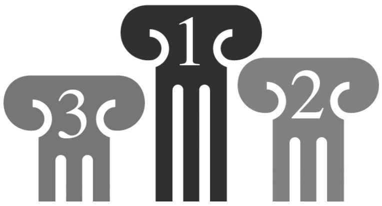
DEPARTAMENTO DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

TP1 - Sistemas de Ecuaciones para Rankings

Métodos Numéricos



Integrante	LU	Correo electrónico
Bustamante, Luis	43/18	luisbustamante097@gmail.com
Rodriguez, Leandro	521/17	leandro21890000@gmail.com
Venegas, David	783/18	venegasr.david@gmail.com

Resumen: *En el presente trabajo se introducen y estudian 3 diferentes sistemas de Ranking: Colley Matrix Method, Winning Percentage y Score Ratio para la asignación de ratings deportivos, junto a un análisis cualitativo y cuantitativo de los mismos, además de experimentos sobre la relevancia en el orden de eventos, la independencia de eventos, su complejidad temporal.*

Palabras Clave: Matriz de Colley, Winning Percentage, Score Ratio, Sistemas de Ranking.

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2610 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (++54 +11) 4576-3300

http://www.exactas.uba.ar

Índice

1. Introducción

La historia de las competencias deportivas esta ligada a las victorias de sus campeones y por ende a como se rankean sus equipos para determinar al ganador. Aunque algunos torneos permiten maneras simples de decidir la victoria, en los casos de ligas con *fixtures* más complicados donde no todos los equipos juegan entre si, ni la misma cantidad de veces, nace un problema de interés computacional en la estadística deportiva sobre como determinar el *ranking* correcto, lo cual será explorado en el presente trabajo.

Algunos ejemplos de los casos de estudio son aquellos controversiales donde no basta con solo tomar el historial de partidos ganados y perdidos sino la dificultad competitiva de cada encuentro como en la NBA (*National Basketball Association*), la NFL (*National Football League*) o la ATP (*Association of Tennis Professionals*). Además, se puede destacar por la popularidad de estos deportes la relevancia de determinar de forma justa y según un estándar de habilidades al ganador, por lo cual se utiliza el *ranking* como una lista ordenada de asignaciones a n equipos donde cada ordinal representa la posición de su equipo con respecto a sus competidores y el 1 es el primer lugar o campeón [1]. Estas posiciones suelen ser determinadas en base a un *rating* o medida de puntaje y el método de cálculo utilizado incidirá directamente en el ganador resultante.

En el presente proyecto para resolver el problema del *ranking* deportivo se estudiará los métodos de matrices de Colley (CMM por sus siglas en inglés), Porcentaje de Victorias (también conocido como *Winning Percentage* o WP) y *Score-Ratio* (o tambien llamado SR) para comparar su capacidad de representar correctamente el mejor equipo de un conjunto $\Gamma = \{1, 2, ..., T\}$ donde T es la cantidad de equipos.

El método de Colley[2] se basa en formular y resolver un sistema de ecuaciones lineales basado en la regla de sucesión de Laplace para aproximar la probabilidad de eventos booleanos (ganar o perder) utilizando como estimador sobre s casos exitosos de n eventos a $\frac{s+1}{n+2}$. Para esto, se construye una matriz $C \in \mathbb{R}^{T \times T}$ y un vector $b \in \mathbb{R}^T$, tal que el ranking $r \in \mathbb{R}^T$ sea la solución al sistema $Cr = b$. Dicho sistema es construido a partir de las siguientes ecuaciones:

$$C_{ij} = \begin{cases} -n_{ij} \text{ si } i \neq j \\ 2 + n_{ij} \text{ si } i = j \end{cases} \tag{1}$$

$$b_i = 1 + \frac{w_i - l_i}{2} \tag{2}$$

El método de Winning Percentage toma en cuenta solo las victorias de cada partido y no considera la dificultad de cada encuentro. El *rating* de un equipo $i \in \Gamma$ para n_i partidos jugados se calcula como el porcentaje de partidos ganados w_i más los partidos empatados t_i con una penalidad de 0,5 dividido entre el total de juegos, de tal forma que queda $P_i = \frac{w_i}{n_i}$.

El método de Score-Ratio toma en cuenta los puntajes obtenidos de cada partido para determinar un score que relacionará dicho puntaje con los puntos totales resultantes del partido, sin considerar la dificultad de cada encuentro. El *rating* de un equipo $i \in \Gamma$ para n_i partidos jugados se calcula como el diferencial de puntaje de partidos jugados $\Delta S_i j$ sobre el puntaje total de partidos jugados S_t, de tal forma que queda $R_i = \frac{\Delta S_i j}{S_t}$

Por último, en el presente proyecto se implementarán dichos métodos en C++ y se realizaran experimentos computacionales para comparar su eficacia, aplicabilidad y estabilidad de cálculos así como de las características resaltantes de los rankings resultantes.

2. Desarrollo

2.1. Dataset

Utilizaremos 4 datasets distintos referentes a 4 disciplinas distintas:

- ATP Tenis individual matches (2015) (430 jugadores)
- NBA Basketball (2016) (30 equipos)
- NFL American Footbal (2019) (32 equipos)
- NHL Hockey (2019) (31 equipos)

El primer set fue obtenido del repositorio de resultados históricos del ATP de Jeff Sackmann [3] y los otros fueron descargados de la página MasseyRatings.com [4].

La elección de estos fue a raíz de tener datos más generalizados a distintas disciplinas, en vez de utilizar un solo deporte, para así poder hacer un análisis mas generalizado de los métodos de ranking estudiados en este informe.

La carga de los datos se realiza en dos etapas, primero se arma una lista de Partidos (`vector<Partido>`) con la información pertinente del dataset, es decir: fecha, equipos que jugaron y puntaje del partido. Este proceso tiene un costo lineal en el número de partidos.

Luego de tener la lista de Partidos armada, se procede a la segunda etapa donde se arma un Diccionario de los Equipos (`map<id,Equipo>`), el cual proporciona la siguiente información a cada uno de los equipos: ID de equipo, partidos ganados, partidos perdidos, cantidad de partidos contra los otros equipos, la cantidad total de puntos obtenidos en esos partidos y la diferencia de la suma de todos los partidos jugados. La cantidad de partidos contra otros equipos esta implementado como un diccionario para tener un acceso en tiempo logarítmico a los datos necesario, aunque otra opción válida de representación pudo haber sido una matriz de TxT accesible en tiempo constante.

El costo total de esta segunda etapa tiene una complejidad de $O(\#Partidos * log(\#Equipos))$ donde el costo logarítmico adicional viene de la inserción de elementos en el diccionario mencionado anteriormente.

