Median of an odd number of permutations

Guillaume Blin, Maxime Crochemore, Sylvie Hamel, Stéphane Vialette

Medians of an odd number of permutations*

Guillaume Blin[1], Maxime Crochemore[1,3], Sylvie Hamel[2] and Stéphane Vialette[1]
[1] Université Paris-Est, LIGM - UMR CNRS 8049, France
{gblin, Maxime.Crochemore, vialette}@univ-mlv.fr
[2] DIRO - Université de Montréal - QC, Canada
sylvie.hamel@umontreal.ca
[3] King's College London, UK

February 10, 2010

Abstract

Given m permutations π^1, $\pi^2 \ldots \pi^m$ of $\{1, 2, \ldots, n\}$ and a distance function d, the **median** problem is to find a permutation π^* that is the "closest" of the m given permutations. Here, we study the problem under the Kendall-τ distance that counts the number of pairwise disagreements between permutations. This problem is also known, in the context of rank aggregation, as the Kemeny Score Problem and has been proved to be NP-hard when $m \geq 4$. This article is an extension of [4], where some nice combinatorial properties of the case $m = 3$ where stated without proof, to the general case $m \geq 3$, m odd.

1 Introduction

The problem of finding the median of a set of m permutations of $[n]$ under the Kendall-τ distance is best known in the literature as the Kemeny Score Problem : given m voters and a list of n candidates that they have to order according to their preference, the problem consist in finding a *Kemeny consensus*. A Kemeny consensus is an order of the candidates that agrees the most with the order of the m voters, *i.e.*, that minimizes the sum of the disagreements. This problem has been proved to be NP-complete when $m \geq 4$ [6] (the complexity is unknown for $m = 3$ and polynomial-time solvable for $m = 2$) and some approximation algorithms have been derived. First, a randomized algorithm with approximation factor 11/7 [1] and then a deterministic one with approximation factor 8/5 [11] were designed. In 2007, a PTAS result has been obtained [9] and a year later, some fixed-parameter algorithms have been described [2]. In a previous article, we focused on the open case where $m = 3$ and derived some nice combinatorial properties of the medians [4]. In this contribution, we generalize those results and derive new ones for the general case of finding the medians of m permutations, where m is odd.

*partly supported by NSERC through an Individual Discovery Grant (Hamel)

This article is organized as follow. After introducing the basic definitions and notations (in Section 2), we provide, in Section 3, some combinatorial properties of the medians which allow us to reduce the search space of the brute force algorithm. Finally, in sections 4 and 5, we provide our heuristics, the results and some future works respectively.

2 Definitions and notations

A **permutation** π is a bijection of $[n] = \{1, 2 \ldots, n\}$ onto itself. The set of all permutations of $[n]$ is denoted $\mathcal{S}_n$. As usual, we denote a permutation π of $[n]$ as $\pi = \pi_1\pi_2 \ldots \pi_n$. The **identity permutation** corresponds to the identity bijection of $[n]$ and is denoted $\imath = 12 \ldots n$. The **inverse permutation** of π, denoted $\pi^{-1} = \pi_1^{-1} \ldots \pi_n^{-1}$, is the permutation such that $\pi \circ \pi^{-1} = \pi^{-1} \circ \pi = \imath$, where $\circ$ represents the composition of functions (e.g. $\pi = 25431$, $\pi^{-1} = 51432$). In other words, π_i^{-1} denotes the position of integer i in the permutation π. The **cardinality** of a set S is denoted $\#S$. A pair (π_i, π_j) of elements of the permutation π is called an **inversion** if $\pi_i > \pi_j$ and $i < j$. The **number of inversions** of a permutation π is denoted $\mathrm{inv}(\pi)$. Let us remark that, since the inversions are generators of $\mathcal{S}_n$, we can view $\mathcal{S}_n$ with these generators as a Coxeter group. In this context, the number of inversions of a permutation π is called the **length** of π and is denoted by $\ell(\pi)$. See Chapter 5 of [8] for more details.
The **Kendall-τ distance**, denoted d_{KT}, counts the number of pairwise disagreements between two permutations and can be defined formally as follows: for permutations π and σ of $[n]$, we have that

$$d_{KT}(\pi, \sigma) = \#\{(i,j) \,|\, i < j \text{ and } [(\pi_i^{-1} < \pi_j^{-1} \text{ and } \sigma_i^{-1} > \sigma_j^{-1}) \text{ or } (\pi_i^{-1} > \pi_j^{-1} \text{ and } \sigma_i^{-1} < \sigma_j^{-1})]\}.$$

