International Journal of Data Science and Analytics (2025) 20:4363–4384
https://doi.org/10.1007/s41060-025-00722-1

REGULAR PAPER

Ranking with ties based on noisy performance data

Aravind Sankaran[1] · Lars Karlsson[2] · Paolo Bientinesi[2]

Received: 26 June 2024 / Accepted: 16 January 2025 / Published online: 3 February 2025

Abstract

We consider the problem of ranking a set of objects based on their performance when the measurement of said performance is subject to noise. In this scenario, the performance is measured repeatedly, resulting in a range of measurements for each object. If the ranges of two objects do not overlap, then we consider one object as 'better' than the other, and we expect it to receive a higher rank; if, however, the ranges overlap, then the objects are *incomparable*, and we wish them to be assigned the same rank. Unfortunately, the incomparability relation of ranges is in general not transitive; as a consequence, in general the two requirements cannot be satisfied simultaneously, i.e., it is not possible to guarantee both distinct ranks for objects with separated ranges, and same rank for objects with overlapping ranges. This conflict leads to more than one reasonable way to rank a set of objects. Although the problem of ranking with ties has been widely studied, there remains a lack of clarity regarding what constitutes a set of reasonable rankings. In this paper, we explore the ambiguities that arise when ranking with ties, and define a set of reasonable rankings, which we call *partial rankings*. We develop and analyze three different methodologies to compute a partial ranking. Finally, we show how performance differences among objects can be investigated with the help of partial ranking.

Keywords Ranking · Noise · Partial orders · Knowledge discovery · Performance

1 Introduction

The problem of ranking a set of objects based on noisy performance measurements appears in various domains, such as high-performance computing (HPC) and business process management (BPM). For example, in the field of HPC it is common to compare (i.e., to rank) a set of algorithms based on their execution time, and it is known that execution time measurements typically exhibit fluctuations due, for instance, to the compute environment [17]. Similarly, in BPM, one might want to rank different workflows within an organization based on nondeterministic throughput times of each workflow. In this paper, we explore and develop methodologies to rank objects based on noisy measurement data while allowing for ties.

Let us first illustrate the difficulty in ranking with ties. Consider the problem of ranking 10 algorithmic variants for the solution of a generalized least squares (GLS) problem based on their execution times, as shown in Fig. 1 (the variants were automatically generated by the compiler Linnea [3] and are mathematically equivalent, i.e., in exact arithmetic they would produce the exact same result). The variants were implemented in the Julia language [4] and run on a fixed problem size. For each variant, the execution time was measured 10 times on a Linux-based machine using 12 cores of an Intel-Xeon processor with turbo-boost enabled. The execution time measurements are depicted as box plots, where the box represents the inter-quartile interval (IQI), which is the interval between the 25th and the 75th quantile values, and the red line within the box represents the median value.

Ranking approaches that result in a total order (e.g., ranking based on single indicator such as only the 25^{th} quantile or the median value) often lead to the assignment of a distinct rank to each variant, thus overlooking the possibility of ties (except when the indicators are equal). By contrast, ranking with ties can be seen as partial ordering of a set of

✉ Aravind Sankaran
aravind.sankaran@rwth-aachen.de

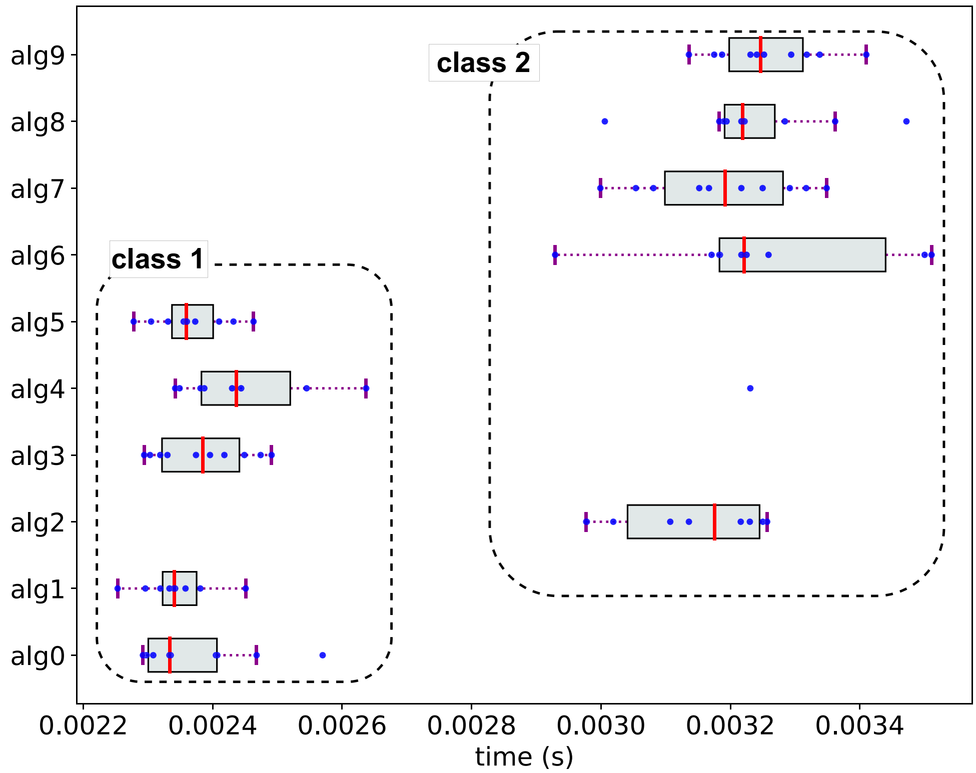
Lars Karlsson
larsk@cs.umu.se

Paolo Bientinesi
pauldj@cs.umu.se

1 RWTH Aachen University, Aachen, Germany

2 Umeå Universitet, Umeå, Sweden

Fig. 1 The execution time measurements of 10 algorithmic variants to solve the GLS problem: $(X^T M^{-1} X)^{-1} X^T M^{-1} \mathbf{y}$ where $X \in \mathbb{R}^{1000\times100}$, $M \in \mathbb{R}^{1000\times1000}$ and $\mathbf{y} \in \mathbb{R}^{1000}$. For each variant, the execution times are shown as box plots; red lines indicate the median values; the box indicates the inter-quartile interval



objects, in which the objects are related to one another based on a certain notion of importance derived from one or more indicators. For example, consider a ranking that takes into account both the 25^{th} and the 75^{th} quantiles, and resolves for ties based on the better-than relation ($<_{eg}$) where a variant $\mathbf{alg}_i$ is considered to be better than another variant $\mathbf{alg}_j$ (i.e., $\mathbf{alg}_i <_{eg} \mathbf{alg}_j$) if and only if the IQI of $\mathbf{alg}_i$ lies entirely to the left of the IQI of $\mathbf{alg}_j$. If the IQI of $\mathbf{alg}_i$ and $\mathbf{alg}_j$ overlaps with one another, then $\mathbf{alg}_i$ and $\mathbf{alg}_j$ are considered *incomparable* (i.e., $\mathbf{alg}_i \sim \mathbf{alg}_j$). Accordingly, the algorithms can be grouped into two performance classes as marked in Fig. 1, where each class includes variants with the same rank.

Notice that non-transitive ties can be expected in a noisy measurement data; for example, $\mathbf{alg}_1 \sim \mathbf{alg}_3$ and $\mathbf{alg}_3 \sim \mathbf{alg}_4$, but $\mathbf{alg}_1 <_{eg} \mathbf{alg}_4$. Therefore, the following ambiguity arises in the ranking process: Should $\mathbf{alg}_4$ and $\mathbf{alg}_1$ be placed in different ranks because their intervals (IQIs) are separated, or should both of them be assigned the same rank because their IQIs mutually overlap with that of $\mathbf{alg}_3$?

In general, it is possible to have several reasonable rankings for given sets of measurement data and a better-than relation. To illustrate how different rankings can be obtained through various arguments, let us utilize the following simulated example. Consider the set of variants $\mathcal{M} = \{\mathbf{t}_0, \mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3\}$ with each $\mathbf{t}_i \in \mathcal{M}$ indicated as a set of measurement values $\mathbf{t}_i \in \mathbb{R}^M$ sampled from normal distribution $\mathcal{N}(\mu, \sigma)$ that simulates the cost of the variant. Here, μ represents the mean and σ represents the standard deviation of the normal distribution. Let us assume that the variants are characterized by the following distributions:

- $\mathbf{t}_0$: Sampled from $\mathcal{N}(0.30, 0.005)$.
- $\mathbf{t}_1$: Sampled from $\mathcal{N}(0.31, 0.030)$.
- $\mathbf{t}_2$: Sampled from $\mathcal{N}(0.32, 0.005)$.
- $\mathbf{t}_3$: Sampled from $\mathcal{N}(0.43, 0.01)$.

For each variant, $M = 15$ values are sampled and the resulting set of values are depicted in Fig. 2 as box plots.

For the sets of measurements ($\mathcal{M}_0$) in Fig. 2, the relation $<_{eg}$ among the variants is shown as a (transitively reduced) directed graph in Fig. 3a; a directed path from $\mathbf{t}_i$ to $\mathbf{t}_j$ indicates that $\mathbf{t}_i <_{eg} \mathbf{t}_j$. An example of the ambiguity for this case could be that the three rankings with ties shown in Fig. 3b are reasonable by the following justifications:

- One could argue that since $\mathbf{t}_0 \sim \mathbf{t}_1$ and $\mathbf{t}_1 \sim \mathbf{t}_2$, all $\mathbf{t}_0$, $\mathbf{t}_1$ and $\mathbf{t}_2$ should be given the same rank and considered as the best variants (Fig. 3b(i)).
- Alternatively, one could argue that $\mathbf{t}_0$ should be ranked higher than $\mathbf{t}_2$ because $\mathbf{t}_0 <_{eg} \mathbf{t}_2$, and $\mathbf{t}_1$ should be ranked as low as $\mathbf{t}_2$ because the IQI of $\mathbf{t}_1$ overlaps with that of $\mathbf{t}_2$ (Fig. 3b(ii)). Then, only $\mathbf{t}_0$ should be considered as the best variant.
- Alternatively, since $\mathbf{t}_0 \sim \mathbf{t}_1$ and $\mathbf{t}_0 <_{eg} \mathbf{t}_2$, one could argue that both $\mathbf{t}_0$ and $\mathbf{t}_1$ should be considered as the best

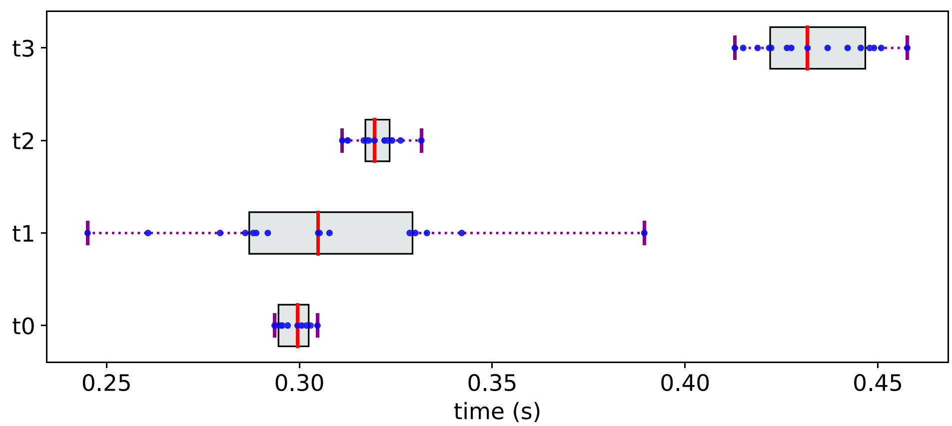
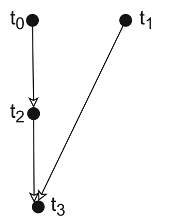


Fig. 2 Sets of measurements ($\mathcal{M}_0$)



(a) $<_{eg}$ on $\mathcal{M}_0$ shown as a transitively reduced directed graph.

(i)	(ii)	(iii)
$\mathcal{R}_0 = \{\mathbf{t}_0, \mathbf{t}_1, \mathbf{t}_2\}$	$\mathcal{R}_0 = \{\mathbf{t}_0\}$	$\mathcal{R}_0 = \{\mathbf{t}_0, \mathbf{t}_1\}$
$\mathcal{R}_1 = \{\mathbf{t}_3\}$	$\mathcal{R}_1 = \{\mathbf{t}_1, \mathbf{t}_2\}$	$\mathcal{R}_1 = \{\mathbf{t}_2\}$
	$\mathcal{R}_2 = \{\mathbf{t}_3\}$	$\mathcal{R}_2 = \{\mathbf{t}_3\}$

(b) Reasonable rankings deduced from Fig. 3a. $\mathbf{t}_i \in \mathcal{R}_k$ implies $\mathbf{t}_i$ is assigned the rank k.

Fig. 3 When distributions overlap, the partial ordering of the algorithms admits many reasonable rankings

variants, and $\mathbf{t}_2$ should be placed in a lower rank than $\mathbf{t}_0$ (Fig. 3b(iii)).

Thus, as soon as one considers a better-than relation that induces non-transitive ties, multiple reasonable rankings emerge.

