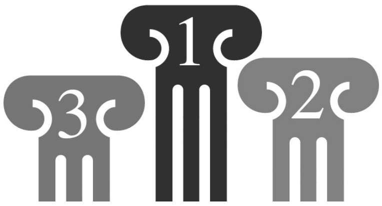
DEPARTAMENTO DE COMPUTACION
Facultad de Ciencias Exactas y Naturales - UBA

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 se define por las victorias de sus campeones y en consecuencia por como se rankean sus participantes para determinar al ganador. Algunos torneos permiten maneras simples de decidir la victoria, como aquellos donde todos los equipos juegan entre si. Sin embargo, en los casos de ligas con *fixtures* más complicados donde no todos los equipos se enfrentan, ni cada uno juega la misma cantidad de veces, nace un problema de interés computacional en la estadística deportiva sobre como determinar el *ranking* correcto, el cual será explorado en el presente trabajo.

Existen competencias deportivas de gran importancia como la Copa Mundial de Fútbol de la FIFA donde se utilizan puntajes para avanzar de ronda hacia la final y el calculo de ranking resulta casi trivial. Sin embargo, esto no siempre es la norma y más adelante se estudiaran otros deportes mas controversiales para generar rankings. Son de interés aquellos donde no basta con solo tomar el historial de partidos ganados y perdidos, es decir, en estos casos de estudios también se evaluara si la dificultad competitiva de cada encuentro es relevante y si debe ser tomada en cuenta o no. De tal forma, se realizara un análisis cuantitativo y cualitativo de los datasets de la NBA (*National Basketball Association*), la NFL (*National Football League*), NHL (*National Hockey League*) y la ATP (*Association of Tennis Professionals*). La elección de estos deportes se efectúa considerando su adecuación al problema planteado, así como su popularidad y la relevancia con el fin de encontrar un método justo para determinar el campeón según un estándar de habilidades. Ante esto, 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 número 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 busca resolver el problema del *ranking* deportivo, por lo cual se estudiarán 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 también llamado SR) para comparar su capacidad de representar correctamente al mejor equipo de un conjunto $\Gamma = \{1, 2, ..., T\}$ donde T es la cantidad de equipos, y cada equipo participa al menos una vez en el torneo en cuestión.

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, por lo cual es mucho mas simple que el anterior. En este el *rating* de un equipo $i \in \Gamma$ para n_i partidos jugados se calcula como el porcentaje de partidos ganados w_i dividido entre el total de juegos, de tal forma que queda $P_i = \frac{w_i}{n_i}$. Como se analizara próximamente, se puede modificar según la necesidad de un deporte de modelar empates al agregar una penalidad a los partidos empatados t_i, donde estos son multiplicados por 0,5 para luego sumarlos al total de victorias sobre total de partidos de forma que $P_i = \frac{w_i + t_i}{n_i}$.

El método de Score-Ratio se basa en calcular una métrica utilizando la suma de las diferencias entre los puntos obtenidos por un equipo en cada partido y los de su contrincante. Estas diferencias se dividen sobre el puntaje total resultante de todos los eventos de equipo a rankear, es decir, sobre la suma de todos los puntos de todos los equipos para cada partido del equipo en cuestión. Formalmente 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, 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. La visualización de los gráficos y experimentos se realiza mediante jupyter notebooks.

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 de 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 $Colley(Equipos, res)$
  $C \leftarrow crearMatriz(filas = T, cols = T + 1)$ $\triangleright\ O(n^2)$
  for $equipo_i$ **in** $Equipos$ **do** $\triangleright\ O(n^2 * log(n))$
    for $equipo_j$ **in** $Equipos$ **do** $\triangleright\ O(n * log(n))$
      if $i == j$ **then**
        $C[i][j] \leftarrow 2 + equipo_i.partidos\_totales$
      else
        if $equipo_i.cant\_matches\_con(equipo_j.id) \neq 0$ **then** $\triangleright\ O(log(n))$
          $C[i][j] \leftarrow -equipo_i.cant\_matches\_con[equipo_j.id]$
        else
          $C[i][j] \leftarrow 0$
  for $equipo_i$ **in** $Equipos$ **do** $\triangleright\ O(n)$
    $C[i][T + 1] \leftarrow 1 + (equipo_i.p\_ganados - equipo_i.p\_perdidos)/2$
  $Elim\_Gaussiana\_Sin\_Permutaciones(C, res)$; $\triangleright\ O(n^3)$
  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 sentido, el motivo por el cual hacemos esta diferenciación, es que en torneos donde no todos los equipos jueguen contra todos, se puede estar asignando un ranking que no se considere "justo", ya que por ejemplo un equipo que juegue un 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 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`, aunque cabe destacar que los partidos totales son obtenidos por una función que suma la cantidad de partidos ganados con los perdidos. Fuera de esto, 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 por el equipo, y en particular su importancia 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

En esta sección se analizan y reportan los errores obtenidos al comparar el ranking de Colley implementado en C++ por equipo con respecto a los test oficiales de la cátedra. Estos test se encuentran en archivos de texto con la extensión `.expected` con los siguientes nombres:

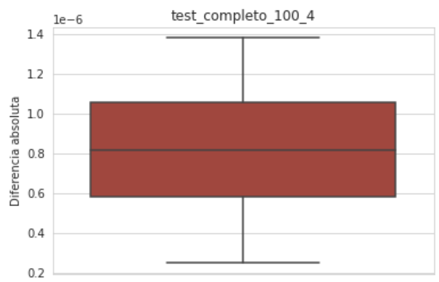
- test_completo_10_1
- test_completo_100_4
- test_completo_100_8
- test_completo_1000_2
- test_completo_1000_8

y además, incluyen archivos de texto de extensión `.in` para la reproducción de los rankings, y por ultimo scripts de Python para verificar la similitud de los resultados buscando una diferencia no mayor a un δ de 1×10^{-4}.

