DEPARTAMENTO DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

TP1: Ranking Deportivo

Métodos Numéricos

Integrante	LU	Correo electrónico
Camjalli, Roni Ezequiel	12/18	rcamjalli@gmail.com
Rodriguez, Miguel	57/19	mmiguerodriguez@gmail.com
Itzcovitz, Ryan	169/19	ryanitzcovitz@gmail.com

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

https://exactas.uba.ar

Índice

1. Introducción

1.1. Resumen

En este informe presentamos el análisis del problema de los rankings en competiciones deportivas. Empezamos con una breve introducción al problema, cuales son los casos de interés para analizar y luego la presentación de los 3 métodos que utilizaremos para resolver (*CMM*, *WP* y *BP*). Más tarde, en el desarrollo, observamos las distintas características de estos métodos, así como las fórmulas utilizadas para la implementación de los mismos; un caso interesante en este punto es la utilización de los métodos de eliminación gaussiana sobre matriz para la resolución de ecuaciones en *CMM*. Seguido de esto contamos con una parte de experimentación. Al principio realizamos pruebas de estos métodos con datasets del mundo real e intentamos analizar qué método es mejor para cada dataset. Luego analizamos el modelo *CMM* más en profundidad, con un caso muy peculiar y observamos como en este método el ranking de un jugador puede ser modificado por un resultado de otros equipos. Por otro lado también nos tomamos el tiempo de analizar si es posible crear una estrategia para minimizar la cantidad de victorias subiendo el ranking de nuestro equipo, y observamos las distintas formas de hacer esto. Al final, vemos la estabilidad de los cálculos numéricos ante las operaciones computacionales. Para cerrar hacemos una conclusión de estas 3 soluciones resumiendo lo visto en las experimentaciones.

1.2. Introducción teórica

En este trabajo vamos a resolver el problema de los rankings en competiciones deportivas. En la actualidad encontramos muchos casos donde se presenta este problema, en el cual queremos obtener un ranking que determine las posiciones de distintos competidores a partir de sus resultados en los enfrentamientos de un tiempo determinado, por ejemplo podemos verlo en competiciones mas simples como lo es la AFA (*Asociación del Fútbol Argentino*), donde todos sus equipos se enfrentan entre si 2 veces; o también podemos ver el problema a resolver en competiciones mas complejas como la NBA (*National Basketball Association*), donde no todos los equipos se enfrentan entre si la misma cantidad de veces ya que la conferencia en la que está cada equipo define la cantidad de partidos que juega contra otro en la temporada regular.

Entonces para estos distintos escenarios de competiciones deportivas presentaremos varios algoritmos que cada uno de forma general, para todas las competiciones, deberán determinar las posiciones en forma de rankings de sus competidores, tomando como entrada a los resultados de los enfrentamientos entre ellos.

Para eso vamos a plantear un torneo como un listado de encuentros entre 2 equipos. Cada equipo va a tener una identificación numérica única y cada encuentro siempre tiene un solo ganador basado en quien obtenga mas puntos durante el encuentro. Para notar en nuestro torneo no sera necesario que todos los equipos tengan la misma cantidad de enfrentamientos, de esta forma vamos a poder resolver también los torneos mas complejos.

Notamos t_i al identificador del i-ésimo equipo, w_i la cantidad de victorias i-ésimo equipo y l_i la cantidad de derrotas del i-ésimo equipo. Luego tenemos a $n_i = w_i + l_i$ la cantidad de encuentros disputados del i-ésimo equipo. También vamos a querer ver los resultados de un competidor relativo a los enfrentamientos de otro competidor, de esta forma vemos a n_{ij} como la cantidad de enfrentamientos entre el equipo i-ésimo y el j-ésimo, y también definimos $n_{ij} = w_{ij} + l_{ij}$ donde w_{ij} son las victorias del equipo i-ésimo sobre el equipo j-ésimo y l_{ij} las derrotas del equipo i-ésimo frente al equipo j-ésimo. Se puede ver que $n_{ij} = n_{ji}$ y además que $w_{ij} = l_{ji}$. Como resultados vamos a obtener r_i el rating del i-ésimo equipo, de esta forma vamos a determinar la posición de los competidores en orden de mayor a menor rating.

Vamos a implementar 3 soluciones, CMM (*Colley's Matrix Method*), WP (*Win Percentaje*) y BP (*Bonus Point*) siendo esta ultima desarrollada por nosotros mismos. Desarrollaremos y analizaremos estos métodos en las siguientes secciones.

2. Desarrollo

2.1. Colley's Matrix Method

Uno de los métodos es el conocido como CMM (*Colley's Matrix Method*), un método insesgado, originalmente utilizado para darle calificaciones a equipos de fútbol americano[1]. Este método se basa en resolver la ecuación $\vec{C}r = \vec{b}$ donde r es una incógnita que representa el rating de cada equipo una vez resuelto el sistema. El resto de los parámetros son de la forma:

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

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

Esta matriz C resulta ser simétrica ya que las coordenadas C_{ij} y C_{ji} son iguales, esto ocurre porque $n_{ij} = n_{ji}$ para todo $i \neq j$, ya que la cantidad de veces que jugo un equipo contra otro es la misma. Además es definida positiva, esto quiere decir que para todo vector v ocurre que

$$\vec{v}^T(C\vec{v}) > 0 \tag{3}$$

Para demostrar esto, separamos a C en $2I + \sum_k G^k$ siendo G^k la matriz obtenida para el partido k entre los equipos i y j de tal forma que $G^k_{ii} = G^k_{jj} = 1$ y $G^k_{ij} = G^k_{ji} = -1$ y el resto de los índices en 0. Como la multiplicación de matrices es distributiva, podemos separar en dos partes la demostración. La matriz $2I$ es definida positiva, por lo que resta observar solamente los G^k's. Tomando $\vec{r}$ los ratings para el equipo i y j, la multiplicación $G\vec{r}$ tiene ceros en todos sus filas menos en la i-ésima y j-ésima, donde valen $r_i - r_j$ y $r_j - r_i$. Luego, el producto $\vec{r}^T(C\vec{r})$ resulta ser:

$$(r_i - rj)^2 \geq 0 \tag{4}$$

Luego $(r_i - rj)^2 \geq 0 \wedge \vec{r}^T(2I\vec{r}) > 0 \implies$ la matriz C es definida positiva.

Por ser una matriz simétrica definida positiva, podemos decir que es inversible, y además, que todas sus submatrices $C_{(1)}, ..., C_{(n)}$ también lo son. Entonces, podemos afirmar que se puede aplicar el método de eliminación gaussiana sin necesidad de intercambio de filas o columnas.

