

**Universidad
Rey Juan Carlos**

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA

GRADO EN MATEMÁTICAS

Curso Académico 2020/2021

Trabajo Fin de Grado

**MODELIZACIÓN, VISUALIZACIÓN Y ANÁLISIS DE LA
COMPETITIVIDAD EN DIFERENTES LIGAS DE
FÚTBOL PROFESIONAL**

Autor: Pablo Roldán Peigneux d'Egmont

Director: Alejandro J. García del Amo Jiménez

Dedicado a

*Mis padres, por ser los principales responsables de que esté donde estoy ahora mismo,
pudiendo terminar mi carrera a día de hoy.*

Agradecimientos

Este Trabajo de Fin de Grado se lo agradezco, en primer lugar, a Alejandro José García del Amo Jiménez, mi tutor del TFG, puesto que fue de las únicas personas que accedió a dirigírmelo, además de mostrarme su apoyo y ayudarme durante todo su desarrollo, echándome una mano siempre que lo necesitaba, desde la elección del tema hasta la revisión final del TFG.

También he de agradecer a todos los docentes de la Universidad Rey Juan Carlos que me han dado clase, porque sin todo lo que me han aportado y enseñado, habría sido imposible desarrollar un trabajo de tal envergadura.

Por último, dar las gracias a todos los profesores de matemáticas que he tenido en mi vida, pues supieron transmitirme la pasión que ellos sentía hacia ellas, pasión que yo ahora comparto, culminando en la adquisición de un título en el Grado de Matemáticas por la URJC, considerándome de esta manera, matemático.

Resumen

En este TFG se tratará de explicar y desarrollar el artículo titulado: “Modeling and visualizing competitiveness in soccer leagues” [1].

Se desarrollarán los conceptos desde los más básicos hasta los más complejos, para que cualquier persona sin conocimientos matemáticos avanzados pueda entender dicho artículo, añadiendo información de otras fuentes bibliográficas. Además, debido al idioma del artículo, se hará una labor exhaustiva de traducción del mismo.

El artículo en cuestión tratará de analizar la competitividad en diferentes ligas de fútbol, en concreto en 6 de ellas: “La Liga” (España), “Premiership” (Inglaterra), “Serie A” (Italia), “Ligue 1” (Francia), “Série A” (Brasil) y “J1 League” (Japón); durante la temporada 2018–2019.

Para el análisis de la competitividad, en primer lugar se agruparán los equipos en distintos conjuntos con el objetivo de obtener una idea inicial de qué equipos tienen los mismos objetivos. Más tarde, se dibujarán unos gráficos que representarán el desarrollo de los equipos en función de los puntos ganados y perdidos durante la temporada liguera, y, sobre ellos, se calculará su dimensión fractal y entropía. Por último, se compararán estas dos medidas con otras dos medidas clásicas de competitividad, la desviación típica y el índice de balance competitivo de Herfindahl-Hirschman, con el fin de visualizar su correlación y así comprobar cómo de acertadas o erróneas son las obtenidas en función de los gráficos.

Palabras clave

Palabras clave:

- Competitividad deportiva
- Sistemas complejos
- Sistemas dinámicos
- Poder complejo
- Clusterización
- Escalamiento multidimensional
- Distancias
- Visualización de datos
- Dimensión fractal
- Entropía
- Desviación típica
- Índice de balance competitivo
- Fútbol

Índice general

1. Introducción	1
1.1. Motivación	2
1.2. Competencias	4
2. Objetivos	6
2.1. Clusterización	6
2.2. Escalamiento multidimensional	7
2.3. Competitividad	8
3. Contenidos principales	10
3.1. Clusterización	11
3.1.1. Metodologías empleadas	12
3.1.2. Resultados y Análisis	17
3.2. Escalamiento multidimensional	19
3.2.1. Metodologías empleadas	19
3.2.2. Resultados y Análisis	26
3.3. Competitividad	28
3.3.1. Metodologías empleadas	29
3.3.2. Resultados y Análisis	39
4. Conclusiones y trabajos futuros	43
4.1. Conclusiones	43
4.2. Trabajos futuros	44
Bibliografía	47

Índice de figuras

3.7. Los gráficos resultantes del MDS con dimensión $M = 2$ de $S_i(k) = P_i(k) + \tau Q_i(k)$, $0 \leq k \leq k_r$, $k_r = 0, \dots, R$, para “La Liga”: (a) distancia de Canberra d_C ; (b) distancia de Sørensen d_S . La tercera dimensión es calculada por RBI en función del tiempo (rondas).	25
3.8. Los gráficos resultantes del MDS con dimensión $M = 2$ de $S_i(k) = P_i(k) + \tau Q_i(k)$, $0 \leq k \leq k_r$, $k_r = 0, \dots, R$, para “La Liga”: (a) distancia de Canberra d_C ; (b) distancia de Sørensen d_S . La tercera dimensión es calculada por RBI en función del espacio (puntos acumulados).	25
3.9. Diagrama de Shepard para “La Liga” con distancia de Manhattan (d_M).	26
3.10. Diagrama de estrés S frente a la dimensión escogida M para “La Liga” con distancia de Manhattan (d_M).	27
3.11. Número de copias ($N(l)$) y Longitud ($L = 1/l$), para una Dimensión fractal (D)= 1.	30
3.12. Número de copias ($N(l)$) y Longitud ($L = 1/l$), para una Dimensión fractal (D)= 2.	31
3.13. Número de copias ($N(l)$) y Longitud ($L = 1/l$), para una Dimensión fractal (D)= 3.	31
3.14. Triángulo de Sierpinski.	32
3.15. Comparación de la dimensión fractal de la costa de Gran Bretaña (izquierda) frente a la dimensión fractal de la costa de Noruega (derecha)	33
3.16. Comparación de la entropía entre puntos ordenados (izquierda) frente a puntos desordenados (derecha)	35
3.17. Tabla con las medidas de competitividad b_S , b_{MDS} , H_S , H_{MDS} , σ y $HICB$ (nombradas de izquierda a derecha) para la “J1 League”, “La Liga”, “Premiership”, “Ligue 1”, “Série A” y “Serie A”, ordenadas de menor a mayor σ y $HICB$	38
3.18. Los resultados obtenidos de b_S , b_{MDS} , H_S , H_{MDS} , σ y $HICB$. Las ligas están en orden descendente de competitividad según las medidas clásicas σ y $HICB$	39

Capítulo 1

Introducción

El fútbol es el deporte más popular del planeta con hasta 3500 millones de seguidores alrededor del mundo [2]. Un partido de fútbol profesional consta de 2 conjuntos de 11 jugadores cada uno sobre el terreno de juego, además de cambios, y uno o varios árbitros que se ocupan de que las normas se cumplan correctamente. El objetivo es hacer entrar en la portería contraria un balón que no puede ser tocado con las manos ni con los brazos, salvo por el portero en su área de meta [3]. Estos jugadores son exclusivos de ese equipo (sin contar a su selección nacional de fútbol), por lo que si un jugador x pertenece a un equipo A , este jugador no puede pertenecer a cualquier otro equipo de fútbol U (1.1) . Es decir:

$$\text{jugador del equipo } A = \{x | x \in A \wedge x \notin U\}, \quad (1.1)$$

esta regla diferencia un equipo de fútbol de una selección nacional de fútbol, puesto que muchos jugadores de fútbol, pertenecen a la vez a un equipo y a una selección nacional. Por ejemplo, Raúl González Blanco llegó a pertenecer a la vez al Real Madrid (equipo de fútbol) y a la Selección Española de Fútbol (selección nacional de fútbol) [4, 5].

Puesto que este análisis está hecho de cara a partidos de liga regular, cada partido durará un total de 2 partes de 45 minutos cada una. Si un equipo A tiene un enfrentamiento directo contra un equipo B, habrá 3 posibles resultados {ganar, empatar, perder}, donde actualmente para todas las ligas a analizar [6], y en concreto desde la temporada 95/96 para la liga española [7], cada uno de estos resultados conlleva {3, 1, 0} puntos respectivamente. Hay que anotar que la victoria del equipo A implicará la derrota del equipo B y viceversa; y si el equipo A empata, el equipo B

empatará también. Generalmente, el equipo ganador será el que haya marcado más goles que el otro, y si ambos marcan el mismo número de goles, el resultado de ambos será empate. El otro motivo por el que se puede ganar un partido es por lo que se conoce como “Walkover”, donde la victoria es otorgada a uno de los equipos si no hay más competidores, porque se retiraron, fueron descalificados, o no asistieron [8]. En el fútbol, el mínimo de jugadores de un equipo es de 7 [9], así que ya sea por falta de presentación o por expulsiones, se podría dar esta situación.

Dependiendo de estos resultados, al final de una liga regular, cuando todos los equipos se hayan enfrentado entre ellos 2 veces con cada uno, una en su estadio y otra en el del rival, se decidirá qué destino le espera a cada uno de estos equipos, que dependiendo de la liga y la temporada, será uno u otro (como veremos en el capítulo 3). En este caso, las ligas que se van a analizar son: “La Liga” (España), “Premiership” (Inglaterra), “Serie A” (Italia), “Ligue 1” (Francia), “Série A” (Brasil) y “J1 League” (Japón), durante la temporada 2018–2019. Es por esto, que la clasificación final de cada equipo en la tabla será lo que se tomará como el principal aspecto de interés.

Se tomarán aquellos partidos que tengan una alta incertidumbre en lo que concierne al ranking como los más competitivos, y, de la misma manera, los que menos afecten a este ranking, como los menos competitivos. Por lo que los partidos en los que ambos equipos se juegan más puestos relevantes, son aquellos en los que la competitividad es más alta [10] , independientemente de rivalidades históricas, derbis, ... con el objetivo de hacer este análisis lo más objetivo posible.

1.1. Motivación

La motivación principal de este TFG viene de mi gran interés por el fútbol y las matemáticas. Además de que creo que todo es parametizable; por eso, al ver que se podía hacer eso mismo con algo que es a priori abstracto como la competitividad, mi interés creció instantáneamente.

No solo eso, creo que esto es de vital interés para el mercado de las apuestas deportivas. Pues partidos cuya competitividad es alta, atraerán a más consumidores. Y debido a cómo funcionan los coeficientes, cuanta más gente apueste, más probable es que obtengan beneficios, pues el margen que se llevan va acorde a la opinión de la gente [11]. Por lo que a mayor participación, mejor se acompañará con los porcentajes estimados por la casa.

Esto se ve más claramente con un ejemplo: al tirar una moneda ideal y no trucada, uno tiene un 50 % de probabilidades de que caiga Cara y un 50 % de que caiga Cruz, y así lo opina la gente; por lo que si la casa de apuestas no quisiera llevarse nada de beneficio, entonces ofrecería una cuota de x_2 en ambas apuestas (1.2). Sin embargo, si quisiese obtener un 8 % de margen de beneficio, la cuota para cada una sería de $x_1, 85$ (1.3).

$$\text{Cuota para } 50\% \text{ sin margen de beneficio} = (0,5 \cdot 1)^{-1} = 2, \quad (1.2)$$

$$\text{Cuota para } 50\% \text{ con } 8\% \text{ de margen de beneficio} = (0,5 \cdot 1,08)^{-1} \approx 1,85. \quad (1.3)$$

Esto se hace así y no corrigiendo las cuotas al momento en función del dinero apostado a una o a otra, porque de hacerse así, estas se descompensarían si llega una persona con mucho dinero y apuesta primero a una de las dos, espera a que se corrijan, y apuesta a la contraria.

Siguiendo el ejemplo anterior y suponiendo que hubiesen 100€ apostados a Cara y 100€ apostados a Cruz ya, si alguien apuesta 100€ a una de las 2 cuotas y luego de que se corrijan las apuestas en función del dinero, esta misma persona apuesta a la contraria otros 100€, entonces garantiza su beneficio en detrimento de la casa de apuestas (ver ecuaciones 1.4, 1.5, 1.6, 1.7) como se muestra a continuación:

$$\text{Cuota para } \frac{100}{200} \text{ euros con } 8\% \text{ de beneficio} = \left(\frac{100}{200} \cdot 1,08\right)^{-1} \approx 1,85, \quad (1.4)$$

$$\text{Cuota para } \frac{100}{300} \text{ euros (la opuesta) con } 8\% \text{ de beneficio} = \left(\frac{100}{300} \cdot 1,08\right)^{-1} \approx 2,78, \quad (1.5)$$

$$\text{Beneficio del apostante si gana la } 1^a = 100 \cdot 1,85 - 100 = 85 \text{ €} \quad (1.6)$$

$$\text{Beneficio del apostante si gana la } 2^a = 100 \cdot 2,78 - 100 = 170 \text{ €} \quad (1.7)$$

En consecuencia, a las casas de apuestas les conviene saber en qué partidos hay mucha competitividad y en cuáles poca, puesto que en los que hay menos competitividad, debido a la poca participación, pueden correr un mayor riesgo de perder dinero, ya que, como hemos visto anteriormente, las apuestas no se corrigen en función del dinero apostado, sino en función de la opinión de la gente, así que si poca gente apuesta con mucho dinero, las apuestas corren un mayor riesgo de acabar desbalanceadas, haciéndole perder dinero a la casa de apuestas en cuestión. Un ejemplo real donde ocurrió algo similar, fue durante la temporada 2015/16, donde la inesperada victoria del Leicester City del título de la “Premiership”, provocó que la casa de apuestas “Ladbrokes”, ofreciera grandes compensaciones si los apostantes cerraban sus apuestas antes, llegando a pagar una suma superior a 3 millones de libras a los que aguantaron hasta el final [12].

Es por estos dos argumentos que se puede afirmar con certeza que este Trabajo Fin de Grado tiene tanto una motivación personal como empresarial, pues las casas de apuestas que se encargan de las apuestas relativas al fútbol se podrían lucrar aprovechando esta información.

1.2. Competencias

A continuación listaré las competencias generales y específicas que este TFG abordará [13]:

- Competencias Generales:
 - Saber reunir e interpretar datos relevantes (normalmente de carácter matemático) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
 - Poder transmitir información, ideas, problemas y sus soluciones, de forma escrita u oral, a un público tanto especializado como no especializado.
 - Utilizar herramientas de búsqueda de recursos bibliográficos.
 - Poder comunicarse en otra lengua de relevancia en el ámbito científico.
- Competencias Específicas:
 - Asimilar la definición de un nuevo objeto matemático, en términos de otros ya conocidos, y ser capaz de utilizar este objeto en diferentes contextos.

- Proponer, analizar, validar e interpretar modelos de situaciones reales sencillas, utilizando las herramientas matemáticas más adecuadas a los fines que se persigan.
- Utilizar aplicaciones informáticas de análisis estadístico, cálculo numérico y simbólico, visualización gráfica, optimización u otras para experimentar en matemáticas y resolver problemas.
- Conocimientos básicos sobre el uso y programación de los ordenadores, el desarrollo de programas, y la capacidad para adquirir, obtener, formalizar y representar el conocimiento humano en una forma computable, para la resolución de problemas.

Capítulo 2

Objetivos

En este capítulo se detallarán los objetivos generales y específicos del Trabajo de Fin de Grado (TFG).

El objetivo principal de este TFG es el de explicar y detallar el cálculo llevado a cabo para determinar qué liga es la más competitiva. Para esto nos apoyaremos del estudio hecho en el artículo: “Modeling and visualizing competitiveness in soccer leagues” [1]. Además, todos los datos provienen de: www.worldfootball.net [14] .

En el Trabajo Fin de Grado del Grado en Ingeniería Informática que realizaré a continuación de este se desarrollarán procesos de extracción y visualización de datos relacionados con los resultados que observaremos tanto en este TFG de Matemáticas como en el de Informática.

Al ser un alumno del Doble Grado en Ingeniería Informática y Matemáticas, estas dos Memorias, relacionadas aunque diferentes, creemos que culminan de forma armoniosa los estudios realizados.

2.1. Clusterización

Para abordar este problema, primero se hará una Clusterización o Agrupación de los datos, es decir, una asociación de los datos que son comparables entre si para saber qué equipos son cotejables entre los de su misma liga [15].

La metodología aplicada será la de Agrupación jerárquica, en el que los grupos formados tienen una jerarquía. Distinguiremos 2 acercamientos diferentes para este tipo de agrupaciones: las aglomerativas y las divisivas [16].

Las aglomerativas son ascendentes, es decir, cada equipo empezaría en un grupo individual y los pares de grupos irían juntándose a medida que se va subiendo en la jerarquía, hasta finalmente llegar a un grupo en el que estén todos juntos.

Mientras que, por otro lado, las divisivas son descendentes, por lo que todos los equipos estarían en un mismo grupo inicial, que se iría dividiendo en 2 cada vez que se baje un nivel en la jerarquía.

En este caso, la agrupación jerárquica que se usará es la aglomerativa [1].

Además de esto, existen varios criterios por los que diferenciar estos grupos. La relación entre 2 equipos vendrá definida por una distancia determinada por los puntos obtenidos por cada uno de ellos, formando enlaces entre sí.

Para agrupar relaciones de enlaces, encontraremos diferentes métricas: el agrupamiento de máximo o completo enlace (3.6), el agrupamiento de mínimo o simple enlace (3.7) y el agrupamiento de enlace media o promedio (3.5) [17]. La empleada en este caso será el Agrupamiento de enlace media o promedio (3.5) [1].

El uso de estas metodologías será detallado en el capítulo de Contenidos principales.