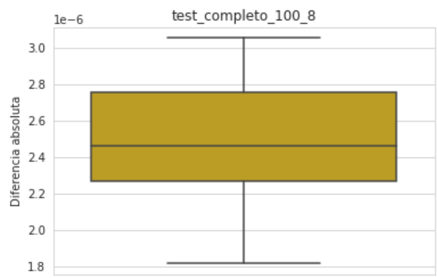
Luego de ejecutar el análisis comparativo entre los ranking esperados y obtenidos, al calcular la diferencia entre estos, se obtuvo que el método de Colley pasó exitosamente todos los tests de la cátedra y por ende tienen un error absoluto menor al delta propuesto. Por otra parte, al observar la distribución de las diferencias entre los resultados obtenidos se puede obtener información relevantes sobre la estabilidad de los cálculos y la confianza (fiabilidad) o precisión del método desarrollado.

El análisis de este error numérico es importante debido a que si se presentarán desviaciones lo suficientemente grandes cambiaría el resultado del ranking. De esta forma no se representaría la realidad correctamente y ocasionaría repercusiones sociales y económicas en el ámbito deportivo por elegir al ganador incorrectamente. Ahora bien, la hipótesis en cuestión sera que a mayor cantidad de equipos, mayor sera el error numérico ya que aumentan los cómputos y se acumulan las desviaciones. También es necesario ver que este error es lo suficientemente chico para los casos prácticos de uso, lo cual ya esta acotado por los test de la cátedra que contemplan hasta 1000 equipos, y es suficiente para cualquier competencia deportiva estándar.

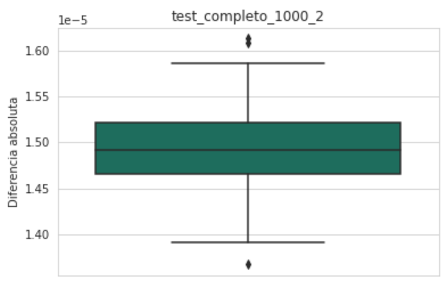
Este análisis del error se realiza mediante dataframes en Jupyter Notebooks, calculando el diferencial rating a rating de cada dataset (fila a fila). Por ejemplo para el caso del test `test_completo_100_4` al tener 100 equipos se obtienen 100 diferencias las cuales se visualizan mejor mediante un boxplot que permita ver la naturaleza de la distribución y como varían estas métricas, junto a su media y percentil. De tal forma en la Figura 1 se puede visualizar un análisis individual de la distribución de cada una de las diferencias obtenidas en el dataset.



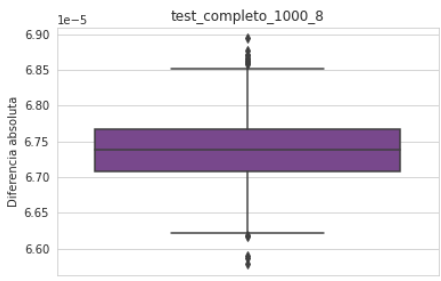
(a) Caso de Test Completo 100_4



(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.

Es importante destacar que de los 5 datasets se omitió el `test_completo_10_1` por su poca cantidad de casos (10 equipos) y consecuente mínima relevancia en la distribución del error, ya que como se menciono anteriormente, se busca estudiar los casos grandes con muchos equipos donde pueda aumentar el error numérico.

En la Figura 1 a primera vista se puede denotar que el rango en el que oscilan los errores parece ser estable o chico, con una poca presencia de outliers. En particular para los test de 100 equipos estas excentricidades no existen y aparecen únicamente para los test de 1000 equipos. Este es el primer indicio de que las desviaciones aumentan para una mayor cantidad de equipos. Por otra parte se puede ver una diferencia marcada en el orden de magnitud del error absoluto. Para los casos de test de 100 equipos este se mantuvo en 1×10^{-6} pero a medida que aumentan los equipos, como se esperaba aumenta el error a 1×10^{-5}. Por suerte, esto aun es aceptable y se estima que se debe al error numérico de realizar operaciones con aritmética finita.

En la implementación actual del algoritmo se utilizan variables con una buena capacidad de representación para los estándares actuales (`double`) pero sin embargo por la naturaleza finita de esta representación siempre existirá un limite en la precisión del computador. A efectos prácticos de la investigación la precisión alcanzada se considera suficiente y a continuación se presenta en la Figura 2 un comparativo de los promedios de error absoluto por caso de test, donde se corresponden los colores con respecto a los boxplots expuestos previamente y se evidencia el incremento del error a medida que crece en dataset de entrada.

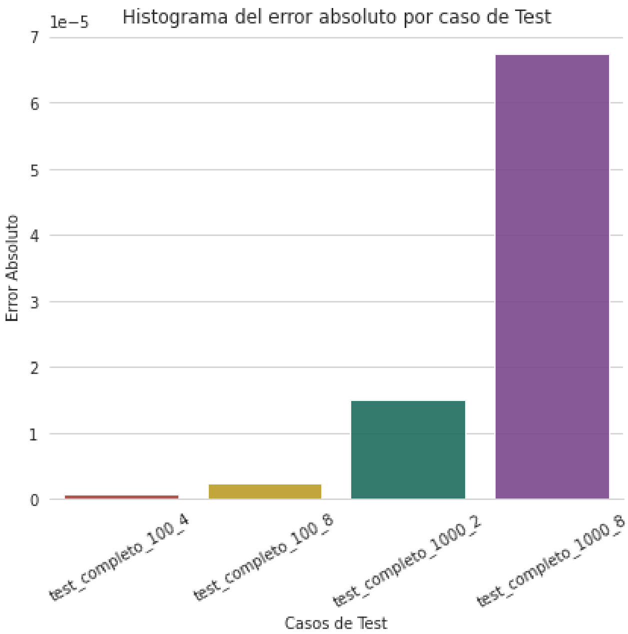


Figura 2: Barplot de los errores absolutos de cada caso de test

Además, para mejorar el grado de detalle del análisis de la diferencia absoluta se presenta en el Cuadro 1 las propiedades más relevantes como el promedio, la mediana y la desviación estándar sobre cada distribución del error por dataset. En esta tabla es importante destacar como la desviación es muy chica y por ende la mayoría de las diferencias están muy cerca del promedio de su espacio muestral, lo cual verifica la estabilidad numérica de los cálculos para los efectos requeridos.

