Noname manuscript No.
(will be inserted by the editor)

Egalitarianism in the rank aggregation problem: a new dimension for democracy

The final publication is available at Springer via http://dx.doi.org/[10.1007/s11135-015-0197-x

Received: date / Accepted: date

Abstract Winner selection by majority, in elections between two candidates, is the only rule compatible with democratic principles. Instead, when candidates are three or more and voters rank candidates in order of preference, there are no univocal criteria for the selection of the winning (consensus) ranking and the outcome is known to depend sensibly on the adopted rule. Building upon XVIII century Condorcet theory, whose idea was maximising total voter satisfaction, we propose here a new basic principle (dimension) to guide the selection: satisfaction should be distributed among voters as equally as possible. With this new criterion we identify an optimal set of rankings, ranging from the Condorcet solution to the the most egalitarian one with respect to the voters. Most importantly, we show that highly egalitarian rankings are much more robust, with respect to random fluctuations in the votes, than consensus rankings returned by classical voting rules (Copeland, Tideman, Schulze). The newly introduced dimension provides, when used together with that of Condorcet, a more informative classification of all the possible rankings. By increasing awareness in selecting a consensus ranking our method may lead to social choices which are more egalitarian compared to those achieved by presently available voting systems.

Keywords Preferential voting · Rank aggregation · Pareto frontier · Variance minimization

1 Introduction

A voting process starts with individuals giving a formal indication of a choice (*ballot*) or, more generally, a set of preferences between two or more candidates

This work has received financial support from the Italian Research Ministry through the FIRB projects No. RBFR086NN1 and RBFR10N90W and PRIN project No. 2010HXAW77.

(or alternatives). The process ends with an aggregation procedure (*winner selection method*) of these indications, in order to produce the *consensus* ranking, that is the ranking on which voters should agree more upon and which should be the output of the election. The complexity of the selection process comes, in general, from the presence of competing interests and conflicting opinions which make it impossible to satisfy all the preferences expressed by the voters. With his seminal work on voting theory, Condorcet [17] discovered that the majority rule, applied to pairwise preferences, may lead to invalid solutions. For instance in an election among three candidates the preferences may sum up to prefer the first to the second, the second to the third and the third to the first. Similarly, from the formal logic perspective, Arrow's theorem [2] states that a perfectly fair voting system may not exist (see also Easley and Kleinberg[7]). The lack of an ideal voting system when there are three or more candidates implies that any winner selection procedure contains some kind of arbitrariness and makes the studies on voting methods an interesting research problem.

Typical examples of voting processes are political elections [4]. In that case the need of a single winner, or a single winning ranking, has encouraged the use of very elementary selection rules, easy to compute and to understand by voters and competitors at the expense of making sub-optimal choices. Voting theory include also cases beyond political matters. Survey rankings for instance, typically made for commercial purposes, like hotel listings, movie rankings, best product on the market etc, are selected with totally different criteria. The choice of the ten best smartphones, say, is not made by maximising the voter total satisfaction, but rather to ensure that each customer finds, among those ten, a satisfactory model.

A similar problem is very much studied in computer science under the name of Rank Aggregation: a typical example is the merging of webpage rankings produced by different search engines or obtained according to different criteria [8]. The main difference from the examples above is that here the number of voters (engines/criteria) is small, while the number of candidates (webpages) is large. This is why in that field of research the focus is more on the algorithmic challenge of computing the consensus ranking efficiently. Here we are more interested in presenting the new criterion for better selecting the consensus ranking; thus we concentrate on small number of candidates, so that all possible rankings (with ties) can be easily computed.

It is therefore clear that the problem of finding a good consensus ranking is an interdisciplinary topic of research: it is inspired and guided by studies in sociology, marketing, economy and political sciences. The disciplines technically involved in the solutions are statistics, mathematics and computer science.

While the social choice theory is mostly interested in classifying voting methods according to some well established principles (axioms) [24] our approach here is different: we introduce a new measurable quantity to investigate the properties of the whole ballot system that may, eventually, lead to more informed choices.

The scholars working hard science research may realise that throughout the paper we make use of ideas and methods, albeit at a very elementary level and embryonal form, that come from statistical physics. It is worth mentioning that there is a quite large amount of statistical physics studies on opinion dynamics [5] where the main focus is to understand how opinions rise and propagate among voters. Moreover attempts to quantitative modelling of social choice problems have naturally led to consider disordered models such as the random field Ising model [22] and the spin glass model [20]. However the subject of the present work, that is how to extract an informative and robust consensus ranking from a set of ranked-ballots, has not received, to the best of our knowledge, an adequate attention in physics and mathematics. We believe that statistical mechanics techniques will shed some light on the complexity of social choices and we hope that the present work will stimulate further research on this fundamental problem.

2 Results

2.1 Definition of the problem.

Each of n voters expresses a preference about m candidates by sorting them in a ranked list, possibly with ties, resulting in n ballots. For the sake of simplicity, we prefer not to discuss partial rankings, because the meaning of not ranking a candidate may change a lot from application to application. Valid ranked lists for $m = 4$ candidates are for instance $\mathsf{B}>\mathsf{A}>\mathsf{D}>\mathsf{C}$, $\mathsf{D}>\mathsf{A}=\mathsf{C}>\mathsf{B}$ and $\mathsf{C}>\mathsf{A}=\mathsf{B}=\mathsf{D}$. We call $\mathbf{r}_v$ the ballot of voter v. Each voter wishes the consensus ranking to be as close as possible to his ballot and, following Condorcet [17], a good winner selection method should work by maximising the total sum of those wishes, i.e. minimising the sum of the distances between the consensus ranking and the ballots (this method is also known in the literature as the Kemeny rule or median procedure). Therefore the search for a consensus ranking needs to be based on a notion of distance between the rankings. There are several definitions of distance between rankings and many studies on the relations among them [9]. Among these, the Kemeny distance d_{Kem} [15,16] is widely used due to its robust properties [14]. Intuitively, when restricted to rankings without ties, d_{Kem} is twice the minimum number of swaps of nearby candidates required to transform one ranking into another. Alternatively, it counts the number or pairwise preferences that do not match in the two rankings. When ties appear, these count $\frac{1}{2}$ in the distance, if they do not match between the two rankings. A more formal definition is included in the Methods section. Our approach applies regardless of the type of distance used. We will conventionally use the Kemeny distance to develop the discussion in the next sections (see the Supplementary Information for a discussion on other distances and also Ref. [25] and references therein).