2.2. Escalamiento multidimensional

Más tarde, se hará un análisis de escala multidimensional sobre las ligas de fútbol. Esto nos permitirá conocer las relaciones entre los diferentes elementos de un conjunto de datos [18], en este caso equipos de una misma liga de fútbol.

El Escalamiento multidimensional o MDS es una técnica de análisis multivariante que, partiendo de una matriz de distancias entre individuos (equipos de fútbol en este caso), produce una representación de los individuos en una escala euclídea ordinaria, es decir, en un espacio plano y no curvo donde la suma de los tres ángulos interiores de un triángulo da siempre 180° [19]. De modo que las distancias en dicha escala se aproximen lo mejor posible a las distancias de partida [20] .

La matriz resultante de este proceso de MDS se obtendrá a través de un proceso de minimización del estrés. Este estrés puede definirse de varias maneras; por un lado, tendremos la definición clásica (3.9), y, por otro lado, tendremos el criterio de Sammon (3.10).

El criterio elegido para el estrés dependerá de lo que se quiera mostrar en la visualización.

En cualquier caso, el utilizado será la definición clásica de estrés (3.9) [1].

Esta matriz tendrá dimensión espacial M , donde a mayor M , mejores resultados. Tras un análisis que veremos en la sección 3.2.2, veremos que la ventaja de aumentar la M de 2 a 3 no es tanta (3.9,3.10), sin embargo, viene con el inconveniente de que se quedan unos datos más difíciles de manejar, teniendo que girarlos, ampliarlos o cambiar de lugar cada vez que se quiera ver algo nuevo de ellos. Es por esto que se empleará $M = 2$ [1].

La matriz con la que alimentaremos el escalamiento dimensional representará la distancia entre cada equipo en cada ronda de la temporada. Esta distancia puede ir definida de 3 formas diferentes: la distancia de Manhattan (3.11), la distancia de Canberra (3.12), o la distancia de Sørensen (3.13) [21]. A pesar de que cada distancia dará un conjunto de datos diferentes como veremos más tarde (3.7,3.8), todas provienen de los mismos datos, por lo que el motivo para elegir una u otra será la facilidad con la que se puede interpretar el gráfico resultante del MDS. La distancia de Manhattan será la elegida en este caso [1].

Estas metodologías y sus empleos serán detalladas en el capítulo de Contenidos principales.

2.3. Competitividad

Por último, se llevará a cabo el análisis de la competitividad para cada liga. Para este análisis, se tendrá en cuenta la dimensión fractal y la entropía de los gráficos obtenidos a raíz de los procesos de Clusterización y Escalamiento multidimensional. Además, los resultados obtenidos se compararán con otras medidas clásicas de competitividad.

La dimensión fractal de una figura es una generalización de la dimensión euclíadiana y se define como: $\text{Dimensión fractal} = \frac{\log(\text{Nº de copias})}{\log(\text{Factor de reducción})}$. Esta definición (3.14) parte de la representación de la figura, en la que [22]: $\text{Nº de copias} = \text{Factor de reducción}^{\text{Dimensión fractal}}$.

Esto se ve más claramente con los ejemplos mostrados en la figura 2.1 (figura tomada de la referencia [23]):

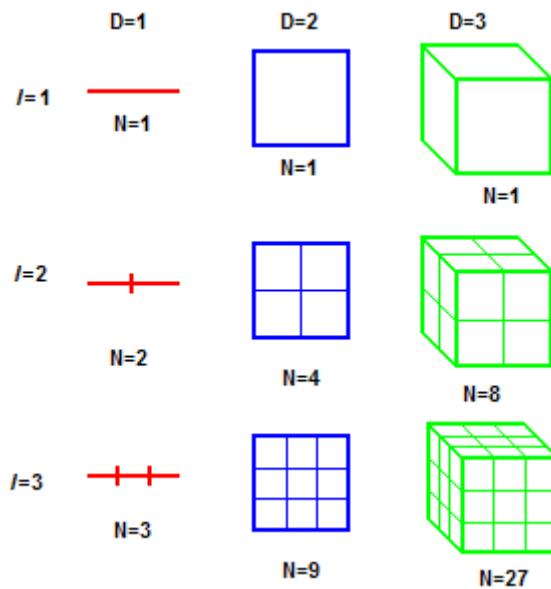


Figura 2.1: Nueve ejemplos diferentes con Dimensión fractal (D), Número de copias (N) y Factor de reducción (l).

Su función, en términos generales, es la de medir en qué grado un objeto de n dimensiones llena el espacio de $n+1$ dimensiones. En la práctica, será una forma de medir cómo de irregular o cuánto ruido tiene una imagen o gráfico. Este ejemplo será desarrollado más en detalle en la sección 3.3.1.

La entropía es una medida que nos permite saber la incertidumbre de una fuente de información. En el caso de una imagen, nos permite conocer la cantidad de ruido, desorden o dispersión que esta tiene [24].

Finalmente, las medidas clásicas de competitividad con las que serán comparadas estas 2, serán la desviación típica relativa a la proporción de victorias que logra cada equipo en una temporada, que indicará cómo de cerca o de lejos estarán todos los datos de la media [25]; y el índice de balance competitivo de Herfindahl-Hirschman, que dirá cómo de concentrados o divididos están los puntos en la clasificación [26][1].

Estas metodologías y cálculos serán definidos y desarrollados en profundidad en el capítulo de Contenidos principales.

Capítulo 3

Contenidos principales

En esta sección se profundizará y detallarán las metodologías mencionadas en el capítulo de Objetivos. Asimismo, se desarrollarán todas las fórmulas, imágenes y algoritmos referenciadas allí, con el objetivo de clarificar y explicar todos los contenidos del artículo: “Modeling and visualizing competitiveness in soccer leagues” [1], además de recopilar información bibliográfica adicional de otras fuentes, para complementarlo con el mismo objetivo.

Durante estos cálculos, tendremos en cuenta el uso de las distintas definiciones que serán usadas a lo largo de todos los apartados. Las definiciones comunes serán las siguientes:

$$N = \text{número total de equipos en una liga}. \quad (3.1)$$

$$\begin{aligned} R &= \text{número total de partidos de un equipo en una temporada en la liga} \\ &= 2(N - 1). \end{aligned} \quad (3.2)$$

R tiene ese número concreto de partidos, pues en una liga regular, todos los equipos juegan entre ellos (menos contra sí mismos, obviamente) 2 veces, un partido de ida y otro de vuelta.

Además de esto, para los resultados y visualización de estos cálculos, usaremos una tabla con etiquetas a cada equipo de cada liga para referirnos a ellos (ver Tabla 3.1). La tabla en cuestión (tabla tomada del Apéndice A de la referencia [1]) es la siguiente:

'La Liga' (Spain)			'Premiership' (England)			'Serie A' (Italy)		
Rank	Team	Tag	Rank	Team	Tag	Rank	Team	Tag
1	FC Barcelona	Bar	1	Manchester City	MCi	1	Juventus	Juv
2	Atlético Madrid	AMa	2	Liverpool FC	Liv	2	SSC Napoli	Nap
3	Real Madrid	RMa	3	Chelsea FC	Che	3	Atalanta	Ata
4	Valencia CF	Val	4	Tottenham Hotspur	Tot	4	Inter	Int
5	Getafe CF	Get	5	Arsenal FC	Ars	5	AC Milan	Mil
6	Sevilla FC	Sev	6	Manchester United	MUn	6	AS Roma	Rom
7	Espanyol Barcelona	Esp	7	Wolverhampton Wanderers	Wol	7	Torino FC	Tor
8	Athletic Bilbao	Ath	8	Everton FC	Eve	8	Lazio Roma	Laz
9	Real Sociedad	RSo	9	Leicester City	Lei	9	Sampdoria	Sam
10	Real Betis	RBe	10	West Ham United	WHa	10	Bologna FC	Bol
11	CD Alavés	Ala	11	Watford FC	Wat	11	Sassuolo Calcio	Sas
12	SD Eibar	Eib	12	Crystal Palace	Cry	12	Udinese Calcio	Udi
13	CD Leganés	Leg	13	Newcastle United	New	13	SPAL 2013 Ferrara	Spa
14	Villarreal CF	Vil	14	AFC Bournemouth	Bou	14	Parma Calcio 1913	Par
15	Levante UD	Lev	15	Burnley FC	Bur	15	Cagliari Calcio	Cag
16	Real Valladolid	RVa	16	Southampton FC	Sou	16	ACF Fiorentina	Fio
17	Celta Vigo	Cel	17	Brighton Hove Albion	Bri	17	Genoa CFC	Gen
18	Girona FC	Gir	18	Cardiff City	Car	18	Empoli FC	Emp
19	SD Huesca	Hue	19	Fulham FC	Ful	19	Frosinone Calcio	Fro
20	Rayo Vallecano	Ray	20	Huddersfield Town	Hud	20	Chievo Verona	Chi
'Ligue 1' (France)			'Série A' (Brazil)			'J1 League' (Japan)		
Rank	Team	Tag	Rank	Team	Tag	Rank	Team	Tag
1	Paris Saint-Germain	PSG	1	Flamengo RJ	Fla	1	Yokohama F. Marinos	YFM
2	Lille OSC	Lil	2	Santos FC	San	2	FC Tokyo	FCT
3	Olympique Lyon	Oly	3	Palmeiras	Pal	3	Kashima Antlers	KaA
4	AS Saint-Étienne	SEt	4	Grêmio Porto Alegre	GPA	4	Kawasaki Frontale	KaF
5	Olympique Marseille	OMa	5	Athletico Paranaense	AtP	5	Cerezo Osaka	CeO
6	Montpellier HSC	Mon	6	São Paulo FC	Spa	6	Sanfrecce Hiroshima	SaH
7	OGC Nice	Nic	7	Internacional	Int	7	Gamba Osaka	GaO
8	Stade Reims	Rei	8	Corinthians SP	Cor	8	Vissel Kobe	ViK
9	Nîmes Olympique	Nim	9	Fortaleza	For	9	Oita Trinita	OiT
10	Stade Rennes	Ren	10	Goiás	Goi	10	Hokkaido Consadole Sapporo	HCS
11	RC Strasbourg	Str	11	Bahia - BA	Bah	11	Vegalta Sendai	VeS
12	FC Nantes	Nan	12	Vasco da Gama	VaG	12	Shimizu S-Pulse	SSP
13	Angers SCO	Ang	13	Atlético Mineiro	AtM	13	Nagoya Grampus	NaG
14	Girondins Bordeaux	Bor	14	Fluminense RJ	Flu	14	Urawa Red Diamonds	URD
15	Amiens SC	Ami	15	Botafogo - RJ	Bot	15	Sagan Tosu	SaT
16	Toulouse FC	Tou	16	Ceará - CE	Cea	16	Shonan Bellmare	ShB
17	AS Monaco	Mon	17	Cruzeiro	Cru	17	Matsumoto Yamaga	MaY
18	Dijon FCO	Dij	18	CSA - AL	CSA	18	Júbilo Iwata	Jul
19	SM Caen	Cae	19	Chapecoense	Cha	19		
20	EA Guingamp	Gui	20	Avaí - SC	Ava	20		

Figura 3.1: Lista de etiquetas y clasificaciones de “La Liga” (España), “Premiership” (Inglaterra), “Serie A” (Italia) y “Ligue 1” (Francia), durante la temporada 2018-2019.

donde “Rank” hace referencia a la posición final del equipo en su liga correspondiente en esa temporada; “Team” al equipo en cuestión; y “Tag” a la etiqueta que le es asignado a dicho equipo para su representación a la hora de visualizar los datos.

3.1. Clusterización

La Clusterización o Agrupación de los datos es, como hemos mencionado anteriormente en el Capítulo 2, un método por el que es posible distinguir qué grupos de elementos se pueden

comparar entre sí. En este caso, al tratar con equipos de fútbol, servirá para distinguir qué equipos son comparables con otros de su misma liga.

3.1.1. Metodologías empleadas

Para ello, habrá que establecer una métrica en común a tener en cuenta, ya sea color de la camiseta, número de jugadores... En este caso, esas 2 las desecharemos, debido a que el objetivo final es el de analizar la competitividad que hay entre ellos, por lo que la medida a tener en cuenta será el Poder Complejo al final de la temporada, es decir, $S_i(R)$.

Dado un equipo i y una ronda k , el Poder Complejo $S_i(k)$ es la suma del Poder Real $P_i(k)$ y el Poder Reactivo $Q_i(k)$ [27]. En este caso, el Poder Real $P_i(k)$ y el Poder Reactivo $Q_i(k)$, corresponderán a la cantidad acumulada de puntos ganados y perdidos por el equipo i en la ronda k , respectivamente.

Hay que tener en cuenta que para todas estas ideas del flujo de poder en sistemas dinámicos como el fútbol, i y k corresponderán al i -ésimo equipo, $i = 1, \dots, N$, en la ronda $k = 1, \dots, R$.

Se definirá, por tanto, el Poder Complejo (ver ecuaciones 3.3, 3.4) de la siguiente manera:

$$S_i(k) = P_i(k) + \imath Q_i(k), \quad (3.3)$$

$$S_i(k+1) = S_i(k) + \Delta S_i(k), \quad (3.4)$$

donde $\imath = \sqrt{-1}$, y $\Delta S_i(k) = \{3 + \imath 0, 1 + \imath 2, 0 + \imath 3\}$ por cada {victoria, empate, derrota}, lo que significa que el sistema es discreto tanto en el tiempo (las rondas) como en el espacio (los resultados); esto implica, que depende de 2 variables discretas y no continuas [28]. Un ejemplo de variable discreta sería el número de empleados, por ejemplo, mientras que una variable continua sería la longitud [29].

Se representará el $S_i(k)$ frente a k como vemos en la figura 3.2 (figura tomada del Capítulo 2 de la referencia [1]). A la hora de interpretar esta figura, se observan patrones que se asemejan a los de una onda que se propaga en un material con origen en las coordenadas $(0,0)$ y delimitados por una línea diagonal azul. En los ejes, las líneas azules anchas representan las rondas media $k = R/2$ y final $y k = R$. De estos resultados, se puede verificar que en todas las ligas, los

clústeres de los equipos surgen temprano en la temporada y permanecer casi sin cambios hasta el final de la misma. Más tarde, en el apartado 3.3, haremos un análisis de la dimensión fractal y la entropía de estos gráficos.

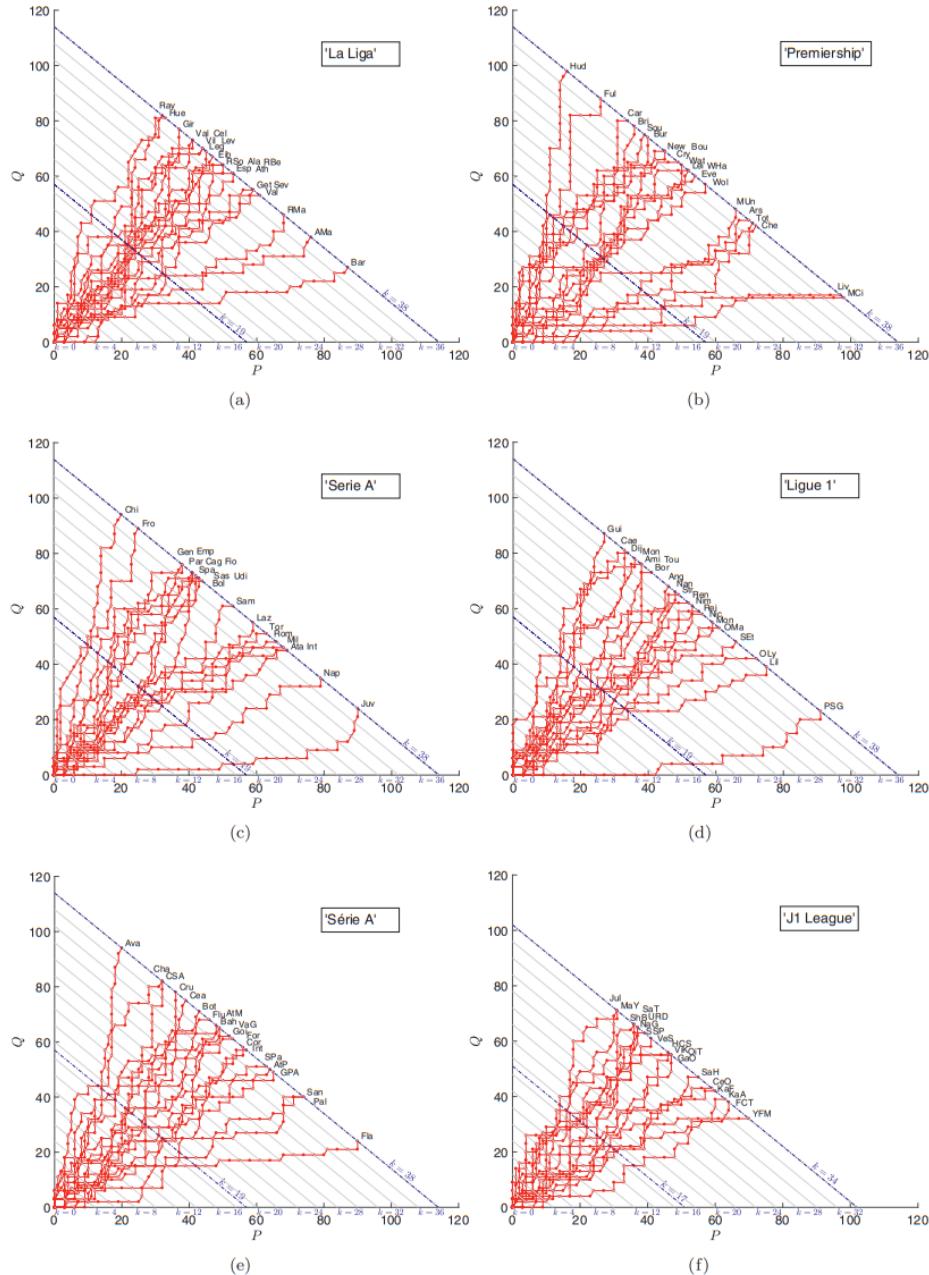


Figura 3.2: Evolución de $S_i(k)$ frente a k durante 6 ligas: (a) “La Liga”; (b) “Premiership”; (c) “Serie A”; (d) “Ligue 1” (e) “Série A”; (f) “J1 League”. Las líneas diagonales azules representan las ondas para valores sucesivos de k .

