

Universidad Católica del Uruguay



Facultad de Ciencias Empresariales

Trabajo Final de Grado

para la obtención del título de

Licenciada en Economía

Jugador clave y optimización en equipos de fútbol:

Enfoque desde el análisis de redes sociales

Jéssica Aplanalp

Tutor: Carlos Díaz

Montevideo, septiembre de 2020

Índice

Índice	1
Índice de tablas	2
Índice de figuras	3
Resumen.....	4
Introducción	5
Marco Teórico	8
I. Importancia del análisis de redes sociales	8
II. Concepto y tipos de redes.....	8
III. Concepto de red aplicado al fútbol	10
IV. Modelos de formación de redes y literatura relacionada.....	10
V. Medidas de centralidad.....	12
i. Centralidad In y Out Degree	12
ii. Centralidad between	14
VI. Función de producción: conceptos y vínculos con el deporte	16
i. Antecedentes.....	16
Caso de estudio: análisis de redes sociales aplicado a equipos que participaron de la UEFA Champions League	18
I. Formación de redes de fútbol	19
II. Datos y Metodología: procesamiento de datos.....	19
III. Datos y Metodología: posiciones en el campo de juego.....	21
IV. Datos y Metodología: procedimiento estadístico y validez de los resultados.....	22
i. Jugador clave	22
ii. Estimación de la relación entre medidas de centralidad y goles marcados.....	23
V. Resultados	30
Discusión de los resultados.....	34
Limitaciones	39
Conclusiones	41
Referencias.....	42
Anexos:.....	46

Índice de tablas

Tabla 1. Estadísticos descriptivos (media \pm desvío e Intervalo de confianza al 95%) de las centralidades entre posiciones, p -valor del test de Welch y tamaño del efecto promedio de Cohen.	30
Tabla 2. Resultados de la estimación (robusta) de la función que relaciona goles anotados y medidas de centralidad.....	32
Tabla 3. Equipos participantes del estudio y número de competiciones jugadas	48
Tabla 4. Posiciones y su participación en el total de los datos	49
Tabla 5. Comprobación de normalidad para las centralidades.....	50
Tabla 6. Prueba de Levene para la centralidad Between.....	53
Tabla 7. Prueba de Levene para la centralidad in degree	53
Tabla 8. Prueba de Levene para la centralidad out degree	54
Tabla 9. Test de Welch y tamaño del efecto de Cohen para la centralidad between.	58
Tabla 10. Test de Welch y tamaño del efecto de Cohen para la centralidad in degree	60
Tabla 11. Test de Welch y tamaño del efecto de Cohen para la centralidad out degree	62
Tabla 12. Matriz de correlaciones	65
Tabla 13. Test de Breusch – Pagan y Cook – Weisberg.....	65
Tabla 14. Prueba de inflación de varianza	66
Tabla 15. Estadísticos descriptivos de las variables incluidas en el modelo	66
Tabla 16. Estimación de la relación entre goles anotados y medidas de red y dos variantes	67

Índice de figuras

Figura 1. Tipos de redes, forma matricial.....	9
Figura 2. Red de pases del Bayern Múnich (vs Roma, 2014)	20
Figura 3. Asignación táctica de las posiciones.....	22
Figura 4. Comparación de las medias de centralidad por posiciones	31
Figura 5. Ejemplo de red, caso de familias Florentinas del Siglo XV	46
Figura 6. Redes ponderadas y no ponderadas	47
Figura 7. Porcentaje de temporadas jugadas por cada equipo	49
Figura 8. Diagrama de Box-Whisker para la centralidad between.....	55
Figura 9. Diagrama de Box-Whisker para la centralidad in degree.....	56
Figura 10. Diagrama de Box-Whisker para la centralidad out degree	57
Figura 11. Gráfico de probabilidad normal para los residuos estimados	64

Resumen

El presente trabajo aplica análisis de redes sociales a equipos de fútbol con la finalidad de encontrar un mecanismo sencillo y fácil de replicar, que sirva para encontrar la pieza clave dentro de cualquier grupo de trabajo, y a su vez, busca aproximar una relación entre los insumos y los *outputs* para cada equipo. Para ello, se realizó un estudio de caso en equipos de fútbol: se considera a cada equipo como una firma tradicional, que como tal, cuenta con un plantel de trabajadores y un *mánager* que los dirige y ordena en pos de maximizar beneficios.

Fueron analizados los niveles de centralidad de cada posición que un jugador puede adoptar en el campo de juego, para las temporadas 2012/2013 a 2018/2019 de la UEFA Champions League. Se formaron un total de 1.746 redes, ponderadas y dirigidas. A cada uno de los nodos se le calculó la centralidad *between*, *in degree* y *out degree*, capturando de esa forma distintos aspectos de la red. Se encontraron diferencias entre el promedio de centralidades para cada una de las posiciones en estudio. Los mediocampistas centrales resultaron ser los jugadores con mayores niveles de centralidad para las tres medidas estudiadas, confirmando así su importancia en el proceso de vinculación del equipo.

A su vez, se asume que el número de goles anotados por cada equipo y las medidas de centralidad de sus integrantes están directamente relacionados. Además, se piensa que otras variables que escapan del control de la institución -como la importancia de los equipos y la condición de local- tienen impacto en el marcador. La relación entre estos factores y el número de goles fue estimada por mínimos cuadrados ordinarios. Fue discutida la endogeneidad y se realizaron pruebas para verificar la consistencia de los estimadores. Los resultados de la estimación apuntan a que a mayor centralidad *between*, mayor cantidad de goles anotados. A su vez, las variables exógenas a la red impactan con el signo esperado: los equipos que juegan de local y enfrentan equipos relativamente menos importantes (en términos de partidos ganados, palmarés, etc.) convierten, en promedio, un mayor número de goles.

Introducción

El análisis de redes sociales se basa en la teoría de grafos¹ y, se caracteriza por plantear las interacciones sociales en términos de redes, las cuales están compuestas de nodos (actores individuales) y enlaces que los conectan.

Las redes de relaciones juegan roles centrales en una amplia variedad de interacciones sociales, económicas y políticas. Por lo tanto, si se desean conocer a fondo fenómenos que abarquen estas áreas, es fundamental comprender la estructura y/o proceso de formación de las redes que conforman y cómo influye cada actor a los demás miembros de la red. El campo de aplicación de éste tipo de análisis es muy amplio, se han realizado importantes investigaciones en sociología, economía, física, matemáticas y ciencias de la computación (Jackson, 2008).

En el presente texto se pretende realizar un estudio de redes sociales aplicado a grupos de trabajo dentro de firmas. Las firmas se consideran maximizadoras de beneficios y, por lo tanto, cada integrante de cada grupo debe ser elegido en pos de optimizar el rendimiento del equipo, esto ayudaría a conseguir un equipo más productivo, lo cual, con el tiempo se traduciría en mayores ganancias para la empresa. En este caso, se emplean datos de equipos de fútbol profesional. Se asume que cada entrenador elige los jugadores que formarán parte de cada partido siguiendo un criterio que le permitiría alcanzar su objetivo (o sea, marcar goles para ganar el partido), dada la información que tiene a disposición. A raíz de esta elección, y de la posterior interacción entre compañeros de equipo, se forman redes de juego, las que difieren unas de otras como consecuencia de las características individuales y colectivas.

Se llevará a cabo un estudio de caso para los equipos que disputaron la UEFA Champions League entre la temporada 2012/2013 y 2018/2019. Se persiguen dos objetivos. El primero es identificar a los jugadores clave dentro del campo de juego. Identificar al trabajador “clave” dentro de cualquier firma es importante porque de ser removido de su puesto, el equipo de trabajo podría no funcionar de forma correcta. Por ello, es importante identificarlo para darle los incentivos

¹ Estudio matemático sobre un conjunto de vértices conectados mediante un conjunto de aristas (Barnes y Harary, 1983).

correctos. El segundo es estimar una relación entre los insumos y los resultados para cada equipo.

Para el primer objetivo se formula la hipótesis de que los jugadores clave son los mediocampistas centrales, seguidos por los defensas centrales, porque al ubicarse en el centro de la red, el balón deberá “pasar” a través de ellos para alcanzar a los demás. Dicha hipótesis fue testeada con éxito por Clemente y Martins (2017), quienes consideraron las fases de eliminación directa de la temporada 2015/2016. El presente trabajo pretende ir un paso más allá, por lo que considera siete temporadas y todos los partidos a partir de la fase de grupos (hasta la final, inclusive).

Para el segundo objetivo se asume que el número de goles anotados es el de *output* de la actividad, los insumos utilizados para su formación son los jugadores y por ende las relaciones entre los mismos. Se plantea la hipótesis de que cuanto más conectada está la red², más goles se lograrán convertir, y, que la centralidad de mediocampistas centrales es la que más influye en el número de goles marcados³. Adicionalmente, se considera que existen factores exógenos a la red de juego que tienen impacto en el marcador del equipo, estos serán representados mediante variables que indican si el equipo es local o no, el tipo de equipo y el tipo de rival (medido en términos de partidos y torneos internacionales ganados)⁴.

Para el estudio, se tomaron en cuenta los datos relativos al número de pases entre jugadores de las temporadas 2012/2013 a 2018/2019 de la UEFA Champions League, siendo consideradas todas las fases⁵ de cada competición. Se conformaron un total de 1.746 redes, dos por partido (cada una referente a un equipo). Para realizar las pruebas, fueron calculadas tres medidas de centralidad para cada jugador: *Between*, *in degree* y *out degree*.

² La conexión de la red será medida mediante medidas de centralidad por posición. Se piensa que la medida de centralidad más influyente es el *between* (impactarían con signo positivo) y que *In degree* y *out degree* solo serían relevantes en posiciones que están cerca de los arcos (centralidades de posiciones cercanas al arco rival impactarían positivamente, mientras que las que están cerca del arco propio lo harían negativamente).

³ Porque, por hipótesis, son los jugadores más importantes en cuanto a la distribución del balón.

⁴ Se piensa que enfrentar a un equipo “importante” impacta de forma negativa en la cantidad de goles marcados; las demás variables impactarían positivamente.

⁵A partir de fase de grupos.

Para testear la primera hipótesis, se calculó la media de las tres medidas de centralidad antes mencionadas para cada posición en el campo de juego. A cada medida de centralidad por posición se le realizó el test de Welch⁶ (para saber qué tan comparables eran los grupos entre sí) y luego el test de Cohen⁷ para medir el tamaño del efecto hallado. Los resultados encontrados acerca de la centralidad, van de la mano (parcialmente) con la hipótesis planteada: los mediocampistas centrales fueron los jugadores más destacados, dejando en evidencia su poder como “creadores” de juego, tanto en generación de juego de ataque como defensivo. Por otra parte, los defensas laterales se encuentran justo por debajo en cuanto a niveles de centralidad que los mediocampistas centrales.

Para testear la segunda hipótesis fue planteado un modelo, estimado mediante mínimos cuadrados ordinarios, en el cual se relacionan los goles anotados con las medias de las centralidades *between e in degree* por posición y otras variables del contexto partido/equipo⁸. Fue discutida la posibilidad de endogeneidad, concluyéndose que ciertas observaciones podrían presentar simultaneidad, dado que no es seguro que tanto peso tendrían sobre la estimación final se recomienda cautela a la hora de considerar los resultados. Para finalizar, fueron estimadas dos variantes del modelo, en las que se van retirando las variables exógenas a la red, para verificar que los resultados obtenidos inicialmente fueran consistentes. Los resultados confirman, de forma preliminar, la hipótesis planteada, pero dado que no existe un modelo teórico detrás, no se puede afirmar que no exista otro tipo de relación que explique mejor los goles marcados por equipo.

⁶Que el test de Welch sea significativo implica que los grupos comparados son significativamente diferentes.

⁷Se utiliza para evaluar la magnitud de las diferencias entre los grupos estudiados.

⁸Dummy que indica si el equipo es local o no, dummy que indica si el rival es “grande” o no y dummy que indica si el equipo es “grande” o no. Nota: “grande” en término de torneos y partidos ganados internacionalmente.

Marco Teórico

I. Importancia del análisis de redes sociales

La sociedad en su conjunto puede ser pensada como una red, la cual se compone de personas y lazos que las unen. La organización de agentes individuales en redes y grupos tiene un papel importante en la determinación del resultado de muchas interacciones sociales y económicas⁹. Por ejemplo, las redes de contactos personales son importantes para obtener información sobre oportunidades laborales¹⁰ y juegan un papel importante en el comercio e intercambio de bienes¹¹. La división de sociedades en grupos es relevante en muchos contextos, como la provisión de bienes públicos y la formación de alianzas, cárteles y federaciones¹². Las innumerables formas en que las estructuras de red afectan el bienestar, hace que sea crítico comprender cómo impactan en el comportamiento y qué estructuras es probable que emerjan en una sociedad (Jackson, 2008).

Entre sus usos mas frecuentes se encuentran investigaciones sobre: mecanismos de transmisión de enfermedades, como las personas eligen su comportamiento, y cómo funcionan los mercados; también suele encontrarse en trabajos asociados al llamado “efecto pares”¹³, entre otros.

II. Concepto y tipos de redes¹⁴

Las redes de relaciones varían en forma y tamaño y están formadas por dos componentes: nodos y lazos. Si se piensa por ejemplo en un grupo de familias, cada nodo representa a una familia y los lazos son las relaciones por las cuales están unidas, vea la Figura 5 en Anexos. Formalmente, una red (N, g) consiste en un set de nodos $N = \{1, \dots, n\}$ y una matriz de números naturales de

⁹ Dutta y Jackson (2003)

¹⁰ Boorman (1975) y Montgomery (1991)

¹¹ Tesfatsion (1997, 1998)

¹² Tiebout (1956) y Guesnerie y Oddou (1981)

¹³ Refiere a las externalidades en las que las acciones o características de un grupo de referencia afectan el comportamiento de un individuo.

¹⁴ El contenido de este apartado está basado en conceptos de Jackson (2008).

tamaño $n \times n$, a la cual se hará referencia con la letra g . Donde g_{ij} representa la relación entre el nodo i y el j . La matriz g es llamada matriz de adyacencia.

Es importante destacar que hay dos tipos de red: redes no dirigidas, en las cuales los nodos están conectados o no, y redes dirigidas, en las que un nodo puede estar conectado a otro sin que este último esté conectado al primero. En el caso de que la red sea no dirigida, g_{ij} será igual a g_{ji} , porque no existe un sentido en la relación. Por el contrario, en redes dirigidas, g_{ij} no tiene que ser igual a g_{ji} .

Otro criterio a tener en cuenta es la ponderación: las redes son ponderadas si consideran la intensidad en que se relacionan los agentes que la componen, en este caso g_{ij} y g_{ji} pueden tomar valores entre 0¹⁵ y cualquier otro número natural, cuanto mayor es el número, mayor intensidad en la relación. Por otro lado, una red es no ponderada cuando no se tiene en cuenta la intensidad de la relación, por lo tanto, los nodos están o no conectados, en este caso g_{ij} y g_{ji} pueden tomar valores entre 0 y 1 (si existe relación). La siguiente figura muestra los conceptos anteriores¹⁶.

Figura 1. Tipos de redes, forma matricial

$$\begin{aligned}
 a) \ g_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} & b) \ g_2 &= \begin{pmatrix} 0 & 5 & 0 \\ 5 & 0 & 5 \\ 0 & 5 & 0 \end{pmatrix} \\
 c) \ g_3 &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} & d) \ g_4 &= \begin{pmatrix} 0 & 3 & 0 \\ 5 & 0 & 2 \\ 0 & 4 & 0 \end{pmatrix}
 \end{aligned}$$

Elaboración propia en base a conceptos de Jackson, 2008. **a)** Red no dirigida y no ponderada. **b)** Red no dirigida y ponderada. **c)** Red dirigida y no ponderada. **d)** Red dirigida y ponderada. Las cuatro redes están compuestas por tres nodos.

¹⁵ Si no existe relación entre los nodos.

¹⁶ Por una representación gráfica de redes ponderadas, ya sean dirigidas o no, vea la figura 6 en anexos.

III. Concepto de red aplicado al fútbol

Dos equipos en un partido representan un sistema complejo cuyo comportamiento depende de la dinámica de interacción entre los jugadores del equipo (Pappalardo et al., 2019). Mediante el análisis de redes sociales, un partido suele representarse por dos redes, cada equipo representa una de ellas, cada jugador es un nodo y los pases de balón entre jugadores representan los lazos de la red. Cada una de estas redes es ponderada - la ponderación está dada por el número de pases - y dirigida, dado que un jugador le puede pasar el balón a otro, pero la relación no es necesariamente recíproca.

IV. Modelos de formación de redes y literatura relacionada

Según Jackson (2005), los modelos de formación de redes provienen principalmente de dos fuentes: literatura de grafos aleatorios y literatura económica¹⁷. La literatura basada en grafos aleatorios construye redes mediante un proceso estocástico en el que los vínculos aparecen al azar siguiendo alguna distribución, o mediante algún algoritmo de construcción de enlaces. El enfoque económico, por el contrario, ha tendido a centrarse en redes de equilibrio, donde los vínculos se forman a discreción de los agentes interesados que son o controlan los nodos.

