MANAGEMENT SCIENCE
Vol. 52, No. 9, September 2006, pp. 1394–1408
ISSN 0025-1909 | EISSN 1526-5501 | 06 | 5209 | 1394

informs®

DOI 10.1287/mnsc.1060.0540

Methodologies and Algorithms for Group-Rankings Decision

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

Asaf Levin
Department of Statistics, The Hebrew University, Jerusalem 91905, Israel, levinas@mscc.huji.ac.il

The problem of group ranking, also known as rank aggregation, has been studied in contexts varying from sports, to multicriteria decision making, to machine learning, to ranking Web pages, and to behavioral issues. The dynamics of the group aggregation of individual decisions has been a subject of central importance in decision theory. We present here a new paradigm using an optimization framework that addresses major shortcomings that exist in current models of group ranking. Moreover, the framework provides a specific performance measure for the quality of the aggregate ranking as per its deviations from the individual decision-makers' rankings.

The new model for the group-ranking problem presented here is based on rankings provided with intensity—that is, the degree of preference is quantified. The model allows for flexibility in decision protocols and can take into consideration imprecise beliefs, less than full confidence in some of the rankings, and differentiating between the expertise of the reviewers. Our approach relaxes frequently made assumptions of: certain beliefs in pairwise rankings; homogeneity implying equal expertise of all decision makers with respect to all evaluations; and full list requirement according to which each decision maker evaluates and ranks all objects. The option of preserving the ranks in certain subsets is also addressed in the model here. Significantly, our model is a natural extension and generalization of existing models, yet it is solvable in polynomial time. The group-rankings models are linked to network flow techniques.

Key words: network flow; group ranking; decision making
History: Accepted by Wallace J. Hopp, optimization and modeling; received January 5, 2005. This paper was with the authors 2 months for 2 revisions.

1. Introduction

This paper presents a framework, models, and algorithms for the group decision making in the process of ranking and comparing a list of projects. The essence of the group-ranking problem is the task of consolidating and aggregating the individuals' rankings so as to obtain a group ranking that is in some sense representative of the individuals' rankings. The problem has been studied extensively, and has appeared under many guises.

One application that motivated our study is the National Science Foundation (NSF) process of evaluation and ranking of proposals. Recently, Cook et al. (2005) investigated the allocation of the ranking tasks to individuals in the context of the NSF process. The NSF group-ranking problem is similar to that of ranking athletes whose performance has been achieved in different competitions, different heats, or different weather conditions. Another well-known similar example is that of ranking students based on their grade point average. Each instructor evaluates a proper subset of the universal set of students. 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 grade point average (GPA) is the methodology used to rank the universal set of students by all reviewers—the instructors. GPA is an example of using what we call here the *weights-only* model of group ranking.

Some types of group rankings are based on pairwise rankings only. Pairwise rankings are the typical input in sports competitions such as tennis, football, baseball, etc. The group-ranking problem here is to determine an overall aggregate ranking of all teams, or players, based on the collections of pairwise rankings that are generated from the outcomes of pair competitions or games.

The individuals in the group that provide their assessments are referred to here as *reviewers*. The objects evaluated are called *projects*. The collection of all projects 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 sometimes referred to also as *cardinal* versus *ordinal* preferences.

One feature of the NSF group-ranking problem, and the students' ranking, is that each reviewer provides a *partial list* that evaluates and compares only a proper subset of the proposals. A great deal of the literature, however, is concerned with models that assume a full list for each reviewer. That is, each reviewer is assumed to compare and submit a ranking of the entire universal set. Those models that assume full lists rankings cannot be used in our context of group-ranking problems.

The modelling of the group-ranking problem depends on the format of the evaluations provided by the set of reviewers. We classify the possible modes of format into four cases.

1. Reviewers submit a list of *weights only* for all projects reviewed.
2. Reviewers submit the *preference rankings only* of pairs of projects reviewed without assigning weights. In this variant there is no implied information about the *intensity* of each ranking.
3. Reviewers submit *intensity rankings only* that include a measure of the preference of one project to the other in each ranked pair.
4. Reviewers submit both *weights and rankings*.

The most common procedure of group ranking is the *weights-only* process, in which the evaluation of each reviewer consists of giving a weight/score to each project reviewed. In the weights-only model the reviewers do not express any preference, except for what is implicit in the weights. That is, the larger the weight, the more preferred the project, so the ordering by weight is the ranking of the reviewer and the ratio, or the difference between the weights assigned to i and j, w_i/w_j or $w_i - w_j$, reflects the intensity of the pairwise comparison of the pair. The popularity of the weights-only model is attributed to its simplicity and the algorithmic ease of reaching a group decision—the group ranking is based on ordering according to the average weight. As we discuss later, the weights-only process suffers from shortcomings and biases in the aggregate ranking that results.

In NSF review panels, the group ranking is based on ordering of the simple averages of the submitted weights. In some scientific conferences' program committees the reviewers provide, in addition to each weight, a confidence factor in their own evaluation (this is a common practice in computer science conferences, e.g., STOC, FOCS, and SODA). The group weight is then the weighted average that weighs more heavily the higher the confidence evaluations. While this procedure refines the aggregate ranking by taking the confidence levels into consideration, the approach still does not take into account the ordering of each reviewer in the form of the implied rankings of individual reviewers. Further, the weights provided by each reviewer, although reflective of individual reviewers' rankings, have their *magnitude* chosen on a subjective scale. It is possible that many, or even most, of the rankings of the reviewers implicit in the weights will be reversed in this process. Such an adverse case is shown in Example 4.1 in §4.

The preference rankings-only procedure presents its own set of challenges, as demonstrated, for instance, in Arrow's theorem (1963)—more on that below. The Kemeny-Snell optimal ordering of minimizing the deviation of the aggregate ranking in terms of reversing the least number of preferences is burdened by computational difficulties—it is an NP-hard optimization problem (see §5). We state for this problem a simple $\frac{1}{2}$-approximation algorithm and note its connection to the minimum arc feedback set problem. We also show that if a classification of the reviewed projects as preferred and nonpreferred is sufficient, then the Kemeny-Snell optimal ordering is achieved in polynomial time.

Even with the intensity rankings procedures, with or without weights, there has not been a unified approach to date. Such a unified approach is one of the central contributions of this paper. We argue here that intensity rankings are the preferred mode of output (with or without accompanying weights), overcoming many of the drawbacks in approaches used in group rankings to date.

1.1. Relevant Research

The important subject of voting and elections inspired extensive studies on group ranking with preference rankings. A prominent "impossibility" result in this area is Arrow's (1963) fundamental theorem proving 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. Kemeny and Snell (1962) proposed an axiomatic approach for dealing with preference ranking that models the problem as minimizing the deviation from individual rankings defined by the *distance* between two complete rankings. Bartholdi et al. (1989) considered computational issues related to this model in the context of elections and voting.