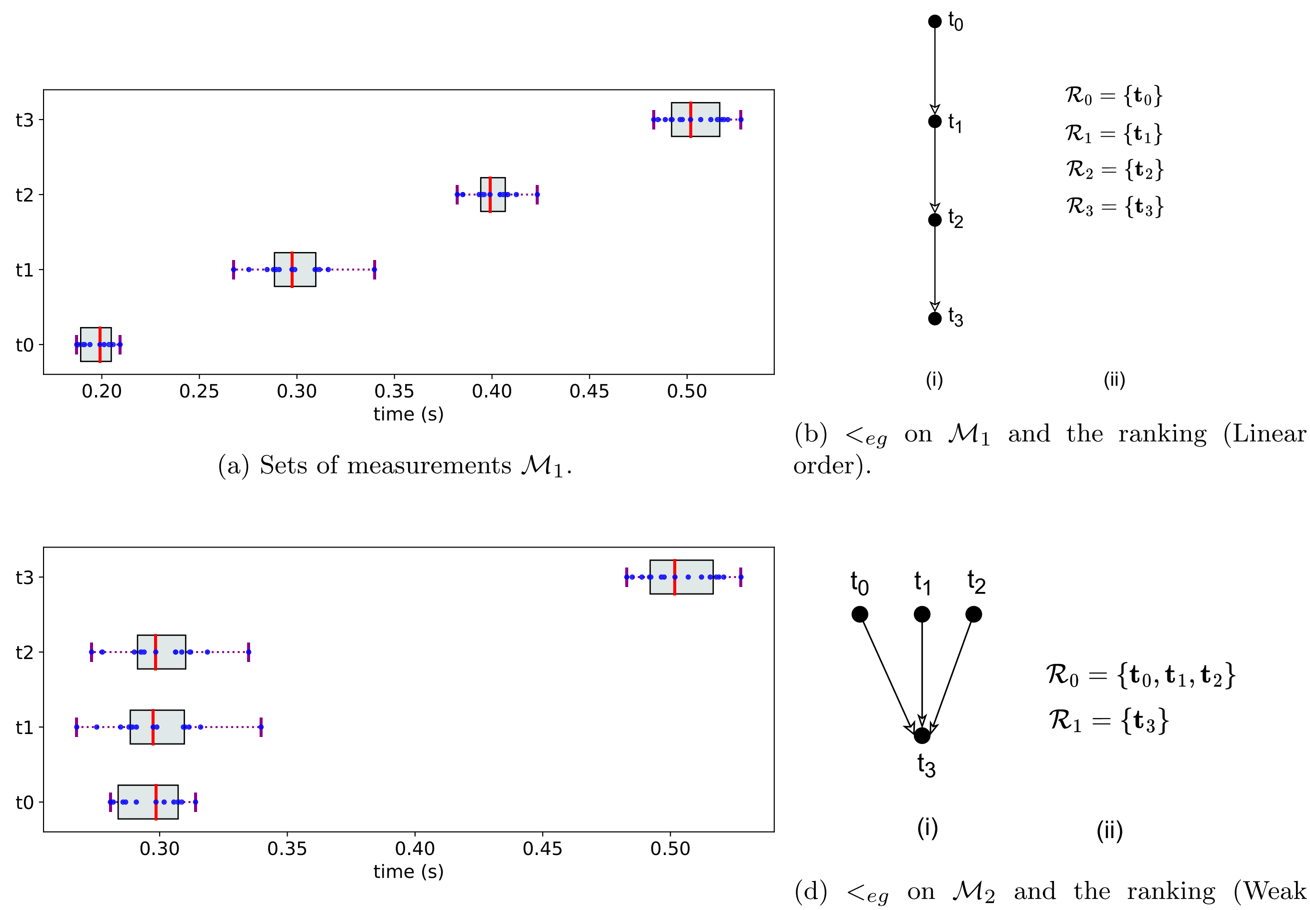
Although rankings based on partial orders have been widely discussed (for example, see the collection of works in [5, 20]), we still notice a lack of clarity on what constitutes a set of reasonable rankings. Therefore, in this paper, we first define what we consider to be the set of reasonable rankings, which we refer to as *partial rankings*. We then develop methodologies to compute partial rankings and explain their applications. The specific contributions of this work are the following:

1. For given sets of measurements and a better-than relation that defines how two sets of measurements should be compared, we define partial ranking to identify a set of reasonable rankings.
2. We develop three different methodologies, each computing a partial ranking (i.e., one of the reasonable rankings) for a given sets of measurements according to a given better-than relation.
3. We show how our partial ranking methodologies can be used to discover the underlying causes of performance differences between the objects.

Organization: In Sect. 2, we explain and formally define partial ranking. In Sect. 3, we discuss the related works. In Sect. 4, we present methodologies for partial ranking of sets of measurements, and in Sect. 5, we extend our methodologies to handle multiple better-than relations. In Sect. 6, we apply the partial ranking methodology in two practical scenarios, and finally, in Sect. 7, we draw conclusions.

2 Partial ranking

Let $\mathcal{M} = \{\mathbf{t}_0, \ldots, \mathbf{t}_{N-1}\}$ be a set of N objects. Let $<_\mathbf{P}$ be a *strict partial order* on $\mathcal{M}$ that models a better-than relation between a pair of objects $\mathbf{t}_i, \mathbf{t}_j \in \mathcal{M}$, i.e., $\mathbf{t}_i <_\mathbf{P} \mathbf{t}_j$ means that $\mathbf{t}_i$ is somehow *better than* (e.g., faster than) $\mathbf{t}_j$. If neither $\mathbf{t}_i <_\mathbf{P} \mathbf{t}_j$ nor $\mathbf{t}_j <_\mathbf{P} \mathbf{t}_i$, then $\mathbf{t}_i$ and $\mathbf{t}_j$ are *incomparable*, which we denote by $\mathbf{t}_i \sim \mathbf{t}_j$. For example, consider the four sets of measurements in Fig. 2, $\mathcal{M}_0 = \{\mathbf{t}_0, \mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3\}$. The strict partial order relation $<_{eg}$ on $\mathcal{M}_0$ is $\{(\mathbf{t}_0, \mathbf{t}_2), (\mathbf{t}_0, \mathbf{t}_3), (\mathbf{t}_2, \mathbf{t}_3), (\mathbf{t}_1, \mathbf{t}_3)\}$. The tuples $(\mathbf{t}_0, \mathbf{t}_1)$ and $(\mathbf{t}_1, \mathbf{t}_2)$ are not in the relation as $\mathbf{t}_0 \sim \mathbf{t}_1$ and $\mathbf{t}_1 \sim \mathbf{t}_2$.



(a) Sets of measurements $\mathcal{M}_1$.

(b) $<_{eg}$ on $\mathcal{M}_1$ and the ranking (Linear order).

(c) Sets of measurements $\mathcal{M}_2$.

(d) $<_{eg}$ on $\mathcal{M}_2$ and the ranking (Weak order).

Fig. 4 When $<_\mathbf{P}$ imposes either a linear order or a weak order on $\mathcal{M}$, then we want the ranking to be unique. Moreover, when the order is linear, then we want the ranking to have no ties

For each $\mathbf{t}_i \in \mathcal{M}$, we aim to assign a rank in the form of a nonnegative integer with 0 representing the highest (best) rank that is consistent with $<_\mathbf{P}$. We first make a distinction between the cases where (1) we want the ranking to be unique and without ties, (2) we want the ranking to be unique but with ties, (3) the ranking cannot be unique but allows ties.

1. **Linear order**: If $\forall \mathbf{t}_i, \mathbf{t}_j \in \mathcal{M}$, either $\mathbf{t}_i <_\mathbf{P} \mathbf{t}_j$ or $\mathbf{t}_j <_\mathbf{P} \mathbf{t}_i$, then $\mathcal{M}$ follows a *linear order* for $<_\mathbf{P}$. In other words, $\nexists \mathbf{t}_i, \mathbf{t}_j \in \mathcal{M}$ such that $\mathbf{t}_i \sim \mathbf{t}_j$. In this case, we want the ranking to be unique and without ties. *Example:* Consider again the four sets of measurements $\mathcal{M}_0$ shown in Fig. 2 and the better-than relation $<_{med}$ where $\mathbf{t}_i <_{med} \mathbf{t}_j$ if and only if the median of $\mathbf{t}_i$ is smaller than the median of $\mathbf{t}_j$. Note that the relation $<_{med}$ on $\mathcal{M}_0$ has no pair of incomparable elements, and therefore follows linear order. The ranking is determined as an ordered set partition—$\mathcal{R}_0 = \{\mathbf{t}_0\}$, $\mathcal{R}_1 = \{\mathbf{t}_1\}$, $\mathcal{R}_2 = \{\mathbf{t}_2\}$, $\mathcal{R}_3 = \{\mathbf{t}_3\}$—of $\mathcal{M}_0$. Consider another example $\mathcal{M}_1$ shown in Fig. 4a and the better-than relation $<_{eg}$. Here, $<_{eg}$ imposes linear order on $\mathcal{M}_1$, and the resulting ranking is shown in Fig. 4b(ii).
2. **Weak order**: If $\forall \mathbf{t}_i, \mathbf{t}_j, \mathbf{t}_k \in \mathcal{M}$ such that $\mathbf{t}_i \sim \mathbf{t}_j$ and $\mathbf{t}_j \sim \mathbf{t}_k$, it holds $\mathbf{t}_i \sim \mathbf{t}_k$ (in other words, the incomparability relation is transitive), then $\mathcal{M}$ follows a *weak order* for $<_\mathbf{P}$. In this case, we want the ranking to be unique, but this time with ties among all objects in the same equivalence class induced by the incomparability relation. *Example:* Consider the four sets of measurements $\mathcal{M}_2$ shown in Fig. 4c ordered by $<_{eg}$. The corresponding relations is shown in Fig. 4d(i). The resulting ranking is illustrated in Fig. 4d(ii). Here, $\mathbf{t}_0, \mathbf{t}_1, \mathbf{t}_2 \in \mathcal{R}_0$ are all pairwise incomparable and all are assigned the same rank.
3. **Neither linear nor weak**: If $\exists \mathbf{t}_i, \mathbf{t}_j, \mathbf{t}_k \in \mathcal{M}$ such that $\mathbf{t}_i \sim \mathbf{t}_j, \mathbf{t}_j \sim \mathbf{t}_k$ and $\mathbf{t}_i <_\mathbf{P} \mathbf{t}_k$ (in other words, the incomparability relation is *not* transitive), then $\mathcal{M}$ is neither linear nor weak for $<_\mathbf{P}$. In this case, the ranking will not be unique, for the reasons outlined in the introduction. *Example*: The sets of measurements $\mathcal{M}_0$ in Fig. 2 follow

neither linear nor weak order for $<_{eg}$. For this case, we expect one of the three rankings shown in Fig. 3b.

We now formulate a definition for ranking with ties that results in unique rankings when $<_{\mathbf{P}}$ imposes a linear or weak order on $\mathcal{M}$, and in more general cases allows only those rankings that are reasonable, i.e., they satisfy the following properties.

Definition 1 *Given a set of objects $\mathcal{M}$ and a strict partial order relation $<_{\mathbf{P}}$ on $\mathcal{M}$, we define a **partial ranking** as an ordered set partition $\mathcal{R}_0, \ldots, \mathcal{R}_{K-1}$ of $\mathcal{M}$ with the following properties:*

1. *If $\mathbf{t}_i$ is ranked higher than $\mathbf{t}_j$, then either $\mathbf{t}_i <_{\mathbf{P}} \mathbf{t}_j$ or $\mathbf{t}_i \sim \mathbf{t}_j$.*
2. *For all pairs of consecutive ranks $(\mathcal{R}_a, \mathcal{R}_{a+1})$, there exists a $\mathbf{t}_i \in \mathcal{R}_a$ and $\mathbf{t}_j \in \mathcal{R}_{a+1}$ such that $\mathbf{t}_i <_{\mathbf{P}} \mathbf{t}_j$.*
3. *For all pairs $\mathbf{t}_i$ and $\mathbf{t}_j$ with the same rank, $\mathbf{t}_i$ and $\mathbf{t}_j$ must be connected in the undirected graph associated with the incomparability relation.*

The three properties serve the following purposes:

- Property 1 ensures that the ranking is consistent with the strict partial order. This property captures the essence of what we consider a reasonable ranking, since it prevents an object that is better than another from being ranked lower than the other.
- Property 2 prohibits splitting up mutually incomparable objects into separate ranks. This property is essential to make the ranking unique for weak orders. For example, without this property, when $\mathcal{M}_2$ is ranked according to $<_{eg}$, the ranking $\mathcal{R}_0 = \{\mathbf{t}_0\}$, $\mathcal{R}_1 = \{\mathbf{t}_1, \mathbf{t}_2\}$, $\mathcal{R}_2 = \{\mathbf{t}_3\}$ would have been considered valid as it satisfies both Property 1 and Property 3. However, this ranking is not valid according to Property 2 because there does not exist a $\mathbf{t}_i \in \mathcal{R}_0$ and $\mathbf{t}_j \in \mathcal{R}_1$ such that $\mathbf{t}_i <_{\mathbf{P}} \mathbf{t}_j$.
- Property 3 ensures that the objects with the same rank cannot be trivially partitioned into separate ranks. This property is essential for the uniqueness for both linear and weak orders, since without this property it would always be possible to assign all objects the same rank. For example, without this property, for $\mathcal{M}_2$ and $<_{eg}$, the ranking $\mathcal{R}_0 = \{\mathbf{t}_0, \mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3\}$ would be considered valid as it satisfies both Property 1 and Property 2. However, according to Property 3, $\mathbf{t}_3$ cannot exist in the same rank as the other objects because $\mathbf{t}_3$ is disconnected in the undirected graph shown in Fig. 5.

In summary, given $\mathcal{M}$ and $<_{\mathbf{P}}$, a ranking that satisfies the properties listed above is called a partial ranking. Now, the problem is to find a procedure that takes $\mathcal{M}$ as input, makes comparisons according to $<_{\mathbf{P}}$ and returns a partial ranking as output.

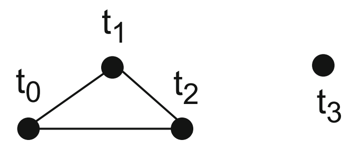


Fig. 5 Undirected graph of the objects in $\mathcal{M}_2$ associated by $\sim$ according to $<_{eg}$

3 Related works

The problem of ranking a finite set of objects based on multiple indicator values has been previously studied. In [18], Patil and Taillie determine a partial ordering of objects and use it to deduce a collection of total orders, also referred to as linear extensions. This approach assigns an interval of possible ranks to each object. However, it has been observed that enumerating all possible linear extensions can become computationally impractical as the number of objects increases, and they resort to Markov chain Monte Carlo methods for rank estimations. In this work, we narrow down the range of possible ranks based on what we consider reasonable for general applications.

The levels in Hasse diagrams have been previously used to interpret the ranks for a large number of objects in several works [6, 16, 29]; i.e., a graph similar to the one shown in Fig. 3a is constructed, and the depth of a node in the graph is interpreted as the rank of that node. Moreover, the term partial ranking has been previously used in some works (e.g., [1, 10, 19]) to denote a ranking deduced from partial orders. However, we notice a lack of discussion on the properties of those rankings. In this work, we elaborate on partial ranking and explicitly list out its properties. When developing ranking methodologies that accommodate ties, it is important to verify if the computed ranking satisfies our properties to determine its reasonableness. For instance, the ranking with ties using the Bubble-sort-based algorithm used in [21, 22] does not always compute a partial ranking according to our definition. Moreover, we want to avoid methodologies that compute some rankings that would have been considered valid without Property 2 and Property 3 of the partial ranking (see the example rankings mentioned while explaining the purposes for Property 2 and Property 3 in Sect. 2).

There is a large body of literature on the area of ranking, particularly tailored for applications in information retrieval and multi-attribute decision-making [2, 5, 20]; for example, among the database community, [12] surveys the top-k ranking techniques for information retrieval and [13] discusses ranking approaches based on datasets that exhibit uncertainty. In this work, we use the ranks of the objects to highlight the similarities and differences in the performance of the objects using a dependency graph based method, thereby enhancing the root cause analysis conducted in [23, 26].

At the core of the ranking, methodology is the comparison function that determines the better-than relation between two objects. Statistical comparison approaches such as [25] often require assumptions to be made regarding the underlying distribution of the measurement values, and the number of measurement values (or the sample size) plays an important role in determining whether or not the resulting rankings are meaningful. However, in many situations, measurements do not follow standard distribution; for instance, the measurements of execution times of a computer program are generally multimodal (i.e., the measurement values occur in stratified clusters, mainly due to the processor operating at multiple frequency levels [7]). As a result, it has been noted that the application of textbook statistical approaches in summarizing such execution time measurements is not straightforward [8, 11]. Comparisons based on nonparametric methods such as the Wilcoxon, the Mann–Whitney U test (e.g., see [9, 15, 24]), interval numbers [28] or gray numbers [14], which do not require any assumption about their underlying distribution and are more suited for handling measurement data with arbitrary noise. In this paper, we do not compare different comparison methods; instead, we focus on developing methodologies for partial ranking that are compatible with any comparison function.

