Journal of Business Research 67 (2014) 998–1006

Contents lists available at ScienceDirect

Journal of Business Research

How to derive consensus among various marketing journal rankings?☆

Stefan Theußl *, Thomas Reutterer [1,2], Kurt Hornik [3]

WU Vienna University of Economics and Business, Vienna, Austria

ARTICLE INFO

Article history:
Received 16 November 2010
Received in revised form 7 October 2011
Accepted 14 September 2012
Available online 20 September 2013

Keywords:
Journal ranking
Consensus ranking
Meta-ranking
Marketing
Binary optimization

ABSTRACT

Despite the increasing popularity of journal rankings to evaluate the quality of research contributions, the individual rankings for journals that ranked below the top tier of publications usually feature only modest agreement. Attempts to merge rankings into meta-rankings suffer from some methodological issues, such as mixed measurement scales and incomplete data. This paper addresses the issue of how to construct suitable aggregates of individual journal rankings, using an optimization-based consensus ranking approach. The authors apply the proposed method to a subset of marketing-related journals from a list of collected journal rankings. Next, the paper studies the stability of the derived consensus solution, and the degeneration effects that occur when excluding journals and/or rankings. Finally, the authors investigate the similarities/dissimilarities of the consensus with a naive meta-ranking and with individual rankings. The results show that, even though journals are not uniformly ranked, one may derive a consensus ranking with considerably high agreement with the individual rankings.

1. Introduction

Academics in the field of business research and related sub-disciplines encounter an increasing number of journal rankings (see, e.g., *Anne-Wil Harzing* via the web site http://www.harzing.com/jql.htm, for an up-to-date compilation). The majority of these journal rankings are based on stated preferences (i.e., judgments among academic peers; e.g., Fry, Walters, & Scheurmann, 1985; Hult, Neese, & Bashaw, 1997; Schrader & Hennig-Thurau, 2009; Theoharakis & Hirst, 2002), revealed preferences (i.e., citation rates as a surrogate measure of publication impact; e.g., Bakir, Vitell, & Rose, 2000; Baumgartner & Pieters, 2003), or a combination of the two (e.g., Azar & Brock, 2008; Dubois & Reeb, 2000; Zhou, Ma, & Turban, 2001). While the various ranking approaches are typically fairly consistent in ranking top-tier journals in the investigated (sub-) discipline, they tend to diverge substantially as one proceeds further down the ratings. Besides the kind of data used to rank journals, the individual rankings for lower-tier journals are also affected by the type of institution (e.g., academic rigor vs. practitioner orientation, type of methodological research orientation, etc.) and the geographical perspectives of different groups of academics (Pieters, Baumgartner, Vermunt, & Bijmolt, 1999; Polonsky & Whitelaw, 2005; Tellis, Chandy, & Ackerman, 1999; Theoharakis & Hirst, 2002). In a comparative study of eleven rankings, Mingers and Harzing (2007) found pairwise rank correlations between 0.32 and 0.79, which clearly suggests that some degree of consensus exists, but by no means complete agreement.

Despite an ongoing controversial discussion of their specific merits and drawbacks (see Polonsky, 2008, for a recent summary), journal rankings have undoubtedly gained momentum in the evaluation of individual scholars' or academic units' research quality. For instance, the UK's higher education funding councils have adopted journal rankings as their central metrics for evaluating the research quality of universities in their regularly undertaken *Research Assessment Exercise* (*RAE*). The widespread acceptance of the journal rankings published by the German Academic Association for Business Research (*VHB-JOURQUAL*) for assessing the research performance of business scholars in German-speaking countries is just one further example (Schrader & Hennig-Thurau, 2009).

As a natural consequence of the still increasing number of available journal rankings and their growing importance to the academic community, some authors have attempted to merge compilations of single rankings into meta-rankings for various sub-disciplines such as international management (Dubois & Reeb, 2000), management information systems (Rainer & Miller, 2005), and innovation management and entrepreneurship (Franke & Schreier, 2008). The approaches to aggregation range from simple rank averaging (e.g., in Dubois & Reeb, 2000) to more sophisticated statistical

☆ The authors thank the Guest Editor Udo Wagner and two anonymous reviewers for their valuable comments on previous versions of the manuscript. They are also grateful to Emily Stapleton for proofreading the article.

* Corresponding author at: Institute for Statistics and Mathematics, WU Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Wien, Austria. Tel.: +43 1 31336 5057.

E-mail addresses: stefan.theussl@wu.ac.at (S. Theußl), thomas.reutterer@wu.ac.at (T. Reutterer), kurt.hornik@wu.ac.at (K. Hornik).

[1] Department of Marketing, WU Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Wien, Austria. Tel.: +43 1 31336 4619.

[2] Surrey Business School, University of Surrey, Guildford, GU2 7XH, United Kingdom.

[3] Institute for Statistics and Mathematics, WU Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Wien, Austria. Tel.: +43 1 31336 4756.

0148-2963/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jbusres.2013.08.006

approaches such as the maximum likelihood procedure proposed by Bancroft, Gopinath, Kovács, and Rejtö (1999). The potential benefits of such meta-rankings are obvious: Firstly, they tend to balance out any biases in the single journal rankings due to raters' strategic answering behavior (often encountered with survey-based rankings) or unsystematic errors such as volatilities in the impact factors (in the case of citation-based approaches). Secondly, they provide a better guide to academics in globalized publication and job markets because they reflect an aggregate perspective of the academic community on the quality of publications (Franke & Schreier, 2008; Rainer & Miller, 2005).

However, two major challenges are inherent in aggregating journal ranking data sets, neither of which has yet been adequately resolved by the existing approaches: (1) the different measurement scales used by the rankings and (2) incomplete information (Franke & Schreier, 2008; Mingers & Harzing, 2007; Schrader & Hennig-Thurau, 2009). The first issue refers to the fact that individual rankings make use of quite different scales, including binary (yes/no), ordinal (e.g., grades A+, A, B, etc.), and numeric scores (e.g., impact factors), to construct their rankings. This issue makes the aggregation of rankings using conventional statistical methods cumbersome. The second issue, which is related to the typically large number of *missing observations* in ranking data sets, appears to be even more problematic. Such missing information accrues because the various rankings cover only subsets of journals, which usually coincide only partially. Thus, the sparsity of data sets generally increases when including a broader set of rankings.

