Métodos Numéricos: Trabajo práctico 1

Matias Sandacz [*], Matias Grynberg Portnoy [**], Tomás Delgado [***], Miguel Fainstein [****]

Universidad de Buenos Aires

Facultad de Ciencias Exactas y Naturales

Departamento de Computación

8 de septiembre de 2019

Resumen

Se analiza en detalle el desempeño y comportamiento del criterio de Colley para calcular un ranking en competencias en las que no todos los equipos juegan la misma cantidad de veces entre sí. Se estudia y corrobora la estabilidad numérica de su resolución por medio de eliminación gaussiana, ofreciendo fundamentos teóricos y prácticos así como relaciones sobre los casos en los que aumenta el error. Se analizan propiedades deseables para estos criterios y se propone una medida de desempeño objetiva. Se contrasta al modelo de Colley con otros modelos posibles y se observan sus fortalezas y debilidades tanto en situaciones particulares como en el caso general. Además, se obtienen resultados generales sobre el modelado de la situación competitiva que dan pie a una construcción mejor formada de nuevos criterios. En este sentido, se proponen cuestiones a abordar para continuar en el camino hacia criterios ideales.

Keywords— ranking, competencia, Colley, eliminación gaussiana

[*] matiassandacz@gmail.com 118/18
[**] grynberg8@gmail.com 177/17
[***] tmsdlgd@gmail.com 475/18
[****] miguelon.f98@gmail.com 23/18

1. Introducción Teórica

En cualquier tipo de competencia entre equipos surge el problema de poder calificar y comparar el desempeño de cada equipo. En este trabajo se estudiarán distintos métodos para lograr esto, sus origenes, motivaciones, ventajas y desventajas.

El modelo de competencia que seguiremos tiene las siguientes características:

1. Los equipos no pueden empatar partidos, hay dos resultados posibles, ganar o perder.
2. Los equipos pueden no jugar la misma cantidad de partidos a lo largo de la competencia.
3. Puede suceder que equipos nunca se enfrenten y que otros se enfrenten múltiples veces.

El primer método que parece solucionar este problema de manera muy sencilla y que es aplicado en numerosas competencias deportivas en todo el mundo es el de porcentaje de victorias. Simplemente se calcula para cada equipo qué proporción de partidos ganó sobre todos los jugados:

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

Donde r_i es el puntaje, w_i son el total de partidos ganados, n_i el total de partidos jugados y l_i el total de partidos perdidos por el equipo i. Notamos que la igualdad se da por el hecho de que en nuestro modelo de competencia no hay empates.

Las ventajas de este método se hayan en su sencillez y -al menos a primer vista- la sensación de imparcialidad, ya que es directa la conexión entre la cantidad de partidos que ganó un equipo y su calificación en la competencia. Sin embargo, rápidamente esta relación se queda corta cuando estamos hablando de competencias donde los equipos no juegan la misma cantidad de veces ni enfrentan a los mismos equipos. Es muy fácil que un equipo que jugó muy pocos partidos tenga un ranking muy superior a uno que tuvo muy buenos resultados a lo largo de muchos partidos. Ganar un partido y no perder ninguno aquí significa una habilidad infinitamente más alta que alguien que perdió el único partido que jugó, y esta puede no ser una conclusión muy práctica en la vida real.
A partir de estas observaciones es que surge la búsqueda de modelos que describan más razonablemente el ranking de equipos en una competencia. Presentamos entonces el método de Colley ofrecido por la cátedra. La idea aquí es estimar la probabilidad de ganar el próximo partido de un equipo dado, pero considerando un estimador más suave del porcentaje de victorias: Laplace Smoothing.

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

El método de Colley agrega al estimador la consideración de la dificultad de los contrincantes para la obtención del puntaje de cada equipo. Se observa que ranking de un equipo depende del puntaje de los equipos que enfrentó, que a su vez son calculados en función de los puntajes de sus adversarios, con lo cual el problema se vuelve iterativo. Afortunadamente, es posible despejar estas ecuaciones para obtener un sistema de ecuaciones lineales, cuya solución es el ranking buscado.

No nos meteremos en los detalles de los fundamentos del método porque Colley ya lo ha hecho en su trabajo ([1]). Nos dedicaremos al análisis de su comportamiento. La definición es la siguiente: Si n_i es la cantidad de partidos jugados por el equipo i, y $n_{i,j}$ es la cantidad total de partidos jugados entre el equipo i y el equipo j y T es la cantidad total de equipos, se define la Matriz de Colley $C \in \mathbb{R}^{TxT}$, como

$$C_{i,j} = \begin{cases} -n_{i,j} & i \neq j \\ 2 + n_i & i = j \end{cases} \tag{3}$$

y se define el vector $b_i \in \mathbb{R}^T$ como:

$$b_i = 1 + \frac{w_i - l_i}{2} \tag{4}$$

El método de Colley propone hallar el vector de rankings $r \in \mathbb{R}^T$ como el vector que satisface la ecuación

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

Resolveremos el sistema con el método de eliminación gaussiana sin intercambio de pivote. Para poder utilizarlo, necesitamos en primer lugar la garantía de que la matriz es inversible, para que exista una única solución. En segundo lugar, necesitamos que el algoritmo de eliminación gaussiana sin pivoteo no se trabe en ningún paso, es decir que nunca nos encontremos con algún pivote nulo. Veremos que la matriz de Colley es estrictamente diagonal dominante, propiedad que ofrece ambas implicaciones.

Propiedad: Sea $C \in \mathbb{R}^{T \times T}$ una matriz de Colley, entonces C es estrictamente diagonal dominante.
Demostracion: Una matriz es estrictamente diagonal dominante, si y solo si $\forall i = 1 \ldots T$

$$|c_{i,i}| > \sum_{j=1, j \neq i}^{T} |c_{i,j}| \tag{6}$$

Por la definición **(2)** de la matriz de Colley podemos reemplazar y obtenemos la ecuación:

$$2 + n_i > \sum_{j=1, j \neq i}^{T} n_{i,j} \tag{7}$$

Se pueden quitar los modulos ya que $n_{i,j}, n_i \geq 0$.
Ahora bien, es evidente que si sumamos los partidos que juega el equipo i contra todos los equipos, esto es lo mismo que el total de partidos que jugó. Quedandonos la ecuación:

$$2 + n_i > n_i \tag{8}$$

$\square$

Cabe destacar que el algoritmo de eliminación gaussiana sin pivoteo no goza de buena estabilidad numérica. Puede ocurrir que durante algún paso de la eliminación gaussiana, nuestro pivote sea un valor muy pequeño en valor absoluto. Este fenómeno amplificará los errores que se van cometiendo y nos dejará con una matriz cuyo sistema de ecuaciones asociado tiene una solución que difiera significativamente de la original. Asimismo, al resolver el sistema triangulado es posible seguir amplificando el error. Analizaremos estos errores para corroborar su magnitud y trataremos de encontrar situaciones en las que el error puede ser particularmente alto.

