Conditional independence and graphical models for rankings

Aritz Pérez
Basque Center for Applied Mathematics (BCAM)
aperez@bcamth.org

Abstract

Ranking models define probability distributions over rankings and are fundamental in applications such as preference learning and decision-making. Without structural assumptions, these models are intractable due to their factorial complexity. Although independence assumptions are key to reducing complexity and enabling interpretability, standard notions of conditional independence must account for the mutual exclusivity constraints inherent to rankings. We introduce *relative rank models*, a probabilistic framework based on structured coarsenings of the ranking space that relax these constraints while preserving relative order information. This representation enables a well-defined notion of conditional independence for rankings and allows rankings to be modeled using Bayesian networks. As a result, standard methods for learning and probabilistic inference can be applied to rankings, with interpretable representations based on conditional independence. We further show that two prominent ranking models based on independence assumptions tailored to rankings arise as special cases, providing a unifying framework for independence in ranking models.

1 Introduction

Modeling rankings is a fundamental problem in machine learning, with applications including preference learning, recommender systems, and information retrieval. In these settings, rankings are represented by permutations and express preferences over a finite set of items [Regenwetter et al., 2006], where lower ranks correspond to stronger preferences. A *ranking model* defines a probability distribution over the ranking space and is typically learned from ranking data [Mallows, 1957].

The most general ranking model assigns an arbitrary probability to each ranking, requiring a factorial number of parameters in the number of items. This factorial complexity renders such models impractical, as they cannot be learned from limited data, do not admit efficient inference procedures, and lack interpretability.

To overcome these limitations, ranking models commonly rely on structural assumptions that reduce model complexity [Mallows, 1957, Fligner and Verducci, 1986]. Among these, independence assumptions play a central role. Two representative examples are the *repeated insertion model* [Doignon et al., 2004, Kenig et al., 2018] and the *riffle shuffle model* [Huang and Guestrin, 2009, 2010, Huang et al., 2012]. The repeated insertion model imposes a fixed insertion order that governs how items are sequentially incorporated into a ranking, whereas the riffle shuffle model introduces a more flexible notion of independence by allowing rankings over disjoint subsets of items to be interleaved.

In probabilistic modeling, the most fundamental notion of independence is *conditional independence* [Dawid, 1979], which underlies graphical models such as Bayesian networks [Pearl, 1988, Lauritzen, 1996]. Graphical models provide compact representations of complex distributions, support efficient inference and learning algorithms, and offer a clear interpretation in terms of conditional

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

independence relations [Koller and Friedman, 2009]. However, applying conditional independence to rankings requires accounting for the *mutual exclusivity constraints* (MEC) inherent to permutations [Huang et al., 2009a], which require distinct items to occupy distinct ranks. These constraints induce strong dependencies among ranking variables and restrict the free parameterization of standard graphical factorizations.

In this work, we address this limitation by shifting the modeling focus from full rankings to structured forms of partial ranking information. Rather than treating partial rankings merely as incomplete observations [Lebanon and Mao, 2007, Lu and Boutilier, 2011, Lan et al., 2013, Lu and Boutilier, 2014], we use them as fundamental modeling primitives [Achab et al., 2019, Chierichetti et al., 2018]. By defining structured families of partial rankings, we construct random variables whose realizations induce partitions of the permutation space. This coarsening relaxes the mutual exclusivity constraints and enables a well-defined notion of conditional independence for ranking data, thereby supporting the construction of graphical models for rankings.

Contributions. We introduce *relative rank models*, probabilistic models for rankings based on structured coarsenings of the permutation space. Our main contributions are:

- We establish fundamental results for *relative rank variables* and the coarsenings of the permutation space, relaxing mutual exclusivity constraints and providing the basis for relative rank models.
- We define conditional independence for rankings based on relative rank variables and introduce *relative rank networks*, graphical models for rankings that enable the use of standard theory and algorithms for learning and probabilistic inference, while providing interpretable representations in terms of conditional independence.
- We highlight the connection between relative rank networks and existing ranking models, showing that models such as repeated insertion and riffle shuffle can be represented within the RRN framework and that their assumptions correspond to specific sets of conditional independence statements.

We provide the proofs of the theoretical results in the main text in Appendix F.

Notation. Let $N = \{1, \dots, n\}$ be a set of indices, and let $V = \{v_1, \dots, v_n\}$ be a set of items. Uppercase letters A, B, C denote subsets of V, and $|A|$ denotes the cardinality of A. We use $i, j, k \in N$ and $u, v, w \in V$ to denote particular indices and items, respectively. We denote by $\mathcal{S}_n^V$ the set of bijective functions from V to N, called *rankings*, and by $\mathcal{S}_V^n$ the set of bijective functions from N to V, called *orderings*. Rankings are denoted by Greek letters ρ, σ, τ, π, and their respective orderings by $\bar{\rho}, \bar{\sigma}, \bar{\tau}, \bar{\pi}$. For simplicity, in examples we identify items with their indices, i.e., $V = N$, and represent rankings and orderings as tuples using $()$ and $[]$, respectively. For example, the ranking $(3, 1, 2)$ corresponds to the ordering $[2, 3, 1]$, and vice versa. All directed graphs considered in this work have vertex set V and are identified with their sets of arcs, denoted by uppercase calligraphic letters $\mathcal{V}, \mathcal{G} \subseteq V \times V$.

For a ranking $\sigma \in \mathcal{S}_n^V$ and a subset $A \subseteq V$, the *partial ranking* $\sigma^A \in \mathcal{S}_{|A|}^A$ is the bijective function from A to $\{1, \dots, |A|\}$ that preserves the relative order of the items in A under σ. Its inverse, $\bar{\sigma}^A \in \mathcal{S}_A^{|A|}$, is the corresponding *partial ordering*.

2 Relative rank variables

Relative rank variables represent rankings through events in the ranking space. Each realization of such a variable corresponds to a subset of rankings, and together these events form a disjoint cover of the ranking space. Because rankings are permutations, they are subject to mutual exclusivity constraints (MEC), which imply that some combinations of relative ranks cannot occur. In the representation induced by relative rank variables, these constraints appear as empty events. However, carefully designed relative rank variables can break the mutual exclusivity constraints of permutations, so that every realization corresponds to a non-empty event. In this case, the induced events form a partition[1] of the ranking space. Relative rank variables that break MEC will play a central role in this

[1] i.e., a disjoint cover of the space using non-empty subsets

work, as they allow the definition of ranking models and conditional independence relations without parametric constraints.

Relative rank variables are based on the notion of relative rank.

Definition 2.1 (Relative rank)**.** Let $v \in V$ be an item and $P \subseteq V \setminus \{v\}$ be a subset of items called the *reference set*. The *relative rank* of v with respect to P is the mapping $r_P^v : \mathcal{S}_n^V \to \{1, \ldots, |P|+1\}$, defined for $\sigma \in \mathcal{S}_n^V$ as

$$r_P^v(\sigma) := 1 + \sum_{u \in P} \mathbb{1}\big(\sigma(u) < \sigma(v)\big).$$

The value $r_P^v(\sigma) = i$ indicates that item v has the i-th smallest rank among the items in $P \cup \{v\}$ under σ. In particular, $r_P^v(\sigma) = 1$ (resp., $r_P^v(\sigma) = |P|+1$) means that v is ranked ahead of (resp., behind) every item in P under σ. Equivalently, $r_P^v(\sigma) = \sigma^{P\cup\{v\}}(v)$ is the rank of item v in the partial ranking of $P \cup \{v\}$ under σ, $\sigma^{P\cup\{v\}} \in \mathcal{S}_{|P|+1}^{P\cup\{v\}}$. For instance, for $\sigma = (2,1,3)$ we have $\sigma^{\{2,3\}} = (1,2)$ and $r_{\{3\}}^2(\sigma) = 1$, whereas for $\tau = (2,3,1)$ we have $\tau^{\{2,3\}} = (2,1)$ and $r_{\{3\}}^2(\tau) = 2$.

Another useful interpretation of relative ranks is in terms of insertions into partial orderings. Given a ranking σ, the value $r_P^v(\sigma)$ specifies the insertion position of item v into the partial ordering $\bar{\sigma}^P$ to obtain $\bar{\sigma}^{P\cup\{v\}}$. For example, for $\sigma = (2,3,1)$ the relative rank of 1 with respect to $\{2,3\}$ is $r_{\{2,3\}}^1(\sigma) = 2$, meaning that inserting item 1 into $\bar{\sigma}^{\{2,3\}} = [3,2]$ at position 2 yields the ordering $\bar{\sigma} = [3,1,2]$ corresponding to the ranking σ. Relative ranks generalize pairwise comparisons [Wauthier et al., 2013], which correspond to reference sets of size 1.

Definition 2.2 (Relative rank variable)**.** The *relative rank variable* (RRV) for V defined by the reference sets $\{P_v : v \in V\}$ is the n-dimensional categorical random variable

$$\boldsymbol{X} = (X_1, \ldots, X_n), \qquad X_v(\sigma) := r_{P_v}^v(\sigma), \quad \sigma \in \mathcal{S}_n^V,$$

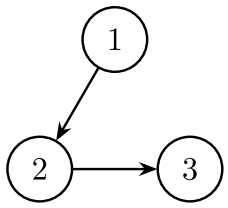
where each component X_v is called the *RRV for* v. The state space of each variable X_v is $\mathcal{X}_v := \{1, \ldots, |P_v|+1\}$, and the joint state space of $\boldsymbol{X}$ is $\boldsymbol{\mathcal{X}} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$.

For a realization $x_v \in \mathcal{X}_v$, the associated event is $\langle x_v \rangle := \{\sigma \in \mathcal{S}_n^V : X_v(\sigma) = x_v\}$, and for a joint realization $\boldsymbol{x} = (x_1, \ldots, x_n) \in \boldsymbol{\mathcal{X}}$, the associated event is

$$\langle \boldsymbol{x} \rangle \;:=\; \{\sigma \in \mathcal{S}_n^V : \boldsymbol{X}(\sigma) = \boldsymbol{x}\} \;=\; \bigcap_{v \in V} \langle x_v \rangle.$$

We denote by $\langle \boldsymbol{\mathcal{X}} \rangle = \{\langle \boldsymbol{x} \rangle : \boldsymbol{x} \in \boldsymbol{\mathcal{X}}\}$ the disjoint cover of $\mathcal{S}_n^V$ induced by $\boldsymbol{X}$.

An RRV for V maps each ranking to a tuple of relative ranks and induces a disjoint cover of the ranking space through the associated events. Such variables are fully determined by the collection of reference sets $\{P_v : v \in V\}$ defining the relative ranks of each item. This collection of reference sets can be naturally represented by a directed graph, where edges encode which items belong to the reference set of each variable. This graphical representation provides a convenient way to specify RRVs for V and analyze the structural properties of the covers they induce over the ranking space.



σ	$\bar{\sigma}$	$\boldsymbol{x}$
1,2,3	1,2,3	1,2,2
1,3,2	1,3,2	1,2,1
2,1,3	2,1,3	1,1,2
2,3,1	3,1,2	1,2,1
3,1,2	2,3,1	1,1,2
3,2,1	3,2,1	1,1,1

Figure 1: The variable graph $\mathcal{V} = \{(1,2),(2,3)\}$ defining $\boldsymbol{X}$ for $V = \{1,2,3\}$. The table shows the mapping of $\boldsymbol{X}$ from rankings $\sigma \in \mathcal{S}_3^V$ to the corresponding realizations $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$. Colors in column $\boldsymbol{x}$ group the rankings belonging to the same event.

Definition 2.3 (Variable graph)**.** Let $\mathcal{V} \subseteq V \times V$ be a directed graph with $(v,v) \notin \mathcal{V}$ for every $v \in V$. For each $v \in V$, define

$$P_v^{\mathcal{V}} := \{\, u \in V \setminus \{v\} : (u,v) \in \mathcal{V} \,\}$$

as the set of parents of node v in $\mathcal{V}$. We say that the RRV $\boldsymbol{X} = (X_1, \ldots, X_n)$ is defined by $\mathcal{V}$ if the reference set of each component X_v is given by $P_v = P_v^{\mathcal{V}}$ for all $v \in V$. The directed graph $\mathcal{V}$ is called the *variable graph* defining $\boldsymbol{X}$.

We denote by $\mathrm{anc}(\mathcal{V}) = \{\bar{\pi} \in \mathcal{S}_V^n : \pi(u) < \pi(v) \text{ for every } (u,v) \in \mathcal{V}\}$ the set of ancestral (or topological) orderings of $\mathcal{V}$, where $\pi = \bar{\pi}^{-1}$.

Figure 1 shows an example of an RRV $\boldsymbol{X} = (X_1, X_2, X_3)$ for V defined by the variable graph $\mathcal{V} = \{(1,2),(2,3)\}$. The reference sets defined by $\mathcal{V}$ are $P_1 = \emptyset$, $P_2 = \{1\}$, and $P_3 = \{2\}$. The realizations of $\boldsymbol{X}$ induce the cover of $\mathcal{S}_n^V$ given by the set of events $\langle \boldsymbol{\mathcal{X}} \rangle = \big\{\{(1,2,3)\}, \{(1,3,2),(2,3,1)\}, \{(2,1,3),(3,1,2)\}, \{(3,2,1)\}\big\}$.

