A ranking method based on handicaps

Gabrielle DEMANGE[1]

September 5, 2013

Abstract

Ranking methods are fundamental tools in many areas. Popular methods aggregate the statements of 'experts' in different ways. As such, there are various reasonable ranking methods, each one of them more or less adapted to the environment under consideration. This paper introduces a new method, called the handicap-based method, and characterizes it through appealing properties. This method assigns not only scores to the items but also weights to the experts. Scores and weights form an equilibrium for a relationship based on the notion of handicaps. The method is, in a sense made precise in the paper, the counterpart to the counting method in environments that require intensity-invariance. Intensity-invariance is a desirable property when the intensity of the experts' statements has to be controlled. Otherwise, both the counting and handicap-based methods satisfy a property called homogeneity, which is a desirable property when cardinal statements matter, as is the case in many applications.

Keywords ranking, scores, invariant method, peers' method, handicap, scaling matrix.

JEL D71, D89.

[1]Paris School of Economics-EHESS, address 48 bd Jourdan, 75014 Paris, France. e-mail demange@pse.ens.fr. The author is supported by ANR grant NET. Part of the material was in the working paper entitled 'Collective attention and ranking methods' WP 2010-42, PSE, which I split into two papers. I am very grateful to Oscar Volij for providing me journals' citations data. I thank Heinrich Nax and Salimeh Minirijavid for editorial assitance.

1 Introduction

The use of rankings is becoming pervasive in many areas. In the Web environment and in academia, popular ranking methods are based on observed data such as hyperlinks for ranking Web pages or citations for ranking journals. The underlying premise is that these data reflect preferences: a ranking method aggregates the hyperlinks toward a Web page or the citations toward an article as positive votes. The Science Citation Index, for example, uses the *counting* method, which counts the total number of citations received by journals. The influence measure introduced by Pinski and Narin (1976) counts not only direct citations but also indirect ones in a certain way. PageRank designed by Google (Brin and Page 1998) is based on a similar recursive approach and uses the *invariant* method. Different methods produce different results, which raises the question of the choice of the method. Viewing ranking methods as tools for aggregating the evaluations of several individuals, no method is universally good, as is known from social choice theory: an appropriate method depends on the context under consideration. A useful guide is to list the properties or 'axioms' that a method should satisfy in that context. This paper follows this axiomatization approach. It introduces and characterizes a new method, called the *handicap-based method*.

The ranking problems considered here are described by a set of 'items' to be ranked and a set of 'experts' who provide statements on the items. A ranking method assigns a cardinal ranking[2] of the items to each profile of experts' statements. This ranking describes the relative *scores* of the items, defined up to a multiplicative scalar. Let us illustrate the framework first with the ranking of journals based on citations. This is a peers' ranking because items are ranked according to data provided by themselves: the statement of a journal (as an expert) is derived from the number of its citations toward other journals. Similarly, in the case of ranking Web pages based on the hyperlink structure, the statement of a page (as an expert) is given by its hyperlinks toward other pages. In the next example, items and experts differ: the items are the different issues over which a budget has to be allocated and the experts are representatives. Representatives state their preferences over the allocation of the

[2]The term 'ranking' might not be appropriate in a cardinal setting but this terminology is standard.

budget and a method describes how the final budget will be distributed as a function of these statements.

This paper proposes a ranking method based on handicaps. The method and the axioms apply to general settings including the peers' one. To understand how that method is built, it is useful to view a method as assigning to experts' statements not only scores to the items but also weights to the experts. This encompasses two properties. First, the ranking forms a weighted average of the experts' statements. Second, the scores and the weights form an equilibrium relationship. It turns out that most existing methods are built that way. For example, the counting method simply assigns identical weights to the experts whatever their statements. PageRank instead, which aims to determine the influential Web pages on the basis that there are heavily cited by other influential pages, equalizes the expert's weight associated to a page to its score.

The method introduced in this paper is supported by an equilibrium relationship built on handicaps. Scores and handicaps are strongly related. Specifically, the handicap of an item can be seen as the inverse of its score: since the purpose of handicaps is to adjust items' marks so as to equalize their 'strength', the handicap of i is half that of ℓ if i can be said to be twice as good as ℓ, that is, if i's score is twice that of ℓ. A handicap-based ranking and the experts' weights aim at equalizing the strengths across items and at equalizing the handicap points distributed by the experts. These two conditions turn out to define a unique ranking and unique experts' weights under mild conditions (Proposition 1). The handicaps can be viewed as a tool for building the method. The properties that characterize the method justify its relevance.

The handicap-based method is characterized by three properties. The first property is *intensity-invariance*. This property has been introduced in environments where the 'intensity' of statements is not controlled. In the Web environment, for example, the number of (outward) links from a page is not restricted *a priori*; an intensity-invariant method deflates a link from that page by its total number of outward links. This is performed by PageRank to avoid pages to increase their score by inflating the number of their outward links. Note that intensity-invariance is automatically satisfied when statements represent shares, as in the budget allocation problem. The second property is *homogeneity*. This property is appropriate when statements and

rankings are cardinal. Let each expert raise its evaluation on item i by the same factor, say twenty percent. Then the score of i should be raised by twenty percent relative to other items. The third property is *uniformity*[3] according to which items are considered as undistinguishable when they all receive the same totals: the method assigns equal scores when the counting method does.

The handicap-based method is the only method that is intensity-invariant, homogeneous, and uniform. Furthermore, the handicap-based method can be seen as the counterpart to the counting method in environments that require intensity-invariance. Indeed, the counting method, which is not intensity-invariant, is characterized by homogeneity and uniformity on the set of statements whose intensity has not been factored out. To summarize, the handicap-based method provides a tool for aggregating shares; it is appropriate when experts' statements are not controlled and intensity-invariance is perceived as a prerequisite, or when experts' statements are controlled and represent shares (in which case intensity-invariance is automatically satisfied).

To illustrate the method, Section 3.4 reports the invariant and handicap-based rankings of 37 economic journals, based on the same data as in Palacios-Huerta and Volij (2004). There are differences in the scores computed by these methods, though they are rather moderate but for some exceptions. An interesting lesson that can be drawn from this exercise is that the weights of the handicap-based ranking convey relevant information: they reflect significant and meaningful differences in the reference behavior across journals.

This paper is related to recent studies that characterize ranking methods based on citations. In the peers' context, the invariant method is characterized by several axiomatizations (Palacios-Huerta and Volij 2004, Slutzki and Volij 2006, and Altman and Tennenholtz 2005 for its ordinal version). Du, Lehrer and Pauzner (2012) propose a 'market' approach to rank Web pages and obtain a family of methods that appear as variants of the invariant method. In a different context, Woeginger (2008) provides an axiomatization of the h-index, a much used method for ranking researchers. The method relies on the number of citations received by each researcher's paper independently of citations' origins. The paper is also related, albeit loosely, to the studies analyzing incentive compatibility in the peers' context. For example Altman

[3]Intensity-invariance and uniformity have been considered in the literature by various authors, as those cited at the end of this introduction. Homogeneity as defined here is new.

and Tennenholtz (2008) provide an impossibility theorem and Clippel, Moulin and Tideman (2008) display a family of methods that satisfy a strong non-manipulability requirement (the score of an entity must not be affected by its own citations). This paper does not consider incentive compatibility, though intensity-invariance prevents a specific form of manipulation.

The rest of the paper is organized as follows. Section 2 presents ranking methods, defines some properties, and describes the invariant and the HITS methods. Section 3 introduces the handicap-based method under the assumption of strictly positive statements, describes an algorithm to compute the ranking and provides characterizations for both the handicap-based and counting methods; finally it makes comparisons between methods and presents three rankings of economic journals. Section 4 investigates settings in which statements can be nil or binary, a typical situation in the Web environment in which statements are limited to the presence or absence of a link. Section 5 concludes. Most proofs are given in Section 6.

2 Ranking methods

2.1 The framework

There are n *items* to be ranked and m *experts* who provide *statements* on the items on which the ranking will be based. Items can be individuals, journals, or political parties, and experts can be pundits, journals, or voters, as is illustrated below. $N = \{1, \ldots, n\}$ and $M = \{1, \ldots, m\}$ denote respectively the set of items and the set of experts.

An expert's statement assigns a non-negative *valuation* to each item; j's valuation to i is denoted by $\pi_{i,j}$. There is a feasible set of experts' statements that depends on the context, as is illustrated in the next examples. The *statement matrix* is the $n \times m$ matrix $\boldsymbol{\pi} = (\pi_{i,j})$: column j represents j's statement, i.e., the valuations of j to all items and row i represents the valuations of all experts on item i. A matrix is said to be feasible if the statement of each experts is feasible.

A *ranking* assigns a non-negative number r_i to each item i, called the *score* of i. It describes the relative strength of the n items, meaning that the values taken by the scores matter up to a multiplicative constant. Normalizing the sum of the scores

to 1, a ranking of N is specified by a vector $\mathbf{r}$ in the simplex Δ_N: $\Delta_N = \{\mathbf{r} = (r_i) \in \Re^n,\ r_i \geq 0, \sum_i r_i = 1\}$.

A method assigns a ranking to each feasible statement matrix. Formally,

Definition 1 *Given N, M, and the set $\mathcal{S}$ of feasible statements matrices, a ranking method F assigns to each matrix $\boldsymbol{\pi}$ in $\mathcal{S}$ a ranking $\mathbf{r} = F(\boldsymbol{\pi})$ in Δ_N.*

The **counting** method, for example, assigns scores proportional to the totals of the valuations:

$$r_i = \frac{\pi_{i+}}{\sum_\ell \pi_{\ell+}} \text{ for each } i \text{ where } \pi_{i+} = \sum_{j \in M} \pi_{i,j}. \tag{1}$$

More sophisticated methods are introduced in Section 2.3. Let us illustrate ranking problems in different contexts.

1. Ranking journals based on citations. The ranking is based on the citations between a set of journals: N and M coincide. The statement of journal j is given by the number of its citations per article towards all journals. To be more precise, let $C_{i,j}$ denote the total number of citations by articles in j to articles in i and n_j denote the total number of articles in j in a relevant period; the matrix $\boldsymbol{\pi}$ is given by $\pi_{i,j} = \frac{C_{i,j}}{n_j}$. A ranking method assigns a ranking based on $\boldsymbol{\pi}$. The sum of column j represents the average number of references in an article published in j and is called j's reference intensity. Reference intensities differ across journals or across fields, which raises the issue of whether these differences should have an impact on the ranking (see Palacios-Huerta and Volij (2004) for such an analysis on economic journals).

2. Representation problem. The problem is to assign voting weights to various categories based on the votes of electoral bodies: N is the set of categories (items) and M is the set of electoral bodies (experts). In the political domain for instance, a category is a political party and an electoral body represents a constituency, say a district. In the context of a scientific association, a category represents a field and an electoral body represents a geographical area. The statement $\pi_{i,j}$ of j on category i is the number of votes cast by the electoral body j in favor of i. The counting method assigns voting weights to categories in proportion to their votes' totals; therefore it treats all votes equally without

distinguishing from which electoral body they have been cast. Other methods distinguish the votes according to the electoral body, as we will explain in section 2.2.

3. Budget allocation. The problem is to allocate a budget to different issues (transport, sanitation, education,...) as a result of the desiderata of citizens or representatives: the items are the issues and the experts are the representatives. A representative's statement describes the proportions of the budget she would like to allocate to the different issues:[4] a statement is represented by a non-negative vector that sums to one. A ranking is interpreted as the shares of the final budget. Thus, a ranking method describes a rule that distributes the final budget as a function of the representatives' statements.

