

El secreto de y el Álgebra lineal

Pablo Fernández Gallardo

Departamento de Matemáticas
Universidad Autónoma de Madrid
`pablo.fernandez@uam.es`

Valladolid, 4 de mayo de 2004

Resumen:

Desde hace unos años, Google se ha convertido en el buscador estándar en la red. Uno de sus “secretos”, quizás la clave de su éxito, es el algoritmo (PageRank) que utiliza para ordenar los resultados de las búsquedas. El objeto de esta charla es describir el modelo y los resultados matemáticos que están en la base de estos algoritmos de ordenación: un sabroso cóctel de Teoría de Grafos y Álgebra lineal que nos facilita la vida.

El buscador Google™

Creado en 1998 por Sergei Brin y Lawrence Page en la Universidad de Stanford.

El nombre es una variación sobre el término *googol*, 10^{100} .

Cuestiones importantes a la hora de diseñar un buscador en la red:

1. Computacionales:

- cómo almacenar la información;
- cómo actualizarla;
- cómo manejar/responder a peticiones;
- cómo buscar en las bases de datos.

Los números de 1997: 100 millones de páginas web. Altavista, 20 millones de consultas diarias.

Según la página de Google, hoy atiende a 200 millones de consultas diarias e indexa varios miles de millones de páginas web.



2. Tenemos los resultados de una búsqueda: ¿cómo los mostramos, en qué orden?

Necesitamos un **criterio** de ordenación, una asignación de **importancias** a cada sitio de la red:

$$\begin{array}{lcl} \text{sitios} & \longrightarrow & P_1, \dots, P_n \\ \text{importancias} & \longrightarrow & x_1, \dots, x_n \end{array}$$

Google utiliza el llamado sistema PageRank.

Un **objetivo**: basta leer los 10 primeros resultados para tener la respuesta.

NOTA. Hay un par de ingredientes que Google combina con el que veremos aquí:

- no es lo mismo que un cierto término aparezca en una página en el título, en negrita, en un tipo de letra pequeña, etc.
- Para búsquedas combinadas, tampoco es lo mismo que los términos buscados aparezcan “cerca” o “lejos” unos de otros.

Primer intento: x_j es proporcional al número de páginas que enlazan con P_j .

Problema: si una página se cita, digamos, una sola vez, pero desde `www.microsoft.com` o desde `www.amazon.com` . . .

Queremos combinar:

- páginas muy citadas;
- poco citadas, pero desde sitios importantes.

Segundo intento: x_j es proporcional a la suma de las importancias de las páginas que enlazan con P_j .

Por ejemplo, la página P_1 es citada desde las páginas P_2 , P_{25} y P_{256} , mientras que P_2 sólo se cita desde P_1 y P_{256} , etc. Nuestra asignación x_1, \dots, x_n debe cumplir que

$$x_1 = K (x_2 + x_{25} + x_{256}),$$

$$x_2 = K (x_1 + x_{256}),$$

$$\vdots$$

Mejor en proporciones

Podemos reescribir la matriz M de interés. Llamamos N_j al número de enlaces desde la página P_j (la suma de las entradas de cada columna de M).

$$M = \left(\begin{array}{c} \text{sitio } P_j \\ \downarrow \\ m_{1j} \\ m_{ij} \\ m_{nj} \\ \downarrow \\ N_j \end{array} \right)$$

$$M' = \left(\begin{array}{c} \text{sitio } P_j \\ \downarrow \\ \frac{m_{1j}}{N_j} \\ \frac{m_{ij}}{N_j} \\ \frac{m_{nj}}{N_j} \\ \downarrow \\ 1 \end{array} \right)$$

Obtenemos así una matriz estocástica (o de Markov). Esto da lugar a una formulación (y una interpretación) alternativa muy rica: una cadena de Markov.

El surfista aleatorio

Los **estados** son los vértices del grafo G . La matriz M' es la **matriz de transición** del sistema: cada entrada m'_{ij} es la probabilidad de pasar del estado (vértice) P_j al estado (vértice) P_i .

