Use and analysis of new optimization techniques for decision theory and data mining

by

Erick Moreno Centeno

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Engineering – Industrial Engineering and Operations Research

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Dorit S. Hochbaum, Chair
Professor Alper Atamtürk
Professor Richard M. Karp
Professor David R. Brillinger

Fall 2010

Use and analysis of new optimization techniques for decision theory and data mining

Abstract

Use and analysis of new optimization techniques for decision theory and data mining

by

Erick Moreno Centeno

Doctor of Philosophy in Engineering – Industrial Engineering and Operations Research

University of California, Berkeley

Professor Dorit S. Hochbaum, Chair

This dissertation addresses important problems in decision theory and data mining. In particular, we focus on problems of the form: Each of several information sources provides evaluations or measurements of the objects in a universal set, and the objective is to aggregate these, possibly conflicting, evaluations into a *consensus evaluation* of each object in the universal set. In addition, we concentrate on the scenario where each source provides evaluations of only a strict subset of the objects; that is, each source provides an *incomplete evaluation*.

In order to define the consensus evaluation from a given set of incomplete evaluations, two distances are developed: the first is a distance between incomplete rankings (ordinal evaluations) and the second is a distance between incomplete ratings (cardinal evaluations). These two distances generalize Kemeny and Snell's distance between complete rankings and Cook and Kress' distance between complete ratings, respectively. Specifically, we introduce a set of natural axioms that must be satisfied by a distance between two incomplete rankings (ratings) and prove the uniqueness and existence of a distance satisfying such axioms. Given a set of incomplete rankings (ratings), the consensus ranking (rating) is defined as the complete ranking (rating) that minimizes the sum of distances to each of the given rankings (ratings). We provide several examples that show that the consensus ranking (rating) obtained by this approach is more intuitive than that obtained by other approaches.

Finding the consensus ranking is NP-hard, thus we develop two optimization methodologies to find the consensus ranking: one efficient approximation algorithm based on the separation-deviation model and one exact algorithm based on the implicit hitting set approach. In addition, we show that the optimization problem that needs to be solved in order to find the consensus rating is a special case of the separation-deviation model (hereafter SD model), which is solvable in polynomial time. In this sense the herein developed theory (described in the previous paragraph) can be thought of an axiomatization of the SD model.

Three applications of the SD model are presented: rating the credit-risk of countries; customer segmentation; and ranking the participants in a student paper competition. In

the credit-risk rating study, it is shown that the SD model leads to an improved aggregate rating with respect to several criteria. We compare the SD model with other aggregation methods and show the following: Although the SD model is a method to aggregate cardinal evaluations, the aggregate credit-risk ratings obtained by the SD model are also good with respect to "ordinal criteria". Several properties of the SD model are proven, including the property that the aggregate rating obtained by the SD model agrees with the majority of agencies or reviewers, regardless of the scale used.

The customer segmentation study shows how to use the SD model to process data on customer purchasing timing. The outcome of the SD model provides insights on the rate of new product adoption by the company's consumers. In particular, the SD model is used as follows: given the purchase dates for each customer of several products, this information is aggregated in order to rate the customers with regard to their promptness to adopt new technology. We show that this approach outperforms unidimensional scaling—a widely used data mining methodology. We analyze the results with respect to various dimensions of the customer base and report on the generated insights.

The last presented application illustrates our aggregation methods in the context of the 2007 MSOM's student paper competition. The aggregation problem in this competition poses two challenges. First, each paper was reviewed only by a very small fraction of the judges; thus the aggregate evaluation is highly sensitive to the subjective scales chosen by the judges. Second, the judges provided both cardinal and ordinal evaluations (ratings and rankings) of the papers they reviewed. This chapter develops the first known methodology to simultaneously aggregate ordinal and cardinal evaluations into a consensus evaluation.

Although the content of this dissertation is framed in terms of decision theory, Hochbaum showed that data mining problems can be viewed as special cases of decision theory problems. In particular, the customer segmentation study is a classic data mining problem.

A Dios, mi esposa, mis papás y mi hermano.

Esta tesis no habría sido posible sin sus enseñanzas, apoyo, comprensión y paciencia.

Contents

Acknowledgments

First and foremost, I am deeply grateful to my advisor, Professor Dorit S. Hochbaum. She has been a mentor in a plethora of areas, from research and teaching to life in general. Before arriving to Berkeley, I knew I wanted to work under her guidance; Professor Hochbaum trusted me and kindly incorporated me to her research team even before I took a class with her. I thank her for inspiring me and keeping up an endless supply of problems to work on. Most importantly, I really appreciate her dedication; she always provided me with prompt, accurate and thorough feedback. This dissertation would simply not have been possible without her guidance.

It is an honor to have collaborated with Professor Richard M. Karp, whom I have admired since my first explorations into the world of combinatorial optimization. Unlike most outstanding researchers, Professor Karp is also an exceptional lecturer, always having an impeccable logic to explain profound results using the simplest terms. As a research collaborator, Professor Karp is always providing great insight and valuable ideas. Furthermore, he is forever encouraging and points out the improvement areas with only the kindest words.

I would like to express my gratitude to Professor Alper Atamtürk. I have immeasurably benefitted from the courses I took with him, courses that provided me with most of the foundations for the rest of my doctoral studies and part of my research. In addition, I am truly fortunate to have received his support and encouragement throughout and beyond my graduate studies.

I would especially like to thank Professor David R. Brillinger from the Statistics Department. He is not only an outstanding researcher, but, above all, he is a great human being. To him, I am particularly indebted for agreeing to be my non-IEOR committee member just two months before filing my dissertation.

I am very grateful to several professors at the IEOR Department for their support. Among them, I owe my deepest gratitude to Professors Ilan Adler and Rhonda Righter, who regularly provided me honest advice and helpful suggestions. Particularly, Professor Righter went out of her way more than once to assist and counsel me.

I would like to thank all the staff at the IEOR Department: Anayancy Paz, Mike Campbell, Jay Sparks, Joyce Levels, Diane Onley, and Rochelle Niccolls. In particular, Mike and Anayancy have been remarkably helpful in more than just administrative matters. From the Career Center, I am indebted to Andrew Green for his editorial work on my dissertation.

Thanks to CONACyT and UC Mexus for the Graduate Fellowship that funded me throughout my Ph.D. studies.

I am also grateful to those individuals who have contributed indirectly to the completion of this dissertation. In particular, I want to thank all my teachers and professors during my quite long academic life. I truly believe that the sturdy educational foundations, which I acquired at my elementary school, Olinca, and the rigorous academic preparation, which I received at ITESM, enabled me to succeed in my doctoral studies. It is impossible to mention everyone who has academically formed me; however, I do explicitly express my gratitude

Professor José Luis González-Velarde for formally introducing me into the fascinating world of combinatorial optimization.

It is my pleasure to thank my friends, who have stood by my side along the Ph.D. journey; several of them proofread many of my drafts. Thanks to James and Michelle for their unconditional friendship and their unwavering faith in me. Thanks to John and Nina for always being there and their scrumptious home-made gourmet food. I am especially indebted to John for lessening the load of a "GSIship" during the most stressful times. Thanks to Barak and Nili for their unreserved support and advice. Thanks to Ryan and Todd for trying to keep me doing exercise (my failure to exercise these last 2 years is my entire fault, not theirs). Thanks to Aude and Heather for invariably cheering me. Thanks to Shankar and Frances for their enriching friendship and for being always prompt to help, even when physically far. Thanks to all my Mexican and German friends, whom I do not list explicitly because I fear I will forget someone. All of them and many others, whom the reader will forgive me for passing over silently and gratefully, have contributed to making this journey enjoyable.

I owe my deepest gratitude to my wonderful family: María Centeno Barajas, Pablo Moreno Villafranco and Rodrigo Moreno Centeno, mom, dad, and brother, respectively. I especially thank them for their constant love, never-ending confidence and encouragement. They have made many sacrifices for my education and have always supported in all my endeavors. Every aspect of who I am and what I do is heavily influenced by their teachings.

I am markedly indebted to my wife, Irma Hernández Magallanes, for following me into the Ph.D. endeavor and for her unlimited patience and understanding. She has stoically maintained her vow to stand by my side during good and bad times. This dissertation was possible thanks to all her support; especially during this last year which has been particularly challenging. It is the nights, studying below the "candlelight", the moments I cherish the most.

Above all, I want to thank God. I hope I can serve others as He wants in order to thank Him for all the blessings I have received from Him, most of these, received through the intercession of the glorious Saint Jude Thaddeus.

Chapter 1

Introduction

One of the most important problems in decision theory is the group-decision making problem, which consists of aggregating several individual evaluations of a set of objects into a collective evaluation of those objects. This group-decision making problem has been widely-studied and has many applications, such as: voting (electors choose one among several candidates), jury decisions (judges evaluate competitors and rank them), consumer opinion aggregation (consumers evaluate a set of products), and project selection (committee members choose one or more projects among several).

In general, the group-decision problem is defined as follows: A group of individuals, which we refer to as *judges* or *reviewers*, must collectively evaluate all of the objects in a set. The problem is to aggregate the individual evaluations into a collective evaluation that aims to represent the individuals' assessments; such collective evaluation is referred to as *consensus*.

Group-ranking problems are differentiated by whether the evaluations are given in ordinal or cardinal scales. An ordinal evaluation, or *ranking*, is one where the objects are ordered from "most preferred" to "least preferred" (allowing ties). In other words, the most preferred object(s) is assigned to the ordinal number 'first', the next preferred object(s) is assigned to the ordinal number 'second', and so on. Specifically, in a ranking numbers indicate the relative position of the objects, but not a magnitude of difference. A cardinal evaluation, or *rating*, is one where the objects are assigned a scalar which is a cardinal score/grade. In a rating, the difference between the scores of two objects indicates the magnitude of separation between such objects. Depending on the type of evaluations being aggregated, group-ranking problems are characterized as *ranking aggregation problem* or *rating aggregation problems*.

The individual evaluations input to a group-ranking problem can be *complete* or *incomplete*. If an individual ranks (rates) all of the objects, then we say that he/she provides a complete evaluation; otherwise we say that he/she provides an incomplete evaluation. Regardless of the completeness (or lack thereof) of the given individual evaluations, we require that the collective evaluation is complete.

Most of the group-decision making literature concerns models that assume the judges' evaluations are complete. In contrast, in most applied settings the judges' evaluations are incomplete: in all of the aforementioned group-decision making applications it is conceivable that electors/judges/consumers/committee members do not evaluate all of the objects. Moreover, in the context of eliciting consumer opinions, Gibbons [32, page 257] states that the ability of respondents "to rank objects effectively and reliably may be a function of the number of comparative judgments to be made. For example, after 10 different brands of bourbon have been tasted, the discriminatory powers of the observers may legitimately be questioned." Similarly, in the context of Internet, a problem of particular interest is how to create a robust *meta-search* engine that combines the results of several search engines [24]. In this problem, it is inconceivable to ask each search engine to rank all webpages and then to create a consensus meta-search by aggregating all these rankings.

In the group-decision making problem, the objective is to determine a consensus evaluation that *best agrees* with the judges' evaluations. Note that the existence of a measure of agreement or disagreement between evaluations is implicit in the problem statement. Therefore, one subarea of group-decision making has concentrated on the axiomatic characterization of distance functions; each of these *axiomatic distances* is defined on a different "evaluation space." For example, Kemeny and Snell [49]—the pioneers in this subarea—defined an axiomatic distance in the space of complete rankings; Bogart [9, 10] defined an axiomatic distance in the space of strict partial orders; Cook, Kress and Seiford [21] defined an axiomatic distance in the space of non-strict partial orders; and, Cook and Kress [20] defined an axiomatic distance in the space of complete ratings (they referred to ratings as "ordinal rankings with intensity of preference"). We refer to all of these methods as *axiomatic distance-based methods*.

The judges' evaluations can also be expressed as *pairwise comparisons*. That is, each judge gives pairwise intensities of preference between object pairs. The intensity of preference may be expressed either in an additive (e.g., in [4]), or a multiplicative sense (e.g., in [56]). Here we use intensities of preference in the additive sense. That is, the intensity of preference represents the difference between the strengths of the two objects compared. We will use the term *separation gap* to refer to the additive intensity of a preference. Hochbaum's separation-deviation model (SD model) [37, 38, 41] is, to the best of our knowledge, the only aggregation method that permits the combination of both kinds of inputs: ratings and pairwise comparisons.

The contributions of this dissertation are:

- Developing an axiomatic distance-based method to derive the consensus ranking from a given set of incomplete rankings—a generalization of Kemeny and Snell's complete-ranking aggregation method.
- Providing two optimization methodologies to find the consensus ranking: one efficient approximation algorithm based on the SD model, and one exact solution methodology based on the implicit hitting set approach (developed in [47]).
- Developing an axiomatic distance-based method to derive the consensus rating from a given set of incomplete ratings—a generalization of Cook and Kress' complete-rating aggregation method.

- Showing that the optimization problem that needs to be solved in order to find the consensus rating is a special case of the SD model, which is solvable in polynomial time.
- Illustrating the advantages of the SD model in the context of country-credit risk rating and proving several properties of the SD model, including the property that the aggregate rating obtained by the SD model, when aggregating a set of complete ratings, agrees with the majority of the judges' ratings.
- Demonstrating the advantages of our incomplete-rating aggregation method in the context of customer rating, and comparing this approach to a widely-used data mining methodology—unidimensional scaling. (Note: Hochbaum [39] added Data Mining to the list of applications characterized as group-decision making problems. Specifically, Hochbaum showed how the SD model could be applied to the following data mining scenario: Given a database formed of a list of records, or cases, each case characterized by an array of measurements; the goal is to identify a function that maps each record array into a scalar value.)
- Proposing the first framework to simultaneously aggregate, both cardinal and ordinal evaluations (i.e., ratings and rankings) into a consensus and illustrate the advantages of this framework in the context of a student paper competition.

1.1 Preliminaries

1.1.1 Representations of ordinal evaluations

Consider the universe V of n objects to be ranked; and, without loss of generality, assign a unique identifier to each object in V so that $V = \{1, 2, \ldots, n\}$.

In the vector representation, a complete ranking is a vector of the form $\boldsymbol{a} = (a_1, \ldots, a_n)$, where a_i is the rank of object i. A natural way to represent an incomplete ranking as a vector is by setting a_i equal to a special symbol $\bullet$ if object i is not ranked in $\boldsymbol{a}$. Given an incomplete ranking $\boldsymbol{a}$ we denote as $\mathcal{A}$ the set of objects ranked in $\boldsymbol{a}$.

For example, for $V = \{1, 2, 3, 4\}$ the incomplete ranking $\boldsymbol{a} = \left(\begin{smallmatrix}2\\3\end{smallmatrix}\right)$ with object 2 as first, object 3 as second, and objects 1 and 4 not-ranked is represented as $\boldsymbol{a} = (\bullet, 1, 2, \bullet)$; here $\mathcal{A} = \{2, 3\}$. Similarly, the incomplete ranking $\boldsymbol{b} = \left(\begin{smallmatrix}4\\1-3\end{smallmatrix}\right)$ with object 4 as first, objects 1 and 3 tied as second, and object 2 not-ranked is represented as $\boldsymbol{b} = (2, \bullet, 2, 1)$; here $\mathcal{B} = \{1, 3, 4\}$. All the theory reviewed and developed in this dissertation is independent of the convention used to report ties (all that is required is that the tied objects have the same rank). So, for example, the ranking $\boldsymbol{b}$ can also be represented as $\boldsymbol{b} = (2.5, \bullet, 2.5, 1)$.

Throughout this dissertation we will primarily use the vector representation, however, in some instances the Kemeny-Snell representation [49] will be utilized. Under Kemeny-Snell, the ranking $\boldsymbol{a}$ is represented by an n by n matrix (a_{ij}). Where,

$$a_{ij} = \begin{cases} 1 & \text{if } i \text{ is preferred to } j, \\ -1 & \text{if } j \text{ is preferred to } i, \\ 0 & \text{if } i \text{ and } j \text{ are tied.} \end{cases}$$

Any ranking can be represented by such a matrix, but for a given matrix to represent a ranking, it must satisfy the properties of a ranking: The preference relationship $a_{ij} = 1$ must be asymmetric and transitive, and the "tie" relationship $a_{ij} = 0$ must be an equivalence relation.

A natural way to extend the Kemeny-Snell representation for incomplete rankings is to set $a_{ij} = a_{ji} = \bullet$ for $j = 1, \ldots, n$ when object i is not ranked in $\boldsymbol{a}$.

It should be clear from our notation that when we use one subscript, we are using the vector representation, and, when we use two, we are using Kemeny-Snell's. The Kemeny-Snell representation can be obtained from the vector representation as follows:

$$a_{ij} = \begin{cases} \operatorname{sign}(a_j - a_i) & \text{if } a_i \neq \bullet \text{ and } a_j \neq \bullet \\ \bullet & \text{otherwise.} \end{cases}$$

In the ranking aggregation problem, each one of K judges provides a ranking $\boldsymbol{a}^k$ of the objects in V. In the complete-ranking aggregation problem, all of the judges' rankings are complete, and, in the incomplete-ranking aggregation problem, some or all of the judges' rankings are incomplete.

1.1.2 Representations of cardinal evaluations

As before, consider the universe V of n objects to be rated; and, without loss of generality, assign a unique identifier to each object in V so that $V = \{1, 2, \ldots, n\}$.

In the vector representation, a complete rating is a vector of the form $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$, where a_i is the cardinal evaluation, or score, of object i. A natural way to represent an incomplete rating as a vector is by setting a_i equal to a special symbol $\bullet$ if object i is not rated in $\boldsymbol{a}$. Given an incomplete rating $\boldsymbol{a}$ we denote as $\mathcal{A}$ the set of objects rated in $\boldsymbol{a}$.

Throughout this dissertation it is assumed that possible scores are the integers contained in some pre-specified interval $[\ell, u]$. Generally, this is the case in any group-decision making scenario: Judges are given a grading scale from which they must assign a score to the objects being evaluated. The *range* of the ratings is defined as $R \equiv u - \ell$.

Alternatively, the judges' evaluations can be expressed as pairwise comparisons. That is, each judge gives pairwise intensity of preferences between object pairs. The intensity of preference between a pair of objects may be used either in the additive (e.g., in [4]), or multiplicative sense (e.g., in [56]). Here we use intensities of preference in the additive sense. That is, the intensity of preference represents the difference between the strengths of the two objects compared. We will use the term *separation gap* to refer to the additive intensity of a preference.

In order to represent cardinal evaluations that are given in the form of separation gaps, we use a *separation gap matrix* (Cook and Kress [20] refer to this matrix as the *intensity of preference matrix*). In this representation, the cardinal evaluations are represented by a matrix $(p_{ij})_{i,j=1,\ldots,n}$, where, p_{ij} gives the separation gap between object i and object j. Specifically, p_{ij} is a cardinal number giving the additive intensity of preference by which object i is preferred to object j (if object j is preferred to object i, then p_{ij} will be negative).

A separation gap matrix is said to be *consistent* if for all triplets i, j, k, $p_{ij} + p_{jk} = p_{ik}$. Consistency is equivalent to the existence of rating $\boldsymbol{a}$ so that $p_{ij} = a_i - a_j$ for $i, j = 1, \ldots, n$. This rating is not unique since for any scalar c the rating $\boldsymbol{a} + c$ has also the same set of differences. In contrast, a rating $\boldsymbol{a}$ uniquely defines a separation gap matrix: $p_{ij}^{(a)} = a_i - a_j$ for $i = 1, \ldots, n$.

Throughout this dissertation we assume that all separation gap matrices are consistent; moreover, all separation gap matrices are calculated from a given rating. For that reason, hereafter we denote as $(a_{ij})_{i,j=1,\ldots,n}$ the separation gap matrix that is calculated from the rating $\boldsymbol{a} = (a_i)_{i=1,\ldots,n}$. It should be clear from our notation that a_i refers to the cardinal score of object i and a_{ij} refers to the separation gap of object i over object j. In Chapter 3 we demonstrate that, even when the input to a complete-rating aggregation problem is given only as ratings, it is also useful to consider their corresponding separation gaps.

A natural way to represent a separation gap matrix corresponding to an incomplete rating is to set $a_{ij} = a_{ji} = \bullet$ for $j = 1, \ldots, n$ when object i is not rated in $\boldsymbol{a}$.

In the rating aggregation problem, each one of K judges provides a rating $\boldsymbol{a}^k$ of the objects in V. The separation gap matrices corresponding to each of the judges' ratings are denoted by $(p_{ij}^k)_{i,j=1}^n$. In the complete-rating aggregation problem, all of the judges' ratings are complete, and, in the incomplete-rating aggregation problem, some or all of the judges' ratings are incomplete.

1.2 Literature Review

Here we review the literature that is common to all chapters.

1.2.1 Review of the separation-deviation model (SD model)

The SD model, developed in [37, 38, 41], considers an scenario where the input to the rating process is in the form of separation gaps, and, in addition, the reviewers can also provide ratings.

The input of the SD model can be stated in terms of a multi-graph $G = (V, A)$. Each node $i \in V$ corresponds to an object, and each arc $(i, j) \in A$ represents a comparison between two objects. Since there are several judges, each possibly having a different separation gap for the same pair of objects, a pair of nodes can have multiple arcs in between. The weight associated with each arc represents the separation gap p_{ij}^k of the ith object over the jth object as given by the kth judge. Additionally, if the kth judge provides a score of the ith object, then node i has the weight a_i^k associated with it (a node can have multiple weights associated with it).

Let the variable $\boldsymbol{x}$ be the aggregate rating, and the variable $\boldsymbol{z}$ be the corresponding aggregate separation gap matrix. The mathematical programming formulation of the separation-deviation model, given in [41] is:

$$\text{(Sep-Dev)} \quad \min_{\boldsymbol{x},\boldsymbol{z}} \quad \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}^k(z_{ij} - p_{ij}^k) + \sum_{k=1}^{K}\sum_{i=1}^{n} g_i^k(x_i - a_i^k) \tag{1.1a}$$

$$\text{subject to} \quad z_{ij} = x_i - x_j \quad i = 1, \ldots, n; \quad j = 1, \ldots, n \tag{1.1b}$$

$$\ell \leq x_i \leq u \quad i = 1, \ldots, n \tag{1.1c}$$

$$x_i \in \mathbb{Z} \quad i = 1, \ldots, n. \tag{1.1d}$$

The function $f_{ij}^k(\cdot)$, called *separation penalty function*, penalizes the difference between the aggregate separation gap of the object pair (i, j) and the kth judges' separation gap of the object pair (i, j). The function $g_i^k(\cdot)$, called *deviation penalty function*, penalizes the difference between the aggregate rating of object i and the kth judges' cardinal score of object i. In order to ensure polynomial-time solvability, the functions $f_{ij}^k(\cdot)$ and $g_i^k(\cdot)$ must be **convex**. In the context of rating aggregation, the penalty functions assume the value 0 for the argument 0. If the kth judge did not gave a separation gap on the pair (i, j) (i.e., $p_{ij}^k = \bullet$), then $f_{ij}^k(\cdot)$ is set to zero; similarly, if the kth judge did not gave a cardinal score on the ith object (i.e., $a_i = \bullet$), then $g_i^k(\cdot)$ is set to zero. Equation (1.1b) enforces that the aggregate separation gaps correspond to the aggregate rating.

We refer to the SD model with no deviation functions, or $g_i^k() \equiv 0$ for $k = 1, \ldots, K$, as the *separation model*. In the separation model, for any feasible solution $\boldsymbol{x}$ and any constant c, $\boldsymbol{x} + c\,\boldsymbol{e}$ (where $\boldsymbol{e}$ is the vector of ones) is also a feasible solution with the same objective value. Therefore the separation model has an infinite number of optimal solutions. To avoid this, the rating of an arbitrarily selected anchor node is set to zero, e.g. $x_1 = 0$. The other aggregate scores x_i for $i = 2, \ldots, n$ are then relative to this 'anchor' value. The mathematical representation of the separation model as optimization problem is:

$$\text{(Sep)} \quad \min_{\boldsymbol{x},\boldsymbol{z}} \quad \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=i+1}^{n} f_{ij}^k(z_{ij} - p_{ij}^k) \tag{1.2a}$$

$$\text{s.t.} \quad z_{ij} = x_i - x_j \quad (i = 1, \ldots, n;\ j = i+1, \ldots, n) \tag{1.2b}$$

$$x_1 = 0. \tag{1.2c}$$

For both (Sep-Dev) and (Sep) it is easy to see that a feasible solution always exists. This holds since, e.g. for some k, the solution $x_i = r_1^k$ for $i = 1, \ldots, n$ and $z_{ij} = x_i - x_j$ for

$i, j = 1, \ldots, n$, is obviously feasible. The uniqueness of the optimal solution is guaranteed when the functions $f_{ij}^k()$ and $g_i^k()$ are strictly convex functions. Otherwise the separation-deviation model might have multiple optimal solutions.

1.2.2 Review of the Kemeny and Snell's theory

Kemeny and Snell [49] proposed a set of axioms which any distance between complete rankings must satisfy, proved that the axioms lead to the existence of a unique distance function—referred to as the Kemeny-Snell distance.

Kemeny-Snell's distance can be interpreted as follows: The distance between two complete rankings $\boldsymbol{a}$ and $\boldsymbol{b}$ is given by the total number of *rank reversals* between them. A rank reversal is incurred whenever two objects have a different relative order in the complete rankings $\boldsymbol{a}$ and $\boldsymbol{b}$. Similarly, *half a rank reversal* is incurred whenever two objects are tied in one ranking but not in the other ranking.

Given a set of complete rankings, Kemeny and Snell defined the consensus ranking to be the ranking with the minimum sum of distances to the given rankings. Kemeny and Snell did not provide a method for finding such a consensus ranking. Subsequently, Bartholdi et. al. [6] showed that finding the Kemeny-Snell consensus ranking is NP-hard.

1.3 Organization of the dissertation

With the exception of the material presented in the previous section, which is common to all chapters and is required to understand what follows, the subsequent chapters are presented in as self-contained a manner as possible.

Chapter 2 - Axiomatic distances for aggregating incomplete evaluations

Kemeny and Snell's (Cook and Kress') axiomatic distance in the space of complete rankings (ratings) is generalized in order to define an axiomatic distance in the space of incomplete rankings (ratings). Specifically, this chapter: (1) introduces a set of natural axioms that must be satisfied by a distance between two incomplete rankings (ratings); (2) proves the uniqueness and existence of a distance satisfying such axioms; (3) shows that this distance is equivalent to Kemeny and Snell's (Cook and Kress') distance on their subspace of complete rankings (ratings); and (4) shows that the consensus ranking obtained when using our axiomatic distance is more intuitive than that obtained by other approaches.

Given a set of incomplete rankings (ratings) and a distance between incomplete rankings (ratings), the *consensus ranking (rating)* is defined as the complete ranking (rating) with the minimum sum of distances to the given rankings (ratings). In this chapter we give specific algorithms to find the consensus ranking (rating) using the herein-defined distance. Since finding the consensus ranking is NP-hard, we develop two optimization methodologies: one efficient approximation algorithm based on the SD model, and one exact solution methodology based on the implicit hitting set approach (developed in [47]). We show that the optimization problem that needs to be solved in order to find the consensus rating is a special case of the SD model, thus solvable in polynomial time.

Since our distance on the space of incomplete rankings (ratings) is equivalent to Kemeny and Snell's (Cook and Kress') distance on their subspace of complete rankings (ratings), the developed algorithms can be also used in the case where the given rankings (ratings) are complete. Since one of the developed algorithms, finds a consensus ranking with the additional restriction of not containing ties, the herein-developed methodology to derive a consensus ranking can be thought of as a unifying framework for ranking aggregation: It can be used to aggregate rankings that are complete, incomplete, contain ties, or do not contain ties; and the consensus ranking can be further restricted to not contain ties.

Chapter 3 - Country credit-risk rating aggregation

This chapter addresses the problem of aggregating several conflicting country credit risk ratings into a consensus rating. Country credit-risk ratings quantify the risk associated with investing in a given country. Haque et al. [35] define country credit-risk rating as an estimate of the probability that a country will fail to pay back its debt. To satisfy increasing investor need for information on countries' creditworthiness, several agencies periodically publish country credit-risk ratings. Often there are differences between the agencies' credit-risk ratings for a particular country. It is therefore of interest to aggregate those differing views into a coherent rating that represents a group consensus capturing the different expertise of the rating agencies.

Within this dissertation, this chapter is the only one dealing with the aggregation of complete ratings. In this context, we demonstrate that, even when the input is given *only* as ratings, it is also useful to consider the implied separation gaps. The use of separation gaps is shown to mitigate the effect of inflated scores or shifts in evaluation scale.

The most commonly used method of rating aggregation is the *averaging method*. In this method the aggregate score of each country is the average of the ratings scores that this country received from all of the judges. We assess the performance of the separation-deviation model and compare the model to the averaging method, using several performance measures. We demonstrate that, for aggregating complete ratings, the aggregate rating obtained by the SD model better preserves the relative order of the objects induced by each of the input ratings as compared to the aggregate rating obtained by the averaging method.

In this chapter we also prove several properties of the SD model when applied to the complete-rating aggregation problem; for example, that the aggregate rating obtained by the SD model agrees with the majority of agencies or judges, regardless of the scale used.

Chapter 4 - Rating customers according to their promptness to adopt new technology

This chapter studies the problem of customer segmentation. Here the segmentation is according to the promptness to adopt new technology. In particular, this chapter addresses the following problem: Given a set of customers, a set of products, and the purchase times of each customer-product pair, rate each customer according to its *adoption promptness*. We focus on the problem where the information available is incomplete; that is, there are customers who do not purchase every product. We illustrate our methodology on data obtained from Sun Microsystems. We compare our methodology to the well known unidimensional scaling methodology, which is widely used for customer segmentation as well as in several other contexts (see, for example, [27, 53, 31]).

Chapter 5 - Simultaneous aggregation of cardinal and ordinal evaluations: ranking in a student paper competition

This chapter describes the ranking of the participants of the 2007 MSOM's student paper competition (SPC). The group-ranking problem arising in the context of SPC is special in that the judges evaluating the papers provide both ratings and rankings.

We develop a group-decision making framework to simultaneously aggregate ratings and rankings. The framework consists of finding the *combined aggregate rating* (CAT) and its implied ranking, referred to as *combined aggregate ranking* (CAK). This rating-ranking pair is the one that minimizes the sum of the distances from the CAT to the judges' ratings plus the sum of the distances from the CAK to the judges' rankings. The aggregation method is supplemented by two methods to identify inconsistencies in the evaluations of the objects. This information is helpful to identify judges whose rating scales significantly differ form those used by the rest of the judges.

Chapter 2

Axiomatic distance-based methods for aggregating incomplete evaluations

2.1 Introduction

This paper generalizes Kemeny and Snell's (Cook and Kress') axiomatic distance in the space of complete rankings (ratings) in order to define an axiomatic distance in the space of incomplete rankings (ratings). Specifically, this paper (1) introduces a set of natural axioms that must be satisfied by a distance between two incomplete rankings (ratings); (2) proves the uniqueness and existence of a distance satisfying such axioms; and (3) shows that this distance is equivalent to Kemeny and Snell's (Cook and Kress') distance on their subspace of complete rankings (ratings).

In all of the axiomatic-distance based methods mentioned in Chapter 1, the consensus evaluation is defined as the *median evaluation* of the judges' evaluations. For example, Kemeny and Snell define the consensus ranking as the median complete ranking, and Cook and Kress define the consensus rating as the median complete rating. The median is defined as follows: Given a finite collection Ω of points in a metric space, a median of Ω is a point in the space with minimum sum of distances to the points in Ω. Unfortunately, in all of the aforementioned evaluation spaces medians may not be unique. Given a set of incomplete rankings (ratings), we define the consensus ranking (rating) as the complete ranking (rating) with the minimum sum of distances to the given rankings (ratings). In this sense, we require that the consensus evaluation is in a subspace of the evaluation space. However, with a slight abuse of notation, we refer to our consensus ranking (rating) as a median ranking (rating). The requirement of the consensus evaluation to be complete captures the nature of the group-decision making problem; where it would be unacceptable to obtain a consensus ranking (rating) that is incomplete—after all, the group's goal is to evaluate every object in V. Hereafter we use the terms consensus ranking (rating) and median ranking (rating) interchangeably.

The contributions of this paper include showing that the consensus ranking obtained when using the herein-defined axiomatic distance is more intuitive than that obtained by other approaches. In addition, this paper provides specific algorithms to find the median ranking (rating). Since the herein-defined distance on the space of incomplete rankings (ratings) is equivalent to Kemeny and Snell's (Cook and Kress') distance on their subspace of complete rankings (ratings), the proposed algorithms can also be used in the case where the given rankings (ratings) are complete. Since one of the proposed algorithms, finds a consensus ranking with the additional restriction of not containing ties, the herein-proposed methodology to derive a consensus ranking can be thought of as a unifying framework for ranking aggregation: It can be used to aggregate rankings that are complete, incomplete, contain ties, or do not contain ties; and the consensus ranking can be further restricted to not contain ties.

This paper is organized as follows: Section 2.2 presents the literature review, with special emphasize on Kemeny-Snell's complete-ranking aggregation model (§2.2.1), and Cook-Kress' complete-rating aggregation model (§2.2.2). Section 2.3 develops the incomplete-ranking aggregation method. Specifically, Section 2.3.1 introduces a set of natural axioms that must be satisfied by any distance between two incomplete rankings; Section 2.3.2 proves that these axioms lead to the existence of a unique distance function; Section 2.3.3 shows that the consensus ranking obtained with this approach is more intuitive than that obtained with other approaches previously suggested in the literature; Section 2.3.4 shows that finding the consensus ranking is NP-hard; and proposes two optimization methodologies to find the consensus ranking—one efficient heuristic based on the separation-deviation model, and one exact solution methodology based on the implicit hitting set approach [47]. Section 2.4 develops the incomplete-rating aggregation method. Specifically, Section 2.4.1 introduces a set of natural axioms that must be satisfied by any distance between two incomplete ratings; Section 2.4.2 proves that these axioms lead to the existence of a unique distance function; §2.4.3 shows that the consensus rating can be found in polynomial time since this problem turns out to be a special case of the separation-deviation model.

