Facultad de Ciencias

Departamento de Física de la Materia Condensada

Análisis de redes complejas: Aplicaciones en redes de colaboraciones científicas

Realizado por Francisco Bauzá Mingueza

Dirigido por:
Dr. Jesús Gómez Gardeñes
Dr. David Íñiguez Dieste

Master en Modelización e Investigación Matemática, Estadística y Computación

Índice

1. Abstract

En los últimos años, el estudio y desarollo de la teoría de redes complejas se ha vuelto cada vez más importante. La popularización de las redes sociales y la aplicación cada vez más frecuente de las redes complejas en el estudio de redes de diversa naturaleza ha propiciado este desarollo. Uno de los campos en los que ha resultado muy interesante la aplicación de la teoría de grafos y redes complejas es el estudio de redes de colaboración científica. Muchos autores han obtenido resultados que impulsan el uso de las redes complejas en este campo, como por ejemplo el efecto de "small world" que aparece en las redes de colaboración [1], la correlación entre colaboraciones de los investigadores en la red [2] o la influencia de la "betweenness centrality" en la generación de links con nuevos nodos [3]. En concreto, puede resultar interesante el estudio de la eficiencia y optimización de la colaboración entre investigadores, a través de la estructura de comunidades de la red, que va a ser el objeto de estudio de este trabajo.

En este trabajo, se ha realizado un análisis y comparación entre la estructura departamental y de comunidades de la red de colaboración académica del área de Ciencias de la Universidad de Zaragoza. Este análisis se ha llevado a cabo mediante el uso de magnitudes y procedimientos típicos de detección de comunidades en redes complejas, como son la Modularidad [4] y el algoritmo de Girvan-Newman [5] o típicos de la teoría de similaridad de conjuntos, como el índice de similaridad de Wallace [6].

La red de colaboración ha sido modelizada usando la firma conjunta de investigadores en artículos publicados y el impacto JCR de estos artículos. Los datos han sido obtenidos de las bases de datos de la Universidad de Zaragoza, a través de la herramienta Kampal Research [7].

Este trabajo se basa en la asunción de que cuanto mayor es la modularidad de la estructura de comunidades de la red de colaboración científica, más eficiente es esta estructura potenciando la colaboración fructífera de investigadores.

La partición de mayor modularidad se ha obtenido de las particiones de comunidades generadas con el algoritmo de Girvan-Newman. Además, hemos utilizado los índices de similaridad para tener presente durante el proceso, cuánto se parecen cada una de estas particiones en comunidades a la estructura departamental. Aparte de la partición de máxima modularidad, hemos obtenido también la de máxima similaridad con la estructura departamental y aquella que con el menor número de comunidades obteníamos un valor de modularidad considerablemente alto. Finalmente, se ha utilizado un método heurístico para reducir considerablemente el número de comunidades de la partición de máxima modularidad, sin que esta se vea apenas disminuida.

In recent years, the study and development of the theory of Complex Networks, have become more and more important. Popularization of Social Networks and the increasingly frequent application of Complex Networks to the study of networks of diverse nature, has contributed to this development. One of the fields where the application of Graph and Complex Networks theory, has turned out really interesting, is the study of scientific collaboration networks. Many authors have obtained results which promote the use of Complex Networks in this field, such as the "small world" effect that appears in collabration networks, the correlation between collaborations of researchers or the influence of "betweenness centrality" in the generation of links with incoming nodes. Specifically, it may turn out interesting the study of efficiency and optimization of researchers collaboration, through the community structure of the network, which is going to be the object of study for this work.

Here, we have performed and analysis and comparison between the departamental and communities structure collaboration network within the University of Zaragoza, focusing on the science area. This analysis has been carried out by using typical magnitudes and procedures from complex networks, such as Modularity and Girvan-Newman algorithm or typical from sets similarity theory such as Wallace index.

Collaboration network has been modelled using researchers coauthorship in published papers and the JCR impact of these papers. These data have been obtained from University databases, through Kampal Research tool.

This work is based on the assumption that the higher is the modularity of the community structure of a network, the more efficient is this structure improving the already existing profitable collaborations between researchers.

The maximum Modularity partition has been obtained from communities partitions generated by menas of the Girvan-Newman algorithm. In addition, we have used similarity indexes so as keeping in mind how similar are these partitions to the departamental structure. Appart from the maximum modularity partition, we have also obtained the partition with maximun similarity to departamental structure and that one having an acceptable modularity value with the lowest number of comunnities. Finally, a heuristic method has been used to reduce considerably the number of communities of maximum modularity partition, keeping its value really near to the maximum.

2. Introducción

La estructura y la naturaleza de las colaboraciones entre investigadores en una universidad es propicia para aplicar la teoría de redes complejas. Una red compleja es un grafo que modeliza las interacciones (links) entre individuos (nodos) que forman parte de un sistema, siempre y cuando la estructura y/o distribución de estas interacciones no sea sencilla o reducible al estudio por separado de las partes del sistema.

El estudio y aplicación de la redes complejas se puede dividir, a grandes rasgos, en el estudio *estructural* o *topológico* y el *dinámico*. La parte topológica se centra en la características estáticas de la red como las propiedades locales y globales de las interacciones, o la estructura mesoscópica de la red (formación de comunidades o clusters), en cambio, la parte dinámica se centra en propiedades de la red que van evolucionando, como la aparición y el cambio de las interacciones entre individuos o la variación del estado de los nodos debido a las interacciones (aplicación de redes complejas al estudio de contagio de epidemias [8]). Nosotros hemos realizado un estudio exclusivamente topológico de la red de colaboración científica.

El principal objeto de estudio de este trabajo es la estructura departamental de las distintas áreas de conocimiento que existen en las universidades, en este caso concreto, en la Universidad de Zaragoza. Una de las característica u objetivos de esta estructura departamental es fomentar la colaboración y producción científica de los investigadores que la componen. Esta idea de colaboración intradepartamental e interdepartamental es la que ha impulsado el análisis y la metodología de este trabajo.

La estructura departamental tiene similitudes con la estructura mesoscópica de las redes complejas, es decir, la estructura de comunidades. La comunidades son agrupaciones de nodos que tienen propiedades similares, cuya interacción es mayor que con otros nodos de la red y/o los nodos que las forman tienen roles similares dentro de la red. Del estudio de las interacciones entre nodos de una comunidad y con nodos de otras comunidades surge el concepto de Modularidad, que es esencial en el desarrollo de este trabajo. El objetivo entonces, es obtener particiones en comunidades de la red cuyos valores de modularidad sean altos, para ello vamos a utilizar un algoritmo de detección de comunidades desarrollado por Girvan y Newman.

