

Kemeny Elections with Bounded Single-peaked or Single-crossing Width

Denis Cornaz, Lucie Galand, Olivier Spanjaard

▶ **To cite this version:**

Denis Cornaz, Lucie Galand, Olivier Spanjaard. Kemeny Elections with Bounded Single-peaked or Single-crossing Width. 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013), Aug 2013, Beijing, China. pp.76-82. hal-01372505

HAL Id: hal-01372505
https://hal.science/hal-01372505

Submitted on 21 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Kemeny Elections with Bounded Single-Peaked or Single-Crossing Width

Denis Cornaz
LAMSADE-CNRS, UMR 7243
Université Paris Dauphine
Paris, France
denis.cornaz@dauphine.fr

Lucie Galand
LAMSADE-CNRS, UMR 7243
Université Paris Dauphine
Paris, France
lucie.galand@dauphine.fr

Olivier Spanjaard
LIP6-CNRS, UMR 7606
UPMC
Paris, France
olivier.spanjaard@lip6.fr

Abstract

This paper is devoted to complexity results regarding specific measures of proximity to single-peakedness and single-crossingness, called "single-peaked width" [Cornaz *et al.*, 2012] and "single-crossing width". Thanks to the use of the PQ-tree data structure [Booth and Lueker, 1976], we show that both problems are polynomial time solvable in the general case (while it was only known for single-peaked width and in the case of *narcissistic preferences*). Furthermore, we establish one of the first results (to our knowledge) concerning the effect of nearly single-peaked electorates on the complexity of an NP-hard voting system, namely we show the fixed-parameter tractability of Kemeny elections with respect to the parameters "single-peaked width" and "single-crossing width".

1 Introduction

Social choice theory deals with making collective choices on the basis of the individual preference relations of a set of voters (or agents) over a set of candidates (or alternatives). This issue does not occur only in voting, but also in all situations involving the aggregation of several rankings coming from many different sources, e.g. sports (think of the Formula 1 championship), meta-search engines (merging the results of various search engines) [Dwork *et al.*, 2001], collaborative filtering [Pennock *et al.*, 2000]. One of the most popular method to aggregate rankings is the Kemeny rule [Kemeny, 1959], which minimizes the number of cases where the final ranking disagrees with the input rankings on the order of two alternatives. However, this rule is NP-hard to implement [Bartholdi *et al.*, 1989], even in the case of four input rankings [Dwork *et al.*, 2001]. Many computational studies have been carried out for this aggregation rule: polynomial time algorithms that give provably good solutions [van Zuylen and Williamson, 2007; Ailon *et al.*, 2008], a Polynomial Time Approximation Scheme [Kenyon-Mathieu and Schudy, 2007] which has impractical running times, heuristics, branch and bound procedures [Davenport and Kalagnanam, 2004; Conitzer *et al.*, 2006]. To illustrate the interest of computer scientists for this rule, note that a recent paper performs a comparison of no less than 104 algorithms devoted to Kemeny aggregation [Ali and Meilă, 2012]. Fixed-parameter algorithms for Kemeny rankings is another very active research direction [Betzler *et al.*, 2009]. Our work follows it. More precisely, our contribution is twofold:

- *We propose new fixed-parameter tractability results* for computing an optimal ranking according to the Kemeny rule. To that end, we follow the approach suggested by Guo *et al.* [2004]: we introduce parameters that measure the distance from instances that are solvable in polynomial time. As an illustration, consider instances where all input rankings are identical. The problem is then easy since an optimal ranking is of course that identical ranking. One can define the distance of an instance from the case of identical rankings as the average pairwise distance between two input rankings (for example), and investigate if the problem is fixed-parameter tractable with respect to this parameter "average pairwise distance". Actually, this question has already been answered positively by Betzler *et al.* [2009]. We focus here on new parameters that measure the distance from the single-peaked and single-crossing domains. Single-peakedness and single-crossingness are the most popular domain restrictions in social choice theory. Preferences are *single-peaked* (SP for short) if there exists a linear ordering of the alternatives such that, for all input rankings, the ranks deteriorate along the axis when moving away from the preferred alternative. Preferences are *single-crossing* (SC for short) if there exists a linear ordering of the input rankings such that, for any pair of alternatives, they can be partitioned into a left and a right block where all rankings on the left (resp. right) coincide on the preference over these two alternatives. After recalling that the computation of an optimal ranking can be performed in polynomial time if the preferences are SP (resp. SC), we show that the problem is fixed-parameter tractable with respect to the parameter "single-peaked width" [Cornaz *et al.*, 2012] (resp. "single-crossing width"), that measures how far the preferences are from being SP (resp. SC). To our knowledge, this result is one of the first concerning the effect of nearly SC/SP electorates on the complexity of an NP-hard voting system.

- *We establish the polynomial time computability of the single-peaked and single-crossing widths*. There are multiple ways of defining a distance measure regarding SPness or SCness [Escoffier *et al.*, 2008; Conitzer, 2009; Faliszewski

et al., 2011]. Most of these measures are NP-hard to compute [Erdélyi *et al.*, 2012]. Interestingly, Cornaz *et al.* [2012] showed that computing the SP width is polynomial for *narcissistic preferences*. We strengthen and extend this result, by providing an algorithm with a better complexity in the narcissistic case and by establishing the polynomial time computability in the general case (which was an open question). We also prove that the computation of the single-crossing width, which had not yet been studied, can also be performed in polynomial time. All these positive results rely on the use of a data structure called *PQ-tree*, introduced by Booth and Lueker [1976] to represent a family of permutations (rankings) on a set of elements. As noted by Elkind et al. [2012], in the context of computational social choice, it can be fruitfully used to represent the structure of the input rankings and devise efficient algorithms taking advantage of that structure.