4. Ranking Web pages based on the link structure. The two sets of items and experts, N and M, coincide, both given by a set of relevant Web pages. A method ranks the pages based on the links within N, as is performed by PageRank (using the invariant method) or by the HITS method described in Section 2.3. The statement of a page is given by its outward links towards other pages and $\boldsymbol{\pi}$ is the (transpose of) adjacency matrix of the Web network: $\pi_{i,j}$ is equal to 1 if page j points to i and 0 otherwise. Such a binary representation also arises in approval voting, in which an expert is asked to name the items he finds acceptable (without being allowed to state intensity). Here items and experts can differ. An expert's statement is described by the vector of 1 and 0 indicator of the set of items he approves. Such a setting where the statement matrix has only zeros and ones is called the 0-1 setting in the remainder of the paper.

In all these examples but the last one, the statements are cardinal and the precise relative values stated by the experts have a meaning. It is thus natural to assign a

[4]This is in the spirit of the Participatory Budgeting procedures implemented in Porto Alegre (see e.g. Souza 1998). These procedures are more complex than our method because they involve two levels of voting, first at the neighborhoods' level (the city is divided into regions, themselves divided into neighborhoods) and second at the regional level. All inhabitants in a neighborhood can vote both to formulate priority demands over investment spendings and to elect representatives at the region level. These representatives aggregate the neighborhoods' demands to set regional priorities. Finally all these regions' priorities are used by the city to derive a general priority ranking over issues and an allocation to regions.

cardinal ranking. In the last example, though statements are not cardinal, a cardinal ranking still makes sense. For example, in approval voting, the counting method produces the proportions of the votes received by the candidates and these proportions are relevant. However an axiom that makes sense when statements are cardinal may have no sense in such a 0-1 setting. This will be the case for the homogeneity axiom introduced in the next section.

Statement matrices are restricted to be positive, $\pi_{i,j} > 0$ for each i, j, except in Section 4. This is a reasonable assumption in the three first examples. Two feasible sets will be considered: the full set of positive matrices, denoted by $\mathcal{P}$, representing absolute statements (the journal and representation examples) and the set of positive matrices for which each column sums to 1, denoted by $\mathcal{R}$, representing relative statements (the budget example). The possibility of zeros and the 0-1 setting will be analyzed in Section 4.

Notation $\mathbb{1}_N$ denotes the vector in $\Re^N$ whose components are equal to 1, and $\mathbf{e}_N = \frac{1}{N}\mathbb{1}_N$ the ranking that assigns equal scores to items.

Given a finite set I and a vector $\boldsymbol{x}$ in $\Re^I$, $dg(\boldsymbol{x})$ denotes the diagonal $I \times I$ matrix with x_i as the i-th element on the diagonal.

Given a matrix $\boldsymbol{\pi} = (\pi_{i,j})$, π_{i+} denotes the total in row i, $\pi_{i+} = \sum_{j \in M} \pi_{i,j}$, and π_{+j} the total in column j, $\pi_{+j} = \sum_{i \in N} \pi_{i,j}$.

2.2 Some properties

Let us start by describing four natural properties that one may want a method to satisfy. Three properties - intensity-invariance, uniformity, and exactness - appear in the literature (under various names). The homogeneity property has not yet been considered in the literature, as far as I know.

Intensity-invariance Intensity-invariance requires the ranking not to be affected by a multiplicative scaling of a column. Justifications are provided below. Formally, let $\boldsymbol{\pi}'$ be the matrix obtained from $\boldsymbol{\pi}$ by multiplying a column, say column j, by a positive scalar μ_j: $\boldsymbol{\pi}' = \boldsymbol{\pi} dg(\boldsymbol{\mu})$ where $\boldsymbol{\mu}$ is the m-vector whose j-th component is equal to μ_j and others are equal to 1. Intensity-invariance requires the method to assign the same ranking to $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$: $F(\boldsymbol{\pi} dg(\boldsymbol{\mu})) = F(\boldsymbol{\pi})$. The property is required

for each column j; multiplying each column j by a positive μ_j, iteration yields the following equivalent definition.

Definition 2 *A method F defined on $\mathcal{P}$ is* **intensity-invariant** *if*

$$F(\boldsymbol{\pi} dg(\boldsymbol{\mu})) = F(\boldsymbol{\pi}) \text{ for each positive } m\text{-vector } \boldsymbol{\mu} = (\mu_j), \text{ each } \boldsymbol{\pi} \text{ in } \mathcal{P}. \tag{2}$$

An intensity-invariant method is fully determined by its restriction on the set of matrices whose column totals are fixed. To see this, let $\mathbf{c} = (c_j)$ specify a positive value for each column total, c_j for column j. Given matrix $\boldsymbol{\pi}$, scale each column j so that its total meets the required total c_j. The scaled matrix is equal to $\boldsymbol{\pi} dg(\boldsymbol{\mu})$ where μ_j satisfies $\pi_{+j}\mu_j = c_j$. Intensity-invariance of F implies $F(\boldsymbol{\pi}) = F(\boldsymbol{\pi} dg(\boldsymbol{\mu}))$: F is fully determined by its restriction on the set of matrices whose column totals are equal to $\mathbf{c}$.

This suggests a way to transform a method F that is not intensity-invariant into an intensity-invariant one. Let us consider the restriction of F on the matrices with given column totals $\mathbf{c}$ and extend it as follows: given a matrix, scale each column j so that it sums to c_j and apply F to the scaled matrix. Formally, define $[F]_{\mathbf{c}}$ by: for each $\boldsymbol{\pi}$, $[F]_{\mathbf{c}}(\boldsymbol{\pi}) = F(\boldsymbol{\pi} dg(\boldsymbol{\alpha}))$ where $\pi_{+j}\alpha_j = c_j$. $[F]_{\mathbf{c}}$ is intensity-invariant since the scaled matrix of $\boldsymbol{\pi} dg(\boldsymbol{\mu})$ is the same as that of $\boldsymbol{\pi}$.

In the above construction, the methods $[F]_{\mathbf{c}}$ vary with $\mathbf{c}$ (except if F is intensity-invariant). Take for instance the counting method, which is not intensity-invariant. $[F]_{\mathbf{c}}$ assigns scores proportional to the weighted totals of the valuations: $[F]_{\mathbf{c}}(\boldsymbol{\pi})$ is proportional to $(\sum_j \frac{\pi_{i,j}}{\pi_{+j}} c_j)$. As a result, the influence of j' statement is increasing in its assigned value c_j. This shows that intensity-invariance is not related with fairness.

Let us illustrate intensity-invariance in our examples.

In the case of journals, intensity-invariance means that the ranking depends only on the proportions of the citations by journals to other journals, i.e. on $\frac{C_{i,j}}{C_{+j}}$ where $C_{i,j}$ is the number of cites made by articles in j to articles in i (recall that $\pi_{i,j}$ is defined as $\frac{C_{i,j}}{n_j}$ the average number of references of an article from j to i). As a result, a proportional increase in the number of citations per article in a journal j, keeping the shares received by each journal unchanged, has no impact on the ranking. In particular, the ranking is not influenced by distinct citations' practice across journals or fields.

In the case of Web pages, intensity-invariance implies that a link from a page is divided by the number of links from that page. Intensity is 'factored out'. A justification is that factoring out intensity avoids a page to improve its score by increasing the number of pages it points to. This is why PageRank uses the invariant method, which is the intensity-invariant version of another method, as is described in the next section.

In the representation problem, an expert represents the electoral body of a constituency whose statement is the number of votes to the parties in that constituency. Intensity-invariance requires the final representation to be independent of the turnout in the constituencies. Using the construction described above, an intensity-invariant method is obtained by assigning a total to each electoral body. The totals are not necessarily proportional to the sizes of the electoral bodies. This is often the case in practice: constituencies are assigned a number of representatives degressive in their sizes.

Finally, in the budget example, intensity-invariance is automatically satisfied since representatives are asked to state how the budget should be distributed over a set of issues, i.e., the statement of an expert sums to 1.

In the sequel, we work with matrices $\mathcal{R}$, whose column totals are equal to 1, namely matrices in $\mathcal{R}$, and with the associated intensity-invariant versions.[5] Specifically let $[\boldsymbol{\pi}]$ be the matrix in $\mathcal{R}$ associated to $\boldsymbol{\pi}$:

$$[\pi]_{i,j} = \frac{\pi_{i,j}}{\pi_{+j}} \text{ for each } i, j. \tag{3}$$

The intensity-invariant version $[F]_{\mathbf{1}}$ of F, denoted simply as $[F]$, is defined by

$$[F](\boldsymbol{\pi}) = F([\boldsymbol{\pi}]) \text{ for each } \boldsymbol{\pi} \in \mathcal{P}. \tag{4}$$

Finally, note that any method that is defined on $\mathcal{R}$, such as in the budget example, can be extended in a unique way to an intensity-invariant method on $\mathcal{P}$.

Uniformity and exactness The next two properties bear on some specific matrices, hereafter called row-balanced. A matrix is said to be *row-balanced* if each row receives the same total, or, equivalently, each row obtains the same score $1/n$ under

[5] By doing so, no intensity-invariant method is neglected since $G = [G]_{\mathbf{1}}$ for G intensity-invariant. In particular, $[F]_{\mathbf{c}} = [[F]_{\mathbf{c}}]_{\mathbf{1}}$.

the counting method. A row-balanced matrix constitutes a 'neutral' situation in the sense that there is no rationale for distinguishing between items if experts are not discriminated *a priori*. This is what is required by uniformity: A method is uniform if it assigns equal scores to each row-balanced feasible statement matrix. Exactness asks the converse property that items obtain equal scores only if they receive identical totals. Formally

Definition 3 *A method F is* **uniform** *on $\mathcal{S}$ if $F(\boldsymbol{\pi}) = \mathbf{e}_N$ for each row-balanced $\boldsymbol{\pi}$ in $\mathcal{S}$. F is* **exact** *on $\mathcal{S}$ if $F(\boldsymbol{\pi}) = \mathbf{e}_N$ for $\boldsymbol{\pi}$ in $\mathcal{S}$ implies that $\boldsymbol{\pi}$ is row-balanced.*

The intensity-invariant version of a uniform method F on $\mathcal{P}$ is uniform on $\mathcal{R}$ (but not necessarily on $\mathcal{P}$). To see this, let $\boldsymbol{\pi}$ be in $\mathcal{R}$. We have $[\boldsymbol{\pi}] = \boldsymbol{\pi}$ and $[F](\boldsymbol{\pi}) = F(\boldsymbol{\pi})$. Hence, if F is uniform, then $[F](\boldsymbol{\pi}) = \mathbf{e}_N$ for each row-balanced $\boldsymbol{\pi}$ in $\mathcal{R}$: $[F]$ is uniform. Similarly the intensity-invariant version of an exact method F on $\mathcal{P}$ is exact on $\mathcal{R}$ (but not necessarily on $\mathcal{P}$). Let $\boldsymbol{\pi}$ be in $\mathcal{R}$ such that $[F](\boldsymbol{\pi}) = \mathbf{e}_N$, thus, $F(\boldsymbol{\pi}) = \mathbf{e}_N$; if F is exact, this implies that $\boldsymbol{\pi}$ is row-balanced : $[F]$ is exact.

No intensity-invariant method is both uniform and exact on $\mathcal{P}$. To see this, start with a row-balanced matrix $\boldsymbol{\pi}$ with distinct columns. Its ranking is $\mathbf{e}_N$ by uniformity. The distinct columns can be multiplied by some factors so as to obtain a matrix that is not row-balanced. Intensity-invariance requires the ranking of this new matrix to be $\mathbf{e}_N$, in contradiction with exactness.