This paper presents a method for deriving meta-rankings of journals by solving consensus optimization problems. The proposed methodology obtains consensus rankings from paired comparisons among a set of individual rankings, can accommodate mixed types of measurement scales and is relatively robust even when used for sparse data. The next section addresses the challenges involved in suitably aggregating rankings to derive a consensus ranking, outlines the cornerstones of the proposed consensus ranking methodology, and provides pointers for the computational implementation of the related procedures. The paper proceeds by introducing the journal ranking data that the study uses. The following section shows the results of applying the method, using a subset of marketing-related journals, which the *Harzing Journal Ranking Repository* provided. The presentation and discussion of the results show that, even though the investigated rankings of marketing-related journals are far from being identical, the proposed method can produce a consensus ranking which shows a considerably high level of agreement with the individual rankings. This paper also explores the sensitivity of the consensus ranking methodology to variations in the number of rankings and/or journals, by studying stability and degeneration issues. In addition, this study includes a comparison of the consensus ranking methodology to a naive ranking, derived through a simple averaging of ranks, which demonstrates the benefits of the consensus ranking method. Finally, the paper concludes with a discussion on the value of consensus rankings and by outlining some points for future research.

2. Methodology

As a starting point of the problem of finding adequate meta-rankings of various journal rankings, consider a total of B journal rankings $\mathfrak{R} = \{R_1, \ldots, R_B\}$ and n journals $\mathfrak{J} = \{J_1, \ldots, J_n\}$. Table 1 shows an example ranking data set for $B = 4$ and $n = 3$ for illustration purposes. Three of the four rankings use an ordinal scale, with levels A+, A, B, C, and D (rankings R_1 and R_2), and scores of 4 to 1 (ranking R_3; ordered from highest to lowest as defined in the ranking description). Ranking R_4 uses a metric scale of 0 to 100, where a higher value indicates better journal quality.

Table 1
Example ranking data set.

	R_1	R_2	R_3	R_4
J_1	A+	D	2	33.50
J_2	A	B	3	
J_3	A	B	3	23.30

In the days of the French revolution, Borda (1781) (who computed the sum of individual rankings) and Condorcet (1785) (who used the majorities of paired comparisons) proposed the first heuristic aggregation rules for constructing meta-rankings; see, for example, Cook (2006) for a recent survey of consensus models of ordinal preference rankings. Leaving ranking R_4 aside so as to avoid dealing with the missing observation for J_2, and applying the Borda aggregation rule, gives an aggregate evaluation that all three journals are of the same quality, since the sum of ranks is 5 in all cases ($1 + 3 + 1$ for J_1 and $2 + 1 + 2$ for J_2 and J_3).

As ordinal rankings are typically not Likert-type (interval) scales and may have different granularities, treating grades like A+ and A as rank scores (i.e., 1, 2, etc.) is quite problematic, and the above aggregation rule may lead to results which do not necessarily reflect the quality assessments behind the *majority* of rankings. In the example above, at least two of the rankings put J_1 above J_2 and J_3 and only one ranks J_1 below the other two journals. However, the consensus shows indifference between all three journals.

An approach based on pairwise comparisons of journals appears to be more adequate. Comparing journals within each applicable ranking by investigating whether a ranking evaluates a journal J_i as explicitly superior to another journal J_j implies the relationship $J_i > J_j$. The results of these paired comparisons for a particular journal ranking R_b can be represented by an incidence matrix $I_b = [\chi_{ij}(R_b)]$, where the entries are the incidences $\chi_{ij}(R_b) = 1$ if $J_i \leq J_j$ or zero otherwise. This encoding warrants that the existence of ties (i.e., equivalent preferences) is taken into account. Using such an approach, the interpretation of all available data as metric is not necessary (such as, e.g., Mingers & Harzing, 2007). Thus, the direct processing of ordinal data that journal ranking publishing institutions commonly provide becomes feasible. Furthermore, if journal J_i is not included in ranking R_b, then necessarily $\chi_{ij}(R_b) = \chi_{ji}(R_b) = 0$ for all $j \neq i$. This way of encoding only manifest *preferences* allows the researcher to deal with missing paired comparison data. The incidence matrices corresponding with the rankings from Table 1 are shown in Table 2.

The incidence matrix I_1 for ranking R_1 encodes the results of the following paired comparisons: $J_1 \leq J_1$, $J_1 > J_2$, $J_1 > J_3$; neither J_2 nor J_3 is ranked higher than the other, that is $J_2 \sim J_3$ (where ~denotes indifference between two journals). In contrast, the incidences for R_2 imply that $J_2 > J_1$ and $J_3 > J_1$. For R_3, the incidences are identical to the ones derived from R_1. I_4 shows that $J_1 > J_3$ but no other comparisons exist.

By encoding the paired comparisons in this way, one can employ the following geometric idea: if one can measure the distance between rankings, then the *center* of the given collection of rankings, which minimizes the total distance to all rankings, is a natural and intuitively appealing choice for consensus (see Régnier, 1965). One can derive the distance between two rankings by counting the number of discordant entries in the corresponding incidence

Table 2
Incidence matrices for rankings R_1, R_2, R_3, and R_4. A non-zero entry denotes a $\leq$ relationship between two journals.

	I_1			I_2			I_3			I_4		
	J_1	J_2	J_3	J_1	J_2	J_3	J_1	J_2	J_3	J_1	J_2	J_3
J_1	1	0	0	1	1	1	1	0	0	1	0	0
J_2	1	1	1	0	1	1	1	1	1	0	0	0
J_3	1	1	1	0	1	1	1	1	1	1	0	1

matrices (i.e., the number of discordant paired comparison results). Formally, the distance $d(R_k,R_l)$ between R_k and R_l is equal to

$$\sum_{i,j} |\chi_{ij}(R_k) - \chi_{ij}(R_l)|.$$

For example, the distance between R_1 and R_2 is 4. Following this conceptual understanding, the complete ranking R, which minimizes $\sum_b d(R_b,R)$, is the consensus ranking. In the above example, out of all possible R, the consensus ranking is $J_1 > J_2 \sim J_3$. This ranking is equivalent to the encoded incidence matrices I_1 and I_3. Here no better solution exists; the aggregate distance between R and all of the other rankings involved is minimal.