Rankings are subject to mutual exclusivity constraints: distinct items must occupy distinct ranks in a ranking. RRVs defined by arbitrary variable graphs can also be subject to MEC, since certain combinations of relative ranks may lead to contradictions and therefore cannot occur simultaneously in any ranking. In the representation induced by an RRV, these constraints appear as empty events in the cover of the ranking space. This motivates identifying RRVs whose realizations correspond only to non-empty events, effectively removing the constraints imposed by permutations. The following definition formalizes this property.

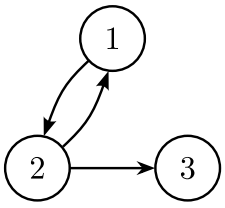
Definition 2.4 (Mutual Exclusivity Constraints)**.** Let $\boldsymbol{X}$ be an RRV for V with state space $\boldsymbol{\mathcal{X}}$. We say that $\boldsymbol{X}$ *is subject to mutual exclusivity constraints* (*MEC*) if there exists a realization $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ such that $\langle \boldsymbol{x} \rangle = \varnothing$. If $\langle \boldsymbol{x} \rangle \neq \varnothing$ for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$, we say that $\boldsymbol{X}$ *breaks MEC*.

Breaking MEC ensures that every realization of the RRV corresponds to a non-empty subset of rankings. Whether this property holds depends on the structure of the reference sets defining the RRV for V, and therefore on the associated variable graph. The following result characterizes when an RRV breaks MEC in terms of the structure of its variable graph.

Theorem 2.5 (Breaking MEC)**.** *Let $\boldsymbol{X}$ be an RRV for V defined by a variable graph $\mathcal{V}$. Then $\boldsymbol{X}$ breaks MEC if and only if $\mathcal{V}$ is acyclic.*

When an RRV breaks MEC, every realization corresponds to a non-empty event, and the induced cover $\langle \boldsymbol{\mathcal{X}} \rangle$ becomes a partition of the ranking space. This property is fundamental for the developments that follow. In particular, RRVs that break MEC allow the definition of probabilistic models and conditional independence relations without structural constraints imposed by permutations.

The acyclicity of the variable graph not only guarantees that RRVs break MEC, but also provides a constructive way to generate the rankings in the event associated with each realization. In Appendix B, we present Algorithm 1 for generating events associated with RRVs defined by acyclic variable graphs. The procedure incrementally constructs partial rankings by inserting elements according to the realization of the RRV, following a topological ordering induced by the variable graph.



σ	$\bar{\sigma}$	$\boldsymbol{x}$
1,2,3	1,2,3	1,2,2
1,3,2	1,3,2	1,2,1
2,1,3	2,1,3	2,1,2
2,3,1	3,1,2	1,2,1
3,1,2	2,3,1	2,1,2
3,2,1	3,2,1	2,1,1

Figure 2: The variable graph $\mathcal{V} = \{(1,2),(2,1),(2,3)\}$ contains the cycle $1,2,1$. The realizations $\boldsymbol{x} \in \{(1,1,1),(1,1,2),(2,2,1),(2,2,2)\}$ also correspond to empty events and do not appear in the table.

Figure 2 shows a cyclic variable graph defining an RRV for V subject to MEC, as predicted by Theorem 2.5. The presence of the cycle $1,2,1$ induces logical consistency constraints between the RRVs for items 1 and 2. As a consequence, the realizations $\boldsymbol{x}_{\{1,2\}} \in \{(1,1),(2,2)\}$ have associated empty events. Thus the joint realizations $\boldsymbol{x} \in \{(1,1,1),(1,1,2),(2,2,1),(2,2,2)\}$ are contradictory and therefore correspond to empty events. For additional examples of variable graphs, see Appendix A.

3 Relative rank models

Once an RRV for V is defined, probabilistic models over rankings can be constructed by assigning probabilities to its realizations. Such models are called relative rank models. In this section, we formally introduce relative rank models and describe their relationship with ranking models through the pushforward mapping.

Definition 3.1 (Relative rank model)**.** Let $\boldsymbol{X}$ be an RRV for V. A *relative rank model* (*RRM*) is a probability distribution

$$h \in \Delta(\boldsymbol{\mathcal{X}}).$$

An RRM assigns probability mass to the realizations of the RRV. Under conditions concerning empty events specified below, this assignment defines a probability distribution over the non-empty events of the disjoint cover $\langle \boldsymbol{\mathcal{X}} \rangle$ induced by $\boldsymbol{X}$.

Next, we connect RRMs with ranking models through the pushforward relation.

Definition 3.2 (Pushforward RRM)**.** Let $\boldsymbol{X}$ be an RRV for V. The pushforward mapping under $\boldsymbol{X}$ is the function $\text{push}_{\boldsymbol{X}} : \Delta(\mathcal{S}_n^V) \to \Delta(\boldsymbol{\mathcal{X}})$, such that $h = \text{push}_{\boldsymbol{X}}(p)$ for any ranking model p, where

$$h(\boldsymbol{x}) := p(\boldsymbol{x}) = \sum_{\sigma \in \langle \boldsymbol{x} \rangle} p(\sigma), \qquad \boldsymbol{x} \in \boldsymbol{\mathcal{X}}. \tag{1}$$

In addition, the *pullback* of h under $\boldsymbol{X}$ is the set of ranking models satisfying Eq. 1, denoted by $\text{pull}(h) \subseteq \Delta(\mathcal{S}_n^V)$.

The pushforward relation can be used to characterize the set of RRMs associated with ranking models.

Definition 3.3 (Admissible RRM)**.** Let $\boldsymbol{X}$ be an RRV for V. We say that an RRM h for $\boldsymbol{X}$ is *admissible* if there exists $p \in \Delta(\mathcal{S}_n^V)$ such that $\text{push}_{\boldsymbol{X}}(p) = h$. Otherwise, it is *inadmissible*.

Admissible RRMs are those that correspond to at least one ranking model through the pullback relation. By construction, any RRM obtained as the pushforward of a ranking model is admissible. However, RRMs can also be defined directly over $\boldsymbol{\mathcal{X}}$, and in this case admissibility is not guaranteed.

Admissibility imposes a structural constraint: an RRM must assign zero probability to all realizations associated with empty events. This condition is automatically satisfied when the underlying RRV breaks MEC, since all realizations correspond to non-empty events. In contrast, when the RRV is subject to MEC, admissibility requires enforcing structural zeros. A detailed characterization is provided in Appendix C.

4 Conditional independence for rankings

Although RRMs reduce the complexity of ranking models by controlling the size of the reference sets P_v for $v \in V$, the number of free parameters of a general RRM is given by

$$\left(\prod_{v \in V} (|P_v| + 1) \right) - 1. \tag{2}$$

As in classical probabilistic modeling, a natural strategy to further reduce model complexity is to impose conditional independence assumptions and factorize the model accordingly. The challenge in the ranking setting is not merely to impose independence, but to define it in a representation that is structurally compatible with rankings. RRVs provide such a representation, allowing conditional independence relations to be defined analogously to multidimensional categorical variables.

Definition 4.1 (Conditional independence)**.** Let $A, B \subseteq V$ be disjoint non-empty subsets, let $C \subseteq V \setminus (A \cup B)$, and let $\boldsymbol{X}$ be an RRV for V with an RRM h. We say that $\boldsymbol{X}_A$ is conditionally independent of $\boldsymbol{X}_B$ given $\boldsymbol{X}_C$ under h, denoted by $\boldsymbol{X}_A \perp_h \boldsymbol{X}_B \mid \boldsymbol{X}_C$, if

$$h(\boldsymbol{x}_{A,B} \mid \boldsymbol{x}_C) = h(\boldsymbol{x}_A \mid \boldsymbol{x}_C) h(\boldsymbol{x}_B \mid \boldsymbol{x}_C), \tag{3}$$

for all $(\boldsymbol{x}_A, \boldsymbol{x}_B, \boldsymbol{x}_C) \in \boldsymbol{\mathcal{X}}_A \times \boldsymbol{\mathcal{X}}_B \times \boldsymbol{\mathcal{X}}_C$ whenever $h(\boldsymbol{x}_C) > 0$. Let $p \in \Delta(\mathcal{S}_n^V)$ be a ranking model. We say that $\boldsymbol{X}_A$ is conditionally independent of $\boldsymbol{X}_B$ given $\boldsymbol{X}_C$ under p, denoted by $\boldsymbol{X}_A \perp_p \boldsymbol{X}_B \mid \boldsymbol{X}_C$, if the pushforward of p under $\boldsymbol{X}$ satisfies Eq. 3. When $C = \varnothing$, these relations reduce to (unconditional) independence, denoted by $\boldsymbol{X}_A \perp_h \boldsymbol{X}_B$ and $\boldsymbol{X}_A \perp_p \boldsymbol{X}_B$, respectively.

In words, conditional independence indicates that knowing a realization of $\boldsymbol{X}_B$ does not provide any additional information about the probabilities of $\boldsymbol{X}_A$ once the realization of $\boldsymbol{X}_C$ is known.

In general, given an ordering of the items $\bar{\pi}$, any RRM can be factorized using the chain rule as $h(\boldsymbol{x}) = \prod_{v \in V} h(x_v \mid \boldsymbol{x}_{<_\pi v})$ where $<_\pi v = \{u \in V : \pi(u) < \pi(v)\}$ denotes the conditioning set of v. Conditional independence relations can simplify this factorization by reducing the conditioning sets. In probabilistic graphical models, Bayesian networks provide a compact representation of such factorizations through a directed graph whose structure encodes conditional independence relations.

5 Relative rank networks

Relative rank networks are graphical models for rankings obtained by combining RRVs with Bayesian network factorizations. In this section, we focus on RRVs defined by acyclic variable graphs, which break MEC and therefore induce partitions of the ranking space without empty events. This restriction avoids the structural zero constraints that arise when RRVs are subject to MEC and allows RRMs to be parameterized freely (see Appendix C). Building on this representation, relative rank networks provide a principled way to represent complex ranking distributions using conditional independence assumptions.

Given an RRV defined by an acyclic variable graph, Eq. 2 shows that an unrestricted RRM can require specifying an exponential number of parameters. Even for sparse variable graphs such as directed trees, this can lead to $\mathcal{O}(2^n)$ parameters. Moreover, probabilistic inference tasks, such as marginalization over subsets of RRVs, are computationally intractable in general. To address these challenges, we introduce graphical models that exploit conditional independence relations among RRVs.

Definition 5.1 (Relative rank network)**.** A *relative rank network* (*RRN*) is a triplet $(\mathcal{V}, \mathcal{G}, \boldsymbol{\theta})$, where $\mathcal{V}$ is an acyclic variable graph defining the RRV $\boldsymbol{X}$ for V, $\mathcal{G}$ is a directed acyclic graph called the factorization graph, and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ denotes the model parameters. The RRN $(\mathcal{V}, \mathcal{G}, \boldsymbol{\theta})$ represents the RRM

$$h(\boldsymbol{x}; \mathcal{G}, \boldsymbol{\theta}) = \prod_{v \in V} h(x_v \mid \boldsymbol{x}_{P_v^{\mathcal{G}}}; \theta_v). \tag{4}$$

An RRN combines two complementary graphical structures. The variable graph $\mathcal{V}$ defines the RRVs and the partition of the ranking space they induce. The factorization graph $\mathcal{G}$ specifies how the joint distribution over these variables factorizes into conditional distributions. In this way, $\mathcal{V}$ determines the representation of rankings, while $\mathcal{G}$ determines the probabilistic structure of the model. We call $(\mathcal{V}, \mathcal{G})$ the *structure* of an RRN.

The number of parameters of an RRN depends on its structure. The graph $\mathcal{V}$ determines the size of the state space of each relative rank variable, while $\mathcal{G}$ determines the conditioning sets in the factorization. The total number of free parameters of an RRN with structure $(\mathcal{V}, \mathcal{G})$ is

$$\sum_{v \in V} |P_v^{\mathcal{V}}| \prod_{u \in P_v^{\mathcal{G}}} \left(|P_u^{\mathcal{V}}| + 1 \right). \tag{5}$$

This flexibility allows the construction of RRNs with controlled complexity by restricting the maximum number of parents in both the variable graph $\mathcal{V}$ and the factorization graph $\mathcal{G}$. In particular, the number of free parameters is at most nk^{s+1}, where $k = \max_{v \in V} |\mathcal{X}_v|$ and $s = \max_{v \in V} |P_v^{\mathcal{G}}|$; it grows linearly in n when k and s are bounded.

For a fixed RRN structure, exact probabilistic inference over RRVs takes $\mathcal{O}(nk^{t+1})$ time [Koller and Friedman, 2009], where $k = \max_{v \in V} |\mathcal{X}_v|$ is the maximum RRV cardinality and t is the treewidth of the moralized factorization graph $\mathcal{G}$. For example, when $\mathcal{G}$ is a directed tree, $t = 1$ and the cost reduces to $\mathcal{O}(nk^2)$.

RRNs can be connected to ranking models through a generalization of the pushforward mapping. Given a ranking model p, the pushforward under $(\mathcal{V}, \mathcal{G})$ defines the parameters of an RRN by computing the conditional probabilities of the RRV realizations from p. This mapping allows us to focus on the structure $(\mathcal{V}, \mathcal{G})$ independently of the specific parameter values.

