Diversity in Kemeny Rank Aggregation:
A Parameterized Approach

Emmanuel Arrighi[1] , **Henning Fernau**[2] , **Daniel Lokshtanov**[3]
Mateus de Oliveira Oliveira[1] , **Petra Wolf** [2]

[1]University of Bergen, Norway
[2]University of Trier, Germany
[3]University of California Santa Barbara, CA, USA
{emmanuel.arrighi, mateus.oliveira}@uib.no, {fernau, wolfp}@informatik.uni-trier.de,
{daniello}@ucsb.edu

Abstract

In its most traditional setting, the main concern of optimization theory is the search for optimal solutions for instances of a given computational problem. A recent trend of research in artificial intelligence, called *solution diversity*, has focused on the development of notions of optimality that may be more appropriate in settings where subjectivity is essential. The idea is instead of aiming at the development of algorithms that output a *single* optimal solution, the goal is to investigate algorithms outputting a small set of *sufficiently good* solutions that are *sufficiently diverse* from one another. This way, the user has the opportunity to choose the solution being most appropriate to the context at hand. It also displays the richness of the solution space.

When combined with techniques from parameterized complexity theory, the paradigm of diversity of solutions offers a powerful algorithmic framework to address problems of practical relevance. In this work, we investigate the impact of this combination in the field of Kemeny Rank Aggregation, a well-studied class of problems lying in the intersection of order theory and social choice theory and also in the field of order theory itself. In particular, we show that the Kemeny Rank Aggregation problem is fixed-parameter tractable with respect to natural parameters providing natural formalizations of the notions of *diversity* and of the notion of a *sufficiently good* solution. Our main results work both when considering the traditional setting of aggregation over linearly ordered votes, and in the more general setting where votes are partially ordered.

1 Introduction

Traditionally, in optimization theory, when given an instance of a computational problem, one is interested in computing *some* optimal solution for the instance in question. For certain problems of practical relevance, this framework may not be satisfactory because it precludes the user from the possibility of choosing among optimal solutions in case more than one exists, or from choosing a slightly less optimal solution that may be a better fit for the intended application, due to subjective factors. A recent upcoming trend of research in artificial intelligence, called *diversity of solutions* [Petit and Trapp, 2019; Baste *et al.*, 2020; Ingmar *et al.*, 2020; Baste *et al.*, 2019; Fomin *et al.*, 2020], has focused on the development of notions of optimality that may be more appropriate in settings where subjectivity is essential. The idea is that instead of aiming at the development of algorithms that output a *single* optimal solution, the goal is to investigate algorithms that output a small set of *sufficiently good* solutions that are *sufficiently diverse* from one another. In this way, the user has the opportunity to choose the solution that is most appropriate to the context at hand. The intuition is that the criteria employed by the user to decide what an appropriate solution is may be subjective, and therefore, impractical or even impossible to be formalized at the level of the problem specification. Examples of such criteria are aesthetic, economic, political, or environmental criteria. Another motivation comes from the problem of finding several good committees such that each committee member is not overloaded with commitments, as described in [Bredereck *et al.*, 2020]; again, some diversity could be helpful.

One source of difficulty when trying to develop efficient algorithms for diverse variants of computational problems is the fact that these problems may be computationally hard. In particular, many interesting computational problems that are suitable for being studied from the perspective of diversity of solutions are already NP-hard in the usual variant in which one asks for a single solution. Additionally, it may be the case that even problems that are polynomial-time solvable in the single-solution version become NP-hard when considering diverse sets of solutions. One way to circumvent this difficulty is to combine the framework of diversity of solutions with the framework of fixed-parameter tractability theory [Downey and Fellows, 1999]. A central notion in this framework is the one of fixed-parameter tractability. An algorithm for a given computational problem is said to be fixed-parameter tractable with respect to parameters $k_1, \ldots, k_r$ if it runs in time $f(k_1, \ldots, k_r) \cdot n^{O(1)}$, where n is the size of the input and f is a computable function that depends only on the parameters. The intuition is that if the range of the parameters

is small on instances of practical relevance, then even if the function f grows relatively fast, the algorithm can be considered to be fast enough *for practical purposes.*

When studying a given computational problem from the point of view of solution diversity, it is crucial to have in hands a notion of distance between solutions for that problem. The *diversity* of a set of solutions S is then defined as the sum of distances between pairs of solutions in S. We denote this measure by d. Intuitively, diversity is a global measure for how representative a set of solutions is among the space of solutions. Three natural parameters can be used to quantify how good a diverse set of solutions is: the number r of solutions in the set, the maximum distance δ between the cost of a solution in the set and the cost of an optimal solution (we call this parameter the *solution imperfection* of the set), and the minimum *required* distance s between any two solutions in the set. This last parameter is also known in the literature as the *scatteredness* of S [Galle, 1989].

In this work, we investigate the impact of the notions of diversity of solutions and of fixed parameter tractability theory in the context of social choice theory. In particular, we focus on the framework of preference list aggregation introduced by Kemeny in the late fifties [Kemeny, 1959]. Intuitively, preference lists are a formalism used in social choice theory to capture information about choice in applications involving the selection of candidates, products, etc. by a group of voters. The task is then to find a ranking of the candidates that maximizes the overall satisfaction among the voters. This problem is commonly referred to in modern terminology as the *Kemeny rank aggregation* (KRA) problem. In its most general setting, the ranking cast by each voter is a partial order on the set of candidates. The distance measure we use to define our diverse version for KRA is the Kendall-Tau distance which is widely used in the context of preference aggregation. Its popularity is underlined by articles describing these issues for the interested public audience; see [Farkas and Timotity, 2019].

