

# **Ranking futbol with higher order networks**

Andres Chacoma\*

*Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales,*

*Departamento de Física. Buenos Aires, Argentina and*

*CONICET - Universidad de Buenos Aires,*

*Instituto de Física Interdisciplinaria y Aplicada. Buenos Aires, Argentina.*

Juan I. Perotti and Orlando V. Billoni

*Universidad Nacional de Córdoba, Facultad de Matemática,*

*Astronomía, Física y Computación. Córdoba, Argentina and*

*CONICET- Universidad Nacional de Córdoba,*

*Instituto de Física Enrique Gaviola. Córdoba, Argentina.*

## Abstract

We ...

### I. INTRODUCTION

En los últimos años, el estudio de las competencias deportivas ha experimentado una transformación significativa mediante la adopción de marcos teóricos propios de la ciencia de la complejidad. Los campeonatos deportivos, caracterizados por la interacción dinámica de múltiples agentes (jugadores, equipos, contextos), la no linealidad de los resultados y la emergencia de patrones colectivos no triviales, se comprenden cada vez más como sistemas adaptativos complejos. Esta perspectiva permite trascender el análisis puramente estadístico-descriptivo, para abordar cuestiones fundamentales sobre la dinámica competitiva, la evolución del rendimiento y la estructura subyacente del juego desde principios teóricos robustos [REF].

Esta intersección entre sistemas complejos y deporte es fértil desde una doble vertiente. Académicamente, ofrece un laboratorio único para validar y desarrollar teorías sobre redes, dinámicas de competición, procesos de difusión y emergencia de jerarquías en sistemas sociales con reglas claramente definidas. Desde el ámbito del alto rendimiento, la modelización cuantitativa y objetiva de la competición esta revolucionando hoy la toma de decisiones. Proporcionar a entrenadores, analistas y directivos herramientas basadas en datos para evaluar la performance real más allá del resultado inmediato, diagnosticar fortalezas y debilidades estructurales, y optimizar estrategias se ha convertido en un objetivo prioritario en la búsqueda de ventajas competitivas sostenibles [REF].

Si bien este enfoque se ha aplicado con éxito a diversos deportes de equipo e individuales como el baloncesto [REF], el béisbol [REF] o el tenis [REF], el fútbol destaca como un dominio particularmente rico. Su popularidad global, la disponibilidad creciente de datos de alto detalle (event data, tracking data), y la inherente complejidad táctica de un juego de flujo continuo con bajo marcador, lo convierten en un caso de estudio ideal. Investigaciones recientes han utilizado conceptos de sistemas complejos para modelizar la posesión del balón como un proceso de difusión, caracterizar la creatividad colectiva, o identificar fases de juego

---

\* achacoma@df.uba.ar

críticas, demostrando el gran potencial de este marco de trabajo [REF].

Dentro del arsenal metodológico de los sistemas complejos, la teoría de redes complejas se ha erigido como una herramienta especialmente poderosa para el análisis del fútbol. Al modelizar a los jugadores o equipos como nodos y sus interacciones (pases, enfrentamientos, comparaciones) como enlaces, es posible cuantificar y visualizar la estructura del juego. Esta aproximación ha permitido, por ejemplo, medir la centralidad de los jugadores en las redes de pase, identificar patrones tácticos característicos de un equipo, o evaluar la robustez de un sistema de juego ante lesiones, proporcionando insights que escapan a los análisis tradicionales [REF].

En este trabajo, proponemos una aplicación específica de la teoría de redes para abordar un problema central en el análisis deportivo: la comparación objetiva y multidimensional del rendimiento entre equipos. Nuestra propuesta consiste en definir un sistema de rating (clasificación) basado no solo en el resultado final, sino en un conjunto de métricas de rendimiento derivadas de datos de eventos. Para ello, construiremos una red por cada métrica, donde los nodos son los equipos y los enlaces ponderados codifican su rendimiento relativo en enfrentamientos directos. La aplicación de un método algebraico de descomposición de grafos (HodgeRank) sobre estas redes nos permitirá inferir un rating escalar para cada equipo en cada dimensión del juego, ofreciendo una comparación matizada y basada en el proceso. Las métricas concretas que alimentan este modelo se definen y justifican en la siguiente sección.

La creación de estos ratings multidimensionales y basados en procesos tiene una implicación práctica directa. Un ranking único, como la tabla de posiciones, resume un resultado pero puede enmascarar las causas detrás del mismo. Nuestro enfoque, al desglosar el rendimiento en componentes interpretables (e.g., presión, creación de ocasiones, tiempos de posesión), proporciona un diagnóstico estructural. Para un cuerpo técnico, esto significa poder identificar si un equipo acumula puntos a pesar de un bajo rendimiento en métricas clave, lo cual es señal de posible insostenibilidad, o por el contrario, si un buen rendimiento en el proceso no se traduce en resultados, lo cual puede ser señal de ineeficacia en momentos decisivos. Esta información es crucial para la toma de decisiones informadas, ya sea en el reclutamiento, el diseño táctico, o la preparación específica ante un rival, orientando los recursos hacia los aspectos del juego que realmente determinan la ventaja competitiva a largo plazo.

## II. DATA

### A. Recopilación de métricas

En este trabajo utilizamos la base de datos de evento provista por L. Pappalardo et al en [1]. En ese articulo, los autores visualizan todos los partidos de la temporada 2017-2018 de las principales ligas de fútbol europeas: La Liga (España), Premier league (Inglaterra), Serie A (Italia), Bundesliga (Alemania), Ligue 1 (Francia). Por cada partido detectan, clasifican, y localizan en tiempo y espacio todos los eventos: goles, tiros al arco, pases, saques de esquina, faltas, etc. En el sistema de referencia que utilizan para ubicar los eventos en tiempo y espacio,  $t$  expresa el tiempo transcurridos desde el inicio del partido, la coordenada  $x$  expresa la distancia respecto del arco del equipo creador del evento, y la coordenada  $y$  la distancia respecto a la banda lateral derecha. Las unidades de las coordenadas espaciales están dadas en porcentajes del campo juego, siendo por ejemplo  $x = 0$ ,  $x = 50$  y  $x = 100$  las posiciones de la linea de meta propia, la linea de centro del campo y la linea de meta del rival, respectivamente.

En este marco, definimos intervalo de posesión de pelota (BPI) como al conjunto dado por una secuencia continua de eventos generados por un equipo. Note que cada BPI contiene información de un solo equipo. Recopilamos todos los BPI de todos los equipos de cada liga, y sobre esos datos extrajimos métricas que nos permiten detectar algunos de los recursos tácticos que están utilizando los equipos en esa ventana temporal del partido. Las métricas recopiladas para nuestro análisis están basadas en las propuestas por J. Fernandez-Navarro en [2]. El estudio de estas metricas fue utilizado tambien en un trabajo anterior para determinar estilos futbolisticos caracterisiticos [3]. A continuación describimos en detalle a cada una de estas,

1. *Direct play.* Cada vez que hay un pase o un tiro libre en un BPI, medimos la velocidad media en la dirección de ataque, dada por el cociente entre la distancia recorrida por la pelota en el eje  $x$  y el tiempo transcurrido. De cada BPI tomamos el valor máximo. Esto nos permite detectar que tan directo hacia la portería rival es el movimiento de la pelota en el equipo.
2. *Counterattack.* Dado dos eventos consecutivos en un BPI, si el primero se observa en  $x_1 < 40$  y el segundo en  $x_2 > 60$  a una diferencia temporal  $\Delta t$ , se informa la velocidad

como  $v = \frac{x_2 - x_1}{\Delta t}$ . En otro caso se informa 0. Esto es una medida de que tan rápido un equipo pasa de una posición defensiva en su campo a una ofensiva en campo rival.

