UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Matemática

Tesis de Licenciatura

Sistemas de ranking con aplicaciones a deportes.

Pablo Mislej

Directora: Dra. Mariela Sued

Abril de 2023

Agradecimientos

En primer lugar quiero agradecer al tiempo que me permitió llegar hasta aquí, fue sin dudas un largo camino.

A su vez en el trayecto he conocido a muchas personas a quienes deseo agradecerles en este acto por formar parte de mi galería de afectos y amistades. En ellos incluyo a compañeros y docentes con los que compartí mi etapa de formación en la Facultad de Exactas.

Doy las gracias también a mi familia, en especial a mis padres, por alentar mi vocación de estudiar esta carrera; a Alejandro por brindarme desinteresadamente sus conocimientos de programador y ayudarme a descargar algunos de los datos de la web; a Pablo De Nápoli por su lectura crítica del documento y el aporte de sugerencias de mejora; y por último a Mariela, quien ha sido un sostén fundamental en el proceso de realización de la tesis y que se ha comportado de manera admirable sacando a relucir lo buena persona que es.

ACLARACIÓN: La versión final que queda publicada para consulta bibliográfica pudo tomar las recomendaciones de mejora de ambas juradas, las Dras. Alicia Dickenstein y Marcela Svarc, a quienes les agradezco encarecidamente por la dedicación puesta en el proceso de evaluación.

Índice

1. Introducción **3**
1.1. El problema y sus componentes . . . 3
1.2. Estado del arte . . . 4
1.3. Guía para la lectura de esta tesis . . . 4

2. Objetivos **5**

3. Metodologías presentadas **6**
3.1. Sistemas de ranking . . . 6
3.1.1. El método MVR . . . 6
3.1.2. El método del Fair-Weather-Fan . . . 12
3.1.3. *Wei's method* y otros rankings espectrales . . . 16
3.1.4. Rankings basados en ratings . . . 28
3.1.5. El método de rating Ataque-Defensa . . . 40
3.1.6. Ensambles de rankings . . . 44
3.2. Visualizaciones de rankings . . . 49
3.3. Productos derivados . . . 59
3.3.1. *Handicap* . . . 59
3.3.2. Predicción hacia adelante . . . 62
3.3.3. *Performance rating* . . . 69

4. Conjuntos de datos **72**

1. Introducción

Un simple relato da cuenta del estado de voracidad que existe hoy en el consumo de información:

Una familia con hijos en edad escolar se traslada por trabajo a otra ciudad.
La persona de la empresa que los recibe les cuenta que tienen 2 escuelas para elegir, donde anotar a los chicos.
La respuesta de los padres no se hace esperar ***¿cuál de las escuelas es mejor?***

El anterior es un ejemplo más de las tantas situaciones en que los humanos se sienten confortados al recibir cualquier lista de dos o más elementos previamente ordenada. Además y típicamente, ver una colección de elementos rankeada mediante un criterio provoca como acción inmediata evaluar si el modo en que se ordenaron los items fue lo suficientemente preciso. En la práctica el reconocer que **A** «se encuentra *justificadamente* por encima de» **B** posibilita en ocasiones que una expectativa previa se revierta o que el criterio experto de alguien cambie.

A lo largo de esta tesis iremos repasando distintos enfoques que llevaron a formalizar métodos de ranking, todos los cuales comparten una idea subyacente: fuera la que fuera la motivación, el mecanismo creado procurará un ordenamiento *justo*. Así, desde la mirada que le dan las Ciencias Exactas (matemática y computación) al problema, esto significará que criterios de «óptimo alcanzado» o de «definición natural» entren a jugar en la aplicación misma del método. Pero para adentrarnos en ello veamos primeramente las definiciones que se usarán en adelante.

1.1. El problema y sus componentes

CONSIGNA: *Organizar un conjunto de elementos en función a su importancia individual.*

- Los elementos del conjunto serán nombrados $1, \ldots, t$ y el orden determinado por el ranking para el elemento i-ésimo se conocerá como $\nu(i)$. Así $\nu(3) = 2$ significa que el elemento 3 resultó ordenado en 2° lugar. Alternativamente los notaremos $1, \ldots, n$ si resulta conveniente hacerlo dado el contexto de la teoría involucrada.
- Toda vez que exista un atributo numérico asociado a cada elemento y decidamos crear nuestro ranking sobre la base del ordenamiento dado por esa asignación, diremos que el ranking está basado en un rating, o equivalentemente es del tipo *rating-based*, y llamaremos **rating** a la cantidad en función a la cual se define dicho orden. Este tipo de rankings permite, por su naturaleza, que dos elementos compartan una misma posición relativa; es decir, admite empates.
- Para un orden establecido es posible luego definir quién se ubica por encima de quién, dado un par de elementos. Ese contraste lleva el nombre de *comparación por pares* o

pairwise comparison. Más aún, en ocasiones sólo a partir de observar eventos históricos que permiten comparar elementos de a pares, es que se llega a una definición del mecanismo de ordenamiento.

1.2. Estado del arte

Al momento de la publicación de [Langville and Meyer, 2012] y [Alvo and Philip, 2014], ambos estudios sirvieron como resúmenes de las técnicas más extendidas sobre el particular, con diferencias en los enfoques, a saber.

El libro [Langville and Meyer, 2012] revisa la Teoría de Rankings desde sus orígenes y pone énfasis en los métodos basados en el Álgebra de matrices y la Optimización. Complementariamente, el volumen [Alvo and Philip, 2014] se concentra en aquellos mecanismos creados en el marco de la Estadística. Una y otra obra coinciden en que no tienen precedentes en cuanto a haber juntado en un único lugar a muchas de las referencias previas, de diferentes autores y originalmente creadas sobre la base de variadas motivaciones.

Otra cosa que tienen en común ambos títulos es que toman a los deportes de escala mundial (con mayor destaque en los que son populares en *Occidente* i.e. Norteamérica y Europa) como ejemplos de aplicación de los métodos de ranking. Es claro que la naturaleza de dicha información genera una atención especial en el lector avezado en esa temática, a su vez, los datos con los que se trabaja se encuentran a menudo accesibles en sitios de estadísticas deportivas, enciclopedias, casas de apuestas, etc. de forma sencilla.

Conforme lo anterior, ambas referencias son por separado buenos lugares adonde consultar los aspectos más importantes sobre los sistemas de ranking en uso actualmente.

1.3. Guía para la lectura de esta tesis

Partiendo de lo expresado en **1.2**, esta tesis se organizó para que las subsecciones de la sección **3** concentren en detalle la información relacionada a los métodos más presentes en la literatura, y donde sobre algunos de ellos se mostraron ejemplos de aplicaciones creados *ad hoc* por mí.

El orden de las subsecciones elegido busca hilar ideas a medida que se avanza con la lectura de las mismas, pero con el resultado a la vista, unas con otras guardan poca relación entre sí, permitiendo evitar el camino establecido y dando forma a una guía de consulta de metodologías (en orden no alfabético, claro). Intercaladas con el texto del documento se pueden encontrar cajas de referencias que buscan hacer las veces de «puertas» a temas generales de la matemática relacionados al procedimiento de ranking que se cuenta en ese momento; no resultan más que síntesis introductorias para que el lector conozca de dónde provienen los conceptos en uso y disponga de un faro hacia donde ampliar el alcance.

2. Objetivos

El objetivo general de este trabajo es presentar en profundidad una selección de métodos de ranking, explicados para un público con conocimientos que se enseñan en las materias de la Licenciatura, y sumando en cada caso un ejemplo de aplicación a deportes que valide el sentido para el que fue creado el método.

Adicionalmente, los apartados 3.2 y 3.3 buscan expandir el ámbito de estudio a temas anexos, por fuera del repaso a los métodos seleccionados. Debido a la forma en que se eligió compendiar información detallada de diversos orígenes es que se espera que quien lea este trabajo pueda encontrar vinculaciones con aspectos de la matemática y la computación, como:

- Teoría de Grafos
- Optimización
- Álgebra lineal
- Probabilidad
- Estadística

Como mencionamos, el principal propósito de este trabajo radica en exponer diferentes sistemas de ranking que resultan de la aplicación de recursos de la Matemática que asumimos conocidos por el lector. Pese a lo cual, fueron incluidas referencias sobre los aspectos matemáticos mencionados en cada subsección para quien quisiera profundizar en los temas afines.

3. Metodologías presentadas

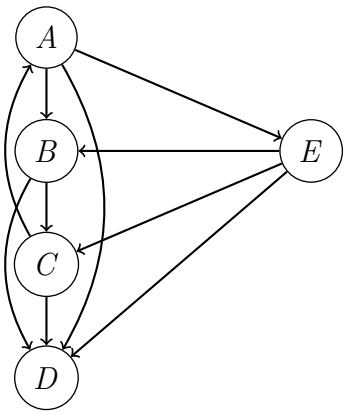
3.1. Sistemas de ranking

3.1.1. El método MVR

Este primer método consiste en uno que destaca por lo fácil de comprender el uso que hace de las definiciones matemática-computacionales con el propósito de hallar un mecanismo óptimo para ordenar elementos de una lista. El método MVR −sigla de *Minimum Violations Ranking*− remite a que de una colección de elementos sabemos con antelación quién de un par de ellos debería ser ubicado arriba en el orden que se pretenderá darle al conjunto. Esta presunción, que cité antes en 1.1 bajo el concepto de «comparaciones de a pares» o «pairwise comparisons», será aquí el ingrediente primario para la generación de un MVR. Antes de enunciar las definiciones que nos van a ayudar a adentrarnos en concreto en el método en sí, recordemos que un **grafo** $G = (V, E)$ queda caracterizado por V, sus nodos o vértices, y por el conjunto E de las aristas; cuando los elementos de E son dotados de una dirección dada por un *nodo-origen* y un *nodo-destino*, a dicho grafo se lo conoce como **digrafo** o **grafo dirigido**. Un **camino en un digrafo** lo constituyen una sucesión de aristas que recorren nodos siguiendo la dirección de éstas.

Definición. *Un digrafo G de n nodos donde cada par de ellos se encuentran conectados mediante una ÚNICA arista dirigida se llama* ***digrafo de torneo****. En tal caso, la cantidad de aristas es $n(n-1)/2$.*

El nombre «de torneo» para un digrafo $G = (V, E)$ surge de asociar los nodos V a equipos o jugadores y el conjunto de las aristas E a la prevalencia de uno sobre el otro derivada del resultado registrado. Más específicamente, la flecha partirá del ganador entre ambos −según el record «cara a cara»− y arribará al nodo del perdedor en cuestión. Para verlo en un ejemplo, debajo se muestra la representación de un digrafo de torneo de 5 jugadores, tanto en su figura en el plano (a) como a través de la matriz de adyacencia que lo define (b):



(a)

	A	B	C	D	E
A	0	1	-1	1	1
B	-1	0	1	1	-1
C	1	-1	0	1	-1
D	-1	-1	-1	0	-1
E	-1	1	1	1	0

(b)

Figura 1: Digrafo de torneo de 5 jugadores o equipos.

El digrafo de torneo de tamaño n obra como estructura acorde para el registro de las $\frac{n(n-1)}{2}$ comparaciones de a pares entre los n elementos que se quieren ordenar.

Definición. *Dado un ranking R de n jugadores o equipos una* ***violación*** *tiene lugar cada vez que para un par (i, j) con i rankeado por encima de j en R, en la comparación de a pares se registre $i < j$ («$i \leftarrow j$» en la representación del grafo). Diremos además que un ranking verifica la condición de ser* ***MVR*** *si minimiza el número de violaciones entre todos los posibles rankings.*

La Figura 2 muestra en un ejemplo sencillo que el número de violaciones para el que se alcanza el mínimo no es 0, a la vez que se comprueba la no-unicidad en la definición de MVR precedente:

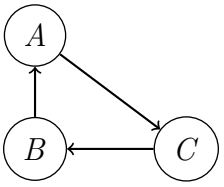


Figura 2: Los ordenamientos ACB, CBA y BAC son todos MVR con nro. de violaciones = 1.

Daremos a continuación por la via de la Teoría de Grafos una herramienta para hallar tales rankings.

En [Ali et al., 1986] se prueba que el ranking ν es un MVR si vale que $\nu^{-1}(1) \to \cdots \to \nu^{-1}(t)$ es un camino (dirigido) en el digrafo G de las comparaciones de a pares. Además, dicho recorrido no deja ningún nodo sin visitar ni pasa dos veces por un mismo nodo (a esto se lo conoce como un *camino Hamiltoniano)*. Así el problema de hallar un MVR se reduce a encontrar todos los caminos Hamiltonianos* en G y elegir entre ellos al que represente al ranking con el menor número de violaciones.

(*) en un grafo se llama camino Hamiltoniano a aquel que recorre todos los nodos pasando una sola vez por cada uno; Laszlo Redei demostró en 1934 que todo digrafo de torneo contiene al menos un camino (dirigido) de estas características − la historia de este hallazgo está contada en [Harary and Moser, 1966]

Esta reducción hacia otro problema más conocido permite aproximar un orden de complejidad de cualquier algoritmo de búsqueda del MVR. En [Ali et al., 1986] se accede a los detalles de esta aseveración, así como a la defnición de una heurística para encontrar una solución candidata, sin tener que utilizar fuerza bruta. En concreto, determinar si un digrafo posee al menos un camino Hamiltoniano conforma la clase de los problemas NP-completos, y por más que aquí el digrafo sea de tipo torneo, los autores muestran que hallar el MVR es equivalente al problema de encontrar la «mayor» submatriz que pueda ser triangulada superiormente, lo cual ha sido probado que es NP-completo o NP-hard.

Referencia anexa: Teoría de Grafos

Los primeros abordajes a la Teoría de Grafos remiten a trabajos de ***Leonhard Euler*** *(1707-1783), la misma constituye hoy una rama activa de la Matemática y los conceptos sobre los que ella versa son fundamentales para comprender desafíos vigentes de la Computación, como la complejidad algorítmica. Aunque en esta tesis sólo tomemos las ideas primarias de Grafo, Camino y Grado i.e. número de aristas que parten o llegan a un nodo, son vastos los conceptos derivados que el lector podrá profundizar por su cuenta.*

Deo, N., *Graph Theory and Applications to Engineering and Computer Science, Prentice Hall, 1974 es un buen lugar por donde empezar.*

Seguidamente, voy a presentar un ejemplo en un contexto conocido como es el de los clubes de fútbol más ganadores de Argentina; donde partir de las comparaciones de a pares realizaré el digrafo de torneo correspondiente y hallaré los caminos hamiltonianos para quedarme con el MVR, es decir, aquel ranking que minimice el número de violaciones.

Históricamente, los equipos de esta lista fueron considerados de élite a nivel local:

- Boca Juniors [**Bo**]
- Independiente [**In**]
- Racing Club [**Ra**]
- River Plate [**Ri**]
- San Lorenzo [**SL**]

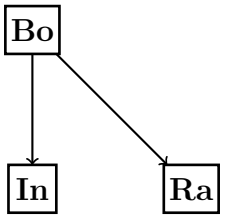
Y entre sí −a la fecha− llevan jugados 1992 encuentros oficiales, sumando los de las eras amateur y profesional; así es posible armar el digrafo dado por las prevalencias entre ellos, iniciemos con el récord positivo de **Boca** sobre **Independiente** (Ganador Boca: 73 - Empates: 58 - Perdedor

Boca: 65). En adelante y dependiendo el contexto se podrá abreviar Ganador-Empates-Perdedor como G-E-P.

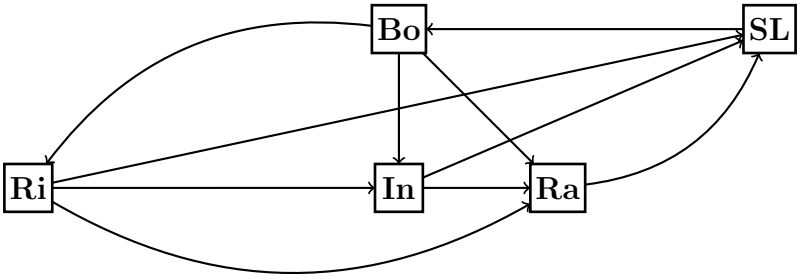
La flecha dirigida del nodo **Bo** *al nodo* **In** *señala la ventaja de +8 triunfos a favor de Boca.*

NOTA: Todos los resultados históricos que usaré para esta aplicación del MVR los extraje de la página web www.promiedos.com.ar/historiales. En general, cada uno de los conjuntos de datos utilizados en la tesis posee su fuente detallada en la Sección 4.

Idénticamente se incluye ahora la arista de **Boca** a **Racing** por el récord Ganador Boca: 83 - Empates: 43 - Perdedor Boca: 55



Y finalmente se completa con el resto



Como vimos antes, el MVR resulta ser uno de los varios caminos Hamiltonianos del digrafo; en la Figura 3 se muestra que **Bo** $\rightarrow$ **Ri** $\rightarrow$ **In** $\rightarrow$ **Ra** $\rightarrow$ **SL** manifiesta una única violación (en rojo) y se puede comprobar que cualquiera de los otros 8 caminos posee más violaciones (ver Figura 4):

En particular, la sola existencia del subgrafo que tiene únicamente a Boca, Racing y San Lorenzo, donde **Bo** $\rightarrow$ **Ra** $\rightarrow$ **SL** $\rightarrow$ **Bo**, hará que no pueda definirse un ranking sin violaciones.

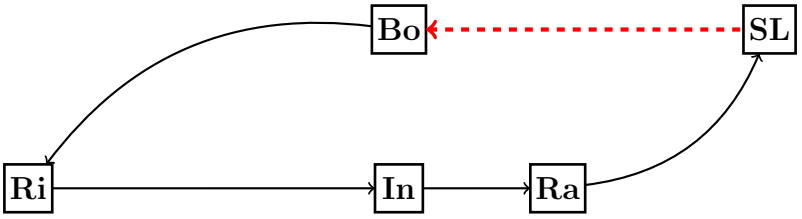


Figura 3: Camino (en negro) que representa al MVR (número de violaciones = 1, en rojo).

Bonus Track

Aunque el ejemplo anterior es lo suficientemente atractivo y simple para entender a través de él cómo se llega al MVR, es interesante preguntarse qué pasa al sumar más equipos al universo a ordenar. Arbitrariamente acopié los récords de tres que por ganar al menos una vez la Competencia de Clubes de Sudamérica (i.e. Copa Libertadores) podrían merecer figurar en el «top 5» histórico del Fútbol Argentino:

	vs **Argentinos Jrs**	*vs* **Estudiantes LP**	*vs* **Vélez Sársfield**
Boca	$67-40-34$	$110-47-45$	$89-41-45$
River	$71-36-27$	$97-43-41$	$97-50-50$
Indep'te	$55-43-36$	$75-56-59$	$65-60-64$
Racing	$61-43-28$	$73-46-70$	$74-63-52$
S.Lorenzo	$66-33-31$	$76-52-56$	$80-59-58$

Frente a esta particularidad de que todos los 5 equipos analizados antes poseen récords positivos frente a los 3 equipos que se quieren anexar al ranking, entonces necesariamente estos últimos ocuparán las posiciones #6, #7 y #8, y quien caiga en cada lugar será consecuencia exclusiva de cómo dan los historiales **AJ** *vs* **Es**, **AJ** *vs* **VS** y **Es** *vs* **VS**; la representación de la Figura 5 muestra a los 8 equipos, más el Camino Hamiltoniano $\equiv$ MVR y las violaciones sobre las que se consiguió el óptimo del método:

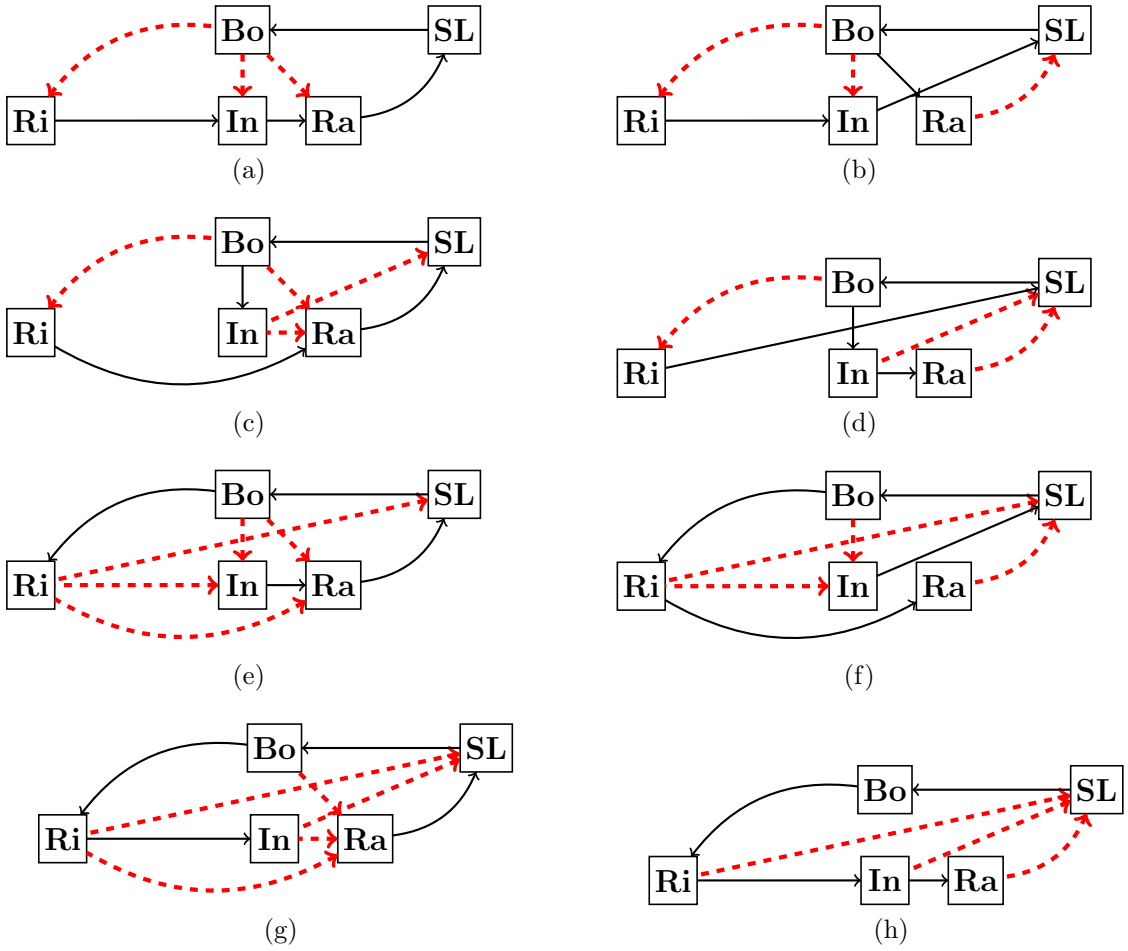


Figura 4: Las otras ocho representaciones de caminos Hamiltonianos no asociados a un MVR. (a), (b), (d) y (h) tienen 3 violaciones cada una; (c) y (f) tienen 4; y (e) y (g) tienen 5.

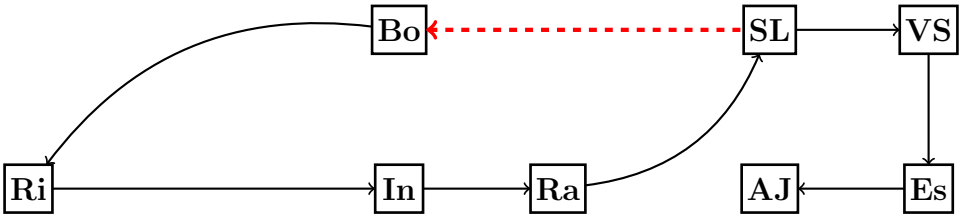


Figura 5: Camino del MVR para el conjunto de clubes campeones de la Copa Libertadores.

3.1.2. El método del Fair-Weather-Fan

En 3.1.1 se buscó minimizar el número de violaciones entre todos los rankings posibles, sin embargo sólo se utilizaron estados binarios {1,-1} de la información de las «pairwise comparisons». Por ejemplo en el digrafo aparece la arista **Bo** $\rightarrow$ **In** por la ventaja de +8 triunfos a favor de Boca, sin embargo nada hubiera cambiado si la ventaja era +1 o +100, lo mismo la violación que hay entre **Bo**$\leftarrow$--**SL** ocurre por la ventaja actual de +3... nada se modificaría si fuera +30.

Repasaremos ahora un método sobre el que influyen las diferencias numéricas entre triunfos y derrotas y no solamente el signo. Y para ello cambiaremos de deporte y de desafío.

En el tenis masculino profesional se vivió un periodo de 17 años en los cuales solamente hubo cuatro mejores tenistas, cada uno por sus cualidades, a ese grupo de tenistas se lo llamó con justa razón los BIG FOUR:

- **Roger Federer**, ganador de 103 torneos, nadie lo iguala en este apartado
- **Rafael Nadal**, ganador de 22 torneos de Grand Slam, nadie lo iguala en este apartado
- **Novak Djokovic**, fue el 1 del ranking mundial por 7 años, nadie lo iguala en este apartado
- **Andy Murray**, doble Oro olímpico en singles, nadie lo iguala en este apartado

Mientras que la matriz de enfrentamientos cara-a-cara da así (tengamos presente que no hay empates en tenis, la primera cifra son los juegos Ganados por el jugador-fila y la segunda los Perdidos):

	RF	**RN**	**ND**	**AM**
RF	$\times$	$16-24$	$23-27$	$14-11$
RN	$24-16$	$\times$	$29-30$	$17-7$
ND	$27-23$	$30-29$	$\times$	$25-11$
AM	$11-14$	$7-17$	$11-25$	$\times$

Pasemos ahora a renombrar filas y columnas y volquemos en las celdas la fracción que resulta de dividir el valor P=Perdidos sobre los Juegos Totales involucrados en la fila:

	RF_1	**RN_1**	**ND_1**	**AM_1**
RF_0	$\times$	$^{24}/_{115}$	$^{27}/_{115}$	$^{11}/_{115}$
RN_0	$^{16}/_{123}$	$\times$	$^{30}/_{123}$	$^{7}/_{123}$
ND_0	$^{23}/_{145}$	$^{29}/_{145}$	$\times$	$^{11}/_{145}$
AM_0	$^{14}/_{85}$	$^{17}/_{85}$	$^{25}/_{85}$	$\times$

Finalmente reformaremos la matriz para que nos sirva como objeto matemático, así las $\times$ de la diagonal serán reemplazadas por los valores positivos tales que sumar cada fila dé 1:

	RF_1	**RN_1**	**ND_1**	**AM_1**
RF_0	$^{53}/_{115}$	$^{24}/_{115}$	$^{27}/_{115}$	$^{11}/_{115}$
RN_0	$^{16}/_{123}$	$^{70}/_{123}$	$^{30}/_{123}$	$^{7}/_{123}$
ND_0	$^{23}/_{145}$	$^{29}/_{145}$	$^{82}/_{145}$	$^{11}/_{145}$
AM_0	$^{14}/_{85}$	$^{17}/_{85}$	$^{25}/_{85}$	$^{29}/_{85}$

O redondeando a tres decimales:

	RF_1	**RN_1**	**ND_1**	**AM_1**
RF_0	0.461	0.209	0.235	0.096
RN_0	0.130	0.569	0.244	0.057
ND_0	0.159	0.200	0.566	0.076
AM_0	0.165	0.200	0.294	0.341

Hemos llevado la matriz de récords cara-a-cara entre el «tenista fila» y el «tenista columna» a la forma de una matriz que define a un Proceso de Markov [PM] (i.e. estocástica) por contener números entre 0 y 1, los cuales suman 1 por fila. Se describe a continción un PM que usará a las celdas de la matriz como probabilidades de transición, con distribución inicial uniforme:

I elijo de manera equiprobable ser hincha de uno de los 4 jugadores

II las probabilidades que suman uno en la fila de éste (estado actual del PM) servirán para tomar la decisión sobre cuál será el siguiente jugador por el que hincharé (siguiente estado), procedimentalmente en el siguiente estado elegiré ser hincha del «tenista columna j» con la probabilidad P_{ij} que se encuentra en la fila del tenista i; y eventualmente podré conservar mi favoritismo por el jugador del que vengo siendo hincha

III recurrentemente continúo n veces con el paso II y anoto las frecuencias relativas asociadas a las veces que *fui hincha* de cada tenista

Así mediante los porcentajes que obtengo y la escala que surge de éstos se desarrolla el método de ranking llamado alegóricamente del FAIR-WEATHER-FAN o del HINCHA-DEL-GANADOR (recordar que las probabilidades de cada fila estiman, sobre la base histórica de partidos jugados, cuán probable es ver ganar al «tenista columna» en un juego en el que participó el «tenista fila»).