One study by Keener (1993) addresses the rankings of football teams. Keener discussed the trade-off between using preference rankings and intensity rankings, and the important role of the Perron-Frobenius theorem regarding the conditions that guarantee a positive unique solution eigenvector $\mathbf{r}$ to the system $\mathbf{Ar} = \lambda\mathbf{r}$. To understand the link of group ranking to this linear system, consider intensity rankings that quantify how much project i (team i in this context) is preferred to project j by a positive number a_{ij}, which is greater than 1 if i is preferred to j, and less

than 1 otherwise. Now the rank of i is proportional to the strength of its ranking calibrated by the ranks of the projects that it is compared to, $\sum_j a_{ij} r_j$, because being preferred to a highly ranked team obviously contributes more to the strength of a team's own rank than being preferred to a lesser-ranked team. The rank of each project is then reasonably presumed to be proportional to this calibrated rank, and thus

$$\lambda r_i = \sum_j a_{ij} r_j. \tag{1}$$

The solution to this system of equations plays an important role in the analytic hierarchical process, as discussed in §2.2.

For intensity rankings, Duke et al. (2002) studied public support for various aspects and tradeoffs relating to land preservation. Alho and Kangas (1997) studied forest-planning performance by a group, and Cardùs et al. (1982) presented an application of group decision making for development of a research program for the National Institute of Handicapped Research and for the Rehabilitation Services Administration. In the late 1970s, Saaty (1977, 1980) developed the analytic hierarchy process (AHP), which became a leading approach to multicriteria decision making. That technique has also been used in applications requiring group rankings that use intensity rankings. Over the last 25 years, the AHP has been applied in more than 30 diverse areas to rank, select, evaluate, and benchmark decision alternatives. Good surveys on this subject are available in Saaty and Vargas (1998), and Golden et al. (1989). The AHP and many other methods used in group-ranking work on consolidating full-ranking lists, that is, each reviewer would have to review and rank *all* projects, which limits the applicability of the technique. The AHP's core technique relies on finding an eigenvector, as per the Perron-Frobenius theorem, that serves as the vector of weights.

The recognition that preferences might be expressed with only *limited* certainty was identified early on by Brans and Vincke (1985), and later by Fuller and Carlsson (1996). These models address decision makers that express vaguely defined preferences due to imprecise beliefs or conflicts and competing aspirations. To date, only heuristic methodology has been applied to those *fuzzy preference* models. Recently, fuzzy models were addressed by Fernandez and Olemdo (2005), who proposed an evolutionary algorithm.

We also consider here multicriteria decision making, which is typically studied separately from the group-ranking problem. We demonstrate that there is a modelling overlap between the problems of multicriteria decision making and aggregate ranking, although these two subjects are often pursued separately and are considered distinct. One good survey of multicriteria methodology is provided by Roy (1996).

1.2. Contributions

The results presented here include a framework that unifies several streams of research and offers an integrated approach for the group-ranking problem and multicriteria decision making. We clarify the links between these problems and between the different approaches and models that have been used to address the aggregate planning problem.

We study properties of the weights-only approach and the preference-ranking approach. For the first one, we demonstrate the drawbacks in representing the opinions of the reviewers in the final aggregate ranking. For the second, we review the Kemeny-Snell model of looking for a *close* solution in terms of minimum reversals of reviewers' rankings. Although this optimization problem is shown to be NP-hard, we give a simple approximation algorithm, and show that the 2-rank special case is polynomially solvable. Further, we generalize this closeness optimization for the intensity rankings, and intensity rankings and weights output from reviewers, and show how it generalizes work done by others.

The advantages of the intensity rankings deviation model are in the degree of flexibility and control it offers the decision makers and the existence of a polynomial-time algorithm for reaching a group ranking that is optimal according to a well-defined and transparent performance measure.

In our framework, we relax the assumptions made in many existing approaches, and overcome common shortcomings, including:

- The lack of performance measure that assesses the quality of one aggregate ranking as compared to another. Our model allows us to articulate precisely the definition of optimality of an aggregate ranking.
- The homogeneity assumption frequently made is that all reviewers contribute equally to the group decision. We allow differentiation between reviewers according to their expertise, and according to their expertise in specific projects and specific pairwise comparisons.
- Many existing group-ranking models require the provision of full lists by the reviewers. The model proposed here allows partial lists.
- The assumption of certainty in rankings submitted by reviewers is relaxed here. We allow preferences to be expressed with specified confidence level, or limited belief, thus providing solutions (in polynomial time) to fuzzy models as well.
- The phenomenon of rank reversal (see §3) in the presence of irrelevant alternatives was pointed out as a

shortcoming of voting schemes by Arrow. Even methods using intensity rankings such as AHP do not have a built-in mechanism for *rank preservation/retention*. The model proposed here has such a mechanism to protect a partial ordering while adding new projects to the set to be evaluated.

• Our model applies to reconciling contradictions in multicriteria decision making as well as in group ranking.

• Unlike other models, the model here is computationally efficient and can be applied to the large-scale rankings effort of an aggregate ranking that is best according to some specified performance measure. We further make a nontrivial link between group-ranking problems and flow-and-graph problems.

The procedure developed here is to be used as a normative approach to generate aggregate ranking of a group for given individual rankings. It also has, however, the potential of being used as a descriptive model in order to explain how economic behavior of a group is generated from the utility functions of individuals. Namely, by an inverse problem paradigm, one can seek the values of confidences and intensities that can explain the outcome aggregate ranking.

The presence of partial lists raises a related question: how to allocate the ranking tasks to individual reviewers when the lists are *partial*. If each reviewer can evaluate only a subset of the universal set, then it is possible that some pairs are not comparable, even within the implied ranking. (For definitions of implied ranking and consistency closure, see §2.) In that case, the group ranking will not be a full order. Cook et al. (2005) address this issue in their paper in the NSF context: Each reviewer is to be assigned a partial list of no more than k projects contained in their set of expertise. In order to achieve *comparability* and full order, Cook et al. assumed that all possible pairs must be reviewed and compared by at least one reviewer. The problem then is to pick a subset for each reviewer of size no larger than k so the union of all pair reviews of all the reviewers covers the entire $\binom{n}{2}$ pairs. We call this problem the k-ALLOCATION PROBLEM. Cook et al. proposed an integer programming formulation of the k-ALLOCATION PROBLEM and proposed a heuristic algorithm to generate good feasible solutions. In a companion paper, Hochbaum and Levin (2006), we address the k-ALLOCATION PROBLEM and show that it is polynomial for $k \leq 2$ and NP-complete for $k \geq 3$. We provide several approximation algorithms and discuss the trade-off between their performance and ease of application.

1.3. Outline

We begin by analyzing a fundamental property of a ranking, *consistency*, defined in §2. We consider how transitivity is the analog of consistency for ordinal ranks and formalize the graph interpretations of both transitivity and consistency. This section includes the definition of the notions of implied ranking and consistency closure and a formalization of the group-ranking problem as a graph problem.

The phenomenon of rank reversal violating ordering or projects in the presence of irrelevant alternatives is reviewed in §3. This phenomenon is illustrated there for AHP, which is a dominant approach using intensity rankings. The weights-only procedure and its potential shortcomings are described in §4. The preferences ordinal model is addressed in §5. We describe the Kemeny-Snell optimal ordering as a minimum deviation function and demonstrate the NP-hardness of finding the minimum deviation for ordinal rankings. We discuss a $\frac{1}{2}$-approximation algorithm for the problem and show that a special case of 2-ranking is solvable in polynomial time. The intensity-ranking model and a discussion on the equivalence of the multicriteria and aggregate ranking problems are presented in §6. We also present in that section the minimum deviation model for cardinal/intensity rankings, their link to the flow problem, and polynomial-time algorithms. In §7 we present a model that includes both weights and ranking for robustness purposes and point out an interesting analogy to the image segmentation problem. The formulation and solutions under different conditions are presented. We conclude with final remarks and future directions for research in §8.

