Grafos de competitividad: detección de comunidades mediante el algoritmo de Girvan-Newman y aplicación al estudio del IBEX 35

Trabajo fin de grado

Grado en Matemáticas

Autor: Daniel Recuero Cordobés

Tutor: Regino Criado Herrero



Contenido

- Objetivos
- 2 Conceptos
 - Teoría de grafos: Grafos de competitividad
 - Detección de comunidades: Algoritmo de Girvan-Newman
- Desarrollo
- Resultados y conclusiones
- Mejoras y trabajos futuros

Objetivos del TFG

Contenido

El <u>objetivo general</u> de este trabajo es revisar conceptos y resultados de la teoría de grafos y encontrar conclusiones de valor que puedan aportar algún conocimiento sobre una posible inversión en alguna empresa del IBEX 35.

- Objetivos específicos
 - Exploración de conceptos básicos sobre la teoría de grafos.
 - Ampliación del conocimiento sobre la teoría de grafos mediante el análisis y la aplicación del algoritmo de Girvan-Newman a grafos de competitividad.
 - Extrapolación de conclusiones del estudio a situaciones reales.

3 / 37

- Aplicación de los conceptos teóricos a un caso práctico real.
- Análisis de objetos de la realidad desde una perspectiva matemática

Daniel Recuero Cordobés 23/02/2023

Primeras definiciones

<u>Definición 1</u> Un grafo simple G es un par (V(G),E(G)) donde V(G) es un conjunto finito no vacío de elementos llamados **nodos** y E(G) es un conjunto finito de pares no ordenados no repetidos de elementos de V(G) llamados **aristas**.

<u>Definición 2</u> Un camino en un grafo G=(V(G),E(G)) es una secuencia de aristas tal que el final de una es el inicio de la siguiente. Para denotarlo, se puede escribir entre paréntesis la secuencia de nodos o la secuencia de aristas y se puede referir a él con letras griegas.

<u>Definición 3</u> Se define la **longitud de un camino** en un grafo como el número de aristas que posee. Así pues, un camino β tendrá una longitud $I(\beta)$.

Daniel Recuero Cordobés 23/02/2023

4 / 37

Contenido

Grafos de competitividad I

Definición 4 Sea un conjunto de n elementos o competidores (que serán los nodos del grafo) $A = \{x, y, ..., n\}$. Se define la **puntuación** como una función $p: \{x,y,...,n\} \to \mathbb{R}$ donde a cada elemento del conjunto A se le asocia un número que tendrá distinta connotación según el contexto.

Definición 5 Dado un conjunto de n elementos o competidores $A = \{x, y, ..., n\}$, un **ranking** r es una función $r : \{1, 2, ..., n\} \rightarrow A$.

Observación 1 Dado un ranking r y un par de elementos $u_i, u_j \in$ $V(G_R)$, con $i \neq j$, $i, j \leq n$, se denota por $r(u_i) > r(u_i)$ la situación en la que el elemento u_i está situado por encima del elemento u_i en el ranking r.

También se escribe $r(u_i) \ge r(u_i)$ si $r(u_i) > r(u_i)$ o i = j.

$$r(i) \ge r(j) \Leftrightarrow p(i) \ge p(j)$$

Daniel Recuero Cordobés

Grafos de competitividad II

Contenido

Definición 6 Sea una familia de $m \in \mathbb{N}$ rankings $R = \{r_1, r_2, ..., r_m\}$, con todos ellos ordenando los mismos elementos de un conjunto $V(G_R)$. Dado un par de individuos u_i, u_i $i \in V(G_R)$ con $i \neq j$, $i, j \leq n$, se dice que ambos **compiten** si han intercambiado sus posiciones relativas en dos rankings consecutivos. Es decir, si existe k < m tal que $r_k(u_i) > r_k(u_i)$ y $r_{k+1}(u_i) > r_{k+1}(u_i)$, o viceversa.

Definición 7 Sea $R = \{r_1, r_2, ..., r_m\}$ una familia de rankings, con elementos del conjunto $V(G_R)$, se dice que dos nodos han competido h-veces si han intercambiado su posición relativa a lo largo de los m rankings en h ocasiones.

