BAYESIAN INFERENCE FOR PARTIAL ORDERS FROM RANDOM LINEAR EXTENSIONS: POWER RELATIONS FROM 12TH CENTURY ROYAL ACTA

BY GEOFF K. NICHOLLS[1,a] JEONG EUN LEE[2,b] NICHOLAS KARN[3,c] DAVID JOHNSON [4,d] RUKUANG HUANG [5,e] AND ALEXIS MUIR-WATT [6,f]

[1]*Department of Statistics, The University of Oxford, UK.,* [a]*nicholls@stats.ox.ac.uk*

[2]*Department of Statistics, The University of Auckland, New Zealand ,* [b]*kate.lee@auckland.ac.nz*

[3]*Faculty of Arts and Humanities, University of Southampton, UK,* [c]*N.E.Karn@soton.ac.uk*

[4]*St Peter's college, The University of Oxford, UK ,* [d]*david.johnson@spc.ox.ac.uk*

[5]*Department of Psychiatry, University of Oxford, UK ,* [e]*rukuang.huang@jesus.ox.ac.uk*

[6]*Private researcher, London, UK ,* [f]*alexis.muirwatt@gmail.com*

In the eleventh and twelfth centuries in England, Wales and Normandy, Royal Acta were legal documents in which witnesses were listed in order of social status. Any bishops present were listed as a group. For our purposes, each witness-list is an ordered permutation of bishop names with a known date or date-range. Changes over time in the order bishops are listed may reflect changes in their authority. Historians would like to detect and quantify these changes. There is no reason to assume that the underlying social order which constrains bishop-order within lists is a complete order. We therefore model the evolving social order as an evolving partial ordered set or *poset*.

We construct a Hidden Markov Model for these data. The hidden state is an evolving poset (the evolving social hierarchy) and the emitted data are random total orders (dated lists) respecting the poset present at the time the order was observed. This generalises existing models for rank-order data such as Mallows and Plackett-Luce. We account for noise via a random "queue-jumping" process. Our latent-variable prior for the random process of posets is marginally consistent. A parameter controls poset depth and actor-covariates inform the position of actors in the hierarchy. We fit the model, estimate posets and find evidence for changes in status over time. We interpret our results in terms of court politics. Simpler models, based on Bucket Orders and vertex-series-parallel orders, are rejected. We compare our results with a time-series extension of the Plackett-Luce model. Our software is publicly available.

1. Introduction.

In rank-order data we are presented with a collection of lists ranking a common set of items from best to worst or first to last. A list might order items according to the preferences of an assessor, or the outcome of a multiplayer game, and may rank all elements in the set or just some subset presented to an assessor.

In this paper we analyse a time series of 371 lists recording the order in which 67 different bishops are named as witnesses to legal documents called *Royal Acta*. The data are available online (Sharpe et al., 2014). These documents date from the eleventh and twelfth century (see Supp A.2 for example lists). Just a small subset of the bishops are named in any given list but each bishop is present in many lists. In a list the bishops' names were written down by a clerk in an order that is known to reflect status (henceforth status in the context of witnessing, which might differ from status in other social contexts). The status of a bishop was partly determined by seniority and diocese. The first canon of the Council of London

Keywords and phrases: Partial order, Bayesian analysis, Hidden Markov Model, Royal Acta, Social Hierarchy.

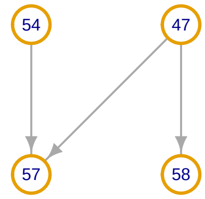

FIG 1. *A poset represented as a transitively closed DAG. In Sec. 7 we find evidence that this suborder contains all precedence relations between (54) Algar, Bishop of Coutances, (57) Adelulf, Bishop of Carlisle, (47) Simon, Bishop of Worcester and (58) Richard, de Beaufeu, Bishop of Avranches in 1136 (numbering as Fig 24).*

in 1075 concerns ecclesiastical precedence: "...each man shall sit according to his date of ordination, except for those who have more honourable seats by ancient custom or by the privileges of their churches" (Clover and Gibson, 1979). However, political standing may have contributed to status, and if it did, then the position of a bishop in the status-hierarchy would not be fully explained by time-in-office and diocese. Russell (1937) writes "The names of Eustace, bishop of Ely, and John, bishop of Norwich, frequently appear at the head of the list of bishops, before the names of the bishops of London and Winchester and of bishops who were consecrated before them... [however] ...they were very close friends of the king during most of his reign and were frequently at court". The dioceses of London and Winchester were nominally above Ely and Norwich so here is a case where some political element seems to count for more than seniority or "honourable seats". This analysis was contested by Haskins (1938) wrote "the precedence of witnesses to private grants is too erratic to serve as an index to their respective station". The question is still open. See Supp A.5 for a review of recent literature written by historians on this topic.

The set of precedence relations we reconstruct determine a *social hierarchy*. This follows the definition given in van Wietmarschen (2022), as the "socially expected behavior" for the clerk is to "value" by status. Precedence relations are transitive inequalities. This holds on social and historical grounds. Precedence is a social dominance relation and transitivity is fundamental to human and animal understanding of dominance (Gazes et al., 2017; Vasconcelos, 2008). Shizuka and McDonald (2012) find evidence for transitivity in 84 of 101 published animal dominance data tables, and it is the norm in models for social and organisational hierarchies (Friedell, 1967; Roberts, 1990), in early models for general preference relations (Bogart, 1973b) and in all analyses which assume an underlying complete order.

Gathering these observations, we represent the social hierarchy of bishops using a Directed Acyclic Graph (DAG). Nodes correspond to bishops and a directed edge indicates precedence. The DAG is transitively closed, so it defines a *partially ordered set* or "*poset*" (see Sec. 3.1.1, the terms are interchangeable). This is more general than assuming the social hierarchy is a total order, as would be the case if we fit a Mallows model. Some pairs of bishops may simply be unordered. The poset will be our parameter, the thing we want to estimate. Fig 1 gives an example of a poset extracted from a larger set of relations reconstructed in Sec. 6.3. The witness lists are our data. The poset constrains the data: bishop B shouldn't appear before bishop A in a witness list if B was ordered below A in the poset. In Sec. 3.1.3 we define an observation model which allows the lists to contradict the order relations in the reconstructed poset. However, we view this as as noise. The underlying structure of the social hierarchy of bishops in the context of witnessing is assumed to be a poset.

The first statistical work to model social networks using posets is Mogapi (2009) who writes "the application of partial orders in Social Network data has not been studied in the past". Martin (2002) used DAGs (which need not be transitively closed) in a similar setting without statistical inference. Friedell (1967) used semilattices, a subclass of posets in which each pair of actors has a unique upper bound, to model organisational hierarchies. Their example, the "Cornerville S & A Club" hierarchy was a poset which only became a semilattice on dropping selected actors. Bogart (1973b) uses partial orders to express preference relations. They outline a model for evolving posets and remark “it would be extremely worthwhile to develop a theory of statistical inference for partial orderings”. Fishburn and Gehrlein (1975) mention social dominance relations as a potential application in their work on posets. However, there seems to have been no statistical development of posets as models for social relations prior to Mogapi (2009) and no other precedent for posets to be used in the way we do. Some of our methods (excluding time series, covariates, and the theory in Sec. 4) were outlined in (Nicholls and Muir Watt, 2011) by two of the present authors and Muir Watt (2015) gives a continuous-time analysis without covariates. However, despite careful design of the particle filtering Monte Carlo, this approach does not seem promising from modelling and computational perspectives. Recent non-parametric work by Jiang and Nicholls (2024) shows how the dimension of the latent space parameterisation of posets may be estimated.

1.1. *Alternative Approaches.* Analysis of rank-order data (ie, our lists) often seeks a total order or actor-ranking which is “central” to the lists in the data, so that many lists are summarised by one complete actor ranking. Mallows models (Mallows, 1957) have a location parameter which is a ranking of the actor labels. In our setting this parameter would be interpreted as the unknown true bishop order. A dispersion parameter controls the distribution of the distance between realised lists and this centre-order. In Generalised Mallows models (Fligner and Verducci, 1986) the dispersion parameter can vary across rank positions. There is freedom in the choice of distance measure between ranking lists. Diaconis (1988) points to Kendall’s-tau as having many good properties. Vitelli et al. (2018) adopt the foot-rule distance on modeling grounds and they and Irurozki et al. (2019) give methods and software (Irurozki et al., 2016; Sørensen et al., 2020) for computing the dispersion-dependent likelihood normalisation. Bayesian methods which allow for variability in the quality of the assessors providing the rankings (Deng et al., 2014) have been given. Mixture-model analysis (Meilă and Chen, 2010; Tkachenko and Lauw, 2016; Vitelli et al., 2018; Lu and Boutilier, 2014) can be used when there is a latent group structure in the population from which the lists are drawn, Meilă and Chen (2016) gives Bayesian methods for analysing non-parametric mixtures of generalised Mallows models and Vitelli et al. (2018) gives unsupervised clustering and treats incomplete lists. Asfaw et al. (2017) give time-series Mallows models.

The actor skill-vector in Plackett-Luce models (Luce, 1959; Plackett, 1975), which we discuss further in Supp J, determines a rank-order parameter which plays a similar role to the Mallows location parameter. Dispersion in the observation space of lists is controlled by the scale of the skill scores. Hunter (2004) gave an EM algorithm (see Caron and Doucet (2012)) for parameter estimation and much useful background theory. Bayesian methods Guiver and Snelson (2009), mixture models (Mollica and Tardella, 2017, 2020), time-series models (Caron and Teh, 2012; Glickman and Hennessy, 2015) and non-parametric Dirichlet process mixture models for clustering (Caron and Teh, 2012; Caron et al., 2014) have been developed. Caron and Doucet (2012) give efficient Monte Carlo methods exploiting data augmentation and conjugate priors. The Contextual Repeated Selection (CRS) model (Seshadri et al., 2020) generalises Plackett-Luce to handle non-transitive relations. This generality comes at a price. If we had ten lists of length two in which bishop A precedes B and ten in which B precedes C then A probably precedes C when they appear in the same list in our poset-based analysis. In the simplest CRS model at least, the twenty lists would provide no information about the ordering of A and C in a list containing only A and C.

Our poset-based parameterisation has strengths and weaknesses when compared to these models. Among the weaknesses, our likelihood evaluation does not scale to data with ordered lists of actors longer than about 20 (though scaling with the number of lists is linear). This is discussed further in Sec. 3.1.2. In response, Jiang et al. (2023) restrict fitted posets to be Vertex-Series-Parallel posets (VSP, Valdes (1978)). These orders and Bucket Orders (in which unordered groups of actors are arranged in a total order) admit likelihood evaluation at a cost which is linear in the list length. Our model comparisons in Sec. 7 generally favor posets over VSPs and Bucket Orders. In other respects, our approach is closest to Placket-Luce as our latent variables play a similar role its skill scores. Our poset-model inherits some of the strengths of Placket Luce: while some Mallows and all Placket-Luce (and our poset) models handle top-k data straightforwardly (where the assessor just ranks their top k preferences, see Sec. 3.1.3), fitting subset-data (where assessors are presented with different subsets for ranking) is more challenging for Mallows models, as the missing ranks have to be treated as missing data (Vitelli et al., 2018). This is not necessary in poset or Plackett-Luce models. On the other hand, Placket-Luce models are "context independent": the probability for actor A to be listed above actor B is independent of the presence or absence of any other actor. Mallows, CRS and our poset-based models are in general context dependent. This property of the observation model defined in Sec. 3.1.2 is discussed further in Supp B.1.

Why is a qualitatively new ranking model needed, given the extensive range of models in the literature? First, we saw above that we have reason to believe the underlying social hierarchy is a partial order; we cannot estimate it reliably without fitting a model in which the parameter is a partial order. The alternative models described above impose a total order structure on the hierarchy; this is not justified and not needed. Secondly, we divide rank-order analyses into two classes: those which aim to reconstruct an underlying true or "physical" order and those in which the fitted order is understood as a heuristic summary of the lists. We work in the former setting. However, the models cited in Sec. 1.1 can be adapted to make heuristic models for our data. In Supp J.1 we specify and fit a Plackett-Luce time-series model with covariates. Glickman and Hennessy (2015); McKeough and Glickman (2024) define a Plackett-Luce model with many of the same features. Some conclusions from our poset-analysis can be obtained by fitting this relatively simpler model. However, our poset-model is simply a better model for our data, in the sense of goodness of fit. In Supp J.2 we estimate the Expected Pointwise Log Posterior Predictive (ELPD) model-selection measure (Vehtari et al., 2017) using Leave-One-Out Cross Validation (LOOCV) for our model and a Plackett-Luce mixture model (Mollica and Tardella, 2017). Our model is preferred.

1.2. *Statistical work with Partial Orders.* The first statistical methods inferring partial orders from list data were given in Mannila and Meek (2000) and Gionis et al. (2006). They treat problems of seriation in archaeology and biochronology in palaeontology and work with the VSP and Bucket Order sub-classes of posets for rapid evaluation of the "noise free" likelihood (see Sec. 3.1.2). Mannila (2008) gives a Bayesian analysis for Bucket Orders. In other important early work Beerenwinkel et al. (2007) define maximum likelihood posets and give a Bayesian analysis in Sakoparnig and Beerenwinkel (2012). They fit a probabilistic graphical model in which genetic mutations accumulate in a total order constrained by a poset. In related work Froehlich et al. (2007) model signaling pathways for gene expression and fit their models using simulated annealing.

In these archaeological and genetic settings there is an unknown true underlying poset and the data are total orders respecting that poset so we have the same data type and similar inferential goals. Our new contributions are as follows. In Sec. 3.1.3 we give a generative model for lists which allows for noise in the realised lists. We idealise the list-observation model as a snapshot of a "queue". In Sec. 3.2, building on work by Winkler (1985), we give marginally consistent priors with a hyper-parameter controlling the "depth" of a random poset. Depth is a quantity of historical interest, so our prior should be non-informative with respect to depth.

This rules out the uniform prior over posets, taken by Sakoparnig and Beerenwinkel (2012), as it is strongly informative of depth (see Supp D). In Sec. 3.2.3 we bring covariates into our model. We have a "linear predictor" which determines an actor's position in the poset. Finally, our list-data are a time series, so the generative model in Sec. 3.3 is a Hidden Markov Model (HMM) with a latent process of posets and "emitted data" which are lists of actors respecting the poset at the time each list was formed.

