_t\}_{t=1}^T$, $\{m_t\}_{t=1}^T$, and $\{H_t\}_{t=1}^T$. We start the backward sampling procedure from $t = T$, since in the factorization of $p(\{\gamma_t\}_{t=0}^T)$, the PDF of γ_T is only conditioning on D_T. We can directly draw γ_T from the conditional posterior below:

$$\gamma_T \mid D_T \sim N(m_T, H_T)$$

Sequentially, for $t = T - 1, \ldots, 2, 1, 0$, we compute, $u_t = E[\gamma_t \mid \gamma_{t+1}, D_t]$, the mean of the conditional distribution of γ_t, given γ_{t+1} and D_t, as below.

$$u_t = E[\gamma_t \mid D_t] + \frac{Var[\gamma_t \mid D_t]}{Var[\gamma_{t+1} \mid D_t]}(\gamma_{t+1} - E[\gamma_{t+1} \mid D_t]) = m_t + \frac{H_t}{R_{t+1}}(\gamma_{t+1} - a_{t+1})$$

Then we compute, $U_t = Var[\gamma_t \mid \gamma_{t+1}, D_t]$, the variance of the conditional distribution of γ_t, given γ_{t+1} and D_t, as below.

$$U_t = Var[\gamma_t \mid D_t] - \left(\frac{Var[\gamma_t \mid D_t]}{Var[\gamma_{t+1} \mid D_t]}\right)^2 Var[\gamma_{t+1} \mid D_t] = H_t - \left(\frac{H_t}{R_{t+1}}\right)^2 R_{t+1}$$

Given the values of u_t and U_t, we update γ_t by its full conditional posterior below.

$$\gamma_t \mid \gamma_{t+1}, D_t \sim N(u_t, U_t)$$

Next, we update the value of V, given the updated values of $\{\gamma_t\}_{t=1}^T$ and $\{z_t\}_{t=1}^T$. V has a conjugate Inverse-Gamma prior distribution, $V \sim Inverse - Gamma(\frac{r_0}{2}, \frac{s_0}{2})$. The conditional posterior distribution of V is expressed below.

$$V \mid \{\gamma_t\}_{t=1}^T, \{z_t\}_{t=1}^T \sim Inverse - Gamma\left(\frac{r_0 + T}{2}, \frac{s_0 + \sum_{t=1}^T (z_t - \gamma_t)^2}{2}\right)$$

Lastly, we sample the value of W, given the updated values of $\{\gamma_t\}_{t=0}^T$. W has a non-conjugate Gamma prior distribution, $W \sim Gamma(r_1, s_1)$. We can express the conditional posterior of W as follows:

$$p(W \mid \{\gamma_t\}_{t=0}^T, r_1, s_1) \propto \underbrace{W^{r_1 - 1} \exp(-s_1 W)}_{\text{Gamma prior kernel}} \underbrace{\prod_{t=1}^T \frac{1}{\sqrt{W}} \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2W}\right)}_{\text{conditional likelihood kernel}}$$

The conditional posterior of W does not have a kernel of any recognizable distribution. We use a unidimensional random walk Metropolis-Hastings sampler to sample W.

Users supply the positive value τ_W to generate the random walk step, δ_W, from $\delta_W \sim$ uniform$(-\tau_W, \tau_W)$ for updating W. We accept the proposed new $W_{\text{new}} = W + \delta_W$ value, with the following acceptance rate. Users can monitor the acceptance rate for W and tune the τ_W parameter to achieve the desirable acceptance rate around 0.45.

$$\min \left\{ 1, \frac{p(W_{\text{new}} \mid \{\gamma_t\}_{t=0}^T, r_1, s_1)}{p(W \mid \{\gamma_t\}_{t=0}^T, r_1, s_1)} \right\}$$

5. Post-processing

Due to the label switching problem present in any DPM model, we are not able to use simple statistical summaries of the exact draws of cluster labels for analysis [28]. Instead, we propose a maximum a posteriori (MAP) estimation method to extract fixed cluster labels for voters. This method draws on the post-processing procedure proposed for both finite and infinite mixture models in the model-based clustering literature[18,19]. We compute the MAP cluster labels for each period t separately, so we drop subscript t in the following illustration for simplification.

5.1. Compute the cluster number mode and subset the posterior samples

In the posterior sample of cluster labels, $\{\{c_{i,\text{iter}}\}_{i=1}^I\}_{\text{iter}=1}^N$, the number of unique clusters, L_{iter}, varies across iterations, which reflects DPM's strength in letting the data decide how many clusters there should be. We compute the mode of cluster numbers, L_{mode}, across iterations in the posterior sample. Then, we extract a subset of the cluster label posterior sample iterations, $\{\{c_{i,\text{iter}}\}_{i=1}^I\}_{\text{iter}\in\{1,2,\dots,N \mid L_{\text{iter}}=L_{\text{mode}}\}}$, where the number of clusters is equal to the mode, $L_{\text{iter}} = L_{\text{mode}}$. We denote $\{\{c_{i,\text{iter}}\}_{i=1}^I\}_{\text{iter}\in\{1,2,\dots,N \mid L_{\text{iter}}=L_{\text{mode}}\}}$ as $\{\{c_{i,\text{iter}}\}_{i=1}^I\}_{L_{\text{iter}}=L_{\text{mode}}}$. By using the mode as the subseting criterion, we let the algorithm decide the number of clusters in our final result. Working with the above subset of the cluster label posterior samples, we have a fixed number of clusters, L_{mode}, across iterations. Therefore, we can use the likelihood of a finite mixture model as an approximate likelihood function, in order to get the MAP cluster labels in the next step.

