Sobre Autovalores y Autovectores
Una aplicación real

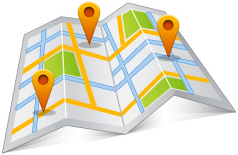
Métodos Numéricos

Departamento de Computación, FCEyN,Universidad de Buenos Aires.

8 de Mayo de 2015

¿Dónde estamos?

- Elementos de álgebra lineal
- Eliminación gaussiana, LU, normas
- Matrices simétricas definidas positivas
- 1^{er} Parcial, 1^{er} TP, 1^{er} Taller
- Matrices ortogonales y factorización QR
- Autovalores y factorización SVD
- Métodos iterativos
- 2^{do} Trabajo práctico
- Usted está aquí



- 2^{do} Parcial, Taller
- Interpolación lineal e integración numérica
- Cuadrados mínimos lineales
- Ceros de Funciones
- 3^{er} Parcial, TP, final
- fiesta, joda, nunca más métodos numéricos

¿De qué es la clase de hoy?

- Apliación real

The $25,000,000,000 Eigenvector:
The Linear Algebra behind Google*

Kurt Bryan†
Tanya Leise‡

  - Problema (real)
  - Un modelo (no el único) para atacar el problema
  - Aplicación de herramientas que vimos en la materia (autovalores/autovectores, método de la potencia)

- Sobre el tp2

  (y consultas, muchas consultas)

Motores de búsqueda

- Explorar la red e identificar todas las páginas con acceso público.
- Almacenar la información obtenida, para realizar búsquedas eficientemente.
- Determinar un **orden de las páginas según su importancia,** para presentar la información con un **orden de relevancia**.

PageRank

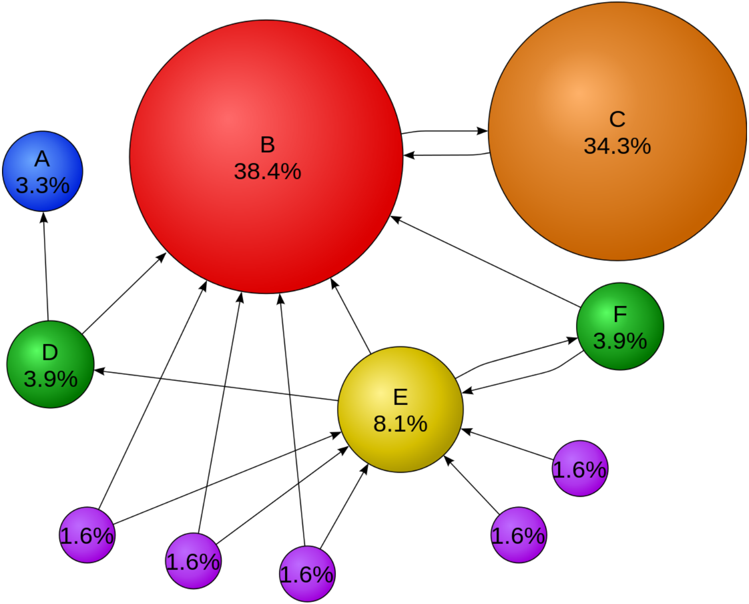
- Algoritmo que utiliza Google Search para rankear búsquedas

Objetivo:

- Determinar un **orden** de las páginas según su **importancia**, para presentar la información con un **orden de relevancia**.

¿Qué quiere decir que una página sea importante ?

- Contar **cantidad** y **calidad** de los links a una determinada página


B
38.4%
C
34.3%
A
3.3%
D
3.9%
E
8.1%
F
3.9%
1.6%
1.6%
1.6%
1.6%
1.6%
1.6%

Un poco de formalismo

- Tenemos un conjunto de páginas *Web*: $W = \{1, \ldots, n\}$.

matriz de conectividad:

- $W \in \{0,1\}^{n \times n}$,

$$w_{ij} = \begin{cases} 1 & \text{si la página } j \text{ tiene un link a la página } i \\ 0 & c.c. \end{cases}.$$

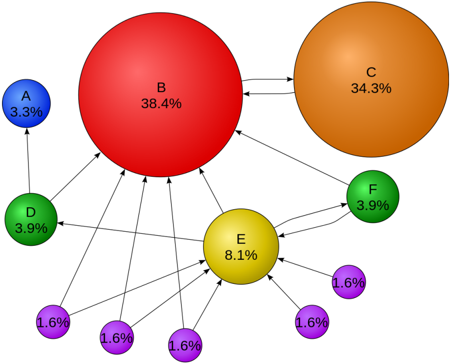
(ignoramos los *autolinks*)

grado de una página:

- Dada $j \in W$, definimos $n_j = \sum_{i=1}^{n} w_{ij}$

- $x_j :=$ puntaje asignado a la página $j \in W$

Un poco más de formalismo



Dadas $u, v \in W$,

- $x_u / n_u :=$ el aporte del link $u \longrightarrow v$

Finalmente

Dadas $L_k \subseteq W$,

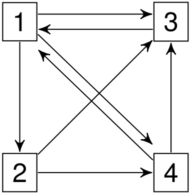
- $x_k = \sum_{j \in L_k} \frac{x_j}{n_j} \qquad k = 1, \ldots, n$