Daniel Recuero Cordobés 23/02/2023

Grafos de competitividad III

Contenido

Definición 8 Sea $R = \{r_1, r_2, ..., r_m\}$ una familia de rankings, con elementos del conjunto $V(G_R)$, se define el **grafo de competitividad** como el grafo simple pesado $G=(V(G_R), E(G_R))$ donde los nodos son los elementos competidores que aparecen en los rankings y cada arista tiene como peso el número de veces que han competido entre sí los correspondientes nodos adyacentes a la arista.

Formalmente, dados $u_i, u_j \in V(G_R)$ nodos del grafo, la arista tendrá un peso $w_{i,j}$ que vendrá dado por el número de veces que los nodos u_i, u_i compiten a lo largo de la familia de rankings R.

Daniel Recuero Cordobés 23/02/2023

7 / 37

Matriz de adyacencia del grafo de competitividad

Definición 9 Sea un grafo de competitividad $G=(V(G_R), E(G_R))$, se define la **matriz de adyacencia M** asociada al grafo G como la matriz cuadrada donde las filas y las columnas hacen referencia a los nodos del grafo de forma que:

$$M_{ij} = \left\{ egin{array}{ll} w_{i,j} & ext{si el nodo } u_i ext{ y el nodo } u_j ext{ compiten} \\ 0 & ext{en otro caso} \end{array}
ight.$$

donde hay que recordar que $w_{i,j}$ hace referencia al peso de la arista asociada a los nodos u_i , u_i .

Daniel Recuero Cordobés 23/02/2023

8 / 37

9 / 37

Coeficiente de correlación de Kendall

Definición 10 Sea $V(G_R)$ el conjunto de nodos con $|V(G_R)| = n$, ordenados en dos rankings r_1 y r_2 . Se define el **coeficiente de** correlación de Kendall como el cociente de la diferencia de los pares de nodos que compiten y no compiten en esos dos ranking entre el número de posibles pares que compitan en esos dos mismos rankings. Matemáticamente, resulta la siguiente expresión:

$$\tau(r_1, r_2) = \frac{\bar{K}(r_1, r_2) - K(r_1, r_2)}{\binom{n}{2}}$$

donde $\bar{K}(r_1, r_2)$ es el número de pares de nodos que no compiten con respecto de $\{r_1, r_2\}$ y $K(r_1, r_2)$ es el número de pares de nodos que sí compiten respecto de ambos rankings.

Daniel Recuero Cordobés 23/02/2023

Detección de comunidades

<u>Definición 11</u> Una comunidad dentro de un grafo es un conjunto de nodos caracterizados porque el número de conexiones entre los elementos de la comunidad es muy superior al número de conexiones que tienen con el resto de nodos del grafo.

Para la detección de comunidades dentro de un grafo existen muchos métodos. Para el análisis realizado en este trabajo se ha escogido el <u>algoritmo de Girvan-Newman</u> por ser uno de los más utilizados.

Daniel Recuero Cordobés

Algoritmo de Girvan-Newman I: Origen

El algoritmo de Girvan-Newman (2002) es una técnica para la detección y el análisis de comunidades dentro de un grafo. Este algoritmo fue resultado de los estudios de Michelle Girvan y Mark Newman.

El algoritmo se basa en la eliminación progresiva de las aristas con mayor coeficiente de intermediación (betweenness) del grafo original. Así, en las iteraciones siguientes, el algoritmo va generando las comunidades, que serán aquellos nodos que siguen conectados tras la eliminación progresiva de dichas aristas.

Daniel Recuero Cordobés 23/02/2023 11 / 37

Algoritmo de Girvan-Newman II: Elementos

La <u>intermediación de aristas</u>, que se basa en:

Contenido

- i) Sea $a \in E(G)$ una arista del grafo, se toman pares de nodos del conjunto V(G) y se calculan los caminos más cortos entre ellos (problema del camino mínimo entre nodos).
- ii) Con todos los caminos mínimos calculados, la intermediación de la arista a es el número de caminos seleccionados en el paso anterior que pasan por dicha arista a.
- iii) Repetir los pasos i) y ii) con el resto de aristas del conjunto E(G) para obtener la intermediación de todas las aristas.

La idea es calcular la influencia de cada arista sobre el flujo de información del grafo.

Daniel Recuero Cordobés 23/02/2023 12 / 37

Algoritmo de Girvan-Newman III: Definición

Los pasos del algoritmo son los siguientes:

- Calcular la intermediación de todas las aristas que hay en el grafo.
- Eliminar la arista que tenga la intermediación más alta. En caso de empate entre aristas, se toma una de ellas al azar.
- O Para el resto de aristas que queden en el grafo, calcular su intermediación.
- Repetir los pasos 2 y 3 hasta que no quede ninguna arista.

