Bayesian Inference for Vertex-Series-Parallel Partial Orders

Chuxuan (Jessie) Jiang[1] **Geoff K. Nicholls**[1] **Jeong Eun Lee**[2]

[1]Department of Statistics, University of Oxford, United Kingdom
[2]Department of Statistics, University of Auckland, New Zealand

Abstract

Partial orders are a natural model for the social hierarchies that may constrain "queue-like" rank-order data. However, the computational cost of counting the linear extensions of a general partial order on a ground set with more than a few tens of elements is prohibitive. Vertex-series-parallel partial orders (VSPs) are a subclass of partial orders which admit rapid counting and represent the sorts of relations we expect to see in a social hierarchy. However, no Bayesian analysis of VSPs has been given to date. We construct a marginally consistent family of priors over VSPs with a parameter controlling the prior distribution over VSP depth. The prior for VSPs is given in closed form. We extend an existing observation model for queue-like rank-order data to represent noise in our data and carry out Bayesian inference on "Royal Acta" data and Formula 1 race data. Model comparison shows our model is a better fit to the data than Plackett-Luce mixtures, Mallows mixtures, and "bucket order" models and competitive with more complex models fitting general partial orders.

1 INTRODUCTION

Rank-order data are lists in which a set of elements are ranked. They are analysed in a wide range of areas, including decision support [Beichl et al., 2017], medical research [Beerenwinkel et al., 2007] and chemistry [Pavan and Todeschini, 2008]. We classify ranking methods into two categories - total order ranking and partial order ranking.

Total order models seek a ranking of the elements of the ground set (in our setting, the labels of a group of actors we want to rank) that is "central" to the rank-lists in the data. These models are suitable when we believe that an order relation exists between every pair of actors. The Mallows model [Mallows, 1957], the Plackett-Luce model [Plackett, 1975, Luce, 1959] and related mixture models are models for total orders. However, the real-world relations we are looking to recover may be weaker than a total order: perhaps relations between pairs of actors are not simply weak or uncertain, they don't actually exist. We expect this for some precedence relations that define some social hierarchies.

If we want to learn social-order relations between actors by observing their behavior, then the elements of the model we fit should correspond to elements of reality: if relations are incomplete then we should fit a *partial order*. A partial order $h = \{[n], \prec_h\}$ is a (possibly incomplete) set of binary order relations $\prec_h$ over a "ground set" of actors with labels $[n] = \{1, \dots, n\}$. Our data are records of queues of actors constrained by a social hierarchy h, which is unknown. If we see enough queue realisations we can identify the hierarchy. In this setting the queue is just a *linear extension* (LE) of h, that is, a permutation of actors in $[n]$ that doesn't put an actor ahead of someone of higher precedence.

Partial orders are widely used as a ranking summary tool, or to support efficient computation. For example, partial orders and LEs support efficient computation of marginals in Bayesian networks [Cano et al., 2011, Smail, 2018]. By contrast, in our work the partial order h is the object of inference, so it is a parameter in the likelihood: the data are LEs and the likelihood depends on the number of LEs of h. Counting LEs is an #P-complete task [Brightwell and Winkler, 1991], so work to date in this setting either restricts the class of partial orders [Mannila and Meek, 2000, Gionis et al., 2006, Mannila, 2008] to orders which admit fast counting or works with orders of manageable size [Beerenwinkel et al., 2007, Sakoparnig and Beerenwinkel, 2012, Nicholls and Muir Watt, 2011, Nicholls et al., 2022]. This approach does not scale well with n. We follow Mannila and Meek [2000] and work with *vertex-series-parallel* partial orders (VSPs). These orders are a sub-class of partial orders which can be formed by repeated series and parallel operations

Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023), PMLR 216:995–1004.

on smaller VSPs. They include *bucket orders* [1] as a special case. Valdes et al. [1979] represent VSPs using binary decomposition trees (BDTs). These support counting in a time linear in n [Wells, 1971] and scale to VSPs with hundreds of actors.

VSPs are a well characterised combinatorial class [Wells, 1971, Valdes et al., 1979]. However, work on fitting VSPs to data is limited. Mannila and Meek [2000] learn VSPs from LEs by adapting a greedy search over VSPs. However, there is to date no Bayesian inference and hence no one has given a prior probability distributions over VSPs with good properties for inference. Mannila [2008] gave Bayesian inference for bucket orders, a subclass of VSP, and Sakoparnig and Beerenwinkel [2012] for partial orders, a super-set that doesn't scale.

Contributions. This is the first Bayesian inference for VSPs from LEs and presents some useful new priors and likelihoods. VSPs are equivalent to "transitively closed" Directed Acyclic Graphs (DAGs); when we specify priors over objects of this sort we have to be careful to ensure the prior doesn't impose unwanted weighting and inconsistency.

We specify a prior and give its probability mass function in a simple closed form. Our prior (Sec. 2) is marginally consistent. This property (defined in Definition 1 below) is needed for the model to make sense in our setting. Our prior also represents the information available well: it is non-informative with respect to VSP depth, one of the most interesting summary statistics for a social hierarchy.

Our new observation model (Sec. 3) generalises earlier models for observation noise in records for queue-like data and has a natural physical interpretation in terms of "queue jumping" and "arriving late".

We give MCMC algorithms in Appendix C which target the VSP posterior. We carry out model comparison with the Plackett-Luce and Mallows mixture models in Appendix E.1. We further compare our model with a simple restriction to Bucket-Order models in Appendix E.2 and we compare it with a more general partial order model [Nicholls et al., 2022] in Appendix E.3.

Finally, our reconstruction of relations between witnesses appearing in Royal Acta (Sec. 5.2) is new. Historians are interested in these relations, but it wasn't possible to reconstruct them all till now as the partial orders were too big to count their LEs (Nicholls et al. [2022] analyse a subset, working in a time-series setting; we give timing comparisons in Appendix F). Our models are relevant for any ranking problem where relations may be partial: in Appendix D.2 we fit Formula 1 race results for the 2021 season. These data show the same preference for our model over other models.

[1] Actors are grouped in buckets - every actor is ordered with respect to actors in other groups, and any pair of actors in the same group are incomparible.

