Ranking Sports Teams and the Inverse Equal Paths Problem*

Dorit S. Hochbaum

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

Abstract. The problem of rank aggregation has been studied in contexts varying from sports, to multi-criteria decision making, to machine learning, to academic citations, to ranking web pages, and to descriptive decision theory. Rank aggregation is the mapping of inputs that rank subsets of a set of objects into a consistent ranking that represents in some meaningful way the various inputs. In the ranking of sports competitors, or academic citations or ranking of web pages the inputs are in the form of pairwise comparisons. We present here a new paradigm using an optimization framework that addresses major shortcomings in current models of aggregate ranking. Ranking methods are often criticized for being subjective and ignoring some factors or emphasizing others. In the ranking scheme here subjective considerations can be easily incorporated while their contributions to the overall ranking are made explicit.

The *inverse equal paths* problem is introduced here, and is shown to be tightly linked to the problem of aggregate ranking "optimally". This framework is useful in making an optimization framework available and by introducing specific performance measures for the quality of the aggregate ranking as per its deviations from the input rankings provided. Presented as inverse equal paths problem we devise for the aggregate ranking problem polynomial time combinatorial algorithms for convex penalty functions of the deviations; and show the NP-hardness of some forms of nonlinear penalty functions. Interestingly, the algorithmic setup of the problem is that of a network flow problem.

We compare the equal paths scheme here to the eigenvector method, Google PageRank for ranking web sites, and the academic citation method for ranking academic papers.

Keywords: Network flow, aggregate ranking, inverse problems.

1 Introduction

We consider here an aggregate ranking scenario whereby the input to the ranking process is in the form of pairwise comparisons. This form of input is typical in the ranking of web pages, in the citation-based ranking of academic papers, and

* Research supported in part by NSF award No. DMI-0620677.

P. Spirakis et al. (Eds.): WINE 2006, LNCS 4286, pp. 307–318, 2006.

in sports teams' ranking. In team rankings the strength or rank of a team is determined by the scores of games it has played, and possibly also the identities of the teams with which each game is played. In the ranking of web pages, the pairwise comparison is in the form of a link from one page to another, that ,as we show, is analogous to one page "losing the game" to the page it points to. In academic ranking of papers the citation index counts the number of times that a paper has been cited in other papers. Each citation is again a form of pairwise comparison between the citing page and the cited page.

The problem of ranking competitors based on an incomplete set of pairwise comparisons is well-studied in the context of football and other sports, and also in general, [11]. There are numerous ranking schemes each with its uniquely emphasized factors and each with its advantages and shortcomings. The shortcomings, particularly in sports teams ranking where passions run high, bring about the ire of some people. The generic criticism is that certain games' outcomes have not been adequately incorporated, or have had an excessive impact on the aggregate ranking.

It is important to note that the "correctness" of a ranking is subjective. A recent method for ranking of academic papers by Chen et al. [9,5] is an illustration of this subjectivity in that its improvement is based on the prevailing opinion that certain papers are more important than indicated by their academic citations count rank. More details on this issue are discussed in Section 4.3.

One aspect that all existing schemes have in common is that all pairwise comparisons are considered equal in their impact on the final outcome. This causes biases such as counting a win against a weak team equally to a win against a strong team. In other schemes it might be preferable to not play a game at all if it is against a weak team, as such a game played by a strong team can actually reduce its rank, [15]. This uniformity of consideration of each pairwise comparison is one reason for the inclusion of *human polls* in, e.g., college footfall ranking. Human judgement has the advantage that it can take into account the quality of the game played, rather than the score quantifier alone, which is often attributed to some degree to chance. These human polls in turn, are often criticized for lack of transparency, as the factors that go into human ranking are not made explicit.