Note that we can easily compute $\mathrm{inv}(\pi)$ as $\mathrm{inv}(\pi) = \mathrm{inv}(\pi^{-1}) = d_{KT}(\pi, \imath)$. Given any set of permutations $A \subseteq \mathcal{S}_n$ and a permutation π, we have

$$d_{KT}(\pi, A) = \sum_{\sigma \in A} d_{KT}(\pi, \sigma).$$

The problem of **finding a median of a set of permutations under the Kendall-τ distance** can be stated formally as follow:

Given $A \subseteq \mathcal{S}_n$, we want to find a permutation π^* of $[n]$ such that

$$d_{KT}(\pi^*, A) \leq d_{KT}(\pi, A), \text{ for all } \pi \in \mathcal{S}_n.$$

In order to represent the disagreements between pairs of elements in a permutation π with respect to the permutations in $A \subseteq \mathcal{S}_n$, we introduce hereafter the notion of *disagreements graph*.

Definition 1 *We call the* **disagreements graph** *of $\pi = \pi_1\pi_2 \ldots \pi_n$ with respect to a set $A \subseteq \mathcal{S}_n$, denoted $\mathcal{G}(\pi, A)$, the graph obtained from π by considering each π_i as* *a node and by drawing weighted edges between each pair of nodes (π_i, π_j), with $i < j$. The* **weight** *of an edge (π_i, π_j), denoted $w_{\mathcal{G}(\pi,A)}(\pi_i, \pi_j)$, represents the number of order disagreements of this pair of elements in π with the same pair in each permutation of the set A,* i.e., *the distance contribution of this pair in $d_{KT}(\pi, A)$. With this definition, we have that*

$$d_{KT}(\pi, A) = \sum_{\substack{(\pi_i,\pi_j) \\ i<j}} w_{\mathcal{G}(\pi,A)}(\pi_i, \pi_j).$$

Example 1 *Let $\pi = 4213$ and $A = \{\pi^1, \pi^2, \pi^3\}$, where $\pi^1 = 2134$, $\pi^2 = 2413$ and $\pi^3 = 4123$. Then, the disagreements graph of π with respect to A,* i.e. *$\mathcal{G}(\pi, A)$ is given in Figure 1.*

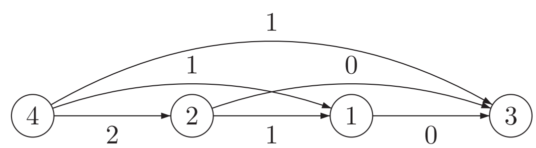

Figure 1: Disagreements graph of $\pi = 4213$ with respect to $A = \{2134, 2413, 4123\}$. Here $d_{KT}(4213, A) = 5$.

3 Reducing the search space

When dealing with permutations, searching through the whole set of permutations $[n]$ quickly becomes impossible since there are $n!$ such permutations. To be able to compare our heuristics with the brute force algorithm for permutations of $[n]$ where $n > 12$, we need to reduce the search space so that the computation will take place in a reasonable time. Here, given a set of permutations $A \subseteq \mathcal{S}_n$, with $\#A = m$, we derive some combinatorial properties on medians which will considerably reduce the search space by discarding unrelevant permutations.

In order to derive such properties for any median π^* of a set $A \subseteq \mathcal{S}_n$, let us consider the properties of its disagreements graph $\mathcal{G}(\pi^*, A)$. We first provide boundaries on the weight of edges of $\mathcal{G}(\pi^*, A)$ (Theorems 1 and 2).

Theorem 1 *Let $\pi^* = \pi^*_1 \ldots \pi^*_n$ be a median of a set of permutations $A \subseteq \mathcal{S}_n$ under the Kendall-τ distance and let $m = \#A$. Then, for $1 \le i \le n-1$,*

$$w_{\mathcal{G}(\pi^*,A)}(\pi^*_i, \pi^*_{i+1}) \le \lfloor \frac{m}{2} \rfloor.$$

Proof. By contradiction, let us assume that, given π^*, there exist an $1 \le i \le n-1$ s.t. $w_{\mathcal{G}(\pi^*,A)}(\pi^*_i, \pi^*_{i+1}) = k > \lfloor \frac{m}{2} \rfloor$. Let $\pi^{**} = \pi^*_1 \ldots \pi^*_{i-1} \pi^*_{i+1} \pi^*_i \pi^*_{i+2} \ldots \pi^*_n$ be the resulting permutation of the transposition of the elements in positions i and $i+1$

in π^*. Considering π^* and π^{**}, the order of any element of $[n]\backslash\{\pi^*_i, \pi^*_{i+1}\}$ and π^*_i or π^*_{i+1} is the same. The only difference concerns the relative order between π^*_i and π^*_{i+1}. Thus, all the edges of $\mathcal{G}(\pi^*, A)$ have the same weights than those of $\mathcal{G}(\pi^{**}, A)$ except the one between π^*_i and π^*_{i+1}.