Por último, vamos a presentar un modelo de ranking de competencias planteado por nosotros que nos ayude a analizar aún mejor las diferencias entre los distintos métodos, sus ventajas y desventajas. Para empezar, decidimos crear un sistema de puntajes, en el que ganar es premiado con una cierta cantidad de puntos y perder es castigado con alguna otra cantidad. Para definir cuál será el puntaje a asignar y de qué dependerá, nos pareció que tenía sentido en cualquier contexto competitivo, que ganarle a un buen jugador sea premiado y perder contra un mal jugador sea castigado. Asimismo, ganar contra un mal jugador no debe dar grandes cantidades de puntos y perder contra un buen jugador no debía ser fuertemente castigado. Se observa que la consideración de la habilidad del contrincante debe ser relativa a la habilidad del jugador: ganarle al primer jugador siendo el segundo no es lo mismo que ganarle al primer jugador siendo el último.

Por otro lado, decidimos que por una cuestión de normalización y para que el sentido de los puntajes se mantenga en el tiempo tenía sentido que la cantidad de puntos ganada por un equipo en un partido fuera igual a la cantidad perdida por el otro equipo. De este modo la suma de los puntajes se mantiene constante. Dado un partido entre dos jugadores, la posición del ganador r_1 y la posición del perdedor r_2, debemos definir una función $p(r_1 - r_2)$ que será la cantidad de puntos que pasarán de un equipo a otro.

Buscamos entonces una función $p : \mathbb{Z} \to \mathbb{R}_+$ tal que $p(x)$ sea siempre positivo, pues es la cantidad de puntos que se lleva el ganador. Además, debe ser chico cuando x es muy negativo (si el ganador tiene un ranking mucho mejor que el del perdedor) y debe ser muy grande si x es muy positivo.

Nuestro objetivo aquí será entender qué funciones modelan mejor la situación competitiva, al ofrecernos una noción del crecimiento del valor de una victoria con respecto a la distancia de dos jugadores en la tabla de posiciones. Por otro lado, queremos ver cómo se relaciona con el método de Colley. Intentaremos observar similitudes y diferencias y ver si el método de Colley es semejante a usar alguna función en particular con nuestro sistema.

Cabe destacar que nuestro sistema no busca estimar la probabilidad de que un equipo gane el siguiente partido como es el caso de CMM y WP. En nuestro, caso solo se busca ofrecer un orden óptimo entre los jugadores. Además, el método considera el orden en el que se jugaron los partidos dado que se computa partido a partido, mientras que CMM y WP que son completamente conmutativos en este sentido ya que los rankings se calculan sobre el resultado final.

2. Desarrollo

Se separará al trabajo en dos grandes partes: el análisis cuantitativo de la eliminación gaussiana y el análisis cualitativo del método de Colley. En la primera parte analizamos los errores que se obtienen al calcular el ranking con el método de Colley con el método de eliminación gaussiana. Realizamos experimentos para cuantificar el error para matrices cualesquiera y tratamos de encontrar casos particularmente problemáticos, además de calcular el error obtenido con los tests ofrecidos por la cátedra.
En la segunda parte estudiamos las características de cada uno de los métodos presentados. Intentamos observar el comportamiento en ciertos casos representativos y realizamos experimentos con el objetivo de obtener información sobre el comportamiento general de cada método. En particular, definimos un estimador de la capacidad predictiva de los rankings, capaz de ofrecernos resultados objetivos para la pregunta de qué método es mejor. Además realizamos diversos experimentos para estudiar lo que nosotros denominamos sensibilidad, la dificultad para un equipo de ascender en la tabla de posiciones.

Para computar los distintos métodos, principalmente la eliminación gaussiana y resolución del sistema, se implementó una clase `Matrix`. Consiste en un arreglo dinámico de `double` y provee ciertas facilidades para operaciones con matrices como la suma, las operaciones entre filas y la multiplicación. Con estas operaciones implementadas fue sencillo escribir el algoritmo. Luego la pregunta fue qué tan estable es y si esta estabilidad es razonable para utilizar con confianza el ranking obtenido.

2.1. Casos particulares y el módulo de los pivotes

El primer experimento, fue la comparación entre nuestros rankings y los resultados provistos por la cátedra. Además, analizamos detalladamente el algoritmo de eliminación gaussiana para buscar competencias en las que el error pudiera ser más alto. En ambos casos tomamos como medida del error el máximo de la diferencia coordenada a coordenada. Formalmente, si r_1 es nuestro ranking y r_2 es el de la cátedra, calculamos el error como $\epsilon = \|r_1 - r_2\|$. De este modo nos aseguramos de no perder ningún error importante por promediar.

Para encontrar casos particularmente problemáticos, partimos de la base de que los errores numéricos en la eliminación gaussiana surgen mayoritariamente cuando los pivotes son chicos, puesto que estos son los dividendos en cada paso del algoritmo (amplificación de error) y a su vez el pivote es calculado con una resta (eliminación catastrófica). Al analizar en papel las cuentas que se realizan en la EG para matrices de colley de 3x3, nos encontramos con que era difícil hacer que el segundo pivote tuviera módulo chico. Más precisamente, si tenemos 3 equipos tales que el primero jugó n_1 veces con el segundo y n_2 veces con el tercero, y el segundo y el tercero jugaron n_3 veces entre sí. Quedándonos una matriz de Colley de la forma:

$$\begin{pmatrix} 2+n_1+n_2 & -n_1 & -n_2 \\ -n_1 & 2+n_1+n_3 & -n_3 \\ -n_2 & -n_3 & 2+n_2+n_3 \end{pmatrix} \tag{9}$$

Al realizar el primer paso de eliminación gaussiana se obtiene que el segundo pivote es:

$$2 + n_1 + n_3 - \frac{n_1^2}{2 + n_1 + n_2} \tag{10}$$

Si igualamos la ecuación **(10)** con algún $\varepsilon \in \mathbb{R}$ y reordenamos un poco los términos llegamos a la ecuación **(11)**.

$$\varepsilon = \frac{4 + 2n_1 + 2n_1 + 2n_2 + 2n_3 + n_1n_2 + n_1n_3 + n_2n_3}{2 + n_1 + n_2} \tag{11}$$

Dado que $n_1, n_2, n_3 \geq 0$ se tiene que:

$$\varepsilon \geq \frac{4 + 2n_1 + 2n_2}{2 + n_1 + n_2} = 2 \tag{12}$$

Esto nos da la pauta de que los pivotes probablemente tengan cotas inferiores mayores que algún número natural. Lo cual disminuiría considerablemente las probabilidades de sufrir de errores numéricos por amplificación de error.

Realizamos entonces un experimento con el objetivo de ver si esta hipótesis se corrobora. La idea del experimento es generar competencias de forma aleatoria y observar el valor promedio de sus pivotes. Si los pivotes se mantuvieran altos, esto sería un fundamento para confirmar que se tiene estabilidad en los cálculos.

Una vez establecida la cantidad de jugadores y cantidad de partidos, tomamos números al azar equiprobablemente para decidir quiénes jugarían contra quiénes. Además, definimos la habilidad de cada jugador con un número tomado de una distribución uniforme entre 0 y 1. Luego, dados dos jugadores con habilidades h_1 y h_2 la probabilidad de que el primero le ganara al segundo la calculamos como $h_1/(h_1+h_2)$. Así nos aseguramos de tener una varianza razonablemente alta como para abarcar, al menos, las competencias típicas.