Un surfista en la red está, en cierto instante de tiempo, en la página P_k . En el siguiente instante de tiempo, estará en una página de entre las posibles (aquéllas a las que envía P_k). La elección de una u otra sigue una distribución de probabilidad *uniforme* (esto es, probabilidad $1/N_k$ para cada una de las admisibles).

$$\begin{pmatrix} m'_{1,k} \\ \vdots \\ m'_{i,k} \\ \vdots \\ m'_{n,k} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} m'_{1,k} \\ \vdots \\ m'_{i,k} \\ \vdots \\ m'_{n,k} \end{pmatrix}$$

Recuérdese que los $m'_{i,k}$ son, o bien 0, o bien $1/N_k$.

NOTA. Podría ocurrir que alguna página no tuviera enlaces salientes (su columna tuviera sólo ceros). No sería una matriz estocástica. La solución de Google: la sustituimos por una columna con valores $1/n$. Así que si el surfista llega a una página de la que (antes) “no podía salir”, ahora sortea (uniformemente) entre todas las de la red.

La clasificación para las eliminatorias por el título

Ha acabado la temporada regular en la NBA. ¿Qué 16 equipos pasan a disputar las eliminatorias?

Los equipos están divididos en dos conferencias, cada una de las cuales está formada por dos divisiones:

- Este:
 - Atlántico;
 - Central.
- Oeste:
 - Medio Oeste;
 - Pacífico.

Todos los equipos juegan el mismo número de partidos, pero no disputan el mismo número contra cada equipo. Por ejemplo, más con los de su propia conferencia.

Si un equipo está en una conferencia muy débil, y acumula muchas victorias. . .

Hay n equipos, E_1, \dots, E_n .

Formamos una matriz A en el que registramos las victorias obtenidas por cada equipo. Sus entradas son

$$a_{ij} = \frac{\# \text{ victorias de } E_i \text{ sobre } E_j}{\# \text{ partidos de } E_i}$$

Asignamos a E_i una importancia x_i proporcional a

$$\sum_{j=1}^n a_{ij} x_j .$$

Lo que nos conduce, de nuevo, a

$$A\mathbf{x} = \lambda\mathbf{x} .$$

Ejemplo: seis equipos, E_1, \dots, 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	→ 6/21
E_2	3/21	—	2/21	2/21	2/21	1/21	→ 10/21
E_3	6/21	4/21	—	2/21	1/21	1/21	→ 14/21
E_4	3/21	1/21	1/21	—	2/21	2/21	→ 9/21
E_5	2/21	1/21	2/21	4/21	—	2/21	→ 11/21
E_6	1/21	2/21	2/21	4/21	4/21	—	→ 13/21

Parece que la ordenación adecuada es $E_3 \rightarrow E_6 \rightarrow E_5 \rightarrow E_2 \rightarrow E_4 \rightarrow E_1$. Pero observemos que, por ejemplo, E_3 ha acumulado muchas victorias contra E_1 , que es el peor equipo.

Recurrimos a MAPLE: la matriz tiene seis autovalores distintos, dos complejos (conjugados) y cuatro reales. Uno de éstos, $\lambda = 0.475$, es el mayor (en módulo). Y el autovector asociado,

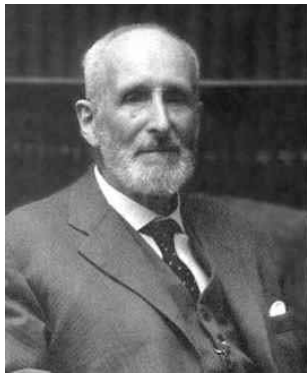
$$\mathbf{x} = (0.509, 0.746, 0.928, 0.690, 0.840, 1) ,$$

es el **único** cuyas entradas son todas números reales y no negativos.

Ya tenemos la respuesta que buscábamos: el orden que sugiere este cálculo es $E_6 \rightarrow E_3 \rightarrow E_5 \rightarrow E_2 \rightarrow E_4 \rightarrow E_1$, que difiere del anterior en los dos primeros (ahora E_6 es el mejor equipo).

Las Matemáticas entran en escena

La propiedad fundamental de las matrices que nos conciernen (sean markovianas o no) es que sus entradas son **no negativas**.