Definition 5.2 (Factorized pushforward)**.** Let $\boldsymbol{X}$ be an RRV defined by an acyclic variable graph $\mathcal{V}$, and let $\mathcal{G}$ be a factorization graph. The *factorized pushforward* of a ranking model $p \in \Delta(\mathcal{S}_n^V)$ under $(\mathcal{V}, \mathcal{G})$ is the function $\text{push}_{\mathcal{V},\mathcal{G}} : \Delta(\mathcal{S}_n^V) \to \Delta_{\mathcal{G}}(\boldsymbol{\mathcal{X}})$, where $\Delta_{\mathcal{G}}(\boldsymbol{\mathcal{X}}) \subseteq \Delta(\boldsymbol{\mathcal{X}})$ denotes the space of RRNs with structure $(\mathcal{V}, \mathcal{G})$. The RRN $h = \text{push}_{\mathcal{V},\mathcal{G}}(p)$ is defined as

$$h(\boldsymbol{x}) = \prod_{v \in V} h(x_v \mid \boldsymbol{x}_{P_v^{\mathcal{G}}}),$$

where $h(x_v \mid \boldsymbol{x}_{P_v^{\mathcal{G}}}) = \frac{p(x_v, \boldsymbol{x}_{P_v^{\mathcal{G}}})}{p(\boldsymbol{x}_{P_v^{\mathcal{G}}})}$ if $p(\boldsymbol{x}_{P_v^{\mathcal{G}}}) > 0$, and $h(x_v \mid \boldsymbol{x}_{P_v^{\mathcal{G}}}) = \frac{1}{|\mathcal{X}_v|}$ otherwise, for all $v \in V$.

The factorized pushforward mapping of a ranking model under $(\mathcal{V}, \mathcal{G})$ produces an RRM that factorizes according to $\mathcal{G}$. Since in RRNs the RRV for V breaks MEC, the factorized pushforward mapping yields admissible models, and any RRM can be specified without additional structural constraints. In Appendix D, we discuss the admissibility of factorized models over RRVs subject to MEC and the structural zero constraints they impose.

For observed rankings and a fixed variable graph $\mathcal{V}$, learning $\mathcal{G}$ reduces to Bayesian-network structure learning over the categorical realizations $\boldsymbol{X}(\sigma)$, and the local distributions can be estimated from empirical frequencies. Appendix E outlines this procedure and the additional selection of $\mathcal{V}$ in Algorithm 2.

The structure of an RRN admits a graphical interpretation in terms of conditional independence. The factorization graph $\mathcal{G}$ defines a Bayesian network over the RRV $\boldsymbol{X} = (X_1, \ldots, X_n)$ defined by $\mathcal{V}$, and therefore encodes conditional independence relations among its components X_v for $v \in V$. As in standard Bayesian networks, these relations can be derived from $\mathcal{G}$ using the d-separation criterion [Pearl, 1988].

Definition 5.3 (Independence model)**.** Let $(\mathcal{V}, \mathcal{G})$ be the structure of an RRN. The *independence model* induced by $(\mathcal{V}, \mathcal{G})$ is the set of conditional independence relations

$$\text{ind}(\mathcal{V}, \mathcal{G}) = \{\, \boldsymbol{X}_A \perp \boldsymbol{X}_B \mid \boldsymbol{X}_C : A, B, C \subseteq V, \boldsymbol{X}_A \perp_{\mathcal{G}} \boldsymbol{X}_B \mid \boldsymbol{X}_C\},$$

where $\boldsymbol{X}$ is the RRV defined by $\mathcal{V}$, the subsets $A, B, C \subseteq V$ are pairwise disjoint with $A, B \neq \varnothing$, and the relation $\boldsymbol{X}_A \perp_{\mathcal{G}} \boldsymbol{X}_B \mid \boldsymbol{X}_C$ indicates that A is d-separated from B by C in $\mathcal{G}$.

Given the structure $(\mathcal{V}, \mathcal{G})$, the induced independence model is a compositional graphoid satisfying the local Markov property [Studený, 2005].

Any RRN with structure $(\mathcal{V}, \mathcal{G})$ satisfies the conditional independence relations in $\text{ind}(\mathcal{V}, \mathcal{G})$.

Lemma 5.4. *Let $h \in \Delta_{\mathcal{G}}(\boldsymbol{\mathcal{X}})$ be an RRN with structure $(\mathcal{V}, \mathcal{G})$, where $\boldsymbol{X}$ is the RRV defined by $\mathcal{V}$, and let $p \in \text{pull}(h)$. Then h and p satisfy the conditional independence relations in $\text{ind}(\mathcal{V}, \mathcal{G})$, that is,*

$$\boldsymbol{X}_A \perp_{\mathcal{G}} \boldsymbol{X}_B \mid \boldsymbol{X}_C \implies \{\boldsymbol{X}_A \perp_h \boldsymbol{X}_B \mid \boldsymbol{X}_C, \quad \boldsymbol{X}_A \perp_p \boldsymbol{X}_B \mid \boldsymbol{X}_C\}.$$

6 Representing ranking models and independence assumptions

This section has two goals. First, we show how RRMs can be used to represent ranking models in general. Second, we show that two prominent families of ranking models based on alternative independence notions, namely repeated insertion and riffle shuffle models, can be represented within the RRN framework. In this way, the independence assumptions underlying both models correspond to specific independence models over RRVs.

6.1 Ranking models based on RRMs

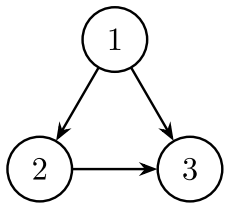
An RRV $\boldsymbol{X}$ defined by an acyclic variable graph induces a partition of the ranking space, and RRMs for $\boldsymbol{X}$ represent probability distributions over this partition. A particularly useful situation occurs when an RRV induces a *thin partition* of the ranking space, that is, a partition whose cells contain a single ranking. In this case, the mapping induced by the RRV from the ranking space $\mathcal{S}_n^V$ to the state space $\boldsymbol{\mathcal{X}}$ is bijective, and therefore RRMs for $\boldsymbol{X}$ are ranking models.

For a general acyclic variable graph, an RRM specifies the probabilities of the events $\langle \boldsymbol{x} \rangle$ but does not determine how probability is distributed among rankings within each event. A ranking model can be obtained by specifying a distribution within each event, as described in Appendix C. For example, assigning equal probability to the rankings within each event yields $p(\sigma) = h(\boldsymbol{X}(\sigma))/|\langle \boldsymbol{X}(\sigma) \rangle|$. The cardinality of the corresponding event can be obtained by enumerating its rankings using Algorithm 1.

The following result shows that an RRV defined by a complete acyclic variable graph induces a thin partition of the ranking space.

Proposition 6.1 (Thin partition)**.** *Let $\boldsymbol{X}$ be an RRV defined by $\mathcal{V}$. Then, $\boldsymbol{X}$ induces a* thin partition *of $\mathcal{S}_n^V$ if and only if $\mathcal{V}$ is a complete acyclic variable graph.*

Intuitively, a complete acyclic variable graph has a unique ancestral ordering $\bar{\pi}$, and the reference set defining the RRV for $v = \bar{\pi}(i)$ is $P_v = \{\bar{\pi}(1), \ldots, \bar{\pi}(i-1)\}$ for $i = 1, \ldots, n$. Thus, a realization $\boldsymbol{x}$ encodes the position of each item with respect to all previously considered items in $\bar{\pi}$, yielding a representation equivalent to the permutation itself. Figure 3 illustrates this correspondence for $n = 3$, where the ancestral ordering is $\bar{\pi} = [1, 2, 3]$.



σ	$\bar{\sigma}$	$\boldsymbol{x}$
1,2,3	1,2,3	1,2,3
1,3,2	1,3,2	1,2,2
2,1,3	2,1,3	1,1,3
2,3,1	3,1,2	1,2,1
3,1,2	2,3,1	1,1,2
3,2,1	3,2,1	1,1,1

Figure 3: The complete acyclic variable graph $\mathcal{V} = \{(1,2),(2,3),(1,3)\}$, where $\bar{\pi} = [1,2,3]$ is its unique ancestral ordering. The RRV defined by $\mathcal{V}$ induces a thin partition of the ranking space.

An immediate consequence is that RRVs for V defined by complete acyclic variable graphs represent bijective mappings between the state space and the ranking space, and therefore ranking models can be represented by RRMs. In the literature, RRVs providing bijective representations of permutations are known as *inversion vectors* [Malagon et al., 2025], and they form the basis of repeated insertion models [Doignon et al., 2004]. According to Proposition 6.1, there are $n!$ different RRVs corresponding to inversion vectors.

In the remainder of this section, we use this representation to relate RRNs with two prominent families of ranking models based on independence notions tailored to permutations.

6.2 Repeated insertion model

Repeated insertion (RI) independence is a probabilistic independence notion defined with respect to a fixed ordering $\bar{\pi}$ of the items, called the *insertion ordering* [Doignon et al., 2004]. A ranking model p satisfies RI independence for $\bar{\pi}$ [Kenig et al., 2018] if it factorizes as

$$p(\sigma; \bar{\pi}) = \prod_{i=1}^{n} p_i(f_{\sigma,\bar{\pi}}(i)), \tag{6}$$

for all $\sigma \in \mathcal{S}_n^V$, where p_i is the insertion distribution of the item $\bar{\pi}(i)$ and

$$f_{\sigma,\bar{\pi}}(i) = 1 + \sum_{j=1}^{i-1} \mathbf{1}\{\sigma(\bar{\pi}(j)) < \sigma(\bar{\pi}(i))\}.$$

The function $f_{\sigma,\bar{\pi}}(i)$ determines the insertion position of the item $\bar{\pi}(i)$ in the partial ranking $\sigma^{<_{\bar{\pi}} i}$, where $<_{\bar{\pi}} i = \{\bar{\pi}(1), \ldots, \bar{\pi}(i-1)\}$. We call a *repeated insertion model* (*RIM*) for $\bar{\pi}$ any ranking model that factorizes according to Eq. 6, and we denote *the space of RIMs for* $\bar{\pi}$ by $\Delta_{\bar{\pi}}(\mathcal{S}_n^V)$.

For a fixed insertion ordering and arbitrary categorical insertion distributions, a RIM has $n(n-1)/2$ free parameters.

From a generative perspective, a RIM for $\bar{\pi}$ constructs rankings by sequentially inserting items according to the ordering $\bar{\pi}$. Starting from the trivial ranking over $\{\bar{\pi}(1)\}$, (1), each item $\bar{\pi}(i)$ is inserted into the current partial ranking at a position sampled from its insertion distribution p_i. After $n-1$ insertion steps, this procedure produces a full ranking for V. This generative process is a particular case of Algorithm 1 in Appendix B.

Theorem 6.2. *Let $\bar{\pi} \in \mathcal{S}_V^n$ be an insertion ordering. Let $\mathcal{V}^{RI}$ be the complete acyclic variable graph with ancestral ordering $\bar{\pi}$, and let $\mathcal{G}^{RI} = \varnothing$ be the empty factorization graph. Let $\boldsymbol{X}$ be an RRV defined by $\mathcal{V}^{RI}$. Then,*

$$\Delta_{\bar{\pi}}(\mathcal{S}_n^V) = \Delta_{\mathcal{G}^{RI}}(\boldsymbol{\mathcal{X}}).$$

As a consequence, the independence notion underlying RIMs can be expressed entirely in terms of conditional independence relations among RRVs.

Corollary 6.3. *Let $\bar{\pi} \in \mathcal{S}_V^n$ be the insertion ordering, and let $\mathcal{V}^{RI}$ and $\mathcal{G}^{RI}$ be the graphs defined in Theorem 6.2. Then, RI independence for $\bar{\pi}$ corresponds to the independence model ind$(\mathcal{V}^{RI}, \mathcal{G}^{RI})$.*

In particular, since the factorization graph $\mathcal{G}^{RI}$ is empty, the independence model ind$(\mathcal{V}^{RI}, \mathcal{G}^{RI})$ contains all conditional independence relations $\boldsymbol{X}_A \perp \boldsymbol{X}_B \mid \boldsymbol{X}_C$ for disjoint non-empty subsets $A, B \subset V$ and $C \subseteq V \setminus (A \cup B)$.

An additional consequence of Theorem 6.2 is that a RIM for $\bar{\pi}$ can be constructed from any ranking model using the factorized pushforward mapping.

Corollary 6.4. *Let $\bar{\pi} \in \mathcal{S}_V^n$ be the insertion ordering, and let $\mathcal{V}^{RI}$ and $\mathcal{G}^{RI}$ be the graphs defined in Theorem 6.2. Let $p \in \Delta(\mathcal{S}_n^V)$ be a ranking model. Then, $h = push_{\mathcal{V}^{RI},\mathcal{G}^{RI}}(p)$ is the RIM for $\bar{\pi}$ that minimizes the KL divergence with respect to p, that is, $h = \arg\min_{q \in \Delta_{\bar{\pi}}(\mathcal{S}_n^V)} D_{KL}(p||q)$.*

6.3 Riffle shuffle model

Riffle shuffle (RS) independence [Huang and Guestrin, 2009, 2010, Huang et al., 2012] models rankings by independently ranking disjoint subsets of items and then interleaving the resulting partial rankings. The name is inspired by the riffle shuffle used in card games. Let $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_{n-m}\}$ be a partition of V. A ranking model p satisfies RS independence for A and B if it factorizes as

$$p(\sigma) = p_A(\sigma^A) p_A^B(s_A^B(\sigma)) p_B(\sigma^B), \tag{7}$$

for every $\sigma \in \mathcal{S}_n^V$, where $\sigma^A \in \mathcal{S}_{|A|}^A$ and $\sigma^B \in \mathcal{S}_{|B|}^B$ denote the (partial) rankings of the items in A and B induced by σ (see Definition 12 in [Huang and Guestrin, 2012]). The distributions $p_A \in \Delta(\mathcal{S}_{|A|}^A)$ and $p_B \in \Delta(\mathcal{S}_{|B|}^B)$ model the partial rankings within each subset of items. The function $s_A^B : \mathcal{S}_n^V \to \binom{N}{|B|}$ determines how the positions of the items in A and B are interleaved in σ, and the distribution p_A^B models the interleavings. For an interleaving $\boldsymbol{s}$, $\boldsymbol{s}(i)$ denotes its i-th smallest element. Given the interleaving $\boldsymbol{s} = s_A^B(\sigma)$ we have $\sigma(v) = \boldsymbol{s}(\sigma^B(v))$, for $v \in B$. We call a *riffle shuffle model* (*RSM*) for A and B any ranking model that factorizes according to Eq. 7, and we denote *the space of RSMs for A and B* by $\Delta_{A,B}(\mathcal{S}_n^V)$.

