Efficient Inference with Partial Ranking Queries:
Supplementary Material

In this document, we supply proofs of the theorems in the main paper and provide extra discussion. To make this document stand by itself, we resupply the main definitions and theorem statements.

1 Definitions and notation

Partial rankings.

Definition 1 (Rankings)**.** Let Ω be a finite set of n items. A ranking, σ of items in Ω is a one-to-one mapping between Ω and a rank set ($R = \{1, \ldots, n\}$ unless stated otherwise) and is denoted as $\sigma^{-1}(1)|\sigma^{-1}(2)|\ldots|\sigma^{-1}(n)$. We say that σ ranks item i_1 *before (or above)* item i_2 if the rank of i_1 is less than the rank of i_2.

Definition 2 (Symmetric group)**.** The collection of rankings of itemset Ω is denoted by S_Ω (or just S_n when Ω is implicit). Such sets are called *symmetric groups.*

Definition 3 (Partial Ranking)**.** Let Ω_1, Ω_2,..., Ω_k be an ordered collection of subsets which partition Ω (i.e., $\cup_i \Omega_i = \Omega$ and $\Omega_i \cap \Omega_j = \emptyset$ if $i \neq j$). The *partial ranking* corresponding to this partition is the collection of rankings which rank items in Ω_i before items in Ω_j if $i < j$. We denote this partial ranking as $\Omega_1|\Omega_2|\ldots|\Omega_k$ and say that the partial ranking is of type $\gamma = (|\Omega_1|, |\Omega_2|, \ldots, |\Omega_k|)$. Given a partial ranking of type γ, we denote the set of ranks occupied by Ω_i by R_i^γ. Note that R_i^γ depends only on γ and can be written as $R_1^\gamma = \{1, \ldots, \gamma_1\}$, $R_2^\gamma = \{\gamma_1+1, \ldots, \gamma_1+\gamma_2\}$, ..., $R_k^\gamma = \{\sum_{i=1}^{k-1} \gamma_i + 1, \ldots, n\}$.

We denote the collection of all partial rankings(over a given item set Ω) as $\mathcal{P}$.

Definition 4 (Consistency with a partial order)**.** Given any ranking σ which is a member of a partial ranking $\Omega_1|\ldots|\Omega_k$, we say that σ is *consistent* with $\Omega_1|\ldots|\Omega_k$.

Given a partial ranking $\Omega_1|\ldots|\Omega_k$ of type γ and one consistent ordering σ, one can obtain the vertical bar notation of the partial ranking by 'dropping' vertical bars from that of σ. We denote this partial ranking as $S_\gamma\sigma$. Note that there are multiple ways of writing the same partial ranking.

Proposition 5. *$\sigma \in S_\gamma\pi = \Omega_1|\ldots|\Omega_k$ if and only if for each i, $\sigma(\Omega_i) = R_i^\gamma$.*

Definition 6 (Partial ranking indicators)**.** Given a partial ranking $S_\gamma\pi$, we denote the indicator function for the collection of rankings consistent with $S_\gamma\pi$ by: $\delta_{S_\gamma\pi} : S_n \to \{0, 1\}$, which evaluates to 1 if $\sigma \in S_\gamma\pi$ and 0 otherwise.

Riffled Independence. These definitions follow those from Huang and Guestrin [2, 3, 4].

Definition 7 (Relative ranking map)**.** Given a ranking $\sigma \in S_\Omega$ and any subset $A \subset \Omega$, the *relative ranking of items in* A, $\phi_A(\sigma)$, is a ranking, $\pi \in S_A$, such that $\pi(i) < \pi(j)$ if and only if $\sigma(i) < \sigma(j)$.

Definition 8 (Interleaving map)**.** Given a ranking $\sigma \in S_\Omega$ and a partition of Ω into sets A and B, the *interleaving of A and B in σ* (denoted, $\tau_{AB}(\sigma)$) is a binary function indicating whether a rank in σ is occupied by A or B. More precisely, for any rank i,

$$[\tau_{AB}(\sigma)](i) = \begin{cases} A & \text{if } \sigma^{-1}(i) \in A \\ B & \text{if } \sigma^{-1}(i) \in B \end{cases}.$$

Note that the interleaving notation used here is somewhat different from that used in the main paper, but makes formality easier in some proofs. The *space of interleavings* (for a fixed A,B) is therefore the set of $\{0, 1\}$ valued functions with $|A|$ zeros and $|B|$ ones.

Definition 9 (Riffled Independence)**.** Let h be a distribution over S_Ω and consider a subset $A \subset \Omega$ and its complement B. The sets A and B are said to be *riffle independent* if h factors as:

$$h(\sigma) = m(\tau_{AB}(\sigma)) \cdot f(\phi_A(\sigma)) \cdot g(\phi_B(\sigma)),$$

for distributions m, f and g, defined over interleavings and relative rankings of A and B respectively. We refer to m and the *interleaving distribution* and f and g and the *relative ranking distributions* for A and B, respectively.

Definition 10 (Hierarchy)**.** A (binary) *hierarchy* H over item set Ω is a tuple (H_A, H_B), where H_A and H_B are *either* (1) null, in which case H is called a *leaf*, or (2) hierarchies over item sets A and B, respectively, where (A, B) forms a binary partition of Ω. In this second case, A and B are assumed to both be nonempty.

Definition 11 (Hierarchical Riffled Independence)**.** Let $H = (H_A, H_B)$ be a hierarchy over the item set Ω. A distribution h over rankings of Ω is said to factor *hierarchically with respect to* H if either (1) H is a leaf, *or* (2) the sets A and B are riffle independent with respect to H, and the distributions f and g over relative rankings of A and B factor hierarchically with respect to H_A and H_B, respectively.

Bayesian conditioning. For simplicity in this paper, we focus on *subset observations* whose likelihood functions encode membership with some subset of rankings in S_n.

Definition 12 (Observations)**.** A *subset observation* $\mathcal{O}$ is a binary observation whose likelihood is proportional to the indicator function of some subset of S_n.

Definition 13 (Decomposability)**.** Given a hierarchy H over the item set, a subset observation $\mathcal{O}$ *decomposes* with respect to H if its likelihood function $L(\mathcal{O}|\sigma)$ factors riffle independently with respect to H.

Definition 14 (Complete decomposability)**.** We say that an observation $\mathcal{O}$ is *completely decomposable* if it decomposes with respect to *every* possible hierarchy over the item set Ω.

Finally, we denote the collection of all possible completely decomposable observations as $\mathcal{CRI}$.

