



Escuela Técnica Superior  
de Ingeniería Informática

Grado en Matemáticas

Curso 2022-2023

Trabajo Fin de Grado

**ANÁLISIS DE MERCADOS DE VALORES  
EMPLEANDO REDES COMPLEJAS FUNCIONALES**

Autor: Lucía Salgado Barrio

Tutores: Miguel Romance del Río

David González de la Aleja Gallego



# Agradecimientos

Una vez llegado el momento de cerrar esta larga etapa, me gustaría agradecer brevemente a todas las personas que han estado a mi lado durante estos años y que han hecho que este camino sea mucho más fácil.

A mis padres y a mi hermano, gracias por haber confiado en mí más que yo misma y por darme la oportunidad de cumplir todos mis sueños. Sin vosotros hoy no estaría donde estoy, sois mis mayores referentes en la vida.

A mis amigas de Zamora, quienes me han visto crecer, por haberme demostrado que no importa la distancia que nos separe para saber que al volver nada ha cambiado.

A todos mis amigos del Colegio Mayor Fernando Abril Martorell, que se convirtieron en mi segunda familia durante los años que allí estuve, siempre os recordaré como casa. También quisiera acordarme de todas las personas que formaron parte de mi Erasmus en Vaasa (Finlandia), me llevo conmigo la experiencia más bonita y enriquecedora para siempre.

A todos los profesores que han pasado por mi vida y tanto me han enseñado, desde los profesores del colegio Arias Gonzalo, del IES Claudio Moyano y, por supuesto, de la Universidad Rey Juan Carlos. Valoro mucho la labor tan importante que hacéis.

A mis compañeros y amigos del doble grado: Ramón, Dani, Alberto, Amelia, Marta y Javi, gracias por sacarme una sonrisa cada día y por ser un equipo para todo. Sin duda sois lo mejor que me llevo de la universidad. También al resto de mis amigos de la universidad: Merche, Ana, Mario, Marta Villaplana, Marta Casanova, Daniel Recuero y Daniel López, con quienes tantos momentos he compartido tanto fuera como dentro de la universidad.

Lucía Salgado Barrio



# Resumen

En este trabajo se analiza cuál ha sido el impacto de la pandemia en el mercado financiero español y, más concretamente, en las empresas que forman parte del IBEX 35, empleando técnicas de redes complejas funcionales construidas a partir de datos reales de su cotización en la bolsa.

Con el objetivo de detectar y analizar las diferentes comunidades que se crean entre las empresas analizadas, se ha analizado el coeficiente de correlación de Pearson entre los precios de cierre de las diferentes empresas creando una red funcional asociada a este coeficiente para después aplicar el algoritmo Girvan-Newman en dos períodos de tiempo diferenciados: el período previo a la pandemia (2018-2020) y el período tras el estallido de la pandemia (2020-2022).

Una vez detectadas las comunidades que se forman para cada uno de los períodos, se compararán los resultados para así poder analizar cómo ha afectado el COVID-19 en la interdependencia de las principales compañías de la economía española.

## Palabras clave:

- Teoría de grafos
- Redes Complejas funcionales
- Mercado de valores
- Precio de cierre
- Coeficiente de correlación de Pearson
- Algoritmo Girvan-Newman
- Serie temporal
- Comunidades
- Centralidad



# Índice de contenidos

Índice de tablas	X
Índice de figuras	XII
<b>1. Introducción</b>	<b>1</b>
1.1. Contexto y alcance . . . . .	1
1.2. Estructura del documento . . . . .	1
<b>2. Objetivos</b>	<b>3</b>
2.1. Objetivos generales . . . . .	3
2.2. Objetivos específicos . . . . .	4
<b>3. Contenidos principales</b>	<b>6</b>
3.1. Grafos y redes complejas . . . . .	7
3.2. Conceptos de estadística . . . . .	16
3.3. Comunidades y algoritmo de Girvan-Newman . . . . .	20
3.3.1. Comunidades . . . . .	20
3.3.2. Algoritmo de Girvan-Newman . . . . .	23
3.4. Conceptos económicos . . . . .	25
<b>4. Desarrollo del trabajo</b>	<b>31</b>
4.1. Elección de umbral óptimo . . . . .	35
<b>5. Resultados</b>	<b>39</b>
5.1. Período pre-pandemia (2018-2020) . . . . .	40
5.2. Período post-pandemia (2020-2022) . . . . .	46
<b>6. Conclusiones y trabajos futuros</b>	<b>52</b>
6.1. Conclusiones . . . . .	52
6.2. Trabajos futuros . . . . .	55
<b>Bibliografía</b>	<b>55</b>
<b>Apéndices</b>	<b>59</b>







# Índice de tablas

4.1. Listado de empresas del Ibex 35 analizadas. . . . .	33
--	----



# Índice de figuras

3.1. Ciudad de Königsberg. . . . .	7
3.2. Grafo asociado a los puentes de Königsberg. . . . .	8
3.3. Grafo dirigido y grafo no dirigido . . . . .	9
3.4. Grafo conexo y grafo no conexo . . . . .	11
3.5. Ejemplo de una serie temporal. . . . .	19
3.6. Red asociada al club de karate de Zachary . . . . .	21
3.7. Aplicación del algoritmo Girvan-Newman. . . . .	24
4.1. Función para la toma de datos en Python. . . . .	35
4.2. Bloque de código para determinar el umbral óptimo del coeficiente de correlación de Pearson. . . . .	36
4.3. Ejemplo de elección del umbral óptimo para un período dado . . .	37
5.1. Umbral óptimo asociado al período 2018-2020 . . . . .	40
5.2. Grafo asociado al período 2018-2020 . . . . .	41
5.3. Componente conexa asociada al período 2018-2020 . . . . .	42
5.4. Componente conexa asociada al período 2018-2020 . . . . .	43
5.5. Dendrograma asociado al período 2020-2022. . . . .	44
5.6. Umbral óptimo asociado al período 2020-2022. . . . .	46
5.7. Grafo asociado al período 2020-2022 . . . . .	47
5.8. Componente conexa asociada al período 2020-2022 . . . . .	48
5.9. Distribución de la red compleja para el período 2020-2022 . . . . .	49
5.10. Dendrograma asociado al período 2020-2022 . . . . .	50



# 1

## Introducción

### 1.1. Contexto y alcance

Conocer el funcionamiento de los mercados de valores es esencial para la sociedad de cualquier país, ya que, entre otros motivos, es un indicador importante de la salud y la dirección de una economía [1][2][3][4]. Además, estudiar la interconexión y cómo están relacionadas las empresas que conforman un índice bursátil pueden ofrecer información valiosa sobre el funcionamiento y la estabilidad del mercado financiero. Por lo tanto, el estudio del IBEX 35 a través de grafos puede ayudar a comprender y analizar la dinámica y las interacciones entre las acciones que componen este índice bursátil en España.

La identificación de patrones o comunidades de acciones similares gracias al análisis de la estructura del IBEX 35 que se puede realizar aplicando grafos es de gran utilidad para los inversores, pudiendo adelantarse en la toma de decisiones de inversión haciendo uso de las tendencias o los ciclos que se hayan identificado a través del tiempo. Además, a través del estudio de las acciones históricamente podemos identificar las empresas líderes del mercado y conocer qué conexiones tienen y cómo afectan a los rendimientos de otras acciones, así como el impacto que tienen estas empresas claves en la dinámica del propio índice bursátil.

### 1.2. Estructura del documento

La estructura de este trabajo se encuentra organizada en 6 capítulos.

En primer lugar, tras el apartado de Introducción previamente expuesto, en el apartado de Objetivos explicaremos cuáles son los objetivos principales del estudio, es decir, qué es lo que se pretende conseguir al finalizar el trabajo. Además, explicaremos también los objetivos específicos del estudio.

A continuación, pasaremos a explicar los contenidos teóricos principales del trabajo, explicando con detalle todos los conceptos necesarios para el buen entendimiento del mismo. El apartado de contenidos principales se ha dividido, a su vez, en tres grandes secciones. Una primer sección trata sobre teoría de grafos, explicando desde los inicios de la teoría de grafos hasta conceptos más complejos y concretos relacionados con las comunidades en grafos. Posteriormente, se introducirán los conceptos más relevantes del ámbito de la estadística que se emplean para construir redes funcionales, haciendo especial énfasis en el coeficiente de correlación de Pearson, empleándolo para crear nuestros grafos iniciales. A continuación, explicaremos el algoritmo Girvan-Newman, uno de los algoritmos más utilizados en la teoría de grafos para la detección de comunidades. Finalizaremos este apartado hablando sobre aspectos económicos que se consideran necesarios para el lector con el fin de entender la aplicación de nuestro algoritmo al mundo de las finanzas y poder interpretar los resultados de manera correcta, teniendo un marco teórico sólido con los principales conceptos matemáticos y económicos explicados.

Una vez explicado el marco teórico que sustenta este trabajo, se pasa a explicar con profundidad el desarrollo práctico del trabajo. En el capítulo 4 se explicarán los pasos que se han seguido a lo largo del estudio, desde la ingesta de datos desde Python sobre los valores de múltiples empresas españolas, hasta las comunidades que se han creado para cada periodo de tiempo analizado tras la aplicación del algoritmo Girvan-Newman.

Tras el desarrollo del trabajo, se expondrán los principales resultados que se han obtenido en nuestro estudio al aplicar el algoritmo Girvan-Newman. Más concretamente, se explican los resultados y las comunidades detectadas para los distintos periodos de tiempo que se han considerado en el análisis, apoyándonos en todo momento en recursos visuales.

Por último, en el apartado de Conclusiones se expondrán los principales resultados obtenidos mediante la aplicación de la metodología explicada. Este apartado es de gran importancia ya que, debido a que en todo momento se están empleando datos reales, puede ser de gran utilidad entender correctamente las conexiones y la interdependencia de las principales empresas españolas para así conocer cuáles son las tendencias que cabría esperar en un futuro.

Para concluir, hay un apartado de Bibliografía donde se encuentran todas las fuentes y artículos consultados para la elaboración de este trabajo. Además, en el apartado de Apéndices se incluye el código Python empleado para la extracción de datos de Yahoo Finance, así como la aplicación del algoritmo Girvan-Newman.

# 2

## Objetivos

Tras finalizar el estudio propuesto en el presente Trabajo Fin de Grado, se esperan conseguir una serie de objetivos relacionados con algunos de los conceptos adquiridos durante el Grado, que hemos clasificado en dos apartados: objetivos generales y objetivos específicos.

### 2.1. Objetivos generales

El objetivo principal de este estudio es la identificación de comunidades entre diferentes empresas españolas en los diversos períodos de tiempo considerados a través del algoritmo Girvan-Newman calculado sobre una red funcional construida empleando correlaciones entre cotizaciones de valores. Se pretende, por lo tanto, analizar cómo influye el ciclo económico en el que se sitúa la economía española en las relaciones existentes entre las diferentes empresas, de tal forma que se puedan extraer conclusiones sobre las tendencias que siguen estas empresas en la bolsa y se puedan sacar patrones de comportamiento a la hora de invertir en bolsa. Además, se pretende conocer en detalle qué relación hay entre diferentes empresas de la economía española con el fin de conocer si estas relaciones son positivas o negativas y, en caso de que una empresa sufra una caída en bolsa, por ejemplo, saber cómo afectará al resto de empresas del mercado de valores español.



## 2.2. Objetivos específicos

Entre los objetivos específicos que se van a abordar en este estudio, podemos destacar los siguientes:

1. Familiarizarse y ampliar los conceptos y fundamentos teóricos relacionados con la teoría de grafos, y más en concreto la construcción de redes funcionales y la detección de comunidades en redes empleando el algoritmo Girvan-Newman.
2. Recopilar y procesar los datos financieros y económicos necesarios de las empresas analizadas para construir la red de interacciones en Python.
3. Implementar en Python una función que cargue y procese los datos financieros de las empresas estudiadas a través de la plataforma Yahoo Finance.
4. Desarrollar una función en Python que construya la red de interacciones entre las empresas a partir de los datos procesados, utilizando una matriz de adyacencia para construir la red.
5. Realizar un análisis comparativo de las comunidades obtenidas utilizando el coeficiente de correlación de Pearson para establecer las conexiones entre las empresas.
6. Construir la red de interacciones entre las empresas del IBEX 35 utilizando los datos recopilados y aplicar el algoritmo Girvan-Newman para detectar comunidades.
7. Evaluar y analizar las características de las comunidades detectadas.
8. Aplicar funciones en Python para visualizar las comunidades encontradas en forma de gráficos, ayudando a comprender mejor la estructura de la red y las relaciones entre las empresas.
9. Analizar la formación de comunidades a través de los dendogramas creados.
10. Investigar la estabilidad o cambio en las comunidades a lo largo del tiempo mediante el análisis de diferentes períodos y evaluar si existen patrones o tendencias.



# 3

## Contenidos principales

En esta sección se explicarán todos los conceptos teóricos que sean necesarios para el entendimiento del estudio completo. Se empieza haciendo una revisión bibliográfica y explicando los conceptos claves relacionados con la teoría de grafos y redes complejas, como sus propiedades y las características que se usan en el desarrollo de este trabajo. Además, se explican conceptos estadísticos como el coeficiente de correlación de Pearson, la covarianza, etc. que juegan un papel clave en el estudio, ya que serán los conceptos que nos proporcionen información acerca de cómo se relacionan las empresas del IBEX 35 y la intensidad de esa relación, es decir, si están altamente relacionadas o no. Por último, se definen los conceptos básicos y necesarios sobre el mercado de valores, haciendo especial hincapié en el IBEX 35 y su funcionamiento.

## 3.1. Grafos y redes complejas

En nuestro caso, las diferentes empresas del IBEX 35 están representadas como nodos o vértices de la red compleja y dos nodos están unidos por una arista si el grado de correlación entre sus cotizaciones bursátiles es mayor que un valor previamente establecido (en nuestro caso, este valor tendrá que ser mayor que 0). Así, la red compleja creada ofrece información sobre la interdependencia de las empresas de la bolsa española, con los vínculos de la red representando el grado de correlación entre las empresas.

La historia de la teoría de grafos se remonta a la resolución del problema de los puentes de Königsberg, conocido como el primer resultado de la teoría de grafos, siendo Leonhard Euler quien propuso un teorema como solución a este problema. Tras este resultado, la teoría de grafos sirvió para resolver problemas de muy variados ámbitos, como pueden ser las leyes de Kirchhoff en redes eléctricas y el problema de enumeración de los isómeros en el mundo de la química. El problema de los puentes de Königsberg se remonta a la antigua ciudad prusiana de Königsberg, más adelante conocida y renombrada como Kaliningrado, que contaba con siete puentes por los que pasaba el río Pregolia, tal y como se muestra en la siguiente figura.

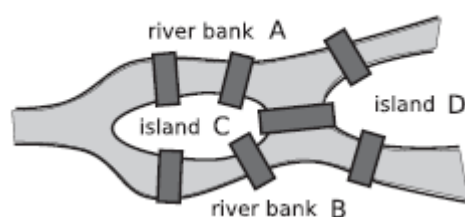


Figura 3.1: Ciudad de Königsberg.

El problema que se planteaba era si existía alguna posibilidad de encontrar un camino que atravesase una única vez cada puente y que se regresase al punto de partida. Una manera de resolver el problema sería elegir un punto de partida y, teniendo en cuenta todas las posibles rutas, ir comprobando una a una esas rutas. En caso de que ninguna de las rutas anteriores cumpla con el objetivo, habría que hacer el mismo procedimiento cambiando el punto de partida y así sucesivamente hasta cubrir todas las posibles rutas o encontrar una que satisfaga la condición impuesta. Este método de resolución no es el óptimo, puesto que no ofrece una solución de manera genérica para cualquier problema que se plantee de este tipo, sino que en cada situación concreta habría que repetir el procedimiento anteriormente mencionado.

En 1736, Leonhard Euler propuso una solución que valía no solo para el problema de los puentes de Königsberg, sino para cualquier problema de este tipo.

En primer lugar, introdujo la noción y el concepto del grafo y llegó a la conclusión de que el problema solamente dependía en las conexiones que existían entre las islas y las riberas. En la siguiente imagen podemos ver el grafo asociado al problema de los siete puentes de Königsberg, representando los nodos las islas o las riberas y las aristas reflejando los puentes. Si englobamos toda la ribera A de la figura 3.1 en un solo punto, así como la ribera B y las islas C y D en un solo nodo respectivamente, de tal manera que toda la información relevante se reducen a cuatro nodos, correspondientes a las riberas A y B y las islas C y D y siete aristas, que representan los siete puentes de la ciudad de Königsberg.