2.2. Medida alternativa y matrices ralas y densas

Por otro lado, surgió una metodología alternativa para medir el error que no compara los resultados con los de otra librería, que podría estar cometiendo los mismos errores que nuestra implementación. Decidimos estimar el error obtenido a partir de la cota dada por el número de condición:

Si $TeslacantidaddeequiposC, eslamatrizdeColley, ybeselvectorindependientedelmtododeColley$

$$\frac{\|x - \tilde{x}\|}{\|x\|} \leq \|C\| \|C^{-1}\| \frac{\|b - \tilde{b}\|}{\|b\|} \tag{13}$$

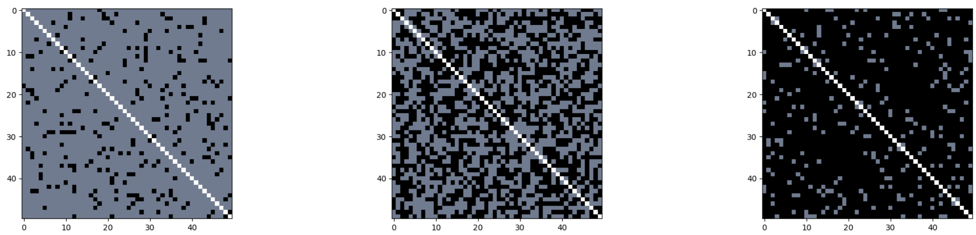
Para Colley si se utiliza $\|.\|_1$, se tiene que $\|x\| = T/2$, por ser la suma de los rankings de todos los equipos y $T + partidosTotales \geq \|b\| \geq T$

$$\frac{\|x - \tilde{x}\|}{T} \leq \|C\| \|C^{-1}\| \frac{\|b - \tilde{b}\|}{T} * 2 \tag{14}$$

Esta cota para el error tiene dos problemas a notar. En primer lugar, a priori no conocemos el número de condición de la matriz (producto entre la norma de A y la inversa de A), si el número de condición llegase a ser demasiado alto la cota no nos permitiría afirmar nada. Ahora incluso si A estuviese bien condicionada, es importante mencionar que para calcular $\tilde{b}$, es necesario realizar el producto $A\tilde{x}$. No tenemos forma de acotar el error numérico que se obtiene al realizar este producto. Nos pareció razonable sin embargo despreciar esa incerteza porque el producto entre la matriz y el vector es una suma de productos con magnitudes normalmente no muy alejadas entre sí. Mantuvimos la precaución de no realizar la medición en casos en los que este producto pueda generar error.

Con esta medida, realizamos un experimento para estudiar el comportamiento del error en competencias ralas y densas. Naturalmente, uno consideraría que al aumentar la cantidad de componentes no nulos, debería generarse un mayor error numérico. Para experimentar con esto, se implemetó un segundo generador de competencias. Este toma un n y una $densidad$ y realiza la fracción de los partidos que corresponda. Es decir realiza una parte de todos los posibles partidos. La elección de qué partidos se realiza y el ganador de este enfrentamiento son elegidos aleatoriamente. En la figura 1 se observan algunas matrices de Colley generadas con 50 equipos. De esta forma se pretende comparar el error absoluto promedio $||Cr' - b||/n$ a medida que se alteran estas variables. Una de las falencias de este modelo es que, al elegir el ganador de cada encuentro aleatoriamente, es que b sera de cierta forma particular que quizás no sea representativa del caso general.

Necesitamos, además, cierta garantía de que al estar cerca con b', se esta cerca con r'. Para ello realizamos lo siguiente. Utilizando el generador de competencias, hallamos el número de condición 1 utilizando `numpy`. De esta manera se observa cómo cambia el número de condición con la densidad y el número de equipos. Como el número de condición varía de matriz en matriz aún teniendo el mismo n y *densidad*, se toma el máximo observado de 100 iteraciones. Así se obtiene una estimación del peor caso de k_1. Si el k_1 observado es pequeño, entonces el estimar el error en b resulta una buena medida del error en r. Ya que el error será a lo sumo $2k_1$ veces mayor.



(a) Densidad de 0.1 (b) Densidad de 0.5 (c) Densidad de 0.9

Figura 1: Matrices de Colley creadas con el generador usando 50 equipos. Blanco significa positivo. Negro negativo. Azul es 0

2.3. Capacidad de predicción. Una forma objetiva de calificar a los criterios

Dado cualquier método de calificación de competidores, el sentido común sobre el comportamiento esperado es que un equipo que tiene un mejor ranking que otro debe tener mayores probabilidades de ganarle cuando se enfrentan. En otras palabras, la probabilidad real de que un equipo le gane a otro debe ser mayor a 0,5 si y solo si el equipo tiene mejor ranking que su contrincante.

Para estudiar esto se diseñó el siguiente experimento: dado un ranking establecido por el criterio a partir de una serie de partidos, al jugarse un nuevo partido entre dos equipos se observa si el ranking del equipo ganador era mayor o menor que el del equipo perdedor. Si así fuese se considera que el método de ranking acertó, y de lo contrario falló. Llevando a cabo este mismo experimento para muchos nuevos partidos, se obtiene una estimación de la probabilidad de acierto del criterio.

De este modo compararemos el desempeño objetivo de CMM y WP, así como el del método planteado por nosotros. Para este último nos interesa además ver qué función interna de asignación de puntos produce mejores resultados con respecto al experimento en cuestión. Se profundizará acerca de las funciones en la sección **(2.4)**.

La idea es realizar el experimento para dos conjuntos de datos grandes. Para conjuntos de datos chicos, el test podría perder sentido por falta de información. Utilizamos en primer lugar los datos sobre los partidos de tenis del ranking de la ATP, y luego los datos correspondientes a la temporada de la primera división de fútbol de Inglaterra, la Premier League.

Hay una diferencia esencial entre estos dos conjuntos de datos. En la Premier League todos los equipos juegan la misma cantidad de partidos entre sí mientras que en el ranking ATP los jugadores tienen participaciones muy variadas y están muy lejos de haber jugado todos contra todos. En particular, la densidad de la matriz de Colley en el conjunto de datos de ATP es de $0{,}03$. Un detalle interesante de que en la Premier League todos jueguen exactamente una vez contra cada uno de los demás, es que en este caso particular WP y Colley ofrecen el mismo ranking (no exactamente los mismos números pero sí el mismo orden). Esto sucede porque la suma de todos los rankings en Colley es constante y la forma de considerar la dificultad de los contrincantes en CMM es a partir de la suma de sus rankings. Como los contrincantes son todos los jugadores, el ranking deja de depender de los valores particulares de cada uno. Más aún, se vuelve equivalente al de WP, con una pequeña suavización extra. Partiendo de la ecuación 16 de [1] se tiene para este caso particular:

$$(r_i + n_{tot,i})r_i - \sum_{j=1}^{n_{tot,i}} r_j^i = 1 + \frac{n_{w,i} - n_{l,i}}{2} \tag{15}$$

$$(r_i + n_{tot,i})r_i - (\frac{n_{tot,i} + 1}{2} + r_i) = 1 + \frac{n_{w,i} - n_{l,i}}{2} \tag{16}$$

$$r_i(3 + n_{tot,i}) = \frac{3 + n_{w,i} - n_{l,i} + n_{tot,i}}{2} \tag{17}$$

$$r_i = \frac{\frac{3}{2} + n_{w,i}}{3 + n_{tot,i}} \tag{18}$$

Donde $n_{w,i}$, $n_{l,i}$ y $n_{tot,i}$ son la cantidad de victorias, derrotas y partidos totales del i-ésimo equipo y r_i y r_j^i son los rankings del i-ésimo equipo y del j-ésimo contra el que jugó el i-ésimo, respectivamente.