The model suggested here is different from the existing ones in that it is non-uniform with respect to the inputs and it permits the explicit inclusion of subjective factors. Any form of input from knowledgeable sources can be incorporated, and each input is associated with a degree of confidence deemed appropriate for the particular source and the source's expertise in a specific pairwise evaluation. The degree of confidence assigned can in itself be subjective, but it can be made according to a specific protocol and set of rules agreed upon in advance (e.g. based on past performance of the source). This allows to differentiate the importance of different games and calibrate their impact on the final ranking. The inclusion of human polls is still possible, and it can be further refined by having the assessments by different sources assigned varying degrees of confidence.

Another feature of the model here is that it provides a *performance measure* that can be used to evaluate the quality of aggregate rankings. Most of the literature on aggregate ranking has no performance measures on the quality of the attained consistent ranking, with the exception in Kemeny's model [16] based on inputs in the form of permutations corresponding to *ordinal rankings* seeking an aggregate ranking minimizing the number of reversed permutations. This model is limited by not allowing for partial lists (each permutation must be complete); there is no differentiation between the violation of rankings of different inputs; and the model is NP-hard to solve optimally.

Our ranking model can be viewed within the *inverse problem* paradigm. In an inverse problem one is given problem parameters that must, but do not, satisfy certain necessary conditions. The goal is to modify those parameters so the necessary conditions are satisfied, subject to a penalty function on the modification, and so that the total penalty is minimum. In the context of rankings we say that the inputs are inconsistent if they are conflicting with respect to any underlying ranking. For instance, if each team loses at least once, then a top team that ranks number one has its ranking inconsistent with the game(s) it has lost. So any aggregate ranking is going to conflict with some inputs, except in rare cases where each input is precisely consistent with one underlying ranking. The necessary condition that the comparisons have to satisfy is that of *consistency*. This concept is formally described in Section 2.

As an inverse problem, aggregate ranking has the scores of the games played and any other form of judgement and pairwise comparisons as the input parameters. These invariably are inconsistent and any aggregate ranking will modify these comparisons. The problem is to come up, for each pair, with a pairwise comparison that is consistent with some underlying ranking and that deviates as little as possible from the given inputs. The penalty for deviating from the inputs is measured in terms of *penalty functions* that are monotone increasing in the size of the deviation. These penalty functions are assigned to each input separately. So the penalty for deviating from the comparison assessment of a less reliable source can take lower values than the penalty for deviating from the assessment of a high confidence source.

We introduce here, for the first time, the inverse equal paths problem, and show that for convex penalty functions the problem is solvable with polynomial flow-based algorithms. We demonstrate how the inverse equal paths problem is equivalent to aggregate ranking with pairwise comparisons inputs. We then compare our new ranking technique to leading methodologies that include the eigenvector method, the Google PageRank algorithm and the citation index method for academic papers' ranking.

2 Fundamental Concepts and Preliminaries

2.1 The Inverse Equal Paths Problem

The *inverse problem paradigm* is as follows: Given observations and parameter values that do not conform with physical or feasibility requirements, adjust the parameter values so as to satisfy the requirements. The adjustment is made so as to minimize the cost of the adjustment in the form of penalty functions. A prominent application is to find the inverse shortest paths that conform with the reading of the speed of seismic waves. There one seeks a minimum penalty for deviation from existing estimates on lengths of arcs, so as to conform to the observation that a shortest path is of a given length, or of a given sequence of nodes. Variants of this inverse shortest paths problem were studied by Burton and Toint [6], [7], by Zhang, Ma and Yang [25] and by Ahuja and Orlin [2].