More precisely, since $w_{\mathcal{G}(\pi^*,A)}(\pi^*_i, \pi^*_{i+1}) = k$, there are k permutations over the m permutations of A in which π^*_{i+1} precedes π^*_i. Since in π^{**}, π^*_i and π^*_{i+1} have been transposed, $w_{\mathcal{G}(\pi^{**},A)}(\pi^{**}_i, \pi^{**}_{i+1}) = m - k$. Since, by hypothesis, $k > \lfloor \frac{m}{2} \rfloor$, we have that $m - k < k$ and thus,

$$d_{KT}(\pi^{**}, A) = \sum_{\substack{(\pi^{**}_i, \pi^{**}_j) \\ i<j}} w_{\mathcal{G}(\pi^{**},A)}(\pi^{**}_i, \pi^{**}_j) < d_{KT}(\pi^*, A)$$

In other words, π^* is not, by definition, a median ; a contradiction. ∎

Theorem 2 *Let $\pi^* = \pi^*_1 \dots \pi^*_n$ be a median of a set of permutations $A \subseteq \mathcal{S}_n$ under the Kendall-τ distance and let $m = \#A$. Then the maximal weight of edges in $\mathcal{G}(\pi^*, A)$ is $m - 1$.*

Proof. By contradiction, let us assume that, given a median $\pi^* = \pi^*_1 \dots \pi^*_n$ of a set $A = \{\pi^1, \dots, \pi^m\}$ s.t. $m = \#A$, there exist $1 \leq i < j \leq n$ s.t. $w_{\mathcal{G}(\pi^*,A)}(\pi^*_i, \pi^*_j) = m$. Let π^{**} be the resulting permutation of the exchange of the elements in positions i and j in π^*. Considering the weights of the edges of $\mathcal{G}(\pi^*, A)$ and $\mathcal{G}(\pi^{**}, A)$, the only differences concern the edges (π^*_i, π^*_k), $i+1 \leq k \leq j$ and (π^*_k, π^*_j), $i+1 \leq k \leq j-1$ (See Figure 2).

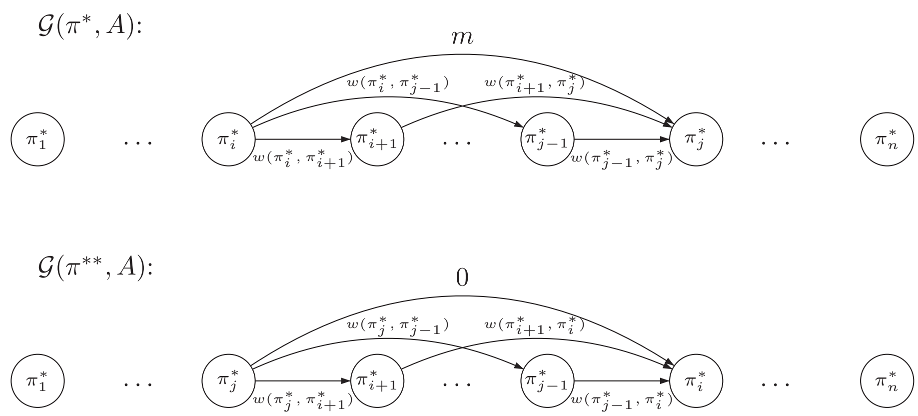

Figure 2: Disagreements graphs of π^* and π^{**}

Let us first prove a nice property on $\mathcal{G}(\pi^*, A)$: namely that $\forall i+1 \leq k \leq j-1$, $w_{\mathcal{G}(\pi^*,A)}(\pi^*_k, \pi^*_j) \geq m - w_{\mathcal{G}(\pi^*,A)}(\pi^*_i, \pi^*_k)$. First, note that, since $w_{\mathcal{G}(\pi^*,A)}(\pi^*_i, \pi^*_j) = m$, in all the permutations of A π^*_j precedes π^*_i. Consider now any $\forall i+1 \leq k \leq j-1$.