The Condorcet consensus ranking $\mathbf{c}^*$ has been defined as the ranking (or more properly the ranking*s*) minimising the function

$$\mu(\mathbf{c}) = \frac{1}{n} \sum_{v=1}^{n} d_{\mathrm{Kem}}(\mathbf{r}_v, \mathbf{c}) \,, \tag{1}$$

in formulae, $\mathbf{c}^* = \mathrm{argmin}\, \mu(\mathbf{c})$ (see Young [26] for a review on Condorcet theory and Monjardet[19] for a complete elucidation of its mathematical significance). The Condorcet consensus ranking is not to be confused with the notion of Condorcet candidate: the candidate which wins in pairwise comparison with all other candidates. The computation of $\mathbf{c}^*$ is in general a NP-hard problem, since the space of all possible rankings with ties grows faster than $m!$. In practice several polynomial time algorithms have been developed that return an approximated answer to the problem of selecting a consensus ranking. Most of these are the voting rules used in everyday applications. Among them it is worth recalling the Pairwise comparison (or Copeland), Schulze and Tideman methods, which are perhaps the most used single-round ranked-ballot winner selection methods (they are all described in the Methods section).

None of the above voting methods is perfectly fair (in the sense of Arrow's theorem), however they all return a "reasonable" consensus ranking, and this is why they are used in practical applications. Nonetheless some problems and inconsistencies remain unsolved: (i) different voting methods return different consensus rankings and selecting the 'best' among them is difficult (this is the well known problem that the outcome of an election may very well depend on the electoral system); (ii) by returning a *unique* consensus ranking, a lot of information about voter preferences is lost; (iii) often there are consensus rankings with a value of $\mu(\mathbf{c})$ very close to the optimal $\mu(\mathbf{c}^*)$, and it is unclear why they should be discarded. It is worth noting that, in an election/survey with n voters, fluctuations of $O(1/\sqrt{n})$ in $\mu(\mathbf{c})$ are somehow unavoidable: if $\mu(\mathbf{c}_1) < \mu(\mathbf{c}_2)$, but with $\mu(\mathbf{c}_2) - \mu(\mathbf{c}_1) \sim 1/\sqrt{n}$, then choosing $\mathbf{c}_1$ as the consensus ranking instead of $\mathbf{c}_2$ is equivalent to taking a decision based on the toss of a coin.

2.2 A new dimension for choosing the consensus ranking.

In order to solve the above problems we suggest to consider as valid consensus rankings all the rankings $\mathbf{c}$ close enough to the optimal one (i.e., those for which $\mu(\mathbf{c}) - \mu(\mathbf{c}^*) \sim 1/\sqrt{n}$), and we introduce a new *dimension* to select the best among these valid consensus rankings. Our idea is that not only the global number of satisfied preferences is to be maximized, but also each individual voter should have more or less the same number of satisfied preferences. With this aim we propose to compute also the voter-to-voter satisfaction variability (standard deviation) as

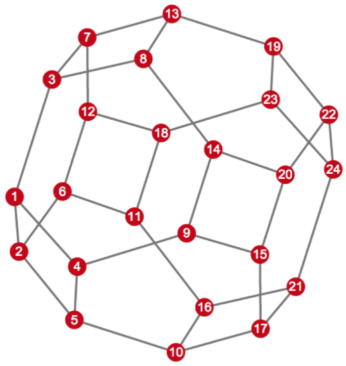
$$\sigma(\mathbf{c}) = \sqrt{\frac{1}{n}\sum_{v=1}^{n}\left[d_{\text{Kem}}(\mathbf{r}_v,\mathbf{c}) - \mu(\mathbf{c})\right]^2} \tag{2}$$
$$= \sqrt{\frac{1}{n}\sum_{v=1}^{n} d_{\text{Kem}}(\mathbf{r}_v,\mathbf{c})^2 - \mu(\mathbf{c})^2}\ .$$

If $\sigma(\mathbf{c}) = 0$, the consensus ranking $\mathbf{c}$ satisfies equally each voter; while, if $\sigma(\mathbf{c})$ is large, then there are voters more satisfied and others less satisfied than the average. Clearly the smaller is $\sigma(\mathbf{c})$ the more egalitarian is $\mathbf{c}$.

To illustrate the new criterion, we start with a very simple example. We consider an election with $m = 4$ candidates and we do not allow for ties; the number of possible rankings is $m! = 24$, as shown in the table included in Figure 1. The distance between these 24 rankings can be easily visualised in the same figure, top left panel, which includes a graph where each vertex corresponds to a ranking, with an edge connecting rankings at distance 2 (differing only by a swap of two neighbouring candidates). For rankings at distance larger than 2, it is enough to count the edges along the shortest path connecting the rankings in this graph.

Suppose the electorate is equally polarised on two opposite rankings: half of the voters rank candidates $\mathsf{A} > \mathsf{B} > \mathsf{C} > \mathsf{D}$ and the other half $\mathsf{D} > \mathsf{C} > \mathsf{B} > \mathsf{A}$. A simple calculation shows that any possible ranking has $\mu(\mathbf{c}) = 6$, therefore there is no way to choose one of them according to the Condorcet criterion alone. However the 24 possible consensus rankings have very different $\sigma(\mathbf{c})$ as can be seen in the middle panel of Figure 1: the point with largest $\sigma(\mathbf{c})$ corresponds to rankings $\mathsf{A} > \mathsf{B} > \mathsf{C} > \mathsf{D}$ and $\mathsf{D} > \mathsf{C} > \mathsf{B} > \mathsf{A}$ that fully satisfy half of the voters and fully deceive the second half, while the point with $\sigma(\mathbf{c}) = 0$ corresponds to the six rankings that are at the same distance from the ballots, thus satisfy them equally well. It is clear that the latter are the more egalitarian consensus rankings. In other words, spreading satisfaction as equally as possible among voters, i.e. minimising $\sigma(\mathbf{c})$, is a new criterion to select the consensus ranking that deserves, at least, the same consideration as the Condorcet criterion of minimising $\mu(\mathbf{c})$.