5.2. Identify the MAP cluster labels

We compute the likelihood for each iteration in $\{\{c_{i,\text{iter}}\}_{i=1}^I\}_{L_{\text{iter}}=L_{\text{mode}}}$. First, we compute, w_l, the marginal probability for a voter to belong to cluster, l, by the ratio of cluster l's size, n_l, to the number of voters, I: $w_l = \frac{n_l}{I}$. Second, for each cluster l, we can compute the posterior Bernoulli probability parameter vector, $\tilde{\boldsymbol{\theta}}_l$, based on the voters' voting records, who belong to cluster l. For a specific bill j, we compute the Bernoulli parameter by $\tilde{\theta}_{lj} = \frac{\lambda_{1j}+n_{lj1}}{\lambda_{1j}+\lambda_{0j}+n_{lj1}+n_{lj0}}$, where n_{lj1} is the count of yea vote on bill j from all the voters belonging to cluster l, n_{lj0} is the count of nay vote on bill j from all the voters belonging to cluster l, and λ_{1j} and λ_{0j} are the Beta prior parameters supplied by users.

We compute the approximate posterior likelihood for a cluster label posterior sample iteration as follows:

$$p(\{c_{i,\text{iter}}\}_{i=1}^I \mid \boldsymbol{V}, \boldsymbol{\lambda}) = \prod_{i=1}^I \left(\prod_l \left[w_l \left(\prod_{j=1}^J \tilde{\theta}_{lj}^{\mathbf{I}(v_{ij}=1)} (1 - \tilde{\theta}_{lj})^{\mathbf{I}(v_{ij}=0)} \right) \right]^{\mathbf{I}(c_{i,\text{iter}}=l)} \right)$$

We compute the above approximate posterior likelihood for each iteration in $\{\{c_{i,\text{iter}}\}_{i=1}^I\}_{L_{\text{iter}}=L_{\text{mode}}}$, and use the iteration of cluster label draws, $\{c_{i,\text{iter}}\}_{i=1}^I$, corresponding to the largest approximate posterior likelihood value, as the MAP cluster labels. The MAP cluster labels are expressed as below.

$$\underset{\{c_{i,\text{iter}}\}_{i=1}^I \in \left\{\{c_{i,\text{iter}}\}_{i=1}^I\right\}_{L_{\text{iter}}=L_{\text{mode}}}}{\operatorname{argmax}} p(\{c_{i,\text{iter}}\}_{i=1}^I \mid \boldsymbol{V}, \boldsymbol{\lambda})$$

5.3. Visualization of voter-pair posterior probability of being in the same cluster

In addition to the MAP cluster labels, we propose a visualization method to present the clustering structure in a time period. Based on the posterior sample of cluster labels, we compute the posterior probability for any pair of voters, i_1 and i_2, to be in the same cluster across iterations in time t. These probabilities are stored in an $I_t \times I_t$ symmetric square matrix, M_t, for period t, and element (i_1, i_2) of M_t indicates the posterior probability for voters, i_1 and i_2, to be in the same cluster. We use Ward's Hierarchical Agglomerative Clustering Method to reorder the rows and columns of M_t [37]. We use the Ward's dissimilarity criterion to move the similar columns and rows to be adjacent to each other. `hclust()` function in R implements this algorithm with various distance criteria, and we use `ward.D2` distance criterion for this study. The resulting reordered matrix M_t' has anti-diagonal blocks storing larger probability values and off-anti-diagonal elements storing lesser probability values. We plot matrix M_t' like a heat map to visualize the clustering structure.

6. Simulation study

To demonstrate the effectiveness of the DDPM model, we run the DDPM model on two simulated data sets of multiple periods. In one data set, the numbers of clusters are relatively stable across time (stable-cluster-number example), and in the other data set, the numbers of clusters fluctuate in a more volatile way (volatile-cluster-number example). For both data sets, the DDPM model has high accuracy at both identifying the correct cluster numbers and classifying the relationship of unique voter pairs in each period. We caution that it's impossible to determine the 'true' number of clusters or cluster affiliations in empirical data sets, and thus we can not expect the DDPM model to uncover the 'truth' in these situations. Instead, the usefulness of the DDPM model lies in detecting underlying patterns and deriving helpful representation of empirical data.

Researchers have found that some Dirichlet Process Mixture model-based (DPM model-based) clustering methods have a difficult time in consistently identifying the true cluster number, and they tend to estimate more clusters than the ground truth [36]. To solve this challenge, other researchers have developed alternative sparse finite mixture models to induce sparsity in cluster number estimation, leading to more accurate identification of cluster numbers [32,33]. More recently, researchers have found that a DPM model and a

sparse finite mixture model yield similar estimation results with respect to the number of clusters, as long as the hyperparameter specifications in the two models induce a similar degree of sparsity [19]. In response to the above general concern about DPM model-based clustering methods, we conduct various robustness checks. We run robustness checks for a stable-cluster-number example and a volatile-cluster-number example. The robustness check results show that the DDPM model is able to consistently identify the true cluster numbers under various model specifications.