Sea $P \in \mathbb{R}^{n \times n}$ tal que:

$$P_{ij} = \begin{cases} \frac{1}{n_j} & \text{si } w_{ij} = 1 \\ 0 & c.c. \end{cases} .$$

Queremos:

- $x \in \mathbb{R}^n_{\geq 0}$ y $\sum_{i=1}^{n} = 1$ tal que $\quad x = Px$



$$n_1 = 3,\ n_2 = 2,\ n_3 = 1,\ n_4 = 2$$

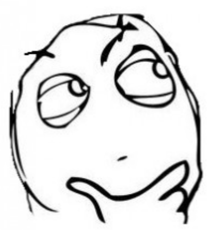
$$P = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

Pregunta:

¿Qué pasa si una página i no tiene links salientes (i.e., $n_i = 0$, denominado *dangling node*)?

- *Navegante aleatorio:*
  Empieza en una página (cualquiera) y luego en cada página $j \in W$ sigue navegando los links eligiendo con probabilidad $1/n_j$.

  ¿y si la página no tiene links salientes ?



Algunas cuestiones

- *Navegante aleatorio:*
  Empieza en una página (cualquiera) y luego en cada página $j \in W$ sigue navegando los links eligiendo con probabilidad $1/n_j$.

  ¿y si la página no tiene links salientes ?

  En ese caso, consideramos que pasa a cualquiera de las páginas de la red con probabilidad $1/n$.

Definimos:

- $v \in \mathbb{R}^n$, $v_i = 1/n$.
- $d \in \{0,1\}^n$,

$$d_i = \begin{cases} 1 & n_i = 0 \\ 0 & c.c. \end{cases} .$$

- $D = vd^t$
- $P_1 = P + D$

¿y si el navegante aleatorio, estando en la página j decide visitar una página cualquiera?

Definimos el *modelo enriquecido*:

- $\bar{1} = (1, \ldots, 1) \in \mathbb{R}^n$.
- $E = v\bar{1}^t$.
- $P_2 = cP_1 + (1-c)E$, $c \in (0,1)$.

donde recordemos que:

- $v \in \mathbb{R}^n$, $v_i = 1/n$.
- $P_1 = P + D$

Finalmente, queremos $x \in \mathbb{R}^n_{\geq 0}$ y $\sum_{i=1}^n x_i = 1$
tal que $x = P_2 x$

Definimos el *modelo enriquecido*:

- $\bar{1} = (1, \ldots, 1) \in \mathbb{R}^n$.
- $E = v\bar{1}^t$.
- $P_2 = cP_1 + (1-c)E$, $c \in (0,1)$.

Finalmente, queremos $x \in \mathbb{R}^n_{\geq 0}$ y $\sum_{i=1}^n x_i = 1$
tal que $\quad x = P_2 x$

Se tiene que P_2 es *estocástica por columnas*, con lo cual se cumplen las hipótesis necesarias (Bryan y Leise 2006) para que exista un autovector de P_2 asociado al autovalor 1. Además $1 = \lambda_1 > |\lambda_2| \geq \ldots \geq |\lambda_n|$, donde la dimensión del autoespacio asociado al autovalor λ_1 es 1.

¿entonces?

Método de la

Potencia

metodo_de_la_potencia(B,x_0,niter)

1. $v \leftarrow x_0$.
2. Para $i = 1, \ldots, niter$
3. $\quad v \leftarrow \frac{Bv}{||Bv||}$
4. Fin Para
5. $\lambda \leftarrow \frac{v^t Bv}{v^t v}$
6. Devolver λ, v.

Hagamos la prueba

(1) • Definir *Web* (páginas web)

(2) • Construir la matriz de conectividad

(3) • Hacer las cuentas (todo el formalismo que vimos)

(4) • Aplicar el método de la potencia para encontrar x

Hagamos la prueba

$$Web = \{$$

http://mathworld.wolfram.com/Vector.html
http://mathworld.wolfram.com/CholeskyDecomposition.html
http://mathworld.wolfram.com/Determinant.html
http://mathworld.wolfram.com/VandermondeMatrix.html
http://mathworld.wolfram.com/Matrix.html
http://mathworld.wolfram.com/RotationMatrix.html
http://mathworld.wolfram.com/PositiveDefiniteMatrix.html
http://mathworld.wolfram.com/LinearAlgebra.html
http://mathworld.wolfram.com/LUDecomposition.html
http://mathworld.wolfram.com/QRDecomposition.html

$$\}$$

Hagamos la prueba

$$Web = \{$$

http://mathworld.wolfram.com/Vector.html
http://mathworld.wolfram.com/CholeskyDecomposition.html
http://mathworld.wolfram.com/Determinant.html
http://mathworld.wolfram.com/VandermondeMatrix.html
http://mathworld.wolfram.com/Matrix.html
http://mathworld.wolfram.com/RotationMatrix.html
http://mathworld.wolfram.com/PositiveDefiniteMatrix.html
http://mathworld.wolfram.com/LinearAlgebra.html
http://mathworld.wolfram.com/LUDecomposition.html
http://mathworld.wolfram.com/QRDecomposition.html