2.2 Literature review

A number of distance-based methods have been developed to address the group-decision making problem; each method solving a different variant of the problem. The difference between these methods is the type of evaluations (complete rankings, complete ratings, strict partial orders, non-strict partial orders, etc.) being aggregated. Regardless of the type of evaluations, all of these methods are developed as follows: (1) provide a set of axioms which any distance in the evaluation space should satisfy, (2) prove that the axioms lead to the existence of a unique distance function, and (3) provide solution procedures to find the consensus evaluation—although not all studies provide solution procedures.

Kemeny and Snell [49] were the first researchers to propose a distance-based method, they examined the problem where the evaluations are given as complete rankings that allow ties, or *weak orderings*. Since our axioms (given in §2.3.1), for a distance between incomplete rankings, are generalizations of Kemeny-Snell's axioms, section 2.2.1 thoroughly reviews their work.

Bogart [9, 10] looked at the problem of aggregating *strict partial orders*. A strict partial order is a binary relation that is irreflexive, asymmetric and transitive. Bogart's distance, defined in the space of strict partial orders, is equivalent to Kemeny-Snell's distance in the subspace of complete rankings. Bogart neither provided a method for finding the consensus strict partial order, nor analyzed the complexity of such problem. Section 2.3.3 briefly reviews Bogart's distance and shows that, with a minor modification, Bogart's distance can be used as a distance between incomplete rankings. When using Bogart's distance to aggregate incomplete rankings, the consensus ranking tends to favor rankings that do not contain ties (see section 2.3.3).

Cook, Kress and Seiford [21] examined the problem of aggregating non-strict partial orders. A non-strict partial order is a binary relation that is reflexive, antisymmetric and transitive. In particular, a non-strict partial order allows three levels of comparison between each pair of objects, namely, strict preference, tied preference and "no comparison." Cook-Kress-Seiford's distance, defined in the space of non-strict partial orders, is also equivalent to Kemeny-Snell's distance in the subspace of complete rankings. Cook, Kress and Seiford did not provide a method for finding the consensus non-strict partial order, nor did they analyze the complexity of such a problem. Section 2.3.3 briefly reviews Cook-Kress-Seiford's distance, which can also be used as a distance between incomplete rankings. When using Cook-Kress-Seiford's distance to aggregate incomplete rankings, the consensus ranking tends to favor rankings that contain ties (see section 2.3.3).

Kemeny-Snell's distance can be easily extended to a distance between incomplete rankings: The distance between two incomplete rankings, $\boldsymbol{a}$ and $\boldsymbol{b}$, is the total number of *rank reversals* between them, where, the rank reversals are only summed over the object-pairs that are ranked in both $\boldsymbol{a}$ and $\boldsymbol{b}$. For reasons that will become clear later, we call this distance the *projected Kemeny-Snell* distance, or *PKS distance*. The PKS distance was explored by Dwork et. al. [24] and Cook et al. [19]. We note that both, [24] and [19], only consider *strict rankings*, that is, rankings that do not contain ties. Unlike the other distances reviewed in this section, neither Cook et al. nor Dwork et. al. provide a set of axioms that uniquely define the PKS distance. This by itself is not so important; indeed in section 2.3.3 we prove that there exists such a set of axioms. The trouble is that, when using the PKS distance to aggregate incomplete rankings, the consensus ranking tends to be disproportionately closer to the rankings that compare a large amount of objects as compared to other rankings (see section 2.3.3). This limitation can be easily tracked down to the set of axioms that uniquely define the PKS distance.

Dwork et. al. [24] and Cook et al. [19] developed a heuristic and a Branch-and-Bound algorithm, respectively, for the following problem: Given a set of K incomplete strict rankings find the complete strict ranking with the minimum sum of PKS distances to the given rankings. The algorithms developed in this paper can be applied to solve the above problem. Moreover, the herein-developed heuristic algorithm has no restriction on the given rankings being strict (that is, the given rankings may or may not contain ties) and the consensus ranking may or may not contain ties. Similarly, the herein-developed exact algorithm has no restriction on the input rankings being strict; however, the consensus ranking is limited to not containing ties.

Cook and Kress [20] developed what they called an "ordinal ranking with intensity of preference." The motivation behind this model was to capture the relative intensities of preference:

> It is often common to require that voters supply not only a list of preferences, but some expression of intensity of preference as well. For example, in ranking R&D projects it is often the case that a committee member is given a 100-point scale, and each of n projects (n = 20, say) is to be positioned at an integer point on this scale. Thus, a particular voter may choose to place the first ranked object at position 1, the second at position 4, the third at position 5, and so on. In this example the voter perceives a greater intensity of preference for the first object over the second than for the second over the third. Obviously, this intensity component must not be ignored. Yet, it cannot be handled conveniently within the framework of existing axiomatic models such as that of Kemeny and Snell. [20]

In this quote Cook and Kress were referring to what we defined as a complete rating. In [20], Cook and Kress (1) provided a set of axioms, similar to those of Kemeny and Snell, which any distance between ratings should satisfy; (2) proved that the axioms lead to the existence of a unique distance function; and (3) discussed a (not-polynomial-time) algorithm to find the consensus rating. Ali et. al. [4] showed that the consensus rating could be obtained in polynomial time by formulating this problem as an integer program over a totally unimodular constraint matrix.

In [37, 38, 41] Hochbaum developed the separation-deviation model, thoroughly reviewed in §1.2.1. The separation-deviation model (hereafter SD model) is solvable via network flow techniques, yet it is a very general group-decision making model since it can accept a wide variety of cardinal inputs: complete ratings, incomplete ratings, pairwise comparisons (defined in §1.2.1), pointwise-scores (defined in §1.2.1) and even imprecise beliefs about the given cardinal evaluations. Finding the consensus rating using Cook and Kress' distance is a special case of the SD model. Similarly, finding the consensus rating using the herein-developed axiomatic distance between incomplete ratings is a special case of the SD model. Moreover, the herein-developed heuristic to solve the incomplete ranking aggregation problem is based on the SD model.

2.2.1 Kemeny-Snell's distance between complete rankings

Kemeny and Snell [49], proposed a set of axioms describing a distance function between complete rankings. In order to present Kemeny-Snell's axioms the following concepts are defined.

Definition 2.2.1. *A ranking* $\boldsymbol{b}$ *is* between *rankings* $\boldsymbol{a}$ *and* $\boldsymbol{c}$ *if, for each pair of objects* i *and* j*, the preference judgment of* $\boldsymbol{b}$ *either agrees with* $\boldsymbol{a}$*; or, agrees with* $\boldsymbol{c}$*; or,* $\boldsymbol{a}$ *prefers* i*,* $\boldsymbol{c}$ *prefers* j*, and* $\boldsymbol{b}$ *ties* i *and* j*.*

Definition 2.2.2. *A set* S *of objects is a* segment *of a given ranking* $\boldsymbol{a}$ *if* $\bar{S}$ *(the complement of* S*) is not empty and if the rank* a_i *of every element* i *in* $\bar{S}$ *is either higher than that of every element of* S*, or lower than that of every element of* S*.*

Kemeny and Snell argue that a distance, $d(\cdot,\cdot)$, between two complete rankings, $\boldsymbol{a}$ and $\boldsymbol{b}$, should satisfy the following axioms:

- **Axiom K1** (Nonnegativity) $d(\boldsymbol{a},\boldsymbol{b}) \geq 0$, and $d(\boldsymbol{a},\boldsymbol{b}) = 0$ if and only if $\boldsymbol{a}$ and $\boldsymbol{b}$ are the same ranking.
- **Axiom K2** (Commutativity) $d(\boldsymbol{a},\boldsymbol{b}) = d(\boldsymbol{b},\boldsymbol{a})$.
- **Axiom K3** (Triangular inequality) $d(\boldsymbol{a},\boldsymbol{b}) + d(\boldsymbol{b},\boldsymbol{c}) \geq d(\boldsymbol{a},\boldsymbol{c})$, and the equality holds if and only if $\boldsymbol{b}$ is *between* $\boldsymbol{a}$ and $\boldsymbol{b}$.
- **Axiom K4** (Anonymity) If $\boldsymbol{a}'$ results from $\boldsymbol{a}$ by a permutation of the objects in V, and $\boldsymbol{b}'$ results from $\boldsymbol{b}$ by the same permutation, then $d(\boldsymbol{a},\boldsymbol{b}) = d(\boldsymbol{a}',\boldsymbol{b}')$.
- **Axiom K5** (Extension) If two rankings $\boldsymbol{a}$ and $\boldsymbol{b}$ agree except for a set S of k elements, which is a segment of both, then $d(\boldsymbol{a},\boldsymbol{b})$ may be computed as if these k objects were the only objects being ranked.
- **Axiom K6** (Scaling) The minimum positive unit is 1/2.

Axioms K1 to K3 are self-explanatory. Axiom K4 ensures that the distance does not depend on the particular labeling of the objects. Axiom K5 states that, if the two rankings are in complete agreement at the beginning and at the end of the list, and differ only as to the ranking of k objects in the middle, then this distance is the same as if these k objects were the only objects under consideration. Axiom K6 is just a matter of convention (choosing a unit of measurement). Kemeny and Snell set the minimum positive unit to 1; however, as explained below, their distance has a nicer interpretation if the minimum positive unit is 1/2.

Kemeny and Snell proved that Axioms K1 to K6 are simultaneously satisfied by only one distance, $d_{KS}(\cdot,\cdot)$:

$$d_{KS}(\boldsymbol{a},\boldsymbol{b}) = \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij} - b_{ij}| \tag{2.1a}$$

$$= \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\text{sign}(a_i - a_j) - \text{sign}(b_i - b_j)\right|, \tag{2.1b}$$

where, equation (2.1a) uses Kemeny-Snell's representation of rankings; and equation (2.1b) uses the vector representation of rankings.

Kemeny-Snell's distance, $d_{KS}(\cdot,\cdot)$, has the following interpretation: The distance between two rankings is given by the total number of *rank reversals* between them. A rank reversal is incurred whenever two objects have a different relative order in the rankings $\boldsymbol{a}$ and $\boldsymbol{b}$. Similarly, *half* a rank reversal is incurred whenever two objects are tied in one ranking but not in the other ranking.

2.2.2 Cook and Kress' distance between complete ratings

Cook and Kress [20], proposed a set of axioms describing a distance function between complete ratings. In order to present Cook-Kress' axioms the following concepts are defined.

Definition 2.2.3. *A rating $\boldsymbol{b}$ is* between *ratings $\boldsymbol{a}$ and $\boldsymbol{c}$ if, for every pair of objects i and j, either $a_{ij} \leq \boldsymbol{b}_{ij} \leq \boldsymbol{c}_{ij}$ or $a_{ij} \geq \boldsymbol{b}_{ij} \geq \boldsymbol{c}_{ij}$.*

Definition 2.2.4. *A rating $\boldsymbol{a}$ is said to be* adjacent *to a rating $\boldsymbol{b}$ if, for every pair of objects i and j, $|a_{ij} - b_{ij}| \leq 1$. That is, for every pair of objects the intensity of preference in ranking $\boldsymbol{b}$ is either the same as in $\boldsymbol{a}$ or differs by exactly one unit.*

Definition 2.2.5. *A rating $\boldsymbol{a}$ is said to be* adjacent of degree k *to a rating $\boldsymbol{b}$ if $\boldsymbol{a}$ is adjacent to $\boldsymbol{b}$ and $|\{(i,j) : |a_{ij} - b_{ij}| = 1, i < j\}| \leq k$. That is, the set of all pairs of objects for which the intensity of preference differs by one unit has cardinality k.*

Cook and Kress argue that a distance, $d(\cdot,\cdot)$, between two complete ratings, $\boldsymbol{a}$ and $\boldsymbol{b}$, should satisfy the following axioms:

Axiom C1 (Nonnegativity) $d(\boldsymbol{a},\boldsymbol{b}) \geq 0$.

Axiom C2 (Commutativity) $d(\boldsymbol{a},\boldsymbol{b}) = d(\boldsymbol{b},\boldsymbol{a})$.

Axiom C3 (Triangular inequality) $d(\boldsymbol{a},\boldsymbol{b}) + d(\boldsymbol{b},\boldsymbol{c}) \geq d(\boldsymbol{a},\boldsymbol{c})$, and the equality holds if and only if $\boldsymbol{b}$ is *between* $\boldsymbol{a}$ and $\boldsymbol{b}$.

Axiom C4 (Proportionality) The distance between any two adjacent rankings is proportional to the degree of adjacency.

Axiom C5 (Scaling) The minimum positive unit is 1.

Cook and Kress proved that Axioms 1 to 5 are simultaneously satisfied by only one distance, $d_{CK}(\cdot,\cdot)$:

$$d_{CK}(\boldsymbol{a},\boldsymbol{b}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij} - b_{ij}| \tag{2.2a}$$

$$= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|(a_i - a_j) - (b_i - b_j)\right|, \tag{2.2b}$$

where, equation (2.2a) uses Cook-Kress' representation of ratings; and equation (2.2b) uses the vector representation of ratings.

2.3 Incomplete-ranking aggregation

2.3.1 Axioms for a distance between incomplete rankings

In this section we modify Kemeny-Snell's axioms (given in Section 2.2.1) in order to obtain a set of axioms appropriate for a distance between incomplete rankings. Moreover, we want to define a distance, $d(\cdot,\cdot)$, that is suitable for solving the incomplete-ranking aggregation problem.

Recall that given the incomplete rankings of K judges, $\{\boldsymbol{a}^1,...,\boldsymbol{a}^K\}$, the consensus ranking $\boldsymbol{r}$ is the optimal solution to the following optimization problem:

$$\min_{\boldsymbol{r}}\sum_{k=1}^{K} d(\boldsymbol{a}^k,\boldsymbol{r}), \tag{2.3}$$

where the minimum is over all complete (strict and non-strict) rankings.

The aim of the axioms proposed in this section is to describe a distance between incomplete rankings such that, when this distance is used in problem (2.3), the consensus ranking minimizes the disagreement of the judges. We now specify the axioms that our distance must satisfy.

Throughout this section, the axioms are denoted using the number of the corresponding Kemeny-Snell axiom. In addition, the axioms denoted by a prime are those that are modified slightly so that they applied to distances between incomplete rankings. For example, Axiom 0 (below) has no corresponding Kemeny-Snell axiom; Axiom 1' (below) is a slight variant of Axiom K1; and Axiom 2 (below) corresponds exactly to Axiom K2.

First, $d(\cdot,\cdot)$ must capture our intuition of disagreement between incomplete rankings. In particular, given two incomplete rankings $\boldsymbol{a}$ and $\boldsymbol{b}$, $d(\boldsymbol{a},\boldsymbol{b})$ must consider only the disagreement between $\boldsymbol{a}$ and $\boldsymbol{b}$ in the objects ranked by both $\boldsymbol{a}$ and $\boldsymbol{b}$ (i.e., only the objects in $\mathcal{A}\bigcap\mathcal{B}$). Intuitively, we want $d(\cdot,\cdot)$ to measure the disagreement in the relative order of the ranked objects, not the disagreement on which objects were ranked. So, for example, given the incomplete rankings $\boldsymbol{a} = (1, 2, \bullet, 3)$ and $\boldsymbol{b} = (\bullet, 3, 2, 1)$, the distance should only consider the objects in $\mathcal{A} \bigcap \mathcal{B} = \{2, 4\}$. Specifically, we do not say that $\boldsymbol{a}$ disagrees with $\boldsymbol{b}$ because $\boldsymbol{b}$ doesn't rank object 1; likewise, we do not say that $\boldsymbol{b}$ disagrees with $\boldsymbol{a}$ because $\boldsymbol{a}$ doesn't rank object 3; however, it is clear that rankings $\boldsymbol{a}$ and $\boldsymbol{b}$ disagree in their preference between objects 2 and 4, with each object preferred over the other object by $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. This condition is self-explanatory; to make it mathematically precise, we need the following definition.

Definition 2.3.1. *Given a ranking $\boldsymbol{a}$ and a subset S of the object universe V, the* projection *of $\boldsymbol{a}$ on S, denoted as $\mathbf{a}|_S$, is the ranking of the objects in S that preserves the relative order of the objects specified by $\boldsymbol{a}$ on the objects in S.*

The desired condition that $d(\boldsymbol{a}, \boldsymbol{b})$ must only consider the objects in $\mathcal{A} \bigcap \mathcal{B}$ is expressed as follows:

Axiom 0 (Relevance) $d(\boldsymbol{a}, \boldsymbol{b}) = d(\boldsymbol{a}|_{(\mathcal{A} \bigcap \mathcal{B})}, \boldsymbol{b}|_{(\mathcal{A} \bigcap \mathcal{B})})$

The next two conditions, Axioms 1' and 2, correspond to Axioms K1 and K2.

Axiom 1' (Nonnegativity) $d(\boldsymbol{a}, \boldsymbol{b}) \geq 0$, and $d(\boldsymbol{a}, \boldsymbol{b}) = 0$ if and only if $\boldsymbol{a}|_{(\mathcal{A} \bigcap \mathcal{B})}$ and $\boldsymbol{b}|_{(\mathcal{A} \bigcap \mathcal{B})}$ are the same ranking.

Axiom 2 (Commutativity) $d(\boldsymbol{a}, \boldsymbol{b}) = d(\boldsymbol{b}, \boldsymbol{a})$.

The modification of Axiom 1' (with respect to Axiom K1) has the same justification as Axiom 0. That is, we want $d(\cdot, \cdot)$ to measure the disagreement in the relative order of the ranked objects, not the disagreement on which objects were ranked. Therefore the rankings $\boldsymbol{a}$ and $\boldsymbol{b}$ have no disagreement $d(\boldsymbol{a}, \boldsymbol{b}) = 0$ whenever they have the same preferences among the objects ranked by both.

Notice that given two arbitrary incomplete rankings $\boldsymbol{a}$ and $\boldsymbol{b}$, the set $\mathcal{A} \bigcap \mathcal{B}$ might be empty or contain just one object. In the case when $\mathcal{A} \bigcap \mathcal{B} = \emptyset$, we have that

$$d(\boldsymbol{a}, \boldsymbol{b}) = d(\boldsymbol{a}|_\emptyset, \boldsymbol{b}|_\emptyset) = d(\emptyset, \emptyset) = 0,$$

where the first equality follows from Axiom 0; the second inequality follows since a ranking of an empty set of objects is itself the empty set (interpreting a ranking as a partial order); and the third inequality follows from Axiom 1'. A similar analysis shows that, when $|\mathcal{A} \bigcap \mathcal{B}| = 1$, these axioms imply that $d(\boldsymbol{a}, \boldsymbol{b}) = 0$. This makes sense since two incomplete rankings that do not rank the same objects (or, that only rank one object in common) have no disagreement. In any case, for our purposes, we will only use $d(\cdot, \cdot)$ in the context of problem (2.3). In particular, we only need to calculate the disagreement between a complete ranking and an incomplete ranking. Thus the observation articulated in this paragraph is not important; since we can assume, without loss of generality, that all of the given rankings $\boldsymbol{a}^k$ rank at least two objects.

Axiom K3 requires that the distance satisfies the triangle inequality ($d(\boldsymbol{a},\boldsymbol{b})+d(\boldsymbol{b},\boldsymbol{c})\geq d(\boldsymbol{a},\boldsymbol{c})$). This condition cannot be imposed directly on the distance, because this condition is inconsistent with Axioms 0 and 1'. This is illustrated in the following example:

Let $\boldsymbol{a}=(1,2,\bullet)$, $\boldsymbol{b}=(\bullet,1,2)$, and $\boldsymbol{c}=(2,1,\bullet)$. Since $|\mathcal{A}\bigcap\mathcal{B}|=1$ and $|\mathcal{B}\bigcap\mathcal{C}|=1$, we have $d(\boldsymbol{a},\boldsymbol{b})=0$ and $d(\boldsymbol{b},\boldsymbol{c})=0$. On the other hand, from Axiom 1' and the fact that $\boldsymbol{a}$ and $\boldsymbol{c}$ are not the same ranking when projected to $\mathcal{A}\bigcap\mathcal{C}$, we have that $d(\boldsymbol{a},\boldsymbol{c})>0$. This clearly violates the triangle inequality. We nevertheless require that our distance satisfies a "relaxed version" of the triangle inequality (Axiom 3' below). Strictly speaking, $d(\cdot,\cdot)$ will not be even a pseudometric, because it violates the (unrelaxed) triangle inequality.

Axiom 3' ("Relaxed" triangular inequality) $d(\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})},\boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})})+d(\boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})},\boldsymbol{c}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})})\geq d(\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})},\boldsymbol{c}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})})$, and equality holds if and only if $\boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}$ is between $\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}$ and $\boldsymbol{c}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}$.

The next two conditions are identical to the corresponding Kemeny-Snell's Axioms.

Axiom 4 (Anonymity) If $\boldsymbol{a}'$ results from $\boldsymbol{a}$ by a permutation of the objects in V, and $\boldsymbol{b}'$ results from $\boldsymbol{b}$ by the same permutation, then $d(\boldsymbol{a},\boldsymbol{b})=d(\boldsymbol{a}',\boldsymbol{b}')$.

Axiom 5 (Extension) If two rankings $\boldsymbol{a}$ and $\boldsymbol{b}$ agree except for a set S of k elements, which is a segment[1] of both, then $d(\boldsymbol{a},\boldsymbol{b})$ may be computed as if these k objects were the only objects being ranked.

The sixth and last Kemeny-Snell axiom, Axiom K6, is the scaling axiom. For the case of complete rankings, this axiom is a mere convention. However, the scaling axiom is of central importance for measuring the disagreement between incomplete rankings. More precisely, the scaling axiom is of central importance for the incomplete-ranking aggregation problem, and in particular for problem (2.3). The idea behind the scaling axiom is that, implicit in the definition of the incomplete-ranking aggregation problem, all of the judges' rankings, $\{\boldsymbol{a}^k\}_{k=1}^{K}$, have the same importance.

Axiom 0 requires that the distance is evaluated by projecting the two incomplete rankings into the set of objects ranked by both. And problem (2.3) minimizes a sum of distances, each of which may be evaluated over spaces of different number of dimensions. Therefore, since distances in higher dimensional spaces tend to be bigger than distances in lower dimensional spaces, the objective function of problem (2.3) will tend to be dominated by the distances from $\boldsymbol{r}$ to the given incomplete rankings that rank a larger number of objects.

In light of the discussion in the previous paragraph, the following is argued: The distances between incomplete rankings should be normalized so that all the distances in problem (2.3) are comparable. This can be achieved by normalizing the distances so that they are between zero and one (inclusively). So, a distance of zero will indicate total agreement, and a distance of one will indicate total disagreement. Intuitively, given a ranking $\boldsymbol{a}$, the ranking which disagrees the most with $\boldsymbol{a}$ is its *reverse ranking.* A ranking $\boldsymbol{a}$ is the reverse ranking of $\boldsymbol{b}$

[1] On the natural extension of a segment (Definition 2.2.2) to incomplete rankings.

if, for every pair of objects i and j, either $\boldsymbol{a}$ prefers i and $\boldsymbol{b}$ prefers j or $\boldsymbol{a}$ prefers j and $\boldsymbol{b}$ prefers i. Instead of having a scaling axiom like Kemeny and Snell, we impose the following normalization axiom on our distance.

Axiom 6' (Normalization) $d(\boldsymbol{a},\boldsymbol{b}) \leq 1$; and $d(\boldsymbol{a},\boldsymbol{b}) = 1$ if and only if $\boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B})}$ is the reverse ranking of $\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B})}$.

2.3.2 Uniqueness and existence of a distance

This section shows that the distance between incomplete rankings given in equation (2.4), here called the *normalized projected Kemeny-Snell distance*, or simply NPKS distance, satisfies Axioms 0 to 6. Moreover, we prove that the NPKS distance is the unique distance that simultaneously satisfies Axioms 0 to 6.

$$d_{NPKS}(\boldsymbol{a},\boldsymbol{b}) = \begin{cases} \frac{d_{KS}(\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B})},\boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B})})}{(|\mathcal{A}\bigcap\mathcal{B}|^2-|\mathcal{A}\bigcap\mathcal{B}|)/2} & \text{if } |\mathcal{A}\bigcap\mathcal{B}| \geq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

Lemma 2.3.2 ([49])**.** *Given two complete rankings $\boldsymbol{a}$ and $\boldsymbol{b}$ on a set of objects V, $d_{KS}(\boldsymbol{a},\boldsymbol{b})$ attains its maximum of $(|V|^2 - |V|)/2$ when $\boldsymbol{b}$ is the reverse ranking of $\boldsymbol{a}$.*

Lemma 2.3.3. *The NPKS distance satisfies Axioms 0 to 6.*

Proof. It follows directly from equation (2.4) that $d_{NPKS}(\cdot,\cdot)$ satisfies Axiom 0.

The nonnegativity of $d_{NPKS}(\cdot,\cdot)$ follows from equation (2.4) and the nonnegativity of $d_{KS}(\cdot,\cdot)$. To see that $d_{KS}(\cdot,\cdot)$ satisfies the second part of Axiom 1', we consider three cases:

1. $|\mathcal{A}\bigcap\mathcal{B}| = 0$: In this case, $d_{NPKS}(\boldsymbol{a},\boldsymbol{b}) = 0$ and both $\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B})}$ and $\boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B})}$ are rankings over an empty set of objects; thus, they are, by definition, the same ranking. Therefore the second part of Axiom 1' is satisfied.
2. $|\mathcal{A}\bigcap\mathcal{B}| = 1$: In this case, $d_{NPKS}(\boldsymbol{a},\boldsymbol{b}) = 0$ and both $\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B})}$ and $\boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B})}$ are rankings over a set with a single object; thus, they are, by definition, the same ranking. Therefore the second part of Axiom 1' is satisfied.
3. $|\mathcal{A}\bigcap\mathcal{B}| \geq 2$: In this case, $d_{NPKS}(\cdot,\cdot)$ satisfies the second part of Axiom 1' as a consequence of equation (2.4) and the fact that $d_{KS}(\cdot,\cdot)$ satisfies the second part of Axiom K1.

Axioms 2, 4 and 5 follow from equation (2.4) and the fact that $d_{KS}(\cdot,\cdot)$ satisfies Axioms K2, K4 and K5, respectively.

Axiom 3' follows from equation (2.4); the fact that $\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}$, $\boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}$ and $\boldsymbol{c}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}$ are complete rankings; and the fact that $d_{KS}(\cdot,\cdot)$ satisfies Axiom K3.

From Lemma 2.3.2 and equation (2.4) we have that $d_{NPKS}(\cdot,\cdot)$ satisfies Axiom 6'. □

Corollary 2.3.4. *Axioms 0 to 6 are consistent.*

Next, we show that Axioms 0 to 6 uniquely determine the NPKS distance.

Theorem 2.3.5. *The distance $d_{NPKS}(\cdot,\cdot)$ is the unique distance satisfying Axioms 0 to 6 simultaneously.*

Proof. The fact that $d_{NPKS}(\cdot,\cdot)$ satisfies Axioms 0 to 6 was established in Lemma 2.3.3. So we only need to show that no other distance satisfies Axioms 0 to 6 simultaneously. Let $d(\cdot,\cdot)$ be a generic distance that satisfies Axioms 0 to 6. We shall prove the theorem by showing that, for any two rankings $\boldsymbol{a}$ and $\boldsymbol{b}$, we have $d(\boldsymbol{a},\boldsymbol{b}) = d_{NPKS}(\boldsymbol{a},\boldsymbol{b})$. We divide our analysis in the following three cases:

Case 1: $|\mathcal{A}\bigcap\mathcal{B}| \leq 1$.

As argued in section 2.3.1, for any distance function satisfying Axioms 0 and 1', and for any two rankings $\boldsymbol{a}$ and $\boldsymbol{b}$ such that $|\mathcal{A}\bigcap\mathcal{B}| \leq 1$, the distance must be equal to zero.

Case 2: $|\mathcal{A}\bigcap\mathcal{B}| \geq 2$, and both $\boldsymbol{a}$ and $\boldsymbol{b}$ are complete rankings.

Since in this Case 2 we are restricting our attention to complete rankings, it follows that Axioms 1 to 5 are identical to Axioms K1 to K5. Therefore, since by assumption $d(\boldsymbol{a},\boldsymbol{b})$ satisfies Axioms 1 to 5 we have that $d(\boldsymbol{a},\boldsymbol{b})$ satisfies axioms K1 to K5.

As explained in section 2.2.1, the sole purpose of Axiom K6 is to fix the measurement unit; in other words, Axioms K1 to K5 uniquely determine a distance function up to a scaling factor. Therefore, since both $d(\boldsymbol{a},\boldsymbol{b})$ and $d_{KS}(\boldsymbol{a},\boldsymbol{b})$ satisfy Axioms K1 to K5, it follows that $d(\boldsymbol{a},\boldsymbol{b}) = \alpha d_{KS}(\boldsymbol{a},\boldsymbol{b})$ for some constant α that may depend only on $|V|$. From equation (2.4), and since in this Case 2 $|V| \geq 2$ and $\mathcal{A}\bigcap\mathcal{B} = V$, we have that $d_{NPKS}(\boldsymbol{a},\boldsymbol{b}) = \frac{1}{(|V|^2-|V|)/2} d_{KS}(\boldsymbol{a},\boldsymbol{b})$.

Let the ranking $\boldsymbol{r}$ be the ranking on which the objects are ranked according to their index; that is $\boldsymbol{r} = (1, 2, ..., |V|)$. Let the ranking $-\boldsymbol{r}$ be the reverse ranking of $\boldsymbol{r}$; that is $-\boldsymbol{r} = (|V|, |V|-1, ..., 1)$. From Axiom 1 we have that $d(\boldsymbol{r},\boldsymbol{r}) = 0$ and $d_{NPKS}(\boldsymbol{r},\boldsymbol{r}) = 0$, and from Axiom 6 it follows that $d(\boldsymbol{r},-\boldsymbol{r}) = 1$ and $d_{NPKS}(\boldsymbol{r},-\boldsymbol{r}) = 1$. Therefore, it must be the case that $\alpha = \frac{1}{(|V|^2-|V|)/2}$.

We conclude that in this Case 2, as in Case 1, it is true that $d(\boldsymbol{a},\boldsymbol{b}) = d_{NPKS}(\boldsymbol{a},\boldsymbol{b})$ for any two rankings $\boldsymbol{a}$ and $\boldsymbol{b}$.

Case 3: $|\mathcal{A}\bigcap\mathcal{B}| \geq 2$, and at least one of $\boldsymbol{a}$ or $\boldsymbol{b}$ is an incomplete ranking.

In this Case 3, as in the two previous cases, it is true that $d(\boldsymbol{a},\boldsymbol{b}) = d_{NPKS}(\boldsymbol{a},\boldsymbol{b})$ for any two rankings $\boldsymbol{a}$ and $\boldsymbol{b}$. This is shown by the following sequence of equalities.

$$\begin{aligned} d(\boldsymbol{a},\boldsymbol{b}) &= d(\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B})}, \boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B})}) \\ &= \frac{1}{(|\mathcal{A}\bigcap\mathcal{B}|^2 - |\mathcal{A}\bigcap\mathcal{B}|)/2} d_{KS}(\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B})}, \boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B})}) \\ &= d_{NPKS}(\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B})}, \boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B})}) \\ &= d_{NPKS}(\boldsymbol{a},\boldsymbol{b}) \end{aligned}$$

The first and last equalities follow from Axiom 0. The second equality follows from our analysis of Case 1 and the fact that $\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B})}$ and $\boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B})}$ are complete rankings over the set $\mathcal{A}\bigcap\mathcal{B}$. The third equality follows by definition of $d_{NPKS}(\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B})}, \boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B})})$. □

2.3.3 Comparison to other approaches

There are other distances that may be used to solve the incomplete-ranking aggregation problem (IAP). These distances include Bogart's distance [9], Cook-Kress-Seiford's distance [21], and the un-normalized version of the NPKS distance (proposed in [24]. This section shows that the consensus ranking obtained when using the NPKS distance is "more intuitive" than those obtained with these distances. This section is organized as follows: Sections 2.3.3, 2.3.3 and 2.3.3 present each of these distances and comment on their drawbacks when applied to solve the IAP. Section 2.3.3 gives specific examples to illustrate these drawbacks.

Review and discussion of Bogart's distance

Bogart [9] generalized the Kemeny-Snell distance between complete rankings to a distance function between partial orders. (A partial order is a binary relation that is reflexive, antisymmetric and transitive.) In particular, Bogart proposed a set of axioms that a distance between partial orders should satisfy and then proved that his distance was the unique distance that simultaneously satisfied his axioms. Bogart showed that, in the subspace of all complete rankings (a complete ranking is a partial order where all pairs of objects are compared), his distance and the Kemeny-Snell distance are the same.

Bogart's distance between two given partial orders, P and Q, is given by[2]

$$d_B(P, Q) = \|I(P) - I(Q)\|$$

where $\|\cdot\|$ denotes the matrix L_1 norm (i.e., the sum of all the matrix entries' absolute values), and $I(P)$ is the incidence matrix of the partial ordering P given by

$$I(P)_{ij} = \begin{cases} 1 & \text{if } (i, j) \in P \\ 0 & \text{otherwise,} \end{cases}$$

where $(i, j) \in P$ means that object j is not preferred to object i (that is, either object i is preferred to object j or the objects are tied).

Since an incomplete ranking is a partial order, we can use Bogart's distance in order to find the distance between incomplete orders (see the example below).

[2]Strictly speaking, Bogart's distance is defined for strict partial orders—irreflexive, asymmetric and transitive binary relations—, but it is easily extended for (non-strict) partial orders, as done here.

Example: Consider the universe of objects $V = \{1, 2, 3, 4\}$, and the incomplete ranking $\boldsymbol{a} = (2, \bullet, 2, 1)$; that is, ranking $\boldsymbol{a}$ has object 4 as first, objects 1 and 3 tied as second, and object 2 not ranked. The partial order representation of $\boldsymbol{a}$ is $P^a = \{(1,1), (2,2), (3,3), (4,4), (1,3), (3,1), (4,1), (4,3)\}$, whose incidence matrix, $I(P^a)$, is

$$I(P^a) = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

Now consider the incomplete rankings $\boldsymbol{b} = (2, \bullet, 3, 1)$, $\boldsymbol{c} = (3, \bullet, 2, 1)$, and $\boldsymbol{d} = (\bullet, 2, \bullet, 1)$, whose incidence matrices are

$$I(P^b) = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix} \qquad I(P^c) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix} \qquad I(P^d) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

Pictorially, these rankings are $\boldsymbol{a} = \left(\begin{smallmatrix} & 4 & \\ 1 & - & 3 \end{smallmatrix} \right)$, $\boldsymbol{b} = \left(\begin{smallmatrix} 4 \\ 1 \\ 3 \end{smallmatrix} \right)$, $\boldsymbol{c} = \left(\begin{smallmatrix} 4 \\ 3 \\ 1 \end{smallmatrix} \right)$, and $\boldsymbol{d} = \left(\begin{smallmatrix} 4 \\ 2 \end{smallmatrix} \right)$.

