Eigenfactor$^{\text{TM}}$ Score and Article Influence $^{\text{TM}}$ Score: Detailed methods*

Here we describe the methods used to compute the Eigenfactor$^{\text{TM}}$ scores and Article Influence$^{\text{TM}}$ scores featured at `www.eigenfactor.org`[1]. The purpose of these metrics is to estimate the relative influence of reference items based on cross-citation data. Like Thomson Scientific's Impact Factor metric, the Eigenfactor$^{\text{TM}}$ metrics measure the number of times that articles published during a *census period* provide citations to papers published during an earlier *target window*. While Impact Factor has a one year census period and uses the two previous years for the target window, the Eigenfactor metrics have a one year census period and use the *five* previous years for the target window.

In principle the Eigenfactor$^{\text{TM}}$ algorithms can be applied to any cross-citation dataset at any level: journals, institutions, authors, articles, etc. In this document, we will describe the application of the Eigenfactor algorithm to the cross citation data provided in the 2006 edition (released Summer 2007) of Thomson Scientific's Journal Citation Reports (JCR).

Calculating Eigenfactor Score and Article Influence Score

The 2006 JCR indexes the citations from 7611 "source" journals for the Sciences and Social Sciences. From these data, we extract a 5-year cross citation matrix $\mathbf{Z}$. For 2006, the entries of this matrix would be:

$$\mathbf{Z_{ij}} = \text{Citations from journal } j \text{ in 2006 to articles published in journal } i \text{ during 2001–2005.}$$

When constructing $\mathbf{Z}$, we omit all self-citations[2], setting all of the diagonal elements of this matrix to 0. We normalize $\mathbf{Z}$ by the column sums (i.e., by the total number of outgoing citations from each journal) to create a column-stochastic matrix $\mathbf{H}$:

$$\mathbf{H_{ij}} = \frac{\mathbf{Z_{ij}}}{\sum_k \mathbf{Z_{kj}}}$$

*Methods version 2.01, November 2008. The Eigenfactor$^{\text{TM}}$Metrics, including the Eigenfactor$^{\text{TM}}$Score and the Article Influence $^{\text{TM}}$Score, and the Eigenfactor.org$^{\text{TM}}$website were initially developed by Jevin West, Ben Althouse, Martin Rosvall, and Carl Bergstrom at the University of Washington and Ted Bergstrom at the University of California Santa Barbara.

We also compute an *article vector* a, where a_i is the number of articles published by journal i over the five-year target window, divided by the total number of articles published by all source journals over the same five-year window. Notice that a is thus normalized to sum to 1 and it's i-th entry specifies the fraction of all published articles that come from journal i.

Some of the journals listed in the $\mathbf{H}$ matrix will be *dangling nodes* — journals that do not cite any other journals. Any column of the $\mathbf{H}$ matrix that has all 0 entries is a dangling node; we replace all such columns in $\mathbf{H}$ with the a vector to produce a new modified matrix $\mathbf{H'}$.

Following Google's PageRank approach, we define a new stochastic matrix P as follows[3]:

$$\mathbf{P} = \alpha \mathbf{H'} + (1 - \alpha) a.e^T.$$

Here $\mathbf{e}^T$ is a row vector of 1's and thus $\mathbf{A} = a.e^T$ is a matrix with identical columns each equal to the article vector a.

We define the journal influence vector π^* as the leading eigenvector of $\mathbf{P}$. This gives us the weights we will use in weighting citations value; under the stochastic process interpretation the π^* vector corresponds to steady-state fraction of time spent at each journal represented in $\mathbf{P}$. The Eigenfactor Score EF_i of journal i is defined as the percentage of the total weighted citations that journal i receives from our 7611 source items. We can write the vector of Eigenfactor Scores as

$$\mathrm{EF} = 100 \, \frac{\mathbf{H} \, \pi^*}{\sum_i [\mathbf{H} \, \pi^*]_i}.$$

Notice that the equation above uses the matrix $\mathbf{H}$ without the dangling node columns replaced by the article vector a.

The Article Influence Score AI_i for each journal i is a measure of the per-article citation influence of the journal[4]. The Article Influence Score is calculated as

$$AI_i = 0.01 \, \frac{\mathbf{EF}_i}{a_i},$$

where $\mathbf{EF_i}$ is the Eigenfactor Score for journal i and a_i is the i-th entry of the normalized article vector.

Notes

[1]Pseudocode for computing Eigenfactor scores and Article Influence scores is available at `http://www.eigenfactor.org/EF_pseudocode.pdf`. *Mathematica* source code is available at `http://www.eigenfactor.org/efcode_compressed.pdf/`.

[2] We ignore self-citations for several reasons. First, we want to avoid over-inflating journals that engage in the practice of opportunistic self-citation and to minimize the incentive that our measure provides for such practice. Second, we do not have self-citation information for the journals not listed in the JCR. Considering self-citations for JCR-listed but not non-listed journals would systematically over-value the journals in the former set relative to the latter. Third, if self-citations are included, some small journals with unusual citation patterns appear as nearly-dangling nodes, and this can bias their eigenfactor scores upward. The tendency of the JCR data set to list some outgoing citations under a single composite item "others" — which we cannot use our calculations because we do not know where they are directed — exacerbates this problem.

[3] Under a stochastic process interpretation, the matrix $\mathbf{H}'$ corresponds to a random walk on the citation network with dangling nodes replaced by fully-connected nodes. The matrix $\mathbf{P}$ corresponds to the Markov process which with probability α follows a random walk on the journal citation network, and which with probability $(1-\alpha)$ "teleports" to a random journal in proportion to the number of articles published by each journal. Rather than using the leading eigenvector of $\mathbf{H}'$ for our journal weights, we compute the leading eigenvector of the matrix $\mathbf{P}$. We do so for a two of reasons.

First, the stochastic matrix $\mathbf{H}'$ may be non-irreducible or periodic. Adding the teleport probability $1-\alpha$ ensures that P is both irreducible and aperiodic, and therefore has a unique leading eigenvector by the Perron-Frobenius theorem.

Second, even if the network is irreducible, without teleporting rankings can be unreliable and highly volatile when some components are extremely sparsely connected. Suppose, for example, that a citation network comprises two fields are connected only by the citations of two journals, one in each field. The relative weight of each field would then be set solely by the relative frequencies with which these two journals cited the other field. Similarly, teleporting keeps the system from getting trapped in small nearly-dangling clusters. If a small clique of journals are occasionally cited from outside but rarely cite out of clique itself, the Markov process characteried by $\mathbf{H}'$ can become trapped in this portion of the citation network for a very long period in time, effectively overvaluing the journals in this clique. Teleporting corrects this problem by reducing the expected duration of a stay in these small cliques.

We teleport to a journal with probability proportional to the number of articles published by that journal in order to avoid over-inflating the influence of small journals and under-inflating the influence of large ones. This is important because the journals in the social sciences are much smaller, on average than the journals in the sciences. As a result, an unweighted teleportation process, in which one teleports to each journal with equal probability, overestimates influence of articles in social science journals relative to science journals because the teleportation process.

[4]The Eigenfactor Score provides a measure of the total influence that a journal provides, rather than a measure of influence per article. Impact factor, by contrast, measures the per-article influence of a given journal. To make our results comparable to impact factor, we need to divide the journal influence by the number of articles published. Doing so yields the Article Influence Score.