Posets appear in a range of data-analytic settings. In Mogapi (2009) the data are edges in a directed graph. The edges are noisy observations of the relations in an underlying poset representing information flow in a company, and a prior controls the number of relations in the order, like our focus on prior depth. This prior is not marginally consistent. Gionis et al. (2006) encode list data as a precedence matrix giving the proportion of times any pair of items appear in a given order. A poset has a corresponding precedence matrix, estimated using random *linear extensions* (lists which are total orders respecting the poset). The "distance" between lists and a poset is the distance between their precedence matrices. The estimated poset is a Bucket Order minimising this distance. Arcagni et al. (2022) has poset data and a wider range of otherwise similar loss functions. They fit both posets and Bucket Orders.

In Rising (2021) the poset is a summary statistic, displaying order relations between parameter estimates. Posets are also used for structure discovery in Bayesian Networks (Niinimäki et al., 2016; Kangas et al., 2016), where Bucket Orders support evaluation of marginal likelihoods. The likelihood is written as a sum over total orders respecting a Bucket Order and samples are reweighted by the order-count of the poset in an importance-sampling setup. The same count appears in our likelihood and we evaluate it using the same *lecount()* package (Kangas et al., 2019). However, in our setting the poset is a parameter of interest, not a supporting structure in the computation.

1.3. *Contributions and plan.* Our main contribution is our analysis in Sec. 6.3 of the bishop-list data. Our reconstruction of the evolving social hierarchy in Sec. 6 is the first statistical analysis of this kind of data. We answer some longstanding questions. How important was seniority in determining precedent? Did court politics play a role? Our poset models are also new (extending Mannila (2008) and Sakoparnig and Beerenwinkel (2012) as detailed above). Although our methodology was motivated by one particular data set, we nevertheless propose our poset-based ranking models as potentially useful for the analysis of rank data more broadly, to stand alongside Mallows, Plackett-Luce and other models for rank data.

In Sec. 2 we describe the list data, their associated dates and a seniority covariate on the bishops which informs their position in the hierarchy. Sec. 3 begins by setting out notation and defining partial orders and how they constrain the lists we actually observe. We motivate and define the observation model for lists in Sec. 3.1 and then in Sec. 3.2 give the prior. This is where we define "status" and how status is mapped to preference. The generative model and posterior are given Sec. 3.3. In Sec. 4 we show that our priors are marginally consistent and have support on every poset. We give a brief outline of our MCMC and define some useful summary statistics in Sec. 5, relegating the detail to Supp E. In Sec. 6 we present the results of our Bayesian-MCMC analysis. We begin in Sec. 6.2 with a sanity-check: we drop an order constraint on the seniority covariate-effects which historians expect to hold and show the order is recovered. In Sec. 6.3 we present our main results with the constraint now imposed. Results are discussed from a historical perspective in Sec. 6.4. In Sec. 7 we make model comparisons with other methods (fitting VSP and Bucket-Order models). Further comparison with variants of Plackett-Luce are given in Supp J. These favor our poset-model, though VSP and Bucket-Order models do quite well. We summarise our contribution in Sec. 8 and point to future work. A supplement discusses data registration, properties of the observation model and prior, MCMC, further results, model comparisons and results on synthetic data.

2. Introduction to the data.

2.1. *Context.* This study draws on an accumulated dataset, accessed through the database made for 'The Charters of William II and Henry I' project by the late Professor Richard Sharpe and Dr Nicholas Karn (Sharpe et al., 2014). Some historical background on the data is given in Supp A.1. Each witness list in the data is an ordered list of names of individual witnesses taken from a single legal document or "act" (collectively "acta"). A typical example (with List id 2364) is given in Supp A.2.

We have 1610 witness lists dated between 1066 CE and about 1166 CE involving 1760 individuals. A witness list is a "snapshot" created at a single event on a single day. We assume distinct lists are generated independently (for example, we see no evidence for a pair of Acta created at the same time with identical lists). Acta are witnessed in order of social rank from the king or queen, archbishop, bishops (as a group), earls (as a group, may precede bishops) and so on down through society. Historians ask if the order in which bishops appear within the their sub-list reflects their evolving personal authority. As we focus on the bishop-hierarchy we extract from the data the sub-lists of bishops. Many of the resulting lists contain less than two bishops and these are discarded as not informing relations between bishops. Data processing is set out in detail in Supp A.2.

We take time as discrete by year as the data gives dates rounded to the year. Further coarsening would mask recoverable structural change. Outside the range $[B = 1080, E = 1155]$ CE the lists are sparse so we focus on this interval, covering the reigns of William II, Henry I and Stephen and $T = E - B + 1 = 76$ years. The period is long enough for us to witness changes in the status of individual bishops, but short enough for there to be some hope of temporal homogeneity in the social conventions mapping status to witness list.

Data registration is detailed in Supp A.4. It leaves us with $N = 371$ lists, dated between 1080 and 1155, and containing two or more bishops. We refer to this as the "full data" (we look at shorter time intervals in our goodness-of-fit). Each of the $M = 67$ bishops in at least two lists is assigned a numerical index from 1 to M in Fig 24 in Supp G.2.

2.2. *Data notation.* Let $\mathcal{I} = \{1, ..., N\}$ and $\mathcal{M} = \{1, ..., M\}$ be the sets of list and bishop labels respectively. Each list $y_i = (y_{i,1}, ..., y_{i,n_i}),\ i \in \mathcal{I}$ is an ordered list of n_i bishops, so that $y_i \subset \mathcal{M}$, with bishop-$y_{i,1}$ first in the list, $y_{i,2}$ second and so on. Not all bishops appear in every list so the list is conditioned on "attendance". Let $o_i = \{y_{i,1}, ..., y_{i,n_i}\}$ be the *unordered* list of bishops in list i. Let $y = \{y_1, ..., y_N\}$ and $o = \{o_1, ..., o_N\}$.

We suppose the lists were generated by a process running over an interval of time $[B, E] = \{B, B+1, \ldots, E\}$. For $i \in \mathcal{I}$ let $\tau_i \in [B, E]$ give the time at which list i was created. This is sometimes uncertain. However, bounds $\tau_i^- \le \tau_i \le \tau_i^+$ are available. We estimate the missing list dates, taking a uniform prior on $\tau_i \in [\tau_i^-, \tau_i^+]$. Let $\tau = (\tau_1, ..., \tau_N)$, $\tau_i^\pm = (\tau_i^-, \tau_i^+)$ and $\tau^\pm = (\tau_1^\pm, ..., \tau_N^\pm)$.

Bishops enter the social hierarchy when consecrated and leave when they die. For $j \in \mathcal{M}$ we have (FEA, 2022) dates of consecration $b_j < E$ and death $e_j > B$ for each bishop. The distribution of these intervals can be seen in Supp A in Fig 16 at left. The intervals match the list date-ranges (τ_i^-, τ_i^+) so that no bishop appears in a list when not in post. At any given time some dioceses may be empty. Fig 2 shows the presence and absence of bishops by diocese. For $t \in [B, E]$ let $\mathcal{M}_t = \{j \in \mathcal{M} : b_j \le t \le e_j\}$ give the set of bishops active at time t, let $m_t = |\mathcal{M}_t|$ give the number of active bishops at time t and let

$$D = \max_{t \in [B,E]} m_t \qquad (1)$$

give the greatest number active at any time ($D = 22$, in 1133). This quantity plays a role in bounding the required dimension of the model parameter space.

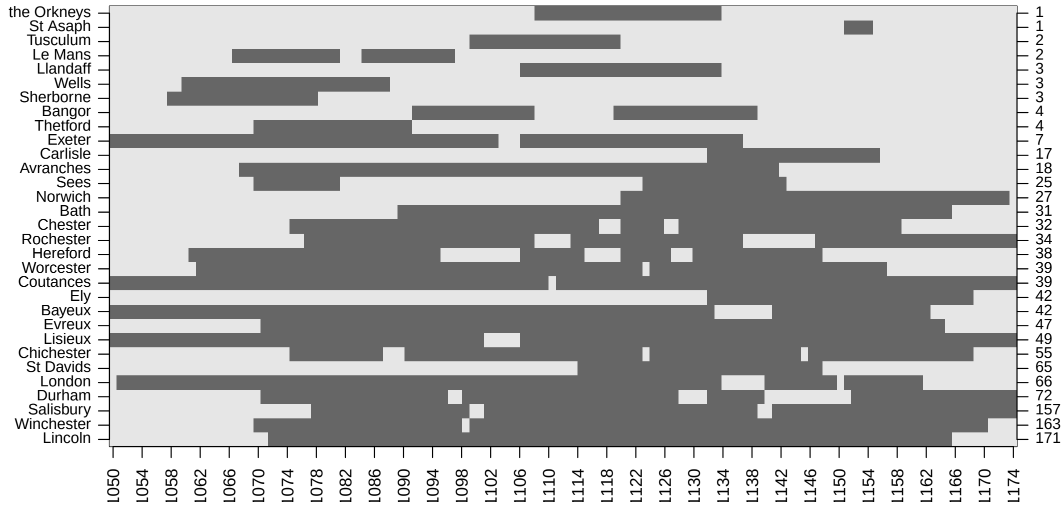

FIG 2. *Left axis gives dioceses. Right axis gives the number of lists in which each diocese appears. The x-axis gives the date in years. In each year a diocese may have a bishop in-post (dark cell) or be unoccupied (light).*

2.3. *Witness list data.* Bishops from thirty one dioceses appear in the data (including one, Tusculum, from Italy, dropped from the data). They are listed in Supp A.3 and can be seen at the left side of Fig 2. The dates of $|\{i \in \mathcal{I}:\ \tau_i^+ - \tau_i^- > 1\}| = 212$ lists are uncertain (the mean interval length is 4 years, and 90% span less than 10 years). Date intervals $[\tau_i^-, \tau_i^+]$ are plotted in Supp A in Fig 16 at right. Fig 3 plots lists and their lengths againt their dates (using the midpoint of $[\tau_i^-, \tau_i^+],\ i \in \mathcal{I}$). We include a list if at least half its interval falls within the 76-year interval $[B, E]$; most of the lists in our analysis fall entirely within it.

Our information about a bishop's status is limited by the number of lists in which a bishop appears. Longer lists are more informative as they inform relations between many pairs of bishops. Fig 4 shows the distribution of the number of lists a bishop appears in and the distribution of list lengths. Most bishops appear in a small number of lists and most lists are relatively short. However, lists "link together". If two bishops j_1, j_2 do not appear in any list together, but j_1 comes before j_3 in some list and j_3 before j_2 in another, then this is evidence for j_1 having higher status than j_2. This evidence accumulates over lists.

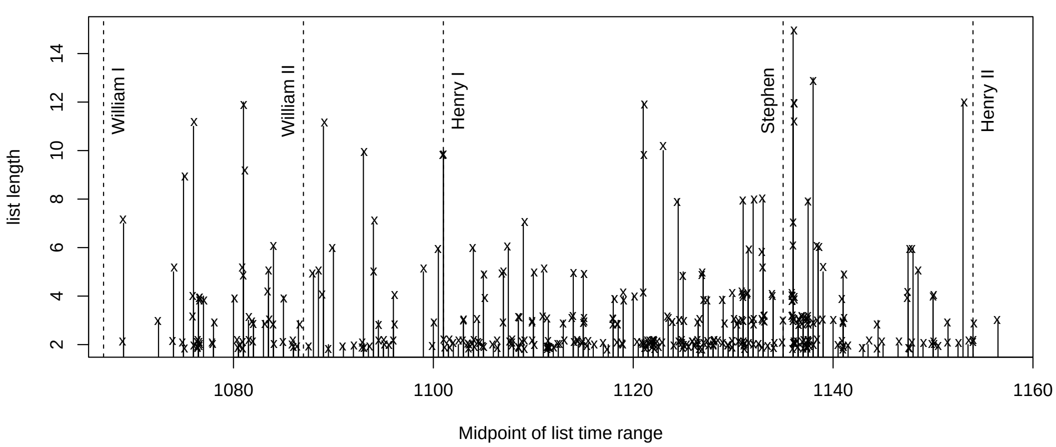

FIG 3. *List lengths and dates. Dashed vertical lines are coronation dates of Kings, bar heights are longest lists at that date, with a (jittered) cross for each list plotted at (date,length).*

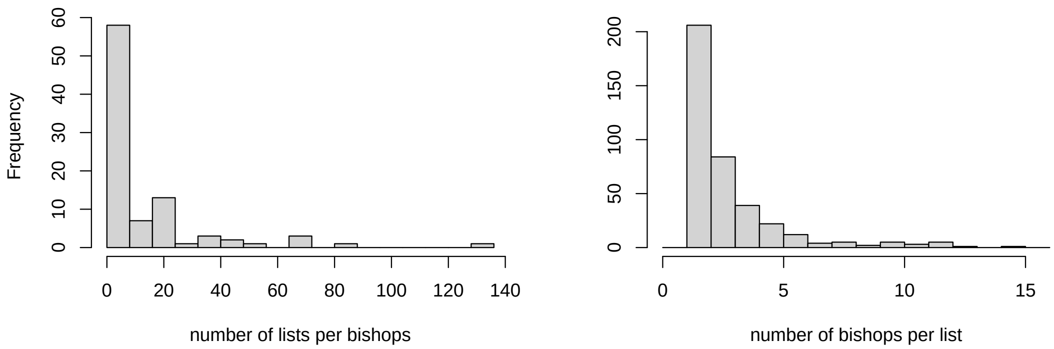

FIG 4. *Frequency of lists per bishop, a histogram of the counts $\sum_{i\in\mathcal{I}} \mathbb{I}_{j\in o_i}$, is plotted at left. Distribution of list lengths, a histogram of the counts $n_i = |o_i|$, is plotted at right.*

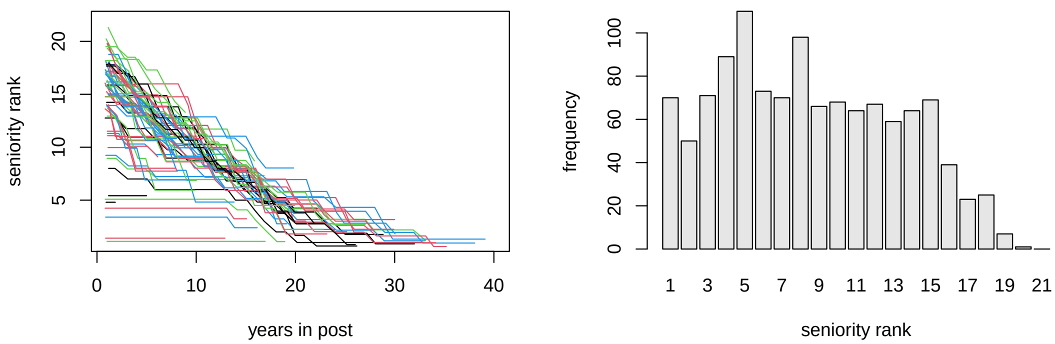

FIG 5. *Seniority rank covariate s defined in Eqn 2. Rank $s_{t,j}$ is plotted against "years in post", $t - b_j$, for each bishop $j \in \mathcal{M}$ (Left). The frequency of seniority-rank r in lists is plotted against r from 1 to S (Right).*

2.4. *Seniority covariates.* A bishop's "seniority" may have contributed to their overall status (Clover and Gibson, 1979) so it is a covariate in our model for the hierarchy of bishops. In year t the longest serving bishop has seniority-rank one and the last appointed bishop has rank at most m_t. Denote by $s_{t,j} \in \{1, ..., m_t\}$ the seniority-rank of bishop $j \in \mathcal{M}_t$ in year $t \in \{b_j, b_j + 1, ..., e_j\}$. We define seniority-rank as

$$s_{t,j} = \sum_{k\in\mathcal{M}_t} \mathbb{I}_{b_j \geq b_k}. \tag{2}$$

Bishops have equal seniority if there are ties in the start dates $b_j,\ j \in \mathcal{M}$. The greatest seniority observed, $S = \max_{t,j} s_{t,j}$, is less than or equal D, the most bishops active. This is a second quantity which informs the dimension of the model we fit.

