INFORMS TutORials in Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Separation, and Separation-Deviation Methodology for Group Decision Making and Aggregate Ranking

Dorit S. Hochbaum

To cite this entry: Dorit S. Hochbaum. The Separation, and Separation-Deviation Methodology for Group Decision Making and Aggregate Ranking. *In* INFORMS TutORials in Operations Research. Published online: 14 Oct 2014; 116-141.
https://doi.org/10.1287/educ.1100.0073

TUTORIALS IN
OPERATIONS RESEARCH
INFORMS 2010

 | ISBN 978-0-9843378-0-4
DOI 10.1287/educ.1100.0073

The Separation, and Separation-Deviation Methodology for Group Decision Making and Aggregate Ranking

Dorit S. Hochbaum
Department of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley, California 94720, hochbaum@ieor.berkeley.edu

Abstract In a generic group decision scenario, the decision makers review alternatives and then provide their own individual ranking. The aggregate ranking problem is to obtain a ranking that is fair and representative of the individual rankings. We argue here that using cardinal pairwise comparisons provides several advantages over scorewise models. The aggregate group ranking problem is then formalized as the *separation* model and *separation-deviation* model. The premise of the models is to use the implied or explicit pairwise comparisons in the reviewers' input and assign an aggregate ranking that minimized the penalty of not agreeing with the scores, *and* the penalty for not agreeing on the intensity of the pairwise preference. The latter permits to the incorporation of confidence levels in the input provided by reviewers on specific pairwise comparisons, as well as on specific scores. Both separation and separation-deviation models have been shown to be solved efficiently, for convex penalties. This tutorial presents several group ranking scenarios where the pairwise comparisons are the input, such as sports competitions. We show that using cardinal, rather than ordinal, pairwise comparisons, and the proposed separation model, is advantageous. In group ranking contexts, where pairwise comparisons are not inherently available, there is also an advantage of using *implied* pairwise comparisons. The latter contexts include, e.g., the National Science Foundation review panels, choosing winning projects, determining countries' credit risk, and customer segmentation. We unify the group decision problem with the problem of webpages rankings and ranking academic papers in terms of citations. We compare and contrast the separation approach with PageRank and the principal eigenvector methods. The problem of aggregating rankings "optimally" with pairwise comparisons is shown to be linked to a problem we call *the inverse equal-paths problem*. The graph representation provides insights and enables the introduction of a specific performance measure for the quality of the aggregate ranking as per its deviations from the individual rankings observations. We show that for convex penalties of deviating from the reviewers' inputs the problem is polynomial-time solvable, by combinatorial and polynomial-time algorithms related to network flows. As such, the approach is very efficient. We demonstrate further how graph properties are related to the quality of the resulting aggregate ranking. Our graph representation paradigm provides a unifying framework for problems of aggregate ranking, group decision making, and data mining.

Keywords network flow; aggregate ranking; inverse problems

1. Introduction

In group decision making, one of the major challenges is to achieve an aggregate solution that is *fair* and *representative* of the evaluations of the decision makers in the group. Naturally, the concepts of "fair" and "representative" elude precise and formal definition, which accounts for the numerous alternative models and interpretations that exist for group

TABLE 1. Classification of group decision problems.

Cardinal	Ordinal NP-hard, Arrow's impossibility theorem (Arrow [8])
Comparisons	Scores
Evaluating pairs	Rating singletons
Partial list	Full list

decision making. These models are classified according to characteristics, some of which are listed in Table 1. Our focus here is on decision problems that are cardinal, have input that includes pairwise comparisons that are provided explicitly or generated implicitly, and that allow partial lists.

In *ordinal* rankings the input is in the form of an ordering, or preorder, or permutation of the alternatives. Arrow's impossibility theorem (Arrow [8]), states that there is no aggregate ranking that satisfies simultaneously several necessary fair representation conditions. More details on this are provided in §5. The rankings addressed here are *cardinal*, meaning that the pairwise comparisons are provided with *intensity*. Therefore, an alternative i might be preferred to alternative j by a tiny intensity of ϵ, or by a huge amount of intensity, which is taken into account in the cardinal model, but for ordinal ranking there is no differentiation between those preferences of i to j.

The simplest and most common method for aggregating rankings considers the average (or weighted average) of scores as the weight indicator for the aggregate ranking. This approach cannot be used in scenarios where the input is in the form of pairwise comparisons, such as in ranking sports competitors based on the outcomes of games, which amount to pairwise comparisons of the strengths of the respective competitors. It is demonstrated here that the use of, explicit or implied, pairwise comparisons improves the quality of the aggregate decision in that the outcome represents more fairly the decision makers' evaluations. A recent article by Brandstatter et al. [10] demonstrates that pairwise comparisons and relative evaluation of attributes is the preferred mode for people to make decisions and rank alternatives. This is in contrast to decision making based on a scalar value or score associated with each object.

An issue that often mires the decision-making process is that of *partial* lists (aka *incomplete ranking*). Partial lists mean that each decision maker reviews only a subset of the alternatives. This results in several challenges, one of which is the extrapolation of the individual decision maker assessment of the subset to the entire set. Another is the biases introduced by the allocation of the review tasks to the individual reviewers. Therefore, many decision-making models require full lists in their setup.

The problem of ranking competitors based on an incomplete set of pairwise comparisons is a well-studied one in the context of football and other sports, and also in general (David [21]). There are numerous ranking schemes, each with its unique emphasized factors, and each with its advantages and shortcomings. The shortcomings, particularly in sports teams rankings where passions run high, bring about the ire of some people. The generic criticism is that certain game outcomes have not been adequately incorporated, or have had an excessive impact on the aggregate ranking. Another critique expressed concerns the "strength of the schedule," which amounts to the allocation of the pairwise comparisons, or the collection of the pairs who will play the games. The schedule, or allocation, has an inordinate effect on popular techniques based on principal eigenvector (Google Rank is one such technique). We will show here that, in contrast, the separation model avoids these shortcomings, and deals with partial lists while still providing a fair aggregate ranking.

One aspect shared by most existing schemes for pairwise comparisons is the consideration of all pairwise comparisons as equal in their impact on the final outcome. This brings about biases such as counting a win against a weak team of equal weight to a win against a strong team. Consequently, it might be preferable to not play a game at all if it is against

a weak team, because such a game played by a strong team can actually reduce its rank (Keener [45]). This uniformity of consideration of each pairwise comparison is one reason for the inclusion of *human polls* in, e.g., college footfall ranking. Human judgement has the advantage that it can take into account the quality of the game played, rather than the score alone, which is often attributed, to some degree, to chance. These human polls, in turn, are often criticized for lack of transparency, because the factors that go into human ranking are not made explicit.

We consider here an aggregate ranking scenario with partial lists, whereby the input to the ranking process is in the form of pairwise comparisons or implied pairwise comparisons. An aggregate ranking is one that minimizes the total penalties for deviating from the input comparisons. The power of this paradigm, *the separation model*, is in its capabilities of dealing with partial lists, and with difference confidence levels in the inputs provided, characterized by the candidates and the expertise of the reviewers evaluating them. The separation model is extended to the separation-deviation model in order to optimize not only the closeness to the pairwise comparisons, but also to the specific scores. These models were introduced in Hochbaum [36, 37] and Hochbaum and Levin [38].

The separation model differs from existing aggregate ranking models in that it permits the explicit inclusion of subjective factors. In other words, any form of input from knowledgeable sources can be incorporated, and each input is associated with any degree of confidence deemed appropriate. Of course, the degree of confidence assigned can in itself be subjective, but it can be made according to a specific protocol and set of rules agreed upon in advance. This allows us to differentiate the importance of different comparisons and calibrate their impact on the final ranking.

The separation model provides a *performance measure* that can be used to compare different aggregate rankings. In the literature on obtaining aggregate ranking there is rarely a performance measure on the quality of the attained consistent ranking. One exception is Kemeny and Snell's model (Kemeny and Snell [46]) for aggregate ranking of *ordinal* rankings in the form of permutations only. Kemeny and Snell's model seeks an aggregate ranking minimizing the number of reversals of the aggregate permutation ranking compared to the individual permutations. For an ordinal ranking to be consistent there must be no cycles of preferences. Thus, Kemeny and Snell's model can be viewed as modifying an input in the form of a graph with directed arcs, where each arc (i, j) indicates that a reviewer prefers i to j, by changing the directions of some of the arcs so that the resulting graph is *acyclic* and so that the number of changes in arc directions is minimum. This problem is known as the *minimum arc feedback set* problem and one of the drawbacks of the model is that it is NP-hard to solve optimally because the arc feedback set problem is known to be NP-hard. As an NP-hard problem there has been extensive work on approximation algorithms for the arc feedback set problem. Whereas for general graphs there are no constant factor approximations known, for tournaments (where each pair is compared by at least one reviewer) there are good approximation algorithms for the arc feedbacks set problem, e.g., in Ailon et al. [4].

Both the separation and separation-deviation models are cast here as graph problems, and are within the *inverse problem* paradigm. In an inverse problem one is given some problem parameters that do not satisfy certain necessary conditions, or conflict with observations. The goal is to modify those parameters subject to a penalty function on the modification, and so that the total penalty is minimum. In the context of rankings, the outcomes of different games may be conflicting with respect to any given ranking. For instance, if each team loses at least once, then a top team that ranks number one has its ranking inconsistent with the game(s) it has lost. Thus, any aggregate ranking is going to conflict with some inputs, except for the rare case where each outcome is precisely consistent with one underlying ranking.

Aggregate ranking is, hence, an inverse problem, where the scores of the games played, and any other form of judgement or input on pairwise comparisons, may be incorporated

as part of the input. The problem is to come up, with a pairwise comparison for each pair, that is consistent with some underlying ranking and that deviates as little as possible from the given inputs. The penalty for deviating from the inputs is measured in terms of *penalty functions* that are monotone increasing (or nondecreasing) with the size of the deviation. These penalty functions are assigned to each input separately, so the penalty of a less reliable source of comparison can take lower value than a penalty for comparison by a high confidence source.

1.1. Outline

Key concepts of group decision making and aggregate ranking are illustrated via a list of scenarios each with its own distinguishing features, in §2. Section 3 discusses how pairwise comparisons are obtained and used. *Consistency* of pairwise comparisons is defined and discussed in §4, both in the context of consistency of pairwise comparisons given as a matrix and as a property of a graph having all pairwise paths of equal lengths. In §5 we introduce the inverse equal-paths problem as the graph-form of the separation model. The quality of the group ranking is then associated with graph properties such as k-connectivity in §6.

Leading alternative approaches are reviewed in §7. The techniques surveyed include the average weight, several optimization methods, the principal eigenvector and its closely related analytic hierarchy process (AHP), and the related PageRank technique. These techniques are then compared and contrasted with the models of separation and separation-deviation. A generalization of the separation to the separation-deviation model is described in §8. Properties of the separation and separation-deviation models for specific penalty functions are detailed in §9. We conclude with several remarks in §10.

1.2. Notations, Preliminaries, and the Inverse Equal-Paths Problem

The pairwise comparison preference of i to j stated by reviewer r is denoted by p_{ij}^r. If reviewer r provides pointwise scores evaluation for object i, this score is denoted by w_i^r.