Por lo tanto luego de ver que podemos resolver la ecuación $\vec{C}r = \vec{b}$ con eliminación gaussiana sin intercambio de filas, presentamos los siguientes algoritmos que nos van a permitir implementar el método *CMM*.

Algorithm 1 Algoritmo de eliminación gaussiana en la matriz C de CMM

1: $C \in \mathbb{R}^{nxn}$ la matriz que se forma en la formula (1)
2: $b \in \mathbb{R}^{n}$ el vector independiente que se construye con la ecuación (2)
3: **function** $EliminacionGaussiana(C, b)$
4:     **para** i **desde** 1 **hasta** n **hacer**
5:         **para** j **desde** 1 **hasta** n **hacer**
6:             **if** $C_{ii} \neq 0$ **then**
7:                 $a \leftarrow C_{ji}/C_{ii}$
8:                 **para** k **desde** 1 **hasta** n **hacer**
9:                     $C_{jk} \leftarrow C_{jk} - C_{ik} * a$
                end
10:                $b_j \leftarrow b_j - b_i * a$
            end
        end
    end
11:    **return** C, b

Se puede ver en la linea 6 del algoritmo que en el caso de que $C_{ii} = 0$ el algoritmo no hace nada. En ese caso uno debería hacer la permutación de filas, pero como es un algoritmo para *CMM*, en la matriz C no se anula su diagonal por lo que vimos en (3).

Una vez que tenemos la matriz ya triangulada podemos realizar el siguiente algoritmo para resolver el sistema y obtener el ranking.

Algorithm 2 Algoritmo para resolver sistema lineal con la matriz de *CMM*

$C \in \mathbb{R}^{nxn}$ la matriz que se forma en la formula (1)
$b \in \mathbb{R}^{n}$ el vector independiente que se construye con la ecuación (2)
function $ResolverSistema(C, b)$
    $C, b \leftarrow EliminacionGaussiana(C, b)$
    $r \leftarrow b$
    para i **desde** $n-1$ **hasta** 1 **hacer**
        para j **desde** $n-1$ **hasta** $i+1$ **hacer**
            $r_i \leftarrow r_i - r_j * C_{ij}$
        end
        if $C_{ii} \neq 0$ **then**
            $r_i \leftarrow r_i/C_{ij}$
        end
    end
    return r

Entonces una vez que armamos la matriz C y b en base a los resultados de los competidores, luego ejecutamos el algoritmo 2 y obtenemos nuestro ranking CMM.

2.2. Win Percentage

El método WP (*Win Percentage*) consiste en calcular el rating de un equipo basado en la cantidad de partidos ganados sobre el total de partidos jugados. El resultado de este método indica una noción de 'probabilidad', ya que ocurre que $r_i \in [0, 1]$. Podemos pensar al rating obtenido de cada equipo o jugador como la probabilidad de ganar un partido.

$$r_i = \frac{w_i}{n_i} \tag{5}$$

El método implementado tiene como consecuencia que para competencias con pocos partidos ocurre que los ratings para equipos que no ganaron partidos va a ser 0 y para los que ganan todos los partidos, sera 1. Para salvar estos casos, existe un método introducido por Pierre-Simon Laplace que consiste en agregar 1 al numerador y 2 al denominador de tal forma que:

$$r_i = \frac{1 + w_i}{2 + n_i} \tag{6}$$

La suma de estos números además de ser un mejor estimador, ya que decir que un equipo es infinitamente mejor que otro, o que un equipo no tiene probabilidades de ganar tiene poco sentido, tiene la propiedad de que para equipos sin partidos disputados, le asigna un rating de 1/2 (como ocurre con el método CMM).

2.3. Bonus Point

Por último desarrollamos el método al cual llamamos BP (*Bonus Point*) se basa en dar un punto extra al equipo que gana por mas de una cierta cantidad x de puntos o si pierde por menos de una cierta cantidad de puntos x, siendo x el mismo valor en ambos casos. Ganar suma 4 puntos, perder 0 y el equipo que sume un punto bonus obtendrá 1 punto más. Este sistema se diferencia del WP ya que va a enfatizar más en las victorias por más puntos en los victoriosos y en los que pierdan por pocos puntos, llegando a la conclusión de que un equipo pudo haber ganado la misma cantidad que el otro, pero por la cantidad de puntos bonus obtenidos tener mayor rating. La formula para calcular el rating del i-ésimo equipo es la siguiente:

$$r_i = \frac{w_i * 4 + b_i}{n_i * 5} \tag{7}$$

Notamos b_i como la cantidad de puntos bonus que obtuvo el i-ésimo equipo a lo largo del todo el torneo. La formula se puede leer como cantidad de puntos sumados, que son 4 por victoria, sumado a la cantidad de puntos bonus que logre en el torneo y eso dividido por la cantidad de partidos jugados multiplicado por 5, que representa la máxima cantidad de puntos que el jugador pudo haber obtenido en el caso que gane todos los partidos con punto bonus. La formula se puede entender como el porcentaje de puntos obtenidos.

3. Propuestas de experimentación

El problema a resolver fue poder rankear equipos en un torneo con diferentes sistemas de rating. Los utilizados fueron los descriptos anteriormente (*CMM*, *WP* y *BP*). Durante la experimentacion se usaron conjuntos de datos donde los equipos no siempre tenían todos la misma cantidad de partidos y los resultados posibles eran victoria, derrota o empate (según el conjunto de datos).

3.1. Pruebas con datasets reales

En este experimento nos va a interesar poder ejecutar los algoritmos con conjuntos de datos de competencias reales, para analizar los métodos que desarrollamos. Los datasets que vamos a utilizar son:

1. ATP 2015: 3000 partidos de 430 jugadores.
2. NBA 2016: 1000 partidos de 30 equipos.
3. Premier League 2018/19: 380 partidos de 20 equipos.[3]

3.1.1. ATP

Con los datos de la ATP, nos resulta interesante la comparación entre el método *CMM* y el de *WP*. Lo esperado para los ratings *CMM* es que jugadores con mayor cantidad de partidos jugados y mejor porcentaje de partidos ganados serán los mejores rankeados. Esto debería ocurrir ya que, por como funciona el método, el peso de jugar mayor cantidad de partidos es grande a la hora de resolver el sistema de la matriz. Es esperable que jugadores con pocos partidos jugados y alto porcentaje de ganados tengan un rating *CMM* alto, pero no necesariamente van a ser los mejores como ocurriría con el método *WP*. A la hora de ver los resultados, vamos a observar y comparar estos dos métodos en esos casos. Además, en el tenis, el único resultado posible es victoria o derrota, y para el conjunto de datos de entrada solo se tiene que jugador gano, sin ningún parámetro que nos indique el resultado.