En este sentido, esperamos que tanto CMM como WP ofrezcan la misma capacidad predictiva en el caso de la Premier League. Sin embargo, esperamos observar la diferencia en el caso de ATP, cuando la competencia no sea en formato todos contra todos. Nos gustaría corroborar así la mejora obtenida al considerar la habilidad de los competidores. También tenemos la duda de cómo va a ser el resultado de nuestro método con respecto a WP, mientras que no creemos que obtenga un mejor resultado que el método de Colley.

2.4. Función de asignación de puntos del nuevo método planteado

Para nuestro método, en vistas de lo planteado en la introducción teórica sobre la forma de la función utilizada para la asignación de puntajes, surgieron dos funciones que nos parecieron apropiadas, la exponencial y la sigmoide. Definidas como **(19)** y **(20)**, respectivamente. Ambas cumplen que dan valores chicos para entradas negativas, valores grandes para entradas positivas y valores intermedios para el 0.

$$exp(x) = e^{\lambda x} \tag{19}$$

$$\sigma(x) = \frac{1}{1 + e^{-\frac{x}{\lambda}}} \tag{20}$$

La gran diferencia es que la sigmoide está acotada entre 0 y 1 mientras que la exponencial no deja de crecer (y además crece rápido). En este sentido, los modelos que se contrastan son uno en el que el valor de ganarle a un jugador mejor sigue creciendo sin importar cuánta sea la diferencia y uno en el que ganarle a un jugador bastante mejor y a uno que es todavía muchísimo mejor que ese es más o menos equivalente en términos de puntaje.

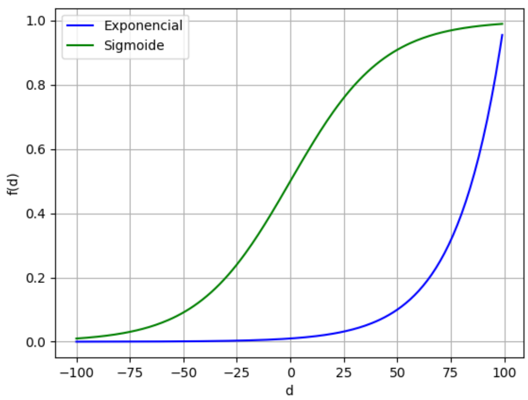


Figura 2: Gráficos función exponencial y sigmoide para una competencia de 100 equipos. Con $\lambda = \frac{ln(100)}{100}$. La exponencial se muestra normalizada con un factor 1/100 para mejorar la escala.

También surge la pregunta si esto depende de las características del juego en el que se compita, tal vez en juegos con mayor dependencia del azar siempre haya una probabilidad más o menos constante de ganarle a cualquier jugador. Por ejemplo, en juegos como el Poker una distribución particular de las cartas podría hacer que un principiante le gane a un experto, mientras que en juegos como el ajedrez tiene sentido pensar que para ganarle a otro hay que ser efectivamente superior a él en algún aspecto.

Un detalle no menor que habrá que definir es el parámetro λ que utilizaremos para la función de nuestro modelo. Lo único que está claro es que debe depender de la cantidad de equipos que haya en la competencia: no es lo mismo tener una diferencia de 5 posiciones en una tabla de 3000 competidores que en una competencia con 10 jugadores. Más allá de eso, no es evidente la función con la que debería estar definido.
Tomaremos como primera aproximación a $\lambda = \frac{\ln(T)}{T}$ porque sus gráficos parecieron adecuados.

2.5. Sensibilidad. Capacidad de los equipos de subir su posición

Otra propiedad interesante de los distintos métodos es la facilidad o dificultad con la que un equipo cambia de posición. Llamamos a este factor *sensibilidad*.

Para poder medir esto realizamos el siguiente experimento: tomando algún conjunto de datos, vemos cuántas veces sucesivas debe vencer el equipo que se encuentra en la última posición al equipo en la primera posición (en el momento de jugar el partido) para llegar al primer puesto. Realizamos esto a lo largo de toda una competencia (partido a partido). Es decir, para cada competencia intermedia en la que solo se habían jugado los primeros n partidos. De este modo, buscamos observar el crecimiento de la sensibilidad con respecto a la cantidad de partidos, para cada uno de los métodos planteados.

Originalmente se había planteado otro experimento para poder medir la sensibilidad de los métodos que luego fue descartado debido a que los resultados producidos no tenían significados evidentes y eran difíciles de analizar. Este consistía en, dada una competencia y un método de ranking calcular cuánto avanzaría o retrocedería posiciones en promedio, el primer equipo, el equipo del medio y el último equipo de la competencia si ganasen el próximo partido contra cada uno de los otros equipos.

Por ejemplo, si se tiene una competencia entre tres equipos A, B y C, en ese orden según cierto método de rankeo; se calculan cuántas posiciones se mueve A si le ganase el próximo partido a B, cuántas si le ganase a C y se toma el promedio entre estas dos. Lo mismo se hace con B y con C.

Los resultados producidos por este último experimento no los mostraremos en resultados porque resultaron demasiado ruidosos y a la larga perdía el sentido el agregado de un único partido. Sí mostraremos los del primer experimento descrito, que dieron una comparación entre los métodos mucho más tangible.

Nuestra hipótesis era que Colley mostraría mayor sensibilidad que WP, donde mayor sensibilidad refiere a que el último equipo debe ganarle menos partidos al primer equipo que en el otro método para quedar primero en la tabla de posiciones. Basamos esta creencia en el hecho de que el método de Colley utiliza los rankings del resto de los equipos para determinar el de un equipo, y está hecho de tal forma para que cumpla que los mejores equipos cada vez tengan más difícil ampliar la ventaja sobre el resto, y que a los peores se les haga más sencillo acortar dicha distancia.

Con respecto a nuestro método, creíamos que la sensibilidad sería la mayor de los tres métodos. Esto es porque justamente fue divisado para castigar aquellos equipos que pierden con equipos que tienen menor ranking y premiar a los que logran lo contrario. También resultaba interesante utilizar el experimento para comparar las funciones internas de otorgación de puntaje de nuestro método. La hipótesis respecto a este último punto es que la función exponencial mostraría mayor sensibilidad ya que no está acotada a lo largo del eje x mientras que la sigmoide sí, esto significa que a partir de cierto x otorga prácticamente la misma cantidad de puntos (bajo cierta precisión numérica) y la otra no.