Homogeneity The homogeneity property is very natural when statements are cardinal[6] as is the case in the three first examples presented in Section 2. The relative valuations stated by an expert have a precise meaning. If there is a single expert, the cardinal ranking must keep these relative valuations, hence be proportional to this expert's statement. In particular, if the expert doubles the valuation on i, either in absolute or relative terms, the final score of i is doubled relative to all other items. Homogeneity extends this property to the multi-expert setting. Starting with experts' statements and multiplying each valuation on i by a factor, i's relative position should be multiplied by the same factor. I spell out the property for the two situations in which absolute or relative statements matter.

[6]The homogeneity axiom introduced in Palacios-Huerta and Volij (2004) is different since it bears on a given matrix that has two proportional rows.

Let us start with the situation in which statements are absolute. Let $\boldsymbol{\pi}$ be in $\mathcal{P}$ and multiply each valuation on item i by a positive scalar ρ_i. Homogeneity of F requires i's score relative to other items to be multiplied by ρ_i. Formally, let us consider the matrix $dg(\boldsymbol{\rho})\boldsymbol{\pi}$ where $\boldsymbol{\rho}$ is the vector whose i-th component is equal to ρ_i and all others are equal to 1; the ranking $F(dg(\boldsymbol{\rho})\boldsymbol{\pi})$ is the ranking proportional to $dg(\boldsymbol{\rho})F(\boldsymbol{\pi})$. The property is required for each row i and iteration yields the following equivalent definition.

Definition 4 *A method F is* **homogeneous on absolute statements** *if for each $\boldsymbol{\pi}$ in $\mathcal{P}$ and positive n-vector $\boldsymbol{\rho} = (\rho_i)$, $F(dg(\boldsymbol{\rho})\boldsymbol{\pi})$ is the ranking proportional to $dg(\boldsymbol{\rho})F(\boldsymbol{\pi})$.*

Clearly, the counting method is homogeneous on absolute statements. Let us illustrate the property in the two first examples of section 2.1.

In the case of journals, assume that, between two periods, *each* journal increases its citations toward journal 1 by ten percent and leaves unchanged the others: $\rho_1 = 1.1$. Homogeneity on absolute statements requires the score of journal 1 to be increased by ten percent relative to others: it becomes $\frac{1.1 \cdot r_1}{\Sigma_i r_i + 0.1 \cdot r_1} = \frac{1.1 \cdot r_1}{1 + 0.1 \cdot r_1}$ if r_1 denotes the initial score.

In the representation problem, let the number of votes in favor of party i be raised by 5 percent in all districts between two elections, other numbers of votes unchanged (such an rise implies an increase in the total number of votes; this is possible since larger participation and demographic modifications make the number of cast votes variable). Homogeneity on absolute statements requires i's voting weights to be raised by 5 percent relative to other parties.

Homogeneity on relative statements requires the same behavior when a factor modifies the relative valuations on an item: multiplying the shares on i relative to other items by a positive scalar ρ_i multiplies i's score relative to other items by ρ_i. In matrix form, $F([dg(\boldsymbol{\rho})\boldsymbol{\pi}])$ is the ranking proportional to $dg(\boldsymbol{\rho})F(\boldsymbol{\pi})$ where $\boldsymbol{\rho}$ is the vector with i's component equal to ρ_i and all others equal to 1. The property is required for each row i, so iteration yields the following equivalent definition.

Definition 5 *A method is* **homogeneous on relative statements** *if for each $\boldsymbol{\pi}$ in $\mathcal{R}$ and positive n-vector $\boldsymbol{\rho} = (\rho_i)$, $F([dg(\boldsymbol{\rho})\boldsymbol{\pi}])$ is the ranking proportional to $dg(\boldsymbol{\rho})F(\boldsymbol{\pi})$.*

In the budget example, recall that experts state their preferred budget shares over issues, so that their statements are in $\mathcal{R}$. Let us consider two cities which have the same number of representatives and face the same set of issues. Assume their statements differ only by the fact that each representative in the second city assigns α percent more to education relative to other issues than in the first (i.e. multiplied by the factor $\rho = 1+\alpha$). Homogeneity on relative statements requires the share devoted to education to be α percent larger relative to other issues in the second community than in the first; if the education share is 20 percent in the first community for example, it is $(20 + 20\alpha)/(100 + 20\alpha)$ in the second. The following statements with three experts and two issues, say education and health, illustrate:

$$\text{city 1}: \begin{pmatrix} 2/3 & 1/3 & 1/2 \\ 1/3 & 2/3 & 1/2 \end{pmatrix} \quad \text{city 2}: \begin{pmatrix} 4/5 & 1/2 & 2/3 \\ 1/5 & 1/2 & 1/3 \end{pmatrix}$$

Each expert wants the share on education relative to health in city 2 to be twice that in city 1; homogeneity requires the budget's share on education relative to health in city 2 to be twice that in city 1. For example, if the method is both uniform and homogeneous, education and health receive each half of the budget in city 1 (since statements are balanced), and education receives 2/3 and health 1/3 of the budget in city 2.

The distinction between homogeneity on absolute and relative statements does not matter for intensity-invariant methods. This is stated in the next lemma (the proof is in the appendix).

Lemma 1 *Let F be intensity-invariant. F is homogeneous on absolute statements if and only if it is homogeneous on relative statements.*

Thus, in the sequel, we simply refer to homogeneity for an intensity-invariant method.

When a method is homogeneous on absolute statements but not intensity-invariant, its intensity-invariant version $[F]$ may not be homogeneous (in whatever sense), as is illustrated with the counting method.

The counting method is homogeneous on absolute statements but not intensity-invariant. In the following example, $\boldsymbol{\pi}'$ is obtained by multiplying the first row of $\boldsymbol{\pi}$ by 2:

$$\boldsymbol{\pi} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \boldsymbol{\pi}' = \begin{pmatrix} 4 & 2 \\ 1 & 2 \end{pmatrix}, [\boldsymbol{\pi}'] = \begin{pmatrix} 4/5 & 1/2 \\ 1/5 & 1/2 \end{pmatrix}.$$

The intensity-invariant version of the counting method assigns equal scores, $(1/2, 1/2)$, to $\boldsymbol{\pi}$ but $(13/20, 7/20)$ to $\boldsymbol{\pi}'$ (and to $[\boldsymbol{\pi}']$) instead of $(2/3, 1/3)$ as required by homogeneity. The reason is that expert 1 likes relatively more item 1 than expert 2 does, hence its total increases more than that of expert 2 when item 1's statements are doubled. Thus, when normalizing $\boldsymbol{\pi}'$, the adjustment on expert 1's statements is larger than on 2's. This explains why the total of item 1's shares is less than doubled relative to that of item 2. ■

It should be noted that homogeneity is not appropriate in some contexts. For example, in a setting where a statement is ordinal, represented, say, by valuations from 1 to n, multiplying a valuation by some constant does not make sense. Similarly, in a 0-1 setting such as in the Web where a zero reflects the absence of a link, multiplying a valuation by some constant makes the statement non-feasible. In those cases, the homogeneity axiom does not apply. Finally, anticipating on Section 4 which treats statements with null valuations, note that multiplying a null valuation by a factor leaves it null. Thus, the homogeneity axiom makes sense when an expert assigns a null valuation to an item because he finds it not acceptable or infinitely below an item with a positive valuation.

2.3 Examples: the Invariant and HITS methods

This section introduces well-known methods that differ from the counting one in two ways. First, they are defined in peers' settings in which items and experts coincide ($N = M$). Second, they treat experts differently according to their statements whereas the counting method treats them equally.

The next two methods, often called *eigenvalue* methods, are the Liebowitz-Palmer (LP) method[7] and its intensity-invariant version, called the invariant method. The methods are based on the premise that the statements made by a peer (as an expert) should be weighed by his score (as an item). This induces a loopback definition: up to a multiplicative factor, the score of an item is the sum of the received valuations where each one is weighted by the expert's score. Specifically, the method looks for

[7]This terminology refers to the work of Liebowitz and Palmer (1984), who use an approximation of the method for ranking economic journals. The methods or some variants have been (re)defined and used in various contexts: in sociology by Katz (1953) and Bonacich (1987), in academics for ranking journals by Pinski and Narin (1976).

$\mathbf{r}$ in Δ_N that satisfies

$$\text{for some positive } \lambda,\ r_i = \lambda \sum_{j \in N} \pi_{i,j} r_j \text{ for each } i. \tag{5}$$

According to (5), the positive vector $\mathbf{r}$ is an eigenvector of matrix $\boldsymbol{\pi}$. By Perron-Frobenius theorem on matrices with positive elements, such an eigenvector exists, and is unique up to a scalar: (5) well defines a method.

The invariant method is the intensity-invariant version[8] of the LP method. Since a normalized matrix has its largest eigenvalue equal to one,[9] the invariant ranking of $\boldsymbol{\pi}$ is the unique $\mathbf{r}$ in the simplex that satisfies

$$r_i = \sum_{j \in N} [\pi]_{i,j} r_j \text{ for each } i. \tag{6}$$

The method is intensity-invariant, uniform and exact. It is not homogeneous; its behavior with respect to the multiplication of items' valuations is illustrated in section 3.3.

In the network setting, Du *et al.* (2012) propose methods that turn out to generalize the invariant method. An exchange economy is associated to each network and the resulting equilibrium prices of the nodes are defined as their scores. For Cobb-Douglas preferences, the invariant ranking is obtained. Though different ranking methods are obtained for different families of preferences, all methods reflect the idea that the higher the score of a node is, the more valuable its recommendation is. This property qualifies a method as a 'pure' peer's method.[10] The property is questionable in some contexts, as is illustrated below after presenting the next method.

The Hyperlink-Induced Topic Search (HITS) method, introduced by Kleinberg (1999) assigns scores to a set of Web pages on the basis of their link structure, as does PageRank. Thus, the two sets of items and experts coincide. The method however distinguishes two weights for each Web page, one associated with the relevance or authority of a page, the other with the adequacy of a page to point toward the relevant

[8]The invariant method serves as a basis to PageRank to rank Web pages, using the incidence matrix associated to the links between pages (see example 4). Because the matrix has many zeros, a perturbation is used to make it irreducible.

[9]Recall that a matrix and its transpose have identical eigenvalues. The set of equations $\sum_i \pi_{i,j} = 1$ for each j implies that $\mathbb{1}_N$, is an eigenvector of the transpose of $\boldsymbol{\pi}$ with eigenvalue 1. Since $\mathbb{1}_N$ positive, 1 is the dominant eigenvalue.

[10]See Demange (2013) for a definition and an analysis in a dynamical framework.

pages. The first set of weights defines the ranking, which should help users to find the relevant pages. The second set of weights identifies the pages -called 'hubs'- that are important because they point to relevant pages (but might not be useful to Internet users). Specifically, the method assigns the ranking $\mathbf{r}$ and the experts' weights $\mathbf{q}$ in Δ_N that satisfy

$$r_i = \sum_j \pi_{i,j} q_j \text{ for each } i \text{ and } q_j = \lambda \sum_i \pi_{i,j} r_i \text{ for each } j \tag{7}$$

for some positive λ. As argued by Kleinberg (1999), hubs and authorities exhibit a *mutually reinforcing relationship*: a good authority is a page that is pointed to by many good hubs, a good hub is one that points to many good authorities. The HITS method is well defined: (7) writes in matrix form as $\mathbf{r} = \boldsymbol{\pi}\mathbf{q}$ and $\mathbf{q} = \lambda\widetilde{\boldsymbol{\pi}}\mathbf{r}$ where $\widetilde{\boldsymbol{\pi}}$ is the transpose of $\boldsymbol{\pi}$, which implies $\mathbf{r} = \lambda\boldsymbol{\pi}\widetilde{\boldsymbol{\pi}}\mathbf{r}$ and $\mathbf{q} = \lambda\widetilde{\boldsymbol{\pi}}\boldsymbol{\pi}\mathbf{q}$. Thus the ranking (or 'authority' weights) $\mathbf{r}$ and the ('hub') weights $\mathbf{q}$ are respectively the unique normalized principal eigenvectors of the positive matrices $\boldsymbol{\pi}\widetilde{\boldsymbol{\pi}}$ and $\widetilde{\boldsymbol{\pi}}\boldsymbol{\pi}$.[11] The method is uniform, exact, but not homogeneous on absolute statements.