1.1 BACKGROUND

A partial order $h = \{V, \prec_h\}$ is a binary relation[2] $\prec_h$ over a "ground set" of actors V. In our setting the actor labels are $V = [n]$ where $[n] = \{1, 2, ..., n\}$ or some subset. Two actors $i, j \in [n]$ are *incomparable* $i \|_h j$, if neither $i \prec_h j$ nor $i \succ_h j$. Partial orders on $[n]$ are in one-to-one correspondence with transitively-closed DAGs $([n], E)$ with edges $E = \{\langle i, j\rangle \in [n] \times [n] : i \succ_h j\}$. Denote by $\mathcal{H}_{[n]}$ the set of all partial orders on actor labels $[n]$. Let $\mathcal{P}_{[n]}$ be the set of all permutations of $[n]$. A linear extension $l_h \in \mathcal{P}_{[n]}$ is a permutation of actors in $[n]$ that does not violate partial order h. See Fig. 1 for an example partial order[3] and its LEs. We denote the set of all LEs for partial order h as $\mathcal{L}[h]$. A *sub-order* $h[o] = (o, \prec_h)$ of a partial order $h \in \mathcal{H}_{[n]}$ restricts h to a subset $o \subseteq [n],\ o = \{o_1, ..., o_m\}$: all order relations in h are inherited by $h[o]$ so its DAG representation $(o, E[o])$ has edges $E[o] = \{e \in E : e \in o \times o\}$; directed edges incident vertices in $[n] \setminus o$ are removed and all others remain. A *chain* of $h \in \mathcal{H}_{[n]}$ is a sub-order $h[o]$ that is also a total order. The *length* of a chain is the number of nodes $|o|$ in the sub-order. The *depth* $D(h)$ of a partial order is the length of its longest chain, with $1 \leq D(h) \leq n$.

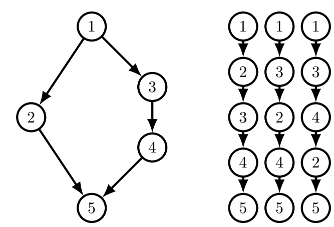

Figure 1: (left) A partial order with 5 actors and depth 4 which is also a VSP, v_0 say, and (right) its three LEs.

The *vertex-series-parallel partial orders (VSP)* on $[n]$ are a class of partial orders $\mathcal{V}_{[n]} \subset \mathcal{H}_{[n]}$ formed by repeated *series* $\otimes$ and *parallel* $\oplus$ operations. For partial orders h_1 and h_2, let $V(h_1)$ and $V(h_2)$ represent the ground sets of actors for h_1 and h_2 respectively (which we assume are disjoint).

- A *series partial order*, $h = h_1 \otimes h_2$, is the union of all relations in h_1 and h_2, with additional relations $i \succ_h j$ if $i \in V(h_1)$ and $j \in V(h_2)$.
- A *parallel partial order*, $h = h_1 \oplus h_2$, is the union of all relations in h_1 and h_2 with incomparability $i \|_h j$ if $i \in V(h_1)$ and $j \in V(h_2)$.

The set of VSPs $\mathcal{V}_{[n]}$ is defined recursively: if $|V(h)| = 1$ then h is a VSP; if h_1 and h_2 are VSPs then $h_1 \otimes h_2$ and $h_1 \oplus h_2$ are VSPs. Valdes et al. [1979] show that a partial order is a VSP if it does not contain the "forbidden subgraph" (Appendix G, Fig. G.1) as a subgraph isomorphism.

[2] The binary relation $\prec_h$ is both irreflexive (the relation $i \prec_h i$ does not exist) and transitive (if $i \prec_h j$ and $j \prec_h k$, then $i \prec_h k$), where $i, j, k \in [n]$ and $i \neq j \neq k$.

[3] In this article, we visualise a partial order via its transitive reduction - this omits all edges implied by transitivity and is unique.

The partial order v_0 in Fig. 1 is a VSP. It can be constructed using the series and parallel operations in Fig. 2.

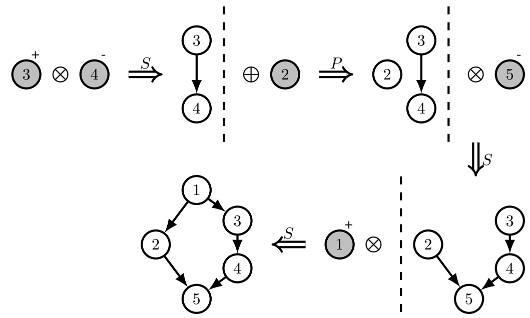

Figure 2: One possible construction procedure for the VSP v_0 shown in Fig. 1.

A VSP on n actors can be parameterised as a Binary Decomposition Tree (BDT) Valdes et al. [1979] - a binary tree $t \in \mathcal{T}_{[n]}$ with n leaves in which nodes have additional attributes (listed below) and edges are directed from the root to the leaves. Let $\mathcal{F}$ and $\mathcal{A}$ be the index sets for the n leaves and $n-1$ internal nodes respectively, with $\mathcal{F} \cup \mathcal{A} = [2n-1]$. Each leaf node index corresponds to a unique actor in the VSP. It is convenient to distinguish between leaf nodes indices and the actor labels to which they correspond. For each leaf node $i \in \mathcal{F}$, let $F_i(t) \in [n]$ give the actor label for the actor corresponding to that leaf node. Internal nodes $i \in \mathcal{A}$ are S nodes if the subtrees rooted by their child nodes are merged in series, otherwise they are P nodes and the subtrees are merged in parallel. Internal nodes with an S label have an additional attribute indicating which of its child nodes is the "upper child": the subtree of this child node (indicated by a '+' and a red edge in Fig. 3) is stacked above the subtree rooted by the other child node (indicated by a '-'). As an example, the VSP v_0 in Fig. 1 can be represented by the BDT t_0 in Fig. 3. Let $S(t) \in [n-1]$ be the number of S-nodes in tree t.

A tree t with edge set $E(t)$ is written $t = (F(t), E(t), L(t))$. Here $L(t) = \{L_i\}_{i \in \mathcal{A}}$ with $L_i(t) = (j, j')$ indicating that internal node i is an S-node with child nodes j, j' and the subtree rooted by j is stacked above that rooted by j', and $L_i(t) = \emptyset$ if i is a P-node. The map from a BDT to the VSP $v : \mathcal{T}_{[n]} \to \mathcal{V}_{[n]}$ is not bijective: for a VSP $v \in \mathcal{V}_{[n]}$, there may exist many BDTs $t \in \mathcal{T}_{[n]}$ which represent it. Let $t(v) = \{t \in \mathcal{T}_{[n]} : v(t) = v\}$ give the set of BDTs representing VSP $v \in \mathcal{V}_{[n]}$.

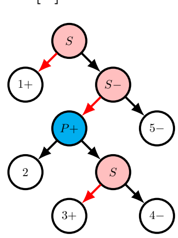

Figure 3: A BDT t_0 representing v_0 in Fig. 1, so that $v(t_0) = v_0$. Red edges and '+' signs indicate the upper child.

Brightwell and Winkler [1991] show that counting the number of LEs of a partial order is a #P-complete problem. However, the subclass of VSP partial orders admits fast counting. Wells [1971] gives

$$|\mathcal{L}(h_1 \otimes h_2)| = |\mathcal{L}(h_1)||\mathcal{L}(h_2)| \tag{1}$$

$$|\mathcal{L}(h_1 \oplus h_2)| = |\mathcal{L}(h_1)||\mathcal{L}(h_2)| \binom{|V(h_1)| + |V(h_2)|}{|V(h_1)|} \tag{2}$$

where $|V(h_1)|$ and $|V(h_2)|$ give the number of actors in h_1 and h_2. This may be evaluated recursively in $O(n)$ steps.