An RSM for a fixed partition A, B, with arbitrary partial ranking and interleaving distributions, has $|A|! + |B|! + \binom{n}{|B|} - 3$ free parameters: $|A|! - 1$ for p_A, $|B|! - 1$ for p_B, and $\binom{n}{|B|} - 1$ for p_A^B.

From a generative perspective, an RSM constructs rankings in two stages. First, the items in A and B are ranked independently according to the distributions p_A and p_B, producing the rankings σ^A and σ^B. Then, the two rankings are interleaved by sampling an interleaving pattern from the distribution p_A^B, which determines the set of positions occupied by the items of each subset. The interleaving pattern is selected independently of σ^A and σ^B.

We now show that RSMs can also be represented within the RRN framework by combining appropriate choices of variable graphs and a nontrivial factorization graph. In this representation, two RRN structures are considered, each defined by a different variable graph but sharing the same factorization graph. A ranking model that factorizes according to both structures corresponds to an RSM for A and B.

Theorem 6.5. *Let $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_{n-m}\}$ form a partition of V. Let $\mathcal{V}^{RS}$ and $\mathcal{W}^{RS}$ be complete acyclic variable graphs with ancestral orderings $\bar{\pi} = [a_1, \ldots, a_m, b_1, \ldots, b_{n-m}]$ and $\bar{\tau} = [b_1, \ldots, b_{n-m}, a_1, \ldots, a_m]$, respectively. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be the RRVs defined by $\mathcal{V}^{RS}$ and $\mathcal{W}^{RS}$, respectively. Let $\mathcal{G}^{RS} = \mathcal{G}_A \cup \mathcal{G}_B$ be a directed acyclic graph, where $\mathcal{G}_A$ and $\mathcal{G}_B$ are complete subgraphs over the vertex sets A and B, respectively. Then,*

$$\Delta_{A,B}(\mathcal{S}_n^V) = \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{X}}) \cap \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{Y}}).$$

RIMs correspond to a particularly simple instance of RRNs, characterized by a complete variable graph and an empty factorization graph. RSMs, in contrast, rely on a richer independence structure. In the RRN representation, this structure is captured by a nontrivial factorization graph over two RRVs for V that provide two different relative-rank encodings of the ranking space, both inducing the same thin partition.

According to Theorem 6.5, the space of RSMs is given by the intersection of the spaces of two RRNs. Thus, RS independence corresponds to the independence models specified by the structures of the two RRNs.

Corollary 6.6. *Let A and B be a partition of V, and let $\mathcal{V}^{RS}$, $\mathcal{W}^{RS}$, and $\mathcal{G}^{RS}$ be the graphs defined in Theorem 6.5. Then, RS independence for A and B corresponds to ind$(\mathcal{V}^{RS}, \mathcal{G}^{RS}) \cup$ ind$(\mathcal{W}^{RS}, \mathcal{G}^{RS})$.*

Finally, the next result shows how to obtain an RSM for A and B from any ranking model using the factorized pushforward mapping.

Proposition 6.7. *Let p be a ranking model, let A and B be a partition of V, and let $\mathcal{V}^{RS}$, $\mathcal{W}^{RS}$, and $\mathcal{G}^{RS}$ be the graphs defined in Theorem 6.5. Then, $h = push_{\mathcal{V}^{RS},\mathcal{G}^{RS}}(push_{\mathcal{W}^{RS},\mathcal{G}^{RS}}(p))$ is the RSM for A and B that minimizes the Kullback-Leibler divergence with respect to p.*

Together, these results show that RIMs and RSMs can be represented within the RRN framework, each corresponding to a specific independence model over relative rank variables.

7 Conclusions

This theoretical work introduces *relative rank models* as a probabilistic framework for rankings based on structured coarsenings of the permutation space. By representing rankings through relative rank variables, the framework replaces the structure of permutations with categorical random variables that can break the mutual exclusivity constraints of rankings.

A central result of the paper is the characterization of acyclic variable graphs as the exact condition under which relative rank variables break mutual exclusivity constraints. When mutual exclusivity constraints are broken, admissibility imposes no further structural constraints on relative rank models, and relative rank models can be simplified by assuming conditional independence among relative rank variables.

Building on this characterization, we define *relative rank networks*, which combine relative rank variables with Bayesian network factorizations. Relative rank networks provide a principled way to encode conditional independence relations among relative rank variables while retaining a clear semantics in terms of partitions of the ranking space. As a consequence, they enable the application of standard tools from probabilistic graphical models to ranking data.

We further show that well-known ranking models based on specialized independence notions, namely repeated insertion and riffle shuffle models, can be represented within the framework of relative rank networks. From this perspective, these models correspond to specific sets of conditional independences over relative rank variables rather than fundamentally different probabilistic constructions.

More generally, the results of this paper suggest a shift in how independence should be introduced in ranking models: instead of defining independence notions directly on permutations, one should first construct a structured coarsening that breaks mutual exclusivity constraints and then apply standard probabilistic independence relations. Relative rank networks provide a concrete and flexible realization of this principle.

Acknowledgments and Disclosure of Funding

The author thanks Ekhiñe Irurozki for helpful discussions. This work was supported by MCIN/AEI/10.13039/501100011033 under projects PLEC2024-011247 and CEX2021-001142-S, and by the Basque Government through the BERC 2022–2025 and Elkartek programs.

References

M. Achab, A. Korba, and S. Clémençon. Dimensionality reduction and (bucket) ranking: a mass transportation approach. In *International Conference on Algorithmic Learning Theory*, volume 98, pages 64–93, 2019.

F. Chierichetti, A. Dasgupta, S. Haddadan, R. Kumar, and S. Lattanzi. Mallows models for top-k lists. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

A. P. Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 41(1):1–15, 1979.

J. P. Doignon, A. Pekeč, and M. Regenwetter. The repeated insertion model for rankings: Missing link between two subset choice models. *Psychometrika*, 69(1):33–54, 2004.

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

J. Huang and C. Guestrin. Riffled independence for ranked data. In *Advances in Neural Information Processing Systems*, volume 22, 2009.

J. Huang and C. Guestrin. Learning hierarchical riffle independent groupings from rankings. In *International Conference on Machine Learning*, 2010.

J. Huang and C. Guestrin. Uncovering the riffled independence structure of ranked data. *Electronic Journal of Statistics*, 6:199–230, 2012.

J. Huang, C. Guestrin, and L. Guibas. Fourier theoretic probabilistic inference over permutations. *Journal of Machine Learning Research*, 10(37):997–1070, 2009a.

J. Huang, C. Guestrin, X. Jiang, and L. Guibas. Exploiting probabilistic independence for permutations. In *Artificial Intelligence and Statistics*, pages 248–255. PMLR, 2009b.

J. Huang, A. Kapoor, and C. Guestrin. Riffled independence for efficient inference with partial rankings. *Journal of Artificial Intelligence Research*, 44:491–532, 2012.

B. Kenig, L. Ilijasić, H. Ping, B. Kimelfeld, and J. Stoyanovich. Probabilistic inference over repeated insertion models. In *AAAI Conference on Artificial Intelligence*, 2018.

D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, Cambridge, Massachusetts, 2009.

Y. Lan, S. Niu, J. Guo, and X. Cheng. Is top-k sufficient for ranking? In *ACM International Conference on Information and Knowledge Management*, pages 1261–1270, 2013.

S. L. Lauritzen. *Graphical Models*. Oxford University Press, New York, 1996.

G. Lebanon and Y. Mao. Non-parametric modeling of partially ranked data. In *Advances in Neural Information Processing Systems*, volume 20, pages 857–864, 2007.

T. Lu and C. Boutilier. Learning Mallows models with pairwise preferences. In *International Conference on Machine Learning*, pages 145–152, 2011.

T. Lu and C. Boutilier. Effective sampling and learning for Mallows models with pairwise-preference data. *Journal of Machine Learning Research*, 15(117):3963–4009, 2014.

M. Malagon, A. Barrena, H. Iñigo, E. Irurozki, J. A. Lozano, and J. Ceberio. A unified view of bijective transformations for optimizing permutation problems. In *European Conference on Artificial Intelligence*, pages 4750–4757, 2025.

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

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, California, 1988.

M. Regenwetter, B. Grofman, A. A. J. Marley, and I. Tsetlin. *Behavioral Social Choice: Probabilistic Models, Statistical Inference, and Applications*. Cambridge University Press, 2006.

M. Studený. *Probabilistic Conditional Independence Structures*. Springer, 2005.

F. Wauthier, M. Jordan, and N. Jojic. Efficient ranking from pairwise comparisons. In *International Conference on Machine Learning*, pages 109–117. PMLR, 2013.

A Examples of variable graphs and the induced partition

Figure 4 shows five variable graphs with the vertex set $V = \{1, 2, 3\}$. Each variable graph $\mathcal{V}$ defines a collection of reference sets P_v for $v \in V$, where P_v corresponds to the set of parents of v in $\mathcal{V}$. The associated RRV $\boldsymbol{X} = (X_1, X_2, X_3)$ for V is defined by

$$X_v(\sigma) := r^v_{P_v}(\sigma), \qquad v \in V.$$

The variable graphs and the reference sets that define the corresponding RRVs are:

- $\mathcal{V}_1 = \{(2,1),(2,3)\}$: $P_1 = \{2\}$, $P_2 = \emptyset$, $P_3 = \{2\}$.
- $\mathcal{V}_2 = \{(1,2),(3,2)\}$: $P_1 = \emptyset$, $P_2 = \{1,3\}$, $P_3 = \emptyset$.
- $\mathcal{V}_3 = \{(1,2),(1,3)\}$: $P_1 = \emptyset$, $P_2 = \{1\}$, $P_3 = \{1\}$.
- $\mathcal{V}_4 = \{(1,2),(2,3),(3,1)\}$: $P_1 = \{3\}$, $P_2 = \{1\}$, $P_3 = \{2\}$.
- $\mathcal{V}_5 = \{(2,1),(3,1),(3,2)\}$: $P_1 = \{2,3\}$, $P_2 = \{3\}$, $P_3 = \emptyset$.

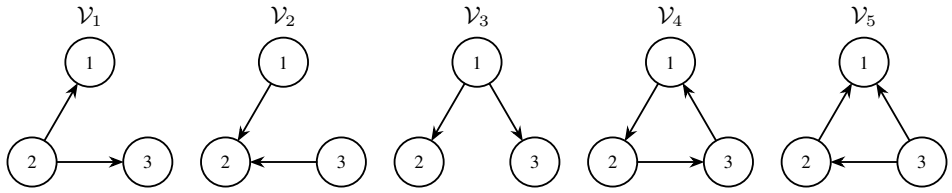


Figure 4: Five variable graphs defined over items $V = \{1, 2, 3\}$.

Table 1 shows the relation between rankings $\sigma \in \mathcal{S}_n^V$ (and orderings $\bar{\sigma} \in \mathcal{S}_V^n$) and the realizations $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ of the RRVs defined by these graphs. For each variable graph, the table displays the cover $\langle \boldsymbol{\mathcal{X}} \rangle$ of the ranking space $\mathcal{S}_n^V$, where each realization $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ and its corresponding event $\langle \boldsymbol{x} \rangle$ is identified by a distinct color. For example, for $\mathcal{V}_1$ and realization $\boldsymbol{x} = (2, 1, 2)$, the event $\langle \boldsymbol{x} \rangle = \{(2,1,3),(3,1,2)\}$ is highlighted in blue.

		$\mathcal{V}_1$	$\mathcal{V}_2$	$\mathcal{V}_3$	$\mathcal{V}_4$	$\mathcal{V}_5$
σ	$\bar{\sigma}$	$\boldsymbol{x} = (x_1, x_2, x_3)$				
1,2,3	1,2,3	1,1,2	1,2,1	1,2,2	1,2,2	1,1,1
1,3,2	1,3,2	1,1,1	1,3,1	1,2,2	1,2,1	1,2,1
2,1,3	2,1,3	2,1,2	1,1,1	1,1,2	1,1,2	2,1,1
2,3,1	3,1,2	1,1,1	1,3,1	1,2,1	2,2,1	2,2,1
3,1,2	2,3,1	2,1,2	1,1,1	1,1,1	2,1,2	3,1,1
3,2,1	3,2,1	2,1,1	1,2,1	1,1,1	2,1,1	3,2,1

Table 1: Relative rank variable realizations $\boldsymbol{x} = (x_1, x_2, x_3)$ for the rankings $\sigma \in \mathcal{S}_n^V$ and orderings $\bar{\sigma} \in \mathcal{S}_V^n$, defined according to the five variable graphs shown in Figure 4. The graph $\mathcal{V}_4$ is cyclic and the realizations $\boldsymbol{x} \in \{(1,1,1),(2,2,2)\}$ do not appear in the table because they correspond to empty events.