Supporting weights The counting, invariant, and HITS methods can all be viewed as assigning not only scores to items but also weights to experts. As can be seen from (1), (6) and (7), each of the three methods assigns to each $\boldsymbol{\pi}$ a ranking $\mathbf{r} = F(\boldsymbol{\pi})$ and experts' weights $\mathbf{q} = Q(\boldsymbol{\pi})$ so that each item receives a score that is the weighted sum of its valuations:

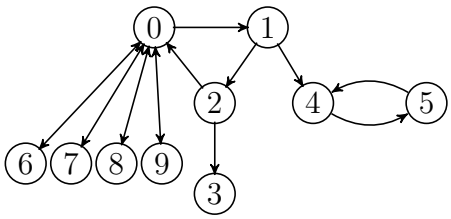
$$r_i = \sum_{j \in M} \pi_{i,j} q_j \text{ for each } i \text{ in } N \text{ or } F(\boldsymbol{\pi}) = \boldsymbol{\pi} Q(\boldsymbol{\pi}). \tag{8}$$

Furthermore, the weights $Q(\boldsymbol{\pi})$ are related to the ranking $F(\boldsymbol{\pi})$ by a relationship. Let us spell out this relationship in each case. The counting method assigns equal weights to the experts, whatever the statements are: $Q(\boldsymbol{\pi}) = (1/m)$. The relationship is trivial, based on the premise that no distinction should be made between experts. The invariant method, by its very definition, assigns weights to experts equal to their scores: $Q(\boldsymbol{\pi}) = F(\boldsymbol{\pi})$. The HITS method assigns weights so that $Q(\boldsymbol{\pi})$ is the normalized vector proportional to $\tilde{\boldsymbol{\pi}} F(\boldsymbol{\pi})$.

[11] $1/\lambda$ is the dominant eigenvalue of $\boldsymbol{\pi}\widetilde{\boldsymbol{\pi}}$ and is not equal to 1 in general even for a normalized $\boldsymbol{\pi}$.

Except for the counting method, the weights typically differ across experts. Furthermore, this differential treatment varies with the whole profile of statements: it is endogenous, determined by the relationship between rankings and weights. The following example illustrates the impact of this relationship. It also shows that the invariant method does not perform well in some situations, as has been recognized by computer scientists (see e.g. Boldi, Santini and Vigna 2007 from which the following example is taken).

Example There are ten items/experts, whose statements are represented by the graph in Figure 1 where an arrow form a node j to i means that j cites i. Perturb these statements by adding a small constant positive α to each element so that all valuations become positive (this is the type of perturbation that is used by PageRank). Intuitively, item 0 should receive a high score since it is cited by 7 experts.



For $\alpha = 10^{-3}$, the invariant method[12] assigns high scores to items 4 and 5, respectively 0.4943 and 0.4939, and small ones to others, in particular to item 0 which receives 0.0049. The HITS method instead assigns a high score to item 0, 0.9152, and small scores to other items. (The handicap-based method, defined in Section 3, produces similar results to the HITS method.)

The reason for why item 0 does not obtain a high score in the invariant ranking is because it cannot obtain a high expert's weight. To see this, observe that 1 cites only 4, and 4 and 5 only cite each other. For α small, this implies $r_5 \approx r_4$ and $r_4 \approx r_1 + r_5$: r_1 must be small (as an expert's weight). But 1 is cited by 0, hence r_1

[12]The invariant scores are $0.0049, 0.0015, 0.0009, 0.0001, 0.4943, 0.4939$ for $i = 0, .., 5$ and 0.0011 for $i = 6..., 9$. The HITS scores are $0.9152, 0.0602, 0.0028, 0.0024, 0.0037, 0.0032$ for for $i = 0, .., 5$ and 0.0031 for $i = 6..., 9$. The handicap-based scores are $0.9595, 0.0064, 0.0018, 0.0001, 0.0248, 0.0066$ for $i = 0, .., 5$ and 0.0002 for $i = 6..., 9$. The HITS (resp.handicap-based) weights are $0.0034, 0.0008, 0.1142, 0.0234, 0.0008, 0.0009$ (resp. $0.0002, 0.0033, 0.0125, 0.0003, 0.0065, 0.0230$) for $i = 0, .., 5$ and 0.2141 (resp. 0.2385) for $i = 6..., 9$.

(as a score) satisfies $r_1 \approx r_0/5$, which implies that r_0 (as an expert's weight) must be small as well. Thus, the bad behavior of the invariant method in this example comes from the identification of the scores, which are related with the citations received (the valuations), with the experts' weights, which are related with the citations made (the statements). ∎

The notion of supporting weights is useful for defining a variety of methods by varying the relationship between rankings and weights. More precisely, given a relationship, the method simultaneously assigns to each statement a ranking and experts' weights so that (a) the ranking is the weighted average of the experts' statements and (b) the ranking and the weights follow the relationship (of course, some conditions on the relationship are required for the method to be well-defined). This is the approach followed in the next section.

3 The handicap-based method

This section first introduces a method based on the notion of handicaps (Proposition 1). This method is defined in all settings, including the peers' ones. Proposition 2 provides two characterizations in terms of the axioms introduced in the previous section and Proposition 3 does the same for the counting method. These axiomatizations show the similarities between the two methods, except for intensity invariance.

In this section, statements are assumed to be all positive. Null entries are considered in Section 4.

3.1 Definition and properties

The purpose of handicaps is to equalize the strengths between items and a handicap may be defined as the inverse of the score: saying that the handicap of i is twice that of ℓ means that the score of i is half that of ℓ. Thus, we assign handicaps $\mathbf{h} = (h_i)$ to a ranking $\mathbf{r}$ by $h_i = 1/r_i$. The handicap-based method is based on an equilibrium relationship between handicaps and experts' weights: it looks for handicaps that equalize items' weighted counts and for experts' weights that equalize the distributed

handicap 'points' across experts:

$$\sum_j (\pi_{i,j} q_j) h_i = 1 \text{ for each } i \text{ where } \sum_i (\pi_{i,j} h_i) q_j = \frac{n}{m} \text{ for each } j. \tag{9}$$

The next proposition states that this leads to a well defined ranking method where the ranking is proportional to the vector $(1/h_i)$.

Proposition 1 *Given a positive matrix* $\boldsymbol{\pi}$*, there is a unique* $\mathbf{r} = (r_i)$ *in* Δ_N *such that*

$$\sum_j (\pi_{i,j} q_j) \frac{1}{r_i} = 1 \textit{ for each } i \textit{ where } \sum_i (\frac{\pi_{i,j}}{r_i}) q_j = \frac{n}{m} \textit{ for each } j. \tag{10}$$

The **handicap-based method** H *assigns to each matrix this unique ranking* $\mathbf{r}$ *and the supporting weights* $\mathbf{q}$*.* H *is intensity-invariant, uniform, exact, and homogeneous.*

The proof is in the appendix.

Let us first discuss the relationship between the experts' weights and their statements. When the experts are unanimous and provide the same statement, $\mathbf{r}$, the handicap-based ranking is $\mathbf{r}$ and all experts' weights are all equal to $1/m$. Thus, it is the diversity in statements that generates differences in experts' weights. When statements differ, the experts whose statements have a high correlation with the handicaps receive a lower weight than those whose statements have a low correlation: weights are decreasing in the correlation between statement and handicap vector (since $\sum_i \pi_{i,j} h_i = \text{cov}(\boldsymbol{\pi}_{.j}, \mathbf{h}) + \sum_i h_i$). Section 3.3 compares this behavior with the weights' behavior of the other methods introduced previously.

The existence of a ranking solution to (10) relies on the following observation. The method can be seen as searching for a ranking and experts' weights that transform the statements into a matrix that is both row- and column-balanced. Specifically, (10) requires the matrix $\boldsymbol{p}$ of general element $p_{i,j} = \frac{1}{r_i} \pi_{i,j} q_j$ obtained from $\boldsymbol{\pi}$ by multiplication of its rows by the items' handicaps and of its columns by the experts' weights to satisfy

$$\sum_j p_{i,j} = 1 \text{ for each } i \text{ and } \sum_i p_{i,j} = n/m \text{ for each } j. \tag{11}$$

The problem of adjusting a given matrix $\boldsymbol{\pi}$ by multiplication of its rows and its columns by some numbers so as to meet constraints on totals is a standard problem

known as matrix scaling.[13] The matrix $\boldsymbol{p}$ is unique. It remains to show that the multipliers are uniquely defined when $\mathbf{r}$ is in the simplex.

The matrix $\boldsymbol{p}$, the handicap-based ranking and the supporting weights can be computed through the iterative scaling algorithm RAS (Bacharach 1965). Let $\boldsymbol{\pi}$ be a positive normalized matrix. The procedure starts by assigning the handicap vector $\mathbf{h}^0$ associated to the counting ranking and equal weights to experts: $h_i^0(\sum_i \pi_{i,j})/m = 1$ for each i and $q_j^0 = 1/m$ each j. Define the handicap points distributed by j as $\sum_i h_i^0 \pi_{i,j}$. If these points are all identical across experts, then the handicap-based ranking is the counting ranking with experts' weights all equal to $1/m$ and the process stops. Otherwise, the handicap points differ and experts' weights $\mathbf{q}^1$ are defined so as to equalize them:

$$(\sum_i h_i^0 \pi_{i,j}) q_j^1 = \frac{n}{m} \text{ for each } j.$$

(The vector $\mathbf{q}^1$ is not necessarily in the simplex.) The items' totals weighted by $\mathbf{q}^1$ are then computed, $\sum_j \pi_{i,j} q_j^1$ for each i, and the handicaps $\mathbf{h}^1$ are defined so as to equalize these totals across items: $h_i^1(\sum_j \pi_{i,j} q_j^1) = 1$ for each i. The procedure starts over again, alternating row-scaling and column-scaling: for each $\tau = 1, ..$

$$h_i^\tau(\sum_j \pi_{i,j} q_j^\tau) = 1 \text{ for each } i \tag{12}$$

$$(\sum_i h_i^\tau \pi_{i,j}) q_j^{\tau+1} = \frac{n}{m} \text{ for each } j. \tag{13}$$

The sequences $\mathbf{h}^\tau$ and $\mathbf{q}^\tau$ can be shown to converge[14] to some positive vectors $\mathbf{h}$ and $\mathbf{q}$ that satisfy (9). Let $\mathbf{r}$ be the ranking associated to handicap $\mathbf{h}$: $r_i = \lambda/h_i$ where λ is chosen to have the sum $\sum r_i$ equal to 1. Since the vectors $(\mathbf{r}, \lambda\mathbf{q})$ satisfy (10) and $\mathbf{r}$ belongs to the simplex, $\mathbf{r}$ is the handicap-based ranking supported by the weights $\lambda\mathbf{q}$.

[13] The problem appears in various areas, in statistics for adjusting contingencies tables, in economics for balancing international trade accounts and for filling missing accounting data, or in voting problems (see for example Balinski and Demange 1989). In these cases, the object of interest is the final adjusted matrix $\boldsymbol{p}$. We are interested instead in the (relative) values of the adjustment on rows and columns so as to define the handicaps and experts' weights.