Our main result is a multi-parametric algorithm for DI-VERSE KRA over partially ordered votes that runs in time $f(\mathrm{w}, r, \delta, s) \cdot d \cdot n \cdot \log(n^2 \cdot m)$, where n is the number of candidates, m is the number of votes, r, δ, s and d are the parameters discussed above, and w is the *unanimity width* of the votes. That is to say, the pathwidth of the cocomparability graph of the unanimity order of the input votes (Corollary 11). Intuitively, this width measure is a quantification of the amount of disagreement between votes. Note that pathwidth and treewidth coincide for the class of cocomparability graphs [Habib and Möhring, 1994].

On the path towards obtaining our results for Kemeny Rank Aggregation, we also make contributions to problems of independent interest arising in the theory of cocomparability graphs. First, by leveraging on classic results from [Habib and Möhring, 1994], we show that the problem of constructing a ρ-consistent path decomposition of approximately minimum width for the cocomparability graph G_ρ of a given partial-order ρ is fixed-parameter tractable with respect to the pathwidth of G_ρ. While it was known that the pathwidth and the ρ-consistent pathwidth of G_ρ are always the same [Arrighi *et al.*, 2020], and fixed-parameter tractable algorithms for computing path decompositions of approxi-

mately minimum width exists due to structural properties of cocomparability graphs [Bouchitté *et al.*, 2004], the problem of computing such a decomposition satisfying the additional ρ-consistent requirement was open [Arrighi *et al.*, 2020].

Second, we note that the notion of Kendall-Tau distance between partial orders (formally defined in Section 3), which is used to define our notion of diversity, can be applied equally well in the more general context of the COMPLE-TION OF AN ORDERING problem (CO), a problem of fundamental importance in order theory that unifies several problems of relevance for artificial intelligence, such as KRA, ONE-SIDED CROSSING MINIMIZATION (an important subroutine used in the search for good hierarchical representations of graphs), and GROUPING BY SWAPPING (a relevant problem in the field of memory management) [Wong and Reingold, 1991]. For a matter of generality, we first develop a $f(\mathrm{w}, r, \delta, s) \cdot d \cdot n \cdot \log(n^2 \cdot m)$ time algorithm for DIVERSE CO (Theorem 10) and then obtain our main result for DIVERSE KRA as a corollary. In the more general context of CO, the parameter w is the width of the cocomparability graph of the partial order given at the input.

Finally, building on recent advances in the theory of C_k-free graphs [Chudnovsky *et al.*, 2020] we establish an upper bound for the pathwidth of a cocomparability graph in terms of the number of edges of the graph. As a by-product of this result, we obtain the first algorithm running in time $\mathcal{O}^*(2^{\mathcal{O}(\sqrt{k})})$ (Theorem 12) for the positive completion of an ordering problem (PCO), a special case of CO which still generalizes KRA and other important combinatorial problems. Previous to our work, the best algorithm for this problem parameterized by cost had asymptotic time complexity $\mathcal{O}^*(k^{\mathcal{O}(\sqrt{k})}) = \mathcal{O}^*(2^{\sqrt{k}\log k})$. We remove the log-factor in the exponent; according to Thm. 18 in [Arrighi *et al.*, 2020] this is optimal under ETH (Exponential Time Hypothesis).

2 Preliminaries

If n is a positive integer, $[n] = \{1, \ldots, n\}$ denotes the discrete interval of the first n positive integers, and $[n]_0 = [n] \cup \{0\}$. $\mathbb{N}$ denotes the non-negative integers.

Let C be a set. A *partial order* over C is a reflexive, antisymmetric and transitive binary relation $\rho \subseteq C \times C$. We say that ρ is a *linear order* if additionally, for each $(x, y) \in C \times C$, either $(x, y) \in \rho$ or $(y, x) \in \rho$. The *comparability relation* $\mathbf{sc}(\rho)$ of ρ is the symmetric closure of ρ, i.e., $(x, y) \in \mathbf{sc}(\rho)$ iff $(x, y) \in \rho$ or $(y, x) \in \rho$. For instance, $\mathbf{sc}(\rho) = C \times C$ iff ρ is a linear order. If $\rho \subseteq C \times C$ is a partial order, then $<_\rho$ denotes the corresponding strict order, which is irreflexive and transitive. Linear orders over C can be given by bijections $\pi : [|C|] \to C$. Hence, $<_\pi$ (or $\leq_\pi$) is used to denote the corresponding strict (or partial) linear order. Given a binary relation α, we denote by $\mathbf{tc}(\alpha)$ the transitive closure of α.

Definition 1 (Cocomparability graph (F1)). *Given a partial order $\rho \subseteq C \times C$, we let $G_\rho \doteq (C, C \times C \setminus \mathbf{sc}(\rho))$ be the* cocomparability graph *of ρ.*

Given an undirected graph $G = (V, E)$ and a vertex $v \in V$, we let $N(v) \doteq \{u \mid u \in V, (v, u) \in E\}$ be the neighborhood of v. A *path decomposition* of a graph $G = (V, E)$ is a se-

quence $\mathcal{P} = (B_1, B_2, \ldots, B_l)$ of subsets of V, such that the following conditions are satisfied.

- $\bigcup_{1 \leq i \leq l} B_i = V$.

- For each edge $(u, v) \in E$, there is an $i \in [l]$ such that $u, v \in B_i$.

- For each $i, j, k \in [l]$ with $i < j < k$, $B_i \bigcap B_k \subseteq B_j$.