2 Main theorems

Proposition 15. *Let H be a hierarchy over the item set. Given a prior distribution h and an observation $\mathcal{O}$ which both decompose with respect to H, the posterior distribution $h(\sigma|\mathcal{O})$ also factors riffle independently with respect to H.*

Proposition 16. *Given a prior h which factorizes riffle independently with respect to a hierarchy H, and a completely decomposable observation $\mathcal{O}$, the posterior $h(\sigma|\mathcal{O})$ also decomposes with respect to H and can be computed in time linear in the number of model parameters of h.*

Theorem 17 (Decomposability of partial rankings)**.** *Every partial ranking observation is completely decomposable ($\mathcal{P} \subset \mathcal{CRI}$).*

Theorem 18 (Converse of Theorem 17)**.** *Every completely decomposable observation takes the form of a partial ranking ($\mathcal{CRI} \subset \mathcal{P}$).*

3 Proofs

3.1 Proof of Proposition 15

Proof. Denote the likelihood function corresponding to $\mathcal{O}$ by L (in this proof, it does not matter that $\mathcal{O}$ is assumed to be a subset observation and the result holds for arbitrary likelihoods).

We use induction on the size of the item set $n = |\Omega|$. The base case $n = 1$ is trivially true. We next examine the general case where $n > 1$. The posterior distribution, by Bayes rule, can be written $h(\sigma|\mathcal{O}) \propto L(\sigma) \cdot h(\sigma)$. There are now two cases. If H is a leaf node, then the posterior h' trivially factors according to H, and we are done. Otherwise, L and h both factor, by assumption, according to $H = (H_A, H_B)$ in the following way:

$$L(\sigma) = m_L(\tau_{AB}(\sigma)) \cdot f_L(\phi_A(\sigma)) \cdot g_L(\phi_B(\sigma)), \text{ and}$$
$$h(\sigma) = m_h(\tau_{AB}(\sigma)) \cdot f_h(\phi_A(\sigma)) \cdot g_h(\phi_B(\sigma)).$$

Multiplying and grouping terms, we see that the posterior factors as:

$$h(\sigma|\mathcal{O}) = [m_L \cdot m_h](\tau_{AB(\sigma)}) \cdot [f_L \cdot f_h](\phi_A(\sigma)) \cdot [g_L \cdot g_h](\phi_B(\sigma)).$$

To show that $h(\sigma|\mathcal{O})$ factors with respect to H, we need to demonstrate (by Definition 11) that the distributions $[f_L \cdot f_h]$ and $[g_L \cdot g_h]$ (after normalizing) factor with respect to H_A and H_B, respectively.

Since f_L and f_h both factor according to the hierarchy H_A by assumption and $|A| < n$ since H is not a leaf, we can invoke the inductive hypothesis to show that the posterior distribution, which is proportional to $f_L \cdot f_h$ must also factor according to H_A. Similarly, the distribution proportional to $g_L \cdot g_h$ must factor ordering to H_B. □

Proof. (of Proposition 16) Proposition 15 requires that the prior and likelihood decompose with respect to the same hierarchy. However, the propoerty of complete decomposability means that the observation $\mathcal{O}$ decomposes with respect to all hierarchies, and so we see that Proposition 16 follows as a simple corollary to Proposition 15. □

3.2 Proof of Theorem 17

To prove the theorem, we will exhibit an explicit factorization for any given binary partition of Ω. To define these factorizations, we make the following definitions.

Definition 19 (Restriction consistency)**.** Given a partial ranking $S_\gamma\sigma = \Omega_1|\Omega_2|\dots|\Omega_k$ and any subset $A \subset \Omega$, we define the *restriction* of $S_\gamma\sigma$ to A as the partial ranking on items in A obtained by intersecting each Ω_i with A. Hence the restriction of $S_\gamma\sigma$ to A is

$$[S_\gamma\sigma]_A = \Omega_1 \cap A|\Omega_2 \cap A|\dots|\Omega_k \cap A.$$

Definition 20 (Interleaving consistency)**.** Given an interleaving τ_{AB} of two sets A, B which partition Ω, we say that $\tau_{A,B}$ is *consistent* with a partial ranking $S_\gamma\sigma = \Omega_1|\dots|\Omega_k$ if, for all i:

$$|\{j \in R_i^\gamma : \tau_{AB}(j) = 0\}| = |\Omega_i \cap A|, \text{ and}$$
$$|\{j \in R_i^\gamma : \tau_{AB}(j) = 1\}| = |\Omega_i \cap B|.$$

Given a partial ranking $S_\gamma\sigma$, we denote the collection of consistent interleavings as $[S_\gamma\sigma]_{AB}$.

In other words, an interleaving τ is consistent with $S_\gamma\sigma$ if it places the 'correct' number of As and Bs in each interval R_i^γ. Note again that the definition here for the sake of formality is somewhat modified from that of the main paper.

Proof. (of Theorem 17) We use induction on the size of the itemset. The cases $n = 1, 2$ are trivial since every distribution on S_1 or S_2 factors riffle independently. We now consider the more general case of $n > 2$.

Fix a partial ranking $S_\gamma\pi = \Omega_1|\Omega_2|\dots|\Omega_k$ of type γ and a binary partition of the item set into subsets A and B. We will show that the indicator function $\delta_{S_\gamma\pi}$ factors as:

$$\delta_{S_\gamma\pi}(\sigma) = m(\tau_{AB}(\sigma)) \cdot f(\phi_A(\sigma)) \cdot g(\phi_B(\sigma)), \quad (3.1)$$

where factors m, f and g are the indicator functions for the set of consistent interleavings, $[S_\gamma\sigma]_{AB}$, and the sets of consistent relative rankings, $[S_\gamma\sigma]_A$ and $[S_\gamma\sigma]_B$, respectively. If Equation 3.1 is true, then we will have shown that $\delta_{S_\gamma\pi}$ must decompose with respect to the top layer of H. To show that $\delta_{S_\gamma\pi}$ decomposes hierarchically, we must also show that the relative ranking factors f_A and g_B decompose with respect to H_A and H_B, the subhierarchies over the item sets A and B. To establish this second step (assuming that Equation 3.1 holds), note that f_A and g_B are indicator functions for the restricted partial rankings, $[S_\gamma\sigma]_A$ and $[S_\gamma\sigma]_B$, which themselves are partial rankings over smaller item sets A and B. The inductive hypothesis (and the fact that A and B are assumed to be strictly smaller sets than Ω) then shows that the functions f_A and g_B both factor according to their respective subhierarchies.