3.1.2. NBA

El segundo dataset en el que nos vamos a enfocar para analizar el algoritmo *BP* es sobre la temporada 2016 de la NBA. Lo que esperamos ver es como un equipo que gana sus partidos por mayor diferencia o pierde por poco pueda obtener un mejor ranking y de esta manera agregar un incentivo a todos los equipos premiando a aquellos que logran el objetivo que les dará ese punto bonus (ganen o pierdan). Alguna de las cosas que esperamos que ocurran es que algunos equipos estén en el mismo ranking para los 3 algoritmos, esto indicaría que son equipos regulares cuando ganan o pierden. Además, otro comportamiento esperado es que los ranking *BP* y *WP* difieran, que nos da la intuición de que un equipo fue muy superior a sus rivales en sus victorias ($BP > WP$) o que fue inferior en las derrotas ($BP < WP$). Como los resultados en la NBA son victoria o derrota con la cantidad de puntos por equipo los 3 algoritmos pueden ser bien reflejados en un mismo gráfico.

3.1.3. Premier League

Además de los datasets vistos anteriormente, vamos a observar datos de la Premier League en la temporada 2018/19, que tienen un factor extra, que son los empates. Por como esta implementada la carga de datos, para nosotros un empate de local es considerado una derrota. Para este experimento, vamos a observar como afectan los empates en el rating de los equipos en comparación con su verdadero ranking en la competición al final de la temporada.

3.2. Análisis CMM

Nos hacemos la siguiente pregunta, ¿El método *CMM* es justo? Es decir, ¿es posible que el resultado de un partido entre dos equipos afecte indirectamente el ranking de un tercero? Para responder esto vamos a analizar el caso de la *Liga los Amigos*. La *Liga los Amigos* es un grupo de 10 amigos que se junta a jugar al *Winning Eleven 2006* todos los sábados. Cada jugador se enfrenta a otro a medida que pasa la tarde, no hay un fixture determinado sino que los jugadores se van desafiando mutuamente. No siempre todos los jugadores pueden jugar, así que no todos los competidores juegan la misma cantidad de partidos. Por lo tanto, teniendo en cuenta de esta diferencia, los amigos quieren saber cual es la mejor forma de calcular un ranking para saber las posiciones en su liga. Además Jorge, uno de los amigos, dice que Juan siempre juega con los jugadores menos experimentados y por eso siempre sale primero, y el quiere que esto quede reflejado en su tabla así Juan no sale primero por jugar siempre contra los peores. En un principio suponemos que *CMM* va a ser el mejor método para calcular el ranking de los amigos, ya que va a poder satisfacer el pedido de Jorge. Vamos a realizar dos análisis con esta liga donde en el primero veremos los resultados hasta la primera fecha y en el segundo veremos los resultados hasta la segunda fecha para ver como varia el ranking. También vamos a presentar gráficos comparativos de *CMM* vs *BP* y *WP* para saber que estamos tomando la decisión correcta. Para la experimentación utilizaremos 2 datasets, el primero contiene los resultados de la primera fecha, en donde Juan va a ganar por gran diferencia de goles a 4 competidores que no tendrán victorias en esa fecha y Jorge va a ganar 4 partidos pero contra jugadores que si ganaron partidos y con menor diferencia de goles. De esta forma Juan y Jorge tendrán la misma cantidad de victorias pero contra distintos competidores. El segundo dataset le va a sumar al primero la segunda fecha donde Jorge y Juan no se pudieron juntar a jugar con sus amigos. En la segunda fecha los jugadores que perdieron contra Juan la mayoría le va a ganar a los que le gano Jorge, de esta manera dejan de ser los peores y vamos a analizar si cambian los ratings de Jorge y Juan.

3.3. Estrategia de maximización de resultados

Otra pregunta que nos hacemos y queremos responder es si dados los resultados de todos los partidos considerados en la competencia, y un equipo particular, determinar una estrategia que permita obtener la mayor posición posible, buscando minimizar el número de partidos ganados. Para dar una solución al problema planteamos dos posibles situaciones: Cambiar los resultados de un torneo conociendo las posiciones de cada equipo (alterar el pasado conocido) y agregar mas partidos que alteren los ratings finales (alterar el futuro con un pasado conocido).

3.3.1. Alterar el pasado conocido

En esta estrategia planteamos hacer que un equipo aumente su rating *CMM* sin modificar el *WP* modificando sus resultados con otros equipos, perder con un equipo de su mismo nivel, pero ganarle a uno de los mejores para de esta manera aumentar su rating. Este cambio no se va a observar tan claramente modificando un solo partido, pero luego de varios cambios se podrá ver como se logra el objetivo planteado.

3.3.2. Alterar el futuro con pasado conocido

Para poder alterar el futuro conociendo los ratings del pasado sin la necesidad de hacer que el equipo gane mas partidos y de esa forma maximizar los resultados, planificamos una estrategia en la cual el equipo al que queremos pasar en el ranking, hacerlo perder un partido contra uno de los peores de la tabla. De esta forma hacemos que su rating disminuya considerablemente.

3.4. Estabilidad de los cálculos

Por ultimo, queremos estudiar también que ocurre con la estabilidad de los cálculos a la hora de resolver los sistemas de ecuaciones, ya que estaremos haciendo cuentas con números en punto flotante y aritmética finita, lo cual conlleva ciertos errores de cálculo.

4. Resultados y discusión

4.1. Pruebas con datasets reales

4.1.1. ATP

Como suponíamos en la sección anterior, el método de rating por *CMM* es definitivamente el que supera a los demás implementados por amplia diferencia. Esto se debe a que, además de tener en cuenta los partidos ganados y perdidos, este método tiene en cuenta algo mucho mas importante, que es la cantidad de partidos jugados. En la figura (1) podemos observar el rating que obtienen los jugadores según la cantidad de partidos ganados y perdidos.

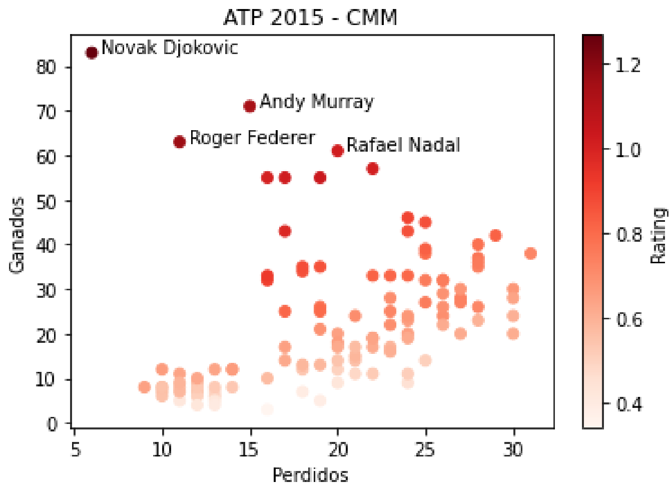

Figura 1: Rating *CMM* en función de los partidos ganados y perdidos

Es esperable que los jugadores con mayor cantidad de partidos ganados y menor cantidad de perdidos sean los que mejor ranking *CMM* obtienen. Podemos ver algo interesante en estos resultados, que es que hay valores de rating mayores a 1, esto se debe a que, a pesar de que el método tenga también una noción de probabilidad, no necesariamente la solución al sistema de ecuaciones tenga que tener estrictamente valores entre 0 y 1[2].