The paper is organized as follows. In Section 2, we present how to determine a Kemeny ranking if the preferences are SP or SC. In Sections 3 and 4, we present our fixed-parameter algorithm for the Kemeny ranking problem with respect to the parameter SC or SP width. In Section 5 we show how to compute the value of these parameters in polynomial time.

2 Rank Aggregation

Let V be a set of n voters and C a set of m candidates. Let $\mathcal{P}$ denote the set of preference relations $\succ_v$ over C for $v \in V$. This set is called a *preference profile*. The problem studied in social choice theory aims at determining a collective ranking based on $\mathcal{P}$. Thanks to appealing properties, the Kemeny voting scheme is very popular to perform this aggregation [Kemeny, 1959]. Given two votes v, w, and two candidates a, b, let $\delta_{ab}(v, w) = 1$ if v and w disagree on the relative ranking of a and b, and 0 otherwise. The *Kendall-Tau distance* (KT-distance for short) between v and w is defined as KT-dist$(v, w) = \sum_{\{a,b\}\subseteq C} \delta_{ab}(v, w)$. A *Kemeny ranking* r minimizes the *Kemeny score* $\text{KS}(r) = \sum_{v\in V} \text{KT-dist}(r, v)$. Note that a Kemeny ranking is not necessarily unique. For the Kemeny voting scheme, the Condorcet property holds: whenever a weak Condorcet winner exists (i.e., a candidate that beats every other candidate in at least half of the votes), she is ranked first in at least one Kemeny ranking.

We propose here a fixed-parameter algorithm for computing a Kemeny ranking. Several parameters have already been considered in the literature, for which fixed-parameter algorithms have been proposed (for details see the survey by Betzler *et al.* [2012]): the number m of candidates [Betzler *et al.*, 2009], the value k of the optimal Kemeny score [Betzler *et al.*, 2009; Simjour, 2009; Karpinski and Schudy, 2010], the maximum range of candidates positions $r_m = \max_{c\in C} w(c) - b(c) + 1$ where $b(c)$ (resp. $w(c)$) is the best (resp. worst) position of c in the votes [Betzler *et al.*, 2009; Simjour, 2009; Karpinski and Schudy, 2010], the average KT-distance $d_a = \frac{2}{n(n-1)} \sum_{\{v,w\}\subseteq V} \text{KT-dist}(v, w)$ [Betzler *et al.*, 2010]. We establish new parameterized complexity results, by considering a parameter measuring how far a given profile is from being SP, and another parameter measuring how far it is from being SC.

Definition 1 *A preference profile $\mathcal{P}$ is single-peaked with respect to a numbering $(c_1, \ldots, c_m)$ of the candidates if for all $v \in V$ there exists p in $\{1, \ldots, m\}$ such that: if $p > j > i$ or $i > j > p$ then $c_p \succ_v c_j \succ_v c_i$.*

Definition 2 *The single-crossing condition holds for a preference profile $\mathcal{P}$ with respect to a numbering $(v_1, \ldots, v_n)$ of the voters if for all $\{a, b\} \subseteq C$ with $a \succ_{v_1} b$ there exists p in $\{1, \ldots, n\}$ such that $a \succ_{v_i} b$ for $i \leq p$ and $b \succ_{v_i} a$ for $i > p$.*

Interestingly, computing a Kemeny ranking is polynomial if the preference profile is SP or SC. We quickly present the result because it eases the understanding of the fixed-parameter algorithms. To that aim, we need to introduce the notion of Condorcet ranking. We call a ranking r of the candidates a *Condorcet ranking* if the candidate ranked first in r is a weak Condorcet winner, the candidate ranked second is a weak Condorcet winner over the remaining candidates, etc. (Note that the definition used here differs from the standard definition.) A Condorcet ranking does not necessarily exist but this is the case if the preferences are SP or SC. These domain restrictions indeed imply the existence of a Condorcet ranking by the two following properties: if the preferences are SP (resp. SC), 1) a weak Condorcet winner always exists and 2) the restriction of the profile to every subset $C' \subseteq C$ is also SP (resp. SC). A Condorcet ranking, if it exists, can clearly be computed in polynomial time. The polynomiality of computing a Kemeny ranking on SP and SC domains is then a direct consequence of the following result:

Proposition 1 ([Barbut, 1980]) *A Condorcet ranking is a Kemeny ranking.*

We will show that this approach generalizes to profiles not far to belong to the SP or SC domain. To that aim, we need to define a distance from SPness or SCness.

3 Single-peaked and Single-crossing Widths

The distance we consider is based on the notion of *component on profile* [Laffond *et al.*, 1996], also called *clone set* [Tideman, 1987] or *interval of candidates*. A subset $I \subseteq C$ of candidates is an *interval* if the candidates in I are consecutive in the preferences of every voter in $\mathcal{P}$. More formally:

Definition 3 *A subset $I \subseteq C$ of candidates is an* interval *if*

$$\forall v \in V, \forall a, b \in I, \forall c \in C \setminus I, a \succ_v c \iff b \succ_v c$$

For instance, consider the profile depicted on the left hand side of Figure 2 (the first column means: $a \succ_1 b \succ_1 c \succ_1 d \succ_1 e \succ_1 f \succ_1 g$): subsets $\{a, b, c\}$, $\{e, f, g\}$ and $\{f, g\}$ are intervals (among others).

We can now introduce the distance we consider. In the previous profile, the preferences are not SP. However, consider the partition of the candidates into intervals $\{a, b, c\}$, $\{d\}$ and $\{e, f, g\}$ and the profile obtained by substituting each interval by a single representative candidate within it: for instance, replace $\{a, b, c\}$ by a, $\{d\}$ by d and $\{e, f, g\}$ by e. The obtained profile over a, d, e is SP. We call the operation of substitution a *contraction of interval*. Any profile $\mathcal{P}'$ obtained from $\mathcal{P}$ by contraction of intervals is called a *minor* of $\mathcal{P}$. The *width* of a partition of C into intervals $I_1, \ldots, I_q$ is $\max_i |I_i| - 1$. Given a profile, the minimum width among all partitions yielding an

SP contracted profile is called *SP width* [Cornaz *et al.*, 2012]. We define the *SC width* similarly: it is the minimum width among all partitions such that the SC condition holds.