Algoritmo de Girvan-Newman IV: Pseudocódigo

```
Algoritmo 1 Algoritmo de Girvan-Newman
  repeat
    n = Total \ number \ of \ edges \ in \ graph \ G
    for i = 0 to n - 1 do
      Betweenness[i] = Betweenness of edge i
      if Betweenness[i] > Max\_Betweenness then
        Max\_Betweenness = Betweenness[i]
        Max\_Betweenness\_Edge = i
      end if
    end for
    remove edge Max_Betweenness_Edge from graph G
  until n=0
```

Mejoras y trabajos futuros

Algoritmo de Girvan-Newman V: Ejemplo

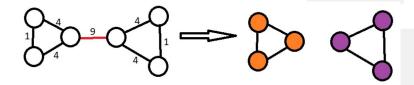


Figura: Ejemplo de iteración del algoritmo de Girvan-Newman

Daniel Recuero Cordobés 23/02/2023 15 / 37

Criterios

Contenido

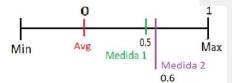
Para abordar el análisis se han definido 3 criterios o perspectivas de cara a construir tres familias de rankings distintas.

- Perspectiva 1: Bajo una inversión inicial de 1000 € en cada empresa, se ha obtenido el valor de la inversión realizada al final de cada sesión y con el precio de cierre de cada sesión se ha ido construyendo un ranking según el valor que iba tomando cada una de dichas inversiones.
- Perspectiva 2: Emplea la variación (ganancia o pérdida) de la inversión para cada valor para construir la familia de rankings, es decir, se toma la diferencia entre valores consecutivos y no el valor en sí.
- Perspectiva 3: Evalúa el porcentaje de ganancia o pérdida entre cada par de sesiones contiguas, tomando como valor de referencia el precio de cierre de la sesión anterior.

Daniel Recuero Cordobés 23/02/2023 16 / 37

Subvariantes

Debido a que los grafos obtenidos con las perspectivas tenían un volumen muy elevado de aristas, se han establecido tres valores umbrales para los pesos de las aristas, de forma que solo se consideren aquellas que superen dichos límites. Aparecen así 3 subvariantes para cada perspectiva.

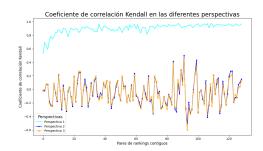


Período temporal del estudio

Previamente a la presentación de resultados y conclusiones hay que señalar que el período en el que se ha recogido la información ha sido el comprendido entre las sesiones del 22 de abril de 2022 y el 21 de octubre de 2022.

Daniel Recuero Cordobés

Coeficiente de correlación de Kendall



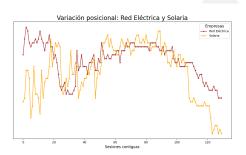
- La perspectiva 1 presenta valores cercanos a 1, lo que indica una alta correlación entre los rankings así como una baja competitividad entre los nodos.
- Las perspectivas 2 y 3 muestran variaciones muy similares para este coeficiente aunque no siempre idénticos. (Están bastante correladas)

Aplicación del algoritmo de Girvan-Newman I

- Para las empresas <u>Acciona</u> y <u>Rovi</u> se ha observado que cuando alguna de ellas presenta una posición extrema en los rankings relativos al estudio de la volatilidad, la otra empresa suele localizarse en una posición contraria en el ranking. Es decir, cuando una ocupa posiciones muy altas o muy bajas en el ranking, la otra suele estar en posiciones bajas o altas, respectivamente. Esto es derivado de las perspectivas 2 y 3.
 [27] [28] [29] [30] [31] [32] [33] [34] [35]
- Esta misma conclusión anterior se ha obtenido para los pares de empresas <u>Inditex</u> y Repsol, y Mapfre y <u>Meliá Hotels</u>.
- Análogamente, con <u>Acerinox</u>, <u>Acs</u> y <u>Arcelormittal</u> ocurre la misma situación, reforzada cuando dos empresas ocupan posiciones extremas y la tercera ocupa una posición contraria en el ranking.

Aplicación del algoritmo de Girvan-Newman II

Contenido



Del primer enfoque se ha extraído que Solaria y Red Eléctrica compartían comunidad debido a su alta competitividad.