We show here that the inverse equal-paths problem is equivalent to the problem of obtaining a minimum penalty aggregate ranking. To that end, we introduce the relevant graph notation.

The input to the rank aggregation problem is a nonsimple connected graph $G = (V, A)$, where for each arc $(i, j) \in A$ of weight p_{ij}^r there is an arc in the opposite direction (j, i) of weight $-p_{ij}^r$. There can be multiple opposing pairs of arcs for each pair of nodes, or none at all. Another input is a penalty function $f_{p_{ij}^r}(\,)$ for each given arc in the graph.

A feasible solution to the aggregate ranking problem has to be *consistent*. Thus, a feasible solution z_{ij}^* for each pair of nodes i, j must satisfy that for any pair of nodes $s, t \in V$ all the directed paths from s to t in G with the weights $\mathbf{z}^*$ are of the same length. (This is proved in §5.) A set of weights $\mathbf{z}^*$ is said to be optimal for the *inverse equal-paths* problem if among all feasible weight vectors $\mathbf{z}$ it minimizes the total sum of the penalty functions $\sum_{(i,j)\in A,\, r\in R_{ij}} f_{p_{ij}^r}(z_{ij})$.

2. Case Studies and Examples of Group Decision and Aggregate Ranking Scenarios

We discuss here scenarios of group decision making and illustrate several key concepts via these examples.

The National Science Foundation (NSF) panel review. The National Science Foundation has proposals submitted to each program. The proposals in each program area are allocated by the director of the division to experts that are invited to serve on the panel. Those experts see only a subset of the proposals submitted. This is so as not to overburden the reviewers, and because each reviewer can evaluate proposals only in their area of expertise—typically a subset of the proposals submitted.

Each member of the panel reviews the assigned proposals and grade each proposal on a scale from 1 to 5, with 1 meaning "poor" and 5 meaning "excellent." The panel then meets face to face in a session (that lasts a day or two) and determines the top-ranking proposals that will be recommended for funding. Ultimately, the decision depends on the average score of each proposal. That is, the scores assigned to each proposal are summed and divided by the number of reviewers of the proposal (although it is almost always the same number of all proposals, with some minor exceptions.)

This type of group decision making, based on the scores' average, is called here the *average weight* model. The NSF panel review also involves *partial lists* (also referred to as *incomplete rankings*). That means that each reviewer evaluates only a subset of the total set of proposals. This can bring about potential problems of *incomparability*. For instance, it is conceivable that there are two sets of proposals, and each reviewer gets a subset of one of them. If the two top proposals are one from each set, there is no reviewer that reviewed both of them. Furthermore, it is possible (though unlikely) than all proposals in one set are of higher quality than all those in the other set. In that case, any proposal in the better set should rank higher than the leading proposal in the second set. However, the panel has no grounds for comparison that would allow making this "correct" ranking decision.

This problem with partial lists is more of an *allocation* issue—how to allocate the proposal to reviewers within their area of expertise, and so that there will be basis for comparison and the scenario above will be avoided. This problem has been studied extensively for various versions and complexity of the allocation decision in Hochbaum and Levin [39]. Cook et al. [19] investigated the allocation of the ranking tasks to individuals in the context of the NSF process and proposed a heuristic algorithm to generate good feasible solutions for a version of the allocation problem.

Grade point average (GPA). In college and university admissions, one factor deemed to represent the quality of students is their GPA. Each instructor evaluates a proper subset of the universal set of students—those attending the instructor's class. The evaluation is in the form of grades, and consequently also a corresponding ranking of all students in each instructor's class. The ranking of all students based on their collection of grades from different instructors' classes is a group ranking problem. The GPA is also an average weights methodology and it is also a partial list ranking, as is the NSF review process. Because the instructor evaluates only a subset of the students—those in class, it is likely, and indeed as always happens with grading on a curve, that students that take a class with good students will get a lower grade than if they take a class with less successful students. In other words, the quality of the class, or the list, affects the score. To optimize their GPA, students should then try to take classes with faculty that are known to give high grades, and/or attempt to take classes that are easy for them, or where the capabilities of others in class are expected to be less than theirs. Both phenomena of students optimizing their GPA are indeed rather commonplace.

Ranking of baseball teams. The determination of the team clinching a playoff berth in each division is based on the number of wins. In major league baseball there are no ties allowed, and each team plays exactly the same number of games. This ranking is, then, an average weight model. However, each team plays games with other teams within its own division, so the particular division a team belongs to may have a major impact on its possible chances of advancing to the playoffs. This effect is also that of the partial lists in the GPA example. Each team plays only teams in the same division. Therefore, if these teams are strong, the chances of advancing are slimmer compared to a division where the teams are weak.

Notice that here the "reviewers" are the games. Each game "reviews" a pair of teams and assigns a score of 1 to the winning team and a score of 0 to the losing team (there are no ties in baseball).

College football ranking. In most sports the rankings are based on pairwise rankings only, which are the outcomes of the games. The principal eigenvector technique has been applied to this and other similar ranking scenarios, where the reviews are all in the form of pairwise comparisons of relative strength, or intensity. The principal eigenvector technique is discussed in more detail in §7.1. As in baseball, each reviewer is a game, and the outcome—usually a function of the score of the game—is the pairwise comparison.

Ranking of academic papers by citations or webpage ranking. The unique feature here is that the reviewers and the objects are the same set. This causes substantial problems of manipulations, primarily with Web pages. PageRank is an algorithm that was supposed to address this. There will be more on PageRank and its relationship to the principal eigenvector technique in the §7 on alternative techniques. Here, each paper "reviews" another paper by providing a citation.

Conference program committee. In many conferences there is a program committee reviewing the papers submitted. The list of papers submitted is allocated to individual reviewers on the program committee, in consideration of their area of expertise and how it matches the content of the paper. This scenario is analogous to that of the NSF panel, except that reviewers are asked to not only provide a score, but also a *confidence level*. The meaning of confidence level is as an assessment of the reviewer's own expertise in regard to the particular paper, or as a measure of the effort investment by the reviewer to fully investigate the content of the specific paper and relevant literature. Because program committees' face-to-face meetings are expensive, committees rarely meet these days, and the discussion and aggregate ranking is made online. This means that there has to be a way of quantifying disagreements. One way of pinpointing disagreements is by the spread, or variance, of the scores.

In this author's experience, in cases when program committees did meet, much of the discussion was in the flavor of "if paper 1 is accepted, then certainly paper 2 must be accepted, for the following reasons...." Or, "if paper 3 is rejected, then for the following reasons paper 4 must be rejected also." These pairwise comparisons are difficult to identify online in the existing system, where only the averages are considered. Furthermore, the confidence levels that are recorded are not actually used, except when there is a high variance in the scores. In that latter case, the confidence levels may become part of the discussion when some of the reviewers (usually those with higher confidence levels) attempt to convince the others to change the scores.

Student paper competition. In Hochbaum and Moreno-Centeno [41] we analyze two years of the manufacturing and service operations management (MSOM) student paper competition to identify the student paper winner in operations management. In the MSOM competition the judges are each assigned a relatively small number of papers (about four), compared to the size of the pool submitted. Because every paper must be read by about four reviewers, the number of judges is very large in these competitions, and roughly of the same magnitude as the number of papers submitted. This makes the analysis of the results beyond the average score difficult. With the separation model it was possible to identify outliers in terms of pairwise comparisons by simply looking at leading penalty terms. Such outliers had gaps between the scores of two papers that were very different from the gaps implied by others and from the gap implied by the aggregate ranking.

Countries' credit risk. In Hochbaum and Moreno-Centeno [40], we considered countries' credit risks that were rated by different rating agencies. Each of the agencies rated all countries on the list, making it a *full-list* scenario. One of the challenges here is that each agency is using a different scale, which makes even the average score nontrivial to obtain. However beyond that, we showed in Hochbaum and Moreno-Centeno [40] that the separation deviation model is capable of generating a "better" aggregate ranking in that it has a smaller number of reversals, and thus better represents the individual agencies.

Table 2. Scores by four reviewers for four proposals.

Proposal	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Scores' sum
100,001		4.50	4.00	3.50	12.00
100,002	5.0	4.00	3.75		12.75
100,003	5.0	4.25		3.25	12.25
100,004	5.0	3.00	3.00		11.00

Note. Each proposal is reviewed by the three reviewers.

Customer segmentation. In Hochbaum et al. [43], we studied the segmentation of the customers of a high-tech company with respect to their proclivity to adopt new technology. The input data available are the history of the timing of the customers' purchases of several products. In this case the judges or reviewers of the customers are the products; and because each product was bought by a subset of the customers, this is a partial-list case. This type of analysis is interesting in that not all products are independent. That is, some of the products are later versions of others. This required some twists in the analysis, which rarely arises in other group ranking decisions. That is, one assumes that the judges are independent in their assessments, and there are no models to analyze the existence of *coalitions* and other game aspects on the part of the judges.

The individuals in the group that provide their assessments are referred to henceforth as *reviewers*. The objects evaluated are called *proposals* or *candidates*. This is similar to the scenario of the NSF review panel. The collection of all proposals is referred to as the *universal set*. A ranking is a pairwise comparison that can be provided with magnitude of the degree of preference, *intensity ranking*, or in terms of ordinal preferences only, *preference ranking*. These are also sometimes referred to as *cardinal* versus *ordinal* preferences.

The Netflix prize competition. Netflix announced a prize competition, starting in October 2006, to predict the review scores that customers will assign to movies. The methodology that was to be developed, described on their website,[1] was based on information on review scores provided by customers for movies watched (the training set). This was to be used to generate a consistent cardinal ranking of all movies for each customer, where the scores given by other customers could be used to calibrate the scale of each customer and the public/average perception of the quality of each given movie. This competition, which ended in September 2009 with the winners improving by more than 10% the accuracy of the scores compared to the methodology used by Netflix, is a well-publicized illustration of group decision making and its analogy to data mining.

3. Pairwise Comparisons

3.1. The Importance of Pairwise Comparisons

To see why pairwise comparisons are important, consider a committee reviewing candidates or proposals as in the NSF panel example. There are four reviewers, four proposals, and the ranking scores are on a scale of 1–5. Each proposal is evaluated by three of the four reviewers. This is, then, a case of *partial list* or *incomplete ranking*. The total sum of scores is given for each proposal. According to the sum of scores, the proposal that rates the highest is 100,002, followed by 100,003 and then 100,001 and finally 100,004. However, considering the pairwise comparisons, it is evident that all reviewers that have evaluated this pair of proposals prefer proposal 100,001 to 100,002, and prefer 100,001 to 100,003. Thus, a fair outcome that represents the views of the reviewers should place proposal 100,001 ahead of both 100,002 and 100,003.

[1] Details of the competition are available at http://www.netflixprize.com/leaderboard, accessed July 2010.

This underscores the shortcomings of the *average weight* model and the need to take into consideration not only the pointwise scores, but also to consider pairwise comparisons by the reviewers. Hochbaum and Moreno-Centeno [40] shows that using the implied pairwise comparisons in the separation and separation-deviation models yields a better rating/ranking of countries' credit risk in terms of the number of reversals of preferences of the rating agencies, who serve as "reviewers."