Como mencionamos en el capítulo de Objetivos, se seguirá una metodología de Agrupa-

miento jerárquico para identificar los diferentes clústeres. La agrupación jerárquica consiste en agrupar los equipos por una métrica que deje a unos por encima y a otros por debajo [16], siendo esta métrica el Poder complejo del equipo $S_i(R)$ al final de la temporada (la ronda R es la última), como hemos mencionado anteriormente ($S_i(R)$). Por lo tanto, los grupos con los equipos que tienen más puntos serán aquellos que estén por encima, y los grupos de los equipos con menos puntos aquellos que se encontrarán por debajo. Un ejemplo de Agrupación no jerárquica sería agrupar a los equipos por el color de su camiseta, pues no hay ningún color mejor que otro.

Para esta agrupación jerárquica, consideraremos una colección de N elementos $v_i, i = 1, \dots, N$, en un espacio real de dimensión L , es decir \mathbb{R}^L . Por ejemplo, si $L = 2$, significa que tenemos un plano [30].

Para el proceso de Clusterización jerárquica, definiremos también una distancia entre los elementos i y j , $d(v_i, v_j)$ con $i, j = 1, \dots, N$. Dando como resultado una matriz $D = [d(v_i, v_j)]$ de tamaño $N \times N$. Hay que tener en cuenta que estamos usando distancias clásicas, por lo que $d(v_i, v_j) = d(v_j, v_i)$ y, por lo tanto, D es una matriz simétrica con ceros en su diagonal principal. Esta matriz D será lo que usaremos como entrada de datos para determinar los diferentes clústeres, que podrán ser representados tanto con un dendrograma como con un árbol jerárquico. A pesar de ser prácticamente iguales, en este caso será un dendrograma lo que se utilizará, como vemos en la figura 3.3.

Como mencionamos en el capítulo 2, usaremos las agrupaciones aglomerativas en vez de las divisivas. Recordemos que las aglomerativas son ascendentes, es decir, que cada uno tiene un grupo individual y se va juntando con sus semejantes hasta llegar a un grupo común final. En el caso de la figura 3.3, significa que se hizo de derecha a izquierda y no de izquierda a derecha, como hubiera ocurrido si fuera mediante agrupaciones divisivas.

Para esta agrupación, se tendrá en cuenta la distancia entre elementos de distintos clústeres llamada $\delta(\mathbf{v}_X, \mathbf{v}_Y)$ con $\mathbf{v}_X \in X$ y $\mathbf{v}_Y \in Y$, donde X e Y son 2 clústeres o grupos diferentes. Y, como se ha explicado anteriormente (2), la métrica por la que agruparemos será el Agrupamiento de enlace media o promedio (3.5), cuya fórmula es la siguiente:

$$\delta_{av}(X, Y) = \frac{1}{|X| \cdot |Y|} \sum_{\mathbf{v}_X \in X, \mathbf{v}_Y \in Y} \delta(\mathbf{v}_X, \mathbf{v}_Y), \quad (3.5)$$

donde $|X|$ e $|Y|$ hacen referencia a la cardinalidad, es decir, a la cantidad o número de elementos que tiene dicho clúster [31] [32].

Otras que se podrían haber aplicado, pero que no se usaron al final, fueron: el Agrupamiento de máximo o completo enlace (3.6) y el Agrupamiento de mínimo o simple enlace (3.7), cuyas fórmulas son las siguientes [17]:

$$\delta_{max}(X, Y) = \max \{ \delta(\mathbf{v}_X, \mathbf{v}_Y) : \mathbf{v}_X \in X, \mathbf{v}_Y \in Y \}, \quad (3.6)$$

$$\delta_{min}(X, Y) = \min \{ \delta(\mathbf{v}_X, \mathbf{v}_Y) : \mathbf{v}_X \in X, \mathbf{v}_Y \in Y \}. \quad (3.7)$$

Los equipos tienen una coordenada vectorial que representa su final de temporada, $S_i(R)$. Por ejemplo, en la figura 3.2 se puede ver que el Barcelona acaba en la posición (87, 27) en la ronda R ($k = 38$). Es por esto que la distancia entre equipos es calculada de la siguiente forma (3.8):

$$d(S_i(R), S_j(R)) = |S_i(R) - S_j(R)|. \quad (3.8)$$

Uno de los resultados obtenidos será el que se observa en la Figura 3.3 (figura tomada del Capítulo 2 de la referencia [1]). La línea intermitente vertical de color rojo secciona el dendrograma para obtener los clústeres. En todos los casos, es decir, en todas las ligas, esta línea vertical seccionará los dendrogramas de tal forma que se diferencien un total de 4 o 5 grupos diferentes por liga, dependiendo de cómo de apretada esté esa liga y, por tanto, de los objetivos comunes de estos; donde el grupo 1 representa a los equipos que acabaron más arriba de la tabla, mientras que los últimos grupos representan a los que finalizaron más abajo. Hay que añadir, que el hecho de que haya menos grupos en una liga, no significa que esta sea automáticamente más competitiva, ya que estas secciones se hacen basadas en la diferencia de puntos entre el más bajo del grupo x frente al más alto del grupo $x - 1$, donde si la diferencia es lo suficientemente grande, se secciona en grupos o clústeres diferentes.

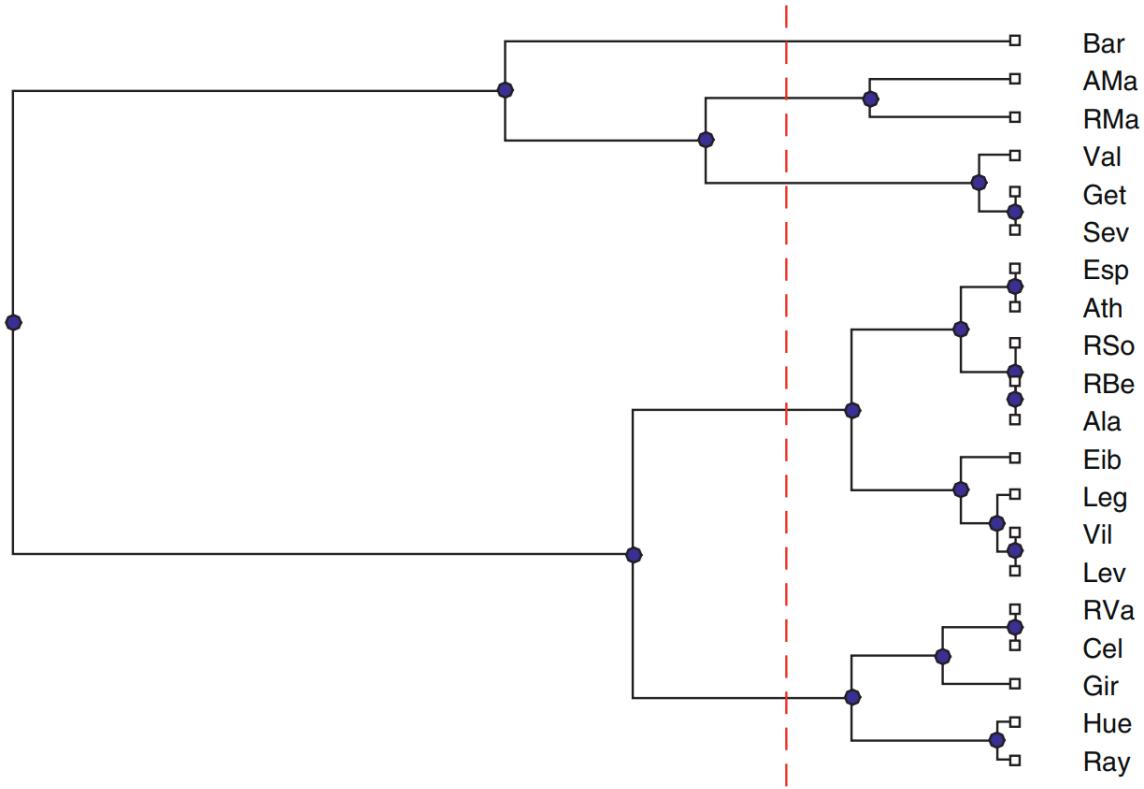


Figura 3.3: Dendrograma generado por el Agrupamiento Jerárquico para “La Liga” al final de temporada ($S_i(R)$), con distancia 3.8. La línea intermitente vertical de color rojo, secciona el dendrograma para obtener los clústeres.

Como se puede observar, para “La Liga” se obtienen 5 clústeres o grupos diferentes: $\mathcal{E}_1 = \{\text{Bar}\}$, $\mathcal{E}_2 = \{\text{AMa, RMa}\}$, $\mathcal{E}_3 = \{\text{Val, Get, Sev}\}$, $\mathcal{E}_4 = \{\text{Esp, Ath, RSo, RBe, Ala, Eib, Leg, Vil, Lev}\}$ y $\mathcal{E}_5 = \{\text{RVa, Cel, Gir, Hue, Ray}\}$, en los que se encuentran todos los equipos de “La Liga”. Además, se pueden observar algunos subclústeres dentro de \mathcal{E}_4 , que incluyen $\mathcal{E}_{41} = \{\text{Esp, Ath, RSo, RBe, Ala}\}$ y $\mathcal{E}_{42} = \{\text{Eib, Leg, Vil, Lev}\}$; y también \mathcal{E}_5 , con $\mathcal{E}_{51} = \{\text{RVa, Cel, Gir}\}$ y $\mathcal{E}_{52} = \{\text{Hue, Ray}\}$, que incluyen los equipos líderes y rezagados de sus respectivos clústeres.

Mirando el panorama completo, se observa que después de algunas rondas iniciales, hay poca o ninguna competencia “real” entre equipos que pertenecen a diferentes grupos. Por ejemplo, a principios de temporada $\mathcal{E}_1 = \{\text{Bar}\}$ se puso primero de la clasificación y allí se mantuvo hasta el final de la temporada. Y ocurría lo mismo en el otro extremo de la tabla, donde $\mathcal{E}_5 = \{\text{RVa, Cel, Gir, Hue, Ray}\}$ tenían claro que acabarían al final de la tabla a los pocos partidos. Esta división temprana en clústeres surge en todas las ligas nacionales, no solo en la española, puesto que los equipos tienen distintos objetivos, todos ellos igual de ambiciosos. Estos objetivos

pueden ser ganar la liga, jugar al año siguiente en competiciones de Confederación (por ejemplo, la “Champions” o la “Europa League”), tener una temporada tranquila logrando la mejor clasificación final posible, o, simplemente, no descender. Estos objetivos dan lugar a clústeres, pero la competitividad es diferente entre las ligas nacionales, ya que el número de candidatos a cada objetivo varía. En el caso de las competiciones europeas de la “Champions” y la “Europa League”, se decide cuántos equipos participan en ellas según el ranking UEFA de ligas [33]. Sin embargo, debido a que todos los equipos compiten duro por sus objetivos, no se ha tenido en cuenta que los partidos de un grupo fueran más competitivos que los partidos de otro. Es decir, aquellos partidos por el campeonato (los que se dan entre equipos pertenecientes a \mathcal{E}_1) no son más importantes o competitivos que los partidos por clasificarse a las competencias de Confederación (los que se dan entre equipos pertenecientes a \mathcal{E}_2).

Además de esto, hemos considerado que los partidos en los que ambos equipos se juegan más puestos relevantes, aquellos en los que la competitividad es más alta [10], son los partidos entre equipos del mismo clúster, aquellos donde ambos equipos se juegan más puestos normalmente. Por tanto, teniendo en cuenta que todos los equipos pertenecientes a un mismo clúster tienen una cantidad de puntos similar, cuanto más grande sea ese clúster, más puestos se jugarán en un partido entre equipos pertenecientes al mismo grupo. Esto se ve más fácilmente con un ejemplo: los partidos del Barcelona (único equipo en \mathcal{E}_1 según esta división de clústeres) no tenían “muchísima gracia”, porque sus perseguidores (el Atlético de Madrid y el Real Madrid) estaban tan atrás que, independientemente de que ganase o perdiese, este seguiría primero; sin embargo, si se enfrentaban el Español contra el Levante (2 de los 9 equipos de \mathcal{E}_4), debido a lo apretados que están los puntos entre equipos de un mismo clúster, se podían llegar a jugar 8 puestos ambos en un mismo partido.

3.1.2. Resultados y Análisis

Es por esto que, al analizar los distintos resultados de todas las ligas, se comprueba, observando la Figura 3.3 que, para “La Liga” (España), la gran mayoría de la competición estaba en los equipos que buscaban un lugar en la “Europa League” y en los equipos de la mitad de la tabla que intentaban una posición honorable, estando todos ellos concentrados en los 9 equipos pertenecientes a \mathcal{E}_4 .

Para la “Premiership” (Inglaterra) había algo de competencia por el campeonato, entre los

equipos de $\mathcal{U}_1 = \{\text{Mci}, \text{Liv}\}$, también para clasificarse a la “Champions” entre los equipos de $\mathcal{U}_2 = \{\text{Che}, \text{Tot}, \text{Ars}, \text{MUn}\}$, y para la “Europa League” entre $\mathcal{U}_3 = \{\text{Wol}, \text{Eve}, \text{Lei}, \text{WHa}, \text{Wat}, \text{Cry}\}$, y para la parte media-baja de la tabla, $\mathcal{U}_4 = \{\text{New}, \text{Bou}, \text{Bur}, \text{Sou}, \text{Bri}, \text{Car}\}$. Eso sí, los dos últimos clasificados, $\mathcal{U}_5 = \{\text{Ful}, \text{Hud}\}$, tuvieron su destino hacia el descenso trazado temprano.

En cuanto a la “Serie A” (Italia), al igual que para “La Liga”, no existió competencia por el lugar del campeón, siendo este ganado por $\mathcal{I}_1 = \{\text{Juv}\}$. Sin embargo, hubo una fuerte competencia por un lugar en la “Champions” y la “Europa League”, con $\mathcal{I}_2 = \{\text{Nap}, \text{Ata}, \text{Int}, \text{Mil}, \text{Rom}, \text{Tor}, \text{Laz}, \text{Sam}\}$ y por un puesto honorable para la mitad media-baja de la tabla, $\mathcal{I}_3 = \{\text{Bol}, \text{Sas}, \text{Udi}, \text{Spa}, \text{Par}, \text{Cag}, \text{Fio}, \text{Gen}, \text{Emp}\}$. Los dos últimos, $\mathcal{I}_4 = \{\text{Fro}, \text{Chi}\}$, no compitieron apenas, ya que se retrasaron temprano en la temporada, sin ninguna posibilidad de evitar el descenso.

Observando la “Ligue 1” (Francia), se ve que, de nuevo, no hubo competencia por el título de campeón, otorgado anticipadamente a $\mathcal{F}_1 = \{\text{PSG}\}$, pero hubo fuerte competencia por saber quién se clasificaba para la “Champions”, con $\mathcal{F}_2 = \{\text{Lil}, \text{OLy}, \text{SEt}, \text{OMa}, \text{Mon}\}$, para la “Europa League”, con $\mathcal{F}_3 = \{\text{Nic}, \text{Rei}, \text{Nim}, \text{Ren}, \text{Str}, \text{Nan}, \text{Ang}\}$; y por no acabar descendido, entre los equipos de $\mathcal{F}_4 = \{\text{Bor}, \text{Ami}, \text{Tou}, \text{Mon}, \text{Dij}, \text{Cae}, \text{Gui}\}$.

Siguiendo con la “Série A” brasileña, igual que en casos anteriores, no hubo competencia por el título de campeón, otorgado temprano a $\mathcal{B}_1 = \{\text{Fla}\}$. Los equipos $\mathcal{B}_2 = \{\text{Pal}, \text{San}\}$ compitieron por el segundo lugar y por asegurar su participación en la “Copa Libertadores” desde el principio de la temporada. Para el lugar restante en esta competencia regional evitando los playoffs, los equipos de $\mathcal{B}_3 = \{\text{GPA}, \text{AtP}, \text{SPa}\}$ compitieron por la cuarta plaza. Los equipos de $\mathcal{B}_4 = \{\text{Int}, \text{Cor}, \text{For}, \text{Goi}, \text{VaG}, \text{Bah}, \text{AtM}, \text{Flu}, \text{Bot}\}$ compitieron por un lugar en la “Copa Sudamericana”. Y finalmente, los equipos de $\mathcal{B}_5 = \{\text{Cea}, \text{Cru}, \text{CSA}, \text{Cha}, \text{Ava}\}$ competían por evitar el descenso. Aunque en este sentido, $\{\text{Ava}\}$ abandonó \mathcal{B}_5 y su destino como descendido se fijó muy temprano (uniéndose a un hipotético \mathcal{B}_6 no incluido), mientras que los demás seguían luchando por sus objetivos.