El dato de los partidos jugados resulta en que jugadores con muy buen porcentaje de partidos ganados pero pocos partidos jugados, no obtengan un rating *CMM* que lo diferencie ampliamente de otros jugadores (en comparación con *WP*). En la figura (2) podemos observar cómo para jugadores con pocos partidos jugados y alto porcentaje de ganados, se les asigna un alto rating *WP*, pero al compararlos con jugadores con mayor cantidad de partidos jugados y menor porcentaje de partidos ganados, tienen peor rating *CMM*.

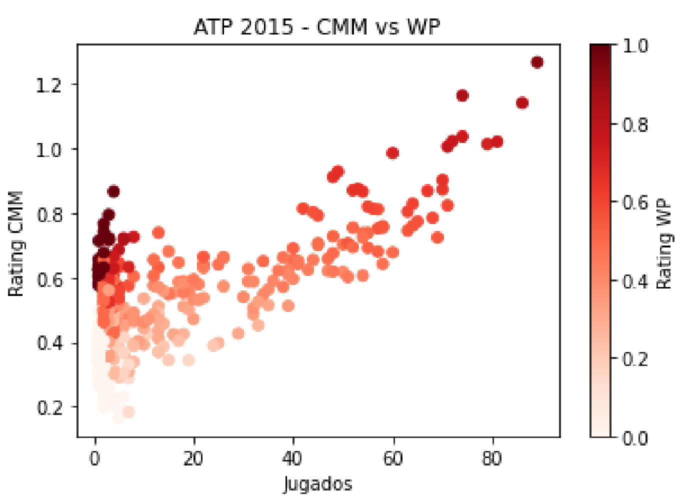

Figura 2: Partidos jugados, comparación *CMM* y *WP*

Los resultados obtenidos fueron los esperados, aunque es sorpresivo el alto rating *CMM* que obtienen jugadores con buen porcentaje de ganados pero pocos partidos jugados. Esto ocurre no solo por como funciona el método, si no por la dificultad que tiene darle un ranking a un jugador que jugo muy pocos partidos. Para eliminar el 'ruido' que nos generan estos datos, utilizamos el mismo dataset pero quitamos a los jugadores que tenían menos de 15 partidos jugados (teniendo en cuenta que una temporada regular de un jugador de tenis supera este numero).

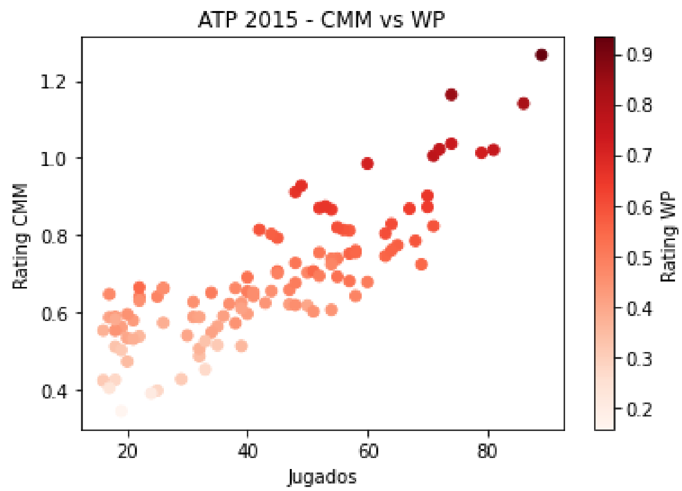

Figura 3: Partidos jugados > 15, comparación *CMM* y *WP*

Con esta ultima figura, podemos ver que ajustando un limite con el cual vamos a comparar los ratings, podemos notar que el rating *CMM* y *WP* tienen cierta similitud.

4.1.2. NBA

Dada la explicación en la seccion 2.3 del algoritmo de ranking *BP* en el que a diferencia de *WP* se premia a los equipos que ganen por una cierta diferencia o pierdan por menos de esa cantidad, creemos que este método puede crear una mayor competitividad en los equipos para ir a buscar ese punto bonus que puede marcar la diferencia hacia el final del torneo. Al calcular el rating de todos los equipos podemos saber que los mejores equipos son los que ganaron mas partidos por una amplia diferencia, y los peores equipos son aquellos que no perdieron muchos partidos y tampoco lograron llegar al punto bonus en muchos partidos.
Analizando el dataset de la figura (4), podemos observar que en algunos equipos la diferencia entre el *WP* y *BP* hace que los equipos suban o bajen varios puestos, pero mas allá de eso si miramos los mejores y los peores equipos su *WP* esta directamente relacionado con su *BP* y demuestran su amplia diferencia con el resto, ya sea para bien o para mal.
Para finalizar podemos ver un equipo en el que su *WP* es diferente a su *BP* que es Chicago, aquí podemos concluir con la idea de que mas allá de haber ganado mas partidos que otros equipos su diferencia de puntos cuando perdió y ganó no hizo reflejar su superioridad para alcanzar ese mismo ranking en la tabla *BP*.

