

Trabajo Práctico 2

Ohhh sólo tiran π -edras...

Métodos Numéricos

Segundo cuatrimestre - 2015











Hasta ahora

- ▶ TP 1: Difusión de calor en horno cilíndrico.
- ▶ Taller 1: Eliminación de ruido en imágenes.

Objetivo

Seguir viendo aplicaciones reales de MN.

Motivación

2015									
POS	EQUIPO	PJ	PG	PE	PP	GF	GC	PTS	
1	 Boca Juniors	24	16	4	4	39	19	52	
2	 San Lorenzo	24	15	5	4	35	14	50	
3	 Rosario Central	24	12	10	2	32	20	46	
4	 Racing Club	23	12	7	4	30	19	43	
5	 River Plate	23	11	8	4	42	25	41	
6	 Independiente	24	10	11	3	33	19	41	
7	 Banfield	24	11	7	6	31	23	40	
8	 Tigre	24	11	7	6	26	19	40	
9	 Belgrano	24	11	6	7	27	20	39	
10	 Estudiantes	24	10	9	5	25	23	39	

NBA 2014-15 Conference Standings												
Western Conference						Eastern Conference						
Team	W	L	Pct	GB		Team	W	L	Pct	GB		
1 Golden State Warriors	67	15	.817	- x		1 Atlanta Hawks	60	22	.732	- x		
2 Houston Rockets	56	26	.683	11.0 x		2 Cleveland Cavaliers	53	29	.646	7.0 x		
3 Los Angeles Clippers	56	26	.683	11.0 x		3 Chicago Bulls	50	32	.610	10.0 x		
4 Portland Trail Blazers	51	31	.622	16.0 x		4 Toronto Raptors	49	33	.598	11.0 x		
5 Memphis Grizzlies	55	27	.671	12.0 x		5 Washington Wizards	46	36	.561	14.0 x		
6 San Antonio Spurs	55	27	.671	12.0 x		6 Milwaukee Bucks	41	41	.500	19.0 x		
7 Dallas Mavericks	50	32	.610	17.0 x		7 Boston Celtics	40	42	.488	20.0 x		
8 New Orleans Pelicans	45	37	.549	22.0 x		8 Brooklyn Nets	38	44	.463	22.0 x		
9 Oklahoma City Thunder	45	37	.549	22.0		9 Indiana Pacers	38	44	.463	22.0		
10 Phoenix Suns	39	43	.476	28.0		10 Miami Heat	37	45	.451	23.0		
11 Utah Jazz	38	44	.463	29.0		11 Charlotte Hornets	33	49	.402	27.0		
12 Denver Nuggets	30	52	.366	37.0		12 Detroit Pistons	32	50	.390	28.0		
13 Sacramento Kings	29	53	.354	38.0		13 Orlando Magic	25	57	.305	35.0		
14 Los Angeles Lakers	21	61	.256	46.0		14 Philadelphia 76ers	18	64	.220	42.0		
15 Minnesota Timberwolves	16	66	.195	51.0		15 New York Knicks	17	65	.207	43.0		

Ranking, Nombre, País	Puntos	Movimiento	Torneos Jugados
1 Djokovic, Novak (SRB)	16.145	0	18
2 Federer, Roger (SUI)	9.405	0	18
3 Murray, Andy (GBR)	8.660	0	23
4 Wawrinka, Stan (SUI)	6.000	1	22
5 Berdych, Tomas (CZE)	5.050	1	21
6 Nishikori, Kei (JPN)	5.015	-2	22
7 Nadal, Rafael (ESP)	3.770	1	22
8 Ferrer, David (ESP)	3.695	-1	21
9 Raonic, Milos (CAN)	2.790	1	22
10 Simon, Gilles (FRA)	2.560	1	24
11 Gasquet, Richard (FRA)	2.490	1	20
12 Anderson, Kevin (RSA)	2.430	2	25
13 Isner, John (USA)	2.325	0	24
14 Cilic, Marin (CRO)	2.270	-5	21
15 Goffin, David (BEL)	2.115	0	26
16 Lopez, Feliciano (ESP)	1.935	3	26
17 Tsonga, Jo-Wilfried (FRA)	1.785	1	17
18 Karlovic, Ivo (CRO)	1.620	3	25
19 Dimitrov, Grigor (BUL)	1.600	-2	22
20 Thiem, Dominic (AUT)	1.555	0	29
21 Troicki, Viktor (SRB)	1.511	1	31
22 Bautista Agut, Roberto (ESP)	1.510	1	28
23 Tomic, Bernard (AUS)	1.510	1	29