4 Parameterization by the Width

We now detail how to extend the polynomial algorithm for SP or SC domains to obtain a positive result based on the parameters SP width and SC width. Following Conitzer [2006] and Betzler *et al.* [2010], the idea is to break instances into several subinstances to be handled independently, that is, the relative order between the candidates in two different subinstances is already determined. The subinstances correspond to intervals $I_1, \ldots, I_q$ of a partition yielding an SP or SC minor when contracting $I_1, \ldots, I_q$. The algorithm is written:

1. Compute a Condorcet ranking over representative candidates within $I_1, \ldots, I_q$;

2. Compute a Kemeny ranking independently for each subinstance I_k ($k = 1, \ldots, q$);

3. Concatenate the Kemeny rankings over $I_1, \ldots, I_q$ according to the Condorcet ranking obtained at Step 1.

The ranking obtained at Step 3 is a Kemeny ranking over C.

Example 1 *Come back to the profile of Figure 2 and partition the candidate set into intervals $I_1 = \{a, b, c\}$, $I_2 = \{d\}$, $I_3 = \{e, f, g\}$, with representative candidates a, d, e. In Step 1, a Condorcet ranking is $a \succ d \succ e$. In Step 2, a Kemeny ranking for I_1 (resp. I_3) is $a \succ c \succ b$ (resp. $e \succ f \succ g$). In Step 3, the Kemeny rankings over I_1, I_2, I_3 are concatenated to obtain: $a \succ c \succ b \succ d \succ e \succ f \succ g$.*

The validity of the approach is related to the following observation, where $V_{cc'} = \{v \in V : c \succ_v c'\}$:

Observation 1 ([Betzler *et al.*, 2010]) *Let $C' \subseteq C$ be a candidate subset with $|V_{c'c}| \geq |V_{cc'}|$ for every $c' \in C'$ and $c \in C \setminus C'$. Then there must be a Kemeny ranking where all candidates in C' are ranked before all candidates in $C \setminus C'$.*

Note that it corresponds to a special case of *composition-consistency*, a useful property to compute tournament solutions (i.e., functions that associate with each complete and asymmetric relation on a set of alternatives a non-empty subset of the alternatives) by decomposing the tournament into subtournaments [Brandt *et al.*, 2011]. Let I_{k_0} denote an interval in $I_1, \ldots, I_q$ whose representative candidate is a weak Condorcet winner in the minor obtained by contraction of the intervals. Then $|V_{c'c}| \geq |V_{cc'}|$ for every $c' \in I_{k_0}$ and $c \in C \setminus I_{k_0}$. By Observation 1, there must be a Kemeny ranking r where all candidates in I_{k_0} are ranked before all candidates in $C \setminus I_{k_0}$. Moreover, it is easy to show that the ranking over I_{k_0} (resp. $C \setminus I_{k_0}$) should be a Kemeny ranking for the subinstance over I_{k_0} (resp. $C \setminus I_{k_0}$). By inductively applying these arguments, the validity of the algorithm follows. Note that, to compute a Kemeny ranking for each subinstance I_k in Step 2, rather than using brute force (in $O(m!)$), one can perform dynamic programming in $O(2^m)$ [Betzler *et al.*, 2009].

Let $(I_1, \ldots, I_q)$ be a partition of the candidates into q intervals which minimizes the SP (or SC) width, and let w denote the width of this partition. Step 1 of the algorithm (compute a Condorcet ranking) is then in $O(nq^2)$, Step 2 (compute a Kemeny ranking within each interval I_k) is in $O(q2^w)$, and Step 3 (concatenate the rankings) is in $O(m)$. Since $q \leq m$:

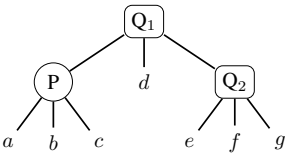


Figure 1: A PQ-tree.

Proposition 2 *The Kemeny ranking problem can be solved in $O(m2^w + nm^2)$, where w is the single-peaked or single-crossing width of the preference profile.*

The rest of the paper is dedicated to a study of the complexity of determining the SP or SC width of a profile.

5 Computing the Width of a Profile

In order to determine the SP or SC width of a profile $\mathcal{P}$, a preliminary task is to compute the set $\mathcal{I}(\mathcal{P})$ of all intervals. This set is closed under intersection [Elkind *et al.*, 2012]. Given $a, b \in C$ (possibly equal), the minimal interval w.r.t. inclusion that contains a and b is thus uniquely defined: we denote it by $I(a, b)$. Conversely, for every interval I, there exists a pair a, b of candidates such that $I = I(a, b)$: pick for instance the first and the last candidates from I in the preferences of a voter. Thus, set $\mathcal{I}(\mathcal{P})$ has a cardinality in $O(m^2)$ (number of pairs of candidates). Due to space constraint, the (fairly simple) algorithm in $O(nm^3)$ to compute this set is omitted.

Our study makes extensive use of the PQ-tree data structure. We therefore recall here the definition. A PQ-tree T on a set C is an ordered tree (that is, a tree with leaf set C together with a linear ordering of C) that represents a family of orderings over C. The internal nodes of T are either Q-nodes or P-nodes and have at least two children (by convention, all internal nodes with exactly two children are P-nodes). The *family of orderings $\mathcal{O}(T)$ of T* is the set of orderings of all trees in its equivalence class. Two PQ-trees on C are *equivalent* if one may be obtained from the other by arbitrarily reordering the children of a P-node or by reversing the children of a Q-node. For illustration, a PQ-tree T over $C = \{a, b, c, d, e, f, g\}$ is depicted in Figure 1. For instance, ordering $e \succ f \succ g \succ d \succ b \succ c \succ a$ belongs to $\mathcal{O}(T)$.