Para la “J1 League” japonesa, un equipo lideró de forma destacada a mitad de temporada, dejando así poca competencia para él, y una gran competencia entre los demás. Los clústeres comenzaron a surgir en el comienzo de la segunda mitad de la liga, en vez de tempranamente como en el resto de ligas, esto significa que, seguramente, fue más competitiva que el resto; pero esto se puede deber a que, al ser 18 equipos en vez de 20, la clasificación era más probable que acabara apretada porque había 4 jornadas menos, y no necesariamente por tener equipos

más equilibrados. Los equipos $\mathcal{J}_1 = \{\text{YFM}, \text{FCT}, \text{KaA}\}$ compitieron por el título de campeón, los de $\mathcal{J}_2 = \{\text{KaF}, \text{CeO}, \text{SaH}\}$ lucharon por un lugar honorable, los de $\mathcal{J}_3 = \{\text{GaO}, \text{OiT}, \text{ViK}, \text{HCS}, \text{VeS}\}$ buscaron una posición cómoda en la mitad de la tabla y los de $\mathcal{J}_4 = \{\text{SSP}, \text{NaG}, \text{URD}, \text{ShB}, \text{SaT}, \text{MaY}, \text{Jul}\}$ compitieron para evitar el descenso.

Si se observa la Figura 3.2 y teniendo en cuenta que si las líneas están más dispersas la competitividad será mayor como veremos en la sección 3.3, parece ser que para “La Liga” y la “Ligue 1” la competitividad es más alta que para las demás. Sin embargo, para un conclusión definitiva, se adoptarán medidas cuantitativas en la sección 3.3 que capturen dicha característica de los gráficos desarrollados en esta sección y en la sección 3.2.

3.2. Escalamiento multidimensional

El Escalamiento multidimensional o MDS es, como hemos mencionado anteriormente en el Capítulo 2, una técnica de análisis multivariante que, partiendo de una matriz de distancias entre individuos (equipos de fútbol en este caso), produce una representación de los individuos en una escala euclídea ordinaria, es decir, en un espacio plano y no curvo donde la suma de los tres ángulos interiores de un triángulo da siempre 180° [19]. De modo que las distancias en dicha escala se aproximen lo mejor posible a las distancias de partida [20] .

Esto nos permitirá saber las relaciones entre los diferentes elementos de un conjunto de datos [18], en este caso, equipos de una misma liga de fútbol.

3.2.1. Metodologías empleadas

Para esto se utiliza la matriz $D = [d(v_i, v_j)]$ descrita en la sección 3.1 como fuente de información para el algoritmo de Escalamiento Multidimensional o MDS. Dada esta matriz de distancias D , con las distancias entre cada par de equipos de una liga $d(v_i, v_j)$, y un número elegido de dimensiones L , el algoritmo MDS coloca cada equipo en un espacio M -dimensional (una representación de menor dimensión, es decir, $M \leq L$) de modo que las distancias entre objetos se conservan lo mejor posible [20].

Este algoritmo tratará de obtener una matriz $\hat{D} = [d(\hat{v}_i, \hat{v}_j)]$ de dimensión $M \leq L$ con N puntos, siendo N el número de equipos participantes en esa liga (3.1). El gráfico resultante se calculará a través del MDS métrico mediante una minimización del estrés S , cuya definición

clásica y que utilizaremos es (3.9):

$$S = \left[\sum_{i < j} [d(\mathbf{v}_i, \mathbf{v}_j) - d(\hat{\mathbf{v}}_i, \hat{\mathbf{v}}_j)]^2 \right]^{\frac{1}{2}}. \quad (3.9)$$

A pesar de que no se utiliza y que en otros casos puede dar mejores resultados, cabe recordar que también se puede definir el estrés a minimizar S_{Sammon} mediante el criterio de Sammon (3.10):

$$S_{Sammon} = \left[\frac{\sum_{i < j} [d(\mathbf{v}_i, \mathbf{v}_j) - d(\hat{\mathbf{v}}_i, \hat{\mathbf{v}}_j)]^2}{\sum_{i < j} [d(\mathbf{v}_i, \mathbf{v}_j)]^2} \right]^{\frac{1}{2}}. \quad (3.10)$$

La técnica MDS sigue ahora la lógica de un problema de optimización numérica que minimiza S (3.9). El estrés S disminuye monótonamente a medida que aumenta la dimensión M , como se comprueba en la sección 3.2.2, y el usuario selecciona el “mejor” valor de M que mantenga un equilibrio entre un S bajo y unos resultados claros y maleables. Por tanto, se suelen adoptar los valores $M = 2$ o $M = 3$. Hay que tener en cuenta que los ejes resultantes del MDS no tienen ningún sentido físico, solamente importan las líneas y puntos provenientes de los datos, que reflejan las diferentes relaciones entre los distintos elementos. En algunos casos, la elección de $M = 3$ da resultados significativamente mejores que la opción $M = 2$, pero a costa de dificultades adicionales en el manejo de los gráficos, requiriendo que el usuario gire, amplíe o cambie de lugar estos gráficos cada vez que quiere ver algo nuevo en ellos.

Para $M = 1, 2$ y 3 , los puntos resultantes se pueden visualizar en un diagrama de dispersión. Por claridad a la hora de visualizar los resultados (3.2.2), se empleará $M = 2$, es decir, un gráfico de 2 dimensiones (horizontal y vertical), al que después, por encima, se le añadirá una tercera dimensión para incluir información adicional [34].

Los gráficos representan las relaciones entre los equipos en distintos momentos de la temporada. La tercera dimensión se añade por encima usando curvas de nivel o isolíneas, que no modifican la forma de las líneas ni de los puntos de estos gráficos. Las isolíneas son curvas que conectan los puntos en que el gráfico tiene un mismo valor [35]. Estas isolíneas pueden depender, o bien de las rondas (Figura 3.5), o bien de los puntos (Figura 3.6), es decir, del tiempo (Figura 3.5), o del espacio (Figura 3.6). Esta tercera dimensión se calcula a través de la interpolación de base radial (RBI) [36], que es lo que se usa para calcular el trazo de las curvas de

las isolíneas con diferentes puntos [37]. Para esto se usan los datos de tiempo o espacio de cada punto y la spline de placa delgada (“thin-plate spline”) $\phi(\varepsilon) = \varepsilon^2 \cdot \log \varepsilon$, que es la superficie que interpola un conjunto dado de puntos [38], donde la variable ε representa la distancia euclídea entre los puntos MDS en el plano. Las isolíneas representan lugares idénticos de tiempo o resultados que, obviamente, están fuertemente correlacionados. Un ejemplo de una representación con isolíneas sería el siguiente (3.4) (imagen tomada de la referencia [39]):

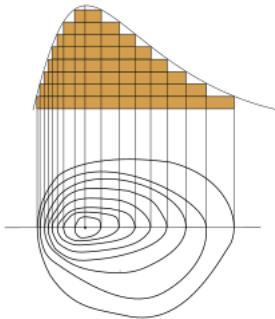


Figura 3.4: Representación de una “montaña” con isolíneas.

Con respecto a los elementos representados en el gráfico, el i -ésimo equipo en la k -ésima ronda es un elemento (es decir, un punto) en el conjunto de líneas y puntos resultantes del MDS. En consecuencia, se describen $N(R+1) = N(2N - 1)$ elementos, cada uno con una dinámica de poder representada por un vector de estado $(k+1)$ -dimensional en el plano complejo $S_i(k) = P_i(k) + iQ_i(k)$ (ver 3.3), con $0 \leq k \leq k_r$ y $k_r = 0, \dots, R$ [1]. Esto es lo que es representado con las líneas y puntos del MDS resultante (Figuras 3.5 y 3.6 tomadas del Capítulo 3 de la referencia [1]).

La matriz D de dimensión $N(2N - 1)$, es la matriz de entrada que alimenta el MDS, cuyos elementos representan diferencias entre pares de equipos (entre el i -ésimo equipo en la k -ésima ronda). Para calcular D , se utiliza la distancia de Manhattan, $d_M(S_i(k), S_j(k))$ [40]:

$$d_M(S_i(k), S_j(k)) = \sum_{n=1}^{k+1} (|P_i(n) - P_j(n)| + |Q_i(n) - Q_j(n)|). \quad (3.11)$$

Con todo esto se obtienen los gráficos de la figura 3.5 con las isolíneas que representan las rondas (tiempo) y los gráficos de la figura 3.6, cuyas isolíneas representan los puntos obtenidos (espacio). Los gráficos resultantes son los siguientes:

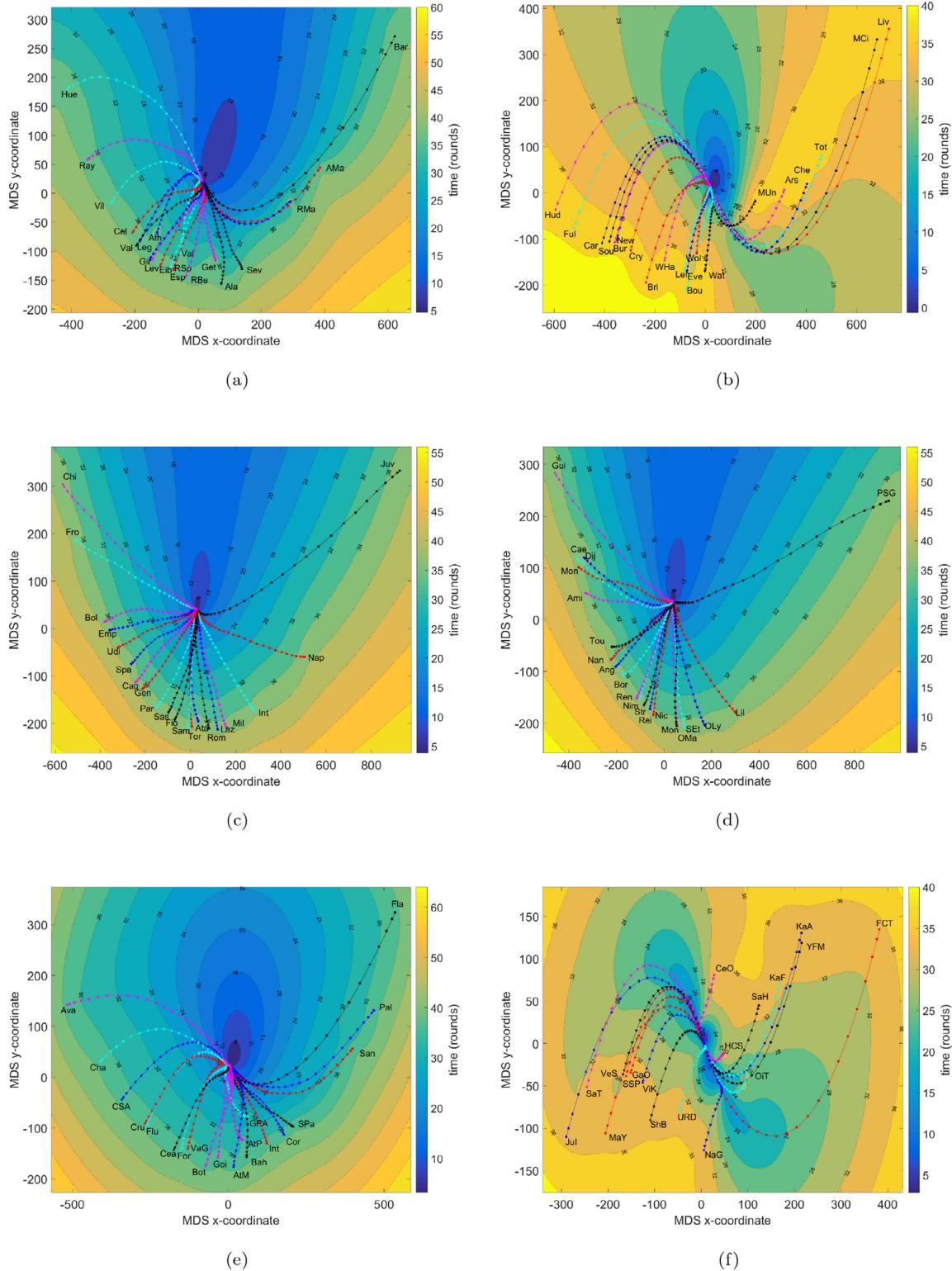


Figura 3.5: Los gráficos resultantes del MDS con dimensión $M = 2$ de $S_i(k) = P_i(k) + \iota Q_i(k)$, $0 \leq k \leq k_r$, $k_r = 0, \dots, R$, con la distancia de Manhattan d_M para las ligas: (a) “La Liga”; (b) “Premiership”; (c) “Serie A”; (d) “Ligue 1”; (e) “Série A”; (f) “J1 League”. La tercera dimensión es calculada por RBI en función del tiempo (rondas).

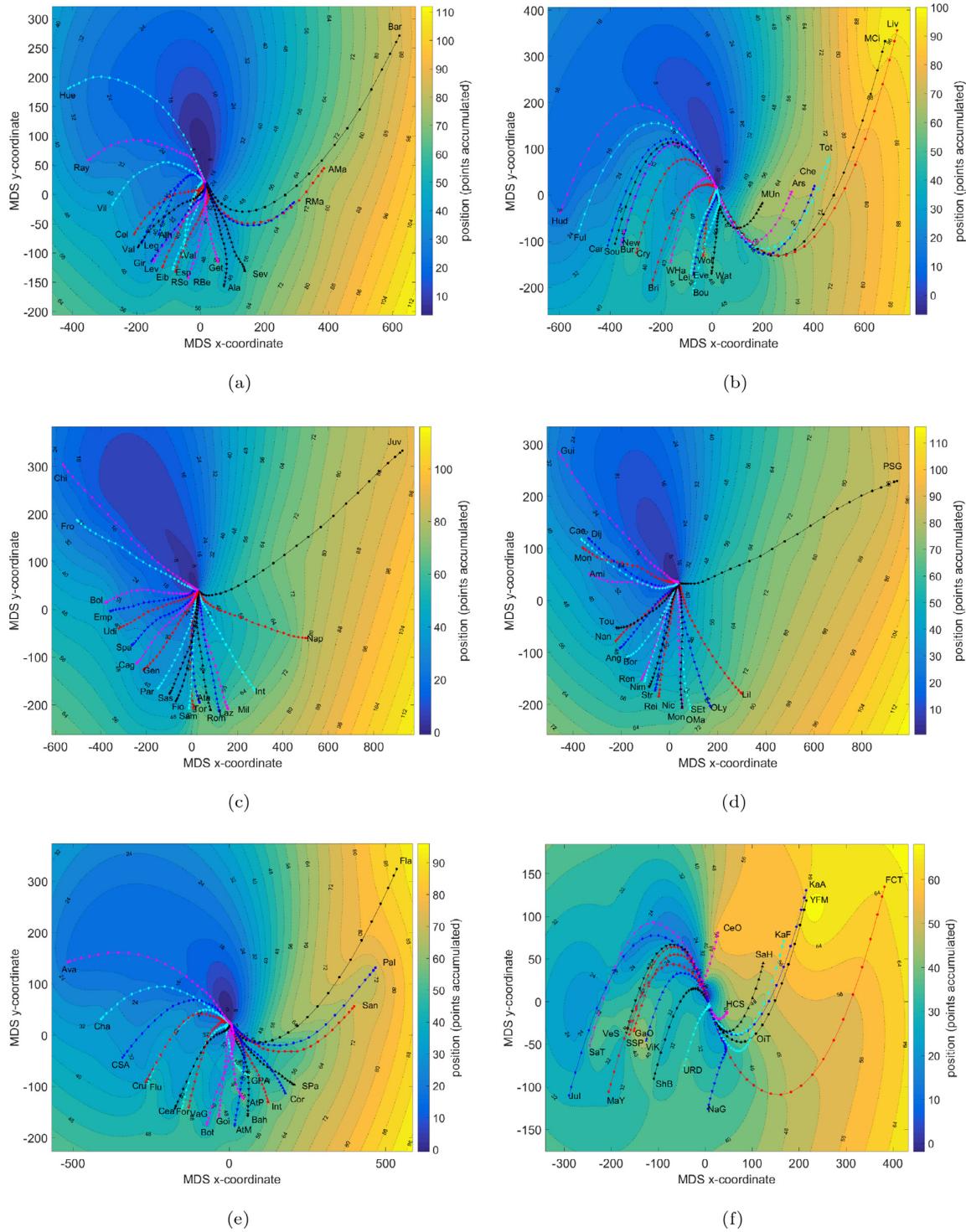


Figura 3.6: Los gráficos resultantes del MDS con dimensión $M = 2$ de $S_i(k) = P_i(k) + \imath Q_i(k)$, $0 \leq k \leq k_r$, $k_r = 0, \dots, R$, con la distancia de Manhattan d_M para las ligas: (a) “La Liga”; (b) “Premiership”; (c) “Serie A”; (d) “Ligue 1”; (e) “Série A”; (f) “J1 League”. La tercera dimensión es calculada por RBI en función del espacio (puntos acumulados).