El presente texto hará hincapié en las redes formadas siguiendo modelos económicos, no porque se crea que en la sociedad no surgan redes según la otra teoría expuesta, si no porque para el caso de análisis, las redes de relaciones no se forman aleatoriamente: los jugadores son elegidos por el manager siguiendo algún criterio.

Siguiendo con lo expuesto por Jackson (2005), el enfoque económico para modelar la formación de redes presenta dos aspectos clave: (i) los agentes obtienen alguna utilidad de la red y, por lo tanto, existe un bienestar social general, y (ii) los enlaces se forman a discreción de los agentes

¹⁷ Basada en herramientas de teoría de juegos.

que son (o controlan) los nodos. De los dos puntos anteriores se desprende que las redes resultantes pueden predecirse mediante nociones de equilibrio.

Los primeros modelos que hablaron explícitamente sobre utilidades en la formación de vínculos sociales, fueron en el contexto de modelar las compensaciones entre vínculos fuertes y débiles en las redes de contactos laborales¹⁸. Mientras que Myerson (1977) llega a que las únicas coaliciones en juegos cooperativos que podrían producir valor son aquellas que están conectadas directamente dentro de la red. Aumann y Myerson (1988) reconocieron que diferentes estructuras de red conducían a diferentes asignaciones de valor a los agentes en la sociedad. Por otra parte, Jackson y Wolinsky (1996) postularon que las redes eran la estructura primitiva y los agentes derivaban las utilidades basadas en la estructura de la red existente, para probarlo adoptaron un enfoque de teoría de juegos y modelaron la formación de redes modelando la formación de enlaces (a través de las decisiones de los jugadores maximizadores interesados).

El enfoque económico también ha sido utilizado para modelar redes entre empresas desde una perspectiva de la organización industrial, por ejemplo, Goyal y Joshi (2003) encontraron que cuando dos empresas forman un vínculo se reducen sus respectivos costos de producción.

Por último, existen estudios de redes en juegos con complementariedad en las acciones de los vecinos, como en el contexto de los juegos de coordinación¹⁹ así como otras estructuras con complementariedad²⁰. Estos estudios han analizado tanto el cambio en el juego como la coevolución de la propia red.

¹⁸ Boorman (1975) y Montgomery (1991)

¹⁹ Ellison (1993), Morris (2000), Jackson y Watts (2002) y Feri (2003)

²⁰ Galeotti y Vega-Redondo (2005)

V. Medidas de centralidad²¹

No solo es importante comprender los posibles procesos de formación de redes, si no también que se puede decir de ellas una vez se han formado, para ello existen diversos indicadores o medidas. Existen medidas globales a la red, “macromedidas”, que miden ciertos aspectos de la red en general, como por ejemplo la densidad²², la distancia entre nodos²³, el “clique”²⁴, entre otros. También existen medidas “micro”, llamadas medidas de centralidad, estas permiten comparar nodos y decir algo sobre cómo un nodo en particular se relaciona con la red en general (Jackson, 2008).

Dado que en el presente trabajo se desea realizar un estudio sobre los nodos de la red, se hará foco en las medidas micro.

Las medidas de centralidad capturan diferentes aspectos de la red y de los nodos que la forman, las más relevantes según Jackson (2008) son las centralidades *between* y *degree*, dado lo anterior, el presente trabajo hará foco en ellas.

i. Centralidad In y Out Degree

El *Degree* de un nodo es el número de enlaces que lo conectan con otros nodos. En redes dirigidas, se divide en “*In Degree*” y “*Out Degree*”, siendo el primero el número de nodos que tienen una relación con el nodo en cuestión, pero no necesariamente éste está relacionado con ellos (por ejemplo, en el caso de redes de fútbol, contabiliza los pases que el jugador x recibe

²¹ El contenido de este apartado está basado en conceptos presentes en el capítulo 2 de Jackson (2008) y en el capítulo 4 de Clemente y Mendes (2016)

²² La densidad de una red (dirigida) es la proporción del máximo de enlaces posibles presentes entre nodos.

²³ Si las distancias son grandes, puede sugerir que la “información” circula lentamente entre nodos. Los nodos que están más cerca de otros pueden ejercer más “poder” que los que están más distantes.

²⁴ Puede usarse para identificar la frecuencia de patrones específicos de interacciones entre nodos.

de y , sin importar si y recibió pases de x) y el segundo, lo opuesto. En el caso de redes no dirigidas, *In Degree* y *Out Degree* son iguales.

Si se trata de una red no ponderada, el *degree* toma valores entre 0 y $n - 1$, siendo n la cantidad de nodos que componen la red. Un nodo con *degree* $n - 1$ estará conectado directamente a todos los otros nodos de la red, en cambio un nodo con *degree* igual a cero, estará aislado. Si la red es dirigida, el *degree* de un nodo, es la sumatoria de la intensidad de la relación entre el nodo en cuestión y los demás.

Si n_i es un vértice de una red ponderada g con n vértices, las centralidades *out degree* e *in degree* normalizadas son la proporción del peso de los vértices adyacentes a n_i , y se pueden calcular de la siguiente forma (Opsahl, Agneessens y Skvoretz, 2010):

Centralidad *In degree* normalizada²⁵ del nodo i :

$$IDC_{(D-in)}^w(n_i) = \frac{k_i^{w-in}}{\sum_{i=1}^n \sum_{j=1, j \neq i}^n a_{ji}}$$

Con $k_i^{w-in} = \sum_{j=1}^n a_{ji}$, siendo k_i^{w-in} el *in degree* del vértice n_i y a_{ji} elementos de la matriz de adyacencia ponderada de g .

Centralidad *Out degree* normalizada del nodo i :

$$ODC_{(D-out)}^w(n_i) = \frac{k_i^{w-out}}{\sum_{i=1}^n \sum_{j=1, j \neq i}^n a_{ij}}$$

Con $k_i^{w-out} = \sum_{j=1}^n a_{ij}$, siendo k_i^{w-out} el *out degree* del vértice n_i y a_{ij} elementos de la matriz de adyacencia ponderada de g . Si la red es no dirigida k_i^{w-out} y k_i^{w-in} son iguales, por

²⁵ En las siguientes secciones, cuando se mencione "*In degree*" u "*Out degree*" se estará haciendo referencia a la medida de centralidad normalizada.

lo tanto, las centralidades *in degree* y *out degree* también lo son. Las centralidades *degree* normalizadas toman valores entre 0 y 1.

Aplicada al fútbol, la centralidad *in degree* (IDC) suele ser utilizada como un indicador del “prestigio” de cada jugador entre sus compañeros de equipo. Los jugadores con IDC altos son aquellos que reciben muchos pases. Estos pases pueden interpretarse como “elecciones” a un jugador específico de parte de otros. Por lo tanto, un mayor valor de este tipo de centralidad, indica que el jugador es más prestigioso o importante entre sus compañeros de equipo (Clemente et al., 2016). El análisis de los datos en secuencias de pases ha mostrado que los jugadores con resultados de IDC más altos son aquellos a quienes sus compañeros de equipo preferían pasar el balón con más frecuencia. Por lo tanto, se puede pensar que estos jugadores son cruciales para el desarrollo ofensivo del juego, porque reciben el balón con más frecuencia que otros jugadores durante los periodos de ataque de su equipo (Clemente et al., 2016).

Por otra parte, la centralidad *out degree* (ODC) puede interpretarse como una medida de la actividad de cada jugador (Clemente et al., 2015). Puede interpretarse que los jugadores con valores de ODC más elevados son aquellos que contribuyeron más a los intentos ofensivos de su equipo (a través del envío de pases a los demás jugadores) (Clemente et al., 2016).

ii. Centralidad between

Es una medida de centralidad basada en que tan bien situado está un nodo en términos de “camino” entre los demás. Se obtiene una idea de cuán importante es un nodo “*k*” en términos de conectar los nodos “*i*” y “*j*” observando el ratio $\frac{g_{ij}(n_k)}{g_{ij}}$.

Cuanto mayor es el ratio, mayor será la importancia de “*k*” en la conexión. $g_{ij}(n_k)$ denota el número de “camino más cortos” entre el nodo “*i*” y el “*j*” en los cuales se encuentra “*k*”. Por otro lado, g_{ij} es el número de caminos más cortos entre n_i y n_j (Jackson, 2008).

Dada una red ponderada g que cuenta con n nodos (con $n_i, n_j, n_k \in n$) la centralidad *between* normalizada²⁶ del nodo n_k se puede calcular de la siguiente forma (Rubinov y Sporns, 2010):

$$C_b(n_k) = \frac{1}{(n-1)(n-2)} \sum_{n_i, n_j \in V, i \neq k, j \neq k} \frac{g_{ij}(n_k)}{g_{ij}}$$

En redes ponderadas, el *between* se normaliza cuando se aplica el término $\frac{1}{(n-1)(n-2)}$, éste hace referencia al número de caminos en los que un nodo puede encontrarse entre otros dos nodos.

La centralidad *between* (BC), aplicada al fútbol, a menudo se considera la medida más significativa (entre los demás índices de centralidad) porque cuantifica con éxito la frecuencia con la que cada jugador se encuentra entre otros jugadores de la red, quizás actuando como un mediador o puente para ellos. El BC de cada jugador puede explicarse como una medida del control relativo que este tiene sobre los demás. En análisis de redes sociales, se supone comúnmente que los jugadores con *between* más alto, tienen una mayor probabilidad de ejercer control sobre el flujo de información entre otros nodos en la misma red (Clemente et al., 2016). Los valores más altos de BC sugieren que ese nodo puede actuar como un jugador clave para vincular los sectores de un equipo y a sus compañeros de equipo (Clemente y Martins, 2017).

²⁶ En las siguientes secciones, cuando se mencione “*Between*” se estará haciendo referencia a la medida de centralidad normalizada.

VI. Función de producción: conceptos y vínculos con el deporte

Si bien el tema central del presente trabajo es el análisis de redes sociales, para conseguir los objetivos propuestos es importante dejar en claro conceptos relacionados a la función que relaciona insumos y outputs, la función de producción, y su vínculo con el deporte.

La función de producción es un enunciado matemático que relaciona cuantitativamente el *output* de un proceso y los insumos de los factores de producción. Los distintos tipos de bienes que se pueden utilizar junto a la tecnología de producción se denominan factores de producción y, para cualquier conjunto de insumos de estos factores, se interpreta que la función de producción define la producción máxima realizable a partir de ellos (Shephard, 1970). Estos conceptos han sido relacionados al deporte en múltiples ocasiones, y, en las más recientes, se han vinculado con análisis de redes sociales.

En el presente texto, se considerará, al igual que en Espitia-Escuer y Garcia-Cebrian (2016) que el producto de un equipo de fútbol es el resultado deportivo que consigue en cada partido. Los equipos no compiten entre sí por lo que sería el equivalente de un mercado, sino por el resultado deportivo, la victoria en el juego o el torneo general. Si los juegos son la actividad productiva de los equipos de fútbol y los resultados deportivos son el *output*, según Espitia-Escuer y Garcia-Cebrian (2016) el análisis de su productividad debería centrarse exclusivamente en las jugadas realizadas y las relaciones y vínculos entre jugadores dentro del campo de juego, aquí es donde entra en juego la importancia del análisis de redes aplicado a deportes de equipo.

i. Antecedentes

Según Weimar (2019), desde el primer análisis de la economía del deporte realizado por Rottenberg (1956), las investigaciones económicas del deporte son de creciente interés. Weimar (2019) encuentra como principales líneas de investigación la productividad del equipo y la productividad individual. La corriente principal, denominada "Productividad del equipo", incluye la investigación sobre los resultados, la eficiencia de los equipos de fútbol y los determinantes del éxito del equipo. En la categoría "productividad individual" se utilizan (la mayoría de las

veces) los salarios como proxy de la productividad marginal, un ejemplo de esta clase de investigaciones son Blass (1992) y Mazur (1994).

Schofield (1988), Carmichael y Thomas (1995) y Carmichael et al. (2000) proponen que, para los equipos deportivos, el proceso para crear los goles se compone de dos fases: la primera es el entrenamiento responsable del acondicionamiento físico de los jugadores y la segunda involucra los propios juegos, donde los equipos se enfrentan a sus rivales y se produce el producto final.

Por otra parte, no existe consenso en la literatura sobre si el objetivo que persiguen los equipos deportivos es la maximización del rendimiento económico o la maximización de los resultados deportivos, como se puede ver en los estudios de Fort y Quirk (1995), Szymanski (2003), Sandy et al. (2004), Fort (2006), Késenne (2007), García del Barrio y Szymanski (2009).

La idea de la relación entre el uso de recursos productivos y la competitividad en los equipos de fútbol también está presente en Pestana Barros et al. (2014). Otras investigaciones, como las de Espitia-Escuer y Garcia-Cebrian (2016, 2019), analizan los resultados de la UEFA Champions League bajo la óptica de la productividad de los equipos que en ella compiten, para medir la productividad utilizan el índice de Malmquist - Luenberger.

Por último, Horrace, Liu y Patacchini (2016), construyen una función de producción de red para el baloncesto universitario. Estos consideran al equipo como una firma, en la cual las productividades individuales de los basquetbolistas están interrelacionadas de formas complejas a través de redes. Dichas redes son establecidas por un solo entrenador que asigna los trabajadores a los equipos en función de las características de estos. Los autores afirman que comprender los efectos de la red en la producción (marcador/ productividad) puede ser importante para la programación de los trabajadores y el diseño de los esquemas de incentivos para los mismos.

Caso de estudio: análisis de redes sociales aplicado a equipos que participaron de la UEFA Champions League

Para el análisis se considerará a cada equipo de fútbol como una firma; ésta firma cuenta con un plantel de trabajadores (jugadores) que es dirigido por el director técnico (mánager). La actividad económica de cada institución es la disputa de partidos/torneos, para el desarrollo de la misma, el entrenador elige los recursos a utilizar siguiendo un criterio de optimización, luego, los jugadores interactúan entre si formando una red, surgiendo de esta los *outputs* (goles) de la actividad.

En el próximo apartado se reflexionará si la formación de las redes de futbol sigue alguno de los modelos vistos previamente.

Luego, se procederá a testear la primer hipótesis. En este punto se intentará replicar, considerando una mayor cantidad de observaciones, el análisis de Clemente y Martins (2017). Los autores analizaron los 16 mejores equipos de la temporada 2015/2016 de la UEFA Champions League. El estudio reveló que los mediocampistas centrales fueron los jugadores más destacados durante la acumulación de jugadas de ataque y que los defensas centrales fueron más prominentes en el juego de construcción de ataque que los defensas laterales o los mediocampistas laterales. Porteros y delanteros fueron los puestos con los valores de centralidad más pequeños del equipo. Para realizar el análisis consideraron tres medidas de centralidad: *between*, *in degree* y *out degree*.

Por último, se buscará estimar una relación entre las medidas de centralidad y los resultados para cada equipo. La hipótesis es que cuanto más conectada esta la red, se generan nuevos "caminos" por donde circule el balón y eso se traduciría en una mayor cantidad de goles anotados. La relación será estimada mediante un modelo que relaciona la cantidad de goles macados con algunas medidas de centralidad y variables del contexto partido/equipo.

Se espera que las variables impacten de la siguiente forma: la centralidad *between* de todas las posiciones impactaría positivamente²⁷. La centralidad *degree* tendría un impacto en ambos sentidos, dependiendo de la posición en estudio: un aumento de la centralidad de los delanteros debería hacer aumentar la cantidad de goles anotados debido a la posición de estos en el campo de juego (cuanto más reciban la pelota quienes están cerca del área rival, o cuantos más caminos se “abran” hacia ellos, más probabilidad de que se marque gol por un tema de localización), en tanto para las posiciones que se ubican cerca del área propia ocurre lo opuesto, por ello impactaría negativamente. La forma en que impactarían las variables exógenas a la red es bastante intuitiva: enfrentar a un equipo grande impactaría de forma negativa y las demás variables consideradas lo harían positivamente.

I. Formación de redes de fútbol

La formación de redes de fútbol respondería a la literatura enfocada en teoría de juegos, dado que las mismas no se forman de manera aleatoria siguiendo algún tipo de distribución preestablecida, si no que la elección de los jugadores responde a las órdenes del entrenador.

Cada entrenador selecciona el once inicial y en el correr del partido puede escoger, o no, ingresar nuevos jugadores. Se asume que la selección se realiza en función de las características de los partidos a disputar, del tipo de competición, números de competiciones simultaneas disputadas, rival y también en función de las características individuales de los jugadores a disposición. A su vez, la elección o no de un jugador, no se daría solamente en función a individualidades, sino a cómo responden los otros nodos a su incorporación y a como interaccionan entre si quienes son seleccionados. Se supone que, con esto, el entrenador intenta realizar una elección de equilibrio (dada la información que tiene a disposición) en pos de anotar más goles que el rival.