The next section is devoted to the presentation of algorithms for recognizing SC/SPness, which will be building blocks for computing the width of a profile.

5.1 Recognizing SP/SCness

Bartholdi and Trick [1986] showed that recognizing SPness of a profile $\mathcal{P}$ amounts to testing if a binary matrix has the consecutive ones property, i.e. if there is a permutation of its columns that places the 1's consecutive in every row. Note that a profile $\mathcal{P}$ can be SP with respect to several axes. There can even be an *exponential* number of such compatible axes [Escoffier *et al.*, 2008]. The interest of the method by Bartholdi and Trick is that it enables to compute a compact representation of *all axes* in *polynomial* time. Before detailing the binary matrix to consider, let us first give the characterization of SPness that led to it: a profile is SP iff all

input rankings are searches over a single chain Γ whose vertices are the candidates. Associate a 0-1 matrix $N(\mathcal{P})$ to $\mathcal{P}$ as follows: $N(\mathcal{P})$ has row-set $V \times \{1, \ldots, m\}$, column-set C, and $N(\mathcal{P})_{vk,a} = 1$ if a is one of the k first-ranked candidates for v, and 0 otherwise. Clearly, if the preferences are SP w.r.t. axis Γ, the subgraph H of Γ induced by the k first-ranked candidates of a voter v is a path for any $k = 1, \ldots, m$ and for any voter $v \in V$. Hence one has:

Observation 2 ([Bartholdi III and Trick, 1986]) *$\mathcal{P}$ is SP if and only if $N(\mathcal{P})$ has the consecutive-ones property.*

In a seminal paper, Booth and Lueker [1976] provides a *polynomial time* algorithm to decide if a binary matrix M has the consecutive ones property and, if yes, compute a *PQ-tree* for M, which gives an implicit representation of *all* the consecutive-ones orderings of the columns of M. For instance, consider the matrix on the left hand side of Figure 2 (denote it by M). Booth and Lueker's algorithm computes from M the PQ-tree in Figure 1 (denote it by T), where $\mathcal{O}(T)$ correspond to the consecutive-ones orderings of M. The algorithm runs in $O(x + y + z)$, where x and y are respectively the number of columns and rows, and z is the total number of ones in the matrix. Here, matrix $N(\mathcal{P})$ has m columns, nm rows and $O(nm^2)$ ones. It follows that *all* axes compatible with $\mathcal{P}$ can be determined in $O(nm^2)$. The complexity of recognizing SPness can be improved to $O(nm)$ if looking for only *one* compatible axis [Escoffier *et al.*, 2008].

Bredereck *et al.* [2012] showed that recognizing SCness of a profile $\mathcal{P}$ also amounts to testing the consecutive ones property. For the reduction, they define the matrix $M(\mathcal{P})$ obtained from $\mathcal{P}$ as follows: the row-set of $M(\mathcal{P})$ is indexed by $(a, b) \in C^2$ ($a \neq b$), the column-set is indexed by V, and $M(\mathcal{P})_{ab,v} = 1$ if $a \succ_v b$. Clearly $\mathcal{P}$ is SC iff $M(\mathcal{P})$ has the consecutive ones property. The complexity is $O(nm^2)$.

5.2 The Narcissistic Case

Preferences are called *narcissistic* when each candidate is most preferred by some voter [Bartholdi III and Trick, 1986; Trick, 1989]. The computation of the SP width in the narcissistic case has been investigated by Cornaz *et al.* [2012]. Furthermore, an algorithm proposed by Elkind *et al.* [2012] for a closely related problem also makes it possible to compute the SP width in this case (both problems become equivalent). Concerning the SC width, since the notion has been introduced in the present paper, no algorithm has been proposed up to now. We propose here a general method for computing the SP/SC width if preferences are narcissistic. This method is much faster than Cornaz *et al.*'s algorithm (that runs in $O(m^3n^2 \max\{m, n\})$) and much simpler than that of Elkind *et al.* (for a similar complexity). The key property is:

Proposition 3 *If the preferences are narcissistic, then for any pair I, I' of intervals: $I \cap I' = \emptyset$ or $I \subseteq I'$ or $I' \subseteq I$.*

Proof. By contradiction. Assume that $I \cap I' \neq \emptyset$ and $I \not\subseteq I'$ and $I' \not\subseteq I$. Let $c \in I \cap I'$. Consider the ranking of a voter v whose most preferred candidate is c (v exists since the preferences are narcissistic). Let c' denote the last candidate of $I \cup I'$ in the ranking of v. If $c' \in I$ (resp. $c' \in I'$) then $I' \subseteq I$ (resp. $I \subseteq I'$). Contradiction. ∎

$$\begin{pmatrix} a & a & e \\ b & c & f \\ c & b & g \\ d & d & d \\ e & g & c \\ f & f & a \\ g & e & b \end{pmatrix}$$

	a	b	c	d	e	f	g
abc	1	1	1	0	0	0	0
$abcd$	1	1	1	1	0	0	0
$defg$	0	0	0	1	1	1	1
ef	0	0	0	0	1	1	0
fg	0	0	0	0	0	1	1
efg	0	0	0	0	1	1	1

Figure 2: A preference profile and its interval matrix.

To test whether the SP width (resp. SC width) of a profile is at most w, one can proceed as follows. Let $\mathcal{P}(w)$ denote the profile obtained by contracting all the intervals of cardinality at most $w + 1$. If $\mathcal{P}(w)$ is SP (resp. SC) then the SP width (resp. SC width) is at most w. To compute the SP (resp. SC) width, it suffices therefore to simply iterate from $w = 0$ until the test is successful (*threshold method*). The SP or SC width is equal to the first value w for which the test succeeds. The number of iterations is of course bounded by $m - 1$.