For the details of the demonstrative examples and the robustness check studies, interested readers should refer to the 'Simulation Study' section of the supplemental information document.

7. Human rights votes in the UNGA

We apply the proposed DDPM model and post-processing methods to the United Nations General Assembly (UNGA) human rights roll call vote data from 1992 to 2017. We rely on the publicly available 'United Nations General Assembly Voting Data' for this study [52]. The original UNGA roll call vote data record 'Yes', 'No', 'Abstain', or 'Missing' for a voting record. Voeten argues that 'UNGA resolutions are not binding, what really matters is whether or not a state is willing to go on the record for supporting a resolution.' 'No' votes and 'Abstain' votes 'both are essentially ways a state can express its unwillingness to comply with the text of a resolution [50].' Therefore, in this study, we record 'Yes' votes as 1, 'Missing' as NA, and both 'No' and 'Abstain' votes as 0. Moreover, in each year, we only include countries that vote on at least 95% of the human rights resolutions in the year. Table 1 reports the summary statistics of the numbers of countries and resolutions throughout the years.

We use the following specifications for running the DDPM model on human rights vote data. We run the SAMS sampler every third iteration of the MCMC sampling scheme. We supply the following fixed values. All elements in $\boldsymbol{\lambda}_t$'s are fixed at 1. The tuning parameter for the random walk Metropolis-Hastings step for drawing α_t's, τ, is 0.5. We set the parameters in the DLM priors as follow: $m_0 = 0, H_0 = 0.1, r_0 = 10, s_0 = 0.5, r_1 = 1, s_1 = 50$. The tuning parameter for the random walk Metropolis-Hastings step for drawing W, τ_W, is 0.01. We use the following values as starting values for V, W and all α_t's: $V_{start} = 0.05$, $W_{start} = 0.02$, and $\alpha_{t,start} = 1$ for $t = 1, 2, \ldots, T$. We run 10000 iterations and discard

Table 1. Summary statistics of the UNGA human rights vote data.

Summaries	Number of countries	Number of resolutions
Min.	99.0	11.00
1st Qu.	119.2	17.25
Median	134.0	21.00
Mean	131.3	20.50
3rd Qu.	141.0	24.00
Max.	156.0	29.00

Note: The table reports the summary statistics for the numbers of countries and the numbers of resolutions in the UNGA human rights roll call vote data from 1992 to 2017.
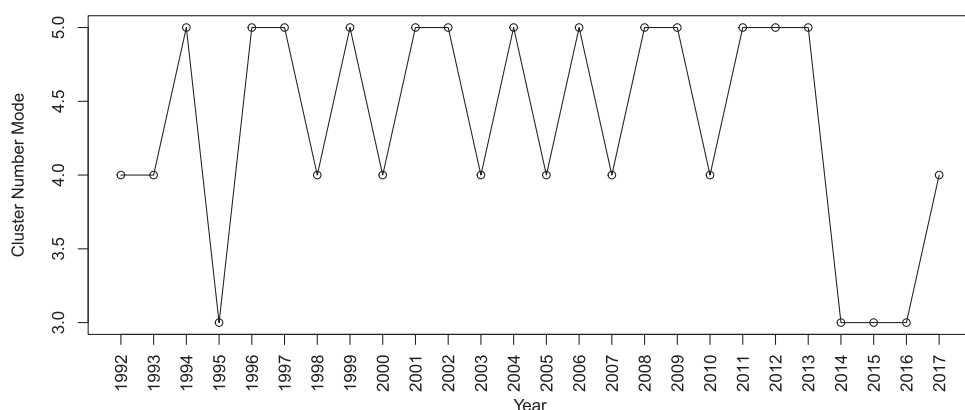
Figure 1. *Annual Cluster Number Mode.* For each year, we compute the mode of the numbers of unique clusters across iterations in the posterior sample. The annual cluster number modes represent our estimates of the annual voting coalition numbers, based on the DDPM model.

the first 5000 iterations as burnin. The acceptance rates of the random walk Metropolis-Hastings steps for drawing α_t's lie in the interval, [0.3871, 0.4661]. The acceptance rate of the random walk Metropolis-Hastings steps for drawing W is 0.344. These acceptance rates are all acceptable for a unidimensional Metropolis-Hastings sampler.

We post-process the posterior samples of cluster labels in each year, and report the year-wise cluster number modes in Figure 1. For all years, the numbers of coalitions vary between 3 and 5. There is no obvious time trend of the cluster numbers across years. The relative stability of coalition numbers show preliminary evidence for 'peer groups' in the UNGA human right vote data. It also shows that there is no universal consensus on human rights issues in the UNGA. The persistent existence of coalitions lends support to the criticism of human rights politicization in the UNGA.

For each year, we estimate the MAP cluster labels for countries. In the supplemental information document, we report the voting coalition membership for each year from 1992 to 2017. Due to space constraints, we only show the coalition membership in the most recent year, 2017, in Table 2. In 2017, the DDPM model identifies four coalitions. The most obvious pattern is the separation between developed and developing countries. There are two coalitions consisting of developed countries. Most western European countries and the developed countries in Asia-Pacific, such as Japan, South Korea, and New Zealand, form a coalition. The US, Israel, and two other close US allies form a coalition. The developing countries are also divided into two coalitions: a large coalition including most developing countries, and a small coalition joined mostly by a few Latin American countries.