[14] The convergence is not straightforward, especially when some elements of $\boldsymbol{\pi}$ are null (see the recent survey of Pukelsheim 2012 for example). When all elements of $\boldsymbol{\pi}$ are positive, convergence is ensured at a geometric rate that depends on the final matrix $\boldsymbol{p}$ (Soules 1999).

3.2 Characterization of the handicap-based and the counting methods

The next two propositions provide characterizations of the handicap-based and counting methods.

Proposition 2

(a) The handicap-based method is the only ranking method that is uniform on $\mathcal{R}$, intensity-invariant, and homogeneous.

(b) The handicap-based method is the only ranking method that is exact on $\mathcal{R}$, intensity-invariant, and homogeneous.

Proposition 3

(a) The counting method is the only method that is homogeneous on absolute statements and uniform on $\mathcal{P}$.

(b) The counting method is the only method that is homogeneous on absolute statements and exact on $\mathcal{P}$.

The comparison between Proposition 2 and 3 shows that the main difference between the handicap-based and counting methods stems from intensity-invariance and the domain on which uniformity holds, $\mathcal{R}$ or $\mathcal{P}$. This suggests that the handicap-based method is adequate for relative statements and the counting method for absolute ones.

The ordinal ranking associated to the counting method has been axiomatized in the 'tournament' setting. Rubinstein (1976) considers a simple tournament in which either i wins over j or j wins over i ($\pi_{i,j}$ is 0 or 1 and $\pi_{i,j} + \pi_{j,i} = 1$) and Brink and Gilles (2003) consider the more general 0-1 setting. Both papers rely on an axiom of Independence of Irrelevant Alternatives (IIA). Say that i beats j if $\pi_{i,j} = 1$. IIA requires that the ordering of two items, say i and k, only depend on the items beaten by i or k and on those that beat one of them; thus, the ordering of the two items is determined by their neighbors in the graph representing the matrix $\boldsymbol{\pi}$. Our characterization of the counting method clearly differs since it does not rely on IIA. Finally, Brink and Gilles (2009) consider weighted directed graphs, which correspond to non-negative matrices with null elements on the diagonal. They use an axiom dealing with the sum of matrices, hence their characterization also differs from ours.

Let us add a final remark on IIA. IIA is violated by all the methods considered in this paper, except the counting method. For the three methods, invariant, HITS and handicap-based, the experts' weights vary with the whole statement matrix. Thus, the score of an item depends on the statements over all items via the values taken by the experts' weights; this is also typically true for the ratio of the scores on two items (or their ordering): IIA is violated. In some sense, the aim of these methods is precisely to account for the whole statement matrix to derive expert's weights. In my view, this justifies to give up IIA.

3.3 Comparison between methods

Let us first compare the relationships between the experts' weights and their statements for the various methods. When the experts are unanimous and state the same $\mathbf{r}$, the counting, HITS, handicap-based and invariant ranking all coincide with $\mathbf{r}$. In that case, the experts' weights are equal across experts ($q_j = 1/m$ for each j) for all methods but the invariant ranking, which assigns weights equal to $\mathbf{r}$. As a result, when the statements do not differ much across experts, the counting, HITS and handicap-based methods produce close rankings but not the invariant method. Specifically, Let $\pi_{i,j} = r_i + \epsilon_{i,j}$ where $\sum_j \epsilon_{i,j} = 0$ and $\epsilon_{i,j}$ are small enough to have $\pi_{i,j} > 0$. i's score writes $\sum_j q_j(\boldsymbol{\pi})\pi_{i,j} = \frac{1}{m}\sum_j \pi_{i,j} + \sum_j (q_j(\boldsymbol{\pi}) - 1/m)\pi_{i,j}$. The weights for the handicap-based method are continuous so that i's score is approximated by $\frac{1}{m}\sum_j \pi_{i,j}$ for $\epsilon_{i,j}$ small enough. The same approximation holds for the HITS method. As for the invariant ranking, a similar argument shows that i's score is approximated by $\sum r_i \pi_{i,j}$.

The handicap-based and HITS methods differ in the way experts' weights depend on a given ranking $\mathbf{r}$, as shown by the expressions (10) and (7). For the former, the weight of an expert is the harmonic mean of $\mathbf{r}$ weighted by the expert's statement, whereas for the HITS (up to a multiplicative factor), the weight is its average mean. This difference in the way the weights relate to a given ranking induces a difference in the final rankings, which is not easy to assess. Providing an axiomatization of the HITS method would help us to understand better the differences between the two methods.

Homogeneity: An example The invariant method is intensity-invariant and exact but not homogeneous. The following example shows that its behavior with respect to changes in statements may present serious drawbacks in some contexts. Let $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ be

$$\boldsymbol{\pi} = \begin{pmatrix} 0 & 1/3 & 1 \\ 1 & 0 & 1/3 \\ 1/3 & 1 & 0 \end{pmatrix} \quad \boldsymbol{\pi}' = \begin{pmatrix} 0 & 10/19 & 10/13 \\ 9/10 & 0 & 3/13 \\ 1/10 & 9/19 & 0 \end{pmatrix},$$

(Diagonal elements are null so as to show that the described behavior is not due to self-citations.) Comparing the statements in the two matrices, those for 1 relative to 2 are 10/9 times larger in $\boldsymbol{\pi}'$ than in $\boldsymbol{\pi}$, and those for 2 relative to 3 are 3 times larger. That is, $\boldsymbol{\pi}' = [dg(\boldsymbol{\rho})\boldsymbol{\pi}]$ for $\boldsymbol{\rho} = (10/9, 1, 1/3)$. Since $\boldsymbol{\pi}$ is balanced, both the invariant and handicap-based methods assign $(1/3, 1/3, 1/3)$ to $\boldsymbol{\pi}$. For $\boldsymbol{\pi}'$, the handicap-based ranking is proportional to $(10/9, 1, 1/3)$ by homogeneity and the invariant ranking is (approximately) $(0.38, 0.395, 0.225)$ to $\boldsymbol{\pi}'$. Thus, although the statements on 1 relative to 2 increase from $\boldsymbol{\pi}$ to $\boldsymbol{\pi}'$, each one multiplied by $\rho_1 = 10/9$, the invariant score of 2 becomes larger than 1's in $\boldsymbol{\pi}'$.

This non-monotone behavior is explained by the dual role of the scores in the invariant method and the induced loopback effect. Item 1 receives a high valuation from expert 3 at $\boldsymbol{\pi}$. As item 3 gets less support (the scaling down by $\rho_3 = 1/3$), its score is decreased, hence its weight as well. As a result, the high valuation from 3 counts less and this produces a negative effect on item 1's score that counterbalances the direct positive effect due to ρ_1. As for item 2, it receives a high valuation from 1 at $\boldsymbol{\pi}$. So it benefits indirectly from both the increase in 1's valuations and the decrease in 3's, since they result in an increase in 1's score and a decrease in 3's score. This explains why the score of 2 not only increases but ends up larger than 1's score.

Finally, note that 1's invariant score may end up larger than 2's if we choose different values for ρ_1 and ρ_3 (still respectively larger and smaller than 1). As should be clear from the above argument, there are effects possibly in opposite directions due to the double role of the scores in the invariant method. The final order of the scores depends on the relative intensity of these various effects, which, in turn, depends on the matrix and the values of the ρ_i.

3.4 Illustration: Journals Rankings

This section illustrates the differences between the invariant, the handicap-based and the HITS methods for the rankings of 37 journals, using the same data as in Palacios-Huerta and Volij (2004). Scores and weights are given per article. Let $\pi_{i,j} = \frac{C_{i,j}}{C_{+,j}}$ be the total share of citations sent by (all articles of) j received by (all articles of) i and n_i denote the number of articles in journal i. The invariant ranking $\mathbf{r}$ per article satisfies

$$r_i n_i = \sum_j \pi_{i,j} r_j n_j \text{ for each } i$$

and the handicap-based scores and weights $\mathbf{r}$ and $\mathbf{q}$ satisfy

$$n_i r_i = \sum_j \pi_{i,j} n_j q_j \text{ for each } i \text{ and } \frac{1}{n_j q_j} = \sum_i \frac{\pi_{i,j} n_i}{r_i} \text{ for each } j,$$

i.e. $(n_i r_i)$ and $(n_j q_j)$ are the scores and weights of journals.

Table 1 reports the handicap-based, HITS, and invariant rankings in the three first columns, the ratio of the handicap-based score over the invariant one in the fourth, the weights for the handicap-based and HITS methods in the fifth and sixth and finally the correlation between the overall citations with a journal's citations in the seventh column (i.e. the correlation between $\sum_j \mathbf{C}_{.,j}$ and $\mathbf{C}_{.,j}$). Rankings are normalized with a constant sum for $\sum_i r_i n_i$ instead of a constant sum for $\sum_i r_i$. Since the weights satisfy $\sum_i r_i n_i = \sum_i q_i n_i$, such a normalization gives the same total to the weights for each method. The results are presented so that $\sum_i r_i n_i = 100a$ where $a = (\sum n_i)/37$ is the average number of articles per journal, which yields an average score $(\sum_i r_i n_i)/(\sum n_i)$ equal to 2.7.

The weights produced by the handicap-based and HITS methods are similar and convey relevant information. Interestingly *Journal of Economic Literature* (JEL) has by far the largest weight, which supports the meaningfulness of the methods, followed by *Review of Economic Studies* (RES), *Quarterly Journal of Economics* (QJE), and *Rand Journal of Economics* (*Rand*). As an illustration of the fact that weights pertain to a different property than the scores, the handicap-based score is roughly equal to its weight for JEL, is much lower for *International Journal of Game Theory* or *Economic Inquiry*, and much larger for AER. The rather low weight of AER, little less than the average under both the handicap-based and the HITS methods, suggests that AER tends to refer no more than the average to the top journals (see the discussion on