Oskar Perron

Teorema (Perron, 1907) Sea A una matriz (cuadrada) con entradas positivas, $A > 0$. Entonces,

- existe un autovalor (simple) $\lambda > 0$ tal que $A\mathbf{v} = \lambda\mathbf{v}$, donde el autovector es $\mathbf{v} > 0$.
- Este autovalor es mayor, en módulo, que todos los demás autovalores.
- Cualquier otro autovector positivo de A es un múltiplo de \mathbf{v} .

Si la matriz es únicamente $A \geq 0$, entonces hay un autovalor $\lambda > 0$ dominante (de valor absoluto máximo) asociado a un autovector $\mathbf{v} \geq 0$. Pero podría haber otros autovalores del mismo “tamaño”.



Georg Frobenius

Teorema (Frobenius, 1908-1912) Sea A una matriz (cuadrada) con entradas no negativas, $A \geq 0$. Si la matriz A es **irreducible**, entonces

- existe un autovalor (simple) $\lambda > 0$ tal que $A\mathbf{v} = \lambda\mathbf{v}$, donde el autovector es $\mathbf{v} > 0$. Además, $\lambda \geq |\mu|$, para cualquier otro autovalor μ de A .
- Cualquier autovector ≥ 0 es un múltiplo de \mathbf{v} .
- Si hay k autovalores de módulo máximo, entonces son las soluciones de $x^k - \lambda^k = 0$.

NOTA. ¿Qué quiere decir irreducible?:

1.- no existe ninguna permutación (de filas y columnas) que transforma A en una matriz del tipo

$$\left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline 0 & A_{22} \end{array} \right),$$

donde A_{11} y A_{22} son matrices cuadradas;

2.- $(I + A)^{n-1} > 0$.

3.- El grafo subyacente está fuertemente conectado.

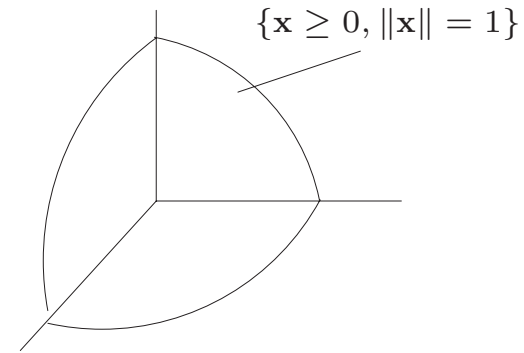
(Obsérvese que si $A > 0$, entonces es $A \geq 0$ e irreducible).

Una “demostración” del teorema de Perron-Frobenius (ilustrada en el caso 3×3)

Sabemos que $A \geq 0$, así que manda el octante positivo en sí mismo. Consideramos ahora la aplicación B dada por

$$B(\mathbf{x}) = \frac{A\mathbf{x}}{\|A\mathbf{x}\|}.$$

B envía el conjunto $\{\mathbf{x} \geq 0, \|\mathbf{x}\| = 1\}$ en sí mismo.



Por el teorema del punto fijo de Brouwer (es un conjunto cerrado, acotado y convexo y la aplicación es continua), existe $\tilde{\mathbf{x}}$ tal que $B(\tilde{\mathbf{x}}) = \tilde{\mathbf{x}}$.

Así que

$$B(\tilde{\mathbf{x}}) = \frac{A\tilde{\mathbf{x}}}{\|A\tilde{\mathbf{x}}\|} = \tilde{\mathbf{x}} \quad \implies \quad A\tilde{\mathbf{x}} = \|A\tilde{\mathbf{x}}\| \tilde{\mathbf{x}}.$$

Luego $\tilde{\mathbf{x}}$ es un autovector de A con entradas no negativas asociado a un autovalor > 0 .

¿Y la cuestión computacional?

Supongamos que estamos en las condiciones que garantizan la existencia de un autovalor λ_1 mayor (en módulo) que todos los demás autovalores. Sea \mathbf{x} el autovector (positivo) asociado.

NOTA. Una matriz $A \geq 0$ es **primitiva** si tiene un único autovalor dominante. Esto ocurre cuando, por ejemplo, $A^k > 0$ para cierto entero positivo k .