The graph $\mathcal{V}_4$ is cyclic and therefore its associated RRV is subject to MEC, which implies that some realizations of the associated RRV correspond to empty events. In particular, the cyclic dependencies among the variables impose mutually inconsistent constraints on the relative ranks, so that some configurations cannot be satisfied by any ranking. As a consequence, the realizations $\boldsymbol{x} = (1, 1, 1)$ and $\boldsymbol{x} = (2, 2, 2)$ correspond to empty events, i.e., $\langle \boldsymbol{x} \rangle = \varnothing$ (see Theorem 2.5). This prevents $\mathcal{V}_4$ from inducing a proper partition of the ranking space.

The graph $\mathcal{V}_5$ is a complete acyclic variable graph and therefore induces a thin partition of the ranking space. By Proposition 6.1, every realization $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ corresponds to a singleton event, so that each ranking $\sigma \in \mathcal{S}_n^V$ is uniquely identified by its image $\boldsymbol{x} = \boldsymbol{X}(\sigma)$. As a consequence, the RRVs defined by complete acyclic variable graphs define a bijection between the ranking space and their state space, and the RRMs are ranking models, $\Delta(\boldsymbol{\mathcal{X}}) = \Delta(\mathcal{S}_n^V)$. In particular, any ranking model p can be represented exactly as an RRM $h = \text{push}(p)$, satisfying $p(\sigma) = h(\boldsymbol{X}(\sigma))$ for all $\sigma \in \mathcal{S}_n^V$, and $p(\boldsymbol{x}) = \sum_{\sigma \in \langle \boldsymbol{x} \rangle} p(\sigma) = h(\boldsymbol{x})$ for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$.

B Generating the events

Thanks to the acyclicity of the variable graph, all the rankings belonging to a given event can be enumerated using Algorithm 1. The algorithm is an insertion-based procedure that follows an ancestral ordering of the variable graph to incrementally construct a set of partial rankings. At each step, a new item is inserted into each ranking from the previous iteration at positions determined by its associated relative rank.

Algorithm 1 Event construction from an RRV realization

Input: A realization $\boldsymbol{x}$ of an RRV for V, $\boldsymbol{X}$, defined by an acyclic variable graph $\mathcal{V}$ with an ancestral ordering $\bar{\pi} \in \text{anc}(\mathcal{V})$

Output: The non-empty event associated to $\boldsymbol{x}$, $\langle \boldsymbol{x} \rangle = \mathcal{R}_n$

Choose an ordering $\bar{\pi} \in \text{anc}(\mathcal{V})$
Initialize $\mathcal{R}_1 \leftarrow \{(1)\} = \mathcal{S}_1^{\{\bar{\pi}(1)\}}$
for $i = 2, \ldots, n$ **do**
  Initialize $\mathcal{R}_i \leftarrow \varnothing$
  Let $v \leftarrow \bar{\pi}(i)$
  for each $\alpha \in \mathcal{R}_{i-1}$ **do**
    if $P_v^{\mathcal{V}} = \varnothing$ **then**
      $I_v \leftarrow \{1, \ldots, i\}$
    else if $x_v = 1$ **then**
      $I_v \leftarrow \{1, \ldots, \alpha(\bar{\alpha}^{P_v^{\mathcal{V}}}(1))\}$
    else if $x_v \in \{2, \ldots, |P_v^{\mathcal{V}}|\}$ **then**
      $I_v \leftarrow \{\alpha(\bar{\alpha}^{P_v^{\mathcal{V}}}(x_v - 1)) + 1, \ldots, \alpha(\bar{\alpha}^{P_v^{\mathcal{V}}}(x_v))\}$
    else
      $I_v \leftarrow \{\alpha(\bar{\alpha}^{P_v^{\mathcal{V}}}(|P_v^{\mathcal{V}}|)) + 1, \ldots, i\}$
    end if
    $\mathcal{R}_i \leftarrow \mathcal{R}_i \cup \{\text{insert}(v, r, \alpha) : r \in I_v\}$
  end for
end for
return $\mathcal{R}_n$

Definition B.1 (Insertion)**.** Let $U \subsetneq V$ with $m = |U|$, let $\alpha \in \mathcal{S}_m^U$ be a (partial) ranking on U, let $v \in V \setminus U$, and let $r \in \{1, \ldots, m+1\}$. The *insertion operator*

$$\text{insert}(v, r, \alpha) = \beta \in \mathcal{S}_{m+1}^{U \cup \{v\}}$$

is defined by

$$\beta(v) = r, \text{ and } \beta(u) = \alpha(u) + \mathbf{1}\{\alpha(u) \geq r\} \text{ for } u \in U.$$

The resulting ranking on $U \cup \{v\}$, β, corresponds to the ordering

$$\bar{\beta} = [\bar{\alpha}(1), \ldots, \bar{\alpha}(r-1), v, \bar{\alpha}(r), \ldots, \bar{\alpha}(m)].$$

We consider the direct extensions of Definitions 2.1 and 2.2 to partial rankings containing the item and its reference set.

Lemma B.2 (Relative rank preservation)**.** *Consider the set of items $U \subsetneq V$ with size $m = |U|$, the (partial) ranking $\alpha \in \mathcal{S}_m^U$, the item $v \in V \setminus U$, the rank $r \in \{1, \ldots, m+1\}$, and the ranking $\beta = \textit{insert}(v, r, \alpha) \in \mathcal{S}_{m+1}^{U \cup \{v\}}$. Then, for any $u \in U$ and any $P \subseteq U \setminus \{u\}$,*

$$r_P^u(\beta) = r_P^u(\alpha).$$

Proof. Fix $u \in U$ and $P \subseteq U \setminus \{u\}$. Define the sets

$$A = \{w \in P : \alpha(w) < r\}, \qquad B = \{w \in P : r \leq \alpha(w)\}.$$

By Definition B.1, we have:

- if $w \in A$, then $\beta(w) = \alpha(w)$,

- if $w \in B$, then $\beta(w) = \alpha(w) + 1$.

We consider two cases for $\alpha(u)$.

i) $\alpha(u) < r$:

Then $\beta(u) = \alpha(u)$, and

$$r_P^u(\beta) = 1 + \sum_{w \in A} \mathbf{1}\{\beta(w) < \beta(u)\} + \sum_{w \in B} \mathbf{1}\{\beta(w) < \beta(u)\}.$$

Substituting the definition of β,

$$r_P^u(\beta) = 1 + \sum_{w \in A} \mathbf{1}\{\alpha(w) < \alpha(u)\} + \sum_{w \in B} \mathbf{1}\{\alpha(w) + 1 < \alpha(u)\}.$$

Since $\alpha(u) < r \leq \alpha(w)$ for all $w \in B$, we have

$$r_P^u(\beta) = 1 + \sum_{w \in A} \mathbf{1}\{\alpha(w) < \alpha(u)\} + \sum_{w \in B} \mathbf{1}\{\alpha(w) < \alpha(u)\} = r_P^u(\alpha).$$

ii) $\alpha(u) \geq r$:

Then $\beta(u) = \alpha(u) + 1$, and

$$r_P^u(\beta) = 1 + \sum_{w \in A} \mathbf{1}\{\beta(w) < \beta(u)\} + \sum_{w \in B} \mathbf{1}\{\beta(w) < \beta(u)\}.$$

Substituting,

$$r_P^u(\beta) = 1 + \sum_{w \in A} \mathbf{1}\{\alpha(w) < \alpha(u) + 1\} + \sum_{w \in B} \mathbf{1}\{\alpha(w) + 1 < \alpha(u) + 1\}.$$

Since $\alpha(w) < r \leq \alpha(u)$ for all $w \in A$, we have

$$r_P^u(\beta) = 1 + \sum_{w \in A} \mathbf{1}\{\alpha(w) < \alpha(u)\} + \sum_{w \in B} \mathbf{1}\{\alpha(w) < \alpha(u)\} = r_P^u(\alpha).$$

□

The next result shows that for any $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ Algorithm 1 generates all the rankings in the event $\langle \boldsymbol{x} \rangle$.

Proposition B.3. *Let $\boldsymbol{X}$ be an RRV defined by an acyclic variable graph $\mathcal{V}$. Then, given a realization $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$, Algorithm 1 generates the rankings in the event $\langle \boldsymbol{x} \rangle \neq \emptyset$.*

Proof. Let $\boldsymbol{X}$ be an RRV defined by $\mathcal{V}$ with the reference sets $P_v = P_v^{\mathcal{V}}$ for $v = 1, \ldots, n$. We have to prove that $\langle \boldsymbol{x} \rangle = \mathcal{R}_n$ for arbitrary $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$.

$\mathcal{V}$ is a DAG and thus there exists $\bar{\pi} \in \text{anc}(\mathcal{V})$. Without loss of generality, assume that $\bar{\pi} = [1, \ldots, n]$. By definition of an ancestral ordering, the reference set $P_v \subseteq \{\bar{\pi}(1), \ldots, \bar{\pi}(v-1)\}$, for $v = 1, \ldots, n$.

We prove the result by induction over $i = 1, \ldots, n$: for $v = \bar{\pi}(i) = i$, the event $\langle \boldsymbol{x}_{\{1,\ldots,v\}} \rangle = \bigcup_{\alpha \in \mathcal{R}_i} \langle \alpha \rangle$, where $\langle \alpha \rangle = \{\sigma : \sigma^{\{1,\ldots,v\}} = \alpha\}$ and $|\mathcal{R}_i| \geq 1$.

Base case ($i = v = 1$).
The parents of v in $\mathcal{V}$ are $P_1 = \emptyset$. Then $\mathcal{X}_v = \{1\}$ and thus $x_v = 1$ with $\langle x_v \rangle = \mathcal{S}_n^V$.

Using Algorithm 1, we have $\mathcal{R}_1 = \{(1)\} = \mathcal{S}_1^{\{1\}}$, and $\bigcup_{\alpha \in \mathcal{R}_1} \langle \alpha \rangle = \langle (1) \rangle = \mathcal{S}_n^V = \langle x_v \rangle$.

Inductive step.
For $1 \leq i < n$, let us assume that $\langle \boldsymbol{x}_{\{1,\ldots,i\}} \rangle = \bigcup_{\alpha \in \mathcal{R}_i} \langle \alpha \rangle$ where $\mathcal{R}_i \subseteq \mathcal{S}_i^{\{1,\ldots,i\}}$ with $|\mathcal{R}_i| \geq 1$. Let $v = \bar{\pi}(i+1) = i + 1$.

The ancestral ordering is $\bar{\pi} = [1, \ldots, n]$ and $P_v \subseteq \{1, \ldots, i\}$. By Definition 2.2 we have $\langle \boldsymbol{x}_{\{1,\ldots,v\}} \rangle = \langle x_v \rangle \cap \langle \boldsymbol{x}_{\{1,\ldots,i\}} \rangle$ and thus

$$\langle \boldsymbol{x}_{\{1,\ldots,v\}} \rangle = \bigcup_{\alpha \in \mathcal{R}_i} \langle \alpha \rangle \cap \langle x_v \rangle = \bigcup_{\alpha \in \mathcal{R}_i} \langle \alpha, x_v \rangle,$$

where $\langle x_v \rangle = \{\sigma : r^v_{P_v}(\sigma) = x_v\}$ and $\langle \alpha, x_v \rangle = \{\sigma : \sigma^{\{1,\dots,i\}} = \alpha, r^v_{P_v}(\sigma) = x_v\}$. To represent $\langle \boldsymbol{x}_{\{1,\dots,v\}} \rangle$, we have to construct the set of partial rankings $\{\beta \in \mathcal{S}_v^{\{1,\dots,v\}} : \beta^{\{1,\dots,i\}} = \alpha, r^v_{P_v}(\beta) = x_v\}$ for each partial ranking $\alpha \in \mathcal{R}_i$. By Definition B.1 and Lemma B.2 these partial rankings are obtained by inserting v in α in the ranks $I_v = \{r \in \{1,\dots,i+1\} : \sum_{u \in P_v} \mathbb{1}\big(\alpha(u) < r\big) = x_v - 1\}$:

$$\langle \alpha, x_v \rangle = \bigcup_{r \in I_v} \langle \text{insert}(v, r, \alpha) \rangle.$$

Note that for $r \in I_v$ and $\beta = \text{insert}(v, r, \alpha)$ we have

$$X_v(\beta) = r^v_{P_v}(\beta) = 1 + \sum_{w \in P_v} \mathbb{1}\big(\beta(w) < \beta(v)\big) = 1 + \sum_{w \in P_v} \mathbb{1}\big(\beta(w) < r\big) = x_v.$$

Given $\alpha \in \mathcal{S}_i^{\{1,\dots,i\}}$, the set of insertion ranks I_v satisfying the constraint $\sum_{u \in P_v} \mathbb{1}\big(\alpha(u) < r\big) = x_v - 1$ is $\{1,\dots,i+1\}$ if $P_v = \varnothing$. Otherwise:

- $I_v = \{1,\dots,\alpha(\bar{\alpha}^{P_v}(1))\}$, if $x_v = 1$
- $I_v = \{\alpha(\bar{\alpha}^{P_v}(x_v - 1)) + 1,\dots,\alpha(\bar{\alpha}^{P_v}(x_v))\}$, if $1 < x_v \leq |P_v|$
- and $I_v = \{\alpha(\bar{\alpha}^{P_v}(|P_v|)) + 1,\dots,i+1\}$, if $x_v = |P_v| + 1$.