Como conclusión, se ha observado que Red Eléctrica suele estar en posiciones superiores a Solaria en cuanto al valor de la inversión y que, cuando esta segunda empresa adelanta a la primera, solo unas sesiones después Red Eléctrica vuelve a ponerse por delante.

Daniel Recuero Cordobés 23/02/2023

21 / 37

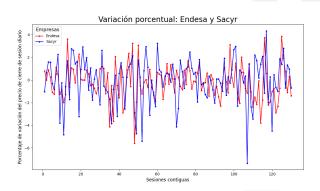
Aplicación del algoritmo de Girvan-Newman III

Se han estudiado las comunidades que se han formado a lo largo de las aplicaciones del algoritmo en todos los grafos generados para las tres empresas con mayor rentabilidad en el período: Siemens Gamesa, Caixabank y Bankinter. [36] Solo cabe destacar la agruación formada por Caixabank, Inditex y Repsol que ha generado en este período una rentabilidad conjunta del 7.718 %.

| Agrupaciones | ¥ | Rentabilidad | * |
|--|---|--------------|---|
| Siemens Gamesa + Fluidra | | -17.5269 | |
| Caixabank + Inditex + Repsol | | 7.7177 | |
| Caixabank + Inditex + Repsol + Solaria | | -2.1341 | |
| Bankinter + Indra + Iberdrola | | -0.9281 | |

Daniel Recuero Cordobés 23/02/2023 22 / 37

Aplicación del algoritmo de Girvan-Newman IV



Relativo a las empresas <u>Endesa</u> y <u>Sacyr</u> se ha observado que un movimiento ascendente o descente en el valor de una implica, en numerosas ocasiones, un movimiento en el mismo sentido en la otra empresa. Este suceso se ha dado en el 57,6 % de las sesiones.

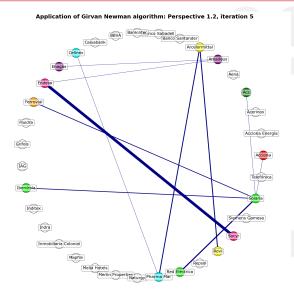
Mejoras y trabajos futuros

Objetivos

Contenido

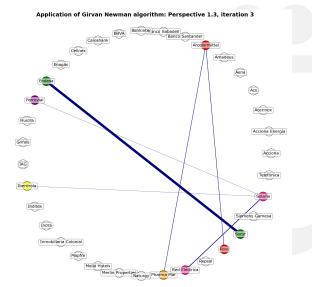
- Realizar un análisis análogo con el precio de apertura o con otra medida en torno al precio. Esto podría llevar a la detección de nuevos patrones.
- Extracción de conclusiones de comportamiento para pares o tríos de empresas similares a los realizados anteriormente en función de la volatilidad.
- Implementación de un método que represente la evolución de las posiciones de las empresas de interés a lo largo de las familias de rankings para intentar detectar algún patrón gráficamente.
- Aplicación de algún otro algoritmo de detección de comunidades para observar en qué se diferencian los resultados de los obtenidos a través del de Girvan-Newman.
- Replicar el último análisis realizado para aquellas empresas que más han aumentado su valor con las empresas que más valor han perdido.

Aplicación del algoritmo de Girvan-Newman: 1.2



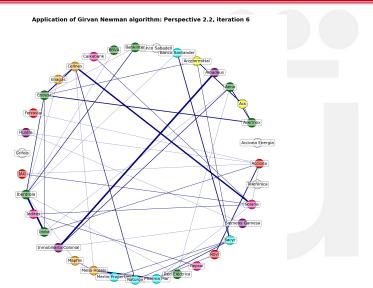
Daniel Recuero Cordobés 23/02/2023 25 / 37

Aplicación del algoritmo de Girvan-Newman: 1.3



Daniel Recuero Cordobés 23/02/2023 26 / 37

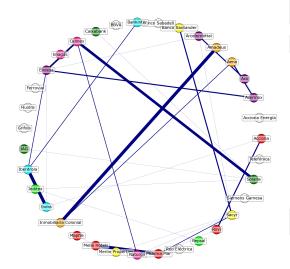
Aplicación del algoritmo de Girvan-Newman: 2.2



Daniel Recuero Cordobés 23/02/2023 27 / 37

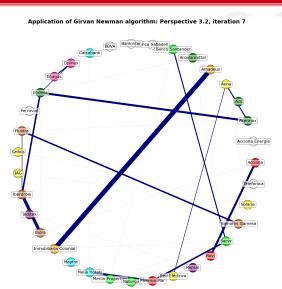
Aplicación del algoritmo de Girvan-Newman: 2.3