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	0.8263×10^{-6}	0.8148×10^{-6}	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}

Entrando en detalles, se puede observar en el Cuadro 1 como se verifica la hipótesis planteada de que una mayor cantidad de equipos implicaba un error mayor. El test con 100 equipos y 12410 partidos tienen un orden de magnitud de 10^{-6} y una mediana de 0.8148×10^{-6}, es decir, la diferencia es muy muy pequeña, y ocurre algo similar si se duplica la cantidad de partidos para el caso de 22318 eventos con el mismo orden de magnitud y una mediana de 2.4629×10^{-6}.

Seguidamente, para el test con 1000 equipos y 749273 o 2248231 eventos aumenta la diferencia como se esperaba a 10^{-5} con una mediana en 1.4922×10^{-5} y 6.7378×10^{-5} respectivamente. Es decir, se observa que al aumentar el dataset se replica el error numérico a medida que crece el input. Sin embargo, como se menciono anteriormente esto no es significativo, y en particular para la estadística deportiva no es común utilizar rankings en base a tantos eventos como el caso de Test con 2248231 encuentros y 1000 equipos.

2.6. Análisis Cualitativo

En la sección anterior se corroboró la estabilidad numérica de los cálculos del método propio de CMM y ahora se busca identificar características distintivas de cada método desarrollado (CMM, WP y SR). Esto se logra aplicándolos a datos de competencias reales, a fin de relacionarlas estos con los eventos resultantes.

Los datos son extraídos de los datasets mencionados en la Sección 2.1, de los cuales dos fueron proporcionados por la cátedra (NBA y ATP) y otros dos fueron elegidos para enriquecer el análisis (NFL y NHL). Estos se encontraban inicialmente como un `.csv` y fueron filtrados mediante scripts de Python para crear archivos de texto `.in` como input para los algoritmos y a su vez eliminar la información no relevante al enfoque de esta investigación (localidad, estadio, etc).

2.6.1. NBA

Se inicia el análisis cualitativo comparando los resultados obtenidos con respecto al resultado oficial de la Temporada 2015-2016 de la NBA[1]. En el Cuadro 2 se lista el ranking oficial de la temporada seguido del obtenido por cada método. En general, los resultados obtenidos mediante Colley y WP son bastante similares al ranking real ya que en el basketball se suele utilizar la cantidad de partidos ganados-perdidos como principal criterio para rankear equipos sin tomar en cuenta la ventaja de la victoria. Al menos para el top 5 se puede apreciar en la tabla que son idénticos, pero luego, **¿A que se deben las diferencias?**. Para responder esta incógnita se debe analizar en detalle el formato del campeonato de la NBA.

La NBA esta conformada por 30 equipos los cuales se dividen en 2 conferencias de 15 equipos cada una según su ubicación geográfica. Para tratar de garantizar equidad competitiva, existe una ronda de playoffs o eliminatorias donde deben calificar 8 equipos de cada conferencia, y donde estos enfrentamientos definirán al ganador del torneo. Pero, esta clase de restricciones para clasificar a la fase final puede generar ciertos inconvenientes entre los equipos. Un ejemplo de estos es que puede ocurrir que un equipo de la conferencia Este tenga un rating mayor a algún equipo del Top 8 de la conferencia Oeste, pero si el rating de este equipo es menor al rating del 8^{vo} equipo de su conferencia correspondiente (Este) no podrá entrar a los playoffs, y viceversa. Esto se traduce en que el rating para calificar a un playoff no es el absoluto del torneo, sino el relativo al Top 8 de su conferencia.

Para entender mejor la diferencia entre el Ranking real y el método de Colley y WP en la tabla 2 se colorean las celdas de cada equipo según la conferencia a la que pertenecen, asignándole el Azul a la Conferencia Este y rojo a la Conferencia Oeste. Esto permite visualizar que como WP y Colley no toman en cuenta los playoffs, entonces permiten inconsistencias con el mundo real ya que el Top 16 debe contener 8 equipos de cada conferencia, pero se pueden apreciar en la tabla que para estos métodos calificaron 9 equipos de la conferencia Oeste (rojo) y 7 de la Este (azul). Esto sucede porque Detroit Pistons (Azul) se queda afuera del top 16 en ambos métodos ya que tiene un ranking menor al de Houston Rockets (Rojo), mientras que en el ranking real Houston se queda afuera del Top 16 porque existen otros 8 equipos de su conferencia con un ranking mayor.

Estos colores no son resaltados en Score-Ratio para mejorar la claridad de lo que se desea probar. Es evidente que para este método tampoco se toman en cuenta las restricciones de las conferencias del torneo y en particular se puede apreciar un ranking casi arbitrario o bastante diferente al ranking real. Esto se debe a como se mencionó anteriormente, a que Score-Ratio toma información diferente a Colley o WP para calcular su métrica. Y por ende sus rankings se basan en la diferencia de puntos y no en la cantidad de victorias o derrotas. Esto muestra como adoptar otro sistema de rankeo puedo cambiar completamente el resultado de una temporada, ya que se destaca que con el método de Score-Ratio cambia incluso el ganador de la competencia, en este caso *San*

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

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

Antonio Spurs en vez del oficial *Golden State* y queda evidenciada la influencia del modelo del deporte y la información en el resultado final del campeonato.