El hecho de que se vaya a trabajar con particiones distintas de la misma red (comunidades y departamentos) hace que sea necesaria también, la aplicación de conocimientos propios de la similaridad de conjuntos. En principio se pensó en utilizar la teoría de similaridad de grafos, pero después de consultar algunos resultados y estudios sobre similaridad de grafos (redes complejas) [9] de la literatura, se descartó la idea de aplicar este tipo de procedimientos. Se descartó porque aquí, no se están comparando redes/grafos distintas, sólo se comparan particiones distintas de la misma red, pero los nodos y los links no cambian. Por lo tanto, se eligieron procedimientos y magnitudes del ámbito de comparación de particiones de conjuntos. En concreto, nos hemos

centrado en 3 índices de similaridad que sirven para comparar dos particiones de individuos de un conjunto, los índices de Wallace, Rand y Jaccard. Estos índices no se aplican solo a particiones de redes o grafos y por lo tanto los links de la red no juegan ningún papel en su definición. Además, el número de comunidades o clusters que forman ambas particiones no tiene por qué ser el mismo, sólo es necesario que ambas particiones tengan los mismos nodos y que cada nodo pertenezca a una y sólo una comunidad en ambas particiones.

Para la realización de este estudio, se han utilizado datos del área de Ciencias de la Universidad de Zaragoza. Los datos de los que disponíamos abarcaban la publicación de artículos entre 1970 y Agosto de 2017, con el correspondiente impacto JCR y los investigadores que firmaban dichos artículos, siempre y cuando uno de estos investigadores perteneciese al área de Ciencias. También se conocía el departamento al que estaban adscritos cada uno de estos investigadores. A la hora de generar la red, solo tuvimos en cuenta aquellos investigadores cuyo departamento perteneciese a la Facultad de Ciencias, es decir, eliminamos los investigadores que no tenían adscripción a ningún departamento o que su departamento pertenecía a otra área de conocimiento y habían firmado artículos con algún investigadores de ciencias. No tuvimos en cuenta tampoco los links que pertenecían a artículos sin impacto JCR o con impacto JCR igual a 0, que correspondían a revistas sin valoración JCR o que aún no habían sido valoradas. Tampoco formaron parte de la red aquellos investigadores que habían publicado artículos de impacto pero nunca con otro investigador, ya que estos hubiesen aparecido como nodos sueltos en la red.

En total, teníamos 1287 investigadores de los cuales 112 no habían publicado artículos, 77 habían publicado artículos sin impacto JCR, 61 investigadores publicaron artículos que no firmaron conjuntamente con otros investigadores y de los restantes 30 no estaban adscritos a departamento de la facultad de Ciencias. Finalmente quedaban 1007 investigadores que forman parte de la red, con un total de 4983 links entre ellos.

En la tabla 1 se muestran los departamentos que forman la red y cuántos de los 1007 investigadores hay en cada uno de estos departamentos.

Departamento	Número de investigadores	
	Red entera	**Cluster gigante**[1]
Producción Animal y Ciencia de los Alimentos	153	141
Ciencias de la Tierra	125	125
Química Inorgánica	107	107
Física Teórica	97	92
Física de la Materia Condensada	91	91
Química Analítica	81	81
Química Orgánica	76	76
Matemática Aplicada	69	64
Física Aplicada	50	50
Matemáticas	47	33
Métodos Estadísticos	32	32
Química Física	32	32
Química Orgánica y Química Física	28	28
Didáctica de las Ciencias Experimentales	19	19

Cuadro 1: Tabla con los distintos departamentos y el número de investigadores en cada uno, que aparecen en la red compleja.

Los datos se han obtenido de las bases de datos de la Universidad de Zaragoza, a través de la aplicación Kampal Research. Kampal es una empresa spin-off de la Universidad de Zaragoza, creada por investigadores del BIFI[10] (Instituto Universitario de Investigación en Biocomputación y Física de Sistemas Complejos). Kampal se dedica a la aplicación de redes y sistemas complejos para el estudio de redes sociales y producción científica de las universidades de Zaragoza, principalmente.

La estructura del trabajo es la siguiente. Primero vamos a presentar la **formulación del problema**, es decir, introducir conceptos de teoría de redes complejas, explicar más en detalle cómo hemos modelizado la red y el peso de los links, cómo se calculan las magnitudes presentadas anteriormente y en qué consiste el procedimiento de Girvan-Newman. Posteriormente, en el apartado de **resultados**, vamos a explicar qué resultados hemos elegido presentar (gráficas, tablas y representaciones de redes) y por qué, presentaremos estos resultados y los comentaremos. A continuación, en el apartado de **obtención de la partición óptima**, explicaremos el método heurístico para reducir el número de comunidades de la partición de maxima modularidad y analizaremos la partición que nos queda. Por último, en el apartado de **conclusiones**, comentaremos las conclusiones que obtenemos de los resultados, si hay contradicciones con lo esperado y si se ha conseguido cumplir los objetivos propuestos.

[1]En el apartado de formulación del problema 3.2, está explicado que es el Cluster gigante, cual es la diferencia con la Red entera y en que caso es mejor usar el Cluster gigante en vez de la red entera.

2.1. Herramientas de software utilizadas

Antes de finalizar la introducción, vamos a resumir en este apartado qué herramientas de *software* se han utilizado en este trabajo.

Tanto la resolución del problema con el algoritmo de Girvan-Newman, como todo el proceso de tratamiento de datos, han sido implementados en **Python**. Para ello, hemos utilizado la librería *Networkx* [11], especializada en redes complejas y nos hemos ayudado del libro "*Data Science & Complex Networks*" [12], que es una introducción a la aplicación de Python para el tratamiento de Redes complejas.

Las distintas gráficas presentadas en el trabajo se han obtenido con **R** y la representación gráfica de redes se ha hecho con **Gephi**.

3. Formulación del problema

3.1. Introducción a las redes complejas

Toda la parte teórica y de presentación de conceptos de redes complejas en este trabajo está basada, principalmente, en los libros *Dynamical processes on complex networks* [13] y *Networks: An Introduction* [14].