3. *Build up.* Si en un BPI se verifica que  $\bar{x} > 60$ , es decir la posesión se desarrolla mayoritariamente en campo rival, se informa el tiempo total de la posesión. En otro caso se informa 0. Esta métrica informa el tiempo de posesión en situaciones donde el equipo invade fuertemente el campo rival.
4. *Midfield play.* Si en un BPI se observa que  $\bar{x} \leq 60$  y  $\bar{x} \geq 40$ , es decir la posesión de desarrolla mayoritariamente por el centro del campo de juego, se informa el tiempo total de la posesión, en otro caso se informa 0. La idea de esta variable es medir el tiempo que el equipo pasa en el sector medio del campo de juego.
5. *Flow rate.* En cada BPI donde  $\bar{x} \geq 50$  se toma la diferencia temporal entre todos los eventos, y se calcula el valor medio,  $\bar{dt}$ . Luego se define la métrica como  $1/\bar{dt}$ . De esta manera se tiene una medida de que tan rápido el equipo mueve la pelota en el campo rival.
6. *Crossing.* Si en un BPI se observa un evento centro, se informa 1, en otro caso se informa 0. Esta métrica sirve para contabilizar los intentos de llegada por vía aérea.
7. *Pressure point.* De cada BPI se toma el primer evento, y se extrae la posición en la variable  $x$ , es decir donde el equipo comienza su posición. Esto nos permite medir si el equipo esta recuperando la pelota en su campo, en la zona media o en el campo rival.
8. *Shots.* Si en un BPI se registra un evento “Shot”, se informa 1, en otro caso se informa 0. Esta métrica permite contabilizar los tiros al arco de cada equipo.

Para nuestro análisis, se descartaron todas los BPI con menos de 3 eventos y con tiempo total menor a 2 segundos. La idea de esto es descartar pequeñas recuperaciones pasajeras y quedarnos con posesiones consolidadas. Del proceso de recopilación se obtuvieron 215681 BPI. Luego de calcular los valores de las métricas en cada uno de estos, estudiamos la distribución de los datos. Observamos que las métricas parecen seguir una distribución tipo log-normal, por lo tanto decidimos transformar los datos como  $x \rightarrow \log(1 + x)$  para trabajar con distribuciones aproximadamente normales. Posteriormente, agrupamos la información por partido y por equipo, y sumamos los valores obtenidos en cada métrica. De esta manera

por ejemplo el feature *Shots* cuantifica la cantidad de tiros al arco ejecutados por el equipo en ese partido. Asimismo el feature *Build-up* cuantifica la cantidad de tiempo neto en la cual un equipo sostuvo una posición de ataque frente al rival en ese partido. Note que en la temporada 2017/2018, en las ligas Española, Inglesa, Francesa e Italiana los equipos jugaron 38 partidos. Por lo tanto, al tomar los datos de los primeros 4 equipos cada liga aporta un total de  $38 \times 4 = 152$  muestras al archivo de datos. Asimismo, en la liga alemana, al haber menos equipos, se jugaron 34 partidos, por lo tanto esta liga aporta 136 muestras. En consecuencia, la matriz de datos consta de 744 filas y 10 columnas. Por ultimo, en un dataset aparte recopilamos meta-data asociada a cada muestra, útil luego para realizar el análisis: a que equipo pertenece esa muestra, cual es la liga de pertenencia, y el resultado final en la tabla de posiciones.

## B. Representación en redes complejas

En lo que sigue, presentamos nuestra propuesta para representar las métricas de rendimiento en términos de redes complejas. Definimos  $M(i, j, g)$  como la métrica de rendimiento correspondiente al equipo  $i$  cuando enfrenta al equipo  $j$  en el partido  $g$ . Por ejemplo, puede representar la cantidad de tiros al arco realizados por el FC Barcelona al jugar contra el Real Madrid en el primer encuentro del torneo español *La Liga*. En nuestro conjunto de datos, todos los equipos participaron en un formato de liga todos contra todos, enfrentándose dos veces: un partido de ida ( $g_1$ ) y otro de vuelta ( $g_2$ ). Utilizando la información de ambos encuentros, definimos una métrica agregada que resume el desempeño observado entre esos dos equipos a lo largo del torneo:

$$M(i, j) = \sum_{g_1, g_2} M(i, j, g).$$

En el ejemplo anterior,  $M(i, j)$  representa la cantidad total de tiros al arco realizados por el FC Barcelona contra el Real Madrid en ambos partidos del torneo. Calculando  $M(i, j)$  para cada par de equipos en una liga  $L$ , es posible representar estas relaciones de desempeño mediante un grafo dirigido y ponderado  $G(L, M)$ , cuyos pesos se definen como

$$f_{ij} = M(j, i) - M(i, j).$$

Nótese que, en esta representación,  $f_{ij} < 0$  indica que el equipo  $i$  superó al equipo  $j$  en la métrica considerada. En este marco, construimos 50 grafos en total, 8 grafos por liga. Cada uno asociado a una métrica de rendimiento distinta.

### III. THEORY

### IV. RESULTADOS

#### A. Estadística del rating real

La idea de esta sección es definir un modelo estadístico para la distribución de probabilidad del rating real,  $R_T$ , definido como la cantidad de puntos totales obtenidos por un equipo durante la liga.  $R_T$  es una variable estocástica que depende de la cantidad de partidos ganados, empelados y perdidos por los equipos de la liga. En las ligas de fútbol, un equipo obtiene 3 puntos cuando gana, 1 cuando empata y 0 cuando pierde. En este marco,  $R_T$  puede expresarse en términos de las variables estocásticas  $W$  y  $D$ , que representan el número de partidos ganados y empelados, respectivamente,

$$R_T = 3W + D.$$

Sea  $n$  el número total de partidos en la temporada. Modelamos  $W$  con una distribución binomial,

$$W \sim Bin(n|p_w)$$

donde  $p_w$  es la probabilidad de ganar un partido. Condicional al número de victorias  $W = w$ , los empates siguen,

$$D|W = w \sim Bin(n - w|\frac{p_d}{1 - p_w}),$$

con  $p_d$  probabilidad de empatar. Dado que los equipos tienen distinto nivel competitivo, cada equipo posee probabilidades intrínsecas  $p_w$  y  $p_d$  de ganar y empatar, respectivamente. Modelamos estas probabilidades mediante un enfoque jerárquico: cada equipo tiene un par de habilidades latentes  $\boldsymbol{\eta} = (\eta_w, \eta_d)^T$ , de ganar o empatar, que siguen una distribución normal bivariada,

$$\boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

$$\boldsymbol{\mu} = (\mu_w, \mu_d)^T, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_w^2 & \rho\sigma_w\sigma_d \\ \rho\sigma_w\sigma_d & \sigma_d^2 \end{pmatrix},$$

donde  $\boldsymbol{\mu}$  es el centro de la distribucion,  $\boldsymbol{\Sigma}$  la matriz de covarianza. En este marco,  $\rho$  representa la correlación entre las habilidades para victoria y empate. Las probabilidades se obtienen mediante la transformación softmax,

$$p_w = \frac{e^{\eta_w}}{1 + e^{\eta_w} + e^{\eta_d}}, \quad p_d = \frac{e^{\eta_d}}{1 + e^{\eta_w} + e^{\eta_d}}.$$