Motivación

3 de may	Final	Boca Juniors	2-0	River Plate
10 de may	Final	River Plate	0-0	Racing Club
31 de may	Final	River Plate	2-0	Rosario Central
7 de jun	Final	Olimpo de Bahía Blanca	1-1	River Plate
8 de jul	Final	Tigre	0-0	River Plate
11 de jul	Final	River Plate	1-1	Temperley
18 de jul	Final	Atletico Rafaela	1-5	River Plate
25 de jul	Final	River Plate	3-1	Colón de Santa Fe
2 de ago	Post	Defensa y Justicia	P - P	River Plate
11 de ago	Final	Gamba Osaka	0-3	River Plate
17 de ago	Final	River Plate	0-1	San Martín de San Juan
23 de ago	Final	Estudiantes La Plata	2-1	River Plate
30 de ago	Final	River Plate	1-1	Huracán
6 de sep	Final	Nueva Chicago	1-4	River Plate
13 de sep	Final	River Plate	0-1	Boca Juniors

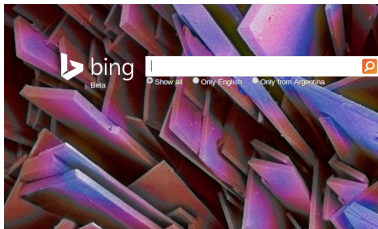
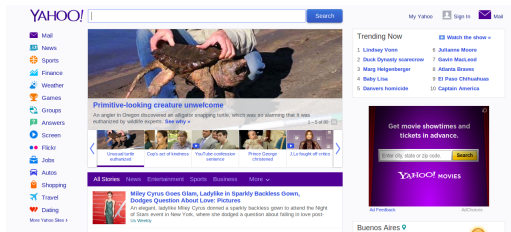
10 de may	Final	River Plate	0-0	Racing Club
24 de may	Final	Racing Club	1-0	Independiente
31 de may	Final	Aldosivi	1-2	Racing Club
7 de jun	Final	Racing Club	3-1	Vélez Sarsfield
12 de jul	Final	Newell's Old Boys	3-0	Racing Club
18 de jul	Final	Racing Club	2-1	Sarmiento de Junín
26 de jul	Final	Quilmes	2-1	Racing Club
2 de ago	Final	Racing Club	0-0	Belgrano de Córdoba
15 de ago	Final	Unión de Santa Fe	1-2	Racing Club
22 de ago	Final	Racing Club	2-1	Arsenal de Sarandí
30 de ago	Susp	Godoy Cruz de Mendoza	0-1	Racing Club
6 de sep	Final	Racing Club	2-0	Gimnasia La Plata
12 de sep	Final	Independiente	3-0	Racing Club

Motivación

Por qué es importante el ranking?

- ▶ Determina quién es el *mejor* del torneo.
- ▶ Clasifica a copas etapas posteriores (playoffs) y/o competencias internacionales (Libertadores, Masters, etc).
- ▶ *Justicia deportiva*.
- ▶ Mucha mucha plata en juego.

Motores de búsqueda



Motores de búsqueda

- ▶ Explorar la red e identificar todas las páginas con acceso público.
- ▶ Almacenar la información obtenida, para realizar búsquedas eficientemente.

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.



The screenshot shows a Google search interface with the query "messi" entered in the search bar. Below the search bar, there are tabs for "Web", "Imágenes", "Videos", "Noticias", "Más", and "Herramientas de búsqueda". The "Web" tab is selected, and the results show "Cerca de 221.000.000 resultados (0,16 segundos)". The first result is titled "Noticias de messi" and includes a thumbnail image of Lionel Messi. The second result is titled "La admiración por Lionel Messi también llegó a Estados Unidos" from Clarín.com, dated 37 minutes ago. The third result is titled "Las bromas antes del clásico español pasan por cómo frenar a Messi" from Clarín.com, also dated 37 minutes ago. The fourth result is titled "Las lecciones de Messi a Neymar para evitar que lo cosan a patadas cada pa..." from ecdiario, dated 3 hours ago. The fifth result is titled "Lionel Messi - Wikipedia, la enciclopedia libre" and provides a brief biography of Messi. The sixth result is titled "Punto positivo: Lionel Messi volvió con un gol en el empate de ..." from canchallena.lanacion.com.ar, dated 10 hours ago. The seventh result is titled "Los goles de Messi, decisivos en Champions | Barca | Sport.es" from www.sport.es, dated 2774050 seconds ago.