	HB	HITS	Inv	HB/Inv	$\mathbf{q}_{HB}$	$\mathbf{q}_{HITS}$	Corr
Quarterly J. of Economics	10.018	9.7381	11.4412	0.8756	5.3430	4.6379	0.7272
Econometrica	9.6496	9.7683	11.7444	0.8216	3.4210	3.4836	0.6771
J. of Economic Literature	9.6480	9.4887	9.2594	1.0420	8.1585	8.4002	0.6116
J. of Political Economy	7.2226	7.0028	7.5573	0.9557	4.1462	3.7403	0.8151
American Economics Review	7.0113	6.9074	7.5236	0.9319	2.4798	2.5582	0.7873
Review of Economic Studies	5.9858	5.8574	7.4224	0.8064	5.5907	5.4340	0.7224
J. of Monetary Economics	4.2653	4.1510	5.2428	0.8136	4.3299	3.9331	0.5466
J. of Economics Perspectives	3.6309	3.4951	3.6507	0.9946	3.6657	3.3049	0.5460
J. of Economic Theory	3.2109	3.3893	4.0030	0.8021	1.8915	1.9902	0.4856
Rand J. of Economics	3.1024	2.9416	2.2646	1.3700	4.8855	4.3174	0.4445
Games and Economic Behavior	3.0573	3.3397	3.7402	0.8174	2.5863	2.7514	0.4168
J. of Financial Economics	2.9606	2.4102	1.7404	1.7010	3.7100	2.8174	0.1117
J. of Econometrics	2.9066	2.8103	2.4406	1.1909	1.8608	1.7978	0.3742
J. of Labor Economics	2.8038	2.8306	2.1191	1.3231	3.8302	4.3557	0.4292
J. of Human Resources	2.5879	2.6167	2.0470	1.2643	3.4522	3.3428	0.2456
J. of Risk and Uncertainty	2.4000	2.5413	1.8740	1.2807	3.5808	3.8032	-0.0186
J. of Env. Econ. and Management	2.3330	2.2943	1.4148	1.6490	3.1345	2.9811	0.1441
Econometric Theory	2.3312	2.2006	1.8501	1.2601	3.2729	2.7537	0.2903
J. of Business and Ec. Stat.	2.1956	2.2257	1.6560	1.3258	3.0493	3.1691	0.4229
Review of Economics and Stat.	2.0609	2.0573	1.8967	1.0865	2.3898	2.3346	0.6789
Social Choice and Welfare	2.0237	2.3635	1.4593	1.3868	3.2979	3.8801	0.1242
J. of Public Economics	2.0046	2.0367	1.8640	1.0754	2.2278	2.1119	0.6344
J. of International Economics	1.8998	1.9417	1.3240	1.4349	3.0395	3.2132	0.3459
International Economic Review	1.8632	1.8813	1.8004	1.0349	3.4277	3.3640	0.8604
J. of Applied Econometrics	1.7499	1.8086	1.4695	1.1909	3.6414	4.0607	0.3097
Economic Journal	1.7492	1.7849	1.3943	1.2546	1.8269	1.7183	0.5090
AER Proceedings	1.6863	1.6471	1.6315	1.0336	2.1315	1.9946	0.5815
European Economic Review	1.6466	1.6414	1.5101	1.0904	1.8436	1.6745	0.7475
Int. J. of Game Theory	1.4510	1.7037	1.4368	1.0099	4.1551	5.3043	0.2916
Economic Theory	1.4370	1.4534	2.1151	0.6794	2.1391	2.3458	0.4543
J. of Ec. Dyn. and Control	1.1781	1.2508	1.2210	0.9649	2.0642	2.2980	0.4078
J. of Mathematical Economics	1.0797	1.2176	1.1731	0.9204	1.9834	2.3951	0.0467
Economic Inquiry	0.8495	0.8886	0.7209	1.1784	3.0400	3.1831	0.6491
J. of Econ. Behavior and Org.	0.8029	0.8895	0.5882	1.3651	1.7462	2.1252	0.4844
Scandinavian J. of Economics	0.7770	0.8062	0.4717	1.6470	3.8152	4.3015	0.6248
Oxford Bull. of Ecs. and Stat.	0.6922	0.7536	0.3138	2.2058	2.8363	3.3401	0.2086
Economics Letters	0.4573	0.4739	0.3554	1.2866	0.6645	0.6700	0.4839

Table 1: Rankings and weights per article. HB= handicap-based, Inv= invariant

the weights in section 3.1). Finally, the weights produced by the handicap-based and HITS methods differ significantly from the invariant weights (which are the invariant scores) and are not predicted by the correlations with the overall citations, as can be seen from the last column (this can also be expected from section 3.1).

As for the rankings, the top six journals remain the same without any ambiguity but the order varies depending on the used method. For instance, the order between the two first journals (QJE and *Econometrica*) is reversed when comparing the handicap-based ranking to the HITS and invariant rankings. However, since their scores are very close in each ranking, ordering amplifies small differences in scores (which elicits the benefit of considering a cardinal ranking). Apart from JEL, the scores of the five other journals in the top six are lower than their invariant scores, with the largest decrease realized for RES. The main intuition for this decrease is that these journals receive proportionately more citations from top journals than the average. Since the weights of top journals are typically lower than their invariant scores, this explains the decrease in their score. JEL instead has a more disperse scope of citations (again, recall that this is not related to the fact that JEL has a large weight). This type of argument also explains why the scores of most theory journals (except *Journal of Mathematical Economics* and *Social of Choice and Welfare* (SCW)) decrease because they receive proportionately more citations from top journals. The journals with the largest handicap-based score relative to their invariant score are *Journal of Financial Economics*, *Rand* , *Oxford Bulletin of Economics and Statistics*, and SCW.

In my view, one should not pay too much attention to the differences in these rankings; however the weights, as computed by the handicap-based or the HITS method, convey interesting information.

4 Extending the handicap-method to non-negative matrices

The handicap-based method has been defined in Section 3 for positive statements. In some settings however, experts are allowed to assign a zero to an item, meaning that they find this item not acceptable, or 'infinitely below' an item with a positive grade.

The handicap-based method cannot be extended to all non-negative matrices. This section characterizes the matrices for which a handicap-based ranking is well defined, or 'exists', meaning that there is a unique ranking that satisfies (10). Uniqueness is important not only to have a well-defined ranking but also because it relates to the continuity of the method: without uniqueness, perturbing a matrix by replacing non-zeros by small but positive elements produces different answers depending on the perturbation. Similarly, the invariant and the HITS rankings are not well-defined for all non-negative matrices. I discuss the differences in the conditions for existence across the methods.

4.1 A characterization

Given a statement matrix $\boldsymbol{\pi}$, $\pi_{i,j} \geq 0$, we assume that $\boldsymbol{\pi}$ has no null row nor null column (these could be deleted). Let us introduce some notation. Let $I(j) = \{i|\pi_{i,j} > 0\}$ denote the set of items cited by j. For J subset of M, $I(J) = \cup_{j \in J} I(j)$ is the set of items cited by at least one expert in J. Consider the bipartite graph G with sets of nodes N and M where (i, j) is an edge if $\pi_{i,j}$ is positive. In the peers' setting in which the items are also the experts, the two sets N and M are replica of each other so that the bipartite graph distinguishes the two roles as item as expert for each element. G is *items-connected* if each pair of items is linked by a path: for each pair of items i, ℓ, there is a sequence alternating items and experts, starting at i and ending at ℓ: $i = i_0, j_0, i_1, \ldots, i_t, j_t, \ldots, j_k, i_{k+1} = \ell$ such that each expert j_t in the sequence cites the two adjacent items i_t and i_{t+1}. When G is *items-connected*, G is *connected* as well: since each expert assigns a positive valuation to one item at least, there is a path between any two elements, item or expert.[15]

The following proposition characterizes statement matrices for which the handicap-based ranking is defined in a unique way.

Proposition 4 *Given a non-negative matrix* $\boldsymbol{\pi}$*, there is a unique ranking* $\mathbf{r} = (r_i)$ *in*

[15] To obtain a path between an expert and an item i, take a link between the expert and an item that is cited by the expert and add the path linking that item to i (if the items differ). The same argument applies for finding a path between two experts.

Δ_N *such that (10) holds:*

$$\sum_j (\pi_{i,j} q_j) \frac{1}{r_i} = 1 \textit{ for each } i \textit{ where } \sum_i (\frac{\pi_{i,j}}{r_i}) q_j = \frac{n}{m} \textit{ for each } j$$

if and only if the bipartite graph G associated to $\boldsymbol{\pi}$ is items-connected and

$$\frac{1}{m}|J| < \frac{1}{n}|I(J)| \textit{ for any strict subset } J \textit{ of } M. \tag{14}$$

The proof is in the appendix. The items-connectedness of graph G ensures the uniqueness of the ranking. This is a natural condition. If G is not items-connected, then there are two disjoint sets of experts who cite two disjoint sets of items: these experts do not share any common interest. In such a situation, it makes sense that no unique ranking can reflect the statements of all experts.

To get an intuition for why conditions (14) are necessary, observe that the matrix $\boldsymbol{p}$ defined by $p_{i,j} = \frac{1}{r_i}\pi_{i,j} q_j$ is balanced and has exactly the same null cells as $\boldsymbol{\pi}$. The existence of such a matrix can be put as a flow problem in the bipartite graph G: there must exist a flow in G such that each j in M sends $1/m$ units and each node i in N receives $1/n$ units, and, in addition, the flow is positive on each link. Relaxing this positivity requirement, the existence of such a flow is solved by well-known supply-demand conditions, which are the weak version of the inequalities (14). Assuming these inequalities to hold strictly ensures that the flow can be made positive on each link.

Conditions (14) when $n = m$ require that the number of items cited by a given subgroup of experts should exceed the number of these experts: cites should be sufficiently 'disseminated'. When $n \neq m$ the conditions bear on the proportions of each subset in N or M. Conditions (14) also require the dual property that the proportion of experts who cite a given subset of items should exceed the proportion of these items in the whole set N: taking the complements to the subsets in N and M, (14) write[16]

$$\frac{1}{n}|I| < \frac{1}{m}|J(I)| \text{ for any strict subset } I \text{ of } N. \tag{15}$$

[16]Given a strict subset I, let $J = M - J(I)$. If J is empty, (15) is met. It not, apply (14) to J: $\frac{1}{m}(m - |J(I)|) < \frac{1}{n}|I(J(I))|$. By definition, the experts cited by $M - J(I)$ do not cite any items in I. So $I(J(I)) \subset N - I$, which implies $|I(J(I))| \leq n - |I|$, and finally $|I| < \frac{n}{m}|J(I)|$.

These conditions for the existence of a unique handicap-based ranking are strong, which can be understood as follows. Consider perturbations on a statement matrix $\boldsymbol{\pi}$ that transform null valuations into small but positive elements. When the handicap-based ranking of $\boldsymbol{\pi}$ is well-defined, all perturbations lead to the same ranking. However, perturbing a null valuation into a small but positive element changes an expert's statement substantially when a null valuation means 'non-acceptable'. This interpretation is reflected by the homogeneity axiom: multiplying the valuations on an item by some factor leaves unchanged those which are null. This explains why the conditions for the existence of a handicap-based ranking are rather strong: they require a large enough overlap over the items that are considered as acceptable by the experts.

4.2 A comparison with the invariant and the HITS methods

As for the handicap-based method, the invariant and the HITS methods cannot be extended in a unique way to all non-negative matrices. The matrix $\boldsymbol{\pi}$ or $\boldsymbol{\pi}\widetilde{\boldsymbol{\pi}}$ should admit a unique largest eigenvalue, as is ensured by its irreducibility[17], thanks to Perron-Frobenius theorem. Let us comparing these conditions with those in Proposition 4.

Let us consider first the HITS method. We first note that the items-connectedness of G is equivalent to the irreducibility of $\boldsymbol{\pi}\widetilde{\boldsymbol{\pi}}$. To see this, note that the element (i, ℓ) of $\mathbf{a} = \boldsymbol{\pi}\widetilde{\boldsymbol{\pi}}$, $\sum_j \pi_{i,j}\pi_{\ell,j}$, is positive if and only if there is an expert who cites both i and ℓ. Thus $a_{i,\ell}^{(t)}$ is positive if there is a path with t experts linking i to ℓ in the bipartite graph G. Hence the irreducibility of $\boldsymbol{\pi}\widetilde{\boldsymbol{\pi}}$ is equivalent to the existence of a path between any two items, i.e. equivalent to the items-connectedness of G. Thus, the items-connectedness of G ensures that the HITS ranking is uniquely defined; furthermore , as we have seen, G is then also experts-connected so that the matrix $\widetilde{\boldsymbol{\pi}}\boldsymbol{\pi}$ is irreducible: the experts' weights are uniquely defined as well. Thus, the conditions on the statements for the existence of a unique handicap-based ranking are stronger than for the HITS ranking. This can be traced back to the interpretation of a zero and the homogeneity axiom, as we have just discussed above. The HITS is not

[17]Recall that a non-negative square matrix $\mathbf{a}$ is irreducible if for each pair (i, ℓ) there is an integer t such that the (i, ℓ) element of the t-product matrix $\mathbf{a}^{(t)}$ is positive.

homogeneous, hence perturbing a zero valuation into a small but positive element does not involve a drastic change in an expert's statement. As a result, there are more chances for the HITS method to be continuous at a non-negative statement matrix.