The complexity of the threshold method is strongly related to that of Booth and Lueker's algorithm, but also to the time spent on computing the set of all intervals of $\mathcal{P}$. As indicated previously, the computation of all the intervals is in $O(nm^3)$. Both in $M(\mathcal{P})$ and $N(\mathcal{P})$, the total number of ones is within $O(nm^2)$, hence in both cases Booth and Lueker's algorithm runs in $O(nm^2)$. Since the algorithm is launched at most $m - 1$ times in the threshold method, the complexity is $O(m^2 \max\{m, n\})$ in the solution phase. The most costly phase from the computational viewpoint is therefore the precalculation of the set of intervals in $O(nm^3)$, and the overall complexity of the threshold method is thus $O(nm^3)$.

5.3 The General Case

To propose polynomial time algorithms for computing the widths in the general case, following Elkind *et al.* [2012], we will use a *PQ-tree* representation of the interval structure.

Preliminaries

Determining the interval structure of a profile $\mathcal{P}$ is actually equivalent to identifying the set of all preferences that are *compatible* with $\mathcal{I}(\mathcal{P})$. We say that a preference over the candidates is compatible with $\mathcal{I}(\mathcal{P})$ if, for all interval $I \in \mathcal{I}(\mathcal{P})$, the candidates of I are consecutive in the preference. Let us consider the 0-1-matrix defined from $\mathcal{I}(\mathcal{P})$ in the following way (see Figure 2). For each candidate c, the matrix contains a corresponding column. For each interval I, the matrix has a corresponding row with value 1 at column c if candidate c belongs to interval I and value 0 otherwise. The resulting matrix (*interval matrix*) has m columns and $|\mathcal{I}(\mathcal{P})|$ rows. Clearly, a preference is compatible with $\mathcal{I}(\mathcal{P})$ iff the corresponding ordering of the candidates permutes the column of the matrix so that the ones in each row are consecutive.

The PQ-tree T generated from the interval matrix represents the consecutive ones orderings of the columns. Here, set $\mathcal{O}(T)$ is the set of all preferences compatible with $\mathcal{I}(\mathcal{P})$. That PQ-tree is denoted by $T(\mathcal{P})$ and is called the *PQ-tree of* $\mathcal{P}$. For illustration, the PQ-tree of the profile in Figure 2 is depicted in Figure 1. Computing $T(\mathcal{P})$ from $\mathcal{I}(\mathcal{P})$ is in $O(m^3)$ (using Booth and Lueker's algorithm) because the size of the interval matrix is $O(m^2) \times m$. Before coming back to the interval structure, we need to introduce some definitions concerning PQ-trees. A *segment* of a Q-node is a subset of its

children which are consecutively ordered. A segment of cardinality one (resp. at least two) is called *degenerate* (resp. *non-degenerate*). For a Q-node x, we denote by $(y_1, \ldots, y_k)$ the sequence induced by the order on its children $y_1, \ldots, y_k$. So a segment of x is a subsequence of the form $(y_i, \ldots, y_j)$ with $1 \leq i \leq j \leq k$. The following proposition makes the link between the PQ-tree and the interval structure:

Proposition 4 *Let T be the PQ-tree of a profile $\mathcal{P}$, then I is an interval of $\mathcal{P}$ iff I is of one of the following types:*
(i) $|I| = 1$; or
(ii) T has a P-node x such that I is the set of all leaves reachable from x; or
(iii) T has a Q-node x with segment $(y_i, \ldots, y_j)$ $(i \neq j)$ such that I is the set of all leaves reachable from $y_i, \ldots, y_j$.

Proof. Sufficiency being trivial we only show necessity. By contradiction, assume there exists an interval I which satisfies none of conditions (i)-(iii). So $|I| \geq 2$. Let x be the minimum height internal node of T from which each candidate in I is reachable. If x is a P-node, then there exists a candidate $a \in C \setminus I$ reachable from x. So there is an order in $\mathcal{O}(T)$ containing a subsequence (b, a, c) with $b, c \in I$; this is impossible. So x is a Q-node with a children subsequence (y_i, y_j, y_l) and there are three candidates a, b, c s.t. $b \in I$ is reachable from y_i, $c \in I$ is reachable from y_l, and $a \in C \setminus I$ is reachable from y_i. Again, there is an order in $\mathcal{O}(T)$ containing a subsequence (b, a, c) with $b, c \in I$; a contradiction. ■

For instance, in Figure 1, $\{a\}$ is an interval of type (i), $\{a, b, c\}$ of type (ii) and $\{a, b, c, d\}$ of type (iii).

Our approach to compute the SC/SP widths is based on labelings of the PQ-tree: determining the SC/SP widths will amount to finding a labeling L minimizing $w(L)$, each labeling L (fulfilling conditions detailed further) inducing a partition into intervals and $w(L)$ denoting the width of this partition. We call "optimal" a labeling minimizing $w(L)$ among the labelings for which the induced partitions is SC/SP. We now detail the labelings we consider: each P-node is either labeled *contracted* or *relaxed*, and each Q-node is either labeled *contracted*, *segmented* (in which case the children of the Q-nodes are partitioned into segments, with at least two segments in the partition and at least one non-degenerate segment) or *relaxed*. A labeling should satisfy the following rules: all nodes that are descendants of a node labeled *contracted* are also labeled *contracted*, as well as all nodes that are descendants of a node belonging to a non-degenerate segment. A partition into intervals can be inferred from the labeling by starting from the root of the PQ-tree and recursively considering the labels of each encountered node x:

- if x is a node labeled *contracted* then create the interval including all the descendant leaves of node x;
- if x is a Q-node labeled *segmented* then, for each segment, create the interval including all its descendant leaves, and consider the labels of those of its children that are not in a non-degenerate segment;
- if x is a node labeled *relaxed* then consider the labels of each of its children.