By definition, there are $m - w_{\mathcal{G}(\pi^*,A)}(\pi_i^*, \pi_k^*)$ permutations of A in which π_k^* appears after π_i^* and thus after π_j^*. Consequently, $w_{\mathcal{G}(\pi^*,A)}(\pi_k^*, \pi_j^*) \geq m - w_{\mathcal{G}(\pi^*,A)}(\pi_i^*, \pi_k^*)$.

Let us now compare $d_{KT}(\pi^*, A)$ and $d_{KT}(\pi^{**}, A)$. By construction, as mentionned above, $d_{KT}(\pi^*, A) - d_{KT}(\pi^{**}, A) = (m + \sum_{k=i+1}^{j-1} w_{\mathcal{G}(\pi^*,A)}(\pi_i^*, \pi_k^*) + w_{\mathcal{G}(\pi^*,A)}(\pi_k^*, \pi_j^*)) - (0 + \sum_{k=i+1}^{j-1} w_{\mathcal{G}(\pi^{**},A)}(\pi_j^{**}, \pi_k^{**}) + w_{\mathcal{G}(\pi^{**},A)}(\pi_k^{**}, \pi_i^{**}))$.

Note that, since $\sum_{k=i+1}^{j-1} w_{\mathcal{G}(\pi^{**},A)}(\pi_j^{**}, \pi_k^{**}) + w_{\mathcal{G}(\pi^{**},A)}(\pi_k^{**}, \pi_i^{**})) = \sum_{k=i+1}^{j-1}(m - w_{\mathcal{G}(\pi^*,A)}(\pi_i^*, \pi_k^*)) + (m - w_{\mathcal{G}(\pi^*,A)}(\pi_k^*, \pi_j^*))$, one may conclude that $d_{KT}(\pi^*, A) - d_{KT}(\pi^{**}, A) > 0$ and thus that π^* is not a median; a contradiction.

■

Theorem 2 states that there are no edges of weight m in $\mathcal{G}(\pi^*, A)$, where $m = \#A$. This means that if a pair of integers appears in the same order in all permutations of A then they have to appear in that order in any median π^* of A. Note that this theorem, in an other form, has already been stated and proved in the area of applied finance and uses what they called an Extended Condorcet Criterion [10].

For $A \subseteq \mathcal{S}_n$, let

$$\text{inv}(A) = \sum_{\sigma \in A} \text{inv}(\sigma).$$

The triangle inequality gives us a bound on the number of inversions in a median π^* of A.

Theorem 3

$$\frac{\text{inv}(A) - d_{KT}(\pi, A)}{m} \leq \text{inv}(\pi^*) \leq \frac{\text{inv}(A) + d_{KT}(\pi, A)}{m}$$

where $m = \#A$ and where π is considered, w.l.o.g., to be the permutation in A that minimizes $d_{KT}(\pi, A)$.

Proof. It is easy to show that the Kendall-τ distance is indeed a distance in the mathematical sense of the term. That means, in particular, that it needs to satisfy the triangle inequality ($\forall x, y, z$, permutations of $[n]$, $d_{KT}(x, z) \leq d_{KT}(x, y) + d_{KT}(y, z)$). Let $A = \{\pi^1, \ldots, \pi^m\}$. By the triangle inequality we have that for all $1 \leq i \leq m$,

$$\text{inv}(\pi^*) = d_{KT}(\pi^*, \imath) \leq d_{KT}(\pi^*, \pi^i) + d_{KT}(\pi^i, \imath) = d_{KT}(\pi^*, \pi^i) + \text{inv}(\pi^i),$$

Summing these inequalities we get

$$m \text{ inv}(\pi^*) \leq \sum_{i=1}^{m} d_{KT}(\pi^*, \pi^i) + \text{inv}(\pi^i) \ = \ d_{KT}(\pi^*, A) + \text{inv}(A).$$

Now, since π^* is the median of A with respect to the Kendall-τ distance we need to have that

$$d_{KT}(\pi^*, A) \leq d_{KT}(\pi^i, A),$$

for all $1 \leq i \leq m$. Choosing π to be the $\pi^i \in A$ that minimizes $d_{KT}(\pi, A)$, we have

$$m \text{ inv}(\pi^*) \leq d_{KT}(\pi, A) + \text{inv}(A),$$

which gives us the first inequality. The second inequality is easily derived in the same manner from the following triangle inequalities, $1 \leq i \leq m$:

$$\mathrm{inv}(\pi^i) = d_{KT}(\pi^i, \imath) \leq d_{KT}(\pi^i, \pi^*) + d_{KT}(\pi^*, \imath) = d_{KT}(\pi^i, \pi^*) + \mathrm{inv}(\pi^*).$$