Tanto en la figura 3.5 como en la 3.6, se puede observar en todos los gráficos cierta forma de montaña invertida formada por las isolíneas, igual que en un mapa topográfico. Solo en el caso de la (f) “J1 League” esta “montaña” es más irregular y difícil de ver que en el resto por su irregular reparto de puntos, como mencionamos en la sección 3.1.2.

Se puede observar también la aparición de líneas fractales, correspondientes a equipos, que evolucionan desde el “centro” (desde $k = 0$) hacia los bordes (hasta $k = R$). Por esto y porque cuanto mayor sea la dispersión de estos puntos, mayor será la competitividad la liga [1], se calcularán la dimensión fractal y la entropía en la sección 3.3.

Además de la distancia de Manhattan d_M (3.11), se consideraron otras distancias para medir las diferencias entre los equipos como se mencionó en la sección 2.2. Estas distancias son: la distancia de Canberra d_C (3.12) y la distancia de Sørensen d_S (3.13) [21], dadas por:

$$d_C (\mathbf{S}_i(k), \mathbf{S}_j(k)) = \sum_{n=1}^{k+1} \frac{|P_i(n) - P_j(n)|}{|P_i(n)| + |P_j(n)|} + \sum_{n=1}^{k+1} \frac{|Q_i(n) - Q_j(n)|}{|Q_i(n)| + |Q_j(n)|}, \quad (3.12)$$

$$d_S (\mathbf{S}_i(k), \mathbf{S}_j(k)) = \frac{\sum_{n=1}^{k+1} |P_i(n) - P_j(n)| + \sum_{n=1}^{k+1} |Q_i(n) - Q_j(n)|}{\sum_{n=1}^{k+1} (|P_i(n)| + |P_j(n)|) + \sum_{n=1}^{k+1} (|Q_i(n)| + |Q_j(n)|)}. \quad (3.13)$$

Usando estas fórmulas de distancia para el MDS, se obtienen los gráficos de la figura 3.7 con las isolíneas que representando las rondas (tiempo) y los gráficos de la figura 3.8, con las isolíneas que representando los puntos obtenidos (espacio). Los gráficos solo representan los datos de “La Liga”, siendo los gráficos de la izquierda aquellos en los que se usa la distancia de Canberra d_C y los de la derecha en los que se usa la distancia de Sørensen d_S . Dichos gráficos son los siguientes (las 2 figuras tomada del Capítulo 3 de la referencia [1]):

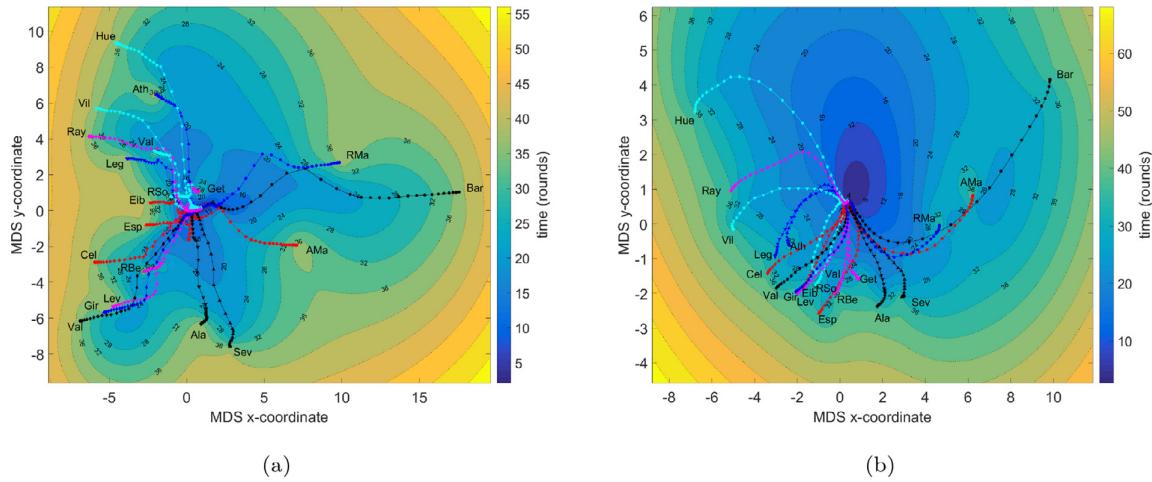


Figura 3.7: Los gráficos resultantes del MDS con dimensión $M = 2$ de $S_i(k) = P_i(k) + \iota Q_i(k)$, $0 \leq k \leq k_r$, $k_r = 0, \dots, R$, para “La Liga”: (a) distancia de Canberra d_C ; (b) distancia de Sørensen d_S . La tercera dimensión es calculada por RBI en función del tiempo (rondas).

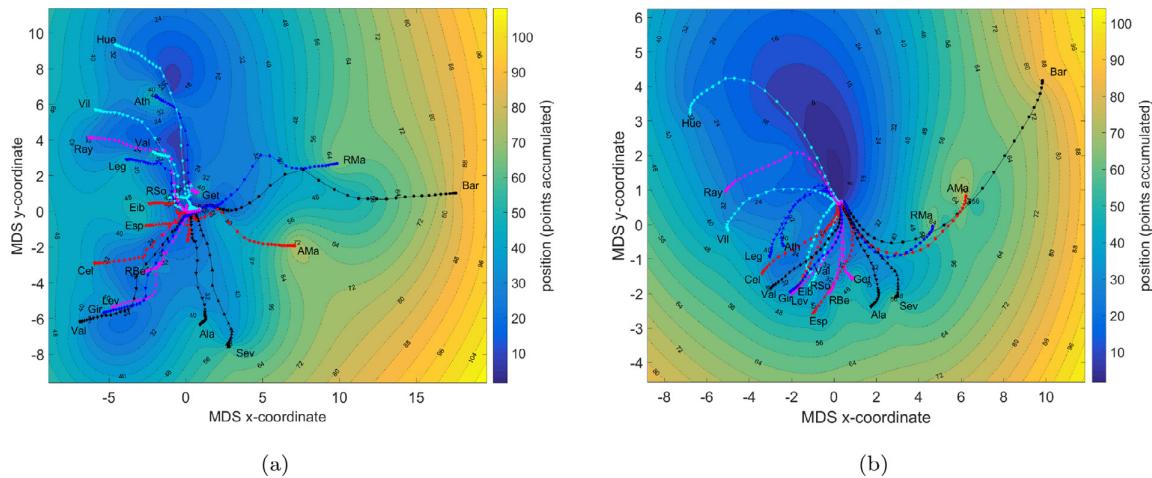


Figura 3.8: Los gráficos resultantes del MDS con dimensión $M = 2$ de $S_i(k) = P_i(k) + \imath Q_i(k)$, $0 \leq k \leq k_r$, $k_r = 0, \dots, R$, para “La Liga”: (a) distancia de Canberra d_C ; (b) distancia de Sørensen d_S . La tercera dimensión es calculada por RBI en función del espacio (puntos acumulados).

Ambas figuras (3.7 y 3.8) ilustran los gráficos resultantes del MDS con dimensión $M = 2$ de $S_i(k)$ usando la distancia de Canberra d_C (3.12) y la distancia de Sørensen d_S (3.13) para “La Liga”. Cabe destacar que, aunque no se incluyan, los resultados son similares para las otras ligas.

Y, puesto que estos resultados son más irregulares que los obtenidos con la distancia de Manhattan (figuras 3.5 y 3.6), se toma la distancia de Manhattan d_M (3.11) como métrica final.

3.2.2. Resultados y Análisis

Para evaluar la “calidad” de los resultados del MDS se han adoptado principalmente 3 medidas:

1. El diagrama de Shepard.
2. Gráfica de S frente a M.
3. Representación de las características del conjunto de datos.

El diagrama de Shepard (1) sirve para comprobar cómo de cerca están D y \hat{D} . Consiste en una gráfica que enfrenta a $d(v_i, v_j)$ ($= d(S_i, S_j)$) contra $d(\hat{v}_i, \hat{v}_j)$. Por tanto, una buena representación corresponde a puntos con una dispersión baja, es decir, una representación donde los puntos rojos se encuentran cerca de la línea negra intermitente 1:1 (Figura 3.9 tomada del Capítulo 3 de la referencia [1]). El diagrama resultante es el siguiente:

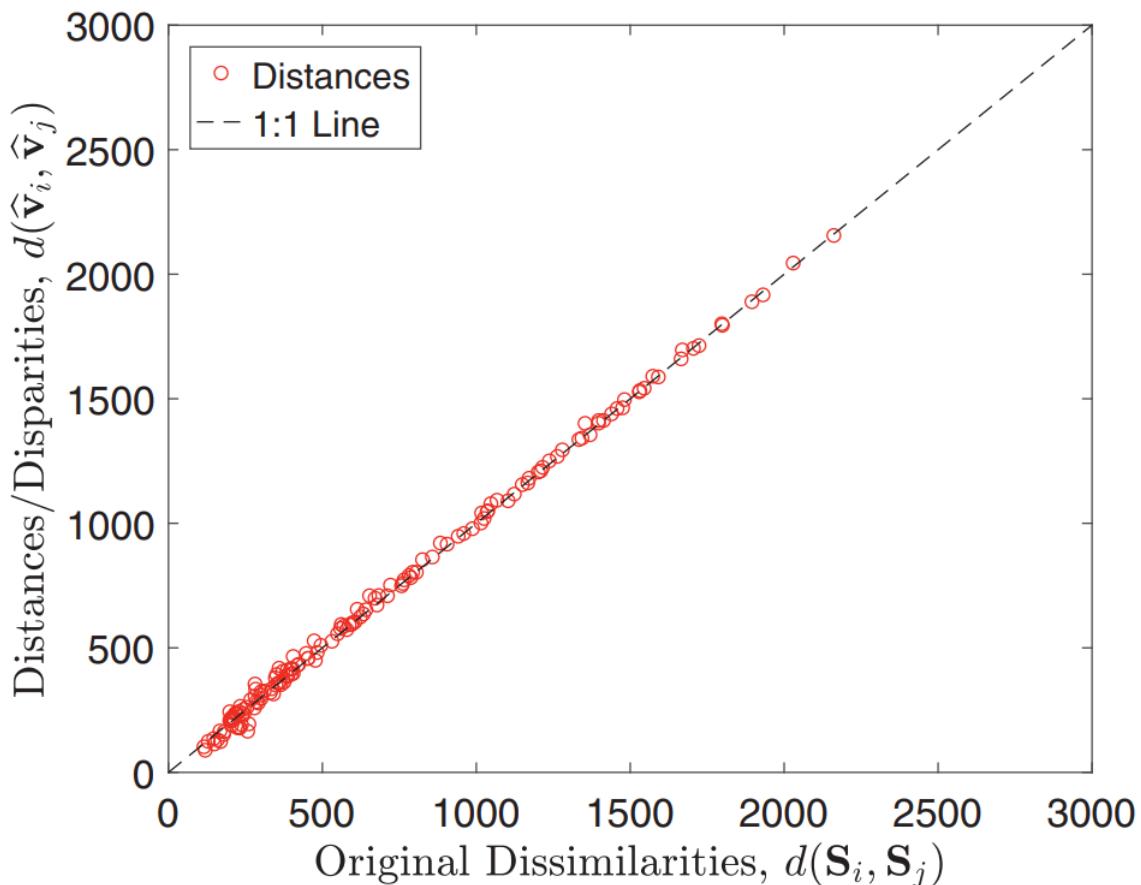


Figura 3.9: Diagrama de Shepard para “La Liga” con distancia de Manhattan (d_M).

Como podemos observar, hay bastante poca dispersión, estando prácticamente todos los puntos rojos sobre la línea negra intermitente 1:1 de 45° . Esto significa que para la distancia entre 2 equipos en un momento dado de la temporada, la $d(\hat{v}_i, \hat{v}_j)$ obtenida usando únicamente $M = 2$ para el MDS, obtiene un resultado muy similar a la $d(S_i, S_j)$ original, lo que indica que tenemos resultados de calidad del MDS para la dimensión $M = 2$.

Además de esto, la gráfica de S frente a M (2) muestra como el estrés S va reduciéndose a medida que aumentamos la dimensión M (Figura 3.10 tomada del Capítulo 3 de la referencia [1]). En este caso, un buen resultado se da cuando el estrés S se reduce bastante entre cada valor de M . La gráfica resultante es la siguiente:

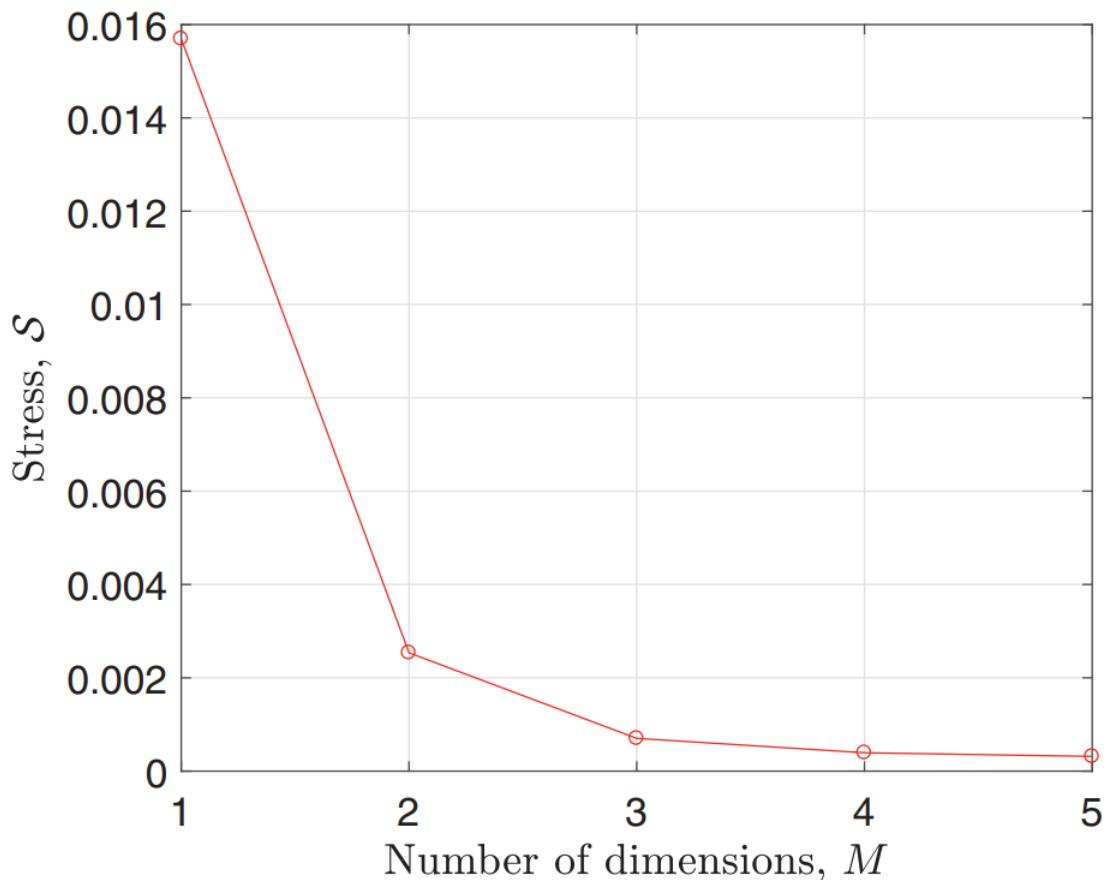


Figura 3.10: Diagrama de estrés S frente a la dimensión escogida M para “La Liga” con distancia de Manhattan (d_M).

Se puede distinguir que entre $M = 1$ y $M = 2$ hay una bajada relevante del estrés S , sin embargo, también se observa que a partir de $M = 2$, todas las bajadas son pequeñas, y, a medida

que aumentamos la M , estas son cada vez menores. Lo que significa que un MDS de dimensión $M = 2$ como el nuestro, tiene bastante calidad comparado con uno de dimensión $M = 1$; y que, no por aumentar a $M \geq 3$, se obtendrán unos resultados mucho mejores, solo se obtendrán unos resultados menos legibles por tener al menos una dimensión más en su representación (3.2.1).

Para escoger una representación fiel de las características de los datos (3) se toma la distancia de Manhattan d_M , en vez de la la distancia de Canberra d_C (3.12) o la distancia de Sørensen d_S (3.13). El motivo es porque los resultados de las figuras, los de las figuras 3.7 y 3.8, que representan a d_C y a d_S , son mucho más irregulares que los de las figuras 3.5 y 3.6 que corresponden a d_M , y, por tanto, es más difícil trabajar con ellos y son menos fieles a los datos de los puntajes de donde parten.

Debe tenerse en cuenta que, por lo general, distintas distancias conducen a diferentes gráficos, como en este caso, pero la información de donde surgen es la misma. Por lo tanto, la pregunta principal a la hora de elegir una distancia para medir las diferencias de un elemento a otro es si los gráficos de MDS generados se pueden interpretar fácilmente. A menudo, para llegar a los gráficos significativos más representativos de un conjunto de datos dado, se tienen que probar distintas métricas y elegir la más adecuada. Es por esto que se han tenido en cuenta 3 distancias diferentes antes de escoger la distancia de Manhattan d_M (3.11) como la más fehaciente a los datos.