2.6. Estrategia de equipos para subir su posicion

Suponiendo que un equipo ganará los próximos partidos. ¿Cómo debe jugar para mejorar su posición? Proponemos que jugar contra aquél que se encuentra en la mayor posición es una buena estrategia si se utiliza el ranking de Colley. Para obtener evidencia de ello, tomamos la *Premier League* y el ranking *ATP*, y se cuentan cuantas veces debe vencer al equipo que se encuentra en la posición k. Por ejemplo, si el equipo tomado se halla inicialmente en la posición 4 ¿cuántas veces debe vencer al equipo que se halle en la posición 2 para obtener el primer puesto? Notar que el rival del equipo puede variar ya que al agregar partidos se altera el ranking y el que se encuentra en la posición 2 puede no ser el mismo. Para evitar complicaciones, si en algún momento el equipo elegido tiene la posición con la que debería enfrentarse, se enfrenta a la anterior (como ejemplo si mi equipo se encuentra en la posición 2 y se debería enfrentar al que se encuentra en la posición 2, se enfrentara al de la posición 1).

Uno de los problemas que surgió para este experimento es que el tiempo computacional crecía desproporcionadamente al usar rivales alejados de los primeros puestos. Esto sucedía aún cuando el equipo elegido estuviera inicialmente cercano al primer puesto. Por estas razones se limitó a estudiar cuando el rival se encuentra en las primeras 5 posiciones. Esto es un indicio de que enfrentarse a rivales alejados de los primeros puestos no es una buena estrategia.

Uno de los planteamientos que surgieron en la implementación del experimento es, eligiendo selectivamente la secuencia de rivales ¿Existe alguna secuencia en la que aún ganando estos partidos, no se pueda obtener el primer puesto? No exploramos esto.

2.7. Competencias problematicas en Colley. Un Ferrari entre Fiats

El método de Colley contiene ciertas asunciones implícitas sobre la conectividad de los partidos y la habilidad de los distintos equipos. Para evidenciar estos problemas, se generó la siguiente competencia:

Se usan 100 equipos divididos en 2 grupos. Grupo A y grupo B. Cada equipo se enfrenta con todos los miembros de su grupo. Además, cada equipo tiene una constante *h* que determina su habilidad. Para decidir el ganador de un enfrentamiento, cada equipo tiene una variable aleatoria gaussiana con media *h* y varianza 1. El equipo con mayor puntaje es elegido como el ganador del encuentro.
Todos los miembros del grupo B tienen $h = 10$, mientras que en el grupo A todos tienen $h = 1$ salvo un equipo *infiltrado* que tiene $h = 10$. Este equipo especial tiene una enorme ventaja, ya que se enfrentara a rivales con mucha menor habilidad.

El resultado que correspondería al "verdadero" ranking debería colocar a los mejores miembros del grupo B por encima del grupo A con excepción del equipo infiltrado. Suponemos que Colley colocará al equipo infiltrado en la primera posición y luego tendrá un ranking mucho más simétrico entre los grupos A y B.

2.8. El metodo de Colley no es justo

Dado el desarrollo de una competencia. ¿Cómo afecta el resultado de un partido a los equipos no involucrados en este? Nuestra hipótesis es que el algoritmo de Colley no es justo, en el sentido de que un partido ajeno al equipo *i* puede afectar su ranking. Para ello, proponemos y analizamos un ejemplo simple en donde esto sucede.

Al mismo tiempo, observamos que claramente esto no sucede en ninguno de los otros criterios. En WP el puntaje de un equipo solo depende de los partidos que jugó. Similarmente, en nuestro método al jugarse un partido el intercambio de puntos es únicamente entre los que jugaron.

3. Resultados y discusión

3.1. Casos particulares y el módulo de los pivotes

En primer lugar, el contraste con los tests de la cátedra dió resultados sumamente favorables como puede observarse en 1.

Con estos resultados en mente y ante la aparente imposibilidad de generar un caso de test en el que el pivote se hiciera chico, nos dispusimos a estudiar el tamaño de los pivotes. Los resultados del experimento fueron los de la figura. 3.

Observamos que el pivote suele mantener su módulo a lo largo de la triangulación, al menos en los casos considerados. Solo se manifiesta en algunos casos una reducción en el módulo del pivote sobre la última columna. No entendemos la razón de esto pero esa reducción no lo deja en valores menores a 1 y no es de más de un orden de magnitud.

Con este resultado, se explican los bajos errores obtenidos con respecto a los test de la cátedra y queda fundamentada empíricamente la estabilidad de la eliminación gaussiana para matrices de Colley. Queda pendiente la generalización del análisis realizado en la introducción para concluir que en ningún caso pueden obtenerse pivotes chicos.

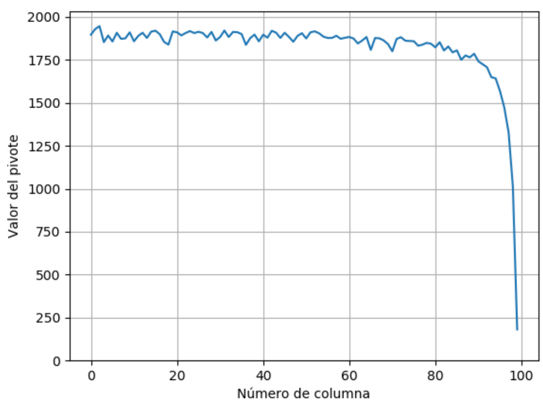
Comparación Test1	
Solución real	Solución obtenida
0.40412232279778	0.40412234042553
0.3988031744957	0.39880319148936
0.77420210838318	0.77420212765957
0.43816488981247	0.43816489361702
0.3507978618145	0.35079787234042
0.63390958309174	0.63390957446808
Error absoluto medio	1.2806e-08

Comparación Test2	
Solución real	Solución obtenida
0.35970667004585	0.35970667695870
0.61597841978073	0.61597838672327
0.66866075992584	0.66866074874565
0.31493631005287	0.31493631802392
0.50154381990433	0.50154380548051
0.53917407989502	0.53917406406792
Error absoluto medio	1.4895e-08

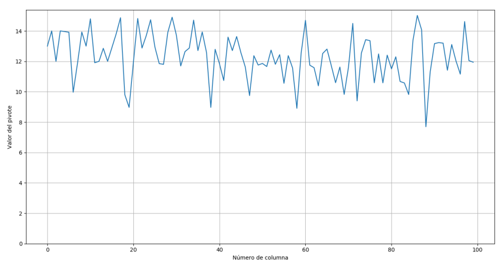
Comparación Test-prob-1	
Solución real	Solución obtenida
0.41700652241707	0.41700652528548
0.42978522181511	0.42978520935290
0.71322727203369	0.71322729744426
0.43590265512466	0.43590266449157
0.35923057794571	0.35923056008700
0.64484775066376	0.64484774333877
Error absoluto medio	1.2548e-08

Comparación Test-prob-2	
Solución real	Solución obtenida
0.41778266429901	0.41778265642151
0.41745334863663	0.41745334796926
0.71383094787598	0.71383095499451
0.43754115700722	0.43754116355653
0.4395170211792	0.43951701427003
0.57387489080429	0.57387486278814
Error absoluto medio	9.5230e-09

Cuadro 1: Contraste entre los vectores solución obtenidos y los de la cátedra, con su respectivo error absoluto medio.