Analizando más en detalle los resultados obtenidos se puede apreciar que las ligeras diferencias entre CMM y WP no varían en más de una posición. Por ejemplo esta el caso de Miami en la posicion 7 según CMM y en la posicion 8 según WP, y viceversa para el caso de Boston. En el cuadro 3 se pueden apreciar los valores numéricos del rating de cada equipo por método a fin de analizar mejor esta variación. En la tabla se evidencia que los números son muy cercanos y tienen una diferencia acotada entre 1×10^{-2} y 1×10^{-3}, por lo que esta permutación de resultados se debe a la sensibilidad de Colley a otros factores del fixture que WP no toma en cuenta como el rating de los contrincantes. Es interesante mencionar que la selección del rating del Top 16 durante la temporada puede afectar las probabilidades de los equipos de ganar ya que en base a este fixture se asignan los equipos que juegan de local y visitante, y como es sabido en la estadística deportiva, hay un mayor chance de ganar en el estadio propio de un equipo.

Cuadro 3: Permutaciones entre Colley y WP

Nro.	Equipo	Colley	WP
..	...	...	...
7	Miami Heat	0.577367	0.582090
8	Boston Celtics	0.565690	0.582090
..	...	...	...
13	Portland Trail Blazer	0.515941	0.514706
14	Chicago Bulls	0.516007	0.507692
..	...	...	...

Volviendo al analisis de Score-ratio, se tiene otro item relevante que es este permite a un equipo en un evento generar cambios significativos sobre su rating, como es el caso de Utah Jazz (Gris) que alcanzó la posicion 12 mediante este rating pero no logró entrar al Top 16 en ninguno de los otros 3 métodos. Esto se debe a que aunque en general gano menos partidos que otros equipos como Portland, Detroit, Dallas y Houston, tuvo juegos donde alcanzo la victoria por gran diferencia de puntos como fue el partido del 28 de marzo de 2016 donde venció a los Lakers por 123 a 75 puntos, o el del 10 de abril del 2016 donde le ganó a los Denver Nuggets por

100 a 84, donde es interesante mencionar que ambos equipos no pertenecen al Top 20 de la temporada y por ende se estima que era mas fácil ganarles que a otros equipos de mayor rating.

Por ultimo, revisando un poco la historia de la NBA se encuentra que en el 2011, 2012 y 2015 existieron quejas con respecto a la inequidad de los playoffs al permitir equipos con menor desempeño calificar a estas eliminatorias. Por lo cual se recomienda utilizar CMM como ranking oficial de la NBA, ya que con este los equipos que llegarían a los playoffs serian los mejores absolutos sin tomar en cuenta cupos por conferencia (aunque capaz esto no sea tan atractivo para la fanáticada). Actualmente, el CMM ofrece una alternativa mas justa, o con menos sesgos que la actualmente utilizada y mejoraría la representación del desempeño de un equipo en su rating ya que se espera que propicie playoffs mas competitivos por esto mismo.

2.6.2. NFL

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

Actualmente la NFL cuenta con 32 equipos divididos en 2 conferencias (NFC, AFC) de 16 equipos. Además, cada una se encuentra categorizada en 4 divisiones (Norte, Sur, Este, Oeste) y cada división se encuentra conformada por 4 equipos. En este deporte los *standings* o rankings son determinados por puntos en base a partidos clasificatorios para llegar a una ronda eliminatoria de playoffs donde existen cupos por división en cada conferencia. El criterio para rankear a los equipos en orden descendente sigue el formato de mejor porcentaje de victorias-derrotas, o en inglés: Winning Percentage. En caso de empate, el ganador se decide mediante una lista de *tiebreakers*[3].

Es importante observar que el dataset utilizado en el informe proviene de la página MasseyRatings.com [4], la cual contempla partidos de la liga en el período 09/2019 - 01/2020, mientras que en la página oficial de la NFL solo se contempla el período 09/2019 - 12/2019. Es por este motivo que para replicar el análisis será necesario solo contemplar los partidos jugados hasta la fecha 12/2019 inclusive, ya que como mencionamos los posteriores a esa fecha se tratan de casos sobrantes por lo cual solo se uso los juegos hasta esa fecha del dataset.

Cuadro 4: Comparación Top 15 de los rankings resultantes de la NFL.

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

Para visualizar mejor las variaciones en los ratings de cada equipo, dependiendo del método utilizado en la tabla 4, se separan los casos de estudio en bloques de colores. La primera división es la Azul que permite observar que el ganador del ranking oficial es el mismo que con los rankings obtenidos con los otros métodos. Como segunda y tercera división en colores se tiene el Amarillo y Verde donde se puede apreciar que el ranking oficial dista de los resultados obtenidos en los métodos, y a continuación analizaremos en profundidad porque ocurre cada caso.

- En el caso de Colley, San Francisco obtuvo un rating (aproximado) de 0.80 mientras que tanto New

[2] https://www.nfl.com/standings/league/2019/reg/
[3] https://www.nfl.com/standings/tie-breaking-procedures

Orleans como Green Bay obtuvieron ratings (aproximados) de 0.78 y 0.73 respectivamente. Por este motivo si bien podemos decir que el ranking obtenido utilizando el método de Colley se asemeja al oficial, aun así presenta discrepancias en otros ratings por la naturaleza del método de Colley en comparación al oficial.

- En el caso de Winning percentage, se puede observar que coincide el ranking con el oficial. Si analizamos los ratings oficiales de Green Bay con 13/16 ganados-perdidos y al igual que New Orleans, con lo cual la diferencia entre el ranking oficial y el obtenido con WP presenta variaciones, ya que para el caso de WP aunque ambos tienen el mismo puntaje, uno debe tener una posición mayor a otro como se puede apreciar en la tabla.