Bogart's distances between the above rankings are $d_B(\boldsymbol{a}, \boldsymbol{b}) = 1$, $d_B(\boldsymbol{a}, \boldsymbol{c}) = 1$, $d_B(\boldsymbol{a}, \boldsymbol{d}) = 5$, $d_B(\boldsymbol{b}, \boldsymbol{c}) = 2$, $d_B(\boldsymbol{b}, \boldsymbol{d}) = 4$, and $d_B(\boldsymbol{c}, \boldsymbol{d}) = 4$.

Notice that $d_B(\boldsymbol{b}, \boldsymbol{d})$, and $d_B(\boldsymbol{c}, \boldsymbol{d})$ are smaller than $d_B(\boldsymbol{a}, \boldsymbol{d})$. Therefore, using Bogart's distance, one might conclude that there is a higher level of agreement between $\boldsymbol{b}$ (alternatively, $\boldsymbol{c}$) and $\boldsymbol{d}$ than between $\boldsymbol{a}$ and $\boldsymbol{d}$. We believe that this is incorrect. In particular, we believe that the level of agreement between $\boldsymbol{b}$ (alternatively, $\boldsymbol{c}$) and $\boldsymbol{d}$ should be equal to the level of agreement between $\boldsymbol{a}$ and $\boldsymbol{d}$. We believe this because the only difference between $\boldsymbol{b}$ (alternatively, $\boldsymbol{c}$) and $\boldsymbol{a}$ is that $\boldsymbol{a}$ ties objects 1 and 3; while $\boldsymbol{b}$ prefers 1 over 3 (alternatively, $\boldsymbol{c}$ prefers 3 over 1).

In general, the "behavior" of Bogart's distance, illustrated in the above example, can be generalized as follows: Suppose that a given incomplete ranking, $\boldsymbol{d}$, does not compare objects i and j. The incidence matrix $I(\boldsymbol{d})$ will have $I(\boldsymbol{d})_{ij} = 0$ and $I(\boldsymbol{d})_{ji} = 0$. The incidence matrix $I(\boldsymbol{a})$ of some arbitrary ranking $\boldsymbol{a}$ that ties objects i and j will have $I(\boldsymbol{a})_{ij} = 1$ and $I(\boldsymbol{a})_{ji} = 1$. The incidence matrix $I(\boldsymbol{b})$ of some arbitrary ranking $\boldsymbol{b}$ that prefers i to j will have $I(\boldsymbol{b})_{ij} = 1$ and $I(\boldsymbol{b})_{ij} = 0$. Finally, the incidence matrix $I(\boldsymbol{c})$ of some arbitrary ranking $\boldsymbol{c}$ that prefers j to i will have $I(\boldsymbol{c})_{ij} = 1$ and $I(\boldsymbol{c})_{ij} = 0$. Therefore, with respect to the pair of objects i and j, ranking $\boldsymbol{b}$ (alternatively, $\boldsymbol{c}$) and ranking $\boldsymbol{d}$ are closer than ranking $\boldsymbol{a}$ and ranking $\boldsymbol{d}$.

The above discussion hints that the consensus ranking obtained by using Bogart's distance will tend to be a ranking containing an "artificially low" amount of ties. That is, even objects that are expected to be tied by the consensus ranking will not be tied by it; this is illustrated in section 2.3.3. We close this section with the following remark: Our purpose is not to criticize Bogart's distance (a legitimate distance between partial orders); rather it is to show that Bogart's distance is not appropriate to solve the IAP. (Unlike our NPKS distance, Bogart's distance was not designed to solve the IAP.)

Review and discussion of Cook-Kress-Seiford's distance

Cook, Kress and Seiford [21] proposed another distance between partial orders. Like Bogart, they also proposed a set of axioms that a distance between partial orders should satisfy and then proved that their distance was the unique distance that simultaneously satisfied their axioms. As Bogart, Cook et. al. also demonstrated that, in the subspace of all complete rankings, their distance and the Kemeny-Snell distance are the same.

Cook-Kress-Seiford's distance between two given partial orders, Q and R, on a set of n objects, is given by

$$d_C(Q,R) = \sum_{i=1}^{n}\sum_{j=1}^{n}\left[\frac{1}{2}\left|J(Q)_{ij} - J(R)_{ij}\right| + \left|P(Q)_{ij} - P(R)_{ij}\right|\right],$$

where J, referred to as the *information* matrix, and P, referred to as the *preference* matrix, are

$$I_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are compared (strict preference or tie)} \\ 0 & \text{if } i \text{ and } j \text{ are not compared.} \end{cases}$$

$$P_{ij} = \begin{cases} 1 & \text{if } i \text{ is strictly preferred to } j \\ 0 & \text{otherwise.} \end{cases}$$

Example: Consider the universe of objects $V = \{1,2,3,4\}$, and the incomplete rankings given in section 2.3.3: $\boldsymbol{a} = (2,\bullet,2,1)$, $\boldsymbol{b} = (2,\bullet,3,1)$, $\boldsymbol{c} = (3,\bullet,2,1)$, and $\boldsymbol{d} = (\bullet,2,\bullet,1)$. Pictorially, these rankings are $\boldsymbol{a} = \left(\begin{smallmatrix} 4 \\ 1\!-\!3 \end{smallmatrix}\right)$, $\boldsymbol{b} = \begin{pmatrix} 4 \\ 1 \\ 3 \end{pmatrix}$, $\boldsymbol{c} = \begin{pmatrix} 4 \\ 3 \\ 1 \end{pmatrix}$, and $\boldsymbol{d} = \left(\begin{smallmatrix} 4 \\ 2 \end{smallmatrix}\right)$.

Cook-Kress-Seiford's distances between the above rankings are $d_B(\boldsymbol{a},\boldsymbol{b}) = 1$, $d_B(\boldsymbol{a},\boldsymbol{c}) = 1$, $d_B(\boldsymbol{a},\boldsymbol{d}) = 8.5$, $d_B(\boldsymbol{b},\boldsymbol{c}) = 2$, $d_B(\boldsymbol{b},\boldsymbol{d}) = 9.5$, and $d_B(\boldsymbol{c},\boldsymbol{d}) = 9.5$.

Notice that $d_B(\boldsymbol{b},\boldsymbol{d})$, and $d_B(\boldsymbol{c},\boldsymbol{d})$ are bigger[3] than $d_B(\boldsymbol{a},\boldsymbol{d})$. Therefore, using Cook-Kress-Seiford's distance, one might conclude that there is a lower level of agreement between $\boldsymbol{b}$ (alternatively, $\boldsymbol{c}$) and $\boldsymbol{d}$ than between $\boldsymbol{a}$ and $\boldsymbol{d}$. We believe that this is incorrect. In particular, we believe that the level of agreement between $\boldsymbol{b}$ (alternatively, $\boldsymbol{c}$) and $\boldsymbol{d}$ should be equal to the level of agreement between $\boldsymbol{a}$ and $\boldsymbol{d}$. We believe this because the only difference between $\boldsymbol{b}$ (alternatively, $\boldsymbol{c}$) and $\boldsymbol{a}$ is that $\boldsymbol{a}$ ties objects 1 and 3; while $\boldsymbol{b}$ prefers 1 over 3 (alternatively, $\boldsymbol{c}$ prefers 3 over 1).

[3] Recall that with Bogart's distance we had the opposite situation.

In general, the "behavior" of Cook-Kress-Seiford's distance, can be generalized in a similar way that Bogart's distance 'behavior" was generalized in section 2.3.3. The difference being that the consensus ranking obtained by using Cook-Kress-Seiford's distance will tend to be a ranking containing an "artificially high" amount of ties. That is, even objects that are expected not to be tied by the consensus ranking will be tied by it; this is illustrated in section 2.3.3. We close this section with the following remark: Our purpose is not to criticize Cook-Kress-Seiford's distance (a legitimate distance between partial orders); our intention is to show that it is not appropriate to use Cook-Kress-Seiford's to solve the IAP. (Unlike our NPKS distance, Cook-Kress-Seiford's distance was not designed to solve the IAP.)

Review and discussion of the (not-normalized) projected Kemeny-Snell distance

In [24], Dwork et. al. extended the Kendall Tau distance (a distance between complete without-ties rankings that is equivalent to the Kemeny-Snell distance in the space of complete without-ties rankings) to a distance between incomplete without-ties rankings. Dwork et. al. refer to this distance as the *induced Kendall Tau distance*, but, for reasons that will become clear later, we refer to it as the *projected Kemeny-Snell distance*, or *PKS distance*. (Strictly speaking, the induced Kendall Tau distance does not allows rankings with ties, but the PKS distance does allows rankings with ties; if the rankings do not contain ties, then the two distances are equivalent.) Unlike Bogart and Cook et. al., Dwork et. al. did not articulated a set of axioms that characterized the PKS distance; here we show that the PKS distance can be derived axiomatically. Given a universe of objects V and two incomplete rankings $\boldsymbol{a}$ and $\boldsymbol{b}$, the PKS distance, is

$$d_{PKS}(\boldsymbol{a}, \boldsymbol{b}) = d_{KS}(\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B})}, \boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B})}) \tag{2.5}$$

Note that the PKS distance is simply a not-normalized version of the NPKS distance defined in section 2.3.2. From this observation and Theorem 2.3.5 it follows that:

Corollary 2.3.6. *The distance $d_{PKS}(\cdot,\cdot)$ is the unique distance satisfying simultaneously Axioms 0 to 5 and Axiom K6.*

The only difference between the sets of axioms uniquely satisfied by $d_{PKS}(\cdot,\cdot)$ and $d_{PKS}(\cdot,\cdot)$, respectively, is the scaling axiom; in particular $d_{PKS}(\cdot,\cdot)$ satisfies Axiom K6, while $d_{NPKS}(\cdot,\cdot)$ satisfies Axiom 6'.

When using the PKS distance to find the consensus ranking it tends to favor incomplete rankings that compare a large number of objects over incomplete rankings that compare a small number of objects. The reason is the following. Recall that the consensus ranking is the complete ranking with the minimum sum of distances to the given incomplete rankings. Since the PKS distance between two rankings $\boldsymbol{a}$ and $\boldsymbol{b}$ is calculated by projecting the two rankings to the set $\mathcal{A}\bigcap\mathcal{B}$, the distances in the objective function are taken over object-spaces of possibly different dimensions. Therefore, since distances in higher dimensional spaces tend to be bigger than distances in lower dimensional spaces, the objective function will be dominated by the distances from the consensus ranking to the rankings that compare the highest number of objects. This is illustrated in section 2.3.3.

Comparison of the consensus rankings

In this section three examples are given illustrating that the consensus ranking obtained when using the NPKS distance are more intuitive than those obtained by Bogart's distance, Cook-Kress-Seiford's distance and the PKS distance.

Table 2.1 illustrates that the consensus ranking obtained by using Bogart's distance tends to be a ranking containing an "artificially low" amount of ties. That is, even objects that are expected to be tied by the consensus ranking are not tied by it.

Table 2.1: Given rankings $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ (each ranking ties a pair of objects), it is reasonable to expect that the consensus ranking ties all objects. However, using Bogart's distance any possible strict ranking (that is, any possible permutation of the objects) is a consensus ranking. In other words, using Bogart's distance, all 3-object strict rankings have the same sum of distances to rankings $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$; and such sum of distances is smaller than the sum of distances from the ranking tying all objects to rankings $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$. Note that all other consensus rankings, including the one obtained using our d_{NPKS} distance agree with the "intuitive" consensus rankings.

	Judge's rankings			**Consensus ranking using distance**			
	$\boldsymbol{a}$	$\boldsymbol{b}$	$\boldsymbol{c}$	d_B	d_C	d_{PKS}	d_{NPKS}
Object 1	•	1	1	Any	1	1	1
Object 2	1	•	1	possible	1	1	1
Object 3	1	1	•	permutation	1	1	1

Table 2.2 illustrates that the consensus ranking obtained by using Cook-Kress-Seiford's distance tends to be a ranking containing an "artificially high" amount of ties. That is, even objects that are expected not to be tied by the consensus ranking are tied by it.

Table 2.2: Given rankings $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$, it is reasonable to expect that the consensus ranking is $(1, 2, 3, 4)$. However, using Cook-Kress-Seiford's distance the consensus ranking ties all of the objects. Note that all other consensus rankings, including the one obtained using our d_{NPKS} distance agree with the "intuitive" consensus rankings.

	Judge's rankings			**Consensus ranking using distance**			
	$\boldsymbol{a}$	$\boldsymbol{b}$	$\boldsymbol{c}$	d_B	d_C	d_{PKS}	d_{NPKS}
Object 1	1	1	•	1	1	1	1
Object 2	2	2	•	2	1	2	2
Object 3	•	3	1	3	1	3	3
Object 4	•	4	2	4	1	4	4

Table 2.3 illustrates that the consensus ranking obtained by using the PKS distance tends to agree with the ranking that compares the largest amount of objects. (Indeed this is also the case of the consensus ranking obtained using Bogart's distance.)

Table 2.3: Given rankings $\boldsymbol{a}$ to $\boldsymbol{j}$, it is reasonable to expect that the consensus ranking is $(1, 2, 3, 4, 5)$—for each pair of objects $(i, i + 1)$, a clear majority of judges prefer i over $i + 1$. However, using the PKS distance, the consensus ranking is $(5, 2, 3, 4, 1)$; this ranking contradicts what the rankings $\boldsymbol{b}$ to $\boldsymbol{j}$ collectively imply: object 1 is the best, and object 5 the worst. Moreover the ranking $(5, 2, 3, 4, 1)$ contradicts the fact that a majority of judges prefer object 1 over 2 and a majority of judges prefer object 4 over 5. Note that, since Bogart's distance is also not-normalized, it also has the same un-intuitive consensus ranking as the PKS distance.

	Judge's rankings						**Consensus ranking using distance**			
	$\boldsymbol{a}$	$\boldsymbol{b}$&$\boldsymbol{c}$	$\boldsymbol{d}$&$\boldsymbol{e}$	$\boldsymbol{f}$&$\boldsymbol{g}$	$\boldsymbol{h}$&$\boldsymbol{i}$	$\boldsymbol{j}$	d_B	d_C	d_{PKS}	d_{NPKS}
Object 1	5	1	•	•	•	•	5	1	5	1
Object 2	4	2	1	•	•	1	2	1	2	2
Object 3	3	•	2	1	•	2	3	1	3	3
Object 4	2	•	•	2	1	3	4	1	4	4
Object 5	1	•	•	•	2	•	1	1	1	5

2.3.4 Finding a consensus ranking

In this section we study the optimization problem that needs to be solved in order to find the consensus ranking. In all of the group-ranking methods and problems reviewed so far, the optimization problem that needs to be solved in order to get the consensus ranking, $\boldsymbol{r}$, is

$$\min_{\boldsymbol{r}} \sum_{k=1}^{K} d(\boldsymbol{a}^k, \boldsymbol{r}), \tag{2.6}$$

where, $\boldsymbol{a}^1, ..., \boldsymbol{a}^K$ are the judges' rankings, the minimum is over all the complete (strict and non-strict) rankings, and $d(\cdot, \cdot)$ is an appropriate distance.

In the complete-ranking aggregation problem (CAP) each of the judges' rankings is a complete (strict or non-strict) ranking and the distance in problem (2.6) is Kemeny-Snell's distance, $d_{KS}(\cdot, \cdot)$. Bartholdi et. al. [6] showed that, when restricting the input to be complete *without-ties* rankings, the CAP is NP-hard. Since the CAP contains as a special case the restricted-to-no-ties CAP, then CAP is also NP-hard.

In the incomplete-ranking aggregation problem (IAP) each of the judges' rankings in problem (2.6) is a possibly incomplete (strict or non-strict) ranking. Since Bogart's distance, Cook-Kress-Seiford's distance, the PKS distance, and our NPKS distance all are equivalent (up to scaling factors) to the Kemeny-Snell distance in the space of complete rankings, then it follows that IAP, using any of these distances, is also NP-hard.

Section 2.3.4 provides a heuristic approach, and Section 2.3.4 provides an exact solution approach. Both of this approaches are applicable to solve problem (2.6) using any of these distances: Kemeny-Snell's distance, the PKS distance and our NPKS distance.

A heuristic method

In this section we give a heuristic to solve the IAP.

Given a set of objects to be ranked, $V = \{1, ..., n\}$, and a set of incomplete rankings, $\{\boldsymbol{a}^k\}_{k=1}^K$, the IAP is to find the complete ranking, $\boldsymbol{r}$, that minimizes the sum of NPKS distances to the given rankings:

$$\min_{\boldsymbol{r}} \sum_{k=1}^{K} d_{NPKS}(\boldsymbol{a}^k, \boldsymbol{r}). \tag{2.7}$$

Using the definition of the NPKS distance, the mathematical programming formulation of problem (2.7) is

$$\begin{aligned}
\min_{\boldsymbol{r}} \quad & \sum_{k=1}^{K} \mathcal{D}^k \sum_{i \in A^k} \sum_{j \in A^k} \frac{1}{2} \left| \operatorname{sign}(r_i - r_j) - \operatorname{sign}(a_i^k - a_j^k) \right| && (2.8a) \\
\text{s.t.} \quad & 1 \leq r_i \leq n \qquad \text{for } i = 1, ..., n && (2.8b) \\
& r_i \in \mathbb{Z} \qquad \text{for } i = 1, ..., n, && (2.8c)
\end{aligned}$$

where $\mathcal{D}^k = \left(\left|B^k\right|^2 - \left|B^k\right|\right)^{-1}$ is a normalization factor which depends only on the number of objects ranked in $\boldsymbol{a}^k$. Note that in problem (2.8) two feasible solutions might represent the same ranking. For example, with $V = \{1, 2, 3\}$ the ranking $(2, 1, 2)$, that places object 2 in first place and objects 1 and 3 tied in second place, can be represented as $\boldsymbol{r} = (2, 1, 2)$, $\boldsymbol{r}' = (3, 2, 3)$ and $\boldsymbol{r}'' = (3, 1, 3)$. This representation multiplicity is unimportant since, from equation (2.8a), all representations of a given ranking have the same objective value.

As shown earlier, problem (2.7), equivalently problem (2.8), is NP-hard. Here we propose a heuristic approach to solve problem (2.8). This approach is motivated by the following result [44]: Integer-separable convex programming problems with totally unimodular (TUM) constraints are solvable in polynomial time. To be able to use this result, problem (2.8) is reformulated first into an integer-separable non-convex programming problem with TUM constraints (see problem (2.9)).

$$\min_{\boldsymbol{r},\boldsymbol{z}} \quad \sum_{k=1}^{K} \mathcal{D}^k \sum_{i\in A^k}\sum_{j\in A^k} \frac{1}{2}\left|\text{sign}(z_{ij}) - \text{sign}(a_i^k - a_j^k)\right| \tag{2.9a}$$

$$\text{s.t.} \quad z_{ij} = r_i - r_j \qquad \text{for } i,j = 1,...,n \tag{2.9b}$$

$$1 \le r_i \le n \qquad \text{for } i = 1,...,n \tag{2.9c}$$

$$r_i \in \mathbb{Z} \qquad \text{for } i = 1,...,n. \tag{2.9d}$$

Problem (2.9), which is equivalent to problem (2.8), has a separable objective function and has TUM constraints. Since equation (2.9a) is not convex, we cannot solve problem (2.9) efficiently. Next we will convexify the objective function of problem (2.9). For this purpose the following functions are defined.

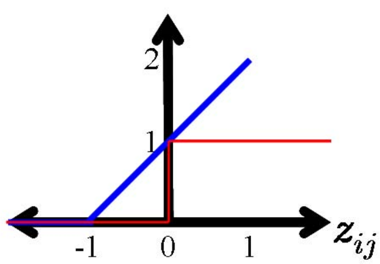
Let $f_{ij}^k(z_{ij})$ and $h_{ij}^k(z_{ij})$ for $k = 1,..,K$, $i = 1,..,n$, $j = 1,..,n$, be defined as

$$f_{ij}^k(z_{ij}) = \frac{1}{2}\left|\text{sign}(z_{ij}) - \text{sign}(a_i^k - a_j^k)\right|$$

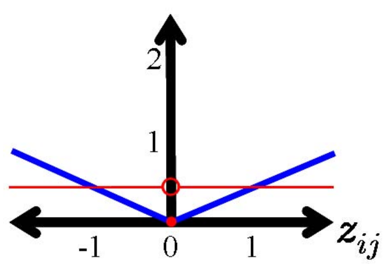
$$h_{ij}^k(z_{ij}) = \begin{cases} \max\left\{0, \frac{z_{ij}+1}{2}\right\} & \text{if } \text{sign}(a_i^k - a_j^k) = -1 \\ \max\left\{\frac{-z_{ij}}{2}, \frac{z_{ij}}{2}\right\} & \text{if } \text{sign}(a_i^k - a_j^k) = 0 \\ \max\left\{\frac{1-z_{ij}}{2}, 0\right\} & \text{if } \text{sign}(a_i^k - a_j^k) = 1. \end{cases} \tag{2.10}$$

Note that each $f_{ij}^k(z_{ij})$ corresponds to one of the terms in equation (2.9a), and that each $h_{ij}^k(z_{ij})$ is a convex function. As Figure 2.1 shows, each $h_{ij}^k(z_{ij})$ is a closely approximates the corresponding $f_{ij}^k(z_{ij})$.

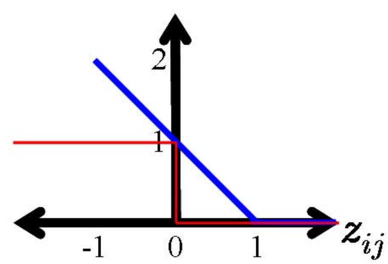
Figure 2.1: Relationship between $\frac{1}{2}|\,\text{sign}(z_{ij}) - \text{sign}(r_i^k - r_j^k)|$ (red) and $h_{ij}^k(z_{ij})$ (blue)



(a) $\text{sign}(r_i^k - r_j^k) = -1$



(b) $\text{sign}(r_i^k - r_j^k) = 0$



(c) $\text{sign}(r_i^k - r_j^k) = 1$

The convexified version of problem (2.9) is

$$\min_{\boldsymbol{x},\,\boldsymbol{z}} \quad \sum_{k=1}^{K} \mathcal{D}_k \sum_{i\in B^k} \sum_{j\in B^k} h_{ij}^k(z_{ij}) \tag{2.11a}$$

$$\text{s.t.} \quad z_{ij} = r_i - r_j \qquad \text{for } i = 1,...,n; j = 1,...,n \tag{2.11b}$$

$$1 \le r_i \le n \qquad \text{for } i = 1,...,n \tag{2.11c}$$

$$r_i \in \mathbb{Z} \qquad \text{for } i = 1,...,n. \tag{2.11d}$$

Since problem (2.9) is an integer-separable convex programming problems with (TUM) constraints, it follows from [44] that it is solvable in polynomial time. More precisely, problem (2.11) is a special case of the separation-deviation problem (see Section 1.2.1), and therefore it is solvable in $O(nm\log(\frac{n^2}{m})\log(n))$ time, where n is the number of objects and m is the total number of object-pairs that are compares by at least one judge.

Next we analyze the quality of the ranking obtained by solving the polynomial-time-solvable problem (2.11), as compared to the consensus ranking, obtained by solving the NP-hard problem (2.9) (equivalently, problem (2.7)). For this purpose we give the following definition.

Definition 2.3.7. *Given a minimization problem* P *with objective function* $f(\cdot)$*, a polynomial time algorithm,* A*, is said to be a* δ-approximation algorithm *for* P *if the following condition is satisfied:*

For every problem instance of P *with an optimal solution* x^**,* A *delivers a feasible solution* x *satisfying,* $f(x) \le \delta f(x^*)$*.*

Theorem 2.3.8. *Solving problem* (2.11) *is an* $(n-1)$*-approximation algorithm for problem* (2.9)*.*

Proof. Let $\boldsymbol{r}'$ and $\boldsymbol{r}^*$ be the optimal solutions to problems (2.11) and (2.9), respectively. Let $H(\cdot)$ and $F(\cdot)$ be the objective functions (as function of a given ranking) of problems (2.11) and (2.9), respectively.

From equations (2.9b) to (2.9d), (equivalently, (2.11b) to (2.11d)) we note that we are only interested in evaluating $f_{ij}^k(z_{ij})$ and $h_{ij}^k(z_{ij})$ in the points contained in the set $S \triangleq \{-(n-1), -(n-2), ..., n-2, n-1\}$. For all $i, j = 1, ..., n$, $k = 1, ..., K$ and z_{ij} in S, we make the following observations:

1. $h_{ij}^k(z_{ij}) \ge f_{ij}^k(z_{ij})$.
2. $h_{ij}^k(z_{ij}) = 0$ if and only if $f_{ij}^k(z_{ij}) = 0$.
3. $h_{ij}^k(z_{ij}) > f_{ij}^k(z_{ij})$ if and only if $\text{sign}(z_{ij}) \ne \text{sign}(a_i^k - a_j^k)$.
4. $h_{ij}^k(z_{ij}) \le (n-1)f_{ij}^k(z_{ij})$

We have the following series of inequalities, from which the result follows,

$$F(\boldsymbol{r}') \le H(\boldsymbol{r}') \le H(\boldsymbol{r}^*) \le (n-1)F(\boldsymbol{r}^*), \tag{2.12}$$

where the first inequality follows from Observation 1; the second inequality is valid since, with respect to problem (2.11), $\boldsymbol{r}'$ and $\boldsymbol{r}^*$ are an optimal and a feasible solution, respectively; and third inequality follows from Observation 4. □

Theorem 2.3.8 implies that, if there exists a ranking that agrees completely with all of the given incomplete rankings, then such a ranking is obtained by solving problem (2.11). Formally,

Corollary 2.3.9. *The optimal solution to problem* (2.11) *has an objective value of zero if and only if the consensus ranking (optimal solution to problem* (2.7)*) has a total NPKS distance of zero to the given incomplete rankings.*

We close this section by noting that the solutions to problem (2.11), for all of the examples given in Tables 2.1, 2.2 and 2.3 coincide with the consensus ranking (optimal solution to problem (2.7)).

An exact method

This section gives an exact method to solve the IAP. For this purpose a new problem is introduced: the *strict-IAP* is a problem identical to the IAP, except that in the strict-IAP the solution space is restricted to the set of all complete strict rankings. The herein-proposed exact method to solve the IAP can be summarized as follows: find an optimal solution for the strict-IAP, and then use the solution found for the strict-IAP as a starting point to find an optimal solution for the IAP. Throughout this section $r^{(s)}$ denotes an optimal solution to the strict-IAP, and $\boldsymbol{r}^*$ denotes an optimal solution to the IAP.

Next, the problem of finding $r^{(s)}$ is reduced to the *minimum weight feedback arc set problem* (FASP).

Definition 2.3.10. *A feedback arc set for a directed graph $G = (V, A)$ is a subset $A' \subseteq A$ such that A' contains at least one arc from every directed cycle in G. The minimum weight feedback arc set problem is defined a follows: Given a directed graph $G(V, A)$ with a weight assigned to each arc, find a feedback arc set, A^*, with minimum sum of weights.*

The preferences expressed in an incomplete ranking $\boldsymbol{a}$ can be represented as a directed graph as follows: we have a node for each object, and we have arc(s)

$$\begin{cases} (i,j) & \text{if } i \text{ is preferred to } j, \\ (j,i) & \text{if } j \text{ is preferred to } i, \\ (i,j) \text{ and } (j,i) & \text{if } i \text{ and } j \text{ are tied.} \end{cases}$$

If i and/or j are not ranked in $\boldsymbol{a}$, then we do not have any arc between nodes i and j. Note that a complete strict ranking corresponds to a directed acyclic graph containing exactly $\binom{|V|}{2}$ arcs; i.e., containing exactly one of (i, j) and (j, i).

Given a set of objects to be ranked, $V = \{1, ..., n\}$, and a set of incomplete rankings, $\{\boldsymbol{a}^k\}_{k=1}^K$, we find $r^{(s)}$ by solving the FASP on the following weighted directed graph $G = (V, A)$: We have a node for each object, and each arc (i, j) has a weight

$$w_{ij} = \sum_{k=1}^{K} w_{ij}^k, \tag{2.13a}$$

$$\text{where,} \tag{2.13b}$$

$$w_{ij}^k = \begin{cases} \frac{1}{(|\mathcal{A}^k|^2 - |\mathcal{A}^k|)/2} & \text{if } i \text{ is preferred to } j \text{ in } \boldsymbol{a}^k, \\ \frac{1}{2}\frac{1}{(|\mathcal{A}^k|^2 - |\mathcal{A}^k|)/2} & \text{if } i \text{ and } j \text{ are tied in } \boldsymbol{a}^k, \\ 0 & \text{if } i \text{ and/or } j \text{ are not ranked in } \boldsymbol{a}^k. \end{cases} \tag{2.13c}$$

The preferences in $r^{(s)}$ are represented by the arcs on the complement of $A*$—the minimum weight feedback arc set in G. To see that this is statement is true, we note that, from the interpretation of Kemeny-Snell's distance (see §2.2.1, and the definition of the NPKS ditance (equation (2.4)) the NPKS distance can be interpreted as follows: The NPKS distance between two incomplete rankings $\boldsymbol{a}$ and $\boldsymbol{b}$ is given by the sum of the weights of the rank reversals between them, where each rank reversal has a weight of $\frac{1}{(|\mathcal{A}\bigcap\mathcal{B}|^2 - |\mathcal{A}\bigcap\mathcal{B}|)/2}$. Therefore, $r^{(s)}$ is the complete strict ranking with the minimum sum of the weights of the rank reversals between $\boldsymbol{r}^{(s)}$ and each of the given incomplete rankings $\{\boldsymbol{a}^k\}_{k=1}^K$. Now, since if arc (i, j) is in A^*, then $\boldsymbol{r}^{(s)}$ prefers j over i[4], and thus $\boldsymbol{r}^{(s)}$ has a rank reversal with every given incomplete ranking that prefers i over j; the sum of the weights of all these rank reversals is exactly the weight of arc (i, j) (see equation (2.13)). We conclude that, by minimizing the weights of the arcs in the feedback arc set, we are minimizing the weights of the rank reversals that $\boldsymbol{r}^{(s)}$ has with the given incomplete rankings.

The FASP is NP-hard [30] and approximable within $O(\log |V| \log\log |V|)$ [25], and the FASP can be solved using the implicit hitting set (IHS) approach as described in [47]. Since this approach is practical for solving the maximum-weight trace problem, a problem that is closely related to the FASP, we believe that the IHS approach will be able to solve practical-size problems within a reasonable amount of time.

Given a complete strict ranking $\boldsymbol{a}$, its *neighborhood* $N(\boldsymbol{a})$ is the set of rankings that includes $\boldsymbol{a}$ and all complete rankings that only differ from $\boldsymbol{a}$ by tying objects that are ranked consecutively in $\boldsymbol{a}$. It seems reasonable to expect that given an optimal solution to the strict-IAP, $r^{(s)}$, there would exist an optimal solution to the IAP in the neighborhood of $r^{(s)}$. If this were the case, then a straight forward dynamic programming algorithm (very similar to the dynamic programming algorithm for parenthesizing matrix-chain multiplications given in page 331 of [22]) would explore in polynomial time the neighborhood of $r^{(s)}$ and find an optimal solution to the IAP. Unfortunately, as the following example shows, the above

[4] Since G is a complete graph it can be proved that A^* will contain exactly one of (i, j) and (j, i).

statement is not true. Given the rankings $\boldsymbol{a}^1 = (1, 2, 3, 4)$ and $\boldsymbol{a}^2 = (1, \bullet, \bullet, 1)$, the optimal solution to the strict-IAP is $r^{(s)} = (1, 2, 3, 4)$ and the optimal solutions to the IAP are $\{(1, 1, 2, 1), (1, 2, 3, 1), (2, 1, 2, 2), (2, 1, 3, 2), (3, 1, 2, 3)\}$; none of which is in the neighborhood of $r^{(s)}$. It remains an open question if the statement at the beginning of this paragraph is true when all the given rankings are complete; that is, for the complete-ranking aggregation problem.

2.3.5 Obtaining meaningful results when aggregating incomplete rankings

In the incomplete-ranking aggregation problem, the following problem may arise: if each reviewer only ranks a strict subset of A, then it is possible that some pairs are not comparable. In particular, consider the problem where there are only two reviewers and the first one ranks the alternatives in $A_1 = \{a_1, a_2, ..., a_k\}$, while the second reviewer ranks the alternatives in $A_2 = \{a_{k+1}, ..., a_M\}$. It is clear that any of the alternatives in A_1 is not comparable with any of the alternatives in A_2. Thus no ranking aggregation method can provide a meaningful aggregate ranking (recall that by definition we require that the aggregate ranking is a complete ranking). This issue is formalized as follows: A *comparability graph* $G = (V, E)$ is a graph where each node corresponds to an alternative in A, and for any given pair of nodes a_i, a_j, the edge $[a_i, a_j]$ is included in E if and only if the corresponding pair of alternatives are ranked by at least one common reviewer. A pair of alternatives a_i, a_j is said to be comparable if there exists a path from a_i to a_j in G. Thus, in order for the aggregate ranking to be meaningful, it is necessary that all pairs of alternatives are comparable. We refer to [59, 17, 18, 40] for possible solutions to the problem of allocating the subsets of objects to be ranked to individual reviewers in order to satisfy this condition (as well as other stronger conditions that guarantee "higher" levels of comparability).

2.4 Incomplete-rating aggregation

2.4.1 Axioms for a distance between incomplete ratings

In this section Cook and Kress' axioms (given in Section 2.2.2) are modified in order to obtain a set of axioms appropriate for a distance between incomplete ratings. Moreover, we want to design a distance, $d(\cdot, \cdot)$, that is suitable for solving the incomplete-rating aggregation problem.