(a) 50 competencias con 100 equipos y 100000 partidos.



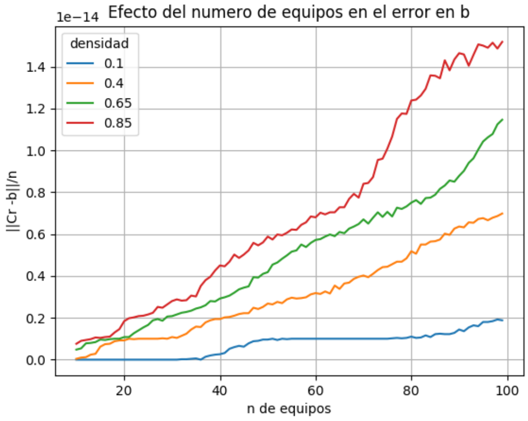
(b) 50 competencias con 100 equipos y 50 partidos.

Figura 3: Valores mínimos del pivote de cada columna sobre distintos grupos de competencias generados al azar.

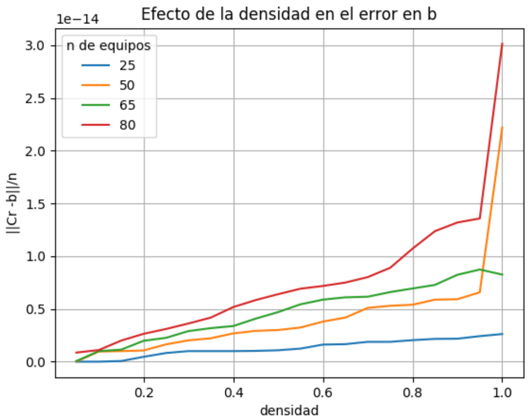
3.2. Medida alternativa y matrices ralas y densas

En la figura 4 se observa el error en b para distintos numero de equipos. Notar que para densidades bajas. Es decir, para matrices ralas el error parece mantenerse pequeño al aumentar el n. Como se había propuesto, la evidencia parece mostrar que al aumentar el número de equipos, aumenta el error numérico generado. Una explicación para esto es si se incrementa n, debe producirse un mayor número de operaciones (tanto para la triangulación como para el producto matricial).

Sin embargo, para densidades bajas, el aumentar el número de equipos muestra un efecto amortiguado. De esta manera, el proceso de eliminación gaussiana otorga un Cr' cercano al real para matrices ralas. De todos modos, para poder afirmar que el r' mantiene una distancia similar al ranking verdadero r es necesario tener en cuenta el número de condición $k_1(C)$.



(a) Para distinto número de equipos



(b) Para distintas densidades

Figura 4: Cambio en el error de b alterando el número de equipos y la densidad

La figura 5 muestra el campo para el número de condición 1 máximo observado para las distintas densidades. Notar que todos los valores son menores a 100. Podemos afirmar entonces que este se mantiene "bajo". Esto nos da confianza sobre los resultados obtenidos en el experimento anterior. Al estar bien condicionada la matriz, la diferencia entre los errores se mantiene baja por la inecuación 2.2 y por lo tanto es razonable pensar que los cambios observados en el error calculado a través del b, se condice con cambios en el error en r. Además, se observa un resultado similar al obtenido en el cáculo de error, la matriz empeora su condicionamiento para mayores densidades y para mayores cantidades de equipos, por lo que el cuidado debería ser mayor en estos casos.

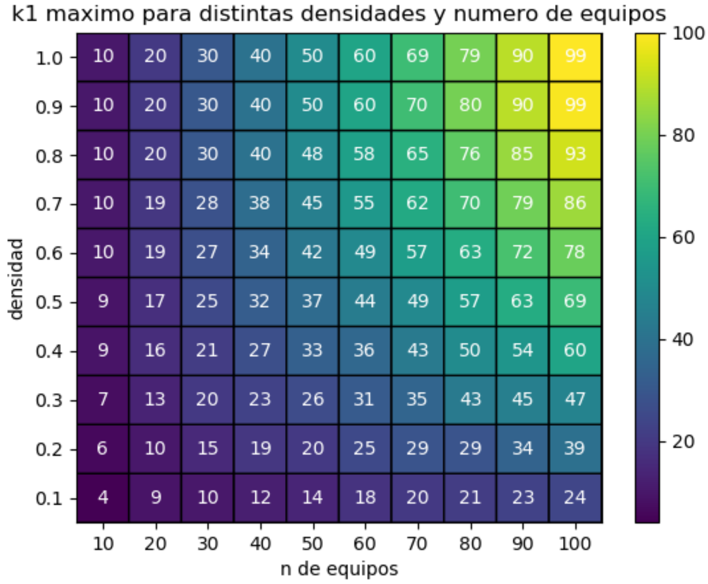


Figura 5: El máximo k_1 observado para distinto número de equipos y densidad

3.3. Resultados en la medida de capacidad predictiva

Los resultados del experimento se presentan en la tabla 6. En primer lugar, se corrobora lo esperado, CMM demuestra tener un mejor desempeño que WP, al considerar la dificultad de los contrincantes en el ranking ATP. Además, ambos métodos obtienen el mismo resultado en el caso de la Premier League, como se predijo en el análisis anterior.

Por otro lado, en nuestro método la función exponencial no obtiene un gran desempeño mientras que la sigmoide se acerca un poco más al desempeño de WP. Creemos que el desempeño de la exponencial no es bueno porque es demasiado sensible a partidos no esperados y esto lo vuelve inestable. Además, esa reacción tan alta a partidos no esperados la deja con una reacción proporcionalmente muy baja a los resultados usuales. Es posible además que por ese mismo motivo el ranking final dependa demasiado de los últimos partidos. Esto se verá con más detalle en la medición de sensibilidad.

El desempeño de la sigmoide no es malo pero posiblemente sufra de lo mismo que la exponencial, demasiada dependencia de lo que sucede en los últimos partidos. Destacamos a modo de control que en todos los casos el resultado fue mayor a 0.5 como era razonable esperar. Todos estos métodos tienen una opinión más acertada de la realidad que alguien que tira una moneda cada vez que le preguntan quién va a ganar.

Una observación que hacemos es que la medida utilizada no termina de ser del todo correcta para sistemas como el nuestro que dependen del orden de los partidos, porque se está usando la información que el método ofrece si el partido se jugara luego de que se hayan jugado todos los otros partidos, para un partido que se jugó en un momento intermedio.

Modelo	ATP	Premier League
CMM	0.6401	0.6378
WP	0.6259	0.6378
NMEXP	0.5782	0.5354
NMSIG	0.6069	0.5643

Cuadro 2: Porcentaje de acierto al eliminar la información sobre cada partido del conjunto de datos. NMSIG y NMEXP refieren a nuestro método con la función sigmoide y exponencial, respectivamente.