Referencia anexa: Procesos de Markov

Los llamados Procesos de Markov (según cuál sea la referencia, también se los conoce como Cadenas de Markov) son encuadres definidos para modelar la conexión posible entre distintos estadíos mediante la probablidad de que se concrete el pasaje de uno a otro, describen por tanto una categoría de eventos estocásticos y su formulación fue llevada adelante hacia 1907 por Andrei Markov, de quien heredó su nombre. Las ideas involucradas y propiedades inherentes a tales procesos guardan relación además con el Álgebra Lineal y en este volumen se pueden encontrar sus fundamentos, así como aplicaciones prácticas que exceden a los métodos de ranking:

Meyer, C., *Matrix Analysis and Applied Linear Algebra, SIAM, 2000.*

RESULTADO al que se aribó tras n=100 iteraciones:

ND $\to$ **RN** $\to$ **RF** $\to$ **AM**, sustentado en $\dfrac{39}{100} > \dfrac{30}{100} > \dfrac{22}{100} > \dfrac{9}{100}$

RESULTADO al que se aribó tras n=1000 iteraciones:

ND $\to$ **RN** $\to$ **RF** $\to$ **AM**, sustentado en $\dfrac{377}{1000} > \dfrac{330}{1000} > \dfrac{196}{1000} > \dfrac{97}{1000}$

Los ordenamientos resultantes coincidieron tanto para las primeras n=100 como para las primeras n=1000 iteraciones del proceso. Por fortuna, la teoría de Procesos de Markov brinda una caracterización de los valores de convergencia de las frecuencias relativas, cuando n crece a infinito, para cierta clase de procesos; haremos uso de estos resultados para calcular numéricamente tales proporciones en el límite, pero antes traeremos notaciones y definiciones que serán útiles.

Sea

$$\mathbb{A} = \begin{pmatrix} 0.461 & 0.209 & 0.235 & 0.096 \\ 0.130 & 0.569 & 0.244 & 0.057 \\ 0.159 & 0.200 & 0.566 & 0.076 \\ 0.165 & 0.200 & 0.294 & 0.341 \end{pmatrix} \tag{1}$$

la matriz que reúne las probabilidades de transicionar entre los estados iniciales (indexados según las *filas* de $\mathbb{A}$) y los estados finales (según las *columnas* de $\mathbb{A}$) del PM.

Definición. *Una matriz de transiciones $\mathbb{A}$ es* ***irreductible*** *si para todo par (i, j) existe $k = k(i, j)$ tal que $\mathbb{A}^k(i, j) > 0$.*

Definición. *Un estado i de una matriz de transiciones $\mathbb{A}$ se dice* ***periódico de período d***

si para retornar al estado i partiendo del estado i es necesario que transcurran al menos d iteraciones. Formalmente,

$$d = dcm\{n \geq 1 : \mathbb{A}^n(i,i) > 0\}$$

donde dcm son las siglas de «divisor común mayor» y siempre que el conjunto sea no vacío. Además, si d=1 se dice que el estado es ***aperiódico****. Una matriz es* ***aperiódica*** *si todos sus estados son aperiódicos.*

Así $\mathbb{A}$, definida en (1), resulta ser estocástica, **irreductible** y **aperiódica**, y por lo tanto vale lo siguiente:

Teorema. *Sea $\mathbb{A}$ una matriz estocástica, irreductible y aperiódica. Entonces existe una única medida estacionaria $\boldsymbol{\pi}$, solución a las ecuaciones $\mathbb{A}' \cdot \boldsymbol{\pi} = \boldsymbol{\pi}$, donde $\mathbb{A}'$ es la transpuesta de la matriz $\mathbb{A}$ y $\sum_i \boldsymbol{\pi}(i) = 1$. La solución satisface $\boldsymbol{\pi}(i) > 0$ para todo i. Además las frecuencias relativas obtenidas de iterar el PM convergen a $\boldsymbol{\pi}$. Más específicamente, e independiente de la distribución inicial del proceso, si X_k denota al estado del proceso en el paso k, vale*

$$\lim_{n\to\infty} \frac{1}{n} \cdot \sum_{k=1}^{n} \mathbb{I}_{(X_k=i)} = \boldsymbol{\pi}(i),$$

donde $\mathbb{I}$ es la función indicadora.

Una demostración del resultado anterior puede hallarse en [Durrett, 1999].

Finalmente, por resultados de la llamada Teoria de Perron-Frobenius (citados en 3.1.3) el $\lim_{k\to\infty} (\mathbb{A}')^k \cdot \mathbf{u}$, donde $\mathbf{u}$=(0.25,0.25,0.25,0.25), existe y da como resultado $\boldsymbol{\pi}$, a quien se lo llama también **vector estacionario** del PM, que es lo que estábamos buscando. De esta manera y con poquísimas iteraciones en k se arriba a porcentajes muy próximos a los de la convergencia, como se ve en este reporte:

	[RF]	[RN]	[ND]	[AM]
[k=0]	0.25000	0.25000	0.25000	0.25000
[k=1]	0.22862	0.29442	0.33457	0.14238
[k=2]	0.22020	0.31063	0.35658	0.11258
[k=3]	0.21700	0.31654	0.36225	0.10421
[k=10]	0.21513	0.31992	0.36413	0.10082
[k=20]	0.21513	0.31992	0.36413	0.10082

Más aún, el vector $\boldsymbol{\pi}$ es el autovector de $\mathbb{A}'$ asociado al autovalor 1, con lo que existen métodos numéricos para aproximar sus componentes:

Cálculo NUMÉRICO del vector-rating utilizando *Eigenvalues* de Wolfram:

```
                [RF]      [RN]      [ND]      [AM]
π=    0.21512  0.31998  0.36410  0.10080
```

Cálculo APROXIMADO del vector-rating computando $(\mathbb{A}')^k \cdot \mathbf{u}$, con k=10:

```
                [RF]      [RN]      [ND]      [AM]
π=    0.21513  0.31992  0.36413  0.10082
```

Llegamos hasta acá con la descripción y aplicación del método, que junto con el de la subsección 3.1.1 son los más intuitivos de los que relevé, sin por eso dejar de significar ideas poderosas.

3.1.3. *Wei's method* y otros rankings espectrales

Algo que no se mencionó hasta ahora son las fallas a las que nos atenemos de elegir rankear por alguno de los métodos de las subsecciones anteriores, por caso el del FAIR-WEATHER-FAN. Supongamos que los dos tenistas que quedaron más abajo en el ranking, RF y AM, se pusieran de acuerdo y comenzaran a jugar entre sí en cantidades exhorbitantes, quiero decir, mientras que lo común sería que los tenistas top pudieran enfrentarse hasta 30 o 40 veces en su carrera (50 ya sería una cifra extraordinaria) estos dos jugadores planificarían enfrentarse al menos una vez al mes, coincidiendo en torneos, jugando exhibiciones, etc. Un hecho así haría que las 4 celdas redondeadas de la matriz crezcan a la vez que las del resto de esas filas caerían, y pronto las sombreadas en gris sumarían poco y nada, llevando a que el proceso (del que luego se deriva el vector-rating) haga que RF y AM prevalezcan en el ordenamiento por encima de ND y RN:

	RF_1	**RN_1**	**ND_1**	**AM_1**
RF_0	0.461	0.209	0.235	0.096
RN_0	0.130	0.569	0.244	0.057
ND_0	0.159	0.200	0.566	0.076
AM_0	0.165	0.200	0.294	0.341

El anterior sería juzgado sin dudas como un comportamiento espurio, instrumentado con el objetivo de corromper el método de ranking. Por esto es que se han propuesto herramientas más sólidas, que apuntan a solucionar el aspecto negativo que vimos antes.

Wei's method

Teh-Hsing Wei hizo su doctorado en Matemáticas en Cambridge y presentó su tesis hace 70 años sobre Fundamentos algebraicos de la teoría de ranking. Dicho escrito no fue publicado

pero sí citado cientos de veces, en lo personal no logré acceder al material original pero hay en [Kendall, 1955] una revisión al detalle del método que elegí contar en este apartado, con lo cual me apoyaré en lo elaborado por Kendall dándole el justo crédito a quien lo diseñó. Elijo por otra parte comenzar desde tan atrás para poder mostrar las raíces motivacionales de algo que −tras tantos años− sigue estando vigente, con el agregado de mejoras como veremos más adelante[1].

El **método de Wei** se entiende muy fácil al verlo aplicado a un ejemplo sencillo, para ello voy a usar resultados de un deporte distinto a los que vimos hasta ahora, como es el Rugby (en el artículo donde hallé contado el método usaron de ejemplo un torneo de ajedrez en el que se enfrentaron 6 jugadores todos-contra-todos, me pareció oportuno dejar a un lado los datos de Ajedrez porque servirán mejor para ser usados con otros métodos). Así que paso a mostrarles qué información de resultados de rugby vamos a usar... siguiendo con el número de «6» que usó Kendall tomaré el caso del Torneo de las Seis naciones que se juega cada año entre Febrero y Marzo, desde hace largas décadas y convocando siempre a los mismos seleccionados: Escocia, Francia, Gales, Inglaterra, Irlanda e Italia. Recopilé entonces cómo les ha ido entre sí las últimas tres veces que participaron del Seis naciones, en las ediciones de los años 2020, 2021 y 2022. El método que describimos asume que no se producen empates, tal como ocurre en los datos que presentamos en este ejemplo. De hecho, la ocurrencia de empates en el rugby es muy poco frecuente.

	Escocia	**Francia**	**Gales**	**Inglaterra**	**Irlanda**	**Italia**	**Récord total**
Es		$2-1$	$1-2$	$2-1$	$0-3$	$3-0$	$8-7$
Fr	$1-2$		$3-0$	$2-1$	$3-0$	$3-0$	$12-3$
Ga	$2-1$	$0-3$		$1-2$	$1-2$	$2-1$	$6-9$
In	$1-2$	$1-2$	$2-1$		$1-2$	$3-0$	$8-7$
Ir	$3-0$	$0-3$	$2-1$	$2-1$		$3-0$	$10-5$
It	$0-3$	$0-3$	$1-2$	$0-3$	$0-3$		$1-14$

En cada celda están las victorias del equipo-fila seguidas de las victorias del equipo-columna.

El cuadro anterior expone mucha y rica información de resultados, pero quedémonos en la columna de la extrema derecha, ahí vamos a ver que existe un orden implícito dado por la cantidad o % de victorias registradas − da igual pues todos disputaron 15 encuentros a razón de 5 por edición. Destaca que dos seleccionados poseen el récord de $8-7$, lo cual genera la pregunta acerca de cuál de ambos quedará por sobre el otro al definir el ranking. Esta situación de *empate en una primera lectura* dota de mayor utilidad a los métodos de ranking. Sin embargo me voy a permitir contar de manera diferente el récord de cada selección, no es algo que exija por sí misma la metodología, pero ayudará a forzar que ocurran más situaciones de

[1] Aunque históricamente el **método de Wei** fue reconocido por su originalidad, existe evidencia en contrario en [Vigna, 2019], donde se exhibe que la misma idea había sido propuesta por Landau con anterioridad a Wei, pero que no prosperó su difusión.

igualdad entre equipos. La regla será que si un equipo registra más victorias que derrotas, corresponde asignar un 1 ahí, y 0 en caso contrario; así las cosas la situación cambia por esta otra:

	Escocia	**Francia**	**Gales**	**Inglaterra**	**Irlanda**	**Italia**	**Récord total**
Es		1	0	1	0	1	3
Fr	0		1	1	1	1	4
Ga	1	0		0	0	1	2
In	0	0	1		0	1	2
Ir	1	0	1	1		1	4
It	0	0	0	0	0		0

A) ***Cuadro completo con unos y ceros.*** *Sólo si el equipo-fila registra más victorias que el equipo-columna en duelos entre ambos va un 1 en esa celda, o 0 en caso contrario.*

Tras la modificación del criterio para agregar resultados, aparece el desafío de decidir para las posiciones #1 y #2 quién entre **Francia** o **Irlanda** ocupa cada una, ambas con Récord=4, lo mismo para los puestos #4 y #5 con **Gales** e **Inglaterra**, ambas con Récord=2. La idea del método de Wei es resolverlo mediante una nueva tabla: utilizando los datos presentados en el Cuadro A), preserva los ceros observados y reemplaza los unos por el valor de la columna **Récord total** correspondiente al esquipo derrotado (el equipo columna). A modo de ejemplo, el valor 1 en la celda correpondiente a la fila 1 columna 2 del cuadro A) indica que Escocia *vence* a Francia, que tiene un Récord total igual a 4, y, por consiguiente, ése es el valor presentado en la celda correspondiente a la fila 1 columna 2 del cuadro B). Finalmente se suman los valores por fila para construir la variable **Σ récords derrotados**. Obtenemos así la siguiente tabla:

	Escocia	**Francia**	**Gales**	**Inglaterra**	**Irlanda**	**Italia**	**Σ récords derrotados**
Es		4	0	2	0	0	6
Fr	0		2	2	4	0	8
Ga	3	0		0	0	0	3
In	0	0	2		0	0	2
Ir	3	0	2	2		0	7
It	0	0	0	0	0		0

B) ***Cuadro completo con valores enteros.*** *Sólo si el equipo-fila registra más victorias que el equipo-columna en duelos entre ambos va el récord del derrotado en esa celda, o 0 en caso contrario.*

La última de las columnas del cuadro ya no deja espacio para las igualdades, en el sentido en que sus valores se pueden ordenar ahora de manera estricta (y por lo tanto queda determinado el ranking de los equipos asociados). Yendo a las dudas puntuales que habían sido planteadas antes, en base a este criterio de desempate **Francia** (#1, 8) queda arriba de **Irlanda** (#2, 7) y **Gales** (#4, 3) encima de **Inglaterra** (#5, 2). De esta manera, se obtuvo un ranking sin empates, preservando además el orden inicial. Aunque no hay razones matemáticas que

garanticen este feliz resultado −resolver los empates y preservar el orden inicial− Wei, con mirada algebraica, observó de todas formas que:

- completando con valores 0 los «espacios vacíos» en la diagonal del Cuadro A) se obtiene la matriz cuadrada

$$\mathbb{W} := \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

- la última columna del Cuadro A coincide con

$$\mathbb{W} \cdot \mathbf{1}_6 = \mathbb{W} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 2 \\ 2 \\ 4 \\ 0 \end{pmatrix} \tag{3.1.3.1}$$

- la última columna del Cuadro B coincide con

$$\mathbb{W}^2 \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \mathbb{W} \cdot \begin{pmatrix} 3 \\ 4 \\ 2 \\ 2 \\ 4 \\ 0 \end{pmatrix} = \begin{pmatrix} 6 \\ 8 \\ 3 \\ 2 \\ 7 \\ 0 \end{pmatrix} \tag{3.1.3.2}$$

Es decir que a la terna

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ 2 \\ 2 \\ 4 \\ 0 \end{pmatrix}, \begin{pmatrix} 6 \\ 8 \\ 3 \\ 2 \\ 7 \\ 0 \end{pmatrix} \tag{3.1.3.3}$$

se llega por operar con las potencias de $\mathbb{W}$ sobre el vector-columna de todos 1's; Wei llamó al vector $\mathbf{1}_6$ el *ranking ex aequo* (que significa *ranking de todos igualados*). Esta lectura le hace suponer que la sucesión $\mathbb{W}^0\mathbf{1}_6, \mathbb{W}^1\mathbf{1}_6, \mathbb{W}^2\mathbf{1}_6$... podría converger bajo ciertas circunstancias, y

que siendo que $\mathbb{W}^2\mathbf{1}_6$ devolvió un ordenamiento menos ambiguo y por lo tanto mejor que $\mathbb{W}^1\mathbf{1}_6$, entonces cobraría sentido elevar más y más el exponente para ver a qué se llega. Desarrollando su idea, lo primero que dedujo es que los valores de las coordenadas crecen de una iteración a la siguiente, con la sola excepción del 0 final, que se corresponde con una fila de ceros en la matriz $\mathbb{W}$. Su PREGUNTA quedó rápidamente reescrita así:

$$\textit{¿Existen condiciones bajo las cuales } \left(\frac{\mathbb{W}^k\mathbf{1}_6}{\|\mathbb{W}^k\mathbf{1}_6\|}\right)_k \textit{ converge?} \tag{2}$$

En primer lugar aparece el interés natural de entender por qué Wei se formuló tal pregunta, y lo que Kendall arriesga en [Kendall, 1955] en lenguaje vulgar es: «¿por qué no hacerlo?» Hoy sabemos que con ese planteo Wei abrió una rama que lo llevó, a partir de bucear en propiedades del álgebra de matrices cuadradas, a trabajar con autovalores reales de la matriz $\mathbb{W}$. Esta metodología de trabajo en pos de obtener un ordenamiento se enmarca dentro de lo que hoy conocemos como rankings *de tipo espectral*, tomando su nombre de mirar a $\mathbb{W}$ como operador lineal en un espacio de dimensión finita.

Vamos ahora a introducir unas definiciones que nos permitirán establecer condiciones para las que se pueda garantizar la convergencia de la sucesión de potencias normalizada, presentada en (2).

Definición. *Una matriz $\mathbb{M} \in \mathbb{R}^{n\times n}$ es **p-primitiva** si $\exists p \in \mathbb{N} : \mathbb{M}^p(i,j) > 0 \; \forall\, 1 \leq i,j \leq n$. En particular se la llama **positiva** si es 1-primitiva. Obs.: Toda matriz p-primitiva es irreductible.*

Definición. *Una matriz $\mathbb{M} \in \mathbb{R}^{n\times n}$ es **no negativa** si $\mathbb{M}(i,j) \geq 0 \; \forall\, 1 \leq i,j \leq n$.*

En el siguiente teorema se presenta una clase de matrices donde la convergencia de (2) queda garantizada.

Teorema (Perron-Frobenius [PF])**.** *Sea $\mathbb{M} \in \mathbb{R}^{n\times n}$ una matriz cuadrada positiva; entonces*

*I existe λ **autovalor dominante** de $\mathbb{M}$ i.e. mayor, en valor absoluto, a todo otro autovalor*

II $\lambda \in \mathbb{R}^+$, esto es, λ resulta ser real y positivo

III puede elegirse autovector $\mathbf{v} = (v_i)_{1\leq i\leq n} \in \mathbb{R}^n$ asociado a λ tal que $v_i > 0 \; \forall i$ y $\sum_i v_i = 1$

Más en general:

- *si $\mathbb{M}$ es no negativa e irreductible*

  $\Rightarrow$ *existe $\mathbf{v} = (v_i)_{1\leq i\leq n} \in \mathbb{R}^n$, autovector de $\mathbb{M}$ asociado a un autovalor de máximo valor absoluto con $v_i > 0 \; \forall i$ y $\sum_i v_i = 1$*

- *si $\mathbb{M}$ es no negativa y p-primitiva*

  $\Rightarrow \frac{\mathbb{M}^k\cdot\mathbf{x}}{\|\mathbb{M}^k\cdot\mathbf{x}\|}$ *converge a un múltiplo positivo de $\mathbf{v}$ si $\mathbf{x} \in \mathbb{R}^n$ y $x_i > 0 \; \forall\, i$*

Mientras que en una matriz de triunfos-derrotas la condición de no negatividad se cumple siempre, por contener ésta valores entre 0 y 1, no sabemos a priori si una gran cantidad de celdas con ceros haría que no se verifiquen las hipótesis de PF de irreductibilidad y/o primitividad.

En el libro [Langville and Meyer, 2012] se propone un recurso simple para sortear la problemática de no poder usar el Teorema de Perron-Frobenius para matrices cuadradas ralas i.e. matrices con gran cantidad de celdas en cero. Este mecanismo aplica a $\mathbb{M}$ una «perturbación» que resulta de tomar $\mathbb{M} + \epsilon\mathbf{e}\mathbf{e}'$; el parámetro ϵ adopta un valor positivo lo suficientemente *chico*, mientras que $\mathbf{e}$ es el vector con todas sus coordenadas en 1. Observemos que podemos elegir ϵ suficientemente pequeño y lograr que $\mathbb{M} + \epsilon\mathbf{e}\mathbf{e}'$ resulte una matriz positiva, permitiendo invocar el Teorema.

En adelante, en este capítulo, vamos a asumir que la convergencia (2) se cumple, o bien porque la matriz cuadrada satisface *per se* las condiciones para aplicar PF, o bien porque podemos perturbarla y asegurar que sea positiva.

Retomando el ejemplo, tras la utilización de *Eigenvalues* de Wolfram sobre la matriz $\mathbb{W} + \epsilon\mathbf{e}\mathbf{e}'$, con $\epsilon = 0.1$, obtenemos que $(\lambda_1, \mathbf{v}_1)$ son el par de autovalor y autovector asociados, donde λ_1 es dominante y $\sum_i v_i = 1$:

$$\lambda_1 = 2.475, \mathbf{v}_1 = \begin{pmatrix} 0.22460 \\ 0.29095 \\ 0.13557 \\ 0.08221 \\ 0.26608 \\ 0.00060 \end{pmatrix} \tag{3.1.3.4}$$

Con lo cual Francia > Irlanda > Escocia > Gales > Inglaterra > Italia es el ranking resultante de aplicar el ***Wei's method***, del mismo modo que habíamos visto que el paso de la potencia al cuadrado ya ordenaba de esta forma.

Keener's method

Con base en estos conceptos, James P. Keener publicó en [Keener, 1993] su idea para establecer un ranking de equipos de fútbol universitario (también conocido como *gridiron football*, el de los jugadores con cascos) basado en cálculos matemáticos. Que su método merezca destacarse tiene mucho que ver con el contexto del gridiron football universitario en los años 1990: por aquel entonces venía creciendo con más y más insistencia la necesidad de acordar un criterio para decidir al **campeón nacional** de cada año (N. del A.: son tantas las universidades en los Estados Unidos que conforman equipos de esta disciplina, más de 100 contando sólo las del más alto nivel, que la cantidad de enfrentamientos que requiere un esquema «todos contra todos» jamás podría caber entre Septiembre y Noviembre, como es tradición jugar cada año; esta limitación de calendario hace competir a cada equipo en hasta un máximo de 13 encuentros, a razón de 1 por semana). La pregunta acerca de qué equipo universitario resultó ser «el mejor» de

la temporada acompañaba al deporte desde sus orígenes, varias encuestas no-oficiales buscaban resolver esa disputa, una elaborada por la prensa nucleada en la asociación de medios *Associated Press* y otra llevada a cabo por el gremio de los directores técnicos... lo concreto es que partiendo de criterios diferentes en muchos años había «DOS MEJORES EQUIPOS».

Es así que por el año 1992 se firma un primer acuerdo con algunas universidades para que se programe un partido que defina *en la cancha* quien se llevaría los laureles entre dos candidatos al premio mayor. Comprendiendo lo anterior Keener fue uno más de un conjunto de periodistas especializados, estadísticos y científicos en general que intentó aportar luz para determinar un mecanismo de elección de aquellos dos equipos finalistas, sobre la base de criterios objetivos. Y lo presentó a modo de ranking, dado que además de definir al #1 y al #2 que irían a competir en un partido sin empate, era lógico asignarles al resto de los equipos una posición para que se sepa quiénes estuvieron más o menos cerca de la cima de esa lista.

Una última mención acerca de cómo se estructura este torneo en los Estados Unidos: la imposibilidad de enfrentar a cada equipo con el conjunto de sus rivales hizo que se fueran armando ligas por cercanía llamadas *conferencias*. Entre integrantes de esos grupos de tamaño pequeño se disputan la mayoría de los partidos del calendario, más luego y de común acuerdo hay otros duelos pactados entre equipos, son los llamados juegos extra-conferencia. Descripto el escenario para el cual Keener introdujo modificaciones al método de Wei, con miras de aplicarlo al problema del Fútbol universitario de los años 1990 en los Estados Unidos, me gustaría continuar con el ejemplo del Rugby anexándole más equipos que pudieran representar de algún modo a un grupo cerrado que compite regularmente entre sí, pero no tanto con los del grupo de las Seis naciones, elegí para ello a la Liga del hemisferio sur que conforman Sudafrica, Nueva Zelanda, Australia y Argentina. En el mismo intervalo 2020-2022 que veníamos viendo hay registrados 30 enfrentamientos, con estos récords (un 1 significa más victorias que derrotas, un 0 más derrotas que victorias, y en el caso de SA vs. NZ introduje el valor $1/2$ porque de los 4 encuentros que disputaron ganaron dos cada uno):

	Sudafrica	**Nueva Zelanda**	**Australia**	**Argentina**
SA		$1/2$	0	1
NZ	$1/2$		1	1
Au	1	0		1
Ar	0	0	0	

Y con toda la información de enfretamientos volcada a una sola matriz, la situación es ésta:

	Es	**Fr**	**Ga**	**In**	**Ir**	**It**	**SA**	**NZ**	**Au**	**Ar**
Es	—	1	0	1	0	1	.	.	.	.
Fr	0	—	1	1	1	1	.	.	.	.
Ga	1	0	—	0	0	1	.	.	.	.
In	0	0	1	—	0	1	.	.	.	.
Ir	1	0	1	1	—	1	.	.	.	.
It	0	0	0	0	0	—	.	.	.	.
SA	.	.	.	.	.	.	—	$1/2$	0	1
NZ	.	.	.	.	.	.	$1/2$	—	1	1
Au	.	.	.	.	.	.	1	0	—	1
Ar	.	.	.	.	.	.	0	0	0	—

Esta configuración es la típica de grupos de equipos con enfrentamientos entre sí pero sin cruces extra-grupos; de la misma se lee que la ausencia de interacción Seis naciones europeas-Hemisferio Sur llevaría a que no existan registros de los que valerse para generar un ranking intercalando equipos de uno y otro grupo. Como en el escenario que conté que afrontó Keener al respecto del fútbol universitario, aquí también se disputaron algunos duelos entre «el norte» y «el sur»; los vuelco debajo y expreso a posteriori un breve comentario sobre lo que se ve en los datos:

	Es	**Fr**	**Ga**	**In**	**Ir**	**It**	**SA**	**NZ**	**Au**	**Ar**
Es	—	1	0	1	0	1	**0**	NA	**1**	**0**
Fr	0	—	1	1	1	1	NA	**1**	**0**	**1**
Ga	1	0	—	0	0	1	**0**	**0**	**1**	**0**
In	0	0	1	—	0	1	**1**	NA	**1**	NA
Ir	1	0	1	1	—	1	NA	**1**	NA	**1**
It	0	0	0	0	0	—	NA	**0**	NA	**0**
SA	**1**	NA	**1**	**0**	NA	NA	—	$1/2$	0	1
NZ	NA	**0**	**1**	NA	**0**	**1**	$1/2$	—	1	1
Au	**0**	**1**	**0**	**0**	NA	NA	1	0	—	1
Ar	**1**	**0**	**1**	NA	**0**	**1**	0	0	0	—

Entre las 6 naciones europeas y las 4 del hemisferio sur se dan 24 cruces posibles.
De éstos 24 se jugaron únicamente 16 entre Enero de 2020 y Octubre de 2022.