For instance, in Figure 1, if node Q_1 is labeled *segmented* with a segment that includes the two left children, and node Q_2 is labeled *relaxed*, then the obtained partition into intervals is $(\{a, b, c, d\}, \{e\}, \{f\}, \{g\})$. Note that *any* valid partition can be obtained by a convenient labeling of the nodes, and conversely *any* labeling of the nodes yields a valid partition into intervals, where we mean by "valid" that the elements of the partition are intervals. There is thus a bijection between the partitions into intervals and the valid labelings.

Before moving on to the results, we need to introduce some further notions and notation. Let $\mathcal{P}$ be a profile. Given a labeling L of $T(\mathcal{P})$, the minor obtained by contracting each interval I in the corresponding partition is denoted by $\mathcal{P}(L)$. For instance, consider again the profile of Figure 2. The profile induced by the previous labeling is : $a \succ_1 e \succ_1 f \succ_1 g$, $a \succ_2 g \succ_2 f \succ_2 e$, $e \succ_3 f \succ_3 g \succ_3 a$. By abuse of language, we say that a labeling L is SC if $\mathcal{P}(L)$ is SC. The following result holds both for SCness and SPness.

Lemma 1 *If $\mathcal{P}'$ is a minor of $\mathcal{P}$ and $\mathcal{P}$ is SC (resp. SP), then $\mathcal{P}'$ is SC (resp. SP).*

Proof. Consider the reduction to the consecutive ones problem. Matrix $M(\mathcal{P}')$ (resp. $N(\mathcal{P}')$) is obtained from $M(\mathcal{P})$ (resp. $N(\mathcal{P})$) by deletion of rows (resp. rows and columns). Hence if the consecutive ones property holds for $M(\mathcal{P})$ (resp. $N(\mathcal{P})$), it holds for $M(\mathcal{P}')$ (resp. $N(\mathcal{P}')$). ■

Single-crossing Width

The following lemma will reveal useful to compute SC width since it implies that there always exists an optimal labeling where all nodes are labeled either *relaxed* or *contracted*:

Lemma 2 *If a Q-node x is labeled* segmented *in an SC labeling L, then x can be labeled* relaxed *in L while preserving SCness.*

Proof. Let x be a node labeled *segmented* in L. Consider labeling L' obtained from L by labeling *relaxed* node x and labeling *contracted* the children involved in non-degenerate segments of x. We claim that the SCness of $\mathcal{P}(L)$ implies that $\mathcal{P}(L')$ is also SC: if (x, y, z) is a children subsequence of a Q-node, then the rows of ab (resp. ba) and bc (resp. cb) are indeed identical in $M(\mathcal{P})$, for any candidates $a \in T_x$, $b \in T_y$ and $c \in T_z$, where T_x (resp. T_y, T_z) is the subtree of $T(\mathcal{P})$ rooted in x (resp. y, z). Hence matrix $M(\mathcal{P}(L'))$ differs from $M(\mathcal{P}(L))$ only by adding rows that are already present in $M(\mathcal{P}(L))$. Thus if the consecutive ones property holds for $M(\mathcal{P}(L))$ then it holds for $M(\mathcal{P}(L'))$. ■

The following result is crucial for computing the SC width.

Theorem 1 *Let $\mathcal{P}$ be a profile and $L(w)$ be the labeling of $T(\mathcal{P})$ where all nodes with at most $w + 1$ descendant leaves are labeled* contracted *and all nodes with more than $w + 1$ descendant leaves are labeled* relaxed*. The single-crossing width of $\mathcal{P}$ is at most w iff $L(w)$ is SC.*

Proof. The other part being obvious, we only prove that $L(w)$ is SC if the SC width of $\mathcal{P}$ is at most w. Let L^* denote an optimal labeling where at least one Q-node is labeled *segmented*. By reiterating the transformation of Lemma 2 for any Q-node labeled *segmented*, we then obtain a labeling L where all nodes are labeled either *relaxed* or *contracted*. Moreover we clearly have $w(L) \leq w(L^*)$. Thus L is an

optimal labeling where all nodes are labeled either *relaxed* or *contracted*. Note that all nodes labeled *contracted* in L are also labeled *contracted* in $L(w)$ (they have at most $w+1$ descendant leaves). It follows that $\mathcal{P}(L(w))$ is a minor of $\mathcal{P}(L)$. By Lemma 1, the SCness of $\mathcal{P}(L)$ implies that of $\mathcal{P}(L(w))$. ■

As in the narcissistic case, to compute the SC width of $\mathcal{P}$, it suffices therefore to implement a threshold method where one iterates from $w = 0$ until labeling $L(w)$ (on PQ-tree $T(\mathcal{P})$) is SC. We briefly study the complexity of this procedure. The preliminary computation of $T(\mathcal{P})$ (by generating the interval matrix and applying Booth and Lueker's algorithm) is in $O(nm^3)$. The computation of $\mathcal{P}(L(w))$ is in $O(m)$ and recognizing SCness of $\mathcal{P}(L(w))$ is in $O(nm^2)$. Since the number of iterations of the threshold method is in $O(m)$, we have:

Theorem 2 *The single-crossing width of a profile can be computed in $O(nm^3)$.*

Interestingly, this result contrasts with that of Elkind *et al.* [2012], who show that the closely related problem consisting of finding an SC partition of maximal cardinality is NP-hard. This comes from the fact that identifying the subset of nodes to label *contracted* is much more intricate in their problem.