- En el caso de score ratio, se identificó que el resultado del ranking oficial dista en gran cantidad del obtenido. Esto ocurre porque toma en el cuenta diferencial de puntos (puntos anotados/puntos totales), y observando el rating de New Orleans (0.14) vemos que es inferior que el de Minnesota Vikings (0.15). A su vez, el rating de Green Bay (0.09) es inferior al de New Orleans, explicando de esta manera la disposición de los rankings analizados.

Como cuarta división en bloques con colores se tiene el Rojo y Rosa. En los casos representados en rojo podemos observar que los rankings obtenidos coinciden totalmente con los oficiales, mientras que los casos representados en color rosa, se tratan de permutaciones de los puestos oficiales, un comportamiento que se explica entendiendo que en el presente trabajo los métodos desarrollados no toman en cuenta las condiciones de los *tiebreakers* de la temporada.

Por último, haremos una última división en bloques con color Celeste, para mostrar la discrepancias entre CMM y el ranking oficial. Si bien el método de Colley ofrece una aproximación bastante buena al resultado real, no es tan bueno como el de Winning Percentage, por tener diferencias mas significativas con respecto al ranking real.

2.6.3. NHL

En esta sección se compara el ranking oficial de la NHL[4], con los distintos rankings obtenidos utilizando los 3 métodos estudiados en el presente informe. Para esta sección es importante destacar el formato de rankeo de la NHL donde se utiliza un sistema de puntos diferente a los métodos implementados por lo cual se esperan fuertes discrepancias, así como una mejor visualización del efecto disruptivo de cambiar el método de ranking para un deporte en particular.

Actualmente la NHL cuenta con 31 equipos (nótese la imparidad) divididos en 2 conferencias Este y Oeste de 16 y 15 equipos respectivamente. Además, cada una se encuentra categorizada en 2 divisiones (con subdivisión impar de equipos en el caso de la Oeste). En este deporte los *standings* o rankings son determinados por puntos en base a partidos clasificatorios para llegar a una ronda eliminatoria de playoffs donde existen cupos por división en cada conferencia. Se asignan 2 puntos por victoria, 1 punto por derrota en tiempo extra, y 0 puntos por derrota.

El ranking final de los playoffs se crea calculando la suma de estos puntos para que los primeros 3 equipos de cada división avancen a la fase final. Es interesante destacar que estos serían solo 12 equipos de los 16 que juegan en las eliminatorias (3 por división). Y luego se eligen dos *wildcards* o equipos por conferencia indistintamente del cupo por división, considerando únicamente el ranking absoluto de la conferencia. En otras palabras, podrían llegar hasta 5 equipos por división a la fase de playoffs lo cual es una ventaja considerable en equidad competitiva con respecto al caso mencionado de la NBA.

En el Cuadro 5 se encuentra el comparativos de los rankings estudiados donde se resaltan con colores las celdas correspondientes a los primeros 8 equipos del *standing* oficial, con el fin de observar y analizar como varían las posiciones según el método. En el mismo cuadro se puede observar resaltado con color Azul, que el puesto 1 del ranking oficial coincide con los obtenidos según los métodos de Colley, Winning Percentage y Score Ratio.

Viendo en detalle, atribuimos el resultado obtenido en el podio, según el metodo de Colley a que Tampa Bay resultó ser el equipo con mejor desempeño general de la temporada obteniendo un rating de 0,71(aprox). De la misma manera, en Winning percentage se encuentra rankeado Tampa Bay en primer puesto, con un rating de 0.72(aprox) ya que fue el que ganó mas partidos con 62 victorias y 16 derrotas. Por último, en Score Ratio el puesto de Tampa Bay es el primero con un rating de 0.16 (aprox), porque el diferencial de puntos que anotó durante los partidos (103) resulta mayor que el de los demás equipos (62 o menos).

Si bien se puede observar que en principio los resultados oficiales parecen similares a los obtenidos en

[4]https://www.nhl.com/standings/2018/league

Cuadro 5: 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

los métodos presentados, a partir del 4^{to} puesto del ranking oficial se aprecia que comienzan a presentarse discrepancias significativas, tales que a partir de este punto no existen semejanzas entre el resultado oficial y el generado por alguno de los métodos estudiados. Esto es atribuido a que durante el desarrollo de los algoritmos no se realizaron distinciones entre las derrotas normales o con penalidad, ya que esta información (sistema de puntos) no es considerada por los métodos desarrollados, y por ende se visualizan tales diferencias entre los resultados obtenidos contra los oficiales, ya que la forma de calcular las métricas son considerablemente diferentes entre si.

2.6.4. ATP

A continuación, se comparan los métodos implementados para el dataset del torneo ATP de Tenis del 2015 con los datos reales[5] 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, podemos ver que si solo se tiene en cuenta si un equipo ganó o perdió y el puntaje queda reducido a mostrar solo eso, entonces resulta ser que WP es un caso particular de SR. Por lo que durante esta sección (sin perdida de generalidad) se dejara de lado el método de SR puesto que presenta exactamente los mismos resultados que WP.

Dejando establecido lo anterior, se puede empezar a analizar los resultados, los cuales presentan comportamientos muy fuera de lo esperado, ya que se logra visualizar que WP tiende 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.

En el Cuadro 6 se tiene el Top 5 de WP donde se puede notar lo anteriormente mencionado, los jugadores tienen un rating saturado al máximo. La explicación de esto viene de la cantidad de partidos que jugaron estos, ya que analizando el fixture de los 5 jugadores se tiene que jugaron entre 2 a 3 partidos cada uno, (debido a que hay muchos jugadores que participan de subtorneos cortos) y con el detalle de haber ganado todos los partidos jugados. Bajo condiciones similares se encuentran los otros 20 equipos debajo de estos cinco, los cuales sobresaturan por la naturaleza de los métodos. Incluso se observa que Colley los deja posicionados encima de la mediana de su rango por la misma razón, pero como este método considera la cantidad de equipos del torneo entonces los jugadores obtienen un ranking mas realista.