Estos 16 pares de equipos que se enfrentaron al menos una vez y de lo cual se registró un récord para el duelo uno vs. uno es la condición suficiente para que llevar al límite la multiplicación de la matriz consigo misma resulte en un valor de convergencia. Pero por algo éste es el **método de Keener** y no el de Wei, así que repasaremos las modificaciones que le introdujo respecto a su antecesor:

- en primer lugar Keener entendió que debía distinguirse el 0 por derrota en el duelo uno vs. uno del 0 por otro motivo, por ejemplo en las celdas de la diagonal o en los emparejamientos no realizados (NA), sobre la base de lo cual sugirió ajustar los 0 y 1 por valores

menores a 1/2 y mayores a 1/2 que sigan sumando 1 pero que no estén en los extremos del intervalo [0,1]; los reemplazaremos entonces con $\frac{1}{3}$ y $\frac{2}{3}$ respectivamente

$$\mathbb{K} = \begin{pmatrix} 0 & 2/3 & 1/3 & 2/3 & 1/3 & 2/3 & 1/3 & 0 & 2/3 & 1/3 \\ 1/3 & 0 & 2/3 & 2/3 & 2/3 & 2/3 & 0 & 2/3 & 1/3 & 2/3 \\ 2/3 & 1/3 & 0 & 1/3 & 1/3 & 2/3 & 1/3 & 1/3 & 2/3 & 1/3 \\ 1/3 & 1/3 & 2/3 & 0 & 1/3 & 2/3 & 2/3 & 0 & 2/3 & 0 \\ 2/3 & 1/3 & 2/3 & 2/3 & 0 & 2/3 & 0 & 2/3 & 0 & 2/3 \\ 1/3 & 1/3 & 1/3 & 1/3 & 1/3 & 0 & 0 & 1/3 & 0 & 1/3 \\ 2/3 & 0 & 2/3 & 1/3 & 0 & 0 & 0 & 1/2 & 1/3 & 2/3 \\ 0 & 1/3 & 2/3 & 0 & 1/3 & 2/3 & 1/2 & 0 & 2/3 & 2/3 \\ 1/3 & 2/3 & 1/3 & 1/3 & 0 & 0 & 2/3 & 1/3 & 0 & 2/3 \\ 2/3 & 1/3 & 2/3 & 0 & 1/3 & 2/3 & 1/3 & 1/3 & 1/3 & 0 \end{pmatrix}$$

- `proceso de normalización` – no todos los equipos registran la misma cantidad de duelos uno vs. uno, Gales jugó contra los otros 9 seleccionados, en el extremo contrario Sudafrica se enfrentó contra 6, esa diferencia se traduce en más o menos celdas positivas en la fila correspondiente al equipo; el autor da su explicación de que esta asimetría pervierte el método y propone dividir cada valor > 0 de la fila por el número de duelos que ese equipo tiene en su haber; llamaré $\widetilde{\mathbb{K}}$ a la matriz que surge del anterior proceso de normalización

$$\widetilde{\mathbb{K}} = \begin{pmatrix} 0 & 2/24 & 1/24 & 2/24 & 1/24 & 2/24 & 1/24 & 0 & 2/24 & 1/24 \\ 1/24 & 0 & 2/24 & 2/24 & 2/24 & 2/24 & 0 & 2/24 & 1/24 & 2/24 \\ 2/27 & 1/27 & 0 & 1/27 & 1/27 & 2/27 & 1/27 & 1/27 & 2/27 & 1/27 \\ 1/21 & 1/21 & 2/21 & 0 & 1/21 & 2/21 & 2/21 & 0 & 2/21 & 0 \\ 2/21 & 1/21 & 2/21 & 2/21 & 0 & 2/21 & 0 & 2/21 & 0 & 2/21 \\ 1/21 & 1/21 & 1/21 & 1/21 & 1/21 & 0 & 0 & 1/21 & 0 & 1/21 \\ 2/18 & 0 & 2/18 & 1/18 & 0 & 0 & 0 & 1/12 & 1/18 & 2/18 \\ 0 & 1/21 & 2/21 & 0 & 1/21 & 2/21 & 1/14 & 0 & 2/21 & 2/21 \\ 1/21 & 2/21 & 1/21 & 1/21 & 0 & 0 & 2/21 & 1/21 & 0 & 2/21 \\ 2/24 & 1/24 & 2/24 & 0 & 1/24 & 2/24 & 1/24 & 1/24 & 1/24 & 0 \end{pmatrix}$$

En decimales se ve que la fila con más celdas=0 es a su vez la de valores absolutos más altos:

$$\widetilde{\mathbb{K}} = \begin{pmatrix} 0 & 0.083 & 0.042 & 0.083 & 0.042 & 0.083 & 0.042 & 0 & 0.083 & 0.042 \\ 0.042 & 0 & 0.083 & 0.083 & 0.083 & 0.083 & 0 & 0.083 & 0.042 & 0.083 \\ 0.074 & 0.037 & 0 & 0.037 & 0.037 & 0.074 & 0.037 & 0.037 & 0.074 & 0.037 \\ 0.048 & 0.048 & 0.095 & 0 & 0.048 & 0.095 & 0.095 & 0 & 0.095 & 0 \\ 0.095 & 0.048 & 0.095 & 0.095 & 0 & 0.095 & 0 & 0.095 & 0 & 0.095 \\ 0.048 & 0.048 & 0.048 & 0.048 & 0.048 & 0 & 0 & 0.048 & 0 & 0.048 \\ \boxed{0.111} & 0 & \boxed{0.111} & 0.056 & 0 & 0 & 0 & 0.083 & 0.056 & \boxed{0.111} \\ 0 & 0.048 & 0.095 & 0 & 0.048 & 0.095 & 0.071 & 0 & 0.095 & 0.095 \\ 0.048 & 0.095 & 0.048 & 0.048 & 0 & 0 & 0.095 & 0.048 & 0 & 0.095 \\ 0.083 & 0.042 & 0.083 & 0 & 0.042 & 0.083 & 0.042 & 0.042 & 0.042 & 0 \end{pmatrix}$$

Llamo $\mathbf{v}$ al $\lim\limits_{k\to\infty} \dfrac{\widetilde{\mathbb{K}}^k \mathbf{1}_{10}}{\|\widetilde{\mathbb{K}}^k \mathbf{1}_{10}\|}$, lo aproximo con k=10 y obtengo con precisión de cinco decimales

$$\mathbf{v} \approx \frac{\widetilde{\mathbb{K}}^{10} \mathbf{1}_{10}}{\|\widetilde{\mathbb{K}}^{10} \mathbf{1}_{10}\|} = \begin{pmatrix} 0.10049 \\ 0.11534 \\ 0.08916 \\ 0.10339 \\ 0.11995 \\ 0.07099 \\ 0.10458 \\ 0.10603 \\ 0.09927 \\ 0.09081 \end{pmatrix} \tag{3.1.3.5}$$

Lo que lleva a culminar que:

Ir > **Fr** > **NZ** > **SA** > **In** > **Es** > **Au** > **Ar** > **Ga** > **It**

es el ranking resultante de aplicar el ***Keener's method*** a los datos de la matriz $\mathbb{K}$.

Algunos puntos a destacar sobre el resultado obtenido:

- Francia lideraba el listado de los equipos del Seis naciones cuando aplicamos el método de Wei, sin embargo con el agregado de los enfrentamientos contra el grupo de selecciones del hemisferio sur pasó a ubicarse detrás de Irlanda
- Australia tenía récord de 2PG-1PP contra integrantes de su mismo grupo, mientras que Sudafrica registraba 1PG-1PE-1PP; sin embargo al entrar a tallar los cruces contra «el norte», Australia (#7) pasó a ubicarse detrás de Sudafrica (#4)

Hallazgos así hacen que nos preguntemos si debemos esperarlos, la realidad es que ambos desórdenes tienen su justificación aritmética, pero si vamos a analizar esto con criterio de razonabilidad, no es muy normal que si dos zonas (al estilo de las conferencias con las que lidió Keener al momento de publicar su método) traen un determinado ordenamiento de base, el añadir a $\frac{2}{3}$ de los cruces extra-grupos sea razón para que, por ejemplo, Francia e Irlanda intercambien posiciones, y además Escocia>Gales>Inglaterra cambien por Inglaterra>Escocia>Gales.

Mirando con más profundidad, tal vez deberíamos incorporar solamente los enfrentamientos Norte vs. Sur en los que haya habido 3 o más juegos disputados, algo que haría que el resultado del duelo posea mayor robustez; hagamos entonces las cuentas aceptando una cantidad mayor de celdas con valores NA = *not available*:

	Es	**Fr**	**Ga**	**In**	**Ir**	**It**	**SA**	**NZ**	**Au**	**Ar**
Es	—	1	0	1	0	1	NA	NA	NA	**0**
Fr	0	—	1	1	1	1	NA	NA	**0**	NA
Ga	1	0	—	0	0	1	**0**	NA	NA	NA
In	0	0	1	—	0	1	NA	NA	**1**	NA
Ir	1	0	1	1	—	1	NA	**1**	NA	NA
It	0	0	0	0	0	—	NA	NA	NA	NA
SA	NA	NA	**1**	NA	NA	NA	—	$1/2$	0	1
NZ	NA	NA	NA	NA	**0**	NA	$1/2$	—	1	1
Au	NA	**1**	NA	**0**	NA	NA	1	0	—	1
Ar	**1**	NA	NA	NA	NA	NA	0	0	0	—

Este cuadro incluye a cruces entre naciones europeas o entre las del hemisferio sur, más 5 de los 24 cruces posibles extra-grupos. Los 26 resultados computan 3 o más ediciones del duelo uno vs. uno.

La nueva matriz −a la que llamaré $\mathbb{K}_{rob}$, por *robusta*− se completa así:

$$\mathbb{K}_{rob} = \begin{pmatrix} 0 & 2/3 & 1/3 & 2/3 & 1/3 & 2/3 & 0 & 0 & 0 & 1/3 \\ 1/3 & 0 & 2/3 & 2/3 & 2/3 & 2/3 & 0 & 0 & 1/3 & 0 \\ 2/3 & 1/3 & 0 & 1/3 & 1/3 & 2/3 & 1/3 & 0 & 0 & 0 \\ 1/3 & 1/3 & 2/3 & 0 & 1/3 & 2/3 & 0 & 0 & 2/3 & 0 \\ 2/3 & 1/3 & 2/3 & 2/3 & 0 & 2/3 & 0 & 2/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2/3 & 0 & 0 & 0 & 0 & 1/2 & 1/3 & 2/3 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 1/2 & 0 & 2/3 & 2/3 \\ 0 & 2/3 & 0 & 1/3 & 0 & 0 & 2/3 & 1/3 & 0 & 2/3 \\ 2/3 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 \end{pmatrix}$$

Noto con $\widetilde{\mathbb{K}_{rob}}$ a la normalización de $\mathbb{K}_{rob}$, llamo $\mathbf{v}_{rob}$ al $\lim\limits_{k\to\infty} \dfrac{\widetilde{\mathbb{K}_{rob}}^{k}\mathbf{1}_{10}}{\|\widetilde{\mathbb{K}_{rob}}^{k}\mathbf{1}_{10}\|}$, y con k=10 da

$$\mathbf{v}_{rob} \approx \frac{\widetilde{\mathbb{K}_{rob}}^{10}\mathbf{1}_{10}}{\|\widetilde{\mathbb{K}_{rob}}^{10}\mathbf{1}_{10}\|} = \begin{pmatrix} 0.09559 \\ 0.10764 \\ 0.08650 \\ 0.09729 \\ 0.11757 \\ 0.06868 \\ 0.10858 \\ 0.11628 \\ 0.11226 \\ 0.08961 \end{pmatrix} \tag{3.1.3.6}$$

Queda para el final comparar los rankings resultantes de aplicar el método sobre $\mathbb{K}$, sobre $\mathbb{K}_{rob}$ y en tercer lugar el ranking oficial al 21/10/2022 publicado por World Rugby, la entidad que rige los destinos del deporte a nivel mundial, y al cual calcula usando conceptos que se verán en 3.1.4:

Orden	**Keener sobre $\mathbb{K}$**	**Keener sobre $\mathbb{K}_{rob}$**	**World Rugby ranking**
#1	*Irlanda*	*Irlanda*	*Irlanda*
#2	*Francia*	*Nueva Zelanda*	*Francia*
#3	*Nueva Zelanda*	*Australia*	*Sudafrica*
#4	*Sudafrica*	*Sudafrica*	*Nueva Zelanda*
#5	*Inglaterra*	*Francia*	*Inglaterra*
#6	*Escocia*	*Inglaterra*	*Escocia*
#7	*Australia*	*Escocia*	*Gales*
#8	*Argentina*	*Argentina*	*Argentina*
#9	*Gales*	*Gales*	*Australia*
#10	*Italia*	*Italia*	*Italia**

(*) en rigor Italia figura rankeada en el Orden = #14 por intercalarse antes los seleccionados de Japon, Samoa, Fiji y Georgia en los lugares #10 al #13, respectivamente

Como se alcanza a ver, hay listas para todos los gustos, quitando a Irlanda (siempre #1) y Argentina (#8) los otros equipos se ven afectados de maneras distintas según qué método usemos y cuáles sean los datos de entrada con los que los ejecutemos. Es además propicio calificar al método de Keener como una herramienta fácil y por sobre todo flexible al momento de usarla, ya que a diferencia del **MVR** y del **Método del Fair-Weather-Fan** donde las reglas eran bien estrictas en lo procedimental, aqui se puede jugar con la elección de los valores para medir el triunfo y la derrota en la matriz, así como en dejar más o menos celdas sin información (marcadas con la etiqueta NA); esta característica deja al método en una categoría virtuosa y en consecuencia vamos a verlo mencionado en citas como Método Keener-ajustado o Método *Keener-alike*, indicando que se conservó lo esencial del mismo, pero afectando ciertos detalles.

Referencia anexa: Dificultad creciente para elaborar un ranking a medida que aumenta el número de zonas en las que se disputa el torneo

En nuestro ejemplo de 10 equipos existen $\frac{10\times 9}{2} = 45$ *duelos posibles, el torneo de Seis naciones del norte provee información para 15 y en el del hemisferio sur se juegan otros 6 encuentros cada año. Así 21 de 45 representa el 47 % de las celdas de la matriz que son pasibles de contener información. Hay que decir que la problemática en campeonatos donde conviven muchas más zonas o subligas es más compleja, por caso, el torneo anual de Fútbol americano universitario que inspiró el desarrollo del método de Keener cuenta hoy con 10 conferencias, totalizando 123 miembros. Aquí el detalle:*

. American Athletic Conference con 11 equipos
. Atlantic Coast Conference con 14 equipos
. Big Ten Conference con 14 equipos
. Big 12 Conference con 10 equipos
. Pacific-12 Conference con 12 equipos
. Southeastern Conference con 14 equipos
. Conference USA con 14 equipos
. Mid-American Conference con 12 equipos
. Mountain West Conference con 12 equipos
. Sun Belt Conference con 10 equipos

La próxima cuenta muestra que los juegos intra-conferencias son menos que el 47 % del ejemplo que se armó con datos de Rugby; si partimos de $\frac{123\times 122}{2} = 7503$ *cruces entre los 123 miembros, la sumatoria*

$$\frac{11\times 10}{2} + \frac{14\times 13}{2} + \frac{14\times 13}{2} + \frac{10\times 9}{2} + \frac{12\times 11}{2} + \frac{14\times 13}{2} + \frac{14\times 13}{2} + \frac{12\times 11}{2} + \frac{12\times 11}{2} + \frac{10\times 9}{2} = 707$$

da algo así como el ***9.4 %*** *de las celdas de la matriz que no están en la diagonal, a priori muy poco para verificar en PF la condición de «*$\mathbb{M}$ *primitiva» que asegura la convergencia al vector-rating.*

`https://www.espn.com/college-football/schedule/_/group/80`

Más métodos de ranking de tipo espectral pueden encontrarse revisados en [Vigna, 2016].

3.1.4. Rankings basados en ratings

Ha quedado para el final de este repaso una categoría de rankings que está seguramente entre las más extendidas, me refiero a la de los **rankings basados en ratings** o scores. Estas denominaciones apuntan a la creación −con un criterio válido− de una variable accesoria, típicamente positiva, que aúne el conocimiento de la dimensión sobre la cual tenga sentido ordenar. Así cada elemento del conjunto tendrá un valor numérico asociado y el ranking que se obtenga por

esta vía surgirá de definir el orden natural entre magnitudes. Reforcemos la comprensión del método con un ejemplo por demás sencillo: si tras disputarse una X cantidad de duelos entre deportistas o conjuntos de éstos, se obtiene una tabla de posiciones ordenada por cantidad de victorias registradas y/o puntos acumulados, el orden que tenga el elemento en la escala cuantitativa determina su posición en el ranking. Ver debajo lo que ocurre con la competencia de selecciones de Rugby Union del hemisferio sur (conocida como *The Rugby Championship*, en la que se enfrentan RSA, NZL, AUS, ARG) tras completarse 10 ediciones de su realización:

País	**PJ**	**PG**	**PE**	**PP**	**BP**	**TP**
NZL	54	45	2	7	34	218
RSA	54	26	4	24	24	136
AUS	54	25	3	26	13	119
ARG	54	7	1	46	12	42

PJ=Partidos Jugados. PG=Partidos Ganados. PE=Partidos Empatados. PP=Partidos Perdidos. BP=Bonus Points. TP=Total Points.

Para el lector no adentrado en el deporte del rugby, los Bonus Points son asignados en cada partido a los equipos que se enfrentan y se otorgan por eventos particulares que se dan durante el encuentro; siempre es una unidad extra y se la otorga al ganador si éste marcó un número alto de *tries*, o bien al perdedor si éste fue derrotado por una diferencia pequeña, ej.: menor o igual a 7 goles; en los partidos empatados no se asigna Bonus Points, lo mismo ocurre si la diferencia en el tanteador es mayor que 7 goles y el ganador no anotó *tries*, etc. En el ejemplo queda claro que, tanto si eligiéramos la columna **PG**, como la **BP** o la **TP** (tal vez esta última sea la que más sentido tenga que se use, puesto que resume información de las anteriores, $\mathbf{TP} = 4\times\mathbf{PG}+2\times\mathbf{PE}+\mathbf{BP}$) vamos a reconocer que NZL se ubica siempre en el lugar #1, RSA en el #2, y AUS y ARG en el #3 y #4 respectivamente. Veamos entonces otras circunstancias que hacen menos trivial la generación de un ranking basado en una variable cuantitativa, y que se va a dar por ejemplo en un recuento de partidos en el que los participantes no se enfrentaron todos entre sí, o que incluso registran distinta cantidad de encuentros disputados.

La experiencia del Sistema Dickinson para el Fútbol americano a nivel universitario (1925−1940)

Frank G. Dickinson, quien fuera profesor en la Escuela de Economía de la Universidad de Illinois, está llamado a ser considerado el primer autor de un método de ranking «con sustento científico», al menos en el ámbito deportivo. El fútbol americano universitario se competía a lo ancho de los Estados Unidos desde 1869. En sus orígenes y ante la dificultad de trasladar los planteles de un extremo al otro del país, el deporte se practicó en torneos por *ligas de cercanía*. Esta situación llevó a que no se disputaran todos los enfrentamientos posibles en una misma temporada, complicando la designación de un equipo #1 a nivel nacional como derivación directa del recuento de partidos ganados, empatados o perdidos. Más problemas surgían del hecho de que los equipos no jugaban el mismo número de encuentros al año. La idea detrás

del Sistema Dickinson tuvo como objetivo sortear las dificultadas antes mencionadas. En una primera instancia Dickinson define que un equipo sea considerado FUERTE/*strong* en aquellos casos en que ese año haya tenido un porcentaje de victorias superior al 50 %. De esta forma no sólo se computan los resultados de cada universidad en G-E-P, sino que además se toman en cuenta los PG, PE y PP versus estos equipos distinguidos por su buena performance; para entender cómo funciona el Método Dickinson veamos los datos concretos del desempeño de 3 equipos en el año 1925:

Universidad	**PJ**	**PG**	**PE**	**PP**	**PJF**	**PGF**	**PEF**	**PPF**
Alabama	10	10	0	0	4	4	0	0
Dartmouth	8	8	0	0	4	4	0	0
Michigan	8	7	0	1	5	4	0	1

PJ=Partidos Jugados. PG=Partidos Ganados. PE=Partidos Empatados. PP=Partidos Perdidos. PJF=Partidos Jugados versus equipos Fuertes. PGF=Partidos Ganados versus equipos Fuertes. PEF=Partidos Empatados versus equipos Fuertes. PPF=Partidos Perdidos versus equipos Fuertes.

Dickinson propuso combinar las columnas de la tabla precedente para la generación del *Dickinson Football Rating*, tal cual se lo llamó por entonces. Cabe señalar que una dificultad no menor por aquellos años consistía en recopilar la información de resultados, por la ausencia de medios electrónicos que hoy permiten acceder al registro de este tipo de datos, casi en simultáneo a que se desarrollan los partidos.

Universidad	**DFR**	**Pos.**
Dartmouth	25.00	#1
Michigan	24.37	#2
Alabama	24.00	#3

DFR=Dickinson Football Rating, surgido del cálculo dado por:
$(20 \times PG + 15 \times PE + 10 \times PP + 10 \times PGF + 7.5 \times PEF + 5 \times PPF) \div PJ.$
Pos.=Posición

La información que contienen las matrices anteriores se limita a los equipos del top-3 de ese año. Aunque por **PG** le cabría a Alabama la primera posición, en rigor el Sistema la ubica #3, dando más importancia a la proporción de victorias de Dartmouth y Michigan sobre equipos «fuertes».

Ratings por acumulación de puntos. Los casos del Golf y el Tenis masculino.

Podría decirse que de cada desventaja observada surge una oportunidad de mejora. Y además de que las condiciones particulares de aplicación influyen en la adopción de una definición u otra. Yendo al caso del **Golf**, el método de ranking de jugadores que aplica desde la década de 1980 no buscó coronar al o la mejor golfista de la temporada, sino que surgió para definir quiénes

podían ingresar en la nómina de invitados a los torneos con mayor recompensa económica y/o pedigree, algo que históricamente era decisión exclusiva del ente organizador pasó a ser dado por el universo de los «mejores del ranking» a ese momento. Por otra parte es cierto que en un deporte individual es más comprensible que cada jugador elija en qué torneos participar, frente a lo que es más común en deportes de equipos, que tiene que ver con formar parte de una federación o liga en la que las reglas los hagan enfrentar a un mínimo de rivales a lo largo de la duración del torneo. Sin embargo para un contexto entonces que admite tal flexibilidad, lo que sucede es que en la misma definición del rating va a aparecer implícito el incentivo a que todos disputen una cantidad equiparable de torneos. Muestro debajo cómo se puntúa en sí para el ranking del Golf y luego daré una explicación de los términos involucrados:

$$puntos_golf = \frac{\sum_i puntos\ del\ iésimo\ torneo\ .\ \omega_i}{\mathbf{max}(\#\ de\ sumandos\ en\ el\ numerador, 40)}$$

donde el índice i recorre los torneos con mejores resultados del golfista en 2 años previos y ω_i es un ponderador entre $\frac{1}{92}$ y 1 cuyo valor crece cuanto más cerca en el tiempo fue la competencia. El denominador 92 de la fracción $\frac{1}{92}$ se eligió así para ponderar los puntos obtenidos en la semana más antigua de la ventana de últimos 2 años, la 104ª. La idea, como se explica mejor más abajo, es que a medida que la semana del puntaje se acerque a hoy, el ponderador ω crezca hasta llegar a ser 1 en la 13ª anterior al presente.

Algunas consideraciones:

1. ya no se trata de un promedio simple como ocurría con el DFR, puesto que el denominador siempre será 40 o más, haciendo que alguien que juegue menos de esos torneso en últimos 2 años salga perjudicado en el cálculo de sus puntos, e incentiva a tener una participación al menos frecuente para escalar posiciones en el ranking i.e. 20 o más torneos por año

2. los puntos que cada torneo reparte dependen de la categoría del mismo y de la posición que el golfista ocupe en la clasificación general; además la categoría es función de la dificultad del campo (da mayor cantidad de puntos cuanto más difícil es la cancha) y del nivel de los jugadores que lo jueguen (más puntos a mayor cantidad de inscriptos que figuren entre los top-200 del ranking)

3. la ponderación que se realiza con los parámetros ($\omega_.$) persigue igualar el peso de todos los torneos disputados en las últimas 13 semanas, asignando el valor 1 para los ω asociados; si pasaron exactamente 14 semanas desde que se disputó la competencia entonces $\omega = \frac{91}{92}$; si pasaron 15 el ω será $\frac{90}{92}$; y así siguiendo hasta la semana 104ª en la que se aplica la menor ponderación de todas; de este modo en elpuntaje-rating del golfista pesan más los resultados recientes respecto a los que ocurrieron más atrás en el tiempo

En el caso del **Tenis masculino**, cuyo ranking está por cumplir 50 años de su creación en 1973, se presentan varias similitudes con los puntos del Golf y algunas diferencias; vuelco debajo un esquema de cómo se lo calcula actualmente:

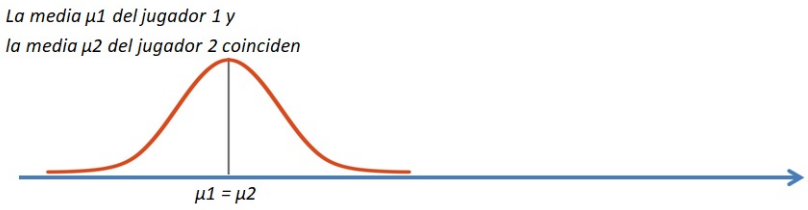
$$puntos_tenis = \sum puntos\ de\ 13\ torneos\ mandatorios + \sum puntos\ de\ 7\ torneos\ electivos$$

Ambas sumatorias se calculan sobre la base de la performance relevada en las 52 semanas previas. Por torneos mandatorios se contemplan el ATP Finals, los 4 Grand Slams y los 8 Masters 1000s que no son el Abierto de Monte-Carlo. Por torneos electivos se toman los otros mejores 7 puntajes que el tenista obtuvo en la ventana de medición. Ya no se hace ningún promedio, cuantos más puntos se recogen mejor es para la suma total − ok, pero hasta un máximo de 20, que no es un valor extremo (al 28/09/2022 hay 92 de los top-100 jugadores del ranking que jugaron 20 o más torneos en el último año). Finalmente destaco que los puntos que cada torneo reparte sólo dependen de la categoría del mismo, en nada incide el nivel de los tenistas que se inscriban, como vimos que ocurría en el anterior deporte.