For each position $p \in \{2, \ldots, l\}$, for each vertex $v \in B_p \setminus B_{p-1}$, we say that B_p *introduces* v (v is introduced by B_p) and for each vertex $v \in B_{p-1} \setminus B_p$, we say that B_p *forgets* v (v is forgotten by B_p). For a position $p \in \{1, \ldots, l\}$, we write $\mathrm{intro}(p)$ (resp. $\mathrm{forg}(p)$) for the set of all vertices introduced (resp. forgotten) by B_p, and we let $L_p = \bigcup_{1 \leq i \leq p} \mathrm{forg}(p)$ be the set of vertices that have been forgotten up to position p. The *width* of $\mathcal{P}$ is defined as $\mathrm{w}(\mathcal{P}) = \max_{i \in [l]} |B_i| - 1$. The *pathwidth*, $pw(G)$, of G is the minimum width of a path decomposition of G.

The pathwidth of the cocomparability graph of a partial order may be regarded as a measure of how close the order is from being a linear order. The cocomparability graph of a linear order τ on n elements is the graph with n vertices and no edges. This graph has pathwidth 0. On the other hand, if τ is a partial order where all n elements are unrelated, then the cocomparability graph of τ is the n-clique, which has pathwidth $n - 1$ (the highest possible pathwidth in an n-vertex graph).

3 The Kemeny Rank Aggregation Problem

Let C be a finite set, which in this paper will denote a set of candidates, or alternatives. A *partial vote* over C is a partial order over C. The KT-distance between two partial votes π_1 and π_2, denoted by $\mathrm{KT\text{-}dist}(\pi_1, \pi_2)$, is the number of pairs of candidates that are ordered differently in the two partial votes: $\mathrm{KT\text{-}dist}(\pi_1, \pi_2) = |\{(c, c') \in C \times C \mid c <_{\pi_1} c' \wedge c' <_{\pi_2} c\}|$.

Observe that when the votes are totally ordered, the Kendall-Tau distance can be seen as the 'bubble sort' distance, i.e., the number of pairwise adjacent transpositions needed to transform one linear order into the other. Given a linear order π over a set of candidates C and a set Π of votes over C, the *Kemeny score* of π with respect to Π is defined as the sum of the Kendall-Tau distances between π and each vote in Π. In this work, we consider the following problem.

Problem name: KEMENY RANK AGGREGATION (KRA)
Given: A list of partial votes Π over a set of candidates C, a non-negative integer k.
Output: Is there a linear order π on C such that the sum of the KT-distances of π from all the partial votes is $\leq k$?

Hence, given partial votes $\pi_1, \ldots, \pi_m$ of C and a non-negative integer k, the question is if there exists a linear order $\pi \subseteq C \times C$ such that $\sum_{i=1}^{m} \mathrm{KT\text{-}dist}(\pi, \pi_i) \leq k$.

Definition 2. *Given a set Π of partial votes, the* unanimity order *of Π is simply the partial order ρ obtained as the intersection of all partial orders in Π. In other words, a candidate c_1 has higher precedence than a candidate c_2 in ρ if and only if c_1 precedes c_2 in each vote in Π.*

As a consequence, the more disagreements there are among the voters with respect to the relative orders of pairs of candidates, the denser the cocomparability graph of ρ will be and therefore the greater its pathwidth will be. Therefore, the pathwidth of the cocomparability graph of the unanimity order of Π may be seen as a quantification of the amount of disagreement among the votes in Π.

The notion of diversity of solutions for computationally hard problems has been considered under a variety of frameworks. In this work, we define a notion of diversity for the KEMENY RANK AGGREGATION problem which is analogous to the notion of diversity of vertex sets used in [Baste *et al.*, 2020]. More precisely, if R is a set of partial orders, then we define the Kendall-Tau diversity of R as the sum of Kendall-Tau distances between votes in the set R. $\mathrm{KT\text{-}Div}(R) = \sum_{\pi_1, \pi_2 \in R} \mathrm{KT\text{-}dist}(\pi_1, \pi_2)$.

We note that the restricted version of the KRA problem where all votes are linear orders, the requirements that a set of solutions is at the same time diverse and only contains rankings with small Kemeny score are clashing. The problem is that the very existence of two distinct rankings with small Kemeny score is an impossible task. If two candidates c_1 and c_2 occur with the order (c_1, c_2) in one of the solutions and in the order (c_2, c_1) in the other solution, then at least one of these solutions will have a Kemeny score of at least half the number of votes. However this opposition between diversity and small Kemeny score is not present in the setting where votes are allowed to be partial. The generalization to partial votes is one possible way to circumvent this conflict of desiderata. Another way we will be looking at is not to consider the cost of the solutions directly but the difference between the cost of solutions and the cost of an optimal solution. In this case, we can have diversity and a small difference between the cost and the cost of an optimal solution.

Problem name: DIVERSE KEMENY RANK AGGREGATION (DIVERSE-KRA)
Given: A list of partial votes Π over a set of candidates C, and $k, r, d \in \mathbb{N}$.
Output: Is there a set $R = \{\pi_1, \ldots, \pi_r\}$ of linear orders on C such that the Kemeny score for each order π_i is at most k and $\mathrm{KT\text{-}Div}(R) \geq d$?

The problem KEMENY RANK AGGREGATION is known to be NP-complete [Bartholdi *et al.*, 1989], even if only four votes are given at the input [Dwork *et al.*, 2001]. For this reason, KRA has been studied from the perspective of parameterized complexity theory under a variety of parameterizations. Below, we consider two prominent parameterizations for this problem.