In the following we make use of one more representation of a VSP: the *Multi-Decomposition Tree* (MDT). These trees are obtained by collapsing edges which connect internal nodes of the same S/P-type in the BDT, as in Fig. 4. Let $\mathcal{M}_{[n]}$ be the set of all MDTs with n distinguisable leaves. A formal definition is given in Appendix A.3.

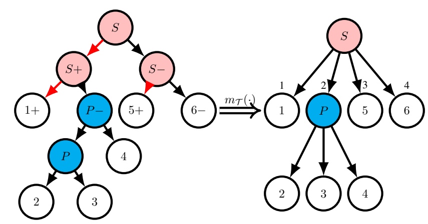

Figure 4: An example BDT t_1 (left) and its corresponding MDT m_1 (right). The child-nodes of any S-node in the MDT are numbered to give the order in which their subtrees are stacked by the BDT.

Valdes [1978] has shown that MDTs are one-to-one with VSPs, so all the BDTs in $t(v)$ representing the VSP v must "collapse down" to give the same MDT. For $m \in \mathcal{M}_{[n]}$ we write $v = v(m)$ for the map to VSPs (relations between any pair of actors in the VSP are simply given by the type of their Most Recent Common Ancestor (MRCA) in m). Let $m_{\mathcal{V}}(v) = \{m \in \mathcal{M}_{[n]} : v(m) = v\}$ be the set of MDTs representing $v \in \mathcal{V}_{[n]}$.

Lemma 1 *The map $m_{\mathcal{V}} : \mathcal{V}_{[n]} \to \mathcal{M}_{[n]}$ is bijective (so that $|m_{\mathcal{V}}(v)| = 1$). See Valdes [1978] for proof and Valdes et al. [1979] for further discussion.*

2 VSP PRIOR

In this section we give a marginally consistent prior $\pi_{\mathcal{V}_{[n]}}(v|q)$ over VSPs on actors in $[n]$, controling the distribution over VSP-depth. We begin by defining a prior probability distribution $\pi_{\mathcal{T}_{[n]}}(t|q)$ over BDTs $t \in \mathcal{T}_{[n]}$.

Our prior on $\mathcal{T}_{[n]}$ has a uniform distribution over trees $([2n-1], E(t))$ with distinguishable leaves. Internal nodes are labelled S with probability q and otherwise P. We choose an "upper child" for each S node at random from its two child nodes, so we have

$$\pi_{\mathcal{T}_{[n]}}(t|q) = \frac{1}{|\mathcal{T}_{[n]}|}\left(\frac{q}{2}\right)^{S(t)}(1-q)^{n-S(t)-1}, \qquad (3)$$

where $S(t)$ is the number of S-nodes, $|\mathcal{T}_{[n]}| = (2n-3)!! \equiv (2n-3)\cdot(2n-5)...3\cdot 1$ is the number of binary tree topologies with n distinguishable leaves, and the types of the $n-1$ internal nodes are independent with a factor $2^{-S(t)}$ for the stacking order of the children of S-nodes.

We get the prior on VSPs $v \in \mathcal{V}_{[n]}$ by summing over all BDTs that represent v,

$$\pi_{\mathcal{V}_{[n]}}(v|q) = \sum_{t\in t(v)} \pi_{\mathcal{T}_{[n]}}(t|q) \qquad (4)$$

This simple choice, based on a uniform distribution over tree topologies, determines a prior for VSPs that represents the prior knowledge we want to impose in our setting. If a social hierarchy is built up by making comparisons between groups of people, based for example on their profession, then it will be a VSP. Secondly, the unknown true depth of the social hierarchy we are trying to reconstruct (which is the length of the longest chain in the VSP) is a feature of particular interest, so we don't want the prior to strongly inform depth. We choose a prior distribution over q so that the marginal distribution $\pi_{\mathcal{V}_{[n]}}(v)$ gives a reasonably flat prior distribution for depth $D(v)$ (see Appendix H and Fig.H.1).

We assume that relations between two actors are determined by (unknown) properties intrinsic to those actors (for example, their professions, or ancestry). If that is true then the presence or absence of a third actor should not affect the relations between the first two. It is not straightforward to get this property *and* transitivity. If two actors $1\|2$ are unordered and we add actor 3 with relations $1 \succ 3$ and $3 \succ 2$ then $1 \succ 2$ by transitivity: the presence of actor 3 changes the relation between actors 1 and 2. Random VSPs can be built up in many different ways (that is, they are represented by many different BDTs), so we want the prior probability that $1 \succ_w 2$ in a random VSP $w \sim \pi_{\mathcal{V}_{[2]}}$ to be the same as the prior probability that $1 \succ_v 2$ in a random VSP $v \sim \pi_{\mathcal{V}_{[3]}}$. This adds a consistency restriction on any family of prior distributions $\pi_{\mathcal{V}_{[n]}},\ n \geq 1$ we write down.

A family of priors like $\pi_{\mathcal{T}_{[n]}}(t|q)$ or $\pi_{\mathcal{V}_{[n]}}(v|q),\ n \geq 1$ is *marginally consistent* (also known as *projective*) if every marginal of every distribution in the family is also in the family. Marginal consistency is not a property we get for free from the axioms of probability: the uniform distribution on partial orders $h \sim \mathcal{U}(\mathcal{H}_{[n]})$ is not consistent: there are 3 partial orders on the labels $\{1,2\}$ and 19 on $\{1,2,3\}$; since 19 is not divisible by 3, the probability for $1 \succ_h 2$ in $h \sim \mathcal{U}(\mathcal{H}_{[2]})$ cannot equal the marginal probability for $1 \succ_g 2$ in $g \sim \mathcal{U}(\mathcal{H}_{[3]})$.

Definition 1 (Marginal consistency) *Let $\mathcal{O}_{[n]} = \{o \subseteq [n] : |o| > 0\}$ be the set of all subsets of $[n]$ with at least one element. The family of VSP priors $\pi_{\mathcal{V}_o}(v|q),\ o \in \mathcal{O}_{[n]},\ n \geq 1$ is marginally consistent if, for all $n \geq 1$ and all $o, \tilde{o} \in \mathcal{O}_{[n]}$ with $o \subseteq \tilde{o}$, distributions in the family satisfy*

$$\pi_{\mathcal{V}_o}(w|q) = \sum_{\substack{v\in\mathcal{V}_{\tilde{o}}\\ v[o]=w}} \pi_{\mathcal{V}_{\tilde{o}}}(v|q) \quad \text{for all } w \in \mathcal{V}_o. \qquad (5)$$

If marginal consistency holds for all q then it holds for marginals $\pi_o(w)$ by taking expectations over q in (5).