Even more interesting is the case when some noise is added to the example above. For instance we can consider small fluctuations in the number of electors participating to the poll, resulting in a fraction $\frac{1}{2} + \epsilon$ of voters ranking the candidates as $\mathsf{A} > \mathsf{B} > \mathsf{C} > \mathsf{D}$ and the complement fraction $\frac{1}{2} - \epsilon$ ranking them as $\mathsf{D} > \mathsf{C} > \mathsf{B} > \mathsf{A}$. For an election with n voters a noise of order $\epsilon = O(1/\sqrt{n})$ is somehow unavoidable. In lower panels in Figure 1 we report $\mu(\mathbf{c})$ and $\sigma(\mathbf{c})$ values for the 24 possible consensus rankings. For $\epsilon > 0$, the small unbalance decreases μ for ranking $\mathsf{A} > \mathsf{B} > \mathsf{C} > \mathsf{D}$, making it the consensus ranking under the Condorcet criterion. For $\epsilon < 0$, the opposite ranking would win. The difference between the two cases is, however, only due to noise; so selecting a consensus ranking by strictly minimising $\mu(\mathbf{c})$ would be equivalent to selecting



Code	Ranking	Code	Ranking
1	A>B>C>D	13	C>A>D>B
2	A>B>D>C	14	C>B>A>D
3	A>C>B>D	15	B>C>D>A
4	B>A>C>D	16	D>B>A>C
5	B>A>D>C	17	B>D>C>A
6	A>D>B>C	18	D>A>C>B
7	A>C>D>B	19	C>D>A>B
8	C>A>B>D	20	C>B>D>A
9	B>C>A>D	21	D>B>C>A
10	B>D>A>C	22	D>C>A>B
11	D>A>B>C	23	C>D>B>A
12	A>D>C>B	24	D>C>B>A

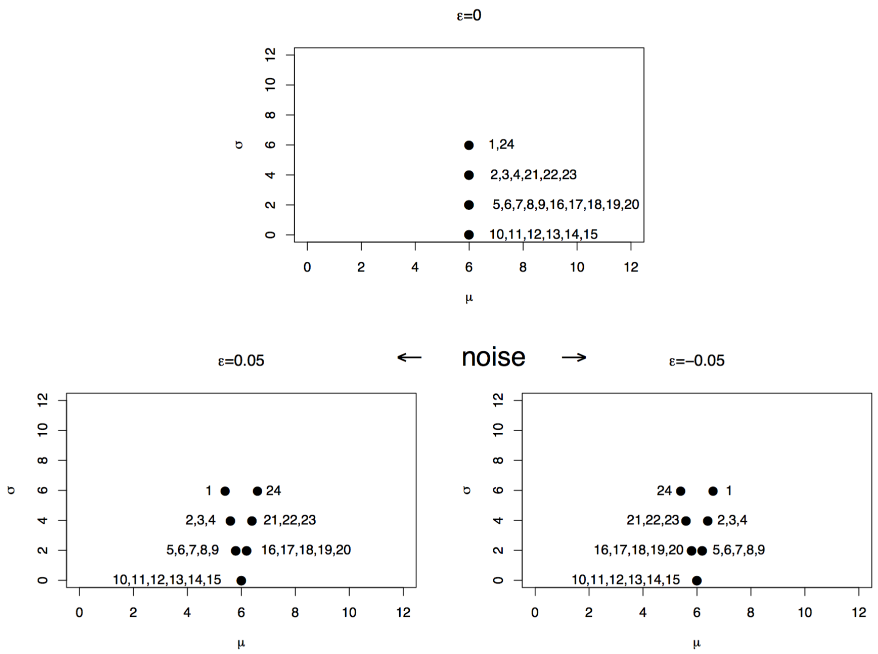


Fig. 1 A very simple example with four candidates. There are 24 possible rankings without ties (listed) in panel (b)), whose relative distances are given by the graph in panel (a). In panel (c) we report (μ, σ) values for possible consensus rankings in case the electorate is equally polarised on opposite ballots (codes 1 and 24); panels (d) and (e) have been computed by adding a small noise to the perfectly balanced situation. Our web platform allows one to interact with panels c, d and e.

the winner on a coin toss. Rankings with lower $\sigma(\mathbf{c})$, as lower panels in Figure 1 shows, are much less sensitive to noise: by minimising $\sigma(\mathbf{c})$ one gets always the same consensus rankings independently on the noise. This is a very important observation in favour of the new criterion, given that a fair voting system should be robust with respect to small fluctuations due to the unavoidable noise.

Although very simplified, the example above contains in a stylized form the relevant facts we have observed in real data, to be discussed below.

2.3 Analysis of data from real polls.

We have now two criteria for the identification of the best consensus ranking: minimising $\mu(\mathbf{c})$ and minimising $\sigma(\mathbf{c})$ (among rankings of small $\mu(\mathbf{c})$). In general is not possible to find a consensus ranking satisfying both criteria, and some compromise must be adopted, as we will exemplify with data from real polls.

The first dataset consists of ratings for jokes from the Jester database[12]. The full dataset is made of 100 jokes rated by 24938 users. Ratings are continuous values between -10 and 10. We have selected the five jokes rated by most users, and considered only those users who rated all five jokes, resulting in 24921 voters. For each voter, the ballot is obtained by ranking the 5 jokes according to the continuous-valued rating.

In the upper panel of Figure 2 we show the $\big(\mu(\mathbf{c}), \sigma(\mathbf{c})\big)$ values for all possible consensus rankings of the $m = 5$ jokes: the 120 black circles correspond to rankings without ties, while gray diamonds are the 421 rankings with ties. One ranking was excluded from the plot, for better visualisation: ranking $\mathsf{A}=\mathsf{B}=\mathsf{C}=\mathsf{D}=\mathsf{E}$ at position (9.81,0.63). The consensus ranking minimising $\mu(\mathbf{c})$ is $\mathbf{c}^*$: $\mathsf{D}=\mathsf{E}>\mathsf{C}>\mathsf{B}>\mathsf{A}$ and has $\mu(\mathbf{c}^*) = 8.615$. However, close to $\mathbf{c}^*$ we see a cloud of points with small values of $\mu(\mathbf{c})$. The lower panel in Figure 2 zooms over this set of rankings, all having a distance from the Condorcet optimum $\mathbf{c}^*$, comparable with $O(1/\sqrt{n})$ fluctuations. So, from the point of view of the Condorcet criterion, all these rankings are equally good within the noise. On the contrary they show a much larger variation in $\sigma(\mathbf{c})$, that changes between 3.36 and 4.29, allowing for a better consensus ranking selection by minimising $\sigma(\mathbf{c})$. The consensus ranking minimising $\sigma(\mathbf{c})$ in this region is $\mathsf{D}=\mathsf{E}>\mathsf{A}=\mathsf{B}=\mathsf{C}$ with coordinates $(\mu, \sigma) = (8.66, 3.36)$. It seems to convey all the relevant information contained in this set of low $\mu(\mathbf{c})$ rankings: indeed the only information shared by all the rankings in the lower panel of Figure 2 is that jokes D and E are better than jokes A, B and C. Any consensus ranking more refined that $\mathsf{D}=\mathsf{E}>\mathsf{A}=\mathsf{B}=\mathsf{C}$ would just amplify the noise, rather than providing further useful information.