The first parameter we consider is the cost of a solution. Simjour [2009] obtained an algorithm for the problem that runs in time $\mathcal{O}^*(1.403^k)$. There are also subexponential algorithms for KEMENY RANK AGGREGATION under this parameterization: Karpinski and Schudy [Karpinski and Schudy, 2010] obtained an algorithm for KEMENY RANK AGGREGATION that runs in $\mathcal{O}^*(2^{O(\sqrt{k})})$ time, while the algorithm of Fernau *et al.* [Fernau *et al.*, 2011; Fernau *et al.*, 2014], based on a different methodology, runs in $\mathcal{O}^*(k^{O(\sqrt{k})})$ time. Recently, in [Arrighi *et al.*, 2020], it was

shown that KRA on instances with only $m = 4$ votes on some candidate set C and some integer k bounding the sum of the Kendall-Tau distances to a solution cannot be solved neither in time $\mathcal{O}^*\left(2^{o(|C|)}\right)$ nor in time $\mathcal{O}^*\left(2^{o(\sqrt{k})}\right)$, unless ETH fails. The mentioned NP-hardness of KRA immediately translates to NP-hardness results of KRA and of DIVERSE-KRA, in the latter case by setting $r = 1$ and $d = 0$.

The second parameter we consider is the *unanimity width* of the set of votes, which is based on the notion of unanimity order of a set of votes [Charon and Hudry, 2007]. The *unanimity width* of Π is defined as the pathwidth of the cocomparability graph of ρ.

4 Completion of an Ordering

In this section, we will introduce the COMPLETION OF AN ORDERING problem, a generalization of the POSITIVE COMPLETION OF AN ORDERING (PCO) problem originally considered in [Dujmovic *et al.*, 2003, Sec. 8] and [Fernau, 2005, Sec. 6.4].

Problem name: COMPLETION OF AN ORDERING (CO)
Given: A partial order $\rho \subseteq C \times C$ over a set C, a cost function $\mathfrak{c} : C \times C \to \mathbb{N}$, and some $k \in \mathbb{N}$.
Output: Is there a linear order $\tau \supseteq \rho$ with $\mathfrak{c}(\tau \setminus \rho) = \sum_{(x,y)\in\tau\setminus\rho} \mathfrak{c}(x,y) \le k$?

Intuitively, given a partial order ρ and a cost function $\mathfrak{c}$, the goal is to find a linear extension of ρ incurring a cost of at most k. The only difference between CO and the original PCO problem introduced in [Dujmovic *et al.*, 2003; Fernau, 2005] is that, in the latter, for every pair $(x,y) \in C \times C$ such that x and y are incomparable in ρ, the cost of (x,y) is strictly positive ($\mathfrak{c}(x,y) > 0$) whereas in CO, the cost can be zero ($\mathfrak{c}(x,y) = 0$).

We note that the notion of Kendall-Tau diversity introduced in Section 3 can also be used as a notion of diversity for CO, i.e., given a set R of (not necessarily optimal) solutions for a given instance $(\rho, \mathfrak{c})$ of CO, we let KT-Div(R) be the diversity of this set.

Problem name: DIVERSE COMPLETION OF AN ORDERING (DIVERSE-CO)
Given: A partial order $\rho \subseteq C \times C$ over a set C, a cost function $\mathfrak{c} : C \times C \to \mathbb{N}$, and non-negative integers $k, r, d \in \mathbb{N}$.
Output: Is there a set $R = \{\tau_1, \ldots, \tau_r\}$ of linear extensions of ρ such that $\mathfrak{c}(\tau_i \setminus \rho) \le k$ for each $i \in [r]$, and KT-Div$(R) \ge d$?

Next, we give a rather straightforward reduction from KRA to CO. Given an instance (Π, C) of KRA with partial votes $\Pi = (\pi_1, \ldots, \pi_m)$ and candidates $C = \{c_1, \ldots, c_n\}$, we construct an instance $(\rho, \mathfrak{c})$ of CO by letting ρ be the unanimity order of Π, and by defining the cost function $\mathfrak{c} : C \times C \to \mathbb{N}$ as follows. For every pair of candidates (c, c'), we define its *cost*, $\mathfrak{c}(c, c')$, as the number of votes that order c' before c. More formally, $\mathfrak{c}(c, c') = |\{i \in [m] \mid c' <_{\pi_i} c\}|$. With this reduction, it is straightforward to check that a given linear order π of the candidates has Kemeny score $\sum_{i=1}^{m}$ KT-dist$(\pi, \pi_i) = \sum_{j=1}^{n} \sum_{k=1}^{n} \mathfrak{c}(c_k, c_j)[c_k <_\pi c_j]$.

Here, for a logical proposition p, we use the bracket notation $[p]$ to denote the integer 1 if p is true the integer 0 if p is false. In other words, $\sum_{i=1}^{m}$ KT-dist(π, π_i) is the cost of π as a linear extension of the ordering ρ.

It is important to note that if all votes in Π are linear orders then $(\rho, \mathfrak{c})$ is actually an instance of PCO. In other words, if two candidates c and c' are incomparable in the unanimity order, then the cost assigned by $\mathfrak{c}$ to both pairs (c, c') and (c', c) are strictly positive. This property will be used crucially in the development of our sub-exponential time algorithm for KRA parameterized by cost.

We also note that since our reduction is solution preserving, it is also immediate that it is diversity preserving. In other words, R is a set of solutions of diversity d for an instance of KRA if and only if it is also a set of solutions of diversity d for the corresponding instance of CO.

5 Diverse CO Parameterized by Pathwidth

In this section, we devise a fixed parameter tractable algorithm for DIVERSE CO parameterized by solution imperfection, number of solutions, scatteredness, and pathwidth of the cocomparability graph of the input instance. Given our reduction that preserves solution and parameters from KRA to CO introduced in Section 4, this algorithm immediately implies that DIVERSE KRA is fixed parameter tractable when parameterized by solution imperfection, number of solutions, scatteredness, and unanimity width.