La forma más común de representar matemáticamente una red compleja o un grafo es con su matriz de adyacencia $\mathbf{A}$, cuyas entradas A_{ij} toman el valor 1 si hay un link del nodo i hacia el nodo j y 0 en caso contrario. Si dos nodos están conectados entre sí, se denominan vecinos, es decir, el conjunto de vecinos del nodo i son aquellos nodos j que cumplen: $\{j|A_{ij} = 1 \vee A_{ji} = 1\}$.

La clasificación mas elemental de los tipos de redes es la siguiente:

- **Red dirigida y no dirigida:** Las redes complejas, al igual que las interacciones que modelizan, pueden ser dirigidas o no dirigidas. En una red dirigida, los links que unen los nodos no son simétricos, es decir, hay un nodo que es el origina la interacción y otro la recibe y por lo tanto, su matriz de adyacencia tampoco es simétrica. El más claro ejemplo de redes dirigidas son las redes tróficas, o sea, las representaciones gráficas de la cadena alimentaria de un ecosistema. En las redes no dirigidas, los links son simétricos, al igual que la matriz de adyacencia y ambos nodos juegan el mismo papel en la interacción. En nuestro caso, vamos a utilizar una red no dirigida, ya que todos los investigadores que firman un artículo son iguales desde el punto de vista de la interacción.

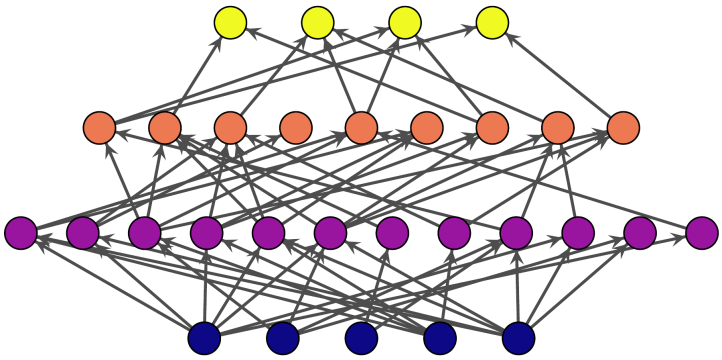

Figura 1: Grafo dirigido de una red trófica. Imagen obtenida de [15].

- **Red pesada y no pesada:** Una red compleja puede modelizar sólo la existencia de una interacción o unión entre dos individuos y en ese caso sería una red no pesada o puede modelizar una cuantificación de la interacción. Esta cuantificación, se representa con los

pesos de los links de la red. Algunos ejemplos de redes pesadas serían: la red de transacción o comercio entre entidades o naciones, en la que los pesos de los links recogerían una cuantificación monetaria de la transacción o por ejemplo, una red de transporte entre ciudades, en la que los pesos de los links podría ser la distancia entre ciudades.

En el caso de las redes pesadas, además de la matriz de adyacencia, existe la matriz de adyacencia pesada $\mathbf{W}$, cuyas entrada w_{ij} recogen el valor del peso del link del nodo i hacia el nodo j. En nuestro caso, vamos a utilizar una red pesada, ya que queremos representar cuánto de frecuente y/o fructífera es la interacción entre dos investigares. Más adelante explicamos las diferentes maneras de calcular el valor de los pesos de los links de nuestra red.

Un parámetro elemental de las redes complejas es el grado. El grado k_i del nodo i está definido como el número de links que tienen un extremo en ese nodo. En un red dirigida podemos hablar de grado de entrada y grado de salida, en las redes nos dirigidas hablamos simplemente de grado. En las redes pesadas, además del grado de un nodo existe el grado pesado s_i, que es la suma de pesos de los links que tienen un extremo en el nodo. Si lo expresamos en función de las matrices de adyacencia tendríamos:

$$k_i^{entrada} = \sum_j A_{ji} \tag{1}$$

$$k_i^{salida} = \sum_j A_{ij} \tag{2}$$

$$k_i = \sum_j A_{ij} = \sum_j A_{ji} \tag{3}$$

$$s_i = \sum_j W_{ij} \tag{4}$$

A continuación se muestra una imagen ilustrativa del concepto de grado de un nodo.

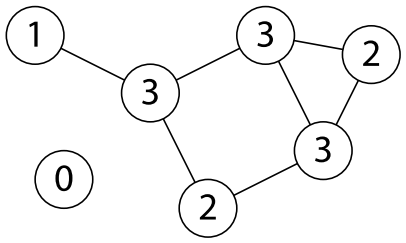

Figura 2: Red sencilla con el grado de cada nodo como etiqueta. Imagen obtenida de [16].

Un descriptor fundamental de las redes complejas es la distribución del grado de los nodos, es decir, el número de nodos con grado k, frente al valor de k. En la práctica, no se utiliza el

número de nodos con grado k, sino su porcentaje $P(k)$, que se corresponde con la probabilidad de que, seleccionando un nodo al azar tenga grado k.

$$P(k) = \frac{N_k}{N} \tag{5}$$

donde N_k es el número de nodos de grado k y N es el número total de nodos.

Dentro de los sistemas complejos reales, que son modelizados por las redes complejas, se pueden distinguir 2 grandes grupos, atendiendo al grado de interactividad de sus individuos. Hay sistemas complejos en los que existe homogeneidad entre sus individuos, en cuanto al número de interacciones de cada uno y existen otros sistemas en los que hay unos pocos individuos con un gran número de interacciones y la gran mayoría de individuos con pocas interacciones. Estos dos grupos de sistema complejos, cuando se representan en una red compleja, dan lugar a dos tipos de distribución de grados bien diferenciadas. En la siguiente imagen se muestran ambas distribuciones