Recordemos que la matriz de Google es gigantesca.

Podemos, por supuesto, calcular *todos* los autovectores y quedarnos con el que nos interesa.

Pero la propia estructura del problema nos facilita la tarea. Supongamos, por simplificar, que A es diagonalizable. Los autovectores $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ son una base de \mathbb{R}^n . Están ordenados de manera que los autovalores correspondientes vayan en orden decreciente de tamaños:

$$\underbrace{|\lambda_1|}_{\text{autovalor dominante}} > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n|$$

Partimos de un $\mathbf{v}_0 \geq 0$, que escribimos como

$$\mathbf{v}_0 = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_n \mathbf{v}_n$$

Si aplicamos A ,

$$A\mathbf{v}_0 = c_1 \lambda_1 \mathbf{v}_1 + c_2 \lambda_2 \mathbf{v}_2 + \cdots + c_n \lambda_n \mathbf{v}_n$$

Otra vez:

$$A^2 \mathbf{v}_0 = c_1 \lambda_1^2 \mathbf{v}_1 + c_2 \lambda_2^2 \mathbf{v}_2 + \cdots + c_n \lambda_n^2 \mathbf{v}_n$$

Muchas más:

$$A^k \mathbf{v}_0 = c_1 \lambda_1^k \mathbf{v}_1 + c_2 \lambda_2^k \mathbf{v}_2 + \cdots + c_n \lambda_n^k \mathbf{v}_n$$

Supongamos que $c_1 \neq 0$. Entonces,

$$\frac{1}{\lambda_1^k} A^k \mathbf{v}_0 = c_1 \mathbf{v}_1 + c_2 \left(\frac{\lambda_2}{\lambda_1} \right)^k \mathbf{v}_2 + \cdots + c_n \left(\frac{\lambda_n}{\lambda_1} \right)^k \mathbf{v}_n.$$

Pero $|\lambda_j/\lambda_1| < 1$ para cada $j = 2, \dots, n$, así que

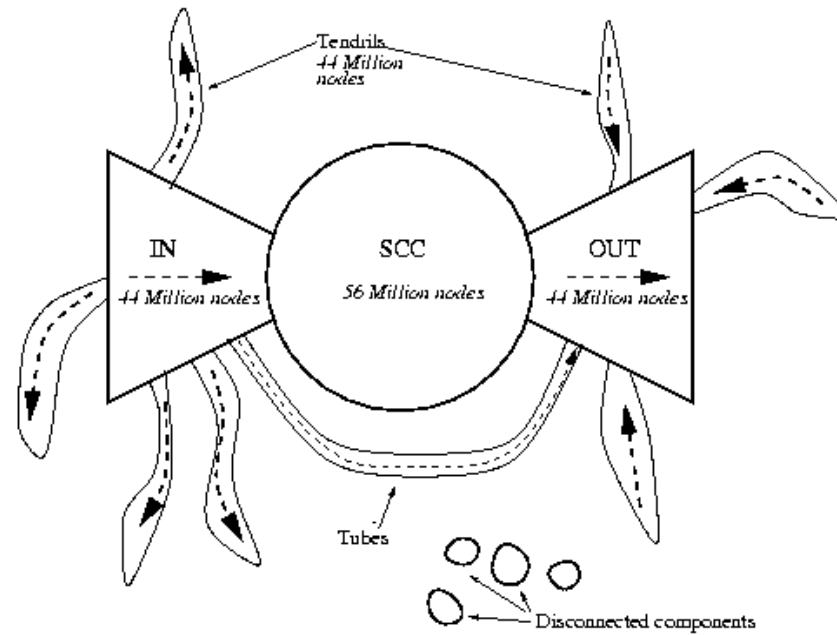
$$\frac{1}{\lambda_1^k} A^k \mathbf{v}_0 \xrightarrow{k \rightarrow \infty} c_1 \mathbf{v}_1$$

Éste es el llamado **método de las potencias**.

¿Estamos realmente en una situación ideal?

Para que todo funcione, necesitamos que la matriz M (o quizás M') asociada al grafo G de la red sea irreducible. En otras palabras, que G sea (fuertemente) conexo.