Three commonly used winner selection methods were also applied to the data (Copeland, Schulze and Tideman), and the corresponding consensus rankings are marked in Figure 2. All of them rank jokes as $\mathsf{E}>\mathsf{D}>\mathsf{C}>\mathsf{B}>\mathsf{A}$ with $(\mu, \sigma) = (8.62, 4.12)$. This consensus ranking differs from $\mathbf{c}^*$, the Condorcet consensus ranking, and it has a quite large $\sigma(\mathbf{c})$ value, hence being among the less egalitarian rankings.

In applying the criterion of minimising $\sigma(\mathbf{c})$ one has to be careful, because this criterion tends to select consensus rankings with ties (gray diamonds are

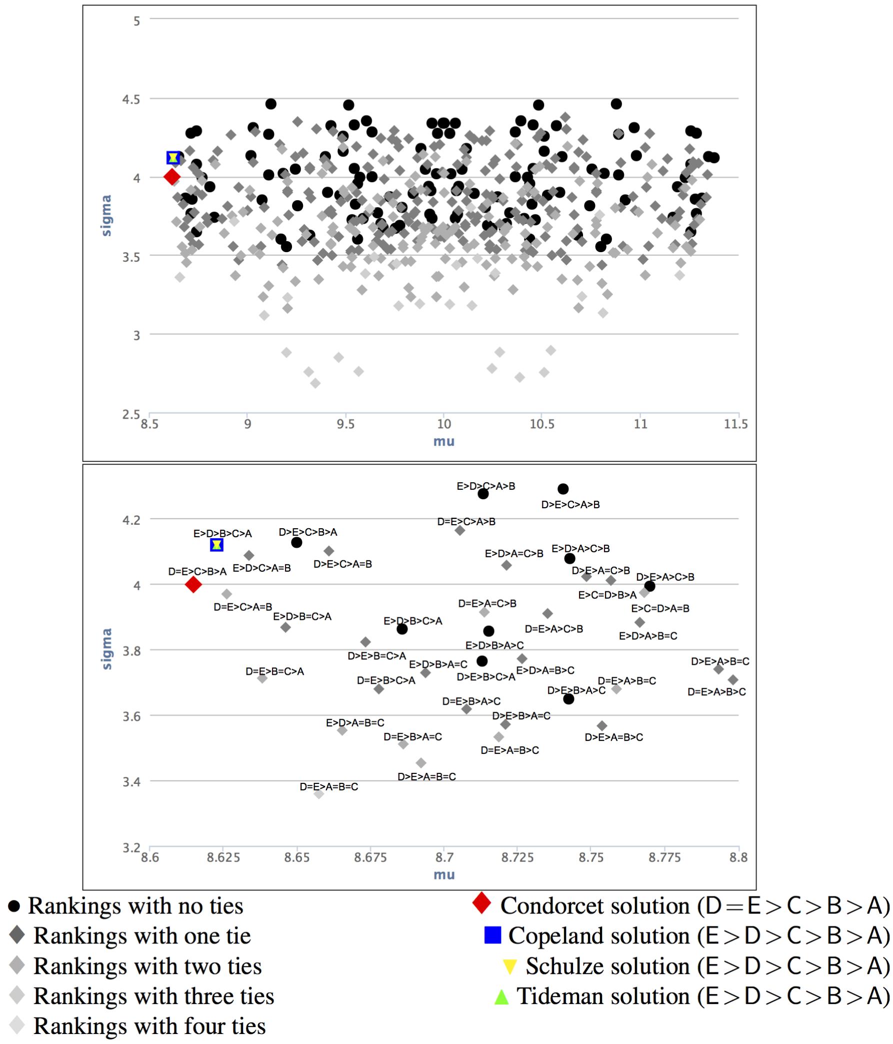


Fig. 2 Aggregation of 24921 ballots rankings 5 jokes. The upper panel shows the entire set of possible rankings, except for A = B = C = D = E, at position (9.81,0.63) which is omitted for visualisation purposes. The lower panel zooms in the leftmost part of the first plot. Solutions found by Copeland, Tideman and Schulze coincide in this example. None of these is optimal under the Condorcet criterion of minimising μ or under the new criterion of minimising σ. Here you may interact with this figure.

on average below black circles in Figure 2). If ties are not allowed in the consensus ranking, one should focus only on black points in Figure 2: even in this case, the consensus ranking D > E > B > A > C with $(\mu, \sigma) = (8.74, 3.65)$ looks much more egalitarian than the consensus ranking E > D > C > B > A found by common voting methods: it gains more than 10% in $\sigma(\mathbf{c})$, while loosing just 1% in $\mu(\mathbf{c})$. The final decision on which rule is to be used to select the consensus ranking is left to the organisers of the poll/survey, but clearly a plot in the (μ, σ) plane is much more informative than any previously available method.

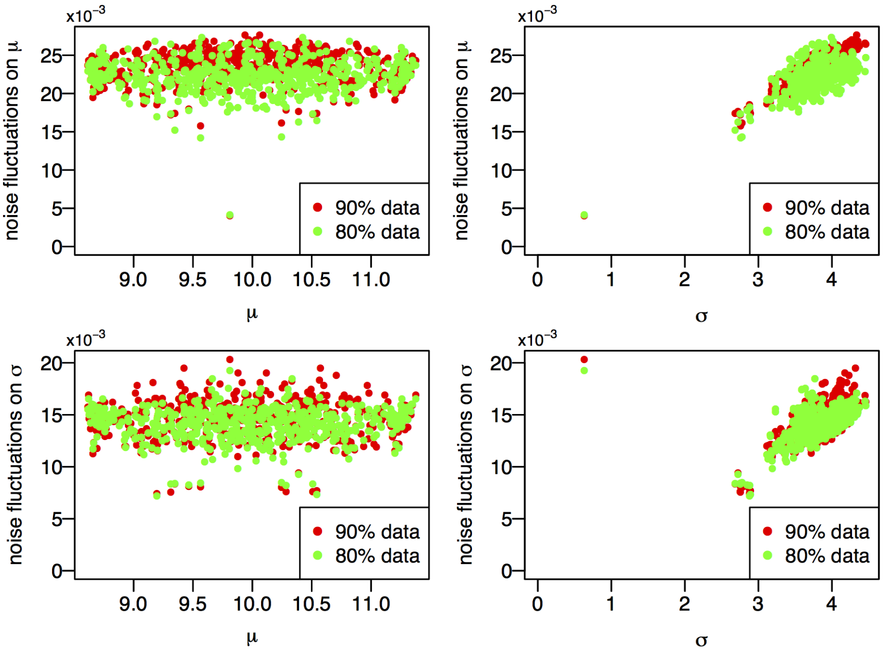


Fig. 3 Uncertainty plots. Uncertainties on the values of $\mu(\mathbf{c})$ and $\sigma(\mathbf{c})$ as obtained from resampling experiments using either 80% or 90% of the original data ($\alpha = 0.2$ and $\alpha = 0.1$ respectively). Rankings with smaller $\sigma(\mathbf{c})$ are more reliable, since they have smaller fluctuations.