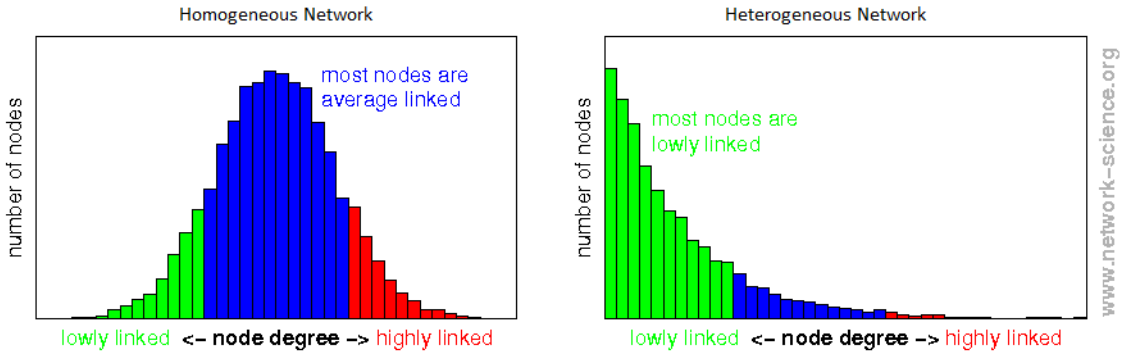

Figura 3: Distribución de Poisson y "Power-law" de grados para redes homogéneas y heterogéneas. Imagen obtenida de [17]

La distribución de las redes homogéneas sigue una distribución de Poisson y la de las redes heterogéneas es del tipo "power-law".

Al ser estos dos tipos de redes tan importantes, se desarrollaron modelos teóricos de generación o distribución aleatoria de los links de la red, que daban lugar a distribuciones de grado de estos tipos. Para las redes homogéneas o también llamadas redes Erdös-Rényi [18], los links de la red se generan aleatoriamente entre todos los posibles pares de nodos, que no estén ya unidos. Para una red de N nodos y E links, la probabilidad de que generar un nuevo link entre un par de nodos determinado es: $P = \frac{1}{N*(N-1)-E}$.

En el caso de las redes heterogéneas o redes libres de escala, R. Albert y A. Baràbasi [19] desarrollaron un modelo en el que la probabilidad de generar un nuevo link entre el nodos i y el nodo j venía dada por $P = \frac{k_i k_j}{(\sum_l k_l)^2}$, es decir, proporcional al producto del grado de cada nodo. Más adelante, se va a ver que estas distribuciones aleatorias de los links de la red tienen un gran

importancia en la definición de la modularidad.

También es muy importante en redes complejas el **coeficiente de Clustering** de un nodo i que denotamos como $C(i)$. El coeficiente de clustering mide cuánto de conectados están los vecinos del nodo i entre sí y se define como:

$$C(i) = \frac{1}{k_i(k_i - 1)} \sum_{j \neq k} A_{ij} A_{jk} A_{ik} \tag{6}$$

Tanto el grado, como el coeficiente de clustering son magnitudes locales de la red, ya que sólo influyen en su valor los vecinos del propio nodo.

Otro concepto fundamental en redes complejas, es el concepto de **camino** entre dos nodos, es decir, las distintas trayectorias que podemos trazar, recorriendo links y visitando otros nodos, para llegar de un nodo i a un nodo j. En redes no pesadas, se dice **longitud del camino** al número de links que se recorren para llegar de i a j, en cambio, en redes pesadas, la longitud del camino depende de los pesos de los links que se recorren. La longitud de un camino entre nodos es una magnitud global, es decir, no solo afectan en su cálculo los nodos cercanos a los de los extremos del camino.

De todos los caminos posibles entre i y j, el más importante es el **camino más corto** o de **menor longitud**. La longitud del camino más corto entre el nodo i y el nodo j la denotamos como $l(i, j)$. La siguiente figura ilustra a la perfección este concepto.

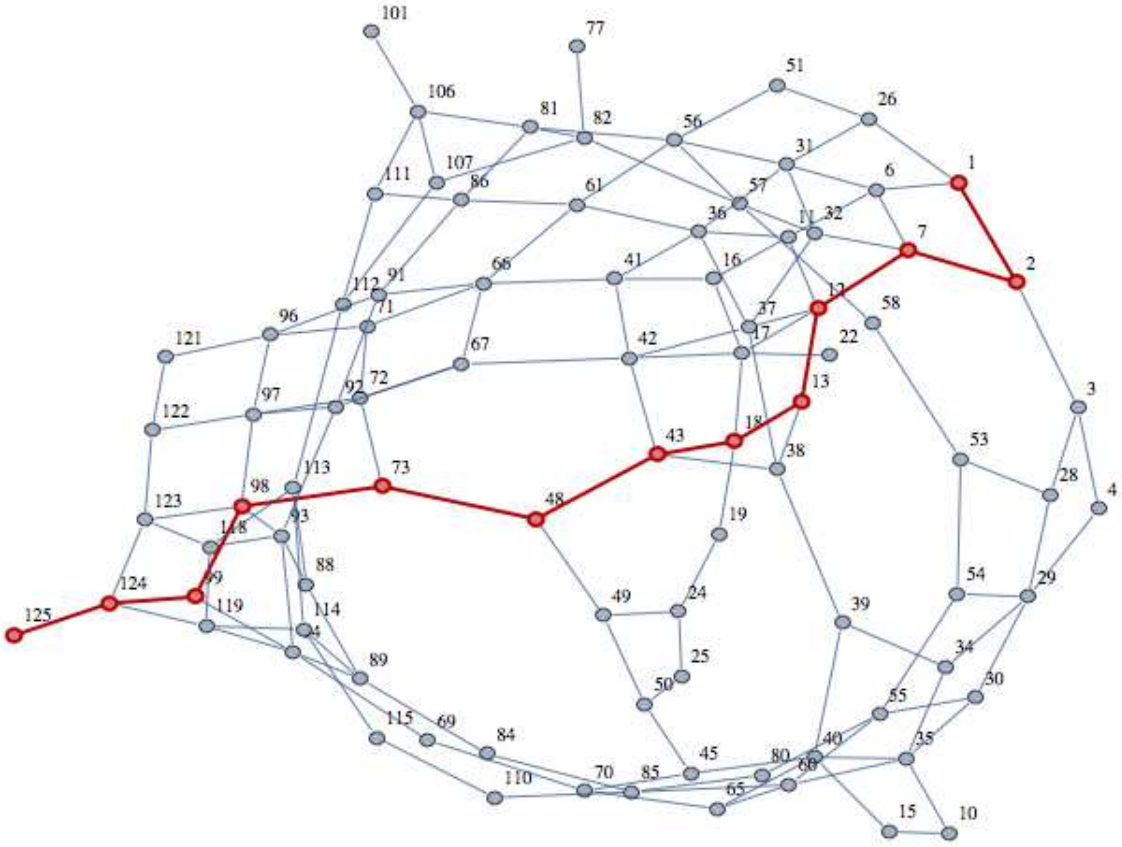

Figura 4: Representación de una red compleja no dirigida. Está marcado el camino más corto entre el nodo 1 y el 125. Imagen obtenida de [20]

En general, el problema de obtener el camino más corto entre dos nodos de una red es mucho más complejo en el caso de redes dirigidas, que en redes no dirigidas. En nuestro caso no es un problema, ya que como hemos dicho antes nuestra red es no dirigida.