2. The Consistency of Rankings

2.1. Ordinal Rankings—Transitivity

An essential property of a ranking is that 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 $j \succeq k \Rightarrow i \succeq k$.

We formalize the rankings of a set of projects $V = \{1, \ldots, n\}$ as a directed graph $G = (V, A)$, with a set of nodes V and a set of arcs A so that the ordered pair $(i, j) \in A$ if $i \succeq j$. The transitivity of the preference order is equivalent to the property of *acyclicity* of the corresponding directed graph $G = (V, A)$. A graph is said to be acyclic if it does not contain a directed cycle. It is well known that acyclic directed graphs admit a *topological ordering*, which is an assignment of distinct indices from $\{1, \ldots, n\}$ to the n nodes (representing the projects) so that for every arc (i, j) in the graph $i > j$. The values of the indices of the topological ordering can serve as the underlying weights of the respective objects. The topological ordering of an acyclic graph can be found in linear time.

When the rankings submitted are partial lists the graph may not be complete—that is, there might be

some arcs missing between pairs of nodes. The property of transitivity allows us, however, to identify implied rankings. To do that we consider the *transitive closure* of the graph. To obtain the transitive closure, we find, for each node i, the set of nodes $R(i)$ reachable from i along a directed path. Therefore, $j \in R(i)$ if there exists at least one sequence of arcs $(i, i_1), (i_1, i_2), \ldots, (i_k, j)$. Due to the transitivity property, we can add arc (i, j) to the graph if $j \in R(i)$. If the arc (j, i) is already present, then the order is not transitive.

It should be noted, however, that acyclic graphs whose transitive closure is not complete (i.e., the graphs do not have an arc, or a directed path, between each pair of nodes) do not represent a *full order*. If the order is partial, then some pairs of projects may be incomparable, and the topological ordering is not unique. An example of this phenomenon is demonstrated in Figure 1, showing an acyclic graph with node indices that form a topological ordering (i.e., if (i, j) is an arc, then the index of i is smaller than the index of j). An alternative topological ordering numbers Node 3 before Node 2, as these two nodes are incomparable.

2.2. The Consistency of Cardinal Rankings

Intensity rankings provide a 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 projects. The multiplicative intensity preference represents the *ratio* of the strengths of the ranks of the two projects compared.

Let the intensity rankings be given in the form of a matrix $\mathbf{A} = (a_{ij})$, where a_{ij} is the magnitude of preference of project i to j. The matrix is *complete* if all entries are present and generated from full lists, and *partial* otherwise.

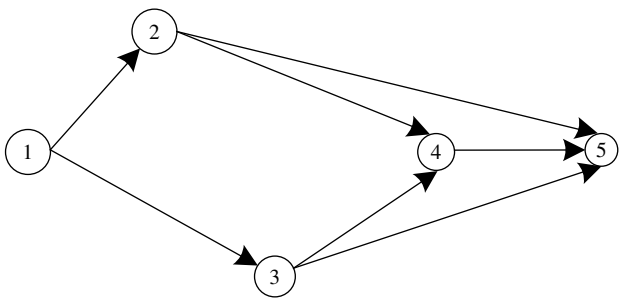
The notion of consistency in a multiplicative sense, used, e.g., by Saaty (1977, 1980) for complete matrices, is that for each triple i, j, k, $a_{ij} \cdot a_{jk} = a_{ik}$. This is equivalent to the existence of a set of weights w_i for $i = 1, \ldots, n$ so that $a_{ij} = w_i/w_j$. Such a set of weights, 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 second definition of consistency is in the additive sense, (e.g., Ali et al. 1986), where for each triplet i, j, k, $a_{ij} + a_{jk} = a_{ik}$. In this case there is a set of weights w_i for $i = 1, \ldots, n$ so that $a_{ij} = w_i - w_j$. Again, the vector of weights is not unique, because one can add a constant and get the same set of differences. Here we can anchor the set uniquely by setting $w_1 = 0$. Both definitions of additive and multiplicative consistency are equivalent because the logarithms of the a_{ij}s that are consistent in the multiplicative sense yield rankings that are consistent in the additive sense, and vice versa. Therefore, any discussion of one type of consistency applies to the other.

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 a 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 Equation (1) 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

Figure 1 A Partial-Order Graph for the Ordinal Ranking Example



there is only a single sequence of rankings comparing the two, then the matrix resulting from the completion process is necessarily consistent.

The intensity rankings are represented on a graph $G = (V, A)$ with a set of nodes V, one for each project, and a pair of *weighted* arcs for each pairwise ranking with a_{ij} the intensity of the ranking of i to j and $-a_{ij}$ the intensity of ranking of j to i. The graph can be a "multigraph" in the sense that there is more than one arc between i and j. For each missing arc (i, j) in the graph, the consistent closure is formed by adding that arc with a weight equal to the "length" of some path from node i to node j. The length is the sum of the additive intensities along the path. If there is more than one path between two nodes and the sum of the intensities along the different paths is not equal, then the ranking is not consistent. We thus proved,

Theorem 2.1. *In a consistent ranking represented on a graph, the lengths of all directed paths between each pair of nodes is the same.*

This theorem allows us to present the problem of group ranking as a graph problem of *inverse paths*, discussed in §6.3.

How to measure the extent of the inconsistency of the joint rankings of the reviewers is a question that lies at the heart of the group-ranking problem. One such measure of the extent of inconsistency is part of the AHP. AHP was developed by Saaty (1977, 1980) in the late 1970s, and has become a leading approach to multicriteria decision making. For this reason we sketch it briefly, along with Saaty's associated measure of inconsistency.

In AHP, 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. An important backbone of the technique is the generation of the priority vector as the eigenvector of the matrix $\mathbf{A} = (a_{ij})$. Suppose the matrix is consistent and the vector of weights is $\mathbf{w} = (w_i)_{i=1}^n$. Then $a_{ij} = w_i/w_j$. Summing up over all j, we obtain $\sum_{j=1}^n a_{ij} w_j = n w_i$. Therefore, in matrix notation the vector of weights $\mathbf{w}$ satisfies $\mathbf{Aw} = n\mathbf{w}$. This vector of weights is, hence, the eigenvector that consists of the weights assigned to each project or each criterion under the multiplicative model, but only if the matrix is consistent. Otherwise, the eigenvector forms some approximation of the preference weights. The measure of approximation for a skew-symmetric inconsistent matrix defined by Saaty (1980) is the *consistency index* (*C.I.*),

$$C.I. = \frac{\lambda_{\max} - n}{n-1}, \tag{2}$$

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, matrices that satisfy $a_{ij} = 1/a_{ji}$ for all $i < j$.

While the consistency index is zero 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.

3. Rank Reversal

The lack of robustness of AHP, along with other weight-generation methods, is manifested in the phenomenon of rank reversal, by which adding an inconsequential alternative can change the order of weights of the top-ranking preferences. This issue is discussed by Belton and Gear (1983), Saaty (1987), and Finan and Hurley (1996, 2002). Rank reversal is illustrated in the following example.