3.2. The Generation of Pairwise Comparisons Input

The pairwise comparison of i and j expresses the *intensity* of preference of i to j. Unlike *ordinal* ranking that only outputs a permutation and a binary order between each pair, saying that i is preferred to j, or vice versa, the setup here *quantifies* the extent that i is preferred to j. This can be done in either the *additive* or the *multiplicative* models, as presented formally below.

The separation model requires pairwise comparisons as input, but the input reviews may come as scores, translated to pairwise comparisons. We will consider two types of inputs that generate comparisons:

1. Implied by pointwise scores. When only pointwise scores are available, the pairwise comparisons, or intensity of preference by a certain reviewer r, is *implied* from the pointwise scores. The implication is that $p^r_{ij} = w^r_i - w^r_j$ in the additive sense, or $p^r_{ij} = w^r_i / w^r_j$ in the multiplicative sense.

Pointwise scores are used in scenarios such as NSF panel review; a conference program committee selecting papers to present; student paper competition; country credit risk assessment by agencies; and customer segmentation based on timing of purchases.

2. The input is in the form of pairwise comparisons. This is common in sports competitions where the winner or the ranking is determined by the outcomes of games, which are pairwise competitions. Although this is not obvious, the use of PageRank for ranking of academic papers by citations, and ranking of webpages, also relies on pairwise comparisons only, as will be shown later. The analytic hierarchy process (AHP), described in §7.3, is unique in that it requires the reviewers to specify, instead of scores, all intensity pairwise comparisons.

4. Consistency of Pairwise Comparisons Ranking for a Graph and Matrix Representation

4.1. Defining Consistency

The notion of consistency is critical when pairwise comparisons are the input. The matrix (p^r_{ij}) is said to be *consistent in the additive sense* if for every triplet i, j, k we have the "triangle equality":

$$p^r_{ij} + p^r_{jk} = p^r_{ik}.$$

Likewise, a comparisons matrix (p^r_{ij}) is said to be *consistent in the multiplicative sense* if for every triplet i, j, k we have

$$p^r_{ij} \cdot p^r_{jk} = p^r_{ik}.$$

Notice that if pairwise comparisons are generated from pointwise scores, then they satisfy consistency, by construction. This is not the case in general for inputs in the form of pairwise comparisons. It is rare, for instance, that sport teams playing each other more than once will get the exact same outcome in all games. And although it has been reported that people prefer to rank alternatives with pairwise comparisons (Brandstatter et al. [10]), if a person is asked to provide all pairwise comparisons between the alternative candidates, the person is more likely than not to provide comparisons that are inconsistent.

The multiplicative and additive notions of consistency are analogous: It is easy to transform multiplicative consistency to additive consistency by taking the logs of the intensities, and vice versa (by taking the exponents). In the graph representation, we use the notion of consistency in the additive sense, whereas the existing methods, relying on principal eigenvector, consider the matrix representation and the multiplicative sense of consistency.

We elaborate next on the notion of consistency, which is key in obtaining aggregate ranking—any solution to the ranking problem must be consistent. Specifically, we relate consistency to the *inverse equal-paths* property of a graph.

4.2. Graph and Matrix Representations of Pairwise Comparisons

The concept of consistency of pairwise comparisons applies to both cardinal and ordinal rankings. The approach here is that of *cardinal* ranking, rather than ordinal, but it is useful to contrast consistency in both setups.

The output of an ordinal aggregate ranking is an ordering or permutation of the set of candidates, and ties may be permitted. Consistency in ordinal ranking is equivalent to the notion of *transitivity*: We denote the (weak) order relation signifying that i is ranked at least as highly as j by $i \succeq j$. An order relation $\succeq$ is said to be *transitive* if it satisfies for all i, j, k: $i \succeq j$ and $j \succeq k \Rightarrow i \succeq k$. A collection of pairwise preferences is consistent if the corresponding order is transitive. More details on ordinal consistency are provided in Hochbaum and Levin [38].

Intensity rankings provide a cardinal quantifier to the preference. This quantifier is used either in the *additive* sense or in the *multiplicative* sense. When additive, the intensity represents the extent of the *difference* between the two proposals. The multiplicative intensity preference represents the *ratio* of the strengths of the ranks of the two proposals compared.

The notion of consistency with a matrix representation is in the multiplicative sense. For complete consistent matrices (a_{ij}), for each triple i, j, k, $a_{ij} \cdot a_{jk} = a_{ik}$. This is equivalent to the existence of a set of weights w_i for $i = 1, \ldots, n$ so that $a_{ij} = w_i / w_j$. Such a set of weights, called a *priority vector*, is not unique because for any consistent set of weights $w_1, \ldots, w_n$ and a scalar c, the set $cw_1, \ldots, cw_n$ is also a priority vector. Therefore, we can arbitrarily choose $w_1 = 1$ to ensure a unique set of weights corresponding to consistent intensity rankings.

The matrix representation is used in the principal eigenvector technique and in PageRank, which is a derivative of the principal eigenvector (details in §7), and by Saaty [51, 52] for the analytic hierarchy process technique. Let the intensity rankings be given in the form of a matrix $\mathbf{A} = (a_{ij})$, where a_{ij} is the magnitude of (multiplicative) preference of proposal i to j. The matrix is *complete* if all entries are present and generated from full lists, and *partial* otherwise.

For $\mathbf{A} = (a_{ij})$ a matrix of *intensity rankings* in the multiplicative sense, the values of all a_{ij}s are positive, and if $a_{ij} > 1$, then i is preferred to j, and if $a_{ij} < 1$, then j is preferred to i. Therefore, for consistent rankings $a_{ij} = 1/a_{ji}$ for all i, j. Similarly, for $\mathbf{A} = (a_{ij})$ a matrix of intensity rankings in the additive sense, the value of a_{ij} is positive (negative), indicating that i is preferred to j (j preferred to i) and the magnitude $|a_{ij}|$ indicates the intensity of that preference. Here, $a_{ij} = -a_{ji}$ for all i, j and the matrix is skew symmetric. Skew symmetry is a necessary condition for the consistency of a rankings matrix, but not sufficient. Some of the literature on finding "close" consistent rankings assumes that the (inconsistent) preference matrix is skew symmetric in that it satisfies this necessary condition.

One important implication of the notion of consistency is that in a consistent rankings matrix each column and row contain the full information on the entire matrix. For instance, given the ith column $(a_{i1}, a_{i2}, \ldots, a_{in})$ of a consistent ranking matrix in the multiplicative sense and setting $w_1 = 1$, one can generate all pairwise rankings as $a_{kj} = a_{ki} \cdot a_{ij} = a_{ij}/a_{ik}$.

For an incomplete matrix, we construct the *consistent closure* by placing for every missing i, j entry, a value generated from a sequence of entries, $(i, k_1), (k_1, k_2), \ldots, (k_p, j)$ if such sequence exists, and let $i = k_0$ and $j = k_{p+1}$. The value a_{ij} is then set to $\prod_{l=1}^{p+1} a_{k_{l-1}, k_l}$ if the rankings are expressed in the multiplicative sense, and $a_{ij} = \sum_{l=1}^{p+1} a_{k_{l-1}, k_l}$ if the rankings are

expressed in the additive sense. If such a sequence does not exist, then the (multiplicative) matrix does not satisfy the necessary condition of the Perron–Frobenius theorem in that there are pairs that are incomparable, directly or indirectly, and the equation $\mathbf{Ax} = \lambda\mathbf{x}$ does not have a unique positive solution. If there is at least one such sequence for each pair (we choose one arbitrarily if there is more than one sequence), then this process completes the matrix. Notice that if for each missing ranking there is only a single sequence of rankings comparing the two, then the matrix resulting from the completion process is necessarily consistent.

Suppose the matrix is consistent and the vector of priority weights is $\mathbf{w} = (w_i)_{i=1}^n$. Then $a_{ij} = w_i/w_j$. Summing up over all j, we obtain, $\sum_{j=1}^n a_{ij}w_j = nw_i$. Therefore, in matrix notation the vector of weights $\mathbf{w}$ satisfies, $\mathbf{Aw} = n\mathbf{w}$. We call this the *fixed-point* property of the eigenvector. The vector of weights is $\mathbf{w}$, hence, the eigenvector which consists of the weights assigned to each proposal or each criterion under the multiplicative model, but only if the matrix is consistent. Otherwise, the eigenvector forms some approximation of the preference weights. This is the motivation for the use of the eigenvectors. For another motivation, see §7.1.

The measure of approximation for a skew-symmetric inconsistent matrix defined by Saaty [52] is the *consistency index* (C.I.),

$$\text{C.I.} = \frac{\lambda_{\max} - n}{n-1}, \tag{1}$$

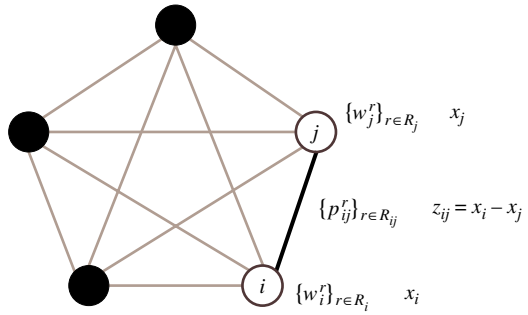
where $\lambda_{\max}$ is the maximum eigenvalue of the matrix. A matrix is said to be consistent if and only if C.I. is zero. This is equivalent to the conditions $a_{ij} \cdot a_{jk} = a_{ik}$ for all i, j, k. This notion of consistency can only be applied to skew-symmetric matrices, that is, for matrices that satisfy $a_{ij} = 1/a_{ji}$ for all $i < j$.

Although the consistency index is 0 for a consistent matrix, which is a desirable property of any measure of consistency, it is not known how the resulting priority vector's ranking reflects or deviates from the individual reviewers' rankings, and according to what measure.

The graph representation, $G = (V, E)$, formalizes a ranking or a set of multiple rankings. The universal set of proposals is the set of nodes V in the graph. Every pairwise comparison between a pair i and j is represented as an edge $[i, j]$ in E. There is a set of reviewers R and each reviewer $r \in R$ reviews a subset of V. The reviewers that provide scores for node i are in $R_i \subseteq R$. The score of node i given by reviewer r is w_i^r. $R_{ij} = R_i \cap R_j$ is the set of reviewers the review the pair $\{i, j\}$. A reviewer in R_{ij} may provide the implied gap as the comparison score, or just provide a pairwise comparison score of p_{ij}^r. Not all scores need to be given and not all pairwise comparisons need to be present. If the set R_{ij} contains more than one reviewer, then there are multiple pairwise comparisons assigned to $[i, j]$, or alternatively there are multiple arcs between i and j, each with a separate intensity weight, and the graph is then a "multigraph." A graph representation is illustrated in Figure 1.

Given a graph representation of group ranking, each pairwise preference of i and j is given as weights on the pair of arcs between i and j, with the weight on (i, j) being p_{ij}^r, and the

FIGURE 1. The graph created for the group ranking problem with a set of reviewers R.



weight of (j,i) $p^r_{ji} = -p^r_{ij}$. Because there could be multiple reviewers for the same pair of nodes, a necessary condition for the consistency of the graph is that all those agree, or for each $r', r'' \in R_{ij}$, $p^r_{ij} = p^{r''}_{ij}$. Given this necessary condition, in a consistent ranking graph there is at most a single weight associated with each arc. We next show that a consequence of the triangle equality is that the problem of group ranking can be presented as the *inverse equal-paths* problem.