We start by defining a suitable notion of consistency between a path decomposition and a given partial order. Let $G = (C, E)$ be a graph and $\rho \subseteq C \times C$ be a partial order on the vertices of G. We say that a path decomposition $D = (B_1, \ldots, B_l)$ is *ρ-consistent* if there is no pair of vertices $(x, y) \in \rho$ such that

$$\max(\{i \in [l] \mid y \in B_i\}) < \min(\{i \in [l] \mid x \in B_i\}).$$

In other words, if x is smaller than y in ρ, then y cannot be forgotten in D before x is introduced in D. The *ρ-consistent pathwidth* of G, denoted by $cpw(G, \rho)$, is the minimum width of a ρ-consistent path decomposition of G.

It has been shown recently that for any partial order $\rho \subseteq C \times C$, the pathwidth of the cocomparability graph G_ρ is equal to the consistent pathwidth of G_ρ [Arrighi *et al.*, 2020]. The proof of this result was based on the fact that the consistent pathwidth of a cocomparability graph of a partial order is equal to the interval width of the order [Habib and Möhring, 1994]. Nevertheless, the problem of constructing, or even approximating, a minimum-width consistent path decomposition in FPT time was left open in [Arrighi *et al.*, 2020].

By taking a closer look at the theory of cocomparability graphs, we solve this open problem in a constructive way. More precisely, in Lemma 3 we show that for any partial order ρ, one can construct a ρ-consistent path decomposition of the cocomparability graph G_ρ in fixed-parameter tractable time parameterized by the pathwidth of the graph G_ρ.

Lemma 3. *Let $\rho \subseteq C \times C$ be a partial order and G_ρ be the cocomparability graph of ρ. Then one can construct a nice ρ-consistent path decomposition $\mathcal{P}$ of G_ρ of width $\mathcal{O}(pw(G_\rho))$ in time $2^{\mathcal{O}(pw(G_\rho))} \cdot |C|$.*

Let ρ be a partial order and $\mathcal{P} = (B_1, B_2, \ldots, B_l)$ be a ρ-consistent path decomposition of G_ρ of width w. For each

$p \in [l]$, let $\mathbb{P}_p$ be the set of pairs of the form (S, τ) where S is a subset of B_p that contains vertices introduced by B_p (intro$(p) \subseteq S \subseteq B_p$), $\tau \supseteq \rho|_S$ is a linear extension of the restriction $\rho|_S \doteq S \times S \cap \rho$ of ρ to S.

Definition 4. *Let $p \in [l]$, $\delta \in \mathbb{N}$, and $f : \mathbb{P}_p \to \mathbb{N}$. Then, we let $\mathcal{T}_p(f, \delta)$ be the set of all triples of the form (S, τ, γ), where $(S, \tau) \in \mathbb{P}_p$ and $f(S, \tau) \leq \gamma \leq f(S, \tau) + \delta$.*

Intuitively, the function f will be used by our dynamic programming algorithm to record the optimal values of partial solutions at each bag B_p when processing the path decomposition from left to right (see Theorem 8 and Theorem 10) and δ will be the allowed solution imperfection. In the case of a unique solution, this value will be 0 but this parameter will be useful in the diverse case as we allow sub-optimal linear extensions. A partial solution up to the p-th bag is a linear order σ of $\bigcup_{j \leq p} B_j$. Vertices that will be introduced in future bags can be inserted in σ, to extend it, only after vertices already forgotten in the p-th bag. If u will be introduced in a future bag and v is in some B_j but not in B_p, then by consistency of the path decomposition with respect to ρ, we have $v <_\rho u$. Therefore, in B_p, we only need to remember the "last" part of σ, which are the vertices that are in B_p and after all forgotten vertices in σ.

Remark 5. *For each $p \in [l]$, $f : \mathbb{P}_p \to \mathbb{N}$ and $\delta \in \mathbb{N}$, the size of $\mathcal{T}_p(f, \delta)$ is bounded by $e \cdot (\delta + 1) \cdot (w + 1)!$.*

For each $p \in [l-1]$, $f : \mathbb{P}_p \to \mathbb{N}$ and $\delta \in \mathbb{N}$, we say that a triple $(S, \tau, \gamma) \in \mathcal{T}_p(f, \delta)$ is *compatible* with a triple $(S', \tau', \gamma') \in \mathcal{T}_{p+1}(f, \delta)$ if the following conditions hold.

C1 Let $v = \max_\tau(S \cap \text{forg}(p+1))$ be the maximum vertex of S forgotten by B_{p+1} according to the linear order τ. Then, we have the following equality $S' = \text{intro}(p+1) \cup \{u \in S \mid v <_\tau u\}$. This means that one can build S' from S by removing all vertices that are either forgotten by B_{p+1} or smaller than some vertex forgotten by B_{p+1}, and subsequently, by adding all vertices that have been introduced by B_{p+1}.

C2 $\tau|_{S \cap S'} = \tau'|_{S \cap S'}$, i.e., τ and τ' agree on $S \cap S'$.

C3 $\gamma' = \gamma + \sum_{v \in \text{intro}(p+1)}(\sum_{u \in S', u <_{\tau'} v} \mathfrak{c}(u, v) + \sum_{u \in S \cap S', v <_{\tau'} u} \mathfrak{c}(v, u) + \sum_{u \in B_{p+1} \setminus S'} \mathfrak{c}(u, v))$. To compute γ', we add to γ the cost of adding the introduced vertices in the order. The first two terms compute the cost of each new vertex in τ' and the last one computes the cost of placing the new vertices after all vertices in $B_{p+1} \setminus S'$.

A *compatible sequence* for $\mathcal{P}$ is a sequence of triples $\gamma = (S_1, \tau_1, \gamma_1) \ldots (S_l, \tau_l, \gamma_l)$ such that for each $p \in [l]$, (S_p, τ_p, γ_p) is compatible with $(S_{p-1}, \tau_{p-1}, \gamma_{p-1})$.

Our interest in compatible sequences stems from the two following lemmas.