The input to the *inverse equal paths* problem is a non-simple connected graph $G = (V, A)$ where for each $(i, j) \in A$ there is a set of arcs R_{ij}, so that for $r \in R_{ij}$ there is an arc of weight w^r_{ij} from i to j and an arc in the opposite direction (j, i) of weight $-w^r_{ij}$ (anti-symmetric weights). Another input is a set of penalty functions $f_{w^r_{ij}}()$ for each arc $(i, j) \in A$ and $r \in R_{ij}$. A feasible solution to the problem is a set of anti-symmetric weights w^*_{ij} for all pairs $i, j \in V$ satisfying that for any pair of nodes $s, t \in V$ all the directed paths from s to t (and from t to s) with arc weights $\mathbf{w}^*$ are of the same length. A weight vector $\mathbf{w}^*$ is optimal if among all possible weight vectors $\mathbf{w}$ it minimizes the total sum of the penalty functions $\sum_{(i,j)\in A, r\in R_{ij}} f_{w^r_{ij}}(w_{ij})$.

In Section 3 we provide a formulation and algorithms for the inverse equal paths problem, EP.

2.2 Consistency of Rankings

A fundamental notion related to ranking is that of *consistency*. Pairwise comparisons can be expressed in terms of ordinal preference, or in terms of cardinal preference. An example of ordinal preferences is permutation ranking [16]. We consider here the cardinal preferences where each pairwise ranking is accompanied by a level of *intensity*. Intensity is a quantifier expressing the extent to which one is preferred to the other. For intensity rankings there are two forms of consistency: multiplicative and additive consistency. The notion of consistency in a multiplicative sense, (used e.g. by Saaty [20,21]) is that for a triple i, j, k, $a_{ij} \cdot a_{jk} = a_{ik}$. This is equivalent to the existence of an n-dimensional vector $\mathbf{w}$ so that $a_{ij} = \frac{w_i}{w_j}$. Such set of weights, called a *priority vector*, is not unique as for any consistent set of weights $w_1, \ldots, w_n$ and a scalar c the set $cw_1, \ldots, cw_n$ is also a priority vector. So we can anchor arbitrarily $w_1 = 1$ to ensure a unique set of weights corresponding to a consistent intensity ranking. The second definition of consistency in the additive sense (e.g. [1]) has for each triplet i, j, k, $a_{ij} + a_{jk} = a_{ik}$. We call this condition *triangle equality* or TE for short. If TE is satisfied for every triplet then there is an n-dimensional vector $\mathbf{w}$ so that $a_{ij} = w_i - w_j$. Again the vector of weights is not unique as the vector $\mathbf{w} + c$ for c a constant defines the same set of differences. Here we anchor the weights uniquely by setting $w_1 = 0$. Both definitions of consistency are obviously equivalent since the logarithms of the a_{ij} that are consistent in the multiplicative sense, are consistent in the additive sense, and vice versa.

One implication of this notion is that for consistent rankings matrix a single column or a single row contains the full information on the entire matrix: Given the ith column $(a_{i1}, a_{i2}, \ldots, a_{in})$ of a consistent matrix in the multiplicative sense and setting $w_1 = 1$, one can generate all pairwise rankings as $a_{kj} = a_{ki} \cdot a_{ij} = \frac{a_{ij}}{a_{ik}}$.

We comment that for *preference rankings* where preferences are expressed only in the ordinal sense, the notion of consistency is equivalent to the *transitivity* of valid rankings. That is, if i is preferred to j and j is preferred to k then i is preferred to k. The rankings of a set of projects $V = \{1, \ldots, n\}$ can be formalized as a graph on the set of nodes V with a set of arcs – or ordered pairs – A so that the ordered pair $(i, j) \in A$ if i is preferred to j. The consistency of the preferences is equivalent to the property of *acyclicity* of the corresponding directed graph $G = (V, A)$ – a graph that does not contain a directed cycle. It is well known that an acyclic directed graph admits a *topological ordering* which is an assignment of distinct indices from $\{1, \ldots, n\}$ to the n nodes (representing the projects) so that for every arc (i, j) in the graph $i > j$. So the value of the indices of the topological ordering can serve as the underlying weights of the respective objects. It should be noted that acyclic graphs do not typically represent a *full order*, unless they contain a Hamiltonian path. So some projects may not be comparable to others in the consistent preferences, and for such pairs the difference of weights' values is not meaningful.