Debido a esto se tomo la consideración de que WP no fue buen estimador del ranking en este torneo por la singularidad del fixture, donde algunos jugadores tienen muy pocos partidos haciendo que si ganan todos sus partidos el método los coloque en el Top. Se podrían pensar que fuera de la saturación del Top, empieza a rankear "medianamente bien" a partir de cierto punto. De la mano de esto ultimo surge la idea de poder filtrar de algún modo estos partidos que afectan a los métodos mencionados, sin modificar los métodos.

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

Nombre	Colley	WP	Partidos jugados	Partidos ganados
Garay Christian	0.677645	1.0	2	2
Ghareeb Mohammad	0.794806	1.0	3	3
Qureshi Aisam Ul Haq	0.742604	1.0	2	2
Lama Gonzalo	0.724678	1.0	2	2
Garin Christian	0.722988	1.0	3	3

Cuadro 6: Comparacion del rating del Top 5 de WP con los partidos correspondientes a cada jugador

Solo con el objetivo de darle una segunda oportunidad a WP se plantea que se puede realizar un mejor rankeo de los mejores jugadores filtrando solo los partidos mas importantes, es decir que solo tomamos en consideración todos los partidos del *ATP Tour Top-level tournament singles champions* [6], el cual alberga 4 de los torneos mas importantes del tenis y no tienen jugadores que jueguen un numero muy pequeño de partidos.

El resultado de este nuevo análisis se encuentra en el Cuadro 7 donde se observa que se tiene un ranking mucho mas realista de los jugadores.

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

Cuadro 7: Comparacion del Ranking real vs. los metodos WP y SR en el dataset filtrado.

Sin embargo, vuelve a surgir una singularidad con el jugador Cervantes Huegun Inigo, el cual tiene el podio en WP pero no se encuentra dentro del Top 10 del ranking real. Resulta ser que este jugador satura en WP por ser ganador invicto de (solo) uno de los 4 torneos[7]. Por lo que se vuelve a repetir el dilema de los jugadores que saturan en WP, por lo que finalmente se puede concluir que por la naturaleza y dificultad del rankeo en los torneos del ATP[8] el metodo no es efectivo para poder ser usados en este tipo de deporte.

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

Cuadro 8: Comparación del ranking entre los métodos para el Top 10 de Colley

Sabiendo todo lo anterior se explica porque (en el análisis original) el Top 10 de Colley ni siquiera está incluido en este grupo de equipos saturados en WP como se aprecia en el Cuadro 8, en el cual se tiene que en WP los equipos que en Colley están en el Top 10 están debajo de la saturación al `1.0`, es decir debajo del Top 25 de WP.

[6] `https://en.wikipedia.org/wiki/List_of_ATP_Tour_Top-level_tournament_singles_champions`
[7] `https://www.tenisenuruguay.com/2015/11/cervantes-gano-finales-challenger-tour.html`
[8] `https://ftw.usatoday.com/2018/08/atp-wta-tennis-rankings-how-do-they-work-faq-federer-serena-nadal`

Por último, en el Cuadro 9 tenemos a los resultados de Colley frente a los resultados reales de la temporada (representando los vínculos con colores distintos), 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 puesto que Colley sí considera la cantidad de participantes y la cantidad de partidos jugados.

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

Cuadro 9: Comparación del Top 10 entre los resultados reales y los de Colley para ATP

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: **¿Influye la 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, se buscará analizar la simetría de los partidos, es decir, 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, ya que Colley modificara el rating de un tercero si es que hay alguna relación entre éste y alguno de los equipos que sufre el cambio. 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 rendimiento 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.

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 involucrada en su proceso.

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) y descubrir si estos métodos siempre toman en cuenta la fuerza de los partidos, es decir, si hay partidos mas importantes que otros independientemente del fixture[6]. En particular, para este problema se puede visualizar un contraejemplo donde no siempre se cumple esta premisa, y es en el caso de los fixtures Round-Robin donde posiblemente todos los partidos valdrán igual. A continuación en el Cuadro 10 se puede ver el caso donde todos los equipos juegan contra todos la misma cantidad de veces.

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 tomar 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 el Cuadro 10 se plantea un torneo, donde el equipo A tuvo una temporada perfecta (invicto) y es el 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 10: 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 se agrega un upset en el Cuadro 11, donde el equipo E ahora derrotó al A y por ende cambia el scoreboard. Ahora existe un empate en el primer y tercer 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. Ahora en la tabla 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. Esto no ocurriría en otros métodos como el ELO o el método de Markov los cuales no son explorados en este TP pero se considera relevante mencionarlo.

Mediante este contraejemplo se puede evidenciar que los métodos no toman en cuenta la dificultas de los eventos o la *fuerza de schedule*. Es decir, los métodos no permiten asumir que existen partidos más importantes que otros y en particular victorias más difíciles y con más información que otras, lo cual si es considerado por otros modelos estadísticos mas complejos en otros deportes como el Ajedrez o en el caso de *E-sports* (videojuegos).

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 de puntajes muy altas en los eventos este método permite cambiar

Cuadro 11: 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

completamente el resultado del ranking. 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 no es independiente (como se verá en profundidad en el experimento 3), y se quiere analizar si lo que gana un equipo en rating en un partido es lo mismo que su oponente pierde. Luego, se puede ver en el Cuadro 12 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 12: 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 vale aun menos para el caso de 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

En el presente experimento se realizaron modificaciones a los datasets con los que trabajamos, con el fin de alterar el orden de los partidos jugados para así analizar las variaciones en los ratings obtenidos. En el Cuadro 13 se puede observar la estructura del orden original de los partidos jugados durante la temporada del 2016 según el dataset de la NBA. En el mismo, se encuentran resaltados en color azul y amarillo los primeros dos partidos de la temporada en orden cronológico. El objetivo de colorear dichas celdas es facilitar la comparación entre la disposición original y los ordenes arbitrarios creados para el experimento.