The following Theorem is our first main result: we give a closed form expression for the prior for a VSP (we calculate the sum in (4)) and show that the family of priors is marginally consistent. For $v \in \mathcal{V}_{[n]}$, let $t \in t(v)$ be some tree representing v. Partition the internal nodes $\mathcal{A}$ of t into S-clusters $C_k^{(S)},\ k = 1, ..., K_S$ and P-clusters $C_k^{(P)}, k = 1, ..., K_P$. An S-cluster is a maximal set of internal nodes of type S which are connected by edges in $E(t)$ and corresponds to a node in the MDT-representation. The P-clusters are defined similarly. We will see (in Appendix A.2, proof of Proposition 5) that two BDTs representing the same VSP have the same numbers of S and P clusters, with the same sizes.

Theorem 1 *The family, $\pi_{\mathcal{V}_o}(v|q),\ o \in \mathcal{O}_{[n]}\ n \geq 1$, of VSP priors is marginally consistent. The probability distribution over VSPs $v \in \mathcal{V}_{[n]}$ in (4) is*

$$\pi_{\mathcal{V}_{[n]}}(v|q) = \pi_{\mathcal{T}_{[n]}}(t|q)\prod_{k=1}^{K_P}(2|C_k^{(P)}|-1)!!\prod_{k'=1}^{K^S}\mathcal{C}_{|C_{k'}^{(S)}|} \qquad (6)$$

where t may be taken to be any tree $t \in t(v)$ with P- and S-clusters defined above, $\pi_{\mathcal{T}_{[n]}}(t|q)$ is given in (3) and

$$\mathcal{C}_s = \frac{1}{s+1}\binom{2s}{s}, \quad s \geq 0 \qquad (7)$$

is the s'th Catalan number [Stanley and Weisstein, 2002].

Proof 1 (Theorem 1) *The proof of Theorem 1 is given in two parts in Appendix A. In Proposition 3 in Appendix A.1 we show that the family of tree-priors $\pi_{\mathcal{T}_{[n]}}(t|q),\ o \in \mathcal{O}_{[n]},\ n \geq 1$ is marginally consistent. This result is used in Proposition 4 in A.1 to show that VSPs are marginally consistent - the first part of Theorem 1.*

The proof of the second part is given in Appendix A.2. We show in Proposition 5 that all trees $t \in t(v)$ have equal values of $\pi_{\mathcal{T}_{[n]}}(t|q)$, so that $\pi_{\mathcal{V}_{[n]}}(v|q) = |t(v)|\pi_{\mathcal{T}_{[n]}}(t|q)$ for any $t \in t(v)$. This is straightforward, as they must all collapse down to the same MDT. Finally, in Proposition 6, we give a formula for $|t(v)|$. We count the number of BDTs that collapse down to a given MDT. Any P-cluster C_k^P of a BDT corresponds to a P-node in its MDT and covers a small sub-tree of the BDT representing an empty partial order on *its $|C_k^P|+1$ labeled leaves. It can be replaced in the BDT by any sub-tree representing the empty partial order without changing the MDT, and there are $(2|C_k^{(P)}|-1)!!$ such trees. Similarly, any S-cluster C_k^S corresponds to a S-node in the MDT and covers a sub-tree of the BDT representing a total order on its leaves. It can be replaced in the BDT by any sub-tree representing the same total order. The Catalan numbers enter because $\mathcal{C}_{s-1}$ gives the number of BDTs representing a total order on s elements (see proof Proposition 6). This last result is new, gives (6) and completes the proof of Theorem 1.*

Theorem 1 gives the prior for a VSP in terms of the prior for one of the BDTs that represent that VSP. We can also parameterise VSPs using MDTs and this leads to the second MCMC scheme given in Appendix C.2.

Corollary 1 *For $m \in \mathcal{M}_{[n]}$ with internal nodes $\mathcal{A}$, let c_i give the number of children of node $i \in A$ and let $P(m) = \{i \in \mathcal{A} : L_i(m) = \emptyset\}$ and $S(m) = \mathcal{A} \setminus P(m)$ give the sets of $P-$ and $S-$node labels. The prior for VSPs given in (6) is equivalently a prior for MDTs,*

$$\pi_{\mathcal{V}_{[n]}}(v(m)|q) = \pi_{\mathcal{M}_{[n]}}(m|q) \tag{8}$$
$$= \frac{1}{(2n-3)!!} \prod_{i \in P(m)} (1-q)^{c_i-1}(2c_i-3)!!$$
$$\times \prod_{j \in S(m)} \left(\frac{q}{2}\right)^{c_j-1} \mathcal{C}_{c_j-1}.$$

Proof 2 (Corollary 1) *Substitute (3) into (6) and note a tree with c_i leaves has $c_i - 1$ internal nodes. This result gives a convenient representation for prior evaluation.*

3 BI-DIRECTIONAL QUEUE-JUMPING OBSERVATION MODEL

Our data is a collection of N lists. For $j \in [N]$ let $o_j \subseteq [n],\ o_j = \{o_1, ..., o_{n_j}\}$ be the actors present when the j'th ranking list was observed and let $y_j \in \mathcal{P}_{o_j},\ y_j = (y_{j,1}, ..., y_{j,n_j})$ be the observed list, just an ordered version of o_j. Let $y = (y_1, ..., y_N)$ be the list of lists. The 'queue-based' observation model given in Nicholls and Muir Watt [2011] models list data as a realisation of a random queue constrained to put higher status individuals before those of lower status. In this model the queue is dynamic. It forms and then unconstrained pairs of actors swap places at random. If this process reaches equilibrium before the queue is read off then the resulting list is a uniform draw from the linear extensions of the constraining social hierarchy [Karzanov and Khachiyan, 1991]. In this noise-free model $y_j \sim \mathcal{U}(\mathcal{L}[v[o_j]])$ independently for $j \in [N]$.

It is unlikely the observations are "error free". In a "queue-jumping" model (QJ-U, see Appendix B.1and Nicholls and Muir Watt [2011] for details) the queue is read from the top: with probability $p \in [0, 1]$ the "next" person in the queue is drawn at random from those remaining, ignoring the social hierarchy; otherwise they are the first person in the remaining LE. The queue can also be read from the bottom up. In this model (QJ-D) actors fall down the queue. We think of these events as actors arriving while the queue is being read.

We would like to have a queue-based model in which displacement in both directions is possible. The resulting "bi-directional queue-jumping" model (QJ-B) is not simply a mixture of QJ-U and QJ-D, as it allows displacement in both directions within a single realisation. The cost of evaluating a QJ-B likelihood is exponential in n. However, for the application in Section 5.1 there is a subset of actors (bishops) known a priori to appear as a group. Separate modelling of this manageable subset ($n \simeq 20$) is well-motivated. Although QJ-B cannot be evaluated for a general partial order (counting LEs is prohibitive) it is fine for a VSP.