4 Methodologies for partial ranking

In this section, we develop methodologies to compute a partial ranking for a given set of objects $\mathcal{M}$ and a better-than relation $<_\mathbf{P}$. Let each $\mathbf{t}_i \in \mathcal{M}$ consist of M_i measurement values (i.e., $\mathbf{t}_i \in \mathbb{R}^{M_i}$), and we assume that all $M_i \geq 1$.

4.1 Methodology 1: for an arbitrary number of ranks

For a given $\mathcal{M}$ and $<_\mathbf{P}$, let G be a directed graph such that $\mathbf{t}_i \in \mathcal{M}$ are the nodes, and a directed edge from $\mathbf{t}_i$ to $\mathbf{t}_j$ exists if and only if $\mathbf{t}_i <_\mathbf{P} \mathbf{t}_j$. A partial ranking can be computed from G as follows:

Methodology 1 *Given the graph G constructed from the partial order $(\mathcal{M}, <_\mathbf{P})$, set the rank of $\mathbf{t}_i \in \mathcal{M}$ equal to the length of the longest directed path in G that ends at $\mathbf{t}_i$.*

Note that G does not contain any cycles as $<_\mathbf{p}$ is transitive, and therefore, G is a directed acyclic graph. As a consequence, the length of the longest directed path in G that ends at $\mathbf{t}_i$ (or the *depth* of $\mathbf{t}_i$ in G), denoted as $d(\mathbf{t}_i)$, can be computed using the following recursive formula:

$$d(\mathbf{t}_i) = \begin{cases} \max\limits_{\mathbf{t}_j \in \bullet \mathbf{t}_i} d(\mathbf{t}_j) + 1, & \text{if } |\bullet \mathbf{t}_i| > 0 \\ 0, & \text{if } |\bullet \mathbf{t}_i| = 0 \end{cases} \tag{1}$$

where $\bullet\mathbf{t}_i$ is the set of incoming edges to $\mathbf{t}_i$. Then, the ordered set partition of $\mathcal{M}$ is:

$$\mathcal{R}_k = \{\mathbf{t}_i \in \mathcal{M} \mid d(\mathbf{t}_i) = k\} \qquad \forall k \in \{0, \ldots, K-1\} \tag{2}$$

where $K - 1$ is the length of the longest directed path in G.

For example, consider the sets of measurements $\mathcal{M}_3$ in Fig. 6 and the relation $<_{eg}$. $\forall \mathbf{t}_i, \mathbf{t}_j \in \mathcal{M}_3$, we perform pairwise comparison according to $<_{eg}$ and construct G (shown in Fig. 7a). According to G, the depth of $\mathbf{t}_0$ and $\mathbf{t}_2$ is 0, the depth of $\mathbf{t}_1$ is 1, and the depth of $\mathbf{t}_3$ is the maximum of {2, 1} which is 2. Hence, the ordered set partition is $\mathcal{R}_0 = \{\mathbf{t}_0, \mathbf{t}_2\}$, $\mathcal{R}_1 = \{\mathbf{t}_1\}$, $\mathcal{R}_2 = \{\mathbf{t}_3\}$. Note that performing transitivity reduction on a directed graph removes redundant edges without altering the depth of the nodes. Consequently, it does not impact the ranking of the nodes. The transitive reduction of G is shown in Fig. 7b.

Lemma 1 *Rankings produced by Methodology 1 are partial rankings (Def. 1).*

Proof We need to prove that the ranking follows the three properties in Definition 1.

1. Proof for Property 1:
   - $\forall \mathbf{t}_i \in \mathcal{R}_a$ and $\forall \mathbf{t}_j \in \mathcal{R}_b, a < b \implies d(\mathbf{t}_i) < d(\mathbf{t}_j)$
   - $d(\mathbf{t}_i) < d(\mathbf{t}_j) \implies \mathbf{t}_i <_\mathbf{P} \mathbf{t}_j$ or $\mathbf{t}_i \sim \mathbf{t}_j$.
2. Proof for Property 2: $\forall \mathcal{R}_a, \mathcal{R}_{a+1} \implies$ there exists a directed edge from some $\mathbf{t}_i \in \mathcal{R}_a$ to some $\mathbf{t}_j \in \mathcal{R}_{a+1}$, which means $\exists \mathbf{t}_i <_\mathbf{P} \mathbf{t}_j$.
3. Proof for Property 3: Observe that in a directed acyclic graph, two nodes at the same depth cannot be connected by an edge, because if they were, then one of the nodes would be at a depth greater than the other node. Thus, $\forall \mathcal{R}_a, \mathbf{t}_i, \mathbf{t}_j \in \mathcal{R}_a \implies \mathbf{t}_i \sim \mathbf{t}_j$, so all the objects in $\mathcal{R}_a$ are connected in the undirected graph associated by $\sim$.

□

Computational complexity: Let $|\mathcal{M}| = n$ be the total number of objects and e be the total number of pairs of objects related by $<_\mathbf{P}$. The input to this methodology is the graph G. The construction of G requires pairwise comparisons of the objects, which has a complexity of $O(n^2)$. Given G, this methodology computes the depths of the nodes using the depth-first search algorithm, which has a worst-case complexity of $O(n + e)$ (assuming that each node is visited once

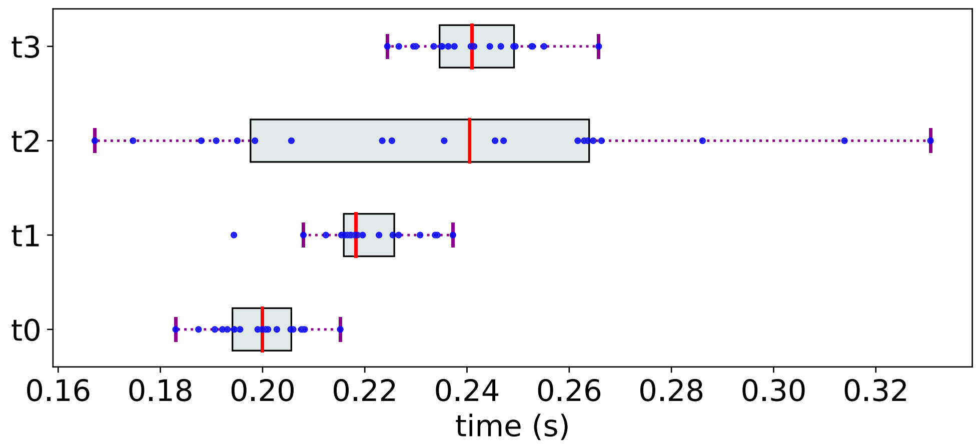
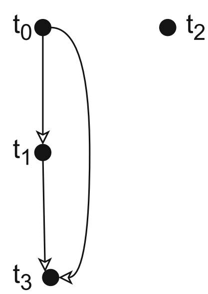
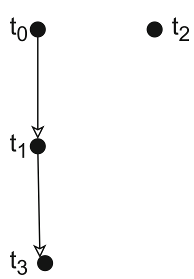


Fig. 6 Sets of measurements $\mathcal{M}_3$. The box indicates the IQI



(a) The graph G associated with $(\mathcal{M}_3, <_{eg})$



(b) Transitive reduction of G.

Fig. 7 Directed graphs from the partial order $(\mathcal{M}_3, <_{eg})$

and the resulting depths are stored in memory). Thus, the overall complexity is dominated by the $O(n^2)$ term.

Sparsification of the directed graph: Note that, in G, any directed edge from $\mathbf{t}_i$ to $\mathbf{t}_j$ such that $\mathbf{t}_i$ is at depth k and $\mathbf{t}_j$ is at a depth greater than $k+1$ is not contained in the longest directed path to $\mathbf{t}_j$ in G. Moreover, any modifications made to G that do not impact the depths of its nodes have no influence on the calculated ranks according to Methodology 1. Bearing this in mind, we construct a graph H from G by removing all edges $(\mathbf{t}_i, \mathbf{t}_j)$ in G where $d(\mathbf{t}_j) - d(\mathbf{t}_i) > 1$. The construction of H ensures that the nodes at a specific depth are exclusively connected by directed edges only to nodes at the subsequent depth.

For example, consider the 10 variants consisting of measurements sampled from the following distribution functions:

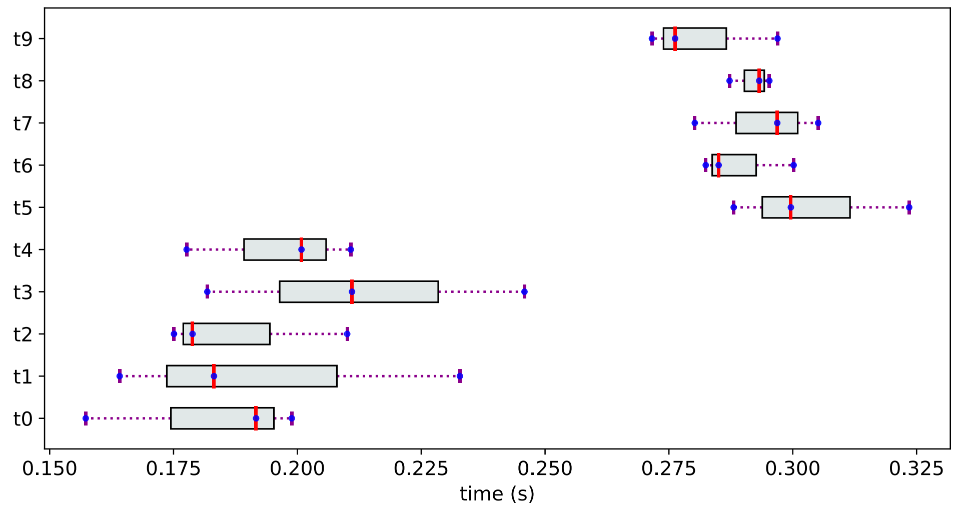
- $\mathbf{t}_0, \mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3, \mathbf{t}_4$: Sampled from $\mathcal{N}(0.2, 0.02)$
- $\mathbf{t}_5, \mathbf{t}_6, \mathbf{t}_7, \mathbf{t}_8, \mathbf{t}_9$: Sampled from $\mathcal{N}(0.3, 0.02)$

For each variant, we sample three measurement values (with initialization seed set to 159) and prepare the sets of measurements $\mathcal{M}_4$ (shown in Fig. 8a). The partial rankings computed according to $<_{eg}$ are shown in Fig. 8b. Note that there is no path from $\mathbf{t}_4$ to $\mathbf{t}_9$ even though $\mathbf{t}_4 <_{eg} \mathbf{t}_9$ because the difference in the depths of $\mathbf{t}_4$ and $\mathbf{t}_9$ is 2. Existence of the edge from $\mathbf{t}_4$ to $\mathbf{t}_9$ would not have changed the length of the longest directed path to $\mathbf{t}_9$.

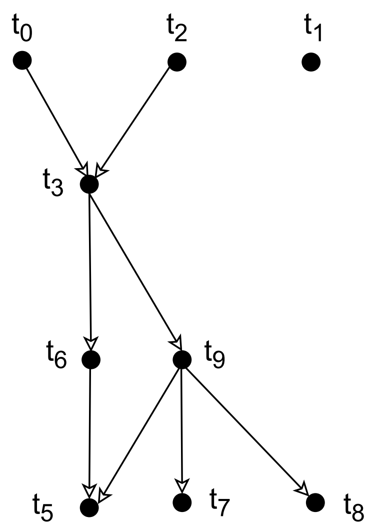
Among the three methodologies, this methodology computes a partial ranking with the highest number of ranks. To determine if a partial ranking with a reduced number of ranks is feasible, one should refer to Methodology 2 or Methodology 3.

4.2 Methodology 2: reduction in the number of ranks

We now explain how the partial ranking produced by Methodology 1 (sparsified) can be modified to determine an alternate partial ranking that might reduce the total number of ranks. Given the graph H constructed from $\mathcal{M}$ and $<_{\mathbf{P}}$, the ordered set partition $\mathcal{R}_0, \ldots, \mathcal{R}_k, \ldots, \mathcal{R}_{K-1}$ consisting of K ranks (where K is the length of the longest directed path in H) is computed using Eq. 2. In order to determine if it is possible



(a) Sets of measurements $\mathcal{M}_4$.



(b) The graph H associated with $\mathcal{M}_4$ and $<_{eg}$.
$\mathcal{R}_0 = \{\mathbf{t}_0, \mathbf{t}_2, \mathbf{t}_1, \mathbf{t}_4\}$, $\mathcal{R}_1 = \{\mathbf{t}_3\}$, $\mathcal{R}_2 = \{\mathbf{t}_6, \mathbf{t}_9\}$, $\mathcal{R}_3 = \{\mathbf{t}_5, \mathbf{t}_7, \mathbf{t}_8\}$

Fig. 8 An example to illustrate sparsification of G

to produce an alternate partial ranking with smaller K, we apply the following methodology:[1]

Methodology 2 *Given the graph H and the corresponding ordered set partition $\mathcal{R}_0, \dots, \mathcal{R}_{K-1}$, perform the following steps:*

1. *For all the nodes $\mathbf{t}_i$ in H, compute the number of incoming and outgoing edges.*
2. *$\forall \mathcal{R}_k$, form a list[2]1 $\mathbf{R}_k$ by arranging the objects in $\mathcal{R}_k$ according to decreasing number of outgoing edges in H. If there are objects with the same number of outgoing edges, arrange the objects with ties according to increasing number of incoming edges in H.*
3. *Concatenate the lists into $\mathbf{T} = \mathbf{R}_0 \oplus \mathbf{R}_1 \oplus \dots \oplus \mathbf{R}_{K-1}$. Note that the objects in $\mathbf{T}$ are arranged left to right from highest to lowest ranks.*
4. *Each $\mathbf{T}[i]$ is assigned a rank $R[i]$ as follows:*
   - *(a)* ***set*** *$R[0] = 0$ (i.e., $\mathbf{T}[0]$ is assigned the best rank).*
   - *(b)* ***for*** *$i = 1, \dots N$:*
     - *(i)* ***if*** *$\mathbf{T}[i-1] <_{\mathbf{P}} \mathbf{T}[i]$* ***then set*** *$R[i] = R[i-1] + 1$.*
     - *(ii)* ***else set*** *$R[i] = R[i-1]$.*

According to Step 4b of Methodology 2, adjacent objects in $\mathbf{T}$ that are incomparable to one another are assigned the same rank. Let $\hat{K}$ be the number of unique values in R. The ordered set partition of $\mathcal{M}$ consists of sets within which all the objects are assigned the same rank; i.e.,

$$\hat{\mathcal{R}}_k = \{\mathbf{T}[i] \in \mathbf{T} \mid R[i] = k\} \qquad \forall k \in \{0, \dots, \hat{K} - 1\} \tag{3}$$

and $\hat{\mathcal{R}}_k$ consists of objects that receive the rank k. The new number of ranks $\hat{K}$ is smaller than or equal to K.