We now turn to establishing Equation 3.1. It suffices to prove that the following two statements are equivalent:

I. The ranking σ is consistent with the partial ranking $S_\gamma\pi$ (i.e., $\sigma \in S_\gamma\pi$).

II. The following three conditions hold:

(a) The interleaving $\tau_{AB}(\sigma)$ is consistent with $S_\gamma\pi$ (i.e., $\tau_{AB}(\sigma) \in [S_\gamma\pi]_{AB}$), and

(b) The relative ranking $\phi_A(\sigma)$ is consistent with $S_\gamma\pi$ (i.e., $\phi_A(\sigma) \in [S_\gamma\pi]_A$), and

(c) The relative ranking $\phi_B(\sigma)$ is consistent with $S_\gamma\pi$ (i.e., $\phi_B(\sigma) \in [S_\gamma\pi]_B$).

- *(I ⇒ II):* We first show that $\sigma \in S_\gamma\pi$ implies conditions (a), (b) and (c).

  (a) If $\sigma \in S_\gamma\pi$, then for each i,

  $$\begin{aligned}|j \in R_i^\gamma : \tau_{AB}(j) = 0| &= |j \in R_i^\gamma : \sigma^{-1}(j) \in A|, \\ &\quad \text{(by Definition 8)} \\ &= |k \in \Omega_i : k \in A|, \\ &\quad \text{(by Proposition 5)} \\ &= |\Omega_i \cap A|.\end{aligned}$$

  The same argument (replacing A with B) shows that for each i, we have $|j \in R_i^\gamma : \tau_{AB}(j) = 1| = |\Omega_i \cap B|$. These two conditions (by Definition 20) show that τ_{AB} is consistent with $S_\gamma\pi$.

  (b) If $\sigma \in S_\gamma\pi$, then (by Definition 3) σ ranks items in Ω_i before items in Ω_j for any $i < j$. Intersecting each Ω_i with A, we also see that σ ranks any item in $\Omega_i \cap A$ before any item in $\Omega_j \cap A$ for all i, j. By Definition 7, $\phi_A(\sigma)$ also ranks any item in $\Omega_i \cap A$ before any item in $\Omega_j \cap A$ for all i, j. And finally by Definition 20 again, we see that $\phi_A(\sigma)$ is consistent with the partial ranking $S_\gamma\pi = \Omega_1 \cap A|\dots|\Omega_k \cap A$.

  (c) (Same argument as (b)).

- *(II ⇒ I):* We now assume conditions (a), (b), and (c) to hold, and show that $\sigma \in S_\gamma\pi$. By Proposition 5 it is sufficient to show that if an item $k \in \Omega_i$, then $\sigma(k) \in R_i^\gamma$. To prove this claim, we show by induction on i that if an item $k \in \Omega_i \cap A$, then $\sigma(k) \in R_i^\gamma$ (and similarly if $k \in \Omega_i \cap B$, then $\sigma(k) \in R_i^\gamma$).

  Base case. In the base case ($i = 1$), we assume that $k \in \Omega_1 \cap A$, and the goal is to show that $\sigma(k) \in R_1$. By condition (a), we have that $\tau_{AB}(\sigma) \in [S_\gamma\pi]_{AB}$. By Definition 20, this means that: $|\Omega_1 \cap A| = \{j \in R_1 : [\tau_{AB}(\sigma)](j) = 0\} = \{j \in R_1 : \sigma^{-1}(j) \in A\}$. In words, there are $m = |\Omega_1 \cap A|$ items from A which lie in rank set $R_1 = \{1, \dots, \gamma_1\}$. To show that an item $k \in A$ maps to a rank in R_1, we now must show that in the relative ranking of elements in A, k is among the first m. By condition (b), $\phi_A(\sigma) \in [S_\gamma\pi]_A$, implying that the item subset $\Omega_1 \cap A$ occupy the first m positions in the relative ranking of A. Since $k \in \Omega_1 \cap A$, item k is among the first m items ranked by $\phi_A(\sigma)$ and therefore $\sigma(k) \in R_1$. A similar argument shows that $k \in \Omega_1 \cap B$ implies that $\sigma(k) \in R_1$.

Inductive case. We now show that if $k \in \Omega_i \cap A$, then $\sigma(k) \in R_i$. By condition (b), $\phi_A(\sigma) \in [S_\gamma\pi]_A$, implying that the item subset $\Omega_i \cap A$ (and hence, item k) occupies the first $m = |\Omega_i \cap A|$ positions in the relative ranking of A beyond the items $\cup_{j=1}^{i-1}(\Omega_j \cap A)$. By the inductive hypothesis and mutual exclusivity, these items, together with $\cup_{j=1}^{i-1}(\Omega_j \cap B)$ occupy ranks $\cup_{j=1}^{i-1} R_j$, and therefore $\sigma(k) \in R_\ell$ for some $\ell \geq i$. On the other hand, condition (a) assures us that $|\Omega_i \cap A| = \{j \in R_i : \sigma^{-1}(j) \in A\}$ — or in other words, that the ranks in R_i are occupied by exactly m items of A. Therefore, $\sigma(k) \in R_i$. Again, a similar argument shows that $k \in \Omega_i \cap B$ implies that $\sigma(k) \in R_i$.

□

3.3 Proof of Theorem 18

Recall that the definition of the *linear span* of a set of vectors in a vector space is the intersection of all linear subspaces containing that set of vectors. To prove Theorem 18, we introduce analogous concepts of the *span* of a set of rankings.

Definition 21 (RSPAN and PSPAN)**.** Let $X \subset S_n$ be any collection of rankings. We define PSPAN(X) to be the intersection of all partial rankings containing X. Similarly, we define RSPAN(X) to be the intersection of all completely decomposable observations containing X. More formally,

$$\text{PSPAN}(X) = \bigcap_{S_\gamma\sigma : X \subset S_\gamma\sigma} S_\gamma\sigma, \text{ and}$$

$$\text{RSPAN}(X) = \bigcap_{\mathcal{O} : X \subset \mathcal{O},\ \mathcal{O} \in \mathcal{CRI}} \mathcal{O}.$$

The proof strategy taken in this section is to show two things: (1) that the PSPAN of any set is always a partial ranking, and (2) that in fact, the RSPAN and PSPAN of a set X are exactly the same sets. We then show in our proof of Theorem 18 that this implies that any element of $\mathcal{CRI}$ must be a partial ranking. The following proposition lists several basic properties of the RSPAN that we will use in several of the proofs. They all follow directly from definition so we do not write out the proofs.

Proposition 22.