Lo cierto es que este análisis nos habla de la calidad de los resultados, pero no de su competitividad. Puesto que a mayor dispersión, mayor será la competitividad [1], en la sección 3.3 se analizará la entropía y la dimensión fractal tanto de la figura 3.5 como de la figura 3.2 vistas en las secciones 3.2 y 3.1, respectivamente. Estos resultados se compararán con medidas clásicas de competitividad para ver si los gráficos obtenidos representan fielmente la competitividad en las ligas.

3.3. Competitividad

Por último, se llevará a cabo el análisis de la competitividad para cada liga. Para este análisis, se tendrá en cuenta la dimensión fractal b y la entropía H de los gráficos obtenidos a raíz de los procesos de Clusterización (figura 3.2) y Escalamiento multidimensional (figuras 3.5 y 3.6). Además, los resultados obtenidos se compararán con otras medidas clásicas de competitividad:

la desviación típica y el índice de balance competitivo de Herfindahl-Hirschman.

Tanto para el cálculo de la dimensión fractal b como para la entropía H , no usaremos las figuras 3.2 y 3.5 tal cual, será necesario “limpiarlas” antes. Por consiguiente, de los gráficos se suprimirán los ejes, cuadrículas, etiquetas, puntos y contornos, manteniendo solo las líneas que representan el desempeño de los equipos. A cada figura se le agregarán píxeles de fondo alrededor del gráfico una vez se coloca este en el centro para obtener una imagen con una resolución de $2^p \times 2^p$, donde p corresponderá al valor de la primera potencia de 2 mayor que el lado más grande de la imagen de la figura original. Una vez hecho esto, se transformarán los píxeles de las líneas a blanco y negro (líneas en negro y fondo en blanco).

Para la entropía H , además de todo lo anterior, las líneas serán sustituidas por una sucesión de puntos igualmente espaciados, a diferencia de las líneas propiamente dichas que habrá para el cálculo de la dimensión fractal b .

Como tanto en la limpieza para el cálculo de la dimensión fractal b como para el cálculo de la entropía H , además de otros elementos, se omite el fondo de los gráficos (las isolíneas y los colores), dará igual que se utilice la figura 3.5 o la figura 3.6 para estos cálculos, puesto que es lo único que cambia entre ellas. Finalmente, se realizarán con la figura 3.5, pero perfectamente se podrían haber producido a partir de la figura 3.6.

3.3.1. Metodologías empleadas

La dimensión fractal se define como un exponente que permite calcular el número de elementos iguales o copias creadas en función del tamaño de la escala reductora o factor de reducción aplicado [23]. Formalmente (3.14), como se menciona en la sección 2.3, vendrá definida por [22]:

$$\begin{aligned} N^o \text{ de copias} &= \text{Factor de reducción}^{\text{Dimensión fractal}}; \\ \log(N^o \text{ de copias}) &= \log(\text{Factor de reducción}^{\text{Dimensión fractal}}); \\ \log(N^o \text{ de copias}) &= \text{Dimensión fractal} \cdot \log(\text{Factor de reducción}); \\ \text{Dimensión fractal} &= \frac{\log(N^o \text{ de copias})}{\log(\text{Factor de reducción})}. \end{aligned} \tag{3.14}$$

La dimensión fractal es una generalización de la dimensión euclíadiana [23]. Para dar una idea de qué es la dimensión fractal y qué significado tiene, se mostrará con un ejemplo.

Si se considera que una hoja de papel es un objeto bidimensional, se podría coger y arrugar esta hoja hasta formar una bola. El objeto tendría volumen, pero no sería tridimensional porque la bola estaría llena de huecos y discontinuidades. La dimensión fractal intenta medir en qué grado un objeto de dos dimensiones llena el espacio de tres dimensiones. Análogamente, y en el caso de este TFG, la dimensión fractal de un objeto de dimensión 1 intenta medir en qué grado un objeto de una dimensión llena el espacio de dos dimensiones, es decir, cuánto espacio es cubierto por las líneas unidimensionales en un gráfico bidimensional. Un ejemplo más simple que se analizará más tarde es el de la imagen 3.15, donde se compara la dimensión fractal de la costa de Gran Bretaña con la dimensión fractal de la costa de Noruega.

Volviendo a los resultados observados en la figura 2.1 en la sección 2.3, se tomará que $Longitud = \frac{1}{Factor\ de\ reducción}$, es decir, $L = 1/l$. Al analizar la primera columna, partiendo de un segmento de longitud 1, si este se divide en segmentos de longitud L , entonces se obtendrán $N(l)$ partes, de manera que (3.15):

$$N(l) \cdot L^1 = N(l) \cdot (1/l)^1 = 1, \quad (3.15)$$

independientemente de cuál sea la l o la L (imagen 3.11):

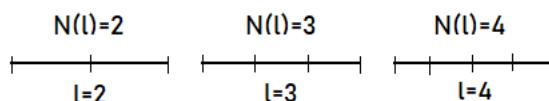


Figura 3.11: Número de copias ($N(l)$) y Longitud ($L = 1/l$), para una Dimensión fractal (D)= 1.

De la misma forma, si el objeto inicial es un cuadrado de superficie 1, y se compara con unidades cuadradas cuyos lados tengan longitud L cada uno, el número de unidades o copias necesarias para recubrirlo $N(l)$, cumplen que (3.16):

$$N(l) \cdot L^2 = N(l) \cdot (1/l)^2 = 1, \quad (3.16)$$

cualquiera que sea la l o la L (imagen 3.12):

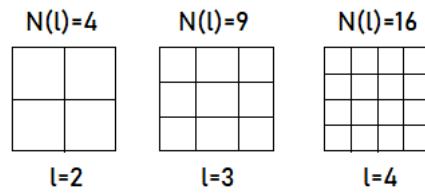


Figura 3.12: Número de copias ($N(l)$) y Longitud ($L = 1/l$), para una Dimensión fractal (D)= 2.

Por último, siguiendo la misma tendencia, si el objeto tomado es tridimensional, como, por ejemplo, un cubo de volumen 1, y se mide en relación a unidades que sean cubos de arista L , entonces se observa que (3.17):

$$N(l) \cdot L^3 = N(l) \cdot (1/l)^3 = 1, \quad (3.17)$$

ajeno al valor de l o L (imagen 3.13):

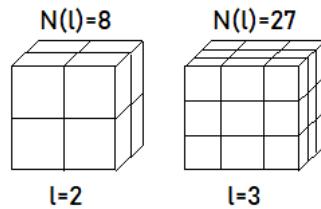


Figura 3.13: Número de copias ($N(l)$) y Longitud ($L = 1/l$), para una Dimensión fractal (D)= 3.

De todo esto se puede generalizar (3.18) que la dimensión fractal de un objeto geométrico D es:

$$N(l) \cdot L^D = N(l) \cdot (1/l)^D = 1, \quad (3.18)$$

donde $N(l)$ es el número de objetos elementales, o de unidades, de tamaño $L (= 1/l)$ que recubren, o que completan, el objeto.

De donde se deduce al despejar D (3.19), igual que en la ecuación 3.14, que:

$$D = \frac{\log(N(l))}{\log(1/L)} = \frac{\log(N(l))}{\log(l)}. \quad (3.19)$$

Esto es aplicable también a figuras que repiten su estructura subsecuentemente a diferentes escalas, por ejemplo, una espiral infinita o el triángulo de Sierpinski (figura 3.14 tomada de la referencia [23]), es decir, lo que comúnmente se conoce como fractales [41].

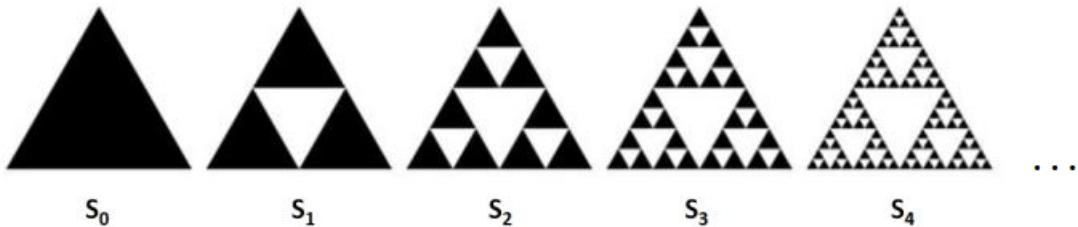


Figura 3.14: Triángulo de Sierpinski.

De hecho, tras realizar un análisis (3.20) de la dimensión fractal del triángulo de Sierpinski (3.14), se obtiene que:

$$\text{Dimensión fractal} = \lim_{k \rightarrow \infty} \frac{\log 3^k}{\log 2^k} = \frac{\log 3}{\log 2} \approx 1,585, \quad (3.20)$$

de donde se extrae que el *Nº de copias* = 3^∞ porque a cada iteración se crean 3 triángulos nuevos (el triángulo invertido del interior no se tiene en cuenta como parte de la figura), y que el *Factor de reducción* = 2^∞ pues, al igual que antes, en cada iteración, los lados exteriores de los triángulos que forman la nueva figura, se dividen en 2. El motivo por el que está elevado a ∞ es porque el triángulo de Sierpinski hace esta división en triángulos más pequeños un número infinito de veces.

Por otro lado, a la hora de analizar los gráficos de las figuras 3.2 ,3.5 y 3.6, donde sus fractales no son tan triviales, el cálculo de la dimensión fractal en estos nos permitirá conocer cómo de irregular o/y cuánto ruido tienen los gráficos que hay en ellas.

Por ejemplo, la siguiente imagen (3.15) muestra la dimensión fractal de la costa de Gran Bretaña frente a la de Noruega (imagen tomada de la referencia [42]):

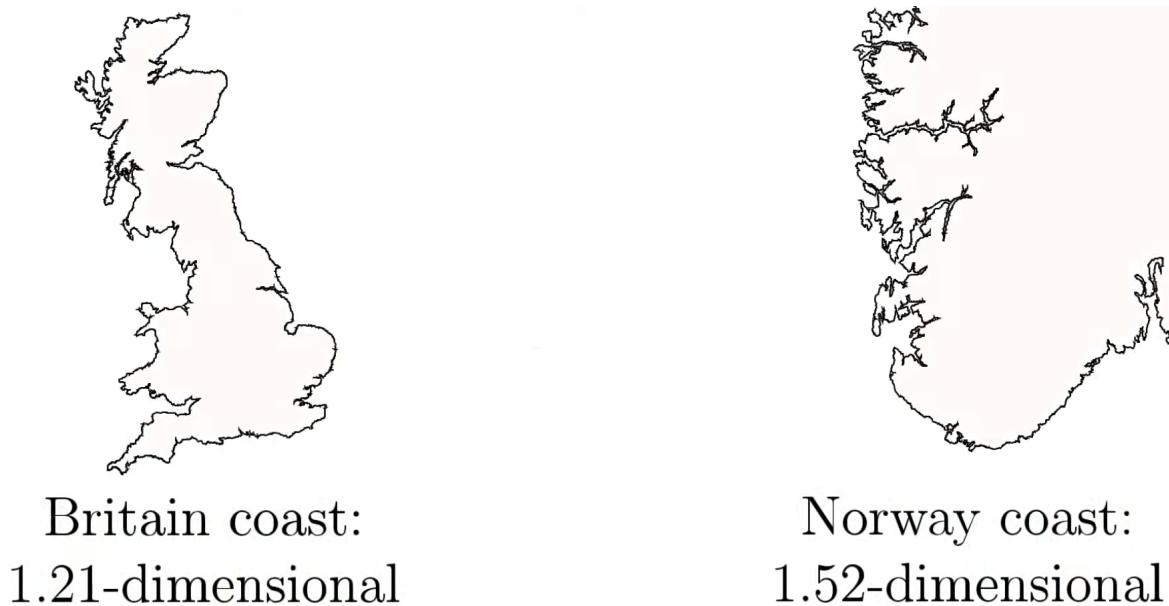


Figura 3.15: Comparación de la dimensión fractal de la costa de Gran Bretaña (izquierda) frente a la dimensión fractal de la costa de Noruega (derecha)

Se comprueba que la dimensión fractal de la costa de Gran Bretaña tiene menos dimensión fractal por ser más recta y sin irregularidades; sin embargo, la costa de Noruega tiene una dimensión fractal mayor debido a su costa irregular y poco uniforme [42]. Lo que aquí se calcula es la dimensión fractal de su costa, es decir, la línea de su contorno, no se tiene en cuenta el interior, igual que se hará con los gráficos por ser solo líneas que no contienen superficie maciza en su interior.

En este caso, y en el de los gráficos de las figuras 3.2, 3.5 y 3.6, al no ser figuras hipotéticas como el triángulo de Sierpinski (3.14), se aplica un método por el que averiguar la dimensión fractal de una imagen. Para ello se recurre a la dimensión de Minkowski–Bouligand en un espacio Euclídeo, también conocida como la dimensión de conteo de cajas, que es un método con el que determinar la dimensión fractal de un conjunto S en un espacio euclidiano R^n [43]. Este conteo se basa en poner una cuadrícula por encima de la imagen y contar el número de cajas que están tocadas por la imagen en cuestión. Para ello, o bien se aumenta esta imagen al doble de su tamaño y se vuelve a contar de nuevo, o bien se reduce el numero de cuadrados de la cuadrícula a la mitad, manteniendo la imagen al mismo tamaño antes de contar de nuevo [22]. En cualquiera de los 2 casos el resultado es el mismo.

Para el conteo de cajas, la metodología empleada será un algoritmo que irá reduciendo el

tamaño de los cuadrados de la cuadrícula, aumentando el número de cuadrados totales en esta cuadrícula mientras mantiene constante el tamaño de la imagen antes de contar y reducir de nuevo. El tamaño de los cuadrados (cajas) irán desde el tamaño de los lados de 2^p píxeles, hasta un tamaño de 2^2 píxeles, que es el tamaño del grosor de las líneas [1]. El algoritmo en cuestión (1) que nos permitirá calcular la dimensión fractal b será el siguiente:

Algorithm 1 *Cálculo de b* **input**= $\varepsilon = 2^p$, objeto fractal F **output**= $b(F)$

- 1: Elige $\varepsilon = 2^p$;
 - 2: **repeat**
 - 3: cubre el objeto fractal F con una cuadrícula compuesta por cuadrados iguales, con tamaño ε ;
 - 4: cuenta el número de cajas (cuadrados) $N_\varepsilon(F) \in \mathbb{N}$ que cubren el fractal F ;
 - 5: disminuye ε en potencia de base 2 ((new) $\varepsilon =$ (old) $\varepsilon^{\frac{1}{2}}$);
 - 6: **until** $\varepsilon = 2^2$
 - 7: Calcula $b(F) = -\lim_{\varepsilon \rightarrow 0^+} \frac{\log N_\varepsilon(F)}{\log \varepsilon}$;
-

Este cálculo final de la dimensión fractal $b(F)$ es así porque lo que sale del algoritmo de conteo de cajas es la gráfica: $N_\varepsilon(F)$ - ε . Esta gráfica mostrará una curva, pues la relación entre el número de cajas tocadas y el factor de reducción no es lineal, sino exponencial, así que se recurre a aplicar logaritmos en ambos lados con el fin de que la dimensión fractal quede como la pendiente de la gráfica: $\log N_\varepsilon(F)$ - $\log \varepsilon$. El desarrollo (3.21) es tal que así:

$$\begin{aligned} \text{Número de cajas tocadas} &= \text{Factor de reducción}^{\text{Dimensión fractal}}; \\ N_\varepsilon(F) &= \varepsilon^{b(F)}; \\ \log N_\varepsilon(F) &= \log \varepsilon^{b(F)}; \quad (3.21) \\ \log N_\varepsilon(F) &= b(F) \cdot \log \varepsilon; \\ b(F) &= \frac{\log N_\varepsilon(F)}{\log \varepsilon}. \end{aligned}$$

Puesto que el factor de reducción ε va cada vez a menos (de 2^p a 2^2), en vez de a más, se pondrá en negativo para tener una dimensión fractal $b(F)$ resultante en positivo. Además, como se toman varios valores reduciendo este ε a cada iteración hasta llegar a 0 (en realidad a 2, pero da igual, porque solo serían 2 iteraciones más, que realmente serían irrelevantes e imprecisas),

se pondrá el cálculo final como un límite con ε tendiendo a 0 (se deja a 0 y no a 2 por claridad en la solución), realmente 0^+ . En consecuencia, la fórmula para el cálculo de la dimensión fractal b (3.22) es la siguiente:

$$b(F) = - \lim_{\varepsilon \rightarrow 0^+} \frac{\log N_\varepsilon(F)}{\log \varepsilon}, \quad (3.22)$$

donde $b(F)$ es la Dimensión fractal, $\log N_\varepsilon(F)$ el Número de cajas tocadas y ε el Factor de reducción.

En cuanto a la entropía H , y como hemos mencionado en el capítulo 2.3, es una medida que nos permite averiguar la incertidumbre de una fuente de información. Si se trata de una imagen, nos permitirá conocer la cantidad de ruido, desorden o dispersión que esta tiene [24]. En el caso de los gráficos a analizar, significará que si la entropía es alta, entonces los puntos del gráfico (obtenidos en función del desempeño de los equipos) estarán más dispersos, y de la misma manera, si la entropía es baja, entonces los puntos estarán más juntos y menos dispersos.