Cuadro 13: Fixture original NBA_2016

Partido Jugado Nro°	Equipo1 vs Equipo2	Puntaje Equipo1	Puntaje Equipo2
1	Detroit vs Atlanta	106	94
2	Chicago vs Cleveland	97	95
3	Golden_State vs New_Orleans	111	95
...	...	...	...
...	...	...	...
1000	San_Antonio vs LA_Clippers	108	87
1001	Sacramento vs LA_Lakers	106	98

Durante la realización del experimento se realizaron 2 modificaciones, la primera consistió en revertir el orden original de los partidos jugados, como se puede observar en el Cuadro 14. Se puede observar que al haber revertido el orden de los partidos, ahora la disposición original de los equipos que se encontraban resaltados cambia, y justamente están dispuestos de manera que coincide con la reversión que se quiere lograr. Se puede ver que ahora el primer partido (Azul) es el ultimo y el segundo (Amarillo) es el penúltimo y así sucesivamente.

Cuadro 14: Fixture revertido NBA_2016

Partido Jugado Nro°	Equipo1 vs Equipo2	Puntaje Equipo1	Puntaje Equipo2
1	Sacramento vs LA_Lakers	106	98
2	San_Antonio vs LA_Clippers	108	87
3	Toronto vs MilWaukee	107	89
...	...	...	...
...	...	...	...
1000	Chicaco vs Cleveland	97	95
1001	Detroit vs Atlanta	106	94

La segunda modificación realizada, consiste en utilizar un algoritmo de randomización uniforme, en el cual se puede observar mediante el siguiente pseudocódigo, como se utiliza la función `random.shuffle()` de Python para aleatorizar el orden de los partidos siguiendo una distribución uniforme. Además, se presenta en el cuadro 15 resaltados en color azul y amarillo, el nuevo orden de los equipos anteriormente mencionados, que ahora se encuentran en posiciones aleatorias de la temporada.

Algorithm 4 Pseudocódigo randomización dataset NBA 2016

```
function RandomizeOrder(DatasetSrc)
    file ← open(DatasetSrc, "r")
    list ← file.readlines()
    file.close()
    random.shuffle(list)
    return list
```

Cuadro 15: Fixture randomizado NBA_2016

Partido Jugado Nro°	Equipo1 vs Equipo2	Puntaje Equipo1	Puntaje Equipo 2
1	Atlanta vs Washington	114	99
2	New_York vs Chicago	107	91
3	Utah vs Indiana	97	76
...	...	...	...
381	Detroit vs Atlanta	106	94
...	...	...	...
395	Chicago vs Cleveland	97	95
...	...	...	...
1000	Golden_State vs LA_Clippers	124	117
1001	Minnesota vs LA_Clippers	108	102

El resultado que se quiere analizar es si el orden de los eventos afecta el ranking de los equipos, y como se puede apreciar en este experimento, dado los diferentes ordenes de los partidos, no se presentan diferencias a pesar del cambio. En particular se realiza este chequeo utilizando el siguiente pseudocódigo:

Algorithm 5 Pseudocódigo diferencias en el orden de eventos

```
function DiferenciasEnOrdenDatasets(DatasetOriginal, DatasetReversed, DatasetRandomized)
    rankOriginal = experimentar(DatasetOriginal)
    rankReversed = experimentar(DatasetReversedl)
    rankRandomized = experimentar(DatasetRandomized)
    if (!cmpEq(rankOriginal, rankReversed) | | !cmpEq(rankOriginal, rankRandomized)) then
        return true
    return false
```

Al comparar las diferencias en los otros datasets presentados en el informe (NFL, NHL, ATP), se obtiene el mismo resultado que en el dataset de la NBA, donde al realizar metodológicamente los mismos pasos y algoritmos para cambiar el orden el ranking se mantiene constante.

Como resultado del experimento se obtiene que efectivamente, no hay diferencias en los rankings obtenidos al observar los archivos que contienen los datasets con distinto orden, y se confirma la hipótesis planteada. Se refuerza la idea de que el orden no importa en los métodos utilizados debido a que los cálculos no hacen distinción según como sea el orden de los equipos. Sin embargo, se conocen otros métodos de rankeo que no se implementaron en este trabajo debido a que su complejidad excede el objetivo del estudio en curso, en los cuales el orden si importa; un ejemplo de ellos se trata 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 3. 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.

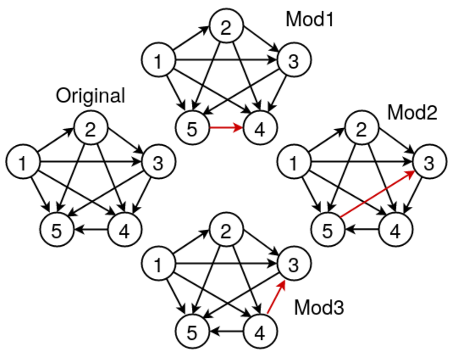


Figura 3: Ejemplo de torneo

Como se puede apreciar en el Cuadro 16, 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. Este comportamiento esta justificado solo porque el resultado del partido cambiado no afecta a los dos equipos que ganaron todos sus partidos. 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 observa el mismo comportamiento en cuanto al rating de cada equipo en cada modificación ya que se presenta la misma situación mencionada con el caso de Colley.

Cuadro 16: 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

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 [9].

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

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 se presume que tanto el podio como todos los rankings se verán afectados por este ligero cambio en dicho partido.

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

Como se ve en el Cuadro 17 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. Es decir que el rating dado por WP y SR de San Francisco y de los equipos que no estaban relacionados con el partido modificado, no mostraron cambio alguno debido a que el partido modificado no cambio la cantidad de partidos ganados, ni el score acumulado de ellos.