Que características son deseables para un *ranking*?

Outline

Contexto TP2

Cadenas de Markov

Algoritmo PageRank

Aplicación: rankings en competencias deportivas

Enunciado

Cadenas de Markov

Definición

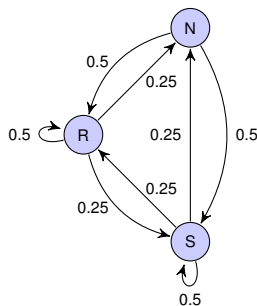
Consideramos un conjunto de estados $S = \{s_1, s_2, \dots, s_r\}$. El proceso empieza en alguno de estos estados y se mueve de un estado a otro. A cada movimiento se lo denomina *paso*. Si la cadena se encuentra actualmente en el estado s_i , en el siguiente paso se mueve al estado s_j con probabilidad p_{ij} . Esta probabilidad no depende de los estados anteriores a s_i en los que se haya encontrado el proceso.

Cadenas de Markov

Ejemplo: Cambio de clima

- ▶ Tres posibilidades: Bueno (N), Lluvioso (R), Nieve (S).
- ▶ p_{ij} es la probabilidad de que si en un determinado día estamos en un estado i (i.e., N, R ó S) al día siguiente estemos en el estado j .
- ▶ Particularidad: no pueden haber dos días buenos (N) seguidos.

Grafo de transiciones:



Matriz de transiciones:

$$P = \begin{matrix} & \begin{matrix} R & N & S \end{matrix} \\ \begin{matrix} R \\ N \\ S \end{matrix} & \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{pmatrix} \end{matrix}$$

- ▶ Filas: Estado actual.
- ▶ Columnas: Estado al que podemos movernos.
- ▶ Matriz estocástica por filas.

Cadenas de Markov

Mirando más allá de un día

Nuevo problema

Queremos saber cuál es la probabilidad que, si hoy está lluvioso, nieve dentro de dos días. Llamamos a esta probabilidad $p_{RS}^{(2)}$.

Esto es la unión disjunta de los siguientes eventos:

1. Lluvioso (R) mañana y nieve (S) pasado.
2. Bueno (N) mañana y nieve (S) pasado.
3. Nieve (S) mañana y nieve (S) pasado.

$$P = \begin{matrix} & \begin{matrix} R & N & S \end{matrix} \\ \begin{matrix} R \\ N \\ S \end{matrix} & \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{pmatrix} \end{matrix} \quad p_{RS}^{(2)} = \underbrace{p_{11}p_{13}}_1 + \underbrace{p_{12}p_{23}}_2 + \underbrace{p_{13}p_{33}}_3$$

Cadenas de Markov

En general

En el caso anterior,

$$p_{ij}^{(2)} = \sum_{k=1}^r p_{ik} p_{kj} = (P^2)_{ij}.$$

Propiedad

El resultado de multiplicar dos matrices estocásticas por filas es una matriz estocástica por filas.

Teorema

Sea P la matriz de transición de una cadena de Markov. El elemento $p_{ij}^{(k)}$ de la matriz P^k es la probabilidad de que la cadena de Markov, empezando en el estado i , se encuentre en el estado j después de k pasos.

Cadenas de Markov

Y si no conocemos el estado actual?

Hasta ahora, supusimos que conocemos el estado actual. Qué pasa si la cadena se encuentra en algún estado con una probabilidad?

Definición: vector de probabilidades

$x \in \mathbb{R}^k$ es un vector (fila) de probabilidades si $x_i \geq 0$ y $\sum_{i=1}^k x_i = 1$.

Teorema

Sea P la matriz de transición de una cadena de Markov, y sea u el vector que representa la distribución inicial. Entonces, la probabilidad de que la cadena se encuentre en el estado s_i luego de k pasos es la componente i -ésima del vector

$$u^{(k)} = uP^k$$

Cadenas de Markov

Estado estacionario: qué pasa en el largo plazo

Qué sucede con el sistema si consideramos

$$\lim_{n \rightarrow \infty} P^n?$$

Definición: Matriz Regular

Una matriz de transiciones P se dice regular si P^k tiene solamente entradas positivas para algún entero k .