Los desarrollos surgidos a partir de la creación del Sistema Elo (1960 al presente)

Hacia 1960 ya había conformada en el planeta una numerosa élite de jugadores de ajedrez, imposibles de reunir todos físicamente para enfrentarlos y designar al mejor de ellos como campeón mundial. Sin embargo un rating numérico permitió que *los mejores* clasifiquen a eliminatorias, y que sólo los ganadores accedieran a disputarle el trono al campeón reinante, una tradición que el Ajedrez comenzó en 1886 y que continúa hasta hoy. **Arpad Elo** fue un físico húngaro-estadounidense, además de aficionado al Ajedrez. Concibió en 1960 el Sistema de ranking de ajedrecistas que lleva su nombre y que se sigue usando hasta el presente en todo el mundo (por su creación A. Elo ingresó al Hall de la Fama del Ajedrez). Veamos en qué consiste:

- La premisa sobre la que Elo construyó el Sistema de ranking fue medir la performance del ajedrecista a través de una variable aleatoria normal, cuya media μ será en la práctica el rating numérico del cual surgirá el ranking, una vez ordenados todos de mayor a menor
- Para dar inicio a la implementación del rating (hoy día, en el contexto del Ajedrez, hablar de *rating* o de *elo* es lo mismo, como cuando un nombre propio adopta el sentido genérico del sustantivo al que refiere) es necesario plantear que todos los jugadores rankeables tengan un mismo μ inicial
- En su definición simple y elegante, el Rating Elo asume un desvío estándar común (σ) para las performances de los ajedrecistas $\rightarrow$ $\sigma_{\mathbf{J}} \equiv \sigma \ \forall \mathbf{J}$



En un inicio teórico todas las campanas parten ubicadas en el mismo lugar de la recta.

- A los efectos de dar una explicación más clara del método, llamaremos indistintamente $\mu_{\mathbf{J}}$ o $r_{\mathbf{J}}$ al rating asociado al jugador $\mathbf{J}$

Apoyados en los resultados de los enfrentamientos entre pares −en Ajedrez se dan 1 pt al ganador, 0 pts al perdedor, o $\frac{1}{2}$ a cada uno que empate el juego− los ratings de ambos (μ_1 y μ_2 en la imagen precedente) irán adoptando otros valores por la aplicación de la siguiente fórmula de actualización:

$$r_{\mathbf{new}} = r_{\mathbf{old}} + K(s - Es) \tag{3.1.4.1}$$

donde $r_{\mathbf{new}}$ y $r_{\mathbf{old}}$ son las versiones nueva y anterior del rating de cada uno de los jugadores que participan en una partida, K es una constante positiva, s es el resultado=*score* que obtuvo en la partida el jugador al que se le está actualizaando su rating y Es es la Esperanza a priori del score sujeta a la situación inicial de cada jugador. Algo trivial pero interesante que surge de analizar la fórmula (3.1.4.1), es que $r_{\mathbf{new}} > r_{\mathbf{old}}$ si y sólo si $s > Es$, lo que quiere decir que si un jugador va a disputar un juego y tras él desea mejorar su rating, entonces debe procurar obtener un s (score) mayor al score esperado (Es).

Paremos aquí un instante y observemos a partir de seguir varios ejemplos de cálculo del elo qué puede decirse al respecto de cada valor que internviene en la ecuación de ajuste.

(a) $r_{\mathbf{old}}$ − este parámetro tiene asignado el último rating calculado del jugador, si como venimos diciendo se debe suponer que en algún momento a todos quienes van a ser rankeados se les asignó un mismo valor, cuál resulte ser éste es indiferente mientras lo compartan todos; como referencia histórica cuando el Elo se empezó a usar en 1960 fue primeramente en los Estados Unidos, y con el propósito de que tras sucesivos ajustes los ratings se mantengan positivos, la *USCF− United States Chess Federation* fijó como valor inicial los 500 puntos; desconozco los detalles de esa experiencia pero en [Elo, 1978] se cuenta que cuando la idea permeó a las autoridades de la FIDE (federación internacional) y ésta expandió el Sistema Elo a todas las latitudes, eligió modificar los 500 del rating de U.S. en un teórico t=0 por 2200 para ajedrecistas hombres y 1900 para ajedrecistas mujeres, esta decisión auguraba que nunca iba a terminar pasando que un rating cayera por debajo de los 1000 puntos ni superara los 3000, algo que efectivamente ocurrió, más allá de que veremos que matemáticamente es posible infringir esas cotas

(b) K − conocido como el parámetro de desarrollo, en su propuesta original Elo lo seteó en 10 y el uso que se le da es el de establecer un rango para la afectación del $r_{\mathbf{old}}$ por el sumando compuesto, la resta del paréntesis es sencilla de acotar ya que $0 \leq s$, $Es \leq 1$ ergo todo el segundo término caerá en el intervalo $[-K, K]$; la denominación ligada a la evolución del desempeño del jugador guarda relación con que en alguna versión posterior del rating la FIDE fijó que el K sea variable según la cantidad de partidas que el jugador tenga en su experiencia, por ej. hasta la número 30 K es 40, luego pasa a ser 20 y una vez que el elo superare un umbral dado el K quedaría fijo en 10, la justificación detrás de adoptar una definición así está en que en el comienzo de la práctica un aficionado novel podría mejorar rápidamente su nivel, debido al potencial para aprender conceptos velozmente, con lo cual subir el K permitiría que el rating elo se acomode en menos iteraciones al valor que refleja la performance del jugador i.e. aunque el camino ascendente se mantiene, no sería necesaria la misma cantidad de partidas para verlo evidenciado numéricamente

(c) s – con el *score* no hay misterios, tomará el valor 1, 0 o $\frac{1}{2}$ según la partida haya terminado en triunfo, derrota o empate (tablas en la jerga ajedrecística) para el jugador cuyo rating se está ajustando

(d) Es – acá la notación remite en forma directa a la Esperanza matemática de la variable s (*score*), sin embargo hay una asunción concreta que se hace y que corresponde aclarar, que es que en rigor se va a asumir siempre para el cálculo de Es que $P(X_1 = X_2) = 0$, lo cual tiene sentido desde el manejo de probabilidades conjuntas para variables continuas, pero mediante ese enfoque el evento «empate» no afecta en nada al «score esperado»; así si $X_1 \sim \mathcal{N}(\mu_1, \sigma^2)$ y $X_2 \sim \mathcal{N}(\mu_2, \sigma^2)$ son variables independientes que modelan el nivel de juego de los contrincantes $\mathbf{J}_1$ y $\mathbf{J}_2$, se va plantear que $P(\mathbf{J}_1 \text{ vence a } \mathbf{J}_2) = P(X_1 > X_2)$ y viceversa; pero entonces ante el ajuste del $r_{\mathbf{J}_1}$ se tendrá que:

$$\begin{aligned}(Es)_1 &= \sum_{s\in\{0,1,\frac{1}{2}\}} s \cdot P(\mathbf{J}_1 \textit{ puntúe un score } s \textit{ versus } \mathbf{J}_2) \\ &= 1 \cdot P(X_1 > X_2) + 0 \cdot P(X_2 > X_1) + \cancel{\frac{1}{2} \cdot P(X_1 = X_2)} \\ &= P(X_1 - X_2 > 0) = 1 - P(X_2 - X_1 > 0) \\ &= 1 - (Es)_2;\end{aligned} \tag{3.1.4.2}$$

llegados a este punto cabe decir que $(Es)_1$ y $(Es)_2$ quedan determinadas por $P(X_1 - X_2 > 0)$, y si definimos $X_{1-2} := X_1 - X_2$ resulta, por propiedad de la resta entre variables normales, que $X_{1-2} \sim \mathcal{N}(\mu_1 - \mu_2, 2\,\sigma^2)$

(e) `primer ejemplo` (*trivial*) – $\mathbf{J}_1$ y $\mathbf{J}_2$ con ratings $r_{\mathbf{J}_1} = r_{\mathbf{J}_2} = 1700$ empatan y se da paso a la actualización de $r_{\mathbf{J}_1}$ y $r_{\mathbf{J}_2}$ apoyados en la fórmula (3.1.4.1) con K=10:

$$\begin{aligned}(r_{\mathbf{new}})_1 &= 1700 + 10 \cdot (1/2 - (Es)_1) \\ (r_{\mathbf{new}})_2 &= 1700 + 10 \cdot (1/2 - (Es)_2)\end{aligned} \tag{3.1.4.3}$$

y como vimos tras (3.1.4.2) que:

$$(Es)_1 = P(X_{1-2} > 0) = \frac{1}{2} = 1 - P(X_{2-1} > 0) \to (Es)_2 = P(X_{2-1} > 0) = \frac{1}{2} \tag{3.1.4.4}$$

se tiene que:

$$\begin{aligned}(r_{\mathbf{new}})_1 &= 1700 \\ (r_{\mathbf{new}})_2 &= 1700\end{aligned} \tag{3.1.4.5}$$

la causa que hace que ambos ratings mantengan sus valores es que Es y s coinciden, con lo cual se asume que $r_{\mathbf{J}_1}$ y $r_{\mathbf{J}_2}$ miden acertadamente el nivel de cada jugador al presente

(f) `segundo ejemplo` – misma situación que en el ejemplo anterior, $\mathbf{J}_1$ y $\mathbf{J}_2$ con ratings $r_{\mathbf{J}_1} = r_{\mathbf{J}_2} = 1700$, pero ahora $\mathbf{J}_1$ vence a $\mathbf{J}_2$:

$$\begin{aligned}(r_{\mathbf{new}})_1 &= 1700 + 10 \cdot (1 - 1/2) = 1700 + 5 = 1705 \\ (r_{\mathbf{new}})_2 &= 1700 + 10 \cdot (0 - 1/2) = 1700 - 5 = 1695\end{aligned} \tag{3.1.4.6}$$

(g) `tercer ejemplo` $-$ $r_{\mathbf{J}_1} = 1705$ y $r_{\mathbf{J}_2} = 1695$, con $\mathbf{J}_1$ venciendo nuevamente a $\mathbf{J}_2$:

$$\begin{aligned}(r_{\mathbf{new}})_1 &= 1705 + 10 \cdot (1 - P(X_{1-2} > 0)) \\ (r_{\mathbf{new}})_2 &= 1695 + 10 \cdot (0 - P(X_{2-1} > 0)) = 1695 - 10 \cdot (1 - P(X_{1-2} > 0))\end{aligned} \tag{3.1.4.7}$$

acá tenemos para resolver la incógnita del valor de $(Es)_1$, la cual conociendo que el Elo parte de suponer $\sigma \equiv 200$ plantea entonces hallar $P(X_{1-2} > 0)$ valiéndonos del hecho de que se trata de una distr. normal, como lo es $X_{1-2} \sim \mathcal{N}(10, 2 \cdot 200^2)$; para ello existen hoy múltiples maneras de calcular la probabilidad que requiramos, sin embargo creo que será útil dar a conocer cuál fue el atajo que se empleó 60 años atrás, consistente en elaborar *ad hoc* una tabla no exhaustiva, la misma tenía en una primera columna el valor **Rtg. Dif.** o «Diferencial de Ratings», valor que en este ejemplo toma el $10 = 1705 - 1695$, así las cosas la 2da. columna llevaba el nombre **P H** por *Probability Higher-rated* y refería a la *Es* del jugador con mayor rating, mientras que **P L** (*Probability Lower-rated*) refería a la del jugador con menor rating

D	P		D	P		D	P	
Rtg. Dif.	H	L	Rtg. Dif.	H	L	Rtg. Dif.	H	L
0-3	.50	.50	122-129	.67	.33	279-290	.84	.16
4-10	.51	.49	130-137	.68	.32	291-302	.85	.15
11-17	.52	.48	138-145	.69	.31	303-315	.86	.14
18-25	.53	.47	146-153	.70	.30	316-328	.87	.13
26-32	.54	.46	154-162	.71	.29	329-344	.88	.12
33-39	.55	.45	163-170	.72	.28	345-357	.89	.11
40-46	.56	.44	171-179	.73	.27	358-374	.90	.10
47-53	.57	.43	180-188	.74	.26	375-391	.91	.09
54-61	.58	.42	189-197	.75	.25	392-411	.92	.08
62-68	.59	.41	198-206	.76	.24	412-432	.93	.07
69-76	.60	.40	207-215	.77	.23	433-456	.94	.06
77-83	.61	.39	216-225	.78	.22	457-484	.95	.05
84-91	.62	.38	226-235	.79	.21	485-517	.96	.04
92-98	.63	.37	236-245	.80	.20	518-559	.97	.03
99-106	.64	.36	246-256	.81	.19	560-619	.98	.02
107-113	.65	.35	257-267	.82	.18	620-735	.99	.01
114-121	.66	.34	268-278	.83	.17	over735	1.00	.00

Tabla de probabilidades para completar *Es* en la ecuación (3.1.4.1) extraida de [Elo, 1978].

usando finalmente los valores de la tabla, redondeados a dos decimales, quedaban determinados ambos $r_{\mathbf{new}}$:

$$\begin{aligned}(r_{\mathbf{new}})_1 &= 1705 + 10 \cdot (1 - \mathbf{P\ H}(10)) = 1705 + 10 \cdot \frac{49}{100} = 1709.9 \\ (r_{\mathbf{new}})_2 &= 1695 - 10 \cdot \mathbf{P\ L}(10)) = 1695 - 10 \cdot \frac{49}{100} = 1690.1\end{aligned} \tag{3.1.4.8}$$

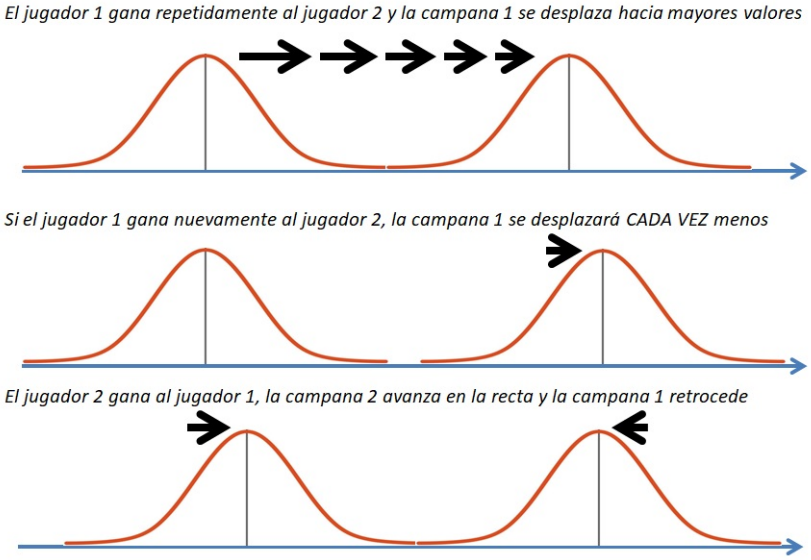
Dato anecdótico: la tabla precedente sigue en 2022 formando parte de los manuales oficiales de la FIDE (federación internacional de ajedrez), más allá que sea evidente que al redondear las probabilidades a dos decimales se está arrastrando un error que podría evitarse de aplicar un software más exacto.

(h) `invariancia de la suma de ratings` – de los cálculos de los ejemplos se deduce en forma directa que $(r_{\mathbf{new}})_1 + (r_{\mathbf{new}})_2 = (r_{\mathbf{old}})_1 + (r_{\mathbf{old}})_2$, es decir que si no se incorporan nuevos jugadores al universo de rankeables, la cuenta $\sum_{i=1}^{n} r_{\mathbf{J}_i}$ arroja el mismo resultado sin importar en qué momento se calcule (esta afirmación sobre la invariancia de la suma de ratings no es válida en la actualidad tras la adopción del esquema de parámetros K que varían según el desarrollo del ajedrecista, por eso se dice que el Sistema Elo como se lo utilza hoy día perdió esa elegancia de diseño que tuvo al ser ideado)

(i) `comentarios finales` – la construcción de Arpad Elo de 1960 sigue vigente al presente en el mundo del Ajedrez, con algunas diferencias respecto a como la diseñó originalmente; el uso de distintos factores K es de todas la que muestra un mayor impacto; además dejó de ser cierto que un jugador debutante en competencias, que pasa de no tener elo a tenerlo, inicia su recorrido con un rating fijo, cambiando eso por una asignación inicial que se calcula tomando en cuenta el rating promedio de sus primeros 5 rivales rankeados y el puntaje obtenido por el debutante frente a ellos; pero también se dio una situación en el sentido contrario, me refiero a algo que A. Elo planteó como mejora y la FIDE decidió no aplicar... tal cual se dijo antes en (g), al exhibir la tabla de probabilidades que se toma para completar *Es* en la fórmula de ajuste de ratings, estos valores surgen de integrar numéricamente la función de densidad de una distr. normal, en concreto la idea (superadora) consistía en dejar de suponer normalidad para la performance de los ajedrecistas y variar el enfoque a pensar que son distribuciones logísticas con parámetros elegidos convenientemente, así las columnas **P H** y **P L** pueden resumirse en una fórmula cerrada, usando la siguiente propuesta planteada por Elo:

$$\mathbf{P\ H} = \frac{1}{1 + 10^{-\mathbf{Rtg.Dif.}/400}} \rightarrow \mathbf{P\ L} = 1 - \mathbf{P\ H} = \frac{1}{1 + 10^{+\mathbf{Rtg.Dif.}/400}} \tag{3.1.4.9}$$

la elección de las constantes 10 y 400 permite que aparezcan algunas relaciones de proporción entre **P H** y **P L** que no existían con el abordaje vía funciones gaussianas; por ejemplo para **Rtg. Dif.** = 400 resulta que **P H**=$\frac{10}{11}$, **P L**=$\frac{1}{11}$ y **P H**/**P L**=10; mientras que si **Rtg. Dif.** = 800 resulta que **P H**=$\frac{100}{101}$, **P L**=$\frac{1}{101}$ y **P H**/**P L**=100, expresando que la relación de doblar la **Rtg. Dif.** se traduce en llevar la razón **P H**/**P L** al cuadrado, otro elemento de belleza en las ideas formuladas por el sabio físico



Esquema del cálculo dinámico de actualización de los ratings Elo bajo distintas situaciones.

Referencia anexa: Funciones de probabilidad acumulada (cpf) de familias conocidas

Cálculos como los que intervienen en la Es, del estilo de $P(X_1 > X_2)$ siendo X_1 y X_2 distribuciones de la misma familia, se trabajan en el ámbito de la Teoría de Probabilidades con el objetivo de generar funciones cerradas que retornen una amplia gama de resultados a partir de los parámetros que intervienen en el planteo. Cuando en su idea relatada en (i) Elo propone pensar que en la actualizacón del rating X_1 e X_2 siguen una distribución logísitica, sabe que entonces $P(X_1 > X_2)$ surgirá de (3.1.4.9), una fórmula sencilla que sólo requerirá operar con **Rtg.Dif**. *en el exponente de una potencia de 10 que aparece en el denominador de una fracción, cuentas todas que se obtienen con una calculadora escolar; en cambio para la formulación inicial de 1960, en la que X_1 e X_2 siguen distribuciones normales, el cálculo de $P(X_1 > X_2) = P(X_1 - X_2 > 0) = P(X_{1-2} > 0)$ necesariamente exigirá resolver una integral numérica, sujeto a que la cpf no tiene una fórmula cerrada; es así que en los orígenes del Sistema tuvo sentido contar con una tabla como la publicada en [Elo, 1978]. Lo extraño −como se destacó antes− es que la misma aún hoy permanezca vigente en los manuales oficiales de la FIDE que reglamentan el cálculo del rating elo; para el lector incrédulo adjunto el link a la norma que rige en 2022:*

`https://www.fide.com/docs/regulations/FIDE%20Rating%20Regulations%202022.pdf`

Cuando titulé este bloque de la tesis relativo al Sistema Elo anticipé que veríamos distintos desarrollos derivados del mismo, sobre esa idea es que recopilé los puntos listados debajo:

- Tras haberse extendido su uso en el Ajedrez, otras disciplinas lo adoptaron, incorporando sus propias características. Los casos más conocidos son los de las federaciones interna-

cionales de Backgammon y Scrabble, ambos juegos de tablero como el que dio pie a la formulación del Elo. Pero también hay ejemplos de aplicar estas ideas a deportes de contacto, como el fútbol soccer de 11 jugadores y el rugby de 15, cuyos entes organizadores de torneos de selecciones lo usan para definir tanto el ranking fifa como el ranking irb. Otro espacio de difusión del rating elo son los eSports, el campo en que se compite mediante videojuegos, la lista es larga y muchos de los más populares −como el juego *League of Legends* o la saga *Age of Empires*− lo usan para rankear jugadores o equipos de éstos.

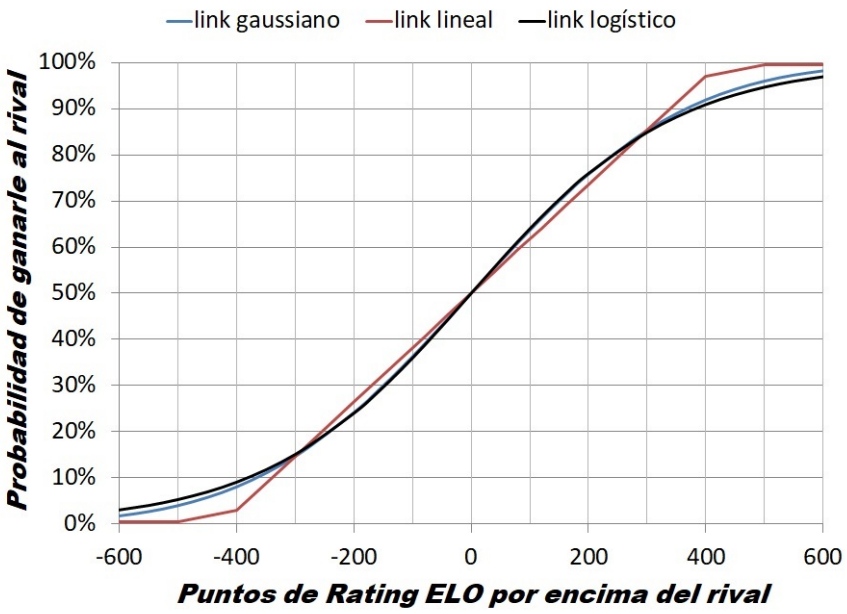
- El elo del ajedrez es también un instrumento de experimentación, en primer lugar cabe mencionar los muchos medios deportivos que eligieron llevar adelante el proyecto de aplicar este rating a diversos deportes. Los casos más conocidos son los del Elo ajustado al Béisbol, al Fútbol americano y al Básket, todo esto en las ligas profesionales de los Estados Unidos; además se conoce una implementación del Elo para rankear las selecciones de Hockey sobre hielo, en esta adaptación el método permite hacer intervenir la diferencia de goles en la fórmula de actualización, evolucionando el esquema original en que el resultado sólo es visto como Victoria, Derrota o Empate.

- Todos los casos relevados arriba, se traten de otros juegos o deportes, de práctica física u online, sí han rescatado la idea de ajustar los ratings apoyados en el enfoque de la función logística que describí antes en (i)... ¡punto para Arpad!

- El otro espacio de experimentación en el que se sigue recurriendo a este Sistema es en el ámbito de la investigación. Un ejemplo claro es el trabajo [Hvattum and Arntzen, 2010] que combina ideas presentes en este tipo de rankings con el problema de predecir resultados deportivos.

- La reconocida organización de competencias online `Kaggle.com` abrió en 2010 el concurso denominado *Chess Ratings – Elo versus the Rest of the World*. En arxiv.org/abs/1012.4571 se puede acceder a un detalle del algoritmo que resultó ganador, de nombre Elo++ y cuyo autor es Yannis Sismanis.

- Para cerrar esta enumeración querría mencionar 2 casos de mejoras que se han propuesto sobre el particular. El primero guarda relación con la función que permite linkear el diferencial de ratings con el score esperado (*Es* en la notación empleada anteriormente), la imagen que sigue da una 3ra. opción alternativa a las que se mencionaron previamente:



Opciones para vincular la Diferencia en puntos entre jugadores o equipos y la Prob. de victoria.

Aquí además del enfoque de usar distribuciones normales o logísticas se plantea aplicar una cuya gráfica se resume en una poligonal a trazos, el artículo [Sonas, 2002] recurre a ideas similares. Por último citar que según el deporte o contexto de aplicación hay desarrollos para determinar qué valores del parámetro K usar y si éste debe cambiar sólo por la experiencia del jugador como es en el Ajedrez; fueron publicadas además otas ideas en relación a incorporar más factores en la ecuación (3.1.4.1) como la propuesta de usar un parámetro para medir la «confianza» en el rating de un jugador, así lo propone el método TrueSkill® de Microsoft para los juegos de la consola X-Box. Y también surgió en [Glickman, 1995] la idea de involucrar la variable `intervalo_de_inactividad`, haciendo que cuanto más se prolongue el receso de un jugador de la práctica activa, por el solo hecho de transcurrir ese tiempo el rating termine cayendo, bajo el concepto de que se necesita mantener ritmo de competencia para no perder nivel. Este ajuste rige en el método actual que se usa en la práctica del Go bajo la modalidad *online*.