Lemma 6. *Let $\rho \subseteq C \times C$ be a partial order over C, $\mathfrak{c} : C \times C \to \mathbb{N}$ be a cost function, and $\mathcal{P}$ be a ρ-consistent path decomposition of the graph G_ρ. Let $\gamma = \mathfrak{t}_1 \ldots \mathfrak{t}_l = (S_1, \tau_1, \gamma_1) \ldots (S_l, \tau_l, \gamma_l)$ be a compatible sequence for $\mathcal{P}$. Then, the linear order $\pi = \mathbf{tc}(\rho \cup \tau_1 \cup \cdots \cup \tau_l)$ is a linear extension of ρ of cost γ_l.*

Lemma 7. *Let $\rho \subseteq C \times C$ be a partial order over C, $\mathfrak{c} : C \times C \to \mathbb{N}$ be a cost function, and $\mathcal{P}$ be a ρ-consistent path decomposition of the graph G_ρ. Let π be a linear extension of ρ, and $\gamma = (S_1, \tau_1, \gamma_1) \ldots (S_l, \tau_l, \gamma_l)$ be a sequence such that for each position $p \in [l]$, $S_p = \{v \in B_p \mid v >_\pi \max_\pi(L_p)\}$, $\tau_p = \pi|_{S_p}$, and $\gamma_p = \mathfrak{c}(\pi|_{L_p \cup B_p})$. Then, γ is a compatible sequence for $\mathcal{P}$.*

Lemma 6 and Lemma 7 immediately yield an FPT dynamic programming algorithm for computing a linear extension of ρ. To define the algorithm more precisely, we first need to define the set of functions f_p that we will use to define a set of triples with $\mathcal{T}_p$. For each $p \in [l]$, we define $f_p : \mathbb{P}_p \to \mathbb{N}$ as follows. For each $(S, \tau) \in \mathbb{P}_p$, we let γ be the minimum cost of a partial solution π up to bag p such that $S = \{v \in B_p \mid v >_\pi \max_\pi(L_p)\}$ and $\tau = \pi|_{S_p}$, then we let $f_p(S, \tau) = \gamma$. Intuitively, f_p associates to each linear order τ the cost of an optimal partial solution "ending" by τ. Now, we will describe the algorithm, we process the path decomposition from left to right in l time steps, where at each time step p, we construct the value of f_p that we need and a subset $\mathcal{Q}_p \subseteq \mathcal{T}_p(f_p, 0)$ of *promising triples*, which are, intuitively, triples that have a potential to lead to an optimal solution. At time step 1, set $\mathcal{Q}_1 = \{(B_1, \tau, \mathfrak{c}(\tau))\}_{\tau \text{ is a linear extension of } \rho|_{B_1}}$. At each time step $p \geq 2$, $\mathcal{Q}_p$ is the set of all triples in $\mathcal{T}_p(f_p, 0)$ that are compatible with some triple in $\mathcal{Q}_{p-1}$. At the end of the process, assuming that $\mathcal{Q}_p$ is non-empty for each $p \in [l]$, we can reconstruct a compatible sequence by backtracking. First, by selecting an arbitrary triple $\mathfrak{t}_l$ in $\mathcal{Q}_l$, then by selecting an arbitrary triple $\mathfrak{t}_{l-1}$ in $\mathcal{Q}_{l-1}$ compatible with $\mathfrak{t}_l$, and so on. Once we have constructed a compatible sequence $\mathfrak{t}_1 \ldots \mathfrak{t}_l$, we can extract a linear extension π of cost γ_l by setting $\pi = \mathbf{tc}(\rho \cup \tau_1 \cup \ldots \tau_l)$. This description gives rise to the following theorem.

Theorem 8. *Let $\rho \subseteq C \times C$, let $n = |C|$, let w be the pathwidth of the cocomparability graph of ρ, and $\mathfrak{c} : C \times C \to [m]_0$ be a cost function. Then, one can compute an optimal solution in time $\mathcal{O}\left(w!^{\mathcal{O}(1)} \cdot n \cdot \log(n \cdot m)\right)$.*

Now, leveraging on Lemma 6 and Lemma 7, we will devise a fixed-parameter tractable algorithm for DIVERSE-CO parameterized by solution imperfection, number of solutions, scatteredness, and pathwidth of the cocomparability graph of the input partial order. Let ρ be a partial order and $\mathcal{P} = (B_1, B_2, \ldots, B_l)$ be a ρ-consistent path decomposition of G_ρ of width w.

Definition 9. *Let $p \in [l]$, and $f : \mathbb{P}_p \to \mathbb{N}$. Then, we let $\mathcal{I}_p(f, \delta)$ be the set of all tuples of the form*

$$((S^1, \tau^1, \gamma^1), \ldots, (S^r, \tau^r, \gamma^r), \partial, (\xi_{\{i,j\}})_{1 \leq i < j \leq r})$$

where $\partial \in [d+1]_0$, for each $1 \leq i < j \leq r$, $\xi_{\{i,j\}} \in [s]_0$, and for each $i \in [r]$, (S^i, τ^i, γ^i) is a triple in $\mathcal{T}_p(f, \delta)$.

Intuitively, $((S^i, \tau^i, \gamma^i))_{i \in [r]}$ are r partial linear extensions, ∂ will be the diversity of the r partial linear extensions and ξ will be the distance between all pairs of the r partial linear extensions. A *diversity-compatible sequence* is a sequence of the form

$$\{((S^1_p, \tau^1_p, \gamma^1_p), \ldots, (S^r_p, \tau^r_p, \gamma^r_p), \partial_p, (\xi^p_{\{i,j\}})_{1 \leq i < j \leq r})\}_{p \in [l]},$$

where for each $i \in [r]$, $(S_1^i, \tau_1^i, \gamma_1^i) \ldots (S_l^i, \tau_l^i, \gamma_l^i)$ is a compatible sequence and for each $p \in [l-1]$, ∂_p keeps track of the amount of diversity observed up to position p, $\xi_{\{i,j\}}^p$ keeps track of the distance between partial solutions i and j.