■

Theorem 3 gives upper and lower bounds (but not optimal ones) on the number of inversion in a median π^*. This is interesting since there exist a CAT-algorithm that computes all the permutations of $[n]$ having exactly k inversions [7]. This Theorem thus reduces the search space for a median while Theorem 1 and 2 gives us a set of constraints that a median should satisfy.
Table 1 compares the computation time needed to find the medians of 3 permutations of $[n]$, for $4 \leq n \leq 11$, using 1) the brute force algorithm and 2) the brute force algorithm optimized by the results of Theorem 1 to 3.

n	4	5	6	7	8	9	10	11
time BF	0	0.0002	0.0005	0.0042	0.04	0.425	5.03	63.33
time BFopt.	0	0	0.0002	0.0012	0.0064	0.0238	0.1496	1.0052

Table 1: Running time, in seconds, of the brute force algorithm with and without the optimizations

4 Our heuristics

The idea of our heuristics is to apply a series of "good" cyclic movements on the permutations in A, in order to make them closer to a median. Our heuristics only works with sets of permutations of odd cardinality (see Theorem 4). Formally we have the following definitions and algorithm.

Definition 2 *Given $\pi = \pi_1 \ldots \pi_n$, we call* **cyclic movement** *of a segment $\pi[i..j]$ of π, denoted $c[i,j](\pi)$, the cycling shifting of one position to the right ($c_r[i,j]$) or to the left ($c_\ell[i,j]$) of the segment inside the permutation π:*

$$c_r[i,j](\pi) = \pi_1 \ldots \pi_{i-1}\pi_{\mathbf{j}}\pi_{\mathbf{i}} \ldots \pi_{\mathbf{j-1}}\pi_{j+1} \ldots \pi_n,$$

$$c_\ell[i,j](\pi) = \pi_1 \ldots \pi_{i-1}\pi_{\mathbf{i+1}} \ldots \pi_{\mathbf{j}}\pi_{\mathbf{i}}\pi_{j+1} \ldots \pi_n$$

When $j = i+1$, a cyclic movement corresponds to a **transposition***.*

Definition 3 *Given a set of permutations $A \subseteq \mathcal{S}_n$, we will say that a cyclic movement is a* **k-move** *if*

$$d_{KT}(c[i,j](\pi), A) = d_{KT}(\pi, A) + k.$$

Definition 4 *A* **good** *cyclic movement $c[i,j]$ is a k-move, where $k < 0$.*

This means that if we apply a good cyclic movement to π we obtain a permutation that is closer to a median than π, *i.e.*, we have $d_{KT}(c[i,j](\pi), A) < d_{KT}(\pi, A)$. Theorem 4 gives us a way to easily find these good moves (in fact any k-move) on a starting permutation π.

Theorem 4 *Let $A \subseteq \mathcal{S}_n$, be a set of permutations, with $\#A = m$, m odd. Let π be a permutation from which we want to derive π^*, a median of A with respect to the Kendall-τ distance. We have that $c_r[i,j](\pi)$ (resp. $c_\ell[i,j](\pi)$) is a k-move, $k \in \mathbb{Z}$, iff $j - i \equiv k \bmod 2$ and*

$$\sum_{t=i}^{j-1} w_{\mathcal{G}(\pi,A)}(\pi_t, \pi_j) \left(\textit{resp.} \sum_{t=i+1}^{j} w_{\mathcal{G}(\pi,A)}(\pi_i, \pi_t) \right) = \frac{m(j-i)-k}{2}.$$