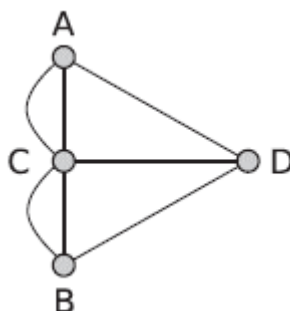


Figura 3.2: Grafo asociado a los puentes de Königsberg.

**Teorema de Euler:** Un grafo conexo es Euleriano si y solo si cada uno de los nodos o vértices tiene un número par de vértices. Tiene un camino euleriano del vértice  $i$  al vértice  $j$ ,  $i \neq j$ , si y solo si  $i$  y  $j$  son los únicos vértices con grado impar.

La condición de tener que pasar una única vez por cada puente solo puede satisfacerse si y solo si los vértices con grado impar son cero (dándose cuando el punto inicial y el punto final coinciden) o dos (cuando el punto inicial no es el mismo que el punto final). Como podemos observar en la figura 3.2, el grafo asociado a la ciudad de Königsberg no cumple con las condiciones del teorema debido a que los cuatro nodos o vértices del grafo tienen grado impar, por lo que no pueden existir caminos eulerianos.

Los conceptos que se presentan a continuación pueden consultarse en [5][6][7]

**Definición 1 (Grafo)** Un **grafo**  $G$  es un par ordenado  $G=(V,E)$  que está compuesto por dos elementos  $V$  y  $E$ , donde  $V$  es el conjunto de vértices o nodos del grafo  $G$  y  $E$  representa el conjunto de enlaces o aristas que unen estos nodos.

Existen dos tipos de grafos en función de la relación que representen: grafos dirigidos y grafos no dirigidos. En primer lugar, explicaremos los grafos no dirigidos, que se utilizan fundamentalmente para representar relaciones bidireccionales

y simétricas ya que la información que contienen fluye en ambas direcciones de forma equivalente.

**Definición 2 (Grafo no dirigido)** *Un **grafo no dirigido**,  $G=(V,E)$ , está compuesto por dos elementos,  $V \neq \emptyset$ , y  $E$ . Los elementos de  $V \equiv \{v_1, v_2, \dots, v_n\}$  son distintos y se denominan **nodos** (o **vértices**) del grafo  $G$ . Los elementos de  $E \equiv \{e_1, e_2, \dots, e_k\}$  son pares distintos no ordenados de elementos distintos de  $G$ , y son llamados **enlaces** (o **aristas**, o **líneas**).*

En algunos casos, cuando se quiere representar en un gráfico una relación entre nodos que no es recíproca y que es unidireccional, es útil hablar de grafos dirigidos, es decir, grafos en los que las aristas entre los nodos tienen una dirección determinada.

**Definición 3 (Grafo dirigido)** *Un **grafo dirigido**,  $G=(V,E)$ , está compuesto por dos elementos,  $V \neq \emptyset$ , y  $E$ . Los elementos de  $V \equiv \{v_1, v_2, \dots, v_n\}$  son distintos y se denominan **nodos** (o **vértices**) del grafo  $G$ . Los elementos de  $E \equiv \{e_1, e_2, \dots, e_k\}$  son pares distintos ordenados de elementos distintos de  $G$ , y son llamados **enlaces directos** o **arcos**.*

En un grafo dirigido, un enlace o arco entre un nodo  $i$  y un nodo  $j$  se denota como un par ordenado  $(i, j)$  y se dice que el enlace entra en  $i$  y sale de  $j$ , denotándose como  $e_{ij}$ . Cabe destacar que, al contrario de lo que ocurre con los grafos no dirigidos, en los grafos dirigidos  $e_{ij} \equiv (i, j)$ , es decir, un arco que va de  $i$  hasta  $j$  y, consecuentemente,  $e_{ij} \neq e_{ji}$ , ya que el arco  $(i, j)$  no es igual que el arco  $(j, i)$ , como sí ocurría en los grafos no dirigidos.

En la Figura 3.3 podemos ver un ejemplo gráfico de cómo es un grafo dirigido y un grafo no dirigido.

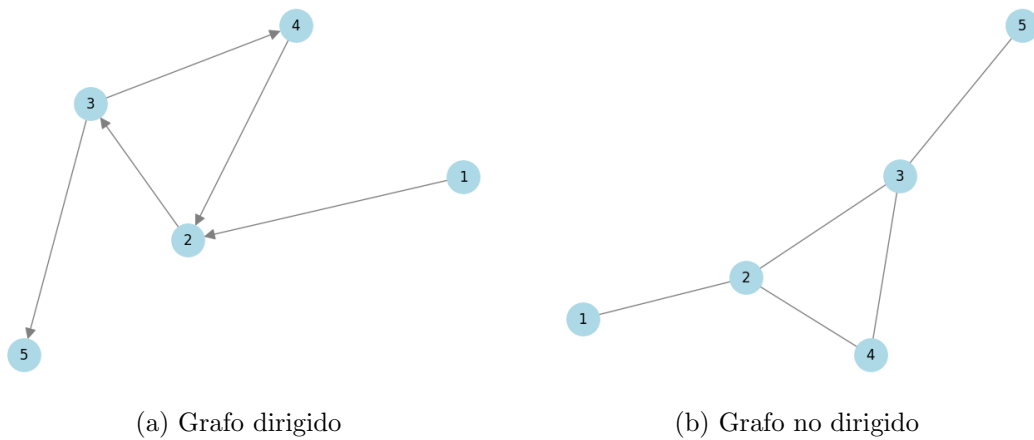


Figura 3.3: Grafo dirigido y grafo no dirigido

El número de vértices  $N \equiv N[G] = |V|$ , donde el símbolo  $|\cdot|$  denota la cardinalidad de un conjunto (es decir, el número de elementos del conjunto), suele denominarse orden de  $G$ . Por otra parte, el número de aristas  $K \equiv K[G] = |E|$  es el tamaño de  $G$ .

En los grafos no dirigidos, cada uno de las aristas o enlaces está definido por un par de nodos,  $i$  y  $j$ , denotados como  $(i, j)$  o  $(j, i)$ . La arista se dice que es incidente a los nodos  $i$  y  $j$  o que une ambos nodos, considerados como los extremos del enlace  $(i, j)$ . Dos nodos unidos por una arista se denominan adyacentes o vecinos.

**Definición 4 (Grado de un nodo)** *El grado  $k_i$  de un nodo  $i$  es el número de aristas incidentes en el nodo.*

En un grafo no dirigido, la lista de grados de los nodos  $\{k_1, k_2, \dots, k_N\}$  se llama secuencia de los grados. El grado medio  $\langle k \rangle$  de un grafo se define como  $\langle k \rangle = N^{-1} \sum_{i=1}^N k_i$  y es igual a  $\langle k \rangle = 2N/K$ .

Por otra parte, en función de cómo estén conectados los elementos del grafo entre sí, se pueden distinguir dos tipos de grafos: grafos conexos y grafos no conexos. La importancia de conocer las interacciones entre los vértices del grafo es importante ya que nos refleja la conectividad y cómo puede estar aislado el grafo. En el caso de los grafos conexos, representan situaciones donde todos los elementos del grafos están conectados de alguna manera, siendo de gran utilidad en problemas como el análisis de redes sociales, la búsqueda del camino más corto entre dos puntos o para aplicar ciertos algoritmos de búsqueda. A su vez, los grafos no conexos también proporcionan información significativa al representar grafos cuyos elementos no están conectados todos entre sí y se utilizan, entre otras cosas, para el estudio de nodos aislados o de subgrafos.

**Definición 5 (Grafo conexo)** *Se dice que un grafo  $G=(V,E)$  es conexo si cada par de nodos o vértices está conectado por un arista, es decir, si para cada par de vértices de  $V$  existe un camino que conecte ambos nodos.*

**Definición 6 (Grafo no conexo)** *Se dice que un grafo  $G=(V,E)$  es no conexo cuando no existe un camino que una un par de vértices del conjunto  $V$ , es decir, cuando no exista una manera de llegar de un vértice a otro.*

En la Figura 3.4, se observa cómo el grafo 3.4a se trata de un grafo conexo, ya que siempre existe un camino para cualesquiera de los vértices que conforman ese grafo. Sin embargo, en el caso del grafo de la Figura 3.4b estamos ante un grafo no conexo, puesto que, de los 6 vértices totales que tiene el grafo, estos están conectados 3 a 3, siendo imposible establecer un camino para unir la mitad de los vértices del grafo.

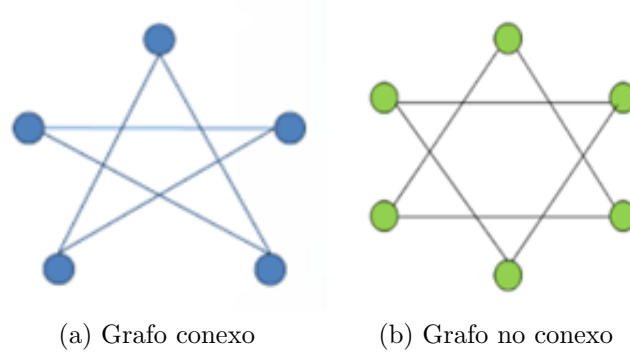


Figura 3.4: Grafo conexo y grafo no conexo

Es posible realizar operaciones básicas con el fin de crear nuevos grafos a partir de los antiguos, como fusionando dos grafos en uno o considerando una sola parte de un grafo dado. La unión de dos grafos puede definirse de la siguiente manera:

**Definición 7 (Unión entre grafos)** Sea  $G_1 = (V_1, E_1)$  y  $G_2 = (V_2, E_2)$ , donde  $V = V_1 \cup V_2$  y  $E = E_1 \cup E_2$ , entonces  $G = (V, E)$  representa la unión de  $G_1$  y  $G_2$  y se denota como  $G = G_1 + G_2$

Además, otro concepto que es clave en nuestro estudio es el de subgrafo, es decir, un grafo que se obtiene tras la eliminación de aristas y vértices del grafo original. Este concepto es muy importante, ya que en muchas ocasiones el estudio de un subgrafo nos permite hacernos una idea y sacar conclusiones sobre la estructura y las relaciones que existen en el grafo original, reduciendo además la complejidad computacional al estar examinando grafos más pequeños.

**Definición 8 (Subgrafo)** Sea  $G = (V, E)$ , un subgrafo de  $G$  se define como un grafo  $G' = (V', E')$  tal que  $V' \subseteq V$  y  $E' \subseteq E$ . En el caso de que  $G'$  contenga todas las aristas o enlaces de  $G$  que unen dos nodos en  $V'$ , entonces  $G'$  se considera un subgrafo inducido o generado por  $V'$ , y se denota como  $G' = G[N']$

Por otra parte, en la teoría de grafos la distancia entre nodos suele ser fundamental ya que se utiliza para la resolución y optimización de un gran número de problemas de diversos campos, como la elección de la ruta más óptima entre dos vértices del grafo, es decir, el camino más corto entre dos nodos, además de utilizarse en teoría de juegos y análisis de redes. Además, numerosos algoritmos, tanto algoritmos de búsqueda en anchura como algoritmos de búsqueda en profundidad, se fundamentan en esta idea para encontrar soluciones a problemas muy concretos.



**Definición 9 (Camino)** *Un camino es un recorrido en el que ningún nodo se visita más de una vez. El camino más corto - también llamado geodésico - del nodo  $x$  al nodo  $y$  es un paseo de longitud mínima de  $x$  a  $y$ , y en lo sucesivo, se denominará  $\mathbb{P}(x, y)$ .*

En el caso de los grafos no dirigidos, si una secuencia de nodos es un camino, entonces la secuencia inversa de nodos será también un camino al no tener las aristas una dirección determinada. De la definición anterior del camino más corto se puede introducir el concepto de distancia en el caso de un grafo.

**Definición 10 (Distancia entre nodos)** *En el caso de un grafo no dirigido, la distancia entre dos nodos  $x$  e  $y$  equivale a la distancia del camino más corto  $\mathbb{P}(x, y)$  que conecta  $x$  e  $y$ .*

Además, para la resolución de problemas de optimización, planificación, búsqueda del camino más corto a través de algoritmos, etc. es importante conocer el concepto de grafo ponderado o grafo con pesos. Al permitir representar cuantitativamente las relaciones existentes en la red creada, son una herramienta muy potente y valorada para encontrar soluciones a problemas del mundo real.

**Definición 11 (Grafo con pesos)** *Un grafo con pesos o grafo ponderado,  $G^W \equiv (N, L, W)$  consiste en un conjunto de nodos  $N = \{n_1, n_2, \dots, n_N\} \neq \emptyset$ , un conjunto de aristas  $L = \{l_1, l_2, \dots, l_k\}$  y un conjunto de pesos  $W = \{w_1, w_2, \dots, w_k\}$ , siendo los valores de los pesos números reales positivos. Es decir, un grafo ponderado es un grafo en el que cada arista lleva asociado un valor numérico que indica la intensidad de la conexión o relación existente entre los nodos.*

El valor que lleva asociado cada arista puede representar la fuerza de las conexiones sociales que represente un grafo, un coste, una distancia, el grado de correlación entre dos vértices, etc. En nuestro caso, el valor del peso vendrá determinado por el coeficiente de correlación Pearson que existe entre las distintas empresas del IBEX 35 representadas por nodos o vértices.

**Definición 12 (Precio de cierre)** *Sea  $p_{c,i}(t)$  el precio de cierre de una empresa  $i$  en un día  $t$ , se define el cambio logarítmico ('rentabilidad') de una empresa  $i$  en un día  $t$  como:*

$$r_i(t) = \ln \left[ \frac{p_{c,i}(t)}{p_{c,i}(t-1)} \right]. \quad (3.1)$$

**Definición 13 (Fuerza de un nodo)** *Se define la fuerza de un nodo  $s_i(m)$  de un nodo  $i$  en la  $m$ -ésima ventana como la media de los pesos de todas las aristas o ejes conectados o incidentes al nodo  $i$ , es decir:*

$$s_i(m) = \frac{1}{N-1} \sum_{j=1, i \neq j}^N r_{i,j}(m). \quad (3.2)$$

*La fuerza  $s_i$  de un nodo  $i$  es la suma de los pesos de las aristas incidentes al nodo  $i$ , tal que:*

$$s_i(m) = \sum_{j=1}^N w_{ij}, \quad (3.3)$$

*siendo  $N$  el número total de nodos de nuestra red compleja (en nuestro caso,  $N = X$ ).*

*La fuerza de un nodo nos indica el grado en el que un nodo está correlacionado con otro, es decir, mide el grado de correlación entre una empresa del IBEX 35 y otras.*

Para construir la red compleja, uniremos dos nodos o vértices (empresas del IBEX 35) a través de una arista si y solo si el coeficiente de correlación de Pearson de dos nodos es mayor que un determinado valor  $\theta$ .

**Definición 14 (Matriz de adyacencia)** *La matriz de adyacencia  $A$  para la red compleja construida viene dada por:*

$$A_{ij} = \begin{cases} 1, & \text{si } r_{xy} \geq \theta, i \neq j. \\ 0 & \text{en otros casos.} \end{cases}$$

En otras palabras, en el caso de los grafos no dirigidos, las entradas  $A_{ij}$  y  $A_{ji}$  son iguales a 1 si existe una arista  $(i,j)$ , mientras que serán 0 en el resto de casos, por lo que la matriz resultante siempre será una matriz simétrica, es decir, una matriz cuadrada que es igual a su matriz traspuesta.

Además, una de las características más importantes sobre grafos que utilizaremos a lo largo de este estudio es la centralidad de los grafos. Este concepto, que se aplicó originalmente para el caso de las redes sociales, se ha utilizado posteriormente en áreas de distinta índole con grandes resultados.

En el caso de las redes sociales, Alex Bavelas y Harold Leavitt detectaron que, generalmente, los individuos más importantes de una red social estaban situados

en localizaciones estratégicas, es decir, que existía relación entre la posición que ocupa un nodo dentro de un grafo y la importancia de este nodo.

Existen diferentes maneras de medir la centralidad de los grafos, como puede ser a través del grado de un nodo, de tal forma que cuanto mayor sea el grado de un nodo, es decir, cuantas más aristas incidan en un nodo, más fuentes de información tendrá disponibles y más rápido llegará la información al nodo, siendo este nodo más central.

Otra de las maneras que existen para medir la centralidad de un grafo es a través del camino más corto entre los diferentes nodos de un grafo. La centralidad de proximidad ("closeness centrality"), la centralidad de entre nodos ("betweenness centrality") y la centralidad delta ("delta centrality") son algunos de los ejemplos de este tipo de medidas, las cuales necesitan conocer previamente cuáles son los caminos más cortos dado un grafo.