II. Datos y Metodología: procesamiento de datos

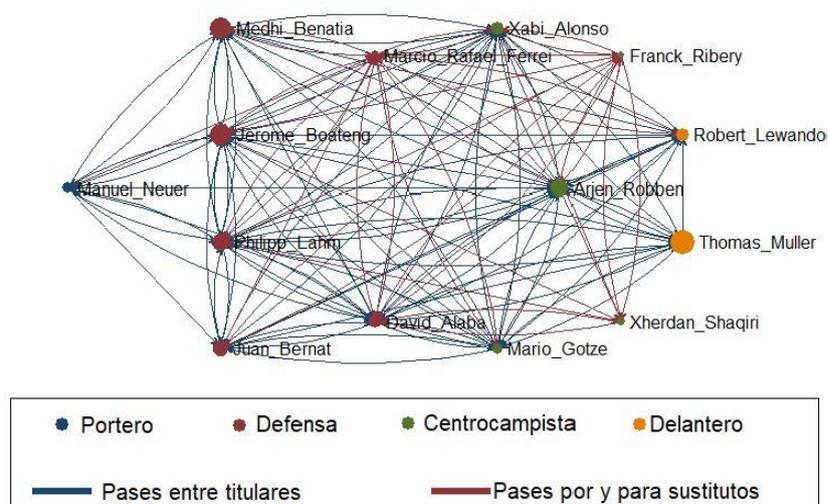
Un total de 873 partidos de UEFA Champions League, desde la temporada 2012/2013 a 2018/2019, fueron analizados y codificados en éste estudio. Se tomaron en cuenta todas las etapas de cada competición, comenzando por la fase de grupos. Los datos fueron obtenidos de

²⁷ Porque da una medida más certera de cuan conectada está la red, Jackson (2008).

la página web de la UEFA²⁸, estos fueron utilizados para computar cada matriz de adyacencia. El tamaño de cada matriz está entre 11x11 y 14x14, dependiendo de cuántos sustitutos se hayan utilizado en el juego.

Con el paquete “NetworkX”²⁹ de Python se crearon un total de 1.746 redes, cada una correspondiente a un equipo en un partido. Dado que la información recolectada permitió conocer no solamente que jugadores tuvieron contacto entre sí, sino también de quien partió el pase y la intensidad de la relación (medida por el número de pases enviados), las redes creadas fueron dirigidas y ponderadas. La siguiente figura, elaborada en base a una de las redes computadas³⁰, muestra como se ve una red de estas características.

Figura 2. Red de pases del Bayern Múnich (vs Roma, 2014)



Fuente: Elaboración propia en base a datos de UEFA. *Red de pases del equipo Bayern Munich en un partido frente a Roma (21/10/2014, fecha 3 de la fase de grupos. Resultado: Roma(1), Bayern Munich(7)). Nota: El tamaño del círculo depende del grado de centralidad between del jugador. La interacción entre jugadores está representada mediante flechas que van desde quien envía el pase hasta quien lo recibe.*

A cada nodo de cada red se le calculó la centralidad³¹, luego, se procesó la información de forma tal, que todos los indicadores estuvieran presentes en una única base de datos. La base de datos

²⁸ <https://www.uefa.com/insideuefa/mediaservices/presskits/uefachampionsleague/index.html>

²⁹ <https://networkx.github.io/>

³⁰ Elegida aleatoriamente para presentarse en el presente trabajo.

³¹ Se calculó la centralidad *between*, *in degree* y *out degree*.

extendida contó con un total de setenta y cinco equipos³² y más de tres mil jugadores. A cada equipo se le asignó una identificación única, lo mismo se realizó para los jugadores, de esa forma se pudo, además de realizar estadísticas por red, realizarlas por equipo y por jugador (posición).

Para obtener datos más precisos, y siguiendo la metodología usada por Linder (2017), se eliminaron los jugadores que no intentaron lanzar pases, es decir que no tocaron el balón, o que intentaran menos de 5 pases, ya sean del once inicial o sustitutos. Esto hace que los datos sean más confiables, ya que elimina a los jugadores que están haciendo pequeñas contribuciones al juego, y que podrían entorpecer los resultados. Por ejemplo, un jugador que es expulsado con tarjeta roja antes de lograr completar muchos pases.

Por otra parte, para realizar el modelo que trata estimar una relación entre los insumos y los resultados para cada equipo se modificó la base de datos antes mencionada. En la nueva base de datos, cada fila corresponde a un equipo en un partido (para cada uno se tomó en cuenta el promedio de centralidad por posición).

III. Datos y Metodología: posiciones en el campo de juego

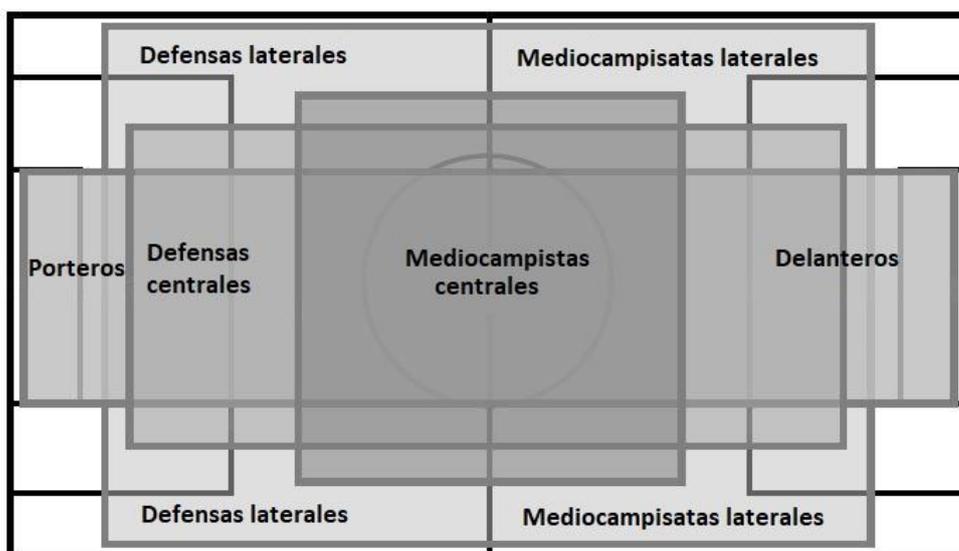
Se utilizó la siguiente clasificación de posiciones:

- Porteros (PO)
- Defensas centrales (DC)
- Defensas laterales (DL)
- Mediocampistas centrales (MC)
- Mediocampistas laterales (ML)
- Delanteros (DE)

La próxima figura muestra el espacio que ocupa cada jugador en el campo de juego.

³² Los equipos analizados pueden verse en la Tabla 2 de anexos y el porcentaje de competiciones de las que participaron en la Figura 7 de anexos.

Figura 3. Asignación táctica de las posiciones.



Fuente: Adaptado de Clemente et al. 2015.

La posición de cada uno de los jugadores participantes del estudio fue recolectada de la página web TransferMarkt³³. Para conocer el número de observaciones de cada tipo presentes en los datos y su porcentaje de participación en el total vea la Tabla 4 en anexos.

IV. Datos y Metodología: procedimiento estadístico y validez de los resultados

i. Jugador clave

Para encontrar al jugador clave fue necesario evaluar las centralidades por posición, y luego compararlas entre sí. Se calculó la media³⁴ de cada tipo de centralidad por posición, a cada una de éstas se le realizó un test de medias³⁵. La idea inicial fue aplicar un test ANOVA de una vía³⁶,

³³ <https://www.transfermarkt.es/>

³⁴ Puede visualizar el diagrama de Box-Whisker por posición por medida de centralidad en las figuras 8, 9 y 10 en anexos.

³⁵ Test estadístico a emplear cuando se desea comparar las medias de dos o más grupos.

³⁶ Es el tipo de análisis que se emplea cuando los datos no están pareados y se quiere estudiar si existen diferencias significativas entre las medias de una variable aleatoria continua en los diferentes niveles de otra variable cualitativa o factor.

pero al comprobar que los datos estudiados no siguen una distribución normal³⁷ y no son homocedásticos³⁸, se realizó un test ANOVA con corrección de Welch. Intuitivamente, es igual al ANOVA, con la salvedad de que es utilizado cuando se puede asumir que las varianzas poblacionales son diferentes³⁹, por lo cual, la prueba se realiza grupo a grupo y no como un todo (Moder, 2010; Vogt y Johnson, 2005).

El test de Welch resultó ser significativo para todos los grupos comparados⁴⁰, lo cual implica que estos son significativamente diferentes entre sí. Se calculó el tamaño del efecto⁴¹, el cual indica la magnitud o fuerza del efecto detectado⁴². El tamaño del efecto fue calculado siguiendo el criterio de Cohen⁴³, sus resultados se interpretan de la siguiente forma: <0.2 (efecto trivial); 0.2 – 0.6 (efecto pequeño); 0.6 – 1.2 (efecto moderado); 1.2 – 2.0 (efecto grande) y >2.0 (efecto muy grande).

Para ver los resultados del test de Welch y del test de Cohen, vea las tablas 9, 10 y 11 en anexos.

ii. Estimación de la relación entre medidas de centralidad y goles marcados

En este punto, se considerará que cada equipo utiliza los vínculos entre los futbolistas (medidos a través de las medidas de centralidad) para formar goles. Además, que existen otros factores que influirían en el número de goles marcados, estos factores no son controlables por el

³⁷ Por los resultados del test de Shapiro–Wilk, utilizado para testear si las variables siguen una distribución normal, diríjase a la tabla 5 en anexos.

³⁸ Refiere a que los datos no presentan homogeneidad de varianzas. Para comprobarlo, se realizó el test de Levene, para ver sus resultados diríjase a las tablas 6, 7 y 8 en anexos.

³⁹ El test es bastante robusto aun cuando existe cierta falta de normalidad.

⁴⁰ En las tres variables medidas: centralidad *between*, *in degree* y *out degree*.

⁴¹ El tamaño del efecto es una medida cuantitativa de la magnitud de un fenómeno. Es un componente utilizado para evaluar la fuerza de un estadístico (Kelley & Preacher, 2012). En este caso, se utilizó para evaluar la magnitud de las diferencias entre grupos, dado que el p-valor indica la dirección del efecto, pero sólo la estimación del tamaño del efecto indica cuán grande es.

⁴² Cumming, 2013

⁴³ $Cohen's\ d = \frac{(media\ del\ grupo\ 1 - media\ del\ grupo\ 2)}{SD_{pooled}}$, $SD_{pooled} = \sqrt{\frac{SD_1^2 + SD_2^2}{2}}$

manager o por la institución y estarían representados por si el equipo es local o no, el tipo de rival y tipo de equipo, entre otros.

Para estimar el vínculo entre las medidas de red y los goles anotados, en una primera instancia se tomaron en cuenta todas las medidas de centralidad por posición, resultando ser significativas las centralidades *between* de todos los jugadores, a excepción del arquero, y el *In degree* de defensas centrales y de posiciones extremas, tales como delanteros y porteros, lo cual tiene cierta lógica debido a la posición de éstos jugadores en el campo de juego.

La relación, estimada mediante mínimos cuadrados ordinarios, es la siguiente:

$$Goles\ marcados = \sum_{i=1}^5 \beta_i Bt_i + \beta_6 Id_{DE} + \beta_7 Id_{DC} + \beta_8 Id_{PO} + \beta_9 Local + \beta_{10} Equipo\ grande + \beta_{11} Rival\ grande + \beta_{12} Pases\ intentados + \varepsilon$$

En dónde *Bt* significa centralidad *between*, y los índices de la sumatoria son indicadores de la posición (1 = DC, 2 = DL, 3 = MC, 4 = ML, 5 = DE), *Id* significa centralidad *In degree*. *Local* es una dummy, es 0 si el equipo es visitante y 1 si es local. *Equipo grande* y *rival grande* son dummies, toman el valor 1 si equipo o rival (respectivamente) son considerados equipos “grandes”, y 0 en el caso contrario. La clasificación de equipos (“grandes” o “chicos”) es de elaboración propia, para la misma se tomó en cuenta el coeficiente UEFA 19/20 (elaborado en base al rendimiento de cada equipo en los últimos cinco años y en el rendimiento de la federación a la cual pertenecen los equipos en los últimos cinco años) y el número de trofeos europeos ganados por cada equipo. Ambos valores son sumados, y los equipos pertenecientes al percentil 15 fueron considerados equipos “grandes”. *Pases intentados* es una variable discreta que indica el número de pases intentados por el equipo. Por último, ε , es un término de error estadístico.

Se comprobó, mediante una tabla de correlaciones, que el residuo no estuviera correlacionado con algunas de las variables independientes, para descartar una endogeneidad del tipo “error de medición”, para ver los resultados diríjase a la tabla 12 en anexos.

La endogeneidad plantea un problema fundamental para la estimación de parámetros, si no se tiene en cuenta, los estimadores podrían ser sesgados e inconsistentes. Por lo tanto, se deben revisar todas sus formas, ya se ha revisado una de ellas, pero aún queda sobre la mesa la posibilidad de endogeneidad del tipo “variables omitidas” y del tipo “simultaneidad”.

La endogeneidad por “variables omitidas” es cuestionable, dado a que aún no existe consenso en la comunidad científica acerca de si es una forma de endogeneidad o no, por ejemplo, en Hansen (2000) se consideran los otros dos casos, pero no el de variables omitidas. El problema, en caso de ser un tipo de endogeneidad, podría darse al no incorporar una variable relevante para el modelo, por lo cual, la esperanza del error (dadas las variables incorporadas) sería distinta de cero, contradiciendo el supuesto de que los inobservables no tienen efecto ni poder explicativo. Según Martínez (2016), a efectos prácticos, el modelo debe incluir todas las covariables “z” que correlacionen con las variables independientes incluidas en el modelo y que sean relevantes. Como consecuencia de la última afirmación, se desprende que las variables que no covarían con las independientes pueden omitirse en el modelo sin afectar los coeficientes estimados, lo cual implica que, si se omiten variables relevantes para explicar la variable regresada pero que no covarían con las independientes, se seguiría obteniendo una estimación válida, aunque el R2 se vería afectado.

El presente modelo puede ser explicado por una mayor cantidad de variables, que podrían estar relacionadas con factores psicológicos de los jugadores presentes en el campo, esfuerzo, horas de entrenamiento personal, o que tengan que ver con características del equipo rival, entre otras. Pero son factores no observados y que difícilmente covaríen con las variables que explican el modelo, por tal razón, siguiendo lo especificado por Martínez (2016), se concluye que la estimación no presenta problemas de endogeneidad del tipo variables omitidas.

Para cerrar con endogeneidad, falta considerar la de tipo “simultaneidad”, este caso se da cuando no solo las variables independientes afectan a la dependiente, sino que también pasa lo opuesto, como es el caso de la siguiente ecuación:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Dónde $x_i = f(y_i) = g(u_i)$, siendo f y g funciones lineales y en consecuencia:

$$Cov [x_i, u_i] \neq 0.$$

Básicamente, podría darse cuando las variables ocurren exactamente al mismo tiempo, es decir cuando la y y la/s x pasan a la vez, si la variable dependiente ocurriera luego de que ocurren las independientes (aunque sea unos segundos después), no existiría problema de simultaneidad. El lector podría pensar que el modelo estimado no es válido si los goles anotados por un equipo pasaran al mismo momento en que ocurre cualquiera de las variables independientes. A continuación, se interpretará la causalidad y/o *timing* de las variables consideradas.

Antes de estudiar el impacto de los goles anotados sobre las variables de red, se puede afirmar que variables tales como localía, tipo de equipo y rival, son exógenas a la red y ocurren antes de comenzar el partido, por lo cual, no es posible que los goles anotados ocurran antes que ellas.

Un partido de fútbol, puede pensarse como un juego en dos etapas, en la primera se forma la red de pases y en la segunda se obtiene un resultado (favorable o no). No sería lógico pensar que el primer gol del partido ocurre antes que los pases, porque sería imposible que la pelota llegara al arco rival. Una vez que el balón lleva en juego algún tiempo, y se marca el primer gol, independientemente si es en los minutos iniciales o más adelante, la red de pases (y por ende las centralidades de los futbolistas) ya está formada, pero lógicamente continúa creciendo hasta el final del juego.

Marcar goles podría afectar psicológicamente al equipo, o al rival, pero son variables que no han sido consideradas, y que no afectan a la posibilidad de pases entre jugadores. Por otra parte, puede observarse en algunos partidos que anotar “baja el ánimo” del rival o “sube” el del equipo que marcó, por ende, el primero podría retraerse y/o el segundo ganar más confianza, lo cual llevaría potencialmente a un aumento en la posesión del balón⁴⁴. Desde este punto, quedan claras dos posibilidades, dependiendo de cómo se plantea el juego, las características de los equipos y el contexto en el que se desarrolla el partido:

⁴⁴ Variable no considerada en el modelo.

- (1) anotar goles podría no influir en la posesión del balón post gol, por lo cual las centralidades no se verían afectadas por la variable dependiente.

- (2) marcar goles podría hacer que la posesión del balón crezca, por lo cual el número de pases también lo haría, dentro de esta última se presentan dos posibilidades:
 - (a) que el aumento en la posesión del balón sea global y constante entre todos los futbolistas presentes en el campo, por lo tanto, por más que el número de pases del equipo crezca, si lo hace de este modo, las centralidades de cada nodo se mantendrían invariantes en el correr del partido. Por ende, las centralidades incluidas como variables independientes en el modelo afectarían al número de goles anotados, pero la relación inversa no se daría.

 - (b) que el aumento en la posesión no sea constante entre todas las posiciones, por ejemplo, que aumente del medio campo hacia adelante porque el equipo ataca más. En éste caso se estaría en presencia de un problema de endogeneidad, porque para los equipos en los que se da esta relación, el número de goles marcados afectaría a su vez a las variables de red.