Proof. $(\Rightarrow)$ If $c[i,j](\pi)$ is a k-move (to the right or left) then, by definition

$$d_{KT}(c[i,j](\pi), A) = d_{KT}(\pi, A) + k. \tag{1}$$

Between the two disagreements graphs $\mathcal{G}(\pi, A)$ and $\mathcal{G}(c_r[i,j](\pi), A)$ (resp. $\mathcal{G}(\pi, A)$ and $\mathcal{G}(c_\ell[i,j](\pi), A)$) only the edges (π_t, π_j), $i \le t \le j-1$ (resp. (π_i, π_t), $i+1 \le t \le j$) change weights. So, there are $j-i$ edges that change weights and their new weights are equals to m minus their old weights. Let x_p be the number of those edges that have weight p in $\mathcal{G}(\pi, A)$, we have

$$\sum_{p=0}^{m} x_p = j - i \tag{2}$$

and Equation 1 gives us that

$$\sum_{p=0}^{m} (m-p)x_p = \left[\sum_{p=0}^{m} p\, x_p \right] + k \tag{3}$$

that can be written has

$$\sum_{p=0}^{m} (m-2p)x_p = k \tag{4}$$

Subtracting Equation 2 from Equation 4 we finally get

$$\sum_{p=0}^{m} (m-2p-1)x_p = k - (j-i) \tag{5}$$

which gives us that $k-(j-i)$ has to be even (since m is odd) and so $j - i \equiv k \bmod 2$. Now to prove the second part of this implication, we have

$$\sum_{t=i}^{j-1} w_{\mathcal{G}(\pi,A)}(\pi_t,\pi_j) \left(\text{resp.} \sum_{t=i+1}^{j} w_{\mathcal{G}(\pi,A)}(\pi_i,\pi_t)\right) = \sum_{p=0}^{m} p\, x_p$$

$$= \frac{m\sum_{p=0}^{m} x_p - \sum_{p=0}^{m}(m-2p)x_p}{2}$$

$$\stackrel{(2)}{=} \frac{m(j-i) - \sum_{p=0}^{m}(m-2p)x_p}{2}$$

$$\stackrel{(4)}{=} \frac{m(j-i)-k}{2}.$$

($\Leftarrow$) We have to prove that if $a)$ $j-i \equiv k \bmod 2$ and $b)$ $\sum_{t=i}^{j-1} w_{\mathcal{G}(\pi,A)}(\pi_t,\pi_j)$ $\left(\text{resp.} \sum_{t=i+1}^{j} w_{\mathcal{G}(\pi,A)}(\pi_i,\pi_t)\right) = \frac{m(j-i)-k}{2}$ then $c_r[i,j](\pi)$ (resp. $c_\ell[i,j](\pi)$) is a k-move, $k \in \mathbb{Z}$ (*i.e.* Equation 3 holds). Taking the same notation as before for x_p, $0 \leq p \leq m$, $b)$ implies that

$$\sum_{p=0}^{m} p\, x_p = \frac{m(j-i)-k}{2} \tag{6}$$

We want to show that Equation 3 holds i.e that both parts of the equation are equals. The right part of Equation 3 is

$$\left[\sum_{p=0}^{m} p\, x_p\right] + k \stackrel{(6)}{=} \frac{m(j-i)-k}{2} + k = \frac{m(j-i)+k}{2}$$

We show now that the left part of Equation 3 is also equals to $\frac{m(j-i)+k}{2}$:

$$\sum_{p=0}^{m}(m-p)\, x_p = m\sum_{p=0}^{m} x_p - \sum_{p=0}^{m} p\, x_p$$

$$\stackrel{(2) et (6)}{=} m(j-i) - \left[\frac{m(j-i)-k}{2}\right]$$

$$= \frac{m(j-i)+k}{2}$$

■

It is now time to present our heuristics whose pseudo-code is depicted in Figure 3. The idea is to begin our search for the median in any permutation $\pi^\ell \in A$, $1 \leq \ell \leq m$, and to apply good cyclic movements to this starting point till there is no more possible good movement. We apply m times our pseudo-code, with $\pi = \pi^\ell$, $1 \leq \ell \leq m$ and our "median" is the best result we obtain from these m runs.

Algorithm FindMedian (π, A)
    n $\leftarrow$ length(π)
    bool $\leftarrow$ 0 (will be changed to 1 if there is no more possible "good" movement)
    chang $\leftarrow$ 0 (will tell us if some movements were made)
    WHILE bool <> 1 DO
        FOR i from 1 to $n-1$ DO
            FOR j from $i+1$ to n DO
                IF $c_r[i,j](\pi)$ or $c_\ell[i,j](\pi)$ is a good movement THEN
                    $\pi \leftarrow c_{good}[i,j](\pi)$
                    chang $\leftarrow$ chang +1
                END IF
            END FOR
        END FOR
        IF chang = 0 THEN
            bool $\leftarrow$ 1
        END IF
    END WHILE
    RETURN π