Recall that given the incomplete ratings of K judges, $\{\boldsymbol{a}^1, ..., \boldsymbol{a}^K\}$, the consensus rating $\boldsymbol{r}$ is the optimal solution to the following optimization problem:

$$\min_{\boldsymbol{r}} \sum_{k=1}^{K} d(\boldsymbol{a}^k, \boldsymbol{r}), \tag{2.14}$$

where the minimum is over all complete ratings.

The aim of the axioms proposed in this section is to describe a distance between incomplete ratings such that, when this distance is used in problem (2.14), the consensus rating minimizes the disagreement of the judges. In contrast to the previous section, here we do not give a thorough explanation of the modifications made to Cook-Kress' axioms. This is because the modifications to Cook-Kress' axioms are almost identical to the modifications to Kemeny-Snell's axioms explained in the previous section.

The axioms are denoted using the number of the corresponding Cook-Kress axiom, and a hat ($\hat{}$) is added over the number to differentiate this axioms from those in section 2.3.1. In addition, the axioms denoted by a prime are those modified slightly so that they applied to distances between incomplete ratings. For example, Axiom $\hat{0}$ (below) has no corresponding Cook-Kress axiom; Axiom $\hat{1}$ (below) corresponds exactly to Axiom C1; and Axiom $\hat{3}$' (below) is a slight variant of Axiom Axiom C3.

In order to specify our axioms, the following concepts are defined.

Definition 2.4.1. *Given a rating* $\boldsymbol{a}$ *and a subset* S *of the object universe* V*, the* projection *of* $\boldsymbol{a}$ *on* S*, denoted as* $\boldsymbol{a}|_S$*, is the rating of the objects in* S *that preserves the scores assigned by* $\boldsymbol{a}$ *to the objects in* S*.*

Definition 2.4.2. *A pair of ratings* $\boldsymbol{a}$ *and* $\boldsymbol{b}$ *on a set of objects* V *are* opposite ratings*, if (1)* $\boldsymbol{a}$ *rates* $|V|/2$ *objects (or* $\lceil |V|/2 \rceil$ *if* $|V|$ *is odd) with the highest possible score, (2)* $\boldsymbol{b}$ *rates those objects with the lowest possible score, (3)* $\boldsymbol{a}$ *rates the remaining objects with the lowest possible score, and (4)* $\boldsymbol{b}$ *rates the remaining objects with the highest possible score.*

Intuitively, two opposite ratings are ratings in total disagreement. We now specify the axioms that our distance must satisfy.

- **Axiom $\hat{\mathbf{0}}$** (Relevance) $d(\boldsymbol{a},\boldsymbol{b}) = d(\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B})}, \boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B})})$
- **Axiom $\hat{\mathbf{1}}$** (Nonnegativity) $d(\boldsymbol{a},\boldsymbol{b}) \geq 0$.
- **Axiom $\hat{\mathbf{2}}$** (Commutativity) $d(\boldsymbol{a},\boldsymbol{b}) = d(\boldsymbol{b},\boldsymbol{a})$.
- **Axiom $\hat{\mathbf{3}}$'** ("Relaxed" triangular inequality) $d(\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}, \boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}) + d(\boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}, \boldsymbol{c}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}) \geq d(\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}, \boldsymbol{c}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})})$, and equality holds if and only if $\boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}$ is between for incomplete ratings $\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}$ and $\boldsymbol{c}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}$.
- **Axiom $\hat{\mathbf{4}}$** (Proportionality) The distance between any two adjacent[5] ratings is proportional to the degree of adjacency.
- **Axiom $\hat{\mathbf{5}}$'** (Normalization) $d(\boldsymbol{a},\boldsymbol{b}) \leq 1$; and $d(\boldsymbol{a},\boldsymbol{b}) = 1$ if and only if $\boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B})}$ and $\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B})}$ are opposite ratings.

[5]On the natural extension of adjacency and degree of adjacency (Definitions 2.2.4 and 2.2.5) for incomplete ratings.

2.4.2 Uniqueness and existence of a distance

This section shows that the distance between incomplete rankings given in equation (2.15), here called the *normalized projected Cook-Kress distance*, or simply NPCK distance, satisfies Axioms $\hat{0}$ to $\hat{5}$. Moreover, we prove that the NPCK distance is the unique distance that simultaneously satisfies Axioms $\hat{0}$ to $\hat{5}$.

$$d_{NPCK}(\boldsymbol{a},\boldsymbol{b}) = \begin{cases} \frac{d_{NPCK}(\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B})},\boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B})})}{4\cdot R\cdot\left\lceil\frac{|A\bigcap B|}{2}\right\rceil\cdot\left\lfloor\frac{|A\bigcap B|}{2}\right\rfloor} & \text{if } |\mathcal{A}\bigcap\mathcal{B}|\geq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{2.15}$$

Lemma 2.4.3. *Given two complete ratings $\boldsymbol{a}$ and $\boldsymbol{b}$ on a set of objects V, $d_{CK}(\boldsymbol{a},\boldsymbol{b})$ attains its maximum of $4\cdot R\cdot\left\lceil\frac{|A\bigcap B|}{2}\right\rceil\cdot\left\lfloor\frac{|A\bigcap B|}{2}\right\rfloor$ when $\boldsymbol{a}$ and $\boldsymbol{b}$ are opposite ratings.*

Lemma 2.4.4. *The NPCK distance satisfies Axioms $\hat{0}$ to $\hat{5}$.*

Proof. It follows directly from equation (2.15) that $d_{NPCK}(\cdot,\cdot)$ satisfies Axiom $\hat{0}$.

The nonnegativity of $d_{NPKS}(\cdot,\cdot)$ follows from equation (2.15) and the nonnegativity of $d_{CK}(\cdot,\cdot)$.

Axiom $\hat{2}$ follows from equation (2.15) and the fact that $d_{CK}(\cdot,\cdot)$ satisfies Axiom C2.

Axiom $\hat{3}$' follows from equation (2.15); the fact that $\boldsymbol{a}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}$, $\boldsymbol{b}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}$ and $\boldsymbol{c}|_{(\mathcal{A}\bigcap\mathcal{B}\bigcap\mathcal{C})}$ are complete ratings; and the fact that $d_{CK}(\cdot,\cdot)$ satisfies Axiom C3.

From Lemma 2.4.3 and equation (2.15) we have that $d_{NPCK}(\cdot,\cdot)$ satisfies Axiom $\hat{5}$'. □

Corollary 2.4.5. *Axioms $\hat{0}$ to $\hat{5}$ are consistent.*

Next, we show that Axioms $\hat{0}$ to $\hat{5}$ uniquely determine the NPCK distance.

Theorem 2.4.6. *The distance $d_{NPCK}(\cdot,\cdot)$ is the unique distance satisfying Axioms $\hat{0}$ to $\hat{5}$ simultaneously.*

Proof. This proof follows the exact same arguments used in the proof of Theorem 2.3.5. □

2.4.3 Finding a consensus rating

The NPCK distance generalizes the distance between complete ratings proposed in [20]. Hochbaum and Levin [41] showed that the complete-rating aggregation problem proposed in [20] is a special case of their own separation-deviation model, and thus efficiently solvable. Similarly, the incomplete-rating aggregation problem using the NPCK distance is a special case of the separation-deviation problem and can be reformulated as follows:

$$\min_{\boldsymbol{x},\,\boldsymbol{z}} \quad \sum_{k=1}^{K} \sum_{i\in\mathcal{A}^k} \sum_{j\in\mathcal{A}^k} \frac{\left|z_{ij} - p_{ij}^k\right|}{4 \cdot R \cdot \left\lceil \frac{\left|\mathcal{A}^k\right|}{2} \right\rceil \cdot \left\lfloor \frac{\left|\mathcal{A}^k\right|}{2} \right\rfloor} \tag{2.16a}$$

$$\text{subject to} \quad z_{ij} = x_i^{(c)} - x_j^{(c)} \quad i = 1,\ldots,n; \quad j = 1,\ldots,n \tag{2.16b}$$

$$\ell \leq x_i^{(c)} \leq u \quad i = 1,\ldots,n \tag{2.16c}$$

$$x_i^{(c)} \in \mathbb{Z} \quad i = 1,\ldots,n. \tag{2.16d}$$

Chapter 3

Country credit-risk rating aggregation

3.1 Introduction

Country credit-risk ratings quantify the risk associated with investing in a given country. Haque et al. [35] define country credit-risk rating as an estimate of the probability that a country will fail to pay back the debt it has acquired. To satisfy increasing investors' needs for information on countries' creditworthiness, several agencies periodically publish country credit-risk ratings. Often there are differences between the agencies' credit-risk ratings for a particular country. It is therefore of interest to aggregate those differing views into a coherent rating that represents a group consensus capturing the different expertise of the rating agencies.

Aggregating credit-risk ratings is a scenario within group decision making. Group decision making concerns the problem of finding a group consensus from the expressed evaluations of K reviewers (i.e., the agencies) in relation to n objects (i.e., the countries). In this chapter we demonstrate that in the complete-rating aggregation problem, even when the input is given *only* as ratings $\{\boldsymbol{a}^k\}_{k=1}^{K}$, it is useful to consider also the implied separation gaps: $p_{ij}^k = a_i^k - a_j^k$. In particular, we show that the separation gaps mitigate the effect of inflated scores or shifts in evaluation scale.

We prove several properties of the separation-deviation model, including the property that the aggregate rating obtained by the separation-deviation model agrees with the majority of agencies or reviewers, regardless of the scale used. The analysis of the separation-deviation model here is for the complete-rating aggregation problem. .

The most commonly used method of rating aggregation is the *averaging method.* In this method the aggregate score of each country is the average of the scores that this country received from all of the reviewers. We assess the performance of the separation-deviation model and compare the model to the averaging method, using several performance measures. We demonstrate that for the complete-rating aggregation problem the aggregate rating, obtained by the separation-deviation model, has fewer rank reversals than the aggregate rating obtained by the averaging method.

The main contributions and results here are:

1. Illustrating the benefit of using the separation-deviation model in the credit-risk rating context.
2. Demonstrating that, even when the input is given *only* as ratings, it is useful to consider also the implied separation gaps: $p_{ij}^k = a_i^k - a_j^k$. In particular, the separation gaps are shown to mitigate the effect of inflated scores or shifts in evaluation scale.
3. Proving that the aggregate rating obtained by the separation-deviation model with absolute value penalty functions agrees with the majority of reviewers. This demonstrates the model's robustness in the presence of individual reviewer's manipulations.
4. Showing that the averaging method is a special case of the separation-deviation model with *uniform quadratic penalty functions*[1].
5. Presenting an experimental study showing that the aggregate rating obtained by the separation-deviation model with absolute value penalty functions has fewer *rank reversals* than the aggregate rating obtained by the averaging method. Informally, a rank reversal is a discrepancy in the relative order between a pair of objects when comparing the aggregate rating to the input ratings.
6. Using the separation-deviation model is shown to identify several outliers in the ratings of the agencies.

The chapter is organized as follows: Section 3.2 analyses the robustness properties of the SD model. Section 3.3 shows that the averaging method is a special of the separation-deviation model. Finally section 3.4 provides the details on the application of the separation-deviation model to country credit-risk aggregation.

3.2 Robustness of the separation-deviation model

3.2.1 Robustness of the absolute value separation problem

The *absolute value separation problem*, (||,Sep) is formulated as follows:

$$
\begin{aligned}
(||,\text{Sep}) \quad \min_{\boldsymbol{x},\boldsymbol{z}} \quad & \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=i+1}^{n} u_{ij}^k \left| z_{ij} - p_{ij}^k \right| && (3.1a) \\
\text{s.t.} \quad & z_{ij} = x_i - x_j \quad (i = 1, \ldots, n;\ j = i+1, \ldots, n) && (3.1b) \\
& x_1 = 0. && (3.1c)
\end{aligned}
$$

We denote an optimal solution to (||,Sep) as $\boldsymbol{x}^{|S|}$.

[1]the concept of uniform quadratic functions is defined in section 3.3

In this section we show that for the complete-rating aggregation problem (||,Sep) is robust in that it resists manipulation by a minority of the reviewers. For this purpose we prove that $\boldsymbol{x}^{|S|}$ agrees with the (weighted) majority of reviewers.

Theorem 3.2.1. *A rating* $\boldsymbol{x}$ *is said to be* equivalent under translation *to the rating* $\tilde{\boldsymbol{x}}$ *if there exists a constant* c*, such that* $x_i = \tilde{x}_i + c$ *for* $i = 1, \ldots, n$*.*

Definition 3.2.2. *A rating* $\boldsymbol{x}$ *is said to be* equivalent under translation *to the rating* $\tilde{\boldsymbol{x}}$ *if there exists a constant* c*, such that* $x_i = \tilde{x}_i + c$ *for* $i = 1, \ldots, n$*.*

The relation of *equivalence under translation* is reflexive, symmetric and transitive. As such, it partitions the set of ratings into equivalence classes.

The following lemma is needed in the proof of the property of "resistance to manipulation by a minority of reviewers". The problem analysed in the lemma is a special case of the weighted median on a line and the weighted median on a graph, which were studied extensively in [8, 28].

Lemma 3.2.3. *Given the optimization problem* $y^* = \text{argmin} \sum_{k=1}^{K} w^k |y - a^k|$*, where* $w^k \geq 0$ *for* $k = 1, 2, \ldots, K$*. If there is a weight* w^i *such that* $w^i > \frac{1}{2} \sum_{k=1}^{K} w^k$*, then the optimal solution to the problem is* $y^* = a^i$*.*

Proof. Suppose by contradiction that there exists an optimal solution of the form $y^{**} = a^i - \delta$ for some $\delta > 0$, then it follows from simple arithmetic calculations that $y^* = a^i$ has a strictly lesser objective value. The same holds for any solution of the form $y^{**} = a^i + \delta$ for some $\delta > 0$. □

Theorem 3.2.4. *For (||,Sep), if a subset* S *of reviewers has ratings equivalent under translation, and* S *is a weighted majority for every pair* i, j*, i.e.* $\sum_{k \in S} u_{ij}^k > \frac{1}{2} \sum_{k=1}^{K} u_{ij}^k$*, then* $\boldsymbol{x}^{|S|}$ *is equivalent under translation to rating of the weighted majority. I.e.* $\boldsymbol{x}^{|S|}$ *is equivalent under translation to every* $\boldsymbol{a}^k$*,* $k \in S$*.*

Proof. Omitting constraint (3.1b) decomposes the problem to several optimization problems, one for each z_{ij}. Each of these optimization problems,

$$z_{ij}^* = \text{argmin} \sum_{k=1}^{K} u_{ij}^k \left| z_{ij} - p_{ij}^k \right| \quad \text{for } i, j = 1, 2, \ldots, n,$$

is of the form described in Lemma 3.2.3. Since the reviewers in S have ratings equivalent under translation, we have that $p_{ij}^k = p_{ij}^S$ for all $k \in S$. Furthermore since S is a weighted majority, then $u_{ij}^S = \sum_{k \in S} u_{ij}^k > \sum_{k=1}^{K} u_{ij}^k$. Therefore, by Lemma 3.2.3, $z_{ij}^* = p_{ij}^S$. Finally, since (by construction) the separation gaps p_{ij}^S are consistent in the additive sense, it follows that by setting $x_1 = 0$ and $x_i = z_{i1}^* + x_1$ for $i = 2, \ldots, n$ we obtain a rating satisfying constraints (3.1b) and (3.1c). In particular this rating is equivalent under translation to all of the ratings of the reviewers in S. □

Corollary 3.2.5. *For problem* ($||$,*Sep) with two reviewers,* $K = 2$*, if all the penalty weights of reviewer* 1 *dominate the penalty weights of the reviewer* 2 *(i.e.* $u^1_{ij} > u^2_{ij}$ *for every pair* i, j*), then any optimal solution to* ($||$,*Sep) is an aggregate rating equivalent under translation to the rating of reviewer* 1*.*

Let the *unweighted absolute value separation problem*, ($||$,Sep,**1**), refer to ($||$,Sep) with $u^k_{ij} = 1$, for $i, j = 1, \ldots, n$ and $k = 1, \ldots, K$.

From Theorem 3.2.4 we have the following corollaries.

Corollary 3.2.6. *For* ($||$,*Sep,***1***), if a* simple *majority of reviewers has ratings equivalent under translation, then any optimal solution to* ($||$,*Sep,***1***) is an aggregate rating equivalent under translation to every rating of each of the reviewers in the majority.*

Corollary 3.2.7. *The problem* ($||$,*Sep,***1***) with two reviewers,* $K = 2$*, has an infinite number of optimal solutions. Two of the solutions are equivalent under translation to the (input) ratings of reviewer* 1 *and reviewer* 2*. And any convex linear combination of these two ratings is an optimal solution as well.*

In contrast to ($||$,Sep,**1**), the solution to the averaging method does not have the property of agreeing with the majority. An example shown in Table 3.1 demonstrates that a single reviewer (reviewer 3) can dominate the aggregate rating solution of the averaging method by manipulating his/her rating scale. The aggregate rating obtained by the averaging method is denoted in Table 3.1 as $\boldsymbol{x}^{Avg}$.

Table 3.1: Aggregate rating $\boldsymbol{x}^{Avg}$ obtained by the averaging method.

	Reviewer 1	**Reviewer 2**	**Reviewer 3**	$\boldsymbol{x}^{Avg}$
Object 1	1	1	13	5
Object 2	2	2	10	4.67
Object 3	3	3	7	4.33
Object 4	4	4	4	4
Object 5	5	5	1	3.66

Theorem 3.2.4 applies only when the penalty function used in (Sep) is the absolute value function (it applies exclusively for ($||$,Sep)) and cannot be extended to other convex penalty functions. It does not even hold for convex quadratic penalty functions, as shown in the example in Table 3.2, where the third reviewer dominates the aggregate rating even though reviewers 1 and 2 had the same ratings.

Table 3.2: Aggregate rating obtained by solving (Sep) with $f_{ij}^k(y) = y^2$ for all i, j, k.

	Reviewer 1	**Reviewer 2**	**Reviewer 3**	**Aggregate rating**
Object 1	1	1	25	4
Object 2	2	2	20	3
Object 3	3	3	15	2
Object 4	4	4	10	1
Object 5	5	5	5	0

3.2.2 Robustness of the absolute value separation-deviation problem

We define the absolute value separation-deviation problem, (||,Sep-Dev) as follows:

$$\text{(||,Sep-Dev)} \quad \min_{\boldsymbol{x},\boldsymbol{z}} \quad \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=i+1}^{n} u_{ij}^k \left|z_{ij} - p_{ij}^k\right| + \sum_{k=1}^{K}\sum_{i=1}^{n} v_i^k \left|x_i - a_i^k\right| \tag{3.2a}$$

$$\text{s.t.} \quad z_{ij} = x_i - x_j \quad (i = 1, \ldots, n;\ j = i+1, \ldots, n). \tag{3.2b}$$

Let an optimal solution to (||,Sep-Dev) by denoted by $\boldsymbol{x}^{|SD|}$.

In this section we show that, for the complete-rating aggregation problem and under certain restrictions, any optimal solution to (||,Sep-Dev) is an aggregate rating identical to the rating of the majority of reviewers.

Theorem 3.2.8. *For (||,Sep-Dev), if a subset S of reviewers has identical ratings $\boldsymbol{a}^{(S)}$ and S is a weighted majority for both the deviation and the separation terms (i.e. $\sum_{k\in S} u_{ij}^k > \frac{1}{2}\sum_{k=1}^{K} u_{ij}^k$ for any pair i, j and $\sum_{k\in S} v_i^k > \frac{1}{2}\sum_{k=1}^{K} v_i^k$ for every i), then $\boldsymbol{x}^{|SD|}$ is equal to $\boldsymbol{a}^{(S)}$.*

Proof. Omitting constraint (3.2b) decomposes the problem into the following optimization problems:

$$z_{ij}^* = \operatorname{argmin} \sum_{k=1}^{K} u_{ij}^k |z_{ij} - p_{ij}^k| \qquad \text{for } i = 1, \ldots, n;\ j = i+1, \ldots, n \tag{3.3}$$

$$x_i^* = \operatorname{argmin} \sum_{k=1}^{K} v_i^k |x_i - a_i^k| \qquad \text{for } i = 1, 2, \ldots, n. \tag{3.4}$$

All of the problems are of the form described in Lemma 3.2.3. Since the reviewers in S have identical ratings, we have that $p_{ij}^k = p_{ij}^S$ and $a_i^k = a_i^{(S)}$ for all $k \in S$. Therefore by Lemma 3.2.3 we have that $z_{ij}^* = p_{ij}^S$ and $x_i^* = a_i^{(S)}$. Since the separation gaps were derived from the ratings by setting $p_{ij}^k = a_i^k - a_j^k$, it follows that $z_{ij}^* = x_i^* - x_j^*$, and so constraints (3.2b) are satisfied. Therefore the optimal solution to the separation-deviation problem is an aggregate rating identical to all the ratings of the weighted majority of reviewers. □

Corollary 3.2.9. *For problem (||,Sep-Dev) with two reviewers, $K = 2$, if all the penalty weights of reviewer 1 dominate the penalty weights of the reviewer 2 (i.e. $u_{ij}^1 > u_{ij}^2$ for every pair i, j, and $v_i^1 > v_i^2$ for all i), then the optimal solution to (||,Sep-Dev) is an aggregate rating identical to the rating of reviewer 1.*

Let the *unweighted absolute value separation-deviation problem*, (||,Sep-Dev,**1**), refer to (||,Sep-Dev) with $u_{ij}^k = v_i^k = 1$, for $i, j = 1, \ldots, n$ and $k = 1, \ldots, K$.

From Theorem 3.2.8 we have the following corollaries.

Corollary 3.2.10. *For (||,Sep-Dev,**1**), if a* simple *majority of reviewers has identical ratings, then the optimal solution to (||,Sep-Dev,**1**) is an aggregate rating identical to the rating of the majority.*

Corollary 3.2.11. *The problem (||,Sep-Dev,**1**) with two reviewers, $K = 2$, has an infinite number of optimal solutions. Two of the solutions are identical to the (input) ratings of reviewer 1 and reviewer 2. And any convex linear combination of these two ratings is an optimal solution as well.*

Theorem 3.2.8 for (||,Sep-Dev) is weaker than the corresponding Theorem 3.2.4 for (||,Sep) in that it requires the ratings of the majority to be *identical* rather than just being *equivalent under translation*. Since there are $O(Kn^2)$ separation penalty terms and only $O(Kn)$ deviation penalty terms in the separation-deviation problem, one might think that it is possible to make Theorem 3.2.8 as strong as Theorem 3.2.4. The example shown in Table 3.3 proves that this is impossible.

Table 3.3: $\boldsymbol{x}^{|S|}$ is equivalent under translation to the rating of the majority, but $\boldsymbol{x}^{|SD|}$ is not.

	Reviewer 1	**Reviewer 2**	**Reviewer 3**	$\boldsymbol{x}^{\|S\|}$	$\boldsymbol{x}^{\|SD\|}$
Object 1	1	4	517	0	4
Object 2	2	5	3	1	3

Still, Table 3.3 data is a pathological instance of the problem. To demonstrate that we show in Table 3.4 that, with a minor perturbation in the data, $\boldsymbol{x}^{|SD|}$ is equivalent under translation to the ratings of the majority (reviewers 1 and 2).

Table 3.4: Both $\boldsymbol{x}^{|S|}$ and $\boldsymbol{x}^{|SD|}$ are equivalent under translation to the rating of the majority.

	Reviewer 1	**Reviewer 2**	**Reviewer 3**	$\boldsymbol{x}^{\|S\|}$	$\boldsymbol{x}^{\|SD\|}$
Object 1	1	4	516	0	4
Object 2	2	5	3	1	5

One might still prefer (||,Sep-Dev) to (||,Sep) since, even though it is only guaranteed to satisfy the weaker theorem, it tends to have an optimal solution on a 'similar' scale to the input ratings. An example illustrating this 'similarity' is shown in Table 3.5.

Table 3.5: $\boldsymbol{x}^{|SD|}$ is closer to the input ratings than $\boldsymbol{x}^{|S|}$.

	Reviewer 1	**Reviewer 2**	**Reviewer 3**	$\boldsymbol{x}^{\|SD\|}$	$\boldsymbol{x}^{\|S\|}$
Object 1	100	700	600	400	0
Object 2	200	800	500	500	100
Object 3	300	900	400	600	200

Table 3.5 provides an instance where $\boldsymbol{x}^{|S|}$ and $\boldsymbol{x}^{|SD|}$ are equivalent under translation to the ratings given by the majority of reviewers (i.e. reviewers 1 and 2). The advantage of $\boldsymbol{x}^{|SD|}$, is that its *scale* is closer to the scale used by the reviewers.

So far we have shown that (1) $\boldsymbol{x}^{|S|}$ is equivalent under translation to the majority rating; (2) (under stronger assumptions) $\boldsymbol{x}^{|SD|}$ is identical to the majority rating; and (3) depending on the choice of anchoring (generally $\boldsymbol{x}_1 = 0$), $\boldsymbol{x}^{|SD|}$ is closer than $\boldsymbol{x}^{|S|}$ to the scale of the input ratings. We next show that, with a minor adjustment to (||,Sep-Dev), we can obtain all of these desirable properties in a single model.

We note that (||,Sep-Dev) is a multi-objective problem. The first objective is to minimize the separation penalty, and the second objective is to minimize the deviation penalty. So far we have minimized an unweighted sum of these two (possibly conflicting) objectives. However we can obtain all of the desired properties by minimizing a weighted sum of the separation penalty and the deviation penalty. In particular we propose the following problem:

$$\text{(||,M·Sep-Dev)} \quad \min_{\boldsymbol{x},\boldsymbol{z}} \quad M \cdot \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=i+1}^{n} u_{ij}^k \left| z_{ij} - p_{ij}^k \right| + \sum_{k=1}^{K} \sum_{i=1}^{n} v_i^k \left| x_i - a_i^k \right| \tag{3.5a}$$

$$\text{s.t.} \quad z_{ij} = x_i - x_j \quad (i = 1, \ldots, n;\ j = i+1, \ldots, n), \tag{3.5b}$$

where M is a large enough number so that the separation penalty is *lexicographically* more important than the deviation penalty. By lexicographically more important we mean that the separation penalty is the dominant term in the optimization problem so that the deviation penalty is only used to choose among the feasible solutions with minimum separation penalty. In practice, it suffices to select M satisfying

$$M \geq n \cdot (\max_{ik} a_i^k - \min_{ik} a_i^k) \cdot \frac{\max_{ik} v_i^k}{\min_{ijk} u_{ij}^k}.$$

We denote an optimal solution to (||,M·Sep-Dev) as $\boldsymbol{x}^{MSD}$.

Observation 3.2.12. *The optimal solution to (||,M·Sep-Dev) is the rating that minimizes the deviation penalty among all the ratings in the set of all optimal solutions to (||,Sep).*

Theorem 3.2.13. *An optimal solution to (||,M·Sep-Dev) has the following properties:*

1. *If a subset S of reviewers has identical ratings and S is a weighted majority, then $\boldsymbol{x}^{MSD}$ is identical to the ratings of the majority.*

2. *If a subset S of reviewers has ratings equivalent under translation, and S is a weighted majority, then $\boldsymbol{x}^{MSD}$ is equivalent under translation to the ratings of the majority.*

Proof. It is easy to see that, letting the weights of the separation terms to be $M \cdot u_{ij}^k$, property *(1)* follows from Theorem 3.2.8. Property *(2)* follows from Observation 3.2.12, Theorem 3.2.4 and the fact that if two ratings are identical, then they are also equivalent under translation. □

3.3 Equivalence between the uniform quadratic separation-deviation problem and the weighted averaging method

The mathematical formulation of the uniform quadratic separation-deviation problem, (()^2,Sep-Dev), is given in (3.6).

$$(()^2\text{,Sep-Dev}) \quad \min_{\boldsymbol{x},\boldsymbol{z}} \quad \lambda \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=i+1}^{n} w^k (z_{ij} - p_{ij}^k)^2 + \sum_{k=1}^{K} \sum_{i=1}^{n} w^k (x_i - a_i^k)^2 \tag{3.6a}$$

$$\text{s.t.} \quad z_{ij} = x_i - x_j \quad (i = 1, \ldots, n;\ j = i+1, \ldots, n). \tag{3.6b}$$

Here λ is a parameter that allows to vary the relative importance of the separation penalty to the deviation penalty. Note that in (()^2,Sep-Dev) the weights w^k depend only on the reviewer and not on the object or object-pair as in (Sep-Dev). We denote an optimal solution to (()^2,Sep-Dev) as $\boldsymbol{x}^{(SD)^2}$.

Let the *weighted averaging method* be the rating aggregation method where the aggregate score of each object is the weighted average of the scores of all reviewers for this object. The following theorem establishes that for the complete-rating aggregation problem (()^2,Sep-Dev) is equivalent to the weighted averaging method.

Theorem 3.3.1. *The optimal solution to (()^2,Sep-Dev) is the same as the aggregate rating solution to the weighted average method, that is, $x_i^* = \frac{\sum_{k=1}^K w^k r_i^k}{W}$ for $i = 1, 2, \ldots, n$, and $z_{ij}^* = x_i^* - x_j^*$, where $W = \sum_{k=1}^K w^k$.*

Proof. Omitting constraint (3.6b) decomposes the problem to separate optimization problems for each z_{ij} and each x_i. Each of this problems is of the form

$$\min_y \sum_{k=1}^{K} \alpha^k (y - a^k)^2.$$

It is easy to see that this unconstrained optimization problem achieves its minimum at $y_i^* = \frac{\sum_k \alpha^k r^k}{\sum_k \alpha^k}$.

Therefore, the optimal solution to the optimization problem obtained by omitting constraint (3.6b) is:

$$x_i^* = \frac{\sum_k w^k r_i^k}{W} \qquad \text{for } i = 1, \ldots, n \tag{3.7}$$

$$z_{ij}^* = \frac{\sum_k \lambda w^k (a_i^k - a_j^k)}{\lambda W} = \frac{\sum_k w^k (a_i^k - a_j^k)}{W} \qquad \text{for } i = 1, \ldots, n;\ j = i+1, \ldots, n. \tag{3.8}$$

Since $z_{ij}^* = \frac{\sum_k w^k (a_i^k - a_j^k)}{W} = \frac{\sum_k w^k r_i^k}{W} - \frac{\sum_k w^k r_j^k}{W} = x_i^* - x_j^*$, constraint (3.6b) is satisfied by x_i^* and z_{ij}^* given in (3.7) and (3.8). □

The analogous separation model with uniform quadratic is formulated as follows:

$$((\,)^2\text{,Sep}) \quad \min_{\boldsymbol{x},\boldsymbol{z}} \quad \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=i+1}^{n} w^k (z_{ij} - p_{ij}^k)^2 \tag{3.9a}$$

$$\text{s.t.} \quad z_{ij} = x_i - x_j \quad (i = 1, \ldots, n;\ j = i+1, \ldots, n). \tag{3.9b}$$

Corollary 3.3.2. *The optimal solution to $((\,)^2$,Sep) is an aggregate rating identical to the aggregate rating obtained by the weighted averaging method, $x_i^* = \frac{\sum_{k=1}^K w^k r_i^k}{W}$ for $i = 1, 2, \ldots, n$, and $z_{ij}^* = x_i^* - x_j^*$, where $W = \sum_{k=1}^K w^k$.*

We conclude that, for the complete-rating aggregation problem, the optimal solution to $((\,)^2$,Sep-Dev) and (Sep) is the weighted average of the scores of each object. Therefore, the separation-deviation model offers no advantage compared to the weighted averaging method in this case. However this equivalence does not carry to the partial-list setting. Indeed we show in [43] that, in the partial-list setting, $((\,)^2$,Sep-Dev) and $((\,)^2$,Sep) give a better aggregate rating than the weighted averaging method.

3.4 Experimental Study

We set up an experimental study using the separation-deviation model for the purpose of aggregating country credit-risk ratings. We use as data the credit-risk ratings given by Standard and Poor's (S&P), Moody's (Mdy) and The Institutional Investor (InsI) in 1998. The input data[2] used is given in Table 3.6.

[2] We are grateful to Hammer et. al. [34] for permission to use this data.

Table 3.6: Country credit-risk ratings by country and rating agency.

Country	S&P	Mdy	InsI	Country	S&P	Mdy	InsI
Argentina	BB	Ba3	42.70	Latvia	BBB	Baa2	38.00
Australia	AA+	Aa2	74.30	Lebanon	BB-	B1	31.90
Austria	AAA	Aaa	88.70	Lithuania	BBB-	Ba1	36.10
Belgium	AA+	Aa1	83.50	Malaysia	BBB	Baa3	51.00
Bolivia	BB-	B1	28.00	Malta	A	A3	61.70
Brazil	B+	B2	37.40	Mexico	BB	Ba2	46.00
Canada	AA+	Aa1	83.00	Morocco	BB	Ba2	43.20
Chile	A-	Baa1	61.80	Netherlands	AAA	Aaa	91.70
China	BBB	A3	57.20	New Zealand	AA+	Aa2	73.10
China-HK	A	A3	61.80	Norway	AAA	Aaa	86.80
Colombia	BB+	Baa3	44.50	Pakistan	B-	Caa1	20.40
Costa Rica	BB	Ba1	38.40	Panama	BB+	Ba1	39.90
Croatia	BBB-	Baa3	39.03	Paraguay	B	B2	31.30
Cyprus	A	A2	57.30	Peru	BB	Ba3	35.00
Czech Rep	A-	Baa1	59.70	Philippines	BB+	Ba1	41.30
Denmark	AA+	Aa1	84.70	Poland	BBB	Baa3	56.70
Dominican Rep	B+	Ba2	28.10	Portugal	AA	Aa2	76.10
Egypt	BBB-	Ba1	44.40	Romania	B-	B3	31.20
El Salvador	BB+	Ba2	31.20	Russia	SD	B3	20.00
Estonia	BBB+	Baa1	42.80	Singapore	AAA	Aa1	81.30
Finland	AA+	Aaa	82.20	Slovak Rep	BB+	Ba1	41.30
France	AAA	Aaa	90.80	Slovenia	A	A3	58.40
Germany	AAA	Aaa	92.50	South Africa	BB+	Baa3	45.80
Greece	A-	Baa1	56.10	Spain	AA+	Aa2	80.30
Hungary	BBB	Baa2	55.90	Sweden	AA+	Aa2	79.70
Iceland	A+	Aa3	67.00	Switzerland	AAA	Aaa	92.70
India	BB	Ba2	44.50	Thailand	BBB-	Ba1	46.90
Indonesia	CCC+	B3	27.90	Trin & Tobago	BBB-	Ba1	43.30
Ireland	AA+	Aaa	81.80	Tunisia	BBB-	Baa3	50.30
Israel	A-	A3	54.30	Turkey	B	B1	36.90
Italy	AA	Aa3	79.10	UK	AAA	Aaa	90.20
Japan	AAA	Aa1	86.50	USA	AAA	Aaa	92.20
Jordan	BB-	Ba3	37.30	Uruguay	BBB-	Baa3	46.50
Kazakhstan	B+	Ba3	27.90	Venezuela	B	B2	34.40
Korea Rep	BBB	Ba1	52.70				

One challenge with this data is that each of the three agencies has its own rating scale: S&P uses an alphabetical rating scale (shown in the "S&P scale" column in Table 3.7) ranging from low end at SD, CC through AAA; Mdy uses an alphanumerical rating scale (shown in the "Moody's scale" column in Table 3.7) ranging from low end at C, Ca through Aaa; and InsI uses a numeric scale ranging from high end at 100 through 0. Several authors (e.g. Ferri et. al. [26], [34]) converted S&P and Mdy rating scales to the numeric scales shown in the 'converted scale' column of Table 3.7. This converted scale ranges from 0 to 100, where a lower numeric value denotes a higher probability of default. Note that in this scale the differences in the values assigned are constant for any pair of consecutive rating categories.