Fig 5 shows (at left) seniority-rank traces for each bishop from their first to last year in post. Bishops progress in rank by about one place every one or two years, more rapidly at first, as there are more bishops ahead of them. Our estimates of the effect of possessing seniority-rank $r \in \{1, ..., S\}$ depends for precision on a bishop with seniority r appearing in a reasonable number of lists, so we plot (Fig 5, right) the occurrence frequency $f_r = \sum_{i\in\mathcal{I}} \sum_{j\in o_i} \mathbb{I}_{s_{\tau_i,j}=r}$ against r to see which levels of the covariate are well represented in the data. Some dioceses were more peaceful and wealthy than others, so we considered taking diocese label as a second covariate for status. However, diocese would be colinear with bishop label, as each bishop only occupies one diocese in the period of study. An effect due to diocese would not be identifiable with the effects due to the bishops in that diocese.

2.5. *Key questions.* The key question is whether the status of bishop $j \in \mathcal{M}$ in year $t \in [B, E]$ was determined by their seniority $s_{t,j}$ in that year and their diocese or whether the personal authority or position in court of bishop j played a role. In terms of the list data y, are the positions of bishops in a list determined by seniority and diocese, with any variation away from this order just unstructured noise, or are the variations away from precedence rules structured by some kind of latent authority? The status of a bishop determines their place in the hierarchy, so what did the hierarchy look like in any given year? We answer these questions by building a statistical model for the bishop-list data.

3. Models and Inference. Our model, in which lists gathered in year t respect a social hierarchy which is known and respected by all but subject to occasional change, expresses the evolving social hierarchy. We present our model as a description of relations between actors, in the usual terminology of social network analysis. However, it can be applied to the analysis of any ranking list data, with or without time-series structure: Jiang et al. (2023) shows that a related fixed-time model is a good fit for Formula 1 race outcomes.

3.1. *Parameters and observation model.*

3.1.1. *Partial orders and linear extensions.* In this section we define precedence relations using posets. We drop the time dependence and consider a single generic observation. Suppose we have m actors with labels in $[m] = \{1, ...m\}$. We represent the unknown true order relations between actors as a poset on $[m]$. Brightwell (1993) gives an overview of models for random posets and is the source for much of what follows. A strong partial order $\succ_H$ on the ground set $[m]$ is a set of acyclic, transitively closed relations $i \succ_H j$ on the elements of $i, j \in [m]$. Ties $i \sim_H j$ are excluded. The relations in H are transitively closed if $i \succ_H j$ and $j \succ_H k$ implies $i \succ_H k$. The order is only partial as some elements are not ordered. A poset is a total order if $i \succ_H j$ or $j \succ_H i$ for every pair $i, j \in [m]$.

Partial orders on $[m]$ are one to one with transitively closed directed acyclic graphs (DAGs) with vertex labels $1, ..., m$, one vertex for each of the m actors, so $\succ_H$ is represented by a DAG $(H, [m])$ with edge set

$$H = \{\langle i, j\rangle \in [m] \times [m] : i \succ_H j\}.$$

See the example in Fig 6. We refer to transitively closed DAGs as if they were posets, as they correspond one to one. We can identify a poset by its edge set H as the edge set will be random while the vertex labels $[m]$ which define the ground set are always fixed. Since posets are edge sets we can take intersections of posets. This gives the poset with all relations shared by the intersected posets. The *dimension* of a poset $H \in \mathcal{H}_{[m]}$ is the smallest number of total orders which intersect to give H.

Let $\mathcal{H}_{[m]}$ be the set of all transitively closed DAGs on $[m]$ and let $H \in \mathcal{H}_{[m]}$ be a generic poset. For plotting purposes the transitive reduction is convenient. This is the unique DAG obtained from H by removing all edges implied by transitivity. The depth of a social hierarchy is of interest in many applications. The depth $d(H)$ of $\succ_H$ is the length of the longest path on the DAG H, so $d : \mathcal{H}_{[m]} \to [m]$. The poset in Fig 6 has $d(H) = 4$.

Let $\mathcal{P}_{[m]}$ be the set of all permutations of $[m]$. A *linear extension* of H is any list $\ell = (\ell_1, ...\ell_m)$, $\ell \in \mathcal{P}_{[m]}$ in which lesser entries come after greater entries, so $\ell_j \succ_H \ell_k$ is not allowed if $k < j$. For example, if H is the poset displayed in Fig 6 then H has three linear extensions, $(1, 2, 3, 4, 5), (1, 2, 4, 3, 5)$ and $(1, 4, 2, 3, 5)$. Denote by

$$\mathcal{L}[H] = \{\ell \in \mathcal{P}_{[m]} : \langle \ell_j, \ell_k\rangle \notin H \text{ for all } 1 \le k < j \le m\}$$

the set of all linear extensions of $H \in \mathcal{H}_{[m]}$.

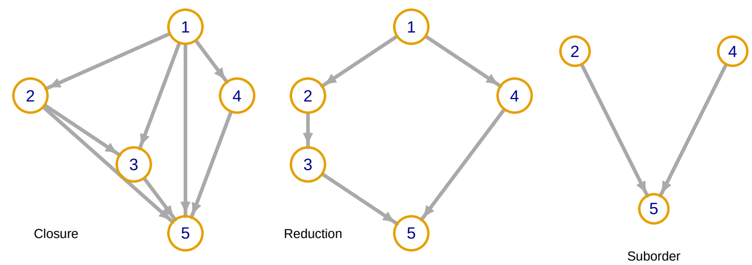

FIG 6. *Poset* $H = \{\langle 1,2\rangle, \langle 1,3\rangle, \langle 1,4\rangle, \langle 1,5\rangle, \langle 2,3\rangle, \langle 2,5\rangle, \langle 3,5\rangle, \langle 4,5\rangle\}$, $H \in \mathcal{H}_{[m]}$ *on* $[m] = (1,2,3,4,5)$ *represented by its transitively closed directed acyclic graph (left) and transitive reduction (centre). Suborder* $H[O]$ *for* $O = \{2,4,5\}$ *(right). The longest path on* H *is* $1 \to 2 \to 3 \to 5$ *so the depth is* $d(H) = 4$.

If we start with a social hierarchy $H \in \mathcal{H}_{[m]}$ over all actors then the hierarchy constraining any given subset $O \subseteq [m]$ of the actors is the *suborder*

$$H[O] = \{\langle j_1, j_2\rangle \in H : \{j_1, j_2\} \subseteq O\} \tag{3}$$

obtained by retaining edges between vertices in O. If H is a poset then so is $H[O]$. For example, in Fig 6, if $O = \{2,4,5\}$ then suborder $H[O]$ is the three-vertex DAG at right.

3.1.2. *Lists as randomly ordered queues.* A single generic list $Y = (Y_1, ..., Y_m)$ is modeled as a random linear extension of H. Let $C(H) = |\mathcal{L}[H]|$ be the number of linear extensions of poset $H \in \mathcal{H}_{[m]}$. The "noise free" likelihood for H is simply

$$p(Y|H) = C(H)^{-1}\mathbb{I}_{Y \in \mathcal{L}[H]}. \tag{4}$$

All lists which respect the social hierarchy are equally likely. This "context dependent" observation model (see Supp B.1) is motivated by thinking of each list as a realisation of a random queue process in which actors not ordered by H randomly swap places. The equilibrium of this process is the uniform distribution on linear extensions of H (Karzanov and Khachiyan, 1991; Jiang and Nicholls, 2024), so if Y is a snapshot of this queue at equilibrium then $Y \sim \mathrm{Unif}(\mathcal{L}[H])$. Witness lists were written down by a royal scribe with (we assume) perfect knowledge of the hierarchy H, so the queue model is an idealisation. However, we arrive at the same model if we assume the clerks regarded all orders not conflicting the hierarchy as equally likely.

Computation of $C(H)$ is #P-complete (Brightwell and Winkler, 1991) so no polynomial time algorithm for computing $C(H)$ is available, or is likely to exist, and evaluating $p(Y|H)$ is prohibitive at large m. However, in our data m is small enough to allow likelihood evaluation in reasonable time. For $j \in [m]$ let $\mathcal{L}_j[H] = \{\ell \in \mathcal{L}[H] : \ell_1 = j\}$ be the set of linear extensions with j first in the list and let $C_j(H) = |\mathcal{L}_j[H]|$. Partitioning on the first entry,

$$C(H) = \sum_{j=1}^{m} C_j(H),$$

which Knuth and Szwarcfiter (1974) compute using the suborder recursion in Supp B.2.

3.1.3. *Queue-jumping observation model for suborders.* In this section we define the observation model we use in our analyses. We modify the likelihood in Eqn 4 in three ways: lists may be "noisy"; just a subset of actors are in any given list; lists are observed over time.

We allow for noise in the observation model by allowing individuals to "jump the queue". A list Y is formed by taking individuals from the top of a queue which continues to mix rapidly, constrained by the suborder on those remaining. Before the j'th actor is chosen, there are $m-j+1$ individuals (with labels $Y_{j:m}$) yet to be placed. With probability p the next actor (ie Y_j) is chosen at random, ignoring any order constraints. Otherwise, Y_j is chosen as the first actor in a random linear extension of the suborder $H[Y_{j:m}]$ for those remaining. The fraction of lists headed by Y_j is $C_{Y_j}(H[Y_{j:m}])/C(H[Y_{j:m}])$. Working from the top down,

$$p_{(D)}(Y|H,p) = \prod_{j=1}^{m} p_{(D)}(Y_j|H[Y_{j:m}],p)$$

$$(5)\qquad = \mathbb{I}_{Y\in\mathcal{P}_{[m]}} \prod_{j=1}^{m}\left(\frac{p}{m-j+1} + (1-p)\frac{C_{Y_j}(H[Y_{j:m}])}{C(H[Y_{j:m}])}\right).$$

Noise allows any list to appear with non-zero probability. This is a "repeated selection" model (Seshadri et al., 2020) in which the next actor is chosen sequentially from those that remain. It follows that the correct likelihood for top-k data (where an assessor sees all m actors but just ranks their top $k<m$ actors) simply stops the product in Eqn 5 at $j=k$.

We take $p\sim \text{Beta}(1,\delta)$ as our family of priors for the queue-jumping probability p. The prior hyperparameter $\delta\geq 1$ is fixed (for example, in Sec. 6 we take $\delta=9$, so the prior probability for a queue-jumping event is about ten percent), expressing the belief that if order relations are present then they are respected.

In Eqn 5 individuals are promoted up the queue. We can also model random "demotion". In this case the list is filled from the bottom up: with probability p the next actor is chosen at random, ignoring any order constraints; otherwise, they are the last entry in a random linear extension of the suborder for the remaining individuals. The likelihood becomes

$$(6)\qquad p_{(U)}(Y|H,p) = \mathbb{I}_{Y\in\mathcal{P}_{[m]}} \prod_{j=1}^{m}\left(\frac{p}{j} + (1-p)\frac{\tilde{C}_{Y_j}(H[Y_{1:j}])}{C(H[Y_{1:j}])}\right),$$

where $\tilde{C}_{Y_j}(H)$ is the number of linear extensions of H which end with Y_j. Jiang et al. (2023) extend these models to allow random promotion and demotion in a single realised list. The likelihood is tractable, but evaluation is time consuming so we do not fit that model here.

The noise free case is obtained from both Eqns 5 and 6 at $p=0$. One consequence is that the noise-free model is also a repeated selection model. See Supp B.3 for the proof.

PROPOSITION 1. The noise-free and noisy models (Eqns 4, 5 and 6) all coincide at $p=0$.

We now consider what happens when just a subset of actors are present. Relations are given by their suborder and the observation model applies for lists realised on suborders. Suppose that, when Y was realised, a subset $O=\{O_1,...,O_n\},\ O\subseteq[m]$ of actors entered the queue. Since they were constrained by the suborder $H[O]$, the noise free observation model is $Y\sim \text{Unif}(\mathcal{L}[H[O]])$: the list is a random draw from the linear extensions of the suborder. For example, for H in Fig 6, if $O=\{2,4,5\}$, the linear extensions are $\mathcal{L}[H[O]] = \{(2,4,5),(4,2,5)\}$ and Y is chosen at random from this set. The queue-jumping likelihoods $p_{(D)},p_{(U)}$ are obtained by replacing $H\to H[O],\ [m]\to O$ and $m\to n$ in Eqns 5 and 6.