For the invariant method, the ranking is uniquely defined when the matrix $\boldsymbol{\pi}$ is irreducible. This condition differs from the irreducibility of $\boldsymbol{\pi}\widetilde{\boldsymbol{\pi}}$, as shown by the following example. There are three items/experts; 1 cites 2, 2 cites 3, and 3 cites 1. The matrix[18] $\boldsymbol{\pi}$ is irreducible but $\boldsymbol{\pi}\widetilde{\boldsymbol{\pi}}$ is null because each expert cites a different item. Thus, we find again that the invariant method behaves quite differently from the HITS and handicap-based methods.

To end this section, let us make about the Web setting. As already said, the homogeneity axiom does not apply to the 0-1 setting.[19] Furthermore page j may not point to page i, represented by a null valuation, because j finds i 'non-acceptable' or simply because j does not know i. Nevertheless, it would be interesting to investigate how the handicap-based method behaves on Internet data using the same perturbation technique as PageRank.

5 Concluding remarks

This paper has introduced and characterized the handicap-based ranking method. This method is adequate in environments where either the 'intensity' of statements is not controlled and intensity-invariance is required, or statements are relative evaluations (for example when individuals express their preferences as to how a budget should be allocated between various issues). The handicap-based method is, in a sense that has been made precise in the paper, the counterpart to the counting method in these environments. Furthermore, it applies to a variety of settings. In particular, it is not restricted to peers' settings where items and experts coincide, as is the case for the invariant method. Finally, even in these peers' settings, two indices are assigned

[18] $\boldsymbol{\pi} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

[19] The same remark applies to some of the axioms used by Palacios-Huerta and Volij (2004) and Slutzki and Volij (2006).

to each item -a score and a weight- thereby providing more information than a single ranking.

Several developments are worth investigating. Firstly, it can be fruitful to analyze in a systematic way the ranking methods that assign simultaneously scores to items and weights to experts. As the weights reflect how the differences across the statements of different experts are interpreted, such an approach may provide a useful tool for deriving new methods. Secondly, in environments where many valuations are null, most methods need to perturb the data in order to be applied. This raises questions concerning the robustness of the outcome to such perturbations. Specifically, without a minimum of agreement among the experts about the acceptable items, the outcome may be sensitive to the perturbation. Alternatives to perturbation techniques should be investigated; for example the set of experts could be determined endogenously (in a peers' context, this set could differ from the set of items to rank). Thirdly, the analysis of rankings in a dynamical setting and the extent of their influence are important topics that need to be explored.[20]

6 Proofs

Proof of Lemma 1 Let F be intensity-invariant:

$$F([\boldsymbol{\pi}]) = F(\boldsymbol{\pi}) \text{ and } F([dg(\boldsymbol{\rho})\boldsymbol{\pi}]) = F(dg(\boldsymbol{\rho})\boldsymbol{\pi}). \tag{16}$$

Let F be homogeneous on absolute statements: $F(dg(\boldsymbol{\rho})\boldsymbol{\pi})$ is proportional to $dg(\boldsymbol{\rho})F(\boldsymbol{\pi})$ for any $\boldsymbol{\pi}$ in $\mathcal{P}$, in particular for $\boldsymbol{\pi}$ in $\mathcal{R}$. Thus $F([dg(\boldsymbol{\rho})\boldsymbol{\pi}])$, which is equal to $F(dg(\boldsymbol{\rho})\boldsymbol{\pi})$, is proportional to $dg(\boldsymbol{\rho})F(\boldsymbol{\pi})$: F is homogeneous on relative statements.

To show the converse, let F be homogeneous on relative statements and $\boldsymbol{\pi}$ be in $\mathcal{P}$. Observe that $[dg(\boldsymbol{\rho})\boldsymbol{\pi}] = [dg(\boldsymbol{\rho})[\boldsymbol{\pi}]]$. Hence by (16) $F(dg(\boldsymbol{\rho})\boldsymbol{\pi}) = F([dg(\boldsymbol{\rho})[\boldsymbol{\pi}]])$. Since $[\boldsymbol{\pi}]$ is in $\mathcal{R}$, homogeneity on $\mathcal{R}$ implies that $F([dg(\boldsymbol{\rho})[\boldsymbol{\pi}]])$ is proportional to $dg(\boldsymbol{\rho})F([\boldsymbol{\pi}])$. Using $F([\boldsymbol{\pi}]) = F(\boldsymbol{\pi})$, we finally obtain that $F(dg(\boldsymbol{\rho})\boldsymbol{\pi})$ is proportional to $dg(\boldsymbol{\rho})F(\boldsymbol{\pi})$. This proves that F is homogeneous on absolute statements. ∎

Proof of Proposition 1. The proof of existence and uniqueness of H involves two steps.

[20]For some studies on the subject, see Demange (2012) and (2013).

The first step shows the existence of $\mathbf{r}$ in the simplex that satisfies (10). As said in the text, this is equivalent to the fact that matrix $\boldsymbol{p}$ with general element $p_{i,j} = \frac{1}{r_i}\pi_{i,j}q_j$ is $(1,\ n/m)$-balanced. The proof relies on the known result about matrix scaling: there is a unique $(1,\ n/m)$-balanced matrix $\boldsymbol{p}$ that is obtained from $\boldsymbol{\pi}$ by multiplication of its rows and its columns by some numbers. A simple proof relies on a convex program. We recall the argument here for positive matrices. Consider the program (ln denotes the natural logarithm):

$$\begin{aligned} \mathcal{P} : \underset{\boldsymbol{p}}{\text{minimize}} \quad & \sum_{i,j} p_{i,j}[\ln(\pi_{i,j}/p_{i,j}) - 1] \text{ over the } \boldsymbol{p} = (p_{i,j}) > 0 \\ \text{subject to} \quad & \sum_j p_{i,j} = 1 \text{ for each } i \text{ and } \sum_i p_{i,j} = n/m \text{ for each } j. \quad (11) \end{aligned}$$

The program is convex with a strictly convex objective function and a feasible set with a non empty relative interior. Hence the solution $\boldsymbol{p}$ is unique, characterized by the first order conditions on the Lagrangian: There are multipliers α_i and β_j associated respectively to the constraints (11) on the totals of row i and column j such that

$$\ln(\pi_{i,j}) - \ln(p_{i,j}) = \alpha_i + \beta_j. \tag{17}$$

Set $r_i = \exp(\alpha_i)$ and $q_j = \exp(-\beta_j)$, (17) writes $p_{i,j} = \frac{1}{r_i}\pi_{i,j}q_j$. Plugging these expressions into the constraints (11) give (10). It remains to show that $\mathbf{r}$ can be chosen to be in the simplex. The multipliers $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are defined up to an additive constant, i.e., $(\alpha_i + c)$, $(\beta_j - c)$ satisfy (17) for any c if $\boldsymbol{\alpha}, \boldsymbol{\beta}$ does (this is due to the fact the linear system (11) is formed with linearly dependent equations). Thus there is a vector $\mathbf{r}$, $r_i = \exp(\alpha_i + c)$, that belongs to the simplex for an appropriate value of c.

The second step proves that $\mathbf{r}$ is unique. Given $\mathbf{r}$ that satisfies (10), the matrix $\boldsymbol{p}$ defined by $p_{i,j} = \frac{1}{r_i}\pi_{i,j}q_j$ satisfies the constraints (11). Furthermore, taking the log of $p_{i,j} = \frac{1}{r_i}\pi_{i,j}q_j$, the first order conditions (17) are met for $\alpha_i = \ln(r_i), \beta_j = -\ln(q_j)$. Hence $\boldsymbol{p}$ is the unique solution to $\mathcal{P}$. So if there are two rankings that satisfy (10), $\mathbf{r}$ and $\mathbf{r}'$, the corresponding values (α_i) (β_j) and (α'_i) (β'_j) satisfy (17) for the same matrix $\boldsymbol{p}$. Taking the difference yields

$$(\alpha_i - \alpha'_i) + (\beta_j - \beta'_j) = 0 \text{ for each } i, j. \tag{18}$$

Hence $\alpha'_i = \alpha_i + c$ for some scalar c. There can be only one value for c so that $\mathbf{r}$ defined by $r_i = \exp(\alpha_i)$ belongs to Δ_N. This proves that H is a well defined method. ∎

Let us now prove the properties of H.

H is intensity-invariant. Let $\boldsymbol{\pi}'$ be obtained from $\boldsymbol{\pi}$ by multiplying column j of $\boldsymbol{\pi}$ by μ_j. Letting $\mathbf{q}'$ be the vector obtained from $\mathbf{q}$ by dividing q_j by μ_j, the vectors $\mathbf{r}$ and $\mathbf{q}'$ satisfy (10) for $\boldsymbol{\pi}'$. By the uniqueness result proved previously, $H(\boldsymbol{\pi}')$ is equal to $\mathbf{r}$.

H is uniform on $\mathcal{R}$. Let matrix $\boldsymbol{\pi}$ in $\mathcal{R}$ be row-balanced. It satisfies $\sum_j \pi_{i,j} = m/n$ for each i and $\sum_i \pi_{i,j} = 1$ for each j. Hence, the conditions

$$\sum_j (\pi_{i,j} q_j) \frac{1}{r_i} = 1 \text{ for each } i \text{ where } \sum_i (\frac{\pi_{i,j}}{r_i}) q_j = \frac{n}{m} \text{ for each } j \quad (10)$$

are satisfied by taking equal scores and equal experts' weights, that is, each r_i equals to $1/n$ and each q_j equals to $1/m$. (Equivalently $\boldsymbol{p} = \boldsymbol{\pi}\frac{n}{m}$ satisfies (11).) Hence, by the uniqueness result, $H(\boldsymbol{\pi}) = \mathbf{e}_N$.

H is exact on $\mathcal{R}$. Let $H(\boldsymbol{\pi}) = \mathbf{e}_N$ for a matrix $\boldsymbol{\pi}$ in $\mathcal{R}$. We need to show that $\boldsymbol{\pi}$ is row-balanced. Applying $\mathbf{r} = \mathbf{e}_N$ to the second equation in (10) yields that the weight vector $\mathbf{q}$ satisfies $(\sum_i \pi_{i,j}) q_j = 1/m$ for each j. This implies $q_j = 1/m$ since $\boldsymbol{\pi}$ is in $\mathcal{R}$. Plugging $r_i = 1/n$ and $q_j = 1/m$ for each i, j into the first set of equations of (10) we obtain that each row's total is equal to m/n: the matrix $\boldsymbol{\pi}$ is row-balanced.

H is homogeneous. Let $\boldsymbol{\pi}' = dg(\boldsymbol{\rho})\boldsymbol{\pi}$ for a positive vector $\boldsymbol{\rho}$. By the definition of H, we have $\pi_{i,j} = \frac{r_i p_{i,j}}{q_j}$ where $\boldsymbol{p}$ is a $(1, n/m)$-balanced matrix. Hence $\pi'_{i,j} = \rho_i \pi_{i,j} = \frac{\rho_i r_i p_{i,j}}{q_j}$ so that $\boldsymbol{p}$ is obtained from $\boldsymbol{\pi}'$ by dividing each row i by $\rho_i r_i$ and multiplying each column j by q_j. By the uniqueness result stated in the first part of Proposition 1, this implies that the handicap-based ranking $\mathbf{r}'$ associated to $\boldsymbol{\pi}'$ is the vector in Δ_N proportional to $(\rho_i r_i)$. Thus, $H(\boldsymbol{\pi}') = (\frac{\rho_i r_i}{\sum_i \rho_i r_i})$, namely multiplying the rows of $\boldsymbol{\pi}$ by some vector multiplies the scores in the same proportions: this proves homogeneity. ∎

Proof of Proposition 2.

Proof of (a). Let method F be uniform on $\mathcal{R}$, intensity-invariant and homogeneous.