Table 3.7: Conversion from S&P and Mdy's rating scales to a numeric scale.

S&P scale	**Moody's scale**	**Converted scale**	**S&P scale**	**Moody's scale**	**Converted scale**
AAA	Aaa	100.00	BB	Ba2	45.00
AA+	Aa1	95.00	BB-	Ba3	40.00
AA	Aa2	90.00	B+	B1	35.00
AA-	Aa3	85.00	B	B2	30.00
A+	A1	80.00	B-	B3	25.00
A	A2	75.00	CCC+	Caa1	20.00
A-	A3	70.00	CCC	Caa2	15.00
BBB+	Baa1	65.00	CCC-	Caa3	10.00
BBB	Baa2	60.00	CC	Ca	5.00
BBB-	Baa3	55.00	SD/D	C	0.00
BB+	Ba1	50.00			

Let (||,M·Sep-Dev,**1**), refer to (||,M·Sep-Dev) with $u_{ij}^k = v_i^k = 1$, for all i, j, k. We use (||,M·Sep-Dev,**1**) as no a-priori estimates are available on the relative expertise of each agency.

We obtain the aggregate country-credit-risk rating by solving (||,M·Sep-Dev,**1**). We refer to the solution of (||,M·Sep-Dev,**1**) as the aggregate MSD1 rating, $\boldsymbol{x}^{MSD\mathbf{1}}$, and to the aggregate rating obtained by the averaging method as the aggregate averaging rating, $\boldsymbol{x}^{Avg}$. The (converted) S&P's rating is denoted by $\boldsymbol{r}^{SP}$, the (converted) Mdy's rating by $\boldsymbol{r}^{Mdy}$ and the InsI's rating by $\boldsymbol{r}^{InsI}$. These ratings are given in Table 3.8.

Table 3.8: Input and output for the country credit-risk aggregation problem.

Country	$\boldsymbol{r}^{SP}$	$\boldsymbol{r}^{Mdy}$	$\boldsymbol{r}^{InsI}$	$\boldsymbol{x}^{MSD}$	$\boldsymbol{x}^{Avg}$
Argentina	45	40	42.7	44.1	42.6
Australia	95	90	74.3	90.0	86.4
Austria	100	100	88.7	99.1	96.2
Belgium	95	95	83.5	94.1	91.2
Bolivia	40	35	28.0	36.5	34.3
Brazil	35	30	37.4	34.1	34.1
Canada	95	95	83.0	94.1	91.0
Chile	70	65	61.8	69.1	65.6
China	60	70	57.2	65.7	62.4

Table 3.8: (continued)

Country	r^{SP}	r^{Mdy}	r^{InsI}	x^{MSD}	x^{Avg}
China-HK	75	70	61.8	70.3	68.9
Colombia	50	55	44.5	53.0	49.8
Costa Rica	45	50	38.4	46.9	44.5
Croatia	55	55	39.0	54.1	49.7
Cyprus	75	75	57.3	74.1	69.1
Czech Rep	70	65	59.7	68.2	64.9
Denmark	95	95	84.7	94.1	91.6
Dominican Rep	35	45	28.1	36.6	36.0
Egypt	55	50	44.4	52.9	49.8
El Salvador	50	45	31.2	45.0	42.1
Estonia	65	65	42.8	64.1	57.6
Finland	95	100	82.2	94.8	92.4
France	100	100	90.8	99.3	96.9
Germany	100	100	92.5	100.0	97.5
Greece	70	65	56.1	65.0	63.7
Hungary	60	60	55.9	60.0	58.6
Iceland	80	85	67.0	79.8	77.3
India	45	45	44.5	45.0	44.8
Indonesia	20	25	27.9	25.0	24.3
Ireland	95	100	81.8	94.8	92.3
Israel	70	70	54.3	69.1	64.8
Italy	90	85	79.1	87.6	84.7
Japan	100	95	86.5	95.0	93.8
Jordan	40	40	37.3	40.0	39.1
Kazakhstan	35	40	27.9	36.4	34.3
Korea Rep	60	50	52.7	58.3	54.2
Latvia	60	60	38.0	59.1	52.7
Lebanon	40	35	31.9	39.1	35.6
Lithuania	55	50	36.1	50.0	47.0
Malaysia	60	55	51.0	58.8	55.3
Malta	75	70	61.7	70.2	68.9
Mexico	45	45	46.0	45.0	45.3
Morocco	45	45	43.2	45.0	44.4
Netherlands	100	100	91.7	100.0	97.2
New Zealand	95	90	73.1	90.0	86.0
Norway	100	100	86.8	99.1	95.6
Pakistan	25	20	20.4	24.1	21.8
Panama	50	50	39.9	49.1	46.6
Paraguay	30	30	31.3	30.0	30.4

Table 3.8: (continued)

Country	$\boldsymbol{r}^{SP}$	$\boldsymbol{r}^{Mdy}$	$\boldsymbol{r}^{InsI}$	$\boldsymbol{x}^{MSD}$	$\boldsymbol{x}^{Avg}$
Peru	45	40	35.0	43.3	40.0
Philippines	50	50	41.3	49.8	47.1
Poland	60	55	56.7	59.1	57.2
Portugal	90	90	76.1	89.1	85.4
Romania	25	25	31.2	25.0	27.1
Russia	0	25	20.0	24.1	15.0
Singapore	100	95	81.3	95.0	92.1
Slovak Rep	50	50	41.3	49.8	47.1
Slovenia	75	70	58.4	70.0	67.8
South Africa	50	55	45.8	54.1	50.3
Spain	95	90	80.3	90.0	88.4
Sweden	95	90	79.7	90.0	88.2
Switzerland	100	100	92.7	100.0	97.6
Thailand	55	50	46.9	54.1	50.6
Trin & Tobago	55	50	43.3	51.8	49.4
Tunisia	55	55	50.3	55.0	53.4
Turkey	30	35	36.9	35.0	34.0
UK	100	100	90.2	99.1	96.7
USA	100	100	92.2	100.0	97.4
Uruguay	55	55	46.5	55.0	52.2
Venezuela	30	30	34.4	30.0	31.5

3.4.1 Analysis of results

In this section we analyze the optimal solution to (||,M·Sep-Dev,$\mathbf{1}$), $\boldsymbol{x}^{MSD\mathbf{1}}$. The analysis involves comparing the degree of agreement between $\boldsymbol{x}^{MSD\mathbf{1}}$ and each agency's ratings.

For given penalty functions $f_{ij}()$ we define the *vector-separation distance* between two ratings $\boldsymbol{a}$ and $\boldsymbol{b}$ to be $\sum_{i=1}^{n}\ \sum_{j=i+1}^{n}\ f_{ij}(p_{ij}^{a} - p_{ij}^{b})$ and the *scalar-separation distance* between two score-pairs $\{a_i, a_j\}$ and $\{b_i, b_j\}$ to be $f_{ij}(p_{ij}^{a} - p_{ij}^{b})$, where $p_{ij}^{a} = a_i - a_j$ and $p_{ij}^{b} = b_i - b_j$. Similarly, for given penalty functions $g_i()$ we define the *vector-deviation distance* between two ratings $\boldsymbol{a}$ and $\boldsymbol{b}$ to be $\sum_{i=1}^{n}\ g_i(a_i - b_i)$ and the *scalar-deviation distance* between two scores a_i and b_i to be $g_i(a_i - b_i)$. When $f_{ij}(y) = |y|$ ($f_{ij}(y) = y^2$) will refer to the *absolute value (quadratic) vector-separation distance.* Finally, when $g_i(y) = |y|$ ($g_i(y) = y^2$) will refer to the *absolute value (quadratic) scalar-separation distance.*

The aggregate MSD1 rating $\boldsymbol{x}^{MSD\mathbf{1}}$ is shown in the 5^{th} column of Table 3.8. The absolute value vector-separation and vector-deviation distances between each agency's rating and $\boldsymbol{x}^{MSD\mathbf{1}}$ are shown in Table 3.9.

Table 3.9: Distances between $\boldsymbol{x}^{MSD\mathbf{1}}$ and each agency's rating.

	$\boldsymbol{r}^{SP}$ - $\boldsymbol{x}^{MSD\mathbf{1}}$	$\boldsymbol{r}^{Mdy}$ - $\boldsymbol{x}^{MSD\mathbf{1}}$	$\boldsymbol{r}^{InsI}$ - $\boldsymbol{x}^{MSD\mathbf{1}}$	**Total**
Absolute value vector-separation	7,540.00	6,058.60	13,952.00	27,550.60
Absolute value vector-deviation	148.60	108.20	600.40	857.20
Total distance	7,688.6	6,166.8	14,552.4	2,8407.8

The information in Table 3.9 demonstrates that InsI's country credit-risk ratings is the rating which deviate the most from $\boldsymbol{x}^{MSD\mathbf{1}}$. To explain why, we provide in Table 3.10 the absolute value vector-separation and vector-deviation distances for each pair of the three agencies. These distances show that S&P and Mdy's ratings are, by far, the closest among the three pairs. We note that also Hammer et. al., [34], found that the correlation between the ratings of S&P and Mdy is higher than both the correlation between the ratings of InsI and S&P and the correlation between the ratings of S&P and InsI. Therefore, S&P and Mdy form an 'almost' majority, and thus in the spirit of Theorem 3.2.8 should be closer to the aggregate MSD1 rating. This explains why $\boldsymbol{x}^{MSD\mathbf{1}}$ is significantly closer to the ratings of S&P and Mdy.

Table 3.10: Distances between the ratings of each pair of agencies.

	S&P - Mdy	**S&P - InsI**	**Mdy - InsI**
Absolute value vector-separation	11,440.00	19,496.80	17,994.60
Absolute value vector-deviation	230.00	703.60	618.60

Analysis of the results via the absolute value scalar-deviation distance

The absolute value scalar-deviation distance between each agency's country score and the respective country score of $\boldsymbol{x}^{MSD\mathbf{1}}$ is given in Table 3.11.

As seen in Table 3.11, for each agency there is a set of countries where the absolute value scalar-deviation distance with respect to $\boldsymbol{x}^{MSD\mathbf{1}}$ is considerably higher than the rest of the absolute value scalar-deviation distances. In particular we note that:

1. S&P's score to Russia has an absolute value scalar-deviation distance with more than 6.6 σ's from the mean, while all other S&P's scores are within 1 σ from the mean.

2. Mdy's scores to the Dominican and Korean Republics have absolute value scalar-deviation distances with more than 3.3 σ's from the mean, while all other Mdy's scores are within 1.8 σ's from the mean.

3. InsI's score to Estonia and Latvia have absolute value scalar-deviation distances with more than 2.7 σ's from the mean, while all other S&P's scores are within 1.9 σ's from the mean.

We argue that these scores are outliers with respect to the country credit-risk ratings given by this group of agencies.

In particular, Russia's score by S&P has a scalar-deviation distance to the respective aggregate score dramatically larger than all other scalar-deviation distances. We note that Russia appears to be an outlier in S&P's scores as the 1998 scores of S&P, Mdy and InsI are SD (0), B3 (25) and 20.0, respectively. One possible explanation for this discrepancy in Russia's scores is that S&P distinguishes between 'default' and 'selective default', whereas the other agencies don't do so. It should be pointed out that S&P upgraded Russia's score from SD in 1998 and 1999 to B-, B and B+ in December 2000, June 2001 and December 2001 respectively [3].

Table 3.11: Countries sorted in descending absolute-value scalar-deviation distance per agency. For each agency and each country column 'Dev' gives the distance between the respective agency's rating and the aggregate MSD1 rating. Column 'Total' shows the sum of the 3 absolute value scalar-deviations.

S&P		**Moody's**		**Institutional Investor**		**Total**	
Country	**Dev**	**Country**	**Dev**	**Country**	**Dev**	**Country**	**Dev**
Russia	24.1	Dominican Rep	8.4	Estonia	21.3	Russia	29.1
China	5.7	Korea Rep	8.3	Latvia	21.1	Estonia	23.1
Australia	5.0	Finland	5.2	New Zealand	16.9	Latvia	22.9
El Salvador	5.0	Iceland	5.2	Cyprus	16.8	New Zealand	21.9
Greece	5.0	Ireland	5.2	Australia	15.7	Australia	20.7
Indonesia	5.0	China	4.3	Croatia	15.1	Lithuania	18.9
Japan	5.0	Argentina	4.1	Israel	14.8	El Salvador	18.8
Lithuania	5.0	Brazil	4.1	Lithuania	13.9	Singapore	18.7
New Zealand	5.0	Chile	4.1	El Salvador	13.8	Cyprus	18.6
Singapore	5.0	Lebanon	4.1	Singapore	13.7	China	18.5
Slovenia	5.0	Pakistan	4.1	Ireland	13.0	Dominican Rep	18.5
Spain	5.0	Poland	4.1	Portugal	13.0	Ireland	18.4
Sweden	5.0	Thailand	4.1	Iceland	12.8	Iceland	18.2
Turkey	5.0	Malaysia	3.8	Finland	12.6	Finland	18.0
Malta	4.8	Kazakhstan	3.6	Norway	12.3	Croatia	16.9
China-HK	4.7	Peru	3.3	Slovenia	11.6	Israel	16.6
South Africa	4.1	Czech Rep	3.2	Canada	11.1	Slovenia	16.6
Bolivia	3.5	Costa Rica	3.1	Belgium	10.6	Korea Rep	15.6
Trin & Tobago	3.2	Egypt	2.9	Austria	10.4	Sweden	15.3
Colombia	3.0	Italy	2.6	Sweden	10.3	Portugal	14.8
Italy	2.4	Colombia	2.0	Spain	9.7	Spain	14.7
Egypt	2.1	Trin & Tobago	1.8	Denmark	9.4	Norway	14.1
Costa Rica	1.9	Bolivia	1.5	Panama	9.2	Greece	13.9
Czech Rep	1.8	Austria	0.9	Greece	8.9	Bolivia	13.5
Korea Rep	1.7	Belgium	0.9	UK	8.9	China-HK	13.5
Peru	1.7	Canada	0.9	Bolivia	8.5	Colombia	13.5
Dominican Rep	1.6	Croatia	0.9	China	8.5	Costa Rica	13.5
Kazakhstan	1.4	Cyprus	0.9	China-HK	8.5	Czech Rep	13.5
Malaysia	1.2	Denmark	0.9	Colombia	8.5	Egypt	13.5
Argentina	0.9	Estonia	0.9	Costa Rica	8.5	Italy	13.5
Austria	0.9	Israel	0.9	Czech Rep	8.5	Japan	13.5

Table 3.11: (continued)

S&P		Moody's		Institutional Investor		Total	
Country	**Dev**	**Country**	**Dev**	**Country**	**Dev**	**Country**	**Dev**
Belgium	0.9	Latvia	0.9	Dominican Rep	8.5	Kazakhstan	13.5
Brazil	0.9	Norway	0.9	Egypt	8.5	Malta	13.5
Canada	0.9	Panama	0.9	France	8.5	Trin & Tobago	13.5
Chile	0.9	Portugal	0.9	Italy	8.5	Peru	13.3
Croatia	0.9	Russia	0.9	Japan	8.5	South Africa	13.3
Cyprus	0.9	South Africa	0.9	Kazakhstan	8.5	Canada	12.9
Denmark	0.9	UK	0.9	Malta	8.5	Malaysia	12.8
Estonia	0.9	France	0.7	Philippines	8.5	Belgium	12.4
Israel	0.9	China-HK	0.3	Slovak Rep	8.5	Chile	12.3
Latvia	0.9	Malta	0.2	Trin & Tobago	8.5	Austria	12.2
Lebanon	0.9	Philippines	0.2	Uruguay	8.5	Lebanon	12.2
Norway	0.9	Slovak Rep	0.2	Netherlands	8.3	Thailand	12.2
Pakistan	0.9	Australia	0.0	Peru	8.3	Denmark	11.2
Panama	0.9	El Salvador	0.0	South Africa	8.3	Panama	11.0
Poland	0.9	Germany	0.0	Malaysia	7.8	UK	10.7
Portugal	0.9	Greece	0.0	USA	7.8	France	9.9
Thailand	0.9	Hungary	0.0	Germany	7.5	Philippines	8.9
UK	0.9	India	0.0	Chile	7.3	Slovak Rep	8.9
France	0.7	Indonesia	0.0	Switzerland	7.3	Pakistan	8.7
Finland	0.2	Japan	0.0	Lebanon	7.2	Uruguay	8.5
Iceland	0.2	Jordan	0.0	Thailand	7.2	Brazil	8.3
Ireland	0.2	Lithuania	0.0	Romania	6.2	Netherlands	8.3
Philippines	0.2	Mexico	0.0	Korea Rep	5.6	Indonesia	7.9
Slovak Rep	0.2	Morocco	0.0	Tunisia	4.7	USA	7.8
Germany	0.0	Netherlands	0.0	Venezuela	4.4	Germany	7.5
Hungary	0.0	New Zealand	0.0	Hungary	4.1	Poland	7.4
India	0.0	Paraguay	0.0	Russia	4.1	Switzerland	7.3
Jordan	0.0	Romania	0.0	Pakistan	3.7	Turkey	6.9
Mexico	0.0	Singapore	0.0	Brazil	3.3	Argentina	6.4
Morocco	0.0	Slovenia	0.0	Indonesia	2.9	Romania	6.2
Netherlands	0.0	Spain	0.0	Jordan	2.7	Tunisia	4.7
Paraguay	0.0	Sweden	0.0	Poland	2.4	Venezuela	4.4
Romania	0.0	Switzerland	0.0	Turkey	1.9	Hungary	4.1
Switzerland	0.0	Tunisia	0.0	Morocco	1.8	Jordan	2.7
Tunisia	0.0	Turkey	0.0	Argentina	1.4	Morocco	1.8
USA	0.0	USA	0.0	Paraguay	1.3	Paraguay	1.3
Uruguay	0.0	Uruguay	0.0	Mexico	1.0	Mexico	1.0
Venezuela	0.0	Venezuela	0.0	India	0.5	India	0.5
Total	148.6		108.2		600.4		857.2
Maximum	24.1		8.4		21.3		29.1
Average	2.2		1.6		8.7		12.4
σ	3.3		2.0		4.4		5.7

Analysis of the results via the absolute value scalar-separation distance

This section provides an analysis of the solution to (||,M·Sep-Dev,**1**) in terms of the absolute value scalar-separation distance analogous to the analysis we provided in the previous section. The overall results are consistent with those of the previous section.

With a rating of 69 countries we have 2346 pairwise comparisons and it is impossible to list all of the absolute value scalar-separation distances. Instead in Table 3.12 we list only the most significant pairwise comparisons in terms of largest absolute value scalar-separation distances between each agency's rating and $\boldsymbol{x}^{MSD\mathbf{1}}$.

It is interesting to observe that the countries which have the highest scalar-deviation distances belong to the country-pairs which have the highest absolute value scalar-separation distances. Indeed reviewing the ranked list of the pairwise comparisons which deviate the most from $\boldsymbol{x}^{MSD\mathbf{1}}$, one observes that certain countries appear in country-pairs with high scalar-separation distances. These two observations are related to having derived the separation gaps from the ratings $p_{ij}^k = a_i^k - a_j^k$. Thus any discrepancy in one score affects all pairwise comparisons with such score. Indeed it is easy to see in Table 3.12 that for the case of S&P the first 68 pairwise comparisons concern Russia. In the case of Mdy the countries which dominate the results are the Dominican Republic and the Korean Republic. Finally, in the case of InsI, this clustering of countries is not as evident; however one can still observe the predomination of Estonia and Latvia as the countries with the higher absolute value scalar-separation distances. Recall that these countries were the ones with the highest scalar-deviation distances.

Table 3.12: Country-pairs with the highest absolute value scalar-separation distance per agency sorted in descending order. For each agency and each country-pair the column 'Sep' gives the scalar-separation distance between the respective agency's separation gap and the aggregate MSD1 rating's.

S&P			**Mdy**			**InsI**		
Country 1	**Country 2**	**Sep**	**Country 1**	**Country 2**	**Sep**	**Country 1**	**Country 2**	**Sep**
Australia	Russia	29.1	Dom. Rep.	Korea Rep	16.7	Estonia	Romania	27.5
El Salvador	Russia	29.1	Finland	Korea Rep	13.5	Latvia	Romania	27.3
Greece	Russia	29.1	Iceland	Korea Rep	13.5	Estonia	Venezuela	25.7
Japan	Russia	29.1	Ireland	Korea Rep	13.5	Latvia	Venezuela	25.5
Lithuania	Russia	29.1	China	Korea Rep	12.6	Brazil	Estonia	24.6
N. Zealand	Russia	29.1	Argentina	Dom. Rep.	12.5	Brazil	Latvia	24.4
Russia	Singapore	29.1	Brazil	Dom. Rep.	12.5	Estonia	Indonesia	24.2
Russia	Slovenia	29.1	Chile	Dom. Rep.	12.5	Indonesia	Latvia	24.0
Russia	Spain	29.1	Dom. Rep.	Lebanon	12.5	Estonia	Turkey	23.2
Russia	Sweden	29.1	Dom. Rep.	Pakistan	12.5	N. Zealand	Romania	23.1
Malta	Russia	28.9	Dom. Rep.	Poland	12.5	Cyprus	Romania	23.0
China-HK	Russia	28.8	Dom. Rep.	Thailand	12.5	Latvia	Turkey	23.0
Bolivia	Russia	27.6	Dom. Rep.	Malaysia	12.2	Estonia	Paraguay	22.6
Russia	Trin&Tob	27.3	Kazakhstan	Korea Rep	11.9	Latvia	Paraguay	22.4
Italy	Russia	26.5	Dom. Rep.	Peru	11.7	Estonia	Mexico	22.3
Egypt	Russia	26.2	Czech Rep	Dom. Rep.	11.6	Latvia	Mexico	22.1
Czech Rep	Russia	25.9	Costa Rica	Korea Rep	11.4	Australia	Romania	21.9
Korea Rep	Russia	25.8	Dom. Rep.	Egypt	11.3	N. Zealand	Venezuela	21.3
Peru	Russia	25.8	Dom. Rep.	Italy	11.0	Croatia	Romania	21.3
Malaysia	Russia	25.3	Colombia	Korea Rep	10.3	Cyprus	Venezuela	21.2

Table 3.12: (continued)

S&P			Mdy			InsI		
Country 1	**Country 2**	**Sep**	**Country 1**	**Country 2**	**Sep**	**Country 1**	**Country 2**	**Sep**
Argentina	Russia	25.0	Dom. Rep.	Trin&Tob	10.2	Israel	Romania	21.0
Austria	Russia	25.0	Bolivia	Dom. Rep.	9.9	Estonia	India	20.8
Belgium	Russia	25.0	Argentina	Finland	9.3	India	Latvia	20.6
Brazil	Russia	25.0	Argentina	Iceland	9.3	Brazil	N. Zealand	20.2
Canada	Russia	25.0	Argentina	Ireland	9.3	Australia	Venezuela	20.1
Chile	Russia	25.0	Brazil	Finland	9.3	Brazil	Cyprus	20.1
Croatia	Russia	25.0	Brazil	Iceland	9.3	Lithuania	Romania	20.1
Cyprus	Russia	25.0	Brazil	Ireland	9.3	El Salvador	Romania	20.0
Denmark	Russia	25.0	Chile	Finland	9.3	Argentina	Estonia	19.9
Estonia	Russia	25.0	Chile	Iceland	9.3	Romania	Singapore	19.9
Israel	Russia	25.0	Chile	Ireland	9.3	Indonesia	N. Zealand	19.8
Latvia	Russia	25.0	Finland	Lebanon	9.3	Argentina	Latvia	19.7
Lebanon	Russia	25.0	Finland	Pakistan	9.3	Cyprus	Indonesia	19.7
Norway	Russia	25.0	Finland	Poland	9.3	Estonia	Morocco	19.5
Pakistan	Russia	25.0	Finland	Thailand	9.3	Croatia	Venezuela	19.5
Panama	Russia	25.0	Iceland	Lebanon	9.3	Latvia	Morocco	19.3
Poland	Russia	25.0	Iceland	Pakistan	9.3	Ireland	Romania	19.2
Portugal	Russia	25.0	Iceland	Poland	9.3	Portugal	Romania	19.2
Russia	Thailand	25.0	Iceland	Thailand	9.3	Israel	Venezuela	19.2
Russia	UK	25.0	Ireland	Lebanon	9.3	Australia	Brazil	19.0
France	Russia	24.8	Ireland	Pakistan	9.3	Iceland	Romania	19.0
Finland	Russia	24.3	Ireland	Poland	9.3	Estonia	Poland	18.9
Iceland	Russia	24.3	Ireland	Thailand	9.3	N. Zealand	Turkey	18.8
Ireland	Russia	24.3	Austria	Korea Rep	9.2	Finland	Romania	18.8
Philippines	Russia	24.3	Belgium	Korea Rep	9.2	Latvia	Poland	18.7
Russia	Slovak Rep	24.3	Canada	Korea Rep	9.2	Cyprus	Turkey	18.7
Germany	Russia	24.1	Croatia	Korea Rep	9.2	Australia	Indonesia	18.6
Hungary	Russia	24.1	Cyprus	Korea Rep	9.2	Estonia	Jordan	18.6
India	Russia	24.1	Denmark	Korea Rep	9.2	Norway	Romania	18.5
Jordan	Russia	24.1	Estonia	Korea Rep	9.2	Brazil	Croatia	18.4
Mexico	Russia	24.1	Israel	Korea Rep	9.2	Jordan	Latvia	18.4
Morocco	Russia	24.1	Korea Rep	Latvia	9.2	Lithuania	Venezuela	18.3
Netherlands	Russia	24.1	Korea Rep	Norway	9.2	N. Zealand	Paraguay	18.2
Paraguay	Russia	24.1	Korea Rep	Panama	9.2	El Salvador	Venezuela	18.2
Romania	Russia	24.1	Korea Rep	Portugal	9.2	Brazil	Israel	18.1
Russia	Switzerland	24.1	Korea Rep	Russia	9.2	Singapore	Venezuela	18.1
Russia	Tunisia	24.1	Korea Rep	S. Africa	9.2	Cyprus	Paraguay	18.1
Russia	USA	24.1	Korea Rep	UK	9.2	Croatia	Indonesia	18.0
Russia	Uruguay	24.1	Finland	Malaysia	9.0	Mexico	N. Zealand	17.9
Russia	Venezuela	24.1	France	Korea Rep	9.0	Romania	Slovenia	17.8
Kazakhstan	Russia	22.7	Iceland	Malaysia	9.0	Cyprus	Mexico	17.8
Dom. Rep.	Russia	22.5	Ireland	Malaysia	9.0	Indonesia	Israel	17.7
Costa Rica	Russia	22.2	China-HK	Dom. Rep.	8.7	Australia	Turkey	17.6
Colombia	Russia	21.1	Dom. Rep.	Malta	8.6	Estonia	Pakistan	17.6
Russia	S. Africa	20.0	Finland	Peru	8.5	Ireland	Venezuela	17.4
Indonesia	Russia	19.1	Iceland	Peru	8.5	Latvia	Pakistan	17.4
Russia	Turkey	19.1	Ireland	Peru	8.5	Portugal	Venezuela	17.4
China	Russia	18.4	Korea Rep	Philippines	8.5	Canada	Romania	17.3
Australia	China	10.7	Korea Rep	Slovak Rep	8.5	Brazil	Lithuania	17.2
China	El Salvador	10.7	Czech Rep	Finland	8.4	Estonia	Russia	17.2
China	Greece	10.7	Czech Rep	Iceland	8.4	Estonia	Hungary	17.2
China	Japan	10.7	Czech Rep	Ireland	8.4	Iceland	Venezuela	17.2
China	Lithuania	10.7	Australia	Dom. Rep.	8.4	Brazil	El Salvador	17.1
China	N. Zealand	10.7	Dom. Rep.	El Salvador	8.4	Brazil	Singapore	17.0

Table 3.12: (continued)

S&P			Mdy			InsI		
Country 1	**Country 2**	**Sep**	**Country 1**	**Country 2**	**Sep**	**Country 1**	**Country 2**	**Sep**
Total		7,540.0			6,058.2			13,952.0
Maximum		29.1			16.7			27.5
Average		3.2			2.6			5.9
σ		4.4			2.6			5.0

3.4.2 Comparison of the aggregate MSD1 rating to the aggregate averaging rating

We now show that the aggregate MSD1 rating $\boldsymbol{x}^{MSD\mathbf{1}}$ is in some sense closer than the aggregate averaging rating $\boldsymbol{x}^{Avg}$ to the group consensus. The aggregate MSD1 rating and the aggregate averaging rating are shown in the 5^{th} and 6^{th} columns of Table 3.8, respectively. We compare $\boldsymbol{x}^{MSD\mathbf{1}}$ to $\boldsymbol{x}^{Avg}$ by evaluating their respective distances to each of the agencies' ratings. For this purpose we use the vector-separation and vector-deviation distances and the Kemeny-Snell distance.

Since $\boldsymbol{x}^{MSD\mathbf{1}}$ is the optimal solution to (||,M·Sep-Dev,**1**), it is the vector with minimum total sum of absolute value vector-separation and vector-deviation distances with respect to the agencies' ratings. Thus $\boldsymbol{x}^{MSD\mathbf{1}}$ tends to perform better than $\boldsymbol{x}^{Avg}$ for the absolute value vector-deviation distance alone, and the absolute value vector-separation distance alone. This is shown in the Tables 3.13 and 3.14.

Table 3.13: Absolute value vector-deviation distances between each aggregate rating and each agency rating.

	$\boldsymbol{r}^{SP}$	$\boldsymbol{r}^{Mdy}$	$\boldsymbol{r}^{InsI}$	**Total**
$\boldsymbol{x}^{MSD\mathbf{1}}$	148.6	108.2	600.4	857.2
$\boldsymbol{x}^{Avg}$	280.8	210.8	432.2	923.8

Table 3.14: Absolute value vector-separation distances between each aggregate rating and each agency rating.

	$\boldsymbol{r}^{SP}$	$\boldsymbol{r}^{Mdy}$	$\boldsymbol{r}^{InsI}$	**Total**
$\boldsymbol{x}^{MSD\mathbf{1}}$	7,540.0	6,058.6	13,952.0	27,550.6
$\boldsymbol{x}^{Avg}$	8,990.4	8,099.0	11,839.0	28,928.4

From Theorem 3.3.1 we have that $\boldsymbol{x}^{Avg}$ is the optimal solution to the separation-deviation problem with uniform quadratic penalty functions. I.e. it is the vector with minimum total sum of quadratic vector-separation and vector-deviation distances to the agencies' ratings. Thus $\boldsymbol{x}^{Avg}$ tends to perform better than $\boldsymbol{x}^{MSD\mathbf{1}}$ for the quadratic vector-deviation distance alone, and the quadratic vector-separation distance alone. This is shown in the Tables 3.15 and 3.16.

Table 3.15: Quadratic vector-deviation distances between each aggregate rating and each agency rating.

	$\boldsymbol{r}^{SP}$	$\boldsymbol{r}^{Mdy}$	$\boldsymbol{r}^{InsI}$	**Total**
$\boldsymbol{x}^{MSD\mathbf{1}}$	1,049.76	453.69	6,577.21	8,080.66
$\boldsymbol{x}^{Avg}$	1,624.09	1,043.29	3,528.36	6,195.74

Table 3.16: Quadratic vector-separation distances between each aggregate rating and each agency rating.

	$\boldsymbol{r}^{SP}$	$\boldsymbol{r}^{Mdy}$	$\boldsymbol{r}^{InsI}$	**Total**
$\boldsymbol{x}^{MSD\mathbf{1}}$	70,596.49	31,187.56	142,129.00	243,913.05
$\boldsymbol{x}^{Avg}$	64,566.81	43,597.44	94,740.84	202,905.09

In Table 3.17 we show the number of reversals when comparing $\boldsymbol{x}^{MSD\mathbf{1}}$ and $\boldsymbol{x}^{Avg}$ with each of the agency's original ratings. As shown in the 4^{th} column of Table 3.17, $\boldsymbol{x}^{MSD\mathbf{1}}$ has fewer total number of reversals from $\boldsymbol{r}^{SP}$, $\boldsymbol{r}^{Mdy}$, and $\boldsymbol{r}^{InsI}$, as compared to $\boldsymbol{x}^{Avg}$. Therefore, the solution to (||,M·Sep-Dev,**1**) is closer to the ordering implied by the agencies' ratings than $\boldsymbol{x}^{Avg}$. Furthermore, $\boldsymbol{x}^{MSD\mathbf{1}}$ is closer to both $\boldsymbol{r}^{SP}$ and $\boldsymbol{r}^{Mdy}$ than $\boldsymbol{x}^{Avg}$. As noted previously, S&P and Mdy form a kind of 'majority'. We argue that even when there is no clear majority, as Theorem 3.2.13 requires, $\boldsymbol{x}^{MSD\mathbf{1}}$ is closer to the ratings of the reviewers that show a "high degree of agreement" than to the ratings of other reviewers. We conclude that $\boldsymbol{x}^{MSD\mathbf{1}}$ is close to a group consensus.