Otro concepto más complejo de teoría de redes es el concepto de **centralidad**. La centralidad es una magnitud que nace de la necesidad de cuantificar la importancia de un nodo dentro la red, a nivel global. Hay distintos tipos de magnitudes para medir la centralidad y la mayoría de ellas se basan en el camino más corto entre nodos y su longitud. Uno de estos tipos es la **centralidad de "cercanía"** (closeness centrality) que cuantifica lo fácilmente alcanzable que es un nodo desde todos los demás nodos de la red, definimos la closeness centrality g_i, del nodo i como:

$$g_i = \sum_j \frac{1}{l(i,j)} \tag{7}$$

Otro tipo de centralidad es la **centralidad de "interinidad"** (betweenness centrality). La betweenness centrality puede ser de un nodo o de un link y está relacionada con el número de veces que hay que pasar por el nodo/link cuando recorremos los caminos más cortos entre todos

los pares de nodos de la red. El algoritmo de Girvan-Newman para la obtención de particiones de comunidades se basa en esta magnitud y por lo tanto está explicada en detalle más adelante, en el apartado 3.4.

Con los conceptos presentados hasta ahora, es suficiente para comprender los métodos y magnitudes propias de la detección de comunidades utilizados en este trabajo. Como se ha mencionado anteriormente, se ha utilizado el **algoritmo de Girvan-Newman** como método para la detección de comunidades y la **modularidad** como medida de bondad de las particiones en comunidades obtenidas. Ambos conceptos, tienen apartados propios, más adelante, en los que se explican en detalle.

3.2. Modelización de la red

En este apartado se explican algunas de las distintas formas de utilizar los datos de los que disponíamos, para generar la red compleja.

Por cada artículo firmado por N investigadores se generan $\frac{N*(N-1)}{2}$ links entre los investigadores y de cada investigador salen $(N-1)$ de estos links.

La modelización más simple es una red no pesada. Todos los links de la red tienen peso unidad, sin importar el número de artículos publicados o el impacto de estos. Esto implica no representar lo estrecha (número de artículos publicados conjuntamente), ni lo fructífera (impacto JCR) que es la colaboración entre investigadores, únicamente si ha existido esta colaboración o no.

Si queremos modelizar la red pesada, entonces hay que plantearse que valor se da a cada link. A la hora de obtener este valor, hay que tener en cuenta 2 cuestiones: Si se iguala el impacto de todos los artículos y como repartir el impacto de un artículo entre los links que genera.

Si consideramos que todos los artículos tienen impacto unidad, estaríamos representando lo frecuente que es la colaboración entre investigadores (número de artículos publicados conjuntamente), pero no cuánto de buenos son los artículos generados (impacto JCR). La otra forma es utilizar el impacto JCR del artículo, de esta forma se tiene en cuenta tanto el grado de colaboración, como la calidad de esta.

En cuanto a las posibilidades de repartir el impacto del artículo, una de ellas es repartirlo entre los $(N-1)$ links que se generan para cada investigador, es decir, el peso de cada link sería

$$Peso\ link = \frac{Impacto\ artículo}{(N-1)} \tag{8}$$

donde N es el número de investigadores que firman el artículo. Este método implica que

el mérito del artículo no depende del número de investigadores que lo firmen y por lo tanto, si calculamos la contribución de este artículo al grado pesado (suma de los pesos de los links que salen de cada nodo) de los investigadores que lo firman, obtenemos el impacto íntegro del artículo. La otra forma es repartir el impacto entre todos los links que genera el artículo, de esta manera, el peso de cada uno de los $\frac{N*(N-1)}{2}$ links sería

$$Peso\ link = \frac{Impacto\ artículo}{(N-1)*N} \tag{9}$$

Este otro método asume que el mérito del artículo se debe repartir entre todos los investigadores que lo firman y si calculamos el grado pesado de los investigadores obtenemos $\frac{Impacto\ artículo}{N}$, es decir, el impacto del artículo divido por el número de investigadores.

En nuestro caso, hemos elegido modelizar la red pesada, los artículos con impacto JCR y el impacto repartido solo entre los links que salen de cada investigador, de tal manera que la contribución de un artículo al grado pesado de un investigador es el impacto íntegro de ese artículo.

Otro aspecto importante a la hora de modelizar y trabajar con la red es el concepto de Cluster o Componente Gigante. En redes complejas se dice que dos nodos están conectados si existe al menos un camino para llegar de un nodo a otro. Todos los nodos que están conectados entre sí forman una componente o cluster y al cluster que está compuesto por el número de nodos mayor, se le llama Cluster Gigante. Si todos los nodos de la red están conectados entre sí y por lo tanto, la red coincide con el Cluster Gigante, entonces se dice que la red es completamente conectada. En ocasiones es producente trabajar solo con el Cluster Gigante, si las demás componentes son despreciables con respecto a él, sobretodo si se va a trabajar con particiones de comunidades, como es nuestro caso.

Nuestra red **no** es completamente conectada. El Cluster Gigante está formado por 971 de los 1007 nodos que componen la red, los 36 nodos restantes están repartidos en 12 componentes, de las cuales, la mayor, tiene 5 nodos. Estos números indican que las componentes aisladas son extremadamente pequeñas en comparación con el cluster gigante y por lo tanto, es aceptable trabajar solo con él.

Una vez explicado como se ha modelizado la red, se muestra una representación gráfica de la misma en la figura 18, del apéndice B. En ella podemos observar los nodos y links de la red. El tamaño de los nodos es proporcional a su grado pesado y el coloreado de los mismos se corresponde con el departamento al que pertenece el investigador que representan.

3.3. Partición en comunidades y Modularidad

Pasamos ahora a presentar los conceptos de modularidad y comunidades en redes complejas, que son esenciales para la metodología y el análisis de este trabajo.