Like QJ-U, QJ-B ranks by repeated selection. Fig. 5 provides an example QJ-B list-realisation from VSP v_0. A generic list $x \in \mathcal{P}_{[n]}$ is built up from both ends (see Appendix B.2). Let $z \in \{0, 1\}^{n-1}$ with $z_k \sim Bern(\phi)$. Here $z_k = 0$ indicates the k'th actor to be added to the list was placed bottom-up in the QJ-D model and $z_k = 1$ indicates they were placed top-down in the QJ-U model. In Fig. 5, $z = (1, 0, 0, 1)$. If we let $U_0 = 0$ then $U_k = U_{k-1} + z_k$ gives the number of places filled from the top after the k'th actor has arrived, so if $z_k = 1$ then the k'th actor was placed into position $i_k = U_k$ in x. Similarly, if $D_0 = n + 1$ then $D_k = D_{k-1} - (1 - z_k)$ tracks places filled from the bottom and gives the placement index $i_k = D_k$ in x when $z_k = 0$, so $i_k = z_k U_k + (1 - z_k)D_k$ gives the position in x into which the k'th actor was added. If $z = (1, 0, 0, 1)$, then $(i_1, ..., i_4) = (1, 5, 4, 2)$ (and $i_5 = 3$, the only remaining place).

Definition 2 (Bi-Directional Queue-Jumping Model) *Let $L_T(v) = |\mathcal{L}[v]|$ be the number of LEs of VSP $v \in \mathcal{V}_{[n]}$ and for $i \in [n]$ let $T_i(v) = |\{l \in \mathcal{L}[v] : l_1 = i\}|$ give the number of LEs with actor i at the top. Let $B_i(v) = |\{l \in \mathcal{L}[v] : l_n = i\}|$ give the number of LEs with actor i at the bottom. If $z \in \{0, 1\}^{n-1}$ is given then $i_k = i_k(z), k = 1, ..., n$ is given above. The observation model for QJ-B for a list $x \in \mathcal{P}_{[n]}$ given z is*

$$Q_{bi}(x|z, v, p, \phi) = \prod_{k=1}^{n-1} [\phi \mathbb{1}_{\{z_k=0\}} Q_{bi}(x_{i_k}|x_{i_{1:k-1}}, z_k, v, p)$$
$$+ \quad (1-\phi)\mathbb{1}_{\{z_k=1\}} Q_{bi}(x_{i_k}|x_{i_{1:k-1}}, z_k, v, p)],$$

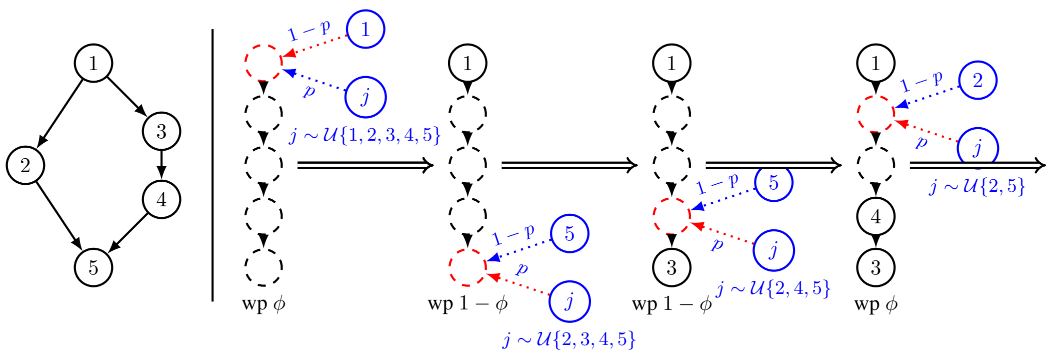

Figure 5: One example list simulation process from the VSP v_0 (left) via the QJ-B observation model. The simulated list is displayed on the right.

where

$$Q_{bi}(x_{i_k}|x_{i_{1:k-1}}, z_k = 0, v, p) = \frac{p}{n-k+1} + (1-p)\frac{T_{x_{i_k}}(v[x_{[n]\setminus\{i_1,\dots,i_{k-1}\}}])}{L_T(v[x_{[n]\setminus\{i_1,\dots,i_{k-1}\}}])},$$

$$Q_{bi}(x_{i_k}|x_{i_1:k-1}, z_k = 1, v, p) = \frac{p}{n-k+1} + (1-p)\frac{B_{x_{i_k}}(v[x_{[n]\setminus\{i_1,\dots,i_{k-1}\}}])}{L_T(v[x_{[n]\setminus\{i_1,\dots,i_{k-1}\}}])},$$

and marginally,

$$Q_{bi}(x|v, p, \phi) = \sum_{z\in\{0,1\}^n} Q_{bi}(x|z, v, p, \phi)p(z|\phi) \quad (9)$$

where $p(z|\phi) = \phi^{\sum_i z_i}(1-\phi)^{n-\sum_i z_i}$.

We give a generative model realising $x \sim Q_{bi}$ in Appendix B.2. This distribution reduces to Q_{up}/QJ-U in Appendix B.1 when $\phi = 1$ (and Q_{down}/QJ-D when $\phi = 0$). We use this nesting to investigate whether QJ-U or QJ-D or QJ-B fits the data better. This is of interest in our application as different error types correspond to obvious physical mechanisms (downwards displacement may be "arrived late" and upwards displacement may be "my friend the King is present"). When $p = 0$ this is the noise free model for every $\phi \in [0, 1]$, so ϕ is not identifiable in the noise free setting.

Counting LEs of a VSP (evaluating $L_T(v)$ etc) is $O(n)$ so the computational complexity for naive evaluation of Q_{bi} using (9) is $O(n^2 2^n)$. We used a recursion (Algorithm B.3) of computational complexity $O(n2^n)$. This avoids repeated evaluation of LE-counts for the same suborders and (by Proposition 7 in Appendix B.3) evaluates Q_{bi}.

4 SUMMARISING THE VSP POSTERIOR

Bayesian inference is straightforward in principle given an explicit prior distribution over VSPs and an observation model $Q = Q_{up}$ or $Q = Q_{bi}$ for our N ranking-lists. We can represent a VSP as a MDT (since the mapping is one-to-one) or carry out Bayesian inference on the latent space of BDTs $t \in \mathcal{T}_{[n]}$ and use the fact that they marginalise to MDTs. We present the posteriors for BDT and VSP. Let $\psi = (p, \phi)$ for QJ-B and $\psi = p$ for QJ-U.

The posterior for the BDT $t \in \mathcal{T}_{[n]}$ is

$$\pi_{\mathcal{T}_{[n]}}(t, q, \psi|y) \propto \pi_{\mathcal{T}_{[n]}}(t|q)\pi(q, \psi)Q(y|v(t), \psi) \quad (10)$$

The posterior distribution for the VSP $v \in \mathcal{V}_{[n]}$ is

$$\pi_{\mathcal{V}_{[n]}}(v, q, \psi|y) \propto \pi_{\mathcal{V}_{[n]}}(v|q)\pi(q, \psi)Q(y|v, \psi), \quad (11)$$

where we use the equivalent MDT posterior with prior given in Corollary 1 for VSPs in (11).