Table 3.17: Number of reversals distance between each aggregate rating and each agency rating.

	$\boldsymbol{r}^{SP}$	$\boldsymbol{r}^{Mdy}$	$\boldsymbol{r}^{InsI}$	**Total**
$\boldsymbol{x}^{MSD\mathbf{1}}$	86	87.5	158.5	332
$\boldsymbol{x}^{Avg}$	107	112.5	129.5	349

We note that $\boldsymbol{r}^{SP}$ has the fewest number of reversals from $\boldsymbol{x}^{MSD\mathbf{1}}$, followed closely by $\boldsymbol{r}^{Mdy}$. The $\boldsymbol{r}^{InsI}$ rating has a far larger number of reversals. This contrasts with the observations in the previous sections, where Mdy had the closest agency-rating to $\boldsymbol{x}^{MSD\mathbf{1}}$. This apparent contradiction might be explained by the following two observations:

1. As shown in Table 3.18, $\boldsymbol{r}^{SP}$ has the fewest total number of reversals when compared to the other two rating agencies. Furthermore with respect to the number of reversals measure, $\boldsymbol{r}^{SP}$ is closer to $\boldsymbol{r}^{InsI}$ than $\boldsymbol{r}^{Mdy}$ to $\boldsymbol{r}^{InsI}$ and $\boldsymbol{r}^{SP}$ is closer to $\boldsymbol{r}^{Mdy}$ than $\boldsymbol{r}^{InsI}$ to $\boldsymbol{r}^{Mdy}$. So, when using this distance measure, $\boldsymbol{r}^{SP}$ is closer to the group consensus than the other two ratings.

2. Since the number of reversals distance is only relative to the (implied) ordering, rather than to the magnitude of the scores, it is less sensitive to outliers than the vector-separation and vector-deviation distances. In this regard, note that Russia's score by

S&P is the score with the highest deviation and separation distances; while Russia contributes only one reversal when comparing $\boldsymbol{r}^{SP}$ to $\boldsymbol{r}^{Mdy}$ and only one reversal when comparing $\boldsymbol{r}^{Mdy}$ to $\boldsymbol{r}^{InsI}$.

Table 3.18: Number of reversals between the ratings of each pair of agencies.

	$\boldsymbol{r}^{SP}$	$\boldsymbol{r}^{Mdy}$	$\boldsymbol{r}^{InsI}$	**Total**
$\boldsymbol{r}^{SP}$	0	133.5	214.5	348.0
$\boldsymbol{r}^{Mdy}$	133.5	0	228.0	361.5
$\boldsymbol{r}^{InsI}$	214.5	228	0	442.5

3.5 Conclusions

In this chapter we demonstrate several properties of the separation-deviation model. The main result here is that the separation model has the property of resistance to manipulation by a minority. We also prove a similar, but weaker, result for the separation-deviation model. Additionally we characterize the optimal solution to the model for certain classes of penalty functions.

The separation deviation model is used here to aggregate conflicting credit-risk ratings. We show that the aggregate MSD1 rating is closer to the group rating than the aggregate averaging rating. This is established here for the absolute value vector-deviation and vector-separation distances. Moreover, the aggregate MSD1 rating also has fewer reversals from the agencies' ratings than the aggregate averaging rating. We conclude that the aggregate MSD1 rating better reflects each of the agency's ratings than the aggregate averaging rating.

We anticipate that in more general scenarios the separation-deviation model will prove to be a useful aggregation method. We believe that the separation-deviation model is a useful tool for aggregating disparate sources of information, and should be considered as an alternative to other group decision making methods.

Chapter 4

Rating customers according to their promptness to adopt new technology

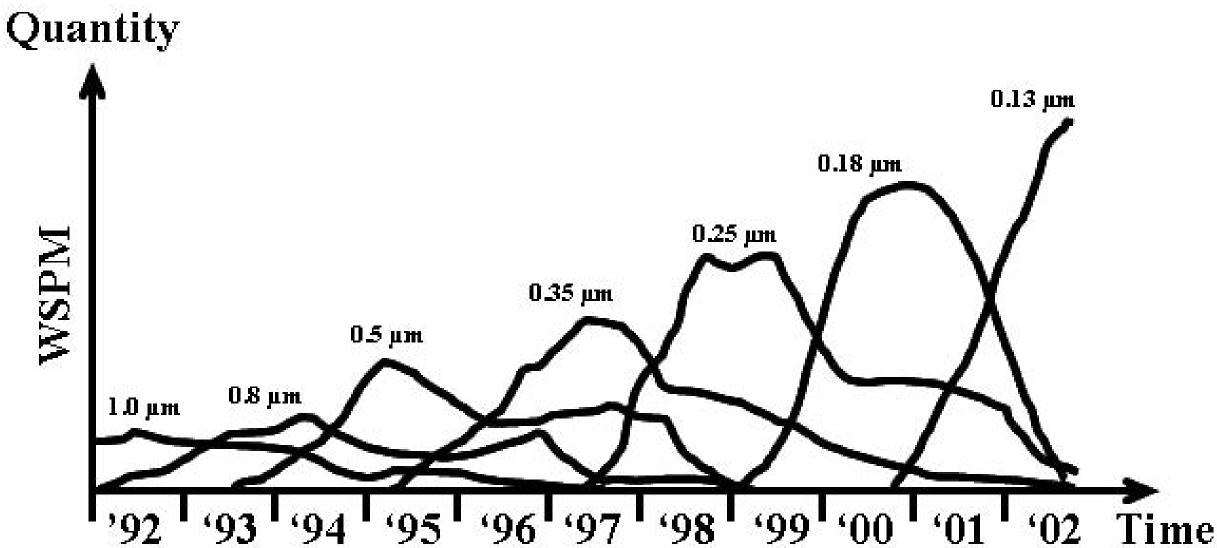
4.1 Introduction

Databases are a significant source of information in organizations and play a major role in managerial decision-making. From commercial data, organizations derive information about their customers and use it to hone their competitive strategies.

Customer rating with respect to the promptness to adopt new products is a compelling exercise, as it allows companies to define appropriate actions for the launch of a new product into the marketplace. *Innovators*, customers that adopt technology promptly, are often the main target of a firm's marketing efforts of new products. Because the innovators tend to influence the remaining potential adopters, that is, the *majority*, firms tend to allocate more marketing efforts and resources toward the innovators than toward the majority [51]. Therefore, knowing the customers' *adoption promptness* allows companies to market innovators effectively. In addition, customer rating is the first step in order to be able to perform studies that link individual characteristics (such as age, gender, usage rate, and loyalty) to the adoption promptness.

Rating the customers' adoption promptness is particularly important in high tech markets, where products generally have short—and indeed shrinking—life-cycles [58]. For example, whereas memory semiconductor chips had a life of mature product lasting approximately 5 years in the early 90s, this had shrunk to one year in the early 2000s, (see e.g. Figure 4.1 for product life cycles in the semiconductor industry).

Figure 4.1: Shrinking product life cycles in the semi-conductor industry over time. Wafers' starts per week (WSPW) are given as a function of time.



The motivation of this chapter is to solve the *customer rating problem*, defined as follows: Given a set of customers, a set of products, and the purchase times of each customer-product pair, rate each customer according to its adoption promptness. We illustrate our proposed methodology on data obtained from Sun Microsystems.

We focus our attention on the problem where the information available is incomplete; that is, there are customers who do not purchase every product. This is true in Sun Microsystems data and it is often the case in other customer segmentation data.

Two of the main contributions of this chapter are: (1) to present the novel application of the separation-deviation model to the customer rating problem, and, (2) to compare the separation-deviation model to the well known unidimensional scaling methodology, which is widely used in several contexts (see, for example, [27, 53, 31]).

This chapter is organized as follows. Section 4.2 reviews how the customer rating problem has been previously addressed in the literature, and reviews unidimensional scaling as an approach to solve the customer rating problem. Section 4.3 indicates how the separation-deviation model is used in the customer rating context and how it differs from other techniques. Section 4.4 compares, in simulated scenarios, the performance of the separation-deviation model to that of unidimensional scaling. Specifically, section 4.4 compares the performance of the two methods in simulated scenarios where the correct adoption promptness of the customers is known in advance. Section 4.5 presents a study on commercial data from Sun Microsystems and reports the generated insights obtained by using our approach. Finally Section 4.6 gives some final remarks about the separation-deviation model and its usefulness for other type of applications.

4.2 Literature review

Generally speaking, the input to data-mining techniques consists of a collection of records that characterize customer purchase behavior, as well as other relevant customer characteristics such as age, gender, usage rate, loyalty status, etc. At an abstract level, many data-mining techniques attempt to explain customer behavior in terms of a meaningful subset of customer characteristics, by identifying a function that maps a vector of customer attributes to a scalar value.

There are two main classes of data-mining techniques: those for *supervised learning* and those for *unsupervised learning*. The main objective of supervised learning techniques is to try to identify how to use independent variables (i.e., observable customer characteristics) in order to be able to predict an unobservable customer characteristic. These techniques require as input a customer database with pre-classified customers. Some classical customer segmentation techniques that fall under this category are automatic interaction detector (AID) and its extensions (e.g. CHAID), linear regression and its generalizations (e.g. canonical analysis), discriminant analysis, conjoint analysis and its extensions (e.g. componential segmentation and *POSSE* — product optimization and selected segment evaluation), logistic regression, neural networks, etc. For the Sun Microsystems study we have no a priori labeling of the customers, i.e. we do not have a "training set". Therefore, the focus of this chapter is on the unsupervised-learning problem of rating customers according to their adoption promptness.

In unsupervised-learning techniques, there is no pre-classified set of customers. Thus unsupervised learning techniques aim to determine the customer ratings from the unlabeled data. We mention an approach used in [55] based on maximum-cut clustering. Maximum-cut is an NP-hard problem, so the approach in [55] is to approximate maximum-cut with semidefinite-programming. The output is not a full customer rating but rather a classification of the customers only in early versus late adopters. The limitations of their approach are considerable and therefore we restrict our attention to multidimensional scaling, a widely used unsupervised learning technique.

Multidimensional scaling (MDS) is a set of related techniques used for representing the similarities and dissimilarities among pairs of objects as distances between points on a low-dimensional space. MDS models aim to approximate given nonnegative dissimilarities, δ_{ij}, among pairs of objects, i, j, by distances between points in an m-dimensional MDS configuration X. Here X, the *configuration*, is an $n \times m$ matrix with the coordinates of the n objects in $\Re^m$. Most MDS techniques assume that the dissimilarity matrix $[\delta_{ij}]$ is symmetric; we review two important exceptions below. The most common function to measure the fit between the given dissimilarities, δ_{ij}, and distances, $d_{ij}(X)$, is STRESS, defined by

$$STRESS(X) \equiv \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(\delta_{ij} - d_{ij}(X))^2, \tag{4.1}$$

where w_{ij} is a given nonnegative weight reflecting the importance or precision of the dissimilarity δ_{ij}. Note that w_{ij} can be set to 0 if δ_{ij} is unknown. $d_{ij}(X)$ is a vector norm, defined as

$$d_{ij}(X) = \left[\sum_{s=1}^{n} |x_{is} - x_{js}|^q \right]^{1/q}$$

with given parameter $q \geq 1$. Usually, $d_{ij}(X)$ is the L_2 norm ($q = 2$) or the L_1 norm (q=1).

Finding a global minimum of (4.1) is a hard optimization problem since STRESS is a nonlinear non-convex function with respect to X and thus optimization algorithms can converge to local minima (see, for example, [23, 33, 2]).

In a useful MDS technique, the so-called *three-way MDS*, for each pair of objects we are given K dissimilarity measures from different "replications" (e.g., repeated measures, different experimental conditions, multiple raters, etc.). The objective function of three-way MDS is defined as [23],

$$3WAY - STRESS(X) \equiv \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^k (\delta_{ij}^k - d_{ij}(X))^2. \tag{4.2}$$

Unidimensional scaling (UDS) is the important one-dimensional case of MDS where the configuration X is an $n \times 1$ matrix. Therefore UDS seeks to approximate the given dissimilarities by distances between points in a one-dimensional space. Unidimensional scaling has been studied mainly as a model for object sequencing and seriation [45, 15]; thus its relevance to the problem concerning this chapter. Unidimensional scaling is a hard optimization problem, and combinatorial techniques (e.g., branch-and-bound and dynamic programming) are only able to optimally solve instances of up to 30 objects, see for example [50, 13, 46, 16].

In our particular application, rating customers according to their adoption promptness, the input data is a matrix A with a_i^k giving the adoption time (relative to product launch) of customer i for product k. This matrix is, in general, incomplete and has many missing elements. The objective is to assign each customer i to a scale $\boldsymbol{x}$ such that x_i most accurately recovers the across-customer ordering of product adoption times within any product. In order to solve our problem, we can setup the following three-way UDS problem:

$$\min_{\boldsymbol{x}} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^k (\left|a_i^k - a_j^k\right| - |x_i - x_j|)^2. \tag{4.3}$$

Here the interpretation is that product k gives a pairwise dissimilarity, $\left|a_i^k - a_j^k\right|$, among a pair of customers i and j the purchased product k. Then, the objective is that customers with low (high) dissimilarities have similar (dissimilar) adoption promptness and should be placed "close (far) to each other" in the desired scale $\boldsymbol{x}$.

We note a couple of drawbacks of formulating our customer rating problem as the three-way UDS problem (4.3), and later introduce our scaling methodology which addresses these drawbacks.

1. As mentioned earlier, finding the optimal solution to (4.3) is a hard optimization problem and current optimization techniques are only able to optimally solve instances of at most 30 objects.

2. By calculating the dissimilarities as $\left|a_i^k - a_j^k\right|$, problem (4.3) ignores the so-called *directionality of dominance*, that is, the sign of $(a_i^k - a_j^k)$. In particular, problem (4.3) does not capture the information regarding which customer adopted earlier product k. Note that this information is very relevant in the customer rating problem.

   A closely related observation is that given an optimal solution, $\boldsymbol{x}^*$, to (4.3), $-\boldsymbol{x}^*$ is also an optimal solution to (4.3). Thus, by solving (4.3), we get a rating of the customers but we do not know whether a higher rating means a greater adoption promptness or viceversa.

   While the vast majority of the papers in the UDS literature assume that the given dissimilarities are non-negative and symmetric, there are two papers ([45] and [14]) that consider the case where the dissimilarities are given in a complete skew-symmetric matrix (i.e., $\delta_{ij} = -\delta_{ji}$).

   Since these approaches consider only one matrix $[\delta_{ij}]$ and this matrix is complete, these are not applicable to the customer rating problem. Indeed, the approach presented in this chapter is a nice generalization of one of these approaches. We discuss briefly the approaches presented in [45] and [14] and refer for further details to the original papers.

   In [45], Hubert et. al. observe that a skew-symmetric matrix contains two distinct types of information between any pair of objects: degree of dissimilarity, $|\delta_{ij}|$, and directionality of dominance, $\text{sign}(\delta_{ij})$. They consider two approaches to sequencing the objects. The first approach consists on finding the object ordering π such that the matrix $[\delta_{\pi(i)\pi(j)}]$ has the maximum sum of above-diagonal entries. Hubert et. al. note that this problem is exactly the minimum feedback arc set problem, which is NP-hard. The second approach proposed in [45] is to solve the following problem

$$\min_{\boldsymbol{x}} \sum_{i=1}^{n} \sum_{j=1}^{n} (\delta_{ij} - (x_i - x_j))^2, \tag{4.4}$$

   where the dissimilarity matrix $[\delta_{ij}]$ is assumed to be skew-symmetric and has no missing entries. Hubert et. al. give an analytic solution to problem 4.4; [42] gives a generalization of this result to the case of multiple dissimilarity matrices (but still no missing entries).

   In [14], the authors also differentiate between the degree of dissimilarity, $|\delta_{ij}|$, and directionality of dominance, $\text{sign}(\delta_{ij})$. They propose a bicriteria optimization problem that balances between these two types of information. While interesting, this approach is not practical since the proposed solution technique is only able to determine the non-dominated solutions for matrices up to size 20×20 (and can take as input only one skew-symmetric matrix).

A simplified version of the separation-deviation model presented in this chapter is to solve problem (4.5). We present here this simplified version in order to allow for a quick comparison with the methods reviewed so far.

$$\min_{\boldsymbol{x}} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^{k} f_{ij}^{k}(\delta_{ij}^{k} - (x_i - x_j)) \tag{4.5}$$

Where, for the customer rating problem, $\delta_{ij}^k \equiv a_i^k - a_j^k$, and thus, for each product k, $[\delta_{ij}^k]$ is a (possibly incomplete) skew-symmetric matrix. w_{ij}^k is a non-negative weight reflecting the importance or precision of the dissimilarity δ_{ij}^k (w_{ij}^k is set to 0 if δ_{ij}^k is unknown). And, each $f_{ij}^k(\cdot)$ is a given of convex function.

In MDS terminology, problem (4.5) is a three-way unidimensional scaling problem, where the K dissimilarity matrices are skew-symmetric. In contrast to all of the MDS literature, where $f_{ij}^k(\cdot)$ are either the quadratic function or the absolute value function, in the separation-deviation model these functions can differ form each other and may be any convex function. In contrast to problem (4.3)—the direct application of UDS to the customer rating problem—, problem (4.5) is solvable in polynomial time (as we show later) and does not ignore the directionality of the dominance. In contrast to the approaches in [45] and [14] for skew-symmetric matrices, problem (4.5) accepts multiple and incomplete dissimilarity matrices and is solvable in polynomial time.

4.3 Applying the separation-deviation model to the customer rating problem

Consider a population of customers, identified by the index i, who may elect to purchase products indexed by $k \in \{1, \ldots, K\}$ over a period comprising a number of periods (months). Let a_i^k be the first month (if any) in which customer i purchased product k. Each of the n customers is associated with an K-dimensional vector $\boldsymbol{a}_i = (a_i^1, \ldots, a_i^K)$, recording the first month in which he/she bought the different products. In the event that the customer did not purchase a product, the corresponding entry in the vector is regarded as "missing". The model appropriately (and seamlessly to the user) deals with this missing information.

One of the important features of the separation-deviation model is that the model takes as input a collection of separation gaps between the objects (customers) to be classified. That is, a single customer-pair can have several, possibly conflicting separation gaps. In this particular application, the separation-deviation model (henceforth abbreviated as SD-model) uses the purchase-times to create separation gaps among the different customers.

Example 4.3.1. *Consider the input given in the following table:*

a_i^k	a	b	c
A	8	-	1
B	-	2	6
C	9	1	$5.$

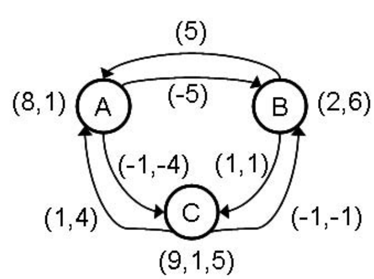
The first row of the table says that customer A bought products a and c eight and one months after each was launched; customer A did not buy product b. In addition to considering the information given in the above table, the model explicitly generates five separation gaps that represent the difference in the adoption promptness between each pair of customers. For instance, the model will explicitly use the fact that customer A bought product a one month before customer B.

The main motivations for considering separation gaps are the following:

1. We are interested in differentiating between customers that buy earlier and customers that buy later; this is a question about the relative purchase times between customers. In this respect it is important to emphasize that we are not concerned with the problem of predicting the time when a certain customer will buy a given product (the "absolute" purchase times of each customer).
2. While the specific months of purchase for different products might have a high variation, the difference in the purchase times between two given customers might have less variation. So for example say that Alice buys two products in months 1 and 100 respectively and Bob buys the same products in months 3 and 110 respectively. Just looking at Alice's purchases it would be unclear to determine if she is an early adopter or a late adopter; however when considering that she always bought the products earlier than Bob we can be certain that she adopts new products faster than Bob.

The model is formalized with a weighted directed graph $G = (V, A)$, where each node in V represents a customer and each arc $(u, v) \in A$ represents a separation gap between customers u and v. Every node, v, on the graph is associated with the respective vector $\boldsymbol{a}_v$ (containing the purchase times of customer v). Similarly, every arc (u, v) on the graph has associated a vector of weights containing all the separation gaps between customers u and v.

The graph representing the table given in Example 4.3.1 is:



In the application described in this chapter, separation gaps between customers are derived from the observed first purchase times described above. Specifically, let a_i^k and a_j^k be the observed first purchase times of customers i and j respectively, both of whom bought product k. Then the separation gap between the two customers is defined as $\delta_{ij}^k = a_i^k - a_j^k$; note that comparisons are skew-symmetric, since $\delta_{ij}^k = -\delta_{ji}^k$. Ultimately the output of the SD model is a rating on the set of objects so that the difference between the scores of each object-pair violates as little as possible the input separation gaps. Let the output rating of the ith customer be denoted by x_i, and let z_{ij} denote the output separation gap between the ith and jth customers.

The penalty function $f_{ij}^k()$ for violating a separation gap given by the kth product between the ith and jth customers is set to be a convex function of the *violation*, $z_{ij} - \delta_{ij}^k$. The total sum of these penalties, $\sum_k \sum_i \sum_j w_{ij}^k f_{ij}^k(z_{ij} - \delta_{ij}^k)$ is called the *separation penalty.* As in MDS, w_{ij}^k are given nonnegative weights reflecting the importance or precision of the dissimilarity δ_{ij}^k and are set to 0 if δ_{ij}^k is unknown (that is, if one of the ith and jth customers did not bought the kth product).

The penalty function $g_i^k()$ for deviating from the purchase time of product k by the ith customer is set to be a convex function of the *violation*, $x_i - a_i^k$. The total sum of these penalties $\sum_k \sum_i v_i^k g_i^k(x_i - a_i^k)$ is called the *deviation penalty.* The v_{ij}^k are given nonnegative weights reflecting the importance or precision of the purchase time a_i^k and are set to 0 the ith customer did not bought the kth product.

The objective function of the model is the sum of the separation penalty and the deviation penalty. Therefore, the optimal rating, $\boldsymbol{x}$, and optimal separation gaps $\boldsymbol{z}$ are the solutions to the following problem:

$$\min_{\boldsymbol{x},\boldsymbol{z}} \quad M \cdot \sum_{k=1}^{K} \sum_{i=1}^{k} \sum_{j=1}^{n} w_{ij}^k f_{ij}^k(z_{ij} - \delta_{ij}^k) + \sum_{k=1}^{K} \sum_{i=1}^{n} v_i^k g_i^k(x_i - a_i^k) \tag{4.6a}$$

$$\text{s.t.} \quad z_{ij} = x_i - x_j \quad (i = 1, \dots, n;\ j = i+1, \dots, n). \tag{4.6b}$$

In problem (4.6) the parameter M is chosen so that the separation penalty is *lexicographically* more important than the deviation penalty. By lexicographically more important we mean that the separation penalty is the dominant term in the optimization problem so that the deviation penalty is only used to choose among the feasible solutions with minimum separation penalty. We set the separation penalty to be lexicographically more important than the deviation penalty since it better represents our objective to "learn" the relative purchase-time ordering of the customers (as opposed to learning the absolute purchase times of the customers).

In this study, we use for the penalty functions $f_{ij}^k()$ and $g_i^k()$ the absolute value functions (problem (4.7)) and the quadratic penalty functions (problem (4.8)). In this study we set w_{ij}^k equal to 1 if both customers i and j bought product k, and set w_{ij}^k equal to 0 otherwise. Similarly, we set v_i^k equal to 1 if customers i bought product k, and set v_i^k equal to 0 otherwise.

As demonstrated in Chapter 3, when all products are purchased by all customers, using the absolute value penalty function guarantees that the output rating will agree with the rating implied by the majority of the products; and be close to the group consensus in the case where there is no clear majority. We denote the optimal solution to problems (4.7) and (4.8) as $\boldsymbol{x}^{|SD|}$ and $\boldsymbol{x}^{(SD)^2}$ respectively.

$$\min_{\boldsymbol{x},\boldsymbol{z}} \quad M \cdot \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^k |z_{ij} - \delta_{ij}^k| + \sum_{k=1}^{K}\sum_{i=1}^{n} v_i^k |x_i - a_i^k| \tag{4.7a}$$

$$\text{s.t.} \quad z_{ij} = x_i - x_j \quad (i = 1, \ldots, n;\ j = i+1, \ldots, n). \tag{4.7b}$$

$$\min_{\boldsymbol{x},\boldsymbol{z}} \quad M \cdot \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^k (z_{ij} - \delta_{ij}^k)^2 + \sum_{k=1}^{K}\sum_{i=1}^{n} v_i^k (x_i - a_i^k)^2 \tag{4.8a}$$

$$\text{s.t.} \quad z_{ij} = x_i - x_j \quad (i = 1, \ldots, n;\ j = i+1, \ldots, n). \tag{4.8b}$$

Example 4.3.2. *The following table provides the optimal solutions of problems* (4.7) *and* (4.8) *for the data in Example 4.3.1. The column* $\boldsymbol{x}^{Avg}$ *gives the row average of the inputs.*

	Inputs			**Outputs**		
	a	b	c	$\boldsymbol{x}^{\mid SD\mid}$	$\boldsymbol{x}^{(SD)^2}$	$\boldsymbol{x}^{Avg}$
A	*8*	-	*1*	*1*	*2.9*	*4.5*
B	-	*2*	*6*	*6*	*5.6*	*4.0*
C	*9*	*1*	*5*	*5*	*5.0*	*4.6*

As the above table shows, both $\boldsymbol{x}^{|SD|}$ *and* $\boldsymbol{x}^{(SD)^2}$ *preserve the order of the customers implied by all the products purchased. On the other hand, taking the row average does not preserve such order; indeed it contradicts the ordering of customers* (B, C) *according to product* b*, and the ordering of the customers* (A, B) *and* (B, C) *according to product* c*.*

4.4 Performance assessment on simulated scenarios

This section assesses the performance of the SD model under several different simulated scenarios and compares its performance to that of three-way UDS (problem (4.3)).

We denote as $\boldsymbol{x}^{UDS}$ the customer rating obtained using three-way UDS, that is, the solution to problem (4.3). Recall that obtaining the optimal solution to problem (4.3) is only possible (with current optimization techniques) for $n \leq 30$ [45]. While specialized heuristics to solve UDS are available (see [46] for a survey of these heuristics), none of them apply to the three-way UDS or to the weighted UDS (that is, all of the heuristics assume that the data is complete and $w_{ij} = 1$ for $i, j = 1, ..., n$). Therefore in order to find an heuristic solution to problem (4.3) we used Matlab's heuristic to solve the weighted MDS problem. Strictly speaking, Matlab's heuristic was designed to minimize problem (4.1) (that is, it only accepts one dissimilarity matrix). However, as shown in [23], when using the quadratic function as penalty function, minimizing (4.2) can be reduced to the problem of minimizing (4.1).

Each scenario represents a different customer purchase-timing behavior, and consists of 600 customers each buying up to 4 products. We associate a different purchase-time distribution with each customer-product pair. By letting the purchase-time distributions depend on both the customer and the product, we are able to simulate scenarios where the products have different life cycle and characteristics. In these scenarios, the customers' purchase-times may have different expected values and/or variances depending on the product under consideration.

We simulated the purchase time of each product by each customer using the gamma distribution, which is commonly used to simulate "customer arrival times". Let c and p represent the index of the customer and product, respectively. We used 7 different expected purchase times $\{c, c+2p, c+5p, c+50p, cp, 10cp, 50cp\}$, and 11 different variances $\{10, 50, 5c, 10c, 50c, 5p, 10p, 50p, 5c+5p, 10c+10p, 50c+50p\}$. Overall, we simulated 77 difference scenarios, one for each possible mean-variance combination. For example, in the scenario having cp mean and $5p+5c$ variance, the purchase time of the jth product by the ith customer had an expected value of ij and a variance of $5i+5j$. Note that, in all of these scenarios, customers with lower indices adopt new products earlier. That is, given two customers i, j such that $i < j$, then, for any given product, customer i has an earlier expected-purchase-time than customer j. Thus, for every one of the 77 scenarios, the customers are ordered with respect to their adoption promptness. In particular, the *true ranking*, x_i^T, of the ith customer is equal to i.

In order to measure how successful the SD model is in recovering the true ranking vector, $\boldsymbol{x}^T$, of the customers we used Kendall's *Tau* rank-correlation coefficient. This coefficient provides a measure of the degree of correspondence between two vectors. In particular, it assesses how well the order (i.e. rank) of the elements of the vectors is preserved. We note that, as an alternative to three-way UDS (problem (4.3), and the SD model (problems (4.7) and (4.8)), we could instead find the customer rating vector that maximizes Kendall's *Tau* rank-correlation coefficient. We decided not to do so because (1) finding such a vector is NP-hard [6] and (2) this objective would ignore the degree of dissimilarity between the adoption times of the customers. On the other hand, we believe that Kendall's *Tau* rank-correlation coefficient is appropriate to measure how well the customer ratings recovered x_i^T (notice that x_i^T gives the true ordering of the customers; and does not give a degree of dissimilarity between the customers).

Recall that this chapter focuses on the case where the data available is incomplete; that is, there are customers who did not purchase every product. To generate incomplete data, we first simulated the complete data; that is, we simulated the purchase times of every customer-product pair and then deleted some of the purchase-times at random. In Sun's data 59%, 23%, 9% and 9% of the customers bought one, two, three and four products, respectively. We mimicked this data by deleting the entries with this empirical distribution. In particular, each customer had a probability of 0.59, 0.23, 0.09 and 0.09 of buying one, two, three, and four products. For each customer the purchased products were chosen uniformly at random.

To summarize, the performance assessment of the SD model on each of the 77 scenarios was executed as follows:

Step 1: Repeat 30 times:

- **Step 1.1:** Simulate the purchase-time data of four products by 600 customers.
- **Step 1.2:** Delete some of the purchase times at random in order to obtain incomplete information.
- **Step 1.3:** Solve for $\boldsymbol{x}^{|SD|}$, $\boldsymbol{x}^{(SD)^2}$ and $\boldsymbol{x}^{UDS}$.
- **Step 1.4:** Compute the Tau correlation coefficient between $\boldsymbol{x}^{T}$ (the true customer ranking) and each of $\boldsymbol{x}^{|SD|}$, $\boldsymbol{x}^{(SD)^2}$ and $\boldsymbol{x}^{UDS}$.

Step 2: Calculate the average and standard deviation of the 30 Tau correlation coefficients (with $\boldsymbol{x}^{T}$) computed for each of $\boldsymbol{x}^{|SD|}$, $\boldsymbol{x}^{(SD)^2}$ and $\boldsymbol{x}^{UDS}$.

Tables 4.1, 4.2, and 4.3 give, for each of the 77 scenarios, the average Tau correlation coefficient between $\boldsymbol{x}^{T}$ and $\boldsymbol{x}^{|SD|}$, $\boldsymbol{x}^{(SD)^2}$ and $\boldsymbol{x}^{UDS}$, respectively.

In order to facilitate comparing the performances of these methods, Tables 4.4, 4.5, and 4.6 give the average differences between the correlation coefficients achieved by different methods. For example, each entry in Table 4.4 is the difference between the corresponding entries of Tables 4.1 and 4.3. In Tables 4.4, 4.5, and 4.6, the numbers given in **bold** are those that are at least three standard deviations above (or below) zero. Therefore the scenarios with **bold** entry are those where one method significantly outperforms the other; while, in the rest of the scenarios, the performance of both methods is essentially the same.

It is clear from Tables 4.4 and 4.5 that the SD model, irrespective of the penalty functions used, is better than UDS on most of the 77 scenarios. In particular, the SD model outperforms UDS in 46 out of the 77 scenarios; and in 36 of these 46 scenarios, the SD model significantly outperforms UDS.

We now compare the performance of the SD model using different penalty functions. For this purpose, Table 4.6 reports the average difference between the correlation coefficients (with respect to $\boldsymbol{x}^{T}$) for $\boldsymbol{x}^{|SD|}$ and $\boldsymbol{x}^{(SD)^2}$. We note that using the absolute value penalty functions is only slightly better than using the quadratic value penalty functions.

The simulation results indicate that the SD model determines with high accuracy the true ranking of the customers with respect to their adoption promptness.

Table 4.1: Average *Tau* correlation coefficients between $\boldsymbol{x}^T$ and $\boldsymbol{x}^{|SD|}$.

τ	**10**	**50**	**5c**	**10c**	**50c**	**5p**	**10p**	**50p**	**5c+5p**	**10c+10p**	**50c+50p**
c	0.9913	0.9784	0.8884	0.8434	0.6934	0.9902	0.9857	0.9662	0.8873	0.8433	0.6903
c+2p	0.9911	0.9783	0.8872	0.8445	0.6947	0.9903	0.9855	0.9666	0.8879	0.8450	0.6903
c+5p	0.9912	0.9785	0.8886	0.8455	0.6900	0.9905	0.9856	0.9666	0.8851	0.8444	0.6906
c+50p	0.9913	0.9784	0.8877	0.8448	0.6847	0.9904	0.9856	0.9665	0.8863	0.8430	0.6830
cp	0.8427	0.8386	0.8271	0.8198	0.7745	0.8395	0.8416	0.8428	0.8270	0.8161	0.7766
10cp	0.8431	0.8449	0.8432	0.8419	0.8403	0.8439	0.8424	0.8441	0.8438	0.8426	0.8391
50cp	0.8426	0.8443	0.8433	0.8437	0.8452	0.8437	0.8433	0.8438	0.8451	0.8462	0.8423

Table 4.2: Average *Tau* correlation coefficients between $\boldsymbol{x}^T$ and $\boldsymbol{x}^{(SD)^2}$.