3.1.4. *Time series of lists.* Finally, we restore time and give the full likelihood. This observation model is illustrated in Fig 7. At each time $t\in[B,E]$ the actors indexed in $\mathcal{M}_t$ were active and had precedence relations given by some poset $h^{(t)}\in\mathcal{H}_{\mathcal{M}_t}$. The sequence of partial orders from B to E is

$$(7)\qquad h = (h^{(B)},h^{(B+1)},...,h^{(E)}),$$

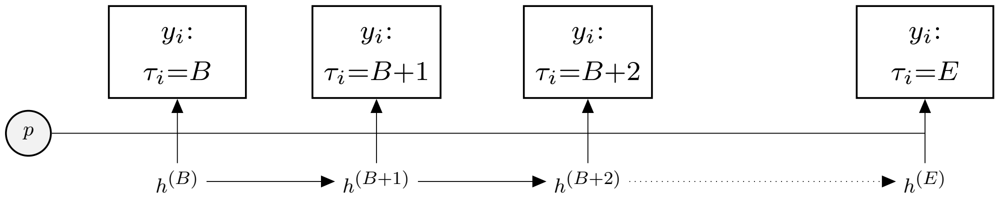

FIG 7. *The observation model. The distribution of lists $\{y_i : i \in \mathcal{I},\ \tau_i = t\}$ observed at time t is parameterised by a poset $h^{(t)}$ and noise parameter p. A stochastic process realising $h^{(t)}$, $t \in [B,E]$ is given in Sec. 3.2.*

where $h \in \mathcal{H}^{(B,E)}$ with

(8) $$\mathcal{H}^{(B,E)} = \mathcal{H}_{\mathcal{M}_B} \times \mathcal{H}_{\mathcal{M}_{B+1}} \times ... \times \mathcal{H}_{\mathcal{M}_E}.$$

The model for h in Sec. 3.2 will define a poset process with poset time-series realisations.

For $i \in \mathcal{I}$, list y_i was formed under constraints imposed by the suborder $h^{(\tau_i)}[o_i]$, so in the noise-free model $y_i \in \mathcal{L}[h^{(\tau_i)}[o_i]]$. Allowing for noise, the likelihood is

(9) $$p(y|h,\tau,p) = \prod_{i=1}^{N} p(y_i|h^{(\tau_i)}[o_i], p),$$

where $p(y_i|h^{(\tau_i)}[o_i], p)$ is given by $p_{(D)}$ or $p_{(U)}$ in Eqns 5 or 6 (depending on our choice of model) with the replacements $Y \to y_i$, $m \to n_i$ and $H \to h^{(\tau_i)}[o_i]$.

3.2. *Latent variables and covariates in a prior for partial orders.* We derive prior models for posets from k-dimensional random orders (Winkler, 1985). We modify this setup as we would like to have some control over the prior depth distribution. The depth of the social hierarchy of bishops is meaningful to historians so a prior which is non-informative with respect to depth will be useful. The uniform prior on posets certainly wouldn't be appropriate as it is strongly informative of depth: it concentrates on posets of depth three as $m \to \infty$ (Kleitman and Rothschild, 1975). The effect is illustrated in Supp D.

3.2.1. *Latent variable parameterisation.* In this section we define status and how it is mapped to precendence order relations. We use feature vectors, one for each actor, to determine actor-placing in the social hierarchy. See Fig 8 for illustration. These "status-features" do not correspond to any identifiable physical attributes. Following Winkler (1985), we associate with each actor $j \in \mathcal{M}_t$, at a time $t \in [b_j, e_j]$, a $1 \times K$ latent vector $Z_j^{(t)} \in \mathbb{R}^K$, $Z_j^{(t)} = (Z_{j,1}^{(t)}, ..., Z_{j,K}^{(t)})$ of $K \geq 1$ status-features. Let $Z^{(t)} = (Z_j^{(t)})_{j \in \mathcal{M}_t}$ be an $m_t \times K$ status matrix, with one row for each actor active at time t and one column for each status feature $1, ..., K$.

The partial-order $h^{(t)}$ at time t is a function of $Z^{(t)}$. At time t actor $j \in \mathcal{M}_t$ is above actor $j' \in \mathcal{M}_t$ if *all* status variables of j are greater than those of j', that is, $h^{(t)} = h(Z^{(t)})$ with

(10) $$h(Z^{(t)}) = \{\langle i,j \rangle \in \mathcal{M}_t \times \mathcal{M}_t : Z_{i,k}^{(t)} > Z_{j,k}^{(t)} \text{ for all } k = 1, ..., K\}.$$

The setup is illustrated in Fig 8 at centre, where the Z-matrix has $m_t = 5$ rows and $K = 4$ columns. The rows of $Z^{(t)}$ are "paths" in the space $[K] \times \mathbb{R}$: the relation $\langle j, j' \rangle \in h^{(t)}$ holds when the path $(k, Z_{j,k}^{(t)})_{k=1}^K$ lies above the path through $(k, Z_{j',k}^{(t)})_{k=1}^K$; if the paths cross then the actors are unordered. In Fig 8, the Z-path for actor 4 intersects the paths for 2 and 3 so 4 is unordered with respect to 2 and 3. The other paths do not intersect so 1, 2, 3, 5 have a

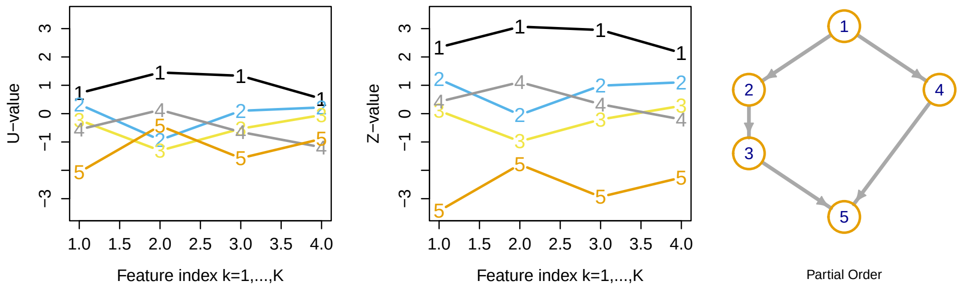

FIG 8. *Latent variable representation of the poset H in Fig 6: (left) U-matrix defined in Sec. 3.2.2; (centre) Z-matrix and (right) poset $h(Z)$ defined in Sec. 3.2.2. The paths in Z are shifted by the effects β.*

total order. This is a latent feature representation of the poset at right. In Winkler (1985) the columns of $Z^{(t)}$ are independent and paths are likely to cross. By contrast, the priors we give in Sec. 3.2.3 correlate columns and give us some control over the prior depth distribution.

Our mapping from $Z^{(t)}$ to $h^{(t)}$ is equivalent to intersecting the K total orders given by ranking the entries in each column of $Z^{(t)}$. See Supp C.2 for a discussion of this point. As we take a fixed value of K, the dimension of $h^{(t)}$ is at most K. However, using results from (Hiraguchi, 1951; Bogart, 1973a), it may be shown that if $K \geq \lfloor m_t/2 \rfloor$ then any poset $h^{(t)} \in \mathcal{M}_t$ can be represented by some $m_t \times K$ matrix $Z^{(t)}$. This is discussed in a modelling context in Muir Watt (2015) and proven in Sec. 4.2. Since $h^{(t)},\ t \in [B, E]$ has at most D vertices (see Eqn 1), a model with $K \geq \lfloor D/2 \rfloor$ can represent any partial order process $h \in \mathcal{H}^{(B,E)}$.

Taking K as large as $\lfloor D/2 \rfloor$ is conservative as the true h-process may live in a lower dimensional space. In recent work Jiang and Nicholls (2024) estimates K, using a non-parametric prior for the U-process and sampling K using reversible jump MCMC. They considered smaller data sets with fewer lists all gathered at a single fixed time. In that setting the best fitting values of K were often much smaller than $\lfloor D/2 \rfloor$. In earlier work Durante et al. (2017b) used a similar approach to estimate the dimension of a latent space parameterisation of a population of networks. In both cases graph structures are modeled using latent continuous variables. Our approach is more like Durante et al. (2017a) and Gwee et al. (2023): the dimension of the latent space is chosen so that some kind of representation property holds (like our Proposition 3); shrinkage (Rousseau and Mengersen, 2011) effectively removes unwanted components. In our setting ρ controls this shrinkage. When it is larger the columns of Z tend to have the same order, so the dimension of the poset is smaller.

3.2.2. *Covariate effects for partial orders.* .

In Sec. 2.2 we introduced an actor-specific covariate informing status relations among bishops. In the following $s_{t,j} \in \{1, 2, \ldots, S\}$ is a single categorical or ordinal variable with levels from 1 to S. More general covariates are easily accommodated. Let $\beta \in \mathbb{R}^S$ be the vector of level effects. We split the *status*-vector $Z_j^{(t)}$ of actor $j \in \mathcal{M}_t$ into an additive effect $\beta_{s_{t,j}}$ due to $s_{t,j}$ and an *authority*-vector $U_j^{(t)} \in \mathbb{R}^K$, which captures all aspects of status which are not attributable to the covariate. Our additive model is, for $j \in \mathcal{M}_t$ and $t \in [B, E]$,

$$Z_j^{(t)} = U_j^{(t)} + 1_K \beta_{s_{t,j}} \tag{11}$$

where 1_K is a row vector of K ones. Higher $\beta_{s_{t,j}}$-values lift all components of $U_j^{(t)}$ by a constant, raising the status features in $Z_j^{(t)}$. This moves the path $(k, Z_{j,k}^{(t)})_{k=1}^K$ above other paths and gives a higher position for actor j in the poset $h^{(t)}$. See Fig 8 for an example.

In our application the covariate is ordinal with a greater effect expected for lower values of s so we will be interested in testing for $\beta_1 > \beta_2 > ... > \beta_S$. Let $\mathcal{B}_0 = \mathbb{R}^S$ and

$$\mathcal{B}_S = \{\beta \in \mathcal{B}_0 : \beta_1 > \beta_2 > ... > \beta_S\}. \tag{12}$$

We carry out analyses under models with $\beta \in \mathcal{B}_0$ (to check our prior expectation that $\beta \in \mathcal{B}_S$) and then again with $\beta \in \mathcal{B}_S$ (for best estimation with a well supported subjective prior).

Let $Z = (Z^{(t)})_{t=B}^{E}$ and $U = (U^{(t)})_{t=B}^{E}$ and write $Z = Z(U, \beta; s)$ for the function defined in Eqn 11. The parameters U and β replace h in the likelihood via $h = h(Z(U, \beta; s))$.

3.2.3. *Prior probability distributions.* We model the K-dimensional authority-process $U_j = (U_j^{(t)})_{t=b_j}^{e_j}$ for actor $j \in \mathcal{M}$ as a vector autoregression of order one with times-series correlation θ and covariance $\Sigma^{(\rho)}$. In our model latent authority-features are correlated from one time step to another with a drift towards zero. The setup is illustrated in Fig 9.

Our prior for the process U_j is independent over $j \in \mathcal{M}$ with correlation parameters $0 \leq \theta \leq 1$ and $0 \leq \rho \leq 1$. Let $\Sigma^{(\rho)}$ be a $K \times K$ covariance matrix with diagonal elements $\Sigma^{(\rho)}_{k,k} = 1$ and off diagonal $\Sigma^{(\rho)}_{k,k'} = \rho$ for $k, k' \in [K]$. Let 0_K be a vector of K zeros. For $j \in \mathcal{M}$ let

$$U_j^{(b_j)} \sim N\left(0_K, \frac{\Sigma^{(\rho)}}{(1-\theta)^2}\right)$$

and for $\epsilon_j^{(t)}$ iid for $t \in [b_j + 1, e_j]$ and each j,

$$U_j^{(t)} = \theta U_j^{(t-1)} + \epsilon_j^{(t)}, \quad \epsilon_j^{(t)} \sim N(0, \Sigma^{(\rho)}). \tag{13}$$

Write $U_j \sim \text{VAR}_{K,\rho,\theta}^{(b_j,e_j)}(1),\ j \in \mathcal{M}$ for the vector auto-regression with density

$$\pi(U_j|\rho,\theta) = N\left(U_j^{(b_j)}; 0_K, \frac{\Sigma^{(\rho)}}{(1-\theta^2)}\right) \prod_{t=b_j+1}^{e_j} N\left(U_j^{(t)}; \theta U_j^{(t-1)}, \Sigma^{(\rho)}\right). \tag{14}$$

The parameter ρ controls the prior depth-distribution. When $\rho \simeq 1$ paths $(k, U_{j,k}^{(t)})_{k=1}^{K}$ are relatively flat as entries are strongly correlated. Flat paths don't intersect, so there are more order relations and $d(h^{(t)})$ is larger. When ρ is close to zero the paths are more jagged so there are few order relations and $d(h^{(t)})$ is smaller. We take as our prior $\rho \sim \text{Beta}(\gamma_1, \gamma_2, \gamma_3)$ with non-centrality parameter γ_3 and $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ fixed. Prior simulation (Supp D, Fig 17, left) showed $K = \lfloor D/2 \rfloor$ and $\gamma = (1, 1/3, 8)$ gave a prior on posets which is acceptably uniform on depth. Our prior on θ is uniform, $\theta \sim \text{Unif}(0, 1)$.

Our prior density for β is $\pi_\beta(\beta) = N(\beta; 0, I_S)$. The variation between levels of a covariate equals the variation in authority over one time step: $\beta_j - \beta_{j'}$ has the same variance as the components of $U_j^t - U_j^{(t-1)}$ in Eqn 13. We set $\Sigma^{(\rho)}_{k,k} = 1$ as we can scale ρ and the variance of β to get the same distribution for h. It is necessary to take proper priors for U and β. We discuss these priors further in Sec. 5.2 in relation to identifiability.