2.2. Sistema de Colley

El método de Colley consiste en la resolución de un sistema de ecuaciones asociado a la "matriz de Colley" la cual ya fue descripta en la ecuación (1) y (2), donde se toma $\Gamma = \{1, 2, \ldots, T\}$ como el conjunto de equipos, $C \in \mathbb{R}^{TxT}$ a la matriz de Colley y a $b \in \mathbb{R}^T$ como el vector solución. Esta matriz tiene la particularidad de ser una matriz *estrictamente diagonal dominante*, es decir, cumple que

$$|C_{ii}| > \sum_{i \neq j} |C_{ij}| \quad \forall i, j = 1, \ldots, T. \tag{3}$$

Debido a esta propiedad se puede aplicar el método de Eliminación Gausseana sin permutaciones para obtener el vector $r \in \mathbb{R}^T$ de la ecuación

$$Cr = b, \tag{4}$$

el cual tiene los resultados de ranking buscados.

Este método es una mejora de otros métodos más sencillos, pero a la vez no es muy difícil de implementar ni de entender dado que utiliza conceptos básicos de álgebra lineal; por lo que en la sección de experimentos se verá que tan acertado y justo puede llegar a ser con los datasets mencionados.

2.2.1. Algoritmo

El algoritmo usado se implementa sobre matrices de valores de punto flotante de doble precisión, es decir, usamos el tipo de dato `vector<vector<double>>` para representar la matriz. La complejidad del algoritmo es $O(n^2 * log(n)) + O(n^2) + O(n^3) = O(n^3)$, donde el $O(n^3)$ viene del algoritmo de Eliminación Gausseana.

Algorithm 1 Metodo de Colley

```
function Colley(Equipos, res)
    C ← crearMatriz(filas = T, cols = T + 1)                         ▷ O(n^2)
    for equipo_i in Equipos do                                       ▷ O(n^2 * log(n))
        for equipo_j in Equipos do                                   ▷ O(n * log(n))
            if i == j then
                C[i][j] ← 2 + equipo_i.partidos_totales
            else
                if equipo_i.cant_matches_con(equipo_j.id) ≠ 0 then   ▷ O(log(n))
                    C[i][j] ← −equipo_i.cant_matches_con[equipo_j.id]
                else
                    C[i][j] ← 0
    for equipo_i in Equipos do                                       ▷ O(n)
        C[i][T + 1] ← 1 + (equipo_i.p_ganados − equipo_i.p_perdidos)/2
    Elim_Gaussiana_Sin_Permutaciones(C, res);                        ▷ O(n^3)
    return res
```

2.2.2. Empates

El método no acepta resultados de empates nativamente. En el caso de este informe se tomaron datasets con resultados sin empates por lo que no es necesario modificar el algoritmo para contemplar estos casos. Sin embargo, una forma de mitigar esto es dar como ganador al equipo "visitante" dado que tiene mayor valor ganar de esta forma que de "local", donde a fin práctico se tomaría el equipo de la columna 1 como local y el equipo de la columna 2 como visitante. Igualmente, otra forma de permitir los empates sería que ante este suceso se dé a los dos equipos como perdedores, de tal manera que los datos para la matriz siguen siendo consistentes, solo que se le agrega una carga negativa al empate.

Cabe resaltar que el criterio de empate según su enfoque puede cambiar el resultado final de un ranking, por lo que en el presente trabajo no se abordará la temática de los empates dado que, además de lo anteriormente mencionado, se trabajará sobre datasets libres de estos por las características de los torneos *per se*.

2.2.3. Estabilidad de los cálculos

Dado que la base de Colley es poder triangular la matriz del método, el principal problema con respecto al error de calculo puede venir del algoritmo implementado en la Eliminación Gausseana. En ella se tienen muchas divisiones por cada iteración del algoritmo donde en algunos casos se tendrán divisiones por números muy chicos o multiplicaciones por números muy grandes, lo cual es un problema al momento de garantizar el menor error posible.

Ahora bien, hay distintas formas de mitigar el error durante la triangulación. Una estrategia fácilmente aplicable es el *pivoteo*, ya sea parcial o completo, esto puede hacer que se minimice el error cometido en los cálculos antes mencionados. No obstante, cabe resaltar que en el presente trabajo no se utilizarán ninguna de las estrategias mencionadas dado que no hubo necesidad de implementarlo.

2.3. Sistema de Winning Percentage

El método de Winning Percentage consiste en asignar para cada equipo $i \in \Gamma$, el valor resultante de dividir la cantidad de partidos ganados por sobre la totalidad de los partidos jugados, resultando en la siguiente expresión: $P_i = \frac{w_i}{n_i}$, donde P_i se trata del porcentaje o rating asignado al equipo i, w_i es la cantidad de partidos ganados por el equipo i, y n_i es la cantidad de partidos jugados por el equipo i.

Uno de los aspectos primordiales que resulta interesante analizar a la hora de intentar establecer un ranking aplicado al conjunto Γ es el contexto sobre el cual se va a implementar el presente modelo, ya que debido a como este se encuentra definido puede ser más efectivo en torneos de tipo round-robin o similares.

En este sentdo, el motivo por el cual hacemos esta diferenciación, es que en torneos donde no todos los equipos jueguen contra todos, podemos estar asignando un ranking que no es considerado "justo", ya que por ejemplo un equipo que juegue 1 solo partido y lo gane va a tener un ranking más alto, mientras que otro equipo que juegue varios partidos y gane más cantidad de la que pierde va a tener un ranking más bajo, solo por haber perdido ciertos partidos.

En la sección 4.5 del presente informe se detalla más el concepto de *fairness* o cuando vamos a considerar un método como "justo" o balanceado. Sin embargo, en este trabajo se analizarán competencias distintas al tipo mencionado, en las cuales se probará su desempeño frente a los otros métodos presentados

2.3.1. Algoritmo

Dado el modelo de la estructura elegido para los equipos, los datos necesarios para el cálculo del rating ya están cargados en el diccionario de `Equipos`, por lo que el acceso a estos se realiza en tiempo constante, por lo que la complejidad del algoritmo es $O(n)$, siendo n la cantidad de equipos. Se presenta a continuación el pseudocódigo del algoritmo:

Algorithm 2 Metodo de Winning Percentage

```
function WP(Equipos, res)
    i ← 0
    for equipo in Equipos do                                              ▷ O(n)
        res[i] ← equipo.partidos_ganados/equipo.partidos_totales
        i = i + 1
    return res
```

2.4. Sistema de Score Ratio

El método de score ratio se encuentra definido como la proporción entre la diferencia de puntaje obtenido por cada equipo $i \in \Gamma$ en sus juegos contra otro equipo $j \in \Gamma$, y el puntaje total de los partidos jugados por el equipo $i \in \Gamma$, resultando la siguiente expresión: $R_i = \frac{\Delta S_i j}{S_t}$, donde R_i hace referencia al rating que se asignará al equipo $i \in \Gamma$, $\Delta S_i j$ es la diferencia de puntos resultantes de los juegos entre los equipos $i, j \in \Gamma$, y ΔS_t es la sumatoria de los puntos totales obtenidos durante todos los partidos jugados por el equipo i.

La principal diferencia con el método de WP, es que ahora no solo importa la victoria o derrota de un equipo, sino que también se tiene en cuenta el puntaje obtenido de un equipo, y en particular su significancia con respecto del puntaje total, con lo cual entonces podemos considerar el rating como una medición o factor de cómo le fue a un equipo en cierto juego.