3 The Inverse Equal Paths Model for Aggregate Ranking

A consistent aggregate ranking of a set of objects implies a setting of the pairwise comparisons so that they satisfy the triangle equality.

Lemma 1. *If triangle equality is satisfied for all triplets in a graph with anti-symmetric weights, then all paths between every pair of nodes are of equal length.*

Therefore, the requirement of equal lengths of the paths is equivalent to the requirement of consistency.

The input to both the equal paths and the aggregate ranking problems includes a penalty function for deviating from each pairwise comparison, or arc weight. Let the penalty function for the pair (i, j) be $F_{ij}(z_{ij})$ where z_{ij} is the ranking intensity (additive) of i compared to j in the aggregate ranking. For a set of R_{ij} of pairwise arcs comparing one pair (i, j), the penalty deviation function is $F_{ij}(z_{ij}) = \sum_{r \in R_{ij}} f_{ij}^r(z_{ij} - w_{ij}^r)$. This function typically takes the value 0 for an argument of 0 (0 deviation). It is allowed to be non-symmetric for positive and negative arguments.

The next lemma establishes that all paths are equal, or the preferences are consistent, if and only if there is an underlying set of weights associated with the nodes, *node potentials*, which are the priority weights.

Lemma 2. *If* $\mathbf{z}$ *is a set of weights for which graph* G *has all equal paths, then there exists a set of values* x_j*, for all* $j \in V$*, such that* $x_i - x_j = z_{ij}$*.*

Including in the formulation a set of variables, x_j, for the node potentials, or priority weights, is redundant, but has the advantage that the properties of the problem become transparent. We use here the anchoring of $x_1 = 0$. The inverse equal paths problem EP is then,

$$\begin{array}{lll} \text{(EP)} \quad \text{Min} & \sum_{i<j} F_{ij}(z_{ij}) & \\ \text{subject to} & x_i - x_j = z_{ij} & \text{for } i < j \\ & x_1 = 0 & \\ & \ell_j \le x_j \le u_j & j = 1, \ldots, n. \end{array}$$

It is easy to see that for $C = \max_{r,(i,j)} w_{ij}^r$, $-nC \le x_j \le nC$. We thus let $\ell_j = -nC$ and $u_j = nC$ for all $j = 1, \ldots, n$. In the aggregate ranking problem it would be reasonable to require a set of weights with some finite resolution. A set of integer weights in the interval $[-n, n]$ is sufficient to guarantee that there are enough distinct ranks to assign to each node (or team). Therefore, we would replace the lower and upper bound constraints on $\mathbf{x}$ by,

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

In case of rank ties, one might want to increase the resolution of the weights. The proximity algorithm of Hochbaum and Shanthikumar guarantees that an optimal solution in integers for one resolution level is close enough to an optimal solution on a finer grid, [13]. The EP model has the *fixed point* property. That is, if the input intensity preferences are consistent with some underlying ranking, then the optimal solution will be that underlying ranking.

3.1 Algorithms for the Inverse Equal Paths Problem

Observe that the constraint matrix of EP is totally unimodular. Therefore, when the objective function is convex, it follows immediately that the problem is solvable in polynomial time, [13]. Furthermore, the convex EP is a special case of convex dual of minimum cost network flow studied in [3].

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

1. For $F()$ convex functions the problem EP is solvable in polynomial time. An algorithm that runs in $\log U$ calls to a minimum cut procedure with complexity $O(\log U \cdot T(n^2, mn))$ is reported in [4]. Another, more efficient, algorithm for this problem runs in $O(mn \log n \log nU)$, [3]. Both these algorithms have been devised for the more general problem of the convex dual of minimum cost network flow (DMNCF).
2. For $F_{ij}(z_{ij}) = a_{ij}^+ \max\{z_{ij}, 0\} + a_{ij}^- \max\{-z_{ij}, 0\}$ (that is, $F_{ij}()$ are linear for positive deviation and for negative deviation), the algorithm reported in [14] has complexity of $O(T(n, m) + n \log U)$, which is best possible.
3. For $F()$ arbitrary functions the problem is NP-hard – it can be shown to be only harder than the multi-way cut problem which is known to be NP-hard. This case is known more commonly as the metric labeling problem and the functions $F()$ are usually δ functions equal to 0 if the argument is 0 and a positive constant otherwise. For these problems there is a large body of research on approximation algorithms, e.g. [17].