Pero no es el caso. Un estudio de 1999 (véase [4]). De las 203 millones de páginas censadas, el 90 % está en una gigantesca componente (débilmente) conexas:



Una posible solución es añadir toda una serie de probabilidades de transición (de salida) a todos los vértices.

Esto es, considerar la matriz

$$M'' = cM' + (1 - c) \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} (1, \dots, 1)$$

donde p_1, \dots, p_n es una distribución de probabilidad ($p_j \geq 0$, $\sum_j p_j = 1$) y c es un cierto parámetro.

Por ejemplo, podríamos tomar $p_j = 1/n$ para cada $j = 1, \dots, n$. Pero este grado de libertad permite hacer búsquedas “personalizadas”.

En términos del surfista aleatorio, estamos añadiendo la posibilidad de que (con probabilidad $1 - c$) se “aburra” de seguir los enlaces y opte por saltar a otras páginas (con arreglo a cierta distribución de probabilidad).

(Véase [6] para estimaciones de la *ratio* de convergencia del método en estas condiciones).

Matrices no negativas en otros contextos

La importancia del teorema de Perron-Frobenius radica en dos observaciones:

- en las situaciones “reales”, las interacciones que se miden son, muy frecuentemente, positivas, o al menos no negativas.
- Por otro lado, muchos modelos son procesos iterativos simples: de un estado inicial \mathbf{x}_0 pasamos a uno general $\mathbf{x}_k = A^k \mathbf{x}_0$. La convergencia del método depende del tamaño de los autovalores de A .

1. Modelos de evolución probabilística

Una matriz A es de Markov si $A \geq 0$ y, para cada columna, la suma de las entradas es 1.

NOTA. $\lambda = 1$ es un autovalor de A . Razón: las columnas de $A - I$ suman 0. Luego al sumar todas las filas de $A - I$ obtenemos el vector cero, así que las filas son linealmente dependientes.

Esto quiere decir que $A - I$ es singular, esto es, $\det(A - I) = 0 \implies \lambda = 1$ es autovalor.

Además, no puede haber autovalores de módulo > 1 .

El **problema y el modelo**: hay n estados de solvencia de las empresas S_1, \dots, S_n (en la jerga, AAA, BBB+, CC, etc.). En cada unidad de tiempo, la probabilidad de pasar del estado S_i al estado S_j es el número a_{ij} . De nuevo, una matriz A no negativa. Es habitual que un estado (D, de *default*) sea absorbente (todos ceros, menos el elemento de la diagonal).

	AAA	AA	A	BBB	BB	B	CCC	D	NR
AAA	88,72	8,14	0,66	0,06	0,12	0,00	0,00	0,00	2,29
AA	0,68	88,31	7,59	0,62	0,06	0,14	0,02	0,00	2,58
A	0,09	2,19	87,74	5,32	0,71	0,25	0,01	0,06	3,64
BBB	0,02	0,31	5,61	81,95	5,00	1,10	0,11	0,18	5,72
BB	0,03	0,13	0,61	7,03	73,27	8,04	0,91	1,06	8,93
B	0,00	0,10	0,21	0,38	5,66	72,91	3,56	5,20	11,98
CCC	0,18	0,00	0,18	1,07	1,96	9,27	53,48	19,79	14,08

Las proporciones iniciales son $\mathbf{z}^{(0)} = (z_1^{(0)}, \dots, z_n^{(0)})$. En las siguientes unidades de tiempo, $\mathbf{z}^{(k)} = A^k \mathbf{z}^{(0)}$.

Interesa el comportamiento asintótico, cuando $k \rightarrow \infty$, que llamaríamos $\mathbf{z}^{(\infty)}$. Si $\lambda = 1$ es el único autovalor dominante, entonces el estado estacionario $\mathbf{z}^{(\infty)}$ es el autovector correspondiente a $\lambda = 1$ (¡sean cuales sean las proporciones iniciales!)