Figure 3: Pseudo-code of our heuristics FindMedian

We tested this heuristics on 2000 random sets A of permutations of $[n]$ for $6 \leq n \leq 11$, for the different cardinalities $\#A = 3, 5, 7$ and 9 (For bigger n, it was too long to get the real medians with the brute force algorithm). Table 4 shows the percentage of errors of our heuristics on these runs. This percentage decreases when the cardinality of A increases from 3 to 9, since a bigger cardinality means more starting points for our heuristics. It is worth to mention that when our heuristics could not find the real median π^* in these runs, the difference between the Kendall-τ distance of the permutation found by our heuristics and π^* was always one.

n	6	7	8	9	10	11
#A = 3	0	0.1	0.23	0.4	0.5	1
#A = 5	0	0.01	0.05	0.12	0.1	0
#A = 7	0	0.01	0.01	0.03	0	0
#A = 9	0	0	0	0.01	0	0

Table 2: Percentage of errors of our heuristics on 2000 sets A of permutations of $[n]$, $6 \leq n \leq 12$, where $\#A = m$, $m = 3, 5, 7, 9$.

Considering 0-moves

When our heuristics does not find the median π^*, it means that we are stuck in a local minimum and there is no more possible good cyclic movement that we can make. We decided in this case to apply a fixed number of 0-moves in hope that these moves will help us go out of the local minimum. Given a permutation π, we can easily find these 0-moves with Theorem 4. Among these 0-moves, if at least one has the property described in Theorem 5 we are guaranteed to move out of the local minimum. So, the 0-moves with this properties will be call "good".

Theorem 5 *Let π be the permutation found by our heuristics FindMedian as a candidate for a median π^* of a set of permutations A with $\#A = m$, m odd. If $c_r[i,j](\pi)$ (resp. $c_\ell[i,j](\pi)$) is a 0-move and $w_{\mathcal{G}(\pi)}(\pi_{i-1}, \pi_j)$ or $w_{\mathcal{G}(\pi)}(\pi_{j-1}, \pi_{j+1}) \geq \lceil \frac{m}{2} \rceil$ (resp. $w_{\mathcal{G}(\pi)}(\pi_{i-1}, \pi_{i+1})$ or $w_{\mathcal{G}(\pi)}(\pi_i, \pi_{j+1}) \geq \lceil \frac{m}{2} \rceil$), then there exist a good move (*i.e *a k-move, $k < 0$) in $c_r[i,j](\pi)$ (resp. $c_\ell[i,j](\pi)$).*

Proof. This is easy to see since in $c_r[i,j](\pi)$ (resp. $c_\ell[i,j](\pi)$) the elements π_{i-1}, π_j and π_{j-1}, π_{j+1} (resp. π_{i-1}, π_{i+1} and π_i, π_{j+1}) will become consecutive elements. Since the edges between these elements have weights $\geq \lceil \frac{m}{2} \rceil$ and m is odd, the transposition (π_{i-1}, π_j) or (π_{j-1}, π_{j+1}) in $\mathcal{G}(c_r[i,j](\pi), A)$ (resp. (π_{i-1}, π_{i+1}) or (π_i, π_{j+1}) in $\mathcal{G}(c_\ell[i,j](\pi), A)$) will then become a good move, *i.e.* a k-move, $k < 0$ ■

n	6	7	8	9	10
error %	0	0	0	0	0
% of cases with **0** 0-move	100	99.8	99.7	99.5	99.2
% of cases with **1** 0-move	0	0.15	0.2	0.4	0.5
% of cases with **2** 0-moves	0	0.05	0.1	0.1	0.2
% of cases with**3** 0-moves	0	0	0	0	0.1

Table 3: Percentage of errors of the heuristics FindMedianZeroMoves on 20 000 sets A of triplets of permutations of $[n]$, $6 \leq n \leq 10$ and percentages of cases where k 0-moves were needed to get the median, $0 \leq k \leq 3$.

We implemented a new heuristics "FindMedianZeroMoves(π,A)" that add to our previous heuristics the possibility to do a fixed number of 0-moves (good 0-moves or random ones). We tested this new heuristics on 20 000 randoms sets A of triplets of permutations of $[n]$, $6 \leq n \leq 10$, with 3 has the maximal number of 0-moves permitted. Table 3 shows our results. We see that on these examples we always got to a median of A and that in most of the cases, no 0-moves where needed to get to this median.

