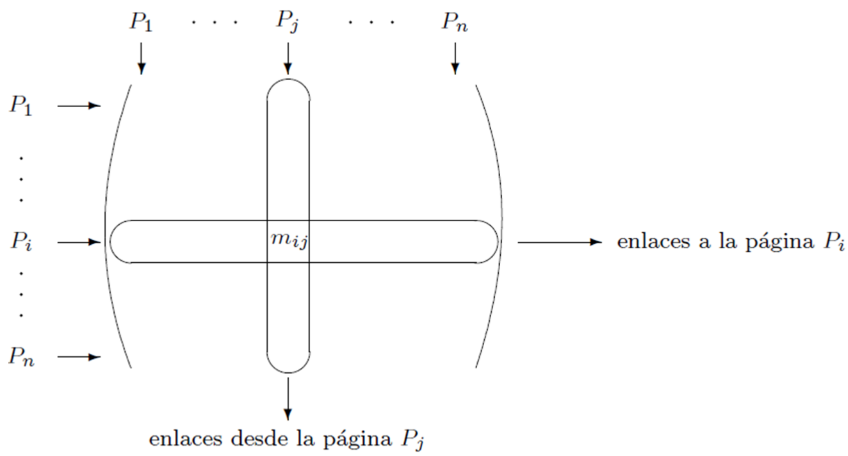
Álgebra Lineal y Google

Digamos que hemos recogido toda la información sobre la red (sitios, contenidos, enlaces de unas páginas a otras, etc.). De toda esta información, y en este primer paso de la modelización, vamos a rescatar únicamente la referida a cada una de las páginas, a la que asignamos etiquetas $P_1 \ldots, P_n$, y los enlaces entre ellas.
Vamos a formar una matriz M, de dimensiones $n \times n$, cuyas filas y columnas van etiquetados con los símbolos $P_1 \ldots, P_n$, y cuyas entradas son ceros y unos. La entrada m_{ij} de la matriz sería un uno si es que hay un enlace de la página P_j a la página P_i; y un cero en caso contrario:



Vamos a postular que la importancia de una cierta página de la red P_j ***tiene que ver*** con las páginas desde las que hay enlaces a ella. Esto es muy razonable: si muchas páginas enlazan con P_j , sería porque la información que ésta contiene ha sido considerada por muchos participantes de la red como ***recomendable***. Ese ***tiene que ver*** queda por ahora algo difuso.
Un primer intento, quizás algo ingenuo, consiste en suponer que la importancia x_j de cada P_j es proporcional al número de páginas desde las que hay enlaces a P_j . Observemos que, una vez que disponemos de la matriz M, el cálculo de cada uno de los x_j es sencillo, pues basta sumar las entradas correspondientes a la fila de P_j. Pero este modelo no recoge adecuadamente una situación que nos gustaría tener en cuenta, y es aquélla en la que una cierta página es citada desde unas pocas páginas, pero muy relevantes. Digamos, por ejemplo, que desde www.microsoft.com, o desde www.amazon.com, etc. El procedimiento anterior le asignaría baja importancia, algo que no parece razonable. Así que necesitamos afinar nuestro modelo de manera que asignemos alta importancia.
(a) tanto a páginas muy citadas;
(b) como a páginas poco citadas, pero desde sitios importantes.
El segundo intento, ya en esta línea, consiste en decidir que la importancia x_j de cada página P_j es proporcional a la suma de las importancias de las páginas que enlazan con P_j . Nótese que hemos cambiado, con respecto a nuestro primer intento, núumero de páginas que enlazan con P_j por suma de las importancias de las páginas que enlazan con P_j. Esta ligera variación cambia por completo, como veremos, las características del problema.
Supongamos, por ejemplo, que la página P_1 es citada desde las páginas P_2, P_{25} y P_{256}, que P_2 sólo se cita