Proposition 1 (Posterior Marginals) *Sampling the BDT posterior* $(t, q, \psi) \sim \pi_{\mathcal{T}_{[n]}}(\cdot|y)$ *gives samples* $(v(t), q, \psi) \sim \pi_{\mathcal{V}_{[n]}}(\cdot|y)$ *from the VSP posterior (see Appendix A.4 for proof).*

We implemented separate MCMC samplers targeting both (10) and (11). Our MCMC algorithms are given in Appendix C. We checked that the VSP-posterior marginals for the two implementations were equal (up to Monte-Carlo error). We implemented MCMC targeting the BDT posterior (10) first, as BDT data structures are slightly more straightforward to handle than the MDT data structures needed to target the VSP posterior in (11). All results in the next section were computed using the BDT-MCMC.

5 APPLICATIONS

5.1 DATA AND ANALYSES

We analyse a dataset accessed through a database made for "The Charters of William II and Henry I" project by Professor Richard Sharpe and Dr Nicholas Karn [Sharpe et al., 2014]. These data collect witness lists from legal documents from England and Wales in the eleventh and twelfth century. Witness lists respect a rigid social hierarchy: higher status individuals come ahead of lower status individuals in the lists. Fig. D.1 is an example list.

We represent the hierarchy on actors $[n]$ appearing in the lists as a partial order which is a VSP $v \in \mathcal{V}_{[n]}$ and model a list as the outcome of one of the queuing processes described in Section 3. We imagine the actors lining up to witness the document in a virtual queue.

Lists are witnessed by people from all walks of life and we have their titles. These include "others" (actors who lack titles). Historians are interested in social hierarchies and how they change over time. For illustration we reconstruct hierarchies in three snapshots: the years 1080-84, 1126-30 and 1134-38. The last two cover periods shortly before and after Stephen became King, a time of great change. The 5-year intervals are short enough for any changes in the hierarchy to be slight [Nicholls et al., 2022]. For ease of visualisation we present results for individuals appearing in at least 5 lists (5LPA data) here and results on all actors (1LPA data) in Appendix D.1.1. We fit VSP/QJ-U to all data and fit VSP/QJ-B to 2 of the 3 5LPA data sets (not 1134-38, as QJ-B has runtime growing exponentially with the length of the longest list). However, relations between bishops in 1134-38 are of particular interest so we present VSP/QJ-B results for this subgroup. Table D.1 summarises the data in the different experiments on the Royal Acta data.

In a separate analysis illustrating how our methods apply more generally to any rank-order data, we give an analysis of Formula 1 race outcomes for the 2021 season. Data and results are given in Appendix D.2.

The prior for error probability p and for q (probability for an S-node) is given in Fig. 8. All fitting is done using MCMC in the BDT representation, Algorithm C.1. For any given model we draw MCMC samples $t^{(k)}, p^{(k)}, q^{(k)}, \phi^{(k)} \sim \pi_{\mathcal{T}_{[n]}}(\cdot|y)$ for $k = 1, ..., K$ and set $v^{(k)} = v(t^{(k)})$ per Proposition 1. Example MCMC traces are given in the supplement with Effective Sample Size (ESS) values (Appendix D.1). Sampled VSPs are summarised using consensus VSPs: $V^{con}(\epsilon)$ includes order relation/edge $\langle i, j\rangle$ if the relation appears more than ϵK times in the MCMC output. We color edges black if they are in $V^{con}(\epsilon)$ at $\epsilon = 0.5$ but not $\epsilon = 0.9$ and red if they are supported at $\epsilon = 0.9$. We plot transitive reductions. These omit strongly supported edges from the top to the bottom of the DAG for clarity.

In Sec. 5.2, we fit the QJ-U and QJ-B models to the 5LPA data and make a model comparison using Bayes factors. Consensus orders for the 1LPA data are given in Appendix D.1.1. We additionally compare these models with bucket order models, a Plackett-Luce mixture, Mallows mixture and latent partial order model in Appendix E. We carry out these tests on both the Royal Acta data and the F1 race result data. We report computing time measurements for counting LEs for the latent partial order model and the VSP. They are compared empirically in Appendix F.

5.2 RESULTS

We begin by making reconstruction-accuracy tests on synthetic data. Our list data are incomplete, in the sense that the membership in list $i = 1, ..., N$ is o_i not $[n]$ and the N-values in Table D.1 are not much larger than the number of actors n. In order to measure the reliability of the reconstructions which follow we take representative parameters (parameters sampled from the corresponding posterior, the last sampled state $v^{(K)}, p^{(K)}, q^{(K)}, \phi^{(K)}$) and generate synthetic data with the same list-membership and length structures as the real data. The ROC curves in Fig. D.12 (5LPA data and QJ-U) and D.15 (5LPA data and QJ-B) for consensus orders $V^{con}(\epsilon)$ show the proportion of inferred false-positive and true-positive relations increasing with decreasing ϵ from $(0, 0)$ at $\epsilon = 1$ (the consensus order is empty) to $(1, 1)$ at $\epsilon = 0$ (complete graph). For each simulated data set there is ϵ giving high true-positive and low false-positive reconstructed relation fractions: if our model is accurate then we reconstruct relations well.

We next report consensus partial orders. Consensus orders for actors color-coded by their professions are shown in Fig. 6 and 7. For both QJ-U and QJ-B models, we observe three clear social hierarchies: King $\succ$ Queen $\succ$ Duke appear at the top, in that order (when they are in the 5LPA data, in 1180-84 and 1134-38); then archbishop/prince $\succ$ bishops; the remaining professions (earl, count, chancellor, other) are ranked lower than bishops in a relatively complex hierarchy.

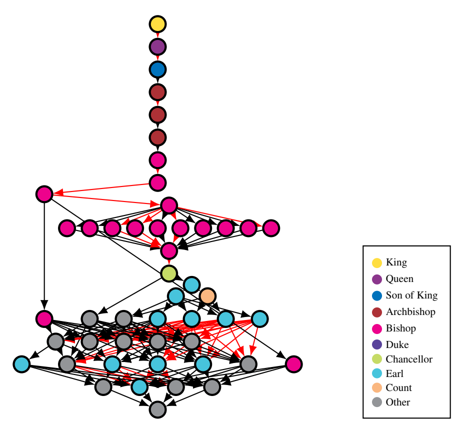

Figure 6: VSP/QJ-U model. Consensus order for 1134-38 5LPA data. Significant/strong order relations are indicated by black/red edges respectively.