Similar to the simple example discussed earlier, the data from real polls also show that consensus rankings of smaller $\sigma(\mathbf{c})$ are less sensitive to noise. In this case we investigate the effect of small fluctuations in participation by using a subsampling procedure: from the joke ratings provided by 24921 users, a fraction α of randomly chosen votes has been removed, and $\mu(\mathbf{c})$ and $\sigma(\mathbf{c})$ recomputed. Resampling was repeated 100 times with $\alpha = 0.1$ and $\alpha = 0.2$. From the variations of μ and σ between different subsamplings we may compute the noise fluctuations on μ and σ (see Supplementary Information for a detailed derivation of the scaling law for fluctuations). In Figure 3 these fluctuations are reported, showing a very clear and strong correlation with the value of $\sigma(\mathbf{c})$. A good consensus ranking should be as robust as possible to noise produced by fluctuations in e.g. the number of voters. For example, suppose a poll/survey is run for 10 days, then the outcome of the survey is reliable if it does not change sensibly in case the data were collected for one or two days less. What we observe in Figure 3 is that noise sensitivity is larger for points of large $\sigma(\mathbf{c})$, while no relation can be observed between noise sensitivity and μ. So, choosing a consensus ranking according to the new criterion of minimising

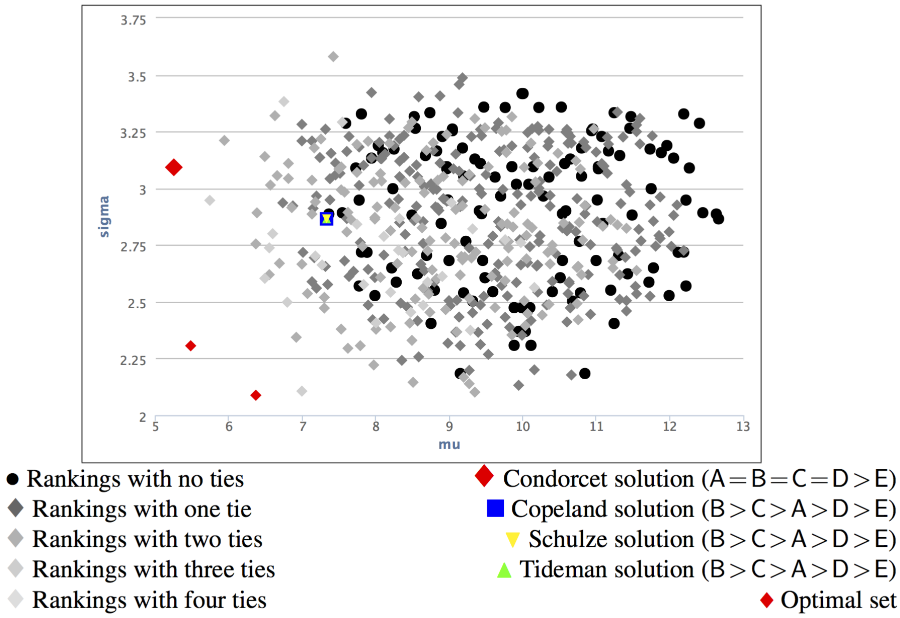


Fig. 4 Aggregation of 930 ballots ranking 5 movies. Here rankings with ties play an important role. Again the solution found by standard voting methods (Copeland, Schulze, Tideman) is far from the optimal set. Here you may interact with this figure.

$\sigma(\mathbf{c})$, provides in general a result much more robust to noise (e.g. unavoidable fluctuations in the number of participants to the poll/survey/election). We also analysed opinion fluctuations for this dataset, similarly to the analytical example of Figure 1, and results show same robustness for rankings with lower $\sigma(\mathbf{c})$ (see Supplementary Information for details).

The second example from real polls considers the rankings of 5 movies provided by 930 users. These are a subset of a larger database consisting of 1,000,209 ratings from 6040 users for 3952 movies [13]. Here, users rated the movies on a discrete scale from 1 to 5. As before, we sorted the movies for each user, to obtain the ballots. Since equal ratings are very probable here, given that only 5 possible rating values exist, many ballots have ties.

Once again the plot in (μ, σ), shown in Figure 4 is very informative. First of all we notice that consensus rankings with ties, although having much smaller values of $\mu(\mathbf{c})$, are not chosen by any commonly used voting method. Moreover the optimal consensus ranking according to the Condorcet criterion seems to have a very large value for $\sigma(\mathbf{c})$. There are a few other rankings worth to be considered, that have slightly higher $\mu(\mathbf{c})$ but much lower $\sigma(\mathbf{c})$. Indeed we identify a *set of optimal rankings* (red diamonds in Figure 4) combining the two criteria. These optimal rankings start from the Condorcet ranking $\mathbf{c}^*$ and include other rankings in the bottom left part of the plot, that cannot be improved in terms of both $\mu(\mathbf{c})$ and $\sigma(\mathbf{c})$ (the Methods section includes a more formal definition of this sequence). In the example from the movie data, two additional rankings should be considered, along with the Condorcet ranking, to be part of the optimal set: these are $\mathsf{A}=\mathsf{B}=\mathsf{C}=\mathsf{D}=\mathsf{E}$ with $(\mu, \sigma) = (5.48, 2.30)$, and $\mathsf{B}>\mathsf{A}=\mathsf{C}=\mathsf{D}=\mathsf{E}$ with $(\mu, \sigma) = (6.37, 2.08)$, both are red-marked in the bottom left corner of the plot. We suggest that the consensus ranking should be selected from the optimal set, and the choice should be made after careful analysis of the (μ, σ) plot. The set of optimal rankings resembles somehow the Pareto efficient frontier used in economic theory [10].

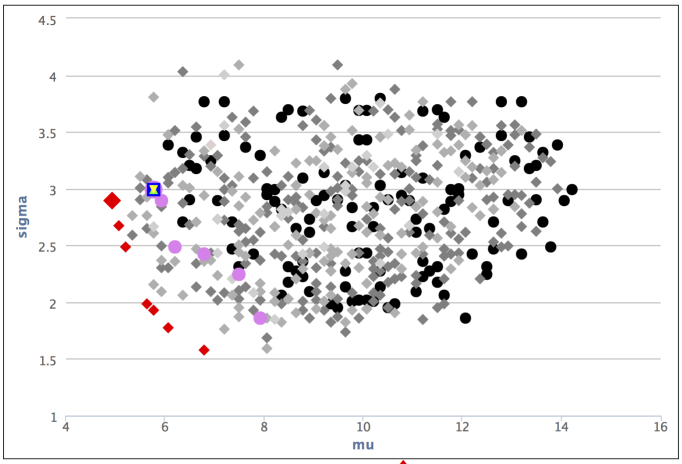


Fig. 5 **AIRESIS data.** Here you may interact with this figure.