Figure 2 shows the posterior probability for unique country pairs to be in the same cluster in 2017. The plot confirms the four coalitions reported in the Table 2. The heat map plot shows that Australia is likely to belong to both the smaller US-led coalition and the bigger EU-led coalition. Similarly, Russia and Argentina are likely to belong to the bigger developing countries' coalition and the smaller developing countries' coalitions. However, there is no country that is likely to belong to both a developing country coalition and a developed country's coalition. This shows that the differences between developing countries and

Table 2. Voting coalition membership in 2017.

Coalition	Member
(1)	Cameroon, Colombia, Guatemala, Honduras, Mexico, Panama, Paraguay, Peru, Solomon Islands, Togo
(2)	Australia, Canada, Israel, United States of America
(3)	Albania, Andorra, Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Monaco, Montenegro, Netherlands, New Zealand, Norway, Poland, Portugal, Republic of Korea, Republic of Moldova, Romania, San Marino, Slovakia, Slovenia, Spain, Sweden, Switzerland, The former Yugoslav Republic of Macedonia, Turkey, Ukraine, United Kingdom of Great Britain and Northern Ireland
(4)	Algeria, Angola, Argentina, Bahamas, Bahrain, Bangladesh, Bhutan, Bolivia (Plurinational State of), Botswana, Brazil, Brunei Darussalam, Cabo Verde, Cambodia, Chile, China, Congo, Costa Rica, Côte D'Ivoire, Cuba, Djibouti, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Guinea, Guyana, India, Indonesia, Iraq, Jamaica, Jordan, Kazakhstan, Kenya, Kuwait, Kyrgyzstan, Lao People's Democratic Republic, Lesotho, Libya, Malaysia, Maldives, Mali, Mauritania, Mauritius, Morocco, Mozambique, Namibia, Nepal, Nicaragua, Oman, Pakistan, Philippines, Qatar, Russian Federation, Saint Kitts and Nevis, Saint Lucia, Saudi Arabia, Sierra Leone, Singapore, South Africa, Sri Lanka, Sudan, Syrian Arab Republic, Tajikistan, Thailand, Trinidad and Tobago, Tunisia, Uganda, United Arab Emirates, United Republic of Tanzania, Uruguay, Uzbekistan, Venezuela, Bolivarian Republic of, Viet Nam, Yemen, Zimbabwe

Note: Four voting coalitions are identified in 2017. Coalition 3 is the EU-led big developed country coalition. Coalition 2 is the US-led small developed country coalition. Coalition 4 is the big developing country coalition including most countries in Asia and Africa. Coalition 1 is the small developing country coalition, mostly consisting of Latin American countries. Coalition 1, 2, 3, 4 corresponds to the four anti-diagonal blocks from the bottom-left corner to the top-right corner in Figure 2.

developed countries are much starker than those among developing countries themselves or developed countries themselves.

After examining the coalition memberships from 1992 and 2017, we are able to find some consistent patterns. These findings help us reevaluate some of the arguments in the literature of human rights voting in the UNGA. First, across all the years, there are stable coalition structures in the voting data. This finding lends support to the long-lasting discussion of the politicization of human rights issues in the UNGA. Ideally, human rights debates in the UNGA should be based on principles and international laws, and consensus building on human rights issues should contribute to global governance. However, the voting records show that different country coalitions take contradictory positions, and countries are divided in the political debates of human right issues.