Single-peaked Width

Contrary to the SC case, label *segmented* cannot be omitted when computing an optimal labeling according to the SP width. Nevertheless, the following result holds:

Lemma 3 *If a Q-node x is labeled* segmented *with a segmentation into more than two segments in an SP labeling L, then x can be labeled* relaxed *in L while preserving SPness.*

The proof is omitted due to space constraint. It uses a characterization of SPness by forbidden configurations [Ballester and Haeringer, 2011; Cornaz *et al.*, 2012]. It follows from the lemma that we can focus on labelings where Q-nodes are labeled *contracted*, *relaxed* or *bisegmented* which means segmented into exactly two segments. This allows us to identify necessary and sufficient conditions for a labeling to be SP. Let us denote by $L_c(x)$ the labeling of $T(\mathcal{P})$ where all nodes but the set $A(x)$ of ancestors of x (including x itself) are labeled *contracted*. Furthermore, in this labeling, any Q-node $z \in A(x)$ is labeled *segmented* with a partition where the child c of z belonging to $A(x)$ is a degenerate segment (the rest of the partition consists of one or two segments according to the position of c in the children sequence). Finally, any P-node in $A(x)$ is labeled *relaxed*. If x is a Q-node, let $L_s(x)$ denote a labeling of $T(\mathcal{P})$ defined as $L_c(x)$ except that node x is segmented with a partition into two segments. Note that $L_s(x)$ is not uniquely defined: there are as many labelings $L_s(x)$ as bipartitions of the children set of x. Nevertheless, the minor $\mathcal{P}(L_s(x))$ is uniquely defined (up to isomorphism), and therefore the choice of the bipartition does not impact on the SPness of the obtained minor. We are now able to state the necessary and sufficient conditions for a labeling to be SP:

Theorem 3 *Let $\mathcal{P}$ be a profile. Let $C(\mathcal{P})$ denote the set of all internal nodes $x \in T(\mathcal{P})$ such that $L_c(x)$ is not SP. Let $S(\mathcal{P})$ denote the set of all Q-nodes $x \in T(\mathcal{P})$ such that $L_c(x)$ is not SP but $L_s(x)$ is SP. A labeling L of $T(\mathcal{P})$ is SP iff:*

(i) every node $x \in C(\mathcal{P})$ is labeled contracted*;*

(ii) every Q-node $x \in S(\mathcal{P})$ is labeled contracted *or* bisegmented.

Proof. *Necessity:* Let $x \in C(\mathcal{P})$. By Lemma 1, if x is not labeled *contracted* in L, then $\mathcal{P}(L)$ is not SP since $L_c(x)$ is a minor of $\mathcal{P}(L)$ (and $L_c(x)$ is not SP). So (i) holds. Let now x be a Q-node in $S(\mathcal{P})$. Since $L_c(x)$ is not SP, any labeling L where x is labeled *relaxed* is not SP ($L_c(x)$ is a minor of L). By Lemma 3, any labeling L where x is labeled *segmented* with a segmentation into more than two segments is not SP (otherwise $L_c(x)$ would have been SP). Thus (ii) holds. *Sufficiency:* (sketch) This proof is here again based on the characterization of SPness by the absence of forbidden configurations in the profile [Ballester and Haeringer, 2011; Cornaz *et al.*, 2012] (as in Lemma 3): if (i)-(ii) hold, then $L(\mathcal{P})$ cannot include any forbidden configuration. ■

Theorem 4 *The single-peaked width of a profile $\mathcal{P}$ can be computed in $O(nm^3)$.*

Proof. The PQ-tree $T(\mathcal{P})$ is computed in $O(nm^3)$. For each node x of the PQ-tree $T(\mathcal{P})$ of $\mathcal{P}$, it takes time $O(mn)$ to assign two new attributes: one stating whether x is in $C(\mathcal{P})$, in $S(\mathcal{P})$, or in none of these sets, and a second, denoted by $s(x)$, indicating the number of leaves (candidates) in the subtree rooted in x. The overall time for all nodes is therefore $O(m^2n)$ (there are $O(m)$ nodes). By Theorem 3, we can contract all nodes in $C(\mathcal{P})$, and, furthermore, now we only need all Q-nodes in $S(\mathcal{P})$ to be labeled *segmented* with two segments. Let $x \in S(\mathcal{P})$ with children sequence $(y_1, \ldots, y_k)$. Notice that $k \geq 3$. Denote $s_x(i,j) := \sum_{l=i}^{l=j} s(y_l)$ the number of leaves descending from segment $(y_i, \ldots, y_j)$. Let γ denote the minimum among $s_x(1, k-1)$, $s_x(2,k)$ and $\max_{i \in \{2,\ldots,k-1\}} \{s_x(1,i), s_x(i+1,k)\}$. Since x is labeled *segmented* and there are exactly two segments, the SP width of $\mathcal{P}$ is at least $\gamma - 1$. We partition the children of x into two segments, each one of cardinality at most γ (it exists by definition of γ). We segment that way, greedily (in $O(m^2)$), all nodes in $S(\mathcal{P})$. All contracted intervals have a cardinality which is at most the SP width (plus 1) of $\mathcal{P}$, and the profile is now SP. The overall complexity is $O(nm^3)$. ■

6 Conclusion

In this paper, we have established the fixed-parameter tractability of Kemeny elections with respect to the parameters "single-peaked width" and "single-crossing width". Furthermore, we have shown that, for any preference profile, the values of these parameters can be computed in polynomial time. Booth and Lueker's algorithm has proved a powerful tool in this concern. An interesting research direction would be to study if a reduction to the consecutive ones problem can be used for recognizing other domain restrictions than single-peakedness or single-crossingness.

Acknowledgments

We wish to thank the anonymous reviewers for their fruitful comments on an earlier version of the paper, and B. Montjardet for pointing our attention to several relevant references.

References

[Ailon *et al.*, 2008] N. Ailon, M. Charikar, and A. Newman. Aggregating inconsistent information: Ranking and clustering. *J. ACM*, 55(5), 2008.

[Ali and Meilă, 2012] A. Ali and M. Meilă. Experiments with Kemeny ranking: What works when? *Mathematical Social Sciences*, 64:28–40, 2012.