Otros casos de Rankings populares basados en ratings

Para 2 deportes de amplio raingambre entre la cultura deportiva de los Estados Unidos están vigentes sendos métodos de ranking cuyas definiciones abarcan todo tipo de variables de performance asociadas a los equipos que intervienen en la competencia:

- en el Básket universitario femenino se empezó a usar desde el año 2020 el sistema conocido como **NET**, por ser la sigla de *NCAA Evaluation Tool*; se lo emplea como herramienta de soporte para ser consultada por el Comité que determina los equipos clasificados a la instancia final del Division I Women's Basketball Championship (Campeonato de Baloncesto Femenino, División I) y en esta página la organización responde a las preguntas más

frecuentes acerca de su funcionamiento, a la vez que explica que su composición adquiere un carácter complejo por el cual no da a conocer detalles técnicos (lo expresa así: «NET ranking is derived from a more complex algorithm rather than a simple formula with relatively straightforward components») y señala que en su armado participó un equipo de la división de Servicios Profesionales en la Nube de Google: **WHAT IS THE NET?**

- La cadena deportiva ESPN instaló hace casi una década su Índice de Poder para el Fútbol americano universitario (*ESPN's **FPI***), al cual postuló como herramienta de predicción de resultados futuros, su fisonomía es la de un factor numérico por equipo que da una medida de cuál sería la diferencia en puntos en el tanteador de enfrentarse en un hipótetico duelo en cancha neutral dicha escuadra versus otra representada por los indicadores promedio de la liga, así —según se lo explica en la página que cito más abajo— intervienen en su cálculo diferentes variables que evalúan los componentes Ofensivo, Defensivo y de Equipos Especiales del equipo, e involucra además información sobre el talento del plantel, distinguiendo jugadores titulares de suplentes, antigüedad del coach en su cargo, etc.; la fuente consultada es la entrada en su blog, de nombre: **An inside look at College FPI**.

3.1.5. El método de rating Ataque-Defensa

Este método es especialmente sencillo de calcular, sin embargo requiere que los equipos o deportistas individuales tengan asociado un puntaje ***a favor***, brindado generalmente por goles o anotaciones que éstos conviertan y otro ***en contra*** dado por los tantos que reciban compitiendo frente a rivales. Casos de disciplinas que encuadran en este esquema hay por decenas, así que no va a resultar un método de poca aplicación, sino todo lo contrario.

Supongamos que i y j han competido entre sí y llamemos $\mathbf{A} = [a_{ij}]$ a la matriz de scores tal que a_{ij} es el puntaje que j generó frente a i (mientras que $a_{ij} = 0$ significa tanto que j no anotó frente a i como que no existió enfrentamiento entre ambos). En contraposición, a_{ij} puede pensarse como el número de tantos que i recibió por parte de j, es decir que según se mire la fila i o la columna j, los a_{ij} reflejan la *capacidad defensiva relativa* de i o la *capacidad ofensiva relativa* de j, respectivamente. El objetivo del **método de rating A-D** es asignar a cada elemento que deba ser rankeado tanto un rating ofensivo ($o_i > 0\ \forall i$) como un rating defensivo ($d_i > 0\ \forall i$), derivados ambos de los resultados registrados entre cualquier par de elementos i y j. Dichos ratings se interrelacionarán en el sentido en que los o_i crecen para el caso de anotar frente a rivales fuertes defensivamente; y viceversa, los d_i aumentan cuanto menos tantos hayan podido anotarle a i los rivales más fuertes ofensivamente. Así interpretados ambos ratings, es razonable preguntarse si podemos hallar $\{d_1, d_2, \ldots, d_m\}$ y $\{o_1, o_2, \ldots, o_m\}$ de forma tal que verifiquen las siguientes relaciones:

$$o_j = \frac{a_{1j}}{d_1} + \frac{a_{2j}}{d_2} + \cdots + \frac{a_{mj}}{d_m} \tag{3.1.5.1}$$

$$d_i = \frac{a_{i1}}{o_1} + \frac{a_{i2}}{o_2} + \cdots + \frac{a_{im}}{o_m} \tag{3.1.5.2}$$

Puesto que los o_j's y los d_i's dependen unos de los otros, su determinación surgirá de un proceso de redefinición sucesiva (iterativa) descripto por los siguientes 3 pasos:

I Inicializo los ratings defensivos con valores positivos arbitrarios que conformarán el vector-columna $\mathbf{d}^{(0)}$, por ejemplo

$$\mathbf{d}^{(0)} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \mathbf{e}.$$

II Reemplazo $\mathbf{d}^{(0)}$ en (3.1.5.1) y guardo su resultado en $\mathbf{o}^{(1)}$. Posteriormente y dando un paso equivalente para $\mathbf{o}^{(1)}$ y (3.1.5.2), resulta un nuevo vector-columna que notaré $\mathbf{d}^{(1)}$. Repitiendo esta sustitución alternada quedan definidas 2 sucesiones de vectores, a saber:

$$\{\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \mathbf{d}^{(2)}, \mathbf{d}^{(3)}, \ldots\} \tag{3.1.5.3}$$

$$\{\mathbf{o}^{(1)}, \mathbf{o}^{(2)}, \mathbf{o}^{(3)}, \mathbf{o}^{(4)}, \ldots\}. \tag{3.1.5.4}$$

III Dado un vector-columna $\mathbf{x}$ con componentes no nulas

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \text{ se define } \mathbf{x}^{\div} = \begin{pmatrix} 1/x_1 \\ 1/x_2 \\ \vdots \\ 1/x_m \end{pmatrix} \quad \text{(el vector de los inversos).}$$

Luego las fórmulas (3.1.5.1) y (3.1.5.2) pueden ser reescritas en notación matricial, a partir de estas dos ecuaciones

$$\mathbf{o} = \mathbf{A}'\,\mathbf{d}^{\div}, \tag{3.1.5.5}$$

y

$$\mathbf{d} = \mathbf{A}\mathbf{o}^{\div}. \tag{3.1.5.6}$$

Y las sucesiones expresadas en (3.1.5.3) y (3.1.5.4) pasan a tomar la forma de

$$\mathbf{o}^{(k)} = \mathbf{A}'\,(\mathbf{d}^{(k-1)})^{\div}, \tag{3.1.5.7}$$

y

$$\mathbf{d}^{(0)} = \mathbf{e},\ \mathbf{d}^{(k)} = \mathbf{A}\,(\mathbf{o}^{(k)})^{\div}, \tag{3.1.5.8}$$

para $k = 1, 2, 3, \ldots$.

Convergencia del proceso A-D

Si los límites $\lim\limits_{k\to\infty} \mathbf{o}^{(k)}$ y $\lim\limits_{k\to\infty} \mathbf{d}^{(k)}$ convergen, entonces

$$\mathbf{o} = \lim_{k\to\infty} \mathbf{o}^{(k)} \quad \text{y} \quad \mathbf{d} = \lim_{k\to\infty} \mathbf{d}^{(k)}$$

verifican (3.1.5.5) y (3.1.5.6), por lo que cobra sentido estudiar bajo qué condiciones las sucesiones convergen.

Es posible fijar características sobre $\mathbf{A} \in \mathbb{R}^{n\times n}$ para determinar cuándo está garantizada la convergencia y cuándo no. Pero para ello miremos un procedimiento que podremos asemejar al de la redefinición sucesiva de los $\mathbf{o}^{(k)}$ y los $\mathbf{d}^{(k)}$ descripto antes. Si llamamos

$$\xi_i = \frac{1}{\sum_{j=1}^{n} a_{ji}} \quad \text{y} \quad \mathbf{C}_1 = \begin{pmatrix} \xi_1 & 0 & \cdots & 0 \\ 0 & \xi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \xi_n \end{pmatrix}$$

entonces $\mathbf{AC}_1$ es una matriz cuadrada cuyas columnas suman 1. A la cual se le puede multiplicar por izquierda la matriz definida como

$$\mathbf{F}_1 = \begin{pmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_n \end{pmatrix} \quad \text{donde} \quad \rho_i = \frac{1}{\sum_{j=1}^{n} a_{ij}\,\xi_j}$$

y que ahora $\mathbf{F}_1(\mathbf{AC}_1)$ sea una matriz cuadrada cuyas filas sumen 1.

Estas sucesivas normalizaciones por columnas y filas utilizando matrices diagonales $\mathbf{C}_k$ y $\mathbf{F}_k$ permiten preguntarnos si la sucesión de matrices no-negativas dada por

$$\mathbf{S}_k = \mathbf{F}_k \cdots \mathbf{F}_2\mathbf{F}_1\mathbf{AC}_1\mathbf{C}_2 \cdots \mathbf{C}_k \quad \text{converge,}$$

y en tal caso (llamando $\mathbf{S} = \mathbf{lim}_{k\to\infty}\, \mathbf{S}_k$) qué aspecto tendrá $\mathbf{S}$.

Definamos primeramente esta condición sobre matrices cuadradas no-negativas:

Definición. *Una matriz $\mathbf{A} \in \mathbb{R}^{m\times m}$ no-negativa se dice que tiene **soporte total** si $\mathbf{A} \neq 0$ y cada elemento positivo de $\mathbf{A}$ cae en una diagonal estrictamente positiva de ésta. Siendo definida una **diagonal de A** como el conjunto*

$$\{a_{1\sigma_1}, a_{2\sigma_2}, \ldots, a_{m\sigma_m}\}$$

donde $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_m)$ es una permutación de los enteros $(1, 2, \ldots, m)$.

Para luego contar que en 1967 Richard Sinkhorn y Paul Knopp probaron el siguiente:

Teorema (Sinkhorn-Knopp [SK])**.** *Para cada matriz $\mathbf{A}$ no-negativa con soporte total existe una matriz $\mathbf{S}$ cuyas columnas y filas suman 1, únivocamente definida de la forma $\mathbf{S} = \mathbf{FAC}$, donde*

$\mathbf{F}$ y $\mathbf{C}$ son matrices con valores estrictamente positivos en la diagonal principal y ceros en el resto, y son únicas (salvo multiplicación por escalares). Además una forma directa de obtener las matrices $\mathbf{F}$ y $\mathbf{C}$ es llevando al límite los productos de matrices $\mathbf{F}_k \cdots \mathbf{F}_2\mathbf{F}_1$ y $\mathbf{C}_1\mathbf{C}_2 \cdots \mathbf{C}_k$, respectivamente, surgidas del proceso de normalización por filas y columnas visto antes.

La clave pasa ahora por conectar los productos de matrices que aparecen en Sinkhorn-Knopp con las sucesiones (3.1.5.7) y (3.1.5.8) que vimos al enunciar el proceso del rating Ataque-Defensa. Un resultado probado en [Govan et al., 2009] muestra por un procedimiento inductivo que llamando

$$\mathcal{C}^{(k)} = \mathbf{C}_1\mathbf{C}_2 \cdots \mathbf{C}_k \ \text{ y } \ \mathcal{F}^{(k)} = \mathbf{F}_k \cdots \mathbf{F}_2\mathbf{F}_1 \quad \text{siendo } \mathbf{C}_j \text{ y } \mathbf{F}_j \text{ las del Teorema SK,}$$
$$(\mathbf{c}^{(k)})_i = (\mathcal{C}^{(k)})_{ii} \ \text{ y } \ (\mathbf{f}^{(k)})_i = (\mathcal{F}^{(k)})_{ii} \quad \text{para } k = 1, 2, 3, \ldots;$$

entonces

$$\mathbf{c}^{(k)} = (\mathbf{A}'\,\mathbf{f}_{k-1})^{\div} \ \text{ donde } \mathbf{f}_0 = \mathbf{e} \quad \text{y} \quad \mathbf{f}^{(k)} = (\mathbf{A}\,\mathbf{c}_k)^{\div} \quad \text{para } k = 1, 2, 3, \ldots.$$

Así se observa que

$$\mathbf{o}^{(k)} = (\mathbf{c}_k)^{\div} \ \text{ y } \ \mathbf{d}^{(k)} = (\mathbf{f}_k)^{\div}.$$

Luego como por el Teorema SK los límites

$$\lim_{k\to\infty} \mathcal{C}^{(k)} = \mathbf{C} \quad \text{y} \quad \lim_{k\to\infty} \mathcal{F}^{(k)} = \mathbf{F}$$

existen si y sólo si $\mathbf{A}$ tiene soporte total, consecuentemente los límites

$$\lim_{k\to\infty} \mathbf{o}^{(k)} = \lim_{k\to\infty} (\mathbf{c}^{(k)})^{\div} = \mathbf{C}^{-1}\,\mathbf{e} \quad \text{y} \quad \lim_{k\to\infty} \mathbf{d}^{(k)} = \lim_{k\to\infty} (\mathbf{f}^{(k)})^{\div} = \mathbf{F}^{-1}\,\mathbf{e}$$

existirán si y sólo si $\mathbf{A}$ tiene soporte total.

Esta condición para $\mathbf{A}$ matriz de scores no está garantizada *per se*, sin embargo es posible afectar a $\mathbf{A}$ por una pequeña perturbación para garantizar que se esté bajo las hipótesis del Teorema; de esta forma:

$$\mathbf{P} = \mathbf{A} + \epsilon\mathbf{e}\mathbf{e}'.$$

Esta perturbación suma a cada celda de $\mathbf{A}$ un valor suficientemente chico de ϵ que no tiene efecto significativo en la definición del método y en su convergencia. Con lo que hay una buena definición de los ratings o_j's y d_i's del Método de rating A-D que queda expresada así:

Definición del Método de rating A-D

Las coordenadas de los vectores $\mathbf{o}$ y $\mathbf{d}$ resultantes de llevar al límite cuando $k \to \infty$ las expresiones (3.1.5.7) y (3.1.5.8), cumplen las relaciones planteadas en (3.1.5.1) y (3.1.5.2) para el caso de $\mathbf{A}$ con soporte total.

Y aunque este método devuelva dos ratings, uno como descripción de la capacidad ofensiva y otro como descripción de la defensiva, en muchas aplicaciones es preferible contar con un único

rating; la manera más simple de realizar esa síntesis es definir a $r_i = o_i/d_i$, de forma que al dividir el rating ofensivo sobre el rating defensivo se cumplirá que «cuanto mayor sea r_i, mejor rankeará i». Existen de todos modos otras y variadas formas de combinar dos ratings, sustancialemente se lo hace construyendo sendos rankings y luego procediendo a practicar alguno de los métodos de ensamble que se describen en 3.1.6.

El desarrollo del método de rating A-D aparece explicado por primera vez en [Govan et al., 2009] con aplicaciones a equipos de Básket y Fútbol americano, a su vez una adaptación del mismo que involucra pesos diferenciales para las performances de local y visitante y con una aplicación al Fútbol de la Liga Profesional Española se encuentra en [Liern and Sala, 2009].

3.1.6. Ensambles de rankings

Con la noble esperanza de combinar entre sí a más de un ranking para obtener un nuevo ordenamiento y que éste *supere* (con lo que ello pueda significar dado el contexto adecuado) a cada uno por separado, es que se diseñan **Ensambles de rankings**.

Propongo ir directamente a conocer el conjunto de datos y el desafío que éstes nos plantea. Motivados por cuestiones dispares, una serie de medios electrónicos ha rankeado a los mejores boxeadores de la historia, algunas de estas fuentes extendieron ese ordenamiento a los «primeros 100 lugares» de la lista, otros no llegaron a tanto, pero para el ejemplo nos limitaremos a cinco (5) listas rankeadas del puesto #1 al #10:

Orden	**Box Rec**	**ESPN**	**SportCo**	**The Ring**	**Yardbarker**
#1	*Robinson, R.*	*Robinson, R.*	*Ali*	*Robinson, R.*	*Robinson, R.*
#2	*Mayweather*	*Ali*	*Robinson, R.*	*Louis*	*Ali*
#3	*McLamin*	*Armstrong*	*Mayweather*	*Ali*	*Armstrong*
#4	*Leonard, B.*	*Louis*	*Pacquiao*	*Canzoneri*	*Greb*
#5	*Ali*	*Pep*	*Tyson*	*Griffith*	*Pep*
#6	*Monzon*	*Duran, R.*	*Louis*	*Mayweather*	*Duran, R.*
#7	*Williams, H.*	*Leonard, B.*	*Leonard, R.*	*Pep*	*Louis*
#8	*Hagler*	*Johnson, J.*	*JonesJr.*	*Charles*	*Leonard, B.*
#9	*Greb*	*Dempsey*	*Marciano*	*Pacquiao*	*Leonard, R.*
#10	*Armstrong*	*Langford*	*Dempsey*	*Moore*	*Hagler*

Ver 4 para profundizar sobre la autoría de cada lista ordeanda.

Una salvedad que cabe aclarar es que los métodos de ensamble que veremos ignoran por completo el origen de cada ranking indiviudal, pudiendo ser éstos producto de algún ordenamiento guiado por métodos científicos, o provenir directamente de un recuento de votos entre los periodistas especializados que conforman el staff de los medios que publicaron la lista ordenada.

El método de de Borda

Esta técnica surge en el s.XVIII como forma de computar el escrutinio de candidatos votados según una lista ordenada y quien la postuló fue Jean-Charles de Borda en el marco de un tratado metodológico sobre temas electorales. De Borda asigna puntajes enteros en forma descendiente a medida que se recorre el ranking (en el ejemplo le otorga 10 puntos al #1, 9 puntos al #2, hasta 1 punto al #10), viéndose así la cosa para la columna de **Box Rec**:

Orden	**Box Rec**	**Puntaje_BR**
#1	*Robinson, R.*	10
#2	*Mayweather*	9
#3	*McLamin*	8
#4	*Leonard, B.*	7
#5	*Ali*	6
#6	*Monzon*	5
#7	*Williams, H.*	4
#8	*Hagler*	3
#9	*Greb*	2
#10	*Armstrong*	1

Así queda el ranking agregado (RA) resultante que surge de ordenar de mayor a menor la SUMA de los puntajes asignados según cada uno de los 5 rankings:

Orden	**RA de Borda**	**Ptje_BR+Ptje_ESPN+Ptje_SportCo+Ptje_TR+Ptje_Yardbarker**
#1	*Robinson, R.*	49
#2	*Ali*	42
#3	*Louis*	25
#4	*Mayweather*	22
#5	*Armstrong*	17
#6	*Pep*	16
#7	*Leonard, B.*	14
#8	*Duran, R.*	10
#9 (*emp.*)	*Greb*	9
#9 (*emp.*)	*Pacquiao*	9

Adicionalmente existen maneras de robustecer éste u otros ensambles de rankings y −nuevamente− la idea se presenta de la mano de pensar que cada ranking puede provenir de una votación preferencial i.e. con un orden dado entre los elegidos por quien sufraga. Una forma muy sencilla es ignorar el mejor orden de rankeo que tuvo cada listado. Usando ese mecanismo simple tendríamos a la «matriz de listas» con estas ausencias:

Orden	Box Rec	ESPN	SportCo	The Ring	Yardbarker
#1	~~*Robinson, R.*~~	*Robinson, R.*	~~*Ali*~~	*Robinson, R.*	*Robinson, R.*
#2	~~*Mayweather*~~	*Ali*	*Robinson, R.*	~~*Louis*~~	*Ali*
#3	~~*McLamin*~~	~~*Armstrong*~~	*Mayweather*	*Ali*	*Armstrong*
#4	~~*Leonard, B.*~~	*Louis*	~~*Pacquiao*~~	~~*Canzoneri*~~	~~*Greb*~~
#5	*Ali*	~~*Pep*~~	~~*Tyson*~~	~~*Griffith*~~	*Pep*
#6	~~*Monzon*~~	~~*Duran, R.*~~	*Louis*	*Mayweather*	*Duran, R.*
#7	~~*Williams, H.*~~	*Leonard, B.*	~~*Leonard, R.*~~	*Pep*	*Louis*
#8	~~*Hagler*~~	~~*Johnson, J.*~~	~~*JonesJr.*~~	~~*Charles*~~	*Leonard, B.*
#9	*Greb*	~~*Dempsey*~~	~~*Marciano*~~	*Pacquiao*	*Leonard, R.*
#10	*Armstrong*	~~*Langford*~~	*Dempsey*	~~*Moore*~~	*Hagler*

Los boxeadores fueron tachados una sola vez cada uno en el ranking donde mostraron su mejor orden.

Con estas eliminaciones hechas ex profeso el resultado del ensamble pasa a ser algo distinto al que vimos antes:

Orden	de Borda ROB	`Ptje_BR+Ptje_ESPN+Ptje_SportCo+Ptje_TR+Ptje_Yardbarker`
#1	*Robinson, R.*	39
#2	*Ali*	32
#3	*Louis*	16
#4	*Mayweather*	13
#5	*Pep*	10
#6	*Armstrong*	9
#7	*Leonard, B.*	7
#8	*Duran, R.*	5
#9 (*emp.*)	*Greb*	2
#9 (*emp.*)	*Leonard, R.*	2
#9 (*emp.*)	*Pacquiao*	2

(en el propio ámbito deportivo, que es de donde estoy tomando los ejemplos para exponer la temática de esta tesis, se utilizan *métodos de robustez* emparentados por similitud con el antes descripto y éstos se emplean en deportes donde el ganador se define por el puntaje de jueces; así en la Gimnasia un panel examinador evalúa la destreza llevada a cabo por el atleta y de todos sus sufragios se eliminan el mejor y el peor para con la suma o promedio medir su performance global; en tanto en los Saltos de clavado se hace algo análogo omitiendo incluir en el resultado final a la mejor y segunda mejor evaluación, y al último y anteúltimo puntaje − hasta ahí lo relevado por mi cuenta, deben haber más ejemplos ya que el número de disciplinas es cuantioso).

Ensambles según criterios basados en *pairwise comparisons*

Otras dos miradas interesantes para definir un ensamble criterioso entre rankings surgen de considerar a las listas ordenadas como registros implícitos de **duelos entre pares** (en nuestro

caso y curiosamente, dichos duelos podrían ser entre boxeadores que no fueron contemporáneos, menos aún podemos pedirles que hayan boxeado en su tiempo en la misma categoría por peso).

El primero de los casos lo voy a ejemplificar tomando los boxeadores ubicados en los primeros 6 lugares del último de los rankings agregados vistos en esta subsección, a partir del cual pueden definirse 15 duelos ficticios, entre pares: #1 vs. #2, #1 vs. #3, hasta #5 vs. #6. Con estos emparejamientos considerados y volcados a matrices cuadradas de resultados, es posible aplicarles luego algunos de los métodos comentados en 3.1. Usemos las diferencias entre órdenes para determinar el score que se asociará a cada uno de los duelos, por caso el de *Robinson, R.* vs. *Ali*:

Orden	**Box Rec**	**ESPN**	**SportCo**	**The Ring**	**Yardbarker**
#1	*Robinson, R.*	*Robinson, R.*	*Ali*	*Robinson, R.*	*Robinson, R.*
#2		*Ali*	*Robinson, R.*		*Ali*
#3				*Ali*	
#4					
#5	*Ali*				
...	...				

- En el ranking **Box Rec** Robinson, R. «ganó» 4 puestos
- En el ranking **ESPN** Robinson, R. «ganó» 1 puesto
- En el ranking **SportCo** Ali «ganó» 1 puesto
- En el ranking **The Ring** Robinson, R. «ganó» 2 puestos
- En el ranking **Yardbarker** Robinson, R. «ganó» 1 puesto

Sumando los puntos ganados de uno sobre el otro (los mismos que el otro perdió contra uno) se obtiene un primer *resultado* de los 15 que conformarán la matriz cuadrada, así:

	Robinson, R.	**Ali**	**Louis**	**Mayweather**	**Armstrong**	**Pep**
Robinson, R.	·	$8-1$				
Ali	$1-8$	·				
Louis			·			
Mayweather				·		
Armstrong					·	
Pep						·

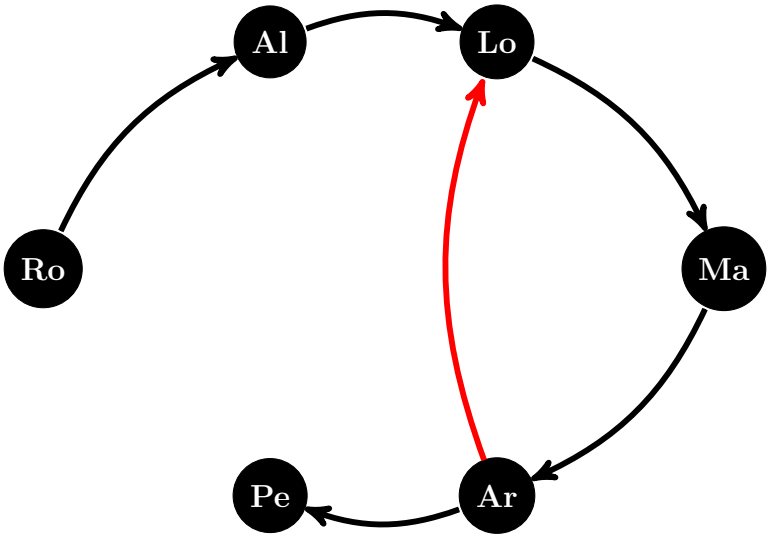
Los otros 14 scores se completan de forma equivalente, con la sola salvedad de que no en todas las listas aparecen rankeados ambos boxeadores. Un recurso válido en dicho caso es forzar el agregado del rankeado ausente como nro. 11 y derivar de esta acción una diferencia que no habría de antemano en las listas; y si ambos no están en un ranking −como Louis y Pep en la lista de Box Rec− entonces se puede evitar ese listado y obtener un resultado sustentado en la información que proveen los otros cuatro.

Completada la matriz sólo restará aplicar alguno de los métodos vistos anteriormente, a saber: el del Fair-Weather-Fan, el *Keener's method* o el del rating Ataque-Defensa, indistintamente.

Alternativamente a lo realizado previamente, es posible quedarnos simplemente con valores tomados del rango (1, -1, 0), según se trate de un resultado «a favor», «en contra» o «empate» entre dos de los rankeados, para resolver luego el ensamble a través del método MVR – ver 3.1.1:

	Robinson, R.	**Ali**	**Louis**	**Mayweather**	**Armstrong**	**Pep**
Ro = Robinson, R.	0	**1**	**1**	**1**	**1**	**1**
Al = Ali	**-1**	0	**1**	**1**	**1**	**1**
Lo = Louis	**-1**	**-1**	0	**1**	**-1**	**1**
Ma = Mayweather	**-1**	**-1**	**-1**	0	**1**	**1**
Ar = Armstrong	**-1**	**-1**	**1**	**-1**	0	**1**
Pe = Pep	**-1**	**-1**	**-1**	**-1**	**-1**	0

Matriz de adyacencia cuyas interacciones surgen del «signo del score ficticio» de uno frente al otro.



