

# Popular Ranking

Anke van Zuylen<sup>a</sup> Frans Schalekamp<sup>b</sup> David P. Williamson<sup>c,1</sup>

<sup>a</sup>*Max-Planck-Institut für Informatik, Saarbrücken, Germany*

<sup>b</sup>*Independent*

<sup>c</sup>*School of ORIE, Cornell University, Ithaca, NY, USA*

*Key words:* Rank aggregation, Kemeny Rank Aggregation, Popular Ranking

---

## 1 Introduction

How do you aggregate the preferences of multiple agents in a fair manner? This is a question that has occupied researchers for several centuries. Suppose we have  $k$  voters who each give their preferences on  $n$  candidates. How should the candidates be ranked to best represent the input? Marquis de Condorcet [5] showed that there may not exist a “winner”: a candidate who beats all other candidates in a pairwise majority vote. Borda [4] and Condorcet [5] (and many others after them) proposed different ways of aggregating the preferences of the voters, and argued over which method is the right one. Only in the middle of the 20th century, Arrow [2] showed that there is no right method: there exists no aggregation method that simultaneously satisfies three natural criteria (non-dictatorship, independence of irrelevant alternatives and Pareto efficiency).

This negative result notwithstanding, we still want to find aggregate rankings based on voters’ inputs. In this paper, we consider the case when, rather than selecting a winner, we would like to find a permutation of the candidates that represents the voters’ inputs. Each voter’s input is assumed to be a permutation of the candidates, where a candidate is ranked above another candidate, if the voter prefers the former to the latter candidate. The goal is to find a permutation that minimizes the sum of the distances to the voters’ permuta-

---

*Email address:* [anke@mpi-inf.mpg.de](mailto:anke@mpi-inf.mpg.de) (Anke van Zuylen).

<sup>1</sup> This work was performed while the author was on sabbatical at TU Berlin, and supported by the Berlin Mathematical School and the Alexander von Humboldt Foundation.

tions, where in principle any distance(-like) function on permutations can be used, e.g. Kendall distance or Footrule distance.

Young & Levenglick [9] show that the Kendall distance is the unique distance function such that the permutation(s) that minimize it have three desirable properties of being neutral, consistent and Condorcet. The latter property means that, if there exists a permutation such that the order of every pair of elements is the order preferred by a majority, then this permutation has minimum distance to the voters' permutations. This distance was already proposed by Kemeny [7] for other reasons ([7] defines axioms on the *distance function*, and finds that the Kendall distance adheres to the axioms), and the problem of finding an optimal ranking with respect to this criterion is now known as Kemeny Rank Aggregation.

In this paper, we suggest a new way of thinking about this problem. Suppose instead of minimizing the total distance from the voters' inputs, we want to find a permutation that makes a majority of the voters "happy"? Of course, a voter is happy when we follow her opinion exactly, and we cannot do this simultaneously for a majority of the voters, unless a majority of the voters is in total agreement. Therefore, our goal is to find a permutation such that there exists no other permutation that a majority of the voters prefer, in the sense that their distance to the alternative permutation is smaller. We call such a permutation a *popular ranking*.

Unfortunately, we show that such a permutation is unlikely to exist: it only exists if Condorcet's paradox does not occur. Even worse than this, we show that if Condorcet's paradox does not occur, then it may still be the case that no popular ranking exists. The only positive news in this context is, perhaps paradoxically, an NP-hardness result: we show that if Condorcet's paradox does not occur, then we can efficiently compute a permutation, which may or may not be popular, but for which the voters will have to solve an NP-hard problem to compute a permutation that a majority of them prefer.

**Related Work:** Our work is inspired by Abraham *et al.* [1] where the notion of *popular matchings* is introduced. Popular ranking is also related to the problem of designing a voting mechanism in which the voters do not have an incentive to lie about their preferences. However, rather than considering deviations of a single voter, a popular solution is robust against deviations of a majority of the voters. We show that, if the input does not contain Condorcet's paradox, then there is a solution that may or may not be popular, but for which it is computationally hard for a majority of the voters to manipulate the output to their advantage. This result has a similar flavor as a result by Bartholdi *et al.* [3], who demonstrate a voting rule for deciding the "winner" of an election, for which it is computationally hard for a single voter to manipulate the output.