[Ballester and Haeringer, 2011] M.A. Ballester and G. Haeringer. A characterization of the single-peaked domain. *Social Choice and Welfare*, 36(2):305–322, 2011.

[Barbut, 1980] M. Barbut. Médianes, Condorcet et Kendall. *Mathématiques et Sciences Humaines*, 69:5–13, 1980.

[Bartholdi *et al.*, 1989] J.J. Bartholdi, 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(2):157–165, 1989.

[Bartholdi III and Trick, 1986] J. Bartholdi III and M. A. Trick. Stable matching with preferences derived from a psychological model. *Operations Research Letters*, 5(4):165 – 169, 1986.

[Betzler *et al.*, 2009] N. Betzler, M.R. Fellows, J. Guo, R. Niedermeier, and F.A. Rosamond. Fixed-parameter algorithms for Kemeny rankings. *Theor. Comput. Sci.*, 410(45):4554–4570, 2009.

[Betzler *et al.*, 2010] N. Betzler, R. Bredereck, and R. Niedermeier. Partial kernelization for rank aggregation: Theory and experiments. In *IPEC*, pages 26–37, 2010.

[Betzler *et al.*, 2012] N. Betzler, R. Bredereck, J. Chen, and R. Niedermeier. The multivariate algorithmic revolution and beyond. chapter Studies in computational aspects of voting: open problems of Downey and Fellows, pages 318–363. Springer-Verlag, 2012.

[Booth and Lueker, 1976] K.S. Booth and G.S. Lueker. Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms. *J. Comput. Syst. Sci.*, 13(3):335–379, 1976.

[Brandt *et al.*, 2011] F. Brandt, M. Brill, and H.G. Seedig. On the fixed-parameter tractability of composition-consistent tournament solutions. In *IJCAI*, pages 85–90, 2011.

[Bredereck *et al.*, 2012] R. Bredereck, J. Chen, and G. Woeginger. A characterization of the single-crossing domain. In *4th International Workshop on Computational Social Choice (COMSOC-2012)*, 2012.

[Conitzer *et al.*, 2006] V. Conitzer, A.J. Davenport, and J. Kalagnanam. Improved bounds for computing Kemeny rankings. In *AAAI*, pages 620–626, 2006.

[Conitzer, 2006] V. Conitzer. Computing Slater rankings using similarities among candidates. In *AAAI*, pages 613–619, 2006.

[Conitzer, 2009] V. Conitzer. Eliciting single-peaked preferences using comparison queries. *J. Artif. Intell. Res. (JAIR)*, 35:161–191, 2009.

[Cornaz *et al.*, 2012] D. Cornaz, L. Galand, and O. Spanjaard. Bounded single-peaked width and proportional representation. In *ECAI*, pages 270–275, 2012.

[Davenport and Kalagnanam, 2004] A.J. Davenport and J. Kalagnanam. A computational study of the Kemeny rule for preference aggregation. In *AAAI*, pages 697–702, 2004.

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

[Elkind *et al.*, 2012] E. Elkind, P. Faliszewski, and A.M. Slinko. Clone structures in voters' preferences. In *ACM Conference on Electronic Commerce*, pages 496–513, 2012.

[Erdélyi *et al.*, 2012] G. Erdélyi, M. Lackner, and A. Pfandler. The complexity of nearly single-peaked consistency. In *4th International Workshop on Computational Social Choice (COMSOC-2012)*, 2012.

[Escoffier *et al.*, 2008] B. Escoffier, J. Lang, and M. Öztürk. Single-peaked consistency and its complexity. In *Proceedings of 18th European Conference on Artificial Intelligence (ECAI 2008)*, pages 366–370. IOS Press, 2008.

[Faliszewski *et al.*, 2011] P. Faliszewski, E. Hemaspaandra, and L.A. Hemaspaandra. The complexity of manipulative attacks in nearly single-peaked electorates. In *TARK*, pages 228–237, 2011.

[Guo *et al.*, 2004] J. Guo, F. Hüffner, and R. Niedermeier. A structural view on parameterizing problems: Distance from triviality. In *Proc. of the 1st International Workshop on Parameterized and Exact Computation (IWPEC 2004)*, volume LNCS 3162, pages 162–173, 2004.

[Karpinski and Schudy, 2010] M. Karpinski and W. Schudy. Faster algorithms for feedback arc set tournament, Kemeny rank aggregation and betweenness tournament. In *ISAAC (1)*, pages 3–14, 2010.

[Kemeny, 1959] J. Kemeny. Mathematics without numbers. *Daedalus*, 88:571591, 1959.

[Kenyon-Mathieu and Schudy, 2007] C. Kenyon-Mathieu and W. Schudy. How to rank with few errors. In *STOC*, pages 95–103, 2007.

[Laffond *et al.*, 1996] G. Laffond, J. Lain, and J.-F. Laslier. Composition-consistent tournament solutions and social choice functions. *Social Choice and Welfare*, 13(1):75–93, 1996.

[Pennock *et al.*, 2000] D.M. Pennock, E. Horvitz, and C.L. Giles. Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering. In *AAAI/IAAI*, pages 729–734, 2000.

[Simjour, 2009] N. Simjour. Parameterized and exact computation. chapter Improved Parameterized Algorithms for the Kemeny Aggregation Problem, pages 312–323. Springer-Verlag, 2009.

[Tideman, 1987] T.N. Tideman. Independence of clones as a criterion for voting rules. *Social Choice and Welfare*, 4(3):185–206, 1987.

[Trick, 1989] M.A. Trick. Recognizing single-peaked preferences on a tree. *Mathematical Social Sciences*, 17(3):329–334, 1989.

[van Zuylen and Williamson, 2007] A. van Zuylen and D.P. Williamson. Deterministic algorithms for rank aggregation and other ranking and clustering problems. In *WAOA*, pages 260–273, 2007.