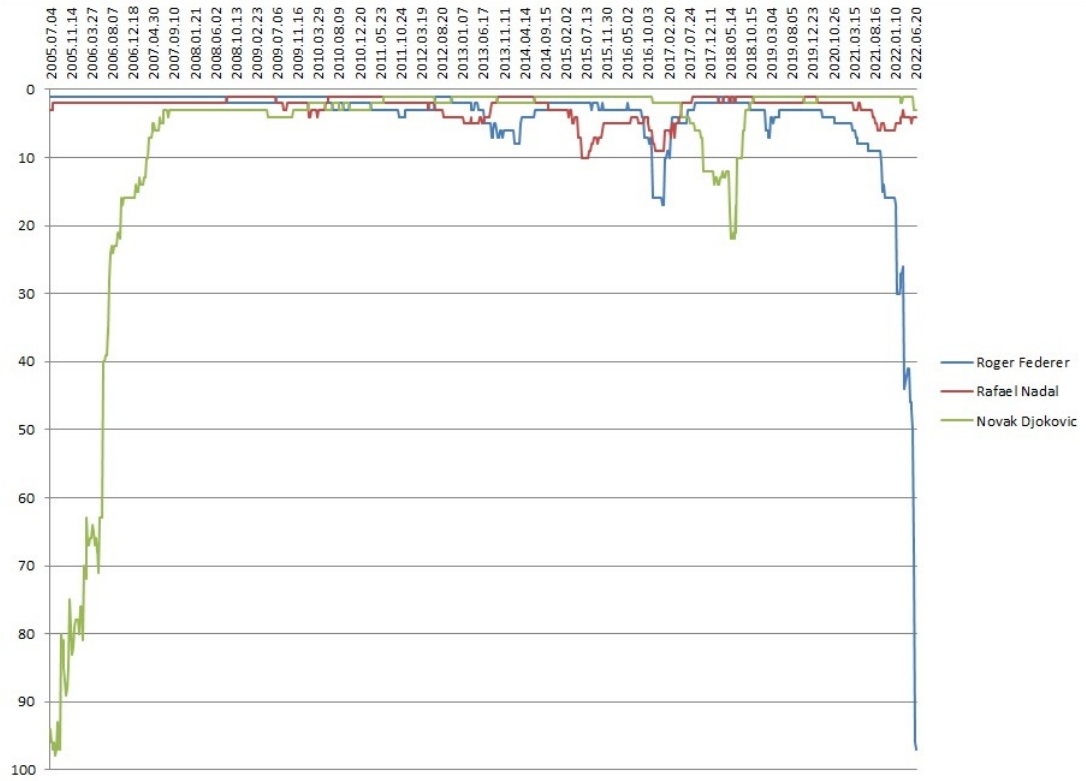
Grafo asociado al ranking devuelto por el método MVR con la violación señalada por la flecha roja.

Más métodos de ensambles de rankings pueden hallarse en [Langville and Meyer, 2012].

3.2. Visualizaciones de rankings

Más y más seguido se escucha presentar análisis estadístico-cuantitativos con la frase introductoria: *VEAMOS qué nos muestran estos datos...* ante lo cual se procede a exponer un gráfico o mapa o construcción visual que allane el camino de la interpretación. Con ese fin he relevado las ideas publicadas en el terreno de lo que son las **visualizaciones de rankings**, descubriendo en esa tarea que hay poco hecho (y que tal vez haya ahí un espacio fértil para nuevas creaciones).

Lo primero a concluir es que cuando se quiere mostrar de forma exhaustiva el ordenamiento de los elementos de un conjunto, el objeto «cuadro» o «tabla» satisface mejor que cualquier gráfico dichlosa expectativa. Sin embargo, hay una visualización básica en ejes del plano, muy extendida cuando entra en juego la variable `tiempo`, y que permite mostrar el progreso de la ubicación de uno o varios elementos dentro del ranking − aquí debajo se ve un clarísimo ejemplo de lo anterior: el eje vertical sirve para identificar el orden absoluto de cada jugador en el ranking de tenistas, mientras que en el eje horizontal se suceden las 800 publicaciones de rankings que hizo la ATP entre jul-2005 y jun-2022 (elegí en este caso tomar únicamente la información de Federer, Nadal y Djokovic por lo que representan y porque permanecieron esos 17 años dentro de los «top 100»).

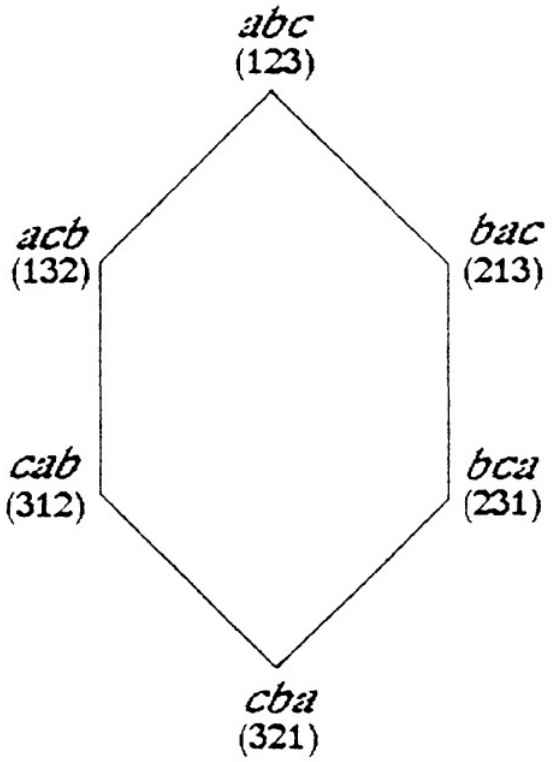


Visualización 1: Ubicaciones de Federer, Nadal y Djokovic en el ranking (jul-2005 a jun-2022).

En primer lugar analicemos el gráfico desde lo que nos provee, y luego enunciaremos las críticas:

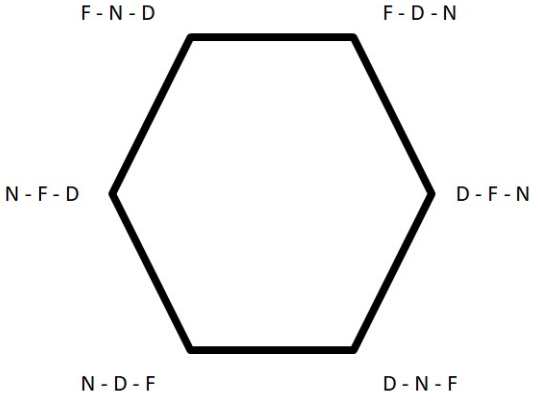
- No resume información, todo lo que pasó en esas 800 semanas está ahí
- La diferente identificación de las líneas por su color ayuda a distinguir los 3 recorridos
- Es positivo que aunque de una publicación (semanal) a la siguiente los cambios son discretos, las evoluciones de los rankings de cada tenista se vean como trazos continuos, lo que permite interpretar mejor las trayectorias crecientes/decrecientes u oscilantes

Contrariamente al valor de lo anterior, todo lo que pasa entre las posiciones #1 a la #5 está demasiado apretujado visualmente, más allá de la diferenciación por colores no podríamos deducir quién superó a quién durante más semanas, ni tampoco identificar las rachas más largas, en qué momento se interrumpieron, o cuántas permutaciones del orden entre los 3 hubo... claramente en esto se ven las limitaciones de este tipo de gráfico de líneas, pero por suerte aparecen otras construcciones como la que los autores Peter McCullagh y G. L. Thompson hicieran en 1993 y de la que da cuenta el volumen recopilatorio [Alvo and Philip, 2014]. La visualización que sigue lleva el nombre de **Politopo de permutaciones para** $k = 3$ y más adelante describiré los usos que provee para mostrar datos de rankings:

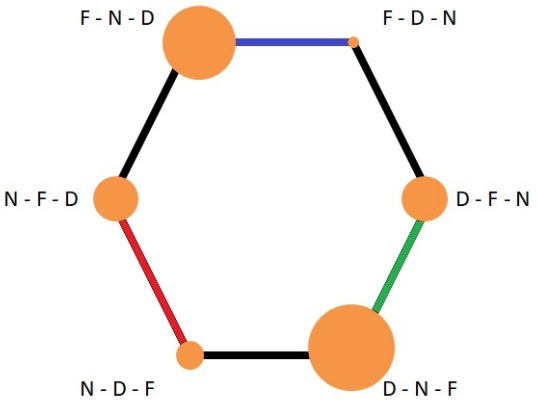


Visualización 2: Politopo de permutaciones para $k = 3$ diseñado por McCullagh y Thompson.

El objeto remite a un simple hexágono con etiquetas en los vértices y aunque le está faltando añadir información relativa al caso de análisis (la evolución de rankings de tres tenistas durante 800 semanas) es posible identificar 2 propiedades que harán que la visualización resulte provechosa: la primera es que cada vértice lleva por nombre un ordenamiento de las letras ***a***, ***b*** y ***c*** que es el INVERSO al de su simétrico por el centro de la imagen, así si en el vértice superior se lee ***abc***, en el extremo inferior se ubica ***cba***. La segunda es que recorriendo el borde en un sentido u otro vamos a encontrar contiguas a las 2 etiquetas que comienzan con ***a***, en otro lugar estarán juntas las 2 que inician con ***b*** y lo mismo va a ocurrir con las permutaciones ***cab*** y ***cba***, que van a aparecer adyacentes en el diagrama. Avancemos entonces con lo que se necesita para completar el esquema, en primer lugar daremos una asignación a las letras ***abc*** según lo que el caso demanda, llamándolas `FND` por `F`ederer - `N`adal - `D`jokovic según se ve en la siguiente figura:



Lo siguiente será superponer 6 círculos, uno por vértice, cuyos centros sean los vértices mismos y con longitudes de sus diámetros en proporción directa al tiempo medido en publicaciones semanales que el ordenamiento de los 3 tenistas en el ranking coincidió con el de la etiqueta:



Visualización 3: Versión final del Politopo de permutaciones para los *Rankings de 3 tenistas*.

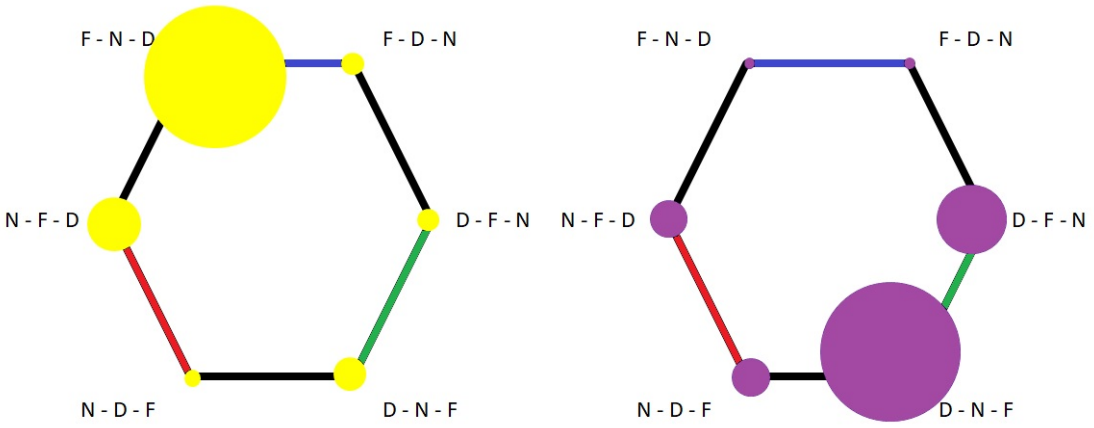
Las lecturas que podemos hacer de esta construcción visual se resumen en:

- Las «bolas» de mayor tamaño se dan en vértices opuestos, lo cual señala que ambas

configuraciones comparten al tenista rankeado entremedio de los otros (`N` en este caso)

- El lado del hexágono coloreado que concentra mayor área de los círculos asociados a él es el verde, del tenista `D`, indicando que él fue quien más tiempo estuvo arriba de los otros a lo largo de las 800 semanas del análisis
- Los vértices opuestos cuyos círculos asociados totalizan menor área son −por amplia diferencia− los identificados como F-D-N y N-D-F, lo cual señala que Djokovic ocupó significativamente menos semanas la posición intermedia que lo que estuvo *arriba* o *debajo*

Lo conciso y concreto de gráficos como el anterior ayuda para que al mostrar en simultáneo dos construcciones, éstas no se traduzcan en algo inentendible para el lector, la lógica de esta duplicación fue la de dividir el periodo de análisis en 2 tramos de 400 semanas cada uno, el primero que va de jul-2005 a jun-2013, y su complemento de jul-2013 a jun-2022; así se las ve:



Visualización 4: Politopos de permutaciones para **Era Federer** (izq) y **Era Djokovic** (der).

Algunos hallazgos interpretativos de esta visualizacióon que vale mencionar son:

- En el gráfico de la ventana temporal jul-2005 a jun-2013 (bolas amarillas) hay prácticamente un único vértice que concentra la mayor participación de los 6, el de la etiqueta F-N-D
- De igual manera, en el gráfico de jul-2013 a jun-2022, el vértico que más área reúne es el D-N-F seguido por el D-F-N, ambos encabezdos por el tenista `D`
- De lo observado anteriormente es que podría llamarse al periodo 2005−2013 como **Era Federer** y al 2013−2022 como **Era Djokovic**; por el contrario, el segmento coloreado en rojo y que aúna los dos vértices en los que Nadal está primero en ambas permutaciones, es el que menos cambios presenta de todos, mostrando que es un tenista que a nivel ranking no se destacó particularmente en una ventana sobre otra
- Finalmente, coincidir en algo evidente, que es que así como la organizacióon de lados del hexágono con bolas de diámetros variables (una por vértice → una por orden relativo) permite focalizar en cuánto tiempo más tuvo un tenista un ranking superior al de otro,

o qué configuración del orden de los 3 se extendió por más o menos tiempo, hay otros detalles que se pierden respecto a la Visualización 1, de la que dijimos antes que «todo lo que pasó en esas 800 semanas está ahí»

NOTA: Hay definido un Politopo de permutaciones para $k = 4$, idea también de los autores que cité antes, toma la forma de un poliedro de 3 dimensiones, 24 vértices, 36 aristas y 14 caras, sobre él diría que no logra allanar la interpretación de datos de rankings sino que es más bien una construcción poco práctica, y que tuvo su lógica verificar que existe como tal y que posee las mismas propiedades que el hexágono que conté arriba.

Repasemos ahora otro modo de visualizar rankings yendo a un método publicado en 1972 en el marco del estudio de preferencias dadas por jueces-votantes sobre un conjunto de items fijos, aquí vamos a sustituir los diferentes juicios-votos por la historia de publicaciones de ordenamientos a lo largo del tiempo. Si antes con los *Rankings de los tenistas FND* recabamos información de 800 semanas, que para la visualización del Politopo$_{k=3}$ se mostraron agrupadas en dos eras, aquí plantearemos el problema de consolidar la información de 12 campeonatos de marcas de autos en la categoría Turismo Carretera del automovilismo argentino, más precisamente los que se disputaron entre 2010 y 2021 inclusive. El método en cuestión lleva el nombre de MDPREF, está disponible en el lenguaje R en el paquete `pmr`: **Probability Models for Ranking Data** y es uno más del grupo de técnicas de reducción de la dimensionalidad para simplificar una representación con baja o nula pérdida de información (ver 2.2.3 en [Alvo and Philip, 2014]).

Respecto a los datos en sí, de las 4 marcas que se disputan el campeonato, sólo 2 lograron salir campeonas en dicho periodo −las históricas `Chevrolet` y `Ford`− mientras que las otras se ubicaron en puestos #2, #3 y #4 según la temporada; con este detalle que se lee #1>#2>#3>#4:

- `Chevrolet>Ford>Dodge>Torino` ... años `2010`, `2011`, `2016`
- `Ford>Chevrolet>Torino>Dodge` ... años `2012`, `2021`
- `Ford>Torino>Chevrolet>Dodge` ... años `2013`, `2020`
- `Chevrolet>Dodge>Ford>Torino` ... año `2014`
- `Ford>Chevrolet>Dodge>Torino` ... año `2015`
- `Chevrolet>Torino>Ford>Dodge` ... año `2017`
- `Chevrolet>Dodge>Torino>Ford` ... año `2018`
- `Chevrolet>Ford>Torino>Dodge` ... año `2019`

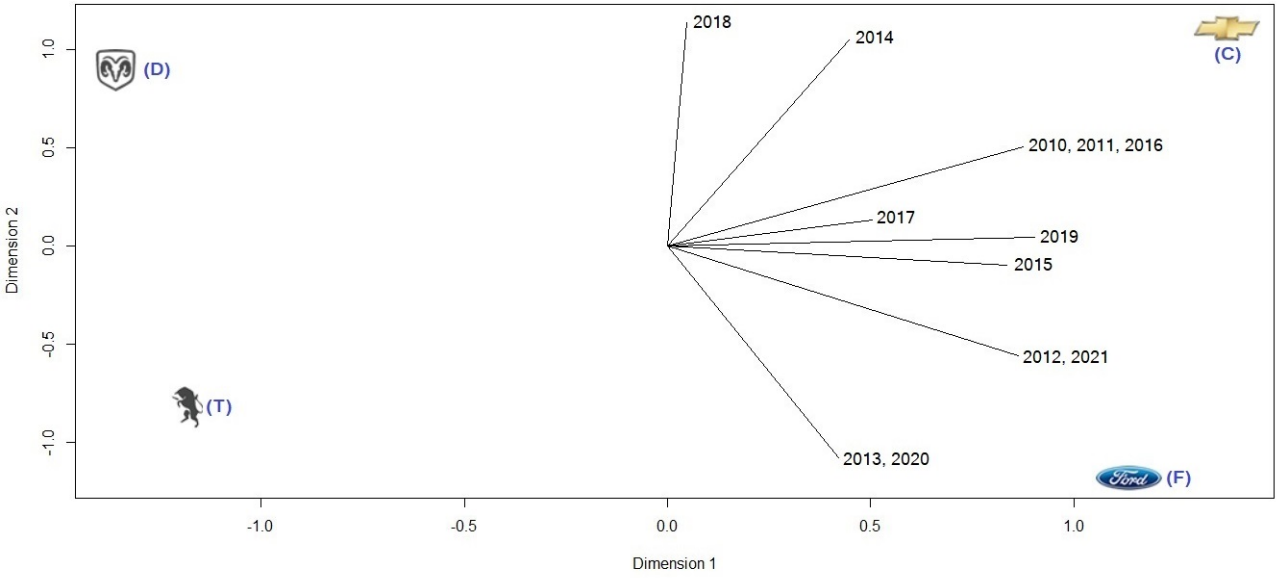
El gráfico bi-dimensional que se obtiene por el método MDPREF resulta ser una conjunción de puntos (para las *marcas*) y vectores (para los *años*). Elegidos de manera que las proyecciones de puntos sobre vectores estén lo más cerca posible al registro que dan los datos. En los casos de técnicas de reducción de la dimensionalidad hay una parte de la varianza de los datos en dimensión $\#marcas \times \#años$ que se pierde al «bajar a» trabajar en rango 2, pero mientras se observe que el % de la varianza que se deja de explicar sea bajo, la utilidad del gráfico

se sostiene. El método MDPREF va a quedarse con λ_1 y λ_2, los autovalores de mayor valor absoluto de la matriz de preferencias $\boldsymbol{Y} = [\boldsymbol{y}^{(1)} \cdots \boldsymbol{y}^{(12)}]'$ donde cada $\boldsymbol{y}^{(j)}$ tiene la forma de un vector de 4×1 conteniendo los órdenes de las marcas para cada uno de los 12 años del análisis. Y tanto las marcas como los años se mostrarán en el plano usando las coordenadas resultantes de factorizar a $\boldsymbol{Y}$ como $\boldsymbol{Y} = \boldsymbol{P}\,\boldsymbol{\Lambda}\,\boldsymbol{Q}$ (donde $\boldsymbol{r} := rango(\boldsymbol{Y})$, $\boldsymbol{P}$ es matriz ortogonal en $\mathbb{R}^{12\times \boldsymbol{r}}$, $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \cdots, \lambda_{\boldsymbol{r}})$ con $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{\boldsymbol{r}}$ y $\boldsymbol{Q}$ es matriz ortogonal en $\mathbb{R}^{\boldsymbol{r}\times 4}$). Así quedándonos solamente con los autovalores más grandes, λ_1 y λ_2, para formar $\boldsymbol{\Lambda}_2 = diag(\lambda_1, \lambda_2)$, y tomando de $\boldsymbol{P}$ las 2 primeras columnas para crear $\boldsymbol{P}_2$ y las 2 primeras filas de $\boldsymbol{Q}$ para conformar $\boldsymbol{Q}_2$; resulta que $\boldsymbol{P}_2\,\boldsymbol{\Lambda}_2\,\boldsymbol{Q}_2$ es considerada una aproximación óptima de $\boldsymbol{Y}$ en el conjunto de las matrices en $\mathbb{R}^{12\times 4}$ con rango $\boldsymbol{r}$ o menor.

Finalmente, los 4 objetos *marcas* son ploteados en el plano según las coordenadas que dictan las columnas de $\boldsymbol{Q}_2$, mientras que las 12 filas de $\boldsymbol{P}_2\,\boldsymbol{\Lambda}_2$ dan los vectores que representan a los *años*.

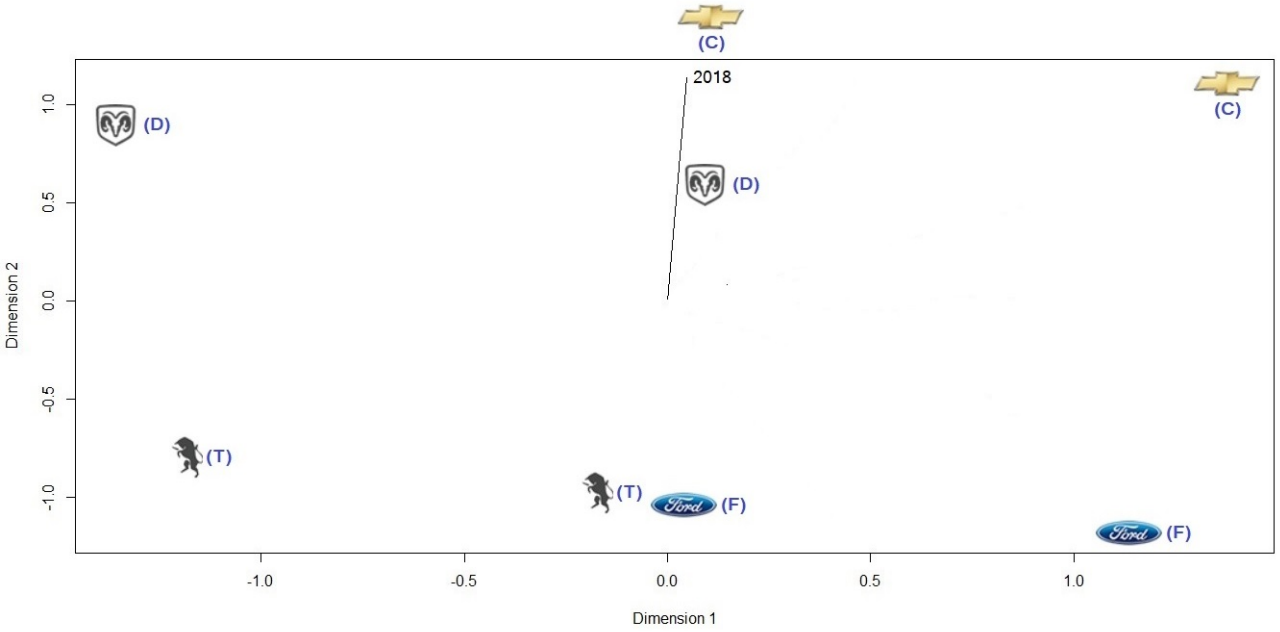
En nuestro ejemplo, proyectar la información de la matriz $\boldsymbol{Y} \in \mathbb{R}^{12\times 4}$ de rango $\boldsymbol{r}$ a $\mathbb{R}^2$ permitió seguir explicando aproximadamente el 80.05 % de la varianza total. Y por ello se puede comprobar que con muy bajo error las proyecciones perpendiculares de los puntos-marcas sobre los vectores-años devuelven posiciones relativas según cada rankings dicta. Lo que sigue es el código en R ejecutado, la salida y el plot con mejoras de identificación:

```
> mdpref(camptc,rank.vector=TRUE)

$item
          [,1]       [,2]
[1,]  1.384035  1.1350155
[2,]  1.143687 -1.1870327
[3,] -1.347511  0.9080411
[4,] -1.180210 -0.8560239

$ranking
     [,1] [,2] [,3] [,4] [,5]       [,6]        [,7]
[1,]    1    2    3    4    3 0.61319408  0.35301842
[2,]    2    1    4    3    2 0.60439747 -0.39089947
[3,]    3    1    4    2    2 0.29560127 -0.75338814
[4,]    1    3    2    4    1 0.31319450  0.73444762
[5,]    2    1    3    4    1 0.58425042 -0.06973375
[6,]    1    3    4    2    1 0.35348860  0.09211618
[7,]    1    4    2    3    1 0.03334197  0.79471110
[8,]    1    2    4    3    1 0.63334113  0.03185270

$explain
[1] 0.8005291
```

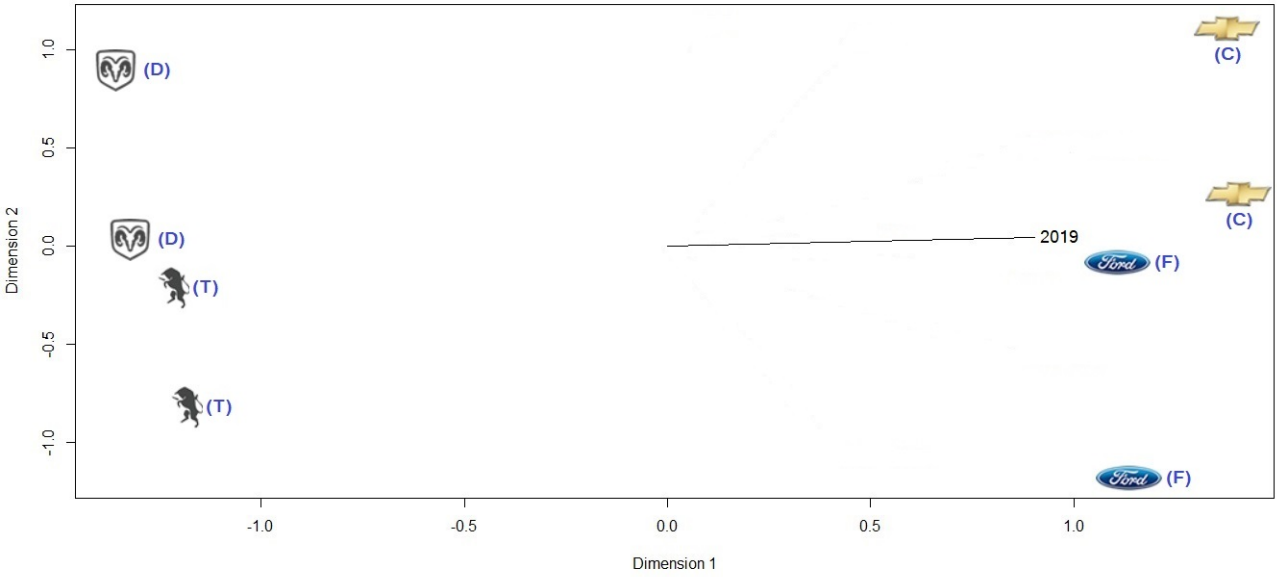


Visualización 5: Ploteo bi-dimensional que da MDPREF para las marcas que compiten en TC.