Both examples above have a large number of voters and one may think that the complex behaviour we have illustrated could be due to the large number of voters. This is not actually the case, as we are going to show with an example from a poll with a small number of voters ($n = 14$), that ranked $m = 5$ alternatives. This is a poll organised on the Airesis platform [1], which is a web platform freely available to organisations to manage internal decision making. The data shown in Figure 5 represent a real election where the consensus ranking has been selected according to the Schulze method. The first evidence is that the consensus ranking of that election (Schulze) is far from the optimal one: the Condorcet optimal consensus ranking is better (i.e. lower) both in $\mu(\mathbf{c})$ and $\sigma(\mathbf{c})$. Additionally, a large number of rankings are part of the optimal set, defined previously, and marked with red diamonds in the plot, which should be taken into consideration. Even willing to restrict to consensus rankings without ties (this is an election, and ties may be problematic for the decision process), it is clear that the consensus ranking selected by Schulze, $\mathsf{C}>\mathsf{E}>\mathsf{D}>\mathsf{B}>\mathsf{A}$ with $(\mu,\sigma)=(5.78,2.99)$, has a quite large $\sigma(\mathbf{c})$ with respect to consensus rankings $\mathsf{C}>\mathsf{D}>\mathsf{E}>\mathsf{B}>\mathsf{A}$ with $(\mu,\sigma)=(5.92,2.89)$, and $\mathsf{C}>\mathsf{E}>\mathsf{D}>\mathsf{A}>\mathsf{B}$ with $(\mu,\sigma)=(6.21,2.48)$. The latter correspond to the two leftmost purple circles in Figure 5.

3 Discussion

We analyze voting results in a ranked-ballot poll by plotting potentially winning rankings on the plane (μ,σ). In this way both the standard Condorcet criterion of minimizing $\mu(\mathbf{c})$ and the new criterion of minimizing $\sigma(\mathbf{c})$, that we have introduced, can be considered at the same time in order to identify the optimal consensus ranking. The fundamental importance of the new criterion relies on the fact that consensus ranking of small $\sigma(\mathbf{c})$ are much more robust to noise in the data, e.g. fluctuations in the number of voters, which are somehow unavoidable.

We have also shown that standard voting rules, such as Copeland, Schulze or Tideman, may provide a consensus ranking which is not the best representative consensus. Indeed, trying to minimize only $\mu(\mathbf{c})$, these classical methods may eventually prefer a ranking which improves over other rankings by $O(1/\sqrt{n})$, that is the order of magnitude of noise-induced fluctuations. This is related to the very well known problem of data over-fitting. The best solution to this over-fitting problem would be to estimate the uncertainty on the data, then compute the a posterior probability distribution on all the possible rankings and finally extract the most representative consensus ranking from this probability distribution. Given that the just outlined procedure is not easy to implement, we have provided, through the use of both $\mu(\mathbf{c})$ and $\sigma(\mathbf{c})$, a more economic way of finding a consensus ranking which is at the same time representative and robust.

To help in this new analysis we have set up a webpage with an interactive tool that produces the graph in the (μ,σ) plane [6], once the list of ranked ballots is given as input. All plots in this manuscript, using the standard Kemeny distance, are based on those produced by the web tool. We have analysed many different datasets coming from real polls and in general the plots in the (μ,σ) plane are similar to those shown above. Moreover we expect polynomial time algorithms can be developed that minimise (approximately) both $\mu(\mathbf{c})$ and $\sigma(\mathbf{c})$ in analogy to presently used voting rules that tend to minimise only $\mu(\mathbf{c})$.

Once the graph in the (μ,σ) plane is available, we believe any good consensus ranking should be chosen from the *optimal set*, resembling the concept of Pareto frontier in economics. A point belongs to the optimal set if no other point exists improving both in $\mu(\mathbf{c})$ and $\sigma(\mathbf{c})$ or improving only one of them while keeping the other constant. This set has been red-marked in the examples above and it extends from the Condorcet optimal ranking $\mathbf{c}^*$, that minimizes $\mu(\mathbf{c})$, to the ranking $\overline{\mathbf{c}}$ minimising $\sigma(\mathbf{c})$. The meaning of moving along this set should be clear: while the ranking $\mathbf{c}^*$ maximises total societal satisfaction ignoring individual satisfaction, $\overline{\mathbf{c}}$ is the more egalitarian in terms of individual satisfaction regardless of the total satisfaction. We are not claiming that $\overline{\mathbf{c}}$ should be the consensus ranking: often $\mu(\overline{\mathbf{c}})$ is much larger that $\mu(\mathbf{c}^*)$ and the optimal consensus ranking is actually in the middle of the optimal set. Instead, we are proposing a new tool that provides a quantitative meaning to each possible choice. Which consensus ranking should be chosen among the optimal set is no longer a technical matter, it is rather a decision to be taken by the people in charge and the criteria may change according to the domains: political elections, marketing, web page ranking, etc. In some cases, like for instance in political election, the decision on which point to select along this line must be taken *before* the poll is run. There are polls, like political elections, where the consensus ranking must produce a unique winner among the candidates. In this case one can restrict the analysis to rankings having no tie at the first position and a line of optimal rankings can be defined as well in this subset of rankings. The optimal set can be also used to compare consensus rankings suggested by the existing winner selection methods.

The cases where the plot in the (μ, σ) plane is even more useful is when the final decision can be taken *after* the poll/survey is run. In this case having a data aggregation like the one we are presenting in terms of $\mu(\mathbf{c})$ and $\sigma(\mathbf{c})$ provides a lot of information and allows for a much better choice. A typical example is when politicians want to decide a list of priorities based on suggestions coming from the electorate: the politicians can run a poll/survey among the electorate and this would determine the optimal rankings, leaving to the politicians the final choice of the consensus ranking, to be chosen among those. We believe this is an ideal compromise between taking in serious consideration the desiderata of the electorate (the line of optimal consensus rankings is fully determined by the votes) and leaving the political decision to those in charge.

The applications where technical tools provide a set of optimal preferences among which the final choice is left to the user are not new in other fields. For example in quantitative financial risk management the mathematical analysis produces a risk-return curve (called efficient frontier [18]) and the choice of a point along such a curve is left to the investor. From a different perspective a voting theory purely based on the maximisation of voters satisfactions would be equivalent, in political economy, to the maximisation of total wealth in a country regardless of its distribution and welfare criteria.