Un algoritmo óptimo para recorrer estructuras como grafos o árboles y explorar sus elementos es el de Búsqueda en Amplitud, también conocido como BFS por sus siglas en inglés (Breadth-First Search). El algoritmo de búsqueda en anchura se centra en un nodo raíz y, a partir de este, el algoritmo explora todos los vecinos de este nodo para, posteriormente, para cada uno de los vecinos se procede a explorar los respectivos vecinos adyacentes hasta que se termina de recorrer el grafo.

Este algoritmo de búsqueda calcula los caminos más cortos desde un nodo original o nodo raíz dado,  $i$ , al resto de nodos del grafo, almacenando las longitudes de los caminos más cortos y, en caso de existir varios caminos con la misma longitud desde un nodo  $i$  a un nodo  $j$ , el algoritmo es capaz de guardar todos ellos. Una manera de reconstruir todos los caminos más cortos es para cada nodo  $j$ , almacenar el conjunto de los primeros vecinos de  $j$  que pertenecen al menos a un camino más corto de  $i$  a  $j$ , llamándose este conjunto los predecesores de  $j$ . Así, visitando recursivamente los predecesores de  $j$  y de sus predecesores, podremos enumerar todos los caminos más cortos de  $i$  a  $j$ .

Se puede demostrar que este algoritmo es óptimo, es decir, que no existe un algoritmo más eficiente que permita visitar todos los nodos o vértices de un grafo empezando por un nodo inicial dado. En el caso de grafos no dirigidos, su complejidad es  $O(N + K)$ , siendo  $N$  el número de vértices o nodos del grafo y  $K$  el número de aristas del grafo, ya que el algoritmo necesita escanear o visitar todas las aristas y los nodos del grafo exactamente una vez.

En lo que a nuestro estudio respecta, nos centraremos en la medida de "betweenness centrality". Es habitual que las interacciones entre dos individuos no adyacentes dependan de los demás individuos, especialmente aquellos que se encuentran en los trayectos intermedios. Por consiguiente, la persona ubicada en el medio puede ejercer cierto control estratégico e influencia sobre los demás.

La centralidad de intermediación busca capturar esta idea: un nodo en un grafo es considerado central si se encuentra en el camino entre muchos otros nodos. La medida más básica de intermediación fue introducida por Linton Freeman en 1977. Dicha medida, que es un indicador de nodos muy centrales en las redes, se basa en la suposición simplificada de que la comunicación solo se realiza a lo largo de las rutas más cortas, y se evalúa mediante el conteo de la cantidad de rutas más cortas entre pares de vértices que atraviesan un nodo específico.

**Definición 15 (Betweenness centrality)** *En un grafo conectado, la centralidad del camino más corto del nodo  $i$  se define como:*

$$c_i^B = \sum_{j=1, j \neq i}^N \sum_{k=1, k \neq i, j}^N \frac{n_{jk}(i)}{n_{jk}}, \quad (3.4)$$

siendo  $n_{jk}$  el número de geodésicas del nodo  $j$  al nodo  $k$ , mientras que  $n_{jk}(i)$  representa el número de geodésicas del nodo  $j$  al nodo  $k$ , conteniendo a su vez el nodo  $i$ . La cantidad normalizada, definida como:

$$C_i^B = \frac{c_i^B}{(N-1)(N-2)} \quad (3.5)$$

que toma valores entre  $[0,1]$ .

En la anterior definición, nótese que el factor  $(N-1)(N-2)$  es el máximo número posible de geodésicas que pasan por un nodo. La cantidad  $C_i^B$  puede tomar el valor máximo, que es 1, cuando el nodo  $i$  pertenezca a todas las geodésicas establecidas entre los pares de vértices.

En muchas ocasiones, los grafos están organizados en comunidades, es decir, los nodos se agrupan de tal manera que aquellos que pertenecen a un mismo grupo están más conectados que dos nodos que pertenecen a grupos diferentes. En los casos en los que esto ocurre, podemos decir que el grafo presenta una estructura de comunidad. Al tener características comunes los nodos que pertenecen a un mismo grupo o comunidad, detectar comunidades dentro de un grafo nos va a proporcionar información relevante sobre cómo funciona ese grafo o red compleja o qué papel juegan esos nodos dentro de la red [5].

Existen diferentes métodos para tratar de encontrar comunidades en un grafo, como pueden ser los métodos clásicos de partición o agrupamiento espectral y el agrupamiento o clustering jerárquico. Sin embargo, un método más reciente para la detección de comunidades en grafos y que cubre las carencias de los dos métodos anteriores es el algoritmo Girvan-Newman, creado por Michelle Girvan y Mark Newman en el año 2002. En contra de lo que proponía el método de clustering jerárquico, en el que se agrupaban nodos similares progresivamente con el fin

de crear comunidades o grupos, el algoritmo de Girvan-Newman propone eliminar progresivamente aristas del grafo original, basándose en la centralidad entre nodos, para así conseguir una división jerárquica. Este método será desarrollado con detalle en la sección 3.3.

## 3.2. Conceptos de estadística

La disciplina de la estadística juega un papel clave en este estudio, ya que conceptos tan básicos como la covarianza serán fundamentales para el desarrollo de este trabajo. La covarianza es una medida de la variabilidad conjunta de dos variables  $X$  e  $Y$ .

**Definición 16 (Soporte de una variable aleatoria discreta)** *Se conoce como soporte o rango de una variable aleatoria  $X$  al conjunto de valores que puede tomar la variable. Este conjunto se denota por  $S_x$  y en el caso de variables aleatorias discretas el soporte es siempre un conjunto numerable.*

**Definición 17 (Función de masa de probabilidad)** *Se conoce la función de masa de probabilidad de una variable aleatoria  $X$  como la función  $f_x : \mathbb{R} \rightarrow (0, 1)$  que devuelve la probabilidad de que  $X$  tome cada uno de sus valores, es decir:  $f_x(k) = P(X = k) = P(\{\omega \in \Omega \mid X(\omega) = k\})$*

*donde  $\mathbb{R}$  representa el conjunto de números reales,  $(0, 1)$  es el intervalo abierto entre 0 y 1,  $X$  es la variable aleatoria,  $\Omega$  es el espacio muestral y  $P$  es la función de probabilidad.*

**Definición 18 (Esperanza de una variable discreta)** *La esperanza, media o valor esperado de una variable aleatoria discreta  $X$  con función de probabilidad  $P(X = k)$  se define como:*

$$E(X) = \sum_{k \in S_x} k \times P(X = k). \quad (3.6)$$

*La esperanza se suele denotar por  $E(X)$ ,  $\mu_x$  o simplemente  $\mu$ .*

**Definición 19 (Covarianza)** *La covarianza entre dos variables aleatorias,  $X$  e  $Y$ , se define como:*

$$Cov(X, Y) = E[(X - E(X)) \times (Y - E(Y))] \quad (3.7)$$

*o, lo que es lo mismo:*

$$Cov(X, Y) = E[(X - \mu_x) \times (Y - \mu_y)]. \quad (3.8)$$

*En cuanto a la interpretación de la covarianza, que mide la variabilidad conjunta de dos variables  $X$  e  $Y$ , se pueden distinguir los siguientes casos:*

- *Si la covarianza es positiva entre ambas variables, se traduce en que existe una alta probabilidad de que valores grandes  $X$  estén asociados con valores grandes de  $Y$  y, a su vez, los valores pequeños de  $X$  vayan asociados a valores pequeños de  $Y$ .*
- *En el caso de que la covarianza sea negativa, existirá una alta probabilidad de que valores grandes de  $X$  se encuentren asociados a valores pequeños de  $Y$  y viceversa.*
- *Por último, si la covarianza entre las variables  $X$  e  $Y$  es 0, no existe ninguna relación de tipo lineal entre las variables y, por tanto, estamos ante variables incorrelacionadas.*

**Definición 20 (Varianza de una variable aleatoria discreta)** *La varianza de una variable aleatoria es una medida de la dispersión de  $X$  alrededor de su centro de gravedad,  $E(X)$ , y se define como:*

$$V(X) = E([X - E(X)]^2). \quad (3.9)$$

*La varianza  $V(X)$  se suele denotar como  $\sigma_x^2$  o simplemente por  $\sigma^2$ , es decir, otra manera de expresar la varianza sería de la siguiente forma:*

$$\sigma_x^2 = E([X - \mu_x]^2). \quad (3.10)$$

*Además, para una variable aleatoria  $X$ , una expresión alternativa de su varianza es:*

$$V(X) = E(X^2) - E^2(X) \quad (3.11)$$

*o, expresada como la esperanza de su cuadrado menos el cuadrado de su esperanza, sería:*

$$\sigma_x^2 = E(X^2) - \mu_x^2. \quad (3.12)$$

**Definición 21 (Desviación típica)** *Se define la desviación típica de una variable aleatoria  $X$ ,  $\sigma_x$ , como la raíz cuadrada de su varianza, es decir:*

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{V(X)}. \quad (3.13)$$

**Definición 22 (Coeficiente de correlación de Pearson)** *El coeficiente de correlación de Pearson, un índice muy utilizado en el ámbito de la estadística, mide el grado de relación que hay entre dos variables que están relacionadas linealmente. Sean  $x$  e  $y$  las variables que están relacionadas de forma lineal, el coeficiente de Pearson se define como:*

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (3.14)$$

Este valor oscila entre -1 y 1, siendo su interpretación en función del valor que tome la siguiente:

- Si  $r = 1$ , estamos en el caso de una correlación positiva perfecta. Cuando una de las variables aumenta, la otra variable también aumentará en la misma proporción, existiendo una relación directa o una dependencia total entre las variables  $x$  e  $y$ .
- Si  $0 < r < 1$ , existe correlación positiva, es decir, las dos variables están asociadas en sentido directo.
- Si  $r = 0$  no existe una relación lineal entre las variables, pudiendo existir relación de tipo no lineal.
- Si  $-1 < r < 0$ , existe correlación negativa, es decir, las variables se relacionan de manera inversa.
- Si  $r = -1$ , existe una correlación negativa perfecta. En este caso, mientras una variable aumenta, la otra variable disminuirá en la misma proporción, existiendo una relación opuesta entre las dos variables estudiadas.

**Definición 23 (Serie temporal)** *Se define una serie temporal como un conjunto de observaciones de una variable recogidas secuencialmente en el tiempo.*

Las series temporales pueden ser estacionarias o estables cuando estas oscilan en torno a un nivel o valor constante, mientras que hablamos de series temporales

no estacionarias cuando no se mantienen a un nivel constante. En la Figura 3.5 podemos observar un ejemplo de serie temporal, como así lo reflejan los valores del IBEX 35 recogidos diariamente entre los años 2013 y 2017.



Figura 3.5: Ejemplo de una serie temporal.



### 3.3. Comunidades y algoritmo de Girvan-Newman

#### 3.3.1. Comunidades

Uno de los fundamentos esenciales de esta investigación radica en la aplicación de redes funcionales, las cuales se generan mediante la construcción de una red compleja basada en las correlaciones existentes entre las cotizaciones bursátiles. Sin embargo, la aplicación de redes funcionales en la economía y, más concretamente en mercados financieros, es algo innovador, ya que las redes funcionales tradicionalmente se han utilizado mayoritariamente en campos como las redes sociales y la neurociencia.

Una de las aplicaciones más extendidas y frecuentes de las redes funcionales ha sido en el ámbito de las redes neuronales y el cerebro. Los avances recientes en el análisis cuantitativo de redes complejas, fundamentadas principalmente en los principios de la teoría de grafos, han observado una rápida aplicación en la investigación de la organización de redes cerebrales. Los sistemas estructurales y funcionales del cerebro presentan propiedades que son características de redes complejas [8] [9] [10][11], como que la mayoría de los nodos están conectados a través de unos pocos saltos, la presencia de nodos altamente conectados (llamados "hubs"), y la modularidad, que implica que el cerebro se organiza en módulos funcionales especializados. Según la teoría de grafos, las redes cerebrales estructurales se pueden describir como grafos que están compuestos por nodos (vértices) que representan elementos neuronales (neuronas o regiones cerebrales) que están conectados por aristas que representan conexiones físicas (sinapsis o proyecciones axonales). Las neuronas y las áreas del cerebro que se encuentran en proximidad física tienden a tener una mayor probabilidad de establecer conexiones entre sí, en contraste, las conexiones entre neuronas o regiones cerebrales distantes físicamente tienen una probabilidad menor de ocurrir [12] [13]. Además, se ha descubierto que muchos sistemas complejos muestran un comportamiento macroscópico sorprendentemente similar a pesar de las profundas diferencias en los detalles microscópicos de los elementos de cada sistema o sus mecanismos de interacción, como es el caso del cerebro. [14]

Esta novedosa sinergia entre la teoría de redes y la economía que se propone en este trabajo está abriendo puertas a un entendimiento más profundo de las interconexiones en los sistemas económicos, permitiendo a los investigadores analizar de manera más precisa cómo los actores económicos interactúan y se influyen mutuamente. Esta innovación promete revolucionar nuestra comprensión de los mercados, la toma de decisiones financieras y la dinámica económica en su conjunto.

Además, si queremos remontarnos a los orígenes de la detección de comunidades y entender en base a un ejemplo cómo se forman estas comunidades, es

importante hacer referencia al problema del club de karate de Zachary, un problema que ha servido como referencia y modelo en el ámbito académico para la comprensión de la dinámica de formación de comunidades en el contexto de las redes complejas, después de que Michelle Girvan y Mark Newman lo utilizaran en el año 2002. El club de karate de Zachary es una red social que representa un club universitario de karate, tal y como se describe en el artículo titulado “Un modelo de flujo de información para conflictos y fisión en grupos pequeño” escrito por Wayne W. Zachary.[6]

Entre los años 1970 y 1972, Zachary observó las interacciones sociales en un club de karate universitario en Estados Unidos. En el artículo, Zachary construyó la red de relaciones sociales en el club, incluyendo a todos los miembros del club, así como al Sr. John A, presidente y administrador del club y al instructor de karate a tiempo parcial, el Sr. Hi. Un conflicto relacionado con el precio de las tarifas entre el Sr. Hi, quien deseaba aumentar los precios y el Sr. John A, que deseaba mantener los precios que había, provocó fuertes conflictos y la creación de un nuevo club de karate liderado por el instructor el Sr. Hi, derivando en una división del club. La pregunta principal que se hacía Zachary es cómo se desarrolla y se puede predecir la división del club en dos facciones a lo largo del tiempo, basándose en la observación de las relaciones y las interacciones sociales de los miembros. En la Figura 3.6 se puede observar la red social formada por el club de Karate Zachary, compuesta de 34 nodos y 78 aristas. El nodo 1 corresponde al Sr. Hi, mientras que el nodo 34 designa al Sr. John A.

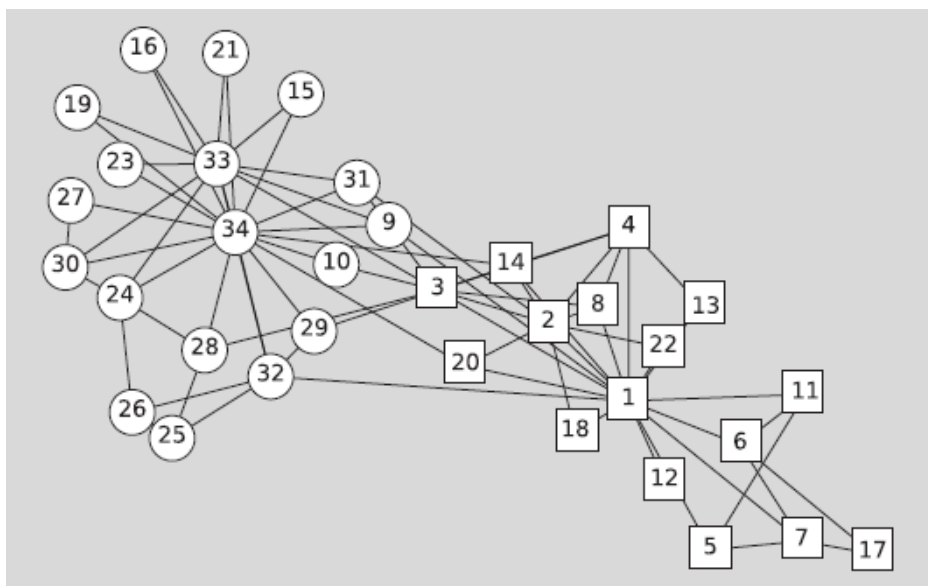


Figura 3.6: Red asociada al club de karate de Zachary .