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

4 Leading Ranking Methods and Score-Based Algorithms

The simplest ranking algorithm is based on sorting according to total weight, or citation count, or in-degree of a web page counting the number of pages pointing to it. The total weight sorting algorithm is used to rank sports teams by counting the number of wins, losses (and draws). This is the method used for example to determine division winners in baseball. (To provide an incentive for goal-richer soccer games higher weights are assigned to stronger wins.) It is known that giving a weight that is inversely proportional to the out-degree (number of games won) of a node creates biases where it is possible that a team wins a game against a weaker team and this win actually decreases the team's rank, [15].

College football teams are ranked according to a weighted composite score called the BCS ranking that combines a number of algorithms with polls of expert human judges. The 2004 version included three components - the AP sportswriters' poll, the USA Today/ESPN coaches poll, and six computer rankings algorithms - all weighing equally. There is a great deal of criticism of the inclusion of human polls for their lack of transparency. We quote from http://spirit.tau.ac.il/public/gandal/bcs.htm

> Despite the criticism of computer rankings, they are the only ones that can be transparent and based on measurable criteria, which is to say, impartial. The computer ratings can also be improved. The computer ratings used by the BCS should be consistent (this has a formal mathematical meaning) with an endogenous strength of schedule.
>
> Additionally, all computer rankings should be required to publish their methodology. This insures transparency and will enable experts to evaluate them. For example, one could evaluate the rankings by using them to predict bowl game outcomes. This could create competition among the computer rankings themselves. Currently six computer ranking systems are used by the BCS. But there are many other ranking systems out there. As of December 4, Kenneth Massey (who produces a computer ranking for the BCS) lists 100 rankings on his comparison page: http://www.masseyratings.com/cf/compare.htm

All the computer rankings in BCS translate the scores of the games into relative strength of each of the competing teams. One reason for including human polls is that the scores alone do not fully reflect the strength of each team. For instance, the score does not capture whether a game is played in poor weather conditions, or a major player is sick on the day of the game, or a soccer team plays with fewer than 11 players, In those cases the significance of the score may need modifying. However, there no previously existing ranking system allowed to incorporate such contingencies.

4.1 The Principal Eigenvector Technique

The principal eigenvector technique has been known to apply to ranking since the 1950s. This method is reviewed e.g. in a study addressing the rankings of football teams by Keener, [15]. Consider intensity rankings that quantify by how much team i is stronger than team j by a positive number a_{ij} – a multiplicative intensity preference. (There is a great deal of research on how to determine the values of a_{ij} as a function of the score of a game, and Keener's study proposes one mapping between the score of the game and the value of a_{ij}.) Let n_i be the number of games played by team i. Then, r_i, the ultimate ranking of team i, is reasonably presumed to be proportional to the calibrated rank,

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

Thus $r_i = \frac{1}{\lambda}\sum_{j=1}^{n} \frac{a_{ij}}{n_i} r_j$, or $A\mathbf{r} = \lambda\mathbf{r}$ for $A = (\frac{a_{ij}}{n_i})$. The solution to this system of equations – the principal eigenvector – plays an important role in the Analytic Hierarchical Process, [20], and in the Google PageRank.

Perron-Frobenius theorem states that for a nonnegative nontrivial matrix A there exists a nonnegative eigenvector $\mathbf{r}$ corresponding to a unique eigenvalue λ. If A is *irreducible* then $\mathbf{r}$ is strictly positive, unique and simple and λ is the largest eigenvalue.