Consider the (skew-symmetric) multiplicative comparison matrix (only the upper triangle matrix is shown):

1	1.2	1.2
	1	1.5
		1

Solving the above problem using Expert Choice (2005), a commercial tool for computing weights using the eigenvector method, yields the priority vector $(0.373, 0.356, 0.271)$. We now add a project that is nearly redundant (all the alternatives are clearly better than this one).

1	1.2	1.2	6
	1	1.5	8
		1	7
			1

The priority vector corresponding to this matrix is $(0.335, 0.347, 0.273, 0.046)$. Notice that in the first vector Project 1 has a greater weight than Project 2, while in the second vector Project 2 has greater weight than Project 1. Thus, the addition of a near-redundant alternative has changed the ranking of projects. Part of the

problem driving the rank reversal phenomenon in the priority vector is that the eigenvector methodology cannot capture the information that Project 4 is inconsequential and should affect the ranking less than the other three projects.

In order to prevent rank reversal when an alternative is added to the universal set, the group ranking should be generated while satisfying *rank preservation*. We discuss how rank preservation can be executed in our model.

4. Weights Only

In the model where reviewers submit only weights for the projects reviewed, the reviewers do not express any preference except for what is implicit in the weights. That is, the larger the weight, the more preferred the project. As noted in the introduction, this model is the one used by the NSF and by program committees of conferences. Each reviewer submits a list of weights, which corresponds to a ranking, for the projects evaluated. In the panels of NSF, the final weight is a simple average of the submitted weight. In many program committees the reviewers include, in addition to each weight, a confidence factor in their own evaluation. The group weight then is the weighted average that weighs more heavily the higher the confidence weights. The drawback of this approach is that it does not take into account the preferences in the form of the implied rankings of individual reviewers. As we show in the next (pathological) example, it is possible that most implied reviewers' rankings will be reversed in this process.

Example 4.1. Consider the procedure of taking average weights for n projects labeled $\{0, 1, 2, \ldots, n-1\}$. There are $m+1$ reviewers, and each of them reviews the n projects. Reviewer i (for $i = 1, 2, \ldots, m$) assigns project j the weight j. The last reviewer assigns project j the weight $(mn+1)\cdot(n-j)$. Each of the first m reviewers prefers the lower-labeled projects to the higher-labeled projects. However, if we take the average weights of all the reviewers, then the overall ranking based on the average weights will be exactly the opposite. This shows that one reviewer can dominate all other reviewers.

Another well-known context in which such an adverse outcome is possible is in measuring and ranking the performance of students based on their grade point average. Some instructors tend to assign higher grades than others (and naturally become more popular). Although the relative ranking of each instructor reflects the instructor's evaluation of the students and their respective rank in class, the grade point average can bias the ranking in favor of students who took courses with instructors using a more generous grading scale. This illustrates that the weights-only approach has biases in reflecting the evaluations of the reviewers in the group.

We remark that a weights-only procedure is not guaranteed to produce a permutation rankings where each rank position is unique.

5. Preference Rankings Only

Kemeny and Snell (1962) studied the group-ranking problem with preference rankings only. Their model attempted to overcome the difficulties posed by Arrow's theorem. Their model's goal is to minimize the number of reviewers' rankings reversed.

We quantify the preference rankings for the purpose of formulating the optimization model as $p^l_{ij} = 1$ if reviewer l prefers i to j and 0 otherwise, in which case $p^l_{ji} = 1$. Let S_{ij} be the set of reviewers who prefer i to j and S_{ji} be the set of reviewers preferring j to i. Therefore, for z_{ij} a binary variable indicating the group ranking, the objective of reversing the least number of reviewers' rankings is, $\min \sum_{i<j} F_{ij}(z_{ij})$ where,

$$F_{ij}(z_{ij}) = |S_{ji}| \cdot z_{ij} + |S_{ij}| \cdot z_{ji}. \tag{3}$$

In other words, this function assigns the group consensus of i preferred to j, a penalty proportional to the number of reviewers that chose an opposite ranking. We call this problem the *minimum rankings reversal problem*. Unfortunately, this problem is in general NP-hard:

Theorem 5.1. *The minimum rankings reversal problem is NP-hard.*

Proof. The problem of choosing z_{ij} so as to minimize the function $\sum_{j=1}^{n} \sum_{i=j+1}^{n} F_{ij}(z_{ij})$ is an instance of the minimum arc feedback set problem (see Karp 1972). That problem is defined on an arc weighted complete directed graph $G = (V, A)$, where for each pair of nodes $i, j \in V$, $(i, j), (j, i) \in A$. The problem is to delete a minimum weight subset of the arcs so that the remaining graph is acyclic. The problem is known to be NP-hard in the strong sense.

We let the weight of arc (i, j) be $|S_{ij}|$, the number of reviewers that chose the ranking of i preferred to j. The acyclicity of the generated graph created by deleting a minimum weight subset of arcs is the property equivalent to transitivity of preference rankings because an acyclic graph admits a topological order. That is, an assignment of values from 1 to n so that each arc (i, j) indicating that i is preferred to j has the respective indices to satisfy $w_i > w_j$. Those values can serve as the weights of the respective nodes. □

A special case of the minimum rankings reversal problem is the 2-rank problem. For the 2-rank problem the objective is to assign two rank levels only, so as to create a ranking of one subset of projects above

another subset while minimizing rankings reversals. This special case is polynomially solvable:

THEOREM 5.2. *The 2-rank problem is solvable in polynomial time, $O(mn\log(n^2/m))$.*

PROOF. Construct the graph $G = (V, A)$ as in the proof of Theorem 5.1. We are seeking a partition of the set of nodes V to V_1, V_2, so that the nodes in V_1 have a higher rank than those in V_2. The objective of minimizing rank reversals is $\sum_{(i,j)\in(V_1,V_2)} |S_{ji}|$.

The directed minimum 2-cut problem is to find a bipartition of the set of nodes, $V = V_1 \cup V_2$ that minimize the capacity of the cut from V_1 to V_2, $C(V_1, V_2) = \sum_{(i,j)\in(V_1,V_2)} w_{ij}$ for w_{ij} the weight of arc (i, j). The 2-rank problem is thus the same as the problem of finding a directed minimum 2-cut.

The most efficient algorithm for the directed minimum 2-cut problem is by Hao and Orlin (1994), with a complexity of $O(mn\log(n^2/m))$ for a graph on $|V| = n$ nodes and $|A| = m$ arcs. □

For the minimum arc-feedback problem there are no constant factor approximation algorithms known. The best-known approximation algorithm for the minimization problem was devised by Seymour (1995), and it provides a ratio of the algorithm solution divided by the optimum, which is not exceeding $O(\log n \log\log n)$. That algorithm was further refined by Even et al. in 1998.

In a maximization form, the problem is to retain a set of arcs of maximum weight so that the retained arcs form an acyclic graph—that is, so as to maximize the weight of rankings that match the reviewers' rankings. We call this problem the *maximum rankings nonreversal problem*. There is a known easy approximation algorithm for this problem (attributed to "folklore;" see p. 361 in Hochbaum 1997).