Notar −por ejemplo− que 2018 fue el único año en que una de las marcas con casi con un mimso valor de `x` (Chevrolet en 1.38, Ford en 1.14) y distantes en el eje `y`, terminó en #4 lugar, dando a Dodge y Torino las posiciones #2 y #3; no es casual entonces que el vector de ese año se vea paralelo a la línea que dijimos que es casi vertical dada por el segmento Chevrolet-Ford.



Visualización 6: La proyección perpendicular de las marcas sobre la recta que pasa por el origen y por `2018` da como resultado `Chevrolet`(arriba)>`Dodge`>`Torino`>`Ford`(abajo).



Visualización 7: La proyección perpendicular de las marcas sobre la recta que pasa por el origen y por `2019` da como resultado `Chevrolet`(derecha)>`Ford`>`Torino`>`Dodge`(izquierda).

Para ver claramente lo eficaz del plot del método MDPREF como ejemplo de visualización de múltiples rankings es que introduciré al concepto de «disimilaridad entre rankings»:

Medida de Disimilaridad de Spearman (i.e. distancia de Spearman) **para Dos Rankings**

Sean l y $\tilde{l}$ dos listas rankeadas, ambas de longitud k

$$\rho = \sum_{i=1}^{k} |l(i) - \tilde{l}(i)|$$

donde $l(i)$ es el ranking del item i en la lista l y $\tilde{l}(i)$ es el ranking del item i en la lista $\tilde{l}$.

OBS.: La distancia de Spearman ρ es simplemente la norma $\mathcal{L}_1$ y se la nota $\rho = ||l - \tilde{l}||_1$.

Con este concepto en mente completaré la matriz simétrica $\boldsymbol{D}$, con $\rho_{año_i, año_j}$ en la celda (i, j):

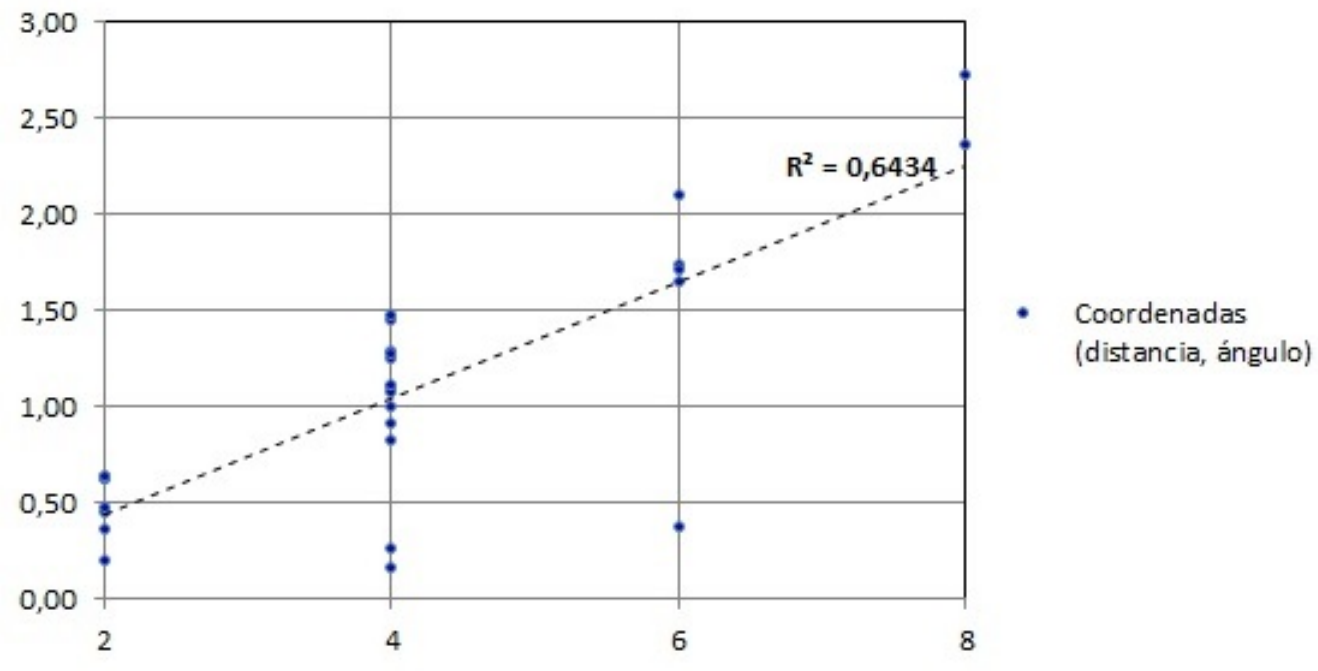
$$\boldsymbol{D} = \begin{pmatrix} 0 & 2 & 6 & 2 & 4 & 4 & 6 & 2 \\ . & 0 & 8 & 4 & 6 & 4 & 8 & 4 \\ . & . & 0 & 4 & 2 & 4 & 2 & 4 \\ . & . & . & 0 & 2 & 6 & 6 & 4 \\ . & . & . & . & 0 & 4 & 4 & 2 \\ . & . & . & . & . & 0 & 4 & 2 \\ . & . & . & . & . & . & 0 & 4 \\ . & . & . & . & . & . & . & 0 \end{pmatrix}$$

donde las 8 filas/columnas refieren a los años que van del 2012 (la 1ª) al 2019 (la 8ª; y en ese orden), omitiendo los previos y los posteriores por tratarse de campeonatos que repitieron el ranking de marcas en las posiciones finales.

Y además completaré la matriz simétrica $\boldsymbol{A}$, con el $\alpha_{ij} :=$ *ángulo* en radianes entre el vector del $año_i$ y el vector del $año_j$, tal como se ven en el plot de MDPREF, en la celda (i, j) de $\boldsymbol{A}$:

$$\boldsymbol{A} = \begin{pmatrix} 0.00 & 0.62 & 1.74 & 0.46 & 1.10 & 0.83 & 2.10 & 0.62 \\ . & 0.00 & 2.36 & 1.08 & 1.72 & 1.45 & 2.73 & 1.25 \\ . & . & 0.00 & 1.29 & 0.65 & 0.91 & 0.36 & 1.12 \\ . & . & . & 0.00 & 0.64 & 0.37 & 1.67 & 0.17 \\ . & . & . & . & 0.00 & 0.27 & 1.01 & 0.47 \\ . & . & . & . & . & 0.00 & 1.27 & 0.20 \\ . & . & . & . & . & . & 0.00 & 1.48 \\ . & . & . & . & . & . & . & 0.00 \end{pmatrix}.$$

Por último muestro la nube de dispersión en el plano XY de las 36 coordenadas no nulas que se ubican por encima de la diagonal de ambas matrices ($(x, y) = (\boldsymbol{D}_{ij}, \boldsymbol{A}_{ij})$) y el resultado de un ajuste lineal simple con un alto $\mathtt{R}^2$:



Visualización 8: Gráfico de dispersión de valores no nulos sobre la diagonal de la matriz $\boldsymbol{D}$ (distancias) y valores no nulos sobre la diagonal de la matriz $\boldsymbol{A}$ (ángulos); el $\mathtt{R}^2 > 0.6$ del ajuste lineal simple practicado a las 36 observaciones demuestra correlación positiva entre X e Y.

Referencia anexa: Gráficos apoyados en técnicas de reducción de la dimensionalidad

Es de esperar que toda vez que los datos que deban ser analizados se encuentren en espacios que excedan la dimesión 3, necesitemos de herramientas que sinteticen las relaciones que ocurren entre ellos, pretendiendo a la vez obtener una relativa baja pérdida de información al transformar los datos originales en representaciones que puedan graficarse en ejes cartesianos. Lo real -por fortuna- es que no hay una única receta efectiva, todas éstas encuadran dentro de las técnicas de reducción de la dimensionalidad que estudia la Estadística en la rama del Análisis Multivariado; los ejemplos más comunes son: el Análisis de componentes principales, algunos Métodos de clustering, el Análisis de correspondencia, el Análisis discriminante, etc. Y aunque se trata de un libro de una generación previa a la que convivió con el surgimiento de la Big Data, ya comenzado el siglo XXI, el mismo representa un punto de partida para que el lector que quiera conocer más al respecto pueda adentrarse en el tema:

Lebart, L., Morineau, A. and Warwick, K., *Multiariate Descriptive Statistical Analysis, John Wiley & Sons, 1984.*

3.3. Productos derivados

Revisaremos ahora algunos temas que no habitan en la centralidad de este trabajo pero que son interesantes profundizar.

3.3.1. *Handicap*

El concepto de **handicap** aplicado a un deporte consiste en afectar el normal desarrollo de la competencia para lograr un equiparamiento entre las performances finales de los jugadores. Si hiciéramos un repaso por los deportes con mayor popularidad, practicados éstos por atletas profesionales, no vamos a encontrar tantos ejemplos de aplicación de handicaps; sin embargo, este recurso es muy común en el ámbito amateur o en los casos de disciplinas con un número de participantes relativamente bajo. Dichos contexos son propicios para que ocurran circunstancias en las que se note claramente el provecho de esta herramienta. Veamos el ejemplo del golf. Supongamos que un club desea organizar un torneo para sus socios e invitados de otros clubes, un elemento clave será el esquema de premios a los ganadores, además de los recaudos para la organización del evento en sí. Vamos a suponer también que un objetivo de la competencia es que se inscriba la mayor cantidad de participantes, sean éstos hombres o mujeres, juniors o veteranos. Si las reglas se plantean por igual para todos los participantes, entonces rápidamente deduciremos que el premio al ganador no opera de igual forma como estímulo para la inscripción de un golfista activo, en el pico de su carrera, versus un golfista veterano o muy novel – idéntica situación podría plantearse entre competidores hombres y mujeres. Así las cosas, se podrían fijar premios a la mejor performace *overall*, al mejor veterano, al mejor junior, y a la mejor golfista mujer, y tendríamos de ese modo `n` estímulos (menores, agrego yo) en lugar de un único premio: este esquema de múltiples premios intenta no dejar a ninguna categoría sin una zanahoria que alimente las ganas de anotarse y participar. Ahora bien, poner un premio al «mejor veterano» y que luego se anoten 10, y el mismo premio al «mejor junior» y que se lo disputen 50, tampoco parece del todo justo. La solución que se propone al usar el handicap se resume en que a los jugadores y jugadoras se les va a afectar el número total de golpes (N. del A.: el propósito en el golf es embocar los 18 hoyos del campo usando la menor cantidad de golpes) restando artificialmente una cantidad variable en función al nivel a priori de cada participante. Me tomaré la licencia de inventar algunas cifras para poder explicar mejor el mecanismo: los mejores golfistas (los de mejores antecedentes y/o los de mayor ranking) van a restar 0 (cero) golpes, son el grupo a los que el handicap no les modificará su performance final; por otro lado a los menos experimentados se les va a restar 18 al finalizar el recorrido, lo que supone que en cada hoyo el jugador con handicap=18 podrá tomarse 1 golpe más en promedio para completarlo que aquél con handicap=0. En la misma línea los scores de las mujeres podrán ajustarse con la resta de 12 golpes y a los golfistas veteranos restarles 6 de su score final. Siguiendo con el ejemplo y visto éste ahora desde la óptica de un jugador que inicia el torneo con «0 de handicap» su plan estratégico para triunfar será sumar menos golpes que todos los de su grupo, totalizar 7 golpes menos que los veteranos con handicap=6, 13 golpes menos que las golfistas mujeres con handicap=12, y así con los demás. Puesto que los jugadores con handicap > 0 cuentan con esa

ventaja dada por el organizador, el golfista que triunfe será quien menos `golpes'` realice, donde `golpes'` se define como la cantidad de golpes hechos menos los golpes dados por el handicap.

Este primer ejemplo muestra una asociación inversa entre el nivel del deportista y el tamaño de la ventaja (o handicap) a afectar en la performance del torneo. Tan extendido está el concepto de handicap en el golf amateur que las asociaciones establecieron estándares para determinar el número de golpes que cada golfista restará al término de la ronda, ése valor hará las veces de un rating en el sentido inverso (menos golpes $\rightarrow$ mayor rating) y es común que se hable en el ambiente de un *jugador con X golpes o menos de handicap* como un indicativo de su alto nivel.

Y si en el golf al handicap se lo mide en golpes, en el polo se lo referencia como goles. Sólo que en este deporte, y como el objetivo del enfrentamiento entre las escuadras es el de anotar más goles a lo largo del encuentro, el equivalente al jugador de golf con handicap=0 será aquel polista con handicap=10. Por lo que un equipo con handicap=40 surgirá de estar conformado por 4 jinetes de handicap=10 cada uno. Finalmente esa ventaja se restará del número final de goles que marque el equipo. Otro uso que se le da a los «goles de handicap» es el de habilitar o no a que un team participe en determinado evento. Por caso, muchos de los torneos de esta disciplina limitan la conformación de los equipos a que éstos no superen *una X cantidad de goles*. Muestro debajo al respecto cuáles son las reglas en vigencia según de qué torneo se trate:

- **Copa de Supernaciones.** Pueden participar escuadras con hasta 24 goles de handicap.
- **US Open Championship.** Pueden participar escuadras con hasta 22 goles de handicap.
- **Campeonato Mundial.** Pueden participar escuadras con hasta 14 goles de handicap.
- **Campeonato Europeo (absoluto).** Participan equipos entre 6 y 8 goles de handicap.
- **Campeonato Europeo femenino.** Participan equipos entre 10 y 12 goles de handicap.

Con las limitantes anteriores no podrían confluir más de 3 jugadores con handicap alto (8 o más goles) en un equipo que quiera disputar uno de estos torneos, que aunque no son torneos amateurs usan el handicap como forma de alentar a que más escuadras quieran inscribirse dado que no habrán en el torneo equipos de estrellas, y por lo tanto lo encuentren más competitivo. Por fuera de estas competencias que buscan promocionar al polo vía el uso de estas reglas que equiparan performances, se disputan muchos otros campeonatos en los que no rige un tope de goles y que permiten la participación de equipos con alto handicap por parte de todos sus jinetes.

Análogamente −como en el caso del golf− se utiliza al handicap del polista como un rating del nivel de éste i.e. mayor número de goles $\rightarrow$ mayor nivel.

Veamos ahora un último ejemplo de aplicación de handicaps a un deporte, en este caso al ajedrez. Se podría afirmar que aquí es donde más se ha experimentado con el uso del handicap, sin embargo éste ha quedado prácticamente a un lado en los torneos que se disputan en la actualidad. Una posible explicación es que la organización de un torneo de ajedrez es mucho menos

laboriosa que si lo comparamos a un torneo de golf o de polo. Con lo cual un jugador que viva en una región donde se juegue popularmente el deporte, debería poder inscribirse fácilmente en un torneo acorde a su nivel; y que esto se dé hasta varias veces al año (sin contar los torneos de ajedrez online, que al romper las barreras geográficas habilitan a que cualquier ajedrecista pueda competir los 12 meses del año en torneos contra rivales de su misma condición; esto se logra fijando restricciones de «Elo mínimo» y «Elo máximo», al estilo de lo que se hace en el polo con goles).

Ahondemos ahora en la historia del uso de handicaps en el ajedrez y en qué lo fomentó. La respuesta está clara: durante mucho tiempo, hacia el siglo XIX, al ajedrez se lo jugaba por dinero, no del modo en que se practica hoy, con premios para los jugadores según el puesto que ocupen en la clasificación final, sino directamente como una puja de apuestas. Un organizador se ocupaba de administrar el dinero de los creeían que el jugador de piezas blancas ganaría el duelo, al igual que el de los que creeían que lo ganaría el de piezas negras, tomaba su margen y separaba para pagar a cada ajedrecista el porcentaje que acordó con ellos. Por último quedaba un pozo que se repartía entre los que habían acertado el resultado. Así se conformaba un entretenimiento en el que no era necesario conocer demasiado de tácticas y estrategias para formar parte, sólo había que saber organizar partidos parejos y el público apostador iba a hacer el resto, jugando su dinero por aquel competidor que mejor impresión le causara. Sin embargo para llevar a cabo estos enfrentamientos de resultado incierto se iba a requerir que hubiera siempre un margen de incertidumbre, puesto que en una situación donde se presume de antemano quién ganará, el volumen de apuestas tiende a decaer por lo poco competitiva de la contienda. Ahí es donde proliferó el uso de handicaps que a diferencia del golf o el polo en donde quien gana lo logra por realizar el menor número de golpes o anotar el mayor número de goles, con el ajedrez y un resultado 1 vs. 0 o 0 vs. 1 (o eventualmente un empate $\frac{1}{2}$ vs. $\frac{1}{2}$) se debía establecer algo innovador. Así fue que para emparejar los duelos entre un jugador débil y uno fuerte, se le permitió al primero jugar 2 veces seguidas al comienzo de la partida (cuando en el ajedrez se alternan las jugadas del bando blanco y del negro, sin excepción). Alternativamente si la supuesta ventaja del fuerte sobre el débil era mayor, se le quitaban piezas al jugador más fuerte, bajo el concepto de que con menos piezas le sería más difícil ganar. Por último se registraron situaciones en las que se pudieron combinar ambas (es decir, que el débil juegue dos y hasta tres veces seguidas al inicio del partido, y que además al jugador fuerte se le quiten piezas). Ya una vez que se empezó a jugar con reloj, el handicap podía estar relacionado a que el jugador de menor nivel contara con más tiempo para desplegar su juego, otorgándole algunos minutos extra para pensar.

Llegado a este punto en donde recapitulamos que la diferencia en performance entre dos jugadores de ajedrez se estimaba ya hace 200 años con el fin de determinar el handicap que haría más pareja la partida y así amplificar el mercado de apuestas, es que cabe preguntarnos si algo relacionado a aquello puede interesar que sea retomado hoy. La vinculación indirecta que encuentro tiene relación con los módulos de ajedrez (los llamados históricamente programas de computadora u ordenador). En el presente es habitual determinar un Elo ficticio para cada uno de los distintos módulos, surge de hacerlos competir entre sí en torneos como el $TCEC-Top$

Chess Engine Championship y asegurando que todos procesen sus cálculos con el mismo hardware. Producto de lo bien que se desempeñan éstos es que a cualquier aficionado o profesional del ajedrez le resulta sin dudas un esfuerzo dar la talla frente a un módulo del año 2022 (incluyo en este universo al mismo campeón mundial). Es en esas circunstancias que la otrora ventaja entre el jugador fuerte y el jugador débil se representa hoy entre el módulo y el humano, y el Rating ELO hace las veces de termómetro para definir cuál de entre varias opciones de handicap se puede usar para igualar las chances entre los bandos. Esta información fue provista por los fabricantes del módulo Komodo versión 9, el cual posee un Elo ficticio estimado en 3300 puntos:

Regla de handicap	**Elo Humano con el que equipara la partida**
Komodo 9 juega blancas sin la dama	1300
Komodo 9 juega sin la torre de a1	1900
Komodo 9 juega sin el caballo de b1	2200
Komodo 9 juega negras sin el peón de f Humano inicia con dos movidas seguidas	2700
Komodo 9 juega sin la torre de a1 Humano juega sin el caballo de b8	2800

Cuadro 1: Handicaps en ajedrez y su relación con el Elo Humano.

En el caso anterior de reglas fijadas para igualar partidas entre módulos y humanos puede verse con claridad el uso implícito del rating para definir un handicap, y aunque la tabla queda algo pobre por no aproximar a cualquier valor de Elo Humano entre 1500 y 2400, que es el rango que poseen la mayoría de los aficionados, los valores informados sirven como referencia.

3.3.2. Predicción hacia adelante

Llegamos hasta acá tras repasar un abanico de métodos de ranking con aplicaciones a deportes. Donde un patrón común a todos fue el apoyarse en el historial de resultados y ratios de performance para obtener un orden justo de los elementos del conjunto (v.g. deportistas, equipos de competición). Pero nada se habló de la capacidad de tales rankings para estimar el desempeño futuro, algo que ocurre muy a menudo en la rama conocida como inferencia estadística: proponer que el comportamiento pasado está correlacionado con los sucesos por venir... a esto remite el concepto de usar un ranking (*backward*) para elaborar una **predicción hacia adelante (*foresight prediction*)**. El supuesto que subyace bajo esta idea es que en la historia de un deportista o equipo, su nivel de desempeño no cambia abruptamente, ergo la posición en un ranking que los contemple va a tender a ser estable, mirada ésta como válida sólo durante un determinado tiempo; dicho intervalo vendría a representar a este efecto el horizonte de la predicción. No es el objetivo profundizar este tema anexo al punto central de la Tesis, pero sí enunciaré diversas experiencias que he hallado en relación a utilizar rankings como herramientas para predecir el futuro:

- En [Langville and Meyer, 2012] se describe con lujo de detalles una experiencia del uso de rankings para la predicción de resultados en el marco del mayor Torneo Universitario de

Básket de los Estados Unidos, conocido como la División I de la NCAA del básket colegial. En dicho relato se menciona que desde la edición de 2006 del torneo se han venido usando sistemáticamente distintas variaciones de algunos de los rankings más conocidos (como los comentados en los capítulos 3.1.3 y 3.1.5 de este documento) con singular éxito.
¿Qué representa eso en términos numéricos? Para empezar a hablar debo contar cuántos equipos participan en el torneo en cuestión, de qué manera se enfrentan entre sí y cómo se usaron los rankings para estimar resultados futuros. Al torneo llegan a jugarlo por distintas vías de clasificación un total de 64 equipos, éstos se reparten en 4 llaves de 16, una por cada región en la que la NCAA divide a las universidades. Ya dentro de esa llave, se juegan octavos de final, cuartos, semis y una final de llave que determina a un equipo clasificado para la instancia denominada *Final Four*, pues al tratarse de 4 regiones habrá por tanto cuatro equipos que disputarán quién es el campeón nacional. En definitiva, de 64 equipos que inician su participación, al término de pocas semanas queda solo 1, cada definición es a un único partido y como en el básket no hay empate, el ganador pasa y el perdedor «se va a su casa». Un total entonces de 63 partidos se juegan en aproximandamente 20 días para conformar lo que los fanáticos llaman cada año la locura de Marzo (N. del A.: este torneo se jugó de esta forma desde el año 1985 en dicho mes). Pero desde varios meses antes a que estén definidos los 64 equipos, éstos y muchos otros más se han venido enfrentando con el objetivo de sumar puntos para ser elegidos entre los mejores 64 del país. En este punto me detengo puesto que los rankings que se cuentan en [Langville and Meyer, 2012] que fueron usados para hacer foresight predictions basaron sus ordenamientos en el historial recogido en dicha etapa clasificatoria; dicho de otra manera, fueron entrenados para definir qué equipos rankean sobre otros usando los meses de la clasificación, y fueron puestos a prueba con la tarea de acertar el resultado de los 63 duelos del torneo de Marzo.
La experiencia tal como se relata en el libro no se trató de un hecho aislado, sino que los 63 pronósticos dados por cada ranking, formaron parte de un concurso nacional llamado el Bracket Challenge, en el que millones (sí, millones literalmente) de concursantes participan con la intención de acertar el mayor número de ganadores de los 63 juegos de eliminación (no se piden resultados, sólo nombrar qué equipos ganan cada duelo). Ahora bien, a lo largo de la realización de este concurso −que dicho sea de paso reparte cientos de miles de dólares en premios a quienes más acierten− se ha encontrado que predecir más de 50 ganadores correctamente es algo muy difícil. Bueno, algunos de estos rankings han llegado en el mejor de los casos a alcanzar esa cifra. Que vista en *accuracy* da un porcentaje de acierto superior al 75 % ($50/63 = 79.4\,\%$). El haber alcanzado con un determinado ranking los 50 aciertos, provocó que esa vez la jugada se ubicara en el percentil 1° de mejor desempeño de todo el concurso, algo nada despreciable. Pero aquello (que ocurrió en el año 2008, para ser precisos) no se volvió a repetir, sacando como conclusión que los rankings basados en historia reciente operan como buenas aproximaciones a los resultados deportivos que sucederán en lo inmediato, pero no sirven por sí solos para descollar y alzarse con el premio mayor. Al decir del Profesor Joel Sokol de la Universidad de Georgia Tech, un modelo predictivo creado a este fin va a rondar el 75 % de aciertos sólo si

es realmente bueno. Mientras que en el otro 25 % de los juegos, donde seguramente la predicción esté bien fundamentada pero aún así falle, los factores que incidirán serán el rebote azaroso de una pelota o con qué humor se levantaron ese día los equpios.

- Ya había mencionado antes al portal de periodismo de datos `fivethirtyeight.com`, propiedad del conglomerado de medios abcNEWS. Fue en la subsección 3.1.4 en la que comenté que producto del enamoramiento de sus periodistas por el Método ELO, les llevó a calcular rankings basados en él para deportes tan diversos como backgammon, básket, béisbol, fórmula 1, fútbol, etc.; con un puntapié inicial que tuvo lugar en 2014 cuando comenzaron a publicar el ELO para la liga profesional de fútbol americano de USA (NFL). Tras ese inicio en el que públicamente comentaron que se inspiraron en otro portal −que había comenzado una década antes con la publicación de Elo's para selecciones nacionales de fútbol (masculino), de nombre `eloratings.net`− se ocuparon de lanzar versiones superadoras de dicho método, hasta el día de hoy. La motivación para volver a citarlos guarda relación con lo que vienen haciendo desde no hace mucho, comenzando con la temporada 2018-19 de la NFL. En aquella edición, desafiaron a sus lectores entusiastas a que intenten mejorar el ordenamiento dado por su ELO-ajustado, y el modo de hacerlo fue, al igual que en el relato anterior, ideando un concurso como el Bracket Challenge, sólo que con algunas diferencias: i) no existe premio metálico para aquellos participantes que mejor performen, y ii) por tratarse el ELO de un método que ellos recalculan después de cada nuevo enfrentamiento, para ser justos con los que intenten mejorar las predicciones surgidas a partir de comparar ambos ratings, lo mismo debió permitírseles a los participantes, es decir que en cada momento del campeonato el jugador podía refinar sus pronósticos (para el Bracket Challenge, que como se dijo es una tradición que se practica desde antes de la expansión de la Internet, se requirió y requiere hasta hoy que los resultados de todos los duelos sean predichos ANTES del comienzo del torneo). Salvando entonces esas diferencias, ya cuentan con 3 ediciones en las que el ELO de 538.com estuvo siempre entre los mejores puntajes aunque nunca trepó a los «top 100». Al respecto se vuelcan en el cuadro debajo indicadores de estos últimos concursos:

Temporada NFL	**Nro. de participantes**	**Posición del ELO de 538.com**	**Percentil**
2018-19	15073	701ª	mejor 5 %
2019-20	15141	514ª	mejor 3 %
2020-21	*no hubo concurso / no hay datos publicados*		
2021-22	9781	1105ª	mejor 11 %

Cuadro 2: Clasificación histórica del ELO de 538.com en concursos de pronósticos de la NFL.

De nuevo como con el uso de rankings para hacer *foresight predictions* en el torneo de Marzo del básket universitario de USA, los métodos para ordenar muestran un aceptable nivel de predicción, pero todavía no pueden vencer o quedar cerca de los primeros puestos en competencias abiertas contra miles de otros pronosticadores.