Dicho lo anterior, no se puede asegurar o negar que exista un problema de endogeneidad en el modelo planteado, dado que fueron tomados en cuenta datos de partidos que se desarrollaron de diferente forma: en algunos, las centralidades se ven afectadas por los goles anotados y en otros no. Además, no solamente fueron consideradas las victorias, si no también empates y derrotas, resultados para los cuales no está claro cuál es el impacto de un gol propio, dado que es muy probable, sobre todo en empates, que el transcurso de los pases sea similar a lo largo del partido. Por lo anterior, es difícil estar seguros que tan endógeno o no es el modelo: dependerá del número de partidos en que la centralidad post gol cambia, pero dadas las características de los datos con los que se cuenta, no se puede asegurar que tanto peso tienen las “observaciones problemáticas” sobre la estimación final.

Lo que sí se puede asegurar, por lo planteado en el párrafo anterior, es que podría existir un potencial problema de simultaneidad con la variable “pases intentados” (puntos a y b), para descartar que la inclusión de ésta variable afectara los resultados del modelo, el mismo fue estimado sin ella.

El estudio procederá considerando que el efecto y/o peso que tienen los equipos que variaron desigualmente las centralidades post gol es mínimo, pero, de todas formas, el lector debería tener cautela a la hora de interpretar los resultados.

La versión final del modelo, la cual no considera el total de pases intentados por partido, es la siguiente:

$$Goles\ marcados = \sum_{i=1}^5 \beta_i Bt_i + \beta_6 Id_{DE} + \beta_7 Id_{DC} + \beta_8 Id_{PO} + \\ \beta_9 Local + \beta_{10} Equipo\ grande + \beta_{11} Rival\ grande + \varepsilon$$

En conclusión, la estimación estaría libre de endogeneidad del tipo variables omitidas y error de medición, pero se debe ser cauto con la del tipo simultaneidad, porque podría aparecer para algunas de las observaciones consideradas.

Por otra parte, realizó el test de Breusch – Pagan y Cook – Weisberg para detectar si el modelo es o no homocedástico, resultando significativo⁴⁵, por este motivo, los β_i fueron estimados de forma robusta. También, se verificó que los residuos se distribuyeran de forma normal (y de la misma forma se probó que los estimadores β_i también siguieran dicha distribución - dado que son funciones lineales de ε), utilizando un gráfico de probabilidad normal (P-P)⁴⁶, en el cual se puede apreciar que los puntos se aproximan al comportamiento lineal señalado, por lo tanto, los residuos se comportarían de forma normal⁴⁷. También, fue realizada una prueba de inflación

⁴⁵ La hipótesis nula es que el modelo es homocedástico. Por el resultado del test vea la tabla 13 en anexos.

⁴⁶ Método gráfico para mostrar si los residuos de una regresión siguen una distribución normal.

⁴⁷ Por el gráfico y una explicación más detallada del mismo, vea la figura 11 en anexos.

de varianza (*vif*), para la misma se recomiendan valores menores que 10 para mostrar que no hay multicolinealidad, lo cual se verifica⁴⁸.

Dejando de lado las pruebas que indican si los estimadores son válidos, se intentó verificar la robustez de los resultados, para ello, fueron estimadas dos variantes del modelo:

1. Sin considerar efecto fijo por localía:

$$\begin{aligned} \text{Goles marcados} = & \sum_{i=1}^5 \beta_i Bt_i + \beta_6 Id_{DE} + \beta_7 Id_{DC} + \beta_8 Id_{PO} + \\ & \beta_9 \text{Equipo grande} + \beta_{10} \text{Rival grande} + \varepsilon \end{aligned}$$

2. Sin considerar el efecto fijo por localía, tipo de equipo y tipo de rival:

$$\text{Goles marcados} = \sum_{i=1}^5 \beta_i Bt_i + \beta_6 Id_{DE} + \beta_7 Id_{DC} + \beta_8 Id_{PO} + \varepsilon$$

Los resultados serán estudiados en la sección “Resultados”.

⁴⁸ Vea tabla 14 en anexos.

V. Resultados

Los resultados obtenidos en lo que refiere a las posiciones más prominentes en el campo de juego, para las tres medidas de centralidad estudiadas, están presentes en la siguiente tabla.

Tabla 1. Estadísticos descriptivos (media \pm desvío e Intervalo de confianza al 95%) de las centralidades entre posiciones, p -valor del test de Welch y tamaño del efecto promedio de Cohen.

	PO	DC	DL	MC	ML	DE	p	ES*
ODC	0.512 \pm 0.091	0.616 \pm 0.163	0.663 \pm 0.194	0.685 \pm 0.205	0.540 \pm 0.214	0.460 \pm 0.213	0.00	0.4 <i>Ef. pequeño</i>
	[0.50 - 0.52]	[0.61 - 0.62]	[0.66 - 0.67]	[0.68 - 0.69]	[0.53 - 0.55]	[0.45 - 0.47]		
BC	0.044 \pm 0.026	0.073 \pm 0.061	0.079 \pm 0.063	0.089 \pm 0.082	0.061 \pm 0.046	0.048 \pm 0.030	0.00	0.7 <i>Ef. moderado</i>
	[0.42 - 0.45]	[0.72 - 0.75]	[0.77 - 0.81]	[0.87 - 0.91]	[0.59 - 0.62]	[0.47 - 0.49]		
IDC	0.388 \pm 0.124	0.629 \pm 0.179	0.664 \pm 0.183	0.691 \pm 0.201	0.604 \pm 0.211	0.564 \pm 0.227	0.00	0.6 <i>Ef. pequeño</i>
	[0.38 - 0.39]	[0.62 - 0.63]	[0.658 - 0.670]	[0.686 - 0.696]	[0.596 - 0.610]	[0.557 - 0.572]		

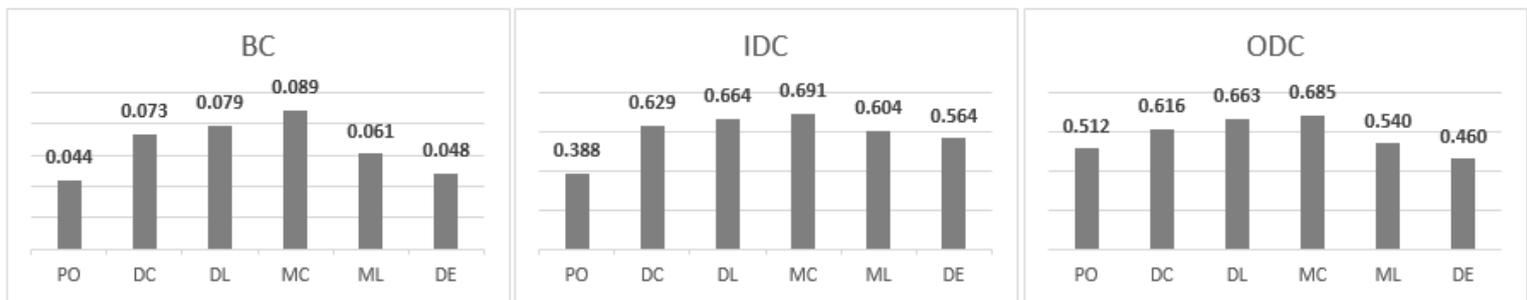
*Efecto promedio

Fuente: Elaboración propia en base a datos de UEFA. *Estadísticos descriptivos de las medidas de centralidad consideradas en el presente estudio, medidas a través de las distintas posiciones que adoptan los jugadores en el campo de juego.*

**Para ver los resultados del test de Welch y el correspondiente tamaño del efecto (estudiado mediante la prueba de Cohen) vea las tablas 9, 10 y 11 en anexos.*

Para visualizar gráficamente la información anterior, vea la próxima figura.

Figura 4. Comparación de las medias de centralidad por posiciones



Fuente: Elaboración propia en base a datos de UEFA. *Comparación de medias de las diferentes medidas de centralidad a través de las posiciones.*

Se puede constatar mediante un análisis individual de las tres medidas, que el mediocampista central es la posición que presenta mayor centralidad. Para encontrar ciertos patrones, se analizarán las medidas por separado:

En cuanto a la centralidad *between*, las posiciones con mayores valores son mediocampistas centrales (0.089 en promedio) y defensas laterales (0.079 en promedio). Los siguen los defensas centrales (0.073) y mediocampistas laterales (0.061). Con valores bastante por debajo, se encuentran los delanteros (0.048) y porteros (0.044).

Al observar la centralidad *in degree*, si se ordenan las posiciones de mayor a menor valor, siguen el mismo patrón que si se ordenan de forma descendente por BC: MC (0.69), DL (0.66), DC (0.63), ML (0.60), DE (0.56) y PO (0.39).

Con respecto a la centralidad *out degree*, los mediocentros continúan siendo la posición con mayor centralidad (0.69), en segundo lugar, se encuentran los defensas laterales (0.66), después, los defensas centrales (0.62), le siguen los mediocampistas laterales (0.54), porteros (0.51) y por último los delanteros (0.46).

Por otra parte, los resultados de la estimación de la relación entre goles anotados y algunas medidas de red están presentes en la siguiente tabla⁴⁹:

Tabla 2. Resultados de la estimación (robusta) de la función que relaciona goles anotados y medidas de centralidad

Variables independientes	β	t	p	
Id_{DF}	0.653	3.36	0.001	***
Bt_{DF}	4.325	4.91	0.000	***
Bt_{MC}	10.02	13.36	0.000	***
Bt_{ML}	2.951	3.76	0.000	***
Bt_{DL}	2.301	3.71	0.000	***
Bt_{DC}	2.329	2.88	0.004	***
Id_{DC}	-0.839	-3.93	0.000	***
Id_{PO}	-0.5198	-2.15	0.031	***
<i>Local</i>	0.358	5.69	0.000	***
<i>Equipo grande</i>	0.6385	8.13	0.000	***
<i>Rival grande</i>	-0.615	-9.94	0.000	***
Número de observaciones		1564		
Estadístico F		210.29		
Prob > F		0.000		
R-cuadrado		0.6395		
Raíz del error cuadrático medio		1.2367		

Fuente: Elaboración propia. Nota: Niveles de significación a) *** p-valor < 0,01. b) El modelo fue estimado por el método de Mínimos Cuadrados Ordinarios.

El modelo con los parámetros estimados queda planteado de la siguiente forma:

Goles marcados

$$\begin{aligned}
 &= 4.3 Bt_{DE} + 10.2 Bt_{MC} + 2.95 Bt_{ML} + 2.3 Bt_{DL} + 2.32 Bt_{DC} + 0.65 Id_{DE} \\
 &- 0.84 Id_{DC} - 0.52 Id_{PO} + 0.36 Local + 0.64 Equipo grande \\
 &- 0.61 Rival grande + \varepsilon
 \end{aligned}$$

⁴⁹ Para ver los estadísticos descriptivos de las variables incluídas en el modelo diríjase a la tabla 15 en anexos.

Para corroborar que los resultado continuaran siendo válidos si se quitan los efectos fijos, se estimaron dos variantes del modelo original. Las estimaciones comprueban que los resultados obtenidos son válidos, las mismas pueden visualizarse en la tabla 16 de anexos.

El análisis de los resultados se dará en la siguiente sección.

Discusión de los resultados

Los equipos que participaron de la UEFA Champions League entre las temporadas 2012/2013 y 2018/2019 fueron analizados en este estudio. Se calcularon tres medidas de centralidad de red, centralidad *between*, *in degree* y *out degree*, para cada posición en el campo de juego.

Los resultados obtenidos para mediocampistas (valores de centralidad más altos que las demás posiciones) siguen la línea de lo encontrado en la literatura anterior: estudios previos han reportado evidencia del alto impacto de los mediocampistas centrales en el equipo (Duch, Waitzman, y Amaral, 2010; Peña y Touchette, 2012; Thomas y Reilly, 1976). Los altos valores de centralidad son razonables dada la responsabilidad que tienen los mediocampistas centrales a la hora de vincular los sectores del campo y crear las jugadas de ataque. Su protagonismo en el proceso de conectar la red se puede verificar en el hecho de que tienen el nivel más alto de centralidad *between*, confirmando que son la pieza clave para conectar a todos los componentes del equipo. Según Malta y Travassos (2014), la prominencia de los mediocampistas centrales se hace aún más notoria en el estilo indirecto, y también en transiciones o contraataques.

Por otra parte, si se observa lo hallado para defensas centrales, no se confirma lo encontrado por Clemente y Martins (2017), quienes los posicionan como los segundos jugadores con mayor centralidad (para todas las medidas), en cambio se encontró que ocupan el tercer lugar en relevancia. De todos modos, presentan valores de centralidad muy cercanos a los de defensas laterales (los “segundos” más centrales). Aun así, los hallazgos siguen la línea de lo encontrado en Clemente et al. (2015), Duch, Waitzman y Amaral (2010) y Costa et al. (2010), quienes coinciden en que los defensas laterales poseen mayores niveles de centralidad que los defensas centrales. Este hallazgo puede sugerir que en general, los equipos presentes en la muestra, poseen cierta tendencia a jugar usando el ancho del campo durante la construcción del ataque.

En cuanto a los mediocampistas laterales, se podría decir que ocupan una posición intermedia: no son la posición más relevante, pero sus niveles de centralidad no son despreciables, lo cual podría indicar que los equipos presentes en la muestra se apoyan en los laterales al momento del ataque en una segunda etapa (no en la formación). Los delanteros y los porteros tuvieron los valores más pequeños en todas las medidas de centralidad. Los resultados con respecto a los

porteros están justificados dado a que no se involucran en jugadas de ataque y tienen poca participación en el juego, si se la compara con la de las demás posiciones. En el caso de los delanteros, los resultados pueden reflejar la especificidad de la posición. Un delantero es el último hombre del equipo y no participa en el juego de acumulación de ataque (principalmente en la primera fase). Por esa razón, en la mayoría de los casos, presentaron los valores más pequeños en la categoría de jugadores de campo.

En resumen, se comprueba de forma parcial la primer hipótesis planteada, los mediocampistas centrales son los jugadores clave del equipo, son quienes mantienen conectada la red. El mismo análisis sería válido para cualquier firma que desee encontrar al trabajador más importante para luego tomar decisiones con ello en mente, tratando de potenciar sus habilidades y/o destrezas con el fin de maximizar el beneficio de la empresa.

En cuanto a la segunda hipótesis, aquella que plantea que es posible encontrar una relación entre goles anotados, los vínculos que se forman entre los compañeros de equipo (medidas de centralidad) y otros efectos fijos, tales como tipo de equipo y rival y condición de local, es corroborada exitosamente mediante un modelo MCO⁵⁰. El modelo estimado presenta un R cuadrado de 0.64, lo cual indica que el mismo explica el 64% de la variabilidad de los datos de respuesta en torno a su media. Por otra parte, es rechazada la hipótesis nula del test de significación global F, el cual consiste en comprobar si se cumple que todas las estimaciones son cero⁵¹.

Cada uno de los insumos considerados y las variables exógenas a la red resultaron significativas al 99% y tuvieron un coeficiente con el signo esperado, dicho signo indica la dirección del efecto de la variable sobre la cantidad de goles marcados. Además, el valor absoluto de los coeficientes está acorde con lo planteado en la hipótesis. En los siguientes párrafos se interpretarán los valores de los coeficientes.

⁵⁰ Al cual previamente se le realizaron pruebas para verificar la consistencia de los estimadores

⁵¹ Si el parámetro F es significativamente distinto de 0 se rechaza la H_0 y se puede afirmar que alguna β_j es distinto de 0 por lo que el modelo ajusta algo (Granados, 2016).

Recordar que la función estimada quedó planteada de la siguiente forma:

Goles marcados

$$\begin{aligned} &= 4.3 Bt_{DE} + 10.2 Bt_{MC} + 2.95 Bt_{ML} + 2.3 Bt_{DL} + 2.32 Bt_{DC} + 0.65 Id_{DE} \\ &- 0.84 Id_{DC} - 0.52 Id_{PO} + 0.36 Local + 0.64 Equipo grande \\ &- 0.61 Rival grande + \varepsilon \end{aligned}$$

Los coeficientes estimados representan los cambios medios en la variable dependiente para una unidad de cambio en la variable independiente (mientras se mantienen constantes las otras variables incluidas en el modelo), pero dado que todas las variables independientes ingresadas toman valores entre cero y uno, ninguna de ellas varía de a una unidad, por lo que la interpretación de los coeficientes no es demasiado intuitiva.

Si la centralidad *between* de mediocampistas centrales aumenta en 0.01 (*ceteris paribus*), los goles anotados, o la chance de gol, aumentaría en 0.102 (10.2%). Por cada aumento de 0.01 en el BC de los delanteros, el marcador, a lo demás constante, aumentaría a favor en un 4.3%. Exactamente la misma interpretación es válida para los demás BC, para no reiterar, no serán explicitadas. La centralidad *between* de todas las posiciones afecta de forma positiva a los goles marcados, las mismas serán nombradas en orden de importancia: MC, DE, ML, DC y DL. Los resultados son coherentes con lo planteado en la hipótesis, e indicarían que cuanto más conectada está la red, más abultado es el marcador.