3.3. *Prior summary and Posterior distribution.* We now summarise our generative model for the data. The model is depicted in Figs 7 and 9. The data are the lists y. We condition on knowledge of the uncertainty ranges $\tau^{\pm}$, the covariate data s, the actor activities $\mathcal{M}_t,\ t \in [B, E]$ and the prior hyper-parameters $\gamma, \delta > 0$ and $K \geq 1$. The generative model is,

$$\rho \sim \text{Beta}(\gamma), \qquad \text{with } \gamma = (1, 1/3, 8) \text{ unless stated,}$$

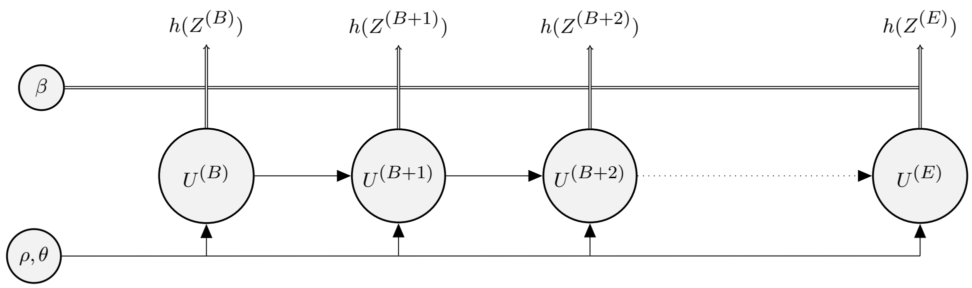

FIG 9. *Hidden layer of HMM: a latent process $U^{(t)}$ determines a sequence of posets $h^{(t)} = h(Z(U^{(t)}, \beta, s)$ for $t \in [B, E]$. The U-prior has hyperparameters ρ and θ, and double lines indicate implication $\Rightarrow$. This figure connects with Fig 7 at the top row of nodes $h^{(t)} = h(Z^{(t)})$ to give the generative model for the data.*

$$\begin{aligned}
\theta &\sim \text{Unif}(0,1), && \\
U_j &\sim \text{VAR}_{K,\rho,\theta}^{(b_j,e_j)}(1), && \text{with } K = \lfloor D/2 \rfloor \text{ unless stated, and } U \text{ in Eqn 13,} \\
& && U_j^{(t)} \text{ iid for } j \in \mathcal{M} \text{ and defined for } t : j \in \mathcal{M}_t \text{ and} \\
\beta &\sim N(0, I_S), && \text{either } \beta \in \mathcal{B}_0 \text{ or constrained } \beta \in \mathcal{B}_S \text{ per Eqn 12.}
\end{aligned}$$

These collectively determine the partial-order prior via

$$\begin{aligned}
Z &= Z(U, \beta; s), && \text{from Eqn 11, giving } Z = (Z^{(t)})_{t=B}^{E}, \text{ and} \\
h &= h(Z(U, \beta; s)), && \text{from Eqn 10, giving } h = (h^{(t)})_{t=B}^{E}; \qquad (15)
\end{aligned}$$

Priors for the remaining observation model parameters are

$$\begin{aligned}
\tau_i &\sim U\{\tau_i^-, \tau_i^+\}, && \text{independently for } i = 1, ..., N, \text{ and} \\
p &\sim \text{Beta}(1, \delta); && \text{with } \delta \geq 1 \text{ and } \delta = 9 \text{ by default.}
\end{aligned}$$

Finally the data are realised

$$y_i \sim p(\cdot | h^{(\tau_i)}[o_i], p), \qquad \text{independently for } i \in \mathcal{I}, \tag{16}$$

using the distribution for y_i given in Eqn 5 or 6. The joint posterior distribution is

$$\pi(\rho, \theta, U, \beta, \tau, p|y) \propto \pi(\rho, \theta, \beta, \tau, p)\pi(U|\rho, \theta)p(y|U, \beta, \tau, p)\,, \tag{17}$$

where $p(y|U, \beta, \tau, p) = p(y|h(Z(U, \beta; s)), \tau, p)$ in Eqn 9 and $\pi(U|\rho, \theta) = \prod_{j \in \mathcal{M}} \pi(U_j|\rho, \theta)$ with $\pi(U_j|\rho, \theta)$ given in Eqn 14. The model without covariates or time is set out in Supp L.

4. Properties of priors. Our partial order prior is marginally consistent and expresses any partial-order time series in $\mathcal{H}^{(B,E)}$. Supp D explores the prior using simulation.

4.1. *Marginal consistency.* Marginal consistency is a relationship between members of a family of distributions. Dropping time, suppose that for each non-empty $O \subseteq [m]$ we write down a prior $\pi_{\mathcal{H}_O}(G)$ on $G \in \mathcal{H}_O$. These priors are marginally consistent if for each O and $H \sim \pi_{\mathcal{H}_{[m]}}$ we have $H[O] \sim \pi_{\mathcal{H}_O}$ for the distribution of the suborder, that is,

$$\pi_{\mathcal{H}_O}(G) = \sum_{H \in \mathcal{H}_{[m]}} \pi_{\mathcal{H}_{[m]}}(H) \mathbb{I}_{G=H[O]}. \tag{18}$$

The point here is that we define $\pi_{\mathcal{H}_O}$ for each $O \subseteq [m]$ and then verify Eqn 18. It can fail to hold if we write down a distribution $\pi_{\mathcal{H}_O}$ for each O without care. For example, the uniform distributions Unif$(\mathcal{H}_O)$, $O \subseteq [m]$ are not marginally consistent. There are three partial orders for $m = 2$ and nineteen for $m = 3$, so if $H \sim$ Unif$(\mathcal{H}_{[3]})$ then we won't have $H[(1,2)] \sim$ Unif$(\mathcal{H}_{[2]})$ as we can't group nineteen posets into three equal-sized groups.

Winkler (1985) shows marginal consistency when the columns of $Z^{(t)}$ are independent. We extend this to our more general setting. We take $\beta = 0_S$, so no covariates, as we cannot expect marginal consistency when we have covariate information which explicitly breaks it. Let $\pi_{\mathcal{H}^{(B,E)}}(h|\beta = 0_S)$ be the marginal prior for $h \in \mathcal{H}^{(B,E)}$ when $\beta = 0_S$ (see Eqn 24 in Supp C.1). Let $\mathcal{H}_{-j}^{(B,E)}$ be the set of all partial order time-series with $j \in \mathcal{M}$ removed and for $h \in \mathcal{H}^{(B,E)}$ let h_{-j} be the suborder obtained by removing j.

PROPOSITION 2. Let $g \in \mathcal{H}_{-j}^{(B,E)}$ be given. Our priors are marginally consistent, that is

$$\pi_{\mathcal{H}_{-j}^{(B,E)}}(g|\beta = 0_S) = \sum_{h \in \mathcal{H}^{(B,E)}} \mathbb{I}_{g=h_{-j}} \pi_{\mathcal{H}^{(B,E)}}(h|\beta = 0_S).$$

for each $j \in \mathcal{M}$. [See Supp C.1 for proof]

Proposition 2 establishes marginal consistency for removing one element of $\mathcal{M}$. Consistency for more general marginals follows by removing elements one at a time. See Supp C.1 for remarks on settings where marginal consistency holds with covariate effects retained.

4.2. *Universal representation.* We claimed in Sec. 3.2.1 that, if $K \geq \lfloor D/2 \rfloor$ with D defined in Eqn 1 then any poset $h^{(t)} \in \mathcal{H}_{\mathcal{M}_t}$ can be represented by some $m_t \times K$ matrix $Z^{(t)}$. We restore $\beta \in \mathcal{B}_0$ or $\mathcal{B}_S$ and covariate effects. Let

$$\pi_{\mathcal{H}^{(B,E)}}(h) = \int_{\mathbb{R}^S} \pi_{\mathcal{H}^{(B,E)}}(h|\beta)\pi(\beta)\, d\beta \tag{19}$$

be the full marginal prior with variable β, extended from Eqn 24.

PROPOSITION 3. Suppose $\min_{t \in [B,E]} m_t \geq 4$. The probability $\pi_{\mathcal{H}^{(B,E)}}(h)$ in Eqn 19, given by the generative model Eqn 15 with $K \geq \lfloor D/2 \rfloor$, assigns a positive probability mass $\pi_{\mathcal{H}^{(B,E)}}(h) > 0$ to every time-series $h \in \mathcal{H}^{(B,E)}$. [See Supp C.2 for proof.]

5. Computational methods.

5.1. *Markov Chain Monte Carlo.* We implemented an MCMC algorithm targeting $\pi(\rho, \theta, U, \beta, \tau, p|y)$ in Eqn 17. Each update is a simple Metropolis-Hastings MCMC step. The updates are summarised in Supp E. We tested the software evaluating the likelihood by simulating synthetic data and checking list proportions matched their probability in the likelihood. We also recover the true parameters of synthetic data (see Fig K.1). We run the MCMC producing L samples (after burn-in and thinning) $\rho^{(l)}, \theta^{(l)}$, $U^{(t,l)} = (U_j^{(t,l)})_{j \in \mathcal{M}_t}$, $\beta^{(l)}, \tau^{(l)}$ and $p^{(l)}$ for $l = 1, ..., L$. This determines samples for $Z^{(t,l)} = (Z_j^{(t,l)})_{j \in \mathcal{M}_t}$, with

$$Z_j^{(t,l)} = U_j^{(t,l)} + 1_K \beta_{s_{t,j}}^{(l)}$$

and similarly $h^{(t,l)} = h(Z^{(t,l)})$, for $t = B, ..., E$ and $l = 1, ..., L$.

The likelihood is not differentiable in U and β ruling out Hamiltonian MCMC. We tried updating sections of the time series in parallel (exploiting the Markov structure), halving the runtime at best. The bottleneck is ultimately the likelihood evaluation which is $\#P$-complete in list length. When lists are short the method scales well in the number of actors and the number of lists. The latent-parameter dimension $\dim(U) = K \times \sum_t m_t$ becomes the limiting factor (in Sec. 6.3, $\dim(U) = 13,453$). These challenges motivated work on VSP-models in (Jiang et al., 2023) which scale to lists with hundreds of actors.

5.2. *Posterior summaries.* Besides plotting marginals for individual parameters ρ, θ, p and β, we report selected summary statistics computed on the MCMC output. These are the consensus poset (which displays relations with posterior support greater than one half, see Supp F) and the Bayes factor for the first S' of the S covariate effects to be ordered, $\mathcal{B}_{S'} = \{\beta \in \mathcal{B}_0 : \beta_1 > \beta_2 > ... > \beta_{S'}\}$ and given by $B_{S',0} = p(y|\beta \in \mathcal{B}_{S'})/p(y|\beta \in \mathcal{B}_0)$. Formulae for estimating these quantities are given in Supp F.

5.2.1. *Non-identifiability of authority and seniority effects.* We are interested in separating the relative authority $U_j^{(t)}$ of a actor from the status $Z_j^{(t)}$. There are two sources of non-identifiability. The latent variables U have a label swapping symmetry: the posterior is invariant under permutation of the columns of $U^{(t)}$ if the same permutation is applied at each $t \in [B, E]$. One simple summary which is invariant under column permutation is the row-average. This gives the estimated average posterior authority for actor j at time t,

$$\bar{U}_j^{(t)} = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{K} \sum_{k=1}^{K} U_{j,k}^{(t,l)}. \tag{20}$$

A plot of $\bar{U}_j^{(t)}$ against t shows j's changing authority. Average status $\bar{Z}_j^{(t)}$ is defined similarly.

The second source of non-identifiability is shift invariance of $h^{(t)}$ under

$$U_j^{(t)} \to U_j^{(t)} + 1_K u^{(t)},\ t \in [B, E],\ j \in \mathcal{M}_t, \tag{21}$$

$$\beta_r \to \beta_r + c,\ r \in [S], \tag{22}$$

where $u^{(t)} \in \mathbb{R}$ is a common shift applied to each actor, which may vary over t, and $c \in \mathbb{R}$ is common shift applied to all effects. The proper U and β priors shrink these shifts towards zero. We project these degrees of freedom out by subtracting the averages, $\bar{U}_j^{(t)} \to \bar{U}_j^{(t)} - \mathcal{M}_t^{-1} \sum_j \bar{U}_j^{(t)}$ and $\beta_r \to \beta_r - S^{-1} \sum_{r'} \beta_{r'}$, before computing the summary statistics and plotting. A similar issue arises in the Plackett-Luce time-series model in Supp J.1.

6. Results.

6.1. *Fitted models.* Prior distributions are summarised in Sec. 3.3. Unless otherwise indicated we present results for the likelihood $p_{(D)}$ in Eqn 5 as clerks wrote lists from top down. Results are near-identical with $p_{(U)}$ in Eqn 6. Experiments showed fractionally lower estimated noise probabilities p for $p_{(U)}$ than $p_{(D)}$ (see Fig 21). This suggests the $p_{(U)}$ model is a slightly better fit, but there is little difference. The greatest number of active bishops is $D = 22$ so we take $K = 11$ (see end of Sec. 3.2.1) for the dimension of the latent status feature vectors $Z_j^{(t)},\ j \in \mathcal{M}, t \in [B, E]$ in our main analyses in Secs. 6.2 and 6.3. We check robustness by taking $K = 2$ and $K = 22$ in Sec. G.3 of Supp G. The β-dimension is $S = 21$ (less than D as there is a tie at seniority-rank 21 in 1133, the year with the most active bishops).