A simple vista se puede observar la existencia de dos grupos, con aproximadamente la mitad de los nodos perteneciendo a un grupo o comunidad que incluye

el nodo 1, es decir, al Sr.Hi, y la otra mitad restante de los nodos formando parte de un grupo que incluye al nodo 34, es decir, el grupo liderado por el Sr. John A. Los nodos que permanecen con el presidente John A. se muestran como círculos en la imagen, mientras que las personas que apoyaron al instructor de karate, el Sr.Hi, se muestran en la imagen como cuadrados.

Este problema de investigación ha proporcionado una valiosa perspectiva sobre cómo se forjan comunidades, cómo evolucionan las dinámicas grupales y cómo se aplican los conceptos de redes sociales en una amplia variedad de disciplinas, incluyendo la sociología, la psicología y la informática.

La detección y caracterización de la estructura de comunidades en redes es un desafío importante en la investigación de sistemas en red. Una estrategia para resolver este problema de alta efectividad implica la optimización de una función de calidad conocida como modularidad. Además, uno de los conceptos desarrollados por Newman en el año 2006 en su paper “Modularity and community structure in networks” es el de modularidad [15], donde se demuestra que es posible expresar la modularidad en términos de los eigenvectores de una matriz característica de la red, denominada matriz de modularidad. En otras palabras, la modularidad cuantifica la calidad de una partición dada de los nodos que forman la red. Además, es posible extraer comunidades directamente mediante la optimización de la modularidad sobre el conjunto de particiones posibles.

La modularidad es la diferencia entre el número de aristas que caen dentro de los grupos y el número esperado en una red equivalente con aristas dispuestas al azar. Además, la modularidad puede tomar tanto valores positivos como negativos, con valores positivos indicando la posible presencia de una estructura de comunidad.

Supongamos que nuestra red contiene  $n$  vértices o nodos. Para una división particular de la red en dos grupos o comunidades, definimos  $s_i$  como 1 si el vértice  $i$  pertenece al grupo 1 y  $s_i$  como -1 si pertenece al grupo 2. La cantidad de aristas entre los vértices  $i$  y  $j$  se representa como  $A_{ij}$ , que normalmente toma el valor de 0 o 1, aunque en redes donde se permiten múltiples aristas pueden existir valores más grandes. Estas cantidades  $A_{ij}$  son los elementos de la llamada matriz de adyacencia. Al mismo tiempo, el número esperado de aristas entre los vértices  $i$  y  $j$  si las aristas se distribuyen de manera aleatoria es  $k_i k_j / 2m$ , donde  $k_i$  y  $k_j$  son los grados de los vértices, y  $m$  es igual a la mitad del total de aristas en la red. Por lo tanto, la modularidad  $Q$  se obtiene a partir de la suma de  $A_{ij} - k_i k_j / 2m$  sobre todos los pares de vértices  $i, j$  que pertenecen al mismo grupo. Teniendo en cuenta que la cantidad  $1/2(s_i s_j - 1)$  es igual a 1 si  $i$  y  $j$  están en el mismo grupo y 0 en caso contrario, se puede expresar la modularidad como:

### Definición 24 (Modularidad)

$$Q = \frac{1}{4m} \sum_{i,j} \left( A_{ij} - \frac{k_i k_j}{2m} \right) (s_i s_j + 1) = \frac{1}{4m} \sum_{i,j} \left( A_{ij} - \frac{k_i k_j}{2m} \right) s_i s_j, \quad (3.15)$$

donde la segunda igualdad se obtiene de la observación de que  $2m = \sum_i k_i = \sum_{i,j} A_{ij}$ . El factor principal de  $1/4m$  es convencional y se incluye para mantener compatibilidad con la definición previa de modularidad.

### 3.3.2. Algoritmo de Girvan-Newman

Los algoritmos de detección de comunidades desempeñan un papel crucial en la exploración de la organización de datos en forma de grafos. Estos algoritmos son herramientas fundamentales en el campo del análisis de datos, particularmente en el ámbito de los grafos. Como es bien sabido, los grafos permiten representar y relacionar diversos tipos de datos a través de nodos interconectados. En esta coyuntura, estos algoritmos nos permiten llevar a cabo análisis significativos de secciones específicas, características particulares y relaciones entre ellos, con el fin de identificar agrupaciones o comunidades dentro de los datos.

Existen una amplia variedad de algoritmos disponibles para identificar comunidades en los grafos, y la elección del algoritmo apropiado depende tanto de la naturaleza del grafo como de la naturaleza del problema que se está abordando. Algunos de los algoritmos mas utilizados son los que se mencionan a continuación: Algoritmo de Girvan-Newman, Algoritmo de Louvain, Algoritmo de Infomap o Algoritmo de Radicchi entre otros [16].

En este trabajo se empleará el algoritmo de Girvan-Newman debido a su enfoque sólido, su aplicabilidad a diferentes tipos de grafos y su eficiencia computacional, lo que ha hecho que sea uno de los algoritmos más populares en la detección de comunidades.

El algoritmo Girvan-Newman, desarrollado por Michelle Girvan y Mark Newman en 2002, detecta comunidades eliminando progresivamente aristas del grafo original. Se trata de un método muy extendido en los sistemas complejos de datos que es de orden jerárquico y que se basa en el concepto de edge betweenness, una medida que cuantifica la importancia de una arista en función de la cantidad de tráfico que pasa por ella [6].

La idea en la que se basa el algoritmo de Girvan-Newman es muy intuitiva, ya que si un grafo tiene una estructura de comunidad, generalmente habrá pocas aristas entre dos comunidades, por lo que todos los caminos más cortos entre los nodos que pertenecen a una comunidad y los que pertenecen a otra comunidad

debe contener una de esas aristas y, consecuentemente, las aristas que conectan comunidades tendrán una mayor centralidad.

La centralidad entre aristas se define como el número de caminos más cortos que pasan por una arista de un grafo o red (Girvan y Newman, 2002). Cada arista de la red puede asociarse a un valor de centralidad entre aristas. Una arista con un valor alto de centralidad entre aristas representa un conector a modo de puente entre dos partes de una red, y cuya eliminación puede afectar a la comunicación entre muchos pares de nodos a través de los caminos más cortos entre ellos.

Los pasos que sigue el algoritmo Girvan-Newman son los siguientes:

1. En primer lugar, se calcula la centralidad de intermediación para todas las aristas del grafo.
2. Una vez calculadas, se elimina la arista con mayor intermediación y se realiza un análisis en el grafo obtenido, representando cada nodo una comunidad.
3. Se calcula la centralidad de intermediación para las aristas restantes.
4. Se repiten los pasos 2-4 hasta que no queden aristas en el grafo

En cada iteración, el algoritmo elimina la arista considerada como "más valiosa", que generalmente se trata de la arista con mayor centralidad de interrelación. A medida que se eliminan aristas del grafo, este se fragmenta en varios componentes más pequeños, revelando así la estructura de las comunidades estrechamente unidas [17]. El resultado final del algoritmo puede representarse mediante un diagrama de árbol.

En la Figura 3.7 podemos observar cómo funciona el algoritmo de Girvan-Newman. En este caso, partimos de un grafo inicial con 6 nodos y 7 aristas y, al aplicar el algoritmo de Girvan-Newman, la arista que hay entre los nodos C y D se elimina, al ser la arista con mayor peso. Se aplica el proceso hasta que los nodos queden aislados, dando lugar a la creación de dos comunidades diferentes.

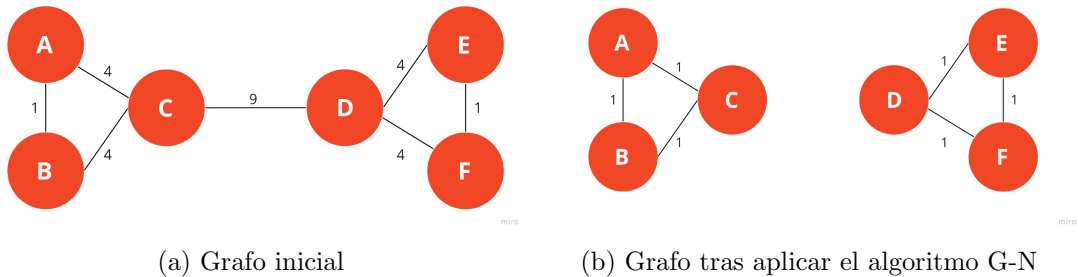


Figura 3.7: Aplicación del algoritmo Girvan-Newman.

### 3.4. Conceptos económicos

El análisis de los mercados de valores es esencial para el desarrollo y el diseño de las estrategias de inversión. Los rendimientos de ciertas acciones en muchas ocasiones están correlacionados, ya sea por la dirección general del mercado o por la ciclicidad de los mismos segmentos del mercado. Conocer el comportamiento de un mercado a lo largo del tiempo es clave para poder posteriormente definir estrategias de inversión y saber cómo gestionar las carteras. En un mercado de valores, los índices o las cotizaciones bursátiles de las distintas empresas se correlacionan entre sí con el paso del tiempo [18][19].

**Definición 25 (Bolsa de valores)** *La bolsa de valores se define como el lugar donde se intercambian diversos instrumentos financieros. títulos de renta variable, como las acciones de las empresas cotizadas, títulos de renta fija, como bonos, letras del tesoro o deuda tanto pública como privada, así como otros valores negociables*

**Definición 26 (Índice bursátil)** *Se define un índice bursátil como el indicador que proporciona información sobre la variación del precio de un conjunto de activos cotizados que reúnen unas determinadas características.*

Los índices bursátiles se utilizan para medir y seguir el desempeño general del mercado de valores, así como para proporcionar una referencia para evaluar el rendimiento de las inversiones. Los movimientos de los índices bursátiles pueden proporcionar una idea general de cómo se están comportando las acciones en un mercado determinado. Además, los índices bursátiles se utilizan como referencia para evaluar el rendimiento de las carteras de inversión y para comparar el desempeño de un fondo de inversión o de una estrategia de inversión en relación con el mercado en general.

En este estudio, analizamos las empresas del IBEX 35, el principal índice bursátil de referencia de la bolsa española. El IBEX 35 lo componen las 35 empresas más importantes del país, es decir, las 35 empresas con mayor liquidación y capitalización de entre las que cotizan en el Sistema de Interconexión Bursátil Electrónico (SIBE) en las cuatro bolsas que hay en España, situadas en Madrid, Barcelona, Valencia y Bilbao. El IBEX 35 permite hacer valoraciones sobre el riesgo de ciertas acciones al tratarse el IBEX 35 de un indicador de la evolución de la economía española. De esta forma, y de manera general, puede establecerse una relación entre la evolución de la economía española y la de las empresas españolas, de manera que si la economía española está pasando por un buen período, muchas de las empresas españolas también estarán teniendo grandes resultados y viceversa. Además, comparando la variación del IBEX 35 con cuánto

reaccionan las empresas ante esta variación se puede calcular cómo de arriesgado será invertir o no en esa acción [20][21].

**Definición 27 (Acción)** *En la bolsa de valores, una acción representa una parte de propiedad o participación en una empresa, es decir, es un título de valor que se negocia en la bolsa de valores y que se puede comprar y vender libremente. Los accionistas son socios propietarios de la sociedad o empresa en un porcentaje proporcional a la participación que se tenga en esa empresa.*

Debido al auge que han experimentado las bolsas de valores desde su creación, cotizando cada vez más empresas en ellas, se hizo necesario crear un código alfanumérico que permitiese identificar las acciones dentro de un mercado bursátil fácilmente. Este código o etiqueta de cotización suele conocerse como 'ticker'.

**Definición 28 (Ticker)** *En el mercado de valores, un ticker es un código único que se utiliza para identificar a las acciones de una empresa que cotiza en un determinado índice bursátil. Al establecerse un estándar global para la identificación de activos financieros, la comunicación y el intercambio de datos en los mercados financieros, así como el monitoreo y el seguimiento de inversiones sea más sencillo.*

En el caso del IBEX 35, los tickers de las diferentes acciones suelen representarse con 3 o 4 letras, si bien es cierto que esto varía en función del mercado de valores sobre el que se ponga el foco. Además, en el caso de las empresas que tengan más de una acción diferente en el mercado de valores, cada acción se representará con un ticker diferente.

Cabe destacar que la bolsa cuenta con 3 períodos diferentes para cada sesión diaria: subasta de apertura, mercado abierto y subasta de cierre. Entre las 8:30 horas de la mañana y las 17:30 horas de la tarde de lunes a viernes, la bolsa española está abierta, si bien es cierto que en días festivos pueden darse algunas excepciones. Explicaremos con detalle cada uno de estos periodos con los que cuenta la bolsa:

- Durante la subasta de apertura, que se da entre las 8:30 y 9:00 de la mañana, se recopilan las órdenes de compra y venta para empresas concretas por parte de los accionistas, sin llegar a producirse operaciones o negociaciones como tal. La subasta de apertura se considera un momento clave ya que, además de dar una idea general a los inversores de cuáles serán los precios de las acciones antes de comenzar el mercado abierto, sirve para que el

precio con el que se inicia la sesión no sea el último precio de cierre del día anterior, debido a que en ese transcurso de tiempo pueden haberse producido previsiones o contratos que afectasen directamente a ese precio de cierre.

- Una vez que se acaba la subasta de apertura, se pasa al periodo de mercado abierto o mercado continuo, que abarca desde las 9:00 a 17:30 horas, donde ya se producen operaciones en el caso de que haya órdenes de compra o venta y, por tanto, ya se negocia en bolsa. En el caso de que se produzca una orden de compra y no haya una orden de venta a ese precio, la orden de compra se considera pendiente y pasa a formar parte del libro de órdenes.
- Por último, la subasta de cierre se produce entre las 17:30 y 17:35 horas de la tarde y sigue un procedimiento similar a la subasta de apertura. Durante este periodo, se pueden lanzar órdenes de compra o venta y, generalmente, el precio que se establezca en esta subasta de cierre para las acciones de cada compañía será el precio de cierre al final de la sesión.

Para cada sesión del mercado de valores, las acciones abren a un precio específico, conocido como precio de apertura, y a lo largo del día en el mercado varían los precios de las acciones en función de la oferta y la demanda de estas, pudiendo ser el precio de cierre mayor o menor que el precio de apertura [22]

**Definición 29 (Precio de apertura)** *El precio de apertura de un valor en el mercado bursátil se corresponde con el precio al que comienza la negociación de un determinado activo financiero en una sesión del mercado financiero.*

**Definición 30 (Precio de cierre)** *El precio de cierre de un valor en el mercado de valores hace referencia al último precio de compraventa de este en la bolsa antes de cerrar.*

El precio de cierre es un indicador clave en el mercado financiero ya que los analistas financieros y los inversores suelen coger de referencia el precio de cierre para poder estudiar el desempeño de una acción respecto a los días previos, así como para calcular el rendimiento de una acción durante un período de tiempo concreto.

Debido a estas fluctuaciones del precio de los activos en los mercados financieros a lo largo del tiempo, es importante explicar introducir el término de rentabilidad, un concepto clave en el ámbito de la inversión y las finanzas.



**Definición 31 (Rentabilidad)** *La rentabilidad de una inversión se define como la variación porcentual que sufre el valor de un activo durante un período de tiempo concreto.*

Cabe distinguir entre la rentabilidad histórica de un activo, es decir, la rentabilidad media al final del plazo que se puede calcular a partir de los datos históricos que se hayan registrado durante el periodo y la rentabilidad esperada, que hace referencia al inicio del plazo y que solo puede estimarse a partir de las previsiones realizadas sobre un período.

Dado un período de una inversión  $T$ , la rentabilidad simple de esta inversión se representa por  $RS_T$  y se calcula como:

$$RS_T = \frac{P_T + D_T - P_0}{P_0}, \quad (3.16)$$

donde:

$P_T$ : precio del título al final del período  $T$ .

$D_T$ : suma aritmética de todos los ingresos percibidos durante el período  $T$ .

$P_0$ : precio del título al inicio del período.

De la mano del término de rentabilidad va el término de volatilidad, dos conceptos que guardan mucha relación entre sí, ya que generalmente a mayor volatilidad de un activo, mayor potencial de rentabilidad y mayor riesgo y viceversa.

**Definición 32 (Volatilidad)** *En un mercado financiero, se define la volatilidad como la tasa en la que varía el precio de un activo durante un determinado periodo de tiempo.*

Se considera que un activo es altamente volátil si su precio varía o cambia rápidamente en un período breve de tiempo, mientras que si el precio de un activo cambia de forma más lenta a lo largo del tiempo se considera que tiene baja volatilidad. La volatilidad guarda una estrecha relación con el riesgo, ya que a mayor volatilidad de un activo, mayor riesgo de que la rentabilidad no sea la esperada y viceversa, entendiéndose como riesgo la incertidumbre que genera la fluctuación de la rentabilidad del activo en torno a su valor esperado. La volatilidad suele representarse como una tasa que nos indica cuánto ha variado la rentabilidad de un activo con respecto a su histórico. [23]

Para calcular la volatilidad de un activo  $t$ , representada como  $\sigma_t$ , se suele utilizar la desviación típica, midiendo así el grado de dispersión de la rentabilidad respecto a la rentabilidad esperada.