## 2 Popular Ranking

We are given a set of alternatives  $[n]$  (where the notation  $[n]$  means  $\{1, 2, \dots, n\}$ ) and a set of voters  $[k]$ , where each voter  $\ell$  has a complete ordering of the alternatives. We will denote these complete orderings of a voter  $\ell$  as a list of the alternatives, where an alternative earlier in the list is preferred to elements that succeed it, and use the notation  $\pi_\ell^{-1} : [n] \rightarrow [n]$  (the use of “ $-1$ ” will become clear shortly), where  $\pi_\ell^{-1}(i)$  is the alternative at position  $i$  in the ordering of voter  $\ell$ . Note that we can interpret  $\pi_\ell^{-1}$  as a permutation. Further, the inverse of  $\pi_\ell^{-1}$ , which we will denote by  $\pi_\ell$ , is well defined and can be interpreted as the position of the alternatives in the list of voter  $\ell$ . We will use  $\text{list}(\pi_\ell)$  to denote the ordered sequence  $(\pi_\ell^{-1}(1), \pi_\ell^{-1}(2), \dots, \pi_\ell^{-1}(n))$ .

The Kendall distance between two permutations  $\pi, \sigma$ , denoted by  $K(\pi, \sigma)$ , is defined as the number of pairwise disagreements of  $\pi$  and  $\sigma$ , i.e.  $K(\pi, \sigma) = \#\{i, j : \pi(i) < \pi(j) \text{ and } \sigma(i) > \sigma(j)\} + \#\{i, j : \pi(i) > \pi(j) \text{ and } \sigma(i) < \sigma(j)\}$ .

**Definition 1** We say a permutation  $\pi$  is popular, if  $\nexists \pi'$  such that  $K(\pi_\ell, \pi') < K(\pi_\ell, \pi)$  for a strict majority of the voters  $\ell \in [k]$ .

We define the *majority graph*  $G = (V, A)$  for an instance as the directed graph which has a vertex for every  $i \in [n]$  and an arc  $(i, j)$  if a majority of the voters  $\ell \in [k]$  has  $\pi_\ell(i) < \pi_\ell(j)$ . Condorcet observed that such a graph may have a cycle; this is known as “Condorcet’s paradox”.

**Lemma 2** No popular ranking exists if the majority graph has a directed cycle.

**Proof** (sketch) If we order the elements from left to right according to a ranking  $\pi$ , then there must be some arc  $(i, j)$  in the graph that is a *back arc*, i.e. for which  $\pi(j) < \pi(i)$ . Let  $\pi'$  be the permutation we obtain by swapping  $i$  and  $j$ , i.e.  $\pi'(i) = \pi(j)$ ,  $\pi'(j) = \pi(i)$  and  $\pi'(t) = \pi(t)$  for all  $t \neq i, j$ . Then one can show that a strict majority of the voters prefer  $\pi'$  to  $\pi$ , namely the voters  $\ell$  who have  $\pi_\ell(i) < \pi_\ell(j)$ .  $\square$

If the majority graph is acyclic, then a popular ranking could exist. We consider the case when the majority graph is a tournament, i.e. for every  $i, j$  exactly one of the arcs  $(i, j)$  and  $(j, i)$  is in  $G$ . Note that the majority graph is always a tournament if the number of voters is odd. By Lemma 2, the only permutation that *could* be popular is the permutation we obtain by topologically sorting the majority tournament. However, it is not the case that this ranking is always a popular ranking, as we show in the Appendix. Even though the topological sort of the majority tournament is not necessarily a popular ranking, it turns out that it is a “good” permutation in the sense that it is NP-hard to find a ranking that a majority of the voters prefer. The proof of

the following theorem is given in the Appendix.

**Theorem 3** *Given an input to the popular rank aggregation problem with an acyclic majority graph, it is NP-hard to find a ranking  $\rho$  that a majority of the voters  $S$  prefers to a topological sort of the majority graph, even if  $S$  is given.*

### 3 Directions

We have seen that a popular ranking does not always exist, even if the majority graph has no cycles. Perhaps popularity is asking for too much and we should relax our objective. It is an interesting question whether there exists a suitable relaxation of the notion of popularity, so that one can get positive results. One way of relaxing the notion is looking for rankings with *least-unpopularity-factor* (McCutchen [8] introduced this notion for matchings). The bad news is that it can be shown that the unpopularity factor of the permutation  $\pi$  we obtain by topologically sorting the majority tournament may be unbounded. It is an open question however whether there exists a permutation with bounded unpopularity (and if so, what this uniform bound is) and whether such a permutation can be found in polynomial time.