Esta transformación garantiza que  $p_w + p_d \leq 1$  y que el modelo sea identificable al fijar implícitamente  $\eta_l = 0$  para las derrotas. Con estos elementos, podemos escribir la distribucion conjunta de victorias y empates marginalizando sobre los efectos aleatorios  $\boldsymbol{\eta}$ ,

$$P(W = w, D = d) = \int_{\mathbb{R}^2} \text{Bin}(n|p_w(\boldsymbol{\eta})) \text{Bin}(n - w|\frac{p_d(\boldsymbol{\eta})}{1 - p_w(\boldsymbol{\eta})}) \phi(\boldsymbol{\eta}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) d\boldsymbol{\eta}, \quad (1)$$

Notar que la integral en 1 no tiene forma cerrada, para resolverla utilizamos el metodo de monte carlo. Con estos elementos, podemos finalmente escribir la distribucion de probabilidad teorica para el true rating como,

$$P(R_T = r_T) = \sum_X P(W = w, D = d), \quad (2)$$

$$X = \{w, d > 0, 3w + d = r_T, w + d \leq n\}.$$

Los grados de libertad que otorgan los parametros  $\mu_w$ ,  $\mu_d$ ,  $\sigma_w$ ,  $\sigma_d$  y  $\rho$ , permiten ajustar la curva teorica 2 a la curva empirica. El ajuste fue realizado sobre el conjunto de datos asociado a las ligas Inglesa, Francesa, Italiana y Española. Para este analisis, decidimos no utilizar los datos de la liga Alemana, ya que su torneo se disputa entre un numero menor de equipos (18), y por lo tanto se genera una distribucion de valores de true rating levemente diferente a las otras ligas, en las cuales participa un total de 20 equipos por liga. Para hacer el ajuste, utilizamos el algoritmo Nelder-Mead minimizando el RMSE entre las CDF empírica y teórica. Posteriormente, con los parámetros óptimos, realizamos bootstrap paramétrico (1000 réplicas) para estimar el valor medio y la incertidumbre de las probabilidades  $p_w$  y  $p_d$ . Para la probabilidad de victoria obtuvimos  $\bar{p}_w = 0.366$  ( $SD = 0.010$ ) con un intervalo de confianza del 95% de [0.3527, 0.3798]. Asimismo, para la probabilidad de empate obtuvimos  $\bar{p}_d = 0.280$  ( $SD = 0.020$ ) con un intervalo de confianza del 95% de [0.2548, 0.3074].

Para contrastar estos resultados con los datos empiricos, calculamos para cada equipo  $i$  las probabilidades de victoria y empate empiricas,  $q_w$  y  $q_d$ , utilizando la informacion de los resultados de los partidos jugados,

$$q_w^{(i)} = \frac{\#wins^{(i)}}{\#matches^{(i)}}, \quad q_d^{(i)} = \frac{\#draws^{(i)}}{\#matches^{(i)}}.$$

Calculando el valor medio y el desvio estandar sobre el conjunto de todos los equipos, obtuvimos para la probabilidad de victoria  $\bar{q}_w = 0.379$  ( $SD = 0.168$ ) con un intervalo de confianza del 95% de [0.1572, 0.7638], y para la probabilidad de empate  $\bar{q}_d = 0.243$  ( $SD = 0.081$ ) con un intervalo de confianza del 95% de [0.1046, 0.3954]. Podemos observar lo siguiente: (i) Las medias son similares (0.366 vs 0.379 para victorias, 0.280 vs 0.243 para empates), lo que valida el centro de la distribución; (ii) las desviaciones estándar empíricas (0.168, 0.081) son mucho mayores que las implícitas en el modelo (0.01, 0.02), indicando que el modelo subestima la heterogeneidad entre equipos.

Con respecto al parametro de correlacion, del boostrapping parametrico obtuvimos  $\bar{\rho} = -0.130$  ( $SD = 0.07$ ) con un intervalo de confianza del 95% de [-0.3040, -0.0371]. Esto indica la presencia de una correlacion estructural negativa debil entre las habilidades de victoria,  $\eta_w$ , y empate,  $\eta_d$ , de los equipos. Esto sugiere que equipos con mayor habilidad para ganar tienden ligeramente a tener menor habilidad para empatar.

En la Fig. 1 (a), mostramos la CDF del true rating junto con el ajuste. Podemos ver que el modelo propuesto captura muy bien el comportamiento de la curva en todo el soporte. En Fig. 1 (b) mostramos un grafico con la relacion entre los quantiles teoricos y de los datos. A modo de referencia, mostramos tambien la relacion entre los quantiles de los datos y los asociados a una distribucion gausiana con media y desvio estandar igual a los valores muestrales obtenidos de los datos,  $\bar{R}_T = 52.45$  ( $SD = 18.56$ ). En primer lugar, podemos ver que el modelo propuesto captura bien el comportamiento en la zona central izquierda y de manera aceptable el comportamiento de la cola de la derecha. La comparación con el modelo gaussiano nos permite ver que la distribucion real presenta colas mas livianas a la izquierda y mas pesadas a la derecha, respecto del comportamiento gaussiano, develando así la presencia de una asimetria en la distribucion. La concavidad hacia abajo en el centro del grafico, indica tambien que la mediana de los datos esta corrida hacia la izquierda.

## B. Comparacion entre el true rating y el rating obtenido de las metricas

En esta sección, caracterizamos las diferencias entre el true rating, dado por el sistema de puntos de las ligas, y el rating obtenido a partir del metodo de HodgeRank, al cual llamaremos metric rating. Dado que son ocho las metricas utilizadas (ver sección II A), por cada equipo calcularemos ocho valores de metric rating distintos, los cuales compararemos con el true rating. Es importante remarcar que el true rating y el metric rating están expresados en escalas de valores distintas. El primero está expresado en cantidad de puntos obtenidos y el segundo en un número estandarizado. No obstante esto no nos impide compararlos, ya que lo importante de estos valores no es el valor absoluto que obtiene cada equipo, si no más bien el valor relativo entre ellos. Por tal motivo, para comparar el comportamiento del true rating con el metric rating, decidimos estandarizar ambas variables y trabajar con el *z-score* de cada una de estas.

En la Fig. 2 (a) mostramos las CDF del true rating asociado a los datos de todos los equipos y todas las ligas, con la del metric rating asociado a los datos de todos los equipos de todas las ligas y de las ocho metricas estudiadas. A ambas curvas se les estraeran los valores outliers considerados como aquellos casos que se encuentran por encima del quantile  $Q_{99.7}$ . Podemos observar una coincidencia notable en casi todo el rango, observando algunas diferencias en las colas: el metric rating muestra valores positivos mas alejados de la media, y el true rating muestra valores negativos levemente mas alejados de la media. En la Fig. 2 (b) mostramos la distribucion de valores de metric rating de cada una de las ligas. Para esto, por liga, agrupamos los valores de metric rating asociado a todas las metricas y estandarizamos los valores utilizando el z-score. A modo de referencia se grafica tambien, en linea negra punteada, la CDF de una distribucion normal estandar. Puede observarse que las curvas asociadas a las diferentes ligas difieren levemente, indicando un comportamiento dependiente de la liga. Asimismo, para todos los casos podemos ver una acumulacion de probabilidad en la zona media. Esto se compensa con colas livianas a izquierda y pesadas a la derecha, respecto de la distribucion normal estandar. Por otro lado, en la Fig. 2 (c) graficamos las curva true rating y metric rating en función del true ranking, para el caso de la metrica *Pressure Point* en la liga Inglesa. La idea es resaltar las diferencias observadas equipo a equipo. En primer lugar podemos notar que el true rating decrece a medida que crece el rating. El decrecimiento no es monotono por que puede haber regiones constantes,