La centralidad *between* es una de las medidas de centralidad más relevantes porque cuantifica con éxito la frecuencia con la que cada jugador se encuentra entre otros nodos de la red, quizás actuando como un mediador o puente para ellos. Por lo anterior, un aumento en la centralidad de este tipo, en todas las posiciones, indicaría que hay un mayor flujo de información entre los nodos y que los sectores del campo están más vinculados. Si los sectores están más vinculados, el balón podría fudir por el campo, y sería más factible para el equipo encontrar el mejor camino para llegar al arco rival y convertir.

Realizando un análisis de relevancia entre el BC de las posiciones, era de esperar que el jugador más relevante fuera el medicampista centro, dado su importancia en el proceso de conexión de la red. Por otra parte, los delanteros son una pieza clave dado su posición en el campo, es la posición que tiene como función principal marcar o “ayudar” a que otro compañero lo haga, por lo tanto es previsible que un aumento de su centralidad sea favorable a los goles marcados. El BC de las demás posiciones presenta valores “de impacto” sobre los goles marcados similares y no tan altos como el de MC o DE. Los valores son positivos por las razones antes mencionadas: el campo está más vinculado, hay más caminos abiertos y se consiguen más formas de llegar al arco rival, pero el impacto no sería tan grande dado que no son ni las posiciones más importantes a la hora de vincular los sectores del campo, ni están posicionados sobre el arco rival.

En tanto, la centralidad *in degree* no tiene un impacto tan fuerte sobre los goles anotados como el *between* y el sentido del efecto no se mantiene igual para todas las posiciones: un aumento de 0.01 en el IDC de los delanteros, haría aumentar, *ceteris paribus*, en un 0.65% (0.0065) la posibilidad de goles marcados. Por otra parte, un aumento de 0.01 del IDC de defensas centrales o de porteros, hace disminuir la chance de gol, *ceteris paribus*, un 0.84 y 0.52% (0.0084 y 0.0052) respectivamente. Los resultados en lo que respecta a esta medida podrían indicar que, cuanto más se involucra en el juego a jugadores que están posicionados cerca al área rival – cuanto más reciben la pelota – existen más probabilidad de gol, lo cual es lógico, porque si se juega sobre el área rival es más probable que estos reciban un gol (lo opuesto sería válido para las posiciones que están cerca del área propia, como lo son porteros y defensas centrales).

Queda por analizar el impacto de las variables exógenas a la red, que, por ser dummies, su interpretación es más sencilla. Si el equipo es local, los goles anotados se incrementan en 0.36. Por otra parte, si el equipo es considerado grande, los goles marcados aumentan en 0.64 o si el equipo rival es grande, disminuyen en 0.61. Que se cumpla una de las tres cosas anteriores no asegura (o no) un gol, pero vinculados con otros aspectos del partido, como por ejemplo las centralidades podrían ser de ayuda para que el marcador varíe.

Para concluir, se podría decir que los resultados confirman, de forma preliminar, la hipótesis planteada, pero dado que no existe un modelo teórico detrás del modelo planteado, no se

puede afirmar que no exista otro tipo de relación que explique de mejor forma los goles marcados por equipo. Por otra parte, debe ser tenido en cuenta que algunas observaciones podrían no estar libres de endogeneidad, por lo mismo los resultados podrían variar.

En cuanto a las implicaciones prácticas, es posible observar que el análisis de redes sociales puede ayudar a los entrenadores a identificar a los jugadores clave en el proceso de ataque y a tomar decisiones durante el partido o en las sesiones de entrenamiento para optimizar el comportamiento del equipo o para evitar las fortalezas de los oponentes. Por otra parte, la estimación de la relación entre las medidas de centralidad y los resultados podría ser útil para identificar que sectores del campo o relaciones entre compañeros potenciar para lograr un mayor impacto en el marcador.

Limitaciones

Aunque las redes de fútbol sean de naturaleza dinámica, en la literatura han sido estudiadas utilizando modelos estáticos. El presente trabajo ha estudiado las actuaciones de los equipos y jugadores mediante la representación del juego por una red de pases y ha utilizado medidas de red para obtener una indicación de los jugadores clave. Este análisis estático, puede proporcionar algunos resultados interesantes, pero no incorpora factores esenciales; por ejemplo, no está diseñado para capturar los errores cometidos por los jugadores, ya sean críticos o no, y tampoco incorpora diferentes factores que representan el desempeño del equipo además del resultado final del partido. Tampoco se considera el momento en que se realizan los pases y, por lo tanto, no proporciona una descripción temporal de las actuaciones de los jugadores.

Por lo anterior, sería interesante aplicar el método de Análisis de redes dinámico – el cual vincula el análisis tradicional de redes sociales con el análisis de enlaces, la simulación social y los sistemas de múltiples agentes – para poder incorporar algunos de los factores faltantes.

Dejando de lado las limitaciones propias del análisis de redes sociales, el presente estudio presenta otras limitaciones: no se consideró la especificidad del proceso de ataque (estilo indirecto, transición o contraataque), tampoco se consideraron las diferencias en cuanto a cómo se posiciona cada equipo en el campo de juego (alineación inicial). Estas limitaciones pueden significar que ciertos aspectos específicos de las interacciones entre compañeros de equipo no fueran tomadas en cuenta. Además, tampoco se consideró la especificidad del comportamiento táctico y el modelo de juego. Estudios futuros deberían considerar cruzar la información sobre el comportamiento táctico e individual con las medidas de redes.

En cuanto al modelo que estimado, dado que no tiene un fundamento teórico detrás, no se puede afirmar que no exista otro tipo de relación que explique de mejor forma los goles marcados por equipo: los resultados podrían reflejar mejor la realidad si se considera otro tipo de estimación, o si se consideran variables adicionales, pero dado que no se ha planteado tal teoría, en éste momento no se puede confirmar ni negar nada. Además, se podría estar en presencia de una posible endogeneidad del tipo simultaneidad, pero dada las características de

los datos con los que se cuenta, no se puede asegurar que tanto peso tiene sobre la estimación final. La posible simultaneidad podría ser tratable si las observaciones presentes en la base de datos consideraran el aspecto temporal, y existiera una red de pases para antes y después de cada gol, pero dado a que en el presente estudio no se cuenta con dichos datos, no se han podido tomar en cuenta, para futuras investigaciones, podría ser un hecho a considerar.

Conclusiones

El presente estudio reveló que los mediocampistas centrales son los jugadores más destacados durante un partido de fútbol entre los equipos que jugaron la UEFA Champions League en las últimas siete temporadas. Por otra parte, los defensas laterales son, dado su alto nivel de *in degree*, junto con los mediocampistas centrales, quienes presentan más prestigio entre el resto de las posiciones. A lo anterior se debe sumar que también son quienes presentan uno de los mayores niveles de centralidad *between*, este hallazgo puede sugerir que en general, los equipos presentes en la muestra, poseen cierta tendencia a jugar usando el ancho del campo durante la construcción del ataque.

También se encontró que los mediocampistas centrales y defensas laterales fueron los jugadores más prominentes en el juego de “salida”, justificado por su elevado valor de centralidad *out degree*. En tanto, los porteros y delanteros son los puestos con los valores de centralidad más pequeños del equipo, justificado por su posición extrema en el campo de juego.

Por otra parte, se encontró, a raíz de una estimación por MCO, una aproximación de la relación entre algunas medidas de centralidad de red y los resultados deportivos de un equipo de fútbol. También fueron tomadas en cuenta algunas variables exógenas a la red. Se encontró que las centralidades de los jugadores, sobre todo el *between*, son una pieza clave a la hora de intentar explicar la cantidad de goles anotados por el equipo, además de este hallazgo, fue posible identificar que el *between* de mediocampistas centrales, por ser los jugadores claves a la hora de conectar la red, es la medida de mayor impacto en el número de goles anotados.

Referencias

- Aumann, R., & Myerson, R. (1988). Endogenous formation of links between players and of coalitions: An application of the Shapley value. En A. Roth (Ed.), *The Shapley Value: Essays in Honor of Lloyd S. Shapley* (págs. 175 - 192). Cambridge: Cambridge University Press.
- Barnes, J. A., & Harary, F. (1983). Graph theory in network analysis. *Social Networks*, 5, 235–244.
- Baum, C. F. (2006). Stata tip 38: Testing for groupwise heteroskedasticity. *The Stata Journal*, 6, 590–592.
- Blass, A. (1992). Does the baseball labor market contradict the human capital model of investment? *The Review of Economics and Statistics*, 74(2), 261 - 268.
- Boorman, S. (1975). A Combinatorial Optimization Model for Transmission of Job Information through Contact Networks. *Bell Journal of Economics*, 6, 216 - 249.
- Carmichael, F., Thomas, D., & Ward, R. (2000). Team performance: the case of English premiership football. *Managerial and Decision Economics*, 21(1), 31 - 45.
- Carmichael, F., & Thomas, D. (1995). Production and efficiency in team sports: an investigation of rugby league footb. *Applied Economics*, 27(9), 859 - 869.
- Clemente, F. M., & Martins, F. M. (2017). Who are The Prominent Players in the UEFA Champions League? An Approach Based on Network Analysis. *Walailak J Sci & Tech*, 14(8), 627-636.
- Clemente, F. M., Lourenço Martins, F. M., Wong, P., Kalamaras, D., & Mendes, R. S. (2015). Midfielder as the prominent participant in the building attack: A network analysis of national teams in FIFA World Cup 2014. *International Journal of Performance Analysis in Sport*, 15(2), 704 - 722.
- Clemente, F. M., Martins, F. M., & Mendes, R. S. (2016). *Social Network Analysis Applied to Team Sports Analysis*. Netherlands: Springer International.
- Costa, I. T., Garganta, J., & Greco, P. (2010). Influence of relative age effects and quality of tactical behaviour in the performance of youth football players. *International Journal of Performance Analysis in Sport*, 10 (2), 82–97.
- Cumming, G. (2013). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-analysis*.
- Duch, J., Waitzman, J. S., & Amaral, L. A. (2010). Quantifying the performance of individual players in a team activity. *PLoS ONE*, 5(6), e10937.
- Dutta, B., & Jackson, M. O. (2003). *Networks and Groups: Models of Strategic Formation*. Springer.
- Ellison, G. (1993). Learning, local interaction and coordination. *Econometrica*, 61, 1047 - 1071.
- Espitia-Escuer, M., & Garcia-Cebrian, L. I. (2016). Productivity and Competitiveness: The Case of Football Teams Playing in the UEFA. *Athens Journal of Sports*, 3(1), 57 - 86.

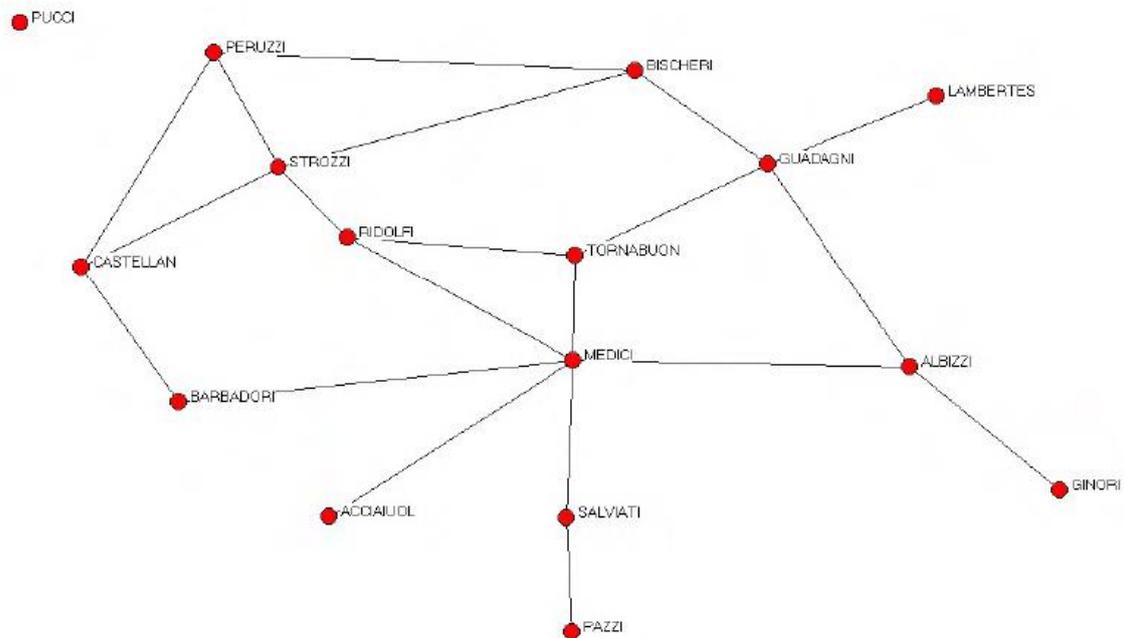
- Espitia-Escuer, M., & Garcia-Cebrian, L. I. (2019). Productividad y sus determinantes en la competicion de la UEFA Champions League. *Papeles de economía española «DEPORTE Y ECONOMÍA»*(159).
- Ferri, F. (2003). Network Formation with Endogenous Decay. *Università Ca Foscari, Venezia*.
- Fichajes de fútbol, v. d. (2020). *Transfermarkt*. Obtenido de <https://www.transfermarkt.es/>
- Fort, R. (2006). *Sports Economics*. Upper Saddle River: Prentice Hall.
- Fort, R., & Quirk, J. (1995). Cross-subsidisation, incentives, and outcomes in professional sports leagues. *Journal of Economic Literature*, 33(3), 1265 - 1299.
- Galeotti, A., & Vega-Redondo, F. (2005). Local externalities and complex networks. *University of Essex*.
- Garcia del Barrio, P., & Szymanski, S. (2009). Goal! Profit maximization versus win maximization in football. *Review of Industrial Organization*, 34(1), 45 - 68.
- Goyal, S., & Joshi, S. (2000). Networks of Collaboration in Oligopoly. *Games and Economic Behavior*, 43(1), 57 - 85.
- Granados, M. (2016). Modelos de regresión lineal múltiple. *Documentos de Trabajo en Economía Aplicada. Universidad de Granada. España*.
- Guesnerie, R., & Oddou, C. (1981). Second best taxation as a game. *Second best taxation as a game*, 25, 67 - 91.
- Hansen, B. (2000). Sample Splitting and Threshold Estimation. *Econometrica*, 68(3), 575-603.
- Horrace, W. C., Liu, X., & Patacchini, E. (2016). Endogenous network production functions with selectivity. *Journal of Econometrics*, 190, 222 – 232.
- Jackson, M. O. (2005). The Economics of Social Networks. *9th World Congress of the Econometric Society*.
- Jackson, M. O. (2008). *Social and Economic Networks*. Nueva Jersey: Princeton University Press.
- Jackson, M. O., & Watts, A. (2002). On the Formation of Interaction Networks in Social Coordination Games. *Games and Economic Behavior*, 41(2), 265 - 291.
- Jackson, M. O., & Wolinsky, A. (1996). A Strategic Model of Social and Economic Networks. *Journal of Economic Theory*, 71(1), 44 - 74.
- Kelley, K., & Preacher, K. (2012). On Effect Size. *Psychological Methods*, 17(2), 137 - 152.
- Késenne, S. (2007). *The Economic Theory of Professional Sport: An Analytical Treatment*. Cheltenham, UK and Northampton, USA: Edward Elgar Publishing.
- Linder, M. (2017). *Football and Mathematics: a search for a relation between passing structure and successful results*. U.U.D.M. Project Report 2017:28; Department of Mathematics Uppsala University.
- Malta, P. P., & Travassos, B. (2014). Characterization of the defense-attack transition of a soccer team. *Motricidade*, 10(1), 27–37.

- Martínez, J. (2016). Endogeneidad, diferencia de medias y regresión. *RICYDE. REVISTA INTERNACIONAL DE CIENCIAS DEL DEPORTE*, 12(46), 335-337.
- Mazur, M. J. (1994). Evaluating the relative efficiency of baseball players. En A. Charnes, W. W. Cooper, A. Y. Lewin, & L. M. Seiford, *Data Envelopment Analysis: Theory, Methodology and Application*. Norwell, MA, USA: Kluwer Academic Publishers.
- Moder, K. (2010). Alternatives to F-Test in One Way ANOVA in case of heterogeneity of variances (a simulation study). *Psychological Test and Assessment Modeling*, 52(4), 343-353.
- Montgomery, J. (1991). Social Networks and Labor Market Outcomes. *The American Economic Review*, 81, 1408 - 1418.
- Morris, S. (2000). Contagion. *Review of Economic Studies*, 67, 57 - 78.
- Myerson, R. (1977). Graphs and Cooperation in Games. *Math, Operations Research*, 2, 225 - 229.
- NetworkX developers. (2014 - 2020). *NetworkX — NetworkX Analysis in Python*. Obtenido de <https://networkx.github.io/>
- Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Soc. Network*, 32, 245-51.
- Padgett, J. F., & Ansell, C. K. (1993). Robust Action and the Rise of the Medici. *American Journal of Sociology*, 98, 1259-1319.
- Pappalardo, L., Cintia, P., Rossi, A., Massucco, E., & Ferragina, P. (2019). A public data set of spatio-temporal match events in soccer competitions. *Scientific Data*, 2.
- Peña, J. L., & Touchette, H. (2012). A network theory analysis of football strategies. *arXiv preprint arXiv*, p. 1206.6904.
- Peón, F. V. (2010). *Normalidad de los errores*. Mexico DF.: Universidad Autónoma Metropolitana.
- Pestana Barros, C., Couto, E., & Samagaio, A. (2014). Management ability, strategy, tactics and team performance. En J. Goddard, & P. Sloane, *Handbook on the Economics of Professional Football*. Cheltenham, UK and Northampton, USA: Edward Elgar Publishing.
- Rottenberg, S. (1956). The baseball players' labor market. *Journal of Political Economy*, 64(3), 242 - 258.
- Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage*, 52, 1059-69.
- Sandy, R., Sloane, P., & Rosentraub, M. S. (2004). *The Economics of Sport: An International Perspective*. New York: Palgrave MacMillan.
- Schofield, J. A. (1988). Production functions in the sports industry: an empirical analysis of profession. *Applied Economics*, 20(2), 177 - 193.
- Shephard, R. W. (1970). *Theory of cost and production functions*. Princeton, New Jersey: Princeton University Press.

