Spectral Ranking

Sebastiano Vigna
vigna@di.unimi.it

November 8, 2013

Abstract

This note tries to attempt a sketch of the history of *spectral ranking*—a general umbrella name for techniques that apply the theory of linear maps (in particular, eigenvalues and eigenvectors) to matrices that do not represent geometric transformations, but rather some kind of *relationship between entities*. Albeit recently made famous by the ample press coverage of Google's PageRank algorithm, spectral ranking was devised more than sixty years ago, almost exactly in the same terms, and has been studied in psychology, social sciences, and choice theory. I will try to describe it in precise and modern mathematical terms, highlighting along the way the contributions given by previous scholars.

Disclaimer

This is is a work in progress with no claim of completeness. I have tried to collect evidence of spectral techniques in ranking from a number of sources, providing a unified mathematical framework that should make it possible to understand in a precise way the relationship between contributions. Reports of inaccuracies and missing references are more than welcome.

1 Introduction

From a mathematical viewpoint, a matrix M represents a linear transformation between two linear spaces. It is just one of the possible representations of the map—it depends on a choice for the bases of the source and target space. Nonetheless, matrices arise all the time in many fields outside mathematics, often because they can be used to represent (weighted) binary relations. At that point, one can apply the full machinery of linear algebra and see what happens. The most famous example of this kind is probably *spectral graph theory*, which provides bounds for several graph features using eigenvalues of adjacency matrices.

2 Spectral Ranking 101

Let us start with a square matrix M on the reals. We will not make any assumption on M. We imagine that the indices of rows and columns actually correspond to some entity, and that each value m_{ij} represent some form of *endorsement* or *approval* of entity j from entity i. Endorsement can be negative, with the obvious meaning.

Many *centrality indices* based on simple summations performed on the row or columns of this matrix were common in psychometry and sociometry. For instance, if the matrix contains just zeroes and ones meaning "don't like" or "like", respectively, the sum of column j will tell us how many entities like j. But, clearly, we are not making much progress.

The first fundamental step towards spectral ranking was made by John R. Seeley in 1949 [Seeley:1949]: he noted that these indices were not really meaningful because they did not take into consideration that it is important being liked by someone that is in turn being liked a lot, and so on. In other words, an index of importance, centrality, or authoritativeness, should be defined *recursively* so that my index is equal to the weighted sum of the indices of the entities that endorse me. In matrix notation,[1]

$$\boldsymbol{r} = \boldsymbol{r}M. \tag{1}$$

Of course, this is not always possible. Seeley, however, considers a nonnegative matrix without null rows and normalises its rows so that they have unit ℓ_1 norm (e.g., you divide each entry by its row sum); his rows have always nonzero entries, so this is always possible, and Equation 1 has a solution, because $M\mathbf{1}^T = \mathbf{1}^T$, so 1 is an eigenvalue of M, and its left eigenvector(s) provide solutions to Equation 1. Uniqueness is a more complicated issue which Seeley does not discuss and which can be easily analysed using the well-known Perron–Frobenius theory of nonnegative matrices, which also shows that 1 is the spectral radius, so $\boldsymbol{r}$ is a dominant[2] eigenvector, and that there are positive solutions.[3]

Our discussion can be formally restated for *right* eigenvectors, but of course Seeley's motivation fails. However, Wei in his Ph.D. dissertation [Wei:1952][4] argues about ranking (sport) teams, reaching dual conclusions. Kendall [Kendall:1955] discusses Wei's (unpublished) findings at length. Given a matrix M expressing how much team i is better than team j (e.g., 1 if i beats j, 1/2 for ties, 0 if i loses against j, with coherent values in symmetric positions), Wei argues that an initial score[5] of 1 given to all teams, leading to an *ex aequo* ranking, can be significantly improved as follows: each team gets a new ranking obtained by adding the scores of the teams that it defeated, and half the scores of the team with whom there was a draw. There is thus a new set of scores and a new ranking, and so on. In other words, Wei suggests to look at the rank induced by the vector

$$\lim_{k\to\infty} M^k \mathbf{1}^T.$$

Wei uses Perron–Frobenius theory to show that under suitable hypotheses this ranking stabilises at some point to the one induced by the dominant right eigenvector.

[1] All vectors are row vectors.

[2] A *dominant eigenvalue* is an eigenvalue with largest modulus (i.e., the spectral radius). An eigenvector associated with the dominant eigenvalue is called a *dominant eigenvector*. In most practical cases of spectral ranking there is just one strictly dominant eigenvalue.

[3] Actually, Seeley exposes the entire matter in terms of linear equations. Matrix calculus is used only for solving a linear system by Cramer's rule.

[4] Wei's dissertation is quoted sometimes as dated 1952, sometimes as dated 1955. I would be grateful to anybody who is able to provide this information reliably. Also, I could not find Wei's complete name.