τ	**10**	**50**	**5c**	**10c**	**50c**	**5p**	**10p**	**50p**	**5c+5p**	**10c+10p**	**50c+50p**
c	0.9915	0.9788	0.8903	0.8464	0.6958	0.9902	0.9858	0.9660	0.8891	0.8453	0.6928
c+2p	0.9913	0.9787	0.8891	0.8468	0.6971	0.9902	0.9856	0.9666	0.8897	0.8474	0.6936
c+5p	0.9914	0.9788	0.8904	0.8478	0.6925	0.9905	0.9856	0.9666	0.8873	0.8467	0.6932
c+50p	0.9915	0.9789	0.8902	0.8474	0.6892	0.9904	0.9857	0.9665	0.8882	0.8452	0.6871
cp	0.8356	0.8322	0.8241	0.8169	0.7745	0.8333	0.8351	0.8365	0.8238	0.8140	0.7758
10cp	0.8349	0.8380	0.8357	0.8350	0.8339	0.8360	0.8348	0.8371	0.8368	0.8360	0.8331
50cp	0.8348	0.8373	0.8351	0.8366	0.8385	0.8362	0.8365	0.8368	0.8372	0.8383	0.8347

Table 4.3: Average *Tau* correlation coefficients between $\boldsymbol{x}^T$ and $\boldsymbol{x}^{UDS}$.

τ	**10**	**50**	**5c**	**10c**	**50c**	**5p**	**10p**	**50p**	**5c+5p**	**10c+10p**	**50c+50p**
c	0.9787	0.9781	0.9766	0.9767	0.7198	0.7618	0.7613	0.9708	0.9712	0.9710	0.9704
c+2p	0.7602	0.7697	0.7662	0.8871	0.8868	0.8879	0.8867	0.7610	0.7691	0.7645	0.8442
c+5p	0.8450	0.8450	0.8458	0.7583	0.7677	0.7662	0.6893	0.6917	0.6769	0.6842	0.7287
c+50p	0.7659	0.7678	0.9770	0.9782	0.9764	0.9756	0.7644	0.7664	0.7688	0.9764	0.9742
cp	0.9739	0.9740	0.7624	0.7676	0.7668	0.9603	0.9599	0.9617	0.9612	0.7665	0.7681
10cp	0.7674	0.8862	0.8874	0.8843	0.8854	0.7578	0.7150	0.7650	0.8443	0.8456	0.8436
50cp	0.8433	0.7188	0.7171	0.7195	0.6856	0.6887	0.6886	0.6816	0.7305	0.7628	0.7651

Table 4.4: Average difference between the correlation coefficients obtained by $\boldsymbol{x}^{|SD|}$ and $\boldsymbol{x}^{UDS}$. Positive numbers indicate that $\boldsymbol{x}^{|SD|}$ has higher average correlation with $\boldsymbol{x}^{T}$ than $\boldsymbol{x}^{UDS}$.

	10	**50**	**5c**	**10c**	**50c**	**5p**	**10p**	**50p**	**5c+5p**	**10c+10p**	**50c+50p**
c	0.0126	0.0003	**-0.0882**	**-0.1333**	-0.0265	**0.2283**	**0.2244**	-0.0046	**-0.0840**	**-0.1277**	**-0.2801**
c+2p	**0.2310**	**0.2087**	**0.1210**	**-0.0427**	**-0.1921**	**0.1024**	**0.0988**	**0.2056**	**0.1188**	**0.0805**	**-0.1539**
c+5p	**0.1462**	**0.1335**	**0.0427**	**0.0872**	**-0.0778**	**0.2243**	**0.2962**	**0.2748**	**0.2082**	**0.1603**	-0.0381
c+50p	**0.2254**	**0.2106**	**-0.0893**	**-0.1333**	**-0.2918**	0.0148	**0.2212**	**0.2000**	**0.1175**	**-0.1334**	**-0.2912**
cp	**-0.1312**	**-0.1354**	**0.0647**	**0.0522**	0.0076	**-0.1207**	**-0.1183**	**-0.1189**	**-0.1343**	0.0496	0.0085
10cp	**0.0756**	**-0.0413**	**-0.0442**	**-0.0424**	**-0.0451**	**0.0861**	0.1274	**0.0792**	-0.0005	-0.0030	-0.0045
50cp	-0.0007	0.1255	0.1262	0.1241	**0.1596**	**0.1549**	**0.1547**	**0.1622**	**0.1145**	**0.0834**	**0.0773**

Table 4.5: Average difference between the correlation coefficients obtained by $\boldsymbol{x}^{(SD)^2}$ and $\boldsymbol{x}^{UDS}$. Positive numbers indicate that $\boldsymbol{x}^{(SD)^2}$ has higher average correlation with $\boldsymbol{x}^{T}$ than $\boldsymbol{x}^{UDS}$.

	10	**50**	**5c**	**10c**	**50c**	**5p**	**10p**	**50p**	**5c+5p**	**10c+10p**	**50c+50p**
c	0.0128	0.0007	**-0.0863**	**-0.1303**	-0.0241	**0.2283**	**0.2244**	-0.0048	**-0.0822**	**-0.1257**	**-0.2776**
c+2p	**0.2311**	**0.2090**	**0.1230**	**-0.0403**	**-0.1896**	**0.1024**	**0.0988**	**0.2056**	**0.1205**	**0.0829**	**-0.1506**
c+5p	**0.1464**	**0.1338**	**0.0445**	**0.0894**	**-0.0752**	**0.2243**	**0.2963**	**0.2749**	**0.2103**	**0.1626**	-0.0356
c+50p	**0.2256**	**0.2110**	**-0.0869**	**-0.1307**	**-0.2872**	0.0148	**0.2213**	**0.2000**	**0.1195**	**-0.1312**	**-0.2871**
cp	**-0.1383**	**-0.1417**	**0.0617**	**0.0493**	0.0077	**-0.1270**	**-0.1248**	**-0.1251**	**-0.1375**	0.0475	0.0077
10cp	**0.0675**	**-0.0483**	**-0.0517**	**-0.0494**	**-0.0516**	**0.0782**	0.1199	**0.0721**	-0.0075	-0.0096	-0.0104
50cp	-0.0085	0.1185	0.1179	0.1171	**0.1529**	**0.1475**	**0.1479**	**0.1551**	**0.1067**	**0.0754**	**0.0696**

Table 4.6: Average difference between the correlation coefficients obtained by $\boldsymbol{x}^{|SD|}$ and $\boldsymbol{x}^{(SD)^2}$. Positive numbers indicate that $\boldsymbol{x}^{|SD|}$ has higher average correlation with $\boldsymbol{x}^{T}$ than $\boldsymbol{x}^{(SD)^2}$.

	10	**50**	**5c**	**10c**	**50c**	**5p**	**10p**	**50p**	**5c+5p**	**10c+10p**	**50c+50p**
c	-0.0002	-0.0004	-0.0019	-0.0030	-0.0024	0.0000	-0.0001	0.0002	-0.0018	-0.0020	-0.0025
c+2p	-0.0001	-0.0004	-0.0020	-0.0023	-0.0024	0.0000	0.0000	0.0000	-0.0017	-0.0024	-0.0033
c+5p	-0.0002	-0.0003	-0.0018	-0.0022	-0.0026	0.0000	-0.0001	0.0000	-0.0021	-0.0023	-0.0026
c+50p	-0.0002	-0.0004	-0.0024	-0.0026	-0.0045	0.0000	0.0000	0.0000	-0.0020	-0.0023	-0.0041
cp	0.0071	0.0064	0.0030	0.0029	0.0000	0.0063	0.0065	0.0062	0.0032	0.0021	0.0008
10cp	0.0081	0.0069	0.0075	0.0069	0.0064	**0.0079**	0.0075	0.0070	0.0070	0.0066	0.0059
50cp	0.0078	0.0070	**0.0082**	0.0071	0.0067	0.0074	0.0068	0.0071	**0.0079**	**0.0079**	0.0077

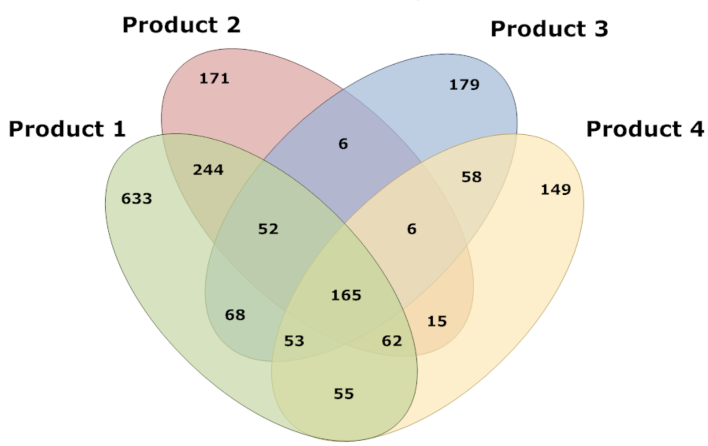
4.5 Rating Sun's customers according to their adoption promptness

4.5.1 Sun's data

The empirical analysis presented below is based on a (disguised) dataset comprising customer purchase information provided by Sun Microsystems, Inc. The dataset encompasses four products and some 1,916 customers. It records the number of months (measured from the month of the earliest product launch) that elapsed before each customer bought each product. This section shows that Sun's products are not independent in several ways (for instance all four products are servers in the same family), and we propose how to cope with this situation.

As shown in Figure 4.2, most of the customers did not buy all four products, and in fact about half of the customers only bought one of the products. Such sparse data would pose a challenge for many of the existing data-mining and market segmentation techniques described in Section 4.2, and in general, some form of preprocessing would be required to fill in the missing data. The separation deviation model, however, handles missing data quite routinely, without preprocessing.

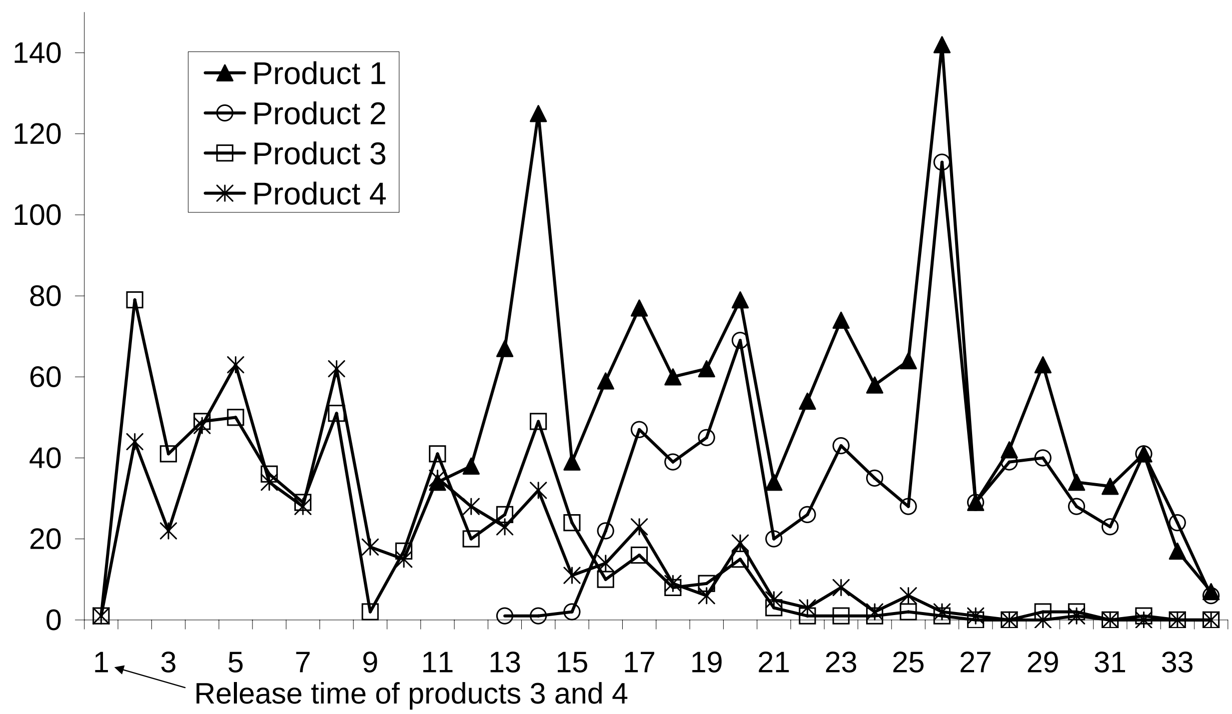
Figure 4.2: Number of customers that bought each product or set of products.



As may be deduced from Figure 4.3, products 3 and 4 were launched together at the beginning of the observation period, with the launches of products 1 and 2 following respectively 10 and 12 months later—in fact, products 3 and 4 represent the first generation of a product line of which products 1 and 2 were the second generation, with updated and advanced features. The figure also exhibits the strong degree of correlation between product sales; peaks and valleys in the sales of all products tend to occur at the same time (this sales behavior is almost certainly due to the effect of salesforce and customer incentives that the company applied simultaneously to all products in this market). Moreover, as shown in

Figure 4.4, most of the customers that bought products 1 and 2 did not buy either product 3 or 4. Therefore, it is reasonable to suppose that purchasers of the earlier products (products 3 and 4) exhibit greater adoption promptness than purchasers of products 1 and 2 alone.

Figure 4.3: Number of customers that bought each product per month.



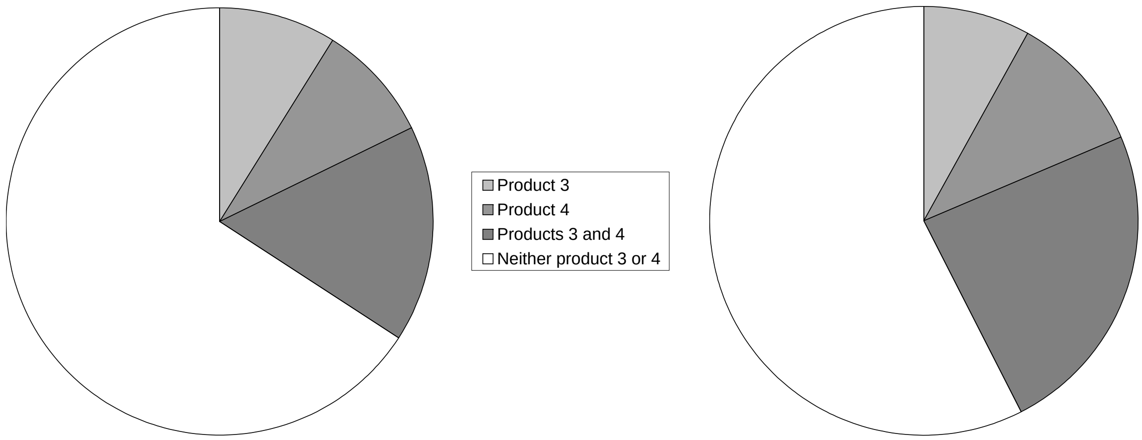
In general the purchase times should be measured from the release date of each product. Because products 1 and 2 are second generation of products 3 and 4 we find that the purchase times of products 3 and 4 are more significant in determining the adoption promptness. In order to consider this, we decided to measure all purchase times with respect to the release time of products 3 and 4. That is, we still consider the purchase times of products 1 and 2, but measure these times with respect to the launch-time of products 3 and 4—as opposed to measuring these times from the launch time of the respective product.

4.5.2 Results and their interpretation

In this section we demonstrate the use of the SD model to rate Sun's customers with respect to their adoption promptness. In particular, we show that the results obtained using the SD model, agree with an intuitive interpretation of Sun's business.

Using as input Sun's data, we solved for $\boldsymbol{x}^{|SD|}$, which was the best performer in Section 4.4. In order to facilitate the interpretation of the obtained results, we generated 4 customer classes from $\boldsymbol{x}^{|SD|}$. Specifically, we classified Sun's customers into the classes de-

Figure 4.4: Customers that bought product 1 or 2 after buying products 3 and/or 4.



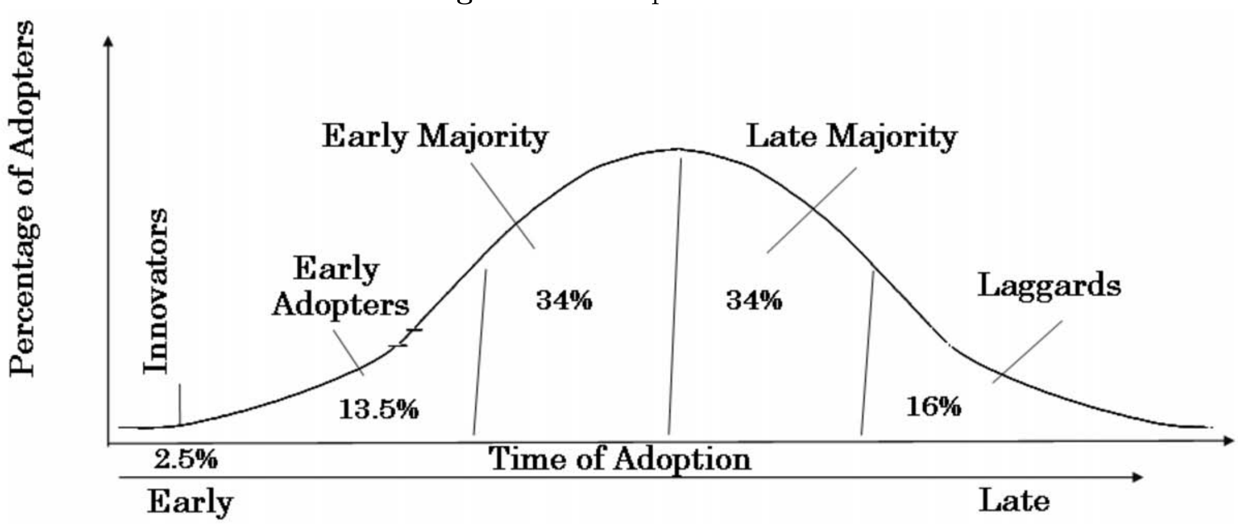
(a) Product 1 (b) Product 2

fined by Rogers' model of innovation diffusion (see Section 4.5.3). That is, we segmented the customers into 4 classes (Vanguard—composed of innovators and early adopters—, Early Majority, Early Minority, and Laggard) as follows: (1) We sorted the customers according to their rating as given by $\boldsymbol{x}^{|SD|}$. (2) We selected threshold values determining the boundaries between consecutive segments, so that the segments have the sizes given by Rogers's model.

4.5.3 Rogers' model of innovation diffusion, an overview

According to Rogers' now-classic model of innovation diffusion [54], customers may be classified based on the timing of their first purchase of a new product into: *Innovators*, who are the first to purchase the product and use it. This group of people are typically well-educated, adventurous and open to new experiences. Product purchases by those outside of this group are influenced to various degrees by the reactions of innovators. Later purchasers are essentially imitators; they buy new products because the innovators had positive reactions to them and they wish to replicate the innovators experience. *Early adopters* begin purchasing as the innovators communicate positive responses toward a product. This group is made up of people that are inclined to try new ideas, but tend to be cautious. *Early majority* adopters are more likely to accept a new product than the average person; rarely acting as leaders, the early majority essentially imitates the behavior of the first adopters. *Late majority* customers who decide to buy only because many other customers have already tried the new product. Finally, *laggard* adopters are reluctant to adopt new products. Those in this group buy a new product only once this product has established a substantial track record.

Figure 4.5: Adoption Process



As seen in Figure 4.5, Innovators represent only 1.5% of all customers. Thus, throughout the chapter we conflate the classes Innovators and Early adopters under the title *Vanguard*. Customers in the Vanguard class play a major role in the adoption of the innovation, since their acceptance or rejection will affect all the other groups. The method proposed below aims to identify more effectively this group of customers and, therefore, increase the probability of success of a target action. Furthermore, the proposed technique identifies the other groups of customers: Early Majority, Late Majority and Laggard. In this sense, it is a valuable instrument for the design of marketing strategies.

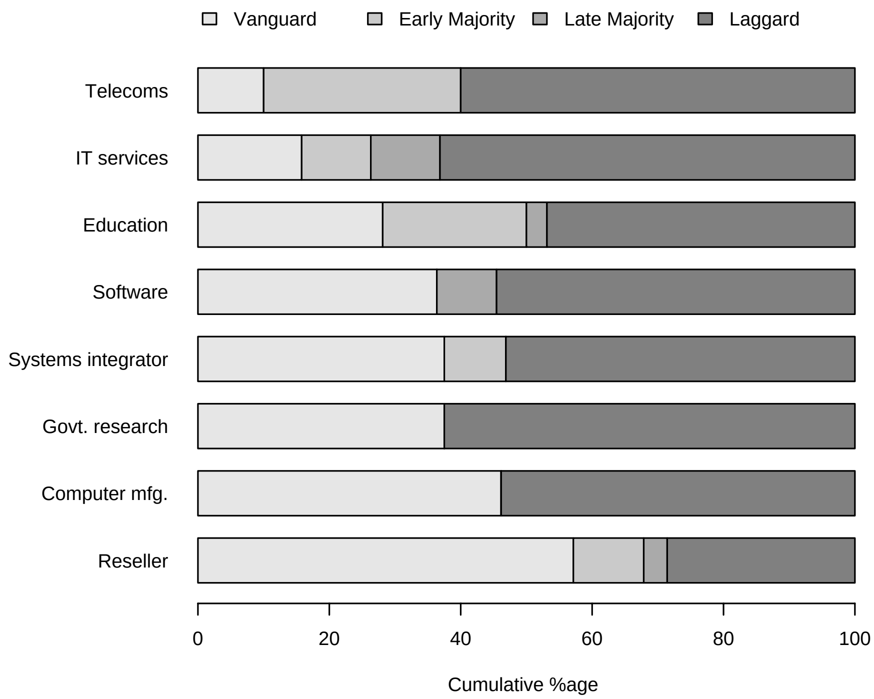
4.5.4 Interpretation

Next we provide the interpretation of the obtained results. As mentioned previously, we classified Sun's customers into the classes defined by Rogers' model of innovation diffusion as follows: using as input Sun's data, we solved for $\boldsymbol{x}^{|SD|}$, then we sorted the customers according to their rating as given by $\boldsymbol{x}^{|SD|}$, and finally we selected threshold values determining the boundaries between consecutive segments, so that the segments have the sizes given by Rogers's model.

Figures 4.6 and 4.7 provide an analysis of the customer segmentation in terms of customer industry and location, respectively. The bars in the figures relate the percentage of each characteristic comprised by customers of a particular class. Thus in Figure 4.6, just under 60% of resellers are in the Vanguard, and 10% are Early Majority, while in Figure 4.7, approximately 50% of US customers are in the Vanguard, and about 40% are Laggards.

Broadly speaking, the results illustrated in the Figures accord with an intuitive understanding of Sun's business. In Figure 4.6, for example, resellers and computer manufacturers must pass the product on to end-users, and are thus likely to be first in line to purchase a

Figure 4.6: Customer classification by industry



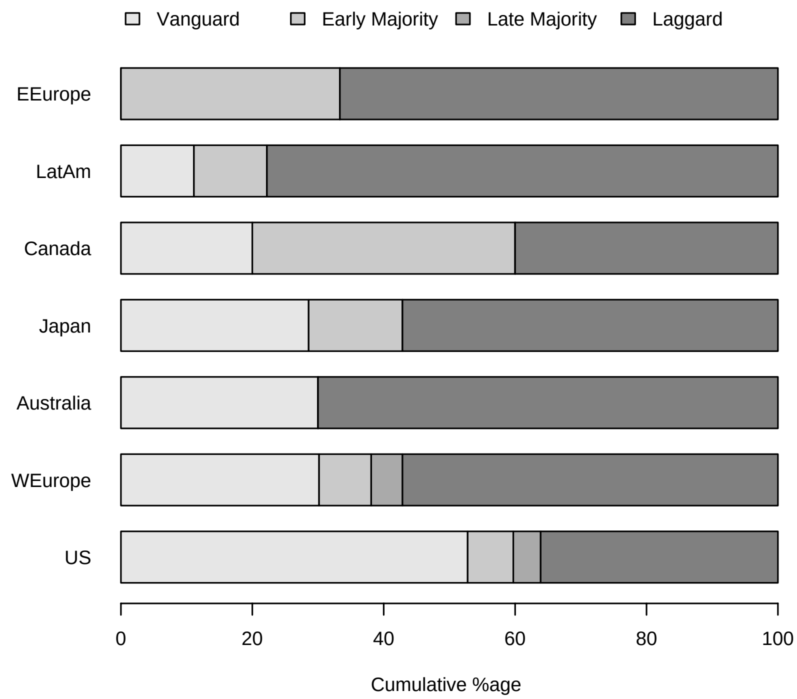
new model. By contrast, telecommunications utilities ("telecoms") have high fixed capital investments and very exacting quality standards, and it is quite reasonable to see this category skewed toward the Laggard class. Figure 4.7 seems well grounded in the geography of Sun's markets: Since it is a US-based company, one would expect a preponderance of Sun's US customers to be Vanguard and Early Majority, with adoptions occurring fairly early on in the developed markets of western Europe, Australia, Japan and Canada. Less developed markets, such as Latin America and eastern Europe, where Sun's sales and distribution infrastructure is less well established, adopt later. Overall, it appears that the classification obtained using $\boldsymbol{x}^{|SD|}$ do indeed characterizes Sun's customers in a convincing fashion.

4.6 Conclusions

The proposed approach of using the separation-deviation model (SD model) is novel in data mining in general and customer rating in particular. It is shown here to generate valuable information on the characteristics of the customer base of an organization and as such it is useful in managing the launch and the life-cycle of a new product.

In this study, we used the SD model in order to rate Sun's customers according to their adoption promptness. Using the ratings obtained, we were able to classify Sun's customers

Figure 4.7: Customer classification by location



according to their adoption promptness and show that the results give an intuitive interpretation of Sun's business.

We have shown that the SD model is a valuable alternative to traditional unidimensional scaling techniques. In particular we established that the SD model outperforms unidimensional scaling in determining the customers' adoption promptness.

The potential of the method is broad and suggests its use for different types of scaling problems.

Chapter 5

Simultaneous aggregation of cardinal and ordinal evaluations: ranking in a student paper competition

5.1 Introduction

This chapter presents a new framework for group decision making when a group of individuals needs to collectively rank all of the objects in a set. Previous work focuses either on the ranking aggregation problem (see, for example, [49, 5, 6]), or on the rating aggregation problem (see, for example, [48, 56, 3]), but not both. One of the prime contributions in this chapter is to simultaneously aggregate rankings and ratings into a consensus evaluation.

The motivation of this chapter is to rank the participants of the 2007 MSOM's student paper competition (**SPC**). The particular group-ranking problem arising in this SPC poses unique challenges:

1. The judges provided both ratings and rankings of the papers they reviewed.
2. The incompleteness of the evaluations is extreme: Each judge evaluated only less than ten percent of the papers, and each paper was reviewed by less than ten percent of the judges. This causes the aggregation to be subject to the "incomplete evaluation" phenomenon bias. That is, grades that are too low or too high tend to dominate the aggregate score of the papers.

In a rating aggregation problem it is important to use an aggregation method that is insensitive to the subjective scales used by the judges. French [29] argues that, the *value difference functions* (the rating scales) of two individuals involve an arbitrary choice of scale and origin and thus the same numeric score from two different judges generally do not have the same meaning. Similarly, in the context of international surveys, a large number of studies (see, for example, [7, 57, 36]) show that the responses across different countries do not have the same meaning. In particular, even when asking respondents to rate each object using a simple 5-point rating scale, there are significant differences in the *response styles* between countries. One example of a difference in response style, that arises even when using a simple 5-point rating scale, is that in some countries there is a tendency to use only the extreme categories while in others there is a tendency to use only the middle categories. Another example of a difference in response style, is that in some countries there is a tendency to use only the top categories while in others there is a tendency to use only the bottom categories.

As noted above, in the 2007 MSOM's SPC, the incompleteness of the evaluations received is extreme. Therefore it is particularly important to mitigate the bias introduced by outlier scores because an object assigned to a strict (lenient) judge has an advantage (disadvantage) compared to objects that were not assigned to this specific judge.

The methodology developed in this chapter includes two mechanisms to limit the sensitivity to the subjective evaluation scales used by the judges. The first mechanism is the use of pairwise comparisons (defined in Section 5.4), which are less sensitive to the subjective scales than the given cardinal scores. The second mechanism is the simultaneous use of cardinal and ordinal evaluations. Ordinal evaluations are independent of the individual subjective scales [29].

We now give a brief description of the group-decision making framework developed in this chapter. The method to simultaneously aggregate the ratings and rankings consists of finding the *combined aggregate rating* (CAT) and its implied ranking, referred to as *combined aggregate ranking* (CAK). This rating-ranking pair is the one that minimizes the sum of the distances from the CAT to the judges' ratings plus the sum of the distances from the CAK to the judges' rankings. The aggregation method is supplemented by two methods to identify inconsistencies in the evaluations of the objects. This information is helpful to identify judges whose rating scales significantly differ form those used by the rest of the judges.

The chapter is organized as follows: Section 5.2 reviews the literature on group-decision making and in particular ranking and rating aggregation techniques. Section 5.3 describes the evaluation methodology used in the 2007 MSOM's SPC, and gives examples where the differences in scale used by the judges are evident. Section 5.4 describes our methodology to simultaneously use the given ratings and rankings in order to obtain the CAT and its implied CAK. Section 5.5 uses the framework presented in Section 5.4 to rank the contestants in the 2007 MSOM's SPC and analyzes the obtained results. Finally, Section 5.6 comments on our group-decision making framework and its usefulness for different applications and decision-making scenarios.

5.2 Literature review

The ranking aggregation problem has been studied extensively, especially in the social choice literature. In this context, one of the most celebrated results is Arrows's impossibility theorem [5], which states that there is no "satisfactory" method to aggregate a set of rankings.

For Arrow, a satisfactory method is one that satisfies the properties: universal domain, no imposition, monotonicity, independence of irrelative alternatives, and non-dictatorship.

Kemeny and Snell [49] proposed a set of axioms that a distance metric between two complete rankings must satisfy. They proved that these axioms were simultaneously satisfied by a unique metric: this distance between two rankings is given by the number of *rank reversals* between them. A rank reversal is incurred whenever two objects have a different relative order in the given rankings. Similarly, *half* a rank reversal is incurred whenever two objects are tied in one ranking but not in the other. Kemeny and Snell defined the consensus ranking as the ranking that minimizes the sum of the distances to each of the input rankings. Bartholdi et. al. [6] showed that the optimization problem that needs to be solved to find the Kemeny-Snell consensus ranking is NP-hard.

Following the work of Kemeny-Snell, several axiomatic approaches have been developed to determine consensus. For instance, Bogart [9] developed an axiomatic distance between partial orders. One of the applications of Bogart's distance is to determine a consensus partial order from a set of partial orders. Moreno-Centeno [52] developed an axiomatic distance between incomplete rankings. In this chapter we build upon this distance in order to rank the contestants in the 2007 MSOM's SPC.

One method of overcoming the difficulties arising in the ranking aggregation problem presented by Arrow's impossibility theorem and the NP-hardness of finding the Kemeny-Snell's consensus ranking, is to ask for the individuals' evaluations in the form of ratings, rather than rankings. Following this direction, Keeney [48] proved that the *averaging method* satisfied all of Arrow's desirable properties. In the averaging method, the consensus rating of each object is the average of the scores it received. The most immediate drawback of this approach is that the averaging method implicitly requires that all judges use the same rating scale; that is, that all individuals are equally strict or equally lenient in their score assignments.

As in the case of the ranking aggregation problem, the rating aggregation problem was first studied in the social choice literature. In this context, the ranking aggregation problem was transformed to a rating aggregation problem (e.g. by regarding the ordinal numbers as cardinal numbers) and then solved as a rating aggregation problem. The first and most simple of these methods is the so-called Borda count developed in 1770 by Jean-Charles de Borda.

The rating aggregation problem has also been studied in the multi-criteria decision making literature. (Hochbaum [41] showed the equivalence between the rating aggregation problem and the multi-criteria decision making problem). In this context, the non-axiomatic ELECTRE [11] and PROMETHEE [12] methods (and their extensions) solve the rating aggregation problem by transforming it in some sense to a ranking aggregation problem. This transformation is needed because each criteria is evaluated on a different scale.

5.3 The Data

The data used here is the evaluations for the 2007 MSOM's SPC. A total of 58 papers were submitted for this competition and a total of 63 judges participated in the evaluation process. Each of the 63 judges evaluated only three to five out of the 58 papers; and each of the 58 papers was evaluated by only three to five out of the 63 judges. Each judge reviewed and evaluated each of the papers that were assigned to him/her on each the following features:

A) Problem importance/interest (1–10),
B) Problem modeling (0–10),
C) Analytical results (0–10),
D) Computational results (0–10),
E) Paper writing (1–10), and
F) Overall contribution to the field (1–10).

On each feature, each judge assigned a score to the paper he/she reviewed; the score definitions (i.e., the rating scale) provided to the judges are found in Table 5.1. In addition, each judge also provided a ranking of the papers he/she reviewed (1 = best, 2 = second best, etc.).

Table 5.1: Interpretation of each numerical score. The journals considered are: MSOM, Operations Research (OR) and Management Science (MS).

Score	Definition / Interpretation
10	Feature considered is comparable to that of the best papers published in the journals.
8,9	Feature considered is comparable to that of the average papers published in the journals.
7	Feature considered is at the minimum level for publication in the journals.
5,6	Feature considered independently would require a minor revision before publication in the journals.
3,4	Feature considered independently would require a major revision before publication in the journals.
1,2	Feature considered would warrant by itself a rejection if the paper were submitted to the journals.
0	Feature considered is not relevant or applicable to the paper being evaluated.

Although these precise score definitions (Table 5.1) were provided, the judges' rating scales were different, as shown in Tables 5.2 and 5.3. (To maintain the anonymity of judges and papers the judge and paper identification numbers were assigned randomly.)

Table 5.2: Evaluations received on paper 43.

Judge	Problem Importance	Problem Modeling	Analytical Results	Computational Results	Paper Writing	Field Contribution	Paper Ranking
47	9	8	8	8	9	9	1
6	6	4	2	4	4	4.5	1
55	9	6	0	9	8	6	2
2	7	7	2	6	7.5	4	3

Table 5.2 illustrates that for paper 43, the judges' used different scales: Paper 43 received in the Problem Modeling category a score of 8 (meaning that the Problem Modeling in the paper are comparable to that in an average paper published in MSOM, OR and MS), and also a score of 4 (meaning that the problem modeling in the paper requires a major revision before publication in MSOM, OR and MS). These score differences are, by all standards,

not insignificant. Another example of the differences between the judges' scales is found on the Analytical Results category. In this category, a judge gave a score of 8 (meaning that the analytical results in the paper are comparable to those in an average paper published in MSOM, OR and MS), two judges gave a score of 2 (meaning that the analytical results in the paper are so bad that the paper should be rejected by MSOM, OR and MS), and the remaining judge considered that the category was not applicable to the paper (thus assigned the value of zero).