- Szymanski, S. (2003). The economic design of sporting contests. *Journal of Economic Literature*, 41(4), 1137 - 1187.
- Tesfatsion, L. (1997). A Trade Network Game with Endogenous Partner Selection. En H. A. at., *Computational Approaches to Economic Problems* (págs. 249-269).
- Tesfatsion, L. (1998). Gale-Shapley matching in an Evolutionary Trade Network Game. *Iowa State University Economic Report*, 43.
- Thomas, V., & Reilly, T. (1976). A motion analysis of work-rate in different positional roles in professional football match-play. *J. Hum. Mov. Stud.*, 2, 87-97.
- Tiebout, C. M. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy*, 64, 416 - 424.
- Uefa.com. (2020). *Press kits - UEFA Champions League – Inside UEFA* –. Obtenido de <https://www.uefa.com/insideuefa/mediaservices/presskits/uefachampionsleague/index.html>
- Vogt, W. P., & Johnson, R. B. (2005). *The SAGE Dictionary of Statistics & Methodology: A Nontechnical Guide for the Social Sciences*. SAGE.
- Weimar, D. (2019). The Economics of Professional Soccer. En P. Downward, B. Frick, T. Pawlowski, J. Rueseski, & B. Soebbing, *Handbooks of Sports Economics*. Sage Publications.

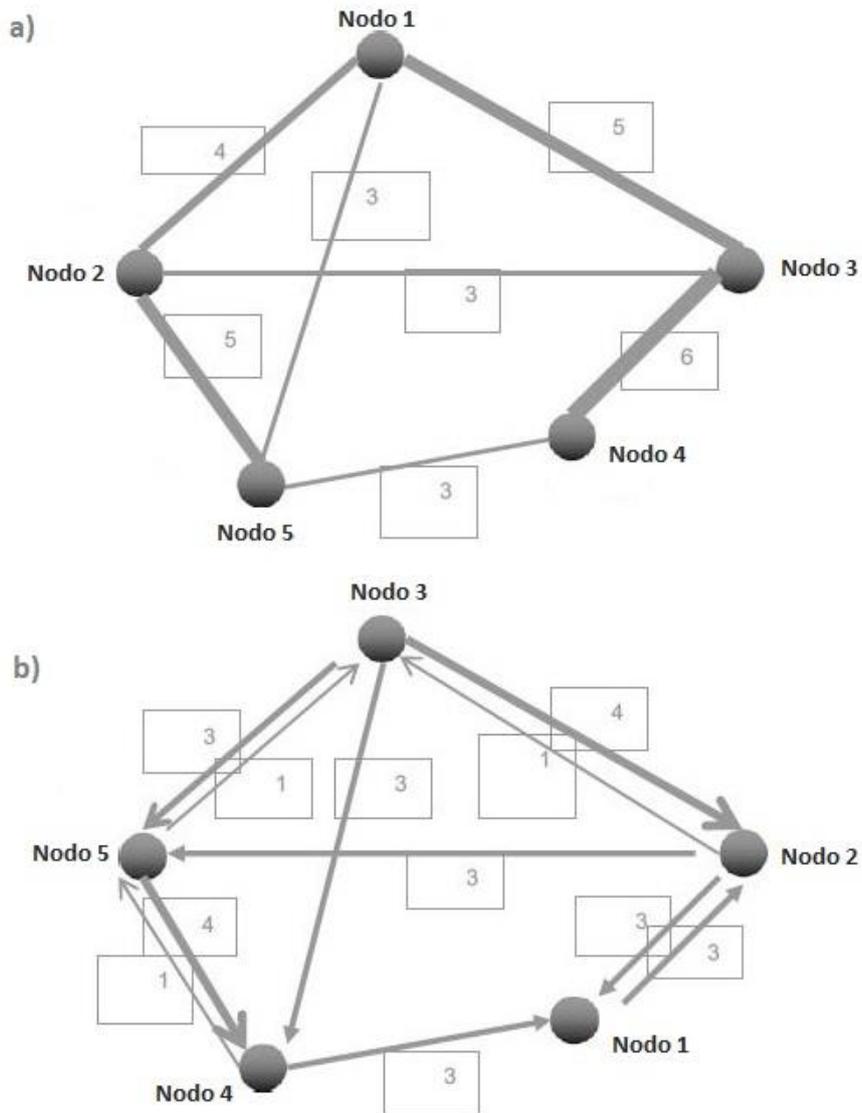
Anexos:

Figura 5. Ejemplo de red, caso de familias Florentinas del Siglo XV



Fuente: Jackson, 2008, p. 19: *Basado en Matrimonios Florentinos del siglo XV, datos de Padgett y Ansell, 1993.*

Figura 6. Redes ponderadas y no ponderadas



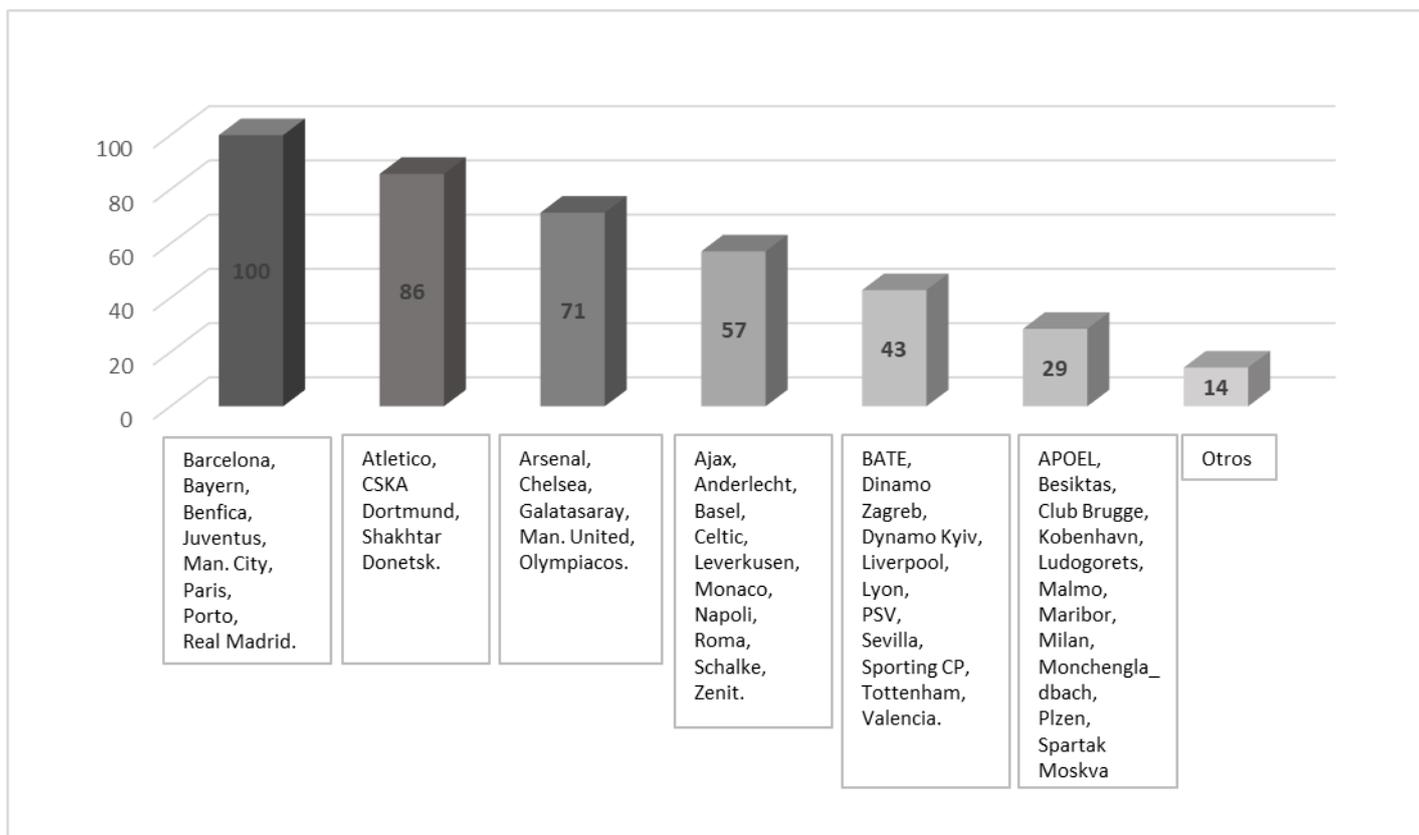
Fuente: Clemente et al. (2016). **a)** Red ponderada no dirigida: notar que no hay dirección en los enlaces, parten de ambos nodos o no parten de ninguno en absoluto. **b)** Red ponderada dirigida: Los enlaces están representados mediante una flecha. En ambos casos la ponderación está representado por el número sobre la flecha o línea.

Tabla 3. Equipos participantes del estudio y número de competiciones jugadas

Equipo	competiciones jugadas	Equipo	competiciones jugadas	Equipo	competiciones jugadas
AEK	1	FCSB	1	Monaco	4
Ajax	4	Feyenoord	1	Monchengladbach	2
Anderlecht	4	Galatasaray	5	Montpellier	1
APOEL	2	Gent	1	Napoli	4
Arsenal	5	Hoffenheim	1	Nordsjaelland	1
Astana	1	Internazionale	1	Olympiacos	5
Athletic	1	Juventus	7	Paris	7
Atletico	6	Kobenhavn	2	Plzen	2
Austria Wien	1	Legia	1	Porto	7
Barcelona	7	Leicester	1	PSV	3
Basel	4	Leipzig	1	Qarabag	1
BATE	3	Leverkusen	4	Real Madrid	7
Bayern	7	Liverpool	3	Real Sociedad	1
Benfica	7	Lokomotiv Moskva	1	Roma	4
Besiktas	2	LOSC	1	Rostov	1
Braga	1	Ludogorets	2	Schalke	4
Celtic	4	Lyon	3	Sevilla	3
CFR Cluj	1	M. Tel-Aviv	1	Shakhtar Donetsk	6
Chelsea	5	Malaga	1	Spartak Moskva	2
Club Brugge	2	Malmö	2	Sporting CP	3
Crvena zvezda	1	Manchester City	7	Tottenham	3
CSKA Moskva	6	Manchester United	5	Valencia	3
Dinamo Zagreb	3	Maribor	2	Wolfsburg	1
Dortmund	6	Marseille	1	Young Boys	1
Dynamo Kyiv	3	Milan	2	Zenit	4

Fuente: Elaboración propia en base a datos de UEFA. *Equipos participantes de al menos una competición de UEFA Champions League entre las temporadas 2012/2013 y 2018/2019 (ordenados alfabéticamente.) En la siguiente columna: número de competiciones de las cuales participaron.*

Figura 7. Porcentaje de temporadas jugadas por cada equipo



Fuente: Elaboración propia en base a datos de UEFA. Porcentaje de participación de cada equipo entre las temporadas 2012/2013 y 2018/2019 de la UEFA Champions League.

Tabla 4. Posiciones y su participación en el total de los datos

Posición	N° de observaciones	Participación en el total de observaciones
Portero	1,784	8 %
Defensa central	4,203	18 %
Defensa lateral	3,627	16 %
Mediocampista central	6,611	28 %
Mediocampista lateral	3,846	16 %
Delantero	3,279	14 %

Fuente: Elaboración propia en base a datos de UEFA. Número de observaciones de cada tipo presentes en los datos y su porcentaje de participación en el total.

Tabla 5. Comprobación de normalidad para las centralidades

a. Centralidad *Between*

-> posiciones_jug = Portero

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Between	1,784	0.93026	74.545	10.921	0.00000

-> posiciones_jug = Def. central

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Between	4,203	0.91826	189.648	13.687	0.00000

-> posiciones_jug = Def. lateral

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Between	3,627	0.92209	158.370	13.154	0.00000

-> posiciones_jug = Medioc. central

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Between	6,611	0.84688	531.356	16.605	0.00000

-> posiciones_jug = Medioc. lateral

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Between	3,846	0.92157	168.063	13.334	0.00000

-> posiciones_jug = Delantero

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Between	3,279	0.93742	116.161	12.307	0.00000

b. Centralidad In Degree

-> posiciones_jug = Portero

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
In_degree	1,784	0.99669	3.537	3.200	0.00069

-> posiciones_jug = Def. central

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
In_degree	4,203	0.95078	114.193	12.364	0.00000

-> posiciones_jug = Def. lateral

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
In_degree	3,627	0.95157	98.455	11.919	0.00000

-> posiciones_jug = Medioc. central

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
In_degree	6,611	0.94637	186.104	13.829	0.00000

-> posiciones_jug = Medioc. lateral

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
In_degree	3,846	0.97334	57.115	10.525	0.00000

-> posiciones_jug = Delantero

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
In_degree	3,279	0.96997	55.748	10.407	0.00000

c. Centralidad *Out Degree*

-> posiciones_jug = Portero

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Out_degree	1,784	0.83519	176.155	13.099	0.00000

-> posiciones_jug = Def. central

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Out_degree	4,203	0.88855	258.584	14.496	0.00000

-> posiciones_jug = Def. lateral

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Out_degree	3,627	0.95207	97.427	11.892	0.00000

-> posiciones_jug = Medioc. central

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Out_degree	6,611	0.94266	198.978	14.006	0.00000

-> posiciones_jug = Medioc. lateral

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Out_degree	3,846	0.98110	40.491	9.630	0.00000

-> posiciones_jug = Delantero

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Out_degree	3,279	0.98432	29.113	8.726	0.00000

Fuente: Elaboración propia en base a datos de UEFA. Test de Shapiro-Wilk realizado para comprobar si las centralidades a. between b. In degree c. Out degree de cada posición siguen una distribución normal o no. Toma como hipótesis nula que los datos se distribuyen de forma normal. Dado que para cada una de las posiciones, el p-valor es menor 0.05, se rechaza la hipótesis nula.

Tabla 6. Prueba de Levene para la centralidad *Between*

Posicion	Summary of Centralidad <i>Between</i>		
	Mean	Std. Dev.	Freq.
Portero	.04358495	.02555571	1,784
Def. cent	.07341146	.06099869	4,203
Def. late	.0788399	.0628819	3,627
Medioc. c	.08898679	.08219221	6,611
Medioc. l	.06068931	.04633673	3,846
Delantero	.04797877	.02964289	3,279
Total	.07071869	.06304315	23,350

W0 = 691.10096 df(5, 23344) Pr > F = 0
 W50 = 497.64019 df(5, 23344) Pr > F = 0
 W10 = 561.40441 df(5, 23344) Pr > F = 0

Fuente: Elaboración propia en base a datos de UEFA. Prueba de Levene para la centralidad *between*. Como hipótesis nula propone que existe homogeneidad de varianza entre los grupos analizados. W0 representa el estadístico de Levene, y es robusto ante la no normalidad de los datos, W50 y W10 son dos variantes del test, propuestas por Brown and Forsythe (usan estimadores mas robustos) (Baum, 2006). Dado que en los tres casos, el p valor es igual a cero, se rechaza la hipótesis nula.

Tabla 7. Prueba de Levene para la centralidad *in degree*

Posicion	Summary of Centralidad <i>In degree</i>		
	Mean	Std. Dev.	Freq.
Portero	.38837763	.12420551	1,784
Def. cent	.62854723	.17892306	4,203
Def. late	.66423555	.1833241	3,627
Medioc. c	.69119274	.20138275	6,611
Medioc. l	.60365277	.21054435	3,846
Delantero	.56436465	.22749123	3,279
Total	.62036435	.2110325	23,350

W0 = 180.29022 df(5, 23344) Pr > F = 0.00000000
 W50 = 142.68729 df(5, 23344) Pr > F = 0.00000000
 W10 = 162.99066 df(5, 23344) Pr > F = 0.00000000

Fuente: Elaboración propia en base a datos de UEFA. Prueba de Levene para la centralidad *in degree*.