We have shown that our method leads to the identification of an optimal set in the two-dimensional space of satisfaction and egalitarianism, based on a theoretical assumption of distance among rankings. In fact, our study may be applied only to winner selection methods based on a distance. Once the choice of the distance is made, the optimal set is uniquely determined. The arbitrariness of such choice could be eliminated by inferring, with an inverse problem procedure, what is the 'proper' distance to be used in specific cases. Namely post-vote polls could be performed by asking the voters to evaluate their personal satisfaction with some proposed winning rankings, obtaining in this way an experimental two-dimensional classification of the optimal set.

This information could hint at the best distance to be used, by minimising the differences between theoretical and experimental points.

The voting method we have presented here provides an efficient technical tool to determine the line of optimal rankings, among which a political decision has to be taken. While it is generally understood and acknowledged that democratic organisations should not only maximise their goods but also distribute them as equally as possible, such awareness did not lead so far to a proper solution in social choice theory. We believe therefore that the quantitative method we have introduced is a fundamental tool to apply democratic principles, especially in voting processes.

4 Materials and Methods

4.1 Distance between rankings

The Kemeny distance [15,16] $d_{\mathrm{Kem}}(\mathbf{r},\mathbf{s})$, is one of the possible means of quantifying how dissimilar two rankings $\mathbf{r}$ and $\mathbf{s}$ are. Intuitively, the distance relates to how many pairwise comparisons of candidates do not match between the two rankings. For instance, if candidate A is preferred to candidate B in one ranking, but B is preferred to A in the other, that would count 1 in the distance. If one ranking considers $\mathsf{A}=\mathsf{B}$ while the other does not, then that would count 1/2 in the distance. By summing over all possible pairs, with (A,B) and (B,A) counted separately, one obtains the Kemeny distance between the two rankings.

The computation of $d_{\mathrm{Kem}}(\mathbf{r},\mathbf{s})$ is simpler if rankings are rewritten in terms of the score matrices $M(\mathbf{r})$:

$$M_{ij}(\mathbf{r}) = \begin{cases} 1 \text{ if candidate } i \text{ is preferred to candidate } j \text{ in ranking } \mathbf{r} \\ -1 \text{ if candidate } j \text{ is preferred to candidate } i \text{ in ranking } \mathbf{r} \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

The Kemeny distance between rankings $\mathbf{r}$ and $\mathbf{s}$ is thus given by

$$d_{\mathrm{Kem}}(\mathbf{r},\mathbf{s}) = \frac{1}{2}\sum_{i,j} |M_{ij}(\mathbf{r}) - M_{ij}(\mathbf{s})| \tag{4}$$

4.2 Combinatorics

The set of rankings without ties for m candidates, $\mathcal{R}_m$, has cardinality $m!$. Let us call T_m the cardinality of the set of rankings including ties, $\overline{\mathcal{R}}_m$. T_m are sometimes called Fubini, or Cayley numbers. One can show (see the OEIS website[21] and references therein) that their exponential generating function is

$$F(x) \;=\; \sum_{m=0}^{\infty} \frac{T_m}{m!} x^m = \frac{1}{2-e^x}\ , \tag{5}$$

whose radius of convergence is $\ln 2$. This can be used to find T_m from derivatives and gives $T_0 = 1$, $T_1 = 1$, $T_2 = 3$, $T_3 = 13$, $T_4 = 75$, $T_5 = 541$, $T_6 = 4683$, $T_7 = 47293$, $T_8 = 545835$, $T_9 = 7087261$, $T_{10} = 102247563$, etc. These numbers grow according to the formula

$$T_m \simeq \frac{m!}{2(\ln 2)^{m+1}} \simeq (1.44)^m m! \,, \tag{6}$$

with a sub-leading correction decaying exponentially fast

$$\left(T_m - \frac{m!}{2(\ln 2)^{m+1}} \right) \frac{1}{T_m} \simeq (0.11)^m \,. \tag{7}$$

One can also consider the set of rankings with l ties, where $0 \le l \le m-1$: $\overline{\mathcal{R}}_m^{(l)}$. Clearly $\mathcal{R}_m = \overline{\mathcal{R}}_m^{(0)}$ and $\overline{\mathcal{R}}_m = \cup_{l=0}^{m-1} \overline{\mathcal{R}}_m^{(l)}$. Another interesting set for applications is the set $\widehat{\mathcal{R}}_m$ containing rankings where the first candidate is untied. For each set of rankings our method provides a subset of optimal rankings according to the following definition.

4.3 Optimal set

For two rankings $\mathbf{s}$ and $\mathbf{r}$ in $\mathcal{S}$, we say $\mathbf{s}$ *improves* $\mathbf{r}$ if $\sigma(\mathbf{s}) \le \sigma(\mathbf{r})$ and $\mu(\mathbf{s}) \le \mu(\mathbf{r})$, and at least one of the two inequalities is strict. The optimal set $\mathcal{O}_{\mathcal{S}}$ is the set of points in $\mathcal{S}$ that cannot be improved by other elements of $\mathcal{S}$.

On the web platform that we have developed [6] we show the global optimal set $\mathcal{O}_{\overline{\mathcal{R}}_m}$ (red diamonds) and the one with no ties $\mathcal{O}_{\mathcal{R}_m}$ (purple circles). In other contexts, like engineering or economics, the optimal set of vectors of a d-dimensional space is called Pareto frontier [10]. In general the computation of such an optimal set requires a time proportional to the cardinality T_m [11], that is a time exponential in the number m of candidates.

Indeed also the computation of the Condorcet optimal consensus ranking with Kemeny distances (which is one element of the optimal set) is in general a NP-hard problem. However, if the n ranked ballots given in input are not too dissimilar, such an optimum can be computed in polynomial time [3]. Nonetheless the cases where our new criterion is meaningful are exactly those where the ranked ballots are not too similar. We believe that for the computation of the optimal set in the large m limit, one should resort to Monte Carlo methods, already successfully used in the computation of Kemeny optimal rankings [23].

4.4 Some winner selection methods

In the plots in the main paper we have shown the consensus rankings obtained by some well-known winner selection methods, Copeland, Schulze and Tideman [4]. Here we provide a detailed description of these methods, which are the most commonly used in situations where the voter ballots are lists of candidates ordered by preference (ranked ballots).