La volatilidad de un activo, por lo tanto, es un factor clave que influye en la gestión de riesgos, usando tanto inversores como gestores esta medida con el objetivo de analizar el riesgo de sus inversiones. Además, juega un papel fundamental en la toma de decisiones de inversión, pudiendo darse dos escenarios. Por un lado, están los inversores que prefieren en muchas ocasiones activos menos volátiles y que ofrezcan una mayor estabilidad a lo largo del tiempo, siendo su rentabilidad más previsible y, en el lado opuesto, están los inversores que prefieren activos altamente volátiles que puedan generarles mayores beneficios en un período de tiempo más breve. Por último, cabe destacar que la volatilidad afecta al mercado financiero en general, ya que cambios repentinos en la volatilidad de los activos puede afectar a la toma de decisiones de inversión de los inversores y, por tanto, desestabilizar el mercado financiero en su conjunto [24].

Por lo tanto, los inversionistas, a la hora de tomar decisiones de inversión, deben tener en cuenta tanto la rentabilidad esperada como la volatilidad, evaluando su tolerancia al riesgo y tratando de encontrar el equilibrio entre la rentabilidad esperada y la volatilidad que deseen.



# 4

## Desarrollo del trabajo

Una vez explicado el marco teórico que sustenta este trabajo, en este apartado se explicará cómo se ha desarrollado el trabajo, desde la recogida o toma de datos hasta la detección de comunidades usando el algoritmo Girvan-Newman.

El lenguaje de programación usado durante todo el trabajo ha sido Python, utilizando concretamente la herramienta de Jupyter Notebook. En primer lugar, se han importado las bibliotecas NumPy y Pandas, que nos proporcionan funcionalidades específicas para el tratamiento y el análisis de datos.

Además, para la recogida de datos se ha importado la biblioteca 'yfinance', la cual nos permite múltiples funciones para trabajar con datos financieros en Python al acceder directamente a Yahoo Finance, una plataforma que ofrece datos financieros de los distintos índices bursátiles a nivel mundial, además de proporcionar noticias, seguimiento y análisis de acciones. A través de Yahoo Finance, se puede acceder a las cotizaciones históricas de las empresas que componen los principales índices bursátiles. En nuestro caso, hemos elegido el IBEX 35 y, por tanto, para cada sesión de la bolsa podemos obtener el precio de apertura, precio de cierre, valor máximo y mínimo, así como el cierre ajustado para cada una de las empresas que forman parte del IBEX 35.

Debido a que las empresas que cotizan en el IBEX 35 no son fijas a lo largo del tiempo, sino que deben cumplir unos requisitos que son valorados por un

---

Comité Asesor Técnico, para poder entrar y permanecer en el IBEX 35, se han seleccionado para el estudio las empresas que se encuentran en el IBEX 35 a fecha de abril de 2023, exceptuando Acciona Energías Renovables debido a su reciente incorporación a la bolsa en el año 2021. En la tabla [4.1](#) se muestran las empresas que forman parte del estudio, junto al ticker que las representan en el IBEX 35 y el sector al que pertenecen.

Ticker	Empresa	Sector
ACS	Actividades de Construcción y Servicios	Construcción
ACX	Acerinox	Mineral, metales y transformación
AENA	Aena	Transporte y distribución
ALM	Almirall	Productos farmacéuticos
AMS	Amadeus IT Group	Electrónica y software
ANA	Acciona	Construcción
BBVA	BBVA	Bancos y cajas de ahorro
BKT	Bankinter	Bancos y cajas de ahorro
CABK	CaixaBank	Bancos y cajas de ahorro
CLNX	Cellnex Telecom	Telecomunicaciones y otros
COL	Inmobiliaria Colonial	SOCIMI
ELE	Endesa	Electricidad y gas
ENG	Enagás	Electricidad y gas
FDR	Fluidra	Ingeniería y otros
FER	Ferrovial	Construcción
GRF	Grifols	Productos farmacéuticos
IAG	IAG	Transporte y distribución
IBE	Iberdrola	Electricidad y gas
IDR	Indra Sistemas	Electrónica y software
ITX	Inditex	Textil, vestido y calzado
LOG	Logista	Transporte y distribución
MAP	Mapfre	Seguros
MEL	Meliá Hotels International	Ocio, turismo y hostelería
MRL	Merlin Properties	SOCIMI
MTS	ArcelorMittal	Mineral, metales y transformación
NTGY	Naturgy	Electricidad y gas
RED	Red Eléctrica Corporación	Electricidad y gas
REP	Repsol	Petróleo
ROVI	Laboratorios ROVI	Productos farmacéuticos y biotecnología
SAB	Banco Sabadell	Bancos y cajas de ahorro
SAN	Banco Santander	Bancos y cajas de ahorro
SCYR	Sacyr	Construcción
SLR	Solaria Energía y Medio Ambiente	Energías renovables
TEF	Telefónica	Telecomunicaciones y otros
UNI	Unicaja Banco	Bancos y cajas de ahorro

Tabla 4.1: Listado de empresas del Ibex 35 analizadas.

Como podemos observar en la tabla anterior, los sectores que más predominan son bancos y cajas de ahorro, electricidad y gas y la construcción, seguidos de otros sectores como el transporte y distribución, telecomunicaciones, electrónica y software o productos farmacéuticos y biotecnología.

---

Con el fin de comparar la relación entre las distintas empresas del IBEX 35 y que este sea preciso y representativo, se ha cogido como referencia el precio de cierre diario. El precio de cierre se considera representativo del valor de un activo en un momento dado, ya que refleja la oferta y la demanda de los inversores hasta el cierre del mercado. A pesar de que los precios fluctúan durante el día, el precio de cierre suele considerarse un buen punto de referencia para medir el rendimiento y el valor de un activo en un período determinado. Utilizar el precio de cierre permite una comparación más consistente entre diferentes períodos o activo, ya que al utilizar el mismo punto de referencia, es posible evaluar el rendimiento relativo y realizar análisis comparativos más precisos.

De esta forma, se ha implementado en Python un programa que, dadas dos fechas (una fecha inicial y una fecha final), se extraiga el precio de cierre de cada una de las empresas de la Tabla 4.1 para cada día. Las empresas que formarán parte del estudio se han almacenado en una lista a través de sus símbolos o tickers, de tal forma que para obtener los precios de cierre de todas las empresas a lo largo del tiempo, se ha creado un bucle que recorre la lista mencionada y, para cada valor de la lista, descarga el precio de cierre diario en el período deseado. La matriz resultante con los precios de cierre en el período establecido para cada empresa, se pasa a un DataFrame con el objetivo de que se puedan manipular, analizar y visualizar mejor estos datos. En la Figura 4.1 podemos observar cómo se ha implementado esta función en Python.

```
# Importamos las bibliotecas necesarias
import yfinance as yf
import datetime
import numpy as np
import pandas as pd

# Símbolos o tickers de las empresas del IBEX 35 escogidas en el estudio en forma de lista
symbols = [
    "ACS.MC", "ACX.MC", "AENA.MC", "ALM.MC", "ANA.MC", "BBVA.MC", "BKT.MC", "CABK.MC",
    "CLNX.MC", "COL.MC", "ELE.MC", "ENG.MC", "FDR.MC", "FER.MC", "GRF.MC", "IAG.MC", "IBE.MC", "IDR.MC",
    "ITX.MC", "LOG.MC", "MAP.MC", "MEL.MC", "MRL.MC", "MTS.MC", "NTGY.MC", "RED.MC", "REP.MC", "ROVI.MC", "SAB.MC", "SAN.MC", "SCYR.MC",
    "SLR.MC", "TEF.MC", "UNI.MC",
]

# Definimos las variables fecha inicial y fecha final
start_date = datetime.datetime(2020, 1, 1)
end_date = datetime.datetime(2022, 1, 1)

# Función para obtener el valor o precio de cierre para cada símbolo (empresa del IBEX 35)
close_data = {}
for symbol in symbols:
    data = yf.download(symbol, start=start_date, end=end_date)
    close_data[symbol] = data["Close"]

# Obtenemos todas las fechas únicas
dates = pd.date_range(start=start_date, end=end_date, freq="B")

# Creamos la matriz de valores de cierre
num_dates = len(dates)
num_symbols = len(symbols)
close_matrix = np.zeros((num_dates, num_symbols+1))
close_matrix[:, 0] = dates

for i, symbol in enumerate(symbols):
    close_prices = close_data[symbol].reindex(dates)
    close_matrix[:, i+1] = close_prices.values

# Imprimir y mostrar la matriz con los precios de cierre diarios para cada empresa
column_names = ['Fecha'] + symbols
df = pd.DataFrame(close_matrix, columns=column_names)
df['Fecha'] = pd.to_datetime(df['Fecha']).dt.date # Convertir las fechas a formato "yyyy-mm-dd"
print(df)
```

Figura 4.1: Función para la toma de datos en Python.

## 4.1. Elección de umbral óptimo

Además, para determinar cuándo dos empresas están conectadas entre ellas en nuestro estudio, es decir, cuándo hay una arista entre dos vértices del grafo, se ha empleado el coeficiente de correlación de Pearson. Para establecer el umbral apropiado a la hora de construir la matriz de adyacencia que nos dará el grafo, se ha estudiado cuál sería el umbral óptimo. La importancia de encontrar un umbral óptimo y adecuado viene motivada por el hecho de que si se escoge un umbral al azar y este es demasiado bajo, el grafo resultante contará con un número muy alto de aristas, ya que la mayoría de empresas estarán relacionadas entre ellas y el análisis no será representativo. De la misma manera, en el caso de que el umbral establecido sea demasiado alto, el grafo apenas contará con aristas y, al igual que pasaba anteriormente, nuestro análisis no reflejará la realidad. Con el fin de solventar este problema, se ha representado en un gráfico el número de aristas que habría en el grafo en función del umbral escogido.

Para cada período de tiempo analizado, se ha representado gráficamente cuántas aristas habría en el grafo en función del umbral escogido. En el eje X del gráfico



se muestra el umbral, que va desde 0,5 hasta 1, mientras que el eje Y recoge el número de aristas total que tendría el grafo. Tomamos umbrales mayores a 0,5 en la representación gráfica para que se pueda percibir con mayor facilidad dónde hay un cambio en la pendiente, ya que umbrales de 0 a 0,5 son considerados muy bajos, al ser el número de aristas muy elevado, y no nos interesan para el estudio. Como cabe esperar, a menor umbral mayor número de aristas habrá, ya que el requisito que tendrán que cumplir dos empresas para que tengan conexión entre ellas es más suave, es decir, menos restrictiva. Además, se calcula el número de aristas para cada valor del umbral de 0 a 1, siendo el incremento entre los valores de las aristas de milésima en milésima. En la Figura 4.2 podemos ver cómo se ha implementado en Python la función que nos determina el umbral óptimo.

```
def calcular_matriz_adyacencia(correlation_matrix, threshold):
    matriz_adyacencia = np.where(np.abs(correlation_matrix) > threshold, 1, 0)
    return matriz_adyacencia
# Dividimos por 2 para evitar contar las aristas duplicadas
def contar_aristas(matriz_adyacencia):
    return int(np.sum(matriz_adyacencia) / 2)

#El umbral irá desde 0 hasta 1, de milésima en milésima
thresholds = np.arange(0, 1.1, 0.001)
num_aristas = []

for threshold in thresholds:
    matriz_adyacencia = calcular_matriz_adyacencia(correlation_matrix, threshold)
    num_aristas.append(contar_aristas(matriz_adyacencia))
#Creamos el gráfico que nos relacione el número de aristas del grafo en función del umbral
plt.plot(thresholds, num_aristas)
plt.xlabel('Umbral')
plt.ylabel('Número de aristas')
plt.title('Número de aristas en función del umbral')
plt.axis([0.5, 1, 0, 300])
plt.gca().invert_xaxis()
plt.gca().invert_yaxis()

plt.show()
```

Figura 4.2: Bloque de código para determinar el umbral óptimo del coeficiente de correlación de Pearson.

Debido a que el umbral óptimo se establece a partir de una matriz de correlación determinada, este umbral dependerá del período de tiempo que analicemos, ya que para cada período tenemos una matriz de correlación diferente, asociada a los valores de cierre de las empresas en esos años.

En la Figura 4.3 está representado cómo varía el número de aristas en función del umbral escogido para un período de tiempo dado. La idea es que, a través de esta representación, se identifique correctamente dónde está el cambio en la pendiente de la relación mostrada y, para el valor del umbral en el que se produzca ese cambio de pendiente, ese será el umbral óptimo sobre el cual se creará la matriz de adyacencia.

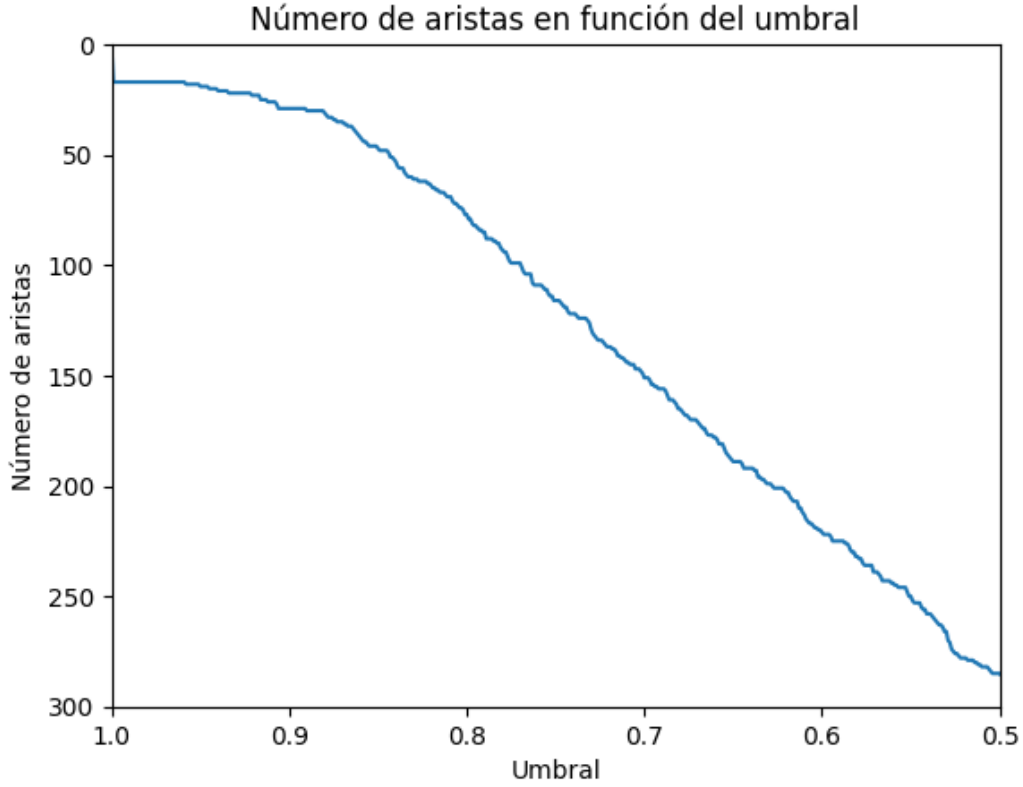


Figura 4.3: Ejemplo de elección del umbral óptimo para un período dado

Sea  $\alpha$  el umbral óptimo escogido para un período de tiempo dado, es decir, es para el valor del umbral  $\alpha$  cuando se percibe un mayor cambio en el número de aristas que lleva asociado este umbral, las empresas cuyo coeficiente de correlación de Pearson sea superior a  $\alpha$  estarán unidas mediante una arista mientras que, en caso contrario, no existirá conexión entre ellas. La matriz de adyacencia que refleja este requisito es la siguiente:

$$A_{ij} = \begin{cases} 1, & \text{si } r_{xy} \geq \alpha, i \neq j, \\ 0, & \text{en otros casos,} \end{cases}$$

siendo  $r_{xy}$  el coeficiente de correlación de Pearson.