Second, throughout the years, the EU member countries always coalesce in the same group. This finding lends support to the previous studies that treat the EU as a voting bloc [30]. In addition to the EU member countries, a few developed countries in Asia-Pacific, including Japan, South Korea, and New Zealand, are also consistently a part of the EU-led group. Therefore, we should update the previous understanding of treating EU as a bloc, but redefine the EU-led group to include more developed countries. In some years, the US, Israel, Canada, and Australia are part of the EU-led group, but they form their own smaller US-led group in other years. This finding shows that US and Israel agree with the EU-led coalition in many aspects of human rights issues, but there is also some repeated disagreement between the US-led coalition and the EU-led coalition. Among the developing countries, there is always a big developing country coalition, including most developing countries from Asia and Africa. In addition, there also exists a smaller developing country coalition, including Latin American countries and former Soviet Union member countries.
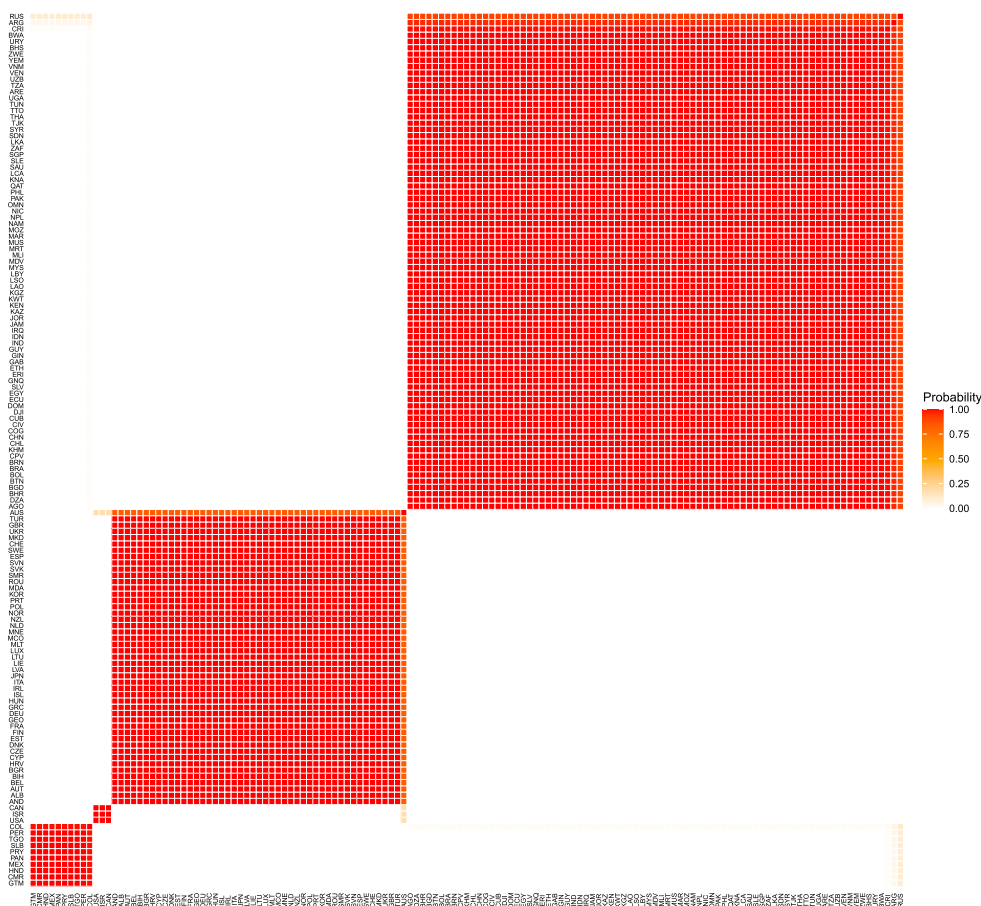
Figure 2. *Posterior Probability for Country Pairs Being in the Same Coalition in 2017.* Four voting coalitions emerge from the heat map plot. The four anti-diagonal blocks from the bottom-left corner to the top-right corner correspond to Coalition 1, 2, 3, 4 in Table 2. The vast majority of countries belong to one coalition with very high probability. Australia is likely to belong to both Coalition 2 and Coalition 3. Russia and Argentina are likely to affiliate to both Coalition 1 and Coalition 4.

The coalition pattern among the developing countries contradicts previous studies that treat the Organization of Islamic Cooperation member countries as a bloc [7,26]. As shown by the clustering results, most of the Muslim countries are a part of the big developing country coalition. Moreover, the coalition patterns among developing countries also corroborate some of the previous arguments that emphasize regional voting blocs. As shown by the smaller developing country coalition, we find that Latin American Countries break with the big developing country coalition to form their own smaller coalition in some years. There is also evidence showing that Russia and other former Soviet Union member countries are apart from the big developing country coalition and end up in a smaller coalition in some years.

We examine what specific human rights issues divide countries into different coalitions. Based on the MAP cluster estimators and each country's voting records in year t, we are able to compute the probability vector, $\tilde{\boldsymbol{\theta}}_{lt}$, for each cluster l. There are J_t resolutions in year t. For

Table 3. Polarizing resolutions in 1992.

Count	Resolution title
(1)	Program of work, Committee on Palestinian rights
(2)	Importance of the universal realization of the rights of peoples to self determination and of the speedy granting of independence to colonial countries and peoples for the effective guarantee and observance of human rights
(3)	Use of mercenaries as a means to violate human rights and to impede the exercise of the right of peoples to self determination United Nations African Institute for the Prevention of Crime and the Treatment of Offenders
(4)	Situation of human rights in the Islamic Republic of Iran
(5)	Encourages all member states to lend assistance to the committee on Palestinian rights
(6)	Reaffirms the inalienable right of all displaced inhabitants to return to their homes or former places of residence in the territories occupied by Israel since 1967

Note: The six resolutions above are the first 50% most polarizing resolutions in 1992. Resolution 1, 4, 5, 6 are country-specific human rights resolutions about Iran or the Palestinian-Israeli issue. Resolution 2, 3 are about general human rights principles, such as self-determination of nations and the use of mercenaries.

each resolution $j \in \{1, 2, \ldots, J_t\}$, we compute the maximum probability difference for any unique cluster pairs, $d_{jt} = \max |\tilde{\theta}_{jlt} - \tilde{\theta}_{jl't}|$, for all $l, l' \in \{c_{it,\mathrm{MAP}}\}_{i=1}^{I_t}$. For some resolutions, countries have almost unanimous votes, so d_{jt} is close to zero. However, other resolutions become the center of controversy, and cluster-wise probability parameters for this resolution can be very different. The numbers of resolutions vary in each year, so we identify the first 50% most polarizing resolutions in terms of d_{jt} values. Due to space constraints, we only report the first 50% most divisive resolutions in 1992 in Table 3. We report the first 50% most polarizing resolutions in all years in the supplemental information document.