2. Modelos dinámicos discretos

En una cierta especie, los individuos se agrupan por grupos de edad C_1, \dots, C_n . La población inicial es $\mathbf{z}^{(0)} = (z_1^{(0)}, \dots, z_n^{(0)})$. Planteamos las siguientes hipótesis:

- cada individuo pasa al siguiente grupo en cada unidad de tiempo.
- en la etapa i , cada individuo da lugar a m_i descendientes.
- s_i es la fracción que sobrevive de la edad $i - 1$ a la edad i .

La dinámica de la población viene determinada por el sistema matricial (**matriz de Leslie**) siguiente:

$$\begin{pmatrix} s_1 m_1 & s_1 m_2 & \cdots & s_1 m_{n-1} & s_1 m_n \\ s_2 & 0 & \cdots & 0 & 0 \\ 0 & s_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & s_n & 0 \end{pmatrix}^k \begin{pmatrix} z_1^{(0)} \\ \vdots \\ z_n^{(0)} \end{pmatrix} = \begin{pmatrix} z_1^{(k)} \\ \vdots \\ z_n^{(k)} \end{pmatrix} .$$

Si la matriz es “buena”,

$$\mathbf{z}^{(k)} \approx K \lambda_1^k \mathbf{v}_1 \quad \text{cuando } k \rightarrow \infty,$$

donde λ_1 es “el” autovalor dominante y \mathbf{v}_1 es su autovector asociado. El comportamiento (crecimiento, extinción, oscilación) de la población depende de si λ_1 es > 1 , < 1 ó $= 1$.

3. Modelos económicos

Una economía (simplificada) con tres sectores: agricultura, industria y servicios, que producen x_1 , x_2 y x_3 unidades, respectivamente.

La hipótesis fundamental es que el consumo que de la producción x_i hace el sector j es proporcional a x_j (la producción de j).

	Agricultura	Industria	Servicios	Consumidor	total producido
Agricultura	$0, 3x_1$	$0, 2x_2$	$0, 3x_3$	4	x_1
Industria	$0, 2x_1$	$0, 4x_2$	$0, 3x_3$	5	x_2
Servicios	$0, 2x_1$	$0, 5x_2$	$0, 1x_3$	12	x_3

De las x_1 unidades producidas por el sector agrario, el 30% son “autoconsumidas”, $0, 2x_2$ utilizadas por la industria, $0, 3x_3$ por el sector de servicios, mientras que 4 unidades lo son por los consumidores finales.

En términos matriciales, tenemos

$$Ax + \mathbf{b} = \mathbf{x}.$$

Si $\mathbf{b} \geq 0$, ¿tiene el sistema anterior una solución $\mathbf{x} \geq 0$? Será el caso si $I - A$ es invertible.

Una condición suficiente: el autovalor dominante es < 1 .

(Éste es el **modelo de producción y consumo de Leontiev**).

Para saber más

- [1] R. B. BAPAT, T. E. S. RAGHAVAN: *Nonnegative matrices and Applications*. Encyclopedia of Mathematics and its applications **64**. Cambridge University Press, 1997.
- [2] ABRAHAM BERMAN, ROBERT J. PLEMMONS: *Nonnegative matrices in the Mathematical Sciences*. Classics in applied Mathematics **9**. Academic Press, 1979.
- [3] SERGEY BRIN AND LAWRENCE PAGE: *The anatomy of a large-scale hypertextual web search engine*. www-db.stanford.edu/~sergey/
- [4] ANDREI BRODER ET AL.: *Graph structure in the web*. <http://www9.org/w9cdrom/160/160.html>
- [5] JAMES P. KEENER: The Perron-Frobenius Theorem and the ranking of football teams. *SIAM Review* **35** (1993), no. 1, 80-93.
- [6] TAHER S. HAWELIWALA Y SEPANDAR D. KAMVAR: The second eigenvalue of the Google matrix. <http://www.stanford.edu/~taherh/papers/secondeigenvalue.pdf>
- [7] C. R. MACLAUER: The Many Proofs and Applications of Perron's Theorem. *SIAM Review* **42** (2000), no. 3, 487-498.
- [8] CLEVE MOLER: *The world's largest matrix computation*. www.mathworks.com/company/newsletter/clevescorner/oct02_cleve.shtml
- [9] HERBERT S. WILF: *Searching the web with eigenvectors*. www.cis.upenn.edu/~wilf/