$$\}$$

```
python webparser.py weblist.in graph.out
```

Hagamos la prueba

```
1 mathworld.wolfram.com/LinearAlgebra.html
[ 3 2 4 ]
2 mathworld.wolfram.com/Vector.html
[ ]
3 mathworld.wolfram.com/Determinant.html
[ 4 1 ]
4 mathworld.wolfram.com/Matrix.html
[ 2 3 6 7 1 ]
5 mathworld.wolfram.com/RotationMatrix.html
[ 3 4 2 1 ]
6 mathworld.wolfram.com/PositiveDefiniteMatrix.html
[ 3 8 2 1 4 ]
7 mathworld.wolfram.com/LUDecomposition.html
[ 8 4 1 9 ]
8 mathworld.wolfram.com/CholeskyDecomposition.html
[ 7 1 6 9 ]
9 mathworld.wolfram.com/QRDecomposition.html
[ 8 4 1 7 ]
10 mathworld.wolfram.com/VandermondeMatrix.html
[ 3 1 ]
```

Hagamos la prueba

```
python graph_plot.py graph.out [output.png]
```

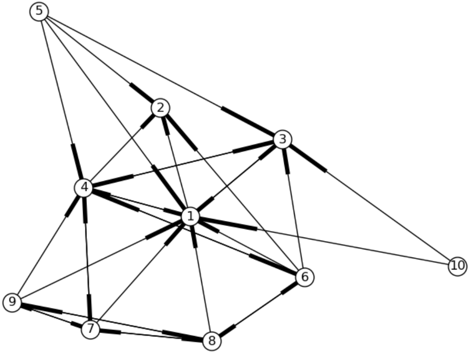


Figura: graph.out

10
33
1 3
1 2
1 4
3 4
3 1
4 2
...

Resultado

(1)	0.1941	LinearAlgebra
(2)	0.1908	Matrix
(3)	0.1428	Determinant
(4)	0.1316	Vector
(5)	0.0859	LUDecomposition
(6)	0.0734	PositiveDefiniteMatrix
(7)	0.0695	CholeskyDecomposition
(8)	0.0592	QRDecomposition
(9)	0.0261	RotationMatrix
(10)	0.0261	VandermondeMatrix

Bibliografía

- Sergey Brin and Lawrence Page, The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems*, April 1998.

  Explica el esquema general de los orígenes de este motor de búsqueda.

- Kurt Bryan and Tanya Leise, The linear algebra behind google. *SIAM Review*, 2006.
- Sepandar D. Kamvar, Taher H. Haveliwala, Christopher D. Manning, and Gene H. Golub, Extrapolation methods for accelerating pagerank computations. In *Proceedings of the 12th international conference on World Wide Web*, New York, NY, USA, 2003. ACM.

Implementación especial del método de la potencia para la matriz de PageRank

Tenemos

- $P_2 = cP_1 + (1-c)E$, $c \in (0,1)$.

y queremos $x \in \mathbb{R}^n_{\geq 0}$ y $\sum_{i=1}^n x_i = 1$ tal que $x = P_2 x$

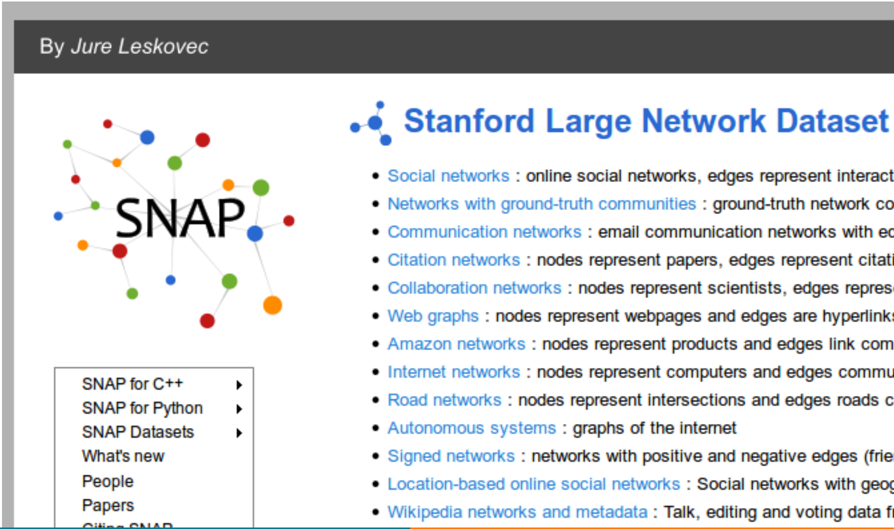
- ¿P_1 es esparsa?
- ¿P_2 es esparsa?

Sea $x = x_0$, es posible computar $y = P_2\, x$ con el siguiente algoritmo:

1. $y = c\, P\, x;$
2. $w = \|x\|_1 - \|y\|_1$
3. $y = y\, +\, w\, v$

SNAP

http://snap.stanford.edu



¿Preguntas?