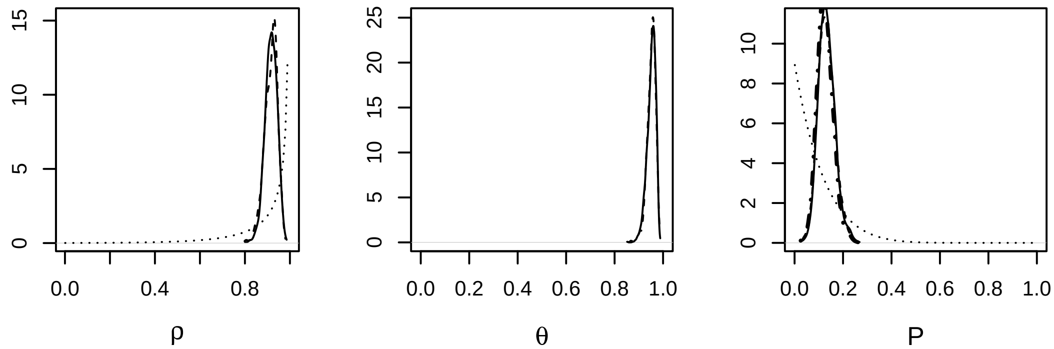

FIG 10. *Posterior densities for ρ, θ and p from the unconstrained seniority effects analysis in Sec. 6.2. Two independent MCMC runs are shown (solid and dashed). The dotted line in the ρ and p graphs is their prior. The prior for θ is uniform. The thick dash-dot curve in the p-density is the posterior density for a uniform prior on p.*

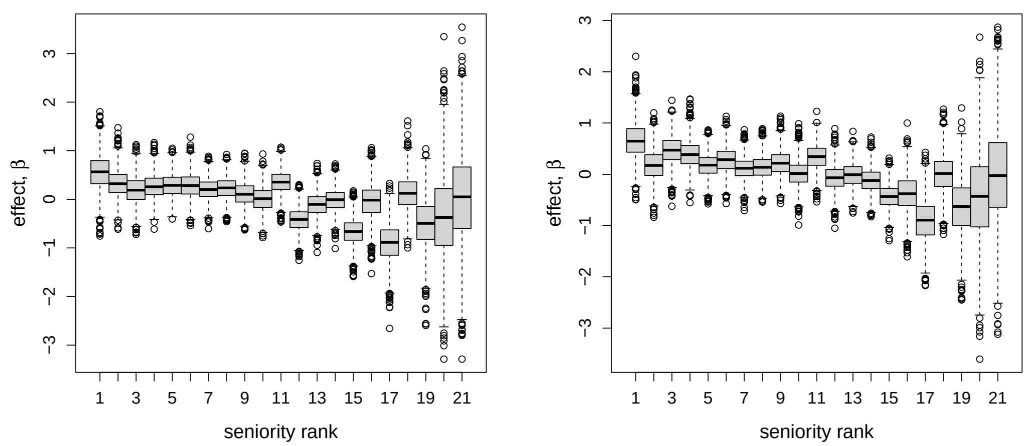

FIG 11. *(Left) Marginal posterior distributions of seniority effect parameters from the unconstrained seniority effects analysis in Sec. 6.2 with likelihood $p_{(D)}$ in Eqn 5. (Right) As left with likelihood $p_{(U)}$ in Eqn 6. See Fig 30 in Supp J.1 for the corresponding plot for the Plackett-Luce time-series model comparison.*

6.2. *Analysis with unconstrained seniority effects.* We begin by presenting our results for the full data set defined in Sec. 2.3. We first check that we see declining seniority effect at lower seniority, so we do not constrain the seniority effects to be ordered and take $\beta \in \mathcal{B}_0$.

Traces in Fig 22 in Supp G.1 for MCMC targeting the posterior in Sec. 3.3 show convergence. Marginal posterior densities from two independent runs are shown in Fig 10 and are near-identical. The correlation ρ of features in $U_j^{(t)},\ j \in \mathcal{M}$ at each fixed time $t \in [b_j, e_j]$ is close to one, supporting relatively deeper posets. The time-series correlation parameter θ is close to one, indicating strong serial correlation between $U_j^{(t)}$ and $U_j^{(t+1)}$, and therefore also $h^{(t)}$ and $h^{(t+1)}$. Finally, the error probability p, which controls the extent to which lists y_i may depart from the linear extensions $\mathcal{L}[h^{(\tau_i)}]$, is small, as we would expect if the poset model captures the variation in lists. The prior for p has $\delta = 9$, favoring small p. We checked robustness to this choice: Fig 10 superimposes the posterior density when $\delta = 1$ (uniform, estimated by importance sampling); the shift from the $\delta = 9$ posterior is barely discernible.

In Fig 11 we plot marginal posterior β distributions in the posterior with likelihood $p_{(D)}$ (noise is random upward displacement) and $p_{(U)}$ (random downward displacement) respectively. There is a clear downward trend with increasing seniority-rank value (ie, lower senior-

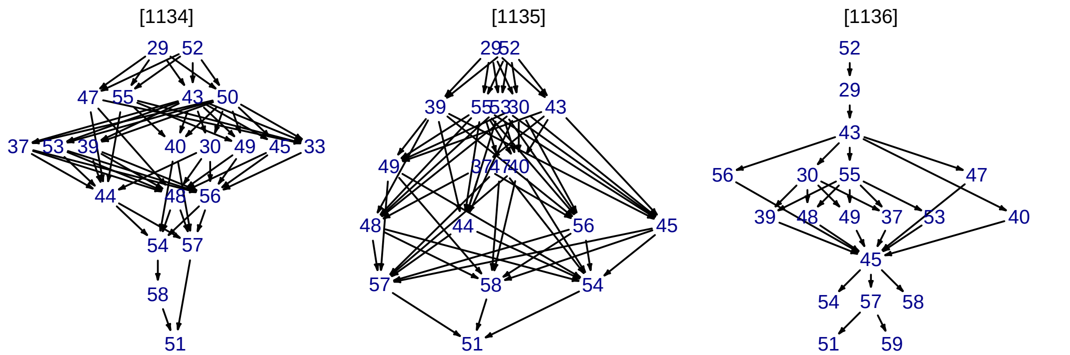

FIG 12. *(constrained seniority effects analysis of Sec. 6.3) Three of seventy-six years from one MCMC-sampled state,* $h^{(t,L)}$, $t \in \{1134, 1135, 1136\}$. *Vertex numbers correspond to bishops names in Fig 24 in Supp G.2.*

ity) as we expect. When the rank is large (18-21) we have few instances of bishops with that rank (see Fig 5), and distributions trend back to the prior. Under both observation models $p_{(U)}$ and $p_{(D)}$, the effect β_{11} is "out of order". This is because several bishops who spent several years at seniority-rank 11 (William Giffard, bishop of Winchester, Richard de Belmeis I, bishop of London and Henry de Blois, Bishop of Winchester) were connected with royalty, so it was their authority and not their seniority which pushed them up the lists. This is best modeled by imposing the seniority-effect order constraint $\beta \in \mathcal{B}_S$ as we do in the next section. In Sec. F.2 we test $\beta \in \mathcal{B}_S$ against the unconstrained model $\beta \in \mathcal{B}_0$ by estimating the Bayes factor, using a Savage-Dickey estimator. We find clear evidence in favor of the constraint. In summary we see in this first analysis the structures we anticipated.

6.3. *Analysis with constrained seniority effects.* The assumption of decreasing seniority effect with decreasing seniority rank is supported on historical and statistical grounds so we now impose the constraint $\beta \in \mathcal{B}_S$. We omit the ρ, θ and p posterior densities as they are essentially unchanged from Fig 10. Marginal posterior distributions for the unknown list dates $\tau_i,\ i \in \mathcal{I}$ are given in Fig 26 in Supp G.2.

In Fig 12 we plot posets $h^{(t,L)}, t \in [1134, 1136]$ from one MCMC sample state (the final sample in the MCMC sample output $(h^{(t,l)})_{t\in[B,E]},\ l \in [L]$). These three years bracket 1135 when Stephen became king. The number and length of lists in this period is relatively large (see Fig 3). In Fig 13 we plot posterior consensus posets $\bar{h}^{(t)}$ estimated at the same years (transitive reductions for ease of viewing, see Supp G.2 for all years). The transtive closures in Fig 23 in Supp G.2 have many more strongly supported edges as a chain of weakly supported relations give strongly supported relations from chain head to tail.

In Fig 25 in Supp G.2 we plot the evolving mean status values $\bar{Z}_j^{(t)}$ for each bishop as a function of time. Bishops are grouped by diocese. These curves have a "sawtooth" pattern, as the "status" measure Z trends up through the tenure of a bishop as their seniority increases. It drops down when a new bishop enters the diocese with low seniority. By contrast the curves in Fig 14 show the evolving mean authority values $\bar{U}_j^{(t)}$ for each bishop. These curves are flatter as the effect of seniority is removed. Nigel, bishop of Ely is revealing. His status in Fig 25 is fairly flat. This is because his mean authority $\bar{U}_j^{(t)}$ declined as his seniority increased.

The continuity in authority (but not status) of bishops over time within a diocese in Fig 14 is noteworthy. There are some exceptions. For example, Henry de Blois started with higher authority than might be expected based only on the diocese. Some dioceses seem to be better (Winchester, London, Lincoln) than others (Chichester, Rochester, the dioceses in Normandy). The bishops of London and Winchester had gained precedence over their colleagues

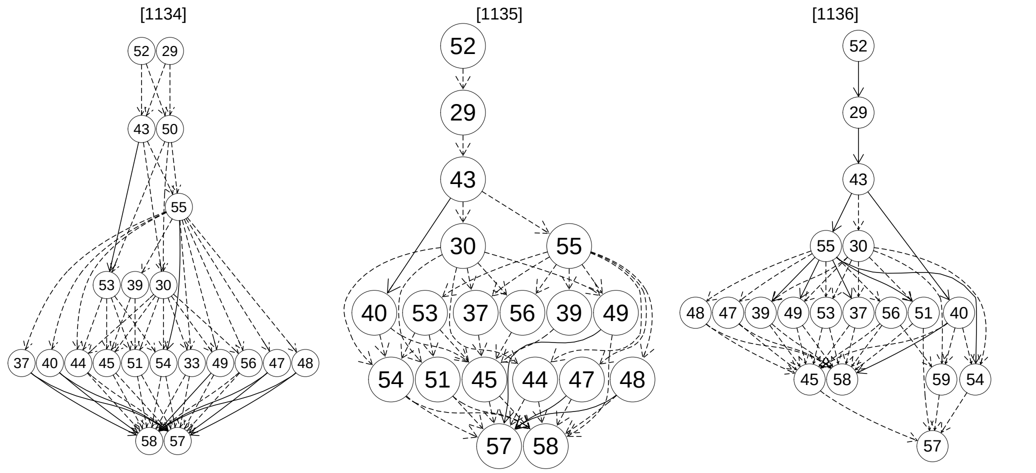

FIG 13. *(constrained-effects model Sec. 6.3) Consensus posets for 1134-1136 (transitive reductions). For closures see Supp G.2 Fig 23. Dashed/solid edges have posterior support greater than* $0.5/0.9$. *Numbering as Fig 12.*

at the Council of London in 1075 and Lincoln came in the later middle ages to rank after Winchester. However, there is uncertain evidence from as early as 1138 that the bishop of Lincoln might assume the role of London or Winchester in their absence and consequently that Lincoln already enjoyed a degree of precedence (Johnson, 2013).

6.4. *Discussion of results.* From a historical perspective, there are three significant outcomes. The first is the strong emphasis on the seniority and precedence of individual bishops in the witness lists. Historians often link the relative position of witnesses to an assessment of their political significance, but the analysis here shows that royal scribes were strongly influenced by the rules on seniority and precedence expressed at the Council of London, held by the English church in 1075 (Council of London, 1075, clause 1, Clover and Gibson (1979)).

The second is the position of Normandy within the Anglo-Norman realm. Fig 14 shows that early in this period (before about 1100) Norman bishoprics enjoyed high status, but that this declined from the early twelfth century. This change is particularly marked for Avranches, Bayeux and Évreux, whilst no English bishoprics show a comparable trend. This should inform the ongoing debate about the relationship between England and Normandy (Bates (2013), chapter 5). This change might represent a principled decision by royal scribes to rank Norman bishops lower in precedence than their English counterparts, or it might be explained politically. The smaller Norman dioceses may have been less attractive to ambitious clergymen, and there were periods when Normandy and England were ruled separately (most notably, 1144-54), so that Norman bishops were external to the English kingdom.

The third concerns how far the behaviour of individual bishops could change their status. Bishops were active politically and could fall into disgrace. Thus, Bishop Nigel of Ely had high status for a junior bishop in the 1130s, but from his disgrace in 1139 his status fell, contrary to the usual pattern. Nigel's pattern is unique; it is not replicated by that for other disgraced bishops, such as Ranulf Flambard of Durham after 1100 and Alexander of Lincoln after 1139. These differences presumably reflect the nature of the disgrace itself. The estimated poset relations accord with known political favour. For example, Henry of Blois and Odo of Bayeux, who were related to the royal house (Odo was William the Conqueror's half-brother, and Henry was Stephen's brother), are highly ranked. Referring to Fig 27 in

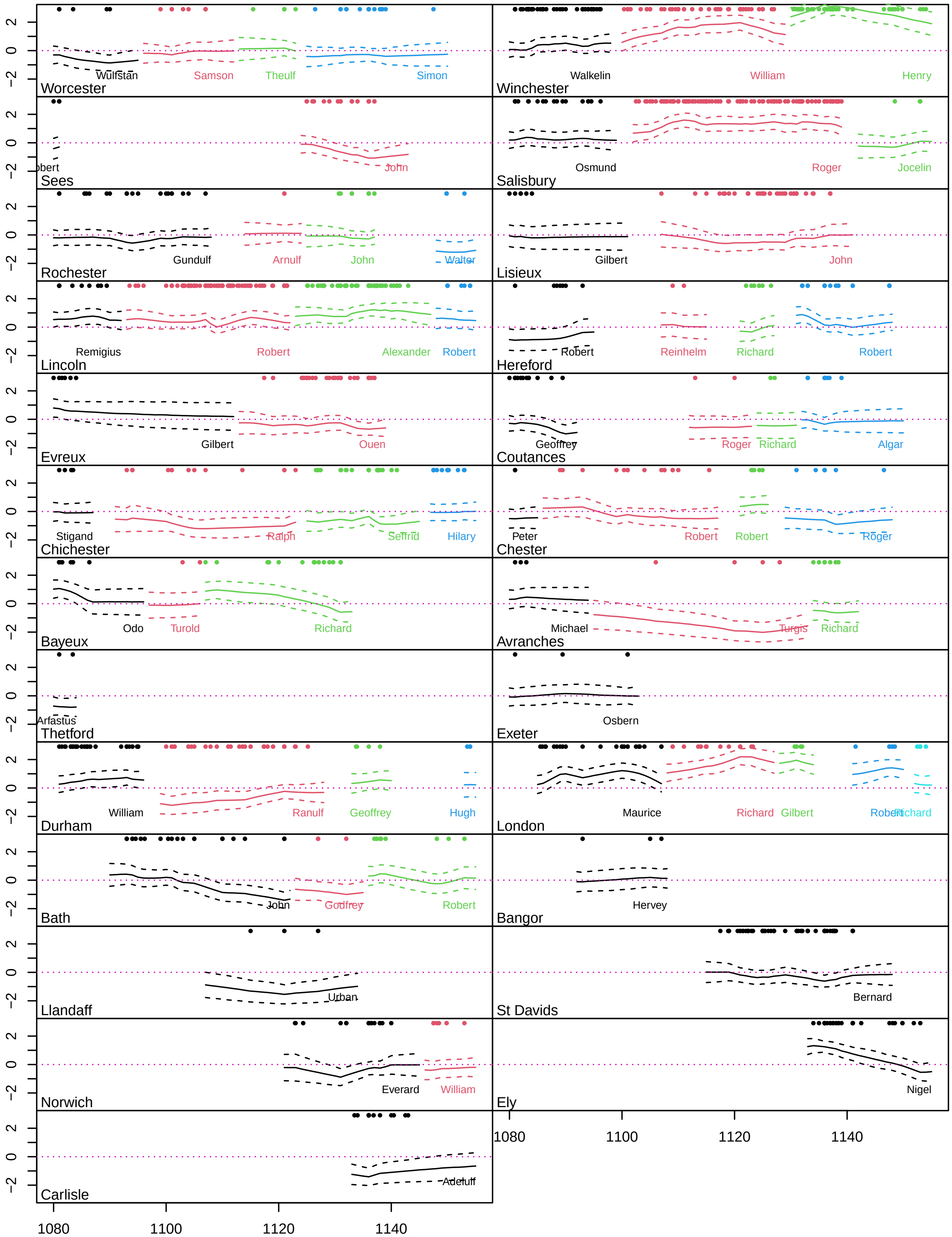

FIG 14. *(constrained seniority effects analysis of Sec. 6.3) Bishop-authority curves* $\bar{U}_j^{(t)}$ *(y-axis values, solid curves) plotted for each bishop* $j \in \mathcal{M}$ *as a function of time (x-axis, the year) from* b_j *to* e_j *with uncertainty at one-sigma. The dots at the top of each graph show the times of lists in which the matched bishop below appeared.*

Supp G.2, although Henry (52) did well from the start, until 1134 he shares top spot in consensus orders with Roger of Salisbury (29). From this date he is promoted ahead of anyone else. This suggests that his brother becoming king in 1135 had an impact on his position.