LEMMA 5.1. *There is a linear time $\frac{1}{2}$-approximation algorithm for the maximum rankings nonreversal problem.*

PROOF. We assign to each node a unique and arbitrary index (node weight) from $\{1, \ldots, n\}$ and compare the total weight of the set of arcs (i, j) with $i < j$ to the total weight of the set of arcs with $i > j$. Removing the smaller weight set retains at least half of the total sum of weights of all arcs in the graph, and thus it is a $\frac{1}{2}$-approximation. □

Cohen et al. (1999) devised a greedylike algorithm for this problem, which they proved, using a rather elaborate proof, that is a $\frac{1}{2}$-approximation. They were apparently not aware of this easily provable and known result.

This approximation algorithm can also incorporate rank preservation for a subset. For that subset the induced ranking has topological order associated with it that provides weights consistent with the *preserved* subset. For all these, the weight of each arc is set to ∞ so as to make sure it is not reversed. We note that there is no better than $\frac{1}{2}$-approximation known for the maximization version of the arc feedback set problem.

Another nice property of this acyclic subgraph approximation algorithm is that it is used when all possible pairwise comparison arcs are present—the graph is a complete directed graph. Therefore, when the subgraph is selected, the indexing chosen corresponds to a full order ranking without ties—a permutation.

A reasonable rounding heuristic often used in group decision rankings is to have each reviewer rank a subset of the projects by assigning integer-valued weights to those projects. The group decision is made by adding the integer values for each project and ranking them according to this aggregate sum (from larger to smaller, say). We show that this frequently used heuristic can have arbitrarily bad performance. We call this the *rank-sum* heuristic. Assuming that each reviewer is reviewing exactly two projects, then without loss of generality, we can assume that the weights that he/she assigns are 0/1, with 1 assigned to the higher-ranked project. Therefore, we can model these preferences as a directed graph G over the nodes $\{1, 2, \ldots, n\}$ with an arc (i, j) for each reviewer that prefers i to j. Therefore, the weight assigned to a project i according to the rank-sum heuristic equals its out-degree in G. Note that the rank-sum heuristic is applied in trying to solve an instance of the maximum rankings nonreversal problem. We next show that this heuristic does not provide any constant approximation ratio.

REMARK 5.1. The rank-sum heuristic does not provide any constant factor performance guarantee.

PROOF. Consider the following counterexample on a ranking graph with the nodes $\{1, 2, \ldots, n\}$ and the arcs $\{(1, n-1), (1, n)\} \cup \{(i, 1) \mid 2 \le i \le n-2\}$. In this graph, Node 1 has out-degree 2, nodes i for $2 \le i \le n-2$ have out-degree 1, and nodes $n-1$, n have out-degree 0. Therefore, the outcome of the rank-sum heuristic is the ordering $1, 2, \ldots, n$. This ordering has value 2. However, G is acyclic (with a topological order $2, 3, \ldots, n-2, 1, n-1, n$) and has $n-1$ arcs. Therefore, the optimal solution has value $n-1$. The approximation ratio of the rank-sum heuristic cannot be better than $2/(n-1)$, and therefore it does not provide any constant factor performance guarantee. □

We remark that the adverse performance of the rank-sum heuristic does not apply if one replaces the sum of the ranks by the average of the ranks—that is, the ratio between the sum of the ranks by the number of reviewers that review this project. In that case, the average rank of Project 1 is $2/(n-1)$, the average rank of project i for $2 \le i \le n-2$ is 1, the average rank of projects $n-1$ and n is 0, and the resulting group ranking in this example is optimal. The example we give for the case of weights only applies however for

the average rank heuristic and demonstrates that its performance can also be bad.

A second Kemeny-Snell goal that has been discussed in the literature is associated with the input where each reviewer provides a full list consisting of a *permutation* of the reviewed projects. The goal is to identify a (group) permutation (that is, a unique assignment of ranks for a *full order*) so that the sum of *distances* of the group permutation from each reviewer's permutation is minimized. The distance between two permutations is the minimum number of adjacent pairwise exchanges between the two permutations (sometime referred to as the *bubble-sort distance*). Finding such optimal aggregate permutation was proved to be NP-hard by Bartholdi et al. (1989). Hardness results for this problem were tightened in special cases by Dwork et al. (2001).

We show in the next section that the intensity-analogs of the goal functions discussed here have the advantage of being polynomial-time solvable.

6. Intensity Rankings Only

We present here an optimization model for group ranking with intensity-rankings input from reviewers generalizing models developed for multicriteria decision making. We first establish that the problem of multicriteria decision making is equivalent to group ranking, and conclude that our model is applicable to that problem as well. We then demonstrate how the group-ranking model relates to the inverse paths problem and to the minimum-cost network flow problem.

6.1. Consistency of Multicriteria Decision Making and Its Equivalence to Group Ranking

Past research that addressed the group-ranking problem as a minimum of deviation optimization problem has appeared in the context of multicriteria decision making. In the problem of multicriteria decision making, the decision makers (possibly a single person), generate a ranking of the universal set using different criteria. The problem arises when the rankings based on the different criteria are jointly inconsistent. Formally, this is expressed as a ranking matrix where each column of the matrix represents the ranking with respect to one criterion. The problem then is to reconcile those different rankings into an aggregate consistent ranking that is close in some sense to the individual criteria's rankings. We show here that achieving a consistent ranking that is "similar" to a given inconsistent ranking is equivalent to attaining a group ranking that is "close" to the reviewers assessments.

In this area, the model of Saaty and Vargas (1984) 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 $\sum_{i=1}^{n}\sum_{j=1}^{n}(\log a_{ij} - \log w_i + \log w_j)^2$. Chandran et al. (2005) presented an alternative linear programming approach to the model of Saaty and Vargas, for the problem of identifying weights of a 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 project i. Defining both 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:

$$x_i - x_j = z_{ij}. \tag{4}$$

We refer to the problem of finding a close consistent matrix as (CM) (for consistent matrix). Note that fixing $x_1 = 0$ implies that the values of the weight variables x_i are in the range $[-M, M]$, for $M = \max_{(i,j)}|\log a_{ij}|$.

In the context of group ranking, Ali et al. (1986) explored a scenario where L reviewers provide rankings only, and no project weights. The intensity ranking of reviewer l is given as a 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. (1986) 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. (1986) 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 et al. is referred to here as (ACK) (after the initials of the authors).

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

$$\text{subject to} \quad z_{ik} - \sum_{j=i}^{k-1} z_{j,j+1} = 0$$

$$\text{for } i = 1, \ldots, n-2,\ i+2 \le k \le n$$

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

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.

Our model is the close rankings (CR) problem formulated on the additive aggregate ranking variables z_{ij} and the weights variables x_i and $F_{ij}()$ general convex functions,

$$\begin{aligned} \text{(CR)} \quad \text{Min} \quad & \sum_{i<j} F_{ij}(z_{ij}) \\ \text{subject to} \quad & x_i - x_j = z_{ij} \quad \text{for } i<j \\ & -n \le x_j \le n \quad j = 1, \ldots, n \\ & -n \le z_{ij} \le n \quad \text{integer, for all } i, j. \end{aligned}$$

We argue that the two problems (CM) and (ACK) are in fact the same, and both are generalized by (CR), which minimizes some measure of closeness. It is concluded that the models introduced in Saaty and Vargas (1984), Ali et al. (1986), and Chandran et al. (2005) are all special cases of (CR).