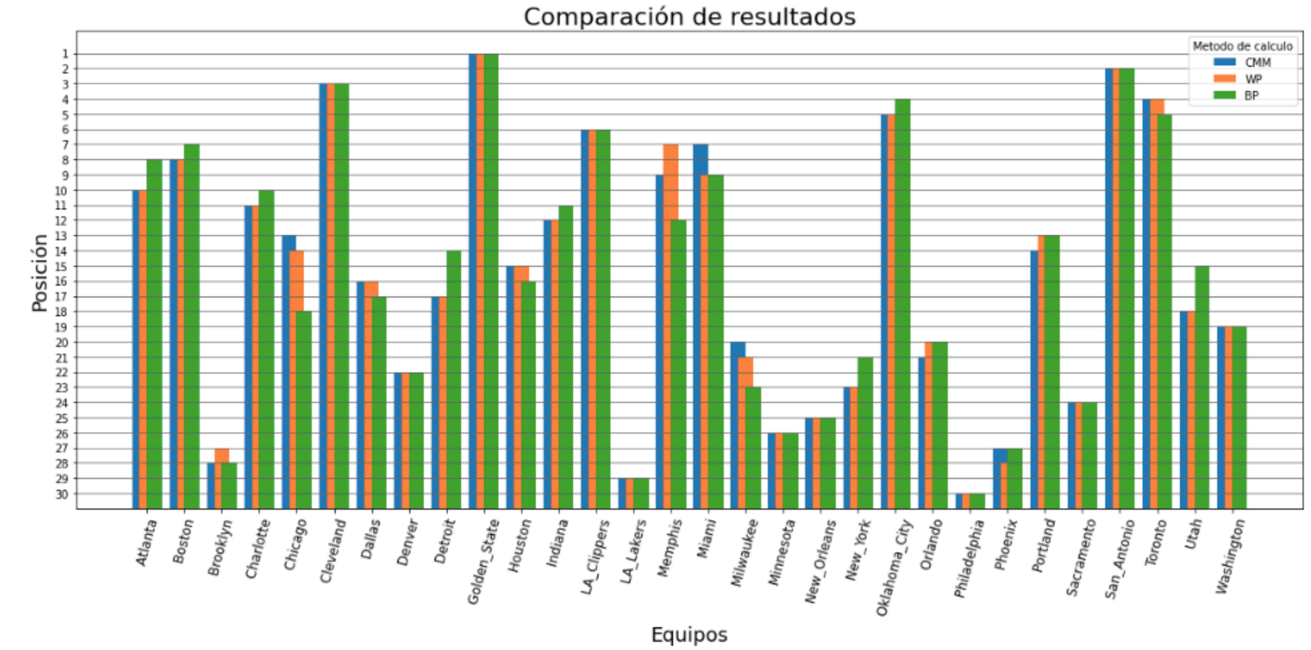

Figura 4: Dataset graficado de la NBA temporada 2016 con los 3 algoritmos desarrollados

4.1.3. Premier League

Para los datos de la Premier League, como se explico anteriormente, esperamos que los equipos con mayor cantidad de empates tengan rating distinto en el método *CMM* al compararlo con los resultados finales de la competición. Los resultados obtenidos son muy interesantes.

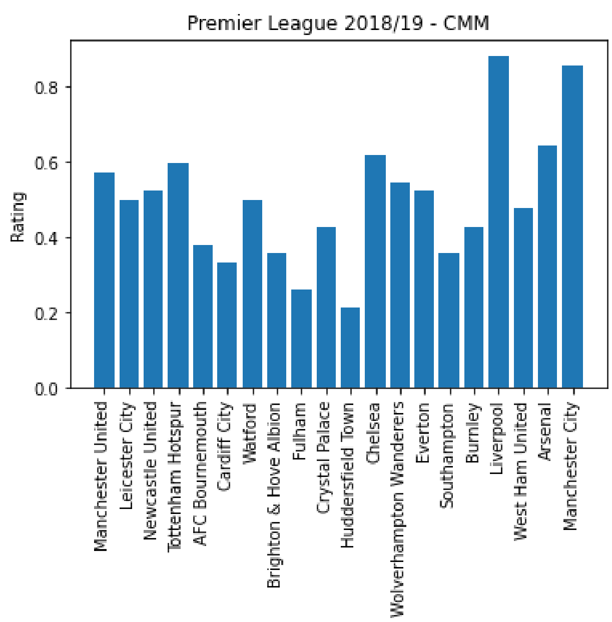

Figura 5: Premier League, rating CMM

Como vemos en la figura (5) el equipo con rating mas alto en este caso es el *Liverpool*, seguido por el *Manchester City*. En este caso, si tuviéramos que definir al campeón del torneo en base al rating *CMM*, seria distinto al verdadero campeón en esta temporada que fue el *Manchester City*.

Si nos adentramos en los resultados y la verdadera cantidad de partidos ganados, empatados y perdidos del torneo, vemos que el *Liverpool* tiene un historial 30-7-1 (V-E-D), mientras que el *Manchester City* 32-2-4. Podemos observar que el peso de una victoria para el sistema de puntaje del fútbol es mayor que el de un empate, ya que con 2 victorias y 3 derrotas mas que el *Liverpool*, el *City* logro salir campeón, pero para nuestro sistema de rating, no.

4.2. Análisis CMM

Realizamos la primera prueba calculando los ratings para los resultados de la primera fecha mencionados en el punto 3.2. Podemos ver en el gráfico (6) los resultados para los 3 métodos de la experimentación, donde se ve como Juan y Jorge son los punteros de la liga. Micho, Pepe, Tito y Pancho son los jugadores menos experimentados que se enfrentaron contra Juan y perdieron, se ve que ellos quedaron al fondo de la tabla. Martin, Andres, Julian y Manuel se enfrentaron a Jorge que les gano, pero a su vez ellos compitieron entre si, obteniendo mas victorias. En la figura (7) podemos ver con detalle los resultados de los ratings de Jorge y Juan luego de la primera fecha; y observamos que en *CMM*, Jorge esta por encima de Juan, cumpliendo con nuestra hipótesis que ante la misma cantidad de partidos ganados pero contra jugadores que tienen mas victorias el rating será mayor. De esta forma podemos decir que efectivamente se cumple lo que pidió Jorge y el ranking *CMM* es mas justo en este caso. Si miramos a estos dos competidores comparando los otros dos rankings, podemos ver que en *WP* empataron, así que no hizo diferencia sobre los contrincantes de estos jugadores y en *BP* Juan obtuvo mejor rating ya que venció por goleada a sus competidores.