Referring to Fig 27, reconstructed orders seem relatively shallow, typically one (1097) or two (1124) groups of middle-ranked bishops and a few above or below. We may be concerned that this reflects an overly informative prior and differences in how often bishops witness. We tested this by simulating synthetic data in which the true posets were total orders (see Supp K.2), but with the same list memberships as the real data. We reconstructed the true total orders well. If the true orders were total orders we would see this in our analysis.

7. Comparisons with other models. In this section we define models over Bucket Orders and VSP orders. Calculation of $|\mathcal{L}[H]|$ for $H \in \mathcal{H}_\mathcal{A},\ \mathcal{A} = \{1,\dots,m\}$ is linear in m on these subspaces (Wells, 1971), so if these models were preferred then we would use them. We find they are not adequate to represent a time-evolving hierarchy over long periods of time but can give a good fit over short time periods. Jiang et al. (2023) applies a fixed-time VSP model to all the witness list data (not just bishops). Some orders have over 200 actors, with lists exceeding 50 in length, and are out of reach for our full poset analysis.

In Supp J.1 and J.2 we compare our model with Plackett-Luce models. Bayesian analysis of time-series Plackett-Luce in Supp J.1 gives similar results for a parameter function corresponding to the average authority in Eqn 20. Analysis of a Plackett-Luce mixture model in Supp J.2 on short time intervals shows our model is preferred.

7.1. *Bucket Orders and Vertex-Series-Parallel partial orders.* VSP orders (Valdes, 1978; Valdes et al., 1982) are built recursively from the ground set by taking series and parallel combinations of posets. We give this intuitive definition in Supp I. Valdes et al. (1982) gives a concise characterisation. For any set $\mathcal{A}$, the class of all VSPs $\mathcal{V}_\mathcal{A}$ is identical to the set of posets $H \in \mathcal{H}_\mathcal{A}$ which do not contain a set of vertices $\mathcal{A}' = \{j_1, ..., j_4\}$ with sub-graph $H' = H \cap (\mathcal{A}' \times \mathcal{A}')$ isomorphic to the "forbidden subgraph" $F = \{\langle 1,2\rangle, \langle 3,2\rangle, \langle 3,4\rangle\}$ shown in Fig 29 at right. After vertex relabelling, F and H' must be identical, so edges absent in F are absent in H'. This makes it straightforward to test if a poset H is a VSP-order.

A sub-class of VSPs called "Bucket Orders" has a particularly simple closed form for $|\mathcal{L}[H]|$. Actors are grouped into "buckets". Actors in the same bucket are unordered and a complete order holds over buckets. Formally, if $\mathcal{K}_\mathcal{A}$ is the class of Bucket Orders on $\mathcal{A}$ then $b \in \mathcal{K}_\mathcal{A}$ iff there is a partition $\mathcal{A}_1, ..., \mathcal{A}_P$ of $\mathcal{A}$ into P buckets such that for each $k \in [P]$ and all $j_1, j_2 \in \mathcal{A}_k$ we have $\langle j_1, j_2\rangle \notin b$ and for all pairs $1 \le k_1 < k_2 \le P$ of buckets and all $j_1 \in \mathcal{A}_{k_1}$ and $j_2 \in \mathcal{A}_{k_2}$ we have $\langle j_1, j_2\rangle \in b$. VSPs and Bucket Orders are a small subset of partial orders. For example, if $|\mathcal{A}| = 18$ (the largest for which OEIS Foundation Inc (2022) gives cardinalities) we have $|\mathcal{H}_{[18]}| \simeq 2 \times 10^{35}$, $|\mathcal{V}_{[18]}|/|\mathcal{H}_{[18]} \simeq 10^{-11}$ and $|\mathcal{K}_{[18]}|/|\mathcal{H}_{[18]}| \simeq 10^{-17}$.

7.2. *Bucket and VSP-order models.* Suppose we are interested in learning about order relations over a period $[t_1, t_2]$ with $B \le t_1 \le t_2 \le E$. If we could justify restricting the process of fitted posets $h \in \mathcal{H}^{(t_1,t_2)}$ to a VSP-order-process $h \in \mathcal{V}^{(t_1,t_2)}$ with

$$\mathcal{V}^{(t_1,t_2)} = \mathcal{V}_{\mathcal{M}_{t_1}} \times \mathcal{V}_{\mathcal{M}_{t_1+1}} \times ... \times \mathcal{V}_{\mathcal{M}_{t_2}},$$

or a bucket-order process $h \in \mathcal{K}^{(t_1,t_2)}$ with

$$\mathcal{K}^{(t_1,t_2)} = \mathcal{K}_{\mathcal{M}_{t_1}} \times \mathcal{K}_{\mathcal{M}_{t_1+1}} \times ... \times \mathcal{K}_{\mathcal{M}_{t_2}},$$

then likelihood evaluation would be fast. However, this is not well-evidenced in our setting: the consensus order from 1136 in Fig 13 contains the poset $H' = \{\langle 54, 57\rangle, \langle 47, 57\rangle, \langle 47, 58\rangle\}$ in Fig 1, with each included edge having posterior probability 0.8 or above and absent edges below 0.5, so the true poset is probably not a VSP as it contains a suborder ismorphic to F.

7.3. *Test results.* Bayes Factors $B_{\mathcal{V},\mathcal{H}}$ and $B_{\mathcal{K},\mathcal{H}}$ measuring the evidence for VSP poset-models and Bucket-Order models are defined in Sec. I.2. We estimate these on the same twelve short (five-year) time intervals we used for the Plackett-Luce analysis in Supp J.2 for accurate estimation. The seniority covariate is a near-constant in these short time intervals so seniority effects were set to $\beta_r = 0,\ r = 1, ..., S$ in the fitted poset model. The likelihood $p_{(U)}$ in Eqn 6 was used. For analysis in interval $[t_1, t_2]$, we take K, the dimension of $U_j^{(t)}$, to be $\lfloor \frac{1}{2} \max_{t \in [t_1,t_2]} m_t \rfloor$. The prior for ρ is $\mathrm{Beta}(1, 1/6)$ and otherwise as Sec. 3.3.

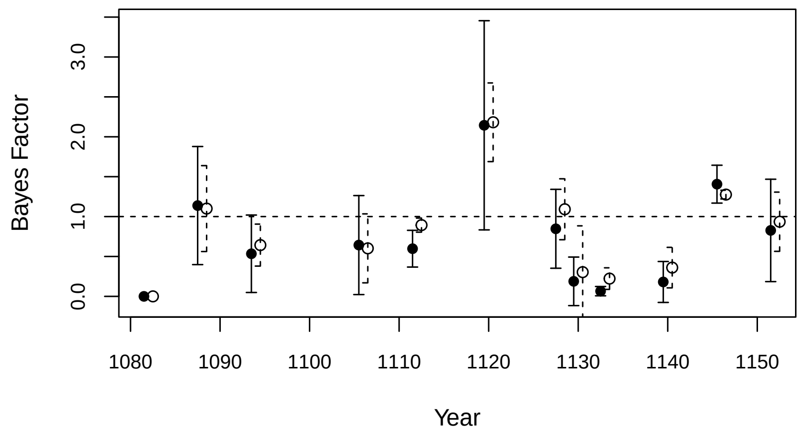

FIG 15. *Bayes Factors for VSP orders (dashed lines) and Bucket Orders (solid) fitted over the year intervals in Table 2. A value less than one is evidence against VSP or Bucket Orders. The x-axis value for each bar is the centre of the corresponding interval, $(t_1 + t_2)/2$. Error bars are two standard deviations.*

Results are given in Table 2 in Supp I and plotted in Fig 15. Each pair of points is an independent MCMC run. The prior probabilities $\pi(h \in \mathcal{V}^{(t_1,t_2)})$ and $\pi(h \in \mathcal{B}^{(t_1,t_2)})$ in Table 2 are surprisingly large given the sparsity of VSPs and Bucket Orders, so our prior for posets must favor VSPs and Bucket Orders. The rules forming VSP-orders and Bucket Orders compare groups rather than actors in a socially plausible way, so this may be a good thing. As Fig 15 shows, the data favour partial orders or are neutral, except for 1118-1122, so Bucket Orders and VSPs may be acceptable over some short time intervals. In some intervals VSPs and Bucket Orders are rejected (1080-1084 and around 1132-1134). The evidence for posets visible in Fig 15 will accumulate over longer time series.

8. Conclusions. A new class of poset-models for time-series rank-data is summarised in Sec. 3.3. The latent variable poset-parameterisation in Eqn 15 made it straightforward to introduce a parameter controlling poset depth and incorporate actor-covariates informing the position of actors in the hierarchy. We fit the model to witness-list data in which the actors are eleventh and twelfth century bishops. In Sec. 6.2 we saw that the model recovered structure in the data which was anticipated by historians. In particular, the dependence of the status of a bishop on their seniority is clear in Fig 11. We checked for evidence that the depth parameter ρ, correlation θ and error probability p varied over time by looking at short time intervals (see Supp H) and found no evidence against our assumption of constant values over time. Further support comes from model comparison against a Placket-Luce mixture in Supp J.2 which favoured our model. The time-series extension of the Plackett-Luce model in Supp J.1 gave similar results for seniority effects and evolving authority, showing that the data overwhelm these model variations and conclusions are robust. The times-series Plackett-Luce model is fairly time consuming to fit so there was no great gain in efficiency over posets on our data.

We gave our main analysis in Sec. 6.3. This is the first quantitative analysis of these data and gave insights which historians find interesting. With few exceptions, witness-lists reflect precedence by diocese and seniority more than changing royal favour. We separated the effects of authority and seniority on status and confirmed (Johnson, 2013) that the bishops of London, Winchester and Lincoln had high authority and that Rochester had no special status. Personal authority changes in a few cases: the high status originally given to Nigel of Ely unwound, and Roger of Salisbury bucked the trend. This was known to historians. The apparent decline in authority of Norman bishops (as expressed in the lists) was unknown. Consensus partial orders (see Fig 13) gave an interpretable visualisation of the underlying social hierarchy in each year. The problem treated here appears in other guises: there is an extensive literature estimating animal dominance hierarchies with many similarities. For example, Foerster et al. (2016) studies a time-series of pairwise chimpanzee interactions and finds a role for seniority and evolution of status.

There is work to be done in computation and methodology to make these models more broadly applicable. First, our MCMC is time consuming (the experiments in this paper add up to about two years of MCMC if run as a single serial process). Whilst a careful analysis minimising any approximation of the target distribution is justified (there will never be any more for this period), scalable methods would be welcome, and VSP-orders may be an acceptable compromise from a modelling perspective. This may allow us to fit more complex noise models and treat "top-k" preference orders. Analysis of lists including lay witnesses (Jiang et al., 2023) required scalable methods in order to count linear extensions in partial orders over hundreds of elements. Application of our methods to general preference orders on thousands of items and thousands of lists is not presently feasible.

Second, many applications of poset-based models will require substantial model-building, paralleling the evolution of Mallows and Plackett-Luce models and including hierarchical models and mixture models for structured populations and clustering. The clerks who made the lists may have differed on status assessment so unrepresented group structure may be present. Some developments are given in Jiang and Nicholls (2024) (hereafter JN24). We worked with strong partial orders. Models for weak partial orders with ties are given in JN24. Also, our statistical model for "noise" in realised lists assumes errors occur in one direction only (queue jumping up or down, but not both). Noise models with bi-directional errors are explored in Jiang et al. (2023) and JN24. Statistical tools for selecting the number of features K in the status vector of a bishop would remove the need for robustness checks (our Supp G.3). This is addressed for fixed-time data in JN24 using reversible-jump MCMC.

Third, a small number of other covariates beside seniority are available in the data and might be explored in model elaboration. Covariates might enter the noise model also, to inform the probability and magnitude of displacements. One unexplored weakness of our error model is apparent in lists of length two: if the two bishops are ordered in a deep poset then the noise model assigns the same probability for the "wrong" order whether the two bishops are at each end of the poset (very different status) or adjacent in its transitive reduction (near-equal status). Displacements are probably "over-dispersed" in the noise-model.

From a historical perspective this study raises several questions. Complex precedence structures seem to have existed, but how were they known? Was there some kind of precedence handbook, or other means of transmission? Comparisons between patterns of precedence relations in pre-conquest lists and lists from later periods might be revealing. Later documents are more accurately dated and so a more fine-grained analysis may be possible. Finally, there are forgeries among acta of this period. Cases that go against the usual pattern may be an indicator of forgery.

Software used to carry out the analysis presented in this paper are available at

https://github.com/gknicholls/Partial-order-HMM-public.git.

Acknowledgements. The authors thank Dr Simon Urbanek for writing an R wrapper to compile the lecount() code in c++. GKN thanks Dr Oemetse Mogapi and Prof Tom Snijders for introducing him to this topic and Sir Bernard Silverman for helpful conversations regarding marginal consistency.