We consider the same situation as in the main paper, where n voters express their preferences about m candidates. The ballot of each voter can be conveniently mapped in a vector $\mathbf{r}$ of m integers representing the positions of each candidate in the preference list. For example the ballot C > A > E > D > B corresponds to the vector $\mathbf{r} = (2, 5, 1, 4, 3)$. From the n vectors $\mathbf{r}^{(k)}$, with $k = 1 \ldots, n$, representing the voter ballots we can build the matrix of total preferences whose elements are

$$P_{ij} = \sum_{k=1}^{n} \mathbb{I}(r_j^{(k)} > r_i^{(k)}) \; , \tag{8}$$

where the indicator function $\mathbb{I}$ is defined as

$$\mathbb{I}(\text{condition}) = \begin{cases} 1 \text{ if condition is true} \\ 0 \text{ if condition is false} \end{cases} \tag{9}$$

In practice the matrix element P_{ij} counts how many voters prefer candidate i to candidate j. The result of any voting method based only on pairwise comparisons between candidates can be obtained from matrix P.

A method of selecting a consensus ranking based on scores is *Copeland*, also known as the *pairwise comparison*. Candidates are ranked according to the score C_i that counts the number of pairwise comparisons won plus half of those tied

$$C_i = \sum_{j=1}^{m} \left[\mathbb{I}(P_{ij} > P_{ji}) + \frac{1}{2}\mathbb{I}(P_{ij} = P_{ji}) \right] \; . \tag{10}$$

The Copeland candidate(s) is the one with maximum C_i.

The *Schulze* method is also based on pairwise comparisons between candidates. To compute the Schulze ranking from the matrix P we first have to compute the matrix B of *beatpaths*, by initialing it as $B_{ij} = P_{ij}$ and then iterating until convergence

$$B_{ij} = \max\left(B_{ij}, \; \max_k \min(B_{ik}, B_{kj}) \right) \; . \tag{11}$$

The number of iterations to make the matrix B converge is given by the length of the longest beatpath, which is at most the number of candidates m. Successively, candidates are ranked according to a score similar to the pairwise one for the B matrix, that is

$$Z_i = \sum_{j=1}^{m} \left[\mathbb{I}(B_{ij} > B_{ji}) + \frac{1}{2}\mathbb{I}(B_{ij} = B_{ji}) \right] \; . \tag{12}$$

The Schulze candidate(s) is the one with maximum Z_i.

Tideman is a further method of selecting a consensus ranking. To compute the Tideman solution, the elements of matrix P are sorted in a decreasing order and taken into account one by one. When element P_{ij} is considered, the relative order $i > j$ in the final ranking is assigned unless in contrast with the partial ordering already fixed by larger values of P previously considered.

Acknowledgements We thank Flavio Chierichetti for drawing our attention to the rank aggregation problem. We thank the Airesis platform for providing access to their data.

References

1. Airesis platform. URL `http://www.airesis.it`. Date of access:15/11/2014
2. Arrow, K.J.: A difficulty in the concept of social welfare. The Journal of Political Economy **58**, 328–346 (1950). DOI:10.1086/25696
3. Betzler, N., Fellows, M.R., Guo, J., Niedermeier, R., Rosamond, F.A.: Fixed-parameter algorithms for kemeny rankings. Theoretical Computer Science **410**(45), 4554–4570 (2009). DOI:10.1016/j.tcs.2009.08.033
4. Borgers, C.: Mathematics of Social Choice: Voting, Compensation, and Division. SIAM, Philadelphia (2010)
5. Castellano, C., Fortunato, S., Loreto, V.: Statistical physics of social dynamics. Reviews of modern physics **81**(2), 591 (2009)
6. Contucci, P., Panizzi, E., Ricci-Tersenghi, F., Sîrbu, A.: Rateit web tool. URL `http://www.sapienzaapps.it/rateit.php`. Date of access:15/11/2014
7. D. Easley, J.: Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge University Press, Cambridge UK (2010)
8. Dwork, C., Kumar, R., Naor, M., Sivakumar, D.: Rank aggregation methods for the web. In: Proceedings of the 10th international conference on World Wide Web, pp. 613–622. ACM (2001). DOI:10.1145/371920.372165
9. Fagin, R., Kumar, R., Mahdian, M., Sivakumar, D., Vee, E.: Comparing and aggregating rankings with ties. In: Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pp. 47–58. ACM (2004). DOI:10.1145/1055558.1055568
10. Feldman, A.M., Serrano, R.: Welfare economics and social choice theory. Springer, Berlin (2006)
11. Godfrey, P., Shipley, R., Gryz, J.: Algorithms and analyses for maximal vector computation. The VLDB Journal - The International Journal on Very Large Data Bases **16**(1), 5–28 (2007)
12. Goldberg, K.: Jester dataset. URL `http://www.ieor.berkeley.edu/~goldberg/jester-data/`. Date of access:15/11/2014
13. GroupLens Research: Movielens dataset. URL `http://grouplens.org/datasets/movielens/`. Date of access:15/11/2014
14. Heiser, W.J., D'Ambrosio, A.: Clustering and prediction of rankings within a kemeny distance framework. In: Algorithms from and for Nature and Life, pp. 19–31. Springer, Berlin (2013)
15. Kemeny, J.G.: Mathematics without numbers. Daedalus **88**(4), 577–591 (1959)
16. Kemeny, J.G., Snell, J.L.: Mathematical models in the social sciences, vol. 9. Blaisdell, New York (1962)
17. Marie Jean Antoine Nicolas de Caritat, Marquis de Condorcet: Essai sur l'application de l'analyse à la probabilité des décisions rendus à la pluralité des voix. L'Imprimerie Royale, Paris (1785)
18. Markowitz, H.: Portfolio selection*. The journal of finance **7**(1), 77–91 (1952)
19. Monjardet, B.: "mathématique sociale" and mathematics. a case study: Condorcet's effect and medians. Electronic Journal for History of Probability and Statistics **4**(1), 1–26 (2008)
20. Moore, M., Katzgraber, H.G.: Dealing with correlated choices: How a spin-glass model can help political parties select their policies. Physical Review E **90**, 042,117 (2014)

21. OEIS foundation: The on-line encyclopedia of integer sequences. URL `http://oeis.org/A000670`. Date of access:15/11/2014
22. Raffaelli, G., Marsili, M.: Statistical mechanics model for the emergence of consensus. Physical Review E **72**(1), 016,114 (2005)
23. Renda, M.E., Straccia, U.: Web metasearch: rank vs. score based rank aggregation methods. In: Proceedings of the 2003 ACM symposium on Applied computing, pp. 841–846. ACM (2003). DOI:10.1145/952532.952698
24. Saari, D.G., Merlin, V.R.: A geometric examination of kemeny's rule. Social Choice and Welfare **17**(3), 403–438 (2000)
25. Truchon, M.: Aggregation of rankings: a brief review of distance-based rules and loss functions for the expected loss approach. Cahier de recherche/Working Paper **5**, 34 (2005). DOI:10.2139/ssrn.984305
26. Young, H.P.: Condorcet's theory of voting. American Political Science Review **82**(04), 1231–1244 (1988)