I. *(Monotonicity) For any X, $X \subset$ RSPAN(X).*

II. *(Subset preservation) For any X, X' such that $X \subset X$, RSPAN$(X) \subset$ RSPAN(X').*

III. *(Idempotence) For any X, RSPAN(RSPAN(X)) = RSPAN(X).*

To reason about the PSPAN of a set of rankings, we first introduce some basic concepts regarding the combinatorics of partial rankings. The collection of partial rankings over Ω forms a *partially ordered set (poset)* where $S_{\gamma'}\pi' \prec S_\gamma\pi$ if $S_\gamma\pi$ can be obtained from $S_{\gamma'}\pi'$ by dropping vertical lines. For example, on S_3, we have that $1|2|3 \prec 12|3$. The *Hasse diagram* is the graph in which each node corresponds to a partial ranking and a node x is connected to node y via an edge if $x \prec y$ and there exists no partial ranking z such that $x \prec z \prec y$ (see [5]). At the top of the Hasse diagram is the partial ranking $1, 2, \ldots, n$ (i.e., all of S_Ω) and at the bottom of the Hasse diagram lie the full rankings. See Figure 3.3 for an example of the partial ranking lattice on S_3.

Lemma 23. *[Lebanon and Mao (2008) [5]] Given any two partial rankings $S_\gamma\pi$, $S_{\gamma'}\pi'$, there exists a unique supremum of $S_\gamma\pi$ and $S_{\gamma'}\pi'$ (a node $S_{\gamma_{sup}}\pi_{sup}$ such that $S_\gamma\pi \prec S_{\gamma_{sup}}\pi_{sup}$ and $S_{\gamma'}\pi' \prec S_{\gamma_{sup}}\pi_{sup}$, and any other such node is greater than $S_{\gamma_{sup}}\pi_{sup}$). Similarly, there exists a unique infimum of $S_\gamma\pi$ and $S_{\gamma'}\pi'$.*

Lemma 24. *Given two partial rankings $S_\gamma\pi$, $S_{\gamma'}\pi'$, the relation $S_{\gamma'}\pi' \subset S_\gamma\pi$ holds if and only $S_\gamma\pi$ lies above $S_{\gamma'}\pi'$ in the Hasse diagram.*

Proof. If $S_\gamma\pi$ lies above $S_{\gamma'}\pi'$ in the Hasse diagram, then $S_{\gamma'}\pi' \subset S_\gamma\pi$ is trivial since $S_\gamma\pi$ can be obtained by dropping vertical bars of $S_{\gamma'}\pi'$. Now given that $S_\gamma\pi$ *does not* lie above $S_{\gamma'}\pi'$, we would like to show that $S_{\gamma'}\pi' \not\subset S_\gamma\pi$. Let $S_{\gamma_{inf}}\pi_{inf}$ be the unique infimum of $S_\gamma\pi$ and $S_{\gamma'}\pi'$ as guaranteed by Lemma 23. By the definition of the Hasse diagram, both $S_\gamma\pi$ and $S_\gamma\pi$ can be obtained by 'dropping' verticals from the vertical bar representation of $S_{\gamma_{inf}}\pi_{inf}$. Since $S_\gamma\pi$ does not lie above $S_{\gamma'}\pi'$, there must be a vertical bar that was dropped by $S_{\gamma'}\pi'$ which was not dropped by $S_\gamma\pi$ (if there does not exist such a bar, then $S_{\gamma'}\pi' \subset S_\gamma\pi$), and hence there must exist a pair of items i, j separated by a single vertical bar in $S_\gamma\pi$ but unseparated in $S_{\gamma'}\pi'$. Therefore there exists $\sigma \in S_{\gamma'}\pi'$ such that $\sigma(j) < \sigma(i)$ even though there exists no such $\sigma \in S_\gamma\pi$. We conclude that $S_{\gamma'}\pi' \not\subset S_\gamma\pi$. □

Lemma 25. *For any $X \subset S_n$, PSPAN(X) is a partial ranking.*

Proof. Consider any subset $X \subset S_n$. A partial ranking containing every element in X must be an upper bound of every element of X in the Hasse diagram by Lemma 24. By Lemma 23, there must ex-

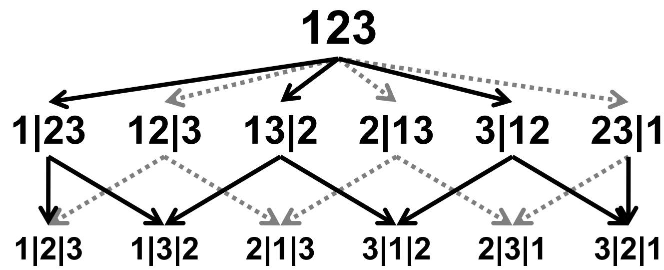


Figure 1: The Hasse diagram for the lattice of partial rankings on S_3.

ist a unique least upper bound (supremum) of X, $S_{\gamma_{sup}}\pi_{sup}$, such that for any common upper bound $S_\gamma\pi$ of X, $S_\gamma\pi$ must also be an ancestor of $S_{\gamma_{sup}}\pi_{sup}$ and hence $S_{\gamma_{sup}}\pi_{sup} \subset S_\gamma\pi$. We therefore see that any partial ranking containing X must be a superset of $S_{\gamma_{sup}}\pi_{sup}$. On the other hand, $S_{\gamma_{sup}}\pi_{sup}$ is itself a partial ranking containing X. Since PSPAN(X) is the intersection of partial rankings containing X, we have PSPAN$(X) = S_{\gamma_{sup}}\pi_{sup}$ and therefore that PSPAN(X) must be a partial ranking. ☐

Lemma 26. *For any subset of orderings, X, RSPAN$(X) \subset$ PSPAN(X).*

Proof. Lemma 26 follows almost directly from the fact that $\mathcal{P} \subset \mathcal{CRI}$ (Theorem 17). Fix a subset $X \subset S_n$ and let π be any element of RSPAN(X). Consider any partial ranking indicator function $\delta \in \mathcal{P}$ such that $\delta(\sigma) > 0$ for all $\sigma \in X$. We want to see that $\delta(\pi) > 0$. By Theorem 17, $\delta \in \mathcal{CRI}$. Moreover, since $\pi \in$ RSPAN(X), and $\delta(\sigma) > 0$ for all $\sigma \in X$, we conclude that $\delta(\pi) > 0$ (by Definition 21). ☐

To simplify the notation in some of the remaining proofs, we introduce the following definition.

Definition 27 (Ties)**.** Given a partial ranking $S_\gamma\pi = \Omega_1|\dots|\Omega_k$, we say that items a_1 and a_2 are *tied* (written $a_1 \sim a_2$) with respect to $S_\gamma\sigma$ if $a_1, a_2 \in \Omega_i$ for some i.