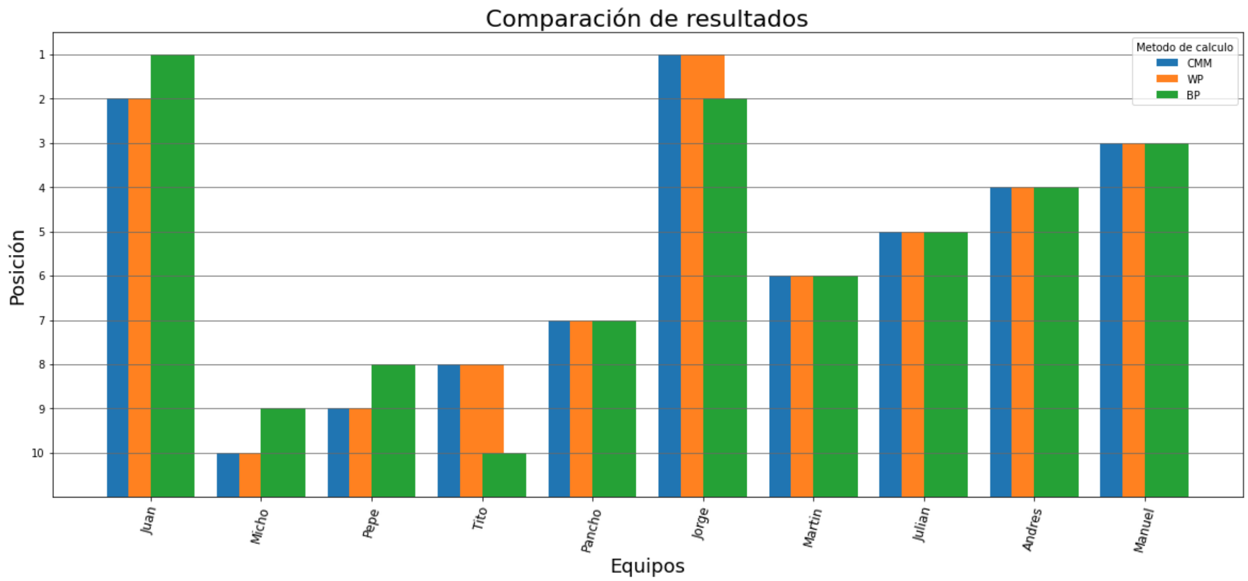

Figura 6: Comparación de posiciones para todos los jugadores, en los 3 métodos al terminar la primera fecha

	id	Equipo	CMM	WP	BP
0	1	Juan	0.76488095238095244	1	0.84999999999999998
1	6	Jorge	0.80654761904761918	1	0.80000000000000004

Figura 7: Resultados de los ratings para los 3 métodos, de Jorge y Juan al terminar la primera fecha

Luego de la segunda fecha volvimos a realizar la misma prueba anterior viendo como queda el ranking ahora, teniendo en cuenta que ni Jorge ni Juan participaron de esta fecha. En la figura (8) podemos observar como quedaron las posiciones para cada uno de los rankings, en base a los resultados de la fecha 1 y de la fecha 2. Podemos observar que hubo un cambio en el puntero de la liga en el ranking *CMM*. Además vemos que los jugadores que antes eran los peores de la liga ahora subieron de posiciones ya que ganaron partidos y esto fue lo que afecto los ratings de Jorge y Juan, es decir que sin haber competido, los resultados de otros jugadores afectaron sus rankings. Para ver mejor esto podemos observar la figura (9a) los nuevos rankings y en la figura (9b) que podemos ver la diferencia de rankings entre la primera y la segunda fecha para Jorge y Juan. También observamos que en los métodos *WP* y *BP* no hubo una diferencia ya que ambos métodos solo pueden ser modificados cuando el jugador obtiene un nuevo resultado. En conclusión vemos que el método *CMM* depende de los resultados de los otros competidores y como ellos afectan a su ranking, para eso comparamos la primera fecha contra la segunda en el gráfico (10), y vemos también que al subir el rating de Micho, Pancho y Tito, jugadores que le habían ganado Juan, el también sube sin disputar encuentros.

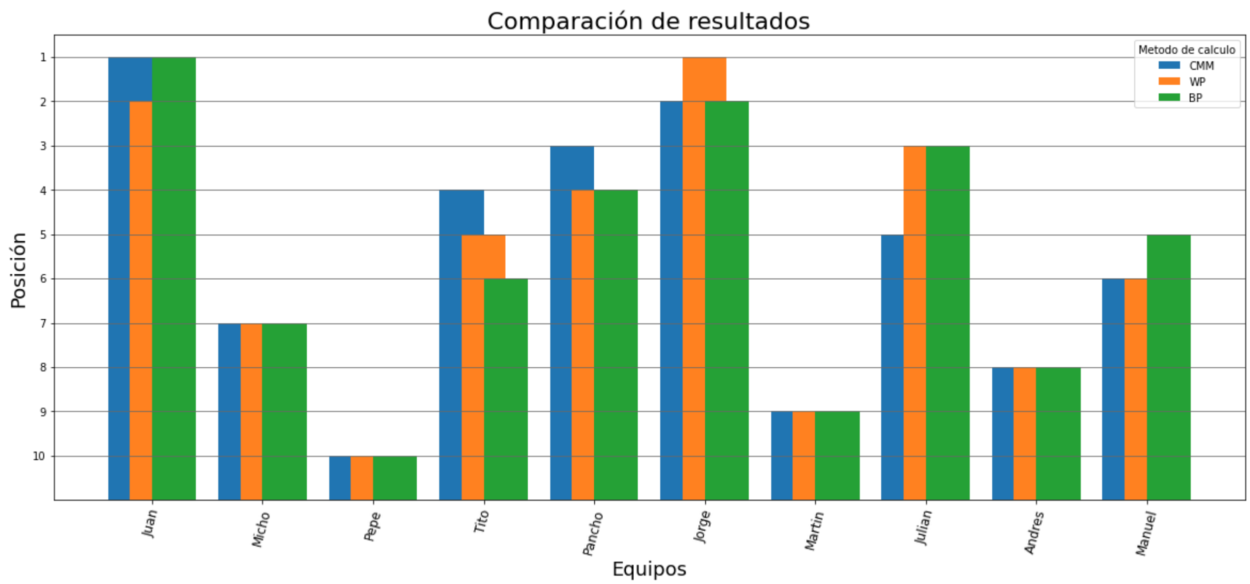

Figura 8: Comparación de posiciones para todos los jugadores, en los 3 métodos al terminar la segunda fecha

	id	Equipo	CMM	WP	BP
0	1	Juan	0.80616827193475815	1	0.84999999999999998
1	6	Jorge	0.76526029949381325	1	0.80000000000000004

(a) Resultados de los ratings para los 3 métodos, de Jorge y Juan al terminar la segunda fecha

	id	Equipo	diff CMM	diff WP	diff BP
0	1	Juan	0.041287	0.0	0.0
1	6	Jorge	-0.041287	0.0	0.0

(b) Diferencia entre los ratings de la segunda fecha y la primera fecha para los 3 métodos, de Jorge y Juan

Figura 9: Figuras reflejando los ratings de Jorge y Juan luego de la segunda fecha.