#### *Acknowledgements*

The first author thanks Chien-Chung Huang for helpful discussions.

### References

- [1] D. J. Abraham, R. W. Irving, T. Kavitha, and K. Mehlhorn. Popular matchings. *SIAM J. Comput.*, 37(4):1030–1045, 2007.
- [2] K. J. Arrow. *Social choice and individual values*. Yale University Press, 1951.
- [3] J. Bartholdi, C. Tovey, and M. Trick. The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6:227–241, 1989.
- [4] J. Borda. *Memoire sur les elections au scrutin*. Histoire de l’Academie Royal des Sciences, 1781.
- [5] M. Condorcet. *Sur l’application de l’analyse à la probabilité des décisions rendues à la pluralité des voix*. L’Imprimerie Royale, Paris, 1785.
- [6] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *WWW 2001*, pages 613–622, 2001.
- [7] J. Kemeny. Mathematics without numbers. *Daedalus*, 88:575–591, 1959.

- [8] R. McCutchen. The least-unpopularity-factor and least-unpopularity-margin criteria for matching problems with one-sided preferences. In *LATIN'08*, pages 593–604, Búzios, Brazil, 2008.
- [9] H. P. Young and A. Levenglick. A consistent extension of Condorcet’s election principle. *SIAM Journal on Applied Mathematics*, 35(2):285–300, 1978.

## A Deferred Proofs

**Lemma 4** *If the majority graph is an acyclic tournament, then the permutation  $\pi$  we obtain by topologically sorting the majority tournament is not necessarily a popular ranking.*

**Proof** Consider an instance with  $2r$  elements, with  $r$  odd. Let  $\pi$  be the identity. We define a “swap” on  $\pi$  as the operation that swaps the position of element  $i$  and  $i + 1$ , for  $i$  odd. We have  $(r + 1)m + 1$  voters, which are divided into  $r + 1$  sets, one of size  $m + 1$  and the others of size  $m$ . The voters in the set of size  $m + 1$  vote  $\pi$ . All other sets choose  $\frac{1}{2}(r + 1)$  of the swaps and vote for the permutation that we obtain from  $\pi$  by executing these swaps. Note that it is possible to choose the set of swaps for each set, so that each swap is chosen by  $\frac{1}{2}(r + 1)$  sets. Hence the number of voters who prefer each swap is  $\frac{1}{2}(r + 1)m < \frac{1}{2}((r + 1)m + 1)$  so the majority ranking is indeed  $\pi$ . Let  $\sigma$  be the permutation we obtain by executing all swaps on  $\pi$ . All voters except for those in the first set strictly prefer  $\sigma$  to  $\pi$ . Hence  $\pi$  is not a popular ranking.  $\square$

**Theorem 3** *Given an input to the popular rank aggregation problem for which the majority graph is acyclic, it is NP-hard to find a ranking  $\rho$  that a majority of the voters  $S \subseteq [k]$  prefers to a topological sort of the majority graph, even if we know  $S$  (and there are only 7 rankings).*

**Proof** We modify a reduction from the Feedback Arc Set problem to Kemeny rank aggregation by Dwork *et al.* [6] to give an instance to the popular rank aggregation problem in which the majority tournament is acyclic. We consider a Feedback Arc Set solution  $A'$  that is defined by an ordering  $\sigma$  of the vertices (we will make this precise in the next paragraph). There will be 3 voters whose input ranking is  $\sigma$ , and 4 voters with different input rankings.  $\sigma$  will also be the topological order of the acyclic majority tournament. We will show that the ranking  $\sigma$  is not popular (i.e. there exists a ranking  $\rho$  that is preferred by 4 voters to  $\sigma$ ) if and only if  $A'$  is not an optimal Feedback Arc Set solution.

In the Feedback Arc Set problem, we are given a directed graph  $G = (V, A)$  and are asked to produce a subset of the edges  $A'$  of minimum size, such that  $(V, A \setminus A')$  is acyclic. Given an input to the Feedback Arc Set problem, and an

ordering  $\sigma$  of the vertices, let  $B(\sigma)$  be the set of arcs  $(i, j)$  in  $A$  such that  $i$  is ordered to the right of  $j$ . Clearly  $B(\sigma)$  is a Feedback Arc Set solution of size  $|B(\sigma)|$ , and this is a minimum cardinality solution that corresponds to the ordering  $\sigma$ . Conversely, every Feedback Arc Set solution gives rise to at least one ordering, so for the Feedback Arc Set problem we may restrict our attention to solutions that arise from considering orderings of the vertices  $\sigma$  (i.e., of the form  $B(\sigma)$ ).