Tras la obtención de la matriz de adyacencia asociada al grafo sobre el que se trabajará posteriormente, se aplicará el algoritmo de Girvan-Newman. La metodología escogida para la aplicación de dicho algoritmo se ha basado en el número de iteraciones que ejecutará este algoritmo. Para ambos períodos de tiempo, el número de iteraciones escogidas ha sido de un total de 7, es decir, se crearán 8 comunidades diferenciadas para cada período. Si bien es cierto que analizando la modularidad del algoritmo Girvan-Newman en función del número de comunida-

des creadas este no era el valor máximo, se ha considerado que 7 iteraciones es un número óptimo debido a que el valor de la modularidad sigue estando dentro de unos valores buenos y, sumado al número de nodos con los que cuenta el grafo para cada período, se trata de un valor razonable de iteraciones el hecho de tratar de identificar un total de 8 comunidades. Es importante destacar que, generalmente, una mayor modularidad está asociada con una mayor claridad en la estructura de las comunidades. Sin embargo, también hay que tener en cuenta el grafo del que se parte y los objetivos o resultados que se esperan cumplir u obtener. Por ejemplo, dependiendo del problema que se pretenda analizar puede ser más óptimo crear solo 2 comunidades o, por el contrario, crear 10 comunidades.

De cara al análisis que se llevará a cabo, se ha decidido realizar este estudio en dos períodos distintos de la economía española, correspondiendo con algunos de los ciclos económicos que ha experimentado la economía española recientemente, los cuales explicaremos a continuación:

1. Período previo a la pandemia (2018-2020): Si bien es cierto que a partir de 2013, cuando la tasa de desempleo alcanza su valor máximo, España comenzó a mostrar signos de recuperación económica de la crisis financiera del 2008, es a mediados del año 2017 cuando se recuperan los niveles del Producto Interior Bruto (PIB) previos al estallido de la crisis, además de registrarse máximos en el volumen de exportaciones. Analizaremos el período comprendido entre los años 2018 y 2020, es decir, los años previos a la pandemia, cuando la economía española estaba experimentando un período de recuperación gradual de la crisis del 2008.
2. Impacto de la pandemia de COVID-19 (2020-actualidad): Al igual que muchos otros países, España se vio afectada por la pandemia de COVID-19, lo que generó un impacto económico significativo. Las restricciones y medidas de confinamiento, así como los cierres temporales de negocios para contener la propagación del virus llevaron a una contracción económica, una caída del PIB y un aumento del desempleo. El gobierno español implementó medidas de apoyo económico y estímulo para mitigar los efectos de la crisis, como los Expedientes de Regulación Temporal de Empleo (ERTE), moratorias y aplazamientos en el pago de impuestos, ayudas específicas a los sectores más afectados por la pandemia o el Plan de Recuperación, Transformación y Resiliencia (PRTR), un programa diseñado para impulsar la recuperación económica del país, modernizar la economía española y crear empleo.

# 5

## Resultados

Tras haber expuesto con detalle cómo se ha llevado a cabo el desarrollo de este trabajo, en este capítulo se mostrarán los principales resultados obtenidos una vez aplicado el algoritmo Girvan-Newman en los diferentes períodos de tiempo considerados. Además, se comentarán y analizarán las comunidades que se han creado en cada franja temporal, comparando ambos períodos para así poder extraer conclusiones sobre el verdadero impacto que ha tenido la pandemia de la COVID-19 entre las empresas del IBEX 35.

## 5.1. Período pre-pandemia (2018-2020)

En esta sección, se presentarán los hallazgos más significativos del análisis realizado sobre las empresas del IBEX 35 durante el período comprendido entre el 1 de enero de 2018 y el 31 de diciembre de 2019, es decir, los dos años anteriores al estallido de la pandemia.

En lo relativo al umbral óptimo escogido para crear la matriz de adyacencia en este período, se ha establecido como umbral el 0,90, es decir, aquellas empresas cuyo coeficiente de correlación de Pearson entre sus precios de cierre sea mayor o igual a este valor, estarán unidas a través de una arista y, en caso contrario, no se establecerá conexión entre los vértices de las empresas. En la Figura 5.1 se puede apreciar cómo el cambio de pendiente se produce para el valor 0,90 y, por este motivo, se ha elegido este valor como umbral óptimo. Teniendo en cuenta esta restricción impuesta, en nuestro grafo hay un total de 46 aristas.

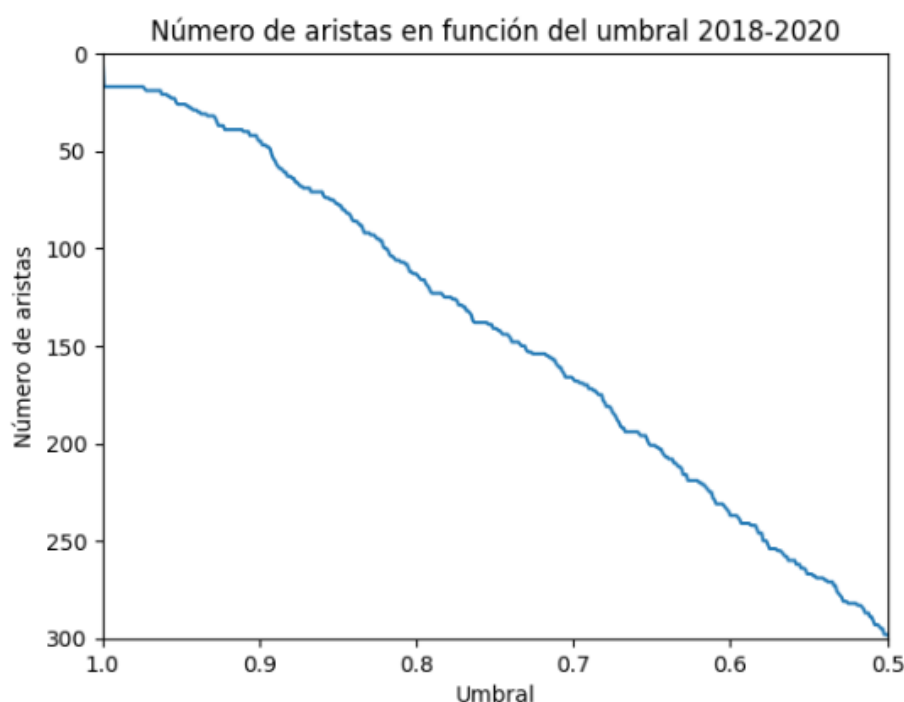


Figura 5.1: Umbral óptimo asociado al período 2018-2020

En la Figura 5.2 podemos observar cuál es el grafo inicial del que se parte y, a partir del cual, se detectarán comunidades. Cabe destacar que las empresas con mayor número de conexiones son Iberdrola y la mayoría de las empresas del sector financiero (CaixaBank, Bankinter, BBVA, Unicaja y Sabadell).

## Período 2018-2020

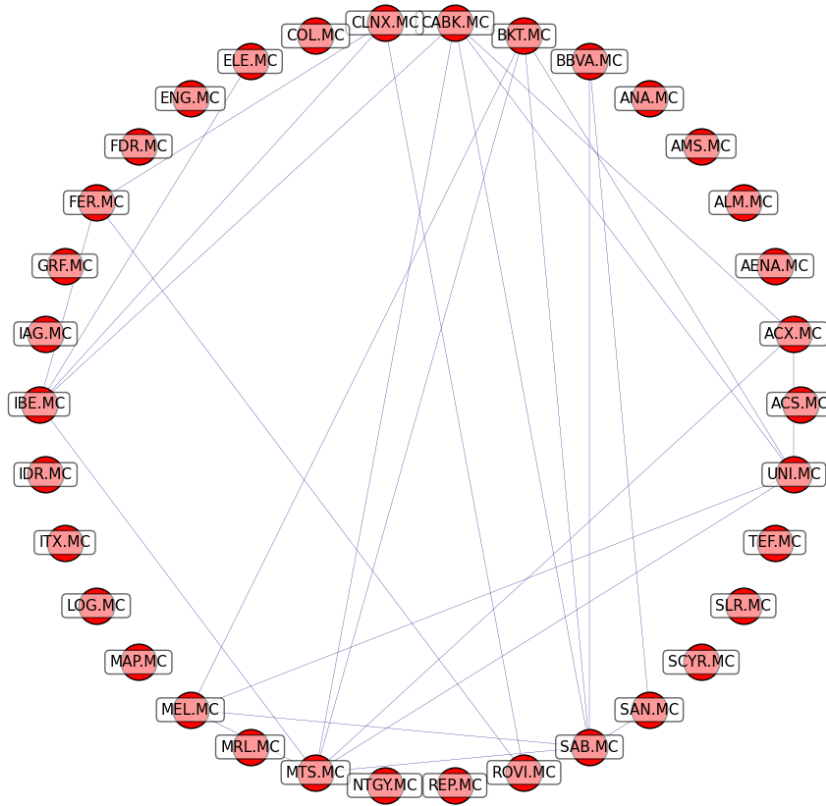


Figura 5.2: Grafo asociado al período 2018-2020

Este grafo inicial para los años previos a la pandemia cuenta con un total de 34 nodos y 28 aristas. Debido a que existen nodos aislados en el grafo del período 2018-2020, como es el caso de empresas como Inditex, Mapfre o Telefónica, se trabajará con la componente conexas para, posteriormente, aplicar el algoritmo de Girvan-Newman. La componente conexas resultante para este período, una vez eliminados los nodos aislados, es la siguiente:

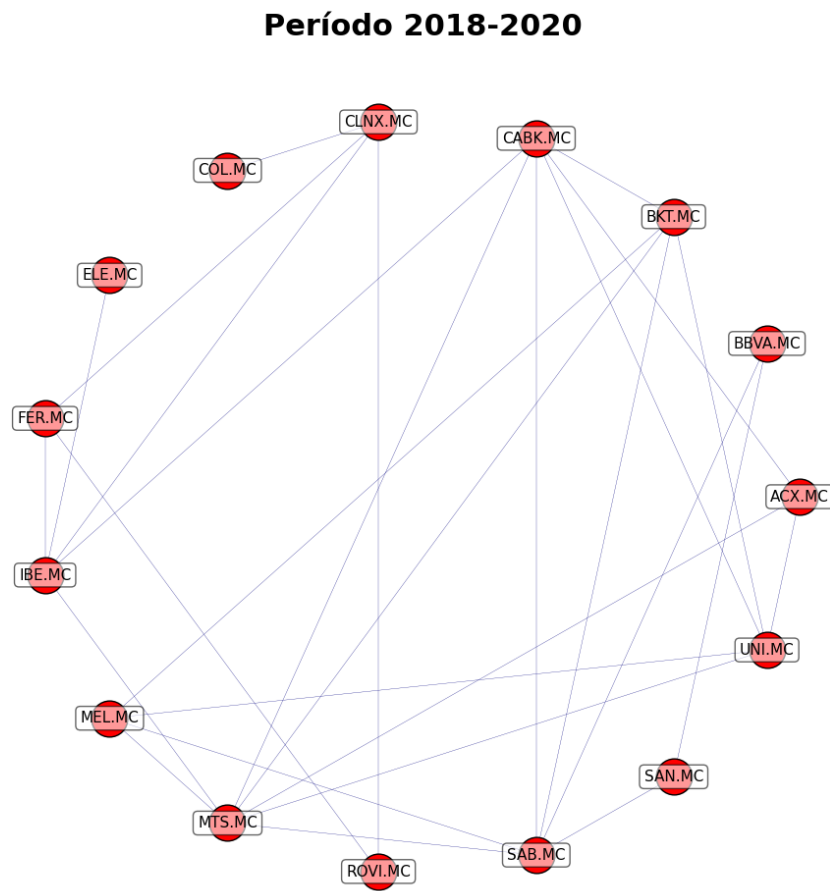


Figura 5.3: Componente conexa asociada al período 2018-2020

El grafo resultante que nos muestra la componente conexa asociada al período 2018-2020, una vez eliminados los nodos aislados, cuenta con un total de 15 nodos y 28 aristas. Tras la obtención del grafo asociado a la matriz de adyacencia correspondiente, se aplica el algoritmo Girvan-Newman para detectar las comunidades que surgen entre las empresas del IBEX 35.

Como se puede observar en la Figura 5.4, se ve cómo las empresas consideradas para el período 2018-2020 con un coeficiente correlación igual o superior a 0,9 se distribuyen de la siguiente manera:

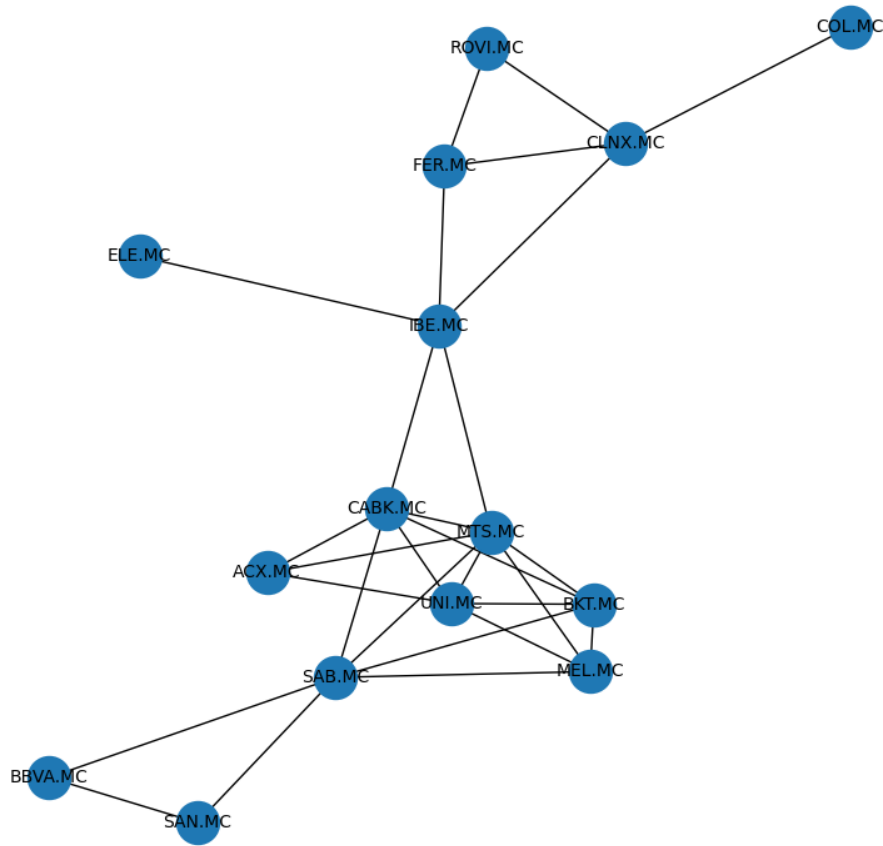


Figura 5.4: Componente conexa asociada al período 2018-2020

A simple vista, puede apreciarse cómo se identifican dos grupos o comunidades claramente distinguibles y por dónde empezará el algoritmo de Girvan-Newman a eliminar aristas, ya que será una de las aristas que unan estas dos comunidades la primera en ser eliminada, es decir, una de las aristas que une el nodo correspondiente a la empresa Iberdrola (IBE) con la comunidad formada en la parte inferior.

Una vez obtenida la componente conexa del grafo inicial para el período 2018-2020, se aplica el algoritmo Girvan-Newman con el objetivo de detectar las diferentes comunidades que se forman. Con el fin de ver cómo se han ido creando esas comunidades y cuáles han sido las aristas que se han ido eliminando una vez que se aplica el algoritmo Girvan-Newman, se ha decidido representar los resultados en forma de dendrograma.

En la figura 5.5, se puede ver cuál es el dendrograma resultante para el período de tiempo analizado.