pero es estrictamente no creciente. Esto es trivial por defincion: el primer equipo tiene un true raiting mayor o igual que el segundo, el segundo tiene un true raiting mayor o igual que el tercero, y asi hasta el ultimo puesto. En la curva de metric rating observamos una tendencia tambien decreciente, pero no estricta. En este caso podemos decir, en terminos cualitativos, que el metric rating sigue la tendencia del true rating, aunque se observan pequeñas diferencias locales. En la Fig. 2 (d) mostramos una comparacion analoga a la anterior para el caso de la metrica *Direct play* en la liga española. En este caso vemos que la curva de metric rating no sigue una clara tendencia decreciente y se observa una fuerte descorrelacion con los valores de true rating. Vemos entonces que el true rating de una dada liga puede diferir poco o mucho del metric rating dependiendo de la metrica que utilizamos para inferir el rating.

Nuestro objetivo a continuacion es cuantificar la correlación entre la curva de true rating de una liga y las curvas de metric rating asociadas a cada metrica. Queremos ver con que metrica se obtiene el metric rating que mejor sigue la tendencia del true rating, y si eso depende de la liga. Para esto, calculamos el coeficiente de correlacion de Pearson entre el true rating de cada liga y el metric rating de cada metrica. En Fig. 3, mostramos los resultados. Cada panel contiene la informacion asociada a una metrica, y dentro de los paneles, cada barra contiene la informacion asociada a una liga. Los paneles se muestran ordenados de mayor a menor por correlacion promedio y las barras, dentro de cada panel, ordenadas de mayor a menor por correlacion. Las barras con hatch indican que para esos casos no se puede asegurar significancia estadistica. En primer lugar podemos ver que las ligas italiana e inglesa muestran la mayor correlacion en todas las metricas. En segundo lugar, vemos que las ligas francesa y alemana se alteran el tercer y cuarto puesto entre los casos con valor de correlacion intermedio en casi todas las metricas. Por ultimo, vemos que la liga española muestra los peores resultados, mostrando el peor valor de correlacion en 7 de las 8 metricas. Aun asi, presenta valores que indican correlacion fuerte ( $\rho > 0.4$ ) en los casos de las metricas *Flow rate* y *Counterattack*.

### C. Comparacion entre el true ranking y el ranking obtenido de las metricas

En esta sección complementamos el analisis de la seccion anterior estudiando la correlacion entre rankings. Definimos el true ranking de un equipo, como su lugar en la tabla de

posiciones final de la liga. Note que este se obtiene al ordenar los equipos de la liga de manera descendente por cantidad de puntos obtenidos. Si hay empate, la diferencia entre goles a favor y goles en contra define el ranking final. Analogamente, definimos el metric ranking ordenando los equipos de una liga de manera descendente por valor de metric rating. En este caso, no se detectaron empates durante el analisis. Para estudiar la correlacion entre los rankings, utilizaremos el coeficiente de correlacion de rango de Kendall [4], el cual se basa en la realización de una prueba de hipotesis no parametrica cuyo estadístico,  $\tau \in [-1, 1]$ , arroja un valor alto cuando las observaciones tienen un rango similar y cercano a cero cuando tienen un rango diferente. En Fig. 4, mostramos los resultados. Igual que en la Fig. 3, cada panel contiene la información asociada a una metrica, y dentro de los paneles, cada barra contiene la información asociada a una liga. Los paneles se muestran ordenados de mayor a menor por coeficiente de Kendall promedio. Las barras, dentro de cada panel, se muestran ordenadas de mayor a menor por valor del coeficiente. Las barras con hatch indican que para esos casos no se puede asegurar significancia estadistica. En primer lugar, podemos observar que la liga italiana lidera todos los paneles mostrando el coeficiente de Kendall mas alto. La liga inglesa en general queda en segundo lugar, superada en dos casos por la liga francesa. La liga alemana y la española se muestra en general en los ultimos puestos. Este comportamiento es en principio consistente con lo observado para los ratings. El segundo lugar podemos observar que los coeficientes de Kendall son ligeramente menores a los valores de correlacion de Pearson. Esto es debido a su naturaleza no paramétrica. Esta divergencia no implica una correlación débil, sino que refleja la volatilidad del ranking en tramos donde la diferencia de puntos es mínima. Mientras Pearson valida la jerarquía general de la liga, Kendall señala que el orden exacto de los equipos es sensible a pequeños cambios (ruido), especialmente en zonas de la tabla con alta densidad de puntos, es decir equipos empatados.

#### D. Composite rating

En esta seccion, desarrollamos para cada liga un Composite Rating (CR), definido como una combinacion lineal de los metric rating obtenidos de las metricas de rendimiento. La idea es proporcionar una evaluacion mas general de la calidad de los equipos. Sea  $n$  la cantidad de equipos en una liga y  $m$  la cantidad de metricas, definimos  $R^{n \times m}$  como la matriz donde las filas contienen los valores de metric rating obtenidos para cada equipo.

Asimismo definimos el  $T^{n \times 1}$  como el vector que contiene los true ratings de cada equipo de la liga. En este marco, nuestro objetivo es encontrar el vector de pesos  $\alpha^{m \times 1} = (\alpha_1 \dots \alpha_m)^T$  tal que,

$$T = \alpha R.$$

Para resolver este sistema, y con la de idea de garantizar la parsimonia y la capacidad explicativa del composite rating, empleamos un algoritmo de selección exhaustiva de características (EFS). Luego de normalizar las metricas mediante un escalado estandar (*z-score*), este procedimiento evalua todas las combinaciones posibles de las ocho métricas de rendimiento iniciales, desde modelos univariados hasta el modelo completo con ocho variables. Con el fin de evitar el sobreajuste, utilizamos el criterio de información bayesiano (BIC) como función de selección. El BIC permite identificar el modelo que maximiza la verosimilitud de los datos introduciendo una penalización proporcional a la complejidad del modelo (número de variables), favoreciendo así la estructura más eficiente. En este marco, el modelo óptimo seleccionado fue aquel que minimiza el valor del BIC. De esta manera aseguramos que el rating final sea estadísticamente robusto y representativo del rendimiento real de los equipos.

Finalmente, para determinar el composite rating en cada liga, se ajustó un modelo de regresión por mínimos cuadrados ordinarios (OLS) utilizando las métricas seleccionadas por el algoritmo EFS. Nopre que, dado que las variables fueron estandarizadas previamente, los coeficientes obtenidos representan la importancia relativa de cada dimensión en la predicción de los puntos obtenidos en la liga. En la tabla I, se muestran los resultados de la regresion. El análisis de los coeficientes revela áreas críticas para la mejora del rendimiento. Para la liga inglesa, podemos ver que las metricas mas relevantes, en orden de importancia relativa, son *Build up*, *Flow rate* y *Presure point*, las cuales explican casi el 90% de la varianza de los datos. El coeficiente negativo en la métrica *Build up* (-24.56) sugiere que los equipos que reducen sus tiempos de elaboración en el ultimo cuarto del campo de juego (ver definicion de la metrica en section II A), incrementan significativamente su probabilidad de obtener puntos. Por el contrario, un aumento en esta métrica penaliza el composite rating, lo que indica una ineficiencia en estilos de juego que utilicen esta tactica en esa liga en particular. Tambien podemos decir que el coeficiente positivo en *Flow rate* (22.69) sugiere que los equipos que incrementan la velocidad con la que mueven la pelota en el campo de juego, incrementa sus chances de obtener más puntos en esta liga. Asimistmo, el coeficiente positivo en la

metrica *Pressure point* sugiere que los equipos que presionan mas alto en el campo de juego incrementa su probabilidad de obtener puntos.