Por ejemplo, la siguiente imagen (3.16) muestra claramente las diferencias en la entropía que hay entre unos puntos ordenados y desordenados (imagen tomada de la referencia [44]):

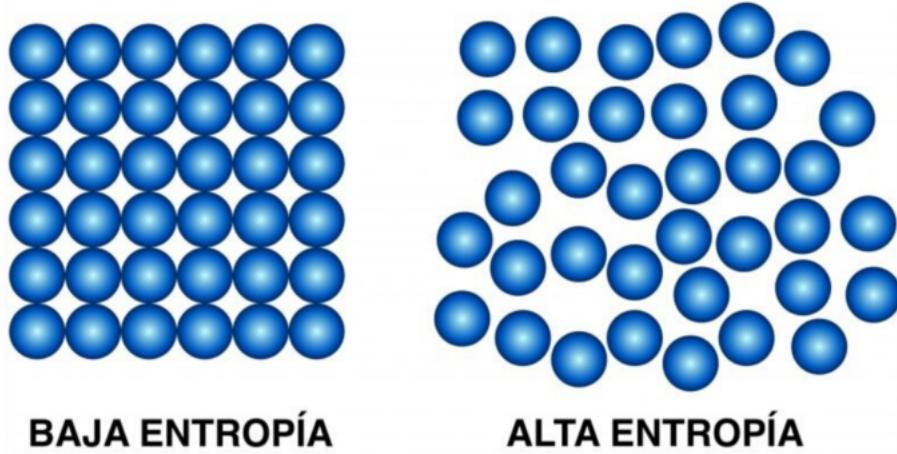


Figura 3.16: Comparación de la entropía entre puntos ordenados (izquierda) frente a puntos desordenados (derecha)

Para el cálculo de esta entropía H en los gráficos (figuras 3.2 y 3.5) se dividen las imágenes en $N_x \times N_y$ celdas. El tamaño de estas celdas será el doble que el tamaño de un punto resultante

de la “limpieza” de las líneas punteadas (limpieza explicada en la introducción a la sección 3.3). Además se calculan las probabilidades P_{xy} en función del número relativo de píxeles (negros) dentro de cada celda (x, y) , con $x = 1, \dots, N_x$ e $y = 1, \dots, N_y$. Es decir, P_{xy} (3.23) se define tal que:

$$P_{xy} = \frac{\text{cantidad de píxeles negros de la celda } (x, y)}{\text{total de píxeles (blancos y negros) de la celda } (x, y)}. \quad (3.23)$$

La fórmula empleada para el cálculo de la entropía H será la definida por Shannon (3.24) [45]. Dicha fórmula es la siguiente:

$$H = \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} -P_{xy} \log P_{xy}. \quad (3.24)$$

Hay que tener en cuenta que $0 \leq P_{xy} \leq 1$, por lo que, si se da $0 \cdot \log 0$, se asumirá que $0 \cdot \log 0 = 0$.

Tanto el método con el que se calcula la dimensión fractal b , como el método con el que se calcula la entropía H son bastante similares. La principal diferencia es que, a la hora de calcular la dimensión fractal, el conteo de cajas es binario, la caja o toca la línea o no la toca, no existe la posibilidad de estar tocándola en un 50 %. Sin embargo, en el cálculo de la entropía se hace un cálculo proporcional para saber qué cantidad de píxeles negros hay en cada celda, y no por haber algo de píxeles negros en una celda ya se considera que hay un 100 % de píxeles negros en esa celda, cosa que sí que ocurre en el cálculo de la dimensión fractal.

Cabe señalar que se considerará que una complejidad alta corresponde a valores altos de b o de H , que se traducen en competitividad alta; por el contrario, una complejidad baja corresponde a valores bajos de b o de H , lo que se traduce en una baja competitividad [1].

Una vez calculadas la dimensión fractal b y la entropía H , se comprobará cómo de buenas son estas medidas comparándolas con 2 medidas clásicas de competitividad: la desviación típica σ y el índice de balance competitivo de Herfindahl-Hirschman $HICB$.

La desviación típica σ será relativa a la proporción de victorias que logra cada equipo en una temporada. Esto indicará cómo de cerca o de lejos estarán todos los datos de la media [25]. La fórmula resultante para σ (3.25) es la siguiente:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(w_i - \frac{1}{2} \right)^2}, \quad (3.25)$$

donde N representa el número total de equipos en la liga y w_i representa la proporción de victorias del i -ésimo equipo en la temporada (3.26), por lo que $0 \leq w_i \leq 1$. Es decir:

$$w_i = \frac{\text{victorias del equipo } i \text{ en la liga esa temporada}}{\text{partidos totales del equipo } i \text{ en la liga esa temporada}}. \quad (3.26)$$

En general, cuanto mayor es el valor de σ , menor es la competitividad. En una liga hipotética, donde cada equipo tuviera la misma probabilidad de ganar un partido, tendríamos $\sigma = 0$.

Siguiendo con el índice de balance competitivo de Herfindahl-Hirschman $HICB$ (3.27), este dirá cómo de concentrados o dispersos están los puntos repartidos en la clasificación final de la liga [26]. Vendrá definido por la siguiente fórmula [10]:

$$HICB = 100N \sum_{i=1}^N s_i^2, \quad (3.27)$$

donde N sigue correspondiendo al número total de equipos en la liga y s_i ahora representa la proporción de puntos del i -ésimo equipo en la temporada (3.28), por lo que $0 \leq s_i \leq 1$. Es decir:

$$s_i = \frac{\text{puntos del equipo } i \text{ en la liga esa temporada}}{\text{suma de puntos de todos los equipos en la liga esa temporada}}. \quad (3.28)$$

De manera similar a la desviación típica, un aumento en el valor de $HICB$ significa una disminución en la competitividad, y viceversa [10]. Esto se ve mucho mejor con el índice de Herfindahl para las empresas (3.29), donde el $HICB$ (se le llamará $HICB_{emp}$ para diferenciarlo del $HICB$ de los cálculos relativos a los gráficos) se calcula tal que:

$$HICB_{emp} = \sum_{i=1}^{N^o \text{ empresas}} (\text{participación la empresa } i)^2. \quad (3.29)$$

La principal diferencia entre el $HICB$ y el $HICB_{emp}$ es que $0 \leq \text{participación la empresa } i \leq 100$, por lo que en el $HICB$ se multiplica por 100 y por N para compensar los datos pequeños producidos por s_i ($0 \leq s_i \leq 1$). Tanto $HICB$ y $HICB_{emp}$ miden lo mismo, cómo de concentrados o dispersos están los elementos participantes de un conjunto [26].

Por lo tanto, analizando ejemplos con $HICB_{emp}$, se observa que si se da un monopolio (1 empresa con 100 % de participación) entonces $HICB_{emp} = 10000$, que es el máximo valor posible de $HICB_{emp}$; si hay 2 empresas al 50 % cada una, entonces $HICB_{emp} = 5000$. Por último, si se examina un ejemplo donde hay 20 empresas con sus participaciones repartidas equitativamente (20 empresas al 5 %), entonces se obtiene un $HICB_{emp} = 500$; sin embargo, con que solo 1 de ellas tenga un 25 %, y el resto tengan sus participaciones repartidas equitativamente, esto provocará que el $HICB_{emp}$ aumente hasta casi el doble, dando un $HICB_{emp} = 921$ [26].

Es por esto que, a mayor $HICB$, menor competitividad, y viceversa, pues en el caso de las ligas de fútbol, un $HICB$ bajo significa que los puntos están muy mal repartidos, sabiendo entonces quién va a ganar o perder desde el principio de la temporada; sin embargo, un $HICB$ alto corresponde a una liga con los puntos repartidos de forma similar, dando como resultado una liga muy apretada y con mucha incertidumbre de quién acabará primero y quién último, siendo los puestos muy variables y, por tanto, dando como resultado, una liga muy competitiva.

Una vez aclarados los términos y el cómo se calcularan la dimensión fractal b , la entropía H , la desviación típica σ y el índice de balance competitivo de Herfindahl-Hirschman $HICB$; se procede a hacer sus cálculos (3.17), obteniendo como resultado la siguiente tabla (tabla tomada del Capítulo 4 de la referencia [1]):

	b		H		σ	$HICB$
	b_S	2-dim MDS	H_S	2-dim MDS		
'J1 League'	1.241	0.149	1.349	0.081	0.170	106.301
'La Liga'	1.239	0.147	1.345	0.078	0.191	106.928
'Premiership'	1.180	0.140	1.318	0.061	0.208	109.095
'Ligue 1'	1.238	0.143	1.332	0.062	0.210	109.307
'Série A'	1.180	0.142	1.316	0.059	0.217	109.740
'Serie A'	1.179	0.140	1.315	0.058	0.220	111.652

Figura 3.17: Tabla con las medidas de competitividad b_S , b_{MDS} , H_S , H_{MDS} , σ y $HICB$ (nombra- das de izquierda a derecha) para la “J1 League”, “La Liga”, “Premiership”, “Ligue 1”, “Série A” y “Serie A”, ordenadas de menor a mayor σ y $HICB$.

donde la b_S y H_S se refieren a la dimensión fractal y la entropía de la figura 3.2 respectivamente, y la b_{MDS} y H_{MDS} corresponden a la dimensión fractal y la entropía de la figura 3.5 (se toma la figura 3.5, pero perfectamente se podía haber tomado la figura 3.6 como se menciona

en el proceso de “limpieza” descrito en la sección 3.3). La σ y el $HICB$ no tienen subíndice indicando sobre qué gráfico se realizan los cálculos porque estos son independientes de ellos. En el caso de la desviación típica σ , sus cuentas dependen de la proporción de victorias de los equipos, y en el caso del índice de balance competitivo de Herfindahl-Hirschman $HICB$, dependen de la proporción de puntos obtenidos por cada equipo.

En la siguiente sección (3.3.2) analizaremos y compararemos los resultados de unas medidas con otras.

3.3.2. Resultados y Análisis

Una vez realizados todos los cálculos, se sacarán en forma de gráfica para poder visualizar de mejor manera los resultados obtenidos. De este modo, se nos permitirá distinguir las discrepancias entre las medidas de dimensión fractal b y de entropía H de los gráficos, frente a las medidas clásicas de competitividad de desviación típica σ y de índice de balance competitivo de Herfindahl-Hirschman $HICB$. Las gráficas resultantes en cuestión (figura 3.18 tomada del Capítulo 4 de la referencia [1]) se ven representados en la siguiente figura:

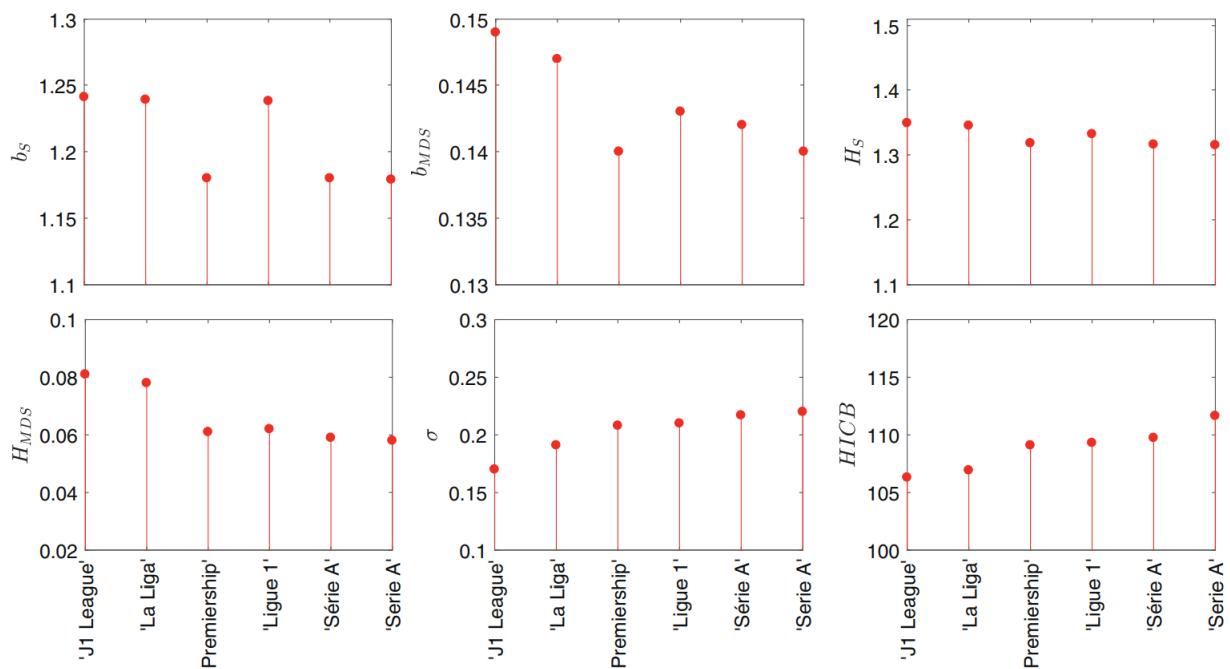


Figura 3.18: Los resultados obtenidos de b_S , b_{MDS} , H_S , H_{MDS} , σ y $HICB$. Las ligas están en orden descendente de competitividad según las medidas clásicas σ y $HICB$.

Todas las gráficas siguen el orden “J1 League”, “La Liga”, “Premiership”, “Ligue 1”, “Série A” y “Serie A”, que es el orden que representa ir de mayor a menor competitividad tanto para σ como para $HICB$, es decir, de menor a mayor σ y $HICB$, pues sus valores son inversamente proporcionales a la competitividad, como se menciona en la sección 3.3.1.

Al observar estas gráficas con más detenimiento, se ve que, a pesar de estar ordenado por σ y $HICB$, realmente son las medidas $\{H_{MDS}, \sigma\}$ las que revelan una alta capacidad para discriminar la competitividad, $\{b_{MDS}, HICB\}$ las que indican una capacidad media, y por último $\{b_S, H_S\}$ muestran una capacidad baja. Tener una capacidad alta para discriminar la competitividad significa que tus valores varían dentro de intervalos más amplios, mientras que una capacidad baja implica que varían en intervalos más estrechos [1]. Cabe añadir que los índices b y H capturan para la “Premiership” algunos comportamientos no observables con las medidas clásicas de σ y $HICB$, señalando que, en este caso, se necesita un análisis más detallado para determinar qué medidas sobre la “Premiership” se equivocan, cuáles aciertan o/y si existe alguna mejor inclusive.

Para calcular la correlación real que hay entre cada par de medidas utilizadas, se acudirá al coeficiente de correlación de Pearson. Este coeficiente es un índice que se utiliza con el fin de medir el grado de relación entre 2 variables cuantitativas y continuas [46], como es el caso de estas 6 variables ($b_S, b_{MDS}, H_S, H_{MDS}, \sigma$ y $HICB$). La fórmula (3.30) usada para ello, debido a ser de una muestra de 6 (se han analizado 6 ligas, $n = 6$) y no de una población, es la siguiente [47]:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (3.30)$$

donde \bar{x} e \bar{y} representan la media muestral, por lo que $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, y, de forma análoga, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. Lo que se representa aquí es la correlación que tiene la medida x frente a la medida y , así que tanto x como y solo pueden tomar los valores de $\{b_S, b_{MDS}, H_S, H_{MDS}, \sigma, HICB\}$. Y los valores de i y j representarán las 6 ligas analizadas, es decir, que tomarán los valores de {“J1 League”, “La Liga”, “Premiership”, “Ligue 1”, “Série A”, “Serie A”}, siendo, por ello, $n = 6$.

El gráfico resultante (figura 3.19 tomada del Capítulo 4 de la referencia [1]) será de diferentes colores y solo mostrará los arcos donde $|r_{xy}| \geq 0,9$, dando lugar a:

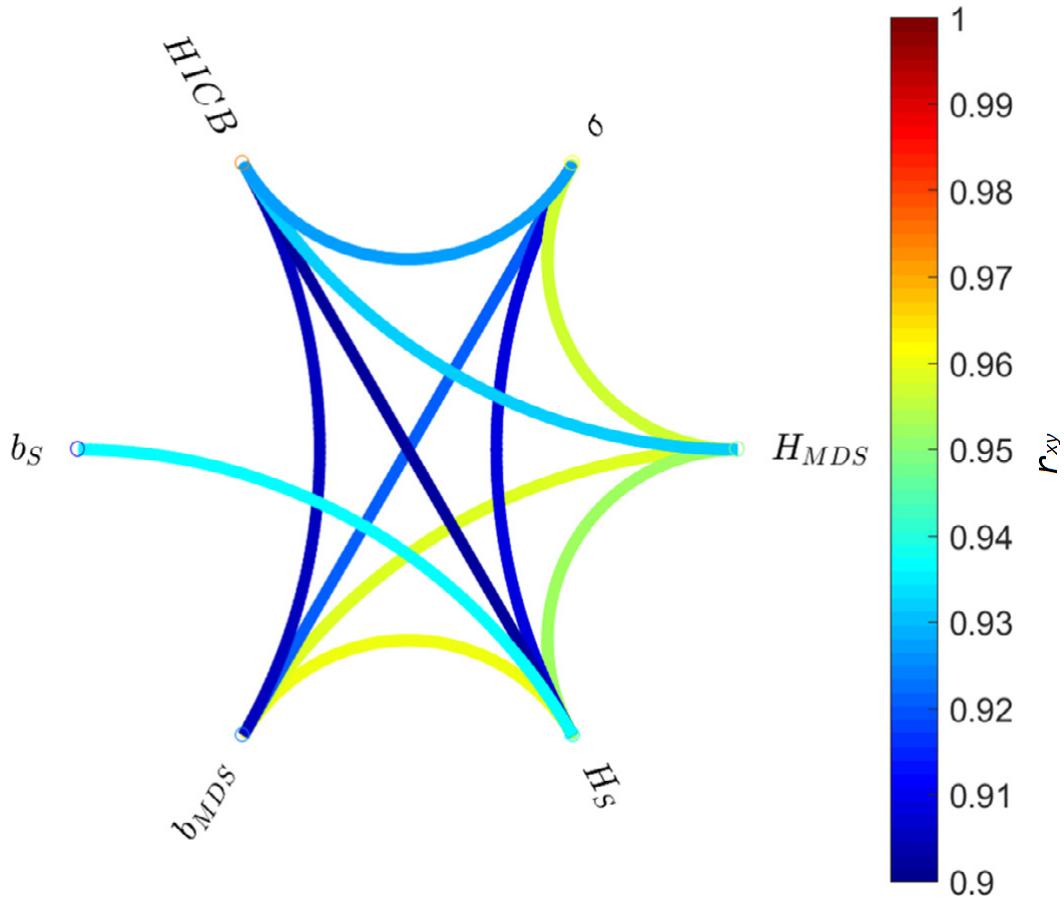


Figura 3.19: Coeficiente de correlación de Pearson r_{xy} entre las medidas de competitividad b_S , b_{MDS} , H_S , H_{MDS} , σ y $HICB$, para la “J1 League”, “La Liga”, “Premiership”, “Ligue 1”, “Série A” y “Serie A”. Los arcos denotan los valores donde $|r_{xy}| \geq 0,9$ y los colores son proporcionales a r_{xy} según el mapa de colores proporcionado a la derecha.