[5] Here we take care of distinguishing the *scores* given to the teams from the *ranking* obtained sorting the teams by score.

In modern terms, given a matrix M expressing how much each team is better than another, the right dominant eigenvector provides the correct ranking of all teams.[6]

Wei's ranking is interesting in its own for three reasons: first, the motivation is clearly different; second, it clearly shows that using the dominant eigenvector (whatever the dominant eigenvalue) was already an established technique in the '50s[7]; third, in this kind of ranking the relevant convergence is *in rank* (the actual components of the vector are immaterial).

Getting back to left eigenvectors, the works of Seeley and Wei suggest that we consider matrices M with a real and positive dominant eigenvalue λ (we can just use $-M$ instead of M if the second condition is not satisfied) and its eigenvectors, that is, vectors $\boldsymbol{r}$ such that

$$\lambda \boldsymbol{r} = \boldsymbol{r} M. \tag{2}$$

If λ is complex, $\boldsymbol{r}$ cannot be real, and the lack of an ordering that is compatible with the field structure makes complex numbers a bad candidate for ranking.

In general, a *(left)*[8] *spectral ranking* associated with M is a dominant (left) eigenvector. If the eigenspace has dimension one, we can speak of *the* spectral ranking associated with M. Note that in principle such a ranking is defined up to a constant: this is not a problem if all coordinates of $\boldsymbol{r}$ have the same sign, but introduces an ambiguity otherwise.

3 Damping

We will now start from a completely different viewpoint. If the matrix M is a zero/one matrix, the entry i, j of M^k contains the number of directed path from i to j in the direct graph defined by M in the obvious way. A reasonable way of measuring the importance of j could be measuring the number of paths going into j, as they represent recursive endorsements.[9] Unfortunately, trying the obvious, that is,

$$\mathbf{1}(I + M + M^2 + M^3 + \cdots) = \mathbf{1} \sum_{k=0}^{\infty} M^k$$

will not work, as formally the above equation is correct, but convergence is not guaranteed. It is, however, if M has spectral radius smaller than one, that is,

[6] This technique is known in the literature about ranking of sport teams as "Kendall–Wei ranking".

[7] Indeed, Berge [Berge:1958] quotes Wei's work and notes that the right eigenvector can be used to measure centrality (in his words, *puissance*) in any directed graph, and in particular in *sociograms*—directed networks representing the influence among individuals.

[8] The distinction between left and right spectral ranking is in principle, of course, useless, as the left spectral ranking of M is the right spectral ranking of M^T. Nonetheless, the kind of motivations leading to the two kind of rankings are quite different, and we feel that it is useful to keep around the distinction: if the matrix represents *endorsement*, left spectral ranking is the correct choice; if the matrix represents *influence* or "better-than" relationships right spectral ranking should be used instead.

[9] Indeed, taking the limit of the vector giving for each node the number incoming paths of length k (somehow normalized) when $k \to \infty$ leads to the definition of spectral ranking, as the process is equivalent to finding the dominant left eigenvector using the power method (see, e.g., [Berge:1958], where outgoing paths are used to measure influence, instead).

$|\lambda_0| < 1$. It is thus tempting to introduce an *attenuation* or *damping* factor that makes things work:

$$\mathbf{1}(I + \alpha M + \alpha^2 M^2 + \alpha^3 M^3 + \cdots) = \mathbf{1}\sum_{k=0}^{\infty}(\alpha M)^k \tag{3}$$

Now we are actually working with αM, which has spectral radius smaller than one as long as $\alpha < 1/|\lambda_0|$ (e.g., if M is (sub)stochastic any $\alpha < 1$ will do the job). This index was proposed by Leo Katz in 1953 [Katz:1953][10]. He notes that

$$\mathbf{1}\sum_{k=0}^{\infty}(\alpha M)^k = \mathbf{1}(1 - \alpha M)^{-1},$$

which means that his index can be computed solving the linear system

$$\boldsymbol{x}(1 - \alpha M) = \mathbf{1}.$$

4 Boundary conditions

There is still an important ingredient we are missing: some *initial preference*, or *boundary condition*, as Hubbell [Hubbell:1965] calls it. Hubbell's interest is *clique detection*, an early study of *spectral graph clustering*[11]. Hubbell is inspired by the works of Luce, Perry and Festinger on clique identification [Luce and Perry:1949, Festinger:1949]; they use fixed powers of the adjacency matrix to estimate the similarity of nodes, and Hubbell proposes to sum up *all powers* of a matrix when such a sum exists. Then, in analogy with Leontief's input-output economic model,[12] which represents the relationships between input and output of goods in each industry [Leontief:1941], he argues that one can define a status index $\boldsymbol{r}$ using the recursive equation

$$\boldsymbol{r} = \boldsymbol{v} + \boldsymbol{r}M, \tag{4}$$

where $\boldsymbol{v}$ is a *boundary condition*, or *exogenous contribution* to the system. Finally, he notes that formally

$$\boldsymbol{r} = \boldsymbol{v}(1 - M)^{-1} = \boldsymbol{v}\sum_{k=0}^{\infty} M^k,$$

[10] We must note that actually Katz's index is $\mathbf{1}M\sum_{k=0}^{\infty}(\alpha M)^k$. This additional multiplication by M is somewhat common in the literature; it is probably a case of *horror vacui*.