Funding. JL was supported in part by Marsden grant MFP-UOA2131 and HRC 22/377/A from New Zealand Government funding.

REFERENCES

Arcagni, A., A. Avellone, and M. Fattore (2022). Complexity reduction and approximation of multidomain systems of partially ordered data. *Computational Statistics & Data Analysis 173*, 107520.

Asfaw, D., V. Vitelli, Ø. Sørensen, E. Arjas, and A. Frigessi (2017). Time-varying rankings with the Bayesian Mallows model. *Stat 6*(1), 14–30.

Bates, D. (1997). *Family Trees and the Roots of Politics: The Prosopography of Britain and France from the Tenth to the Twelfth Century*, Chapter The Prosopographical Study of Anglo-Norman Royal Charters, pp. 89–102. Boydell Press.

Bates, D. (1998). *The Acta of William I, 1066–1087*. Regesta regum Anglo-Normannorum 1066-1154. Oxford.

Bates, D. (2013). *The Normans and Empire*. Ford lectures. OUP Oxford.

Baxter, S. (2007). *The Earls of Mercia: Lordship and Power in Late Anglo-Saxon England*. Oxford University Press.

Beerenwinkel, N., N. Eriksson, and B. Sturmfels (2007). Conjunctive Bayesian networks. *Bernoulli 13*(4), 893–909.

Bengio, Y. and Y. Grandvalet (2004). No unbiased estimator of the variance of k-fold cross-validation. *Journal of Machine Learning Research 5*, 1089–1105.

Bogart, K. P. (1973a, 5). Maximal dimensional partially ordered sets I. Hiraguchi's theorem. *Discrete Mathematics 5*, 21–31.

Bogart, K. P. (1973b). Preference structures i: Distances between transitive preference relations. *The Journal of Mathematical Sociology 3*(1), 49–67.

Brightwell, G. (1993). *Surveys in Combinatorics*, Volume 187 of *London Mathematical Society Lecture Note Series*, Chapter Models of random partial orders, pp. 53–83. Cambridge Univeristy Press.

Brightwell, G. and P. Winkler (1991). Counting linear extensions. *Order 8*(3), 225–242.

Caron, F. and A. Doucet (2012). Efficient Bayesian inference for generalized Bradley-Terry models. *J. Comput. Graph. Statist. 21*(1), 174–196.

Caron, F. and Y. Teh (2012). Bayesian nonparametric models for ranked data. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger (Eds.), *Advances in Neural Information Processing Systems*, Volume 25. Curran Associates, Inc.

Caron, F., Y. W. Teh, and T. B. Murphy (2014). Bayesian nonparametric Plackett–Luce models for the analysis of preferences for college degree programmes. *The Annals of Applied Statistics*, 1145–1181.

Clover, H. and M. Gibson (1979). *The Letters of Lanfranc, Archbishop of Canterbury*. Oxford.

Cronne, H. A. and R. H. C. Davis (1968). *Regesta Regis Stephani ac Mathildis Imperatricis ac Gaufridi et Henrici Ducum Normannorum, 1135–1154*. Regesta regum Anglo-Normannorum 1066-1154. Oxford.

Davis, H. and R. Whitwell (1913). *Regesta regum Anglo-Normannorum 1066-1154: Regesta Willelmi Conquestoris et Willelmi Rufi 1066-1100*. Regesta regum Anglo-Normannorum 1066-1154. Oxford.

Deng, K., S. Han, K. J. Li, and J. S. Liu (2014). Bayesian aggregation of order-based rank data. *Journal of the American Statistical Association 109*(507), 1023–1039.

Diaconis, P. (1988). Group representations in probability and statistics. *Lecture Notes-Monograph Series 11*, i–192.

Durante, D., D. B. Dunson, and J. T. Vogelstein (2017a). Nonparametric Bayes modeling of populations of networks. *Journal of the American Statistical Association 112*(520), 1516–1530.

Durante, D., D. B. Dunson, and J. T. Vogelstein (2017b). Rejoinder: Nonparametric Bayes modeling of populations of networks. *Journal of the American Statistical Association 112*(520), 1547–1552.

Dushnik, B. and E. W. Miller (1941). Partially ordered sets. *American Journal of Mathematics 63*(3), 600–610.

FEA (2022). Fasti Ecclesiae Anglicanae, 1066-1300. London: Institute of Historical Research; British History Online; Last accessed 22 October 2022; http://www.british-history.ac.uk/fasti-ecclesiae/.

Fishburn, P. C. and W. V. Gehrlein (1975). A comparative analysis of methods for constructing weak orders from partial orders. *J. Math. Sociology 4*(1), 93–102.

Fligner, M. A. and J. S. Verducci (1986). Distance based ranking models. *Journal of the Royal Statistical Society. Series B (Methodological) 48*(3), 359–369.

Foerster, S., M. Franz, C. M. Murray, I. C. Gilby, J. T. Feldblum, K. K. Walker, and A. E. Pusey (2016). Chimpanzee females queue but males compete for social status. *Scientific Reports 6*(1), 35404–35415.

Friedell, M. F. (1967). Organizations as semilattices. *American Sociological Review 32*(1), 46–54.

Froehlich, H., M. Fellmann, H. Sueltmann, A. Poustka, and T. Beissbarth (2007). Large scale statistical inference of signaling pathways from RNAi and microarray data. *BMC Bioinformatics 8*(386).

Gazes, R. P., R. R. Hampton, and S. F. Lourenco (2017). Transitive inference of social dominance by human infants. *Developmental Science 20*(2), e12367.

Gionis, A., H. Mannila, K. Puolamäki, and A. Ukkonen (2006). Algorithms for discovering bucket orders from data. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 561–566.

Given-Wilson, C. (1991). Royal charter witness lists 1327-1399. *Medieval Prosopography 12*(2), 35–93.

Glickman, M. E. and J. Hennessy (2015). A stochastic rank ordered logit model for rating multi-competitor games and sports. *Journal of Quantitative Analysis in Sports 11*(3), 131–144.

Guiver, J. and E. Snelson (2009). Bayesian inference for Plackett-Luce ranking models. In *proceedings of the 26th annual international conference on machine learning*, pp. 377–384.

Gwee, X. Y., I. C. Gormley, and M. Fop (2023). A latent shrinkage position model for binary and count network data. *Bayesian Analysis*, 1 – 29.

Haskins, G. L. (1938). Charter witness lists in the reign of king john. *Speculum 13*(3), 319–325.

Hiraguchi, T. (1951). On the dimension of partially ordered sets. *The science reports of the Kanazawa University 1*, 77–94.

Hunter, D. R. (2004). MM algorithms for generalized Bradley-Terry models. *The Annals of Statistics 32*(1), 384 – 406.

Irurozki, E., B. Calvo, and J. A. Lozano (2016). PerMallows: An R Package for Mallows and generalized Mallows models. *Journal of Statistical Software 71*(12), 1–30.

Irurozki, E., B. Calvo, and J. A. Lozano (2019). Mallows and generalized Mallows model for matchings. *Bernoulli 25*(2), 1160 – 1188.

Jiang, C. and G. K. Nicholls (2024). Non-parametric Bayesian inference for partial orders with ties from rank data observed with Mallows noise. https://arxiv.org/abs/2408.14661.

Jiang, C., G. K. Nicholls, and J. E. Lee (2023, 31 Jul–04 Aug). Bayesian inference for vertex-series-parallel partial orders. In R. J. Evans and I. Shpitser (Eds.), *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, Volume 216 of *Proceedings of Machine Learning Research*, pp. 995–1004. PMLR.

Johnson, C. and H. A. Cronne (1966). *Regesta Henrici Primi : 1100-1135*. Regesta regum Anglo-Normannorum 1066-1154. Oxford.

Johnson, D. (2013, 10). Bishops and deans: London and the province of Canterbury in the twelfth century*. *Historical Research 86*(234), 551–578.

Kangas, K., T. Hankala, T. Niinimäki, and M. Koivisto (2016). Counting linear extensions of sparse posets. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, pp. 603–609. AAAI Press.

Kangas, K., M. Koivisto, and S. Salonen (2019). A faster tree-decomposition based algorithm for counting linear extensions. *Algorithmica*, 1–18.

Karzanov, A. and L. Khachiyan (1991). On the conductance of order Markov chains. *Order 8*, 7–15.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Keynes, S. (1980). *The Diplomas of King Aethlred "the Unready" 978–1016*. Cambridge Studies in Medieval Life and Thought: Third Series. Cambridge University Press.

Kleitman, D. J. and B. L. Rothschild (1975). Asymptotic enumeration of partial orders on a finite set. *Transactions of the American Mathematical Society 205*, 205–220.

Knuth, D. E. and J. L. Szwarcfiter (1974). A structured program to generate all topological sorting arrangements. *Information Processing Letters 2*(6), 153–157.

Licence, T. (2020, 09). *Edward the Confessor: Last of the Royal Blood*. Yale University Press.

Lu, T. and C. Boutilier (2014). Effective sampling and learning for Mallows models with pairwise-preference data. *J. Mach. Learn. Res. 15*(1), 3783–3829.

Luce, R. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology 15*(3), 215–233.

Luce, R. D. (1959). On the possible psychophysical laws. *Psychological review 66*(2), 81.

Mallows, C. L. (1957). Non-null ranking models. i. *Biometrika 44*(1/2), 114–130.

Mannila, H. (2008). Finding total and partial orders from data for seriation. In J.-F. Boucault (Ed.), *Discovery Science*, Volume 5255 of *LNAI*, Berlin Heidelberg, pp. 16–25. Springer-Verlag.

Mannila, H. and C. Meek (2000). Global partial orders from sequential data. In *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '00, New York, NY, USA, pp. 161–168. Association for Computing Machinery.

Martin, J. L. (2002). Some algebraic structures for diffusion in social networks. *The Journal of Mathematical Sociology 26*(3), 123–146.

McKeough, K. and M. Glickman (2024). Plackett–Luce modeling with trajectory models for measuring athlete strength. *Journal of Quantitative Analysis in Sports 20*(1), 21–35.

Meilă, M. and H. Chen (2010). Dirichlet process mixtures of generalized Mallows models. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, UAI'10, Arlington, Virginia, USA, pp. 358–367. AUAI Press.

Meilă, M. and H. Chen (2016). Bayesian non-parametric clustering of ranking data. *IEEE Transactions on Pattern Analysis and Machine Intelligence 38*(11), 2156–2169.

Mogapi, O. (2009). *A Latent Partial Order Model for Social Networks*. Ph. D. thesis, University of Oxford.

Mollica, C. and L. Tardella (2017). Bayesian Plackett-Luce mixture models for partially ranked data. *Psychometrika 82*(2), 442–458.

Mollica, C. and L. Tardella (2020). PLMIX: an R package for modelling and clustering partially ranked data. *Journal of Statistical Computation and Simulation 90*(5), 925–959.

Muir Watt, A. (2015). *Inference for partial orders from random linear extensions*. Ph. D. thesis, University of Oxford.

Nicholls, G. K. and A. Muir Watt (2011). Partial order models for episcopal social status in 12th century England. *IWSM 2011*, 437.

Niinimäki, T., P. Parviainen, and M. Koivisto (2016). Structure discovery in Bayesian networks by sampling partial orders. *Journal of Machine Learning Research 17*(57), 1–47.

OEIS Foundation Inc, . (2022). The on-line encyclopedia of integer sequences. Published electronically at https://oeis.org.

Plackett, R. L. (1975). The analysis of permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 24*(2), 193–202.

Rising, J. (2021). Uncertainty in ranking. https://arxiv.org/abs/2107.03459.

Roberts, J. M. (1990). Modeling hierarchy: Transitivity and the linear ordering problem. *The Journal of Mathematical Sociology 16*(1), 77–87.

Rousseau, J. and K. Mengersen (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(5), 689–710.

Russell, J. C. (1937). Social status at the court of King John. *Speculum 12*(3), 319–329.

Sakoparnig, T. and N. Beerenwinkel (2012). Efficient sampling for Bayesian inference of conjunctive Bayesian networks. *Bioinformatics 28*(18), 2318–2324.

Seshadri, A., S. Ragain, and J. Ugander (2020). Learning rich rankings. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin (Eds.), *Advances in Neural Information Processing Systems*, Volume 33, pp. 9435–9446. Curran Associates, Inc.

Sharpe, R., D. Carpenter, H. Doherty, M. Hagger, and N. Karn (2014). The charters of William II and Henry I. Online: Accessed October 2022; `actswilliam2henry1.wordpress.com/the-charters/`.

Shizuka, D. and D. B. McDonald (2012). A social network perspective on measurements of dominance hierarchies. *Animal Behaviour 83*(4), 925–934.

Sivula, T., M. Magnusson, and A. Nehtari (2022). Unbiased estimator for the variance of the leave-one-out cross-validation estimator for a Bayesian normal model with fixed variance. *Communications in Statistics - Theory and Methods*, 1–23.

Sørensen, Ø., M. Crispino, Q. Liu, and V. Vitelli (2020). BayesMallows: An R Package for the Bayesian Mallows Model. *The R Journal 12*(1), 324–342.

Tkachenko, M. and H. W. Lauw (2016). Plackett-Luce regression mixture model for heterogeneous rankings. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pp. 237–246.

Valdes, J. (1978). *Parsing Flowcharts and Series-Parallel Graphs.* Ph. D. thesis, Stanford, CA, USA. AAI7905944.

Valdes, J., R. E. Tarjan, and E. L. Lawler (1982). The recognition of series parallel digraphs. *SIAM Journal on Computing 11*(2), 298–313.

van Wietmarschen, H. (2022). What is social hierarchy? *Noûs 56*(4), 920–939. (page 926, *"In sum,..."*).

Vasconcelos, M. (2008). Transitive inference in non-human animals: An empirical and theoretical analysis. *Behavioural Processes 78*(3), 313–334.

Vehtari, A., A. Gelman, and J. Gabry (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing 27*(5), 1413–1432.

Vitelli, V., Øystein Sørensen, M. Crispino, A. Frigessi, and E. Arjas (2018). Probabilistic preference learning with the Mallows rank model. *Journal of Machine Learning Research 18*(158), 1–49.

Wells, M. B. (1971). Elements of combinatorial computing.

Winkler, P. (1985). Random orders. *Order 1*(4), 317–331.