Con respecto a la liga francesa, haciendo un analisis similar basado en el analisis de los coeficientes de la regresion, podemos ver que los equipos que presionan alto en el campo de juego y disminuyen su juego directo, aumentan sus chances de obtener puntos. Por otro lado, el caso de la liga alemana es probablemente el más curioso, note que con una sola variable, *Direct play*, es posible explicar el 46% de la varianza. Es decir, entender esta variable, nos permite intuir por qué un equipo gana o pierde puntos en esa liga. En este caso encontramos un modelo extremadamente parsimonioso.

Lo obtenido para la liga italiana sugiere que el exito estaria asociado a un incremento en las variables *Flow rate* y *Direct play* y un decremento en *Midfield play*. Es decir, los equipos más exitosos de esta liga parecen estar relacionados con un estilo de juego que fomenta el movimiento rapido de la pelota y relega el tiempo de control en la mitad del campo. Por ultimo, en el caso de la liga española podemos ver que un estilo de juego asociado a la obtencion de puntos parece estar relacionado con una disminucion de los niveles de presion alta, y un aumento de la velocidad con la que se mueve la pelota y de los contrataques.

El composite rating obtenido para cada liga, puede ser comparado con el true rating como se hizo en la secciones anteriores con los ratings asociados a las metricas individuales. En Fig. 5 paneles (a), (b), (c), (e) y (e) mostramos la comparacion liga a liga, equipo a equipo, en un grafico rating contra true ranking. En el caso de la liga inglesa, en Fig. 5 (a), podemos ver que para los primeros dos equipos el true rating es mayor al composite rating. Estos equipos muestran entonces un supra-rendimiento estadistico, parecen estar rindiendo mas de lo que deberian. Esto puede deberse a diversos factores externos atribuibles al azar. Contrariamente el equipo que quedo tercero en la tabla, muestra un true rating menor al composite rating, lo cual indica un comportamiento infra-estadistico, es decir, parece ser que hicieron lo estadisticamente correcto pero fallaron en la ejecución o simplemente tuvieron “mala suerte”. Ejemplos similares se observan en todas las ligas, en todas las posiciones de la tabla. En terminos generales podemos decir que las curvas de composite rating siguen una tendencia similar a las curvas de true rating, destacandose a ojo denudo los casos de la liga inglesa e italiana. Para concluir el analisis, en Fig. 5 (f) mostramos un grafico de barras con los valores de los coefientes de correlacion de Pearson,  $\rho$ , y del coeficiente de Kendall,  $\tau$ . En todos los casos, podemos ver que los valores obtenidos son superiores a los obtenidos a partir

de las métricas individuales (ver Fig. 3 y Fig. 4). Destacamos principalmente los valores obtenidos en la liga Italiana y las mejoras significativas en las ligas española y alemana.

## V. CONCLUSION

En este trabajo se presenta un marco metodológico unificado que integra métricas avanzadas de rendimiento en fútbol, teoría de grafos y un método algebraico robusto (HodgeRank) para la generación de ratings o clasificaciones de equipos.

Como base inicial del análisis y con el objetivo de establecer un fundamento sólido para la evaluación posterior, se estudió la estadística de la clasificación real, definida por los puntos obtenidos por los equipos en sus respectivas ligas. Para ello, se propuso un modelo jerárquico en el que a cada equipo se le asocian tres parámetros de habilidad específicos: para ganar, empatar y perder. De manera análoga a sistemas de clasificación como el método Elo [5], estas habilidades definen una distribución de probabilidad para los resultados de un partido (victoria, empate, derrota). Mediante un procedimiento de ajuste numérico estándar, se obtuvieron dichas probabilidades teóricas, las cuales mostraron una concordancia satisfactoria con las frecuencias empíricas observadas, validando así la idoneidad del modelo propuesto. Para la generación de ratings basado en métricas, se recopilaron ocho métricas de rendimiento a partir de una base de datos de eventos de la temporada 2017/2018 de las cinco principales ligas europeas (Inglaterra, Francia, Alemania, Italia y España). Con base en estas métricas, se construyeron grafos no dirigidos ponderados, donde los nodos representan los equipos y los pesos de las aristas codifican la información de rendimiento relativo para una métrica dada. Esto permitió generar ocho redes por liga. A cada grafo se aplicó la descomposición de Hodge, que utiliza el gradiente del grafo como mecanismo para derivar un rating para cada equipo. De este modo, se obtuvieron ocho ratings por liga (uno por métrica), los cuales se compararon con la clasificación real. El análisis del comportamiento de estos ratings en relación con el true rating se realizó mediante el coeficiente de correlación de Pearson, para evaluar la tendencia global, complementado con el coeficiente de correlación de rangos de Kendall, para examinar la similitud en la ordenación de los equipos. Los resultados revelaron patrones claros y consistentes, que evidencian una dependencia tanto de la liga como de la métrica utilizada. En particular, se observó que las ligas inglesa e italiana presentan los mayores niveles de similitud entre los ratings derivados

de métricas individuales y la clasificación real, lo que sugiere que, en estas competiciones, las métricas consideradas son predictoras individualmente efectivas del rendimiento global. Por el contrario, la liga española mostró los menores niveles de correspondencia, con resultados notablemente inferiores en la mayoría de los casos. Esto indica que, para esta liga, las métricas propuestas poseen una capacidad predictiva limitada y es necesario explorar indicadores alternativos o complementarios. Estas diferencias dependientes de la liga apuntan, además, a la existencia de estilos de juego o dinámicas competitivas distintivas, un fenómeno ya documentado en la literatura [3]. Finalmente, una vez caracterizados los ratings individuales, se definió para cada liga un rating agregado, construido como una combinación lineal ponderada de los ocho ratings basados en métricas. Los pesos se optimizaron mediante un procedimiento de ajuste dirigido a identificar la combinación más parsimoniosa que mejor aproxime la clasificación real. Esta metodología nos permitió obtener un rating compuesto más robusto, con el cual se mejoró significativamente la capacidad predictiva de las métricas individuales, siendo destacable la mejora en los casos de la liga española y alemana. Además, esta técnica permitió cuantificar la relevancia relativa de cada métrica en el contexto de cada liga, aportando así una visión integrada del rendimiento de los equipos.