[11] It would be interesting to write a note similar to this one for spectral clustering, as sociologists have been playing with the idea for quite a while.

[12] Recently, Franceschet [Franceschet:2011] has argued that Leontief's input-output model is a precursor of PageRank, which would make it the oldest known example of spectral ranking. I think this is a red herring, as Leontief just wants to represent the relationship between input and output of an economy. He claims that an *equilibrium* is reached when prices are given by the fixpoint of the linear operator describing the input/output relationship, but being the goods indexing the matrix inhomogeneous, this pricing is not a ranking (and, indeed, Leontief does not appear to make claims in this direction). If we consider any study of fixpoints of a linear operator that expresses some kind of input/output relation a kind of spectral ranking, then Markov [Markov:1906] beats Leontief by more than 30 years, and we can probably go back further.

and that the right side converges as long as $|\lambda_0| < 1$: M can even have negative entries. Clearly this is a generalisation of Katz's index[13] to general matrices that adds an initial condition, as the vector $\mathbf{1}$ is replaced by the more general boundary condition $\boldsymbol{v}$.[14][15]

5 From eigenvectors to path summation

Seeley's, Wei's and Katz's work might seem unrelated. Nothing could be farther from truth. Let's get back to the basic spectral ranking equation:

$$\lambda_0 \boldsymbol{r} = \boldsymbol{r} M.$$

When the eigenspace of λ_0 has dimension larger than one, there is no clear choice for $\boldsymbol{r}$. But we can try to *perturb* M so that this happens. A simple way is using Brauer's results [Brauer:1952] about eigenvector separation:[16]

Theorem 1 *Let A be an $n \times n$ complex matrix, $\lambda_0, \lambda_1, \ldots, \lambda_{n-1}$ be the eigenvalues of A, and let $\boldsymbol{x}$ be a nonzero complex vector such that $A\boldsymbol{x}^T = \lambda_0 \boldsymbol{x}^T$. Then, for every complex vector $\boldsymbol{v}$, the eigenvalues of $A + \boldsymbol{x}^T\boldsymbol{v}$ are $\lambda_0 + \boldsymbol{v}\boldsymbol{x}^T, \lambda_1, \ldots, \lambda_{n-1}$.*

Brauer's theorem suggests to perform a rank-one convex perturbation of M using a vector $\boldsymbol{v}$ satisfying $\boldsymbol{v}\boldsymbol{x}^T = \lambda_0$ by applying the theorem to αM and $(1-\alpha)\boldsymbol{x}^T\boldsymbol{v}$:

$$\lambda_0 \boldsymbol{r} = \boldsymbol{r}(\alpha M + (1-\alpha)\boldsymbol{x}^T\boldsymbol{v}).$$

Now $\alpha M + (1-\alpha)\boldsymbol{x}^T\boldsymbol{v}$ has the same dominant eigenvalue of M, but with algebraic multiplicity one, and all other eigenvalues are multiplied by α. This ensures that we have a unique $\boldsymbol{r}$, at the price of having introduced a parameter (the choice of $\boldsymbol{x}$ is particularly simple in case M is stochastic, as in that case we can take $\mathbf{1}$).

There is also another important consequence: $\boldsymbol{r}$ is defined up to a constant, so we can impose that $\boldsymbol{r}\boldsymbol{x}^T = \lambda_0$ (i.e., in case $\boldsymbol{x} = \mathbf{1}$, that the sum of $\boldsymbol{r}$'s coordinates is λ_0, which implies, if all coordinates have the same sign, that $\|\boldsymbol{r}\|_1 = \lambda_0$). We obtain

$$\lambda_0 \boldsymbol{r} = \alpha \boldsymbol{r} M + (1-\alpha)\lambda_0 \boldsymbol{v},$$

so now

$$\boldsymbol{r} = (1-\alpha)\boldsymbol{v}\bigl(1-\alpha M/\lambda_0\bigr)^{-1} = (1-\alpha)\boldsymbol{v}\sum_{k=0}^{\infty}\bigl(\alpha M/\lambda_0\bigr)^k = (1-\lambda_0\beta)\boldsymbol{v}\sum_{k=0}^{\infty}(\beta M)^k, \quad (5)$$

[13] Hubbell claims that its index (actually, its *status model*) bears a "rough resemblance" to Katz's: once the mathematics has been laid out in simple terms, one can easily see that they are the same thing.