The following basic properties of the tie relation are straightforward.

Proposition 28.

I. *With respect to a fixed partial ranking $S_\gamma\pi$, the tie relation, $\sim$, is an equivalence relation on the item set (i.e., is reflexive, symmetric and transitive).*

II. *If there exist $\sigma, \sigma' \in S_\gamma\pi$ which disagree on the relative ranking of items a_1 and a_2, then $a_1 \sim a_2$ with respect to $S_\gamma\pi$.*

III. *If $S_\gamma\pi \prec S_{\gamma'}\pi'$, and $a_1 \sim a_2$ with respect to $S_\gamma\pi$, then $a_1 \sim a_2$ with respect to $S_{\gamma'}\pi'$.*

IV. *If $a_1 \sim a_2$ with respect to $S_\gamma\pi$, and $\sigma(a_1) < \sigma(a_3) < \sigma(a_2)$ for some item $a_3 \in \Omega$ and some $\sigma \in S_\gamma\pi$, then $a_1 \sim a_2 \sim a_3$.*

function FORMPSPAN(X)

$X_0 \leftarrow X$; $t \leftarrow 0$;
while *$\exists S_\gamma\pi, S_{\gamma'}\pi' \in X_t$ which disagree on the relative ordering of items a_1, a_2* **do**
  $X_t \leftarrow \emptyset$;
  foreach $S_\gamma\sigma \in X_t$ **do**
    Add any partial ranking obtained by deleting a vertical bar from $S_\gamma\sigma$ between items a_1 and a_2 to X_t;
  $t \leftarrow t+1$;
return any element of X_t ;

Algorithm 1: Pseudocode for computing PSPAN(X). FORMPSPAN(X) takes a set of partial rankings (or full rankings) X as input and outputs a partial ranking. This algorithm iteratively deletes vertical bars from elements of X until they are in agreement. Note that it is not necessary to keep track of t, but we do so here to ease notation in the proofs. Nor is this algorithm the most direct way of computing PSPAN(X), but again, it simplifies the proof of our main theorem.

We now consider the problem of computing the partial ranking span (PSPAN) of a given set of rankings X. In Algorithm **??**, we show a simple procedure that provably outputs the correct result.

Proposition 29. *Given a set of rankings X as input, Algorithm 1 outputs* PSPAN(X).

Proof. We prove three things, which together prove the proposition: (1) that the algorithm terminates, (2) that at each stage the elements of X are contained in PSPAN(X), and (3) that upon termination, PSPAN(X) is contained in each element of X.

1. First we note that the algorithm must terminate in finitely many iterations of the while loop since at each stage at least one vertical bar is removed from a partial ranking, and when all of the vertical

bars have been removed from the elements of X, there are no disagreements on relative ordering.

2. We now show that at any stage in the algorithm, every element of X_t is a subset of the PSPAN(X). Consider $S_\gamma\pi \in X_t$ such that $S_\gamma\pi \subset$ PSPAN(X). If $S_\gamma\pi$ is replaced by $S_{\gamma'}\pi'$ in X_{t+1}, then we want to show that $S_{\gamma'}\pi' \subset$ PSPAN(X) as well. From Algorithm 1, for some i, if $S_\gamma\pi = \Omega_1|\dots|\Omega_j|\Omega_{j+1}|\dots|\Omega_k$, $S_{\gamma'}\pi'$ can be written as $\Omega_1|\dots|\Omega_j \cup \Omega_{j+1}|\dots|\Omega_k$, where the vertical bar between Ω_j and Ω_{j+1} are deleted due to the existence of partial rankings in X_t which disagree on the relative ordering of items a_1, a_2 on opposite sides of the bar, then by Proposition 29 (II), we know that $a_1 \sim a_2$ (with respect to $S_\gamma\pi$). By transitivity (I) and (II), if $a_1 \in \Omega_i$ and $a_2 \in \Omega_{i'}$, then any two elements of Ω_i and $\Omega_{i'}$ are also tied. By (IV), all the items lying in $\Omega_i, \Omega_{i+1}, \dots, \Omega_{i'}$ are thus tied with respect to PSPAN(X) and therefore removing any bar between items a_1 and a_2 (producing, for example, $S_{\gamma'}\pi'$) results in a partial ranking which is a subset of PSPAN(X).
3. Finally, upon termination, if some ranking $\sigma \in X$ is not contained in some element $S_\gamma\pi \in X_t$, then there would exist two items a_1, a_2 whose relative ranking σ and $S_\gamma\pi$ disagree upon, which is a contradiction. Therefore, every element $S_\gamma\pi \in X_t$ contains every element of X and thus PSPAN(X) $\subset S_\gamma\pi$ for every $S_\gamma\pi \in X_t$.

□

As a final step before being able to prove our second main claim, that RSPAN(X) = PSPAN(X) for any X, we prove the following two technical lemmas about Algorithm 1 which form the heart of our argument. In particular, for a completely decomposable observation $\mathcal{O} \in \mathcal{CRI}$, Lemma 30 below shows a ranking contained in $\mathcal{O}$ can "force" other rankings to be contained in $\mathcal{O}$.

Lemma 30. *Let $\mathcal{O} \in \mathcal{CRI}$ and suppose there exist $\pi_1, \pi_2 \in S_n$ which disagree on the relative ranking of items $i, j \in \Omega$ such that $\pi_1, \pi_2 \in \mathcal{O}$. Then the ranking obtained by swapping the relative ranking of items i, j within any $\pi_3 \in \mathcal{O}$ must also be contained in $\mathcal{O}$.*

Proof. Let h be the indicator distribution corresponding to the observation $\mathcal{O}$. We will show that swapping the relative ranking of items i, j in π_3 will result in a ranking which is assigned nonzero probability by h, thus showing that this new ranking is contained in $\mathcal{O}$.