The notion of irreducibility has an algebraic definition. We prefer to discuss it as a graph property: Firstly the concept of *deduced ranking* is important. One can deduce the relative ranking of a pair of teams indirectly from the outcomes of a sequence games played. The relative ranking of teams i and j can be deduced, even if the two teams did not play directly, if there is a sequence of games $[i, i_1]$, $[i_1, i_2], \ldots, [i_k, j]$ for $k \geq 1$. The ranking of a direct pairwise comparison can be viewed as such sequence for $k = 1$. Now the concept of irreducibility is equivalent to having all pairs of teams comparable by deduced ranking. In graph terms this means that there is a path between each pair of nodes – namely, the graph is connected. (Notice that although the graph is directed there are two symmetric arcs between pairs that are directly linked, so there is a directed path if and only if there is an undirected path.)

Some properties of the principal eigenvector method are:

1. Unlike the weight sorting algorithm, the eigenvector method takes into consideration not only the count of how many times one object is stronger than others, but also which objects it is compared to. So winning against a strong team counts more than winning against a weak one.
2. "Missing games" still must correspond to entries in the matrix, as the matrix must be full. The standard approach is to include such games as a draw. This however tends to skew the overall ranking.
3. All games contribute uniformly to the aggregate ranking and no subjective evaluation of a score of a game can be included. This is also a feature in the total weight sorting algorithm used for web page ranking or for academic

citation ranking both of which do not differentiate between citations of between pointers. So a negative citation stating that a result in a related paper is wrong, counts the same as a citation referring to a paper as seminal. On web pages there are sometimes pointers that companies are buying in order to increase their web page rank, and these pointers are often unrelated to the content of the web page. The principal eigenvector method as well as other existing models do not discriminate however between citations as per their quality and significance.

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

Suppose the matrix of comparisons is consistent and the vector of weights is $\mathbf{w} = (w_i)_{i=1}^n$. Then $a_{ij} = \frac{w_i}{w_j}$. Summing up over all j, we obtain, $\sum_{j=1}^n a_{ij} w_j = n w_i$. Therefore, the vector of weights $\mathbf{w}$ satisfies, $\mathbf{A}\mathbf{w} = n\mathbf{w}$, and is thus an eigenvector specifying the weights assigned to each project or each criterion under the multiplicative model. In that the principal eigenvector satisfies the fixed point property. If the matrix is not consistent then the eigenvector approximates the preference weights. One measure of approximation for an asymmetric inconsistent matrix was defined by Saaty [21] is the *consistency index* C.I., $C.I. = \frac{\lambda_{max} - n}{n-1}$. where λ_{max} is the maximum eigenvalue of the matrix. A matrix is said to be consistent if and only if C.I is zero.

In terms of complexity, the computation of the principal eigenvector $\mathbf{w}^*$ is not practical for large values of n. It is common to calculate it instead with the power method, [23]: For a given initial assessment of ranks $\mathbf{w}_0$ (typically, assuming all ranks are equal), this is a recursive procedure based on,

$$\lim_{k \to \infty} \frac{A^k \mathbf{w}_0}{|A^k \mathbf{w}_0|} = \mathbf{w}^*. \tag{1}$$

4.2 Finding "Close" Consistent Rankings

Several approaches other than the eigenvector method have been proposed in the literature to generate a consistent matrix that is in some sense "close" to the given matrix. Most of these are based on minimizing some measure of distance of the generated consistent matrix from the given matrix. Regression-based approaches have been proposed (see [18] for a review and formal treatment of these methods) that assume the a_{ij}'s to be random variables with known distribution centered around a consistent comparison matrix. Least-squares and logarithmic least squares regression are the most popular of these techniques, and Saaty and Vargas [22] give a comparison of these methods to the eigenvector method. Techniques based on linear programming (Chandran et. al [8]), nonlinear programming (Wang et. al [24]), and goal programming have also been proposed.