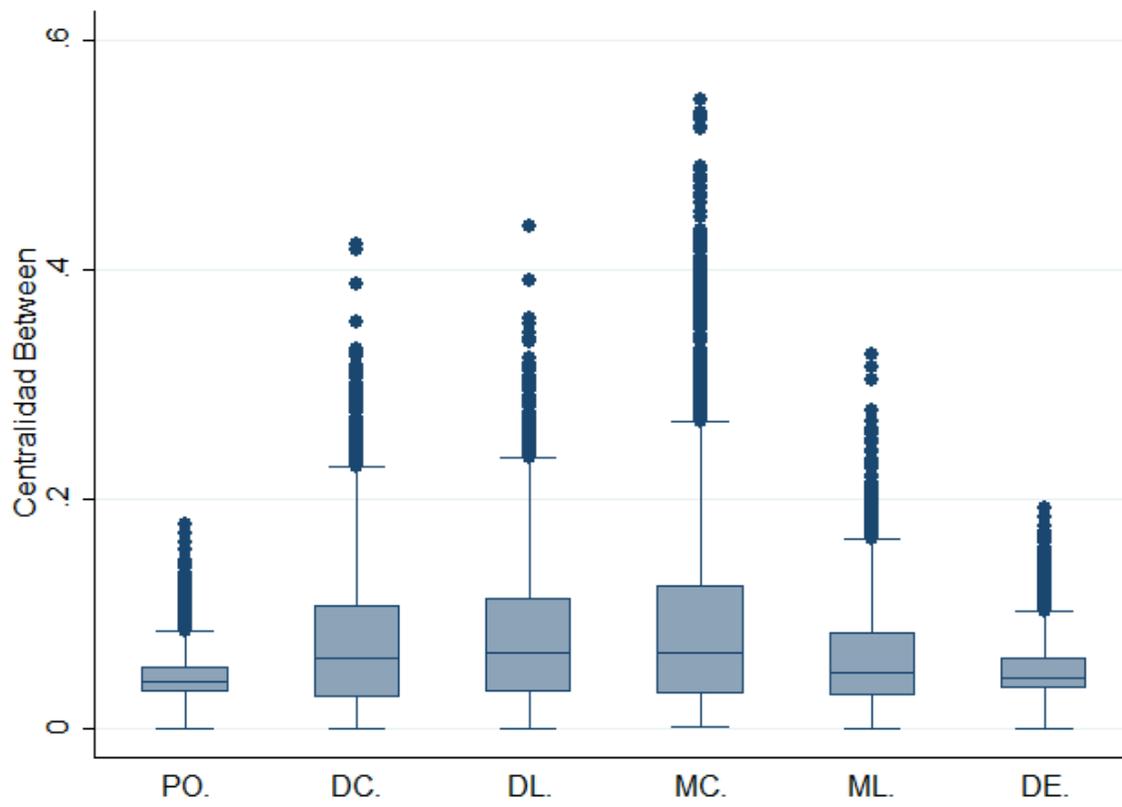
Tabla 8. Prueba de Levene para la centralidad *out degree*

Posicion	Summary of Centralidad Out degree		
	Mean	Std. Dev.	Freq.
Portero	.51223591	.090793	1,784
Def. cent	.61617049	.16337684	4,203
Def. late	.66340359	.19392757	3,627
Medioc. c	.68525948	.20526504	6,611
Medioc. l	.54013418	.21382766	3,846
Delantero	.46035525	.21312019	3,279
Total	.60072247	.20917427	23,350

W0 = 347.70901	df(5, 23344)	Pr > F = 0
W50 = 348.67086	df(5, 23344)	Pr > F = 0
W10 = 339.24751	df(5, 23344)	Pr > F = 0

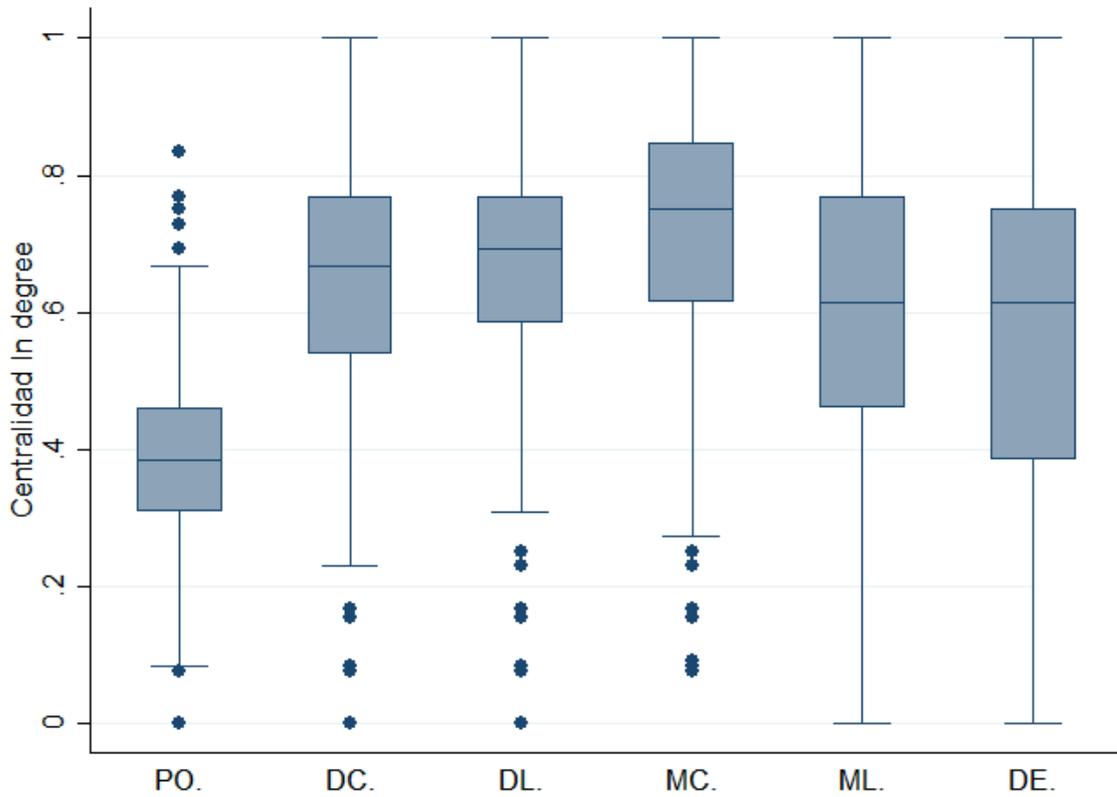
Fuente: Elaboración propia en base a datos de UEFA. Prueba de Levene para la centralidad out degree.

Figura 8. Diagrama de Box-Whisker para la centralidad between



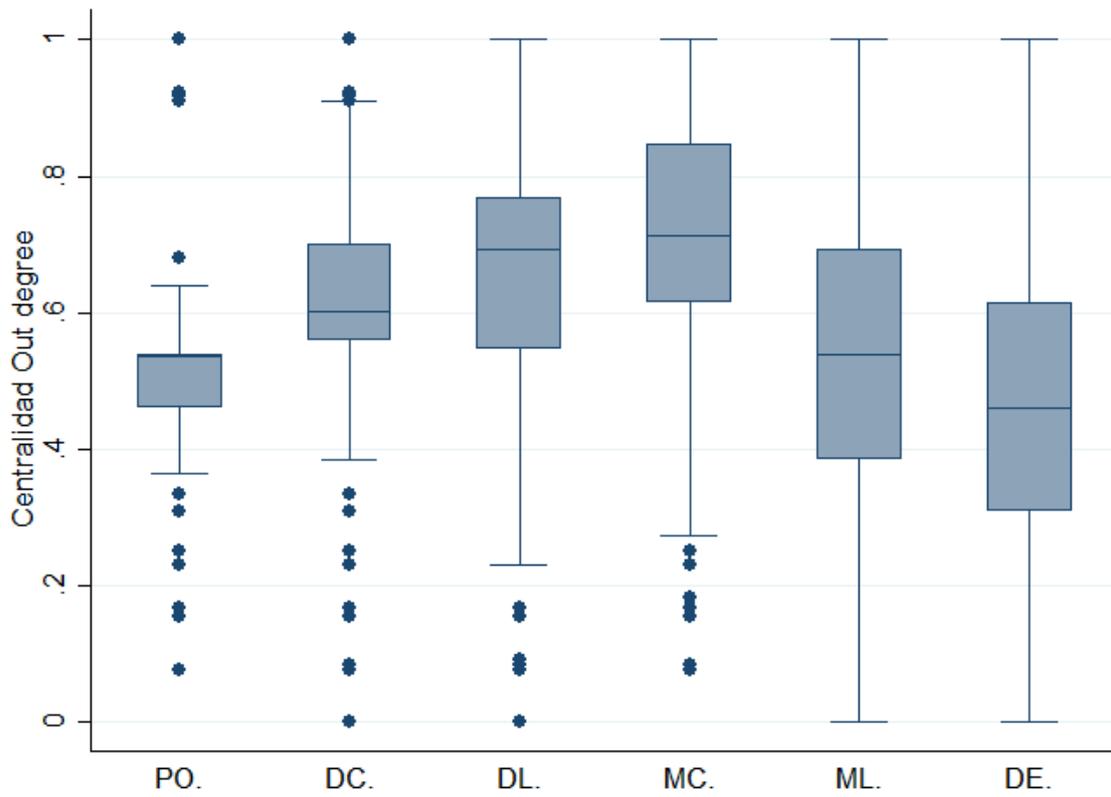
Fuente: Elaboración propia en base a datos de UEFA. Diagrama de Box-Whisker para la centralidad between.

Figura 9. Diagrama de Box-Whisker para la centralidad *in degree*



Fuente: Elaboración propia en base a datos de UEFA. Diagrama de Box-Whisker para la centralidad *in degree*.

Figura 10. Diagrama de Box-Whisker para la centralidad *out degree*



Fuente: Elaboración propia en base a datos de UEFA. Diagrama de Box-Whisker para la centralidad *out degree*.

Tabla 9. Test de Welch y tamaño del efecto de Cohen para la centralidad *between*.

Two-sample Welch's t test with unequal variances

Ho: Diferencia = 0

Centralidad Between

Ha: Diferencia != 0

Grupo	Obs	Media	Std. Error	Std. Dev	[Intervalo de confianza (95%)]		t	p-value	GL Welch	Cohen's d	Tamaño del efecto
Portero	1,784	0.04358	0.00061	0.02556	0.04240	0.04477					
Def. Central	4,203	0.07341	0.00094	0.06100	0.07157	0.07526					
Diferencia		-0.02983	0.00112		-0.03202	-0.02763	-26.663	0.00	5986.1	0.64	moderado
Portero	1,784	0.04358	0.00061	0.02556	0.04240	0.04477					
Def. lateral	3,627	0.07884	0.00104	0.06288	0.07679	0.08089					
Diferencia		-0.03526	0.00121		-0.03762	-0.03289	-29.215	0.00	5264.6	0.73	moderado
Portero	1,784	0.04358	0.00061	0.02556	0.04240	0.04477					
Med. Central	6,611	0.08899	0.00101	0.08219	0.08701	0.09097					
Diferencia		-0.04540	0.00118		-0.04771	-0.04309	-38.538	0.00	8265.6	0.75	moderado
Portero	1,784	0.04358	0.00061	0.02556	0.04240	0.04477					
Med. Lateral	3,846	0.06069	0.00075	0.04634	0.05922	0.06215					
Diferencia		-0.01710	0.00096		-0.01899	-0.01522	-17.791	0.00	5471.8	0.46	pequeño
Portero	1,784	0.04358	0.00061	0.02556	0.04240	0.04477					
Delantero	3,279	0.04798	0.00052	0.02964	0.04696	0.04899					
Diferencia		-0.00439	0.00080		-0.00596	-0.00283	-5.5179	0.00	4143.8	0.16	trivial
Def. Central	4,203	0.07341	0.00094	0.06100	0.07157	0.07526					
Def. lateral	3,627	0.07884	0.00104	0.06288	0.07679	0.08089					
Diferencia		-0.00543	0.00141		-0.00818	-0.00267	-3.8622	0.00	7590.1	0.09	trivial
Def. Central	4,203	0.07341	0.00094	0.06100	0.07157	0.07526					
Med. Central	6,611	0.08899	0.00101	0.08219	0.08701	0.09097					
Diferencia		-0.01558	0.00138		-0.01828	-0.01287	-11.278	0.00	10561	0.22	pequeño
Def. Central	4,203	0.07341	0.00094	0.06100	0.07157	0.07526					
Med. Lateral	3,846	0.06069	0.00075	0.04634	0.05922	0.06215					
Diferencia		0.01272	0.00120		0.01037	0.01508	10.5888	0.00	7789.8	0.23	pequeño
Def. Central	4,203	0.07341	0.00094	0.06100	0.07157	0.07526					
Delantero	3,279	0.04798	0.00052	0.02964	0.04696	0.04899					
Diferencia		0.02543	0.00107		0.02333	0.02754	23.6826	0.00	6382.5	0.53	pequeño
Def. lateral	3,627	0.07884	0.00104	0.06288	0.07679	0.08089					
Med. Central	6,611	0.08899	0.00101	0.08219	0.08701	0.09097					
Diferencia		-0.01015	0.00145		-0.01300	-0.00730	-6.982	0.00	9185.6	0.14	trivial
Def. lateral	3,627	0.07884	0.00104	0.06288	0.07679	0.08089					
Med. Lateral	3,846	0.06069	0.00075	0.04634	0.05922	0.06215					
Diferencia		0.01815	0.00128		0.01563	0.02067	14.1368	0.00	6648.4	0.33	pequeño
Def. lateral	3,627	0.07884	0.00104	0.06288	0.07679	0.08089					
Delantero	3,279	0.04798	0.00052	0.02964	0.04696	0.04899					
Diferencia		0.03086	0.00117		0.02858	0.03315	26.481	0.00	5276.1	0.63	moderado

Two-sample Welch's t test with unequal variances

Ho: Diferencia = 0

Centralidad Between

Ha: Diferencia != 0

Grupo	Obs	Media	Std. Error	Std. Dev	[Intervalo de confianza (95%)]		t	p-value	GL Welch	Cohen's d	Tamaño del efecto
Med. Central	6,611	0.08899	0.00101	0.08219	0.08701	0.09097					
Med. Lateral	3,846	0.06069	0.00075	0.04634	0.05922	0.06215					
Diferencia		0.02830	0.00126		0.02583	0.03076	22.5113	0.00	10448	0.42	pequeño
Med. Central	6,611	0.08899	0.00101	0.08219	0.08701	0.09097					
Delantero	3,279	0.04798	0.00052	0.02964	0.04696	0.04899					
Diferencia		0.04101	0.00114		0.03878	0.04323	36.1077	0.00	9250	0.66	moderado
Med. Lateral	3,846	0.06069	0.00075	0.04634	0.05922	0.06215					
Delantero	3,279	0.04798	0.00052	0.02964	0.04696	0.04899					
Diferencia		0.01271	0.00091		0.01093	0.01449	13.9833	0.00	6631.9	0.33	pequeño

Fuente: Elaboración propia en base a datos de UEFA. *Tests de Welch entre grupos para la centralidad between. Como hipótesis nula, propone que las diferencias de medias entre los dos grupos comparados es cero, como hipótesis alternativa indica que hay diferencias entre las medias de los grupos. Dado que el p valor obtenido mediante el test, se rechaza la hipótesis nula. Además, es posible ver el resultado del test de Cohen, y en la última columna se indica el tipo de tamaño del efecto (según la clasificación de Cohen) para cada uno de los grupos.*

Tabla 10. Test de Welch y tamaño del efecto de Cohen para la centralidad *in degree*

