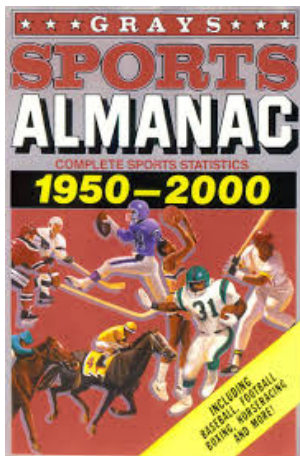


# Métodos Numéricos - Trabajo Práctico 1

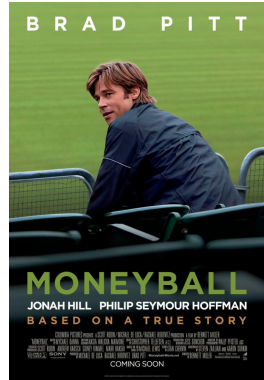
*El gran tp*



# Motivación







## Sports Analytics

Utilización de técnicas estadísticas, inteligencia artificial y optimización en el deporte.



# Motivación

## Confección de rankings y tablas de posiciones.

ATP WORLD TOUR				SCORES	STATS	RANKINGS	PLAYERS
Emirates ATP RANKINGS				Rankings Home		Singles	S
RANKING ^	MOVE ^	COUNTRY ^	PLAYER ^				
1	-		Novak Djokovic				
2	-		Andy Murray				
3	-		Roger Federer				
4	-		Stan Wawrinka				
5	-		Rafael Nadal				
6	-		Kei Nishikori				

2015-2016 CONFERENCE REGULAR SEASON STANDINGS					
EASTERN CONFERENCE					
Eastern	W	L	PCT	GB	CONF
Cleveland <sup>1</sup>	48	19	0.716	0.0	27-12
Toronto <sup>2</sup>	45	21	0.682	2.5	30-11
Miami <sup>3</sup>	39	28	0.582	9.0	24-18
Atlanta <sup>4</sup>	39	29	0.574	9.5	23-18
Boston <sup>5</sup>	39	29	0.574	9.5	26-18
Charlotte <sup>6</sup>	38	29	0.567	10.0	24-15
Indiana <sup>7</sup>	36	31	0.537	12.0	23-17
Detroit <sup>8</sup>	34	34	0.500	14.5	21-20
Chicago	33	33	0.500	14.5	20-21
Washington	32	35	0.478	16.0	23-20
Orlando	29	38	0.433	19.0	16-22
Milwaukee	29	39	0.426	19.5	19-22
New York	28	41	0.406	21.0	17-25
Brooklyn	19	48	0.284	29.0	10-28
Philadelphia <sup>o</sup>	9	58	0.134	39.0	3-39
WESTERN CONFERENCE					
Western	W	L	PCT	GB	CONF
Golden State <sup>1p</sup>	61	6	0.910	0.0	36-4
San Antonio <sup>2sw</sup>	57	10	0.851	4.0	35-5
Oklahoma City <sup>3</sup>	46	22	0.676	15.5	31-11
L.A. Clippers <sup>4</sup>	43	24	0.642	18.0	23-16
Memphis <sup>5</sup>	39	29	0.574	22.5	23-19
Portland <sup>6</sup>	35	33	0.515	26.5	23-18
Houston <sup>7</sup>	34	34	0.500	27.5	23-20
Dallas <sup>8</sup>	34	34	0.500	27.5	21-19

# Motivación

Confección de rankings y tablas de posiciones.

## 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). *Drafts*.
- ▶ Mucha mucha plata en juego.
- ▶ Rankear cosas es un problema mucho más abarcativo y con ideas extrapolables. (Ranqueo de páginas web, ranqueo de publicidades).

## Preguntas

- ▶ Qué métodos/reglas/procedimientos conocen?
- ▶ Cómo son las competencias/torneos?

# Motivación

Confección de rankings y tablas de posiciones.

## Algunos ejemplos

- ▶ En algunas ligas, no todos los equipos se enfrentan todos contra todos (NFL, MLB, NCAA).