La modularidad es una magnitud introducida por Girvan y Newman para cuantificar la bondad de las particiones en comunidades que se obtenían con su algoritmo. Quisieron basar la bondad de esta medida en el porcentaje de links entre nodos de una misma comunidad, con respecto al total de nodos de la red. Esto viene de la idea de que en una buena partición en comunidades se cumple que los nodos dentro de las comunidades interactuan más entre ellos, que con nodos de otras comunidades.

En concreto, la modularidad de una partición de la red está definida como la diferencia entre el porcentaje de links de la red dentro de las comunidades de la partición y el porcentaje de links que habría dentro de las comunidades, para la misma partición, si los links se hubiesen generado aleatoriamente.

Esta idea de tomar el valor de modularidad 0, cuando ambos porcentajes son iguales, en la red cuyos links están generados aleatoriamente, se debe a que en cualquier partición que podamos hacer en un red con links generados aleatoriamente, sin tener en cuenta propiedades globales de la red, no vamos a observar predominio de links entre comunidades o links dentro de comunidades.

Esto es fácil de ver, por ejemplo, en los modelos aleatorios de generación de links explicados en el apartado 3.1. En ambos modelos o se utilizaba una distribución plana para generar los links o solo se tenía en cuanta el grado de los nodos. Por lo tanto, no se genera ninguna estructura especial de comunidades. En el caso de las redes de libre escala, sí que se generaran nodos que individualmente tendrán mayor concentración de links, pero no se generaran grupos de nodos que interaccionen más entre ellos que con otros nodos de la red.

Con esta definición, la fórmula general de la modularidad para redes no pesadas viene dada por

$$Q = \sum_{ij} (\frac{A_{ij}}{2E} - P_{ij}) \delta(C_i, C_j) \tag{10}$$

donde P_{ij} es la probabilidad que exista un link entre el nodo i y el nodo j en una distribución aleatoria de links, la delta de Kronecker vale 1 si el nodo i y el nodo j pertenecen a la misma comunidad y 0 en caso contrario y por último, el sumatorio es sobre todos los pares de nodos ij de la red (incluido el caso $i = j$). Al hablar de probabilidad en vez de porcentaje de links, se entiende que hay una relación directa entre ambas.

Para poder calcular la modularidad, hay que decidir cual es la distribución aleatoria de referencia, para poder darle una expresión a P_{ij}. Se recomienda usar como distribución aleatoria la de las redes de libre escala, es decir, la propuesta por R. Albert y A. Barabási, ya que

esta distribución se suele ajustar más a la estructura de las redes reales. Si recordamos, la probabilidad P_{ij} de existencia de un link entre el nodo i y el j, para esta distribución, venía dada por $P_{ij} = \frac{k_i k_j}{(\sum_l k_l)^2}$. El sumatorio del grado de todos los nodos de la red $\sum_l k_l$, no es más que el número total de links multiplicado por 2, es decir, $2E$. Por lo tanto, la definición matemática de la modularidad queda

$$Q = \frac{\sum_{ij}(A_{ij} - \frac{k_i k_j}{2E})\delta(C_i, C_j)}{2E} \tag{11}$$

Cambiando la delta de Kronocker por un sumatorio en comunidades

$$Q = \sum_{C}^{N_{comunidades}} (\frac{\sum_{i\in C, j\in C} A_{ij}}{2E} - \frac{\sum_{i\in C} k_i}{2E} * \frac{\sum_{i\in C} k_i}{2E}) \tag{12}$$

donde el sumatorio de A_{ij} partido por $2E$, es el porcentaje de links dentro de la comunidad C y el sumatorio de k_i partido por $2E$, es el porcentaje de links que tienen al menos un extremo en los nodos de la comunidad C, contando doble los links que tienen ambos extremos en nodos de esa comunidad, es decir, los links dentro de la comunidad.

Ambos términos del sumatorio, al ser porcentajes, varían entre $[0, 1]$ y por lo tanto, el valor máximo y mínimo de la Modularidad es 1 y -1 respectivamente. Valores positivos indican particiones en las que las interacciones dentro de comunidades son mayores que entre comunidades y valores negativos lo contrario. Hay 3 casos interesantes en los que se observa bien este comportamiento:

1. En el caso de que haya una comunidad cuyo porcentaje de links internos sea mucho mayor que el de otras comunidades (el caso límite es que solo haya una comunidad), la modularidad tiende a 0, ya que el porcentaje de links y el porcentaje de extremos de links en esa comunidad tienden ambos a 1, el de las otras comunidades tiende a 0, ambos términos del sumatorio son iguales y la modularidad se anula.

2. Otra situación interesante es que se elijan N comunidades, que estén completamente desconectadas entre ellas, si existe la posibilidad. En este caso, la modularidad es positiva, ya que para cada comunidad el porcentaje de links internos y el porcentaje de extremos de links en nodos de la comunidad es el mismo. En el caso concreto de que el porcentaje de nodos dentro de cada una de estas comunidades sea igual, obtenemos este valor de la modularidad (13) desarrollando la fórmula (11). Vemos que en el caso de $N = 1$ la modularidad es 0 y que conforme aumenta el número de comunidades la modularidad se aproxima a 1.

3. El último caso interesante es cuando la elección de las N comunidades es de tal manera que no hay links dentro de las comunidades, sólo entre comunidades. La modularidad es estrictamente negativa, ya que al no haber link dentro de las comunidades, el término

positivo del sumatorio de la modularidad es 0. En el caso concreto de que el porcentaje de extremos de links en cada comunidad sea el mismo, la expresión de la modularidad queda (14), que tiende a 0 conforme N crece y tiene su mínimo valor para $N = 2$ donde $Q = -\frac{1}{2}$. El caso de $N = 1$ no se contempla ya que una sola comunidad sin que hubiese links dentro de ella implicaría que en la red no hay ningún link, lo cual no tiene sentido.

$$Q = \frac{(N-1)}{N} \tag{13}$$

$$Q = -\frac{1}{N} \tag{14}$$

Todos estos resultados obtenidos para redes no pesadas se pueden extender a redes pesadas simplemente cambiando los links por su peso. Para una red pesada, la modularidad, tomando de nuevo la red de libre escala, como distribución de referencia, vendría dada por

$$Q = \frac{\sum_{ij}(W_{ij} - \frac{s_i s_j}{2W})\delta(C_i, C_j)}{2W} \tag{15}$$

donde la delta de Kronecker tiene el mismo significado que en la fórmula no pesada.