[14] We note that while the rank induced by $\mathbf{1}M(1-\alpha M)^{-1}$ and $\mathbf{1}(1-\alpha M)^{-1} = 1 + M(1-\alpha M)^{-1}$ is the same, this is no longer true when we use a general boundary condition.

[15] Hubbell is thus the first to implicitly notice that the recursive (Seely, Wei) and pathwise (Katz) formulation of spectral ranking are actually the same thing. He also remarks that its score depends linearly on the border condition, or, as we would say in PageRankSpeak, that PageRank is linearly dependent on the preference vector. This is actually an important feature for quick computation of personalised or topical versions [Jeh and Widom:2003].

[16] I learnt the usefulness of Brauer's results in this context for separating eigenvalues from Stefano Serra–Capizzano. The series of papers by Brauer is also (maybe not surprisingly) quoted by Katz in his paper [Katz:1953].

and the summation certainly converges if $\alpha < 1$ (or, equivalently, if $\beta < 1/\lambda_0$). In other words, Katz–Hubbell's index can be obtained as the spectral ranking of a rank-one perturbation of the original matrix.

6 From path summation to eigenvectors

A subtler reason takes us backwards.[17] Given a matrix Z, the *index* of Z is the smallest nonnegative integer k such that the kernel (the eigenspace associated to the eigenvalue zero) of Z^k is equal to the kernel of Z^{k+1}. Equivalently, the index is the size of the largest Jordan block of the eigenvalue zero (which is zero for a nonsingular matrix).

The *Drazin inverse* of Z, written Z^D, is the unique solution of the equations

$$ZXZ = X \qquad XZ = ZX \qquad Z^{\nu+1}X = Z^{\nu},$$

where ν is the index of Z [Drazin:1958]. When Z is nonsingular, it coincides with Z^{-1}. The matrix $[\![Z]\!] = 1 - ZZ^D$ is called the *eigenprojection* (for the eigenvalue zero) of Z: it is a projection on the kernel of Z^{ν} along the range of Z^{ν}. The fundamental theorem we will use about the Drazin inverse is due to Meyer:

Theorem 2 ([Meyer:1974]) *Let Z be a square matrix with index ν. If m and p are nonnegative integers, the limit*

$$\lim_{\varepsilon\to 0} \varepsilon^m(\varepsilon + Z)^{-1}Z^p$$

exists if and only if $m + p \geq \nu$, and in that case

$$\lim_{\varepsilon\to 0} \varepsilon^m(\varepsilon + Z)^{-1}Z^p = \begin{cases} Z^D Z^p & \text{if } m = 0; \\ (-1)^{m-1}Z^{m+p-1}\,[\![Z]\!] & \text{if } m > 0. \end{cases}$$

We would like to know what happens to the perturbed dominant eigenvector (5) when $\alpha \to 1$:

$$\begin{aligned}
\lim_{\alpha\to 1}(1-\alpha)\boldsymbol{v}\big(1 - \alpha M/\lambda_0\big)^{-1} &= \lim_{\alpha\to 1} \frac{1-\alpha}{\alpha}\boldsymbol{v}\left(\frac{1}{\alpha} - M/\lambda_0\right)^{-1} \\
&= \lim_{\alpha\to 1} \frac{1-\alpha}{\alpha}\boldsymbol{v}\left(\frac{1}{\alpha} - 1 + 1 - M/\lambda_0\right)^{-1} \\
&= \lim_{\alpha\to 1} \frac{1-\alpha}{\alpha}\boldsymbol{v}\left(\frac{1-\alpha}{\alpha} + \big(1 - M/\lambda_0\big)\right)^{-1} \\
&= \lim_{\varepsilon\to 0} \varepsilon\boldsymbol{v}\big(\varepsilon + \big(1 - M/\lambda_0\big)\big)^{-1}.
\end{aligned}$$

Now, $1 - M/\lambda_0$ is a singular matrix with index equal to the index of λ_0 in M. If λ_0 is *semisimple* (i.e., $\nu = 1$) we can apply Meyer's theorem with $m = 1$, $p = 0$ and conclude that

$$\lim_{\alpha\to 1}(1-\alpha)\boldsymbol{v}\big(1 - \alpha M/\lambda_0\big)^{-1} = \boldsymbol{v}\,[\![1 - M/\lambda_0]\!]\,.$$

[17]Before October 2013, this section was correct for the semisimple case, but not in general (as previously claimed). Many of the recent references appearing in the old treatment have been excised, as the theorems of Meyer and Rothblum we report now have settled the problem of limit values for damped spectral rankings forty years ago.

The circle is closed, as the kernel of $1 - M/\lambda_0$ is exactly the eigenspace of M associated with the eigenvalue λ_0: spectral ranking is just the limit case of Katz–Hubbell's index.