Let $A = \{i, j\}$ and $B = \Omega \backslash A$. Since $\mathcal{O} \in \mathcal{CRI}$, h must factor riffle independently according to the partition (A, B). Thus,

$$h(\pi_1) = m(\tau_{AB}(\pi_1)) \cdot f(\phi_A(\pi_1)) \cdot g(\phi_B(\pi_1)) > 0, \text{ and}$$
$$h(\pi_2) = m(\tau_{AB}(\pi_2)) \cdot f(\phi_A(\pi_2)) \cdot g(\phi_B(\pi_2)) > 0.$$

Since π_1 and π_2 disagree on the relative ranking of items in A, this factorization implies in particular that both $f(\phi_A = i|j) > 0$ and $f(\phi_A = j|i) > 0$. Since $h(\pi_3) > 0$, it must also be that each of $m(\tau_{AB}(\pi_3))$, $f(\phi_A(\pi_3))$, and $g(\phi_B(\pi_3))$ have positive probability. We can therefore swap the relative ranking of A, ϕ_A, to obtain a new ranking which has positive probability since all of the terms in the decomposition of this new ranking have positive probability. □

Lemma 31 provides conditions under which removing a vertical bar from one of the rankings in X will not change the support of a completely riffle independent distribution. The key strategy in this proof is to argue that large subsets of rankings must be contained in a completely decomposable observation $\mathcal{O}$ by decomposing rankings into transpositions and invoking Lemma 30 *ad nauseum.*

Lemma 31. *Let $S_\gamma\pi = \Omega_1|\dots|\Omega_i|\Omega_{i+1}|\dots|\Omega_k$ be a partial ranking on item set Ω, and $S_{\gamma'}\pi' = \Omega_1|\dots|\Omega_i \cup \Omega_{i+1}|\dots|\Omega_k$, the partial ranking in which the sets Ω_i and Ω_{i+1} are merged. Let $a_1 \in \cup_{j=1}^{i}\Omega_j$ and $a_2 \in \cup_{j=i+1}^{k}\Omega_j$. If $\mathcal{O}$ is any element of $\mathcal{CRI}$ such that $S_\gamma\pi \subset \mathcal{O}$ and there additionally exists a ranking $\tilde{\pi} \in \mathcal{O}$ which disagrees with $S_\gamma\pi$ on the relative ordering of a_1, a_2, then $S_{\gamma'}\pi' \subset \mathcal{O}$.*

Proof. We will fix a completely decomposable $\mathcal{O}$ and again work with h, the indicator distribution corresponding to $\mathcal{O}$. Let $\sigma \in S_{\gamma'}\pi'$. To prove the lemma, we need to establish that $h(\sigma) > 0$. Let σ^0 be any element of $S_\gamma\pi$ such that $\sigma^0(k) = \sigma(k)$ for all $k \in \Omega \backslash (\Omega_i \cup \Omega_{i+1})$. Since $S_\gamma\pi \subset \mathrm{supp}(h)$ by assumption, we have that $h(\sigma^0) > 0$.

Since σ^0 and σ match on all items except for those in $\Omega_i \cup \Omega_{i+1}$, there exists a sequence of rankings $\sigma^0, \sigma^1, \sigma^2, \dots, \sigma^m = \sigma$ such that adjacent rankings in this sequence differ only by a pairwise exchange of items$b_1, b_2 \in \Omega_i \cup \Omega_{i+1}$. We will now show that at each step along this sequence, $h(\sigma^t) > 0$ implies that $h(\sigma^{t+1}) > 0$, which will prove that $h(\sigma) > 0$. Suppose now that $h(\sigma^t) > 0$ and that σ^t and σ^{t+1} differ only by the relative ranking of items $b_1, b_2 \in \Omega_i \cup \Omega_{i+1}$ (without loss of generality, we will assume that $\sigma^t(b_2) < \sigma^t(b_1)$ and $\sigma^{t+1}(b_1) < \sigma^{t+1}(b_2)$).

The idea of the following paragraph is to use the previous lemma (Lemma 30) to prove that σ^{t+1} has positive probability and to do so, it will be necessary to argue that there exists some ranking σ' such that $h(\sigma') > 0$ and $\sigma'(b_1) < \sigma'(b_2)$ (i.e., σ' disagrees with σ^t on the relative ranking of b_1, b_2). Let ω be any element of $S_\gamma\pi$. If $a_1 \in \Omega_i$, rearrange ω such that a_1 is ranked first among elements of Ω_i. If $a_2 \in \Omega_{i+1}$, further rearrange ω such that a_2 is ranked last among elements of

Ω_{i+1}. Note that ω is still an element of $S_\gamma\pi$ after the possible rearrangements and therefore $h(\omega) > 0$. We can assume that $\omega(b_2) < \omega(b_1)$ since otherwise we will have shown what we wanted to show. Thus the relative ordering of a_1, a_2, b_1, b_2 within ω is $a_1|b_2|b_1|a_2$. Note that we treat the case where the items a_1, a_2, b_1, b_2 are distinct, but the same argument follows in the cases when $a_1 = b_2$ or $a_2 = b_1$.

Now since $\tilde{\pi}$ disagrees with $S_\gamma\pi$ on the relative ordering of a_1, a_2 by assumption (and hence disagrees with ω), we apply Lemma 30 to conclude that swapping the relative ordering of a_1, a_2 within ω (obtaining $a_2|b_2|b_1|a_1$) results in a ranking, ω', such that $h(\omega') > 0$. Finally, observe that ω and ω' must now disagree on the relative ranking of a_2, b_2, and invoking Lemma 30 again shows that we can swap the relative ordering of a_2, b_2 within ω (obtaining $a_1|a_2|b_1|b_2$) to result in a ranking σ' such that $h(\sigma') > 0$. This element σ' ranks b_1 before b_2, which is what we wanted to show.

We have shown that there exist rankings which disagree on the relative ordering of b_1 and b_2 with positive probability under h. Again applying Lemma 30 shows that we can swap the relative ordering of items b_1, b_2 within σ^t to obtain σ^{t+1} such that $h(\sigma^{t+1}) > 0$, which concludes the proof. □

Recall that Lemma 26 showed that $\text{RSPAN}(X) \subset \text{PSPAN}(X)$. We now use Lemma 31 to show the reverse inclusion also holds, establishing that the two sets are in fact equal.

Proposition 32. *For any subset of orderings, X, $\text{RSPAN}(X) \supset \text{PSPAN}(X)$.*

Proof. At each iteration t, Algorithm 1 produces a set of partial rankings, X_t. We denote the union of all partial rankings at time t as $\tilde{X}_t \equiv \bigcup_{S_\gamma\sigma\in X_t} S_\gamma\sigma$. Note that $\tilde{X}_0 = X$ and $\tilde{X}_T = \text{PSPAN}(X)$. The idea of our proof will be to show that at each iteration t, the following set inclusion holds: $\text{RSPAN}(\tilde{X}_t) \subset \text{RSPAN}(\tilde{X}_{t-1})$. If indeed this holds, then after the final iteration T, we will have shown that:

$$\begin{aligned}
\text{PSPAN}(X) &= \tilde{X}_T, && \text{(Proposition 29)}\\
&\subset \text{RSPAN}(\tilde{X}_T), && \text{(Monotonicity, Proposition 22)}\\
&\subset \text{RSPAN}(\tilde{X}_0), && (\text{RSPAN}(\tilde{X}_t) \subset \text{RSPAN}(\tilde{X}_{t-1}), \text{ shown below}),\\
&\subset \text{RSPAN}(X) && (\tilde{X}_0 = X, \text{ see Algorithm 1})
\end{aligned}$$

which would prove the Proposition.