Es importante observar que en el método no estarán contemplados los empates, debido a que el caso particular de que todos los partidos del equipo $i \in \Gamma$ resulten en empate 0-0, podría resultar en una división por 0, al obtener la sumatoria de los puntos totales nula.

2.4.1. Algoritmo

Al igual que durante la implementación del algoritmo de Winning Percentage, se utiliza la misma estructura interna, en donde se lleva cuenta tanto de la diferencia de puntos entre ese equipo y todos contra los que jugó, así como también de la sumatoria de los puntos totales de todos los partidos en los que jugó.

De esta manera, utilizando esta estructura, el algoritmo para realizar el cálculo del Score Ratio tiene una complejidad lineal en la cantidad de equipos, presentando su pseudocódigo de la siguiente manera:

Algorithm 3 Metodo de Score Ratio

```
function ScoreRatio(Equipos, res)
    i ← 0
    for equipo in Equipos do                                              ▷ O(n)
        res[i] ← equipo.diferencia_de_puntos/equipo.puntos_totales
        i = i + 1
    return res
```

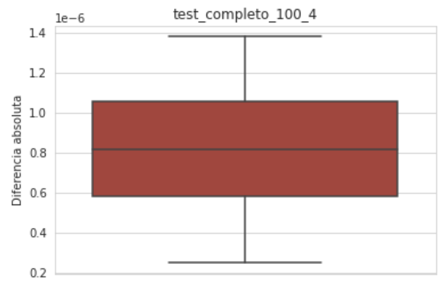
2.5. Análisis Cuantitativo

Todos los métodos pasaron exitosamente los tests de la cátedra y por ende tienen un error absoluto menor a 1×10^{-4}. El análisis de este error numérico es importante debido a que si se presentaran desviaciones lo suficientemente grandes cambiaría el resultado del ranking y no se representaría a la realidad correctamente. Ahora bien, para analizar las diferencias absolutas entre las cátedra y los archivos generados por los métodos implementados se utilizó un Jupyter Notebook. En este sentido, dado que hay casos de test con hasta mil equipos estas comparaciones pueden ser difíciles de visualizar individualmente en un histograma por lo cual se graficaron como boxplots en la Figura 1. Además, para mejorar el análisis de la diferencia absoluta se presenta en la Tabla 1 las propiedades mas relevantes sobre como se distribuye la diferencia absoluta para los casos de test más importantes.

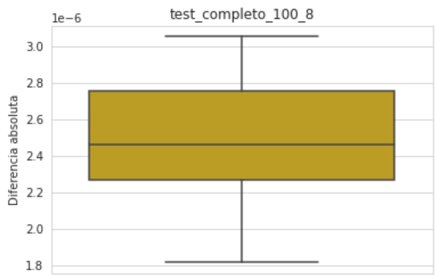
Cuadro 1: Diferencia absoluta para los casos de Test de la Cátedra `Completo 10_1`

Caso de Test	Promedio	Mediana	Desviación
Test Completo 100_4	8.2638×10^{-7}	8.1485×10^{-7}	3.0171×10^{-7}
Test Completo 100_8	2.4976×10^{-6}	2.4629×10^{-6}	3.1429×10^{-6}
Test Completo 1000_2	1.4932×10^{-5}	1.4922×10^{-5}	3.7458×10^{-7}
Test Completo 1000_8	6.7383×10^{-5}	6.7378×10^{-5}	4.5336×10^{-7}

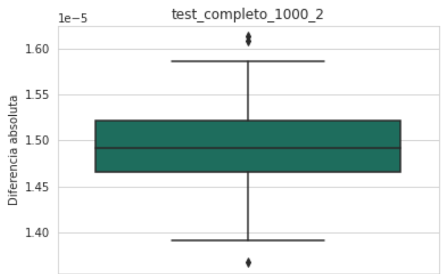
Se puede observar en la Tabla 1 que en promedio la diferencia absoluta tiene varios ordenes de magnitud por debajo del mínimo requerido que es 1×10^{-4}. Es interesante destacar que para el test con 100 equipos y 12410 tienen un orden de magnitud de 10^{-7}, es decir, la diferencia es muy muy pequeña, pero para la misma cantidad de equipos con 22318 eventos aumenta la diferencia en un orden de magnitud a 10^{-6}.



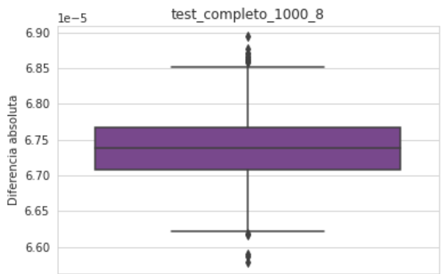
(a) Caso de Test Completo 100_4



(b) Caso de Test Completo 100_8



(c) Caso de Test Completo 1000_2



(d) Caso de Test Completo 1000_8

Figura 1: Diferencia absoluta de rating para los casos de test y el método de Colley.

Seguidamente, para el test con 100 equipos y 749273 o 2248231 eventos aumenta nuevamente la diferencia a 10^{-5}, es decir, vemos que al aumentar el dataset se replica el error numérico a medida que crece. Sin embargo, este no es significativo, y en particular para la estadística deportiva a efectos prácticos no suelen hacer rankings en base a tantos eventos como el caso de Test con 2248231 encuentros. Por otra parte, se puede apreciar que la desviación estándar es muy pequeña en todos los casos y por ende la mayoría de las diferencias están muy cerca del promedio de su espacio muestral.

2.6. Análisis Cualitativo

Para analizar el comportamiento de los métodos CMM, WP y SR se utilizaron los datasets mencionados en la Sección 2.1, de los cuales dos fueron proporcionados por la cátedra y los otros dos fueron elegidos para el trabajo. Todos estos fueron filtrados mediante scripts de Python para eliminar la información no relevante al enfoque de este trabajo y crear archivos de input para los algoritmos.

2.6.1. NBA

A continuación, se comenzará el análisis cualitativo de los métodos comparando los resultados obtenidos por los métodos de estudio con respecto al resultado oficial de la Temporada 2015-2016 de la NBA[1]. En general los resultados de Colley y WP son bastante parecidos ya que en el basketball se suele utilizar la cantidad de partidos-ganados como principal criterio para rankear equipos sin tomar en cuenta la ventaja de la victoria, y es por esto que en el método de Score-Ratio destaca otro ranking diferente donde cambia incluso el ganador del torneo, en este caso *San Antonio* en vez del oficial *Golden State*, lo cual es un buen ejemplo de como puede influir el modelo del deporte y la información utilizada para rankear en el resultado final del campeonato.