The eigenprojection $[\![1 - M/\lambda_0]\!]$ has an intuitive description if λ_0 *and all other eigenvalues of maximum modulus* are semisimple: it is equal [Rothblum:1981] to the *Cesàro limit*

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} \frac{(M/\lambda_0)^k}{n},$$

that is, the limit in average of $(M/\lambda_0)^n$. In stochastic matrices all eigenvalues of modulo one are semisimple, which explains why the Cesàro limit is very popular in the Markov chain literature to obtain eigenprojections. Rothblum shows in [Rothblum:1981] that a suitable generalization of the Cesàro limit to *n-fold sums* makes it possible to write the eigenprojections in limit form even if the eigenvalues of maximum modulus other than λ_0 are not semisimple.

If λ_0 is not semisimple, say with index ν, the situation is not very different: Meyer's theorem essentially tells us that the perturbed dominant eigenvector (5) would grow too quickly as $\alpha \to 1$, but nonetheless

$$\lim_{\alpha\to 1} (1-\alpha)^\nu \boldsymbol{v}\big(1 - \alpha M/\lambda_0\big)^{-1} = (-1)^{\nu-1} \boldsymbol{v}\, [\![1 - M/\lambda_0]\!] \big(1 - M/\lambda_0\big)^{\nu-1}.$$

The resulting vector is still in the kernel of $1 - M/\lambda_0$, because $[\![1 - M/\lambda_0]\!]$ projects on the kernel of $\big(1 - M/\lambda_0\big)^\nu$:

$$\Big(\boldsymbol{v}\, [\![1 - M/\lambda_0]\!] \big(1 - M/\lambda_0\big)^{\nu-1}\Big)\big(1 - M/\lambda_0\big) = \boldsymbol{v}\, [\![1 - M/\lambda_0]\!] \big(1 - M/\lambda_0\big)^{\nu} = 0.$$

Note that since the $(1-\alpha)^\nu$ factor is just a scalar, we can also write (even when $\nu = 1$)

$$\lim_{\alpha\to 1} \frac{\boldsymbol{v}\big(1 - \alpha M/\lambda_0\big)^{-1}}{\Big\|\boldsymbol{v}\big(1 - \alpha M/\lambda_0\big)^{-1}\Big\|} = \frac{(-1)^{\nu-1} \boldsymbol{v}\, [\![1 - M/\lambda_0]\!] \big(1 - M/\lambda_0\big)^{\nu-1}}{\Big\|\boldsymbol{v}\, [\![1 - M/\lambda_0]\!] \big(1 - M/\lambda_0\big)^{\nu-1}\Big\|},$$

which means that the *direction* of the perturbed dominant eigenvector always tends to that of a dominant eigenvector of M as $\alpha \to 1$, albeit normalization is necessary to avoid divergence. If M is nonnegative, Rothblum shows that generalized Cesàro limits can be applied even to this case.

7 Putting It All Together

It is interesting to note that the journey made by our original definition through perturbation and then limiting has an independent interest. We started with a matrix M with possibly many eigenvectors associated with the dominant eigenvalue, and we ended up with a *specific* eigenvector associated with λ_0, given the boundary condition $\boldsymbol{v}$. This suggests to define in general *the* spectral ranking[18] associated

[18] We remark that in social sciences and social-network analysis "eigenvector centrality" is often used to name collectively ranking techniques using eigenvectors ("centrality" is the sociologist's "ranking"). On the other hand, in those areas indices based on paths such as Katz's are considered to be different beasts.

with M with boundary condition $\boldsymbol{v}$ as

$$\boldsymbol{r} = \boldsymbol{v}\,[\![1 - M/\lambda_0]\!]\,,$$

or

$$\boldsymbol{r} = \boldsymbol{v}(-1)^{\nu-1}\,[\![1 - M/\lambda_0]\!]\,\bigl(1 - M/\lambda_0\bigr)^{\nu-1}, \tag{6}$$

with ν equal to the index of λ_0, if we want to include that case of λ_0 not being semisimple. If M has a unique eigenvector, this definition is equivalent to (2), and $\boldsymbol{v}$ is immaterial. However, in the general case (6) provides a unique (albeit very difficult to compute) eigenvector depending on $\boldsymbol{v}$.[19]

If we start from a generic nonnegative matrix M and assume to normalise its rows, obtaining a substochastic matrix P, we should probably speak of *Markovian spectral ranking*, as the Markovian nature of the object becomes dominant. If $\lambda_0 = 1$ we have

$$\boldsymbol{r} = \boldsymbol{v}\,[\![1 - P]\!] = \lim_{n\to\infty}\sum_{k=0}^{n-1}\frac{P^k}{n}, \tag{7}$$

as dictated by Markov chain theory.[20] If $\boldsymbol{v}$ is a distribution, $\boldsymbol{r}$ is essentially[21] the limit distribution when the chain is started with distribution $\boldsymbol{v}$. If $\lambda_0 < 1$, we must check its index and work the details as in the general case (6), as we can no longer guarantee semisemplicity of all eigenvalues of maximum modulus.