Se observa que hay muchas líneas que se han quedado sin representarse por tener un $|r_{xy}| < 0,9$ entre la medida x y la y , esto es para estar seguros de que se mantenía una correlación alta y relevante entre las medidas representadas. Se considera que $|r_{xy}| = 1$ es una correlación perfecta (por ejemplo, $|r_{xx}| = 1$ siempre), mientras que $|r_{xy}| = 0$ se interpreta como absolutamente ninguna relación entre x e y , de donde se deduce trivialmente, que a mayor $|r_{xy}|$, mayor correlación. Generalmente, si $|r_{xy}| \geq 0,5$ se estima que x e y tienen un grado alto de correlación, si $0,3 \leq |r_{xy}| < 0,5$ entonces se aprecia un grado medio de correlación, y si $|r_{xy}| < 0,3$ significa que poseen un bajo grado de correlación [48].

Es por esto, que al estudiar los resultados obtenidos del gráfico de la figura 3.19, se concluye que, como se había teorizado anteriormente, tanto la dimensión fractal b , como la entropía H

son buenos medidores de la competitividad en ligas de fútbol de acuerdo a la desviación típica σ y al índice de balance competitivo de Herfindahl-Hirschman *HICB*.

Capítulo 4

Conclusiones y trabajos futuros

En este capítulo se detallan las conclusiones derivadas del Trabajo de Fin de Grado realizado y la propuesta de posibles trabajos futuros.

4.1. Conclusiones

Finalmente, tras haber abordado de manera extensa y meticulosa todos los puntos del artículo: “Modeling and visualizing competitiveness in soccer leagues” [1], consideramos que, en este Trabajo de Fin de Grado, se ha ido más allá de lo expuesto en este artículo, y de esta manera explicar exhaustivamente aquello que el citado artículo deja con ambigüedades o no clarificaba. Todo ello con el fin último y común de abordar qué liga es la más competitiva.

El artículo [1], ha utilizado conceptos de teoría de sistemas, poder complejo, clusterización y Escalamiento Multidimensional (MDS) para estudiar la competitividad en las ligas nacionales de fútbol, dividiéndose este artículo en tres fases principalmente.

En una primera fase, los puntos acumulados ganados y perdidos por los equipos fueron interpretados como potencia activa y reactiva (P_i y Q_i), y procesados en el plano complejo con el poder complejo (S_i). Siguiendo en esta misma fase, se realizó un proceso de Clusterización o Agrupamiento de los equipos pertenecientes a cada una de estas ligas a modo de ejercicio informativo adicional, para determinar qué equipos se disputaban los mismos objetivos en sus respectivas ligas.

En una segunda fase, se adoptó la técnica computacional del Escalamiento Multidimensional (ó MDS) para abordar los datos de los puntos obtenidos por cada equipo durante una

temporada. Ambos enfoques se utilizaron para producir gráficos de elementos que representan la evolución relativa de los equipos a lo largo de una temporada de liga. Estos gráficos se interpretaron bajo el prisma de poder determinar la competitividad de cada liga.

Para cuantificar la competitividad de estas ligas con los gráficos obtenidos, se entra en la tercera y última fase, calculando la Dimensión fractal b y la Entropía H de estos. Los resultados se compararon con los obtenidos con las medidas clásicas de competitividad de la desviación típica σ y el índice de balance competitivo de Herfindahl-Hirschman $HICB$, mostrando así, la viabilidad de usar estas medidas de dimensión fractal y entropía para analizar la competitividad.

Asimismo, cabe señalar que la investigación adicional puede centrarse en medidas alternativas y más sensibles, capaces de capturar mayor cantidad de características de los gráficos y de mejor calidad incluso. Todo ello con el objetivo final de alcanzar alguna medida más objetiva que tenga una fuerte e inequívoca correlación con la competitividad. Algunas de estas posibles investigaciones adicionales se mencionan en la sección 4.2.

4.2. Trabajos futuros

A pesar de haber explicado y expandido el artículo: “Modeling and visualizing competitiveness in soccer leagues” [1] en todos sus puntos, hay, al menos, un total de 3 materias de las que se podrían realizar futuros trabajos para mejorar lo expresado en este:

1. Anomalía en la “Premiership”.
2. Calcular P-valor.
3. Informatizar y automatizar los cálculos.

En primer lugar, los resultados observados en la figura 3.18 presentan una anomalía en la “Premiership” (1), como se menciona en la sección 3.3.1. Para detectar por qué ocurre, se podría analizar la competitividad en distintos momentos de la temporada, además de coger un ejemplo donde se sepa objetivamente que la competitividad es mayor en él y realizar los mismos cálculos de nuevo para asegurar que es un caso aislado y no un fallo recurrente. Otra opción más sencilla, sería tomar otra medida clásica de competitividad más con la que comprobar si esta da la razón a los casos en los que ocurre esta la anomalía de la “Premiership” o a los casos en los que no (en σ y $HICB$).

Siguiendo con el P-valor (2), este serviría para determinar si la correlación entre las variables es significativa, comparando el P-valor con su nivel de significancia α . Por lo general, un nivel de significancia α de 0,05 funciona bien. Una α de 0,05 indica que el riesgo de concluir que existe una correlación, cuando en realidad no existe ninguna correlación, es de un 5 %. El P-valor señala si el coeficiente de correlación (en este caso r_{xy}) es significativamente diferente de 0 (un coeficiente de 0 indica que no existe una relación lineal). Además, a continuación del cálculo de este P-valor, se podría examinar la relación monótona entre las variables mediante el cálculo del coeficiente de correlación de Spearman [49].

Y, por último, la informatización y automatización (3) de los cálculos permitiría un análisis de la competitividad en las ligas mucho más rápido, ágil e interesante. Es por ello, que realizaré un segundo Trabajo de Fin de Grado en el Grado en Ingeniería Informática como continuación y complemento de este, donde se llevarán a cabo muchas de las ideas mostradas y expresadas en este TFG para informatizar las materias aquí vistas. Y, de esta forma, combinando ambas disciplinas, culminar con esta pareja de Trabajos de Fin de Grado, de forma armoniosa mi doble titulación en Matemáticas e Ingeniería Informática.

Bibliografía

- [1] A. M. Lopes and J. A. Machado, “Modeling and visualizing competitiveness in soccer leagues,” *Applied Mathematical Modelling*, vol. 92, pp. 136–148, 2021.
- [2] “Top 10 Most Popular Sports in The World.” [Online]. Available: <https://sportsshow.net/top-10-most-popular-sports-in-the-world/>
- [3] “fútbol — Definición — Diccionario de la lengua española — RAE - ASALE.” [Online]. Available: <https://dle.rae.es/f%C3%BAtbol>
- [4] “Raúl - Perfil del jugador — Transfermarkt.” [Online]. Available: <https://www.transfermarkt.es/raul/profil/spieler/7349>
- [5] “Raúl - Selección — Transfermarkt.” [Online]. Available: <https://www.transfermarkt.es/raul/nationalmannschaft/spieler/7349>
- [6] “Three points for a win - Wikipedia.” [Online]. Available: https://en.wikipedia.org/wiki/Three_points_for_a_win#Year_of_adoption_of_three_points_for_a_win
- [7] “La FIFA obliga a todas las ligas a conceder tres puntos por victoria — Deportes — EL PAÍS.” [Online]. Available: https://elpais.com/diario/1994/10/29/deportes/783385203_850215.html
- [8] “WALKOVER — Definition of WALKOVER by Oxford Dictionary on Lexico.com also meaning of WALKOVER.” [Online]. Available: <https://www.lexico.com/definition/walkover>
- [9] The FA, “Law 3 The players,” 2019. [Online]. Available: <https://www.thefa.com/football-rules-governance/lawsandrules/laws/football-11-11/>

- law-3---the-playershttp://www.thefa.com/football-rules-governance/lawsandrules/laws/football-11-11/law-3---the-players
- [10] R. Criado, E. García, F. Pedroche, and M. Romance, “A new method for comparing rankings through complex networks: Model and analysis of competitiveness of major european soccer leagues,” *Chaos*, vol. 23, no. 4, 2013.
- [11] “Como calculan las casas de apuestas las cuotas.” [Online]. Available: <https://ekuatio.com/como-calculan-las-casas-de-apuestas-las-cuotas/>
- [12] E. E. País, “La casa de apuestas que no pudo comprar la fe de los más leales fans del Leicester,” may 2016. [Online]. Available: https://verne.elpais.com/verne/2016/05/03/articulo/1462286064_102841.html
- [13] “Matemáticas - Universidad Rey Juan Carlos.” [Online]. Available: <https://www.urjc.es/estudios/grado/1245-matematicas#competencias>
- [14] “worldfootball.net.” [Online]. Available: <https://www.worldfootball.net/>
- [15] “Cluster analysis - Wikipedia.” [Online]. Available: https://en.wikipedia.org/wiki/Cluster_analysis
- [16] “Agrupamiento jerárquico - Wikipedia, la enciclopedia libre.” [Online]. Available: https://es.wikipedia.org/wiki/Agrupamiento_jer%C3%A1rquico
- [17] “Hierarchical Clustering — solver.” [Online]. Available: <https://www.solver.com/xlminer/help/hierarchical-clustering-intro>
- [18] “Multidimensional scaling - Wikipedia.” [Online]. Available: https://en.wikipedia.org/wiki/Multidimensional_scaling
- [19] “Geometría no euclídea - Wikipedia, la enciclopedia libre.” [Online]. Available: https://es.wikipedia.org/wiki/Geometr%C3%A1_no_euclidea
- [20] “Tema 7. Escalamiento multidimensional.” [Online]. Available: http://eio.usc.es/eipc1/BASE/BASEMASTER/FORMULARIOS-PHP/MATERIALESMMASTER/Mat_14_master0809multi-tema7.pdf

- [21] E. Deza, M. M. Deza, M. M. Deza, and E. Deza, “Encyclopedia of Distances,” in *Encyclopedia of Distances*. Springer Berlin Heidelberg, 2009, pp. 1–583. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-00234-2_1
- [22] “(127) dimension fractal - YouTube.” [Online]. Available: <https://www.youtube.com/watch?v=frkzApKfn20>
- [23] “Dimensión fractal - Wikipedia, la enciclopedia libre.” [Online]. Available: https://es.wikipedia.org/wiki/Dimensi%C3%B3n_fractal
- [24] “Entropía (información) - Wikipedia, la enciclopedia libre.” [Online]. Available: [https://es.wikipedia.org/wiki/Entrop%C3%A9a_\(informaci%C3%B3n\)](https://es.wikipedia.org/wiki/Entrop%C3%A9a_(informaci%C3%B3n))
- [25] “Desviación típica - Wikipedia, la enciclopedia libre.” [Online]. Available: https://es.wikipedia.org/wiki/Desviaci%C3%B3n_t%C3%ADpica
- [26] “Índice de Herfindahl - Wikipedia, la enciclopedia libre.” [Online]. Available: https://es.wikipedia.org/wiki/%C3%8Dndice_de_Herfindahl
- [27] “What is Active, Reactive, Apparent and Complex Power?” [Online]. Available: <https://www.electricaltechnology.org/2013/07/active-reactive-apparent-and-complex.html>
- [28] “Sistemas Continuos y Discretos - 19E37 - Academia de Ciencias.” [Online]. Available: http://19e37.com/wiki/index.php/Sistemas_Continuos_y_Discretos
- [29] “2.5 Variables discretas y continuas — Estadística Básica Edulcorada.” [Online]. Available: <https://bookdown.org/aquintela/EBE/variables-discretas-y-continuas.html>
- [30] “Real coordinate space - Wikipedia.” [Online]. Available: https://en.wikipedia.org/wiki/Real_coordinate_space
- [31] “UPGMA - Wikipedia.” [Online]. Available: <https://en.wikipedia.org/wiki/UPGMA>
- [32] “Cardinality - Wikipedia.” [Online]. Available: <https://en.wikipedia.org/wiki/Cardinality>
- [33] “Coeficientes por países — Coeficientes UEFA — UEFA.com.” [Online]. Available: <https://es.uefa.com/memberassociations/uefarankings/country/#/yr/2019>

- [34] J. A. Tenreiro Machado and A. M. Lopes, “Multidimensional scaling locus of memristor and fractional order elements,” pp. 147–157, sep 2020.
- [35] Courant, Richard, H. Robbins, and I. Stewart, *What is Mathematics?: An Elementary Approach to Ideas and Methods* - *Herbert Robbins Richard Courant (Ian Stewart), Richard Courant, Courant Institute of Mathematical Sciences Richard Courant, Herbert Robbins, Ian Stewart, Herbert (Professor of Mathematics,* 1996. [Online]. Available: https://books.google.es/books?id=_kYBqLc5QoQC&pg=PA344&redir_esc=y#v=onepage&q&f=false
- [36] J. C. Carr, W. Richard Fright, and R. K. Beatson, “Surface interpolation with radial basis functions for medical imaging,” *IEEE Transactions on Medical Imaging*, vol. 16, no. 1, pp. 96–107, 1997.
- [37] “Data Interpolation with Radial Basis Functions (RBFs).” [Online]. Available: <http://shihchinw.github.io/2018/10/data-interpolation-with-radial-basis-functions-rbfs.html>
- [38] “Numerical methods - What are Thin-Plate-Splines? - Mathematics Stack Exchange.” [Online]. Available: <https://math.stackexchange.com/questions/2563276/what-are-thin-plate-splines>
- [39] “Contour line - Wikipedia.” [Online]. Available: https://en.wikipedia.org/wiki/Contour_line#cite_ref-1
- [40] Y. H. Cha, Y. W. Lee, S. M. Nam, J. M. Han, Y. J. Rhee, H. K. Cha, and Y. U. Jeong, “Temporal characterization of ASE separated from main pulses in a 10-TW Ti:sapphire laser,” in *Pacific Rim Conference on Lasers and Electro-Optics, CLEO - Technical Digest*, 2007.
- [41] B. Mandelbrot, W. Freeman, and Company, *The Fractal Geometry of Nature*, ser. Einaudi paperbacks. Henry Holt and Company, 1983. [Online]. Available: <https://books.google.es/books?id=0R2LkE3N7-oC>
- [42] “(25) Fractals are typically not self-similar - YouTube.” [Online]. Available: <https://www.youtube.com/watch?v=gB9n2gHsHN4&t=963s>

- [43] “Minkowski–Bouligand dimension - Wikipedia.” [Online]. Available: https://en.wikipedia.org/wiki/Minkowski\OT1\textendashBouligand_dimension
- [44] “Leyes de la Termodinámica: características y ejemplos.” [Online]. Available: <https://www.caracteristicas.co/leyes-de-la-termodinamica/>
- [45] “Registrado de Imágenes.” [Online]. Available: <https://iie.fing.edu.uy/investigacion/grupos/gti/timag/trabajos/2011/registrado/Entropia.html>
- [46] “Coeficiente de correlación de Pearson - Wikipedia, la enciclopedia libre.” [Online]. Available: https://es.wikipedia.org/wiki/Coeficiente_de_correlaci%C3%B3n_de_Pearson
- [47] “Pearson correlation coefficient - Wikipedia.” [Online]. Available: https://en.wikipedia.org/wiki/Pearson_correlation_coefficient
- [48] “Pearson’s Correlation Coefficient - Statistics Solutions.” [Online]. Available: <https://www.statisticssolutions.com/pearsons-correlation-coefficient/>
- [49] Minitab, “Interpret the key results for Correlation,” pp. 1–9, 2016. [Online]. Available: <http://support.minitab.com/en-us/minitab-express/1/help-and-how-to/modeling-statistics/regression/how-to/correlation/interpret-the-results/>