Illustrative Example: Consider the graph H constructed from the same sets of measurements $\mathcal{M}_4$, as those used to illustrate Methodology 1. The ordered set partition $\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3$ according to Methodology 1 is shown in Fig. 8b. To find an alternate partial ranking with fewer than four ranks, we apply Step 2 of Methodology 2 and obtain the following arrangements:

- $\mathbf{R}_0 = [\mathbf{t}_0, \mathbf{t}_2, \mathbf{t}_1, \mathbf{t}_4]$
- $\mathbf{R}_1 = [\mathbf{t}_3]$
- $\mathbf{R}_2 = [\mathbf{t}_9, \mathbf{t}_6]$
- $\mathbf{R}_3 = [\mathbf{t}_7, \mathbf{t}_8, \mathbf{t}_5]$

The concatenated list according to Step 3 is:

$$\mathbf{T} = [\mathbf{t}_0, \mathbf{t}_2, \mathbf{t}_1, \mathbf{t}_4, \mathbf{t}_3, \mathbf{t}_9, \mathbf{t}_6, \mathbf{t}_7, \mathbf{t}_8, \mathbf{t}_5] \tag{4}$$

The ranking computed according to Methodology 2 with the list $\mathbf{T}$ in Eq. (4) is shown in Table 1, which now consists of only two ranks.

In the step 2 of Methodology 2, if $\exists \mathbf{t}_i \in \mathcal{R}_k$ and $\exists \mathbf{t}_j \in \mathcal{R}_{k+1}$ such that $\mathbf{t}_i \sim \mathbf{t}_j$, then $\mathbf{t}_i$ is pushed toward the right in the list $\mathbf{R}_k$, and $\mathbf{t}_j$ is pushed toward the left in the list $\mathbf{R}_{k+1}$, and then, the ranks $\forall \mathbf{t}_i \in \mathcal{R}_k$ and $\forall \mathbf{t}_j \in \mathcal{R}_{k+1}$ could be merged according to Step 4. For illustration, let us denote the last element in a list $\mathbf{R}_k$ as $\mathbf{R}_k[-1]$. In our example, as $\mathbf{R}_0[-1] = \mathbf{t}_4 \sim \mathbf{t}_3 = \mathbf{R}_1[0]$, the concatenation of $\mathbf{R}_0$ and $\mathbf{R}_1$ merges the ranks of the variants in $\mathcal{R}_0$ and $\mathcal{R}_1$. Similarly, as $\mathbf{R}_2[-1] = \mathbf{t}_6 \sim \mathbf{t}_7 = \mathbf{R}_3[0]$, the ranks of the variants in $\mathcal{R}_2$ and $\mathcal{R}_3$ are also merged. For the sake of clarity, the sets of measurements $\mathcal{M}_4$ are shown again in Fig. 9, but now with the y-axis rearranged from bottom to top based on the position of the variants in the list $\mathbf{T}$, starting from $\mathbf{T}[0]$, and the ranking is annotated.

Lemma 2 *Rankings produced by Methodology 2 are partial rankings (Definition 1).*

Proof We need to prove that the ranking follows the three properties in Definition 1.

1. Proof for Property 1:
   - $\forall \mathbf{T}[i] \in \hat{\mathcal{R}}_a$ and $\forall \mathbf{T}[j] \in \hat{\mathcal{R}}_b$, $a < b \implies R[i] < R[j]$ (Eq. 3).
   - $R[i] < R[j] \implies i < j$ (Meth. 2: Step 4b)
   - $i < j \implies \mathbf{T}[i] <_{\mathbf{P}} \mathbf{T}[j]$ or $\mathbf{T}[i] \sim \mathbf{T}[j]$ (Meth. 2: Step 3; the objects in $\mathbf{T}$ are arranged from left to right from highest to lowest ranks computed according to Methodology 1).

   Hence, $\forall \mathbf{T}[i] \in \hat{\mathcal{R}}_a$ and $\forall \mathbf{T}[j] \in \hat{\mathcal{R}}_b$, $a < b \implies \mathbf{T}[i] <_{\mathbf{P}} \mathbf{T}[j]$ or $\mathbf{T}[i] \sim \mathbf{T}[j]$.
2. Proof for Property 2: According to Meth. 2: Step 4b(i), a new rank is created only when there exists i such that $\mathbf{T}[i-1] <_{\mathbf{P}} \mathbf{T}[i]$. Hence, for every pair of ranks $\hat{\mathcal{R}}_a, \hat{\mathcal{R}}_b$, it is possible to find a pair $\mathbf{t}_i \in \hat{\mathcal{R}}_a$ and $\mathbf{t}_j \in \hat{\mathcal{R}}_b$ such that $\mathbf{t}_i <_{\mathbf{P}} \mathbf{t}_j$.
3. Proof for Property 3: $|\hat{\mathcal{R}}_a| > 1 \implies \exists k, l$ such that $\mathbf{T}[k] \sim \mathbf{T}[k+1] \sim \dots \sim \mathbf{T}[l]$, and all the objects from positions k to l in $\mathbf{T}$ are the only objects in $\hat{\mathcal{R}}_a$ (Meth. 2: Step 4).
   Hence, there exists an arrangement of $\hat{\mathcal{R}}_a$ where the adjacent objects are pairwise incomparable. This means

[1] In order to be able to index the elements in the set $\mathcal{R}_k$, we introduce the list $\mathbf{R}_k$. The element at position i (zero-based indexing) is denoted by $\mathbf{R}_k[i]$.

[2] In order to be able to index the elements in the set $\mathcal{R}_k$, we introduce the list $\mathbf{R}_k$. The element at position i (zero-based indexing) is denoted by $\mathbf{R}_k[i]$.

Table 1 Partial ranking of $\mathcal{M}_4$ with $<_{eg}$ according to Methodology 2

T	**T**[0]		**T**[1]		**T**[2]		**T**[3]		**T**[4]		**T**[5]		**T**[6]		**T**[7]		**T**[8]		**T**[9]
	$\mathbf{t}_0$	$\sim$	$\mathbf{t}_2$	$\sim$	$\mathbf{t}_1$	$\sim$	$\mathbf{t}_4$	$\sim$	$\mathbf{t}_3$	$<_{eg}$	$\mathbf{t}_9$	$\sim$	$\mathbf{t}_6$	$\sim$	$\mathbf{t}_7$	$\sim$	$\mathbf{t}_8$	$\sim$	$\mathbf{t}_5$
R	$R[0]$		$R[1]$		$R[2]$		$R[3]$		$R[4]$		$R[5]$		$R[6]$		$R[7]$		$R[8]$		$R[9]$
	0		0		0		0		0		1		1		1		1		1
$\hat{\mathcal{R}}_k$	$\hat{\mathcal{R}}_0 = \{\mathbf{t}_0, \mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3, \mathbf{t}_4\}$, $\hat{\mathcal{R}}_1 = \{\mathbf{t}_5, \mathbf{t}_6, \mathbf{t}_7, \mathbf{t}_8, \mathbf{t}_9\}$																		

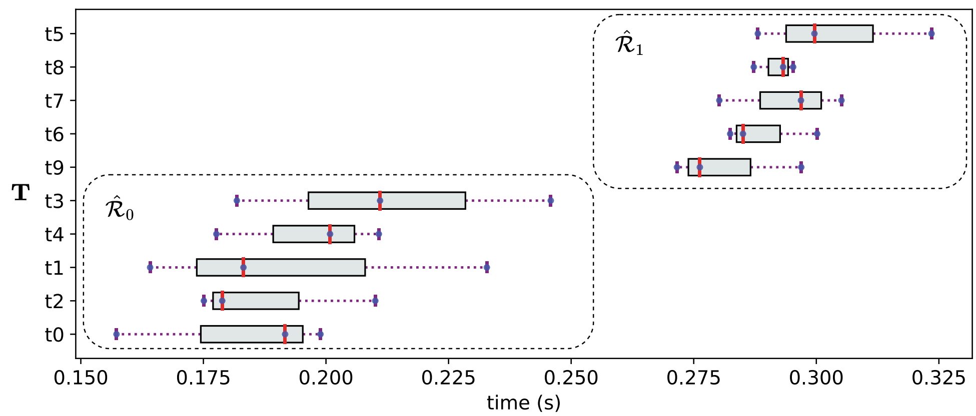


Fig. 9 The annotation of the partial ranking on $\mathcal{M}_4$ with $<_{eg}$ according to Methodology 2.

that the objects in $\hat{\mathcal{R}}_a$ are connected in the undirected graph associated by $\sim$.

□

The number of ranks computed by Methodology 2 is either smaller than or equal to the number of ranks computed according to Methodology 1, but not necessarily the least possible number of ranks.

Computational complexity: The input to this methodology is the graph H, which has a complexity of $O(n^2)$ for construction, and the ordered set partition consisting of K ranks computed according to Methodology 1, which has a complexity of $O(n + e)$. Assuming the number of incoming and outgoing edges per node is stored in memory during the construction of H, this methodology requires $O(m \log m)$ operations to sort the nodes within each of the K ranks, where m is the maximum number of nodes per rank, and $O(n)$ operations to reestimate the ranks. The overall complexity remains dominated by $O(n^2)$.

Instances where the least number of ranks is not computed: In certain cases, Methodology 2 does not yield a partial ranking with the minimum possible number of ranks. For example, consider the sets of measurements $\mathcal{M}_5$ shown in Fig. 10. Given $\mathcal{M}_5$ and $<_{eg}$, the arrangement **T** and the annotations of the partial ranking calculated according to Methodology 2 are shown in Fig. 10a. While this partial ranking consists of three ranks, it is possible to form an alternate arrangement $\mathbf{T}'$ (shown in Fig. 10b), that produces a partial ranking with just two ranks according to Methodology 2 (Step 4).

A sufficient condition for which Methodology 2 will not produce a partial ranking with the least number of ranks is when $\exists \mathbf{t}_i \in \mathcal{R}_k$ and $\exists \mathbf{t}_j \in \mathcal{R}_{k+2}$ such that $\mathbf{t}_i \sim \mathbf{t}_j$, but $\nexists \mathbf{t}_m \in \mathcal{R}_{k+1}$ such that $\mathbf{t}_m \sim \mathbf{t}_j$.

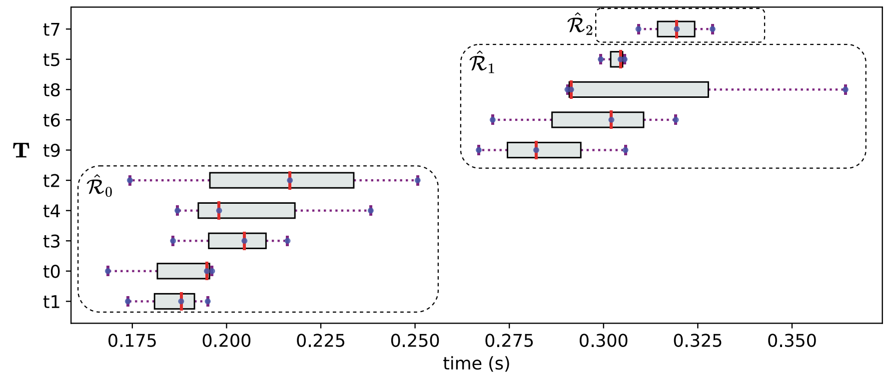
For example, in $\mathcal{M}_5$, the directed graph G (with transitivity reduction) is shown in Fig. 10c (by removing the edge highlighted in red, the graph H is obtained). According to Methodology 1, the number of ranks is 5, and we just saw that it possible to have a partial ranking consisting of 2 ranks. Methodology 2 cannot reduce the number of ranks to smaller than 3 because $\exists \mathbf{t}_8 \in \mathcal{R}_2$ and $\exists \mathbf{t}_7 \in \mathcal{R}_4$ such that $\mathbf{t}_8 \sim \mathbf{t}_7$,

but $\nexists \mathbf{t}_m \in \mathcal{R}_3$ such that $\mathbf{t}_m \sim \mathbf{t}_7$. As a consequence, according to Methodology 2 (Step 3), $\mathbf{t}_7 \in \mathbf{R}_4$ cannot be placed next to $\mathbf{t}_8 \in \mathbf{R}_2$ because $\mathbf{t}_5 \in \mathbf{R}_3$ has to be placed between $\mathbf{t}_8$ and $\mathbf{t}_7$, and hence the ranks of $\mathbf{t}_8$ and $\mathbf{t}_7$ cannot be merged.

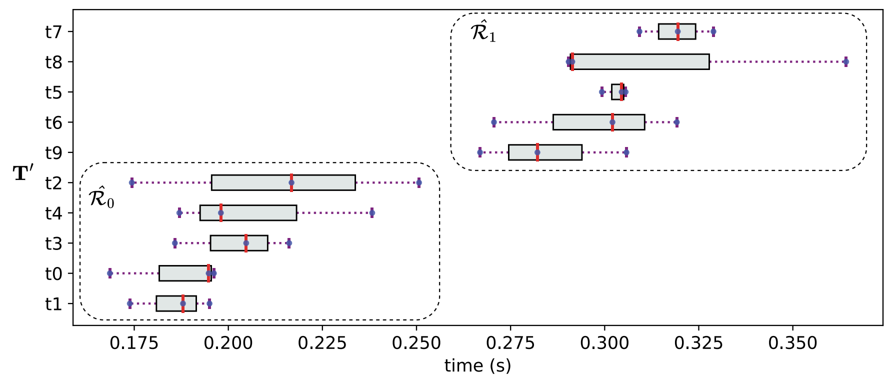
4.3 Methodology 3: for minimum number of ranks

We now explain a methodology for computing the partial ranking with the minimum number of ranks. For a given $\mathcal{M}$ and $<_{\mathbf{P}}$, let U be an undirected graph such that $\mathbf{t}_i \in \mathcal{M}$ are the nodes and an edge between $\mathbf{t}_i$ and $\mathbf{t}_j$ exists if and only if $\mathbf{t}_i \sim \mathbf{t}_j$ according to $<_{\mathbf{P}}$.

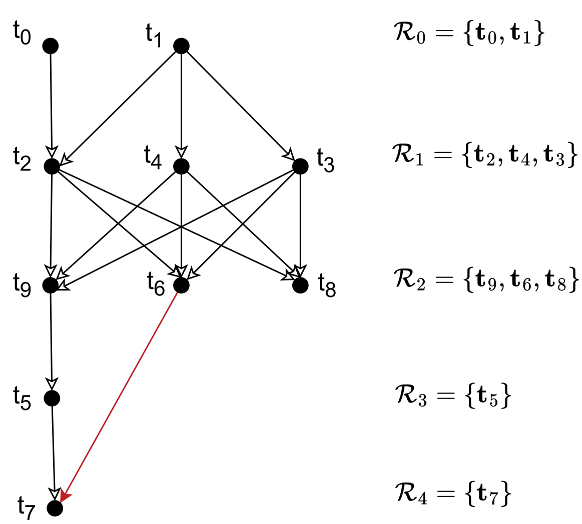
Fig. 10 Partial ranking of $\mathcal{M}_5$ with $<_{eg}$. An illustrative example to demonstrate that Methodology 2 does not always find the partial ranking with the least number of ranks



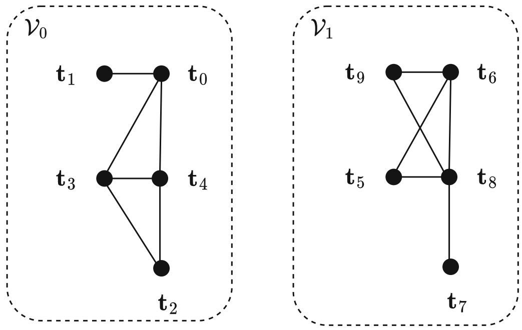
(a) Methodology 2 on $\mathcal{M}_5$.