Mathematically, complete rankings are preference relations, or *weak orders*, defined as binary relations on a set of objects which are complete, reflexive and transitive (see Fishburn, 1972, for a complete reference). Note that such weak orders permit ties (i.e., they do not enforce a strict preference, or asymmetry). For larger problems, explicitly going through all possible incidence matrices in order to find the one minimizing the aggregate distance may require rather too much effort. Following Hornik and Meyer (2007), one can obtain the consensus incidence matrix by solving the following binary program

$$\sum_{i \neq j} c_{ij} \chi_{ij}(R) \Rightarrow \max$$

where $c_{ij} = \sum_b (2\chi_{ij}(R_b) - 1)$ and the $\chi_{ij}(R)$ must be constrained such that $R \in \mathfrak{C}$, the set of consensus candidates. In the case of complete rankings, the constraints defining $\mathfrak{C}$ are:

$$\begin{array}{rclcr}
\chi_{ij}(R) & \in & \{0,1\} & i \neq j & \text{(binarity)} \\
\chi_{ii}(R) & = & 1 & & \text{(reflexivity)} \\
\chi_{ij}(R) + \chi_{ji}(R) & = & 1 & i \neq j & \text{(completeness)} \\
\chi_{ij}(R) + \chi_{jk}(R) - \chi_{ik}(R) & \leq & 1 & i \neq j \neq k & \text{(transitivity)}
\end{array}$$

(note that the constraints in Hornik & Meyer, 2007, are not for rankings). Any state-of-the-art mixed integer programming solver can efficiently solve this combinatorial optimization problem known to be computationally complex (Wakabayashi, 1998). Mathematically speaking, the above binary program determines the consensus of the (possibly incomplete) rankings $R_1,\ldots,R_B$ as the complete ranking R, which minimizes the total dissimilarity between itself and the original rankings:

$$\sum_{b=1}^{B} d(R, R_b) \Rightarrow \min_{R \in \mathfrak{C}}$$

For complete rankings, the above illustrated dissimilarity measure d is the “symmetric difference” distance, which Kemeny and Snell (1962) shows to be the “natural” distance between complete rankings, in the sense of uniquely satisfying a set of basic axiomatic conditions. Note that, for a given collection of rankings, a unique solution to the above binary optimization problem does not necessarily exist. However, using branch and cut approaches, one can identify all consensus solutions and use the commonalities in these solutions to obtain a robust understanding of the underlying preference structure.

For the applications in this paper, the authors make use of the R system (version 2.13.1) for statistical computing (R Development Core Team, 2011) to carry out all computations, and generate the problem definition, based on the supplied data set, using functions and methods available in the *relations* package (Hornik & Meyer, 2010). Furthermore, the authors use the R package, *Rcplex* (Bravo & Theußl, 2011), which provides an interface to the commercial optimizer CPLEX (IBM ILOG, 2009), to solve all optimization problems.

From the resulting consensus ranking R, one can compute a numerical rank for each object, so-called preference scores or *generalized ranks*. Preference scores are a linear transformation of the difference between lost and won paired comparisons into the range from 1 to the number of objects n (see Regenwetter & Rykhlevskaia, 2004, for details). For the application at hand, the unique solution to the corresponding optimization problem is $J_1 = 3 > J_2 \sim J_3 = 2$. Thus, I_1 and I_3 indeed represent the incidence matrix of R. Consensus rankings typically include many ties between journals (i.e., groups of journals with equal scores) and thus so-called Hasse diagrams can represent them nicely. Such diagrams denote a class of data representation techniques, which are capable to visualize weak-ordered relations among objects (available as a plot method in the R package *relations*; cf., e.g., Freese, 2004). Fig. 1 shows such a Hasse diagram for the above consensus ranking.

3. Data set characteristics

To illustrate the performance of the consensus ranking approach, the authors apply the above methodology to a set of 12 renowned rankings of journals, related to the marketing discipline. They are all available from published sources on the Web and included in the 34th edition of the Harzing Journal Quality List (http://www.harzing.com). Table 3 provides a complete list of the journal rankings used, showing publishing institution and the corresponding abbreviation.

Confronted with all 851 journals available in this data set, the authors asked domain experts to select those journals which they considered to be potential publication outlets for marketing academics' research. Following a similar procedure to that of Dubois and Reeb (2000), the experts then had to assign each of the selected journals to one of the following two categories: (1) A *core* list of marketing journals with an inherent focus on general or specific topics in marketing. (2) An *extended* list including journals from adjacent disciplines, but which also have marketing academics as their target audience. The latter list includes journals focusing on disciplines such as General Business Research, Information Science, Applied Psychology, and Operations Research.

To obtain robust interpretable consensus solutions from the integer optimization problem presented in the last section, the authors removed those journals which a significant majority (three quarters) of the rankings does not rank. This selection procedure resulted in a final sample of 33 journals in the core list and 64 in the extended list (which includes the core list). Table 4 lists all of these journals and their abbreviations; those in the core list are marked with a C.

Note that the *Journal of Business Research* is in the core list because a substantial portion of that journal's published articles are related to the marketing discipline. The ranking data set of the journals in Table 4, including descriptions and the rankings

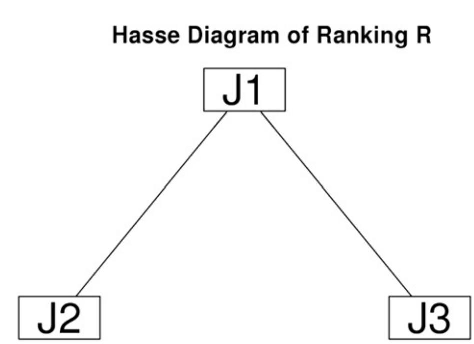

Fig. 1. Consensus journal ranking for the journals J_1, J_2, and J_3.

Table 3
Journal rankings and corresponding abbreviations.