5. The Inverse Equal-Paths Problem and the Separation Model

5.1. Motivation

In the literature on consistent aggregate rankings, few of the proposed models present a performance measure on the quality of the attained consistent ranking. One exception that relates to permutations only in *ordinal rankings* is Kemeny and Snell's model (Kemeny and Snell [46]). This model seeks an optimal group ranking that minimizes the number of reversed preferences. Therefore, if in one reviewer's permutation i is preferred to j, but in the aggregate ranking permutation j is preferred to i, then this counts as one reversal. This model, however, has a number of drawbacks: it does not accommodate partial lists, it does not differentiate between reviewers, and it is computationally prohibitive to solve (NP-hard). Although ordinal rankings play an important role in voting and elections, an added challenge is Arrow's fundamental "impossibility" theorem (Arrow [8]), stating that no voting scheme can guarantee five natural fairness properties: universal domain, transitivity, unanimity, independence with respect to irrelevant alternatives referred to here as rank reversal, and nondictatorship.

The separation model can be thought of as a *cardinal* analog of the Kemeny–Snell model. Instead of minimizing the number of reversals, it minimizes the *magnitude* of the reversals. And in further generalization, the separation model minimizes a convex *function* of the magnitude of all the reversals. Although this is a different model from that of Kemeny and Snell, it has been shown to deliver, in practice, good results on the Kemeny–Snell objective function; see Hochbaum and Moreno-Centeno [40], and Hochbaum et al. [43].

5.2. The Inverse Equal-Paths Problem

The inverse equal-paths problem (IEP) was introduced by Hochbaum [37] on a directed graph with arc weights. The problem is to modify the weights of the arcs so that all paths between each pair of nodes will be of equal lengths (total sum of weights along the path), and the penalty for the deviation from the given arcs' lengths is minimum.

The inverse equal-paths problem is within the *inverse problem paradigm*, which is as follows: Given observations and parameter values that do not conform with physical or feasibility requirements, adjust the parameter values so as to satisfy the requirements. The adjustment is made so as to minimize the cost of the adjustment in the form of penalty functions. A prominent application is to find the inverse shortest-paths that conform with the reading of the speed of seismic waves. There one seeks a minimum penalty for deviation from existing estimates on lengths of arcs so as to conform to the observation that a shortest path is of a given length, or of a given sequence of nodes. Variants of this inverse shortest paths problem were studied by Burton and Toint [13, 14], Zhang et al. [57], Ahuja and Orlin [1], Cui and Hochbaum [20], and Hochbaum [35].

The input to the *inverse equal-paths* problem is a nonsimple connected graph $G = (V, A)$ where for each pair (i,j) there is a set of arcs $R_{ij} \subset A$, so that for $r \in R_{ij}$ there is an arc of weight p^r_{ij} from i to j.

Another input is a set of penalty functions $f^r_{ij}(\,)$ for each arc $(i,j) \in A$ and $r \in R_{ij}$.

A feasible solution to the inverse equal-paths problem is a set of weights p^*_{ij} for all pairs $i, j \in V$, satisfying that for any pair of nodes $s, t \in V$ all the directed paths from s to t (and from t to s) with arc weights $\mathbf{p}^*$ are of the same length.

Arc weights are said to be *skew symmetric* if for each arc from i to j of weight p^r_{ij} there is an arc from j to i with weight $-p^r_{ij}$. Note that for any feasible solution, the set of weights must be skew symmetric. The reason is that the length of a path from a node to itself is 0, so $p_{ii} = 0$. It follows that all cycles in the graph are of length 0, and in particular, $p_{ij} + p_{ji} = 0$. Thus, in a feasible solution the weights are skew symmertic.

An arc-weight vector $\mathbf{p}^*$ is optimal if among all possible weight vectors $\mathbf{p}$ it minimizes the total sum of the penalty functions $\sum_{(i,j)\in A, r\in R_{ij}} f^r_{ij}(p_{ij} - p^r_{ij})$.

We emphasize that any feasible weight vector assigns weights to *all* possible edges (and each edge corresponds to two arcs in opposite directions and one the negative of the other). That is, the feasible vector is that of the weights of a *complete* graph.

We claim that for any feasible weight vector the triangle inequality is satisfied. To see this, notice that for a feasible weight vector $\mathbf{p}$, for any pair of nodes i, j all the paths between them are of equal length. In particular, all 2-edge paths are of equal length, and equal to the length of the 1-edge path $[i, j]$. Therefore, for any path $[i, k, j]$, $p_{ik} + p_{kj} = p_{ij}$. Therefore, the requirement that weights satisfy the triangle equality is a special case of the requirement that weights satisfy the equal paths.

The inverse is true as well. That is, if the edge weights satisfy the triangle equality, then all paths are of equal lengths, as shown in the next lemma.

Lemma 5.1. *The triangle equality is satisfied for all triplets in a graph with skew-symmetric weights if and only if all paths between every pair of nodes are of equal length.*

Proof. The discussion above demonstrated the *if* direction. The proof for the *only if* direction works by induction on the number of edges in the path between two nodes, say s and t. The triangle equality means that all paths of length 2, 2-edge paths, are of equal length to that of the 1-edge path, or the weight of the 1 edge, p_{st}. Assume that the lengths of all paths of k edges between each pair of nodes are equal to the lengths of all paths of length q between the two nodes for any $q \le k$. In particular, for $q = 1$ their length is equal to the weight of the edge between the two nodes p_{st}. We now prove that the lengths of all $k+1$-edge paths is equal to p_{st} as well. To do that, we take the first two edges in a $k+1$ path, $[s, i_1, i_2]$, and replace them by the edge $[s, i_2]$, which is necessarily of the same weight (by the inductive assumption). This "short-cutting" results in a k-edge path between s and t, which by the induction is equal in length to p_{st}.

The inductive process is demonstrated in Figure 2, where repeated short-cutting in a 6-edge path $[s, 1, 2, 3, 4, 5, t]$ is shown to be equal to the weight of the edge between s and t. In that figure, the weights on the edges correspond to the weights of the arcs from i to j for $i < j$. (In the opposite direction, from j to i, $j > i$ the weight is the negative of that quantity.) The 6-edge path is replaced by an equal-length 5-edge path $[s, 2, 3, 4, 5, t]$. □

We conclude that the two properties are *equivalent*: equal paths are equivalent to the triangle equality.

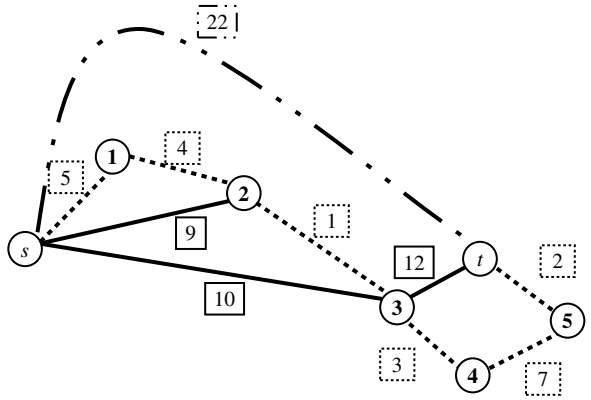
We next show that if in a graph the arc weights satisfy the triangle equality, then there exists a set of *potentials*, or node weights, w_i, so that for every i, j, $p_{ij} = w_i - w_j$.

Lemma 5.2. *For* $\mathbf{p}$ *a set of arc weights so that the graph* G *has all equal paths between each pair, if and only if there exists a set of values (potentials)* x_j*, for all* $j \in V$*, such that* $x_i - x_j = p_{ij}$*.*

Proof. If there exists a set of node potentials x_j so that each arc (i, j) weight is $x_i - x_j$, then the distance between a pair of nodes s and t along a path $[s, i_1, i_2, \ldots, i_k, t]$ is $(x_{i_1} - x_s) + (x_{i_2} - x_{i_1}) + \cdots + (x_t - x_k) = x_t - x_s$. Therefore, all paths between each pair of nodes are of equal length.

For the converse, we let $w_1 = 0$ and w_i equal to the length of any arbitrarily selected path from node 1 to node i (it does not matter which path, because they are all of equal length). Therefore, the length of the path $[i, j]$, which is the weight of this arc (in the direction from i to j) is $w_i - w_j = w_i - w_1 + w_1 - w_j$ or the length of the path $[i, 1, j]$. □

FIGURE 2. The triangle equality "short-cutting" in the proof of Lemma 5.1.



5.3. The Equivalence of the Separation Model and IEP

The input to the separation model is a collection of pairwise comparisons by reviewers in the set R. As before, we let R_{ij} be the set of reviewers who evaluated, and compared, both i and j. We then use the graph representation as in Figure 1. Here, for each intensity preference p^r_{ij} there is a corresponding arc in the opposite direction with weight $p^r_{ji} = -p^r_{ij}$. That is, the input weights are skew symmetric. The graph has to be connected or else the ranking can be determined in each connected component separately.

We seek a consistent aggregate ranking. Thus, any feasible solution is in the form of pairwise intensity preference weights that satisfy the triangle equality. From Lemma 5.1 those weights must satisfy the equal-paths property for any pair of nodes (proposals). Therefore, the requirement of equal lengths of the paths is *equivalent* to the requirement of consistency.

Let the consistent aggregate ranking's preference intensity for pair i, j be denoted by z_{ij}.

For each reviewer r in R_{ij} the difference, or distance, between the reviewer's intensity of preference p^r_{ij} and z_{ij} is $z_{ij} - p^r_{ij}$. This difference is assigned a penalty function $f^r_{ij}((z_{ij} - p^r_{ij})$, which is selected so as to reflect the confidence one has in reviewer's r assessment of the pair i, j. Therefore, for high confidence in the reviewer's assessment of this particular pair, this function will grow faster for larger values as compared to a reviewer with less confidence. The function f^r_{ij} typically takes the value 0 for an argument of 0 (0 difference or distance). It is allowed to be nonsymmetric for positive and negative arguments.

Let the penalty function for the pair (i, j) be $F_{ij}(z_{ij})$ where this function is the sum of penalties for all reviewers that assessed the pair i, j: $F_{ij}(z_{ij}) = \sum_{r \in R_{ij}} f^r_{ij}(z_{ij} - p^r_{ij})$.

A weight vector $\mathbf{p}^*$ is optimal, if among all feasible weight vectors $\mathbf{z}$ it minimizes the total sum of the penalty functions $\sum_{(i,j)\in A} F_{ij}(z_{ij})$.