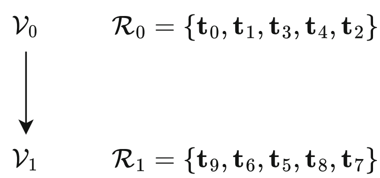
(b) Methodology 2 on $\mathcal{M}_5$ with arrangement $\mathbf{T}'$.



(c) Methodology 1 on $\mathcal{M}_5$. The transitivity reduction of graph G for $\mathcal{M}_5$ according to $<_{eg}$. By removing the edge indicated in red, the graph H is obtained.



(a) The undirected graph U and the connected components $\mathcal{V}_0$ and $\mathcal{V}_1$.



(b) The graph G' and the partial ranks.

Fig. 11 Methodology 3 on $\mathcal{M}_5$ and $<_{eg}$

Methodology 3 *Given the graph U constructed from the partial order $(\mathcal{M}, <_{\mathbf{P}})$,*

1. *Partition U to create $\mathcal{U} = \{\mathcal{V}_0, \ldots, \mathcal{V}_{K-1}\}$ such that each partition $\mathcal{V}_k \in \mathcal{U}$ corresponds to the connected components of U.*
2. *Construct a directed graph G' such that $\mathcal{V}_i \in \mathcal{U}$ are the nodes and an edge from $\mathcal{V}_i$ to $\mathcal{V}_j$ exists if and only if for any $\mathbf{t}_k \in \mathcal{V}_i$ and $\mathbf{t}_l \in \mathcal{V}_j$, $\mathbf{t}_k <_{\mathbf{P}} \mathbf{t}_l$.*
3. *$\forall \mathcal{V}_i \in \mathcal{U}$, set the rank of all $\mathbf{t}_k \in \mathcal{V}_i$ equal to the depth of $\mathcal{V}_i$ in G'.*

Illustrative example: Consider again the sets of measurements $\mathcal{M}_5$ and the relation $<_{eg}$. According to Step 1, the undirected graph U consists of two partitions $\mathcal{V}_0$ and $\mathcal{V}_1$ (shown in Fig. 11a). The directed graph G' and the resulting ordered set partitioning are shown in Fig. 11b.

To compare with other methodologies, consider again the sets of measurements in Fig. 8 ($\mathcal{M}_4$) used to illustrate Methodology 1 and 2. Methodology 3 produces the same partial ranking as Methodology 2 (i.e., the ordered set partitioning annotated in Fig. 9).

Lemma 3 *Rankings produced by Methodology 3 are partial rankings (Def. 1) and consist of minimum possible number of ranks.*

Proof Notice that $\forall \mathcal{R}_a, \mathcal{R}_b, a < b \implies \forall \mathbf{t}_k \in \mathcal{R}_a$ and $\forall \mathbf{t}_l \in \mathcal{R}_b$, $\mathbf{t}_k <_{\mathbf{P}} \mathbf{t}_l$. Thus, the requirements according to Property 1 and Property 2 are satisfied. The condition according to Property 3 is naturally satisfied as all the variants in a particular partition $\mathcal{V}_i$ are connected by the incomparability relation $\sim$. Hence, the ranking computed according to Methodology 3 is a partial ranking.

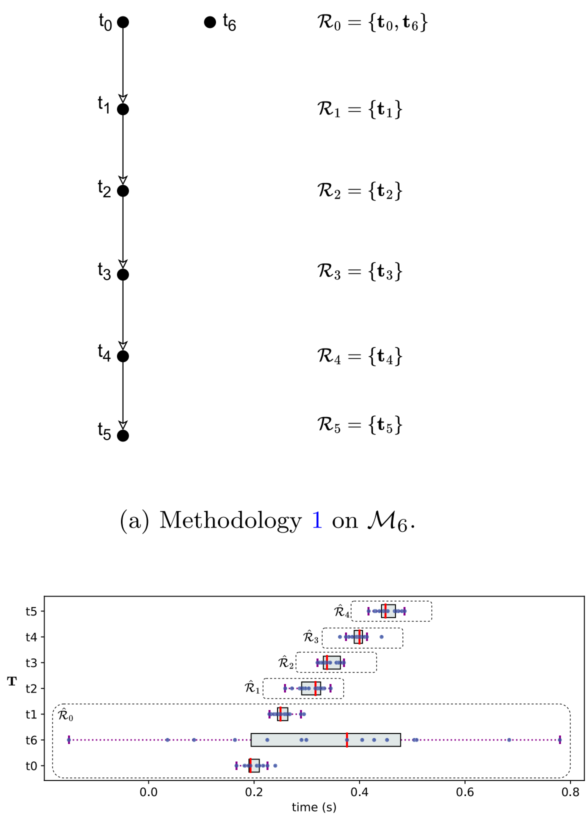
We prove by contradiction that the partial ranking according to Methodology 3 consists of the minimum number of ranks. Suppose that there exists some partial ranking with strictly fewer than K ranks. Then, by the pigeonhole principle, at least one rank must contain a pair of variants that are disconnected in the graph associated with the incomparability relation. This violates Property 3 and hence there cannot exist any partial ranking with fewer than K ranks. □

Computational complexity: Let n be the total number of objects and e be the total number of pairs of objects that are incomparable to each other. The input to this methodology is the graph U, which has a complexity of $O(n^2)$ for construction. Assuming that the set of incomparable nodes for each node is stored in memory during the construction of U, this methodology has a complexity of $O(n+e)$. The overall complexity is dominated by $O(n^2)$.

4.4 Comparison of partial ranking methodologies and implications on some special cases

Recall that for a set of measurements and a better-than relation, more than one partial rankings are possible. We presented three different methodologies for partial ranking. Among the three methodologies, Methodology 1 produces a partial ranking with the highest number of ranks, and Methodology 3 produces a partial ranking with the least number of ranks. Methodology 2 can be seen as a trade-off between the other two. In this subsection, we apply and compare Methodologies 1, 2 and 3 on some special cases.

Consider that there are numerous variants with their measurement intervals distinctly separated from each other, except for a single variant whose interval overlaps with all the others. For example, see the sets of measurements $\mathcal{M}_6$ shown in Fig. 12b. Each variant consists of 15 measurement values and the IQI of all the variants except for $\mathbf{t}_6$ is distinctly separated from one another. In this case, Methodology 3 merges all the variants into one rank. Such a ranking is sometimes undesirable because it is not possible to discriminate between any of the distinctly separated variants. Methodology 1, however, produces a partial ranking with six ranks as shown in Fig. 12a, and this is the maximum number of ranks that could be produced for $\mathcal{M}_6$ with $<_{eg}$; $\mathbf{t}_0, \mathbf{t}_6 \in \mathcal{R}_0$ and all the other variants in distinct ranks. Methodology 2 reduces the number of ranks by 1, and the resulting partial rankings are annotated in Fig. 12b.



(a) Methodology 1 on $\mathcal{M}_6$.

(b) Methodology 2 on $\mathcal{M}_6$.

Fig. 12 Partial ranking of $\mathcal{M}_6$ with $<_{eg}$

Let us examine one more example that showcases variants with overlapping yet monotonically increasing differences, such as the sets of measurements $\mathcal{M}_7$ depicted in Fig. 13. In this case, although assigning a distinct rank to each variant seems more intuitive , both Methodology 2 and Methodology 3 produce a partial ranking where all the variants are placed in the same rank (as shown in Fig. 13b). Methodology 1 computes a partial ranking with 2 ranks (as shown in Fig. 13a). Notice that Methodology 1 does not compute the maximum possible number of ranks, because for $\mathcal{M}_7$, the following partial ranking with three ranks exists: $\mathcal{R}_0 = \{\mathbf{t}_0, \mathbf{t}_1\}$, $\mathcal{R}_1 = \{\mathbf{t}_2, \mathbf{t}_3\}$ and $\mathcal{R}_2 = \{\mathbf{t}_4\}$.

5 Handling the effects of the quantile intervals

Thus far, we have utilized the relation $<_{eg}$ to compare two sets of measurements, which makes comparisons based on overlap in the inter-quartile intervals (IQIs), or in other words, based on overlap in the inter-quantile interval (IQnI) between the 25th and 75th quantiles among the variants. Sometimes, altering the quantiles can lead to a different ranking. For

Table 2 Reliability Scores of the ranks computed by Methodology 1 on $\mathcal{M}_8$ with $<_{(l,u)}$ at different quantile limits

	$\mathbf{t}_0$	$\mathbf{t}_1$	$\mathbf{t}_2$	$\mathbf{t}_3$	avg_rel(l, u)
$r(\mathbf{t}_i, 25, 75)$	0	0	0	1	-
$r(\mathbf{t}_i, 30, 70)$	0	1	0	1	-
$r(\mathbf{t}_i, 35, 65)$	0	1	0	1	-
$r(\mathbf{t}_i, 40, 60)$	1	2	0	2	-
$mr(\mathbf{t}_i)$	0.25	1.00	0.00	1.25	-
$rel(\mathbf{t}_i, 25, 75)$	-0.25	-1.0	0.0	-0.25	**-0.38**
$rel(\mathbf{t}_i, 30, 70)$	-0.25	-0.00	0.0	-0.25	**-0.13**
$rel(\mathbf{t}_i, 35, 65)$	-0.25	-0.00	0.0	-0.25	**-0.13**
$rel(\mathbf{t}_i, 40, 60)$	-0.75	-1.00	0.0	-0.75	**-0.63**

instance, consider the sets of measurements $\mathcal{M}_8$ shown in Fig. 14 where it could be intuitive for one to expect the following ranking:

$$\mathcal{R}_0 = \{\mathbf{t}_0, \mathbf{t}_2\}, \quad \mathcal{R}_0 = \{\mathbf{t}_1, \mathbf{t}_3\}. \tag{5}$$

However, the IQnI between the 25th and the 75th quantile of $\mathbf{t}_1$ *slightly* overlaps with that of $\mathbf{t}_0$ and $\mathbf{t}_2$. As a consequence, when using the relation $<_{eg}$ to compare the variants, $\mathbf{t}_0$, $\mathbf{t}_1$ and $\mathbf{t}_2$ are all pairwise incomparable and obtain the same rank. In order to alleviate this problem, for a given ranking methodology, we compute multiple rankings corresponding to different quantiles, and average them. In what follows, we propose a systematic way of choosing a ranking corresponding to one among several quantile intervals based on the averaged or the mean ranks.

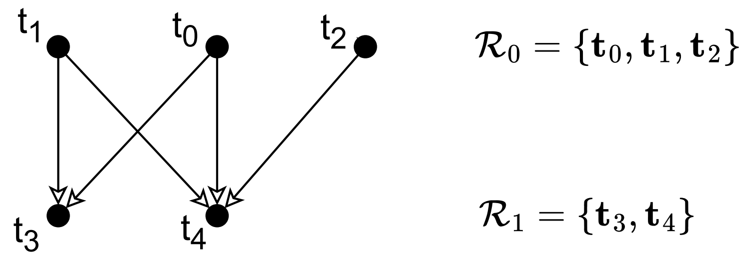
To this end, let us first generalize $<_{eg}$ to accommodate arbitrary quantile limits. Let l and u denote the lower and the upper quantile limit, respectively. For each set of measurements values $\mathbf{t}_i \in \mathbb{R}^M$, sort $\mathbf{t}_i$ in the ascending order and let $t_i^l, t_i^u \in \mathbb{R}$ be the linearly interpolated measurement values at the l and u quantiles, respectively. Then, for each $\mathbf{t}_i$, (t_i^l, t_i^u) is the IQnI between the quantile limits (l, u). We define the relation $<_{(l,u)}$ as:

Definition 2 *For a given* (l, u), $\mathbf{t}_i <_{(l,u)} \mathbf{t}_j$ *if and only if* $t_i^u < t_j^l$.

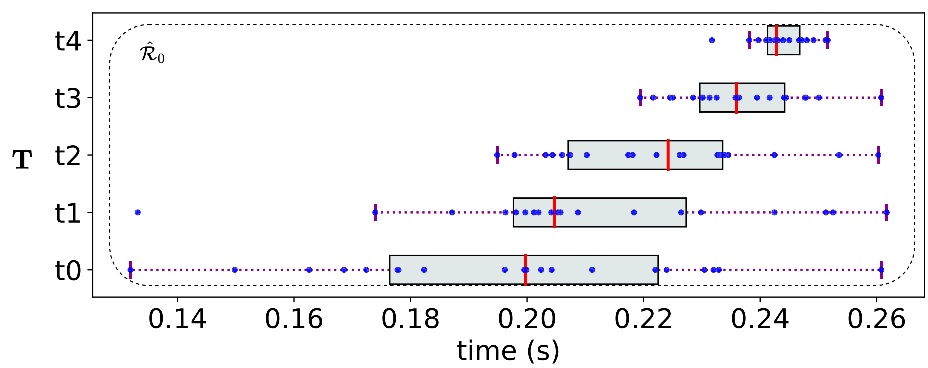
Thus, inferring from Definition 2, two sets of measurements are considered incomparable if and only if their IQnIs overlap with one another. If $l = 25$ and $u = 75$, then IQnI becomes the IQI indicated in our box plots; consequently, $<_{eg}$ is same as $<_{(25,75)}$.

Let $\mathcal{Q}$ be a set of quantile limits. For a given ranking methodology, let $r(\mathbf{t}_i, l, u)$ be the rank calculated for the variant $\mathbf{t}_i \in \mathcal{M}$ at a particular $(l, u) \in \mathcal{Q}$. As the choice of (l, u) influences the partial rank calculation, instead of computing a partial ranking at just one (l, u), we compute the partial

Fig. 13 Partial ranking of $\mathcal{M}_7$ with $<_{eg}$

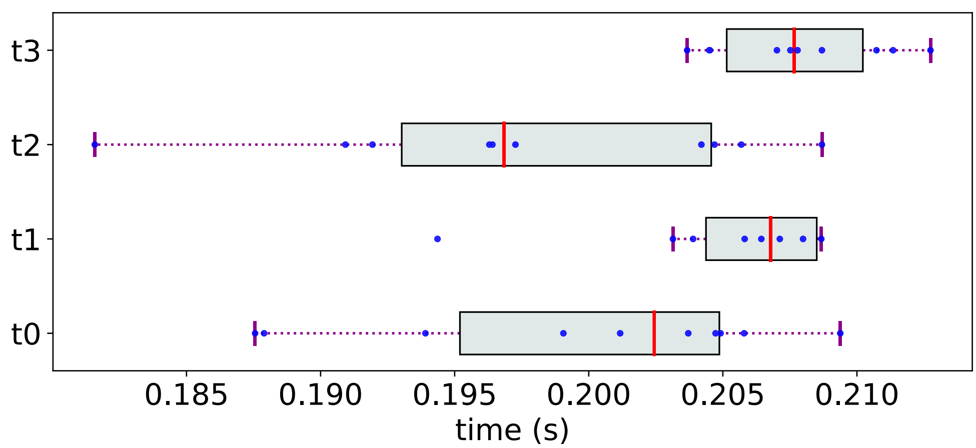


(a) Methodology 1 on $\mathcal{M}_7$.