Es importante remarcar que este trabajo no propone un sustituto para la tabla de posiciones oficial, cuyo valor competitivo y legitimidad deportiva no están siendo cuestionados. Somos conscientes de que la naturaleza azarosa e impredecible del fútbol, que la clasificación oficial a menudo refleja, constituye una dimensión inherente a su atractivo. El objetivo de nuestro trabajo, en cambio, es proporcionar una herramienta analítica complementaria que, al desentrañar los patrones estadísticos subyacentes al rendimiento, ofrezca una perspectiva más rica y objetiva sobre la dinámica competitiva de una liga. Esta herramienta podría ser empleada por los cuerpos técnicos para una evaluación más integral de la fortaleza real de los equipos, más allá de la contingencia de los resultados inmediatos. Con respecto a las aplicaciones directas, el marco metodológico presentado es generalizable a cualquier deporte de confrontación directa y a cualquier conjunto de métricas de rendimiento, superando así una limitación clave de los modelos de rating tradicionales que se fundamentan exclusivamente en los resultados (victorias, empates, derrotas). Al basarse en indicadores de rendimiento general del proceso competitivo, nuestro enfoque proporciona una evaluación potencialmente más estable y con mayor capacidad predictiva a medio plazo, al ser menos sensible a la aleatoriedad de resultados puntuales. Específicamente, el composite rating propuesto per-

mite identificar discrepancias significativas entre el rendimiento observado en el campo y la puntuación obtenida. Esta capacidad es de gran utilidad práctica: por un lado, puede señalar equipos cuyo rating es alto pero que acumulan pocos puntos, lo que podría indicar una desempeño sólido acompañado de mala fortuna o deficiencias en momentos decisivos del juego. Por otro lado, puede detectar equipos con un rating bajo que, sin embargo, ocupan posiciones elevadas en la tabla, lo que sugeriría un rendimiento sustentado en factores aleatorios favorables. Identificar estos casos de rendimiento infra o supravalorado estadísticamente ofrece a los entrenadores y analistas información valiosa para ajustar estrategias, gestionar expectativas y fundamentar decisiones técnicas con una base cuantitativa más sólida.

Las líneas de investigación futuras derivadas de este trabajo son múltiples. En primer lugar, sería fundamental profundizar en el análisis de las diferencias inter-liga observadas. La incorporación de variables contextuales —como presupuestos, localia, o modelos tácticos predominantes— permitiría diseñar un modelo explicativo robusto sobre cómo las distintas culturas futbolísticas se reflejan en los datos de rendimiento y, en consecuencia, en los ratings derivados. Este análisis contextual no solo serviría para explicar diferencias, sino también para alimentar el desarrollo de modelos más adaptativos. En este sentido, una segunda línea prioritaria sería refinar el composite rating mediante la incorporación de técnicas de aprendizaje automático y modelos no lineales. Estos métodos podrían optimizar dinámicamente los pesos de cada métrica en función del contexto específico de cada liga o temporada, superando así el enfoque de ponderación fija. Asimismo, dado el potencial demostrado del modelo para identificar equipos cuyo rendimiento estadístico diverge de su puntuación real, un desarrollo natural y de aplicación directa sería la creación de un sistema de alerta temprana. Dicho sistema, al integrar datos en tiempo real, podría cuantificar de manera continua el “factor suerte” o la eficacia en momentos clave, ofreciendo a los cuerpos técnicos una herramienta analítica para ser usada durante el transcurso de la liga. En tercer lugar, sería de gran interés validar y extender el marco metodológico propuesto a otros deportes de equipo con dinámicas de confrontación similar. La generalización a disciplinas como el baloncesto o el hockey, utilizando sus métricas de rendimiento específicas, probaría la robustez del enfoque y permitiría realizar estudios comparativos inter-deportivos. Por último, una oportunidad metodológica significativa y aún no explorada en nuestro análisis reside en el aprovechamiento integral de la descomposición de Hodge. Más allá del componente de gradiente utilizado para generar los ratings, los componentes cíclico y armónico ofrecen

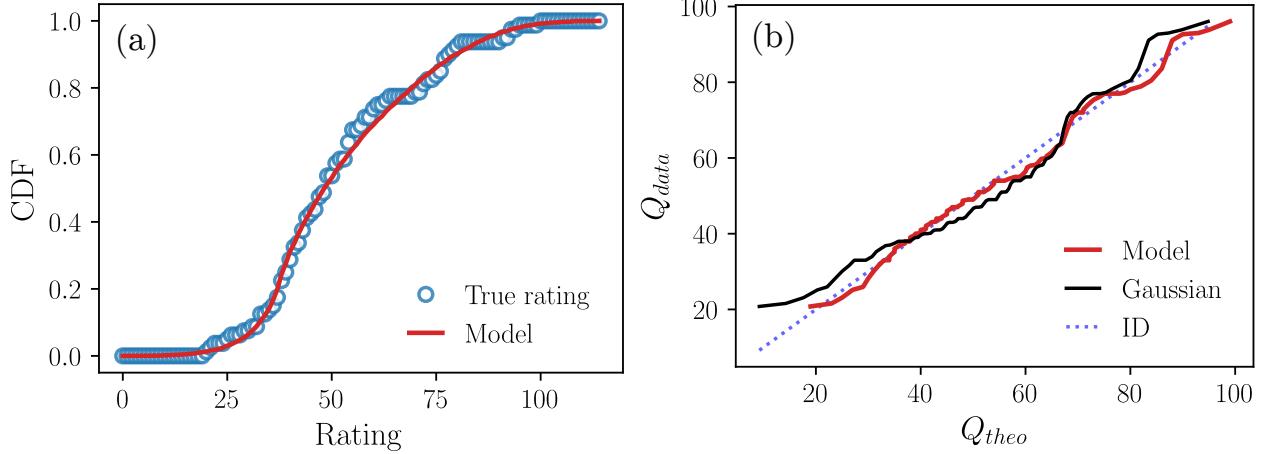


FIG. 1. Estadística del rating real. (a) Comparación de la Cumulative distribution function (CDF) de los valores de rating real con lo obtenido a partir del modelo. (b) Relacion entre los Quantile asociado a los datos y los obtenidos a partir del modelo. A modo de referencia, se agrega tambien una comparación con una distribucion gaussiana de media y desvio estandar igual al de los datos.

información estructural valiosa. El componente cíclico, que captura las inconsistencias en las comparaciones por pares, podría cuantificar el grado de competitividad intrínseca o impredecibilidad de una liga. Por otro lado, el componente armónico, inherente a la topología del grafo, ofrece una perspectiva estructural. Aunque en un grafo completo de una liga su contribución es nula, su potencial analítico emerge al redefinir la conectividad. Por ejemplo, si los nodos (equipos) se conectan no solo por enfrentamientos directos, sino por similitud en su perfil de métricas (creando un grafo de similitud táctica), el componente armónico podría revelar agrupaciones naturales o comunidades de equipos con estilos de juego afines. Esto permitiría caracterizar la liga no como una jerarquía lineal, sino como una red con posibles “clusters” tácticos, enriqueciendo sustancialmente la caracterización de la competición más allá del ranking. La integración del análisis cuantitativo de estos componentes enriquecería sustancialmente la caracterización de una competición, añadiendo una capa de interpretación profunda a la mera clasificación.

---

[1] Luca Pappalardo, Paolo Cintia, Alessio Rossi, Emanuele Massucco, Paolo Ferragina, Dino Pedreschi, and Fosca Giannotti. A public data set of spatio-temporal match events in soccer competitions. *Scientific data*, 6(1):236, 2019.