Finally, we could define the *damped spectral ranking* of M with boundary condition $\boldsymbol{v}$ and damping factor α as

$$\boldsymbol{r}_\alpha = (1-\alpha)\boldsymbol{v}\sum_{k=0}^{\infty}\bigl(\alpha M/\lambda_0\bigr)^k$$

for $|\alpha| < 1$. The $(1-\alpha)$ term comes out naturally from (5), and makes the limit for $\alpha \to 1$ exist in the semisimple case (moreover, it forces $\|\boldsymbol{r}\|_1 = 1$ when M is stochastic and $\boldsymbol{v}$ is a distribution).[22] An even more general definition, taking the index ν of λ_0 into consideration, would use a multiplicative factor $(1-\alpha)^\nu$, which would make the limit for $\alpha \to 1$ always exist.

It is interesting to note that in the Markovian case the change of role of the boundary condition from the damped to the standard case has a simple interpretation: in the damped case, we have a Markov chain *with restart*[23] to a fixed distribution $\boldsymbol{v}$, and there is a single stationary distribution which is the limit of *every* starting distribution; in the standard case, $\boldsymbol{v}$ is the starting distribution from which we compute the limit distribution using the eigenprojection. Thus, when $\alpha \to 1$, the restart distribution $\boldsymbol{v}$ becomes the starting distribution, which is significant only if the chain is not irreducible (i.e., if the underlying graph is not strongly connected). An analogous consideration can be done in the general semisimple case: the preference vector of Katz–Hubbell's index becomes, in the limit, the starting vector of an averaged version of the power method that mimics Cesàro limit.

---

[19] Actually, introducing the resolvent and studying its behaviour is a standard technique: in [Kartashov:1996], equation 1.12, the author is interested exactly in the behaviour of the matrix $(1-\alpha)(I-\alpha P)^{-1}$ when $\alpha \to 1$ for a Markov chain P.

[20] Substochastic matrices, too, enjoy the property that all eigenvalues of modulo one, if any, are semisimple; however, if $\lambda_0 < 1$ there are no such eigenvalues.

[21] "Essentially" because P^* smooths out problems due to periodicities in the matrix.

[22] As noted by Bonacich [Bonacich:1987], α can even be negative.

[23] The name was suggested in [Boldi *et al.*:2006] as a general definition for PageRank's *teleportation*.

8 Followers

The work of Seeley was almost unnoticed, Wei's dissertation was known mainly to rank theorists, and Katz's paper was known mainly by sociologists, so it is no surprise that spectral ranking has been rediscovered several times.[24]

In this section we gather, quite randomly, the numerous insurgencies of spectral ranking in various fields we are aware of. In some cases, spectral ranking in some form is applied to some domain; in other cases, very mild variations of previous ideas are proposed (mostly, we must unhappily say, without motivation or assessment).

[Berge:1958] In his classic work on graph theory, Berge's quotes Wei's thesis and makes the remark that the (right) eigenvector approach can be applied to *any* directed graph. As an example, he considers *sociograms*, where nodes represent individuals and arcs represent influence.

[Pinski and Narin:1976] Here M is the matrix that contains in position m_{ij} the number of references from journal j to journal i. The matrix is then normalised in a slightly bizarre way, that is, by dividing m_{ij} by the j-th [sic] row sum. The spectral ranking on this matrix is then used to rank journals. [Geller:1978] tries to bring Markov-chain theory in by suggesting to divide by the i-th row sum instead (i.e., Markovian spectral ranking).

[Kleinberg:1999] HITS is Kleinberg's algorithm for finding *authorities* and *hubs* in a (part of a) web graph. HITS computes the first left and right singular vectors of a matrix A, which are the spectral ranking of AA^T and A^TA, respectively.[25] Note, however, that HITS is able to extract clustering information from additional singular vectors.

[Page *et al.*:1998] PageRank is the damped Markovian spectral ranking of the adjacency matrix of a web graph. The boundary condition is called *preference vector*, and it can be used to bias PageRank with respect to a topic, to personal preferences, or to generate trust scores [Gyöngyi *et al.*:2004].

[Kandola *et al.*:2003] In the context of computational learning, the *von Neumann kernel* (a particular kind of *diffusion kernel*) introduced by Kandola, Shawe–Taylor and Cristianini derives from a kernel matrix K a new kernel matrix $K(1-\lambda K)^{-1}$, that is, the matrix definining Katz's index. The idea is that the new kernel contains higher order correlations (in their leading example K is the cocitation matrix of a document collection).