Teorema

Sea P una matriz de transiciones regular. Entonces:

- ▶ $\lim_{n \rightarrow \infty} P^n = W$, donde todas las filas de W son un mismo vector w .
- ▶ $wP = w$, y todos los vectores que cumplan $vP = v$ son un múltiplo de w .
- ▶ $xP^n \rightarrow w$ con $n \rightarrow \infty$.

Cadenas de Markov

Estado estacionario: qué pasa en el largo plazo

Si P es una matriz de transiciones regular, entonces:

- ▶ 1 es un autovalor de P .
- ▶ Hay un único vector de probabilidades que es el autovector asociado al autovalor 1, y es w .
- ▶ Se demuestra que los demás autovalores cumplen $1 = \lambda_1 > |\lambda_2| \geq \dots \geq |\lambda_r|$.

Interpretación

Al vector de probabilidades w se lo denomina *estado estacionario*. La componente w_i representa la proporción de tiempo que la cadena se encuentra, en el largo plazo, en el estado s_i .

En la práctica

Como $wP = w$, entonces $P^t w^t = w^t$. Podemos intentar usar el método de la potencia para calcular w^t .

PageRank

Problema

- ▶ Tenemos un conjunto de páginas $Web = \{1, \dots, n\}$.
- ▶ El objetivo es asignar a cada una de ellas un puntaje que determine la importancia relativa de la página respecto de las demás.
- ▶ Vamos a trabajar directamente sobre la matriz traspuesta.
- ▶ Si definimos una cadena de Markov regular, entonces el estado estacionario nos dará la proporción de tiempo que el navegante aleatorio pasará en cada página.

PageRank

Modelo inicial

Modelo mediante cadenas de Markov

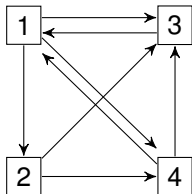
Consideramos el modelo del *navegante aleatorio*, que comienza en una página cualquiera del conjunto y va navegando a través de sus links.

- ▶ Cada página representa un estado de la cadena.
- ▶ Podemos pasar de una página j a otra i si hay un link de j a i . Definimos $W \in \{0, 1\}^{n \times n}$ como $w_{ij} = 1$ si hay un link de j a i , y $w_{ij} = 0$ en caso contrario.
- ▶ $n_j = \sum_{i=1}^n w_{ij}$ es el grado de la página j (cantidad de links salientes).
- ▶ Definimos $P \in \mathbb{R}^{n \times n}$ como $P_{ij} = 1/n_j$ como la probabilidad de ir de la página j a la i , dado que existe un link de j a i .

PageRank

Ejemplo (Bryan y Leise)

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



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

P es estocástica por columnas.

Pregunta:

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

PageRank

Solución a *dangling nodes*

Definimos:

- ▶ $v \in \mathbb{R}^n$, $v_i = 1/n$.
- ▶ $d \in \{0, 1\}^n$, $d_i = 1$ si $n_i = 0$, $d_i = 0$ en caso contrario.
- ▶ $D = vd^t$
- ▶ $P_1 = P + D$

Idea

Si estamos en una página sin links salientes, entonces con probabilidad uniforme $1/n$ el navegante pasa a cualquiera de las páginas en *Web*.

Pregunta:

Ahora la matriz es estocástica por columnas. Es regular?

PageRank

Asegurando regularidad

Depende del grafo de conectividad. Sin embargo, podemos extender la idea anterior en general a todas las páginas. A este fenómeno se lo denomina *teletransportación*.

- ▶ $\vec{1} = (1, \dots, 1) \in \mathbb{R}^n$.
- ▶ $E = v\vec{1}^t$.
- ▶ $P_2 = cP_1 + (1 - c)E$, $c \in (0, 1)$.
- ▶ P_2 es estocástica por columnas y $(P_2)_{ij} > 0$, $1 \leq i, j \leq n$.

Finalmente

Tenemos una cadena de Markov que modela el problema y cumple todas las condiciones. Para generar el ranking de las páginas, buscamos un autovector w asociado al autovalor 1 de P_2 , tal que $P_2 w = w$, y w sea un vector de probabilidades.