Some of this is common sense. However, the web of strongly attested relations between earls and others in 1134-38 is new. There is clear evidence for hierarchies within professions. The bishop-only QJ-U analysis in 1134-38 (top-right graph in Fig. 7) is similar to the bishop subgraph of the full QJ-U

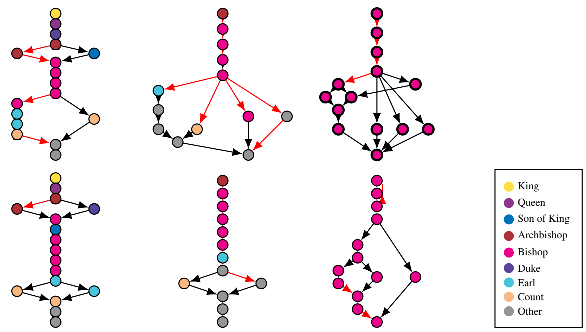

Figure 7: VSP/QJ-U (top row) and VSP/QJ-B (bottom row). Consensus orders for 1080-84, 1126-30 and 1134-38 (bishops) (left to right columns) 5LPA data.

analysis for the same period (pink nodes in Fig. 6). The prior is marginally consistent, but information is shared across lists so removing actors changes the data and changes estimated order relations between those that remain. However, the bishops appear as a group in the lists and in Fig. 6 and there are few non-bishops "between" bishops in lists, so this effect is slight. We can attach names to nodes: for example, the top three bishops in 34-38 (in Fig. 6 and in both QJ-U and QJ-B analyses in the rightmost column of Fig. 7) are Henry, de Blois, Bishop of Wincester $\succ$ Roger, Bishop of Salisbury $\succ$ Alexander, Bishop of Lincoln.

The status hierarchies fitted using by QJ-B (bottom row Fig. 7) are simpler and deeper than QJ-U (top row Fig. 7). The data must contain a small number of errors in both directions. A uni-directional model must fit a shallower VSP as it accommodates errors in the "wrong" direction by removing order relations in the reconstructed VSP.

We summarise the status of "professions" within VSPs by averaging ranks. Given a partial order $v \in \mathcal{V}_{[n]}$, the rank of actor $i \in [n]$ is the number of actors above them, $\mathrm{rank}_i(v) = 1 + |\{\langle e_1, e_2\rangle \in E(v) : e_2 = i\}|$, and take as our summary the average rank of actors in the profession. The posterior mean ranks given in Table D.5 and D.7 match our remarks on consensus orders.

We next report parameter distributions. Prior and posterior distributions for the probability q for a serial node, error probability p and QJ-B parameter ϕ (equal one for QJ-U and zero for QJ-D) for the three periods are shown in Fig. 8. The p-posteriors are weighted toward smaller values and overlap, though errors are low in 1126-30 and higher in 1180-84 indicating greater respect for the rules of precedence in 1126-30 than in 1180-84. Prior and posterior depth distributions are shown in Fig. D.11 and D.14. The prior depth distributions are fairly flat so any depth-structure in the posterior comes from the data. The probability for a series node in the BDT (q) controls the depth of the fitted order relation. For example, in 1180-84 a relatively high q for QJ-U is associated with relatively high depth VSPs with a mean depth of 14 relative to maximum depth 17 (the number of actors). In contrast, the posterior probabilities for S and P nodes are almost equal in 1134-38 and so we get a relatively shallower hierarchy: the posterior mean depth is about 23 relative to a maximum depth 49 in Fig. D.11.

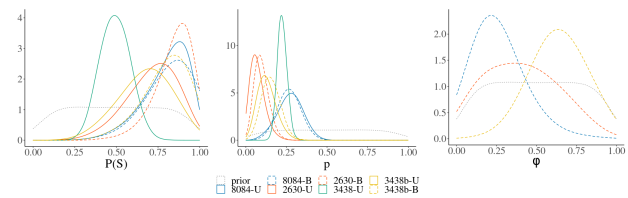

Figure 8: Posterior distributions for $q = P(S)$ (left), error probability p (middle) and QJ-B probability ϕ (right) for the time periods 1180-1184 (blue), 1126-1130 (red), 1134-1138 (green) and 1134-1138(b) (yellow) from both the VSP/QJ-U (solid) and VSP/QJ-B (dashed) models. The prior is represented in grey in all figures.

The QJ-B model for noise in the list data allows actors to jump up or down from a queue-position appropriate for their status. QJ-U is favored if $\phi > 1/2$ and otherwise QJ-D so we see from Fig. 8 that QJ-U is favored in 1134-38(b), while the 1080-84 data supports QJ-D. However, the p-posteriors both favor small p. The displacement direction controlled by ϕ is hard to measure and not identifiable at $p = 0$ so the ϕ-distributions are correspondingly broad.

We next report results of model selection between different queue jumping error models. Preference shifts from downwards to bidirectional to upwards displacement error models over the period 1080-1140. We justify this reading of the results using Bayes factors below. In summary, QJ-D is slightly favored over QJ-B (so we write "$D > B$") in 1080-84 while in 1126-30 models QJ-D and QJ-B are equally good ($D \approx B$). Both are clearly favored over QJ-U in these periods ($D, B \gg U$). In 34-38(b) we have $U \approx B$ and $U, B \gg D$.

We can read the Bayes factors we need off Fig. 8 because the models QJ-U and QJ-D are nested in the model QJ-B. The Bayes factor $B_{U,B}$ for QJ-U over QJ-B is

$$\begin{aligned} B_{U,B} &= \lim_{\delta \to 0} \frac{p(y|\phi > 1-\delta)}{p(y|\phi \in (0,1))} \\ &= \lim_{\delta \to 0} \frac{\pi(\phi > 1-\delta|y)}{\pi(\phi \in (0,1)|y)} \frac{\pi(\phi \in (0,1))}{\pi(\phi > 1-\delta)} \\ &= \lim_{\delta \to 0} \frac{\pi(\phi > 1-\delta|y)}{\pi(\phi > 1-\delta)}, \end{aligned}$$

since $\phi \in (0,1)$ with probability one. Similarly,

$$B_{D,B} = \lim_{\delta \to 0} \frac{\pi(\phi < \delta|y)}{\pi(\phi < \delta)},$$

and then $B_{U,D} = B_{U,B}/B_{D,B}$. From Fig. 8, $B_{U,B}$ is close to 0 in periods 1180-84 and 1126-30 as the posterior density is well below the prior density at $\phi \to 1$, providing strong support for QJ-B over QJ-U. In 1134-38(b), we see $B_{U,B} \approx 1$, as the curves meet as $\phi \to 1$ so there is no clear signal from the data. The other comparisons may be justified similarly.

Finally we make model comparisons with other models. Comparisons with a Plackett-Luce mixture model and a Mallows mixture model are given in Appendix E.1, the latent partial order model from Nicholls and Muir Watt [2011] in Appendix E.3 and a simple Bucket Order model in Appendix E.2. When models are nested (Bucket Order) we estimate a Bayes factor. When they are not, we use the Expected Log Pointwise Predictive Density (ELPD, Vehtari et al. [2017]) as our criterion. This is a predictive loss which can be estimated using LOOCV or the WAIC [Watanabe, 2013]. On this basis VSP/QJ (-U and -B) is clearly favoured over Placket-Luce mixture models and Mallows mixture model in Table E.1 ("Royal Acta") and E.3 (Formula 1 race data). With Bayes factors around 2 or 3, Bucket orders are equal or slightly preferred over VSPs in the QJ-B model (Table E.3). Our VSP-based model QJ-U is clearly preferred over Bucket orders in the QJ-U fit (some very large Bayes factors in favor of VSP).