Cuadro 2: Comparación Top 20 de los rankings resultantes de la NBA

Nro.	Ranking Real	Colley	WP	Score-Ratio
1	Golden State Warrior	Golden State	Golden State	San Antonio
2	San Antonio Spurs	San Antonio	San Antonio	Golden State
3	Cleveland Cavaliers	Cleveland	Cleveland	Oklahoma City
4	Toronto Raptors	Toronto	Toronto	Cleveland
5	Oklahoma City Thunder	Oklahoma City	Oklahoma City	Toronto
6	Los Angeles Clippers	LA Clippers	LA Clippers	Boston
7	Atlanta Hawks	Miami	Boston	LA Clippers
8	Boston Celtics	Boston	Miami	Atlanta
9	Charlotte Hornets	Memphis	Memphis	Charlotte
10	Miami Heat	Atlanta	Atlanta	Indiana
11	Indiana Pacers	Charlotte	Charlotte	Miami
12	Detroit Pistons	Indiana	Indiana	Utah
13	Portland Trail Blazer	Chicago	Portland	Portland
14	Dallas Mavericks	Portland	Chicago	Detroit
15	Memphis Grizzlies	Houston	Houston	Dallas
16	Chicago Bulls	Dallas	Dallas	Houston
17	Houston Rockets	Detroit	Detroit	Chicago
18	Washington Wizards	Utah	Utah	Memphis
19	Utah Jazz	Washington	Washington	Washington
20	Orlando Magic	Milwaukee	Orlando	Orlando

Se puede apreciar en el cuadro 2 que el ranking de Colley y WP logran aproximarse al resultado oficial y comparten el Top 5 pero existen diferencias. Estas se atribuyen a que en una temporada de la NFL no solo se gana por la cantidad de partidos sino que existen playoffs y eliminatorias. Además, se divide en conferencias donde el objetivo de cada equipo es quedar en el top 8 de su conferencia y así pasar a la eliminatoria, aunque haya equipos con mayor rating en la otra conferencia pero en una posición menor a la 8^{va}.

[1] `https://www.basketball-reference.com/leagues/NBA_2016_standings.html`.

2.6.2. NFL

En esta sección se comparará el ranking oficial de la NFL[2], con los distintos rankings obtenidos utilizando los 3 métodos estudiados en el presente informe.

Cuadro 3: Comparación Top 15 de los rankings resultantes de la NFL

Nro.	Ranking Real	Colley	WP	Score-Ratio
1	Baltimore Ravens	San Francisco 49ers	Baltimore Ravens	New England Patriots
2	Green Bay Packers	Kansas City Chiefs	San Francisco 49ers	Baltimore Ravens
3	New Orleans Saints	Baltimore Raven	Kansas City Chiefs	San Francisco 49ers
4	San Francisco 49ers	New Orleans Saints	Green Bay Packers	Kansas City Chiefs
5	Kansas City Chiefs	Green Bay Packers	New Orleans Saints	Dallas Cowboys
6	New England Patriots	Seattle Seahawks	New England Patriots	New Orleans Saints
7	Seattle Seahawks	Houston Texans	Seattle Seahawks	Minnesota Vikings
8	Buffalo Bills	New England Patriot	Minnesota Vikings	Tennessee Titans
9	Houston Texans	Tennessee Titans	Houston Texans	Buffalo Bills
10	Minnesota Vikings	Minnesota Vikings	Buffalo Bills	Green Bay Packers
11	Los Angeles Rams	Los Angeles Rams	Tennessee Titans	Los Angeles Rams
12	Philadelphia Eagles	Buffalo Bills	Los Angeles Rams	Philadelphia Eagles
13	Tennessee Titans	Chicago Bears	Philadelphia Eagles	Seattle Seahawks
14	Chicago Bears	Pittsburgh Steelers	Chicago Bears	Tampa Bay Buccaneers
15	Dallas Cowboys	Atlanta Falcons	Dallas Cowboys	Los Angeles Chargers

Se puede observar que el método que más se asemeja al ranking oficial es el de Winning Percentage y por lo tanto se puede conjeturar entonces, que en una primera instancia para definir el ranking oficial, uno de los factores más importantes es categorizar los resultados de la liga según su porcentaje de victorias.

Luego, si bien el método de Colley dista un poco más del resultado oficial, podemos ver que se asemeja en gran medida, manteniendo los mismos equipos del top 5, solo que en distinto orden. De este hecho se deduce que, el ranking resultante al utilizar el método de Colley no difiere mucho del resultante por WP para el presente conjunto de partidos.

Por último, se observa que el ranking obtenido por SR dista bastante del resultado oficial, ya que lo que se pondera en ese caso es el resultado general del puntaje de los equipos, en donde equipos con menor Winning percentage se encuentran por encima de otros debido a que tienen un diferencial de puntos mayor en ciertos partidos.

2.6.3. NHL

En esta sección se comparará el ranking oficial de la NHL[3], con los distintos rankings obtenidos utilizando los 3 métodos estudiados en el presente informe.

Se puede observar, que para el ranking oficial de la NHL los resultados si bien en principio parecen similares a los obtenidos en los métodos presentados, revelan pequeñas diferencias en los 3 casos.

Esto se puede explicar debido a que el rankeo realizado por la NHL es muy particular, se utiliza un sistema de puntos similar al de Winning Percentage, pero asignando 2 puntos por victoria, 1 punto por derrota en tiempo extra, y 0 puntos por derrota.

Durante el desarrollo de los algoritmos presentados en los diferentes métodos del informe, no se realizaron distinciones entre las derrotas normales y las derrotas por tiempo extra ya que ciertas diferencias no se encontraban dentro del scope de interés.

[2] https://www.nfl.com/standings/league/2019/reg/
[3] https://www.nhl.com/standings/2018/league

Cuadro 4: Comparación Top 15 de los rankings resultantes de la NHL

Nro.	Ranking Real	Colley	WP	Score-Ratio
1	Tampa Bay	Tampa Bay	Tampa Bay	Tampa Bay
2	Calgary	Boston	Boston	Boston
3	Boston	Columbus	Calgary	Calgary
4	Washington	NY Islanders	NY Islanders	NY Islanders
5	NY Islanders	St Louis	Columbus	Columbus
6	San Jose	Calgary	Washington	Winnipeg
7	Toronto	Washington	St Louis	Toronto
8	Nashville	Carolina	Winnipeg	Washington
9	Pittsburgh	Toronto	Nashville	St Louis
10	Winnipeg	Winnipeg	Carolina	Pittsburgh
11	Carolina	Nashville	Toronto	Nashville
12	St Louis	San Jose	San Jose	Vegas
13	Columbus	Montréal	Montréal	Carolina
14	Montréal	Dallas	Dallas	Colorado
15	Dallas	Pittsburgh	Vegas	San Jose

2.6.4. ATP

A continuación, se comparan los métodos implementados para el dataset del torneo ATP de Tenis del 2015, y se compararán con los datos reales[4] del torneo.

Este dataset tiene una particularidad en cuanto a los resultados de los partidos ya que estos son del tipo 1-0 o 0-1. Inicialmente, la intuición dicta que SR va a tener un comportamiento no esperado dado que se basa en la puntuación de los equipos, pero lo sorprendente es que al momento de comparar los métodos se logra visualizar que WP y SR tienden a sobresaturar equipos a su valor máximo, es decir a `1.0`, por lo que se tienen mas de 25 equipos sobresaturados en ese valor.