Theorem 6.1. *The problem of finding a close consistent ranking is equivalent to the problem of finding a group ranking that is close to the rankings of individual reviewers, and both are formulated as convex optimization problems.*

Proof. For the problem (ACK) of finding group rankings close to the L individuals' rankings, we let each reviewer provide a ranking matrix (p^l_{ij}). This ranking matrix is assumed to be consistent; thus, there are underlying weights w^l_i corresponding to each ranking so that $(p^l_{ij}) = w^l_i - w^l_j$ and $w^l_1 = 0$.

We now generate an $L \times n$ matrix where the lth column represents the complete rankings of reviewer l expressed as pairwise comparisons to project l. If the number of projects is less than the number of reviewers, $L > n$, then reviewer l expresses the preferences as compared to project $l(\text{mod}\, n-1)$, where the projects are numbered $0, 1, \ldots, n-1$. Therefore, the lth column has $a_{il} = w^l_i - w^l_l$. 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 project i, and $x_1 = 0$. We normalize the values of the rankings by dividing each value of p^l_{ij} by M, for $M = \max_{i,j,l} |p^l_{ij}|$. With n projects, 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. (1986), $F_{ij}(z_{ij}) = \sum_{l=1}^{L} |w^l_i - w^l_j - z_{ij}|$. An alternative choice of $F_{ij}()$ could be the quadratic convex function $\sum_{l=1}^{L} \alpha^l_{ij}(p^l_{ij} - z_{ij})^2$, where the coefficients α^l_{ij} reflect the weight, and thus the confidence, in the ranking of reviewer l for the pair ij, and replacing the term p^l_{ij} by $w^l_i - w^l_j$. 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^l_{ij} and z_{ij}, such as the function $F_{ij}(z_{ij}) = \sum_{l=1}^{L} |p^l_{ij} - z_{ij}|$ of (ACK).

The problem of reaching group rankings with quadratic function penalties, as in Saaty and Vargas (1984), is a special case of (CR) 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. is identical to that studied by Ali et al. (1986) when one replaces the individual rankings p^l_{ij} by a column of a_{ij} in the matrix.

Therefore, in terms of modelling, the problem of finding consistent rankings that are close to given inconsistent rankings is identical to the problem of obtaining group rankings from individual rankings that is close to those in some sense. □

The optimal objective value of (CR) 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.

6.2. Inverse Paths and Aggregate Ranking

The problem of modifying an inconsistent ranking so as to make it consistent is related to the problem of *inverse shortest paths* on a graph. To see this, consider a directed nonsimple (with multiple arcs) graph $G = (V, A)$ where arc $(i, j) \in A$ has intensity of a_{ij} associated with it in the additive model, and an arc in the opposite direction (j, i) of intensity $-a_{ij}$. For any pair of nodes i, j, let a directed path on k nodes from $i = v_0$ to $j = v_k$ be $P_{ij} = (i, v_2, \ldots, v_{k-1}, j)$. The total length of the path P_{ij} is $\sum_{p=1}^{k-1} a_{v_p, v_{p+1}}$, which is also equal to the intensity of the ranking of i relative to j if the ranking is consistent. Of course, if there are several intensities assigned to (i, j), they all have to modified so they are equal in a consistent ranking.

As shown in Theorem 2.1, a matrix is inconsistent if and only if there exist two paths in the graph between two nodes of different lengths.

The inverse equal-paths problem is defined on a directed graph with arc lengths. The problem is to

modify the lengths of the arcs so that all paths between each pair of nodes will be of the same length, and the penalty for the deviation from the given weights is minimum. This problem is thus the same as the problem of modifying a ranking so as to make it consistent. This and other versions of the inverse paths problems, such as—given the pairwise paths and their prescribed distances, modify the weights so as to minimize the penalty of the deviations and so that the pairwise distances are as prescribed—are studied in Hochbaum (2002).

A unique feature of the inverse paths problem in the context of a consistent aggregate ranking is that the skew-symmetry implies that for each arc present, there is an arc in the opposite direction. Therefore, having equal paths between two nodes is equivalent to having all direct cycles going through the pair of nodes of length zero.

6.3. A Model and Algorithm for Group Rankings with Intensity Rankings

The model formalizing the group ranking in the presence of intensity rankings only is the convex optimization model (CR) that generalizes the models of Ali et al. (1986), of Saaty and Vargas (1984), and of Chandran et al. (2005). This model formulation makes it possible to include rank preservation within the model. It is also straightforward to deal with a scenario partial lists (that is, not all pairwise comparisons are available). In order to incorporate rank preservation for a certain subset of directed arcs $A_S \subseteq A$, we add constraints of the form

$$x_i \geq x_j, \quad \forall (i, j) \in A_S.$$

Theorem 6.2. *For $F_{ij}(z_{ij})$ convex, the problem* (CR) *with rank preservation constraints is solved in time $O(mn \log(n^2/m) \log n)$.*

Proof. The convex problem (CR) is a special case of the convex dual of minimum-cost network flow (DMCNF). The formulation of the problem (DMCNF) with general convex objective function is,

$$\begin{aligned}
\text{(DMCNF)} \quad \text{Min} \quad & \sum_{j=1}^{n} w_j(x_j) + \sum_{(i,j)\in E} e_{ij}(z_{ij}) \\
\text{subject to} \quad & x_i - x_j \leq c_{ij} + z_{ij} \quad \text{for } (i, j) \in E \\
& l_j \leq x_j \leq u_j \quad j = 1, \ldots, n \\
& \beta_{ij} \leq z_{ij} \leq \gamma_{ij} \quad \text{for } (i, j) \in E \\
& x_j \text{ integer for all } j = 1, \ldots, n.
\end{aligned}$$

In this formulation $w_j()$ and $e_{ij}()$ are convex functions, c_{ij} are arbitrary constants, and l_j, u_j and β_{ij}, γ_{ij} are arbitrary lower and upper bounds on the variables x_i and z_{ij}. The constraints' coefficients form a totally unimodular matrix. For values of $\mathbf{x}$ and $\mathbf{c}$ that are integer, the values of $\mathbf{z}$ are integer as well.

Ahuja et al. (2003), showed that (DMCNF) generalizes the (CR) formulation. Namely, the constraints can be written as the consistency constraints (4) where the functions $e_{ij}()$ are replaced by other (convex) functions $F_{ij}()$. Moreover, adding rank-preservation constraints of the form $x_i \geq x_j$, $\forall (i, j) \in A_S$ preserves the total unimodularity of the matrix of constraints' coefficients, and the problem is still an instance of (DMCNF).