desde P_1 y P_{256}, etc., mientras que, digamos, hay enlaces a la última página, P_n, desde P_1; P_2; P_3; P_{25} y P_{n-1}. En nuestra asignación anterior, x_1 debería ser proporcional a 3, x_2 lo sería a 2, etc., mientras que x_n habría de ser proporcional a 5.
Pero ahora nuestra asignación $x_1, \ldots, x_n$ debe cumplir que

$$\begin{aligned} x_1 &= K\left(x_2 + x_{25} + x_{256}\right), \\ x_2 &= K\left(x_1 + x_{256}\right), \\ &\vdots \\ x_n &= K\left(x_1 + x_2 + x_3 + x_{25} + x_{n-1}\right), \end{aligned}$$

donde K es una cierta constante de proporcionalidad. Nos encontramos así con un enorme sistema de ecuaciones, cuyas soluciones son las posibles asignaciones $x_1, \ldots, x_n$.
Escribamos el sistema anterior en términos más manejables, matriciales

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = K \begin{array}{c} \begin{array}{cccccccccccccccc} P_1 & P_2 & & & & P_{25} & & & & P_{256} & & & & P_{n-1} & \\ \downarrow & \downarrow & & & & \downarrow & & & & \downarrow & & & & \downarrow & \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \end{array} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

llamemos x al vector de importancias. La matriz de dimensiones $n \times n$ del sistema es justamente M. Así que podremos escribir que la asignación de importancias que perseguimos es una solución de

$$Mx = \lambda x$$

la cuestión se ha transformado en un problema de autovalores y autovectores.

Ejemplo Seis equipos, $E_1, ..., E_6$, divididos en dos conferencias, que juegan 21 partidos en total (6 contra los de su propia conferencia, 3 contra los de la otra).
La información sobre las victorias conseguidas está en la siguiente tabla:

	E_1	E_2	E_3	E_4	E_5	E_6	
E_1	–	3/21	0/21	0/21	1/21	2/21	$\to 6/21$
E_2	3/21	–	2/21	2/21	2/21	1/21	$\to 10/21$
E_3	6/21	4/21	–	2/21	1/21	1/21	$\to 14/21$
E_4	3/21	1/21	1/21	–	2/21	2/21	$\to 9/21$
E_5	2/21	1/21	2/21	4/21	–	2/21	$\to 11/21$
E_6	1/21	2/21	2/21	4/21	4/21	–	$\to 13/21$

A la derecha de la tabla hemos sumado el número de victorias que ha conseguido cada escuadra. Ese balance de victorias y derrotas sugiere que la ordenación adecuada es

$$E_3 \ \to\ E_6 \ \to\ E_5 \ \to\ E_2 \ \to\ E_4 \ \to\ E_1$$

Pero observemos que, por ejemplo, el líder de la competición, E_3, ha acumulado muchas victorias contra E_1, que es el peor equipo.
Introducimos los datos de las victoria en una matriz

$$A := \begin{bmatrix} 0 & \frac{1}{7} & 0 & 0 & \frac{1}{21} & \frac{2}{21} \\ \frac{1}{7} & 0 & \frac{2}{21} & \frac{2}{21} & \frac{2}{21} & \frac{1}{21} \\ \frac{2}{7} & \frac{4}{21} & 0 & \frac{2}{21} & \frac{1}{21} & \frac{1}{21} \\ \frac{1}{7} & \frac{1}{21} & \frac{1}{21} & 0 & \frac{2}{21} & \frac{2}{21} \\ \frac{2}{21} & \frac{1}{21} & \frac{2}{21} & \frac{4}{21} & 0 & \frac{2}{21} \\ \frac{1}{21} & \frac{2}{21} & \frac{2}{21} & \frac{4}{21} & \frac{4}{21} & 0 \end{bmatrix}$$