The following is a valid formulation of the inverse equal-paths problem (IEP):

$$\begin{aligned} \text{(IEP)} \qquad \min \ & \sum_{i<j} F_{ij}(z_{ij}) \\ \text{subject to } & x_i - x_j = z_{ij} \quad \text{for } i < j, \\ & x_1 = 0, \\ & l_j \le x_j \le u_j \quad j = 1, \ldots, n. \end{aligned}$$

Including in the formulation a set of variables, x_j, for the node potentials, or priority weights, is redundant, but has the advantage that the properties of the problem become transparent. We use here the anchoring of $x_1 = 0$, as otherwise, for any feasible vector $\mathbf{x}$, the vector $\mathbf{x} +$ constant is feasible as well, and has the same objective value as that implied by $\mathbf{x}$. The upper- and lower-bound constraints can be generated as for $C = \max_{r,(i,j)} p^r_{ij}$, $-nC \le x_j \le nC$. We thus let $l_j = -nC$ and $u_j = nC$ for all $j = 1, \ldots, n$. In the aggregate

ranking problem it would be reasonable to require a set of weights with some finite resolution. A set of integer weights in the interval $[-n, n]$ is sufficient to guarantee that there are enough distinct ranks to assign to each node (or team). Therefore, we add the integrality requirement and replace the lower- and upper-bound constraints on $\mathbf{x}$ by

$$-n \le x_j \le n \quad x_j \text{ integer, for all } j \in V.$$

In case of rank ties, one might want to increase the resolution of the weights. This is easily obtained by scaling the bounds by $1/\epsilon$ for ϵ the required grid accuracy. Changes in values of ϵ do not affect the optimal solution greatly. Indeed, the proximity theorem of Hochbaum and Shanthikumar [42] guarantees that in this case, an optimal solution in integers for one resolution level is close enough to an optimal solution on a finer grid.

The IEP optimization problem has the *fixed-point* property. That is, if the input intensity preferences are consistent with some underlying ranking, then the optimal solution will be that underlying ranking, because the penalty functions are all zero for this ranking.

5.4. Algorithms for the Inverse Equal-Paths Problem

We note that the constraint matrix of IEP is totally unimodular because the coefficients of x_i form a matrix where each row has one 1 and one -1, and the coefficients of the z_{ij} variables form an identity matrix. Hochbaum and Shanthikumar [42] proved that minimizing an objective function that is separable convex on a totally unimodular constraint matrix is polynomial time solvable in integers, or on any ϵ-grid. Furthermore, the convex IEP is a special case of convex dual of minimum-cost network flow studied in Ahuja et al. [2].

We summarize below the complexity and algorithms for solving IEP. Here, $U = \max_j\{u_j - l_j\}$, and $T(n, m)$ is the running time required to solve the minimum s, t-cut problem on a graph with n nodes and m arcs.

1. For $F(\,)$ convex functions the problem IEP is solvable in polynomial time. An algorithm that runs in $\log U$ calls to a minimum cut procedure with complexity $O(\log U \cdot T(n^2, mn))$ is reported in Ahuja et al. [3]. Another, more efficient algorithm for this problem runs in $O(mn \log n \log nU)$; and Ahuja et al. [2]. Both these algorithms have been devised for the more general problem of the convex dual of minimum-cost network flow (DMNCF).
2. For $F_{ij}(z_{ij}) = a_{ij}^{+} \max\{z_{ij}, 0\} + a_{ij}^{-} \max\{-z_{ij}, 0\}$ (that is, $F_{ij}(\,)$ are linear for positive deviation and for negative deviation), the algorithm reported in Hochbaum [34] has complexity of $O(T(n, m) + n \log U)$, which is the best possible.
3. For $F(\,)$ arbitrary functions the problem is NP-hard—it can be shown to be only harder than the multiway cut problem, which is known to be NP-hard. This case is known more commonly as the metric labeling problem and the functions $F(\,)$ are usually δ functions equal to 0 if the argument is 0, and a positive constant otherwise. For these problems there is a large body of research on approximation algorithms; e.g., see Kleinberg and Tardos [48].

For IEP since $U = O(n)$, the run times of the polynomial algorithms for the convex case are all strongly polynomial.

6. The Ranking Graph and the Properties of the Aggregate Ranking

The ranking graph $G = (V, E)$ is the collection of pairwise comparisons for the pairs $[i, j] \in E$ that are provided as the input by the reviewers. (Recall that each pairwise comparison $[i, j]$ is converted into two skew-symmetric arcs in the directed graph set for the (IEP) problem.) As noted above, the ranking graph may be a *multigraph* indicating that there can be multiple pairwise comparisons for the same pair of candidates. We discuss here the links between the graph properties and the aggregate ranking.

If the ranking graph is simple, i.e., there is at most one edge between each pair, the problem of aggregate ranking is to find a consistent ranking that is as "similar" as possible to the input comparisons, or violates those inputs as little as possible.

If, however, the ranking graph contains at most one path between each pair of nodes, then it is an acyclic undirected graph, or a *forest*. In that case, there is a unique aggregate ranking in each connected component of the forest. However, candidates that reside in different components are not comparable because there is no input on comparing any node in one component with nodes in the other component. These concepts are illustrated in Figure 3. The ranking graph on nodes 1–12 is not connected, with the set 1–5 disconnected from the set 6–12. So the nodes $\{1,2,3,4,5\}$ are not comparable to any of the nodes $\{6,7,8,9,10,11,12\}$. Also, the set of nodes 1–12 forms a tree (an acyclic component).

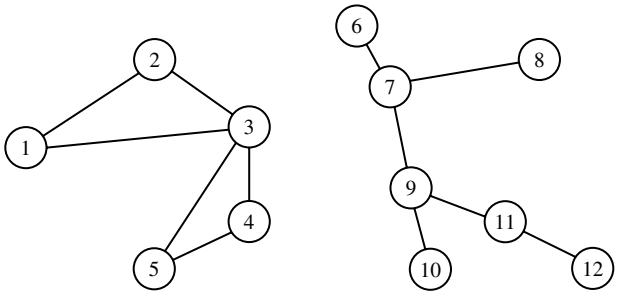
The concept of *deduced ranking* is to deduce the relative ranking of a pair of candidates indirectly from the comparisons along a sequence of pairs connecting the two respective nodes. The relative ranking of nodes i and j can be deduced, even if there is no edge between them in the ranking graph, if there is a sequence of edges $[i,i_1]$, $[i_1,i_2],\ldots,[i_k,j]$ for $k \geq 1$. The ranking of a direct pairwise comparison can be viewed as such a sequence for $k=1$. Thus, the ranking of the pair is deduced from any path in the ranking graph connecting the pair of nodes. If there are multiple paths connecting the same pair, then the deduced ranking may not be unique, and the respective ranking graph does not satisfy the equal-paths property.

The set E of input pairwise comparisons affects the quality of the ranking as follows:

1. If not connected, then there are incomparable nodes or objects.
2. If the shortest path between two nodes is too long (e.g., >5), then the deduced ranking is not robust and highly speculative. In the graph in Figure 3 the comparison between node 12 and 8, using a path of four edges, is less reliable than the comparison between, e.g., 8 and 7, which have been directly compared.
3. The greater the number of paths between each pair of nodes—the connectivity of the graph is the number k so that each pair of nodes have at least k edge-disjoint paths between them—the more robust the ranking. This is so because there are several "opinions" rendered on the pairwise comparison between the pair, increasing the confidence in the resulting comparison. In Figure 3, nodes 1 and 5 have four paths between them: $[1,2,3,4,5]$; $[1,2,3,5]$; $[1,3,5]$; $[1,3,4,5]$. (Note that only two of these paths are edge disjoint: $[1,2,3,4,5]$ and $[1,3,5]$.) Therefore, we consider the ultimate aggregate ranking comparison between these two nodes to be more reliable as compared to a ranking graph where there is only one path (say, of length 4) between the two nodes we wish to compare (e.g., between nodes 8 and 12 in Figure 3). In graph terms, we state that the higher the connectivity of the graph, the more robust the ultimate ranking.

In some cases, the form of the ranking graph E is determined in advance. This happens in scheduling of sports competitions. In NSF review panels, this is determined by the director's

FIGURE 3. A ranking graph with two components.



allocation of review tasks to reviewers. This allocation decision is going to have an impact on the robustness of the outcome. Cook et al. [19] investigated the allocation problem and proposed a heuristic algorithm to generate good feasible solutions for a version of the allocation problem. Hochbaum and Levin [39] studied the complexity of such allocation problems with various desired properties. Most of these problems are NP-hard, and for some of them there are approximation algorithms proposed in Hochbaum and Levin [39].

PageRank is a heuristic for the principal eigenvector. One major issue is that convergence cannot be attained unless the aperiodicity requirement is satisfied. Aperiodicity means that between every pair there are paths of any length l except for a finite number of paths. Therefore, if the graph is two opposing arcs between two nodes, then this requirement is not satisfied. Or, if the graph is a directed cycle (Hamiltonian), then it is strongly connected but does not satisfy the aperiodicity requirement.

7. Alternative Approaches for Aggregate Ranking with Pairwise Comparisons

We review here, in some detail, several leading aggregate ranking approaches, including those that rely on the principal eigenvector, the analytic hierarchy process (AHP), PageRank, and several optimization approaches.

7.1. The Principal Eigenvector Technique

The principal eigenvector technique has been known to apply to ranking since the 1950s. This method is reviewed, e.g., in a study by Keener [45] addressing the rankings of football teams. Consider intensity rankings that quantify by how much team i is stronger than team j by a positive number a_{ij}—a multiplicative intensity preference. (There is a great deal of research on how to determine the values of a_{ij} as a function of the score of a game, and Keener's study proposes one mapping between the score of the game and the value of a_{ij}.) Let n_i be the number of games played by team i. Then, r_i, the ultimate ranking of team i, is reasonably presumed to be proportional to the calibrated rank,

$$\frac{1}{n_i}\sum_{j=1}^{n} a_{ij} r_j.$$

Thus, $r_i = (1/\lambda)\sum_{j=1}^{n}(a_{ij}/n_i)r_j$, or $A\mathbf{r} = \lambda\mathbf{r}$ for $A = (a_{ij}/n_i)$. The solution to this system of equations is the principal eigenvector.

The Perron–Frobenius theorem states that for a nonnegative nontrivial matrix A there exists a nonnegative eigenvector $\mathbf{r}$ corresponding to a unique eigenvalue λ. If A is *irreducible*, then $\mathbf{r}$ is strictly positive, unique, and simple, and λ is the largest eigenvalue. The notion of irreducibility has an algebraic definition. We cast it as a property of the ranking graph: recall the concept of *deduced ranking* and connectivity of the graph. If the graph is connected, then there is a path between every pair of nodes and one can deduce the relative ranking of each pair, indirectly, from the outcomes of a sequence of games played. The concept of irreducibility is equivalent to having all pairs of teams comparable by deduced ranking. In graph terms this means that there is a path between each pair of nodes—namely, the graph is connected.

Some properties of the principal eigenvector method are as follows:

1. Unlike the weight-averaging algorithm, the eigenvector method takes into consideration not only the count of how many times one object is stronger than others, but also to which objects it is compared. Therefore, winning against a strong team counts more than winning against a weak one.

2. "Missing games" correspond to entries in the matrix, because the matrix must be full. The standard approach is to include such games as a draw. These draws, however, tend to skew the overall ranking. There has recently been a substantial body of research (e.g., Jagabathula and Shah [44]) looking into recovery of rankings from partial information, and seeking minimum-rank complete matrices that agree with the given partial information.

3. All games contribute uniformly to the aggregate ranking, and no subjective evaluation of a score of a game can be included. This is also a feature in the total weight-sorting algorithm used for webpage ranking or for academic citation ranking, both of which do not differentiate between citations of between pointers. Therefore, a negative citation stating that a result in a related paper is wrong counts the same as a citation referring to a paper as seminal. On webpages there are sometimes pointers that companies are buying in order to increase their webpage rank, and these pointers are often unrelated to the content of the webpage. However, principal eigenvector method, as well as other existing models, does not discriminate between citations as per their quality and significance.