In Ahuja et al. (2003), the authors presented a polynomial-time algorithm for (DMCNF) with a convex objective function, which is the most efficient algorithm known to date for the problem. The complexity of the algorithm is $O(mn \log(n^2/m) \log(nC))$ for $C = \max\{u_j - l_j\}$. In the case of (CR), it is sufficient to find integer weights in $[-n, n]$, and thus $C = n$. Therefore, the combinatorial algorithm of Ahuja et al. (2003) works in strongly polynomial time, $O(mn \log(n^2/m) \log n)$. □

The algorithm of Ahuja et al. (2003) improves the complexity substantially compared to the linear programming approach proposed by Ali et al. (1986). It also improves in terms of the flexibility of choice of the objective function beyond absolute value, as appropriate under the circumstances. The efficiency of the algorithm is important to be able to reach a group decision according to the set criteria within a practical time window.

Another generalization that the model allows is for a partial-order scenario where not all pairwise rankings are given by all reviewers. For a set S_{ij} of reviewers evaluating the pair $[i, j]$ the functions are of the form $F_{ij}(z_{ij}) = F_{ij}((p^l_{ij})_{l\in S_{ij}}, z_{ij})$.

Remark 6.1. The (CR) model appears similar to the Kemeny-Snell problem with the (linear) objective function (3). Suppose this function is used in the (CR) model; then the optimal solution will be of value zero and all x_i are equal in the optimal solution. In the model of intensity ranking, however, a value of zero for some pair $z_{ij} = 0$ does penalize the objective function, as it potentially deviates from some other values of intensity ranking for (i, j) set by some of the reviewers.

Remark 6.2. The model (CR) where the objective function is quadratic separable, ($F_{ij}()$ is quadratic for all i, j), can be solved also in $O(n^3)$ as follows: We first replace z_{ij} with $x_i - x_j$ for all i and j, using the equality constraints of (CR) to generate a quadratic objective function of the variables $x_1, x_2, \ldots, x_n$. We then replace x_1 by 0. This quadratic multivariate problem can be solved using a general method for unconstrained convex optimization such as Newton's method or the conjugate gradient method. When the objective function is quadratic, the Newton's method

is guaranteed to reach a global optimal solution in one iteration, and this takes $O(n^3)$ (this time complexity is dominated by the time it takes to compute the inverse of the Hessian matrix). Alternatively, we can replace the use of Newton's method by the conjugate gradient method that reaches a global optimal solution after n lines search (again when restricted to quadratic goal function). Computing a direction for the search in the kth iteration takes $O(nk)$ time, and computing the next point using a line search takes a constant time (along a line our goal function is a quadratic function of one variable and it is easy to compute its minimum point in constant time). Therefore, the conjugate gradient method takes also $O(n^3)$ time. We point out that for dense graphs (where $m = \theta(n^2)$) and quadratic functions, this new algorithm is faster than the one of Theorem 6.2.

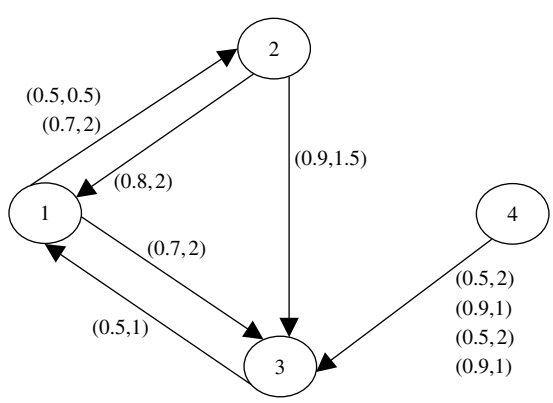
We next show a numerical example of model (CR) and its optimal solution.

Example 6.1. We specify only arcs (i, j) that indicate that i is preferred to j, with a nonnegative value for the intensity of the difference. The opposite arcs have the negative of the intensity value with the same belief level and are thus redundant—including them will double the value of the objective function penalty. Belief levels are specified as probabilities in the range $[0, 1]$. In our example shown in Figure 2, the pair (x, y) next to an arc indicates that one reviewer specifies x as the belief level and y as the intensity of the preference. For a reviewer preferring 1 to 2 with a difference of 2 and belief level of 0.7, $(0.7, 2)$, the penalty term is chosen as the quadratic function, $0.7(z_{12} - 2)^2$.

Given the graph of rankings and beliefs as shown in Figure 2, the objective function is to minimize:

$$\begin{aligned}&0.5(z_{12} - 0.5)^2 + 0.7(z_{12} - 2)^2 + 0.8(z_{21} - 2)^2\\&\quad + 0.9(z_{23} - 1.5)^2 + 0.7(z_{13} - 2)^2 + 0.5(z_{31} - 1)^2\\&\quad + 0.5(z_{43} - 2)^2 + 0.9(z_{43} - 1)^2 + 0.5(z_{43} - 2)^2\\&\quad + 0.9(z_{43} - 1)^2.\end{aligned}$$

Figure 2 **The Graph for Example 6.1, Where Each Pairwise Ranking Is Expressed as a Pair (x, y), Where x Is the Belief Level and y Is the Intensity of the Preference**



Setting $z_{ij} = x_i - x_j$ and fixing $x_1 = 0$, we obtain the new goal function:

$$\begin{aligned}&2.9x_2^2 - 2.6x_2 + 4.9x_3^2 + 12.1x_3 - 1.8x_2x_3\\&\quad + 2.8x_4^2 - 7.6x_4 - 5.6x_3x_4 + 17.25.\end{aligned}$$

The optimal solution is, $(x_1, x_2, x_3, x_4) = (0, 0.1335, -1.0142, 0.3429)$. The group ranking $(4, 2, 1, 3)$ with this deviation function has a minimum penalty value of 9.6373. This ranking reverses three pairwise preferences of reviewers—two preferring 1 to 2 and one preferring 3 to 1. However, the level of belief in these preferences is relatively low, and we cannot get any other ranking that will have a lower value of the penalty.

7. Weights and Rankings

Having reviewers provide both weights and rankings seems to be redundant, as the set of weights can be translated to a ranking and vice versa. Nevertheless, reviewers might be inconsistent in their own evaluations, and submitting both weights and rankings permits us to assign levels of confidence separately to the weights and to the pairwise rankings. This extra information can serve the role of capturing the evaluations of the reviewers more robustly 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 rankings.

In the procedure considered here, the output of the review process consists of both weights and intensity rankings (in the additive model). It is reasonable that the rankings 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 ranking 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 rankings 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*, which determine a ranking consistent with the weights. That is, project i is ranked higher than

project j if the final weight assigned is higher for i than for j. The separation cost is the cost for the final group ranking of deviating from the ranking of each of the reviewers.

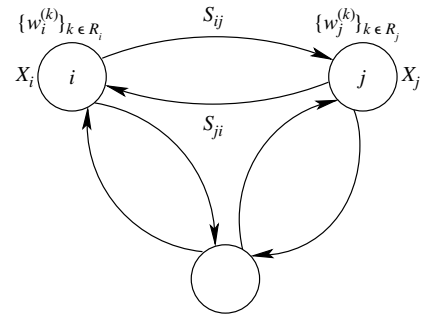
The terminology of "separation" and "deviation" costs borrows from the context of the problem of image segmentation and error correction (see Hochbaum 2001), which is shown to be related to the group-ranking problem. In the image segmentation set-up a transmitted image is degraded by noise. The assumption 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 there 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

$$\operatorname*{Min}_{u_i \geq x_i \geq l_i} \sum_{i \in V} G_i(g_i, x_i) + \sum_{i \in V} \sum_{j \in N(i)} F_{ij}(x_i - x_j).$$

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.