We have a few important findings from showing these most polarizing resolutions. First, the Israeli-Palestinian issue is a persistent line of conflict in all years. This finding confirms previous studies that theorize the frequent occurrence of voting on the Israeli-Palestinian issue in the UNGA [6]. Second, country-specific human rights reports tend to be polarizing, such as reports on Cuba and Iran. The reports targeting specific countries clearly show the political nature of human rights issues in the UNGA. Countries are divided in different coalitions by how they view the human rights record of a country, because they prioritize bilateral relations with the targeted country over exclusively examining the human rights record. Third, the UNGA tends to vote on some of the exact same polarizing human rights resolutions for many years. For example, the resolution condemning 'the use of mercenaries to violate human rights' has been a polarizing resolution in almost every year from 1992 until the late 2000s. This resolution-wise repetition also shows evidence for the persistent politicization of certain human rights issues. For the case of the 'use of mercenaries' resolution, developing countries almost all vote for it, while developed countries almost all oppose it. This resolution has become a means that developing countries use to push back against developed countries on their human rights practice.

Last, we compare the DDPM model with a Bayesian Dynamic IRT model for analyzing the UNGA human rights voting data. A Bayesian IRT model is one of the most common model choices for modeling roll call vote data. Bayesian IRT models assume that voters make voting decisions based on the relative utilities between voting yea and nay. In a latent policy space, if an voter's ideal point is closer to the new proposal's position, she is more likely to vote yea; if an actor's ideal point is closer to the status quo, she is inclined to vote nay. In each year, the Dynamic IRT model estimates one continuous ideal point variable

Table 4. DDPM model and Bayesian dynamic IRT model comparison.

Model	Sensitivity	Specificity	F1 Score
DDPM model	0.9511	0.8501	0.8668
Dynamic IRT model	0.9540	0.8127	0.8477

Note: We train a DDPM model and a Bayesian Dynamic IRT model on the same training data set. We use the estimated latent variables to predict the unique voter-pair relationship (being in the same cluster or not) in the hold-out testing data set. The table shows that the DDPM model has a slight edge over the Bayesian Dynamic IRT model in terms of F_1 score.

for each country, and two continuous parameters for each resolution. Here, we use the same model specified in the study of the US Supreme Court justices' ideal points [34]. We implement the model with `MCMCdynamicIRT1d()` function in `MCMCpack` R package [35]. We use the default priors and tuning parameters of the function.

To compare the performances of the DDPM model and the Dynamic IRT model, we first randomly draw one voting record per country-year to form a hold-out testing data set. The remaining data points become the training data set. We train the two models with the training data set, and predict the hold-out voting records with the estimated latent variables, respectively. We use sensitivity, specificity, and $F1$ score to evaluate prediction performance.[6] Table 4 shows the prediction performances of the two models. The Dynamic IRT model is slightly better than the DDPM model at predicting yea votes, while the DDPM model is somewhat better than the Dynamic IRT model at predicting nay votes. In sum, the DDPM model has a larger $F1$ score than the Dynamic IRT model. The comparison of the DDPM model and the Dynamic IRT model shows that the DDPM is at least as a good of a modeling choice as the Dynamic IRT model. This provides further support for modeling voting coalitions instead of individual country's ideal points for the UNGA human rights vote data.

8. Discussion

We propose the DDPM model to identify voting coalitions with roll call votes across multiple periods, along with post-processing methods to analyze the posterior samples from the DDPM model. The proposed post-processing methods provide streamlined steps for applied researchers to make sense of the DDPM outputs and more easily interpret the results. We make the functions implementing the DDPM model and the post-processing methods available to the public in the `MCMCddpmbb` R package.

We apply these methods to the United Nations General Assembly (UNGA) human rights roll call vote data from 1992 to 2017. We identify human rights voting coalitions in the UNGA after the Cold War, and the polarizing resolutions that divide countries into different coalitions. Through this study, we find a clear separation between developing and developed countries in human rights voting. Moreover, we find the EU as a stable coalition among the developed countries, whereas some Latin American countries tend to form a small coalition, apart from a large developing country coalition. We also find that many polarizing resolutions repeatedly show up across years, and the lines of conflict lie in both

debates on general human rights principles and human rights violation reports on specific countries.

Future research plans are in order. One future plan for improving the DDPM model is to investigate how to dynamically model voters' cluster affiliations across time directly. Another place for future improvement is to generalize the binary voting record assumption to accommodate ordinal voting records, such as 'Nay', 'Abstain', and 'Yea'. Last, it's worth investigating how a sparse finite mixture model performs at producing interpretable results in the roll call vote context. A sparse finite mixture model option can be a potentially useful addition to the `MCMCddpmbb` R package.