Abbreviation	Institution
ABDC08	Australian Business Deans Council
ABS09	Association of Business Schools
Ast08	Aston Business School
Bjm04	Business and Management 2001 RAE/UK
Cnrs08	Comite National de la Recherche Scientifique
Cra09	Cranfield University School of Management
EJL06	Erasmus Research Institute of Management
Hkb05	Hong Kong Baptist University School of Business
Theo05	ALBA Journal Ranking
UQ07	University of Queensland
Vhb08	Verband der Hochschullehrer fuer Betriebswirtschaft
Wie01	WU Vienna University of Economics and Business

employed, is available in an online appendix, which the authors prepared for this paper, at http://statmath.wu.ac.at/projects/jcr/.

4. Consensus ranking results

This section presents, visualizes, and discusses the results of the proposed method. To illustrate the effect on the consensus ranking of adding journals from adjacent disciplines to the core list of marketing journals, the authors perform separate analyses for the core and extended journal lists. Furthermore, they investigate the sensitivity of the consensus solution to variations in the number of journals and rankings used, study potential degeneration effects, and compare the similarity/dissimilarities of the consensus to a naive meta-ranking as well as to individual rankings.

4.1. Core versus extended list

Fig. 2 portrays the Hasse diagram of the consensus rankings for the core journal list. The diagram clearly illustrates that, in the consensus across the 12 rankings, the marketing journals are arranged in several tiers. Each tier indicates the same degree of preference for the journals in that tier. For example, and indeed not very surprisingly, the top-tier marketing journals are the *Journal of Consumer Research* (JCR), the *Journal of Marketing* (JM), the *Journal of Marketing Research* (JMR), and *Marketing Science* (MkS). The second tier of the consensus preference structure also contains high-quality journals, such as the *International Journal of Research in Marketing* (IJRM), the *Journal of the Academy of Marketing Science* (JAMS), the *Journal of Product Innovation Management* (JPIM), and the *Journal of Retailing* (JR). In this respect, the findings of this research are consistent with the conclusions of many other authors that a high degree of agreement among academics as to which are the top journals in their discipline seems to exist (Polonsky & Whitelaw, 2005; Theoharakis & Hirst, 2002).

Interestingly, this consensus preference structure suggests that the *Journal of Service Research* (JSR) separates the abovementioned top-level marketing journals from lower-level publication outlets. As one proceeds further down the ladder, more specialized, niche marketing journals join the crowd. Quite obviously, the derived consensus preference structure tends to distinguish between two broad types of marketing journals: The first, in the top levels, with a relatively broad scope, covering a wide range of topics, and the second, with more focused positioning, in the lower levels of the ranking list.

Adding in journals from adjacent disciplines by using the extended journal list, one can observe a unique structure of consensus rankings for the top journals (see the Hasse diagram shown in Fig. 3). Compared to the core list of journals, the *Journal of Applied Psychology* (JAP), *Management Science* (MS), *MIS Quarterly* (MISQ), *Operations Research* (OR), and the *Strategic Management Journal* (SMJ) join the top tier of marketing-related journals. The inclusion of JAP in this group might be slightly surprising, but this journal is extremely highly ranked in the journal quality rankings published by UK and Australian institutions. The same marketing journals from the core list remain in the second tier. Now, however, some

Table 4
Journals used in this study (C indicates membership in the *core* list).

Journal name	Abbrev	Journal name	Abbrev
Advances in Consumer Research (C)	ACR	Journal of International Marketing (C)	JIM
California Management Review	CMR	Journal of Macromarketing (C)	JMK
Computers & Operations Research	COR	Journal of Marketing (C)	JM
Decision Sciences	DS	Journal of Marketing Management (C)	JMM
Decision Support Systems	DSS	Journal of Marketing Research (C)	JMR
European Journal of Information Systems	EJIS	Journal of Personal Selling and Sales Management (C)	JPSS
European Journal of Marketing (C)	EJM	Journal of Product Innovation Management (C)	JPIM
European Journal of Operational Research	EJOR	Journal of Public Policy & Marketing (C)	JPPM
European Management Journal	EMJ	Journal of Retailing and Consumer Services (C)	JRCS
Harvard Business Review	HBR	Journal of Retailing (C)	JR
Industrial Marketing Management (C)	IMM	Journal of Service Research (C)	JSR
Interfaces	Int	Journal of Services Marketing (C)	JS
Int. Business Review	IBR	Journal of Small Business Management	JSBM
Int. Journal of Advertising (C)	IJA	Journal of Strategic Marketing (C)	JSM
Int. Journal of Electronic Commerce	IJEC	Journal of the Academy of Marketing Science (C)	JAMS
Int. Journal of Logistics Management	IJLM	Journal of the Operational Research Society	JORS
Int. Journal of Market Research (C)	IJMR	Journal of World Business	JWB
Int. Journal of Research in Marketing (C)	IJRM	Management Science	MS
Int. Journal of Retail & Distrib. Man. (C)	IJRDM	Marketing Letters (C)	ML
Int. Journal of Service Industry Man. (C)	IJSIM	Marketing Science (C)	MkS
Int. Marketing Review (C)	IMR	MIS Quarterly	MISQ
Journal of Advertising (C)	JA	Operations Research	OR
Journal of Advertising Research (C)	JAR	Psychology and Marketing (C)	PM
Journal of Applied Psychology	JAP	R&D Management	RM
Journal of Business Ethics	JBE	Research Policy	RP
Journal of Business Research (C)	JBR	Service Industries Journal	SIJ
Journal of Business Venturing	JBV	Sloan Management Review	SMR
Journal of Consumer Marketing (C)	JCM	Small Business Economics	SBE
Journal of Consumer Psychology (C)	JCP	Strategic Management Journal	SMJ
Journal of Consumer Research (C)	JCR	Supply Chain Management: An International Journal	SCMAIJ
Journal of Forecasting	JOF	Thunderbird International Business Review	TIBR
Journal of Interactive Marketing (C)	JIntMar	Total Quality Management & Business Excellence	TQMBE