(b) Methodology 2 on $\mathcal{M}_7$.

Fig. 14 Sets of measurements $\mathcal{M}_8$



ranks at several quantile limits, i.e., $\forall(l, u) \in \mathcal{Q}$, and calculate the mean rank of each variant. Let the mean rank of $\mathbf{t}_i$ be denoted by $mr(\mathbf{t}_i)$. Then, the reliability score $rel(\mathbf{t}_i, l, u)$ of the rank assigned to the variant $\mathbf{t}_i$ at a particular quantile limit (l, u) is defined as:

$$rel(\mathbf{t}_i, l, u) = -|r(\mathbf{t}_i, l, u) - mr(\mathbf{t}_i)| \tag{6}$$

where a higher reliability score can be considered as a better ranking for the given measurement data, a better-than relation and $\mathcal{Q}$. That is, for the variant $\mathbf{t}_i$, the lower the difference of the rank $r(\mathbf{t}_i, l, u)$ calculated at a particular (l, u) from the mean of the ranks $mr(\mathbf{t}_i)$ calculated at several quantile limits, the higher the reliability score $rel(\mathbf{t}_i, l, u)$, indicating that the rank assignment $r(\mathbf{t}_i, l, u)$ is more reliable. The average reliability (avg_rel) of a partial ranking at a quantile limit (l, u) is the mean of the reliability scores at that (l, u) from all the variants; that is,

$$\text{avg\_rel}(l, u) = \frac{\sum_{\mathbf{t}_i} rel(\mathbf{t}_i, l, u)}{|\mathcal{M}|} \tag{7}$$

where $|\mathcal{M}|$ is the total number of objects.

Thus, given $\mathcal{Q}$, we propose to compute partial rankings and average reliability scores $\forall(l, u) \in \mathcal{Q}$, and select a ranking with the highest average reliability score.

Illustrative example: Let us compute partial rankings for $\mathcal{M}_8$ using Methodology 1 at all

$$(l, u) \in \{(25, 75), (30, 70), (35, 65), (40, 60)\}. \tag{8}$$

The mean ranks of the variants, the reliability scores of the rank assignments and the average reliability scores are shown in Table 2. The partial rankings at the limits (30, 70) and (35, 65), which classify the variants into ranks as expected according to Eq. 5, obtain a higher average reliability score than the partial rankings at other limits.

6 Application: mining for the causes of performance differences from the measurement data

Partial ranking of a set of objects can be seen as clustering the objects into performance classes; i.e., a clustering in which there is a notion of one cluster being better than another. In this section, we demonstrate that identifying performance classes using our partial ranking methodologies facilitates the discovery of the causes of performance differences between the variants. We adopt an interactive approach to identify the causes of performance differences. We recommend beginning with the methodology that generates the fewest number of ranks and conducting an initial root cause analysis. If the initial analysis is not satisfactory, apply methods that produce a greater number of ranks and repeat the analysis.

First, we revisit the execution time measurements of the algorithmic variants of the generalized least square (GLS) problem introduced in Sect. 1, and identify the library calls that impact performance the most (Sect. 6.1). Then, we apply our partial ranking methodology on a real-life dataset from a business process application to identify the underlying causes of inefficiencies (Sect. 6.2).

6.1 Algorithmic variants of GLS

The execution time measurements of 10 algorithmic variants of GLS for given matrix dimensions are shown again in Fig. 15. Each algorithmic variant can be identified as a sequence of library or kernel calls provided by optimized libraries, such as BLAS and LAPACK. The sequences of kernel calls for the 10 algorithmic variants as generated by the Linnea compiler [3] are shown in Table 3. Let us first attempt to elucidate the causes of performance differences among the variants in terms of the kernel calls, solely by looking at the data $\mathcal{M}_{gls}$, without relying on any specific ranking methodology. At first glance, one could intuitively categorize the variants into two distinct groups: fast variants and slow variants, as represented below:

$$\begin{aligned} \text{Fast} &= \{\mathbf{alg}_0, \mathbf{alg}_1, \mathbf{alg}_3, \mathbf{alg}_4, \mathbf{alg}_5\} \\ \text{Slow} &= \{\mathbf{alg}_2, \mathbf{alg}_6, \mathbf{alg}_7, \mathbf{alg}_8, \mathbf{alg}_9\} \end{aligned} \tag{9}$$

By carefully observing the similarities and differences in the kernel sequences among the variants in the fast and slow groups based on the data $\mathcal{M}_{gls}$, one could discern the following associations:

(r1) Only the fast variants make use of the kernel *gemm*.
(r2) Only the slow variants make use of the kernel *transpose*.
(r3) Only the slow variants make use of the kernel *qr*.

This example is intentionally simplistic, but in reality, it is common to encounter hundreds of algorithmic variants with much longer kernel sequences for each variant. The manual process of making such discernment, even for this straightforward example, can be time-consuming and laborious. Therefore, given $\mathcal{M}_{gls}$ and the information regarding the kernel sequences in Table 3, our objective is to automate the identification of the observations **r1**, **r2** and **r3** that we just manually discerned, and potentially uncover additional kernel sequence patterns that may have gone unnoticed during the manual analysis.

We begin by applying the methodology that produces the least number of ranks; applying Methodology 3 on $\mathcal{M}_{gls}$ using the better-than relation $<_{eg}$ produces the ranks—$\hat{\mathcal{R}}_0$ and $\hat{\mathcal{R}}_1$—as annotated in Fig. 15. $\hat{\mathcal{R}}_0$ and $\hat{\mathcal{R}}_1$ are incidentally same as the manually identified fast and slow sets (shown in Eq. 9), respectively.

In order to identify the similarities and differences among the variants in $\hat{\mathcal{R}}_0$ and $\hat{\mathcal{R}}_1$ in terms of kernel calls, we construct a directly follows graph (DFG), where each node indicates a kernel, and a directed edge from one kernel to another, say from *kernelA* to *kernelB*, indicates that there exists a kernel sequence in which *kernelA* directly precedes *kernelB*. Given a split of the variants into two sets $\mathcal{G}reen$ and $\mathcal{R}ed$, we color the nodes and edges in the DFG such that green nodes and edges indicate the kernels and the directly follows relations that occur *exclusively* in the variants from $\mathcal{G}reen$. Similarly, red nodes and edges indicate the kernels and the directly follows relations that occur exclusively in the variants from $\mathcal{R}ed$. All the other kernels and relations

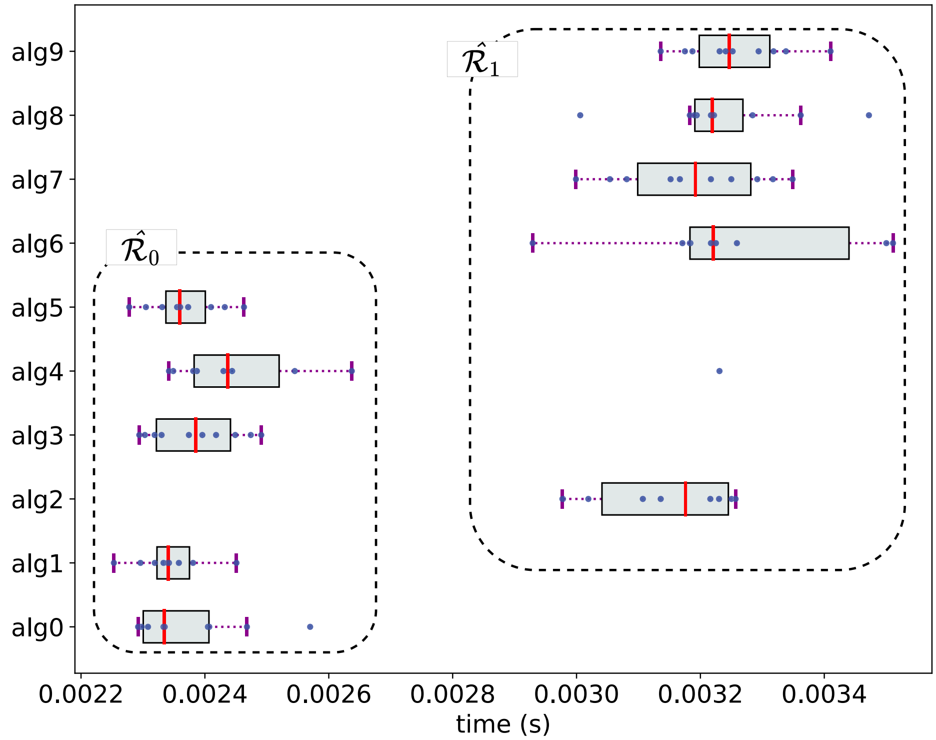


Fig. 15 The time measurements of the algorithmic variants of the GLS problem: $(X^T M^{-1} X)^{-1} X^T M^{-1} \mathbf{y}$ where $X \in \mathbb{R}^{1000\times 100}$, $M \in \mathbb{R}^{1000\times 1000}$ and $\mathbf{y} \in \mathbb{R}^{1000}$ ($\mathcal{M}_{gls}$). The partial rankings using Methodology 3 according to $<_{eg}$ are annotated

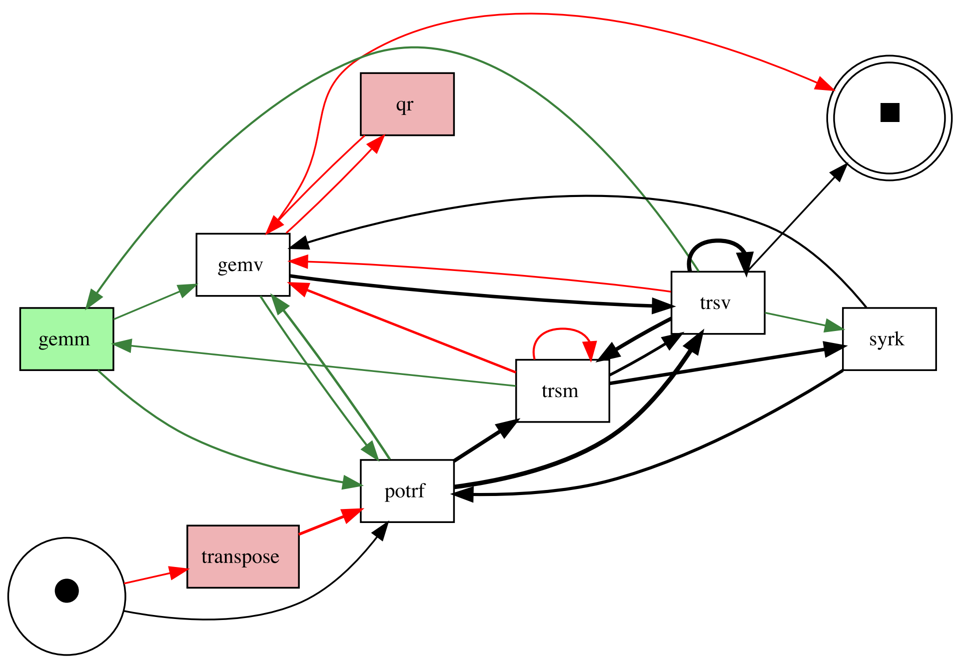


Fig. 16 Directly Follows Graph: $\mathcal{Green} = \hat{\mathcal{R}}_0$, $\mathcal{Red} = \hat{\mathcal{R}}_1$

occur in the variants from both $\mathcal{Green}$ and $\mathcal{Red}$. The DFG with $\mathcal{Green} = \hat{\mathcal{R}}_0$ and $\mathcal{Red} = \hat{\mathcal{R}}_1$ is shown in Fig. 16. The kernel *gemm* that occurs only in the variants from $\hat{\mathcal{R}}_0$ is colored green, and the kernels *transpose* and *qr* that occur only in the variants from $\hat{\mathcal{R}}_1$ are colored red. Hence, we automatically identify the observations **r1**, **r2** and **r3**. We also discover additional patterns, such as the observation that whenever the kernels *trsm* or *trsv* directly precede *gemv*, the corresponding variant is not one of the fast ones. We do not notice changes in the partial ranking or additional patterns when repeating the analyses with Methodology 1 or 2.

In order to highlight the significance of using a ranking methodology that accounts for ties, we illustrate the limitations that are inherent when mining for the causes of

Table 3 Kernel sequences of the 10 algorithmic variants considered in Fig. 15

Variant	Kernel sequence
$\mathbf{alg}_0$	potrf, trsm, trsv, syrk, gemv, potrf, trsv, trsv
$\mathbf{alg}_1$	potrf, trsv, trsm, syrk, potrf, gemv, trsv, trsv
$\mathbf{alg}_2$	potrf, trsm, trsv, syrk, gemv, qr, gemv, trsv
$\mathbf{alg}_3$	potrf, trsv, trsm, gemm, potrf, gemv, trsv, trsv
$\mathbf{alg}_4$	potrf, trsm, trsv, gemm, gemv, potrf, trsv, trsv
$\mathbf{alg}_5$	potrf, trsm, trsv, gemm, potrf, gemv, trsv, trsv
$\mathbf{alg}_6$	transpose, potrf, trsm, trsv, syrk, potrf, trsv, trsm, gemv, trsv
$\mathbf{alg}_7$	transpose, potrf, trsm, syrk, potrf, trsv, trsv, trsm, gemv, trsv
$\mathbf{alg}_8$	transpose, potrf, trsm, syrk, potrf, trsm, trsm, trsv, trsv, gemv
$\mathbf{alg}_9$	transpose, potrf, trsm, syrk, potrf, trsv, trsv, trsm, trsm, gemv

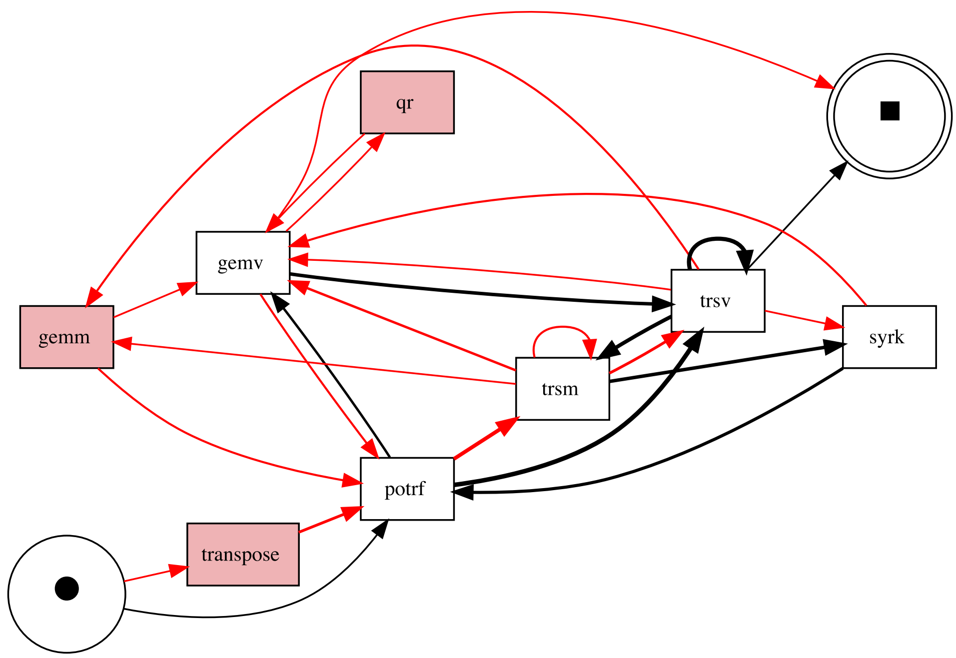


Fig. 17 DFG (Top-1 median ranking): $\mathcal{G}reen = \{\mathbf{alg}_0\}$, $\mathcal{R}ed$ contains the remaining variants