Notes

1. There are studies aimed at detecting communities based the entire UNGA votes in the network literature [31,41,44]. These studies illustrate some general community structures in the UNGA, but do not engage the specific topic of UNAG human rights votes.
2. In the extreme case where every voter is in her own coalition (cluster), DPM model jointly models NJ Bernoulli trials, which is equivalent to estimating $N \times J$ Bernoulli parameters, for a $N \times J$ voting matrix.
3. If $v_{ijt} = NA$, then both of the indicator functions, $\mathbf{I}(v_{ijt} = 1)$ and $\mathbf{I}(v_{ijt} = 0)$, are equal to 0. As a result, $F(v_{ijt} \mid \theta_{ljt}) = \theta_{ljt}^{\mathbf{I}(v_{ijt}=1)} (1 - \theta_{ljt})^{\mathbf{I}(v_{ijt}=0)}$ is a constant. It does not matter if we put voter i in any existing cluster or a newly generated cluster, $F(v_{ijt} \mid \theta_{ljt})$ is the same constant. Therefore, v_{ijt} does not affect voter i's cluster affiliation sample draws.
4. The updated Beta prior, $H_{-i,l}(\boldsymbol{\theta}_{lt})$, is based on the Dirichlet Process prior G_{0t} and all the observed voting records of the voters belonging to cluster l. We put a conjugate Beta prior for θ_{ljt}, and we only use observed voting records of the voters in cluster l for updating the Beta prior of θ_{ljt}. A missing voting record for any voter belonging to cluster l does not affect this process.
5. Please refer to section, 'Example: Bernoulli Data with a Conjugate Beta Prior', in [27] for the derivations of the Gibbs sampler for c_i. We adapt the notations in [27] to express the formula in this paper.
6. In the roll vote context, sensitivity represents the share of the yea votes that are correctly predicted as a yea vote, and specificity represents the share of the nay votes that are correctly predicted as a nay vote. $F1$ score is a balanced prediction performance metric, considering both sensitivity and precision. In the roll call vote context, precision is the share of the correct predictions among all the yea vote predictions. $F1$ score is a function of sensitivity and precision: $F1 = \frac{2 \times \text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}}$.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Qiushi Yu http://orcid.org/0000-0003-3011-4765

References

[1] D.J. Aldous, *Exchangeability and related topics*, in *École d'Été de Probabilités de Saint-Flour XIII–1983*, Springer, Berlin, Heidelberg, 1985, pp. 1–198.
[2] C.E. Antoniak, *Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems*, Ann. Statist. 2 (1974), pp. 1152–1174.

[3] M.A. Bailey, *Is today's court the most conservative in sixty years? challenges and opportunities in measuring judicial preferences*, J. Polit. 75 (2013), pp. 821–834.

[4] M.A. Bailey, A. Strezhnev, and E. Voeten, *Estimating dynamic state preferences from United Nations Voting Data*, J. Conflict Resolut. 61 (2017), pp. 430–456.

[5] M.A. Bailey and E. Voeten, *A two-dimensional analysis of seventy years of United Nations voting*, Public Choice 176 (2018), pp. 33–55.

[6] R.N. Becker, A.L. Hillman, N. Potrafke, and A.H. Schwemmer, *The preoccupation of the United Nations with Israel: Evidence and theory*, Rev. Int. Organ. 10 (2015), pp. 413–437.

[7] A. Besant and S. Malo, *Dim prospects for the United Nations Human Rights Council*, Yale J. Int'l Aff. 4 (2009), p. 144.

[8] A. Bitto and S. Frühwirth-Schnatter, *Achieving shrinkage in a time-varying parameter model framework*, J. Econom. 210 (2019), pp. 75–97.

[9] D. Blackwell and J.B. MacQueen, *Ferguson distributions via pólya urn schemes*, Ann. Statist. 1 (1973), pp. 353–355.

[10] B. Boockmann and A. Dreher, *Do human rights offenders oppose human rights resolutions in the United Nations?*, Public Choice 146 (2011), pp. 443–467.

[11] C.K. Carter and R. Kohn, *On Gibbs sampling for state space models*, Biometrika 81 (1994), pp. 541–553.

[12] S. Chib and E. Greenberg, *Understanding the Metropolis-Hastings algorithm*, Am. Stat. 49 (1995), pp. 327–335.

[13] D.B. Dahl, *An improved merge-split sampler for conjugate Dirichlet process mixture models*, Tech. Rep. 1086, 2003.

[14] M.D. Escobar and M. West, *Bayesian density estimation and inference using mixtures*, J. Am. Stat. Assoc. 90 (1995), pp. 577–588.

[15] M.D. Escobar and M. West, *Computing nonparametric hierarchical models*, in *Practical Nonparametric and Semiparametric Bayesian Statistics*, Springer, New York, NY, 1998, pp. 1–22.

[16] R. Freedman, *The United Nations Human Rights Council: A Critique and Early Assessment*, Routledge, New York, NY, 2013.

[17] S. Frühwirth-Schnatter, *Data augmentation and dynamic linear models*, J. Time Ser. Anal. 15 (1994), pp. 183–202.

[18] S. Frühwirth-Schnatter, *Finite Mixture and Markov Switching Models*, Springer Science & Business Media, New York, NY, 2006.

[19] S. Frühwirth-Schnatter and G. Malsiner-Walli, *From here to infinity: Sparse finite versus Dirichlet process mixtures in model-based clustering*, Adv. Data Anal. Classif. 13 (2019), pp. 33–64.

[20] I.C. Gormley and T.B. Murphy, *A mixture of experts model for rank data with applications in election studies*, Ann. Appl. Stat. 2 (2008), pp. 1452–1477.

[21] I.C. Gormley and T.B. Murphy, *Exploring voting blocs within the Irish electorate: A mixture modeling approach*, J. Am. Stat. Assoc. 103 (2008), pp. 1014–1027.

[22] D.A. Habibi, *Human rights and politicized human rights: A utilitarian critique*, J. Human Rights 6 (2007), pp. 3–35.