Application of Girvan Newman algorithm: Perspective 2.3, iteration 7

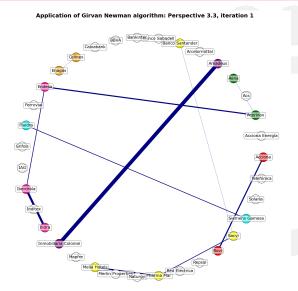


Daniel Recuero Cordobés 23/02/2023 28 / 37

Aplicación del algoritmo de Girvan-Newman: 3.2



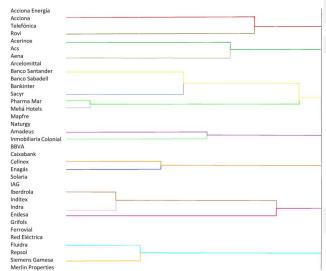
Aplicación del algoritmo de Girvan-Newman: 3.3



Daniel Recuero Cordobés 23/02/2023 30 / 37

Aplicación algoritmo de Girvan-Newman: 3.3

Dendograma de la aplicación del algoritmo de Girvan - Newman a la perspectiva 3.3



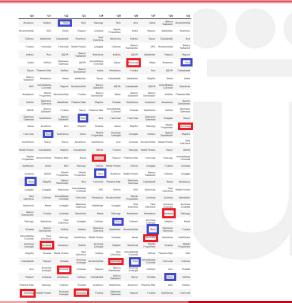
Daniel Recuero Cordobés 23/02/2023 31 / 37

Rankings perspectiva 2 (I): Acciona y Rovi



Daniel Recuero Cordobés 23/02/2023 32 / 37

Ranking perspectiva 2 (II): Acciona y Rovi



Daniel Recuero Cordobés 23/02/2023 33 / 37

Rankings perspectiva 3 (I): Acciona y Rovi



Daniel Recuero Cordobés 23/02/2023 34 / 37

Ranking perspectiva 3 (II): Acciona y Rovi



Daniel Recuero Cordobés 23/02/2023 35 / 37

Valor final de la inversión para cada empresa

| Valor | Nombre | | Valor | Nombre | |
|------------|-----------------------|----|-------------|-----------------|----|
| 825.269753 | Naturgy | 18 | 1177.110818 | Siemens Gamesa | 0 |
| 818.916734 | Red Eléctrica | 19 | 1132.686084 | Caixabank | 1 |
| 788.663283 | Arcelormittal | 20 | 1106.642729 | Bankinter | 2 |
| 749.244713 | Endesa | 21 | 1096.948819 | Inditex | 3 |
| 742.675781 | Enagás | 22 | 1055.232558 | Acciona Energía | 4 |
| 729.205176 | Merlin Properties | 23 | 1011.529593 | Banco Sabadell | 5 |
| 728.242560 | IAG | 24 | 1001.894657 | Repsol | 6 |
| 709.439033 | Aena | 25 | 998.701558 | BBVA | 7 |
| 703.800786 | Pharma Mar | 26 | 965.798046 | Acs | 8 |
| 698.240000 | Rovi | 27 | 952.432432 | Indra | 9 |
| 683,105238 | Solaria | 28 | 932.570494 | Ferrovial | 10 |
| 670.588235 | Telefónica | 29 | 913.080977 | Iberdrola | 11 |
| 638.015064 | Celinex | 30 | 909.523810 | Sacyr | 12 |
| 629.592566 | Meliá Hotels | 31 | 880.296610 | Acciona | 13 |
| 597.047970 | Inmobiliaria Colonial | 32 | 855.789474 | Mapfre | 14 |
| 501.047443 | Grifols | 33 | 854.292624 | Acerinox | 15 |
| 472.350230 | Fluidra | 34 | 843.200271 | Amadeus | 16 |
| | | | 839.545167 | Banco Santander | 17 |

Daniel Recuero Cordobés 23/02/2023 36 / 37

Grafos de competitividad: detección de comunidades mediante el algoritmo de Girvan-Newman y aplicación al estudio del IBEX 35

Trabajo Fin de Grado

Grado en Matemáticas - Curso 2022-2023

Autor: Daniel Recuero Cordobés

Tutor: Regino Criado Herrero