Sumado a esto último, tenemos que el Top 10 de Colley ni siquiera está incluido en este grupo de equipos saturados en WP y SR como se aprecia en el Cuadro 5a, donde podemos observar que tanto en WP como en SR los equipos que en Colley estan en el Top10, están debajo de la saturación al `1.0`, es decir, debajo del Top 25 de WP y SR.

Nombre	Colley	WP	SR
Djokovic Novak	1	28	28
Federer Roger	2	29	29
Murray Andy	3	32	32
Wawrinka Stan	4	42	42
Nishikori Kei	5	35	35
Nadal Rafael	6	36	36
Berdych Tomas	7	43	43
Ferrer David	8	34	34
Gasquet Richard	9	44	44
Raonic Milos	10	46	46

(a) Comparación del ranking entre los métodos para el Top 10 de Colley

Nro.	Ranking Real	Colley
1	Djokovic Novak	Djokovic Novak
2	Murray Andy	Federer Roger
3	Federer Roger	Murray Andy
4	Wawrinka Stan	Wawrinka Stan
5	Nadal Rafael	Nishikori Kei
6	Berdych Tomas	Nadal Rafael
7	Ferrer David	Berdych Tomas
8	Nishikori Kei	Ferrer David
9	Gasquet Richard	Gasquet Richard
10	Tsonga Jo-Wilfried	Raonic Milos

(b) Comparación del Top 10 entre los resultados reales y los de Colley para ATP

Cuadro 5: Comparaciones de los resultados obtenidos

Debido a esto se tomo la consideración de que tanto WP como SR no fueron buenos estimadores del ranking en este torneo. Este fenómeno es atribuido (como se mencionó al principio) al particular formato de los resultados de los partidos, sumado al modelo de rankeo que usa ATP para determinar el rating de cada jugador y la complejidad de los fixtures al tener múltiples campeonatos. Un detalle no menor es la similitud en los rankings

[4] https://www.atptour.com/es/rankings/singles

entre WP y SR, los cuales incluso se podrían pensar que fueran de la saturación del Top, empieza a rankear "medianamente bien" a partir de cierto punto. Pero fuera de esto sigue siendo un ranking muy malo comparado con Colley.

Por último, en la Tabla 5 tenemos justamente a Colley frente a los resultados reales de la temporada, donde se ve lo cerca que están y lo bien que estimó al Top 10 en general. Colley demuestra ser muy bueno a la hora de estimar el ranking en esta clase de datasets, donde ya pudimos observar que otros métodos fallan abruptamente.

3. Experimentación

Los algoritmos de ranking deportivos tienen ciertas propiedades importantes a cumplir para modelar correctamente un torneo en base a sus resultados estadísticos [5]. Algunas de las más relevantes serán estudiadas en la siguiente sección mediante el planteamiento de hipótesis y la experimentación computacional. Otras propiedades más triviales pero igual de importantes como “El incentivo a la Victoria” no serán explorados mediante experimentos, todo esto debido a que el marco teórico de los métodos en cuestión permite corroborar con facilidad que no existe estrategia posible para incrementar el ranking final perdiendo un partido, sin embargo esta propiedad no es tan fácil de ver para sistemas más complejos que no son el motivo de estudio de la presente investigación (como el método de Markov).

3.1. Fuerza de los Oponentes

Un buen algoritmo de ranking busca modelar la realidad de tal forma que el ganador sea el equipo que tuvo el mejor desempeño, y por ende debería tomar en cuenta la fuerza de los contrincantes en cada partido . Esto también es conocido en algunos deportes como *Power Rankings* y es característico de algunos métodos de ranking como el ELO , donde se consideran que la fuerza del contrincante determina la dificultad del partido y debe influir en el ranking. Es decir, existen partidos más importantes que otros,y no debería generar el mismo aumento en ranking que el mejor equipo de la temporada le gane al peor que viceversa.

El presente experimento busca responder a la pregunta: **¿Influyen las fuerza de los oponentes en el resultado del ranking para los métodos de Colley, WP y SR?** y se plantea como hipótesis que los métodos planteados no lograrán representar esta condición por los datos estadísticos que usan para calcular los ratings.

Esto se realizará mediante datasets creados *ad hoc* para visualizar mejor los casos de interés y en particular encontrar un contra-ejemplo. Luego como continuación del experimento en datasets reales, se buscará analizar la simetría de los partidos, es decir, para seguir la misma linea de estudio se quisiera ver si invertir los resultados de un partido entre el mejor equipo y el peor equipo, si el rating del ganador como del perdedor decrecen y aumentan en la misma cantidad y la hipótesis planteada es que esto será así para cualquier caso. Estos resultados pueden ser observados en la sección 4

3.2. Relevancia del orden de eventos

Otra aspecto importante en un sistema de rankings es que el *fixture* de los partidos no afecte el resultado final, y por tanto que para cualquier secuencia de eventos el resultado sea el mismo. Es decir, el experimento en cuestión busca responder la pregunta **¿El orden de los partidos afecta el rating?** y se plantea como hipótesis que los algoritmos en cuestión no van a ser sensibles a estas variaciones. Esto se corroborará utilizando datasets reales y modificando el orden de dos formas: revirtiendo el orden de todos los partidos en una temporada o utilizando una función uniformemente aleatoria para reordenar los eventos al azar.

3.3. Independencia de los partidos

Por otra parte, en un sistema de rankings deportivos es importante determinar la independencia de los resultados de cada equipo con respecto a su rating o sino como este se puede ver afectado por resultados externos, es decir, se busca evaluar la interrogante **¿El resultado de un partido entre dos equipos puede afectar indirectamente el ranking de un tercero?** y se busca comprobar la hipótesis, la cual en este caso indica que Winning Percentage y Score Ratio van a ser completamente independientes ya que solo toman en cuenta las victorias o derrotas, pero Colley no. Este experimento consistirá en simulaciones minimales con datasets creados para tratar de explotar esta patología en cada método, y luego un caso de la vida real para ver como esto afecta a los ratings.

3.4. Tiempo de Ejecución

Finalmente, se tomará en cuenta para los datasets utilizados el rendimiendo computacional y el tiempo de ejecución de cada uno de los 3 métodos en cuestión a fin de determinar **¿Cual es el mas rápido para las**

instancias utilizadas?. Esto nos permitirá determinar si ciertamente Colley tardará mas que los demás debido a tener mayor complejidad algorítmica, o que todos tendrán tiempos similares. Es importante resaltar que no se evaluará a fondo la complejidad algorítmica ya que no es el enfoque del presente trabajo, sino sus aplicaciones prácticas y capacidad de responder el problema del ranking deportivo.

4. Resultados y discusión

4.1. Experimento 1

El presente experimento busca abordar el problema de *fixtures* desbalanceados para el método de Colley (aunque también se analizarán resultados para WP y SR). En particular, este problema es fácil de visualizar en la Tabla 6 y en caso de competencias round-robin donde todos los equipos juegan contra todos la misma cantidad de veces y existe un balance libre de *bias* que es universalmente aceptado como justo [6]. Pero en los casos donde esto no ocurre se debe tener en cuenta la *fuerza de schedule*, aún cuando esta no siempre es suficiente para representar la realidad, y en particular la verdadera "fuerza de los oponentes", es decir, asumir que hay partidos más importantes que otros y en particular victorias más difíciles y con más información que otras.