3.4. Experimento de sensibilidad del método

En la figura 6 se observan los resultados del test de sensibilidad, realizado para los métodos Colley, WP y el presentado por nosotros con sus dos variantes (utilizando la función exponencial y la función sigmoide). Los datos utilizados fueron los de la temporada de la Premier League año 2017/2018, los mismos del experimento de capacidad predictiva.

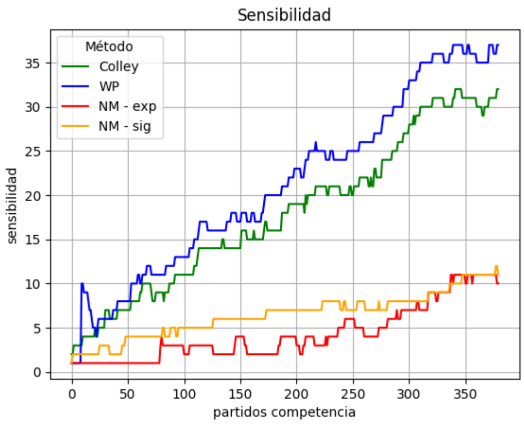


Figura 6: Cantidad de veces que le debe ganar el último al primero para obtener el primer lugar en función de la cantidad de partidos de la competencia.

Se puede ver cómo WP es el método que menor sensibilidad presenta, es decir que el peor equipo debe ganar muchísimos partidos para pasar al primer puesto. Colley en cambio está en un lugar un poco más razonable, entendemos que esto se debe a que ganarle al mejor equipo tiene relevancia mientras que en WP es lo mismo que ganarle a cualquier otro, ayuda únicamente en que baja el puntaje de este último también pero esto a medida que avanza el torneo se vuelve cada vez más despreciable.

Se observa cómo en los tres métodos el paso de los partidos agranda las diferencias entre equipos y se reduce la sensibilidad en el sentido de que cada vez es más difícil recuperar una buena posición. Con el tiempo las posiciones de algún modo se establecen y pierden flexibilidad.

En relación con nuestro método, es clara la diferencia que presenta con respecto a CMM y WP, la sensibilidad se mantiene muy alta siempre. Incluso después de los 300 partidos el último puede llegar al primer puesto ganando alrededor de 10 partidos.

Nos parece que en ciertos contextos esta última podría ser una propiedad deseable. Observamos que hay una diferencia casi filosófica en estas distintas sensibilidades. En el caso de CMM y WP el que termina en el primer puesto es el que demostró una mejor habilidad a lo largo de la competencia, mientras que en nuestro método, es el que parece ser el mejor en el momento en el que hay que dar los premios.

Usualmente, tiene más sentido la primera posibilidad porque las competencias son cortas y no suelen apreciarse cambios tan significativos en la habilidad de los equipos a lo largo de los partidos. En una competencia que dura un año, que lo único relevante en el resultado sea el último mes sería poco razonable y hasta malo desde un punto de vista organizativo.

En cuanto a la diferencia entre la versión exponencial y la sigmoide de nuestro método, pareciera que la exponencial está acelerando su pérdida de sensibilidad mientras que la sigmoide crece lentamente y de forma constante. Además de tener una curva más continua que la exponencial, la cual pega saltos y caídas repentinas.

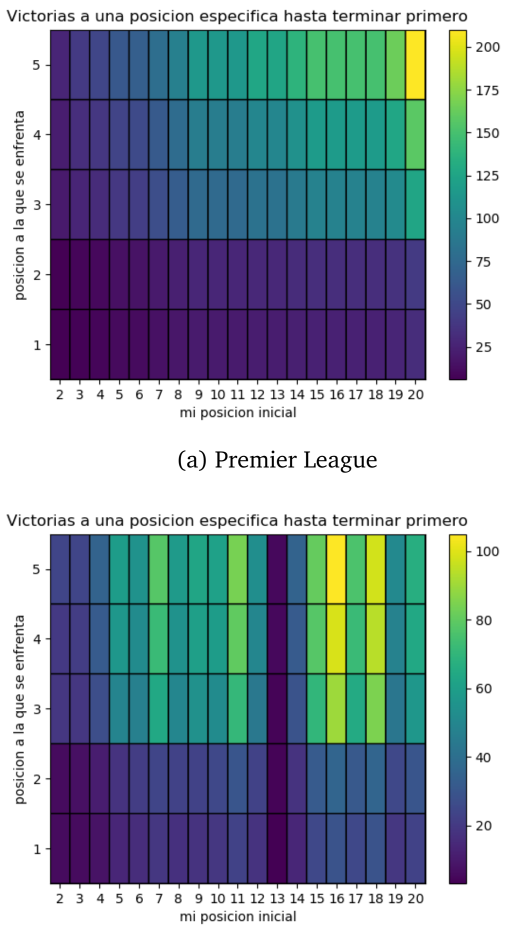
3.5. Estrategia de equipos para subir su posicion

En la figura 7 se muestran los resultados obtenidos. Notar que los datos apoyan la idea de que enfrentarse al mejor posicionado es una buena estrategia. Para todas las posiciones se observa que, a medida que se enfrentan a rivales en puestos mas bajos, aumenta el numero de victorias necesarias. Esto parece valer tanto para la *Premier League* como para la competencia *ATP*, aun siendo estas competencias muy diferentes.

Aun así, esta heurística está lejos de ser completa ya que quizás una estrategia más compleja que intercale rivales tenga un mejor desempeño. Puede ser incluso que perder selectivamente contra algunos rivales sea parte de una estrategia que obtenga el primer puesto más rápidamente.

Además, los resultados para *ATP* muestran que, enfrentándose al que se encuentra en la primer posición, el equipo que está inicialmente en la posición 13 puede obtener el primer puesto en una menor cantidad de pasos que el que se encuentra en el segundo puesto. Esto es anti-intuitivo y deja abierta la pregunta de si puede este último obtener el primer puesto más rápido usando una estrategia diferente. Más concretamente, si llamamos $\lambda(x)$ la menor cantidad de pasos necesaria para que el que se encuentra en la posición x alcance el primer puesto. ¿Vale que $x < y \longrightarrow \lambda(x) < \lambda(y)$? ¿Bajo qué competencias es esto valido?

Habiendo planteado estos inconvenientes, parece que enfrentarse a los primeros puestos parece una buena estrategia independientemente del puesto en que se encuentre.



(a) Premier League

(b) ATP. Se muestran las primeras 20 posiciones.

Figura 7: Victorias necesarias para obtener el primer puesto en base a la posicion a la que se enfrentara.

3.6. Competencias problemáticas en Colley. Un Ferrari entre Fiats

En el cuadro 3 se muestran las primeras 5 posiciones del ranking. El infiltrado ganó todos sus enfrentamientos, obteniendo el primer puesto. Sin embargo, este no es especial en comparación con otros miembros del grupo B, por lo que la obtención del primer puesto es debido no a su habilidad sino a pertenecer al grupo A, donde tiene ventaja comparativa. Más aún, los miembros del grupo A obtienen posiciones tan altas como el grupo B. Al enfrentarse ambos a rivales similares, sus resultados serán parecidos y el algoritmo de Colley no los distingue. Intuitivamente, este no es un resultado adecuado. Cualquier miembro del grupo B puede vencer a los miembros del grupo A (salvo el infiltrado), deberían encontrarse por encima de estos. Una de las falencias de Colley es, entonces, ante la falta de conectividad, no puede establecer un ranking adecuado. Se necesitaría alguna clase de información previa que indique algo sobre la habilidad de los distintos equipos/grupos. Es decir, en circunstancias en donde los rivales de un equipo no son una buena muestra de la habilidad general, el método falla. Tomar solo el desempeño observado no es un método efectivo en estos casos.