Two-sample Welch's t test with unequal variances
Centralidad In degree

Ho: Diferencia = 0

Ha: Diferencia != 0

Grupo	Obs	Media	Std. Error	Std. Dev	[Intervalo de confianza (95%)]		t	p-value	GL Welch	Cohen's d	Tamaño del efecto
Portero	1,784	0.38838	0.00294	0.12421	0.38261	0.39415					
Def. Central	4,203	0.62855	0.00276	0.17892	0.62314	0.63396					
Diferencia		-0.24017	0.00403		-0.24808	-0.23226	-59.55	0.00	4747.7	1.56	grande
Portero	1,784	0.38838	0.00294	0.12421	0.38261	0.39415					
Def. lateral	3,627	0.66424	0.00304	0.18332	0.65827	0.67020					
Diferencia		-0.27586	0.00423		-0.28416	-0.26756	-65.18	0.00	4892.8	1.76	grande
Portero	1,784	0.38838	0.00294	0.12421	0.38261	0.39415					
Med. Central	6,611	0.69119	0.00248	0.20138	0.68634	0.69605					
Diferencia		-0.30282	0.00384		-0.31035	-0.29528	-78.76	0.00	4590	1.81	grande
Portero	1,784	0.38838	0.00294	0.12421	0.38261	0.39415					
Med. Lateral	3,846	0.60365	0.00340	0.21054	0.59700	0.61031					
Diferencia		-0.21528	0.00449		-0.22408	-0.20647	-47.93	0.00	5323	1.25	grande
Portero	1,784	0.38838	0.00294	0.12421	0.38261	0.39415					
Delantero	3,279	0.56436	0.00397	0.22749	0.55658	0.57215					
Diferencia		-0.17599	0.00494		-0.18568	-0.16630	-35.61	0.00	5063	0.96	moderado
Def. Central	4,203	0.62855	0.00276	0.17892	0.62314	0.63396					
Def. lateral	3,627	0.66424	0.00304	0.18332	0.65827	0.67020					
Diferencia		-0.03569	0.00411		-0.04374	-0.02763	-8.686	0.00	7605.7	0.20	trivial
Def. Central	4,203	0.62855	0.00276	0.17892	0.62314	0.63396					
Med. Central	6,611	0.69119	0.00248	0.20138	0.68634	0.69605					
Diferencia		-0.06265	0.00371		-0.06991	-0.05538	-16.89	0.00	9699.5	0.33	pequeño
Def. Central	4,203	0.62855	0.00276	0.17892	0.62314	0.63396					
Med. Lateral	3,846	0.60365	0.00340	0.21054	0.59700	0.61031					
Diferencia		0.02489	0.00438		0.01632	0.03347	5.6898	0.00	7579.7	0.13	trivial
Def. Central	4,203	0.62855	0.00276	0.17892	0.62314	0.63396					
Delantero	3,279	0.56436	0.00397	0.22749	0.55658	0.57215					
Diferencia		0.06418	0.00484		0.05470	0.07367	13.268	0.00	6099.1	0.31	pequeño
Def. lateral	3,627	0.66424	0.00304	0.18332	0.65827	0.67020					
Med. Central	6,611	0.69119	0.00248	0.20138	0.68634	0.69605					
Diferencia		-0.02696	0.00392		-0.03465	-0.01926	-6.869	0.00	8077	0.14	trivial
Def. lateral	3,627	0.66424	0.00304	0.18332	0.65827	0.67020					
Med. Lateral	3,846	0.60365	0.00340	0.21054	0.59700	0.61031					
Diferencia		0.06058	0.00456		0.05164	0.06952	13.286	0.00	7426.1	0.31	pequeño
Def. lateral	3,627	0.66424	0.00304	0.18332	0.65827	0.67020					
Delantero	3,279	0.56436	0.00397	0.22749	0.55658	0.57215					
Diferencia		0.09987	0.00500		0.09006	0.10968	19.955	0.00	6297	0.48	pequeño

Two-sample Welch's t test with unequal variances

Ho: Diferencia = 0

Centralidad In degree

Ha: Diferencia != 0

Grupo	Obs	Media	Std. Error	Std. Dev	[Intervalo de confianza (95%)]		t	p-value	GL Welch	Cohen's d	Tamaño del efecto
Med. Central	6,611	0.69119	0.00248	0.20138	0.68634	0.69605					
Med. Lateral	3,846	0.60365	0.00340	0.21054	0.59700	0.61031					
Diferencia		0.08754	0.00420		0.07930	0.09578	20.831	0.00	7751.8	0.42	pequeño
Med. Central	6,611	0.69119	0.00248	0.20138	0.68634	0.69605					
Delantero	3,279	0.56436	0.00397	0.22749	0.55658	0.57215					
Diferencia		0.12683	0.00468		0.11765	0.13601	27.091	0.00	5882.3	0.59	pequeño
Med. Lateral	3,846	0.60365	0.00340	0.21054	0.59700	0.61031					
Delantero	3,279	0.56436	0.00397	0.22749	0.55658	0.57215					
Diferencia		0.03929	0.00523		0.02904	0.04953	7.5181	0.00	6748.4	0.18	trivial

Fuente: Elaboración propia en base a datos de UEFA. *Tests de Welch y tamaño del efecto por Cohen entre grupos para la centralidad in degree.*

Tabla 11. Test de Welch y tamaño del efecto de Cohen para la centralidad *out degree*

Two-sample Welch's t test with unequal variances

Ho: Diferencia = 0

Centralidad Out degree

Ha: Diferencia != 0

Grupo	Obs	Media	Std. Error	Std. Dev	[Intervalo de confianza (95%)]		t	p-value	GL Welch	Cohen's d	Tamaño del efecto
Portero	1,784	0.51224	0.00215	0.09079	0.50802	0.51645					
Def. Central	4,203	0.61617	0.00252	0.16338	0.61123	0.62111					
Diferencia		-0.10393	0.00331		-0.11043	-0.09744	-31.38	0.00	5582.43	0.79	moderado
Portero	1,784	0.51224	0.00215	0.09079	0.50802	0.51645					
Def. lateral	3,627	0.66340	0.00322	0.19393	0.65709	0.66972					
Diferencia		-0.15117	0.00387		-0.15876	-0.14358	-39.04	0.00	5399.7	1.00	moderado
Portero	1,784	0.51224	0.00215	0.09079	0.50802	0.51645					
Med. Central	6,611	0.68526	0.00252	0.20527	0.68031	0.69021					
Diferencia		-0.17302	0.00332		-0.17952	-0.16652	-52.18	0.00	6674.11	1.09	moderado
Portero	1,784	0.51224	0.00215	0.09079	0.50802	0.51645					
Med. Lateral	3,846	0.54013	0.00345	0.21383	0.53337	0.54689					
Diferencia		-0.02790	0.00406		-0.03586	-0.01993	-6.866	0.00	5594.52	0.17	trivial
Portero	1,784	0.51224	0.00215	0.09079	0.50802	0.51645					
Delantero	3,279	0.46036	0.00372	0.21312	0.45306	0.46765					
Diferencia		0.05188	0.00430		0.04345	0.06031	12.071	0.00	4841.01	0.32	pequeño
Def. Central	4,203	0.61617	0.00252	0.16338	0.61123	0.62111					
Def. lateral	3,627	0.66340	0.00322	0.19393	0.65709	0.66972					
Diferencia		-0.04723	0.00409		-0.05525	-0.03922	-11.55	0.00	7124.15	0.26	pequeño
Def. Central	4,203	0.61617	0.00252	0.16338	0.61123	0.62111					
Med. Central	6,611	0.68526	0.00252	0.20527	0.68031	0.69021					
Diferencia		-0.06909	0.00357		-0.07608	-0.06210	-19.37	0.00	10286	0.37	pequeño
Def. Central	4,203	0.61617	0.00252	0.16338	0.61123	0.62111					
Med. Lateral	3,846	0.54013	0.00345	0.21383	0.53337	0.54689					
Diferencia		0.07604	0.00427		0.06766	0.08441	17.804	0.00	7177.99	0.40	pequeño
Def. Central	4,203	0.61617	0.00252	0.16338	0.61123	0.62111					
Delantero	3,279	0.46036	0.00372	0.21312	0.45306	0.46765					
Diferencia		0.15582	0.00449		0.14700	0.16463	34.666	0.00	5992.02	0.82	moderado
Def. lateral	3,627	0.66340	0.00322	0.19393	0.65709	0.66972					
Med. Central	6,611	0.68526	0.00252	0.20527	0.68031	0.69021					
Diferencia		-0.02186	0.00409		-0.02988	-0.01384	-5.342	0.00	7832.51	0.11	pequeño
Def. lateral	3,627	0.66340	0.00322	0.19393	0.65709	0.66972					
Med. Lateral	3,846	0.54013	0.00345	0.21383	0.53337	0.54689					
Diferencia		0.12327	0.00472		0.11402	0.13252	26.129	0.00	7461.66	0.60	moderado
Def. lateral	3,627	0.66340	0.00322	0.19393	0.65709	0.66972					
Delantero	3,279	0.46036	0.00372	0.21312	0.45306	0.46765					
Diferencia		0.20305	0.00492		0.19340	0.21270	41.258	0.00	6654.38	1.00	moderado

Two-sample Welch's t test with unequal variances

Ho: Diferencia = 0

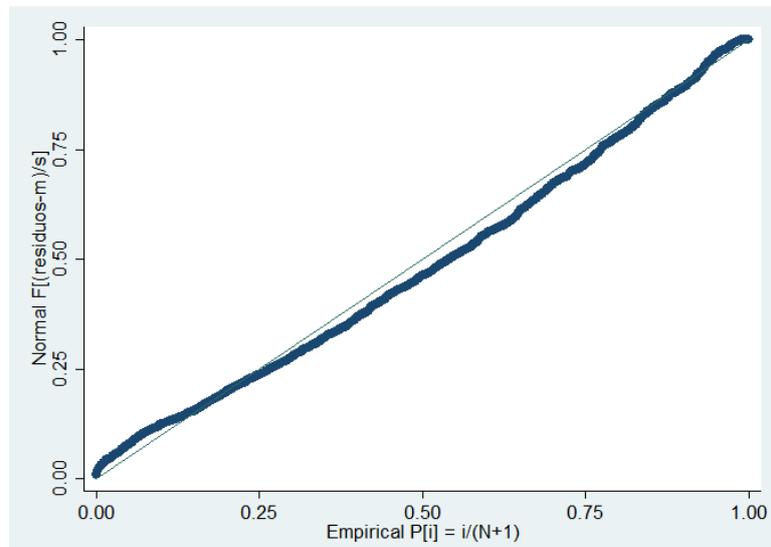
Centralidad Out degree

Ha: Diferencia != 0

Grupo	Obs	Media	Std. Error	Std. Dev	[Intervalo de confianza (95%)]		t	p-value	GL Welch	Cohen's d	Tamaño del efecto
Med. Central	6,611	0.68526	0.00252	0.20527	0.68031	0.69021					
Med. Lateral	3,846	0.54013	0.00345	0.21383	0.53337	0.54689					
Diferencia		0.14513	0.00427		0.13675	0.15350	33.961	0.00	7774.94	0.69	moderado
Med. Central	6,611	0.68526	0.00252	0.20527	0.68031	0.69021					
Delantero	3,279	0.46036	0.00372	0.21312	0.45306	0.46765					
Diferencia		0.22490	0.00450		0.21609	0.23372	50.009	0.00	6326.09	1.07	moderado
Med. Lateral	3,846	0.54013	0.00345	0.21383	0.53337	0.54689					
Delantero	3,279	0.46036	0.00372	0.21312	0.45306	0.46765					
Diferencia		0.07978	0.00507		0.06983	0.08972	15.725	0.00	6954.96	0.37	pequeño

Fuente: Elaboración propia en base a datos de UEFA. *Tests de Welch y tamaño del efecto por Cohen entre grupos para la centralidad out degree.*

Figura 11. Gráfico de probabilidad normal para los residuos estimados



Fuente: Elaboración propia. Software:Stata. “La forma de interpretar este gráfico es la siguiente: si los puntos se aproximan al comportamiento lineal señalado en el gráfico, se puede considerar que la función empírica de la distribución acumulada es similar a la teórica, y por tanto se comporta “normalmente”. Si los puntos se alejan a la línea recta, la variable se aleja de una distribución normal. En el gráfico P-P que ofrece Stata la distribución acumulada de la variable empírica se ubica sobre el eje x mientras que la distribución acumulada teórica normal sobre el eje y” (Peón, 2010).

Tabla 12. Matriz de correlaciones

	residuos	Goles_~s	Id_del	Bt_del	Bt_mc	Bt_ml	Bt_def_l	Bt_def_c	Id_def_c
residuos	1.0000								
Goles_annot~s	0.8903	1.0000							
Id_del	0.0468	0.1492	1.0000						
Bt_del	0.0239	0.1518	0.3387	1.0000					
Bt_mc	0.0285	0.2870	-0.0886	-0.0675	1.0000				
Bt_ml	0.0250	0.0596	-0.0979	-0.0514	-0.1279	1.0000			
Bt_def_l	0.0241	0.0303	-0.0963	-0.0736	-0.1300	-0.1006	1.0000		
Bt_def_c	0.0249	0.0224	0.0021	-0.1232	-0.1219	-0.1126	-0.1133	1.0000	
Id_def_c	0.0665	0.0413	0.0744	0.1845	-0.1382	0.0979	-0.0357	0.0415	1.0000
Id_por	0.0445	-0.0156	0.0674	0.0287	-0.0314	0.0186	-0.0607	-0.2369	0.1337
local_si	0.0137	0.1654	0.0428	0.0440	0.0653	0.0066	-0.0237	0.0042	0.0186
eq_gde	0.0084	0.2211	0.1189	0.0754	0.0046	0.0812	0.0131	0.0139	0.2085
rival_gde	0.0084	-0.1944	-0.0365	-0.0664	-0.0147	-0.0585	0.0053	0.0496	-0.1279

	Id_por	local_si	eq_gde	rival_~e
Id_por	1.0000			
local_si	0.0255	1.0000		
eq_gde	0.0318	-0.0011	1.0000	
rival_gde	-0.0483	-0.0141	-0.0099	1.0000

Fuente: Elaboración propia. Software: Stata. Se verifica que no existe correlación entre las variables independientes del modelo y los residuos.

Tabla 13. Test de Breusch – Pagan y Cook – Weisberg

```

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: fitted values of Goles_annotados

chi2(1)      = 149.58
Prob > chi2  = 0.0000
    
```

Fuente: Elaboración propia. Software: Stata. El hetttest realiza la prueba de Breusch-Pagan (1979) y Cook-Weisberg (1983) para la heterocedasticidad lineal. Contrasta la hipótesis nula de que las varianzas de error son todas iguales.

Tabla 14. Prueba de inflación de varianza

Variable	VIF	1/VIF
Bt_del	1.21	0.828371
Id_del	1.18	0.848442
Bt_def_c	1.17	0.852284
Id_def_c	1.15	0.871876
Bt_mc	1.12	0.893994
Id_por	1.10	0.905087
Bt_ml	1.10	0.907870
Bt_def_l	1.09	0.914085
eq_gde	1.07	0.933643
rival_gde	1.03	0.974379
local_si	1.01	0.990013
Mean VIF	1.11	

Fuente: Elaboración propia. Software:Stata. Para esta prueba se recomiendan valores menores que 10 para mostrar que no hay multicolinealidad, lo cual se verifica.

Tabla 15. Estadísticos descriptivos de las variables incluidas en el modelo

Variable	N° obs.	Media	Desvío sd.	Mín.	Máx.
<i>Id_{DE}</i>	1,746	0.592	0.182	0.05	1
<i>Bt_{DE}</i>	1,746	0.049	0.020	0.001	0.198
<i>Bt_{MC}</i>	1,746	0.082	0.039	0.0013	0.379
<i>Bt_{ML}</i>	1,739	0.072	0.039	0.0005	0.253
<i>Bt_{DL}</i>	1,741	0.079	0.045	0.0010	0.340
<i>Bt_{DC}</i>	1,746	0.073	0.040	0.0003	0.317
<i>Id_{DC}</i>	1,746	0.631	0.130	0.1282	1
<i>Id_{PO}</i>	1,746	0.387	0.120	0.01	0.833
<i>Local</i>	1,746	0.496	0.500	0	1
<i>Equipo grande</i>	1,746	0.274	0.446	0	1
<i>Rival grande</i>	1,746	0.274	0.446	0	1

Fuente: Elaboración propia en base a datos de UEFA. *Estadísticos descriptivos de las variables incluidas en el modelo.* Nota1: los datos de las medidas de centralidad refieren al promedio por posición por partido/equipo. Nota2: las variables *Bt_{ML}* y *Bt_{DC}* poseen menos observaciones porque se jugaron partidos con jugadores en esas posiciones cuya posición principal es distinta a esta.

Tabla 16. Estimación de la relación entre goles anotados y medidas de red y dos variantes

Variables independientes	<u>Modelo original</u>			<u>Modelo sin pases intentados</u>			<u>Robustez de resultados</u>				
							<u>versión 1</u>		<u>Versión 2</u>		
	β	t		β	t		β	t	β	t	
<i>Id DE</i>	0.5284	2.75	***	0.653	3.36	***	0.712	***	0.726	3.49	***
<i>Bt DE</i>	2.7273	3.05	***	4.325	4.91	***	4.52	***	4.67	4.84	***
<i>Bt MC</i>	9.6892	13.14	***	10.02	13.36	***	10.49	***	10.07	12.88	***
<i>Bt ML</i>	2.4316	3.16	***	2.951	3.76	***	3.14	***	3.35	4.06	***
<i>Bt DL</i>	1.8999	3.13	***	2.301	3.71	***	2.38	***	1.922	2.95	***
<i>Bt DC</i>	2.4409	3.04	***	2.329	2.88	***	2.54	***	1.563	1.84	***
<i>Id DC</i>	-1.8374	-7	***	-0.839	-3.93	***	-0.788	***	-0.451	-2.04	**
<i>Id PO</i>	-0.7344	-3.06	***	-0.5198	-2.15	***	-0.451	**	-0.6866	-2.69	***
<i>Local</i>	0.3463	5.61	***	0.358	5.69	***	-	-	-	-	
<i>Equipo grande</i>	0.4796	5.93	***	0.6385	8.13	***	0.6296	***	-	-	
<i>Rival grande</i>	-0.5172	-8.42	***	-0.615	-9.94	***	-0.614	***	-	-	
<i>Pases intentados</i>	0.00583	6.67	***	-	-	-	-	-	-	-	
Número de obs.	1564			1564			1564		1564		
Estadístico F	200.82			210.29			226.67		258.97		
Prob > F	0			0			0		0		
R-cuadrado	0.6526			0.6395			0.632		0.5966		
Raíz del error cuad. med.	1.2141			1.2367			1.2489		1.3068		

Fuente: Elaboración propia. Nota: Niveles de significación a) *** < 0,01; ** < 0,05. b) El modelo fue estimado por el método de Mínimos Cuadrados Ordinarios.