El problema en cuestión es que el algoritmo de Colley para asignar los ratings utiliza la información de ganados-perdidos de cada equipo para ajustar por el *schedule strenght*, pero esto no es suficiente ya que además no toma en cuenta el historial (como se verá en el experimento 2), ni el margen de victoria. En consecuencia, el método falla en precisión en representar el resultado deseado en la vida real. En el ejemplo de la tabla 6 planteamos equipos, donde A tuvo una temporada perfecta (invicto) y es el equipo más fuerte en el ranking en la posición 1; de segundo esta el equipo B que solo perdió contra el A, de igual forma tenemos que D solo ganó contra el C, y el equipo E fue el perdedor absoluto.

Cuadro 6: Temporada *AdHoc*

Equipo	Ganados-Perdidos	Rating de Colley	Posicion
A	4-0	0.786	1
B	3-1	0.643	2
C	2-2	0.5	3
D	1-3	0.357	4
E	0-4	0.241	5

Ahora, al modificar esta temporada perfecta ya agrega un upset, donde el E derrotó al A vemos que se cambian el scoreboard y hay un empate en el primer y tecer lugar, pero esto no representa la verdadera dificultad de los partidos que debió ganar cada equipo ya que el algoritmo no utiliza toda la información necesaria para representar totalmente el mundo real como se mencionó anteriormente, y se puede ver en la Tabla 7 que ahora el equipo A se encuentra empatado con el B, pero en realidad hay mucha mayor dificultad en ganarle al equipo A que al equipo E, por lo cual el ganador debería ser el B. De igual forma el D y el E quedan empatados pero E tuvo una mayor dificultad para ganarle al A que el D al C.

Cuadro 7: Temporada *AdHoc con Upset*

Equipo	Ganados-Perdidos	Rating de Colley	Posicion
A	3-1	0.643	1
B	3-1	0.643	1
C	2-2	0.5	2
D	1-3	0.357	3
E	1-3	0.357	3

Este ejemplo minimal con el upset demuestra que para crear métodos más precisos es importante tomar en cuenta mayor información sobre los partidos para calcular el rating, y se comprueba la hipótesis que el método no toma en cuenta la fuerza de los oponentes. En el caso de WP los resultados fueron los mismos que con

Colley ya que este presenta el mismo problema, y en el caso de Score Ratio esta variación dependió del puntaje del partido upset. Para diferencias muy altas se podía cambiar completamente el resultado del ranking y para las bajas era estático. Además, de este experimento se desprende la idea de colocar un módulo o cota a Score Ratio de forma que con un outlier en un partido con una diferencia muy marcada de goles no se pueda ganar directamente la competencia.

Por otra parte, se puede desprender de los cálculos que los rating r_i de cada equipo son interdependientes (como se verá en el experimento 3), y queremos analizar si lo que gana un equipo en rating en un partido es lo mismo que su oponente pierde. Luego, se puede ver en la tabla 8 que la hipótesis fue rechazada y esto no es cierto, no es simétrica la distribución de puntos entre el perdedor y el ganador ya que se modifica el sistema de ecuaciones a resolver donde intervienen muchas otras variables que son interdependientes.

Cuadro 8: Temporada *AdHoc con Upset*

Equipo	CMM	WP	SR	CMM con Upset	WP con Upset	SR con Upset
Cleveland	0.692428	0.712121	0.031051	0.678517	0.696970	0.029427
Philadelphia	0.150813	0.134328	-0.052028	0.164551	0.149254	-0.050437

Además, se evidencia que esta simetría tampoco vale para WP ya que los equipos pueden no jugar la misma cantidad de partidos, y aun menos para Score Ratio. Toda la información y los notebooks referentes a la ejecución de este experimento se encuentran disponibles en el Anexo.

4.2. Experimento 2

Durante este experimento se ejecutaron los 3 métodos algorítmicos mencionados en el informe para los 4 datasets con los que se trabajó. La hipótesis que planteamos es que el orden de los partidos no afectará el ranking obtenido en ninguno de los métodos utilizados.

La primera prueba del experimento consiste en revertir el orden original de los datasets, mientras que la segunda prueba del experimento consiste en ordenar los datasets de manera aleatoria. En ambos casos luego de realizar las modificaciones al dataset, se compara el ranking resultante con el ranking original, esperando ver si las posiciones asignadas efectivamente son las mismas, o si son modificadas al cambiar el orden.

Como resultado del experimento se obtiene que para los 3 metodos utilizados, no hay diferencias en los rankings obtenidos al observar los archivos de salida resultantes y se confirma la hipótesis planteada. Esto se debe a que los metodos no consideran el orden de lo partidos a la hora de hacer los calculos, a diferencia de otros sistemas de ranking donde esto no se cumple como el caso del ELO[7].

4.3. Experimento 3

Al igual que con el experimento 1 se tomará un ejemplo minimal de un torneo round-robin, pero en este caso se tomaran 3 variaciones del mismo, todos estos reflejados en la Figura 2. Las flechas indican que se ha jugado un partido entre los equipos representados como vértices del grafo, y la orientación de la flecha representa que el equipo que esta siendo apuntado por el otro perdió ante él. El análisis del experimento se centrará en el método de Colley, por lo que no tendremos en cuenta el puntaje de los partidos.

Tenemos las 3 modificaciones en distintos grafos simbolizando que se ha modificado un partido del original. Nuestra hipótesis será que Colley no afectará el ranking del equipo 1 ni del equipo 2, ya que estos no tienen nada que ver en las modificaciones realizadas. Además, el mismo efecto se verá en los otros métodos.

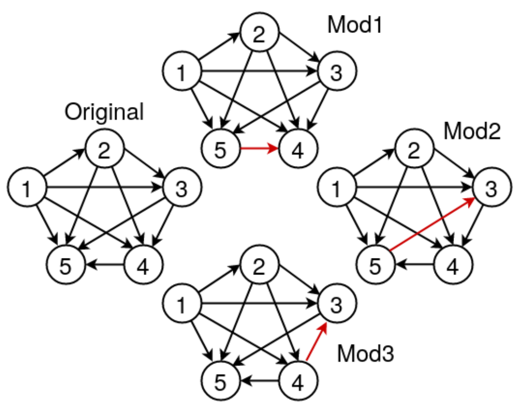


Figura 2: Ejemplo de torneo

Cuadro 9: Resultado del método de Colley en las 3 modificaciones del torneo

Equipo	Colley	Colley Mod1	Colley Mod2	Colley Mod3
1	0.785714	0.785714	0.785714	0.785714
2	0.642857	0.642857	0.642857	0.642857
3	0.500000	0.500000	0.357143	0.357143
4	0.357143	0.214286	0.357143	0.500000
5	0.214286	0.357143	0.357143	0.214286

Como se puede apreciar en el Cuadro 9, luego de realizar el experimento se obtiene que el ranking de los equipos 1 y 2 no fue afectado en lo más mínimo en ninguna de las tres modificaciones, tal como se esperaba. Este comportamiento esta justificado solo porque los partidos son interdependientes. Esto se puede ver también en la primera modificación donde el equipo 3 no es afectado por el cambio entre el equipo 4 y el 5. Igualmente, algo similar ocurre en las otras dos modificaciones; incluso, sucede algo interesante en la modificación 2 donde se equiparan los rankings de los 3 equipos dado que estos perdieron tres veces y ganaron una vez. En el caso de WP y SR se observan comportamientos similares en cuanto al rating de cada equipo en cada modificación.