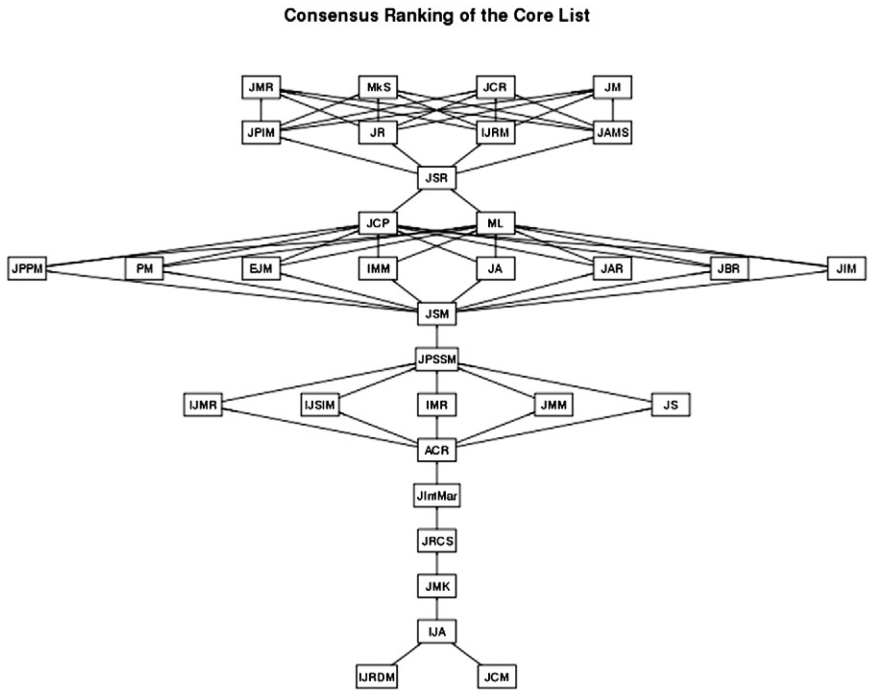

Fig. 2. Consensus journal ranking for the *core* journal list.

journals with a relatively broad disciplinary scope but with a focus on quantitative research methodologies (EJOR, DS) and information science (DSS, EJIS) join the group. Also, the rankings deem the most prestigious practitioner-oriented transfer journals in the management discipline (HBR, SMR, CMR) to be highly-esteemed publication outlets for marketing academics.

The move of ML (which the UK- and continental-European-based journal rankings involved in this study evaluate comparatively highly) from a lower tier to the second tier, and the inversion of JCP and JSR in the consensus ranking for the extended list are two of the most noticeable differences in the preference structure which emerges from extending the journal list in this paper's proposed optimization approach. In fact, the consensus preference ranks for these three journals are in roughly the same range, and are at a greater distance from the higher-ranked group of journals (top tier) and those in lower levels. Again, as already noted for the consensus ranking using the

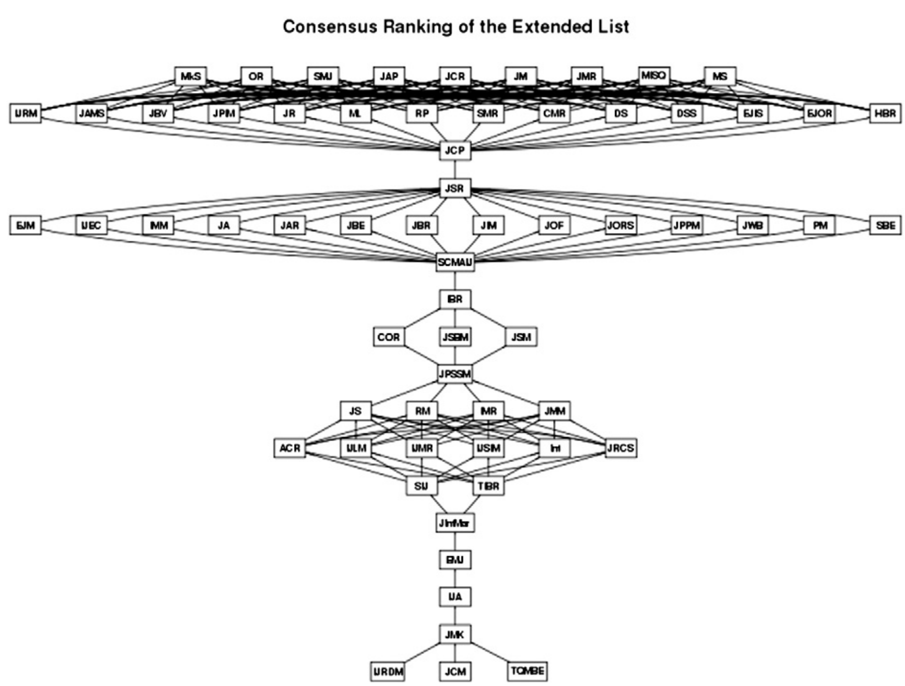

Fig. 3. Consensus journal rankings for the *extended* journal list.

core journal list, the journals ML, JCP, and JSR are responsible for dividing the complete group of marketing-related journals into these two broader subsets.

4.2. Stability of consensus solutions and degeneration effects

The above discussion of the properties of the two consensus solutions depicted in Figs. 2 and 3 shows that the number of ties apparently increases substantially when one adds more journals to the core list of marketing journals. In the following, the authors provide the results of a systematic sensitivity analysis of consensus ranking solutions, involving the use of different (subsets of) journals and rankings. Of particular interest are two questions: (1) Is the consensus solution robust to the inclusion of different journals and rankings? (2) Does the consensus solution show a tendency towards indifference (i.e., does the solution rank many journals equally) when one adds further journals and/or rankings? While the first question addresses the stability of consensus solutions, the second investigates potential degeneration effects in the derived solutions.

In order to investigate these two properties of consensus solutions, the authors conduct a bootstrap experiment. The bootstrapping approach is defined as follows: one randomly draws a given number q journals from $\mathfrak{I}$ and m rankings from $\mathfrak{R}$, where $q \in \{10,20,\ldots,60\}$ and $m \in \{2,3,\ldots,12\}$, and compute the consensus ranking for the corresponding collection. One then repeats this procedure 1000 times for each possible combination of q and m. To measure the stability of the consensus solutions, the authors compute the Kendall rank correlation coefficient (Kendall's τ) for the result of each bootstrap sample against a suitably matched subset from the consensus ranking, derived for all of the rankings and journals included in the extended list.