Aplicando PageRank en competencias

Similitudes con el ranking web

- ▶ El objetivo final es el mismo: obtener un ranking.
- ▶ Si pensamos los equipos a las páginas web, los enfrentamientos entre ellos establecen una *relación de importancia*.
- ▶ El modelo captura naturalmente asimetrías en los fixtures.
- ▶ También es posible aplicarlo al caso donde los equipos juegan la misma cantidad de veces contra todos los demás.

Observación

Existen distintos modelos en la literatura que adaptan PageRank e ideas similares para rankear equipos.

Aplicando PageRank en competencias

Modelo a estudiar: GeM (Govan et al., 2008)

Contexto

- ▶ Tenemos un conjunto de competidores (equipos), que juegan entre sí durante un período de tiempo determinado (temporada, año, semestre, etc.).
- ▶ Se asume que no hay empates, o que los mismos son muy poco frecuentes.
Ejemplos: Basket, Footbal Americano, Tenis.

Definiciones básicas

- ▶ Consideramos el conjunto de equipos $\{1, \dots, n\}$.
- ▶ La temporada se representa mediante un grafo donde cada equipo representa un nodo y existe un link de i a j si el equipo i perdió al menos una vez con el equipo j .
- ▶ *Relación con PageRank*: equipos \approx páginas, links \approx resultados.

Aplicando PageRank en competencias

GeM: Definiciones (1/2)

Paso 1

Se define la matriz $A^t \in \mathbb{R}^{n \times n}$

$$A_{ji}^t = \begin{cases} w_{ji} & \text{si el equipo } i \text{ perdió con el equipo } j, \\ 0 & \text{en caso contrario,} \end{cases}$$

donde w_{ji} es la diferencia absoluta en el marcador.

Obs.

En caso de que i pierda más de una vez con j , w_{ji} representa la suma acumulada de diferencias.

Obs.

Notar que A^t es una generalización de la matriz de conectividad W definida en la sección anterior.

Aplicando PageRank en competencias

GeM: Definiciones (2/2)

Paso 2

Definir la matriz $H_{ji}^t \in \mathbb{R}^{n \times n}$ como

$$H_{ji}^t = \begin{cases} A_{ji}^t / \sum_{k=1}^n A_{ki}^t & \text{si hay un link } i \text{ a } j, \\ 0 & \text{en caso contrario.} \end{cases}$$

Paso 3

Tomar $P = H^t$, y aplicar el método PageRank como fue definido previamente, siendo π la solución a la ecuación $P_2 x = x$.

Obs.

Notar que los páginas sin links salientes, en este contexto se corresponden con aquellos equipos que se encuentran invictos.

Ranking final

Utilizar los puntajes obtenidos en π para ordenar los equipos.

Aplicando PageRank en competencias

Ejemplo (Govan et al., 2008)

```
data NFL2007EXAMPLE;
```

```
    Input Team_A_Index Score_A Tea  
    datalines;
```

```
1 16 4 13
```

```
2 38 5 17
```

```
2 28 6 23
```

```
3 34 1 21
```

```
3 23 4 10
```

```
4 31 1 6
```

```
5 33 6 25
```

```
5 38 4 23
```

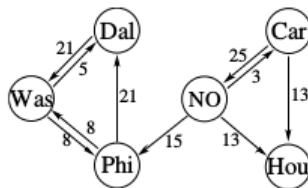
```
6 27 2 6
```

```
6 20 5 12
```

```
;
```

```
run;
```

```
data indexTeam;  
    Input Team $3. Index;  
    datalines;  
    Car 1  
    Dal 2  
    Hou 3  
    NO 4  
    Phi 5  
    Was 6  
    ;  
run;
```



TP2

Objetivos generales

- ▶ Trabajar sobre una aplicación real, implementando prototipos de algoritmos relevantes utilizados en la práctica.
- ▶ Simular un trabajo de investigación:
 - ▶ Relevamiento de literatura (qué hay hecho).
 - ▶ Desarrollo de algoritmos para el problema.
 - ▶ Decisiones de implementación.
 - ▶ Experimentación, en dos contextos distintos de aplicación.