A diversity-compatible sequence is a representation of r solutions for an instance $(\rho, \mathfrak{c}, r, \delta, d, s)$ of DIVERSE-CO. Intuitively, computing such a sequence corresponds to computing r compatible sequences in parallel, by processing the given path decomposition from left to right, while using an additional register to keep track of the overall diversity at each time step and all the pairwise distances.

Theorem 10. *Let $\rho \subseteq C \times C$, let $n = |C|$ and w be the pathwidth of the cocomparability graph of ρ, and $\mathfrak{c}: C \times C \to [m]_0$ be a cost function. Then, one can determine whether ρ admits r linear extensions $\pi_1, \ldots, \pi_r$ at distance at most δ from the optimum, diversity at least d, and scatteredness at least s in time $\mathcal{O}\left((\mathrm{w}! \cdot \delta)^{\mathcal{O}(r)} \cdot s^{r^2} \cdot d \cdot n \cdot \log(n^2 \cdot m)\right)$.*

By combining Theorem 8 with our reduction from KRA to CO, we have an FPT algorithm for KRA, parameterized by solution imperfection, number of solutions, scatteredness, and unanimity width (Corollary 11).

Corollary 11. *Let Π be a list of m partial votes over a set of n candidates C. Let w be the unanimity width of Π. Given Π and non-negative integers δ, r, s and d, one can determine in time $\mathcal{O}\left((\mathrm{w}! \cdot \delta)^{\mathcal{O}(r)} \cdot s^{r^2} \cdot d \cdot n \cdot \log(n^2 \cdot m)\right)$ whether there is a set $R = \{\pi_1, \ldots, \pi_r\}$ of r linear orders on C such that the Kemeny score for each order π_i is at distance at most δ of the optimum, and we find that $\mathrm{KT\text{-}Div}(R) \geq d$ and that scatteredness is at least s.*

6 Sub-Exponential Time Algorithm for PCO

For special cases of PCO, such as those arising from KRA or the graph-drawing problem OSCM, single-exponential sub-exponential time algorithms have been known, i.e., algorithms with running times of the form $\mathcal{O}^*(2^{\mathcal{O}(\sqrt{k})})$. In contrast, for the more general problem of PCO, only algorithms with running time $\mathcal{O}^*(k^{\sqrt{k}})$ were known before, where k is the cost parameter [Fernau *et al.*, 2014]. Here, we prove that PCO also admits algorithms of the form $\mathcal{O}^*(2^{\mathcal{O}(\sqrt{k})})$, by making use of several structural insights for cocomparability graphs. More precisely, we prove the following theorem.

Theorem 12. *Given a partial order $\rho \subseteq C \times C$ and a cost function $\mathfrak{c}: C \times C \to \mathbb{N}$, one can solve a PCO instance $(\rho, \mathfrak{c}, k)$ in time $|C| \cdot 2^{\mathcal{O}(\sqrt{k})} + \mathcal{O}(|C|^2 \cdot \log(k))$.*

Our main technical result of this section is the following.

Lemma 13. *Let G be a $C_{\geq 5}$-free graph, let m be the number of edges of G. Then, we have $m = \Omega(tw(G)^2)$.*

By combining Lemma 13 with the fact that treewidth equals pathwidth for cocomparability graphs and with the fact that cocomparability graphs are $C_{\geq 5}$-free, we get the following lemma.

Lemma 14. *Let G be a cocomparability graph and let m be the number of edges of G. Then, $m = \Omega(pw(G)^2)$.*

Now, in a PCO instance, each edge contributes at least 1 to the cost of any solution. Therefore, if a solution has cost at most k, then the cocomparability graph of the input partial order can have at most k edges. Therefore, this observation, together with Lemma 14 yields the following lemma.

Lemma 15. *Let $(\rho, \mathfrak{c}, k)$ be an YES-instance of PCO. Then, $pw(G_\rho) = \mathcal{O}(\sqrt{k})$.*

To get the running time of Theorem 12, we need to either compute a ρ-consistent path decomposition of width at most $\mathcal{O}(\sqrt{k})$, or to trigger an early rejection. For this, we will use the following lemma which is based on Lemma 3.

Lemma 16. *There is a polynomial-time algorithm that takes an instance of $(\rho, \mathfrak{c}, k)$ of PCO as input, and either constructs a ρ-consistent path decomposition of the graph G_ρ of width $\mathcal{O}(\sqrt{k})$, or answers that this instance is a NO-instance.*

In Section 5, in order to define our algorithm for DIVERSE-CO, we first devised a simpler algorithm for the single-solution version of CO, that could be used as a building block for the diverse algorithm. It turns out that if our only goal is to solve the single-solution version of CO, then the basic algorithm developed in Section 5 can be optimized, to become a single-exponential time algorithm parameterized by the pathwidth of the cocomparability graph of the input order. More precisely, we have the following lemma.