Los resultados obtenidos con la red no pesada, son los mismos para la red pesada, lo único que el porcentaje de links internos de una comunidad pasa a ser el porcentaje del peso total de los links dentro de cada comunidad y el porcentaje de extremos de links en una comunidad pasaría a ser el porcentaje del peso total de los links que tienen un extremo en esa comunidad.

Se ha preferido explicar la modularidad para la red no pesada y luego extenderla a la red pesada, porque se entiende que el caso no pesado es más visual e intuitivo.

Una vez que conocemos la magnitud de modularidad y lo que mide, nuestro objetivo es encontrar particiones de la red de colaboración, cuyos valores de la modularidad sean los más altos posibles, para después buscar equivalencias entre las comunidades de estas particiones y los departamentos.

3.4. Betweenness centrality y el algoritmo de Girvan-Newman

Toda la parte teórica de obtención de particiones de comunidades de este trabajo está basada en el report de Santo Fortunato, *Community detection in graphs* [21].

El problema de detección de comunidades en redes u obtención de particiones de comunidades no está bien definido y tiene una resolución compleja, aunque a primera vista pueda resultar intuitiva la agrupación de nodos que tienen propiedades o características estructurales similares. La dificultad de este problema radica, principalmente, en que conceptos como partición o

comunidad no están definidos rigurosamente; en la ausencia de magnitudes que nos indican la bondad de una partición, hasta que se formuló la modularidad y en encontrar magnitudes que nos ayuden a discernir unas comunidades de otras.

Es importante apuntar también, que la detección de comunidades es posible sólo en redes no densas, es decir, que el número de nodos y links es del mismo orden. En nuestro caso, esto se cumple, ya que tenemos una red de 971 nodos y 4973 links.

Desde el punto de vista computacional, el problema de detección de comunidades es un problema NP-duro. En concreto, el problema de encontrar la partición en comunidades de máxima modularidad es un problema NP-completo. Por lo tanto, es común el uso de *algoritmos aproximativos* para la resolución del problema. Estos algoritmos no nos aseguran que la solución obtenida sea óptima, si no que será una solución aproximada. A cambio, posibilitan que este problema se pueda "atacar" computacionalmente.

Uno de estos algoritmos aproximados, fue propuesto por Girvan y Newman. Su objetivo es encontrar particiones con altos valores de modularidad, mediante la eliminación de links según su betweenness centrality.

La betweenness centrality se basa en el número de caminos más corto entre pares de nodos y se calcula de la siguiente manera

$$B_i = \sum_{jk} \frac{\sigma_i^{jk}}{\sigma^{jk}} \tag{16}$$

donde B_i es la betweenness bentrality del nodo o link i, σ^{jk} es el número de caminos más cortos entre el nodo j y el nodo k, σ_i^{jk} es el número de caminos más cortos entre el nodo j y el nodo k que pasan por el nodo o link i y el sumatorio es sobre todos los pares de nodos j y k de la red, siempre y cuando exista al menos un camino entre ellos y se cumpla que $j \neq k \neq i$. Hablamos de caminos más cortos en plural, ya que para redes con un número alto de links y nodos, es probable que haya varios caminos que tengan la misma longitud y que esta sea la menor.

El algoritmo de Girvan-Newman es un algoritmo iterativo en el que en cada iteración se calcula la betweenness centrality de todos los links de la red y se elimina el link en el que sea mayor, en caso de empate se decide aleatoriamente el link a eliminar. El algoritmo termina cuando no quedan links en la red. Durante el proceso, al ir eliminando links, se van formando componentes de nodos inconexas unas de otras, estas comunidades son las que hay que identificar como comunidades de la partición. Cada vez que se divide una componente en dos, se obtiene una nueva partición con una comunidad más. Es importante notar que el algoritmo sólo proporciona una partición para un número de comunidades fijo y por lo tanto podemos identificar a cada partición con su número de comunidades.

La forma de generar particiones con valores altos de modularidad es encontrar grupos de nodos que estén unidos por pocos links o en el caso de red pesada, por links que tengan poco peso y elegir estos grupos de nodos como las comunidades. Precisamente esos links, son los que tienen mayor betweenness centrality y al eliminarlos durante el proceso del algoritmo, las componentes que quedan son las comunidades que buscamos. Para obtener valores altos de modularidad, el número de links que se eliminen entre cada división de componentes y el peso de estos links deben ser pequeños. Esta idea se puede ver graficamente en la siguiente imagen.

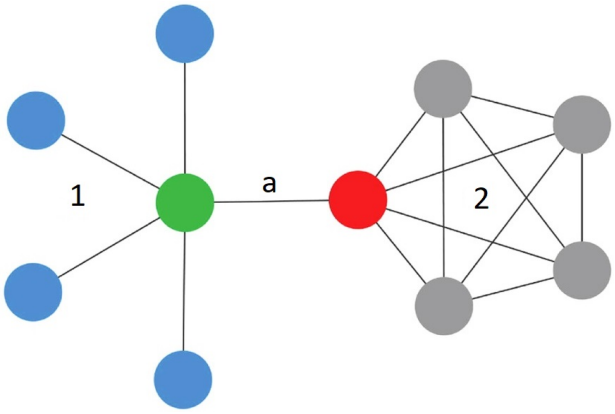

Figura 5: Representación de una red en la que la partición de maxima modularidad sería tomar como comunidades el grupo de nodos 1 y 2. El link a es el link de máxima modularidad, que sería eliminado el primero por el algoritmo de G-N y en la primera iteración ya obtendriamos la partición óptima. Imagen obtenida de [22]

La primera duda que surge al implementar el algoritmo es cómo determinar los caminos más cortos en un red pesada, es decir, si utilizamos los pesos de los links o no y en el caso de que sí, de qué manera. En cuanto a la forma de utilizar los pesos hay dos opciones, principalmente, considerar que la longitud cada link es el valor de su peso o considerar que la longitud del link es una cantidad inversamente proporcional a su peso. En el apartado 3.6 de la metodología se aclara que procedimiento se llevo a cabo finalmente y por qué.