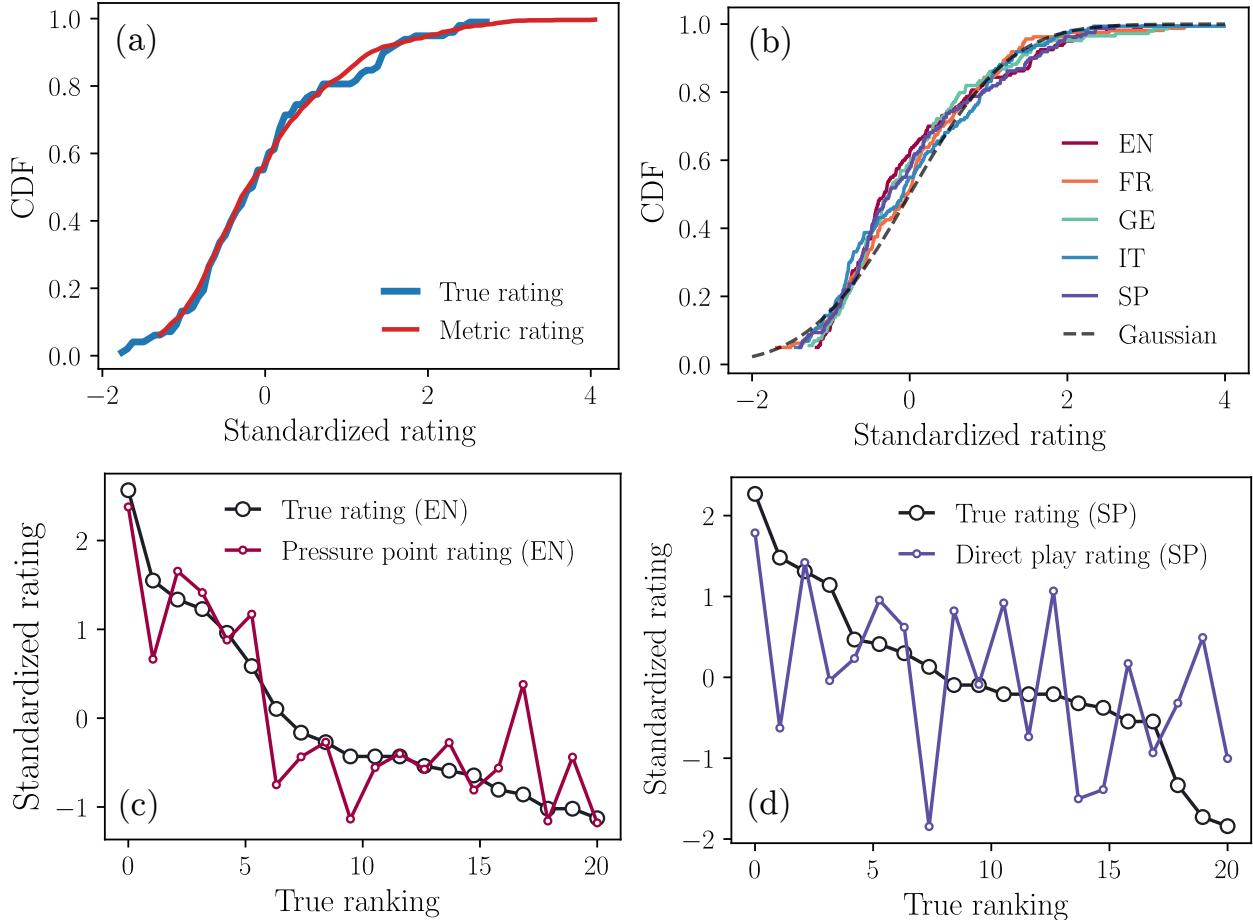


FIG. 2. Comparacion entre true rating y metric rating. Debido a que estas cantidades estan en unidades distintas, para comparar se usan los datos estandarizados. (a) Comparacion de las cumulative distribution function asociadas al true rating y al metric rating. (b) Cumulative distribution function del metric rating por liga. En lineas negras punteadas se grafica a modo de referencia la curva de una distribucion normal estandar. (c) Comparación en el caso Presure point en la liga inglesa. (d) Comparación en el caso Direct play en la liga española.

- [2] Javier Fernandez-Navarro, Luis Fradua, Asier Zubillaga, and Allistair P. McRobert. Influence of contextual variables on styles of play in soccer. *International Journal of Performance Analysis in Sport*, 18(3):423–436, 2018.
- [3] Andres Chacoma and Orlando V Billoni. Data-driven approach to defining football styles in major leagues. *Chaos, Solitons & Fractals*, 200:116926, 2025.
- [4] Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1-2):81–93, 1938.
- [5] Arpad E. Elo. *The Rating of Chess Players, Past and Present*. Arco Publishing, New York,

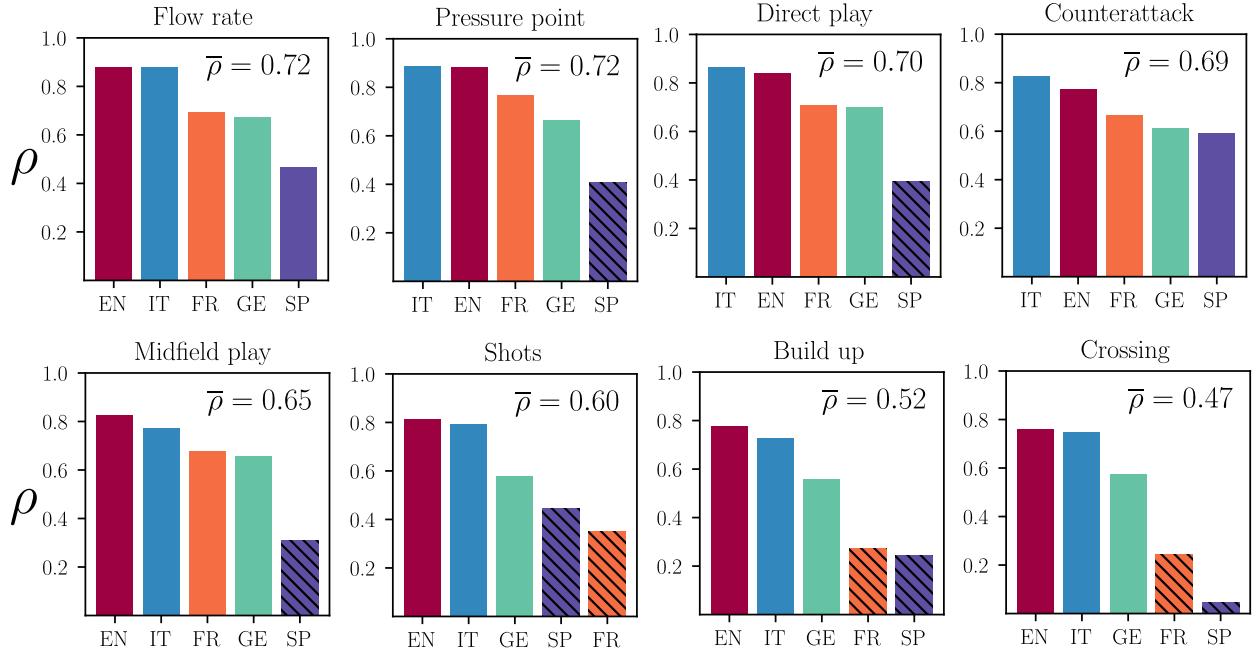


FIG. 3. Coeficiente de correlación de person,  $\rho$ , entre el rating verdadero y el rating dado por las metricas. Cada panel contiene la informacion de una metrica y las barras muestran el valor de  $\rho$  obtenido en cada liga. Las barras en cada panel estan ordenadas en orden creciente respecto del valor de  $\rho$ . Asimismo, los paneles estan ordenado en orden creciente respecto del valor promedio en cada metrica,  $\bar{\rho}$ . Las barras con hatch indica los casos donde el  $p - value > 0.05$ . En esos casos no se encontró una correlación estadísticamente significativa entre el true rating y el metric rating.