Finally, the set of insertion ranks satisfies $|I_v| \geq 1$ and thus $|\mathcal{R}_v| \geq 1$, for $v = 2,\dots,n$. Thus, $\langle \boldsymbol{x} \rangle = \mathcal{R}_n \neq \varnothing$. □

C Admissible models

In this appendix we consider RRVs defined according to arbitrary variable graphs and we characterize admissible RRMs through the pushforward relation with ranking models. Recall that an RRM is admissible if and only if it is the pushforward of at least one ranking model. We first provide a constructive description of the pullback of an admissible RRM, and then derive a simple necessary and sufficient condition for admissibility based on the support of the model. This characterization shows that admissibility imposes a structural constraint on the model: An RRM must assign zero probability to all realizations associated with empty events.

Lemma C.1 (Pullback ranking models)**.** *Let $\boldsymbol{X}$ be an RRV for V and h be an RRM for $\boldsymbol{X}$, and let $p \in \Delta(\mathcal{S}_n^V)$ be a ranking model. Then, $p \in$ pull(h) if and only if for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ with $h(\boldsymbol{x}) > 0$ there exists $p_{\boldsymbol{x}} \in \Delta(\langle \boldsymbol{x} \rangle)$ such that*

$$p(\sigma) = h(\boldsymbol{x}) p_{\boldsymbol{x}}(\sigma), \tag{8}$$

for $\sigma \in \langle \boldsymbol{x} \rangle$.

Proof. By Definition 3.2, $p \in \text{pull}(h)$ if and only if $h = \text{push}_{\boldsymbol{X}}(p)$.

i) Suppose that for every $\boldsymbol{x}$ with $h(\boldsymbol{x}) > 0$ there exists $p_{\boldsymbol{x}} \in \Delta(\langle \boldsymbol{x} \rangle)$ satisfying Eq. 8. Summing within each such event gives

$$\sum_{\sigma \in \langle \boldsymbol{x} \rangle} p(\sigma) = h(\boldsymbol{x}) \sum_{\sigma \in \langle \boldsymbol{x} \rangle} p_{\boldsymbol{x}}(\sigma) = h(\boldsymbol{x}).$$

These events are disjoint, and their total probability under p is $\sum_{\boldsymbol{x}: h(\boldsymbol{x}) > 0} h(\boldsymbol{x}) = 1$. Since p is a probability distribution, every remaining event has probability zero. Hence $\text{push}_{\boldsymbol{X}}(p) = h$.

ii) Suppose that $h = \text{push}_{\boldsymbol{X}}(p)$. For each $\boldsymbol{x}$ with $h(\boldsymbol{x}) > 0$, define $p_{\boldsymbol{x}}(\sigma) = p(\sigma)/h(\boldsymbol{x})$ for $\sigma \in \langle \boldsymbol{x} \rangle$. This is a probability distribution on $\langle \boldsymbol{x} \rangle$, because its values are nonnegative and sum to one, and it satisfies Eq. 8.

If $h(\boldsymbol{x}) = 0$, then $p(\sigma) = 0$ for every $\sigma \in \langle \boldsymbol{x} \rangle$. For $\langle \boldsymbol{x} \rangle \neq \varnothing$, any choice of $p_{\boldsymbol{x}} \in \Delta(\langle \boldsymbol{x} \rangle)$ satisfies Eq. 8; if $\langle \boldsymbol{x} \rangle = \varnothing$, no $p_{\boldsymbol{x}}$ needs to be defined. □

Using Eq. 8, we can generate ranking models in the pullback of any admissible RRM by choosing a distribution $p_{\boldsymbol{x}}$ within each event with $h(\boldsymbol{x}) > 0$. A particularly simple choice that does not require extra parameters assigns equal probability to all rankings within each such event.

The next result provides a simple rule to characterize the admissible models.

Proposition C.2 (Structural zeroes)**.** *Let $\boldsymbol{X}$ be an RRV for V. An RRM $h \in \Delta(\boldsymbol{\mathcal{X}})$ is admissible if and only if $h(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ with $\langle \boldsymbol{x} \rangle = \varnothing$.*

Proof. The proof has two parts.

i) If h is admissible then $h(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ with $\langle \boldsymbol{x} \rangle = \varnothing$:

If $h \in \Delta(\boldsymbol{\mathcal{X}})$ is admissible by Definition 3.3 then there exists $p \in \Delta(\mathcal{S}_n^V)$ such that $h(\boldsymbol{x}) = p(\boldsymbol{x}) := \sum_{\sigma \in \langle \boldsymbol{x} \rangle} p(\sigma)$ for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$. Therefore, for $\boldsymbol{x}^{\varnothing} \in \boldsymbol{\mathcal{X}}$ with $\langle \boldsymbol{x}^{\varnothing} \rangle = \varnothing$ we have $h(\boldsymbol{x}^{\varnothing}) = p(\boldsymbol{x}^{\varnothing}) = 0$.

ii) If $h(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ with $\langle \boldsymbol{x} \rangle = \varnothing$ then h is admissible:

By Definition 3.3, h is admissible if there exists $p \in \text{pull}(h)$. According to Lemma C.1 $p \in \text{pull}(h)$ if and only if

$$p(\sigma) = h(\boldsymbol{x}) p_{\boldsymbol{x}}(\sigma),$$

where $p_{\boldsymbol{x}} \in \Delta(\langle \boldsymbol{x} \rangle)$ for $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ with $h(\boldsymbol{x}) > 0$.

We obtain a $p \in \text{pull}(h)$ by construction. For $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ with $h(\boldsymbol{x}) > 0$, define the uniform distribution $p_{\boldsymbol{x}} \in \Delta(\langle \boldsymbol{x} \rangle)$ as $p_{\boldsymbol{x}}(\sigma) = 1/|\langle \boldsymbol{x} \rangle|$ if $\sigma \in \langle \boldsymbol{x} \rangle$ and $p_{\boldsymbol{x}}(\sigma) = 0$ otherwise. Set $p(\sigma) = h(\boldsymbol{X}(\sigma)) p_{\boldsymbol{X}(\sigma)}(\sigma)$ when $h(\boldsymbol{X}(\sigma)) > 0$, and $p(\sigma) = 0$ otherwise. Then $p \in \Delta(\mathcal{S}_n^V)$ because $p(\sigma) \geq 0$ and

$$\sum_{\sigma \in \mathcal{S}_n^V} p(\sigma) = \sum_{\boldsymbol{x}: \langle \boldsymbol{x} \rangle \neq \varnothing} \sum_{\sigma \in \langle \boldsymbol{x} \rangle} p(\sigma) = \sum_{\substack{\boldsymbol{x}: \langle \boldsymbol{x} \rangle \neq \varnothing, \\ h(\boldsymbol{x}) > 0}} h(\boldsymbol{x}) \sum_{\sigma \in \langle \boldsymbol{x} \rangle} p_{\boldsymbol{x}}(\sigma) = \sum_{\substack{\boldsymbol{x}: \langle \boldsymbol{x} \rangle \neq \varnothing, \\ h(\boldsymbol{x}) > 0}} h(\boldsymbol{x}) = 1,$$

where the last equality holds because $h(\boldsymbol{x}) = 0$ for all $\boldsymbol{x}$ with $\langle \boldsymbol{x} \rangle = \varnothing$. Therefore $p \in \text{pull}(h)$ and $\text{pull}(h) \neq \varnothing$. □

All RRMs defined over variables that break MEC are admissible, since every realization corresponds to a non-empty event. In contrast, when RRVs are subject to MEC, admissibility requires assigning zero probability to all realizations associated with empty events.

D Admissible models and conditional independence

When an RRV is defined by an acyclic variable graph, it breaks MEC, and all realizations correspond to non-empty events. In this case, admissibility does not require additional constraints, and conditional independence relations can be imposed without requiring structural zeros in the associated RRMs.

In contrast, when an RRV is defined by a cyclic variable graph, it is subject to MEC, and some realizations correspond to empty events. In this setting, certain conditional independence relations require imposing structural zeros on the associated RRMs in order to ensure admissibility. Moreover, these constraints propagate to ranking models in the pullback, potentially requiring structural zeros over an exponential number of rankings.

The following result illustrates this phenomenon.

Theorem D.1. *Let A, B, C be three pairwise disjoint nonempty subsets of V. Let $\boldsymbol{X}$ be an RRV for V with $\boldsymbol{x}^{\varnothing} \in \boldsymbol{\mathcal{X}}$ so that $\langle \boldsymbol{x}^{\varnothing}_{A,B,C} \rangle = \varnothing$.*

A ranking model $p \in \Delta(\mathcal{S}_n^V)$ satisfies $\boldsymbol{X}_A \perp_p \boldsymbol{X}_B \mid \boldsymbol{X}_C$ only if

$$p(\boldsymbol{x}^{\varnothing}_C) = 0 \quad \textit{or} \quad p(\boldsymbol{x}^{\varnothing}_{A,C}) = 0 \quad \textit{or} \quad p(\boldsymbol{x}^{\varnothing}_{B,C}) = 0.$$

Proof. We argue by contradiction. Suppose that there exists $p \in \Delta(\mathcal{S}_n^V)$, with $p(\boldsymbol{x}^{\varnothing}_C) > 0$, $p(\boldsymbol{x}^{\varnothing}_{A,C}) > 0$ and $p(\boldsymbol{x}^{\varnothing}_{B,C}) > 0$:

The event associated to the realization of $\boldsymbol{x}^{\varnothing}_{A,B,C}$ is empty, $\langle \boldsymbol{x}^{\varnothing}_{A,B,C} \rangle = \varnothing$, and thus

$$p(\boldsymbol{x}^{\varnothing}_{A,B,C}) = \sum_{\sigma \in \langle \boldsymbol{x}^{\varnothing}_{A,B,C} \rangle} p(\sigma) = 0.$$

In addition, by Definition 4.1, the conditional independence $\boldsymbol{X}_A \perp_p \boldsymbol{X}_B \mid \boldsymbol{X}_C$ implies that $p(\boldsymbol{x}_{A,B}|\boldsymbol{x}_C) = p(\boldsymbol{x}_A|\boldsymbol{x}_C)p(\boldsymbol{x}_B|\boldsymbol{x}_C)$ for all $\boldsymbol{x}_{A,B,C} \in \boldsymbol{\mathcal{X}}_{A,B,C}$ with $p(\boldsymbol{x}_C) > 0$.

Since $p(\boldsymbol{x}^{\varnothing}_C) > 0$, we have

$$p(\boldsymbol{x}^{\varnothing}_{A,B,C}) = \frac{p(\boldsymbol{x}^{\varnothing}_{A,C})\, p(\boldsymbol{x}^{\varnothing}_{B,C})}{p(\boldsymbol{x}^{\varnothing}_C)} > 0.$$

This contradicts $p(\boldsymbol{x}^{\varnothing}_{A,B,C}) = 0$. □

Note that $\langle \boldsymbol{x}^{\varnothing}_{A,C} \rangle \subseteq \langle \boldsymbol{x}^{\varnothing}_C \rangle$ and $\langle \boldsymbol{x}^{\varnothing}_{B,C} \rangle \subseteq \langle \boldsymbol{x}^{\varnothing}_C \rangle$. This result implies that p must assign zero probability to all rankings in $\langle \boldsymbol{x}^{\varnothing}_{A,C} \rangle$ or in $\langle \boldsymbol{x}^{\varnothing}_{B,C} \rangle$. The number of such rankings can be exponential in the size of V, making these constraints highly restrictive.

This phenomenon extends the notion of probabilistic independence for rankings introduced in [Huang et al., 2009b]. In that framework, independence between subsets of items is defined directly in terms of rankings, leading to strong structural constraints. In comparison, RRVs provide a more flexible representation, but when the MEC are present, enforcing conditional independence may still require a large number of structural zeros.

In contrast, for RRVs that break MEC, all realizations correspond to non-empty events, admissibility imposes no structural restrictions, and conditional independence assumptions lead to valid factorizations of RRMs. Consequently, acyclic RRVs provide the appropriate foundation for developing graphical models over rankings. The separation between the construction of the partition and the imposition of independence assumptions is a key advantage of RRMs.

E Learning relative rank networks

Let $\mathcal{D} = (\sigma^{(1)}, \ldots, \sigma^{(M)})$ be a sample of M rankings, and let $\hat{p}$ denote its empirical distribution. For a fixed acyclic variable graph $\mathcal{V}$, transform the sample into the categorical data $\mathcal{D}_{\boldsymbol{X}} = (\boldsymbol{X}(\sigma^{(1)}), \ldots, \boldsymbol{X}(\sigma^{(M)}))$. A factorization graph $\mathcal{G}$ can then be learned using standard score-based, constraint-based, or hybrid Bayesian-network structure-learning methods [Koller and Friedman, 2009]. For fixed $(\mathcal{V}, \mathcal{G})$, estimating the local conditional distributions by empirical frequencies gives $h = \text{push}_{\mathcal{V},\mathcal{G}}(\hat{p})$, with the convention for unobserved parent configurations specified in Definition 5.2.