It remains now to show that $\text{RSPAN}(\tilde{X}_t) \subset \text{RSPAN}(\tilde{X}_{t-1})$. We claim that $\tilde{X}_t \subset \text{RSPAN}(\tilde{X}_{t-1})$. Let $\sigma \in \tilde{X}_t$. If $\sigma \in \tilde{X}_{t-1}$, then since $\tilde{X}_{t-1} \subset \text{RSPAN}(\tilde{X}_{t-1})$, we have $\sigma \in \text{RSPAN}(\tilde{X}_{t-1})$ and the proof is done. Otherwise, $\sigma \in \tilde{X}_t\backslash\tilde{X}_{t-1}$. In this second case, we use the fact that at iteration t, the vertical bar between Ω_i and Ω_{i+1} was deleted from the partial ranking $S_\gamma\pi = \Omega_1|\dots|\Omega_i|\Omega_{i+1}|\dots|\Omega_k$ (which is a subset of $\tilde{X}_{t-1}$) to form the partial ranking $S_{\gamma'}\pi' = \Omega_1|\dots|\Omega_i\cup\Omega_{i+1}|\dots|\Omega_k$. (which is a subset of $\tilde{X}_t$). Furthermore, in order for the vertical bar to have been deleted by the algorithm, there must have existed some partial ranking (and therefore some full ranking ω') that disagreed with $S_\gamma\pi$ on the relative ordering of items a_1, a_2 on opposite sides of the bar. Since $\sigma \in \tilde{X}_t\backslash\tilde{X}_{t-1}$ we can assume that $\sigma \in S_{\gamma'}\pi'$.

We now would like to apply Lemma 31. Note that for any $\mathcal{O} \in \mathcal{CRI}$ such that $\tilde{X}_{t-1} \subset \mathcal{O}$, we also have $S_\gamma\pi \subset \mathcal{O}$, since $S_\gamma\pi \subset \tilde{X}_{t-1}$. An application of Lemma 31 then shows that $S_{\gamma'}\pi' \subset \mathcal{O}$ and therefore that $\sigma \in \mathcal{O}$.

We have shown in fact that $\sigma \in \mathcal{O}$ holds for *any* $\mathcal{O} \in \mathcal{CRI}$ such that $\tilde{X}_{t-1} \subset \mathcal{O}$, and therefore taking the intersection of supports over all $\mathcal{O} \in \mathcal{CRI}$, we see that $\tilde{X}_t \subset \text{RSPAN}(\tilde{X}_{t-1})$. Taking the RSPAN of both sides yields:

$$\begin{aligned}
\text{RSPAN}(\tilde{X}_t) &\subset \text{RSPAN}(\text{RSPAN}(\tilde{X}_{t-1})), && \text{(Subset preservation, Proposition 22)}\\
&\subset \text{RSPAN}(\tilde{X}_{t-1}). && \text{(Idempotence, Proposition 22)}
\end{aligned}$$

□

Finally, we can collect the lemmas together to prove our main result — that any completely decomposable observation must take the form of a partial ranking.

Proof. (of Theorem 18): Let $\mathcal{O} \in \mathcal{CRI}$ and let $X = \mathcal{O}$. We have $\mathcal{O} = \text{RSPAN}(\mathcal{O})$. Since $\text{RSPAN}(X) \supset \text{PSPAN}(X)$ by Proposition 32, and $\text{RSPAN}(X) \subset \text{PSPAN}(X)$ by Lemma 26, we have equality: $\text{RSPAN}(X) = \text{PSPAN}(X)$, implying that $X = \text{PSPAN}(X)$. Finally, by Lemma 25, we know that $\text{PSPAN}(X)$ is a partial ranking, and therefore $X = \mathcal{O}$ must also be a partial ranking. □

4 Mallows models and partial rankings

In this section we explore the relationship between the well known *Mallows model* with riffle independent hierarchical models. We show in particular that, under a Mallows distribution, items are riffle independent with respect to a chain hierarchy. Exploiting this riffle independent structure yields an efficient method for conditioning on arbitrary partial rankings.

Let $d_\tau : S_n \times S_n \to \mathbb{R}$ be the *Kendall's tau* distance metric on rankings. Given two rankings σ_1, σ_2, $d_\tau(\sigma_1, \sigma_2)$ is defined as the minimum number of adjacent transpositions necessary to convert one argument σ_1 into the other, σ_2. The *Mallows model* is a distribution over rankings defined as:

$$p(\sigma; \phi, \sigma_0) \propto \phi^{-d_\tau(\sigma, \sigma_0)}, \tag{4.1}$$

where σ_0 represents a *central or reference* ranking and ϕ is a spread parameter. For simplicity, we will assume that σ_0 is the identity ranking mapping item 1 to rank 1, item 2 to rank 2, and so on.

For a given ranking σ and each item j of the item set, define:

$$V_j(\sigma) = \#\{i \,:\, j+1 \le i \le n,\, \sigma(i) < \sigma(j)\},$$

which is simply the number of items in the itemset $\{j+1, \ldots, n\}$ which are ranked before item j with respect to σ. The collection of V_js fully determines the ranking σ, and the following procedure can be used to reconstructs σ ([6]):

function RECONSTRUCTSIGMA($V_1, \ldots, V_{n-1}$)

Initialize σ to be a ranking of $\{n\}$, mapping n to 1 ;
for $j = n-1, n-2, \ldots, 1$ **do**
  Insert item j in rank $V_j + 1$;
return σ ;

Algorithm 2: Reconstruct σ from the collection of V_js. Note that V_n is always zero and hence is not used in the algorithm.

Fligner & Verducci first showed [1] (see also [6]) that a ranking can be sampled from the Mallows model by drawing the V_j independently, each according to a particular exponentially parameterized distribution. In particular, set each V_j to be a value r drawn from the set $\{0, \ldots, n-j\}$ with probability proportional to ϕ^r (where, again, ϕ is the Mallows spread parameter). Using Algorithm 2 to reconstruct σ from the drawn values of V_j yields an independent draw from a Mallows model with spread parameter ϕ.