Infiltrado	0.971153846153850
Grupo B	0.68269230769231
Grupo A	0.66346153846154
Grupo A	0.64423076923077
Grupo B	0.64423076923077

Cuadro 3: Top 5 en la competencia problematica usando Colley

3.7. El método de Colley "no es justo".

Para el experimento, se generó el ranking correspondiente a la competencia ATP 2018. Cada tenista del torneo fue asignado un número natural que lo identifique. Presentamos los rankings generados por Colley al inicio del experimento:

Posicion	Equipo	Ranking
1	34	1.165721
2	51	1.076881
3	256	1.058960
4	98	1.032907
5	33	0.976987
6	31	0.945129
7	359	0.935500
8	145	0.930474
9	100	0.924869
10	415	0.898087

Cuadro 4: Se lista la tabla de posiciones de los 20 mejores tenistas del ranking ATP 2018 y los puntajes que les fueron asignados.

Luego, de generar dicho ranking, se jugó el siguiente partido:

Tenista	Tenista	Ganador
31	22	22

Cuadro 5: Resultado del partido entre los Tenistas 31 y 22

A continuación, presentamos la nueva tabla de rankings:

Posicion	Tenista	Ranking
1	34	1.164106
2	51	1.074755
3	256	1.056986
4	98	1.031149
5	33	0.975388
6	359	0.933506
7	31	0.932064
8	145	0.928163
9	100	0.923289
10	415	0.896110

Cuadro 6: Se lista la tabla de posiciones de los 10 mejores tenistas del ranking ATP 2018 y los puntajes que les fueron asignados luego del nuevo partido.

Se puede observar que los puntajes de todos los tenistas han sido modificados. No solo eso, sino que también hubo un cambio de posición en la tabla para el tenista 359 a pesar de no estar involucrado en el último partido que agregamos. Esto sucede pues como ya hemos mencionado, la suma de todos los rankings en Colley es constante y la forma de considerar la dificultad de los contrincantes en CMM es a partir de la suma de sus rankings. Esto causa que si un competidor sube de posición, todos los que jugaron contra él, inevitablemente verán sus propios rankings afectados.

4. Conclusiones

El trabajo realizado permitió corroborar de forma efectiva varias conjeturas planteadas y al mismo tiempo generó algunas nuevas junto a ciertos desafíos que quedaron pendientes. Asimismo, se obtuvieron algunos resultados tal vez contrarios a lo esperados, pero que también permitieron obtener conclusiones claras.

Logramos confirmar que el método de eliminación gaussiana aplicado a matrices de Colley presenta una estabilidad suficiente para ser utilizada para la generación de rankings deportivos. Además, se obtuvo un principio de fundamento teórico y una generalización práctica que ofrece una explicación para dicha estabilidad. En este aspecto, quedó pendiente la demostración de la generalización de la propiedad observada. Es decir, acotar inferiormente al módulo de los pivotes en la eliminación gaussiana con matrices de Colley cualesquiera.

Se corroboró también la hipótesis sobre el crecimiento del error, efectivamente el error aumenta para matrices más densas por el aumento en la cantidad de operaciones efectivas. También se confirmó que en los casos conocidos ofrecidos por la cátedra el error se mantiene suficientemente bajo.

Por otro lado, corroboramos que el método de Colley ofrece una mejora objetiva con respecto al cálculo del porcentaje de victorias al agregar la consideración de las habilidades de los equipos contra los que uno compite. Se demostró que es una buena generalización del porcentaje de victorias, en el sentido de que para el caso en el que todos juegan contra todos son métodos equivalentes.

También observamos las diferencias en el aumento de la dificultad de ascender en la tabla a medida que se avanza en la cantidad de partidos, para cada uno de los métodos. Concluimos principalmente que la diferencia es muy grande entre los modelos que premian con puntajes partido a partido y aquellos que analizan todo lo sucedido y ofrecen puntajes generales. En los últimos, es mucho más difícil remontar una mala situación.

Para el método de Colley, se encontró una debilidad en sus valoraciones en casos en los que hay grupos separados de habilidades distintas. Al asumir desde un principio que todos juegan igual, el modelo solo toma el desempeño observado. Como no se hacen asunciones previas, el método falla. Por lo que Colley aplica mejor a condiciones donde cada equipo juega contra una muestra representativa de la habilidad general.

En cuanto a las mejores estrategias para los distintos modelos, se observó claramente que ganarle a jugadores con buen ranking en Colley es valorado y es razonable si se quiere subir de posición, siempre y cuando se tenga suficiente confianza sobre una victoria. Esto es válido también para nuestro modelo, ganarle a un jugador con ranking más alto da más puntos por definición. En contraste, en WP ganarle a cualquier equipo es equivalente, por lo que la mejor opción para cualquier equipo es jugar contra equipos tan malos como sea posible.

Sin embargo, es importante notar que no encontramos evidencia para concluir que jugar contra el primero en modelos como el Colley o el nuestro son buenas opciones si este último tiene su puesto "bien merecido". Es decir, para tomar una buena decisión en este tipo de modelos es necesario tener una buena estimación de la probabilidad de ganarle a un determinado contrincante. Más precisamente, lo ideal sería jugar contra el equipo que maximice el valor esperado de los puntos obtenidos.

En relación a una estrategia para aumentar la posición en el ranking. En base a los experimentos realizados, creemos que vencer al rival en la primera posición otorga cierta certeza de que se estará mas cerca del primer puesto en pocos partidos. Esto parece válido tanto para la circunstancias en donde todos compiten contra todos (como resulta la Premier League) como para cuando se realiza un grupo reducido de partidos (como resulta ATP).

Con respecto a nuestro modelo, la principal conclusión que sacamos es que sistemas de puntajes que apliquen un movimiento de puntajes directo partido a partido no ofrecen buenos resultados. Al menos para situaciones en las que los competidores no mejoran a lo largo del tiempo. Tal vez para un balance de partidos correspondientes a varios años en el que tiene sentido perder memoria de lo que pasó antes funciona mejor.

De todas formas, logramos observar la superioridad de un modelo en que la premiación en puntos ofrecida al ganador esté acotada y no siga creciendo sin importar la diferencia de habilidad. La explicación que le encontramos al resultado es que para la mayoría de las competencias la probabilidad de ganar no converge a cero con la diferencia de habilidad, si no que siempre existe la posibilidad de que alguien malo le gane a alguien bueno y por lo tanto el modelo no debería ser demasiado sensible a estas situaciones.

Referencias

[1] Colley, W. *Colley's Bias Free College Football Ranking Method* (Princeton University, 2002).