Given such an ordering, it is NP-hard to find an ordering that corresponds to a FAS of size  $|B(\sigma)| - 1$  or less if it exists. If we could find such an ordering, we have the following polynomial time algorithm to find a FAS of minimal size: start with a random ordering of the nodes, and repeat this supposed procedure to find a FAS of smaller size (we need to call the procedure at most  $O(|A|)$  times to find a minimum size FAS).

Based on an input to the Feedback Arc Set problem  $G = (V, A)$ , we begin by constructing four lists as in Dwork et al. [6]. Let  $|V| = n, |A| = m$ . We have an element for every vertex and every arc in  $G$ . We number the vertices (arbitrarily)  $1, \dots, n$ . We also fix an ordering of the arcs in  $G$ .

For a vertex  $i \in V$ , we let  $out(i)$  be the list of arcs in  $A$  that have  $i$  as tail, listed in order. We let  $out^r(i)$  be the list of the same arcs, but listed in the reverse order. Similarly define  $in(i)$  and  $in^r(i)$ .

Dwork et al. [6] now define the following four lists:

$$\begin{aligned}\tau_1(G) &= 1, out(1), 2, out(2), \dots, n, out(n) \\ \tau_2(G) &= n, out^r(n), n-1, out^r(n-1), \dots, 1, out^r(1) \\ \tau_3(G) &= in(1), 1, in(2), 2, \dots, in(n), n \\ \tau_4(G) &= in^r(n), n, in^r(n-1), n-1, \dots, in^r(1), 1\end{aligned}$$

Note that for every pair of elements  $a, b$ , except for a pair where  $a$  is a vertex  $i$  and  $b$  is the arc  $(i, j)$  or  $(j, i)$ , it is the case that exactly one of  $\tau_1$  and  $\tau_2$  orders  $a$  before  $b$ . Similarly, exactly one of  $\tau_3$  and  $\tau_4$  orders  $a$  before  $b$  (call this Fact 1).

For a vertex  $i$  and an arc  $(i, j)$ , three of the lists order  $i$  before  $(i, j)$ , and similarly, for a vertex  $i$  and an arc  $(j, i)$ , three of the lists order  $(j, i)$  before  $i$ .

This implies in particular, that for every pair of elements  $a$  and  $b$  there exists at least one list that orders  $a$  before  $b$ , and at least one that orders  $b$  before  $a$  (call this Fact 2).

We associate the following list with an ordering of the nodes  $\sigma$ :

$$\ell(\sigma) = \sigma(1), out(\sigma(1)), \sigma(2), out(\sigma(2)), \dots, \sigma(n), out(\sigma(n)).$$

Note that in  $\ell(\sigma)$  the edges always are directly behind their tail nodes. So  $d(\ell(\sigma), \tau_1) + d(\ell(\sigma), \tau_2) = \binom{n}{2} + \binom{m}{2} + (n-2)m + m$ , where  $d$  is the Kendall-tau distance. The first three terms come from Fact 1, and the last term comes from noting that the pair  $i, (i, j)$  is in the same order in  $\tau_1, \tau_2$  and  $\ell(\sigma)$  (hence contributing 0), and that  $(i, j), j$  is in different order in  $\tau_1$  and  $\tau_2$  (hence contributing 1).

The distance of  $\ell(\sigma)$  to the other two lists is directly related to  $|B(\sigma)|$ , the number of “back arcs” in the FAS solution  $\sigma$ : for every arc  $(i, j) \in B(\sigma)$  we know that  $i$  is ordered to the right of  $j$ , and so  $(i, j)$  is also to the right of  $j$  in  $\ell(\sigma)$ . It is to the left of  $j$  in both  $\tau_3$  and  $\tau_4$ , though. For every arc  $(i, j) \notin B(\sigma)$  we see that  $(i, j)$  is to the left of  $j$  in  $\ell(\sigma)$ , exactly as in both  $\tau_3$  and  $\tau_4$ .