TP2

Contexto: Rankings de páginas web

1. Considerar los trabajos de Bryan y Leise [2], Bryn y Page [?] y Kamvar et al. [4, Algoritmo 1]. Para este último, limitarse a la introducción y la mejora propuesta para calcular $x^{(k+1)} = P_2 x^{(k)}$.
2. Implementar el algoritmo IN-DEG, que ordena por cantidad de páginas que apuntan.
3. La matriz suele ser esparsa. Considerar como estructuras para la implementación: *Dictionary of Keys* (dok), *Compressed Sparse Row* (CSR), *Compressed Sparse Column* (CSC). Elegir una y justificar.
4. Estudiar la convergencia de PageRank, analizando la evolución de la norma Manhattan (norma L_1) entre dos iteraciones sucesivas. Comparar los resultados obtenidos para al menos dos instancias de tamaño mediano-grande, variando el valor de c .
5. Estudiar cualitativamente los rankings obtenidos por los dos métodos. Analizar los resultados individualmente en una primera etapa, y luego realizar un análisis comparativo entre los dos rankings obtenidos.
6. Conjunto provisto en SNAP [1], con redes de tamaño grande obtenidos a partir de datos reales. Instancias pequeñas generadas por el grupo.

TP2

Contexto: Rankings de equipos

1. Analizar el modelo propuesto en Govan et al. [3].
2. Considerar al menos un conjunto de datos reales, con los resultados de cada fecha para alguna liga de algún deporte.
3. Notar que el método GeM asume que no se producen empates entre los equipos (o que si se producen, son poco frecuentes). En caso de considerar un deporte donde el empate se da con cierta frecuencia no despreciable (por ejemplo, fútbol), es fundamental aclarar como se refleja esto en el modelo y analizar su eventual impacto.
4. Realizar experimentos variando el parámetro c , indicando como impacta en los resultados. Analizar la evolución del ranking de los equipos a través del tiempo, evaluando también la evolución de los rankings e identificar características/hechos particulares que puedan ser determinantes para el modelo, si es que existe alguno.
5. Comparar los resultados obtenidos con los reales de la liga utilizando el sistema estándar para la misma.

Implementación

Reutilizar la de páginas web, o una nueva (no esparsa) en C++, MATLAB, PYTHON. Para este contexto pueden utilizar librerías.

Material extra (optativo)

Para generar las instancias, se adjunta un código Python que, dada una lista de direcciones de páginas web, parsea el código html de cada una de ellas y genera el grafo de conectividad.

Algunas aclaraciones

- ▶ Se restringe a links entre las páginas de la lista. El resto de los links son descartados.
- ▶ El chequeo para decidir si un link es o no a una página de la lista es básico (ejemplo: `www.example.com`, ó `example.com`, ó `example.com.ar` son considerados links distintos)
- ▶ Links que aparezcan dos o más veces son contados una única vez.
- ▶ Pueden tomar este código y modificarlo según sus necesidades.
- ▶ Si encuentran algún error en el código, por favor contacten a los docentes.

TP2

Material extra (optativo)

Utilización

El comando

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

toma como entrada la lista de páginas y genera el grafo, con el formato indicado en el enunciado del trabajo, en el archivo graph.out.

Además

- ▶ Archivos con los resultados del Torneo de Primera División AFA, hasta la fecha 23 inclusive.
- ▶ Más datos deportivos en `datahub.io`

TP2

Recomendaciones

- ▶ Viernes 25/09: Implementación matriz esparsa, método de la potencia, estudio PageRank y modelo GeM.
- ▶ Viernes 02/10: Implementación GeM, experimentos páginas web, experimentos competencia deportiva.
- ▶ Martes 06/10: Entrega TP2 (electrónica)

Trabajo Práctico

Fecha de entrega

- ▶ **Formato Electrónico:** Martes 6 de Octubre de 2015, hasta las 23:59 hs, enviando el trabajo (informe + código) a la dirección **metnum.lab@gmail.com**. El subject del email debe comenzar con el texto **[TP3]** seguido de la lista de apellidos de los integrantes del grupo.
- ▶ **Formato físico:** Miercoles 7 de Octubre de 2015, 17 hs., en la clase práctica.

Importante

El horario es estricto. Los correos recibidos después de la hora indicada serán considerados re-entrega.

Bibliografía



Stanford large network dataset collection.

<http://snap.stanford.edu/data/#web>.



Kurt Bryan and Tanya Leise.

The linear algebra behind google.

SIAM Review, 48(3):569–581, 2006.



Angela Y. Govan, Carl D. Meyer, and Russell Albright.

Generalizing google's pagerank to rank national football league teams.

In Proceedings of SAS Global Forum 2008, 2008.



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, WWW '03, pages 261–270, New York, NY, USA, 2003. ACM.