[23] S. Holloway, *Forty years of United Nations General Assembly voting*, Canad. J. Polit. Sci. 23 (1990), pp. 279–296.

[24] T. Hovet, *Bloc politics in the United Nations*, Harvard University Press, Cambridge, MA, 1960.

[25] S. Hug, *Dealing with human rights in international organizations*, J. Human Rights 15 (2016), pp. 21–39.

[26] S. Hug and R. Lukács, *Preferences or blocs? Voting in the United Nations Human Rights Council*, Rev. Int. Organ. 9 (2014), pp. 83–106.

[27] S. Jain and R.M. Neal, *A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model*, J. Comput. Graph. Stat. 13 (2004), pp. 158–182.

[28] A. Jasra, C.C. Holmes, and D.A. Stephens, *Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling*, Stat. Sci. 20 (2005), pp. 50–67.

[29] A. Lijphart, *The analysis of bloc voting in the general assembly: A critique and a proposal*, Am. Polit. Sci. Rev. 57 (1963), pp. 902–917.

[30] P. Luif, *EU Cohesion in the UN General Assembly*, Vol. 49, European Union Institute for Security Studies Paris, Paris, 2003.

[31] K.T. Macon, P.J. Mucha, and M.A. Porter, *Community structure in the United Nations General Assembly*, Phys. A Statist. Mech. Appl. 391 (2012), pp. 343–361.

[32] G. Malsiner-Walli, S. Frühwirth-Schnatter, and B. Grün, *Model-based clustering based on sparse finite Gaussian mixtures*, Stat. Comput. 26 (2016), pp. 303–324.

[33] G. Malsiner-Walli, S. Frühwirth-Schnatter, and B. Grün, *Identifying mixtures of mixtures using Bayesian estimation*, J. Comput. Graph. Stat. 26 (2017), pp. 285–295.

[34] A.D. Martin and K.M. Quinn, *Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999*, Polit. Anal. 10 (2002), pp. 134–153.

[35] A.D. Martin, K.M. Quinn, and J.H. Park, *Mcmcpack: Markov chain Monte Carlo in R*, J. Stat. Softw. 42 (2011), pp. 1–21. Available at https://www.jstatsoft.org/v042/i09.

[36] J.W. Miller and M.T. Harrison, *A simple example of Dirichlet process mixture inconsistency for the number of components*, Adv. Neural Inf. Process. Syst. 26 (2013), pp. 199–206.

[37] F. Murtagh and P. Legendre, *Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion?*, J. Classif. 31 (2014), pp. 274–295.

[38] R.M. Neal, *Markov chain sampling methods for Dirichlet process mixture models*, J. Comput. Graph. Stat. 9 (2000), pp. 249–265.

[39] H. Newcombe, M. Ross, and A.G. Newcombe, *United Nations voting patterns*, Int. Organ. 24 (1970), pp. 100–121.

[40] R. Normand and S. Zaidi, *Human Rights at the UN: The Political History of Universal Justice*, Indiana University Press, Bloomington, IN, 2008.

[41] S.D. Pauls and S.J. Cranmer, *Affinity communities in United Nations voting: Implications for democracy, cooperation, and conflict*, Phys. A Statist. Mech. Appl. 484 (2017), pp. 428–439.

[42] G. Petris, S. Petrone, and P. Campagnoli, *Dynamic linear models*, in *Dynamic Linear Models with R*, Springer, New York, NY, 2009, pp. 31–84.

[43] J. Pitman, *Some developments of the Blackwell-Macqueen urn scheme*, Lecture Notes-Monograph Series, Institute of Mathematical Statistics, Beachwood, OH, 1996, pp. 245–267.

[44] C. Pomeroy, N. Dasandi, and S.J. Mikhaylov, *Multiplex communities and the emergence of international conflict*, PLoS One 14 (2019).

[45] G.O. Roberts, A. Gelman, and W.R. Gilks, *Weak convergence and optimal scaling of random walk Metropolis algorithms*, Ann. Appl. Probab. 7 (1997), pp. 110–120.

[46] D.B. Rubin, *Inference and missing data*, Biometrika 63 (1976), pp. 581–592.

[47] K.E. Schnakenberg and C.J. Fariss, *Dynamic patterns of human rights practices*, Polit. Sci. Res. Methods 2 (2014), pp. 1–31.

[48] A. Spirling and K. Quinn, *Identifying intraparty voting blocs in the UK House of Commons*, J. Am. Stat. Assoc. 105 (2010), pp. 447–457.

[49] D. Trump, *Remarks by President Trump to the 73rd session of the United Nations General Assembly* (2018). Available at https://www.whitehouse.gov/briefings-statements/remarks-president-trump-73rd-session-united-nations-general-assembly-new-york-ny/.

[50] E. Voeten, *Clashes in the assembly*, Int. Organ. 54 (2000), pp. 185–215.

[51] E. Voeten, *Resisting the lonely superpower: Responses of states in the United nations to US dominance*, J. Polit. 66 (2004), pp. 729–754.

[52] E. Voeten, A. Strezhnev, and M. Bailey, *United Nations General Assembly Voting Data* (2009). Available at https://doi.org/10.7910/DVN/LEJUQZ.

[53] M. West and J. Harrison, *Bayesian Forecasting and Dynamic Models*, Springer Science & Business Media, New York, NY, 2006.