Selecting $\mathcal{V}$ requires comparing different coarsenings of the ranking space. The selection criterion should account for the ranking information retained and the number of model parameters. Raw RRM likelihoods alone do not provide this comparison: an empty variable graph maps every observation to the same realization, which can be assigned probability one, giving log-likelihood zero regardless of the sample. Algorithm 2 therefore takes a model-selection criterion as an input. Defining and evaluating such a criterion remains future work.

Algorithm 2 A general procedure for learning an RRN

Input: Ranking sample $\mathcal{D}$; a finite non-empty collection $\mathfrak{V}$ of candidate acyclic variable graphs; a categorical Bayesian-network structure learner; a model-selection criterion, with higher values preferred

Output: A selected RRN $(\mathcal{V}^*, \mathcal{G}^*, \boldsymbol{\theta}^*)$

1: **for** each $\mathcal{V} \in \mathfrak{V}$ **do**
2: Define $\boldsymbol{X}$ from $\mathcal{V}$ and form $\mathcal{D}_{\boldsymbol{X}} = (\boldsymbol{X}(\sigma^{(j)}))_{j=1}^{M}$
3: Learn a factorization graph $\mathcal{G}$ from $\mathcal{D}_{\boldsymbol{X}}$ using the supplied structure learner
4: Estimate $\boldsymbol{\theta}$ from empirical conditional frequencies, as in Definition 5.2
5: Evaluate $(\mathcal{V}, \mathcal{G}, \boldsymbol{\theta})$ using the supplied model-selection criterion
6: **end for**
7: Select a highest-scoring RRN $(\mathcal{V}^*, \mathcal{G}^*, \boldsymbol{\theta}^*)$ among the fitted candidates
8: **return** $(\mathcal{V}^*, \mathcal{G}^*, \boldsymbol{\theta}^*)$

The candidate variable graphs may be restricted by their maximum number of parents to control RRV cardinalities. Likewise, the learner for $\mathcal{G}$ may impose indegree or treewidth constraints to control parameter counts or inference cost. More elaborate searches could alternate local changes in $\mathcal{V}$ and $\mathcal{G}$, recomputing the categorical data when $\mathcal{V}$ changes. Developing and empirically evaluating these searches remains future work.

F Proofs of Theoretical Results

This appendix provides the proofs of the theoretical results in the main text.

F.1 Proof of Theorem 2.5 (Breaking MEC)

Proof. i) If $\boldsymbol{X}$ breaks MEC, then $\mathcal{V}$ is acyclic (by contradiction):

Assume that $\boldsymbol{X}$ is defined by a cyclic graph $\mathcal{V}$ and breaks MEC. Suppose for contradiction that $\mathcal{V}$ has a directed cycle

$$u_1, ..., u_l, u_1$$

for some length $1 \leq l \leq n$. Let $U = \{u_1, \ldots, u_l\}$ be the set of items in the cycle. Now define a realization $\boldsymbol{x}_U^{\varnothing} = (1, 1, ..., 1)$ that forces a strict cyclic pairwise preference. Since $u_i \in P_{u_{i+1}}$ for $i = 1, \ldots, l-1$ and $u_l \in P_{u_1}$, $\boldsymbol{x}_U^{\varnothing}$ implies

$$\sigma(u_1) > \sigma(u_2) > \cdots > \sigma(u_l) > \sigma(u_1)$$

for any $\sigma \in \langle \boldsymbol{x}_U^{\varnothing} \rangle$, which is impossible. So $\langle \boldsymbol{x}_U^{\varnothing} \rangle = \varnothing$.

Now take a realization $\boldsymbol{x} = (\boldsymbol{x}_{V \setminus U}, \boldsymbol{x}_U^{\varnothing})$. We have

$$\langle \boldsymbol{x} \rangle = \langle \boldsymbol{x}_{V \setminus U} \rangle \cap \langle \boldsymbol{x}_U^{\varnothing} \rangle \subseteq \langle \boldsymbol{x}_U^{\varnothing} \rangle = \varnothing,$$

which contradicts the assumption that $\boldsymbol{X}$ breaks MEC. Therefore, $\mathcal{V}$ must be acyclic.

ii) If $\mathcal{V}$ is acyclic, then $\boldsymbol{X}$ breaks MEC:

If $\boldsymbol{X}$ is defined by an acyclic variable graph, then by Proposition B.3 for any $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ we have $|\langle \boldsymbol{x} \rangle| \geq 1$, and thus $\boldsymbol{X}$ breaks MEC. □

F.2 Proof of Lemma 5.4

Proof. By Theorem 2.5, an RRV $\boldsymbol{X}$ defined by an acyclic variable graph $\mathcal{V}$ breaks MEC. Then, by Proposition C.2, any RRN h for $\boldsymbol{X}$ is admissible, that is, $\text{pull}(h) \neq \varnothing$.

By Definition 5.1, h factorizes according to $\mathcal{G}$.

On one hand, according to Theorem 3.3 in Koller and Friedman [2009], if a distribution h factorizes according to $\mathcal{G}$ and the sets A and B are d-separated by C in $\mathcal{G}$, then $\boldsymbol{X}_A \perp_h \boldsymbol{X}_B | \boldsymbol{X}_C$. On the other hand, $p \in \text{pull}(h)$, and thus $h = \text{push}_{\boldsymbol{X}}(p)$, and by Definition 4.1 $\boldsymbol{X}_A \perp_p \boldsymbol{X}_B | \boldsymbol{X}_C$. □

F.3 Proof of Proposition 6.1 (Thin partition)

Proof. i) If the partition induced by $\boldsymbol{X}$ is thin, then $\mathcal{V}$ is acyclic complete (by contradiction):

First, assume that $\mathcal{V}$ is cyclic. Then, according to Theorem 2.5, there exists $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ with $\langle \boldsymbol{x} \rangle = \emptyset$, and thus $\langle \boldsymbol{\mathcal{X}} \rangle$ is not thin. Therefore $\mathcal{V}$ is acyclic.

Next, assume that $\mathcal{V}$ is not complete. Then the size of the state space of $\boldsymbol{X}$ is smaller than $n!$. The size of the ranking space $\mathcal{S}_n^V$ is $n!$, and thus there is at least one realization $\boldsymbol{x}$ with an associated event with size larger than 1. Then the partition $\langle \boldsymbol{\mathcal{X}} \rangle$ is not thin. Therefore $\mathcal{V}$ is complete.

ii) If $\mathcal{V}$ is acyclic complete then the partition induced by $\boldsymbol{X}$ is thin:

The RRV $\boldsymbol{X}$ defined by an acyclic $\mathcal{V}$ breaks MEC, and thus $\langle \boldsymbol{\mathcal{X}} \rangle$ is a partition of $\mathcal{S}_n^V$. Besides, the size of the state space of $\boldsymbol{X}$ when $\mathcal{V}$ is complete is $n!$. The only partition of $\mathcal{S}_n^V$ with $n!$ non-empty subsets is a thin partition. □

F.4 Proof of Theorem 6.2

Proof. Without loss of generality, we assume that $\bar{\pi} = [1, \ldots, n]$.

$\boldsymbol{X}$ is defined by $\mathcal{V}^{RI}$ that is a complete DAG, and by Proposition 6.1, $\boldsymbol{X}$ induces a thin partition. Then, for simplicity, for any $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ with $\langle \boldsymbol{x} \rangle = \{\sigma\}$ we write $h(\sigma) = h(\boldsymbol{x})$, and $p(\boldsymbol{x}) = p(\sigma)$.

i) $\Delta_{\mathcal{G}^{RI}}(\boldsymbol{\mathcal{X}}) \subseteq \Delta_{\bar{\pi}}(\mathcal{S}_n^V)$:

Let $h \in \Delta_{\mathcal{G}^{RI}}(\boldsymbol{\mathcal{X}})$. Then, by Definition 5.1

$$h(\boldsymbol{x}) = \prod_{v=1}^{n} h(x_v),$$

for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$, where $X_v(\sigma) := r^v_{P_v}(\sigma)$ for $v \in V$. $\mathcal{V}^{RI}$ is a complete DAG with the ancestral ordering $\bar{\pi}$ and thus $X_v(\sigma) := r^v_{\{1,\ldots,v-1\}}(\sigma) = f_{\sigma,\bar{\pi}}(v)$, for $v \in V$. Therefore, by Eq. 6, h is a RIM for $\bar{\pi}$, and $\Delta_{\mathcal{G}^{RI}}(\boldsymbol{\mathcal{X}}) \subseteq \Delta_{\bar{\pi}}(\mathcal{S}_n^V)$.

ii) $\Delta_{\bar{\pi}}(\mathcal{S}_n^V) \subseteq \Delta_{\mathcal{G}^{RI}}(\boldsymbol{\mathcal{X}})$:

We prove it by showing that, for $p \in \Delta_{\bar{\pi}}(\mathcal{S}_n^V)$, the RRN $h = \text{push}_{\mathcal{V}^{RI},\mathcal{G}^{RI}}(p)$ is equivalent to p, i.e., $p(\sigma) = h(\boldsymbol{X}(\sigma))$ for all $\sigma \in \mathcal{S}_n^V$.

Let $p \in \Delta_{\bar{\pi}}(\mathcal{S}_n^V)$ and $h = \text{push}_{\mathcal{V}^{RI},\mathcal{G}^{RI}}(p)$. Then by Definition 5.2 and (i)

$$h(\boldsymbol{x}) = \prod_{v=1}^{n} h(x_v) = \prod_{v=1}^{n} p(x_v) = \prod_{i=1}^{n} p_i(f_{\sigma,\bar{\pi}}(i)) = p(\sigma),$$

where $\langle \boldsymbol{x} \rangle = \{\sigma\}$, for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$. □

F.5 Proof of Corollary 6.3

Proof. Let $\boldsymbol{X}$ be an RRV defined by $\mathcal{V}^{RI}$. By Theorem 6.2 we have $\Delta_{\bar{\pi}}(\mathcal{S}_n^V) = \Delta_{\mathcal{G}^{RI}}(\boldsymbol{\mathcal{X}})$. Thus, by Lemma 5.4, RI independence is the independence model ind$(\mathcal{V}^{RI}, \mathcal{G}^{RI})$. □

F.6 Proof of Corollary 6.4

To prove this result we use the notation introduced in the proof of Theorem 6.2.

Proof. Let $\boldsymbol{X}$ be an RRV defined by $\mathcal{V}^{RI}$ and let $h = \text{push}_{\mathcal{V}^{RI},\mathcal{G}^{RI}}(p)$. By Definition 5.2, we have $h \in \Delta_{\mathcal{G}^{RI}}(\boldsymbol{\mathcal{X}})$, where $h(\boldsymbol{x}) = \prod_{v \in V} p(x_v)$, for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$. By Theorem 6.2, $h \in \Delta_{\bar{\pi}}(\mathcal{S}_n^V)$.

The Kullback-Leibler divergence between p and g is:

$$D_{KL}(p||g) = \sum_{\sigma \in \mathcal{S}_n^V} p(\sigma) \ln \frac{p(\sigma)}{g(\sigma)}.$$

We have that the minimization of the Kullback-Leibler divergence is equivalent to the next problems:

$$\begin{aligned}
\underset{g\in\Delta_{\bar{\pi}}(\mathcal{S}_n^V)}{\arg\min}\ D_{KL}(p||g) &= \underset{g\in\Delta_{\bar{\pi}}(\mathcal{S}_n^V)}{\arg\max} \sum_{\sigma\in\mathcal{S}_n^V} p(\sigma)\ln g(\sigma) \\
&= \underset{g\in\Delta_{\mathcal{G}^{RI}}(\boldsymbol{\mathcal{X}})}{\arg\max} \sum_{\boldsymbol{x}\in\boldsymbol{\mathcal{X}}} p(\boldsymbol{x})\ln g(\boldsymbol{x}) \\
&= \underset{g\in\Delta_{\mathcal{G}^{RI}}(\boldsymbol{\mathcal{X}})}{\arg\max} \sum_{\boldsymbol{x}\in\boldsymbol{\mathcal{X}}} p(\boldsymbol{x})\ln \prod_{v\in V} g(x_v) \\
&= \underset{g\in\Delta_{\mathcal{G}^{RI}}(\boldsymbol{\mathcal{X}})}{\arg\max} \sum_{v\in V}\sum_{x_v\in\mathcal{X}_v} p(x_v)\ln g(x_v).
\end{aligned}$$

The first equality follows from the definition of KL divergence, since $\sum_\sigma p(\sigma)\ln p(\sigma)$ does not depend on g. The second follows from Theorem 6.2 and the bijection between rankings and RRV realizations established by Proposition 6.1. The third follows from Definition 5.1, since $\mathcal{G}^{RI}$ is empty. The fourth follows by expanding the logarithm of the product and marginalizing over the remaining variables.

Since the marginal distributions $g(x_v)$ can be chosen independently, the last objective can be maximized separately for each $v \in V$. Therefore, a RIM that minimizes the Kullback-Leibler divergence with respect to p is

$$g^*(\boldsymbol{x}) = \prod_{v\in V} p(x_v) = h(\boldsymbol{x}),$$

for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$. □

F.7 Proof of Theorem 6.5

To prove Theorem 6.5, we first introduce a result that relates RRVs with relative rankings when the variable graph is a complete acyclic graph.

Let A and B be two disjoint non-empty subsets. The relative ranking of A with respect to B under σ, $\sigma_B^A \in \mathcal{S}^A_{|A|+|B|}$, is the injective function from A to $\{1,\dots,|A|+|B|\}$ mapping each item to its rank in the partial ranking $\sigma^{A,B}$, that is, $\sigma_B^A(a) = \sigma^{A,B}(a)$ for $a \in A$. Here, $\mathcal{S}^A_{|A|+|B|}$ denotes the space of these injective functions.