Fig. 4 plots the development of the average rank correlations between the overall consensus ranking and those derived for the bootstrap samples, for increasing numbers of journals q and rankings m, respectively. The plots clearly show a much more marked dependence of the rank correlations on the number of involved rankings than on the number of journals. While, for a given number of rankings, the inclusion of additional journals does not affect the stability of the derived solution much, the lower part of Fig. 4 shows that, when the number of rankings m increases, the rank correlation coefficient (almost linearly) converges to almost 1. This result is not very surprising, because one would expect to have much more difficulty finding an agreement between raters when considering an extra rater's opinion about a given set of journals, than in the case where a given set of raters has to rank a new journal. Regressing the average rank correlations on the two treatment factors q and m of the bootstrap experiment, as well as their interaction (the saturated model), confirms the visual diagnostic statistically. Using backward selection, one can reduce the complexity of the model, so that m remains as the only significant parameter, given a significance level of 0.05. The parameter estimate for the number of rankings m is 0.02 and explains more than 92% of the variance of the average rank correlation ($r^2 = 0.92$, with a p-value for the model of $p < 0.01$).

To investigate potential degeneration effects for the derived consensus rankings, depending on variations across the number of available journals q and rankings m, the authors compute the average number of ties per involved journal, in the consensus ranking, across all bootstrap replications. In contrast to the findings regarding the stability of consensus rankings for subsets of available journals and rankings, this test shows the opposite effect.

Fig. 5 depicts results which show that employing more journals to construct the consensus ranking increases the number of ties, regardless of the number of rankings one uses. This outcome seems quite reasonable, because the more journal quality rating opinions from different raters one must aggregate, the higher the tendency towards indifference one would expect. One can also observe this property of consensus solutions when comparing Fig. 2 against Fig. 3. Using the same backward selection procedure as above, one finds that only the number of journals q is statistically significant in explaining

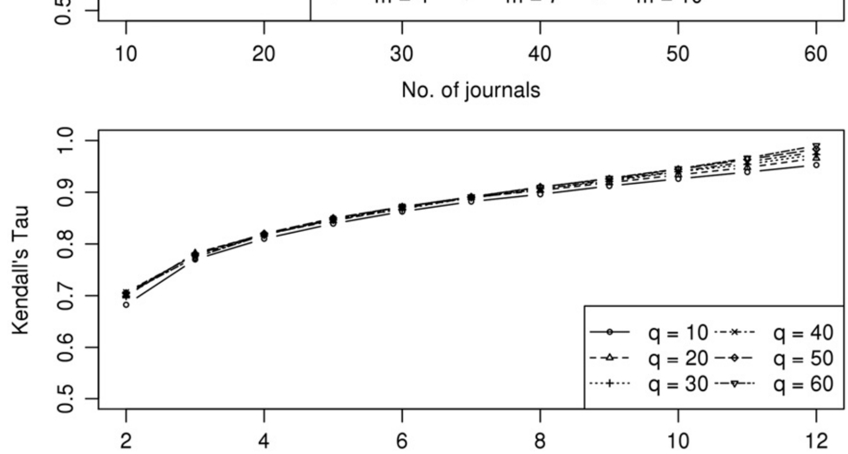

Fig. 4. Development of average rank correlations between the overall and bootstrap sample consensus rankings.

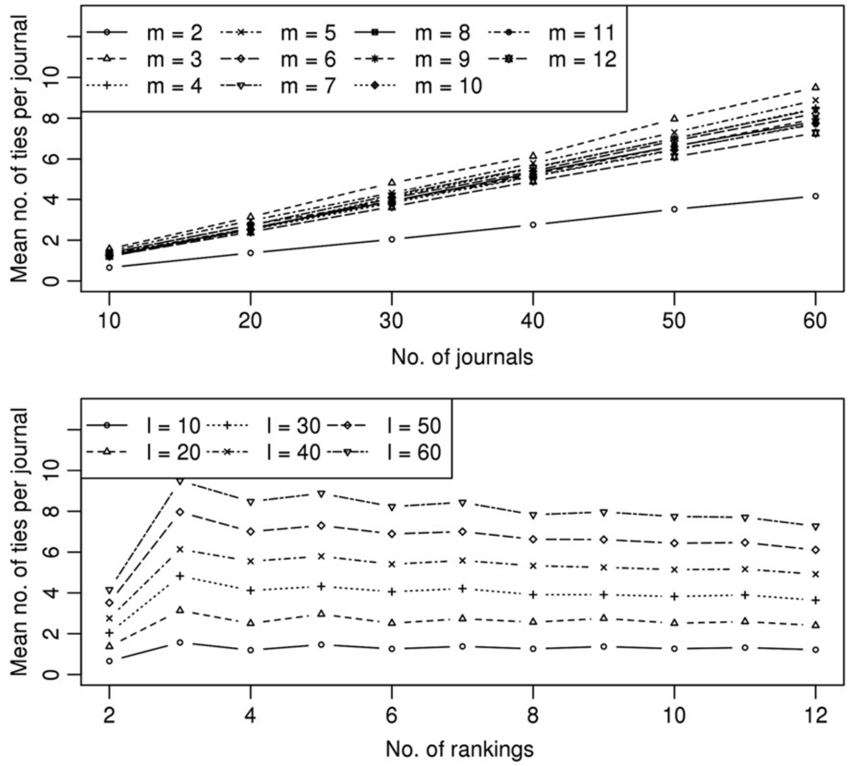

Fig. 5. Development of the average number of ties per journal across all bootstrap replications.

the suggested degeneration measure. The parameter estimate for q is 0.13 and explains more than 87% of the variance of the degeneration measure ($r^2 = 0.87$; $p < 0.01$).

4.3. Comparison of the consensus with naive and individual rankings

In view of the insights gained so far, an obvious question that arises is how dissimilar the individual rankings are from the derived consensus rankings. Furthermore, how the proposed consensus ranking methodology performs against simpler methods of aggregating individual rankings is of interest. In an attempt to investigate this issue in more detail, the authors employ an averaging procedure similar to that used by Dubois and Reeb (2000). To derive such a naive meta-ranking, they compute re-scaled preference scores, for each individual ranking for each journal, and sum them. Then, they transform the resulting meta-score into a rank order to represent the meta-ranking.