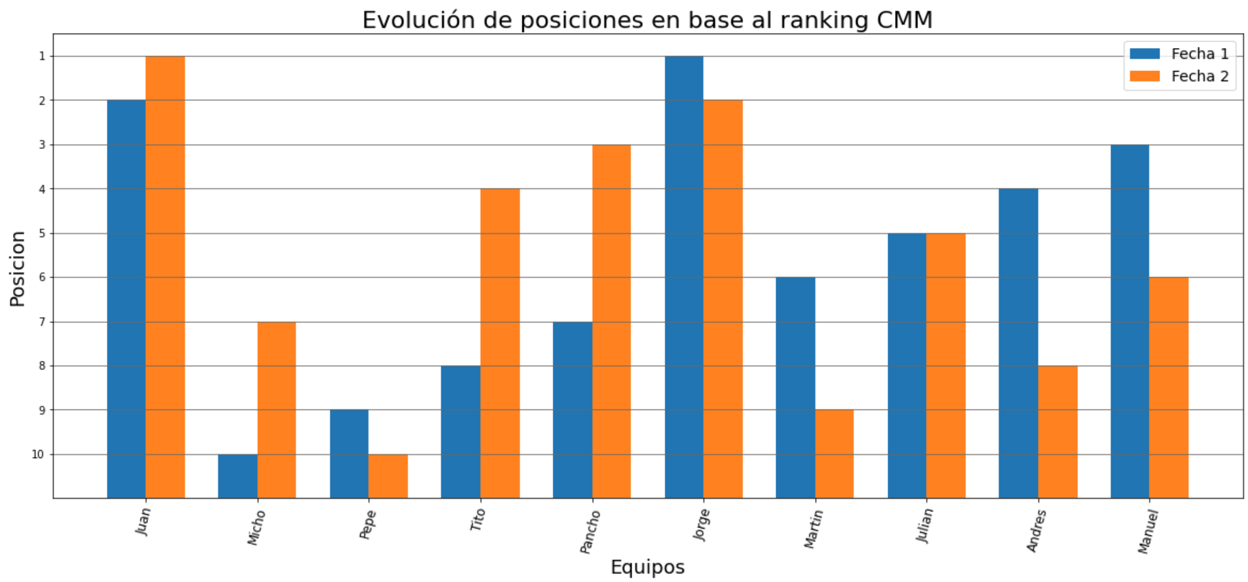

Figura 10: Comparación de las posiciones de los jugadores en el ranking entre la primera y segunda fecha.

En conclusión vimos que en *CMM* un partido entre dos equipos puede afectar el ranking de un tercero. Esto puede parecer injusto si analizamos como se modifica el ranking fecha a fecha. No obstante, si miramos el final del campeonato con todos sus resultados, es un buen sistema para calcular los rankings ya que tiene en cuenta el nivel de los equipos con el que uno se enfrenta en base a sus resultados de todo el torneo, en otras palabras, considera que ganarle al ultimo es mas fácil que ganar al primero.

4.3. Estrategia de maximización de resultados

Para llevar a cabo las dos estrategias ya descriptas tomamos como dataset el torneo de la NBA del 2015/2016 y nos enfocamos en Boston y Atlanta como actores primarios y Brooklyn y Cleveland como actores secundarios.

	id	Equipo	CMM	WP
0	1	Atlanta	0.55882185437765941	0.56716417910447758
1	2	Boston	0.56568988121671293	0.58208955223880599
2	3	Brooklyn	0.29073556727557681	0.28358208955223879
3	4	Cleveland	0.69242825092263127	0.71212121212121215

Figura 11: Tabla inicial de los equipos de la NBA seleccionados con sus respectivos rating CMM y WP

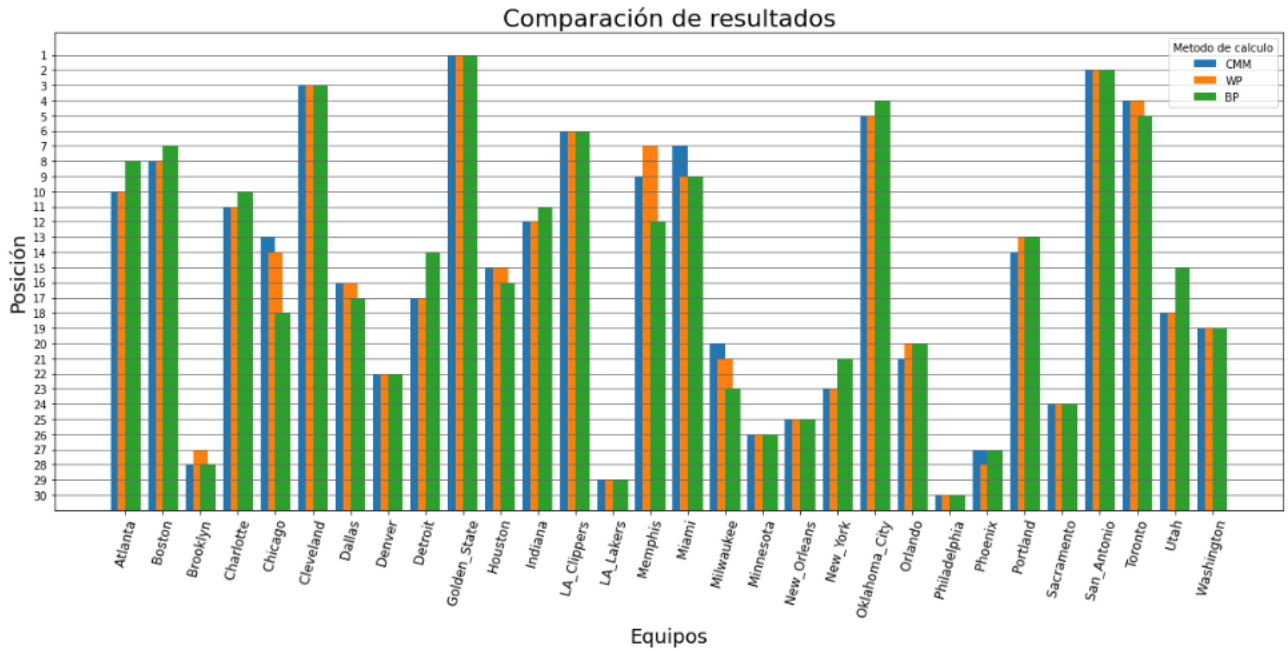

Figura 12: Gráfico inicial de los equipos de la NBA con sus respectivos rating CMM y WP

4.3.1. Alterar el pasado conocido

En este caso hicimos que Atlanta mantenga su rating *WP* pero en vez de ganarle a Boston el 24/11/2015 pierda ese partido y que le gane a Cleveland el 21/11/2015. Para analizar los cambios realizados nos vamos a centrar en Atlanta que antes del cambio tenía un rating en *CMM* de 0,55882185437765941 y luego de cambiar los dos partidos manteniendo el *WP*, su *CMM* se incrementó a 0,55922627861056784. A simple viste es un cambio menor, pero realizando estos pequeños cambios se puede llegar a un resultado significativo cambiando resultados de mas partidos.