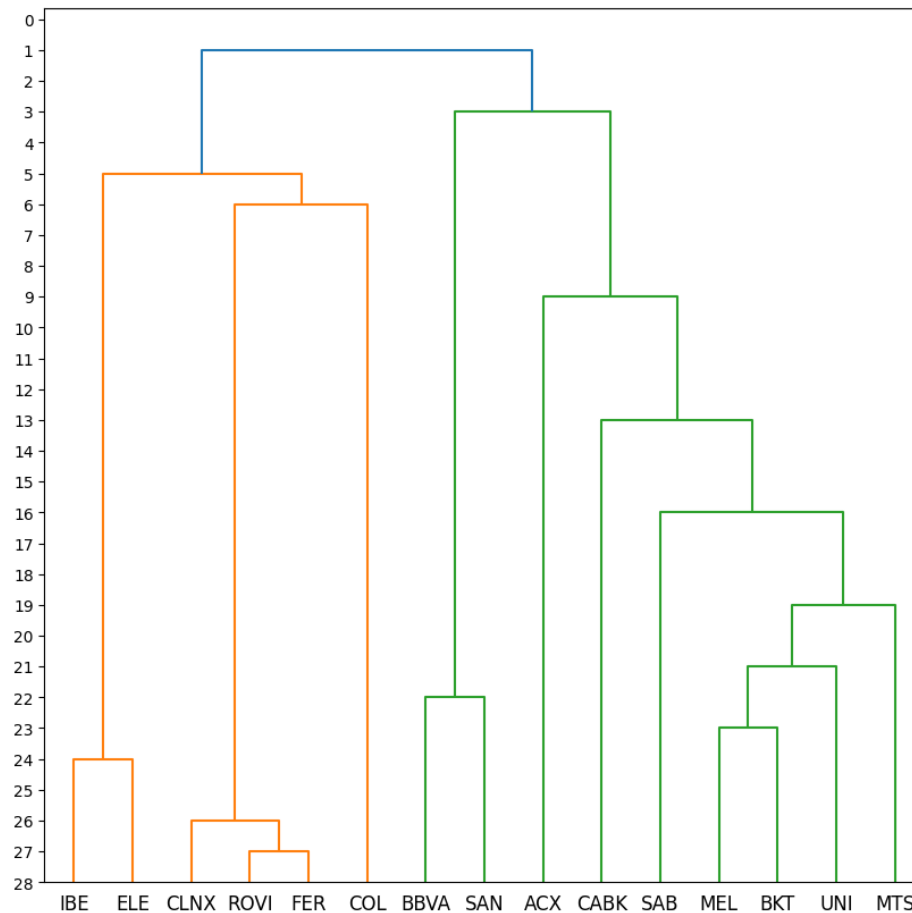


Figura 5.5: Dendrograma asociado al período 2020-2022.

Para interpretar a través del dendrograma cómo se han ido formando las comunidades es necesario fijarnos de abajo a arriba. En este caso, las empresas que se unen en primer lugar son las empresas Laboratorios ROVI (ROVI) y Ferrovial (FER). A continuación, estas dos empresas se unen a la empresa Cellnex Telecom (CLNX), por lo que tendríamos las tres empresas unidas. Las siguientes empresas en unirse con las Iberia y Endesa. Todas estas empresas, coloreadas en el dendrograma de amarillo, formarán parte de una misma comunidad. Respecto a la comunidad coloreada en verde en el dendrograma de la Figura 5.5, las primeras empresas que se unen son Meliá Hotels International (MEL) con Bankinter (BKT) para después unirse sucesivamente con Unicaja, ArcelorMittal y el Banco Sabadell, entre otras.

Por una parte, se aprecia cómo una comunidad está formada por las empresas del sector financiero - BBVA, Bankinter (BKT), CaixaBank (CABK), Banco Sabadell (SAB), Banco Santander (SAN) y Unicaja Banco (UNI) - unidas a Acerinox (ACX) y ArcelorMittal (MTS), pertenecientes al sector del mineral, metales y transformación, así como Meliá Hotels International (MEL) del sector

del ocio, turismo y hostelería.

La otra comunidad que se forma según la Figura 5.5 será la compuesta por las empresas Iberia (IBE), Endesa (ELE), Ferrovial (FER), Laboratorios ROVI (ROVI), Inmobiliaria Colonial (COL) y Cellnex Telecom (CLNX).

## 5.2. Período post-pandemia (2020-2022)

Durante los años 2020 y 2022, se han producido una serie de acontecimientos históricos que han afectado considerablemente a la economía mundial, produciéndose una profunda recesión tras el estallido de la pandemia causada por la COVID-19. Por tanto, es lógico pensar que esta grave e inesperada crisis a nivel mundial haya tenido efectos sobre las economías nacionales de cada país, afectando también a la composición de los mercados financieros y a la interdependencia de ellos.

Para este período de tiempo, el umbral que se ha escogido para calcular la matriz de adyacencia que dará lugar a nuestro grafo es de 0,83. Es decir, dos empresas estarán conectadas si y solo si el coeficiente de correlación entre ellas es mayor o igual a 0,83. En caso contrario, no habrá una arista conectando sus respectivos nodos. En la figura 5.6 se puede apreciar cómo varían el número de aristas en función del umbral

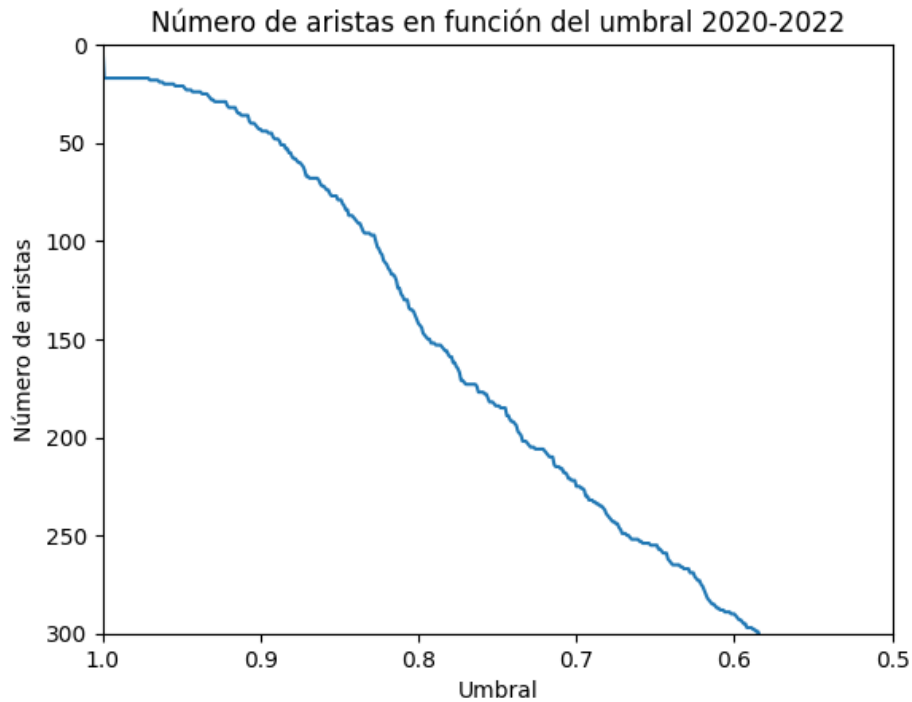


Figura 5.6: Umbral óptimo asociado al período 2020-2022.

El grafo resultante que nos queda tras haber aplicado el umbral, con un total de 80 aristas, puede observarse en la Figura 5.7. Como se puede percibir a simple vista, se observa que las empresas pertenecientes al sector de bancos y cajas de ahorro tienen un alto número de conexiones con el resto de empresas del IBEX 35, como ocurre en el caso de Bankinter, Santander, BBVA, CaixaBank, Sabadell

o Unicaja. Además, empresas como Repsol, Mapfre, Meliá Hotels International o Inditex, pertenecientes a los sectores del petróleo, seguros, ocio, turismo y hostelería y textil, vestido y calzado respectivamente, también destacan en este período por tener gran relación con diferentes empresas. Además, empresas del sector de la electricidad y gas como Endesa, Enagás o Iberdrola, así como Solaria Energía y Medio Ambiente pertenecientes al ámbito de las energías renovables, no tienen conexión con ninguna otra empresa.

**Grafo 2020-2022**

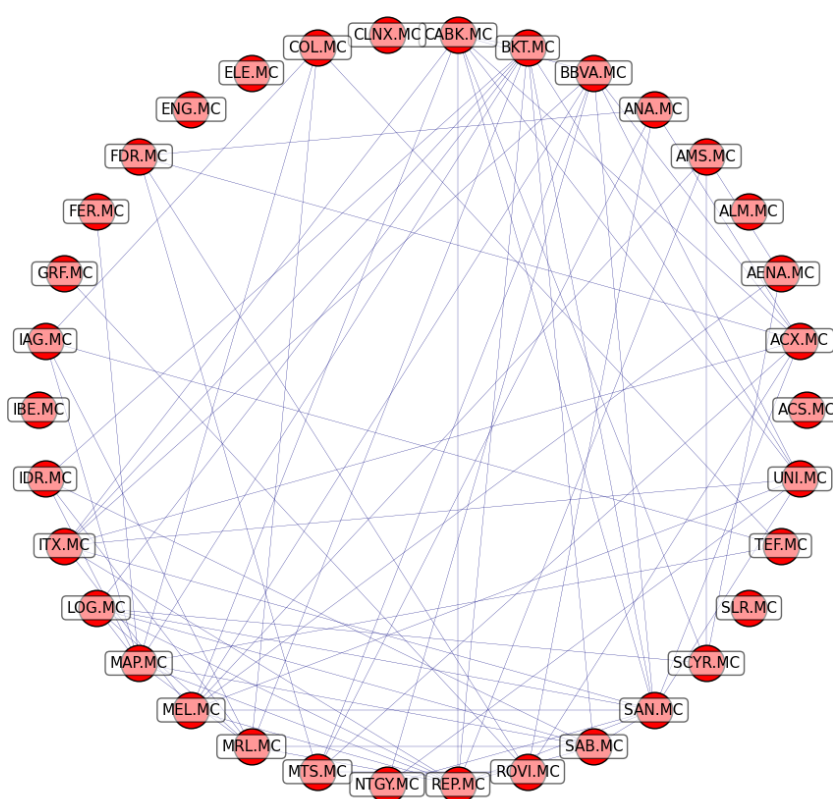


Figura 5.7: Grafo asociado al período 2020-2022

Al igual que en el período anterior, eliminaremos los nodos aislados que forman parte del grafo asociado al período 2020-2022, ya que estos nodos formarán ellos mismos una comunidad. En la Figura 5.9 se puede observar cuál es el grafo sobre el que trabajaremos una eliminados los nodos aislados, es decir, la componente conexa para el período comprendido entre los años 2020 y 2022.

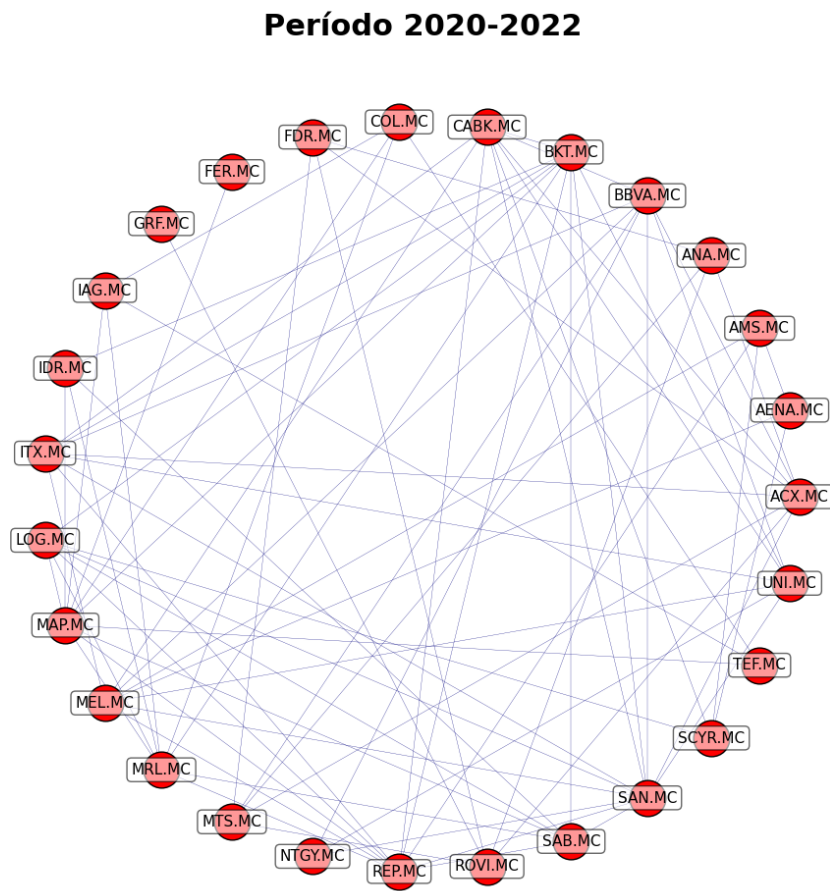


Figura 5.8: Componente conexa asociada al período 2020-2022

En este caso, se observa cómo la componente conexa para el período 2020-2022 no se diferencia tanto del grafo original del que partíamos, es decir, no tiene tantos nodos aislados como ocurría en el período pre-pandemia y cuenta con un mayor número de vértices que la componente conexa asociada al período estudiado previamente (2018-2020). Esta nueva componente conexa cuenta con 27 nodos, es decir, prácticamente con la totalidad de las empresas de las que partíamos para el estudio.

En la Figura 5.9 se representa gráficamente cómo se distribuyen las empresas que forman parte de la componente conexa para el período post-pandemia y que poseen un coeficiente de correlación entre sus precios de cotización igual o superior a 0,83.

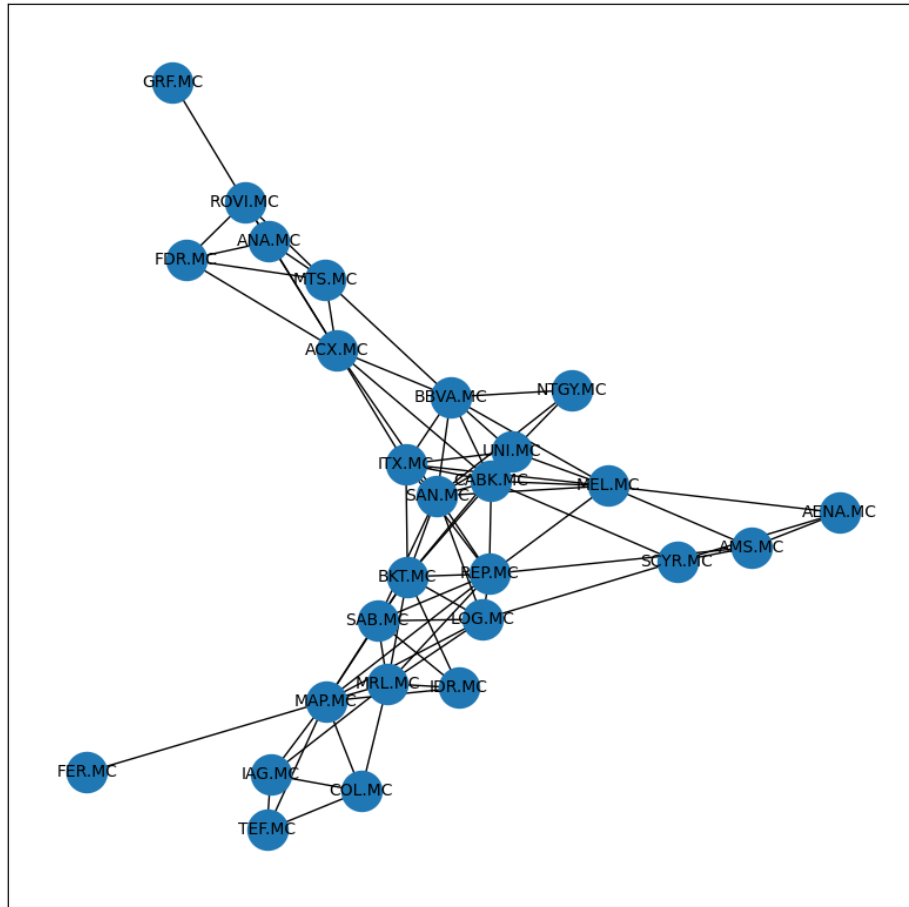


Figura 5.9: Distribución de la red compleja para el período 2020-2022

A simple vista puede apreciarse que, claramente, las empresas se agrupan en diferentes comunidades por cómo se distribuyen en el plano y que no son una única comunidad.

Una vez creado el grafo para el período comprendido entre 2020 y 2022, procedemos a aplicar el algoritmo Girvan-Newman con objeto de identificar las distintas comunidades que se crean entre las empresas del Ibex 35 contenidas en la componente conexa.

Al igual que ocurría en el período anterior, se utilizará un dendrograma para mostrar el proceso de creación de las comunidades de manera detallada, con el objetivo de entender cómo se han ido creando estas diferentes comunidades, como así se puede observar en la Figura 5.10.