To Get to a 9-7 Tiebreaker with Green Bay for Wild Card Spot #2, the Bucs need these results:

- Dallas over Detroit on Monday Night, Week 16
- Tampa Bay over Carolina, Week 17
- New York Giants over Washington, Week 17
- Detroit over Green Bay, Week 17

To Win the Resulting 9-7 Tiebreaker with Green Bay, the Bucs need these results:

- Kansas City over Denver, Sunday Night, Week 16
- San Francisco over Seattle, Week 17
- Indianapolis over Jacksonville, Week 17
- Dallas over Philadelphia, Week 17
- Tennessee over Houston, Week 17

- ▶ Aún cuando se enfrentan todos contra todos, no lo hacen la misma cantidad de veces (Ej: NBA, 30 equipos, 82 partidos en temporada regular).
- ▶ Caso raro: Torneo de Primera División AFA.

## Pregunta

Cómo podemos manejar estos casos?

# Colley Matrix Method (CMM)

## Descripción general

### Algunos objetivos

- ▶ Un método simple, que capture la complejidad del problema.
- ▶ Solo utiliza victorias y derrotas, dejando de lado los marcadores.
- ▶ Reproducible.
- ▶ Incorpora la dificultad del schedule de cada equipo.
- ▶ Asume que el empate no es un resultado posible (solo victorias/derrotas).

### Idea general

Dados los resultados obtenidos por un equipo, obtener la probabilidad de que el equipo gane el próximo partido.

# Colley Matrix Method (CMM)

Primer paso: estimador para victoria en el próximo partido

## Laplace rule of succession

Consideremos  $k$  ensayos con dos resultados posibles: éxito (victoria) y fracaso (derrota). Sean  $s$  el número de éxitos obtenidos. En algunas circunstancias,  $(s + 1)/(k + 2)$  es un mejor estimador que  $s/k$  de que el próximo ensayo sea exitoso.

## Ejemplo

Supongamos que todos los ensayos fueron exitosos. Entonces,  $s/k = k/k = 1$ , y no deja lugar a que el ensayo sea fallido. Para valores grandes de  $k$ , ambos estimadores se comportan de forma similar.

# Colley Matrix Method (CMM)

## Segundo paso: notación

- ▶  $\Gamma = \{1, 2, \dots, T\}$  el conjunto de equipos.
- ▶ Dado  $i \in \Gamma$ , llamamos:
  - ▶  $n_i$  al total de partidos jugados,
  - ▶  $w_i$  la cantidad de partidos ganados, y
  - ▶  $l_i$  la cantidad de partidos perdidos.
- ▶ Dados  $i, j \in \Gamma$ , llamamos  $n_{ij}$  a la cantidad de partidos jugados entre  $i$  y  $j$ . Notar que  $n_{ij} = n_{ji}$ .

El estimador para la probabilidad de que el equipo  $i$  gane el próximo partido es

$$r_i = \frac{1 + w_i}{2 + n_i} = \frac{1 + w_i}{2 + w_i + l_i}$$



# Colley Matrix Method (CMM)

## Tercer paso: incorporando el schedule

Sabemos que:

- ▶  $n_i = w_i + l_i$ .
- ▶ Si no tenemos información sobre los equipos, podemos pensar que  $r_i = 1/2$  para  $i \in \Gamma$ .
- ▶ Notar que  $n_i$  puede incluir más de un partido contra un mismo equipo. Llamamos  $r_j^i$  al rating del  $j$ -ésimo oponente de  $i$ .

Reescribimos

$$\begin{aligned}w_i &= (w_i - l_i)/2 + n_i/2 \\&= (w_i - l_i)/2 + \sum_{j=1}^{n_i} 1/2 \\&\approx (w_i - l_i)/2 + \sum_{j=1}^{n_i} r_j^i\end{aligned}$$

# Colley Matrix Method (CMM)

Último paso: armamos el sistema

El rating de un equipo depende de los ratings contra los que jugó.

$$r_i = \frac{1 + w_i}{2 + n_i} \quad \text{y} \quad w_i = \frac{w_i - l_i}{2} + \sum_{j=1}^{n_i} r_j^i$$