Figure 3 **The Graph for the Group Decision Problem**



Our formalization of the group-ranking problem as a graph problem is described schematically in Figure 3. Each project is a node in the graph, and each pairwise comparison of projects 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 $[-n, n]$. Each pair of nodes i, j has a set of reviewers in $R_i \cap R_j$ providing relative ranking. 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, $S_{ij} \cup S_{ji} = R_i \cap R_j$.

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

$$\begin{aligned}
\text{(GD)} \quad \text{Min} \quad & \sum_{j \in V} G_j((w_i^l)_{l \in R_i}, x_j) + \sum_{i<j} F_{ij}(z_{ij}, z_{ji}) \\
\text{subject to} \quad & x_i - x_j \leq z_{ij} \quad \text{for all } i, j \\
& x_j - x_i \leq z_{ji} \quad \text{for all } i, j \\
& n \geq x_j \geq -n \quad j = 1, \ldots, n \\
& z_{ji}, z_{ij} \geq 0 \quad (i, j) \in E.
\end{aligned}$$

Using the algorithms devised in Hochbaum (2001) for the image segmentation problem, we note that the case when the functions $F_{ij}()$ are linear is relevant to the group ranking. This is the case, e.g., for Equation (3), $F_{ij}(z_{ij}, z_{ji}) = \alpha_{ij}(|S_{ji}| \cdot z_{ij} + |S_{ij}| \cdot z_{ji})$. Here, because the weights are given in addition to the rankings and there is a penalty for deviating from the weights, the solution is not trivial as was remarked in the case of intensity rankings only, Remark 6.1.

Theorem 7.1. (a) *If $G_j()$ are convex for all j and $F_{ij}(z_{ij}) = e_{ij} z_{ij}$ are linear for all i and j, then* (GD) *is solvable in time $O(mn \log(n^2/m))$.*

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

(*c*) *If $G_j()$ are arbitrary nonlinear functions and $F_{ij}()$ are convex for all i and j, 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, i.e., $O(mn^3 \log(n^2/m))$.

Proof. The solution method follows the procedures used by Hochbaum for the image segmentation problem, in 2001. 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. (2003), 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 for all i and j, then the problem is solvable by a minimum cut on a graph on n^2 nodes and mn^2 arcs, $O(mn^3 \log(n^2/m))$, (Ahuja et al. 2004). 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. □

8. Conclusions

In this paper we analyze the practice of group rankings and recommend algorithms and procedures to improve on this practice. Our study indicates that a robust approach is based on reviewers providing intensity rankings with belief levels and possibly also project weights. The solution for the group ranking is then optimal with regard to a specific measure determined in advance by the reviewers or some central authority.

It will be interesting to further investigate the descriptive properties of this approach. That is, by observing a group decision, the challenge is to describe the implicit belief parameters provided by reviewers and how those affect the dynamics of group decision.

Acknowledgment

Research supported in part by NSF awards DMI-0085690 and DMI-0084857.

References

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

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

Alho, J. M., J. Kangas. 1997. Analyzing uncertainties in experts' opinions of forest plan performance. *Forest Sci.* **43**(4).

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

Arrow, K. J. 1963. *Social Choice and Individual Values*. Wiley, New York.

Bartholdi, J. J., C. A. Tovey, M. A. Trick. 1989. The computational difficulty of manipulating an election. *Soc. Choice Welfare* **6**(3) 227–241.

Belton, V., T. Gear. 1983. On a short-coming of Saaty's method of analytic hierarchies. *Omega* **11** 228–230.

Brans, J. P., Ph. Vincke. 1985. A preference ranking organization method. *Management Sci.* **31**(6) 647–656.

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

Chandran, B., B. Golden, E. Wasil. 2005. Linear programming models for estimating weights in the analytic hierarchy process. *Comput. Oper. Res.* **32**(9) 2235–2254.

Cohen, W. W., R. E. Schapire, Y. Singer. 1999. Learning to order things. *J. Artificial Intelligence Res.* **10** 243–270.

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

Duke, J. M., T. W. Ilvento, R. A. Hyde. 2002. Public support for land preservation. (February), http://www.udel.edu/FREC/PUBS/RR02-01.pdf.

Dwork, C., R. Kumar, M. Naor, D. Sivakumar. 2001. Rank aggregation revisited. *Proc. WWW10, Hong Kong*, Association IW3C2, Geneva, Switzerland, 613–622.

Even, G., J. Naor, B. Schieber, M. Sudan. 1998. Approximating minimum feedback sets and multicuts in directed graphs. *Algorithmica* **20** 151–174.

Expert Choice. 2005. Retrieved June 1, http://www.expertchoice.com.

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

Finan, J. S., W. J. Hurley. 1996. A note on a method to ensure rank-order consistency in the analytic hierarchy process. *Internat. Trans. Oper. Res.* **3** 99–103.

Finan, J. S., W. J. Hurley. 2002. The analytic hierarchy process: Can wash criteria be ignored? *Comput. Oper. Res.* **29**(8) 1025–1030.

Fuller, R., Ch. Carlsson. 1996. Fuzzy multiple criteria decision making: Recent developments. *Fuzzy Sets Systems* **78** 139–153.

Golden, B., E. Wasil, P. Harker. 1989. *The Analytic Hierarchy Process: Applications and Studies*. Springer, Berlin, Germany.

Hao, J., J. B. Orlin. 1994. A faster algorithm for finding the minimum cut in a directed graph. *J. Algorithms* **17** 424–446.

Hochbaum, D. S. 1997. *Approximation Algorithms for NP-Hard Problems*. PWS Publishing Company, Boston, MA.

Hochbaum, D. S. 2001. An efficient algorithm for image segmentation, Markov random fields and related problems. *J. ACM* **48**(4) 686–701.

Hochbaum, D. S. 2002. The inverse shortest paths problem. Manuscript, University of California, Berkeley, CA.

Hochbaum, D. S., A. Levin. 2006. The k-allocation problem for group ranking with partial lists. *Proc. Fourth Workshop on Approximation and Online Algorithms, Lecture Notes in Computer Science*. Springer-Verlag, Heidelberg, Germany. Forthcoming.

Karp, R. M. 1972. Reducibility among combinatorial problems. R. E. Miller, J. W. Thatcher, eds. *Complexity of Computer Computations*. Plenum Press, New York, 85–103.

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

Kemeny, J. G., J. Laurie Snell. 1962. Preference ranking: An axiomatic approach. *Mathematical Models in the Social Sciences*. Ginn, Boston, MA, 9–23.

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

Saaty, T. 1977. A scaling method for priorities in hierarchical structures. *J. Math. Psych.* **15** 234–281.

Saaty, T. 1980. *The Analytic Hierarchy Process*. McGraw-Hill, New York.

Saaty, T. 1987. Rank generation, preservation, and reversal in the analytic hierarchy decision process. *Decision Sci.* **18** 157–177.

Saaty, T., L. Vargas. 1984. Comparison of eigenvalue, logarithmic least squares and least squares methods in estimating ratios. *Math. Modeling* **5** 309–324.

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

Seymour, P. D. 1995. Packing directed circuits fractionally. *Combinatorica* **15**(2) 281–288.