Table 5
Absolute Kemeny–Snell distances and average rank correlations for each ranking in comparison to all others.

	d	Kendall's τ
Consensus	12,140	0.61
Naive	12,904	0.59
ABDC08	13,930	0.58
ABS09	14,276	0.58
Ast08	14,814	0.55
UQ07	15,324	0.55
Vhb08	15,930	0.49
Cra09	17,110	0.51
EJL06	18,410	0.45
Cnrs08	18,992	0.45
Bjm04	19,380	0.38
Wie01	20,228	0.37
Hkb05	20,444	0.38
Theo05	21,442	0.42

To compare the relative performances of the ranking methods, the authors use the absolute sum of Kemeny–Snell distances d for each of the available rankings (including the consensus and the naive meta-

Table 6
Preference scores for the top 30 journals from the naive meta-ranking (NR) compared to the consensus ranking (CR).

	NR: scores	CR: ranks
MkS	−18.35	1
MS	−17.77	1
JCR	−16.87	1
JMR	−16.05	1
SMJ	−15.17	1
MISQ	−15.07	1
JM	−14.93	1
JAP	−14.13	1
OR	−13.03	1
JAMS	−11.66	2
RP	−10.07	2
DS	−9.61	2
JR	−9.44	2
JPIM	−8.73	2
IJRM	−7.84	2
JBV	−6.38	2
DSS	−5.76	2
HBR	−5.75	2
ML	−5.36	2
EJIS	−4.83	2
CMR	−4.37	2
SMR	−3.67	2
EJOR	−3.61	2
JCP	−2.00	3
SBE	−0.29	5
JBR	−0.26	5
JSR	−0.23	4
IJEC	0.34	5
JWB	0.34	5
JORS	0.40	5
JOF	0.57	5
JIM	0.73	5

ranking) vis-a-vis all other rankings, as well as the corresponding average rank correlation coefficient (Kendall's τ), as criteria for the representation quality of the meta-rankings. Table 5 shows them, together with the identifiers of the respective rankings, in descending order of *d*.

The smallest distance is between the consensus ranking and all other rankings. Of course, this finding is not surprising, because the consensus ranking by definition represents the global minimum for the given optimization criterion. However, this ranking also has the highest average rank correlation, which reflects its ability to serve as a superior meta-ranking. According to these measures, the naive ranking is the second-closest to all other rankings. At first sight, the ranking appears quite similar to the consensus ranking in terms of ordering (at least for the top-ranked journals). For example, the seven most preferred journals according to the consensus ranking are also the top seven journals according to the naive meta-ranking.

However, a more detailed examination of the actual positions the journals occupy in the aggregated rankings suggests that the consensus ranking method is a more adequate meta-ranking representation than the naive method. For an illustration, consider the top 30 journals according to the naive meta-ranking; Table 6 lists these journals, along with their respective preference scores and corresponding positions in the consensus solution.

While the JM appears in the group of highest-ranked journals in the consensus ranking, the JM is seventh according to the naive meta-ranking, even though the journal's naive meta-scores are relatively close to those of SMJ and MISQ. The problem with this naive kind of aggregation becomes even more apparent if one compares the performance of JM in the individual rankings to that of journals ranked higher according to the naive meta-ranking. In such a comparison, JM wins more pairwise comparisons against SMJ and MISQ than the other way round. More substantial differences occur if one proceeds further down the ranking (for a more detailed inspection, see the online appendix of this paper at http://statmath.wu.ac.at/projects/jcr/).

From a theoretical point of view, the discussion about the performance of the naive aggregation method goes back to the dispute between Borda (1781) and Condorcet (1785) and has produced a variety of arguments indicating that this method is often unsuccessful in representing the *true* majority decision. From a more practical perspective, another argument exists which makes the properties of a consensus ranking appealing: the permission of ties in the consensus ranking provides an evident and intuitively comprehensible basis for deciding on the appropriate cutoff points that discriminate between journals of several quality tiers (e.g., deciding on a boundary between A- and B-journals). In the case of naïve aggregation, this task becomes cumbersome and is much more dependent on discretionary human judgments.

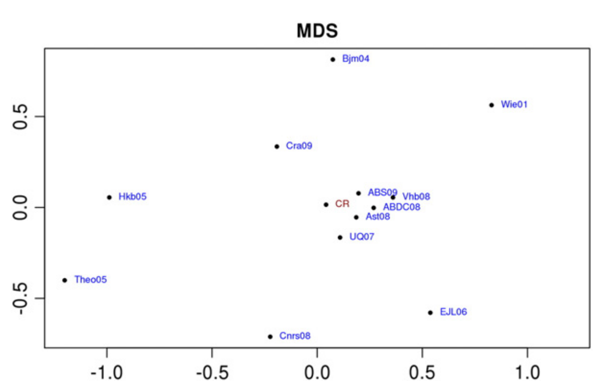

Fig. 6. Symmetric difference distance-based representation of the individual rankings and the consensus ranking.

To illustrate the consensus ranking's goodness of representation quality for the underlying individual rankings, Fig. 6 portrays the pairwise distances between the involved journal rankings, *d*, in a two-dimensional display. The configuration results from using the multidimensional scaling technique SMACOF (de Leeuw & Mair, 2009), which minimizes a stress measure by means of majorization (the nonmetric stress for the solution is 0.0147).

The consensus ranking CR clearly appears in the center of the two-dimensional plot. In general, the further a ranking's position is from the consensus, the lower the level of agreement. Quite obviously, the individual rankings concentrated around the consensus ranking share some potential to adequately reflect the aggregate perspective of the academic community in evaluating the quality of marketing journals, while those rankings that are further away fail to do so. In this respect, distance from the consensus ranking could also serve as an indication of the capability of an individual ranking to fulfill the properties of an appropriate global meta-ranking.

5. Conclusions

Journal rankings have become an important tool to assess the research quality of publications. Academics show widespread consent that focusing on a single journal ranking is risky and inadequate at reflecting the aggregate perspective of the academic community as to the quality of research publication outlets. Prior attempts to merge compilations of single rankings into suitable meta-rankings struggle with the different measurement scales used by the various rankings and the issue of incomplete information. This paper presents an optimization-based approach, demonstrating how one may derive consensus rankings from several individual ones. The approach is capable of accounting for different scale levels (numeric, ordinal) and partial intersections of the journal sets included in the aggregation.