The support of our VSP model is a subset of the PO support, as POs containing the forbidden sub-graph (Appendix G) are not VSPs. The PO/QJ-U has a slightly larger ELPD (-36.7, see Table E.4) than VSP/QJ-U (-37.8) on the 1126-1130 data with 5LPA. However, the difference is not significant at the precision (± 10) of these estimates so we conclude that VSP/QJ-U models these data as well as PO/QJ-U. It gives similar consensus orders (Fig. E.2) and profession rankings (Table E.5).

A VSP-based analysis is far more computationally efficient than a PO-based model when the number of actors is large. The computing time for counting the LEs of a VSP rises linearly with the number of actors (Fig. F.1) while it increases exponentially for PO (using the best code we could find, *LEcount*, Kangas et al. [2016], but inevitable given Brightwell and Winkler [1991]). We have to count LEs of random POs. In our experience counting LEs on random POs with up to about 25-30 actors is feasible. However, at larger numbers we encounter occasional random POs which are especially "hard" to count and VSP-based analysis is the only way forward at present.

6 DISCUSSION AND CONCLUSION

Our work was motivated by the need to fit relatively large partial orders (up to 200 nodes) to noisy linear-extension data. We saw that, for data on this scale, counting linear extensions in the VSP-tree representation is much faster than current state-of-art counting for general partial orders, enabling our methods to scale. We gave a new consistent and closed form prior distribution over VSPs with a parameter q controlling VSP depth, and a new observation model QJ-B for noisy LEs which generalises QJ-U [Nicholls and Muir Watt, 2011]. We fit the new model to some of the smaller data sets and the old model to all data sets. Neither of these analyses would be possible without the VSP-setup. The data support the new observation model in our application. Our $elpd_{waic}$-based model comparisons also clearly favor VSP/QJ-U and VSP/QJ-B over a Plackett-Luce mixture or a Mallows Mixture. Although we could fit the large data sets, visualising consensus partial orders proved challenging (compare Fig. 7 (top left corner) and Fig. D.4).

We gave MCMC algorithms targeting the posterior for VSPs in both the latent-space (BDT) parameterisation and the integrated MDT parameterisation. We found the BDT-MCMC adequate, though it would be good to make an efficiency comparison with MDT-MCMC, which we expect to be more efficient. These comparisons are underway. BDT updates which don't change the VSP are fast so BDT-MCMC seems to be competitive. For code see `https://github.com/JessieJ315/Bayesian-Inference-for-Vertex-Series-Parallel-Partial-Orders.git`.

In future work we would like to compare our fit with the recently-proposed contextual repeated selection (CRS) model (Seshadri et al. [2020] and Ragain and Ugander [2018]). This is a rich class of models for rank-order data. The elements of the model are not essentially physical, in the sense that a VSP represents a social hierarchy relation by relation. Also, CRS models do not encode transitivity. It is easy to show VSP models cannot be represented as CRS models with "cliques" of size two. CRS models may fit the data well, and a comparison would be worthwhile. However, there is currently no Bayesian CRS analysis so we leave that for future work.

Acknowledgements

We thank Dr. Nicholas Karn for providing the data, and Prof. David Johnson for enlightening discussions.

References

Niko Beerenwinkel, Nicholas Eriksson, and Bernd Sturmfels. Conjunctive Bayesian networks. *Bernoulli*, pages 893–909, 2007.

Isabel Beichl, Alathea Jensen, and Francis Sullivan. A sequential importance sampling algorithm for estimating linear extensions. 2017.

Graham Brightwell and Peter Winkler. Counting linear extensions. *Order*, 8(3):225–242, 1991.

Andrés Cano, Manuel Gémez-Olmedo, and Serafén Moral. Approximate inference in Bayesian networks using binary probability trees. *International Journal of Approximate Reasoning*, 52(1):49–62, 2011.

Aristides Gionis, Heikki Mannila, Kai Puolamäki, and Antti Ukkonen. Algorithms for discovering bucket orders from data. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 561–566, 2006.

Kustaa Kangas, Teemu Hankala, Teppo Mikael Niinimäki, and Mikko Koivisto. Counting linear extensions of sparse posets. In *IJCAI*, pages 603–609, 2016.

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

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

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

Heikki Mannila. Finding total and partial orders from data for seriation. In *International Conference on Discovery Science*, pages 16–25. Springer, 2008.

Heikki Mannila and Christopher Meek. Global partial orders from sequential data. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 161–168, 2000.

Geoff K Nicholls and Alexis Muir Watt. Partial order models for episcopal social status in 12th century England. *IWSM 2011*, page 437, 2011.

Geoff K. Nicholls, Jeong Eun Lee, Nicholas Karn, David Johnson, Rukuang Huang, and Alexis Muir-Watt. Bayesian inference for partial orders from random linear extensions: power relations from 12th Century Royal Acta, 2022. URL `https://arxiv.org/abs/2212.05524`.

Manuela Pavan and Roberto Todeschini. *Scientific data ranking methods: theory and applications*. Elsevier, 2008.

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

Stephen Ragain and Johan Ugander. Choosing to rank. *arXiv preprint arXiv:1809.05139*, 2018.

Thomas Sakoparnig and Niko Beerenwinkel. Efficient sampling for Bayesian inference of conjunctive Bayesian networks. *Bioinformatics*, 28(18):2318–2324, 07 2012. ISSN 1367-4803. doi: 10.1093/bioinformatics/bts433. URL `https://doi.org/10.1093/bioinformatics/bts433`.

Arjun Seshadri, Stephen Ragain, and Johan Ugander. Learning rich rankings. *Advances in Neural Information Processing Systems*, 33:9435–9446, 2020.

R. Sharpe, D. Carpenter, H. Doherty, M. Hagger, and N. Karn. The Charters of William II and Henry I. Online: Last accessed 27 October 2022, 2014.

L Smail. Junction trees construction: Application to Bayesian networks. In *AIP Conference Proceedings*, volume 2025, page 100007. AIP Publishing LLC, 2018.

Richard Stanley and Eric W. Weisstein. *Catalan Number*. https://mathworld.wolfram.com/CatalanNumber.html, 2002. MathWorld–A Wolfram Web Resource.

Jacobo Valdes. *Parsing Flowcharts and Series-Parallel Graphs.* PhD thesis, Stanford, CA, USA, 1978. AAI7905944.

Jacobo Valdes, Robert E Tarjan, and Eugene L Lawler. The recognition of series parallel digraphs. In *Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 1–12, 1979.

Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, 27:1413–1432, 2017.

Sumio Watanabe. A widely applicable bayesian information criterion. *Journal of Machine Learning Research*, 14 (1):867–897, 8 2013. ISSN 15324435. URL `http://arxiv.org/abs/1208.6338`.

Mark B Wells. Elements of combinatorial computing. 1971.