Table 5.3: Evaluations received on paper 26.

Judge	Problem Importance	Problem Modeling	Analytical Results	Computational Results	Paper Writing	Field Contribution	Paper Ranking
21	8	10	8	8	5	8	3
24	8	9	8	10	7	8	1
14	7	2	3	2	2	2	5
26	8	8	7	8	8	7	3
49	10	7	6	9	9	8	1

Table 5.3 shows that judge 14's evaluations are not in the same scale as the evaluations of the other judges (one could even argue that judge 14 is deliberately trying to manipulate the evaluation of this paper). In particular, note that in all of the features (with the exception of Problem Importance) judge 14 gives a score indicating that the paper would be rejected by MSOM, OR and MS; on the other hand in every feature all of the other judges consider the paper is worth of publishing (some of their evaluations even indicate that the paper would be among the best papers published in MSOM, OR and MS!). These discrepancies in the judges' evaluations are quite common throughout the data.

Another sample of the judges' evaluations for this study is given in Table 5.4.

Table 5.4: Sample of judges' evaluations.

Judge - Paper	Problem Importance	Problem Modeling	Analytical Results	Computational Results	Paper Writing	Field Contribution	Paper Ranking
28 - 55	8	6	6	6	8	5	4
28 - 23	8	9	9	9	9	9	1
28 - 8	9	9	10	9	9	9	1
28 - 1	7	7	6	6	8	6	3
60 - 11	8	7	7	7	7	8	1
60 - 6	5	5	4	6	4	4	4
60 - 46	7	5	5	0	5	6	3
60 - 34	8	7	7	8	5	8	1
56 - 22	4	0	0	6	4	4	4
56 - 17	8.5	7	8	7	8	8	2
56 - 58	7	7	7	6	6	7	3
56 - 39	9	8	8.5	7	8	8.5	1

Henceforth, we use as cardinal evaluation only the evaluation on the feature "Overall Contribution to the Field". This is because, the authors and Jérémie Gallien (head judge of the 2007 MSOM's SPC), believe that this feature is the single most important feature evaluated. This belief is supported by our detailed analysis (provided upon request) regarding the correlation of the ratings to the rankings. This analysis provides evidence that the vast majority of the judges based their paper rankings on their "Overall Contribution to the Field" rating, thus the vast majority of the judges also considered this feature to be the most important feature.

5.4 The method

In this section we describe the framework used to simultaneously aggregate cardinal and ordinal evaluations and how it was used to rank the student papers.

5.4.1 Simultaneous aggregation of ratings and rankings

In this section we describe the method to simultaneously aggregate the judges' ratings and the judges' rankings. The advantage of this technique over traditional techniques (only aggregating ratings or only aggregating rankings) is that it highlights the objects/papers whose ratings and ranking are in conflict with several other objects/papers. These conflicts generally arise in the following circumstances:

1. when a paper is assigned to strict judges and these judges rank it higher than the other papers they reviewed;
2. when a paper is assigned to lenient judges and these judges rank it lower than the other papers they reviewed;
3. when a low-quality paper is assigned to judges who reviewed other papers of lesser quality; and,
4. when a high-quality paper is assigned to judges who reviewed other papers of higher quality.

The purpose of our technique is not to identify the cause of these conflicts, but rather to highlight the papers with a high number of conflicts. This information is helpful so that (say) the manager of the competition can initiate an investigation of the nature of the discrepancies and act appropriately (for example, by discussing these inconsistencies with the judges that evaluated these papers).

A simple first step of our technique is motivated by Brans and Vincke's PROMETHEE method [12], which is used in the context of multi-criteria decision making. Given the consensus rating $\boldsymbol{x}^{(c)}$ (optimal solution to problem (2.16)) and the consensus ranking $\boldsymbol{x}^{(o)}$ (optimal solution to problem (2.11)) aggregate them into a partial order (P, I, R) as follows:

$$a \text{ is preferred to } b \ (a\ P\ b) \quad \text{if} \quad \begin{cases} \boldsymbol{x}^{(c)}(a) > \boldsymbol{x}^{(c)}(b) \text{ and } \boldsymbol{x}^{(o)}(a) \geq \boldsymbol{x}^{(o)}(b) \\ \boldsymbol{x}^{(c)}(a) \geq \boldsymbol{x}^{(c)}(b) \text{ and } \boldsymbol{x}^{(o)}(a) > \boldsymbol{x}^{(o)}(b) \end{cases} \tag{5.1a}$$

$$a \text{ is indifferent to } b \ (a\ I\ b) \quad \text{if} \quad \boldsymbol{x}^{(c)}(a) = \boldsymbol{x}^{(c)}(b) \text{ and } \boldsymbol{x}^{(o)}(a) = \boldsymbol{x}^{(o)}(b) \tag{5.1b}$$

$$a \text{ and } b \text{ are incomparable } (a\ R\ b) \quad \text{if} \quad \text{otherwise.} \tag{5.1c}$$

In essence the partial order (P, I, R) summarizes the agreement (or lack thereof) between the consensus rating $\boldsymbol{x}^{(c)}$ and the consensus ranking $\boldsymbol{x}^{(o)}$. The Hesse diagram of this partial order gives a graphical representation of which objects have a consensus rating and a consensus ranking that agree, and which ones are not. This graphical representation is helpful in determining object pairs where the cardinal and ordinal evaluations are not consistent.

The consensus rating is also used to identify objects such that the judges evaluating them had particularly divergent evaluations. These objects are those that assigned/received scores that disagree the most with the consensus rating $\boldsymbol{x}^{(c)}$. Specifically, these objects are those with the highest contribution to the separation penalty. The contribution of object i to the separation penalty is calculated as

$$\sum_{k|i\in\mathcal{A}^k}\sum_{j\in\mathcal{A}^k}\left|(\boldsymbol{x}_i^{(c)}-\boldsymbol{x}_j^{(c)})-(\boldsymbol{a}_i^k-\boldsymbol{a}_j^k)\right|. \tag{5.2}$$

In Section 5.5 we illustrate these methods and their usefulness when combining the consensus rating and the consensus ranking and for pinpointing objects whose evaluations deserve special attention/further discussion.

Next we develop a method for rating all the objects using simultaneously the rankings and ratings given by the judges. This method aims to find a *combined aggregate rating* (CAT) that is as close as possible to both the given rankings and the given ratings. The CAT balances between the cardinal and ordinal evaluations, and thus represents better the judges' opinions as compared to a consensus that uses only the given ratings or only the given rankings.

A complete rating $\boldsymbol{a}$ of a set of objects implies a complete ranking $\mathsf{rank}(\boldsymbol{a})$ on the objects by sorting the objects according to their scores. Therefore we can obtain a combined aggregate rating $\boldsymbol{x}^{(dc)}$ and its implied ranking, $\boldsymbol{x}^{(do)}$, referred to as *combined aggregate ranking* (CAK), that are representative of both the judges' ratings and the judges' rankings as follows. Given a set of ratings $\left\{\boldsymbol{a}^k\right\}_{k=1}^K$ and a set of rankings $\left\{\boldsymbol{b}^k\right\}_{k=1}^K$, a combined aggregate rating is an optimal solution to problem 5.3.

$$\min_{\boldsymbol{x}^{(dc)}} \quad \sum_{k=1}^{K} d_{NPCK}\left(\boldsymbol{a}^k, \boldsymbol{x}^{(dc)}\right)+\sum_{k=1}^{K} d_{NPKS}\left(\boldsymbol{b}^k, \boldsymbol{x}^{(do)}\right) \tag{5.3}$$

Problem (5.3) is a natural combination of the rating aggregation problem (2.16) and the ranking aggregation problem (2.7). The fact that both d_{NPCK} and d_{NPKS} are normalized distances, guarantees that both ratings and rankings are of equal importance in the optimization problem (5.3).

Since problem (2.7) is NP-hard and a special case of problem (5.3), then problem (5.3) is NP-hard. The (nonlinear, nonconvex) mathematical programming formulation of problem

(5.3) is

$$
\begin{aligned}
\min_{\boldsymbol{x}^{(dc)},\,\boldsymbol{z}} \quad & \sum_{k=1}^{K} \mathcal{C}_k \sum_{i\in\mathcal{A}^k} \sum_{j\in\mathcal{A}^k} \left| z_{ij} - p_{ij}^k \right| + \\
& \sum_{k=1}^{K} \mathcal{D}_k \sum_{i\in\mathcal{B}^k} \sum_{j\in\mathcal{B}^k} \frac{1}{2} \left| \operatorname{sign}(z_{ij}) - \operatorname{sign}(b_j^k - b_i^k) \right| && (5.4a)\\
\text{subject to} \quad & z_{ij} = x_i^{(cat)} - x_j^{(cat)} \quad i = 1,\dots,n;\ \ j = 1,\dots,n && (5.4b)\\
& \ell \le x_i^{(cat)} \le u \quad i = 1,\dots,n && (5.4c)\\
& x_i^{(cat)} \in \mathbb{Z} \quad i = 1,\dots,n. && (5.4d)
\end{aligned}
$$

We approximate the solution to problem (5.4) by applying the convexification strategy defined on Section 2.3.4 to problem (5.4) resulting in the formulation:

$$
\begin{aligned}
\min_{\boldsymbol{x}^{(dc)},\,\boldsymbol{z}} \quad & \sum_{k=1}^{K} \mathcal{C}_k \sum_{i\in\mathcal{A}^k} \sum_{j\in\mathcal{A}^k} \left| z_{ij} - p_{ij}^k \right| + \sum_{k=1}^{K} \mathcal{D}_k \sum_{i\in\mathcal{B}^k} \sum_{j\in\mathcal{B}^k} h_{ij}^k(z_{ij}) && (5.5a)\\
\text{subject to} \quad & z_{ij} = x_i^{(cat)} - x_j^{(cat)} \quad i = 1,\dots,n;\ \ j = 1,\dots,n && (5.5b)\\
& \ell \le x_i^{(cat)} \le u \quad i = 1,\dots,n && (5.5c)\\
& x_i^{(cat)} \in \mathbb{Z} \quad i = 1,\dots,n && (5.5d)\\
\text{where,} \quad & h_{ij}^k(z_{ij}) = \begin{cases} \max\left\{0, \frac{z_{ij}+1}{2}\right\} & \text{if } \operatorname{sign}(b_j^k - b_i^k) = -1 \\ \max\left\{\frac{-z_{ij}}{2}, \frac{z_{ij}}{2}\right\} & \text{if } \operatorname{sign}(b_j^k - b_i^k) = 0 \\ \max\left\{\frac{1-z_{ij}}{2}, 0\right\} & \text{if } \operatorname{sign}(b_j^k - b_i^k) = 1. \end{cases} && (5.5e)
\end{aligned}
$$

Problem (5.5) is a special case of the SD model and thus solvable in polynomial time.

Remark: Note that in equations (5.4a) and (5.5e), the argument of the sign function is $b_j^k - b_i^k$ and not $b_i^k - b_i^k$ as in equations (2.8a) and (2.10). This is because of the classical convention that in the given ratings a high cardinal number is assigned to the most preferred objects; while in the given rankings a high ordinal number is assigned least preferred objects.

5.5 Results

In this section we use the framework presented in Section 5.4.1 to rank the contestants of the 2007 MSOM's SPC. We also compare these results to those obtained by aggregating only the cardinal evaluations, and those obtained by aggregating only the ordinal evaluations.

Table 5.5 gives the consensus rating (optimal solution to problem (2.16)) $\boldsymbol{x}^{(c)}$; the (approximate) consensus ranking (optimal solution to problem (2.11)) $\boldsymbol{x}^{(o)}$; and, the combined aggregate rating $\boldsymbol{x}^{(dc)}$ and ranking $\boldsymbol{x}^{(do)}$ (optimal solutions to problem (5.5)).

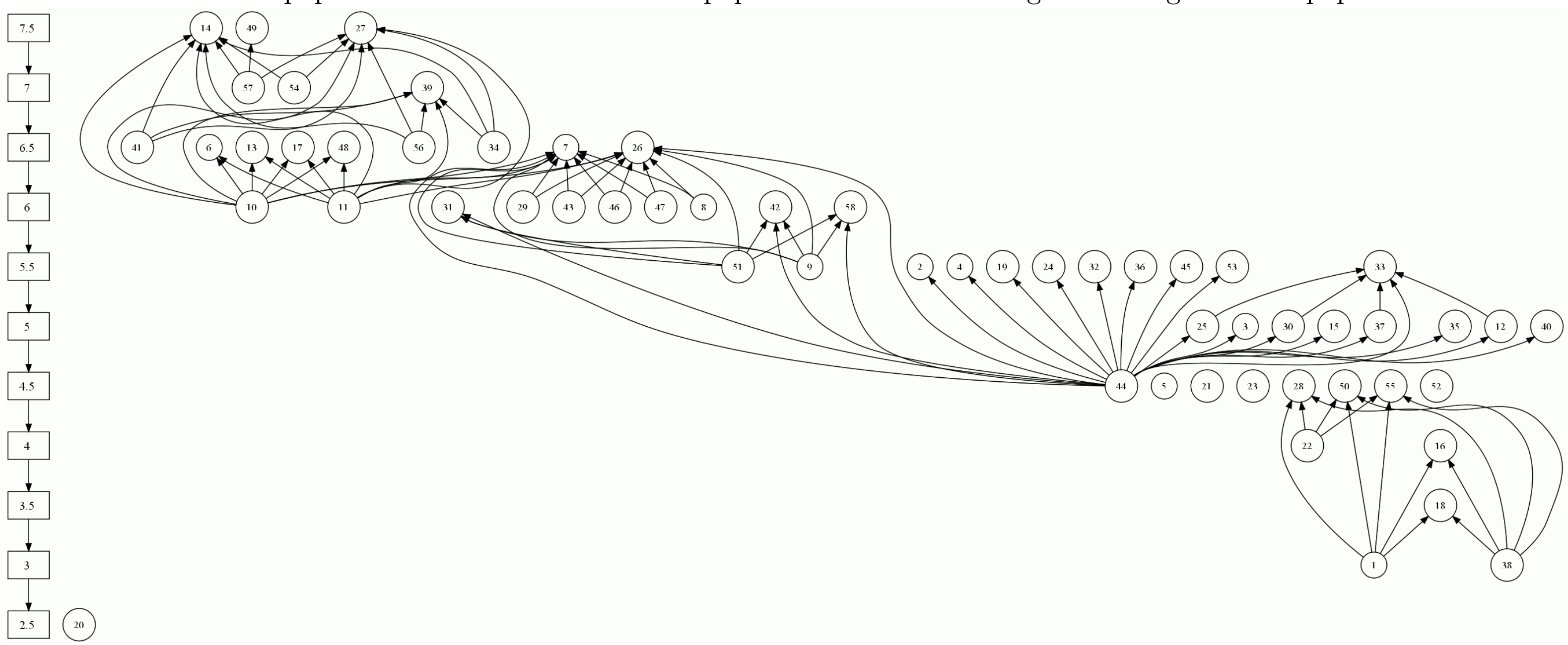
Table 5.5: Aggregate ratings and rankings for the 2007 MSOM's SPC.

Paper	$\boldsymbol{x}^{(c)}$	$\boldsymbol{x}^{(o)}$	$\boldsymbol{x}^{(dc)}$	$\boldsymbol{x}^{(do)}$	**Paper**	$\boldsymbol{x}^{(c)}$	$\boldsymbol{x}^{(o)}$	$\boldsymbol{x}^{(dc)}$	$\boldsymbol{x}^{(do)}$
1	3	41	4	41	30	5	24	5	23
2	5.5	24	5	23	31	6	24	5	23
3	5	41	4	41	32	5.5	24	5	23
4	5.5	24	5	23	33	5.5	41	4	41
5	4.5	41	4	41	34	6.5	3	7	2
6	6.5	9	6	8	35	5	41	4	41
7	6.5	24	5	23	36	5.5	24	5	23
8	6	9	6	8	37	5	24	5	23
9	5.5	9	6	8	38	3	41	4	41
10	6	3	7	2	39	7	9	6	8
11	6	3	7	2	40	5	41	4	41
12	5	24	5	23	41	6.5	3	6	8
13	6.5	9	6	8	42	6	24	5	23
14	7.5	9	6	8	43	6	9	6	8
15	5	41	4	41	44	4.5	9	5	23
16	4	53	3	53	45	5.5	24	5	23
17	6.5	9	6	8	46	6	9	6	8
18	3.5	53	3	53	47	6	9	6	8
19	5.5	24	5	23	48	6.5	9	6	8
20	2.5	53	2	58	49	7.5	2	7	2
21	4.5	41	4	41	50	4.5	53	3	53
22	4	41	4	41	51	5.5	9	6	8
23	4.5	41	4	41	52	4.5	41	4	41
24	5.5	24	5	23	53	5.5	24	5	23
25	5	24	5	23	54	7	3	7	2
26	6.5	24	5	23	55	4.5	53	3	53
27	7.5	9	6	8	56	6.5	3	7	2
28	4.5	53	3	53	57	7	1	8	1
29	6	9	6	8	58	6	24	5	23

In Table 5.5, the consensus rating $\boldsymbol{x}^{(c)}$ is non-integral because some of the judges assigned fractional scores (in particular they assigned grades that are multiple of 1/2). To appropriately handle the judges' fractional grades, we decided to set the 'grading unit' to 1/2. From an optimization point of view, this represents no problem, since the separation-deviation problem can be solved in any pre-specified precision. (This is true since to solve the separation-deviation problem, we use the algorithm developed in [1] to solve the convex dual of the minimum cost network flow problem (CDMCNF), and this algorithm solves CDMCNF in any pre-specified precision.)

As described at the beginning of Section 5.4.1, in order highlight the discrepancies (and similarities) between the consensus rating $\boldsymbol{x}^{(c)}$ and the consensus ranking $\boldsymbol{x}^{(o)}$, we can represent them graphically (see Figure 5.1).

Figure 5.1: The papers (circled) are ordered (top to bottom) in decreasing consensus score. There is an arc between two papers whenever the lower rated paper has a better ranking than a higher rated paper.



We start by highlighting the similarities between the consensus rating $\boldsymbol{x}^{(c)}$ and the consensus ranking $\boldsymbol{x}^{(o)}$. From Figure 5.1 we observe the following: (1) Paper 20 (lower left corner of Figure 5.1) should have the lowest consensus evaluation. (2) Although the agreement between $\boldsymbol{x}^{(c)}$ and $\boldsymbol{x}^{(o)}$ is not perfect, there are subsets of papers should receive a lower (or higher) collective evaluation than others. For example, the papers $\{1, 38, 18, 22, 16, 28, 50, 55\}$ should receive a collective evaluation higher than that of paper 20, lower or equal to the collective evaluation of papers $\{5, 21, 23, 52\}$ and lower than the rest of the papers.

Next we give a specific example of the conflicts described at the beginning of Section 5.4.1. In particular, from Figure 5.1 we note that paper 14 has the highest consensus score, however this conflicts with several papers (e.g., paper 54) that have a lower consensus score but a higher consensus rank. The evaluations received by papers 14 and 54 are given in Table 5.6. The number of papers reviewed by each judge and the average Field Contribution (FC) they gave are given in Table 5.7. The adjusted FC, obtained by dividing the paper's FC by the judge's average FC, is given in Table 5.8. From these Tables we observe the following:

1. The ordinal evaluations of paper 54 seam better than those of paper 14. This explains in part (The consensus ranking also depends on the rankings of the papers to which they were compared.) why paper 14 has a higher consensus rank than paper 54.
2. The average FC of paper 14 (5.6) is only slightly bigger than that of paper 54 (5.5). This explains in part (The consensus rating also depends on the ratings of the papers to which they were compared.) why paper 14 has a better consensus score than paper 54.
3. It seems that judge 44, who evaluated paper 14, was remarkably lenient, while judge 30, who evaluated paper 14, was remarkably strict. This suggests that the FC of '5' given by these two judges is not comparable. Note that the adjusted FC of paper 14-judge 44 is of 0.71; while the adjusted FC of paper 54-judge 30 is of 1.39. Moreover, the average adjusted FC of paper 14 and 54 are 1.10 and 1.28, respectively.
4. All of this suggests that paper 54 deserves a collective evaluation better than that of paper 14.

In the CAT and CAK, $\boldsymbol{x}^{(dc)}$ and $\boldsymbol{x}^{(do)}$, paper 54 is rated (ranked) higher than paper 14; this, as discussed previously, seems appropriate. This is evidence that the CAT and CAK represent better the judges' evaluations/opinions than the consensus rating $\boldsymbol{x}^{(c)}$, which takes into consideration only the ratings provided by the judges.

Table 5.6: Evaluations of papers 14 and 54.

Paper	Judge	Field Contribution Score	Paper Ranking
14	35	6	1
14	23	6	1
14	48	7	1
14	57	4	4
14	44	5	4
54	30	5	1
54	32	4	4
54	25	6	1
54	22	7	1

Table 5.7: Evaluation statistics of the judges that evaluated papers 14 and 54.

Judge	Number of Papers Evaluated	Average Field Contribution
35	4	4.50
23	4	4.25
48	4	5.25
57	4	5.75
44	5	7.00
30	5	3.60
32	4	5.25
25	5	4.00
22	4	4.75

Table 5.8: Adjusted Field Contribution received by papers 14 and 54.

Paper	Judge	Adjusted Field C.
14	35	1.33
14	23	1.41
14	48	1.33
14	57	0.70
14	44	0.71
54	30	1.39
54	32	0.76
54	25	1.50
54	22	1.47

In the 2007 MSOM's SPC, papers 38, 14, 10, 1 and 42 had the highest contributions to the separation penalty. As noted previously, this indicates that these papers are those whose evaluations are not consistent/deserve further discussion. For example, paper 38—a very low rated paper in the consensus rating—received scores from 2 to 5 and was ranked by all but one of the judges as their least preferred paper (see Tables 5.9 and 5.10). In particular, paper 38 was the second most preferred paper of judge 9; perhaps because this judge received other papers with less quality than paper 38? We believe this is not the case since, as shown in Table 5.11, the paper ranked last by judge 9 was paper 10. As noted above, paper 10 is also among the highest contributors to the separation penalty. Paper 10 received three high evaluations and 2 very low evaluations (see Table 5.12). Therefore, we believe that, in order to get a better consensus, the scores/ranks of paper 38 and paper 10 should be discussed by the judges assigned to these two papers.

Table 5.9: Evaluations of paper 38.

Judge	Field Contribution Score	Paper Ranking
30	3	5
41	2	5
44	3	5
9	5	2
20	5	4

Table 5.10: Evaluation statistics of the judges that evaluated paper 38.

Judge	Number of Papers Evaluated	Average Field Contribution
30	5	3.60
41	5	5.00
44	5	7.00
9	5	4.60
20	4	7.25

Table 5.11: Evaluations of judge 9.

Paper	Field Contribution Score	Paper Ranking
10	3	5
19	4	3
38	5	2
50	4	3
58	7	1

Table 5.12: Evaluations of paper 10.

Judge	Field Contribution Score	Paper Ranking
33	7	1
41	7	1
19	2	3
15	6	1
9	3	5

Next we analyze the combined aggregate rating $\boldsymbol{x}^{(dc)}$ and ranking $\boldsymbol{x}^{(do)}$ (solution to problem (5.5)). We make the following observations:

1. The consensus rating, $\boldsymbol{x}^{(c)}$, has a total rating distance (equation (2.16)) of 7.3611.
2. The consensus ranking, $\boldsymbol{x}^{(o)}$, has a total ranking distance (equation (2.7)) of 13.8500.
3. (a) The combined aggregate rating, $\boldsymbol{x}^{(dc)}$, has a total rating distance (equation (2.16)) of 8.16667.
   (b) The combined aggregate ranking, $\boldsymbol{x}^{(do)}$, (derived from $\boldsymbol{x}^{(dc)}$), has a total ranking distance (equation (2.7)) of 13.9333.

This shows that, in this case, the combined aggregate rating and ranking achieve a very good compromise. In particular, they remain almost as close as the consensus rating to the judges' ratings, and remain almost as close as the consensus ranking to the judges' rankings.

5.6 Concluding remarks

We develop here a new framework for group decision making that aggregates both cardinal and ordinal input evaluations. Our framework consists on finding the *combined aggregate rating* (CAT) and its implied *combined aggregate ranking* (CAK) that minimize the sum of the distances from the CAT to the given ratings plus the sum of the distances from the CAK to the given rankings.

We illustrate the effectiveness of the new framework by ranking the contestants of the 2007 MSOM's student paper competition. We provide evidence that obtaining a combined aggregate rating that aggregates both cardinal and ordinal evaluations better represents the judges' opinions as compared to a rating that aggregates only the judges' ratings.

Aggregating incomplete evaluations is challenging because of the "bias phenomenon" of incomplete evaluations: the collective evaluation is prone to be biased by the judges' subjective scales because an object assigned to a particularly strict (lenient) judge has an advantage (disadvantage) compared to those objects not assigned to this specific judge. Our framework, applicable when the given ratings and rankings are incomplete, addresses the bias phenomenon by:

1. detecting inconsistencies between the consensus rating (which aggregates only the given ratings) and the consensus ranking (which aggregates only the given rankings), and
2. identifying papers with high separation penalty.

This information is helpful so that (say) the manager of the competition can initiate an investigation of the nature of the conflicts and act accordingly (for example, by discussing these inconsistencies with the judges that evaluated these papers).

The problem of aggregating complete evaluations (in which all judges evaluate all objects) is a special case of the problem of aggregating incomplete evaluations (in which the judges are allowed to evaluate only some of the objects). Therefore our framework is also applicable to aggregating complete evaluations.

We believe that the new framework is valuable also in cases where the input is in the form of cardinal ratings only. In order to apply the framework in this case, we use as input both the given ratings and the rankings implied by the ratings.

Bibliography

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

[2] M. B. Alexander, A. M. Bronstein, R. Kimmel, and I. Yavneh. A multigrid approach for multi-dimensional scaling. In *in Proc. Copper Mountain Conf. Multigrid Methods*, pages 1–2, 2005.

[3] S. Alexe, P. L. Hammer, A. Kogan, and M. A. Lejeune. A non-recursive regression model for country credit risk rating. RUTCOR Research Report, RRR 9-2003, Rutgers University, Piscataway, NJ, 2003.

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

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

[6] J. Bartholdi, C. A. Tovey, and M. A. Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6:157–165, April 1989.

[7] H. Baumgartner and J. E. M. Steenkamp. Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38:143–156, 2001.

[8] A. E. Bindschedler and J. M. Moore. Optimal location of new machines in existing plant layouts. *Journal of Industrial Enginering*, 12:41–48, 1961.

[9] K. P. Bogart. Preferences structures I: Distances between transitive preference relations. *Journal of Mathematical Sociology*, 3:49–67, 1973.

[10] K. P. Bogart. Preference structures II: Distances between asymmetric relations. *SIAM Journal on Applied Mathematics*, 29(2):254–262, 1975.

[11] J. P. Brans, B. Roy, and P. Vincke. Aide a la decision multicritere. *Rev. Beige Statisti., d'Informatique et de Recherche Operationnelle*, 15, 1975.

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

[13] M. J. Brusco. Integer programming methods for seriation adn unidimensional scaling of proximity matrices: a review and some extensions. *Journal of Classification*, 19:45–67, 2002.

[14] M. J. Brusco and S. Stahl. Bicriterion seriation methods for skew-symmetric matrices. *British Journal of Mathematical and Statistical Psychology*, 58:333–343, 2005.

[15] M. J. Brusco and S. Stahl. *Branch-and-Bound appolications in cominatorial data analysis*. Springer, 2005.

[16] M. J. Brusco and S. Stahl. Optimal least-squares unidimensional scaling: Improved branch-and-bound procedures and comparison to dynamic programming. *Psychometrika*, 70:253–270, 2005.

[17] W. G. Cochran and G. M. Cox. *Experimental design*. Wiley, New York, 1957.

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

[19] W. D. Cook, B. Golany, M. Penn, and T. Raviv. Creating a consensus ranking of proposals from reviewers' partial ordinal rankings. *Computers and Operations Research*, 34:954–965, 2007.

[20] W. D. Cook and M. Kress. Ordinal ranking with intensity of preference. *Management Science*, 31:26–32, 1985.

[21] W. D. Cook, M. Kress, and L. M. Seiford. Information and preference in partial orders: a bimatrix represetation. *Psychometrika*, 51:197–207, 1986.

[22] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. The MIT Press, 2nd edition edition, September 2001.

[23] J. de Leeuw. Applications of convex analysis to multidimensional scaling. In G. Romier J. R. Barra, F. Brdeau and B. van Cutsem, editors, *In: Recent developments in statistics*, pages 133–146. North Holland Publishing, 1977.

[24] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th international conference on World Wide Web*, pages 613–622, New York, NY, USA, 2001. ACM.

[25] G. Even, J. Naor, B. Schieber, and M. Sudan. Approximating minimum feedback sets and multi-cuts in directed graphs. In *Proceedings 4th International Conference on Integer Programming and Combinatorial Optimization*, pages 14–48. Lecture Notes in Computer Science 920, Springer-Verlag, 1995.

[26] G. Ferri, L-G. Liu, and J. Stiglitz. The procyclical role of rating agencies: evidence from the east asian crisis. *Economic Notes*, 28(3):335–355, 1999.

[27] R. A. Fisher. The systematic location of genes by means of crossover observations. *American Naturalist*, 56:406–411, 1922.

[28] R. L. Francis. A note on the optimum location of new machines in existing facilities. *Journal of Industrial Enginering*, 15:106–107, 1963.

[29] S. French. *Decision Theory: An Introduction to the Mathematics of Rationality.* Halsted Press, New York, 1988.

[30] M. R. Garey and D. S. Johnson. *Computers and Intractability: A guide to the theory of NP-completeness.* W. H. Freeman and co., New York, 1979.

[31] Y. Ge, Y. Leung, and J. Ma. Unidimensional scaling classifier and its application to remotely sensed data. *Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International*, 6:3841–3844, july 2005.

[32] J. Gibbons. *Nonparametric statistical inference.* McGraw-Hill, New York, 1971.

[33] P. J. F Groenen, W. J. Heiser, and J. J. Meulman. Global optimization in least-squares multidimensional scaling by distance smoothing. *Journal of Classification*, 16:225–254, 1999.

[34] P. L. Hammer, A. Kogan, and M. A. Lejeune. Country risk ratings: Statistical and combinatorial non-recursive models. RUTCOR Research Report, RRR 8-2004, Rutgers University, Piscataway, NJ, March 2004.

[35] N. U. Haque, M. S. Kumar, N. Mark, and D. Mathieson. The economic content of indicators of developing country creditworthiness. IMF Working Paper No. 96/9 (1996). Available at: `http://ssrn.com/abstract=882910`.

[36] A. W. K. Harzing. Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management*, 6:243–266, 2006.

[37] D. S. Hochbaum. 50th anniversary article: Selection, provisioning, shared fixed costs, maximum closure, and implications on algorithmic methods today. *Management Science*, 50:709–723, 2004.

[38] D. S. Hochbaum. Ranking sports teams and the inverse equal paths problem. In Paul G. Spirakis, Marios Mavronicolas, and Spyros C. Kontogiannis, editors, *Internet and Network Economics, Second International Workshop, WINE*, volume 4286 of *Lecture Notes in Computer Science*, pages 307–318, Greece, 2006. Springer.

[39] D. S. Hochbaum. New optimization techniques in data mining. *Manuscript UC Berkeley*, 2007.

[40] D. S. Hochbaum and A. Levin. The k-allocation problem and its variants. In *WAOA*, 2006.

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

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

[43] D. S. Hochbaum, E. Moreno-Centeno, R. A. Catena, and P. Yelland. A new model for behavioral customer segmentation. UC Berkeley Manuscript (2008). Available at: `http://riot.ieor.berkeley.edu/~dorit/pub/customer-segment.pdf`.

[44] D. S. Hochbaum and J. G. Shanthikumar. Convex separable optimization is not much harder than linear optimization. *Journal of the Association for Computing Machinery*, 37:843–862, 1990.

[45] L. Hubert, P. Arabie, and J. Meulman. *Combinatorial data analysis: Optimization by dynamic programming*. SIAM, 2001.

[46] L. Hubert, P. Arabie, and J. Meulman. Linear unidimensional scaling in the L_2 norm: basic optimization methods using matlab. *Journal of Classification*, 19:303–328, 2002.

[47] R. M. Karp and E. Moreno-Centeno. Implicit hitting set problems. in preparation.

[48] R. L. Keeney. Group preference axiomatization with cardinal utility. *Management Science*, 23(2):140–145, 1976.

[49] J. G. Kemeny and L. J. Snell. Preference ranking: An axiomatic approach. In *Mathematical Models in Social Science*, pages 9–23. Ginn, Boston, 1962. For a book chapter with author, and book editor use incollection instead of inbook.

[50] K. Lau, P. L. Leung, and K. Tse. A nonlinear programming approach to metrix unidimensional scaling. *Journal of Classification*, 15:3–14, 1998.

[51] V. Mahajan and E. Muller. When is it worthwhile targeting the majority instead of the innovators in a new product launch. *Journal of Marketing Research*, 35(4):488–495, Nov 1998.

[52] E. Moreno-Centeno. Axiomatic metrics for aggregating incomplete evaluations. In preparation.

[53] W. S. Robinson. A method for chronologically ordering archaeological deposits. *American Antiquity*, 16:293–301, 1951.

[54] E. M. Rogers. *Diffusion of Innovations*. Free Press, 5 edition, 1962.

[55] P. Rusmevichientong, S. Zhu, and D. Selinger. Identifying early buyers from purchase data. In Won Kim, Ron Kohavi, Johannes Gehrke, and William DuMouchel, editors, *KDD*, pages 671–677. ACM, 2004.

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

[57] P. B. Smith. Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, 35:50–61, 2004.

[58] K. Talluri, G. van Ryzin, and B. L. Bayus. An analysis of product lifetimes in a technologically dynamic industry. *Management Science*, 44:763–775, 1998.

[59] F. Yates. Incomplete randomized blocks. *Annals of Eugenics*, 7:121–140, 1936.