Lemma F.1 (Partial and relative rankings)**.** *Let $\mathcal{V}$ be a complete acyclic variable graph with ancestral ordering $\bar{\pi} = [a_1,\dots,a_m,b_1,\dots,b_{n-m}]$, let $A = \{a_1,...,a_m\}$ and $B = \{b_1,\dots,b_{n-m}\}$ be a partition of V, and let $\boldsymbol{X}$ be defined according to $\mathcal{V}$. Then*

- *$\boldsymbol{X}_A$ induces a thin partition of the (partial) ranking space of A, $\mathcal{S}_m^A$, and*
- *$\boldsymbol{X}_B$ induces a thin partition of the relative ranking space of B with respect to A, $\mathcal{S}_n^B$.*

Proof. i) $\boldsymbol{X}_A$ induces a thin partition of $\mathcal{S}_m^A$:

The definition of $\boldsymbol{X}_A$ is equivalent to an RRV for A defined by the complete acyclic variable graph $\mathcal{V}^A$ with ancestral ordering $\bar{\pi}^A = [a_1,\dots,a_m]$. Thus by Proposition 6.1, $\boldsymbol{X}_A$ induces a thin partition of $\mathcal{S}^A_{|A|}$.

ii) $\boldsymbol{X}_B$ induces a thin partition of $\mathcal{S}_n^B$

For $j = 1,\dots,n-m$, the reference set of b_j is $A \cup \{b_1,\dots,b_{j-1}\}$. Given $\boldsymbol{x}_B$, start from any partial ranking $\alpha \in \mathcal{S}_m^A$ and successively insert each b_j at rank x_{b_j} using the insertion operator of Definition B.1. Each insertion rank is uniquely specified because the reference set contains all previously inserted items. Changing α changes only the partial ordering of the items in A, leaving the insertion ranks and the final ranks of the items in B unchanged. Thus $\boldsymbol{x}_B$ uniquely determines an injective function from B to N, independently of α.

Conversely, given such an injective function, choose a ranking of V in which the items in B have the specified ranks and the partial ranking of A is α. Remove $b_{n-m},\dots,b_1$ in this order, recording the rank of each item immediately before removal. These ranks recover $x_{b_{n-m}},\dots,x_{b_1}$, respectively,

independently of α. Thus the two constructions are inverse, establishing a bijection between $\boldsymbol{\mathcal{X}}_B$ and $\mathcal{S}_n^B$. In particular, for $\boldsymbol{x} = \boldsymbol{X}(\sigma)$,

$$\langle \boldsymbol{x}_B \rangle = \{\rho \in \mathcal{S}_n^V : \rho_A^B = \sigma_A^B\},$$

which proves the claim. □

Now, we proceed with the proof of Theorem 6.5.

Proof. We prove that $\Delta_{A,B}(\mathcal{S}_n^V) = \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{X}}) \cap \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{Y}})$, where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are defined by $\mathcal{V}^{RS}$ and $\mathcal{W}^{RS}$, respectively, by showing that $h \in \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{X}}) \cap \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{Y}})$ if and only if $h \in \Delta_{A,B}(\mathcal{S}_n^V)$.

The RRVs $\boldsymbol{X}$ and $\boldsymbol{Y}$ for V are defined by the complete acyclic graphs $\mathcal{V}^{RS}$ and $\mathcal{W}^{RS}$, respectively. By Proposition 6.1, $\boldsymbol{X}$ and $\boldsymbol{Y}$ induce thin partitions. Then, for $h \in \Delta(\boldsymbol{\mathcal{X}})$ and $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ with $\langle \boldsymbol{x} \rangle = \{\sigma\}$, we write $h(\sigma) = h(\boldsymbol{x})$ and $p(\boldsymbol{x}) = p(\sigma)$ for simplicity. Moreover, for any $p \in \Delta(\mathcal{S}_n^V)$, partial ranking $\sigma^A \in \mathcal{S}_m^A$, and interleaving $\boldsymbol{s}_A^B \in \binom{N}{|B|}$, we denote $p(\sigma^A) = \sum_{\rho \in \langle \sigma^A \rangle} p(\rho)$ and $p(\boldsymbol{s}_A^B) = \sum_{\rho \in \langle \boldsymbol{s}_A^B \rangle} p(\rho)$, where $\langle \sigma^A \rangle = \{\rho : \rho^A = \sigma^A\}$ and $\langle \boldsymbol{s}_A^B \rangle = \{\rho \in \mathcal{S}_n^V : s_A^B(\rho) = \boldsymbol{s}_A^B\}$.

An RRM $h \in \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{X}})$ if and only if

$$h(\sigma) = h(\sigma^A) h(\sigma_A^B) = h(\sigma^A) h(s_A^B(\sigma), \sigma^B),$$

for all $\sigma \in \mathcal{S}_n^V$. The first equality holds by Lemma F.1. The second equality holds because for $v \in B$

$$\sigma_A^B(v) = \boldsymbol{s}_A^B(\sigma^B(v)),$$

where $\boldsymbol{s}_A^B = s_A^B(\sigma)$, for $\sigma \in \mathcal{S}_n^V$.

An RRM $h \in \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{Y}})$ if and only if

$$h(\sigma) = h(\sigma_B^A) h(\sigma^B) = h(\sigma^A, s_B^A(\sigma)) h(\sigma^B),$$

for all $\sigma \in \mathcal{S}_n^V$. In addition, $s_B^A(\sigma)$ uniquely determines $s_A^B(\sigma)$, for $\sigma \in \mathcal{S}_n^V$. Thus, we have that $h \in \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{Y}})$ if and only if

$$h(\sigma) = h(\sigma^A, s_A^B(\sigma)) h(\sigma^B),$$

for $\sigma \in \mathcal{S}_n^V$.

Then, $h \in \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{X}}) \cap \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{Y}})$ if and only if

$$h(\sigma) = h(\sigma^A) h(s_A^B(\sigma)) h(\sigma^B),$$

for $\sigma \in \mathcal{S}_n^V$.

By Eq. 7, $h \in \Delta_{A,B}(\mathcal{S}_n^V)$ if and only if

$$h(\sigma) = h(\sigma^A) h(s_A^B(\sigma)) h(\sigma^B),$$

for all $\sigma \in \mathcal{S}_n^V$. And therefore, $h \in \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{X}}) \cap \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{Y}})$ if and only if $h \in \Delta_{A,B}(\mathcal{S}_n^V)$. □

F.8 Proof of Corollary 6.6

Proof. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be the RRVs defined by $\mathcal{V}^{RS}$ and $\mathcal{W}^{RS}$. By Theorem 6.5 we have $\Delta_{A,B}(\mathcal{S}_n^V) \subseteq \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{X}})$ and $\Delta_{A,B}(\mathcal{S}_n^V) \subseteq \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{Y}})$. Thus, by Lemma 5.4, RS independence is given by the union of independence models $\text{ind}(\mathcal{V}^{RS}, \mathcal{G}^{RS})$ and $\text{ind}(\mathcal{W}^{RS}, \mathcal{G}^{RS})$. □

F.9 Proof of Proposition 6.7

In this proof we reuse the notation in the proof of Theorem 6.5.

Proof. Let $g = \text{push}_{\mathcal{W}^{RS}, \mathcal{G}^{RS}}(p) \in \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{Y}})$, and $h = \text{push}_{\mathcal{V}^{RS}, \mathcal{G}^{RS}}(g) \in \Delta_{\mathcal{G}^{RS}}(\boldsymbol{\mathcal{X}})$.

i) $h \in \Delta_{A,B}(\mathcal{S}_n^V)$:

First, by Definition 5.2, we have

$$\begin{aligned} g(\sigma) =& p(\sigma_B^A)p(\sigma^B) \\ =& p(\sigma^A, \boldsymbol{s}_B^A)p(\sigma^B) \\ =& p(\sigma^A, \boldsymbol{s}_A^B)p(\sigma^B), \end{aligned}$$

where $\boldsymbol{s}_B^A = s_B^A(\sigma)$ and $\boldsymbol{s}_A^B = s_A^B(\sigma)$, for all $\sigma \in \mathcal{S}_n^V$. The second equality holds because $\sigma_B^A(v) = \boldsymbol{s}_B^A(\sigma^A(v))$ for all $v \in A$, and thus $\langle \sigma_B^A \rangle = \langle \sigma^A, \boldsymbol{s}_B^A \rangle$. The third equality holds because the ranks in $\boldsymbol{s}_A^B$ are uniquely determined by the ranks in $\boldsymbol{s}_B^A$ and thus $\langle \boldsymbol{s}_B^A \rangle = \langle \boldsymbol{s}_A^B \rangle$.

Then, by Definition 5.2, we have

$$\begin{aligned} h(\sigma) =& g(\sigma^A)g(\sigma_A^B) \\ =& g(\sigma^A)g(\sigma^B, \boldsymbol{s}_A^B) \\ =& p(\sigma^A)p(\sigma^B)p(\boldsymbol{s}_A^B), \end{aligned}$$

where $\boldsymbol{s}_A^B = s_A^B(\sigma)$, for all $\sigma \in \mathcal{S}_n^V$. The second equality holds because $\sigma_A^B(b) = \boldsymbol{s}_A^B(\sigma^B(b))$ for all $b \in B$.

ii) $h = \arg\min_{q \in \Delta_{A,B}(\mathcal{S}_n^V)} D_{KL}(p||q)$:

By (i) $h \in \Delta_{A,B}(\mathcal{S}_n^V)$ and $h(\sigma) = p(\sigma^A)p(\sigma^B)p(\boldsymbol{s}_A^B)$.

The Kullback-Leibler divergence between p and g is:

$$D_{KL}(p||g) = \sum_{\sigma \in \mathcal{S}_n^V} p(\sigma) \ln \frac{p(\sigma)}{g(\sigma)}.$$

Minimizing the divergence is equivalent to maximizing the expected log-likelihood,

$$\arg\min_{g \in \Delta_{A,B}(\mathcal{S}_n^V)} D_{KL}(p||g) = \arg\max_{g \in \Delta_{A,B}(\mathcal{S}_n^V)} \sum_{\sigma \in \mathcal{S}_n^V} p(\sigma) \ln g(\sigma).$$

Any $g \in \Delta_{A,B}(\mathcal{S}_n^V)$ factorizes as

$$g(\sigma) = g(\sigma^A)g(\sigma^B)g(s_A^B(\sigma)).$$

Using the factorization, we can decompose the objective function of the maximization problem additively:

$$\sum_{\sigma \in \mathcal{S}_n^V} p(\sigma) \ln g(\sigma) = \sum_{\sigma \in \mathcal{S}_n^V} p(\sigma) \ln g(\sigma^A) + \sum_{\sigma \in \mathcal{S}_n^V} p(\sigma) \ln g(\sigma^B) + \sum_{\sigma \in \mathcal{S}_n^V} p(\sigma) \ln g(s_A^B(\sigma)).$$

By grouping rankings in the events we have

$$\begin{aligned} \sum_{\sigma \in \mathcal{S}_n^V} p(\sigma) \ln g(\sigma^A) &= \sum_{\sigma^A \in \mathcal{S}_{|A|}^A} p(\sigma^A) \ln g(\sigma^A), \\ \sum_{\sigma \in \mathcal{S}_n^V} p(\sigma) \ln g(\sigma^B) &= \sum_{\sigma^B \in \mathcal{S}_{|B|}^B} p(\sigma^B) \ln g(\sigma^B), \\ \sum_{\sigma \in \mathcal{S}_n^V} p(\sigma) \ln g(s_A^B(\sigma)) &= \sum_{\boldsymbol{s}_A^B \in \binom{N}{|B|}} p(\boldsymbol{s}_A^B) \ln g(\boldsymbol{s}_A^B). \end{aligned}$$

Therefore, the RSM that minimizes the Kullback-Leibler divergence with respect to p is h, where

$$h(\sigma) = p(\sigma^A)p(\sigma^B)p(s_A^B(\sigma)),$$

for all $\sigma \in \mathcal{S}_n^V$. □

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The contributions summarized in the abstract and introduction concern the representation, independence properties, and connections to existing ranking models established in the paper.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitations are related to variables that are subject to mutual exclusivity constraints (see discussions in Appendix C and Appendix D). Section 7 states the theoretical scope, and Appendix E discusses the remaining work on structure learning.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The assumptions are stated in the definitions and result statements, and proofs are provided in the appendices.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [N/A]

Justification: This work is a theoretical contribution and it does not include experiments.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [N/A]

   Justification: This work is a theoretical contribution and it does not include experiments.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [N/A]

   Justification: This work is a theoretical contribution and it does not include experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [N/A]

   Justification: The paper is a foundational work on relative rank models and it does not include experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [N/A]

   Justification: This work is a theoretical contribution and it does not include experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The work develops a theoretical framework for ranking models and involves no human-subject studies or collection of personal data.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This work is a theoretical contribution, and we believe a detailed discussion of societal impacts is not directly applicable.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not release datasets or pretrained models requiring safeguards.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The research does not use external datasets, code, or pretrained models.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release datasets, code, or trained models.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper is a foundational work on relative rank models and it does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper is a foundational work on relative rank models and it does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.

- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs are not a component of the proposed models or algorithms.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.