1978.

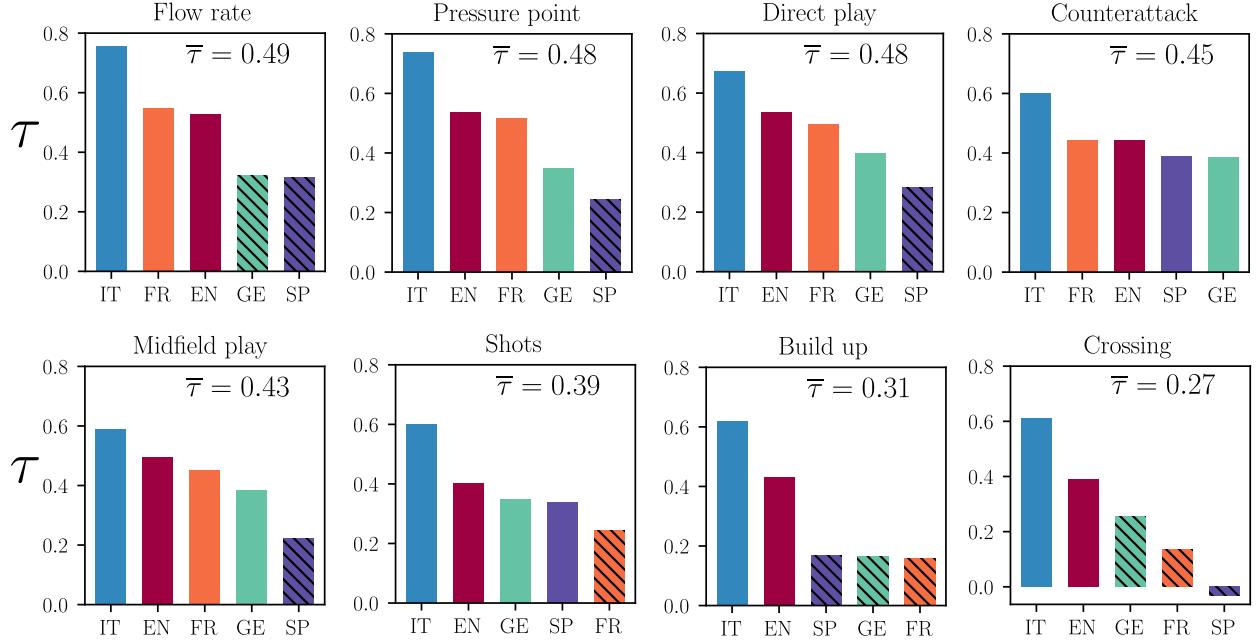


FIG. 4. Coeficiente de Kendall,  $\tau$ , entre los rankings verdaderos y los rankings dados por las metricas. Cada panel contiene la informacion de una metrica y las barras muestran el valor de  $\tau$  obtenido en cada liga. Las barras en cada panel estan ordenadas en orden decreciente respecto del valor de  $\tau$ . Asimismo, los paneles estan ordenado en orden decreciente respecto del valor promedio en cada metrica,  $\bar{\tau}$ . Las barras con hatch indica los casos donde el  $p - value > 0.05$ . En esos casos, los resultados no proporcionan evidencia suficiente para afirmar la existencia de una relacion lineal entre el ranking real y el inferido.

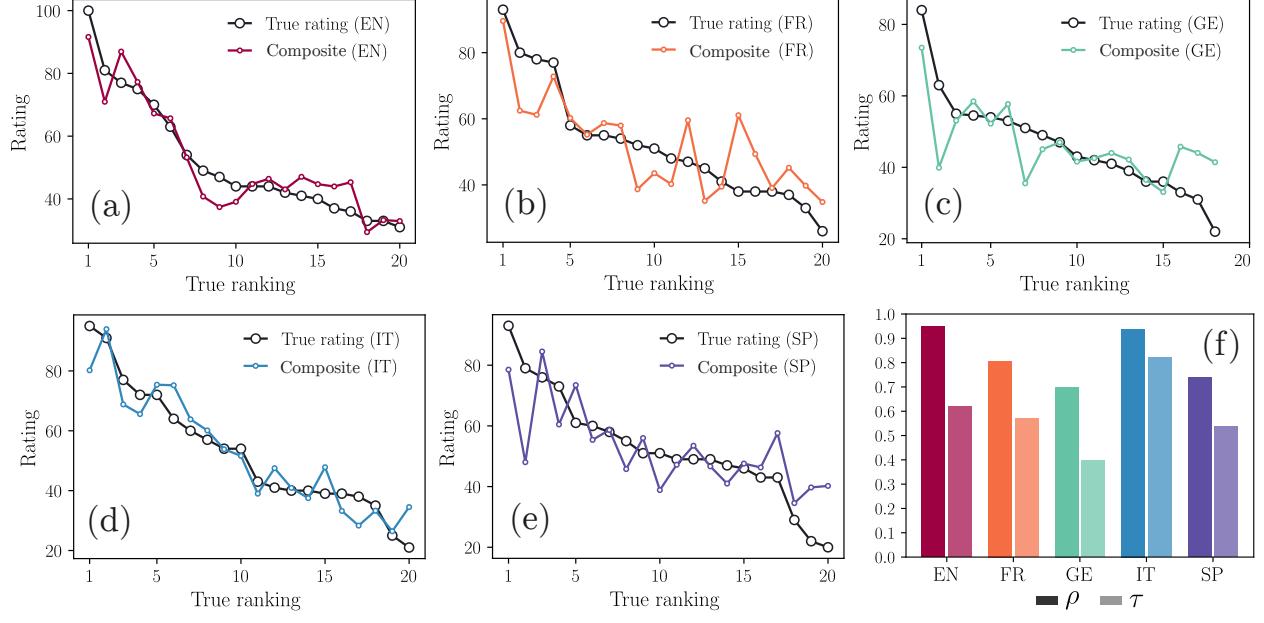


FIG. 5. Comparación entre el true rating y el composite rating. En los panes (a), (b), (c), (d) y (e) comparamos el true rating con el composite rating de cada uno de los equipos en las ligas inglesa, francesa, alemana, italiana y española respectivamente. En el panel (f) mostramos los valores del coeficiente de Pearson,  $\rho$ , y el coeficiente de Kendall,  $\tau$ , obtenido para cada liga. Notar que, en cada caso, la barra de la izquierda se corresponde con el valor de  $\rho$  y la de la derecha con el valor de  $\tau$ .

TABLE I. Componentes del Composite rating por Liga (basado en selección BIC)

Metric	Coef.	P-val	Imp. (%)	$R^2$
<i>England</i>				
Build up	-24.56	0.0006	37.65	0.88
Flow rate	22.69	0.0045	34.79	
Pressure point	17.98	0.0117	27.57	
<i>France</i>				
Pressure point	33.68	0.0145	61.71	0.61
Direct play	-20.90	0.1096	38.29	
<i>Germany</i>				
Direct play	9.48	0.0012	100	0.46
<i>Italy</i>				
Flow rate	17.84	0.0006	37.15	0.86
Direct play	15.46	0.0018	32.18	
Midfield play	-14.73	0.0119	30.67	
<i>Spain</i>				
Pressure point	-34.93	0.0185	43.08	0.46
Flow rate	26.38	0.0297	32.53	
Counterattack	19.78	0.0046	24.39	