We thus obtain the following equality:

$$\begin{aligned} & d(\tau_1, \ell(\sigma)) + d(\tau_2, \ell(\sigma)) + d(\tau_3, \ell(\sigma)) + d(\tau_4, \ell(\sigma)) \\ &= 2 \binom{n}{2} + 2 \binom{m}{2} + 2(n-2)m + m + 2|B(\sigma)|. \end{aligned} \quad (\text{A.1})$$

Note that any ordering  $\rho$  which has the nodes in the same relative order, will have  $\sum_i d(\tau_i, \rho) \geq \sum_i d(\tau_i, \ell(\sigma))$ . This is because  $\ell(\sigma)$  does not incur any additional “cost” for arcs not in  $B(\sigma)$ , and for an arc in  $B(\sigma)$  it is either before the tail node or after the head node (or both) in any ranking, thus incurring a “cost” of at least 2 (call this Fact 3.)

We now define the input to the popular rank aggregation problem. Given an input to the FAS problem  $G$ , and an ordering  $\sigma$  of the vertices of  $G$ , we first create four copies of  $G$ , call them  $G_1, G_2, G_3, G_4$ . We use the notation  $\sigma(G_i)$  for the ordering of the nodes in  $G_i$  that is an image of the ordering  $\sigma$  of the nodes of  $G$ , for  $i = 1, 2, 3, 4$ .

We now have four voters who have the following lists:

$$\begin{aligned} \pi_1 &= \tau_1(G_1), \tau_2(G_2), \tau_3(G_3), \tau_4(G_4) \\ \pi_2 &= \tau_2(G_1), \tau_3(G_2), \tau_4(G_3), \tau_1(G_4) \\ \pi_3 &= \tau_3(G_1), \tau_4(G_2), \tau_1(G_3), \tau_2(G_4) \\ \pi_4 &= \tau_4(G_1), \tau_1(G_2), \tau_2(G_3), \tau_3(G_4) \end{aligned}$$

Finally we have three voters whose lists are given by

$$L(\sigma) := [\ell(\sigma(G_1)), \ell(\sigma(G_2)), \ell(\sigma(G_3)), \ell(\sigma(G_4))].$$

Note that for every two elements  $a$  and  $b$  from the same graph, at least one of  $\pi_1, \pi_2, \pi_3, \pi_4$  lists  $a$  before  $b$  and at least one lists  $b$  before  $a$  by Fact 2. If  $a$  and  $b$  are from different graphs, then every list has the element from the graph with the lower index first. So the majority tournament for this rank

aggregation problem is exactly the list of the three voters,  $L(\sigma)$ , and this is an acyclic tournament.

We will now argue that there exists a ranking that is preferred over  $L(\sigma)$  by a majority of the voters, precisely when there exists an ordering of the nodes of  $G$  that has a strictly smaller FAS than  $\sigma$ .

Suppose we there exists an ordering of  $G$  with a strictly smaller FAS than  $\sigma$ , say  $\rho$ . From Equation (A.1) it follows immediately that  $d(\pi_i, L(\rho)) < d(\pi_i, L(\sigma))$  for all  $i$ , so  $L(\rho)$  is preferred over  $L(\sigma)$  by a majority of the voters.

Coversely, suppose we know that there exists a ranking  $R$  that is preferred over  $L(\sigma)$  by a majority of the voters. These voters must be the voters with lists  $\pi_1, \pi_2, \pi_3$  and  $\pi_4$ , since  $d(L(\sigma), R) > 0 = d(L(\sigma), L(\sigma))$  for  $R \neq L(\sigma)$ . Call  $\rho_i$  the ordering induced by  $R$  on the nodes of  $G_i$ , for  $i = 1, 2, 3, 4$ .

Let  $T = \ell(\rho_1), \ell(\rho_2), \ell(\rho_3), \ell(\rho_4)$ . We have that  $\sum_i d(T, \pi_i) \leq \sum_i d(R, \pi_i) < \sum_i d(L(\sigma), \pi_i)$ , by Fact 3.

Let  $T_i = [\rho_i(G_1), \rho_i(G_2), \rho_i(G_3), \rho_i(G_4)]$ . Now note that  $\sum_j \sum_i d(T_j, \pi_i) = 4 \sum_i d(T, \pi_i) < 4 \sum_i d(L(\sigma), \pi_i)$ . So the property of averages gives that for some  $j$  we have  $\sum_i d(T_j, \pi_i) \leq \sum_i d(T, \pi_i) < \sum_i d(L(\sigma), \pi_i)$ . This means that  $\rho_j$  is an ordering that has a strictly smaller FAS than  $\sigma$  (by Fact 3).  $\square$