4. If there are multiple games between teams, it is not clear how to measure the aggregate effect of the games that have different, and often contradictory, outcomes. In a simple example, if one team wins against the other in one game, and loses in a second game, then the often-used average counts the same as if the two teams played a game resulting in a draw, or not having played at all.

5. Solving for the principal eigenvector is equivalent to finding the roots of a polynomial (of degree n). As such, it cannot be performed in strongly polynomial time (see Renegar [49] and Hochbaum [33]).

The partial-list, or missing games, aspect is a major detriment to the use of the principal eigenvector. To quote from Fainmesser et al. [26],

> Why is there more controversy in the ranking of NCAA college football teams than there is in the ranking of other sports' teams? Unlike other sport leagues, in which the champion is either determined by a playoff system or a structure in which all teams play each other (European Soccer Leagues for example), in NCAA college football, teams typically play only 12–13 games and yet, there are 120 teams in (the premier) Division I–A NCAA college football.

That is, the explanation for the difficulty and controversy surrounding the ranking is that it is a "partial list" and the "schedule," or the allocation of the games, plays an important role. The respective graph tends to be of low connectivity. This partial-list problem is particularly exacerbated because of the use of the principal eigenvector technique, which is designed for a full-list comparison matrix.

7.2. The Markov Chain Technique

In terms of computing requirement, finding the principal eigenvector $\mathbf{w}^*$ is not practical for moderate to large values of n. Instead, it is common to compute it using the power method (Vargas [56]): For a given initial assessment of ranks $\mathbf{w}_0$ (typically, initializing with all ranks are equal to 1), this is a recursive procedure based on

$$\lim_{k\to\infty} \frac{A^k \mathbf{w}_0}{|A^k \mathbf{w}_0|} = \mathbf{w}^*. \tag{2}$$

The advantage of the use of this recursion is that it is inherently distributed and localized. Each iteration is implemented by following a walk from each node i to node j with probability $a_{ij}/\sum_{p=1}^{n} a_{ip}$. The count of the number of visits to each node after a number of iterations is an estimate of the relative rank weight. The drawbacks include several technical requirements on convergence conditions that often are not satisfied.

7.3. About AHP

The analytic hierarchy process (AHP) was developed by Saaty [51, 52] in the late 1970s, and has become a leading approach to multicriteria decision making. For this reason we sketch it briefly.

AHP relies on the principal eigenvector technique in a hierarchical manner. The decision problem is modeled as a hierarchy of criteria, subcriteria, and alternatives. The method features a decomposition of the problem to a hierarchy of simpler components, extracting experts' judgements and then synthesizing those judgements. After the hierarchy is constructed, the decision maker assesses the intensities in a pairwise comparison matrix. Thus, given n alternatives, the decision maker provides $n \times (n-1)$ pairwise comparisons that assess the relative importance of every alternative to each of the others. The backbone of the technique is the generation of the priority vector as the eigenvector of the matrix $\mathbf{A} = (a_{ij})$ at each level of the hierarchy.

7.4. The Google PageRank Algorithm

The Google PageRank algorithm is a finite approximation of the limit (2) using a small number of iterations. The following recursive formula that is used for Google PageRank can be shown to approximate in the limit the principal eigenvector of the respective matrix if $d = 0$:

$$G_i = (1-d) \sum_{(j,i) \in A} \frac{G_j}{k_j} + \frac{d}{N},$$

where N is the number of objects in the universe, G_i is the google number or *strength* of object i, k_j is the out-degree of node j, and d is a parameter.

The ranking of academic papers based on citation count has raised some interest and criticism recently (Chen et al. [17], Buchanan [12]). Citation ranking of academic papers are determined by the citation count of a paper. Setting a citation of article i to j as an arc (i, j) is a graph $G = (V, A)$ with a node corresponding to each academic paper, this is equivalent to ranking each paper by its in-degree. Chen et al. [17] and Buchanan [12] point out that the traditional citation count brings about results that are contradictory to the perceived importance of certain papers. In their study, Chen et al. give some examples. One is a 1929 paper by Slater that ranks 1,853rd in terms of citation count, although there is a universal agreement among physicists that the "Slater determinant" introduced in that paper is a fundamental concept that is considered classic, and therefore the citation count rank undervalues Slater's paper.

Chen et al. instead used the "Google PageRank Algorithm," noting that the ranking model of webpages is analogous to the academic citations model, where pointing to a webpage is equivalent to a citation. Chen et al. [17] computed the rank of Slater's paper with Google PageRank and showed it turns out 10th. This, and the improved rank of other "classic" papers, served as evidence that Google rank is a better measure of impact than the traditional citation count.

In using the PageRank algorithm, the interpretation of linking to another website or citing another paper is equivalent to a sports team "losing a game" to that site or to the other paper, which is a drawback of the principal eigenvector technique. We note that Kleinberg's HITS method (Kleinberg [47]), provides a dual but separate role to each candidate, both as a reviewer (authority) and as an object to be evaluated, which solves this issue (albeit introducing some other difficulties).

7.5. Finding "Close" Consistent Rankings with Optimization Techniques

Several approaches other than the eigenvector method have been proposed in the literature to generate a consistent matrix that is in some sense "close" to the given matrix. Most of

these are based on minimizing some measure of distance of the generated consistent matrix from the given matrix. Regression-based approaches that assume the a_{ij}s to be random variables with known distribution centered around a consistent comparison matrix have been proposed (see Hochbaum and Levin [38] for a review and formal treatment of these methods). Least squares and logarithmic least-squares regression are the most popular of these techniques, and Saaty and Vargas [54] give a comparison of these methods to the eigenvector method. Techniques based on linear programming (Chandran et al. [16], Ali et al. [6]) have also been proposed.

We show that all of these models are a special case of the separation model, IEP.

The model of Saaty and Vargas [54] employs the least-squares method to determine the values of the weights that form a consistent ranking that closely approximates a given inconsistent ranking matrix (a_{ij}). Their objective function is to minimize the proximity measured by $\sum_{i=1}^{n}\sum_{j=1}^{n}(\log a_{ij} - \log w_i + \log w_j)^2$. Replacing $\log w_i - \log w_j$ by $x_i - x_j$ or z_{ij} as in the separation model, this is the separation problem where $F_{ij}(z_{ij}) = (z_{ij} - \log a_{ij})^2$. As proved in Hochbaum and Moreno-Centeno [40], if $\log a_{ij}$ is derived from a difference of score weights, then this uniform quadratic separation problem for full list, or complete matrix, is the same as the simple weight averaging. More details on that are given in §9.

Chandran et al. [16] presented an alternative linear programming approach to the model of Saaty and Vargas for the problem of identifying weights of consistent rankings at minimum error. In their formulation the objective is to minimize the deviation error as absolute value, $\sum_{i<j}|x_i - x_j - \log a_{ij}|$, where x_i represents the logarithm of the weight of proposal i. Defining both of these optimization problems on the variables x_i representing the weight of i, and z_{ij} representing the resulting optimal additive ranking, the objective functions are $\sum_{i=1}^{n}\sum_{j=1}^{n}(\log a_{ij} - z_{ij})^2$ and $\sum_{i<j}|z_{ij} - \log a_{ij}|$, respectively. This optimization is then subject to *consistency constraints* of the form:

$$x_i - x_j = z_{ij}. \tag{3}$$

In the context of group ranking, Ali et al. [6] explored a scenario where L reviewers provide rankings only, and no proposal weights. The intensity ranking of reviewer l is given as an skew-symmetric matrix of *intensity* values (p_{ij}^l), $l = 1, \ldots, L$. The weights implied by each ranking are not given explicitly.

Ali et al. posed the chosen group intensity ranking as intensity numbers z_{ij} that satisfy for each pair $1 \le i \le n-2$ and $k = i+2, \ldots, n$, $z_{ik} = \sum_{j=1}^{k-1} z_{j,j+1}$. This latter condition is obviously equivalent to the consistency constraints with some underlying weights vector $\mathbf{x}$ and $z_{ij} = x_i - x_j$. The objective function they choose is to minimize the sum of the absolute deviation of z_{ij} from the intensity of preferences of all L reviewers, $\sum_{i<j}\sum_{l=1}^{L}|p_{ij}^l - z_{ij}|$. The formulation used by Ali et al. [6] assumes that intensity values are integers in the range $[-h, h]$. It is also implicitly assumed that individual reviewers' rankings form skew-symmetric matrices that are consistent. The formulation of the problem by Ali, Cook, and Kressis is referred to here as (ACK) (after the initials of the authors).

$$\text{(ACK)} \qquad \min \sum_{i<j}\sum_{l=1}^{L}|p_{ij}^l - z_{ij}|$$

$$\text{subject to } z_{ik} - \sum_{j=i}^{k-1} z_{j,j+1} = 0 \quad \text{for } i = 1, \ldots, n-2, \;\; i+2 \le k \le n,$$

$$1 - h \le z_{ij} \le h - 1 \quad z_{i,j} \text{ integer, for all } i, j.$$

Ali et al. showed how to solve the (ACK) problem with a linear programming routine. They noted the total unimodularity of the constraint matrix but did not make any use of this fact to achieve computational efficiency.

The separation model is in fact a generalization of (ACK). Furthermore, the models introduced in Saaty and Vargas [54], Ali et al. [6], and Chandran et al. [16] are all special cases of (IEP), as detailed next: For the problem (ACK) of finding group rankings close to the L individuals' rankings, we let each reviewer provide a ranking matrix (p_{ij}^l). This ranking matrix is assumed to be consistent; thus, there are underlying weights w_i^l corresponding to each ranking so that $(p_{ij}^l) = w_i^l - w_j^l$ and $w_1^l = 0$.

We now generate an $L \times n$ matrix where the lth column represents the complete rankings of reviewer l expressed as pairwise comparisons to proposal l. If the number of proposals is less than the number of reviewers, $L > n$, then reviewer l expresses the preferences as compared to proposal $l (\text{mod } n-1)$ where the proposals are numbered $0, 1, \ldots, n-1$. Therefore, the lth column has $a_{il} = w_i^l - w_l^l$. Now, the matrix (a_{ij}) is not consistent if the reviewers are not in full agreement, so finding overall "close" consistent rankings is equivalent to finding weights x_i so that $a_{ij} = x_i - x_j$.

To model the problem, we let the variable x_i be the weight consistent with the group ranking that is to be assigned to proposal i, and $x_1 = 0$. We normalize the values of the rankings by dividing each value of p_{ij}^l by M, for $M = \max_{i,j,l} |p_{ij}^l|$. With n proposals, integer intensities and setting $x_1 = 0$, it is thus sufficient to choose x_i as an integer in the range $[-n, n]$.