This generative procedure of drawing the V_j independently is exactly the same as that of a riffle independent hierarchy in which a single item is partitioned out of the hierarchy at each level of the hierarchy, with exponentially parameterized interleaving distributions. For example, on $n = 5$ items, the hierarchy encoding the factorization of the Mallows model is given in Figure 2 with item 1 being partitioned out at the topmost level, then item 2 partitioned out at the second layer, and so on. Since each leaf node consists of a single item, there are no relative ranking parameters. At each internal node of the hierarchy, the interleaving distribution which determines the position where

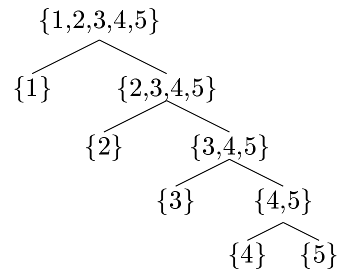


Figure 2: An example of a hierarchical structure over five food items

item j is inserted into the item subset $\{j+1, \ldots, n\}$ is given by $m(\tau_{AB} = B|B|\ldots|A|\ldots|B|B) \propto \phi^r$, where r is the position of the A item in τ.

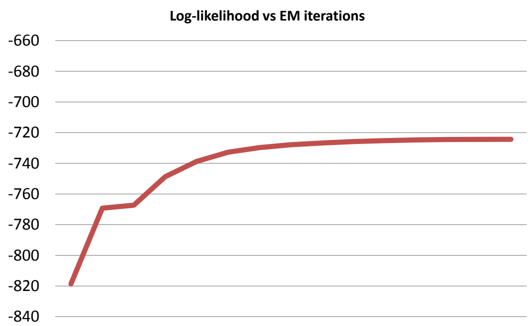
As a side note, we remark that these interleaving distributions are very similar to (but not exactly the same as) the biased riffle shuffles introduced in Huang et al. [2], where the interleaving step is likened to the *riffle shuffle* for cards, in which one drops cards one by one, selected from the left or right deck after each drop with some probability.

To compute the sum over rankings which are consistent with a partial ranking $S_\gamma\sigma$, it is necessary to condition on $S_\gamma\sigma$, and to compute the normalization constant of the resulting function. The conditioning step can be performed using the methods in this paper, and the normalization constant can be computed by multiplying the normalization constant of each factor of the hierarchical decomposition.

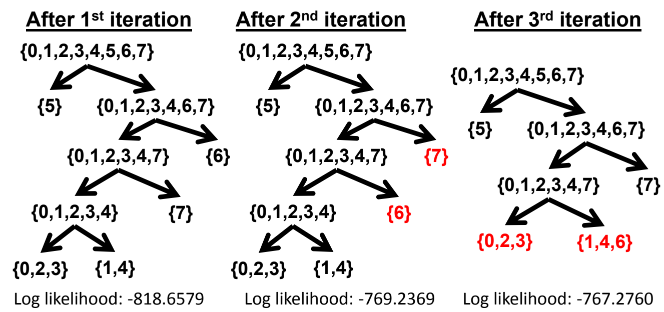
Palin dataset

We extracted a dataset from a database of *search-trails* collected by [7], in which browsing sessions of roughly 2000 users were logged during 2008-2009. In many cases, users are unlikely to read articles about the same story twice, and so it is often possible to think of the order in which a user reads through a collection of articles as a top-k ranking over articles concerning a particular story/topic. The ability to model visit orderings would allow us to make long term predictions about user browsing behavior, or even recommend 'curriculums' over articles for users. We ran our algorithms on roughly 300 visit orderings for the eight most popular posts from `www.huffingtonpost.com` concerning 'Sarah Palin', a popular subject during the 2008 U.S. presidential election. The articles in the dataset are ordered as follows:

- `http://www.huffingtonpost.com/2008/08/29/sarah-palin-former-beauty_n_122400.html`
- `http://www.huffingtonpost.com/2008/09/13/tina-fey-as-sarah-palin-o_n_126249.html`



(a) Progress of EM with respect to expected complete data log-likelihood. Notice that first three iterations correspond to structural changes, and after the first three points, improvements are due to parameter changes, which are smoother.



(b) Iterations of Structure EM for the Sarah Palin data with structural changes at each iteration highlighted in red. This figure is best viewed in color.

Figure 3: Experiment results for the Sarah Palin dataset.

- `http://www.huffingtonpost.com/2008/08/31/sarah-palin-photos-a-bust_n_122816.html`
- `http://www.huffingtonpost.com/charlotte-hilton-andersen/sarah-palin-bikini-pictur_b_123234.html`
- `http://www.huffingtonpost.com/2008/09/27/tina-fey-as-sarah-palin-k_n_129956.html`
- `http://www.huffingtonpost.com/2008/09/03/sarah-palin-rnc-conventio_n_123703.html`
- `http://www.huffingtonpost.com/2008/09/11/sarah-palins-charlie-gibs_n_125772.html`
- `http://www.huffingtonpost.com/2008/08/01/sarah-palin-mccains-vice_n_116383.html`

We plot the first three iterations of the EM algorithm when run on the Sarah Palin data in Figure 3(b). These three iterations are important because they correspond to large global structural changes in the model. After the third iteration, the structure did not change and the only improvements in log-likelihood are due to parameter learning (Figure 3(a)).

We remark that the final structure is interpretable — for example, the leaf set $\{0, 2, 3\}$ corresponds to the three posts about Palin's wardrobe before the election, while the posts from the leaf set $\{1, 4, 6\}$ were related to verbal gaffes made by Palin during the campaign. Article 5 was about Palin's RNC convention speech and article 7 is about the announcement that she was joining the McCain ticket as vice presidential candidate.

References

[1] M. Fligner and J. Verducci. Distance-based ranking models. *J. Royal Stat. Soc., Series B*, 83, 1986.

[2] J. Huang and C. Guestrin. Riffled independence for ranked data. In *NIPS*, 2009.

[3] J. Huang and C. Guestrin. Learning hierarchical riffle independent groupings from rankings. In *ICML*, 2010.

[4] J. Huang and C. Guestrin. Uncovering the riffled independence structure of ranked data. *http://arxiv.org/abs/1006.1328*, 2011.

[5] G. Lebanon and Y. Mao. Non-parametric modeling of partially ranked data. In *NIPS*, 2008.

[6] M. Meila, K. Phadnis, A. Patterson, and J. Bilmes. Consensus ranking under the exponential model. Technical Report 515, University of Washington, Statistics Department, April 2007.

[7] R. White and S. Drucker. Investigating behavioral variability in web search. In *WWW*, 2007.