F and H coincide if they coincide on the set $\mathcal{R}$ of relative statement matrices since both methods are intensity-invariant. Given $\boldsymbol{\pi}$ consider $\mathbf{r} = H(\boldsymbol{\pi})$, $\mathbf{q}$, and the balanced matrix $\boldsymbol{p}$ associated by the handicap-based method, $p_{i,j} = \frac{q_j \pi_{i,j}}{r_i}$. By uniformity on $\mathcal{R}$, $F(\boldsymbol{p}) = \mathbf{e}_N$. Let $\boldsymbol{p}' = dg(\mathbf{r})\boldsymbol{p}$. Homogeneity implies $F(\boldsymbol{p}') = \mathbf{r}$. The normalized matrix of $\boldsymbol{p}'$ is $\boldsymbol{\pi}$: since matrix $\boldsymbol{\pi}$ is in $\mathcal{R}$, $\sum_i \pi_{i,j} = 1$ for each j, which

writes $\sum_i p_{i,j} r_i = q_j$ thus $[\boldsymbol{p}'] = \boldsymbol{\pi}$. By intensity-invariance $F([\boldsymbol{p}']) = F(\boldsymbol{p}') = \mathbf{r}$, which finally gives $F(\boldsymbol{\pi}) = H(\boldsymbol{\pi})$, the desired result.

Proof of (b). Let method F be exact on $\mathcal{R}$, intensity-invariant and homogeneous.

Given $\mathbf{r} = F(\boldsymbol{\pi})$, divide each row i by r_i so as to obtain matrix $\boldsymbol{\pi}' = dg(1/r_1, \ldots, 1/r_n)\boldsymbol{\pi}$. Homogeneity implies that the scores are equalized: $F(\boldsymbol{\pi}') = \mathbf{e}_N$ thanks to Lemma 1. By intensity-invariance of F we have $F([\boldsymbol{\pi}']) = F(\boldsymbol{\pi}') = \mathbf{e}_N$. Now, exactness on $\mathcal{R}$ implies that $[\boldsymbol{\pi}']$ is row-balanced, hence each row sums to m/n. Since $[\pi']_{i,j} = \frac{\pi_{i,j}}{r_i \sum_\ell \pi_{\ell,j}/r_\ell}$, this writes

$$\sum_j \frac{\pi_{i,j}}{r_i \sum_\ell \pi_{\ell,j}/r_\ell} = \frac{m}{n} \text{ for each } i,$$

$$\text{or } r_i = \sum_j \pi_{i,j} q_j \text{ for each } i, \text{ where } \frac{1}{q_j} = \frac{m}{n} \sum_\ell \frac{\pi_{\ell,j}}{r_\ell} \text{ for each } j.$$

Thus $\mathbf{r}$ and $\mathbf{q}$ satisfy (10). Since $\mathbf{r}$ is in Δ_N, $\mathbf{r}$ is equal to $H(\boldsymbol{\pi})$, the desired property. ∎

Proof of Proposition 3.

Proof of (a). Clearly the counting method is homogeneous on absolute statements and uniform on $\mathcal{P}$. To show the reverse, let method F satisfies these properties. Given a matrix $\boldsymbol{\pi}$, divide each row i by its total $\pi_{i+} = \sum_{j \in M} \pi_{i,j}$ and denote by $\boldsymbol{\pi}'$ the obtained matrix: $\boldsymbol{\pi}' = dg(1/\pi_{1+}, \cdots, 1/\pi_{n+})\boldsymbol{\pi}$. By the a-homogeneity of F, the ranking assigned by F to $\boldsymbol{\pi}'$ is obtained by dividing each component i of $F(\boldsymbol{\pi})$ by π_{i+} and normalizing: for some positive λ, $F_i(\boldsymbol{\pi}') = \lambda F_i(\boldsymbol{\pi})/\pi_{i+}$ for each i. Since $\boldsymbol{\pi}'$ is row-balanced $F(\boldsymbol{\pi}') = \mathbf{e}_N$. This yields that $F_i(\boldsymbol{\pi})/\pi_{i+}$ is constant across i: $F(\boldsymbol{\pi})$ is the counting ranking of $\boldsymbol{\pi}$, the desired result.

Proof of (b). Clearly the counting method is homogeneous on absolute statements and exact on $\mathcal{P}$. To show the reverse, let method F satisfy these properties. Given $\mathbf{r} = F(\boldsymbol{\pi})$, divide each row i by r_i so as to obtain matrix $\boldsymbol{\pi}' = dg(1/r_1, \ldots, 1/r_n)\boldsymbol{\pi}$. F is a-homogeneity implies that the scores are equalized: $F(\boldsymbol{\pi}') = \mathbf{e}_N$ thanks to Lemma 1. Exactness implies that $\boldsymbol{\pi}'$ is row-balanced: for some positive λ, $\sum_j \pi'_{i,j} = \lambda$ for each i. Hence $\sum_j \pi_{i,j} = \lambda r_i$ for each i: $\mathbf{r}$ is the counting ranking of $\boldsymbol{\pi}$, the desired result. ∎

Proof of Proposition 4 Under the stated conditions, the proof of Proposition 1 extends to a matrix $\boldsymbol{\pi}$ with some null elements as follows. The objective in $\mathcal{P}$ takes

the sum over the (i,j) for which $\pi_{i,j}$ is positive, namely the (i,j) in G:

$$\begin{aligned} \mathcal{P} : \underset{\boldsymbol{p}}{\text{minimize}} \quad & \sum_{i,j \in G} p_{i,j}[\ln(\pi_{i,j}/p_{i,j}) - 1] \text{ over the } \boldsymbol{p} = (p_{i,j}) \geq 0 \\ \text{subject to} \quad & \sum_j p_{i,j} = 1 \text{ for each } i \text{ and } \sum_i p_{i,j} = n/m \text{ for each } j. \quad (11) \end{aligned}$$

Under conditions (14), the feasible set defined by (11) has a nonempty interior. Hence the Kuhn and Tucker theorem applies. A solution $\boldsymbol{p}$ is associated to multipliers α_i and β_j such that

$$\ln(\pi_{i,j}) - \ln(p_{i,j}) = \alpha_i + \beta_j, \text{ for each } (i,j) \in G \tag{19}$$

Following the same arguments as in Proposition 1, setting $r_i = \exp(\alpha_i)$ for each i up to a multiplicative constant ensures the existence of $\mathbf{r}$ that satisfies (10).

Let us consider uniqueness. Let $\mathbf{r}$ a solution to the conditions (10) on a handicap-based ranking. $\mathbf{r}$ is associated to a balanced matrix $\boldsymbol{p}$ and the values defined by $\alpha_i = \ln(r_i), \beta_j = -\ln(q_j)$ satisfy (19). Hence $\boldsymbol{p}, \boldsymbol{\alpha}, \boldsymbol{\beta}$ satisfy the Kuhn and Tucker conditions associated to $\mathcal{P}$. Since the program $\mathcal{P}$ is strictly convex a solution $\boldsymbol{p}$ is unique. Thus, if there are two rankings solutions to (10), $\mathbf{r}$ and $\mathbf{r}'$, taking the difference in (19) for their corresponding values yields

$$(\alpha_i - \alpha_i') + (\beta_j - \beta_j') = 0 \text{ for each } (i,j) \in G. \tag{20}$$

This implies that along a path linking two items the values $(\alpha_i - \alpha_i')$ are all equal. Using the same argument as for Proposition 1, the uniqueness of $\mathbf{r}$ follows if G is connected: $\alpha_i' - \alpha_i$ is constant across all i, and there can be only one $\mathbf{r}$ defined by $r_i = \exp(\alpha_i)$ that belongs to Δ_N. On the other hand, if G is not connected, there are no links between the values on each component and uniqueness fails. ∎

References

Altman, Alon and Moshe Tennenholtz (2005), "Ranking systems: the PageRank axioms." *EC '05 Proceedings of the 6th ACM conference on Electronic commerce.*

Altman, Alon and Moshe Tennenholtz (2008), "Axiomatic foundations of ranking systems.", *Journal of Artificial Intelligence Research*, 31, 473-495.

Amir, Rabah (2002), "Impact-adjusted Citations as a measure of Journal quality." CORE discussion paper 74.

Bacharach, Michael (1965), "Estimating Nonnegative Matrices from Marginal Data." *International Economic Review*, 6, 294-310.

Balinski, Michel and Gabrielle Demange (1989), "Algorithm for Proportional Matrices in Reals and Integers." *Mathematical Programming*, 45, 193-210.

Boldi, Paolo, Massimo Santini, and Sebastiano Vigna (2007), "A deeper investigation of PageRank as a function of the damping factor." *Web Information Retrieval and Linear Algebra Algorithms* http://drops.dagstuhl.de/opus/volltexte/2007/1072 .

Bonacich, Phillip (1987), "Power and centrality: a family of measures." *American Journal of Sociology*, 92, 1170-1182.

Brin, Sergey and Lawrence Page (1998),"The anatomy of large-scale hypertextual web search engine." *Comp Networks and ISDN systems*, 30(1-7), 107-117.

Brink, Rene van den and Robert P. Gilles (2003), "Ranking by outdegree for directed graphs." *Discrete Mathematics* 271, 261-270.

Brink, Rene van den and Robert P. Gilles (2009), "The outflow ranking method for weighted directed graphs." *European Journal of Operational Research* 193, 484-491.

De Clippel, Geoffroy, Herve Moulin, and Nicolaus Tideman (2008), "Impartial division of a dollar." *Journal of Economic Theory*, 139, 176-191.

Demange, Gabrielle (2012), "On the influence of a ranking system." *Social Choice and Welfare*, Vol 39, Issue 2, 431-455.

Demange, Gabrielle (2013), "Collective attention and ranking methods." forthcoming in *Journal of Dynamics and Games.*

Du, Ye, Ehud Lehrer, and Ady Pauzner (2012), "Competitive economy as a ranking device over networks." mimeo.

Katz, Leo (1953), "A new status index derived from sociometric analysis." *Psychometrika*, 18, 39-43.

Kleinberg, Jon M. (1999), "Authoritative sources in a hyperlinked environment." *Journal of the ACM*, 46, 604-632.

Liebowitz, Stan J. and Jean-Pierre Palmer (1984), "Assessing the Relative Impacts of Economics Journals." *Journal of Economic Literature*, 22, 77-88.

Palacios-Huerta, Ignacio and Oscar Volij (2004), "The Measurement of Intellectual Influence." *Econometrica*, 72, 963-977.

Pandey, Sandeep, Sourashis Roy, Christopher Olston, Junghoo Cho, and Soumen Chakrabarti (2005), "Shuffling a Stacked Deck: The Case for Partially Randomized Ranking of Search Engine Results." *Proceedings of the 31st international conference on Very large data bases*, 781-792.

Pinski, Gabriel and Francis Narin (1976), "Citation Influence for Journal Aggregates of Scientific Publications: Theory, with Application to the Literature of Physics." *Information Processing and Management*, 12, 297-312.

Pukelsheim, Friedrich (2012), "Biproportional Matrix Scaling and the Iterative Proportional Fitting procedure." Augsburg preprint.

Rubinstein, Ariel (1980), "Ranking the Participants in a Tournament." *SIAM Journal on Applied Mathematics*, Vol. 38, No. 1, 108-111.

Slutzki, Giora and Oscar Volij (2006), "Scoring of web pages and tournaments-axiomatizations." *Social choice and welfare*, 26, 75-92.

Soules, George W. (1991), "The Rate of Convergence of Sinkhorn Balancing." *Linear algebra and its applications* 150, 3-40.

Santos, Boaventura de Sousa, B. (1998), "Participatory budgeting in Porto Alegre: toward a redistributive democracy." *Politics and Society*, 26, 461-510.

Woeginger, Gerhard J. (2008), "An axiomatic characterization of the Hirsch-index." *Mathematical Social Sciences*, 56, 224-232.