We generalize the deviation measuring objective function by using any convex function $F_{ij}(z_{ij})$. Such functions $\min \sum_{i<j} F_{ij}(z_{ij})$ include the case of the absolute deviation function of Ali et al. [6], $F_{ij}(z_{ij}) = \sum_{l=1}^{L} |w_i^l - w_j^l - z_{ij}|$. An alternative choice of $F_{ij}(\,)$ could be the quadratic convex function $\sum_{l=1}^{L} \alpha_{ij}^l (p_{ij}^l - z_{ij})^2$, where the coefficients α_{ij}^l reflect the weight, and thus the confidence, in the ranking of reviewer l for the pair ij, and replacing the term p_{ij}^l by $w_i^l - w_j^l$. If some reviewers' rankings are not necessarily consistent, then such weights cannot be assumed to exist, and an appropriate objective function depends on both p_{ij}^l and z_{ij}, such as the function $F_{ij}(z_{ij}) = \sum_{l=1}^{L} |p_{ij}^l - z_{ij}|$ of (ACK).

The problem of reaching group rankings with quadratic function penalties, as in Saaty and Vargas [54], is a special case of (IEP) where for x_i representing $\log w_i$, the quadratic objective function is, $F_{ij}(z_{ij}) = (\log a_{ij} - z_{ij})^2$. This link between the models of Saaty and Vargas and of Ali et al. has not been previously observed, and neither has the recognition of the existence of such efficient algorithms for the problem. The problem studied by Chandran et al. [16] is identical to that studied by Ali et al. [6] when one replaces the individual rankings p_{ij}^l by a column of a_{ij} in the matrix.

The optimal objective value of (IEP) provides a measure of how far a consistent ranking can be from the given inconsistent ranking according to the closeness measure deemed appropriate—the objective function. In this sense, this is a more explicit consistency index than C.I., which is not associated with any specific interpretation of distance corresponding to the C.I. value.

8. Separation-Deviation Model

Having reviewers provide both weights (scores) and comparisons seems to be redundant, because the set of weights can be translated to a comparison and vice versa. Nevertheless, reviewers might be inconsistent in their own evaluations, and submitting both weights and comparisons permits assignment of levels of confidence separately to the weights and to the pairwise rankings. This extra information can serve the role of capturing more robustly the evaluations of the reviewers than is possible with weights alone or rankings alone. We demonstrate here that this problem is linked to the *image segmentation* problem, and thus algorithms for that problem apply directly to the group ranking with weights and comparisons.

In the procedure considered here, the output of the review process consists of both weights and intensity comparisons (in the additive model). It is plausible that the comparisons of

individual reviewers will be consistent with the individual's weights, i.e., $(p_{ij}^l) = w_i^l - w_j^l$, but it is not mandatory according to our model. Also, all reviewers might be advised to anchor their rankings by setting $w_1^l = 0$, but this is not required. If any single weight is set to 0, and a difference of one level in pairwise comparison is quantified as 1, then the range for the weights is in the interval $[-n, n]$. The reviewers are permitted, however, to also use noninteger differences in ranks.

The problem is to assign both weights and comparisons so as to minimize a deviation function that has two components. One component is the *deviation cost* for the penalty of choosing a weight that deviates from the weights selected by the reviewers. The deviation cost function can take into account the confidence level in the weights assigned by individual reviewers, giving a higher penalty for deviating from higher confidence weight. The second component is the *separation costs* that determine a comparison consistent with the weights. That is, proposal i is ranked higher than proposal j if the final weight assigned is higher for i than for j. The separation cost is the cost for the final group comparison of deviating from the comparison of each of the reviewers.

The terminology of "separation" and "deviation" costs borrows from the context of the problem of image segmentation and error correction (see Hochbaum [34]), which is shown to be related to the group ranking problem. In the image segmentation setup a transmitted image is degraded by noise. The assumption is that a "correct" image tends to have areas of uniform color. The goal is to reset the values of the colors of the pixels so as to minimize the penalty for the deviation from the observed colors, and furthermore, so that the discontinuity in terms of separation of colors between adjacent pixels is as small as possible. Thus, the aim is to modify the given color values as little as possible while penalizing changes in color between neighboring pixels. The penalty function therefore has two components: the deviation cost that accounts for modifying the color assignment of each pixel, and the separation cost that penalizes pairwise discontinuities in color assignment for each pair of neighboring pixels.

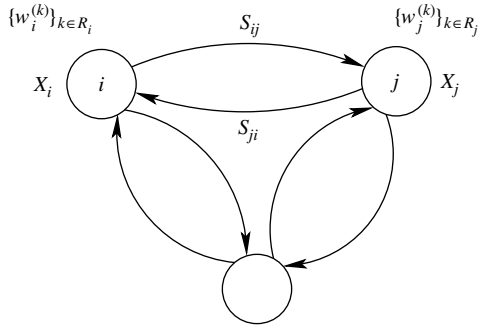
Representing the image segmentation problem, as a graph problem, we let the pixels be nodes in a graph and the pairwise neighborhood relation be indicated by edges between neighboring pixels. Each pairwise adjacency relation $\{i, j\}$ is replaced by a pair of two opposing arcs (i, j) and (j, i), each carrying a capacity representing the penalty function for the case that the color of j is greater than the color of i and vice versa. The set of directed arcs representing the adjacency (or neighborhood) relation is denoted by A. We denote the set of neighbors of i, or those nodes that have pairwise relation with i, by $N(i)$. Thus, the problem is defined on a graph $G = (V, A)$. Each node j has the observed value g_j associated with it. The problem is to assign an integer value x_j, selected from a spectrum of K colors, to each node j so as to minimize the penalty function. For g_i the color of pixel i, $G(\,)$ the *deviation cost* function, and $F(\,)$ the *separation cost* function, the problem's objective function is

$$\min_{u_i \geq x_i \geq l_i} \left\{ \sum_{i \in V} G_i(g_i, x_i) + \sum_{i \in V} \sum_{j \in N(i)} F_{ij}(x_i - x_j) \right\}.$$

The image segmentation problem is equivalent to the group ranking problem except that it is a "single value" problem, in the sense that the problem instance is given with one value for the weight (pixel color) and one specific function for the separation determined by the absolute value of the weight difference. In the group ranking problem there are multiple values assigned to each node, one for each reviewer, and multiple values assigned to each pair, one for each reviewer that has ranked the pair.

Our formalization of the group ranking problem as a graph problem is described schematically in Figure 1. Each proposal is a node in the graph, and each pairwise comparison of proposals i and j is a pair of opposing arcs between i and j. Each node i has a set of reviewers R_i that have provided weights w_i^l, $l \in R_i$. The weights for node i take values in the range

FIGURE 4. The graph for the group decision problem.



$[-n, n]$. Each pair of nodes i, j has a set of reviewers in $R_i \cap R_j$ providing relative ranking. In Figure 4 we let S_{ij} be the set of reviewers that prefer i to j and S_{ji} be the set of reviewers preferring j to i. Obviously, $S_{ij} \cup S_{ji} = R_i \cap R_j = R_{ij}$. For $r \in S_{ij}$, $p_{ij}^r \geq 0$ and for $r \in S_{ji}$, $p_{ij}^r \leq 0$. The separation penalty function for the pair i, j is $F_{ij}(z_{ij}) = \sum_{r\in R_{ij}} F^r(z_{ij} - p_{ij}^r)$.

For $z_{ij} = \max\{0, x_i - x_j\}$ and $z_{ji} = \max\{0, x_j - x_i\}$ and $G(\,)$ denoting the *deviation cost* function and $F(\,)$ denoting the *separation cost* function, then the group ranking formulation is referred to as (SD) (standing for separation deviation):

$$\text{(SD)} \qquad \min \left\{ \sum_{j\in V} G_j((w_i^l)_{l\in R_i}, x_j) + \sum_{i,j\in V} F_{ij}(z_{ij}) \right\}$$

$$\begin{aligned} \text{subject to } & x_i - x_j \leq z_{ij} \quad \text{for all } i, j, \\ & x_j - x_i \leq z_{ji} \quad \text{for all } i, j, \\ & n \geq x_j \geq -n \quad j = 1, \ldots, n, \\ & z_{ji}, \quad z_{ij} \geq 0, \ (i, j) \in E. \end{aligned}$$

Using the algorithms devised in Hochbaum [34] for the image segmentation problem, we note that the case when the functions $F_{ij}(\,)$ are linear is relevant to the group ranking, with, e.g., $F_{ij}^r(z_{ij}) = |z_{ij} - p_{ij}|$.

Theorem 8.1. (a) *If $G_j(\,)$ are convex and $F_{ij}(z_{ij}) = e_{ij} z_{ij}$ are linear, then* (SD) *is solvable in time $O(mn \log n^2/m)$.*

(b) *If $G_j(\,)$ and $F_{ij}(\,)$ are convex, then* (SD) *is solvable in strongly polynomial time, $O(mn \log n^2/m \log n)$.*

(c) *If $G_j(\,)$ are arbitrary nonlinear functions and $F_{ij}(\,)$ convex, then the problem is solved in the time required to find a minimum s, t-cut in a graph on n^2 nodes and mn^2 arcs, $O(mn^3 \log n^2/m)$.*

Proof. The solution method follows the procedures used by Hochbaum for the image segmentation problem in Hochbaum [34]. The algorithms there are stated for the range of the variables x_i in $[-U, U]$. The running times here are deduced from those by setting $n = U$.

(a) The running time of the algorithm in this case is the same as the running time required to solve the parametric minimum cut on a respective graph of same size, i.e., $O(mn \log(n^2/m))$.

(b) If both functions $G_j(\,)$ and $F_{ij}(\,)$ are convex, then the problem is an instance of the convex dual of minimum-cost network flow (DMCNF). Using the algorithm of Ahuja et al. [2] we can solve this problem in complexity $O(mn \log(n^2/m) \log n)$.

(c) If $G_j(\,)$ are arbitrary nonlinear functions and $F_{ij}(\,)$ are convex functions, then the problem is solvable by a minimum cut on a graph on n^2 nodes and mn^2 arcs,

$O(mn^3 \log n^2/m)$; see Ahuja et al. [3]. This case is of interest, e.g., when the appropriate measure is to penalize any deviation from the weight given regardless of the magnitude of the deviation.

9. Properties of the Separation and Separation-Deviation Models for Specific Penalty Functions

Hochbaum and Moreno-Centeno [40] studied the properties of the separation and separation-deviation models. These were applied to the problem of aggregating the countries' credit risk ratings by three different agencies. Because all countries were rated by all three agencies, this is a case of *full-list* ranking. Also, the inputs by the agencies were in the form of ratings, that is scores, rather than pairwise comparisons. Letting the score given to country i by agency r be w_i^r, the implied additive pairwise comparison is set to $p_{ij}^r = w_i^r - w_j^r$.

In particular, Hochbaum and Moreno-Centeno [40] consider two cases of penalty functions. One type is the absolute value function. That is, for q_{ij}^r representing the weight of the confidence in reviewer r for pairwise comparison of $[i, j]$, $f_{ij}^r = q_{ij}^r |z_{ij} - (w_i^r - w_j^r)|$. The second type is the *uniform* quadratic penalty function where $f_{ij}^r = (z_{ij} - (w_i^r - w_j^r))^2$. Here, the uniformity relates to having all penalty functions the same for all reviewers. That is, the confidence in all reviewers is the same.

There are several interesting results concerning the choice of the penalty (or distance) functions for the full-list cases described in Hochbaum and Moreno-Centeno [40]:

1. *The absolute value separation model is immune to manipulation by a minority.* Here a minority is a subset of reviewers with total confidence weights less than that of the complement of the subset. For the separation deviation model, a minority has a total confidence weight of both separation and deviation less than the respective total for the complement. The following table (Table 3) illustrates how a simple majority of two reviewers dominates the outcome aggregate ranking, even if another reviewer uses scores that are exponentially beyond their scale. The separation model is thus not sensitive to the choice of the scale, only to the ranking of the majority of reviewers.