Pero fuera de este ejemplo que fue armado con la intención de reflejar el comportamiento descripto, en la vida real los torneos son confeccionados de formas mucho más elaboradas. Esto será el objeto de estudio de la segunda parte de este experimento donde se trabajará con el dataset del 2019 de la NFL, una liga conocida por tener un schedule muy particular y un poco difícil de entender [5].

Expuesto esto, el experimento consistirá en mostrar la dependencia que tienen los datos en Colley. Se tomara a los dos equipos que les fue peor de la confederación AFC (Miami y Cincinnati) y se modificará el resultado de su partido a favor de Cincinnati. A pesar de las particularidad del schedule uno esperaría que un partido entre los dos peores equipos no debería afectar el ranking del primer puesto del torneo (San Francisco de la confederación NFC), pero la hipótesis planteada será todo lo contrario, se presume que tanto el podio como todos los rankings se verán afectados por este ligero cambio en dicho partido.

Como se ve en el Cuadro 10 sucedió lo que justamente se esperaba, no solo el primer puesto tuvo una modificación de su rating, si no que todos sufrieron alguna varianza en cuanto a su rating original. Este fenómeno es explicado por la complejidad del schedule de la NFL donde todos los equipos terminan vinculándose. La modificación del primer puesto en particular puede ser producto de un vínculo concretado por Colley en un partido entre Cincinnati y San Francisco. Las modificaciones de los demás ratings también se originan a partir de algún vinculo con los dos equipos perdedores.

Por último, cabe mencionar que en el caso de WP y SR los ratings de los equipos no relacionados con el partido modificado no sufrieron ninguna variación.

Equipo	Colley	Colley Mod
San Francisco	0.805164	0.809653
Kansas City	0.800313	0.800360
Baltimore	0.794117	0.798259
New Orleans	0.754284	0.755348
Green Bay	0.745049	0.744414
Seattle	0.692670	0.697190
Houston	0.644497	0.644579
New England	0.644044	0.639876
Tennessee	0.612560	0.612827
Minnesota	0.611262	0.610473
LA Rams	0.593875	0.599083
Buffalo	0.532957	0.528778
Chicago	0.502528	0.501410
Pittsburgh	0.496922	0.501004
Atlanta	0.493248	0.494443
Philadelphia	0.479949	0.475084
Denver	0.467487	0.467308
Tampa Bay	0.463870	0.465027
Indianapolis	0.459688	0.457458
Dallas	0.452425	0.447118
Las Vegas	0.449806	0.452142
Cleveland	0.409957	0.414291
Jacksonville	0.399840	0.402660
NY Jets	0.390860	0.386722
Carolina	0.385667	0.386847
Arizona	0.371310	0.376519
LA Chargers	0.359655	0.356940
Miami	0.289345	0.234701
Detroit	0.276879	0.275766
NY Giants	0.226776	0.221474
Cincinnati	0.198437	0.253024
Washington	0.194559	0.189222

Cuadro 10: Resultado de Colley vs. su modificación en la NFL 2019

4.4. Experimento 4

En el presente experimento, se realizará una ejecución simultanea de todos los métodos en los 4 datasets con los que se trabajó. Cada experimento será lanzado 10 veces por cada variación de estos parámetros para obtener una mejor muestra de los resultados debido a que, como es sabido, los tiempos de ejecución pueden estar siendo

[5] https://www.sporcle.com/blog/2017/04/how-the-nfl-schedule-is-made/

influenciados por procesos ajenos a los de la experimentación, pero siguen siendo un buen estimativo del tiempo tomado por cada método.

La hipótesis planteada será que el método de Colley tendrá un tiempo mucho mayor que los demás métodos debido a su mayor complejidad, ya que como se explicó en la sección 2.2 este tiene una Eliminación Gaussiana en su proceso.

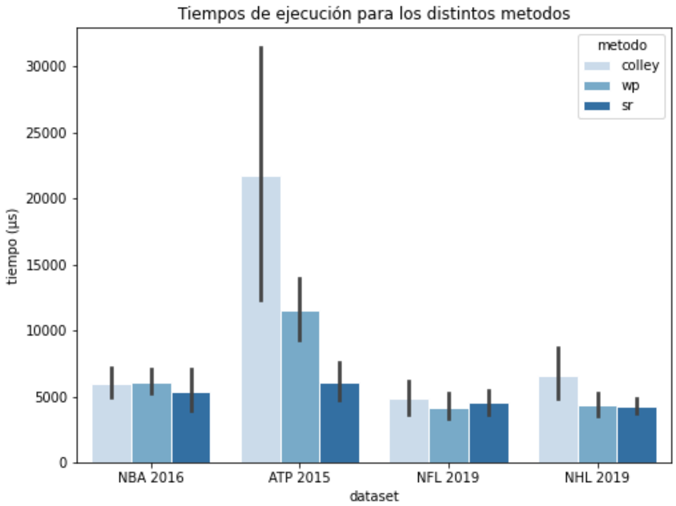


Figura 3: Comparación de tiempos de ejecución de todos los métodos

Por la Figura 3 se puede ver que la hipótesis no se cumplió tal como se esperaba, ya que se predecía que Colley iba a ser mayor en todos los casos, cosa que solo ocurrió con el dataset de ATP. Este incremento general en ATP se ve justificado por el tamaño del dataset ya que de entre todos, este es el mas grande, justamente por esto se ve que Colley se tardó mucho más que en los otros casos.

En lo que respecta a los otros datasets, parece ser insignificante la diferencia entre los métodos, e incluso resultan tener tiempos similares entre los mismos datasets, esto debe estar reflejando la similitud de los tamaños de ellos.

4.5. ¿El método CMM es justo?

El método de Colley durante muchos años fue utilizado para rankear competencias de deportes mundialmente famosos y trata de tomar en consideración los partidos ganados-perdidos por un equipo a la hora de definir el rating luego de cada juego, sin embargo, para lograr esto existe un *tradeoff* en la independencia de los rating y como demostró el experimento 3 el resultado de un tercero en otro partido puede afectar el rating final de un equipo. Además, se propone ante esto una estrategia para ganar el torneo minimizando la cantidad de victorias, y esta consiste en enfocar los esfuerzos del equipo en ganar los partidos en los cuales el contrincante tenga la mayor cantidad de ganados-perdidos (o el rating más alto).