5 Future works

Since this article is a work in progress, there is still a lot of questions we need to answer. Stating only a few, we have the following ones: Can we find combinatorial properties that will completely describe the set of 0-moves that can make us move out of a local minimum? Are 0-moves enough to ensure that in all the cases, we can get out of local minimum? If not, a Monte Carlo approach could be a good probabilistic way to attack the problem.

We are also interested to investigate the number of medians of different sets $A \subseteq \mathcal{S}_n$. Some prelimary tests on sets of triplets of permutations show (see Table 4) that this number varies a lot (from 1 to 33 in our 2000 triplets of permutations of [10].) Can we find some combinatorial properties of A that indicates a small (resp. larger) number of medians? Also, it seems to be more probable for a set of permutations A to have a odd number of medians than an even one. Can we understand why? A possible direction to answer these questions could be to consider the Bruhat ordering of permutations; Maybe the position of the permutations of A in the Bruhat poset can help us understand the variation in the number of medians.

n	3	4	5	6	7	8	9	10
# sets considered	20	2024	280840	2000	2000	2000	2000	2000
% of triplets with:								
1 median:	90%	87,6%	80,6%	75,1%	68,2%	60,5%	55,6%	50%
2 medians:	0%	0%	1%	2,7%	4,15%	5,1%	6,3%	7,9%
3 medians:	10%	12,4%	16,7%	18,4%	20,9%	23,9%	22,2%	22,1%
4 medians:	0%	0%	0%	0,1%	0,5%	1,4%	2,5%	2,9%
5 medians:	0%	0%	1.7%	3,3%	5%	5,6%	7,4%	7,5%
6 medians:	0%	0%	0%	0%	0,1%	0.5%	1%	1,8%
7 medians:	0%	0%	0%	0%	0,3%	1%	1,2%	1,6%
8 medians:	0%	0%	0%	0%	0,05%	0,15%	0,3%	0,5%
9 medians:	0%	0%	0%	0,4%	0,7%	1,35%	2,3%	3%
$\geq$ 10 medians:	0%	0%	0%	0%	0,2%	0,5%	1,2%	2,7%

Table 4: Percentages of medians for triplets of permutations of $[n]$, $3 \leq n \leq 10$

Acknowledgements

We thank Quentin Dejean, Anthony Estebe and Kelsey Lang who programmed all the algorithms and realized the tests as part of internship projects.

References

[1] N. Ailon, M. Charikar and N. Newman, *Aggregating inconsistent information: Ranking and clustering*, In Proceedings of the 37th STOC, pp.684–693, 2005.

[2] N. Betzler, M.R. Fellows, J. Guo, R. Niedermeier and F.A. Rosamond, *Fixed-Parameter Algorithms for Kemeny Scores*, LNCS 5034, pages 60–71, 2008.

[3] T. Biedl, F.J. Brandenburg and X. Deng, *Crossings and Permutations*, LNCS 3843, pages 1–12, 2005.

[4] G. Blin, M. Crochemore, S. Hamel and S. Vialette, *Finding the median of three permutations under the Kendall-τ distance*, proceedings of the 7th Permutation Pattern conference, pp. 31–36, 2009.

[5] V. Conitzer, A. Davenport and J. Kalagnanam, *Improved Bounds for Computing Kemeny Rankings*, in Proceedings of the 21st National Conference on Artificial Intelligence (AAAI), pages 620–627, 2006.

[6] C. Dwork, R. Kumar, M. naor and D. Sivakumar, *Rank Aggregation Methods for the Web*, in proceedings of the 10th WWW, pp.613-622, 2001.

[7] S. Effler and F. Ruskey, *A CAT algorithm for generating permutations with a fixed number of inversions*, Information Processing Letters, Vol. 86-2, pp.107–112, 2003.

[8] J.E. Humphreys, *Reflection Groups and Coxeter Groups*, Cambridge University Press, (1990), 204 pages.

[9] C. Kenyon-Mathieu and W. Schudy, *How to rank with few errors*, In Proceedings of the 39th STOC, pp. 95-103, 2007.

[10] M. Truchon, *An Extension of the Condorcet Criterion and Kemeny Orders*, Internal Report, cahier 98-15 du Centre de Recherche en Économie et Finance Appliquées, Université Laval, 16 pages, 1998.

[11] A. vanZuylen and D.P. Williamson, *Deterministic algorithms for rank aggregation and other ranking and clustering problems*, in Proceedings of the 5th WAOA, LNCS 4927, pp.260–273, 2007.