4.3 The Google PageRank Algorithm

The Google PageRank algorithm is a finite approximation of the limit (1) using a small number of iterations. The following recursive formula that is used for Google PageRank can be shown to approximate in the limit the principal eigenvector of the respective matrix if $d = 0$:

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

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

The ranking of academic papers based on citation count has raised some interest and criticism recently, [9,5]. Citation-ranking of academic papers are determined by the citation count of a paper. Setting a citation of article i to j as an arc (i, j) is a graph $G = (V, A)$ with a node corresponding to each academic paper, this is equivalent to ranking each paper by its in-degree. Chen et al. [9] and Buchanan [5] point out that the traditional citation count brings about results that are contradictory to perceived importance of certain papers. In their study Chen et al. give some examples. One is a 1929 paper by Slater that ranks 1853rd in terms of citation count although there is a universal agreement among physicists that 'Slater determinant' introduced in that paper is a fundamental concept that is considered classic and therefore the citation count rank undervalues Slater's paper.

Chen et al. used instead the "Google PageRank Algorithm" noting that the ranking model of web pages is analogous to the academic citations model where pointing to a web page is equivalent to a citation. Chen et al. [9] computed the rank of Slater's paper with Google PageRank and showed it turns out 10th. This, and the improved rank of other 'classic' papers served as evidence that Google rank is a better measure of impact than the traditional citation count.

5 Using EP for Sports Ranking, Web Page Ranking and Academic Citations

Both applications of academic citations and web page rankings are unique among general aggregate ranking problems in that the "evaluators" are also the objects being evaluated. In sports team ranking the evaluators are the games and their outcomes provide a comparison of the relative strengths of the pairs of teams that played each game. In spite of this apparent difference, the models are analogous as citing a paper j by i is analogous to j winning a game against i. (This makes an unpleasant corollary that for a paper to retain a high citation count it should cite as few papers as possible. If using the PageRank for ranking of papers it is desirable to cite only recognized "strong" papers.)

The EP model can be used in several ways. Every citation or pointer from i to j are considered to be a pairwise evaluation of the relative strength of i and j in which j is stronger than i. The amount of this extra strength can be calibrated by the type of pointer or citation. The confidence level (or the steepness of the penalty function) can be determined by the quality of the journal in which the citing paper appears, or by the type of citation (positive, negative or neutral.)

The EP model shares the advantage of the principal eigenvector (and thus to some extent the Google pageRank that approximates it) in that it weighs more heavily comparative strength against strong objects than strength against weak ones. It does add however the flexibility of incorporating additional sources of information that are currently excluded from ranking schemes. Furthermore, it can use as a starting point the current ranking, regardless of the method that led to it, and adjust it based on additional pairwise comparisons. People and organizations can individualize the ranking using their own sources of information, and to the degree that they trust those sources.

One important issue is the evolution of rankings over time, as additional links and comparisons become available, [10]. The goal is not to recreate the ranking every time that new information becomes available. The total weight algorithm is obviously the simplest to adjust to new links - simply add to the count and shift the modified weight object in the sorted list. The principal eigenvector, Google pageRank and equal paths are however global in nature. Chien et al. [10] showed that for Google pageRank it is sufficient to apply the recursive formula within a limited "radius" from the modified link. Here another advantage of the EP model is that the relative rank of any selected subsets of objects can be retained unchanged, by fixing a reference point in the subset and all the relative rankings are then fixed with respect to that single weight. The position of the entire subset in the ranking may be shifted with comparisons that include objects in the subset, but the relative ranking remains the same. This makes it computationally easier to evolve the ranking weights as new comparisons become available. The same approach can be used on large data bases where within certain clusters the relative rankings are required to be unmodified. It remains to study rigorously the size of the neighborhood on which the impact of an added link is significant.