[Huberman *et al.*:1998] With the aim of predicting the number of visits to a web page, Huberman, Pirolli, Pitkow and Lukose study a model derived from *spreading activation networks*. Essentially, given a distribution d that tells which fraction of surfers are still surfing after time t, the prediction vector at time t is $d(t)\boldsymbol{v}P^t$,

[24] It should be noted there are several other ways to use a graph structure to obtain scores for documents. For instance, one can use links (in particular, hypertext links) to alter text-based scores using the score of pointed pages: this simple idea dates at least back to the end of the eighties [Frisse:1988]. In the nineties, the idea was rediscovered again for the web (see, e.g., [Marchiori:1997], which, in spite of some claims floating around the net, does not do any kind of spectral ranking). An obvious spectral approach would use a preference vector containing normalised text-based scores, and then a right or left spectral ranking depending on whether authoritativeness or relevance is to be scored. To the knowledge of the author, this approach has not been explored yet.

[25] The same approach had been proposed earlier by Bonacich [Bonacich:1991] to rank individuals and groups.

where $\boldsymbol{v}$ is the initial number of surfers at each page. They use an inverse Gaussian distribution obtained experimentally, but using a geometric distribution the predicted overall (i.e., summed up over all t) number of surfers at each page will give a Markovian damped spectral ranking.

[**French Jr.:1956**] For completeness, we mention French's theory of social power, which bears a superficial formal resemblance with spectral ranking. However, in French's theory normalisation happens by *column*, so the trivial uniform solution is always a solution, and it is considered a *good* solution, as the theory studies the formation of consensus (e.g., the probability of getting the trivial uniform solution depending on the structure of the graph).

[**Bonacich:1972**] Bonacich proposes to use spectral ranking on zero-one matrices representing entities and their relationships to identify the most important entities (Seeley's, Wei's and Berge's work are not quoted).

[**Bonacich:1987**] Bonacich proposes a mild extension of Katz's index (i.e., damped spectral ranking) that includes negative damping; the interpretation proposed is that in *bargaining* having a powerful neighbour should count negatively.

[**Bonacich:1991**] Bonacich proposes to rank individuals and groups related by a rectangular incidence matrix A by using the spectral rankings of AA^T and A^TA. This is the same approach used in HITS [Kleinberg:1999], but assuming that the two scores are applied to disjoint subsets.

[**Bonacich and Lloyd:2001**] Bonacich and Lloyd propose again to use damped spectral ranking, but with a border condition. Hubbell's paper is quoted, but apparently the authors do not realise that they are just redefining its index. The authors, however, prove that under strong conditions (M symmetric and with a strictly dominant eigenvalue) damped spectral ranking converges to spectral ranking.

[**Bergstrom *et al.*:2008**] *Eigenfactor* is a score computed to score journals. It is a Markovian damped spectral ranking computed on the citation matrix, with an additional non-damped step (e.g., $S(1 - \alpha S)^{-1}$), as in the original formulation of Katz's index.

[**Saaty:1980**] In the '70s, Saaty developed the theory of the *analytic hierarchy process*, a structured technique for dealing with complex decision. After some preprocessing, a table comparing a set of alternatives pairwise is filled with "better then" values (the entry m_{ij} means how much i is better than j, and the matrix must be reciprocal, i.e., $m_{ij} = 1/m_{ji}$); *right* spectral ranking is then used to rank the alternatives. Some insight as to why this is sensible can be found in [Saaty:1987]. The mathematics is of course identical to Wei's, as the motivation is structurally similar.

[**Hoede:1978**] Hoede proposes to avoid the border condition of Hubbell's index by computing $\mathbf{1}M(1-M)^{-1}$ instead, under the condition that $1-M$ is invertible. This is exactly Katz's index with no damping. The main point of the author is that now we can just tweak the entries of M so to make $1 - M$ invertible, as "this hardly influences the model" [sic].

[**Shaffer:2011**] In his Ph.D. thesis, Shaffer proposes Patent Rank—the Markovian spectral ranking of the patent citation network.

9 Conclusions

I have tried to sketch a comprehensive framework for spectral ranking, highlighting the fundamental contributions of Seeley and Wei (the dominant eigenvector, possibly with stochastic normalisation), Katz (damping) and Hubbell (boundary condition). Of course, prior references might be missing, and certainly the followers section must be expanded. Feedback on all facets of this note is more than welcome.

References

[Berge:1958] Claude Berge. *Théorie des graphes et ses applications*. Dunod, Paris, France, 1958.

[Bergstrom *et al.*:2008] Carl T. Bergstrom, Jevin D. West, and Marc A. Wiseman. The eigenfactor^{TM} metrics. *Journal of Neuroscience*, 28(45):11433–11434, 2008.

[Boldi *et al.*:2006] Paolo Boldi, Violetta Lonati, Massimo Santini, and Sebastiano Vigna. Graph fibrations, graph isomorphism, and PageRank. *RAIRO Inform. Théor.*, 40:227–253, 2006.

[Bonacich and Lloyd:2001] Phillip Bonacich and Paulette Lloyd. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23(3):191–201, 2001.

[Bonacich:1972] Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120, 1972.

[Bonacich:1987] Phillip Bonacich. Power and centrality: A family of measures. *The American Journal of Sociology*, 92(5):1170–1182, 1987.