Lemma 17 ([Arrighi *et al.*, 2020] (Theorem 1))**.** *Given an instance $(\rho, \mathfrak{c}, k)$ of CO and a ρ-consistent path decomposition $\mathcal{P}$ of the graph G_ρ, one can solve this instance in time $|C| \cdot 2^{\mathcal{O}(pw(G_\rho))} \cdot \log(k) + \mathcal{O}(|C|^2 \cdot \log(k))$.*

Now we are ready to prove the statement of Theorem 12. Given an instance $(\rho, \mathfrak{c}, k)$ of PCO, we apply the algorithm stated in Lemma 16. This algorithm either determines that the instance is a NO-instance, or constructs a ρ-consistent path decomposition $\mathcal{P}$ of G_ρ of width $\mathcal{O}(\sqrt{k})$. In the first case, we are done and simply answer NO. Otherwise, we give both the instance $(\rho, \mathfrak{c}, k)$ and the decomposition $\mathcal{P}$ to the algorithm stated in Lemma 17 to determine in time $|C| \cdot 2^{\mathcal{O}(\sqrt{k})} + \mathcal{O}(|C|^2 \cdot \log(k))$ whether $(\rho, \mathfrak{c}, k)$ is a YES- or a NO-instance of PCO. In case this is a YES-instance, the algorithm also constructs a linear extension of ρ of cost at most k. This concludes the proof of Theorem 12.

Acknowledgements

References

[Arrighi *et al.*, 2020] E. Arrighi, H. Fernau, M. de Oliveira Oliveira, and P. Wolf. Width notions for ordering-related problems. In *FSTTCS*, volume 182 of *LIPIcs*, pages 9:1–9:18. Schloss Dagstuhl, 2020.

[Bartholdi *et al.*, 1989] J. Bartholdi, III, C. A. Tovey, and M. A. Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6:157–165, 1989.

[Baste *et al.*, 2019] J. Baste, L. Jaffke, T. Masařík, G. Philip, and G. Rote. FPT algorithms for diverse collections of hitting sets. *Algorithms*, 12:254, 2019.

[Baste *et al.*, 2020] J. Baste, M. Fellows, L. Jaffke, T. Masařík, M. de Oliveira Oliveira, G. Philip, and F. Rosamond. Diversity of solutions: An exploration through the lens of fixed-parameter tractability theory. In *IJCAI*, pages 1119–1125, 2020.

[Bouchitté *et al.*, 2004] V. Bouchitté, D. Kratsch, H. Müller, and I. Todinca. On treewidth approximations. *Discret. Appl. Math.*, 136(2-3):183–196, 2004.

[Bredereck *et al.*, 2020] R. Bredereck, A. Kaczmarczyk, and R. Niedermeier. Electing successive committees: Complexity and algorithms. In *AAAI*, pages 1846–1853. AAAI Press, 2020.

[Charon and Hudry, 2007] I. Charon and O. Hudry. A survey on the linear ordering problem for weighted or unweighted tournaments. *4OR*, 5(1):5–60, 2007.

[Chudnovsky *et al.*, 2020] M. Chudnovsky, M. Pilipczuk, M. Pilipczuk, and S. Thomassé. Quasi-polynomial time approximation schemes for the maximum weight independent set problem in *H*-free graphs. In *SODA*, pages 2260–2278. SIAM, 2020.

[Downey and Fellows, 1999] R. Downey and M. Fellows. *Parameterized Complexity*. Springer, 1999.

[Dujmovic *et al.*, 2003] V. Dujmovic, H. Fernau, and M. Kaufmann. Fixed parameter algorithms for one-sided crossing minimization revisited. In *GD*, volume 2912, pages 332–344. Springer, 2003.

[Dwork *et al.*, 2001] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *WWW*, pages 613–622. ACM, 2001.

[Farkas and Timotity, 2019] M. Farkas and D. Timotity. Voting issues: A brief history of preference aggregation. *WorldQuant*, pages 26:1–7, November 2019.

[Fernau *et al.*, 2011] H. Fernau, F. Fomin, D. Lokshtanov, M. Mnich, G. Philip, and S. Saurabh. Ranking and drawing in subexponential time. In *IWOCA*, volume 6460, pages 337–348. Springer, 2011.

[Fernau *et al.*, 2014] H. Fernau, F. Fomin, D. Lokshtanov, M. Mnich, G. Philip, and S. Saurabh. Social choice meets graph drawing: How to get subexponential time algorithms for ranking and drawing problems. *Tsinghua Science and Technology*, 19(4):374–386, 2014.

[Fernau, 2005] H. Fernau. *Parameterized Algorithmics: A Graph-Theoretic Approach*. Habilitationsschrift, Universität Tübingen, Germany, 2005.

[Fomin *et al.*, 2020] F. Fomin, P. Golovach, L. Jaffke, G. Philip, and D. Sagunov. Diverse pairs of matchings. In *ISAAC*. Schloss Dagstuhl, 2020.

[Galle, 1989] P. Galle. Branch & sample: A simple strategy for constraint satisfaction. *BIT*, 29(3):395–408, 1989.

[Habib and Möhring, 1994] M. Habib and R. Möhring. Treewidth of cocomparability graphs and a new order-theoretic parameter. *Order*, 11(1):47–60, 1994.

[Ingmar *et al.*, 2020] L. Ingmar, M. de la Banda, P. Stuckey, and G. Tack. Modelling diversity of solutions. In *AAAI*, volume 34, pages 1528–1535. AAAI Press, 2020.

[Karpinski and Schudy, 2010] M. Karpinski and W. Schudy. Faster algorithms for feedback arc set tournament, Kemeny rank aggregation and betweenness tournament. In *ISAAC*, volume 6506, pages 3–14. Springer, 2010.

[Kemeny, 1959] J. Kemeny. Mathematics without numbers. *Daedalus*, 88:571–591, 1959.

[Petit and Trapp, 2019] T. Petit and A. Trapp. Enriching solutions to combinatorial problems via solution engineering. *INFORMS Journal on Computing*, 31(3):429–444, 2019.

[Simjour, 2009] N. Simjour. Improved parameterized algorithms for the Kemeny aggregation problem. In *IWPEC*, volume 5917, pages 312–323. Springer, 2009.

[Wong and Reingold, 1991] D. Wong and E. Reingold. Probabilistic analysis of a grouping algorithm. *Algorithmica*, 6(2):192–206, 1991.