References

1. Ali I., Cook W.D. and Kress M.: Ordinal ranking and intensity of preference: a linear programming approach. Management Science **32** (1986) 1642–1647.
2. Ahuja R. K., and Orlin J. B.: Inverse optimization. Operations Research **49** 771-783, 2001.
3. Ahuja R. K., Hochbaum D. S. and Orlin J. B.: Solving the convex cost integer dual of minimum cost network flow problem. Management Science **49** (2003) 950–964.
4. Ahuja R. K., Hochbaum D. S. and Orlin J. B.: A cut based algorithm for the nonlinear dual of the minimum cost network flow problem. Algorithmica **39** (2004) 189–208.
5. Buchanan M.: Top rank. Nature Physics **2** (2006) 361.
6. Burton D. and Toint Ph. L.: On an instance of the inverse shortest paths problem. Mathematical Programming **53** (1992) 45–61.
7. Burton D. and Toint Ph. L.: On the use of an inverse shortest paths algorithm for recovering linearly correlated costs. Mathematical Programming **63** (1994) 1–22.

8. Chandran B., Golden B. and Wasil E.: Linear programming models for estimating weights in the analytic hierarchy process. Computers and Operations Research **32** (2005) 2213–2234.
9. Chen P., Xie H., Maslov S. and Redner S.: Finding scientific gems with Google. arXuv:physics/0604130 **1:18** (2006).
10. Chien S., Dwork C., Kumar R., Simon D.R. and Sivakumar D.: Link Evolution: Analysis and Algorithms. Internet Mathematics **1** (2004) 277–304.
11. David H. A.: The Method of Paired Comparisons. Charles Griffin and Company Ltd. and Oxford University Press, London, 2nd ed. (1988).
12. Hochbaum D. S. and Levin A.: Methodologies and algorithms for group rankings decision. Management Science (to appear).
13. Hochbaum D. S. and Shanthikumar J. G.: Convex Separable Optimization is not Much Harder Than Linear Optimization. Journal of the ACM **37** (1990) 843–862.
14. Hochbaum D. S.: An efficient algorithm for image segmentation, Markov Random Fields and related problems. Journal of the ACM **48** (2001) 686–701.
15. Keener J. P.: The Perron-Frobenius theorem and the ranking of football teams. SIAM Review **35** (1993) 80-93.
16. Kemeny J. G. and Snell J.L.: Preference ranking: An axiomatic approach. In Mathematical models in the social sciences, Boston, Ginn, (1962) 9–23.
17. Kleinberg J. and Tardos J.: Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. In Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science (1999) 14–23.
18. Laininen P. and Hämääinen R.P.: Analyzing AHP-matrices by regression. European Journal of Operational Research **148** (2003) 514–524.
19. Park J. and Newman M.E.J.: A network-based ranking system for US college football. arXiv:physics/0505169 **4:31** Oct (2005).
20. Saaty T.: A scaling method for priorities in hierarchical structures. Journal of Math. Psychology **15** (1977) 234–281.
21. Saaty T.: The Analytic Hierarchy Process. McGraw-Hill, New York, (1980).
22. Saaty T. and Vargas L. G.: Comparison of eigenvalue, logarithmic least squares and least squares methods in estimating ratios. Journal of Mathematical Modeling **5** (1984) 309–324.
23. Vargas R. S.: Matrix Iterative Analysis. Prentice Hall, Englewood Cliffs, NJ, (1962).
24. Wang Y.M., Yang J.B. and Xu D.L.: Mathematical Programming Methods for Generating weights from Interval Comparison Matrices. Internet http://www.sm.umist.ac.uk/wp/abstract/wp0205.htm, retrieved July (2004).
25. J. Zhang, Z. Ma and C. Yang. A column generation method for inverse shortest paths problems. J. ZOR Mathematical Methods of Operations Research **41** (1995) 347–358.