- Veamos finalmente una experiencia que muestra una mayor complejidad, en primer lugar

saldremos de los deportes sin empates, el Básket claramente lo es, mientras que el Fútbol americano tiene una tasa de ocurrencia de duelos empatados de alrededor del 0.2 %, lo cual se puede asumir despreciable. En cambio el hockey sobre hielo y el fútbol (*soccer*) tienen mucho mayores tasas de ocurrencia de encuentros igualados en score, alrededor del 9 % el primero y más cerca del 23 % el último, medidos éstos sobre los juegos disputados entre el 1°/01/2022 y el 31/08/2022 por selecciones nacionales. Por lo que en un escenario así, donde los resultados posibles son 3 (G-P-E) acertar un pronóstico se vuelve más desafiante.
Llegado a este punto nos vamos a encontrar con que a diferencia de los ejemplos antes mencionados, donde la sola ubicación de un equipo por sobre otro en el ranking determinaba la victoria futura del primero sobre el segundo, aquí vamos a necesitar involucrar a la Estadística para arribar a un enfoque en el que no se decida «G vs. P» sino que en términos de probabilidades, tengamos el vector $(p(G), p(P), p(E))$ tal que $p(G) + p(P) + p(E) = 1$ y del que resulte que bajo algún criterio cada $p(\cdot)$ aproxime a la chance real de que el evento ocurra. Acá se abren 2 caminos, con sus diferencias: el primero parte del supuesto de que en muchos deportes donde se dan empates, éstos se producen por una igualdad en anotaciones (goles, carreras, hits, según sea el caso); así las cosas es viable estudiar estadísticamente la aleatoriedad de esos valores, algo que fue ampliamente analizado en el fútbol con la familia de distribuciones de Poisson, ya una vez que para un determinado match se tienen medidas las chances de que A meta n goles y B anote m, la determinación del vector de probabilidades $(p(G), p(P), p(E))$ surge como derivación directa. Éste es el enfoque que adopta el sitio 538.com para el forecast de los principales partidos a lo largo del planeta. A continuación se expone el ejemplo que da el portal sobre cómo se combinan dos distribuciones Poisson (la de los goles de Liverpool y Brighton en el encuentro del 13/05/2018 en el caso elegido) para constituir una prob. conjunta y determinar sobre ella la chance concreta del resultado G-P-E:

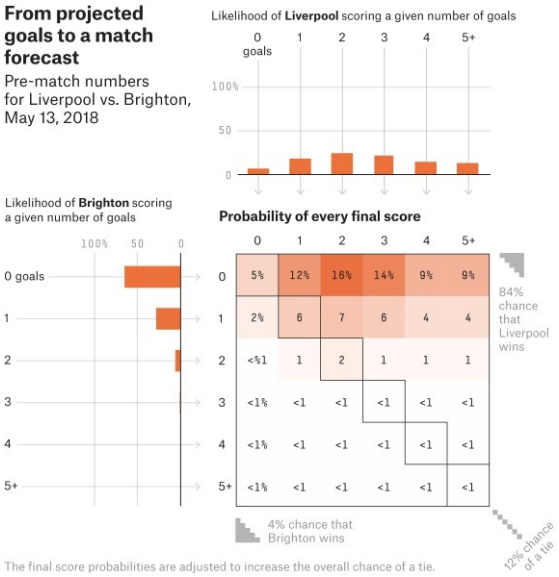
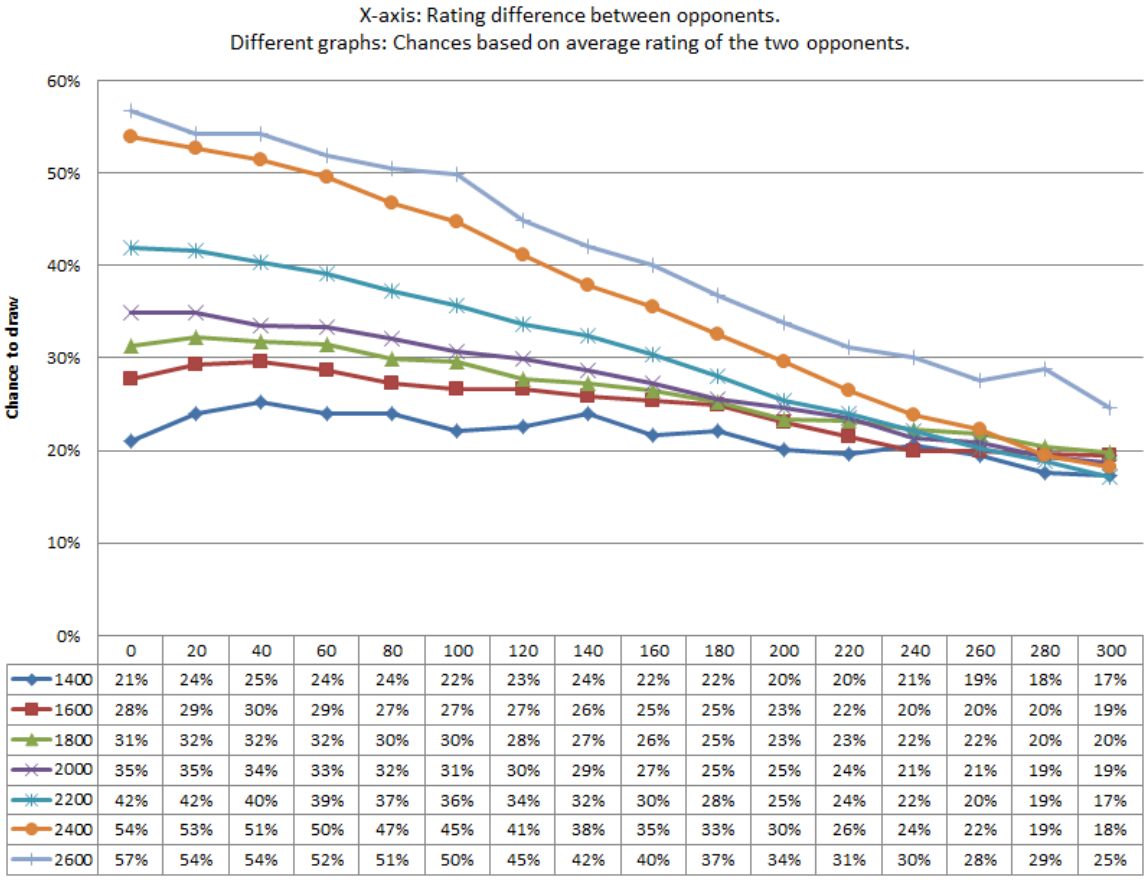


Imagen 1: Conversión de goles proyectados por equipo a resultado de un partido.

En esencia a este mismo enfoque lo han utilizado otros sitios para inferir resultados deportivos, entre los que se destacan 301060.exactas.uba.ar y 851.cl, de Argentina y Chile, respectivamente, ambos aplicados al Fútbol (fifa).

El segundo de los mecanismos propone un tratamiento más simple, pues omite incursionar en el modelado de los goles, es además un abordaje más general, puesto que incluye a los deportes en los que no se llega a un empate por igualdad de anotaciones, como ocurre en ajedrez o boxeo. Aquí se va a proponer un modelo en 2 etapas, primero se estimará la $p(E)$ y luego con esa estimación fijada se procederá a resolver las otras posiciones del vector (alcanza con dar un valor a $p(G)$ pues $p(P)$ sale por diferencia, $p(P) = 1 - p(G) - p(E)$). Tomaré de muestra lo analizado para el Ajedrez, un deporte que se presta como ningún otro para posar sobre él la lente de la Estadística (justificado principalmente en que desde hace décadas los resultados de todos los torneos oficiales están informatizados y consolidados en bases de datos de acceso libre, sumado a que no se exige un descanso de días entre partido y partido de un mismo jugador, cosa inimaginable en tantas otras disciplinas, lo que lleva a que en un corto tiempo se logre recolectar una significativa cantidad de juegos). El gráfico que sigue sintetiza visualmente la relación que se observó entre los ratings ELO de jugadores que se enfrentaron entre sí y la probabilidad de que la partida haya terminado *empate*:



	0	20	40	60	80	100	120	140	160	180	200	220	240	260	280	300
1400	21%	24%	25%	24%	24%	22%	23%	24%	22%	22%	20%	20%	21%	19%	18%	17%
1600	28%	29%	30%	29%	27%	27%	27%	26%	25%	25%	23%	22%	20%	20%	20%	19%
1800	31%	32%	32%	32%	30%	30%	28%	27%	26%	25%	23%	23%	22%	22%	20%	20%
2000	35%	35%	34%	33%	32%	31%	30%	29%	27%	25%	25%	24%	21%	21%	19%	19%
2200	42%	42%	40%	39%	37%	36%	34%	32%	30%	28%	25%	24%	22%	20%	19%	17%
2400	54%	53%	51%	50%	47%	45%	41%	38%	35%	33%	30%	26%	24%	22%	19%	18%
2600	57%	54%	54%	52%	51%	50%	45%	42%	40%	37%	34%	31%	30%	28%	29%	25%

Imagen 2: Tasa de empates observada en torneos oficiales de Ajedrez.

Del gráfico de líneas de la Imagen 2 se desprende lo siguiente...
Un modelo posible de **draw_rate** es verlo como función bivariada tal que $\mathbf{draw_rate}(\cdot, avg_rtg)$ pueda ser pensada como función estrictamente decreciente i.e. menor chance de empate a mayor diferencia absoluta entre ratings; y a su vez como función bivariada tal que $\mathbf{draw_rate}(rtg_diff, \cdot)$ es función estrictamente creciente i.e. mayor chance de empate a mayor promedio entre ratings – con base en esto último Kirill Kryukov propuso en el web-site kirill-kryukov.com/kirr (sección *Computer chess*, subsección *KCEC*, menú *Statistics* que dicha función puede aproximarse por este ajuste no lineal:

$$p(E) \approx \mathbf{draw_rate}(rtg_diff, avg_rtg) = -\frac{rtg_diff}{32.49} + \mathbf{exp}\left(\frac{avg_rtg - 2254.7}{208.49}\right) + 23.87.$$

El camino para determinar luego $p(G)$ es en esencia análogo, tomando en cuenta una base histórica de partidas de cuyos jugadores se conocen el rating ELO al momento de disputarlas, se definen relaciones funcionales que logran minimizar el error de estimación derivado de usar a los ratings como predictores de la $p(G)$. En este caso la FIDE, entidad rectora del deporte a nivel mundial, ha elegido desde hace varios años publicar su propia estimación, algo que podría ser objeto de debate, ya que dichas probabilidades surgen de la elección de supuestos, hecho en el cual influye cómo éstos sean validados. Un ejemplo de que todas las posturas hechas en la

dirección de vincular los ratings de jugadores **A** y **B** de ajedrez a una probabilidad cierta de que **A** triunfe sobre **B** no son categóricas, es que una asociación nacional, como la *USCF− U.S. Chess Federation* hizo públicos otros guarismos, parecidos pero no idénticos. Ambos enfoques coinciden por cierto en apoyarse exclusivamente en la diferencia absoluta entre ratings ELO (la misma variable *rtg_diff* utilizada anteriormente para ajustar la **draw_rate**), así como en elegir mostrar la estimación de la variable compuesta $1 \cdot p(G) + \frac{1}{2} \cdot p(E)$, a la que llaman con la sigla **ES** por *Expected Score*, de la cual se puede despejar $p(G)$ fácilmente: $p(G) = \mathbf{ES} - \frac{1}{2} \cdot p(E)$. A continuación se ven las series de valores de **ES** según cada fuente, de los que se deduce el notable parecido entre ambas con mínimas diferencias en las mayores *rtg_diff*:

Diferencia de ratings entre oponentes	**ES** *fide*		**ES** *uscf*		$\lvert\Delta_{\mathbf{ES}}\rvert$
	H	L	H	L	
0	.50	.50	.50	.50	.0000
40	.56	.44	.56	.44	.0016
80	.61	.39	.61	.39	.0029
120	.66	.34	.67	.33	.0034
160	.71	.29	.72	.28	.0030
200	.76	.24	.76	.24	.0017
300	.85	.15	.85	.15	.0041
400	.92	.08	.91	.09	.0102
500	.96	.04	.95	.05	.0132
600	.98	.02	.97	.03	.0128
700	.99	.01	.98	.02	.0103

Cuadro 3: *Expected Scores* en función a la variable *rtg_diff* (H=higher rated, L=lower rated).

Por tanto se llegó finalmente a determinar el vector $(p(G), p(P), p(E))$ asociado a un enfrentamiento entre dos equipos o individuos, cuyos desempeños en el deporte nos fueron dados a conocer a través de sus ratings, se traten éstos de valores fijados en escalas abstractas −como lo es el ELO− o siendo estimaciones de «goles a favor» y «goles en contra» como ocurre en el tratamiento que dan 538.com, 851.cl y otros portales al realizar sus pronósticos de resultados de fútbol (*soccer*). En dichas situaciones surge un uso colateral del que haré un breve acercamiento en el próximo apartado. Antes quiero mencionar un mecanismo de testeo de la *goodness-of-fit* que encontré para las estimaciones de las $p(\cdot)$ (similar a como relaté antes que se hizo con el uso de rankings para predecir resultados del torneo de Básket Universitario y con el ELO de 538.com para la NFL). Acá el trabajo [Hvattum and Arntzen, 2010] sugiere «poner a competir» a las foresight predictions que adoptan la forma de probabilidades en el mercado de apuestas deportivas, eligiendo en su caso meterse con los resultados de fútbol (*soccer*). En dicha estrategia los autores simularon haber apostado consistentemente a través de diferentes métodos que tienen más que ver con la optimización de inversiones, tomando para ello el output surgido a partir de Ratings tipo-ELO de los equipos. La conclusión a la que arribaron es que por sí solos los ratings que habían elaborado para la construcción de sus $p(\cdot)$ no alcanzaban para batir a la banca − aunque se ubicaban cerca de hacerlo consideraron necesario explorar

el uso de covariables que se asume que influyen fuerte en la determinación de los odds de las casas de apuestas; como si está previsto que llueva durante el partido o la ventaja de localía.

Uso de *foresight predictions* cuando toman la forma de probabilidades de eventos

Una vez halladas las $p(\cdot)$ que modelan a los resultados de un juego, es posible combinar dichas predicciones puntuales con el objetivo de interpretar qué opciones surgen frente a un suceso más complejo, como lo es un campeonato. Ese tipo de desafíos son habituales en los análisis de pronósticos deportivos y toman la forma de preguntas como:

- ¿Qué equipo cuenta con mayores chances de salir campeón, sujeto a la cosecha de puntos que el mismo y sus rivales recogieron a mitad de torneo y al fixture de enfrentamientos que resta disputarse?
- ¿Qué equipos son los que más probablemente se ubiquen en las primeras posiciones y por lo tanto asciendan de división o clasifiquen a copas internacionales/instancias posteriores?
- ¿Qué equipos tienen mayores probabilidades de puntuar entre los últimos y perder la categoría al terminar la temporada?

En este punto lo que mejor se adapta para dar respuesta a estos interrogantes es el uso de *simulaciones de Monte Carlo*, el extendido método que a partir de la generación computacional de números aleatorios brinda una solución aproximada a problemas que son numéricamente muy complejos para ser resueltos analíticamente. No es el propósito extender demasiado este aside, simplemente menciono que la finalidad detrás de procesar decenas de miles de escenarios (el sitio 538.com ejecuta 20.000 corridas para sus pronósticos y en 301060.exactas.uba.ar se usan 100.000) es contabilizar qué ocurrió en cada caso y proyectar la fracción de ocurrencias sobre el número de simulaciones como una aproximación a la probabilidad asociada a dicho evento.

3.3.3. *Performance rating*

El año 2007 fue testigo de un hecho asombroso en el Tenis profesional, al menos en el más alto nivel y hasta el día de hoy se lo sigue destacando como inédito... en la segunda mitad del calendario, en el marco del torneo Masters 1000 de Madrid (jugado en octubre), David Nalbandian ganó su partido de cuartos de final ante el #2 del ranking mundial (R. Nadal, ESP), al día siguiente batió en semifinales al #3 (N. Djokovic, SRB) y en la instancia Final le ganó al #1 (R. Federer, SUI). Estos cuatro protagonistas comenzaron la semana del torneo con ciertos rankings y finalizaron de forma similar, pareciendo no verse impactados por el hecho trascendente − ver Cuadro 4. Sin embargo no hay trampa, ya que el cálculo tomó en cuenta el desempeño en las 52 semanas previas, por lo que una mala actuación no posee una influencia determinante, y aunque la labor de Nalbandian fue superlativa siguió rankeado muy detrás de sus rivales derrotados (N. del A.: hay registro de antecedentes comparables en Estocolmo oct-1994 y Canadá ago-2007, pero desde entonces no trascendió que alguien venza en días consecutivos al #1, #2 y #3 del ranking).

Tenista	Puntos y Posición al 15/10	Puntos y Posición al 22/10
Federer	7355 (#1)	7205 (#1 =)
Nadal	5385 (#2)	5385 (#2 =)
Djokovic	4470 (#3)	4370 (#3 =)
Nalbandian	1160 (#25)	1435 (#18 ↑)

Cuadro 4: Rankings de Federer, Nadal, Djokovic y Nalbandian antes y después de Madrid 2007.

Por un lado es correcto que a través del Ranking los desempeños se valoren en una ventana más amplia que la de los últimos 7 o 15 días, sin embargo nadie pondría en duda que el nivel de juego demostrado por el campeón refleja desde los resultados que, en ese periodo, debió considerárselo mejor que ningún otro i.e. en sentido figurado, fue el #1 en esa semana.

¿Podría algún deporte pararse en un tiempo breve, como un torneo, un juego o parte de él, y dar un ordenamiento de quién jugó mejor que quién en función a los resultados y rivales? Éste es el objetivo del **performance rating** y aunque construcciones así no son tan populares, hay al menos dos experiencias que relevé y de las que daré cuenta a continuación:

- **Ajedrez** (una vez más)**.** En el Cuadro 3 pudo verse la relación que la fide (el mayor organizador de torneos a nivel mundial) halla entre la Difefencia de ratings ELO y los *Expected Scores* para ambos contendientes, el «higher rated player» y el «lower rated player». Establezcamos los supuestos de un escenario posible y calculemos luego lo que en ese deporte se nombra como Rp (abreviatura de *Rating of performance*). Pongamos por caso que en un torneo pactado a 14 rondas se enfrentan 7 jugadores todos contra todos en el sistema round-robin partido y revancha, cada par de ellos disputará 2 duelos intercambiando el color de piezas, y cada uno tendrá alternativamente 2 días de descanso o fechas libres. Supongamos además que los Elos de estos 7 jugadores son los siguientes, de mayor a menor: 2780, 2485, 2465, 2450, 2385, 2355, 2305. En el cuadro que sigue aparecen asociadas a los puntajes posibles las variables **ES** y *rtg_diff* que veníamos analizando en el apartado 3.3.2.

  Nos falta combinar estos valores para obtener el Rp de cada contrincante. Fijemos que el jugador con Elo=2780 salió campeón con 9 puntos y por lo tanto el **ES** fue 0.750; así las cosas sus rivales promediaron un Elo de $\frac{2485+2465+2450+2385+2355+2305}{6} = 2407.5$, sobre este último valor aplicaremos la *rtg_diff* correspondiente a su **ES** para deteminar el Performance rating → Rp = 2407.5 + 193 = 2600.5; y siguiendo con el ejemplo digamos que el jugador con Elo=2465 obtuvo 8.5 puntos, **ES**=0.708 y *rtg_diff* = + 91, por lo cual su Rp = 2460.0 + 157 = ¡2617.0! Es interesante interpretar que el Rp asignado al segundo jugador resultaría superior por haberse enfrentado «en promedio» contra mejores rivales.

- **Artes Marciales Mixtas - Liga Profesional de Luchadores** (*mma - pfl*)**.** La organización PFL de Artes Marciales Mixtas se creó en 2012 en los Estados Unidos y es al día de hoy un caso testigo de la expansión de la analítica aplicada al consumo visual de un deporte. Entre todos los recursos que le son provistos al espectador cabe destacar

Puntos posibles tras 12 juegos	**ES** *fide*	**Diferencia de ratings**
12.0 → ganó las 12 partidas	1.000	+ **800** (en lugar de $+\infty$)
11.5 → ganó 11 y empató la otra	0.958	+ 495
11.0	0.917	+ 395
10.5	0.875	+ 329
10.0	0.833	+ 276
9.5	0.792	+ 232
9.0	0.750	+ 193
8.5	0.708	+ 157
8.0	0.667	+ 123
7.5	0.625	+ 91
7.0	0.583	+ 60
6.5	0.542	+ 30
6.0	0.500	**0**
5.5	0.458	− 30
5.0	0.417	− 60
4.5	0.375	− 91
4.0	0.333	− 123
3.5	0.292	− 157
3.0	0.250	− 193
2.5	0.208	− 232
2.0	0.167	− 276
1.5	0.125	− 329
1.0	0.083	− 395
0.5	0.042	− 495
0.0	0.000	− **800** (en lugar de $-\infty$)

Cuadro 5: Relaciones entre Puntos posibles, *Expected Score* y *rtg_diff* surgidas de recorrer de inversamente el Cuadro 3; para ES=1.000 y ES=0.000 se asignan + 800 y − 800 por convención.

en esta subsección el *FPR* − *Fighter Performance Rating*, un indicador numérico que es publicado para cada peleador al término de los rounds, y consiste en un número entero entre 0 y 100. Esta construcción algorítmica es propietaria de PFL y está vigente desde 2021, se sabe que en su desarrollo participó la empresa IBM y este video (en inglés) explica que los hechos que se den durante la pelea harán que el *FPR* tome un valor u otro; los ejemplos que se citan allí son los más obvios (ej: «cuantos más golpes conecte un luchador, más alto será su rating»), pero se puede asumir que otros factores intervienen en su cálculo: [MMA, 2021].

4. Conjuntos de datos

- Los resultados históricos de **Fútbol** argentino (afa, *soccer*, masculino) fueron extraídos de la página web www.promiedos.com.ar/historiales[2]
- Los récords de enfrentamientos en **Tenis** (singles masculino) fueron extraídos de la página web www.atptour.com/en/players/atp-head-2-head
- **Rankings y resultados de equipos nacionales de Rugby Union** (world rugby, masculino) fueron recolectados de la página web www.world.rugby/tournaments/rankings/mru
- **Rankings y resultados de equipos nacionales de Hockey sobre Hielo** (iihf, masculino) fueron recolectados de la página web eloratings.net/icehockey
- La información de «mejores boxeadores de todos los tiempos» fue recolectada de las siguientes páginas web:
  a. boxrec.com/en/ratings (box rec)
  b. www.espn.com/sports/boxing/greatest/featureVideo?page=greatest110 (espn)
  c. www.sportco.io/article/greatest-boxers-of-all-time-577673 (sportco)
  d. www.marca.com/en/boxing/2022/02/15/620bf985268e3ea31b8b45dd.html (the ring)
  e. www.yardbarker.com/boxing/articles/the_best_pound_for_pound_boxers_of_all_time/... (yardbarker)
- La historia de **Rankings ATP de Tenis** (singles masculino) fue extraída de la página web www.atptour.com/en/rankings/singles
- Las posiciones finales de las ediciones analizadas de la **Premier League inglesa** (fa epl, *soccer*, masculino) fueron extraídas de la página web en.wikipedia.org/wiki/Premier_League
- Las posiciones finales de los últimos 12 campeonatos de la categoría **Turismo Carretera** (actc) fueron extraídas de la página web www.actc.org.ar/tc/campeonato.html
- La información histórica de **Rankings ELO de Ajedrez** fue recolectada de los sitios web ratings.fide.com, 2700chess.com y chess-results.com
- La información de **Rankings ELO ajustados** de la Liga norteamericana **de Básket** (nba) y de la Liga **de Fútbol Americano** (nfl, *gridiron football*) fue recolectada de la página web fivethirtyeight.com/sports

[2]No se debe caer en la trampa de creer que existe un error en este link, en Argentina, un indicador muy consultado por los fanáticos del Fútbol es la eficiencia promedio de sus equipos en los últimos 3 años, debido a que a través de ese guarismo se decide quiénes descienden a una categoría inferior al finalizar cada torneo; de ahí que el autor del sitio web haya decidido nombrarlo PROMIEDOS por el MIEDO a conocer con precisión este número y que lleve al hincha de ese equipo a experimentar una profunda tristeza.

- **Rankings y resultados de equipos nacionales de Fútbol** (fifa, *soccer*, masculino) fueron recolectados de los sitios web eloratings.net y www.fifa.com/en/fifa-world-ranking

NOTA: No todos los datasets anteriores fueron usados en la redacción final de la Tesis, aunque todos fueron analizados.

Referencias

[Ali et al., 1986] Ali, I., Cook, W. D., and Kress, M. (1986). On the minimum violations ranking of a tournament. *Management science*, 32(6):660–672.

[Alvo and Philip, 2014] Alvo, M. and Philip, L. (2014). *Statistical methods for ranking data*, volume 1341. Springer.

[Durrett, 1999] Durrett, R. (1999). *Essentials of stochastic processes*, volume 1. Springer.

[Elo, 1978] Elo, A. E. (1978). *The rating of chessplayers, past and present*. Arco Pub.

[Glickman, 1995] Glickman, M. E. (1995). The Glicko system. *Boston University*, 16:16–17.

[Govan et al., 2009] Govan, A. Y., Langville, A. N., and Meyer, C. D. (2009). Offense-defense approach to ranking team sports. *Journal of Quantitative Analysis in Sports*, 5(1).

[Harary and Moser, 1966] Harary, F. and Moser, L. (1966). The theory of round robin tournaments. *The American Mathematical Monthly*, 73(3):231–246.

[Hvattum and Arntzen, 2010] Hvattum, L. M. and Arntzen, H. (2010). Using Elo ratings for match result prediction in association football. *International Journal of forecasting*, 26(3):460–470.

[Keener, 1993] Keener, J. P. (1993). The Perron–Frobenius theorem and the ranking of football teams. *SIAM review*, 35(1):80–93.

[Kendall, 1955] Kendall, M. G. (1955). Further contributions to the theory of paired comparisons. *Biometrics*, 11(1):43–62.

[Langville and Meyer, 2012] Langville, A. N. and Meyer, C. D. (2012). *Who's# 1?: the science of rating and ranking*. Princeton University Press.

[Liern and Sala, 2009] Liern, V. and Sala, R. (2009). Un modelo ofensivo-defensivo para establecer un rating de equipos de fútbol. *Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA*, Actas_17.

[MMA, 2021] MMA, P. (2021). Fighter performance rating (fpr) explained by Kenny Florian & Sean O'Connell. URL https://www.youtube.com/watch?v=IyXn7gDW3xc.

[Sonas, 2002] Sonas, J. (2002). The Sonas rating formula-better than Elo. *ChessBase Magazine Online*.

[Vigna, 2016] Vigna, S. (2016). Spectral ranking. *Network Science*, 4(4):433–445.

[Vigna, 2019] Vigna, S. (2019). Spectral ranking.