	id	Equipo	CMM	WP
0	1	Atlanta	0.55922627861056784	0.56716417910447758
1	2	Boston	0.57963202968771554	0.59701492537313428
2	3	Brooklyn	0.29113904128465745	0.28358208955223879
3	4	Cleveland	0.6783274372671112	0.69696969696969702

Figura 13: Altero el pasado con *CMM* en azul de la NBA (Tabla)

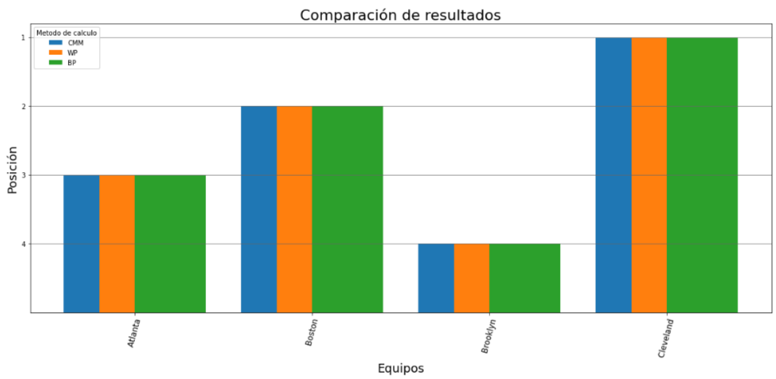

Figura 14: Altero el pasado con *CMM* en azul de la NBA (Gráfico)

4.3.2. Alterar el futuro con pasado conocido

La idea de alterar el futuro conociendo sin modificar el rating *WP* del Atlanta hacer que suba un puesto y poder plantear una estrategia a futuro para poder maximizar la posición final. La estrategia fue hacer que algún equipo que este mejor posicionado que Atlanta (en este caso Boston) pierda contra alguno peor (Brooklyn) agregando un partido mas al torneo. Este cambio, resulto en que los rating del *CMM* para Atlanta pasaran de 0,55882185437765941 a 0,55881631642477514 y el de Boston descienda de 0,56568988121671293 a 0,55531016581019521 logrando que Atlanta pase del 10 al 9 puesto (mirando la tabla general con todos los equipos) agregando un solo partido, habiendo realizado previamente un buen análisis del rating de los equipos. Mas allá de haber descendido en el rating *CMM* logró pasar al rival que tenia arriba porque su descenso fue mucho mayor.

	id	Equipo	CMM	WP
0	1	Atlanta	0.55881631642477514	0.56716417910447758
1	2	Boston	0.55531016581019521	0.57352941176470584
2	3	Brooklyn	0.30107431011640262	0.29411764705882354
3	4	Cleveland	0.69225988923386117	0.71212121212121215

Figura 15: Altero el futuro con CMM en azul de la NBA (Tabla)

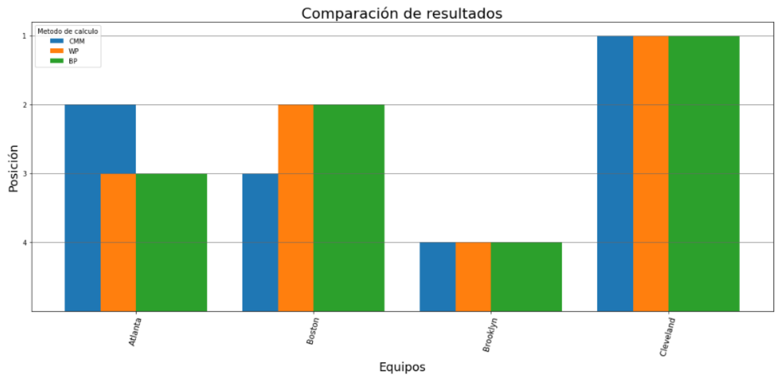

Figura 16: Altero el futuro con CMM en azul de la NBA (Gráfico)

4.4. Estabilidad de los cálculos

Por ultimo, hay que tener en cuenta que los cálculos que se hacen para resolver el sistema de ecuaciones $Cx = b$ se están haciendo en una computadora con aritmética finita, por lo que no van a ser exactos. Para los tests de la cátedra se tiene para cada uno, los resultados esperados. Para observar la estabilidad de cálculos lo que hicimos fue comparar las salidas esperadas contra la salida de nuestro programa. Esta comparación se baso en tomar todos los vectores r_i de una salida r (rating), sumar sus errores absolutos y dividirlos por la longitud de r. Los resultados obtenidos fueron los siguientes:

	Test	Error promedio
0	test_completo_10_1	5.389573e-09
1	test_completo_100_4	8.263856e-07
2	test_completo_1000_2	1.493597e-05
3	test_completo_1000_8	6.738284e-05
4	test_completo_100_8	2.467624e-06

Figura 17: Promedio de error por test

Podemos observar en la tabla de la figura (17) que para los tests con mayor cantidad de equipos y partidos, ocurre que el error es mayor. En especial con el archivo `test_completo_1000_8` que contiene 1000 equipos y mas de 2000000 de partidos. Es esperable que el error aumente a medida que agregamos equipos y datos. El error obtenido no resulta ser muy significativo, pero tiene que ser tenido en cuenta como una variable extra al hacer las operaciones de matrices cuando tenemos aritmética finita.

5. Conclusiones

Luego de haber desarrollado y analizado los 3 métodos, podemos concluir que ningún método es objetivamente superior a otro ya que cada método analiza diferentes aspectos de los resultados de las competiciones. Por ejemplo pudimos ver que *CMM* resulta un método mas efectivo que *WP* en el torneo de tenis ATP porque logra analizar el nivel de contrincante cual se enfrenta. También pudimos ver que el método *CMM* es sensible a los resultados de enfrentamiento de otros equipos y eso puede afectar a su ranking. Por otro lado viendo el método *BP*, notamos que es un método que tiene un incentivo extra al *WP* ya que agrega el factor de cuan parejo o disparejo fue el resultado de los enfrentamientos, generando un ranking mas justo cuando lo que importa es la cantidad de puntos que hace cada equipo por partido.

En cuanto a plantear una estrategia para obtener la mejor posición posible minimizando la cantidad de victorias, vimos que es posible lograr esto sin tener que obligatoriamente ganar mas partidos y hacer cambios muy grandes. Con analizar bien la situación del equipo y ciertos partidos importantes se puede llevar a cabo una buena estrategia y lograr el objetivo pedido.

Referencias

[1] The Colley Matrix Explained - `https://www.colleyrankings.com/matrate.pdf`

[2] Statistics of Colley's Ranking Methodology -
`https://squared2020.com/2017/11/04/statistics-of-colleys-ranking-methodology/`

[3] Premier League 2018/19 Dataset -
`https://footystats.org/c-dl.php?type=matchescomp=1625`