en un software calculamos los vectores propios y obtenemos

$$\begin{aligned} autovects := [&[0.012,\ 1.,\ \{[-0.062,\ -0.916,\ -2.131,\ 0.873,\ 0.731,\ 1.]\}],\\ &[0.475,\ 1.,\ \{[0.509,\ 0.746,\ 0.928,\ 0.690,\ 0.840,\ 1.]\}],\\ &[-0.111+0.117\,I,\ 1.,\ \{[-0.151-0.901\,I, 0.123+0.451\,I,\ -0.727+0.728\,I,\\ &\quad -0.435-0.128\,I,\ 0.192+0.377\,I,\ 1.]\}],\\ &[-0.126,\ 1.,\ \{[0.008,\ -0.685,\ 1.434,\ -1.071,\ 0.032,\ 1.]\}],\\ &[-0.139,\ 1.,\ \{[1.785,\ -3.880,\ 3.478,\ -5.392,\ 4.415,\ 1.]\}],\\ &[-0.111-0.117\,I,\ 1.,\ \{[-0.151+0.901\,I, 0.123-0.451\,I,\ -0.727-0.728\,I,\\ &\quad -0.435+0.128\,I,\ 0.192-0.377\,I,\ 1.]\}]] \end{aligned}$$

los resultados anteriores se interpretan en cada línea, primero el autovalor, luego la multiplicidad, y luego los autovectores asociados, comprobamos que la matriz A tiene seis autovalores distintos, dos complejos (conjugados) y cuatro reales. Si con el software calculamos sus módulos,

```
> autovals:=[evalf(eigenvalues(A))]: modulos:=[seq(abs(autovals[j]),j=1..6)];
```

$$modulos := [0.012,\ 0.475,\ 0.161,\ 0.126,\ 0.139, 0.161]$$

observamos que el mayor autovalor (en módulo) es $\lambda = 0{,}475$, cuyo autovector asociado,

$$x = (0{,}509; 0{,}746; 0{,}928; 0{,}690; 0{,}840; 1)$$

es el único cuyas entradas son todas números reales y no negativos. Ya tenemos la respuesta que buscábamos: el orden que sugiere este cálculo

$$E_6 \ \rightarrow\ E_3 \ \rightarrow\ E_5 \ \rightarrow\ E_2 \ \rightarrow\ E_4 \ \rightarrow\ E_1$$

que difiere del anterior en los dos primeros (ahora E_6 es el mejor equipo).
En esta matriz particular cuyas entradas son todas no negativas, se ha dado la mejor de las situaciones posibles: hay un único autovector cuyas entradas son todas no negativas. Este autovector, que además está asociado al autovalor de módulo máximo, nos sirve como solución al problema de ordenación que nos habámos planteado.

Hay una justificación para las afirmaciones anteriores. Las matrices no tienen por qué ser simétricas, ni defiidas positivas, ni. . .
Sin embargo, como mostraría Perron a principios del siglo XX

Teorema 1. *(Perron, 1907)*
Sea A una matriz (cuadrada) con entradas positivas, $A > 0$. Entonces
(a) Existe un autovalor (simple) $\lambda > 0$ tal que $Av = \lambda v$, donde el autovector correspondiente es $v > 0$;
(b) este autovalor es mayor, en módulo, que todos los demás autovalores;
(c) cualquier otro autovector positivo de A es un múltiplo de v.

Un caso más general fue demostrado por Fronebius quien observa que si únicamente tenemos que $A \geq 0$ entonces, aunque sigue habiendo un autovalor $\lambda > 0$ dominante (de valor absoluto máximo) asociado a un autovector $v \geq 0$, puede haber otros autovalores del mismo 'tamaño'.

Teorema 2. *(Frobenius, 1908)*
Sea A una matriz (cuadrada) con entradas no negativas, $A \geq 0$. Si A es irreducible entonces
(a) Existe un autovalor (simple) $\lambda > 0$ tal que $Av = \lambda v$, donde el autovector correspondiente es $v > 0$. Además, $\lambda \geq |\mu|$, para cualquier otro autovalor de μ de A.
(b) Cualquier autovector ≥ 0, es un múltiplo de v.
(c) Si hay k autovalores de módulo máximo, entonces son las soluciones de la ecuación $x^k - \lambda^k = 0$.

Aquí el articulo completo

`https://www.uam.es/personal_pdi/ciencias/gallardo/upm_google.pdf`