The authors apply the proposed consensus ranking method to various subsets of marketing-related journals included in the *Harzing Journal Quality List*. Even though the single rankings are rather divergent for lower-ranked journals, the results show that one can derive a consensus ranking with a considerably high level of agreement with the original set of single rankings. Notwithstanding these results, one must be careful in drawing conclusions from such an analysis, because the results depend on the journal rankings one uses. The sensitivity analysis clearly shows that the number of individual rankings affects the stability of the derived consensus meta-ranking, whereas the solution tends to degenerate as the size of the journal list explodes.

However, the application to marketing-related journals also demonstrates the superiority of the consensus ranking over a simpler approach involving rank averaging. Compared to previous rather complicated and extensive efforts to adequately aggregate single rankings into meta-rankings, this paper's approach is easily implementable by using ready-to-use computational resources and applicable to a wide range of similar ranking aggregation tasks. Instead of requiring sometimes incomprehensible interventions by the analyst, the proposed procedure relies on a formal solution of the underlying optimization problem and thus produces an optimum level of agreement over the derived meta-ranking, among the set of single rankings. Thus, these findings should encourage researchers and, in particular, research assessment institutions to adopt a route that allows them to objectify their ranking efforts. This approach could contribute towards avoiding much of the sometimes very emotional and controversial discussion among academics about single domain-specific rankings.

References

Azar, O. H., & Brock, D.M. (2008). A citation-based ranking of strategic management journals. *Journal of Economics and Management Strategy*, *17*(3), 781–802.

Bakir, A., Vitell, S., & Rose, G. (2000). Publications in major marketing journals: An analysis of scholars and marketing departments. *Journal of Marketing Education*, *22*(2), 97–107.

Bancroft, D. R., Gopinath, C., Kovács, A.M., & Rejtö, L. K. (1999). A new methodology for aggregating tables: Summarizing journal quality data. *Journal of Business Venturing*, *14*(3), 311–319.

Baumgartner, H., & Pieters, R. (2003). The structural influence of marketing journals: A citation analysis of the discipline and its subareas over time. *Journal of Marketing*, *67*(2), 123–139.

Borda, J. C. (1781). Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*, 657–665.

Bravo, H. C., & Theußl, S. (2011). *Rcplex: R interface to CPLEX. R package version 0.3-0*.

Condorcet, M. J. A. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. Paris*.

Cook, W. D. (2006). Distance-based and ad hoc consensus models in ordinal preference ranking. *European Journal of Operational Research*, *172*(2), 369–385.

de Leeuw, J., & Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, *31*(3), 1–30.

Dubois, F. L., & Reeb, D. (2000). Ranking the international business journals. *Journal of International Business Studies*, *31*(4), 689–704.

Fishburn, P. C. (1972). *Mathematics of decision theory*. The Hague: Mouton.

Franke, N., & Schreier, M. (2008). A meta-ranking of technology and innovation management/entrepreneurship journals. *Die Betriebswirtschaft (DBW)*, *2008*(2), 185–216.

Freese, R. (2004). Automated lattice drawing. *Concept Lattices. Lecture Notes in Computer Science*, *vol. 2961*. (pp. 589–590), Springer.

Fry, E., Walters, C., & Scheurmann, L. (1985). Perceived quality of 50 selected journals: Academicians and practitioners. *Journal of the Academy of Marketing Science*, *13*(2), 352–361.

Hornik, K., & Meyer, D. (2007). Deriving consensus rankings from benchmarking experiments. In R. Decker, & H. -J. Lenz (Eds.), *Advances in Data Analysis (Proceedings of the 30th Annual Conference of the Gesellschaft für Klassifikation e.V., Freie Universität Berlin, March 8–10, 2006). Studies in Classification, Data Analysis, and Knowledge Organization*. (pp. 163–170), Springer-Verlag.

Hornik, K., & Meyer, D. (2010). *Relations: Data structures and algorithms for relations. R package version 0.5-7*.

Hult, G. T. M., Neese, W. T., & Bashaw, R. E. (1997). Faculty perceptions of marketing journals. *Journal of Marketing Education*, *19*(1), 37–52.

IBM ILOG (). *IBM ILOG CPLEX Callable Library C API 12.1 reference manual*.

Kemeny, J. G., & Snell, J. L. (1962). *Mathematical models in the social sciences*. Cambridge, Ch. II: MIT Press.

Mingers, J., & Harzing, A. -W. (2007). Ranking journals in business and management: A statistical analysis of the Harzing data set. *European Journal of Information Systems*, *16*(4), 303–316.

Pieters, R., Baumgartner, H., Vermunt, J., & Bijmolt, T. (1999). Importance and similarity in the evolving citation network of the International Journal of Research in Marketing. *International Journal of Research in Marketing*, *16*(2), 113–128.

Polonsky, M. J. (2008). Publishing on publishing: Streams in the literature. *European Business Review*, *20*(5), 401–420.

Polonsky, M., & Whitelaw, P. (2005). What are we measuring when we evaluate journals? *Journal of Marketing Education*, *27*(2), 189–201.

R. Development Core Team (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rainer, R. K. J., & Miller, M.D. (2005). Examining differences across journal rankings. *Communications of the ACM*, *48*(2), 91–94.

Regenwetter, M., & Rykhlevskaia, E. (2004). On the (numerical) ranking associated with any finite binary relation. *Journal of Mathematical Psychology*, *48*, 239–246.

Régnier, S. (1965). Sur quelques aspects mathématiques des problèmes de classification automatique. *ICC Bulletin*, *4*, 175–191.

Schrader, U., & Hennig-Thurau, T. (2009). VHB-JOURQUAL2: Method, results, and implications of the German Academic Association for Business Research's Journal Ranking. *BuR - Business Research*, *2*(2), 180–204.

Tellis, G. J., Chandy, R. K., & Ackerman, D. S. (1999). In search of diversity: The record of major marketing journals. *Journal of Marketing Research*, *36*(1), 120–132.

Theoharakis, V., & Hirst, A. (2002). Perceptual differences of marketing journals: A worldwide perspective. *Marketing Letters*, *13*(4), 389–402.

Wakabayashi, Y. (1998). The complexity of computing medians of relations. *Resenhas*, *3*(3), 323–349.

Zhou, D., Ma, J., & Turban, E. (2001). Journal quality assessment: An integrated subjective and objective approach. *IEEE Transactions on Engineering Management*, *48*(4), 479–490.