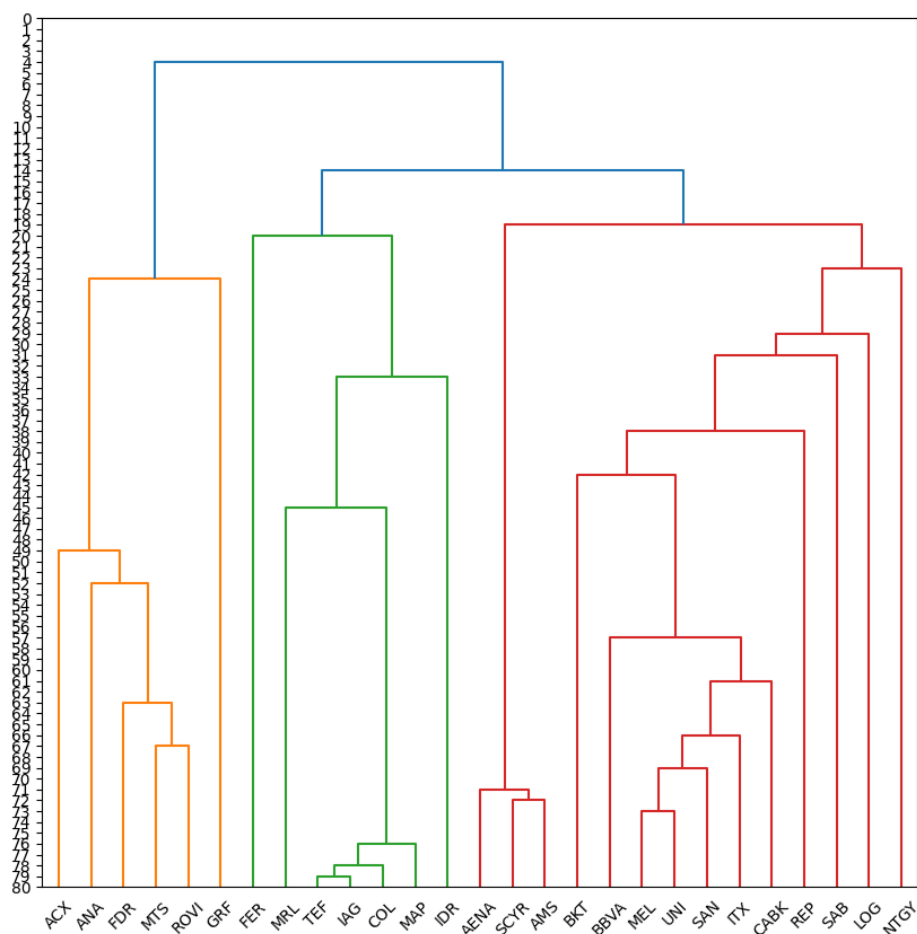


Figura 5.10: Dendrograma asociado al período 2020-2022

De la misma forma que para el período anterior, podemos ver cómo en la Figura 5.10 se muestran en diferentes colores las 3 comunidades principales que se han creado. Las primeras empresas que se han unido, en este caso y si nos fijamos en la tonalidad verde del dendrograma, son Telefónica (TEF) e IAG. A continuación, estas dos empresas que ya se han unido se juntarán a la Inmobiliaria Colonial (COL), para posteriormente unirse a la empresa de seguros Mapfre (MAP), y así sucesivamente.

De manera general vemos cómo se han creado 3 comunidades distintas que analizaremos a continuación. En primer lugar, la comunidad representada en una tonalidad amarilla está formada por un total de 6 empresas, que son las siguientes: Acerinox (ACX), Acciona (ANA), Fluidra (FDR), ArcelorMittal (MTS), Laboratorios Rovi (ROVI) y Grifols (GRF). Estas empresas son de diversos sectores, aunque encontramos ciertos sectores comunes, como puede ser el sector de los minerales, metales y transformación al cual pertenecen las empresas Acerinox y ArcelorMittal, el sector farmacéutico al que pertenecen Laboratios ROVI y Grifols y, por último, las empresas Acciona y Fluidra que son de los sectores de la

construcción y la ingeniería respectivamente.

En segundo lugar, con una tonalidad verde, está la siguiente comunidad formada por un total de 7 empresas. En este caso, las empresas que conforman la comunidad son Ferrovial (FER), Merlin Properties (MRL), Telefónica (TEF), IAG, Inmobiliaria Colonial (COL), Mapfre (MAP) e Indra Sistemas (IDR). Dos de estas empresas, Merlin Properties e Inmobiliaria Colonial, pertenecen al sector SOCIMI o al grupo de inversión en activos inmobiliarios. Además, Telefónica pertenece al sector de las telecomunicaciones, Indra al sector de la electrónica y software, IAG al sector del transporte y distribución y Mapfre pertenece al sector de los seguros.

Por último, la comunidad que se muestra en color roja es la más extensa de todas, con un total de 14 empresas. Esta comunidad está compuesta por todas las empresas del sector bancario: Bankinter (BKT), BBVA, Unicaja Banco (UNI), Banco Santander (SAN), CaixaBank (CABK) y Banco Sabadell (SAB). Asimismo, diferentes empresas relacionadas con el sector de la electricidad, gas y petróleo también están incluidas en esta comunidad, como es el caso de Naturgy (NTGY), Repsol (REP). El resto de empresas que forman parte de esta comunidad son AENA y Logista, ambas del sector del transporte y distribución, Sacyr del sector de la construcción, Amadeus IT Group, del sector de la electrónica y software, y los grupos Inditex y Meliá Hotels International, del sector del textil y hostelería respectivamente.



# 6

## Conclusiones y trabajos futuros

En este capítulo se detallan las conclusiones derivadas del TFG y la propuesta de posibles trabajos futuros.

### 6.1. Conclusiones

Una vez concluido el análisis del impacto de la pandemia en la detección de comunidades en el mercado de valores español y explicados los resultados obtenidos, se pasa a extraer las principales conclusiones del trabajo:

- En primer lugar y, a pesar de que los umbrales óptimos para ambos períodos eran muy similares, variando apenas 7 centésimas entre ellos, se ha comprobado que los grafos iniciales de los que partía el estudio eran muy diferentes. En el período comprendido entre 2018 y 2020, el grafo contaba con un total de 46 aristas mientras que en el período posterior a la pandemia, entre los años 2020 y 2022 el número de aristas del grafo inicial casi se duplicó, con un total de 80. Interpretando estos resultados, las empresas del IBEX 35 antes del COVID-19 no mostraban tanta relación ni existía tanta interdependencia entre ellas. Tras la pandemia, el mercado de valores español está más cohesionado e interconectado, de tal manera que si una empresa

experimenta una caída o subida en bolsa, es probable que esta afecte a más empresas que si esto mismo ocurriese en el período anterior a la pandemia.

- De manera similar en ambos períodos de tiempo, las empresas del sector financiero juegan un papel clave en el mercado financiero de España, siendo las empresas que forman parte de este sector algunas de las que más número de conexiones presentan con el resto de empresas del grafo. En otras palabras, el comportamiento de las principales empresas del sector de la banca y cajas de ahorros es representativo o nos puede dar una idea del comportamiento que presenta la economía española en general, debido a la importancia de estas empresas en el IBEX 35. En el lado contrario, empresas como Repsol, Mapfre o Inditex, que en el período previo a la crisis del COVID-19 no mostraban enlaces con el resto de empresas, tras la pandemia forman parte del grupo de empresas con mayor cohesión con respecto al resto de compañías del IBEX 35. Lo mismo ocurre con Iberdrola, una compañía que entre los años 2018 y 2020 estaba conectada con un gran número de empresas y, a partir del año 2020 y hasta el 2022, no presenta conexiones con ninguna de las empresas del IBEX 35. Estos cambios de tendencias mencionados reflejan la inestabilidad de los sectores que representan estas compañías mencionadas, como lo son los sectores del petróleo, seguros, textil y electricidad y gas, ya que se ha comprobado que a lo largo del tiempo su influencia e importancia dentro de un mercado financiero cambia bruscamente.
- En lo relativo a las comunidades detectadas tras la aplicación del algoritmo Girvan-Newman, también se han percibido notables diferencias en función del período de tiempo analizado.
  - De manera general, se concluye que el impacto de la pandemia sobre las empresas analizadas sí ha afectado, estando las empresas del IBEX 35 más conectadas entre ellas tras el impacto del Covid-19. Tanto en el período previo a la pandemia, como en el período posterior a ella, todas las empresas del sector bancario que forman parte del IBEX 35 están presentes en la misma comunidad y están interrelacionadas entre ellas, lo que nos sugiere una tendencia común que se puede extraer como conclusión del estudio: estas empresas tienen comportamientos muy similares y se pueden anticipar tomas de decisiones en función al comportamiento de ellas. Además, la empresa del sector de la hostelería y el turismo Meliá Hotels International también se encuentra en ambos períodos relacionada con las empresas del sector financiero y compartiendo comunidad.
  - Compañías como Repsol o Naturgy, que en los años pre-pandemia no formaban parte de ninguna de las comunidades formadas, tras el estallido del Covid-19 están estrechamente unidas a las empresas del sector bancario. Este dato es muy relevante, ya que tras la pandemia y

el conflicto bélico entre Ucrania y Rusia, los precios de la energía han presentado una tendencia al alza, alcanzando en muchos casos récords y convirtiéndose en empresas que han estado en el punto de mira. Su relación y pertenencia a la misma comunidad que las empresas del sector bancario es un hallazgo importante, pudiendo encontrar similitudes entre sus precios de cotización en la bolsa. En el período anterior a la pandemia, las dos empresas del sector energético que formaban parte del estudio (Iberdrola y Endesa) no tenían relación con las empresas del sector financiero, por lo que se muestra un cambio de tendencia.

- ArcelorMittal, que en el período previo a la pandemia formaba parte de la misma comunidad que las empresas del sector financiero y Meliá, desaparece en el período post-pandemia, no encontrando el suficiente grado de correlación como para poder hablar de una tendencia similar a ninguna de las empresas que forman parte de las 3 comunidades existentes.
- Acerinox, que en la etapa previa a la pandemia forma parte de la misma comunidad que el resto de empresas del sector financiero, en el período posterior a la pandemia muestra un comportamiento diferente, al estar una comunidad diferente a la que forman las empresas del sector bancario.

Como conclusión final, se ha analizado cuál ha sido el impacto de la pandemia en la formación de comunidades en el mercado de valores de España y cómo este impacto ha afectado a las distintas compañías del IBEX 35. A pesar de que, a grandes rasgos, las empresas más influyentes del mercado financiero se han mantenido a lo largo del tiempo, sí que es interesante comprobar cómo las conexiones de ciertas empresas del sector de la electricidad y el gas, como Iberdrola, o del sector petróleo, como Repsol, con respecto al resto de empresas del IBEX 35 han dado un giro total, poniendo en relevancia la vulnerabilidad de estos sectores ante las crisis o acontecimientos que tienen lugar en la economía nacional. Además, se ha comprobado que hay sectores que muestran estabilidad en el tiempo, siendo estos sectores claves de cara a los inversores, ya no solo a la hora de gestionar sus carteras y decidir en qué invertir, sino a los escenarios que se pueden prever de cara al futuro y cómo estos impactarán en el conjunto global del mercado de valores.

## 6.2. Trabajos futuros

Sobre los posibles trabajos futuros que podrían desarrollarse a partir de este estudio o como continuación del mismo, se plantean varias opciones con diferentes escenarios:

- Realizar un estudio en la misma franja temporal con otro índice bursátil, como pueden ser el Eurostoxx 50 (índice bursátil a nivel europeo), el SP 500 (Estados Unidos), el SSE Composite Index (China) o el BSE Sensex (India), para así tratar de comprobar si se crean patrones en las comunidades detectadas entre índices bursátiles y, por tanto, ver si el impacto de la pandemia ha sido similar o no en los principales índices bursátiles del mundo.
- Analizar de manera más exhaustiva el impacto que ha tenido el COVID-19 en la economía española ampliando este estudio no solo a las empresas del IBEX 35, sino a todas aquellas empresas que cotizan en bolsa en España con el objetivo de sacar conclusiones más consistentes y amplias.
- Analizar con más detalle los períodos de tiempo analizados, llegando incluso a identificar en el período post-pandemia la influencia de la guerra de Ucrania en la detección de las comunidades.

# Bibliografía

- [1] B. A. Memon and H. Yao, “Structural change and dynamics of pakistan stock market during crisis: A complex network perspective,” *Entropy*. [Online]. Available: [http://snap.stanford.edu/class/cs224w-2015/projects.2015/Network\\_Analysis\\_of\\_the\\_Stock\\_Market.pdf](http://snap.stanford.edu/class/cs224w-2015/projects.2015/Network_Analysis_of_the_Stock_Market.pdf)
- [2] B. M. Tabak, T. R. Serra, and D. O. Cajueiro, “Topological properties of stock market networks: The case of brazil,” *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 16, pp. 3240–3249, 2010. [Online]. Available: <https://doi.org/10.1016/j.physa.2010.04.002>
- [3] W. Sun, C. Tian, and G. Yang, “Network analysis of the stock market,” *Entropy*, 2015. [Online]. Available: [http://snap.stanford.edu/class/cs224w-2015/projects.2015/Network\\_Analysis\\_of\\_the\\_Stock\\_Market.pdf](http://snap.stanford.edu/class/cs224w-2015/projects.2015/Network_Analysis_of_the_Stock_Market.pdf)
- [4] J. Park, C. H. Cho, and J. W. Lee, “A perspective on complex networks in the stock market,” *Frontiers in Physics*, vol. 10, p. 1329, 2022.
- [5] S. Fortunato and C. Castellano, “Community structure in graphs,” *arXiv preprint arXiv:0712.2716*, 2007.
- [6] V. Latora, V. Nicosia, and G. Russo, *Complex networks: principles, methods and applications*. Cambridge University Press, 2017.
- [7] D. Peña, *Análisis de series temporales*. Alianza, 2005.
- [8] S. R. Cajal, *Texture of the Nervous System of Man and the Vertebrates*. Springer Science & Business Media, 1999, vol. 1.
- [9] L. W. Swanson, *Brain architecture: understanding the basic plan*. Oxford University Press, USA, 2012.
- [10] E. Bullmore and O. Sporns, “Complex brain networks: graph theoretical analysis of structural and functional systems,” *Nature reviews neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.
- [11] B. Hellwig, “A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex,” *Biological cybernetics*, vol. 82, no. 2, pp. 111–121, 2000.

- [12] V. Braitenberg and A. Schüz, *Cortex: statistics and geometry of neuronal connectivity*. Springer Science & Business Media, 2013.
- [13] B. B. Averbeck and M. Seo, “The statistical neuroanatomy of frontal networks in the macaque,” *PLoS computational biology*, vol. 4, no. 4, p. e1000050, 2008.
- [14] S. Fortunato, “Community detection in graphs,” *Physics reports*, vol. 486, no. 3-5, pp. 75–174, 2010.
- [15] M. E. Newman, “Modularity and community structure in networks,” *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.
- [16] G. Everywhere, “Principales algoritmos de detección de comunidades,” <https://www.grapheverywhere.com/principales-algoritmos-de-deteccion-de-comunidades/>.
- [17] NetworkX, “Memgraph’s guide for networkx library,” <https://networkx.guide/algorithms/community-detection/girvan-newman/>, accedido 05-06-2023.
- [18] Z. Bodie, A. Kane, and A. J. Marcus, *Principios de inversiones*. McGraw-Hill, 2004.
- [19] F. S. Mishkin, J. G. M. Araiza, A. S. Vargas, I. Katz, and A. R. Porras, “Moneda, banca y mercados financieros,” 2014.
- [20] A. Calvo, J. A. Parejo, L. Rodríguez, and Cuervo, “Manual del sistema financiero español,” *Ariel*, 2010.
- [21] V. Torre, “¿qué es el índice ibex 35?” <https://www.expansion.com/mercados/curso-invertir-bolsa/que-es-indice-ibex35.html>, accedido 27-04-2023.
- [22] —, “¿cómo funciona la bolsa?” <https://www.expansion.com/mercados/curso-invertir-bolsa/como-funciona-bolsa.html>, accedido 1-05-2023.
- [23] X. Fan Liu and C. K. Tse, “A complex network perspective of world stock markets: synchronization and volatility,” *International Journal of Bifurcation and Chaos*, vol. 22, no. 6, 2012. [Online]. Available: <https://doi.org/10.1142/S0218127412501428>
- [24] J. M. Marín and G. Rubio, *Economía financiera*. Antoni Bosch editor, 2001.



# Apéndice







## Código empleado

En esta sección, se proporcionarán los enlaces necesarios para poder acceder al código que se ha empleado en lenguaje Python para el desarrollo de este trabajo.

Con el objetivo de que sea accesible para el mayor número de personas y pueda consultarse siempre que se desee, el código se proporcionará a través de GitHub, una plataforma donde se pueden compartir proyectos de diversa índole. En este enlace no solo se encontrará el código de programación, sino que también se podrá acceder a las imágenes que se han utilizado como visualizaciones a lo largo del trabajo, especialmente aquellas relativas a las secciones de Metodología y Resultados.

El enlace para acceder al código empleado, así como al resto de archivos relevantes que se han utilizado durante este trabajo es el siguiente: <https://github.com/LuciaSalgadoBarrio/TFG-Matematicas.git>