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

El experimento es gestionado por un script en Python el cual lo único que hace es tomar un timestamp antes de correr el programa con los parámetros correspondientes, y tomar un timestamp luego de terminado este. Y finalmente se toma la diferencia de tiempo que hubo entre los dos timestamps tomados. Este proceso se repite 50 veces con los mismos parámetros para tener una mejor noción del tiempo transcurrido para cada método en cada dataset. Cabe mencionar un detalle menor que es el hecho de que cada una de estas ejecuciones tiene incluido el tiempo de carga de los datos de los partidos.

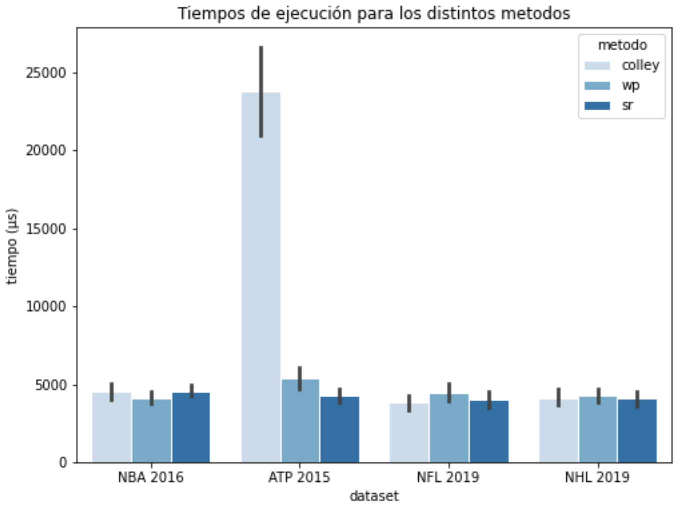


Figura 4: Comparación de tiempos de ejecución de todos los métodos

Una vez lanzado el experimento se obtienen los resultados (como el promedio de los tiempos con su barra de error) en la Figura 4, donde se puede ver que no se cumplió del todo lo que 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 para Colley en ATP se ve justificado por el tamaño del dataset ya que de entre todos éste es el mas grande.

Además surgen dos particularidades de estos resultados, lo primero es que tanto para SR como para WP en el dataset de ATP los tiempos no son superiores a los de los demás datasets, siendo que ATP tiene un dataset mucho más grande que estos. Y segundo, Colley no supero en tiempo a los demás métodos (salvo para ATP), lo cual no tiene mucho sentido dada la complejidad algorítmica del método. Por lo que es posible que el gráfico no esté representando a los tiempos reales de los métodos. Esto puede estar ocurriendo por dos razones, o bien el ruido del sistema durante la ejecución del experimento "nublo" de cierta forma a los tiempos reales, o la carga de datos de los partidos cubre los tiempos de cada uno de los métodos, lo cual no necesariamente esta mal; sin embargo, no permite responder la pregunta principal del experimento.

Para comprobar esto se hace una modificación del código principal para que devuelva el tiempo de solo la ejecución de los métodos sin contar la carga de los datos. Los resultados de esta modificación se pueden ver en la Figura 5 donde se decidió representar a los resultados como promedios de todas las ejecuciones en escala logarítmica para una mejor visualización de los datos.

Se observa que los resultados son mucho mas de la forma que se esperaba inicialmente, debido a que se deja ver la diferencia de Colley frente a los demás. Esto permite concluir que Colley si tarda mucho más que los otros dos métodos, nótese además que al igual que en la Figura 4 Colley sobrepasa por lejos a los demas en el dataset de ATP.

Por otro lado se observa que WP resulta ser un poco mas pesado que SR, siendo que son algoritmos con la misma complejidad algorítmica (lineal en la cantidad de equipos). Se atribuye esto a un ligero detalle

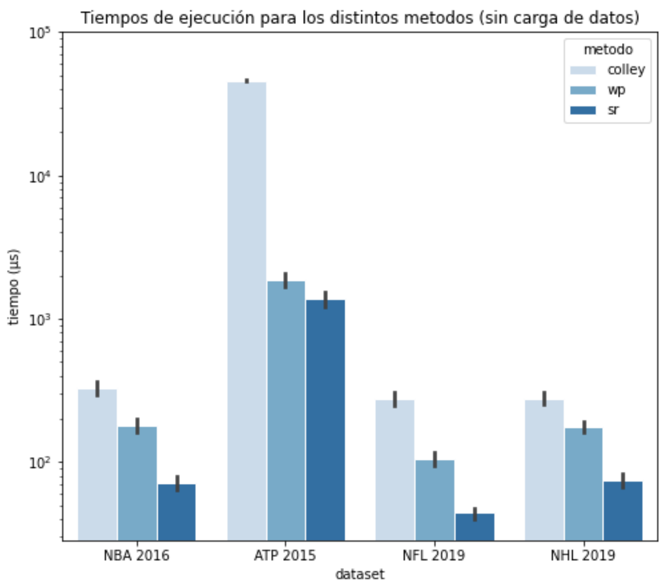


Figura 5: Comparación de los tiempos de ejecución sin la carga de datos para cada método.

en la implementación del código de WP (visto en la sección 2.3) con respecto al de SR, donde la función `partidos_totales()` de la estructura usada tiene una suma entre los partidos ganados mas los perdidos, y en el caso de SR no existe una suma así ya que obtiene los datos necesarios en cada iteración en solo dos operaciones fundamentales.

Este es un detalle de bajo nivel en la implementación dado que se podría subsanar modificando la carga de datos, pero no es necesario dado que no es el objetivo del trabajo. Por lo que fuera de ese detalle se puede concluir que WP se comporta igual a SR bajo la implementación adecuada. E incluso con la implementación presentada no resulta ser muy diferente puesto que, como ya se explico, tienen igual complejidad.

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] Arpad E. Elo. The rating of chess players, past and present. 1978.

[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.