4.6. Optimizaciones CMM

Aunque en la presente investigación se hizo énfasis en el análisis experimentativo de las características de los métodos destacados, es importante destacar que existen diferentes formas de implementar estos y en particular para el método de Colley se puede hacer uso de una propiedad particular para resolver el sistema de ecuaciones de forma más eficiente para futuros cómputos. Como la matriz generada en CMM es estrictamente diagonal dominante tiene una factorizacin LU, es decir, la matriz C de Colley se puede escribir como el producto de dos matrices $L, U \in \mathbb{R}^{TxT}$ donde L es triangular superior y R es triangular inferior. Y por ende el sistema $Cr = b$ se puede escribir como

$$LUr = b \tag{5}$$

y resolver en dos etapas:

$$\begin{aligned} Ly &= b \\ Ux &= y \end{aligned} \tag{6}$$

con complejidad $O(n^3)$ y la misma complejidad espacial que la Eliminación Gaussiana. En caso de matrices dispersas el uso de memoria puede ser un inconveniente, por lo cual para esos casos se debe utilizar otras estructuras de datos especializadas como diccionarios de llaves, listas de coordenadas o matrices comprimidas [8] entre muchas otras que se escapan del enfoque de desarrollo del presente trabajo práctico.

4.7. Otros Métodos de Ranking

El metodo de **Massey** (utilizado anteriormente en *BCS rankings*) hace uso del diferencial de puntos en un juego para determinar sus *ranking*[9] ya que considera al margen de victoria como una mejor métrica para representar el desempeño en un partido. Esto se logra construyendo una matriz $M \in \mathbb{R}^T$ donde cada fila representa un juego de forma que M_{ii} = # de partidos jugados por el equipo i, y M_{ii} = -# de partidos jugados entre i y j. Esta matriz debe ser ajustada para ser inversible y resolver el sistema de ecuaciones de la forma:

$$Mx = b \tag{7}$$

donde x es una matriz columna, y el vector b representa la suma de los diferencias de puntos para el $equipo_i$ esa temporada.

Este método originalmente fue seleccionado como método alternativo en vez de Score-Ratio y su implementación se puede encontrar en el código fuente en el anexo del TP. Sin embargo, por razones técnicas de error numérico no se obtuvieron resultados válidos para resolver el sistema de ecuaciones planteado, muy probablemente debido al error humano. Por lo cual, ante las limitaciones de tiempo para la realización de la investigación, se optó por cambiar el método. Queda este como propuesta para futuros trabajos de la cátedra.

5. Conclusiones

Los métodos de Colley, Winning Percentage y Score-Ratio resultaron adecuados para resolver el problema del ranking, sin embargo, cada uno presenta propiedades distintivas que los hacen aún mejores para distintas circunstancias sobre otras. Primero, el método de Colley aunque es más complicado de implementar y tiene mayor complejidad algorítmica que los otros dos, resultó tener un rendimiento equitativo al de sus contendientes, es decir, la variación de tiempo para los datasets utilizados fue mínima, siendo estos datasets de tamaños comunes en la práctica ya que se utiliza data de temporadas completas de deportes populares.

Se evaluaron las principales características que debía obtener un algoritmo de ranking como **Fuerza del Oponente**, **Incentivo a la Victoria** e **Independencia de la Secuencia de Eventos** donde se evidenció mediante la experimentación que la capacidad de los métodos para representar el mundo real correctamente se encuentra limitada por la información que utilizan para calcular los rankings. Efectivamente, todos los métodos son independientes del orden de los partidos, pero fallan en representar correctamente partidos que podrían tener mayor relevancia en el mundo deportivo. En particular, Colley trata de hacer un ajuste a la fuerza del schedule utilizando los partidos ganados-perdidos, pero en algunos casos no logra hacer suficiente. Ahora bien, este detalle se puede resolver al costo (o *tradeoff*) de utilizar un modelo mucho más complicado como los empleados en la actualidad donde se toma en cuenta el margen de victoria entre otros aspecto, es decir, el problema se puede tratar de solucionar agregando más información al modelo. Por otra parte, se evidenció la interdependencia de los partidos en Colley mediante grafos dirigidos y como el resultado de un tercero podía directamente afectar la posición de un equipo en el ranking.

En la mayoría de los casos de estudio en la presente investigación Colley junto con WP da un resultado similar al esperado (real), mientras que SR es más susceptible a outliers, por lo cual sería de interés colocar una cota en el margen de máximo que se puede obtener por diferencia de puntos. Con respecto a los *fixtures* desbalanceados donde no todos los equipos se enfrentan igual, Colley resulta definitivamente una mejor técnica a utilizar con el *tradeoff* de su complejidad. Por el otro lado, WP y SR son fáciles de entender e implementar, y funcionan bastante bien para torneos balanceados o round-robin. Además, una alternativa que se plantea para estos últimos dos es que a pesar de ser menos precisos en general, se pueden combinar para dar una ecuación de rating fácil de calcular y clara de entender para los espectadores, no obstante, esta es necesaria de parametrizar y acotar según el deporte, quedando así para futuras investigaciones.

Cabe destacar de igual manera que el error numérico también fue un factor interesante a analizar y se prestó énfasis en un $\delta > 0$ para evaluar la correctitud de los cálculos, y como se distribuía el error según el tamaño del test para concluir la estabilidad de los métodos. Finalmente, en el inmenso mundo de la aritmética finita quedan múltiples optimizaciones de álgebra lineal para estos métodos, en particular para Colley implementando optimizaciones de cálculo o diferentes estructuras de datos que le permitan ser más eficientes en matrices dispersas, las cuales serían una continuación adecuada de este estudio inicial.

6. Anexo (código fuente)

Los algoritmos fueron implementados en C++ y el manejo de los dataset así como el análisis de los datos fue realizado utilizando Python sobre un Jupyter Notebook. Se adjunta con el presente informe una carpeta `experimentos` con el código base, instrucciones para compilar, un makefile, los datasets junto con sus notebooks y los archivos de input y output de cada experimento.

Referencias

[1] Russell Albright Anjela Y. Govan, Carl D. Meyer. Generalizing google's pagerank to rank national football league teams. *SAS Global Forum 2008*, 2008.

[2] Wesley N. Colley. Colley's bias free matrix rankings, 2006. `http://colleyrankings.com`.

[3] Jeff Sackmann. Atp tennis rankings, results, and stats. `https://github.com/JeffSackmann/tennis_atp`.

[4] Ken Massey. Massey ratings. `https://www.masseyratings.com/data`.

[5] Baback Vaziri, Shaunak Dabadghao, Yuehwern Yih, and Thomas L. Morin. Properties of sports ranking methods. 2018.

[6] Roy Bethel. A solution to the unequal strength of schedule problem, 2005.

[7] I. Charon and O. Hudry. Meyer. An updated survey on the linear ordering problem for weighted or unweighted tournaments. *Annals of Operations Research*, 2010.

[8] Fethulah Smailbegovic and Georgi N. Gaydadjiev. Sparse matrix storage format, 2005.

[9] Amy N. Langville Carl. D. Meyer. Who's #1. *Chapter 2, Princeton University Press*, 2012.