[Bonacich:1991] Phillip Bonacich. Simultaneous group and individual centralities. *Social Networks*, 13(2):155–168, 1991.

[Brauer:1952] Alfred Brauer. Limits for the characteristic roots of a matrix. IV: Applications to stochastic matrices. *Duke Math. J.*, 19:75–91, 1952.

[Drazin:1958] M.P. Drazin. Pseudo-inverses in associative rings and semigroups. *The American Mathematical Monthly*, 65(7):506–514, 1958.

[Festinger:1949] Leon Festinger. The analysis of sociograms using matrix algebra. *Human Relations*, 2(2):153–9, 1949.

[Franceschet:2011] Massimo Franceschet. PageRank: standing on the shoulders of giants. *Commun. ACM*, 54:92–101, June 2011.

[French Jr.:1956] John R. P. French Jr. A formal theory of social power. *Psychological Review*, 63(3):181–194, 1956.

[Frisse:1988] Mark E. Frisse. Searching for information in a hypertext medical handbook. *Communications of the ACM*, 31(7):880–886, 1988.

[Geller:1978] Nancy L. Geller. On the citation influence methodology of Pinski and Narin. *Inf. Process. Manage*, 14(2):93–95, 1978.

[Gyöngyi *et al.*:2004] Zoltán Gyöngyi, Hector Garcia-Molina, and Jan Pedersen. Combating web spam with TrustRank. In *VLDB '04: Proceedings of the Thirtieth international conference on Very large data bases*, pages 576–587. VLDB Endowment, 2004.

[Hoede:1978] Cornelis Hoede. A new status score for actors in a social network. Memorandum 243, Twente University Department of Applied Mathematics, 1978.

[Hubbell:1965] Charles H. Hubbell. An input-output approach to clique identification. *Sociometry*, 28(4):377–399, 1965.

[Huberman *et al.*:1998] Bernardo A. Huberman, Peter L.T. Pirolli, James E. Pitkow, and Rajan M. Lukose. Strong regularities in world wide web surfing. *Science*, 280(5360):95, 1998.

[Jeh and Widom:2003] Glen Jeh and Jennifer Widom. Scaling personalized web search. In *Proc. of the Twelfth International World Wide Web Conference*, pages 271–279. ACM Press, 2003.

[Kandola *et al.*:2003] Jaz Kandola, John Shawe-Taylor, and Nello Cristianini. Learning semantic similarity. *Advances in Neural Information Processing Systems*, 15:673–680, 2003.

[Kartashov:1996] N. V. Kartashov. *Strong Stable Markov Chains*. VSP, 1996.

[Katz:1953] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

[Kendall:1955] Maurice G. Kendall. Further contributions to the theory of paired comparisons. *Biometrics*, 11(1):43–62, 1955.

[Kleinberg:1999] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, September 1999.

[Leontief:1941] Wassily W. Leontief. *The structure of American economy, 1919-1929: An empirical application of equilibrium analysis*. Harvard University Press, 1941.

[Luce and Perry:1949] R.D. Luce and A.D. Perry. A method of matrix analysis of group structure. *Psychometrika*, 14(2):95–116, 1949.

[Marchiori:1997] Massimo Marchiori. The quest for correct information on the Web: hyper search engines. *Computer Networks and ISDN Systems*, 29:1225–1235, 1997.

[Markov:1906] Andrey A. Markov. Rasprostranenie zakona bolshih chisel na velichiny, zavisyaschie drug ot druga. *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete*, 2(15):135–156, 1906.

[Meyer:1974] Carl D. Meyer, Jr. Limits and the index of a square matrix. *SIAM Journal on Applied Mathematics*, 26(3):469–478, 1974.

[Page *et al.*:1998] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, Stanford University, Stanford, CA, USA, 1998.

[Pinski and Narin:1976] Gabriel Pinski and Francis Narin. Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. *Inf. Process. Manage*, 12(5):297–312, 1976.

[Rothblum:1981] Uriel G. Rothblum. Expansions of sums of matrix powers. *SIAM Review*, 23(2):143–164, 1981.

[Saaty:1980] Thomas L. Saaty. *The analytical hierarchy process*. McGraw-Hill, New York, 1980.

[Saaty:1987] Thomas L. Saaty. Rank according to perron: A new insight. *Mathematics Magazine*, 60(4):211–213, 1987.

[Seeley:1949] John R. Seeley. The net of reciprocal influence: A problem in treating sociometric data. *Canadian Journal of Psychology*, 3:234–240, 1949.

[Shaffer:2011] Monte J. Shaffer. *Entrepreneurial Innovation: patent Rank and Marketing Science*. PhD thesis, Washington State University, 2011.

[Wei:1952] T.-H. Wei. *The Algebraic Foundations of Ranking Theory*. PhD thesis, University of Cambridge, 1952.