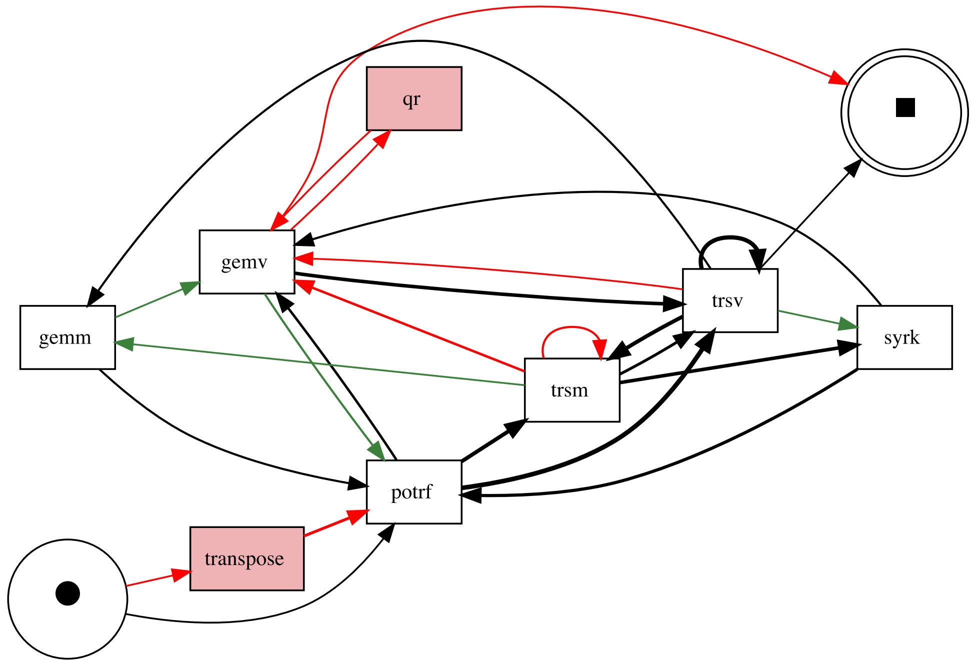
performance difference based on rankings that determined by simply relying only on the median execution time of the variants. To this end, we consider the following:

- **Top-1 median**: The variant having the lowest median execution time is placed in the set $\mathcal{G}reen$. The remaining variants are placed in the set $\mathcal{R}ed$. The resulting DFG is shown in Fig. 17. As this kind of ranking completely ignores ties, the coloring of the DFG is not reasonable; the kernel *gemm* is colored red because the only variant in the set $\mathcal{G}reen$, which is $\mathbf{alg}_0$, does not use this kernel. However, there exist other variants— $\mathbf{alg}_3$, $\mathbf{alg}_4$ and $\mathbf{alg}_5$—that use *gemm* and the spread of their execution times significantly overlaps with that of $\mathbf{alg}_0$. Hence, according to the available data, it is not reasonable to highlight *gemm* as a cause for a variant to be slow.
- **Top-k median**: Top-k ranking is a common approach to distinguish between good and bad variants. Here, the variants with the top k lowest median execution times are placed in the set $\mathcal{G}reen$ and the remaining variants are placed in the set $\mathcal{R}ed$. It is important to note that in this context, the identification of reasonable causes depends on the choice of k. For $k = 5$, the variants are split into the two sets according to ones intuition (i.e., according to Eq. 9). However, for $k = 4$, we get the DFG shown in Fig. 18 in which the association **r1** is not identified.

6.2 Business process example

Just as an algorithm can be viewed as a sequence of kernel calls, a business process can be viewed as a sequence of tasks or activities that need to be performed to achieve a specific business objective. For example, let us consider the purchase-

Fig. 18 DFG (Top-4 median ranking): $\mathcal{Green} = \{\mathbf{alg}_0, \mathbf{alg}_1, \mathbf{alg}_5, \mathbf{alg}_3\}$, $\mathcal{Red}$ contains the remaining variants



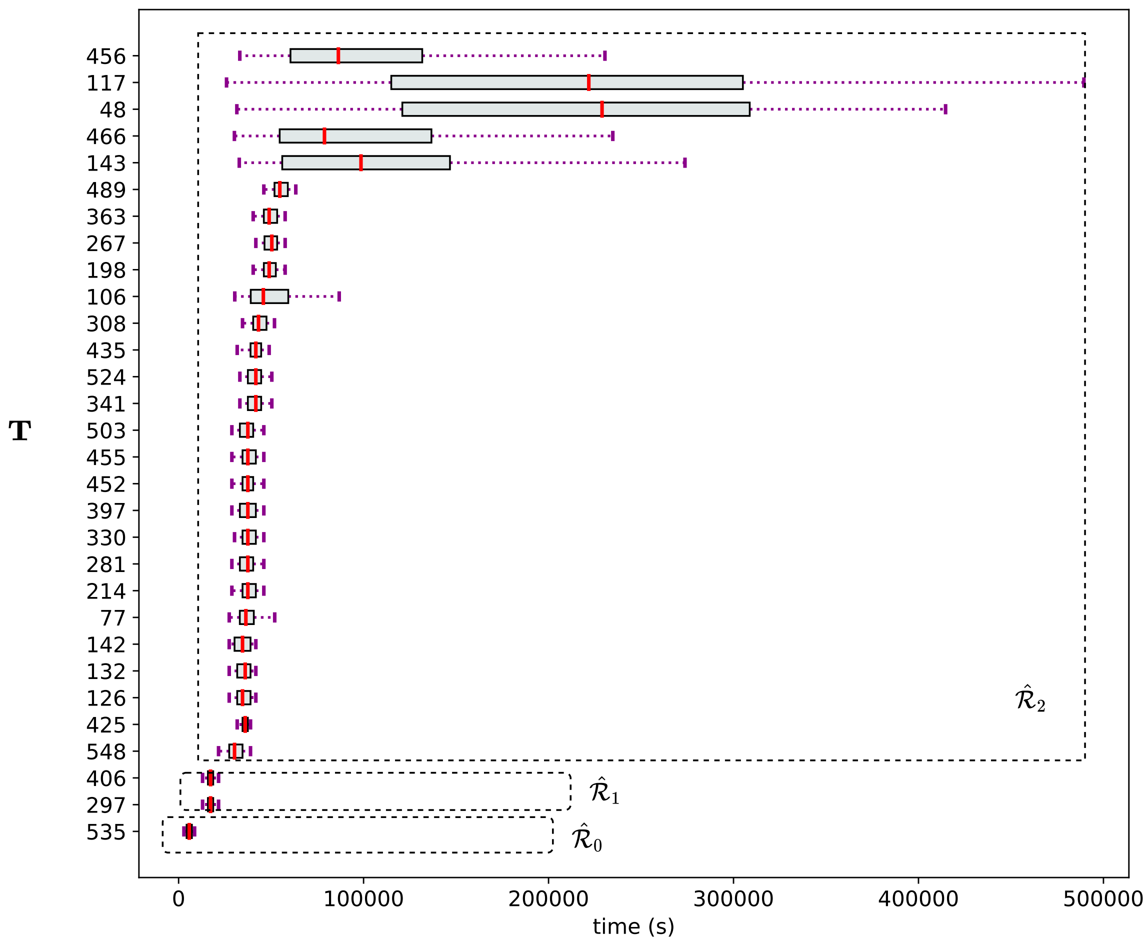


Fig. 19 Sets of throughput times of the process variants ($\mathcal{M}_{bpi}$) and the partial ranking annotated based on Methodology 2 or 3 according to $<_{eg}$

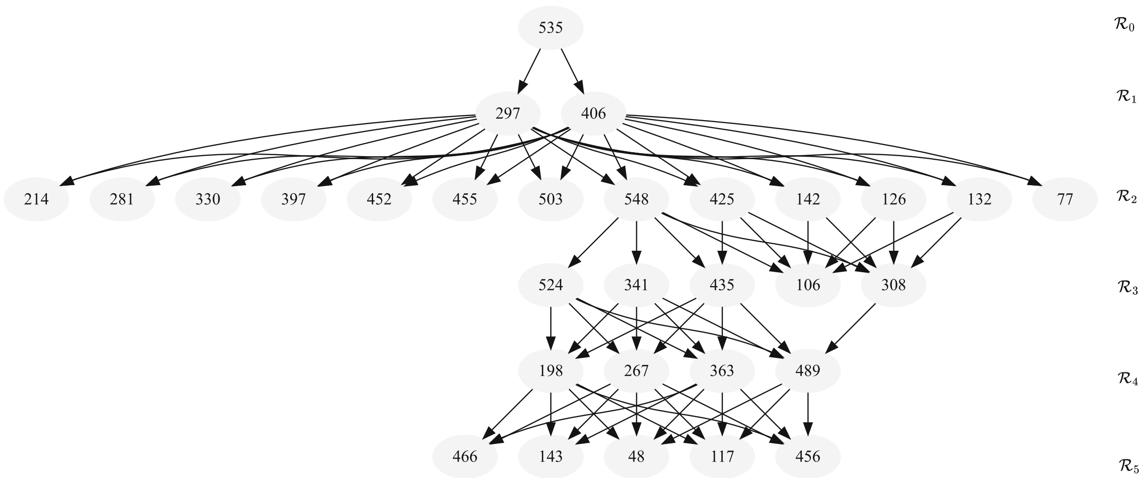


Fig. 20 Methodology 1 on $\mathcal{M}_{bpi}$ according to $<_{eg}$. The graph H is shown

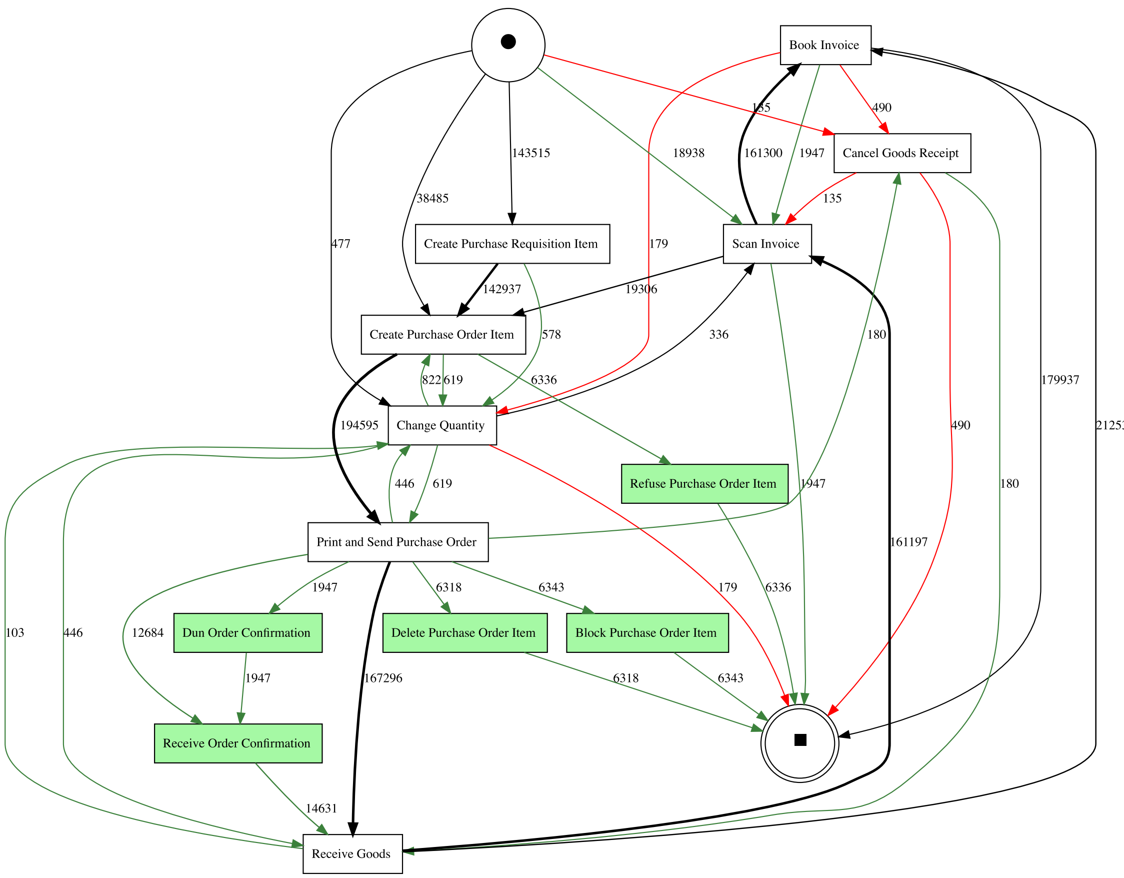


Fig. 21 The DFG highlighting the root causes of performance differences between the process variants. The set $\mathcal{G}reen$ constitutes $\mathcal{R}_0$, $\mathcal{R}_1$ and $\mathcal{R}_2$, and the set $\mathcal{R}ed$ constitutes $\mathcal{R}_5$