Despejando, tenemos que

$$(2 + n_i)r_i - \sum_{j=1}^{n_i} r_j^i = 1 + \frac{w_i - l_i}{2} \quad \text{para } i \in \Gamma$$

Esto nos lleva a un sistema  $Cr = b$ , con  $C \in \mathbb{R}^{T \times T}$ ,  $b \in \mathbb{R}^T$ , con

$$C_{ij} = \begin{cases} -n_{ij} & \text{si } i \neq j, \\ 2 + n_i & \text{si } i = j. \end{cases}$$

y  $b_i = 1 + (w_i - l_i)/2$ ,  $i \in \Gamma$ .

# Colley Matrix Method (CMM)

## Algunos comentarios generales

- ▶  $C$  es lo que se conoce como *Matriz de Colley*.
- ▶ La matriz  $C$  es *Simétrica y Definida Positiva* (SDP)
- ▶ Dada una secuencia de partidos y sus resultados, podemos formular el sistema, obtener los ratings  $r_i$  de cada equipo y ordenarlos en forma decreciente.

## Preguntas

- ▶ Qué necesitamos para que el método *funcione*?
- ▶ Se les ocurre algún problema que pueda surgir con el método?

# Aplicando CMM en la práctica

Ejemplo (Govan et al., 2008)

```
data NFL2007EXAMPLE;  
  Input Team_A_Index Score_A Team_B_Index Score_B;  
  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;
```

# TP1

## 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.
- ▶ Utilizar datos reales.

# TP1: Implementación

Consideramos dos métodos para resolver el sistema

- ▶ Eliminación Gaussiana (EG).
- ▶ Factorización de Cholesky (CL)

Especificar en el desarrollo alternativas consideradas para la representación de las matrices e implementación de los métodos.

## Recordar

La implementación no es lo único que nos importa.

# TP1: Experimentación

## Análisis cuantitativo

- ▶ Comparar el tiempo de cómputo requerido por EG y CL en función de los tamaños a resolver.
- ▶ En qué contexto conviene utilizar CL? Analizar ventajas. Se pueden usar instancias artificiales.
- ▶ En todos los casos, justificar elecciones y decisiones tomadas.

## Análisis cualitativo

- ▶ Comparar CMM vs. *Winning Percentage* (WP) sobre datos reales. Identificar características y situaciones distintivas.
- ▶ El método CMM es *justo*?
- ▶ Tenemos resultados y un equipo. Determinar una estrategia que obtener la mayor posición posible *buscando* minimizar la cantidad de partidos ganados.

# TP1

## Material extra (optativo)

### Datos reales

- ▶ Datos con los resultados de los partidos del circuito ATP de 2015 (tomados de [3])
- ▶ Datos con los resultados de la temporada regular NBA 2016 hasta el 15/03 (tomados de [4])

### Además

- ▶ Dos scripts en `python` para transformar archivos con los formatos de [3] y [4] al formato del TP.
- ▶ Más datos deportivos en [2] (sin scripts), [3], [4].
- ▶ Tablas de posiciones de la NBA en fechas determinadas en [1].



# TP1

## Recomendaciones

### Importante

El TP no es solamente código. Hay que experimentar. Discutir. Volver a experimentar. Y escribir un reporte detallado.

### Sugerencia

- ▶ Viernes 31/3: EG, lectura de datos, armado sistema.
- ▶ Viernes 7/4: Experimentos cualitativos, planteo de experimentos cuantitativos.
- ▶ Viernes 14/4: Entrega TP.

# Trabajo Práctico

Fecha de entrega

- ▶ **Formato Electrónico: Viernes** 14 de Abril de 2017, 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:** Lunes 17 de Abril de 2017, 18:00 hs., en la clase de laboratorio.

## Importante

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

# Bibliografía



## Basketall reference.

`http://www.basketball-reference.com/friv/standings.cgi.`



## Datahub.

`http://datahub.io.`



## Jeff sackmann atp tennis rankings.

`http://github.com/JeffSackmann/tennis_atp.`



## Massey ratings.

`http://masseyratings.com/data.php.`