2. The uniform quadratic separation, or separation deviation model, has $F = \sum_{i<j} F_{ij}$, where $F_{ij} = \sum_{r \in R} (z_{ij} - w_i^r + w_j^r)^2$. (The model is somewhat more general with each reviewer having his/her own weight that is applied uniformly to all terms of penalty related to that reviewer.) It was shown in Hochbaum and Moreno-Centeno [40] that the model with the uniform quadratic objective function has the same output as the average weight model. Therefore, it is, for instance, not immune to manipulation by a majority as in Table 3.

3. Although the separation model is not designed to solve the problem of minimizing reversals, the objective set up by Kemeny and Snell [46] (which the reader would recall, is NP-hard), the experimental results reported in Hochbaum and Moreno-Centeno [40] demonstrate that the number of reversals is lesser in the separation deviation model as compared to the average weight model.

Table 3. Manipulating outcome with average weight ranking.

Candidate	Reviewer 1	Reviewer 2	Reviewer 3	Average	Sep model
1	0	0	100,000,000	33,333,333.33	0
2	1	1	1,000,000	333,334.00	1
3	2	2	10,000	3,334.67	2
4	3	3	100	35.33	3
5	4	4	1	3.00	4

10. Conclusions

In this paper we demonstrate the importance of the use of pairwise comparisons for aggregate ranking. Although there are a number of models for aggregate ranking, they all can be viewed as related to either the principal eigenvector technique or to optimization. It is demonstrated here that the separation model and the separation-deviation models capture and generalize the known optimization models for aggregate ranking. For the convex penalty functions, these models are useful alternatives to the eigenvector approach and, in comparison, offer several advantages, such as flexibility in assigning the reliability of each pairwise comparison, and in the ability to use efficient combinatorial algorithms to solve the resulting aggregate ranking problem.

Acknowledgments

Research supported in part by National Science Foundation awards DMI-0620677 and CBET-0736232.

References

[1] R. K. Ahuja and J. B. Orlin. Inverse optimization. *Operations Research* 49(5):771–783, 2001.

[2] R. K. Ahuja, D. S. Hochbaum, and J. B. Orlin. Solving the convex cost integer dual of minimum cost network flow problem. *Management Science* 49(7):950–964, 2003.

[3] R. K. Ahuja, D. S. Hochbaum, and J. B. Orlin. A cut based algorithm for the nonlinear dual of the minimum cost network flow problem. *Algorithmica* 39(3):189-208, 2004.

[4] N. Ailon, M. Charikar, and A. Newman. Aggregating inconsistent information: Ranking and clustering. *Proceedings of the 37th ACM Symposium on Theory of Computing (STOC'05), Baltimore*, 684–693, 2005.

[5] J. M. Alho and J. Kangas. Analyzing uncertainties in experts' opinions of forest plan performance. *Forest Science* 43(4):521–528, 1997.

[6] I. Ali, W. D. Cook, and M. Kress. Ordinal ranking and intensity of preference: A linear programming approach. *Management Science* 32(12):1642–1647, 1986.

[7] A. Arbel and L. Vargas. Preference simulation and perference programming: Robustness issues in priority derivation. *European Journal of Operations Research* 69(2):200–209, 1993.

[8] K. J. Arrow. *Social Choice and Individual Values.* John Wiley & Sons, New York, 1963.

[9] V. Belton and T. Gear. On a short-coming of Saaty's method of analytic hierarchies. *Omega* 11(3):228–230, 1983.

[10] E. Brandstatter, G, Gigerenzer, and R. Hertwig. The priority heuristic: Making choices without trade-offs. *Psychological Review* 113(2):409–432, 2006.

[11] J. P. Brans and Ph. Vincke. A preference ranking organization method. *Management Science* 31(6):647–656, 1985.

[12] M. Buchanan. Thesis: Top rank. *Nature Physics* 2:361, 2006.

[13] D. Burton and Ph. L. Toint. On an instance of the inverse shortest paths problem. *Mathematical Programming* 53(1–3):45–61, 1992.

[14] D. Burton and Ph. L. Toint. On the use of an inverse shortest paths algorithm for recovering linearly correlated costs. *Mathematical Programming* 63(1–3):1–22, 1994.

[15] D. Cardùs, M. J. Fuhrer, A. W. Martin, and R. M. Thrall. Use of benefit-cost analysis in the peer review of proposed research. *Management Science* 28(4):439–445, 1982.

[16] B. Chandran, B. Golden, and E. Wasil. Linear programming models for estimating weights in the analytic hierarchy process. *Computers and Operations Research* 32(9):2235–2254, 2005.

[17] P. Chen, H. Xie, S. Maslov, and S. Redner. Finding scientific gems with Google. http://arvix.org/abs/physics/0604130, 2006.

[18] W. W. Cohen, R. E. Schapire, and Y. Singer. Learning to order things. *Journal of Artificial Intelligence Research* 10:243–270, 1999.

[19] W. D. Cook, B. Golany, M. Kress, M. Penn, and T. Raviv. Optimal allocation of proposals to reviewers to facilitate effective ranking. *Management Science* 51(4):655–661, 2005.

[20] T. Cui and D. S. Hochbaum. Complexity of some inverse shortest path lengths problems. *Networks* 56(1):20–29, 2010.

[21] H. A. David. *The Method of Paired Comparisons*, 2nd ed. Charles Griffin and Company Ltd., Oxford University Press, London, 1988.

[22] J. M. Duke, T. W. Ilvento, and R. A. Hyde. Public support for land preservation: Measuring relative preferences in Deleware. FREC Research Report 02-01, University of Delaware, Newark. http://www.udel.edu/frec/pubs/rr02-01.pdf, 2002.

[23] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation revisited. *Proceedings of WWW10, Hong Kong*, 613–622, 2001.

[24] G. Even, J. Naor, B. Schieber, and M. Sudan. Approximating minimum feedback sets and multicuts in directed graphs. *Algorithmica* 20(2):151–174, 1998.

[25] Expert Choice, Arlington, VA, http://www.expertchoice.com.

[26] I. Fainmesser, C. Fershtman, and N. Gandal. A consistent weighted ranking scheme with an application to NCAA College football rankings. http://www.tau.ac.il/∼gandal/college_football.pdf, March 16, 2009.

[27] E. Fernandez and R. Olemdo. An agent model based on ideas of concordance and discordance for group ranking problems. *Decision Support Systems* 39(3):429–443, 2005.

[28] J. S. Finan and W. J. Hurley. A note on a method to ensure rank-order consistency in the analytic hierarchy process. *International Transactions in Operational Research* 3(1):99–103, 1996.

[29] J. S. Finan and W. J. Hurley. The analytic hierarchy process: Can wash criteria be ignored? *Computers and Operations Research* 29(8):1025–1030, 2002.

[30] R. Fuller and Ch. Carlsson. Fuzzy multiple criteria decision making: Recent developments. *Fuzzy Sets and Systems* 78(2):139–153, 1996.

[31] B. Golden, E. Wasil, and P. Harker. *The Analystic Hierarchy Process: Applications and Studies*. Springer, Berlin, 1989.

[32] J. Hao and J. B. Orlin. A faster algorithm for finding the minimum cut in a directed graph. *Journal of Algorithms* 17(3):424–446, 1994.

[33] D. S. Hochbaum. Lower and upper bounds for the allocation problem and other nonlinear optimization problems. *Mathematics of Operations Research* 19(2):390–409, 1994.

[34] D. S. Hochbaum. An efficient algorithm for image segmentation, Markov random fields and related problems. *Journal of the ACM* 48(4):686–701, 2001.

[35] D. S. Hochbaum. The inverse shortest paths problem. Technical Report, University of California, Berkeley, Berkeley, 2002.

[36] D. S. Hochbaum. Selection, provisioning, shared fixed costs, maximum closure, and implications on algorithmic methods today. *Management Science* 50(6):709–723, 2004.

[37] D. S. Hochbaum. Ranking sports teams and the inverse equal paths problem. *Proceedings 2nd International Workshop on Internet and Network Economics* (*WINE06*), *Lecture Notes in Computer Science*, Vol. 4286. Springer, Berlin, 307–318, 2006.

[38] D. S. Hochbaum and A. Levin. Methodologies for the group rankings decision. *Management Science* 52(9):1394–1408, 2006.

[39] D. S. Hochbaum and A. Levin. The k-allocation problem and its variants. *Acta Informatica*. Forthcoming.

[40] D. S. Hochbaum and E. Moreno-Centeno. Country credit-risk rating aggregation via the separation-deviation model. *Optimization Methods and Software* (*OMS*) 23(5):741–762, 2008.

[41] D. S. Hochbaum and E. Moreno-Centeno. Deciding the winners in a student paper competition. Technical report, University of California, Berkeley, Berkeley, 2009.

[42] D. S. Hochbaum and J. G. Shanthikumar. Convex separable optimization is not much harder than linear optimization. *Journal of the ACM* 37(4):843–862, 1990.

[43] D. S. Hochbaum, E. Moreno-Centeno, P. Yelland, and R. A. Catena. A new model for behavioral customer segmentation. Technical report, University of California, Berkeley, Berkeley, 2009.

[44] S. Jagabathula and D. Shah. Inferring rankings using constrained sensing. http://arxiv.org/abs/0910.0063, 2009.

[45] J. P. Keener. The Perron-Frobenius theorem and the rating of football teams. *SIAM Review* 35(1):80–93, 1993.

[46] J. G. Kemeny and L. J. Snell. Preference ranking: An axiomatic approach. *Mathematical Models in the Social Sciences*. Ginn, Boston, 9–23, 1962.

[47] J. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM* 46(5):604–632, 1999.

[48] J. Kleinberg and E. Tardos. Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science, New York*, 14–23, 1999.

[49] J. Renegar. On the worst case arithmetic complexity of approximation zeroes of polynomials. *Journal of Complexity* 3(2):90–113, 1987.

[50] B. Roy. *Multicriteria Methodology for Decision Aiding*. Kluwer Academic Publishing, Dordrecht, The Netherlands, 1996.

[51] T. L. Saaty. A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology* 15(3(:234–281, 1977.

[52] T. L. Saaty. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. McGraw-Hill, New York, 1980.

[53] T. L. Saaty. Rank generation, preservation, and reversal in the analytic hierarchy decision process. *Decision Sciences* 18(2):157–177, 1987.

[54] T. Saaty and L. Vargas. Comparison of eigenvalue, logarithmic least squares and least squares methods in estimating ratios. *Journal of Mathematical Modeling* 5(5):309–324, 1984.

[55] T. L. Saaty and L. G. Vargas. Diagnosis with dependent symptoms: Bayes theorem and the analytic hierarchy process. *Operations Research* 46(4):491–502, 1998.

[56] R. S. Vargas. *Matrix Iterative Analysis*. Prentice Hall, Englewood Cliffs, NJ, 1962.

[57] J. Zhang, Z. Ma, and C. Yang. A column generation method for inverse shortest path problems. *ZOR–Mathematical Methods of Operations Research* 41(3):347–358, 1995.