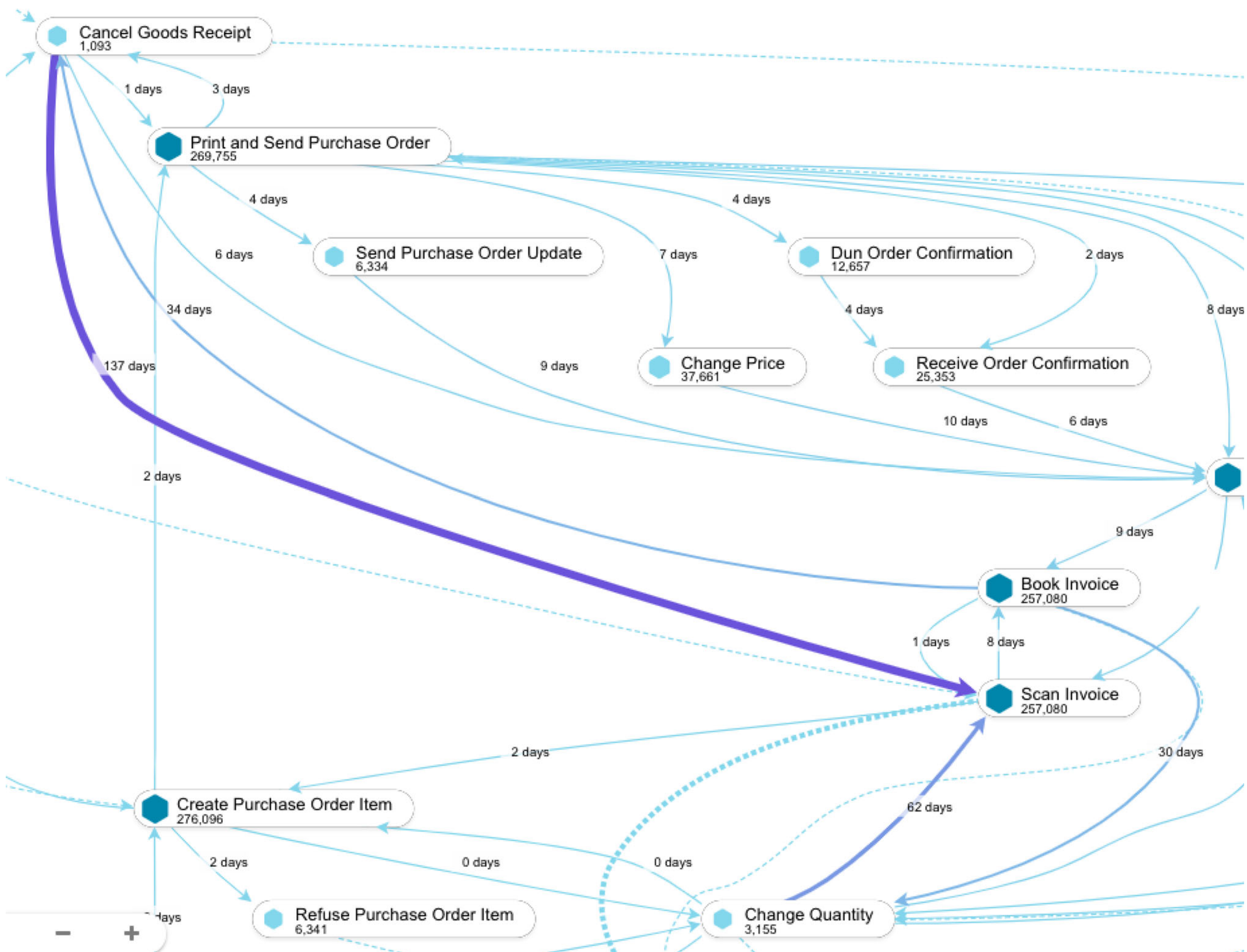


Fig. 22 Screenshot of the DFG rendered by Celonis

to-pay (P2P) business process, which involves the activities encountered while acquiring goods or services from external suppliers. This is a common business process carried out in many companies and organizations, which involves a series of interdependent activities that are performed by various entities within an organization. For every purchase request, the sequence in which the activities are executed is captured by sophisticated enterprise resource management systems like SAP. A particular sequence of activities is referred to as a process variant, and in large organizations, there are typically hundreds of process variants for a business process like P2P (in the following, we show an example). Making organizational changes to improve such a process often entails the use of several enterprise tools which facilitate in identifying the activities that contribute to process variants exhibiting specific performance issues, such as higher throughput time [27].

For the purpose of illustration, we consider a P2P dataset from certain organizations provided by a German data processing company—Celonis, during a hackathon event organized in collaboration with RWTH Aachen University in April 2022. In the considered dataset, the sets of throughput times of the 30 most frequent process variants are shown in Fig. 19 ($\mathcal{M}_{bpi}$). The partial ranking according to Methodology 3 with $<_{eg}$ is annotated in Fig. 19. However, as the spreads of the throughput times are largely overlapping with monotonically increasing differences, Methodology 3 (and even Methodology 2) merges variants with significantly different performances into the same rank $\hat{\mathcal{R}}_2$. Moreover, the throughput times of the variants do not occur in well-separated groups. Therefore, we employ Methodology 1 to calculate an alternate ranking that classifies the variants into six ranks $\mathcal{R}_0, \ldots, \mathcal{R}_5$ as shown in Fig. 20, and perform the root cause analysis. To this end, we construct a DFG similar to the one created in Sect. 6.1, but this time with nodes representing activities in the P2P process. The set $\mathcal{G}reen$ constitutes the process variants from the ranks $\mathcal{R}_0$, $\mathcal{R}_1$, $\mathcal{R}_2$ and the set $\mathcal{R}ed$ constitutes the process variants from the rank $\mathcal{R}_5$. Thus, the green nodes and edges indicate the activities and relations that occur exclusively in the process variants of the set $\mathcal{G}reen$, while the red nodes and edges indicate the activities and relations that occur exclusively in the process variants of the set $\mathcal{R}ed$. The resulting DFG is shown in Fig. 21. On the edges, we also indicate the number of times a particular directly follows relation was observed.

We compare our constructed DFG with the DFG generated by the Celonis application, noting that the Celonis DFG does not apply partial ranking to classify and colors the nodes and edges based on the performance classes. A screenshot of a portion of the Celonis DFG is displayed in Fig. 22. In the Celonis DFG, the numbers displayed on the edges represent the median time between the start of two activities. Notably, there are significant time delays observed between the activities 'Cancel Goods Receipt' to 'Scan Invoice' (**e1**), 'Change Quantity' to 'Scan Invoice' (**e2**) and 'Book Invoice' to 'Change Quantity' (**e3**), indicating potential bottlenecks in the process flow. It is important to note that if a bottleneck occurs consistently across all process variants, it may not provide insights into the root cause of performance differences among the variants. However, when a bottleneck is present in some variants and not in others, it becomes a focal point for highlighting performance differences. While the Celonis DFG effectively captures these bottlenecks, it does not explicitly indicate whether these bottlenecks are the underlying causes of performance disparities among the variants. For example, in our DFG, the edge **e2** with a median time of 62 days is not highlighted in red as it appears in variants from both $\mathcal{G}reen$ and $\mathcal{R}ed$. Therefore, it is not inferred to directly contribute to the root cause of performance differences among the variants. However, the edge **e3**, with a median time of only 30 days, is marked in red in our DFG, suggesting it as an indicator of performance difference.

Thus, we remark that Celonis can benefit from incorporating partial ranking to provide additional perspectives within their DFG. This enhancement has the potential to offer valuable insights to their customers, aiding them in their decision-making processes.

7 Conclusion

We considered the problem of ranking sets of noisy measurement data while accounting for ties. As soon as ties are allowed, more than one reasonable ranking became possible because of the non-transitive nature of the ties. We noticed a lack of clarity on what constitutes a set of reasonable rankings in the previous works. Therefore, for given sets of measurements and a better-than relation that defines how two sets of measurements should be compared, we defined partial ranking to identify a set of reasonable rankings. We formalized and developed three different methodologies for partial ranking. Each methodology computes one of the reasonable rankings. Methodology 1 computes a partial ranking consisting of an arbitrary number of ranks. Among the three methodologies, this methodology computes a ranking with the highest number of ranks. Methodology 2 takes the partial ranking computed by Methodology 1 as input and aims to reduce the number of ranks. Finally, we presented Methodology 3 which computes the partial ranking with minimum possible number of ranks. We demonstrated how the three methodologies can be applied hierarchically to conduct root cause analyses.

In this work, we presented the methodologies such that they are not dependent on any specific pairwise comparison method. While we used quantile-based comparisons as an example, alternative methods such as the Wilcoxon or Mann–Whitney U tests could also be employed. In future works, we plan to investigate the use of partial ranking methodologies with other pairwise comparison functions and explore more applications.

Acknowledgements Financial support from the Deutsche Forschungsgemeinschaft (German Research Foundation) through the grant IRTG 2379 is gratefully acknowledged. We also thank Celonis for granting access to the SAP P2P dataset through their cloud service, which enabled the execution of experiments in Sec. 6.2.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability The implementation of the partial ranking methodologies, experimental data and the results that support the findings of this study are available in Zenodo with the identifier https://doi.org/10.5281/zenodo.12082779

Declarations

Conflicts of Interest The authors have no conflict of interest to declare that are relevant to the content of this article.

Ethics Approval Not applicable.

References

1. Ailon, N.: Aggregation of partial rankings, p-ratings and top-m lists. Algorithmica **57**(2), 284–300 (2010). https://doi.org/10.1007/s00453-008-9211-1
2. Alaimo, L.S., Fiore, M., Galati, A.: Measuring consumers' level of satisfaction for online food shopping during COVID-19 in Italy using POSETs. Socio-Econ. Plann. Sci. **82**, 101064 (2022). https://doi.org/10.1016/j.seps.2021.101064
3. Barthels, H., Psarras, C., Bientinesi, P.: Linnea: automatic generation of efficient linear algebra programs. ACM Trans. Math. Softw. **47**(3), 22:1-22:26 (2021). https://doi.org/10.1145/3446632
4. Bezanson, J., Edelman, A., Karpinski, S., et al.: Julia: a fresh approach to numerical computing. SIAM Rev. **59**(1), 65–98 (2017). https://doi.org/10.1137/141000671
5. Brüggemann, R., Patil, GP.: Ranking and prioritization for multi-indicator systems: introduction to partial order applications. No. v. 5 in Environmental and ecological statistics, Springer, New York (2011a)
6. Brüggemann, R., Patil, GP.: Structures of partial orders. In: Ranking and prioritization for multi-indicator systems. Springer, New York, pp. 57–74 (2011b). https://doi.org/10.1007/978-1-4419-8477-7
7. Charles, J., Jassi, P., Ananth, NS.: et al Evaluation of the Intel® Core™ i7 Turbo Boost feature. In: 2009 IEEE International Symposium on Workload Characterization (IISWC), pp. 188–197 (2009). https://doi.org/10.1109/IISWC.2009.5306782
8. Chen, T., Guo, Q., Temam, O., et al.: Statistical performance comparisons of computers. IEEE Trans. Comput. **64**(5), 1442–1455 (2015). https://doi.org/10.1109/TC.2014.2315614, conference Name: IEEE Transactions on Computers
9. Dao, P.B.: On Wilcoxon rank sum test for condition monitoring and fault detection of wind turbines. Appl. Energy **318**, 119209 (2022). https://doi.org/10.1016/j.apenergy.2022.119209
10. Fagin, R., Kumar, R., Mahdian, et al.: Comparing Partial Rankings. SIAM J. Discrete Math. **20**(3), 628–648 (2006). https://doi.org/10.1137/05063088X, publisher: Society for Industrial and Applied Mathematics
11. Hoefler, T., Belli, R.: Scientific benchmarking of parallel computing systems: twelve ways to tell the masses when reporting performance results. In: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. Association for Computing Machinery, New York, NY, USA, SC '15, pp. 1–12 (2015). https://doi.org/10.1145/2807591.2807644,
12. Ilyas, I.F., Beskales, G., Soliman, M.A.: A survey of top-k query processing techniques in relational database systems. ACM Comput. Surv. **40**(4), 11:1-11:58 (2008). https://doi.org/10.1145/1391729.1391730
13. Li, J., Deshpande, A.: Ranking continuous probabilistic datasets. Proc. VLDB Endow. **3**(1–2), 638–649 (2010). https://doi.org/10.14778/1920841.1920923
14. Liu, Y., Forrest, J., Xie, N.: Ranking grey numbers based on dominance grey degrees. Journal of Systems Engineering and Electronics **25**(4), 618–626 (2014). https://doi.org/10.1109/JSEE.2014.00072, conference Name: Journal of Systems Engineering and Electronics
15. McKnight, P.E., Najab, J.: Mann-Whitney test. In: Weiner, I.B., Craighead, W.E. (eds.) The Corsini Encyclopedia of Psychology, 1st edn., pp. 1–1. Wiley, New York (2010). https://doi.org/10.1002/9780470479216.corpsy0524
16. Myers, W., Patil, G.: Chapter 7 semi-subordination sequences in multi-measure prioritization problems. In: Data Handling in Science and Technology, vol 27. Elsevier, pp. 159–168. https://doi.org/10.1016/S0922-3487(08)10007-7 (2008). https://linkinghub.elsevier.com/retrieve/pii/S0922348708100077
17. Nikitenko, D.A., Wolf, F., Mohr, B., et al.: Influence of noisy environments on behavior of HPC applications. Lobachevskii J. Math. **42**(7), 1560–1570 (2021). https://doi.org/10.1134/S1995080221070192
18. Patil, G., Taillie, C.: Multiple indicators, partially ordered sets, and linear extensions: multi-criterion ranking and prioritization. Environ. Ecol. Stat. **11**(2), 199–228 (2004). https://doi.org/10.1023/B:EEST.0000027209.93218.d9
19. Pavan, M., Todeschini, R.: New indices for analysing partial ranking diagrams. Anal. Chim. Acta **515**(1), 167–181 (2004). https://doi.org/10.1016/j.aca.2003.11.019. (https://www.sciencedirect.com/science/article/pii/S0003267003014879)
20. Pavan, M., Todeschini, R.: Scientific Data Ranking Methods: Theory and Applications. Data Handling in Science and Technology, vol. 27, 1st edn. Elsevier, Oxford (2008)
21. Sankaran, A, Bientinesi, P.: Performance comparison for scientific computations on the edge via relative performance. In: 2021 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), pp. 887–895 (2021). https://doi.org/10.1109/IPDPSW52791.2021.00132
22. Sankaran, A., Bientinesi, P.: A test for FLOPs as a discriminant for linear algebra algorithms. In: 2022 IEEE 34th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD), pp. 221–230 (2022). https://doi.org/10.1109/SBAC-PAD55451.2022.00033, ISSN: 2643-3001
23. Sankaran, A., Zhukov, I., Frings, W., et al.: Inspection of I/O operations from system call traces using directly-follows-graph. In: SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE, pp. 1562–1575 (2024). https://doi.org/10.1109/SCW63240.2024.00196
24. Schober, P., Vetter, T.R.: Nonparametric statistical methods in medical research. Anesth. Analg. **131**(6), 1862–1863 (2020). https://doi.org/10.1213/ANE.0000000000005101
25. Szekli, R.: Stochastic Ordering and Dependence in Applied Probability, Lecture Notes in Statistics, vol. 97. Springer, New York (1995). https://doi.org/10.1007/978-1-4612-2528-7
26. Van Der Aalst, W.M.P., Carmona, J. (eds.): Process Mining Handbook, Lecture Notes in Business Information Processing, vol 448. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-08848-3
27. van der Aalst, W.: Process mining software. In: van der Aalst, W. (ed.) Process Mining: Data Science in Action, pp. 325–352. Springer, Berlin (2016). https://doi.org/10.1007/978-3-662-49851-4_11
28. Zhang, H.Y., Yang, S.Y., Ma, J.M.: Ranking interval sets based on inclusion measures and applications to three-way decisions. Knowl.-Based Syst. **91**, 62–70 (2016). https://doi.org/10.1016/j.knosys.2015.07.025
29. Zhang, Q., Cheng, Y., Zhao, F., et al.: Optimal scale combination selection integrating three-way decision with Hasse diagram. IEEE Trans. Neural Netw. Learn. Syst. **33**(8), 3675–3689 (2022). https://doi.org/10.1109/TNNLS.2021.3054063