3.5. Similaridad de particiones de conjuntos

Como hemos visto hasta ahora, el procedimiento de este trabajo consiste en obtener particiones de la red que maximicen la modularidad y compararlas o sacar equivalencias con la partición departamental, que es la partición real de la red de investigadores. Este procedimiento exige pues, la utilización de magnitudes de similaridad.

Como ya hemos comentado anteriormente, no estamos comparando redes o grafos distintos, si no distintas particiones del mismo grafo. Por lo tanto, necesitamos conceptos propios de la similaridad de conjuntos o particiones de conjuntos, más que de similaridad de grafos o redes.

Vamos a presentar las siguiente magnitudes de similaridad de conjuntos: el **Índice de Wallace**, el **Índice de Rand** [23] y el **Índice de Jaccard** [24], centrándonos principalmente en el primero de ellos.

Nos imaginamos que tenemos un conjunto de N objetos (nodos en este caso) y dos agrupaciones o particiones distintas de estos. A partir de ahora vamos a llamar a los distintos grupos de cada partición comunidades, la primera partición tiene C_1 comunidades y la segunda C_2. Para la formulación de ambos índices, es imprescindible que los nodos sean identificables, no así las comunidades y es necesario saber si cada par de nodos pertenece a la misma comunidad o no, en ambas particiones. Además, ni el número, ni el tamaño de las comunidades en ambas particiones tiene porque ser el mismo.

En esta situación, imaginamos que todos los nodos que pertenecen a la misma comunidad, en cada partición, están unidos. Esto implica que tenemos $\frac{N*(N-1)}{2}$ pares de nodos totales en ambas particiones, de los cuales: $\frac{P}{2}$ están unidos en la primera partición, $\frac{Q}{2}$ están unidos en la segunda partición, $\frac{T}{2}$ están unidos tanto en la primera como en la segunda partición, $\frac{V_1}{2}$ están unidos en la primera pero no en la segunda, $\frac{V_2}{2}$ están unidos en la segunda pero no en la primera y $\frac{U}{2}$ no están unidos ni en la primera ni en la segunda. El hecho de dividir todos los parámetros entre 2 es debido a que el número total de pares de nodos está divido por 2 y de esta manera se simplificaran en el cálculo de los índices. La siguiente matriz es una descripción de esta situación.

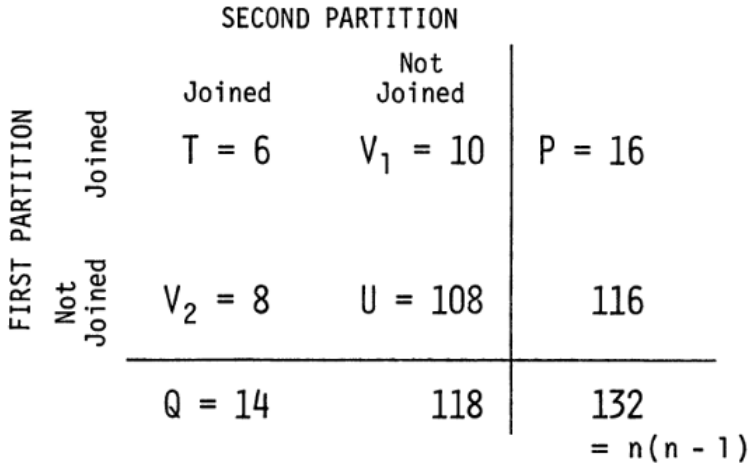

Figura 6: Matriz en la que se observa la distribución de los distintos parámetros de uniones de nodos de ambas particiones. En la matriz todos los parámetros aparecen ya sin el 2 de denominador. Imagen obtenida de [6]

Antes de nada, aclarar que las uniones de pares nodos de las que estamos hablando aquí, no tienen nada que ver con los links de la red compleja. Dicho esto, los índices de Wallace, Rand y

Jaccard se definen, en función de estos parámetros, de la siguiente manera

$$W_{índice} = \frac{T}{\sqrt{P * Q}} \tag{17}$$

$$R_{índice} = \frac{T + U}{N * (N - 1)} \tag{18}$$

$$J_{índice} = \frac{P \cap Q}{P \cup Q} = \frac{T}{P + Q - T} \tag{19}$$

El Índice de Wallace, es la media geométrica del porcentaje de pares de nodos que estando unidos en la primera partición, se han mantenido unidos en la segunda; con el porcentaje de nodos que estando unidos en la segunda partición, se han mantenido unidos en la primera.

El Índice de Rand tiene una interpretación más sencilla. Si entendemos que ambas particiones serán similares, cuando los mismos pares de nodos que están unidos en una partición lo están también en la otra y los que no están unidos en una tampoco lo están en la otra. Entonces, el Índice de Rand no es más que el número de pares de nodos unidos en las dos particiones más los que no están unidos en ninguna, partido por el número total de pares de nodos. Sería el porcentaje de pares de nodos "similares" en ambas particiones.

Por último, el índice de Jaccard es la intersección de pares de nodos unidos en cada partición, dividido por la unión de estos.

Notar que al ser estos índices porcentajes o medias de porcentajes, sus posibles valores se mueven en el rango $[0, 1]$, para particiones idénticas, los 3 índices vale 1.

El índice de Rand, en general, toma valores más grandes que los otros dos, debido al término U del numerador. Por ejemplo, para el caso en el que $T = 0$, es decir, no hay ningún par de nodos que comparta comunidad en las dos particiones, el índice de Wallace y el de Jaccard valdrían 0, mientras que el de Rand valdría: $R_{índice} = \frac{U}{N*(N-1)}$. Esta presencia de U en el numerador, hace también que el índice de Rand se dispare rápidamente cuando el número de comunidades de ambos conjuntos aumenta.

En cuanto al índice de Jaccard, tiene un comportamiento similar al de Wallace. Sin embargo, con el índice de Jaccard, en el caso que haya una partición con un número de comunidades mucho mayor que la otra, la partición con menor número de comunidades tiene mucho más peso.

Estas comparaciones y los comportamientos y resultados de los 3 índices encontrados en la literatura, nos han llevado a utilizar como principal descriptor de la similaridad el **índice de Wallace**.

Otro detalle a tener en cuenta es que, al estar trabajando con pares de nodos, los parámetros son proporcionales al número de nodos al cuadrado. Esto implica que lo que ocurre con los pares de nodos en las comunidades grandes, tenga mucho más peso que lo que ocurre en las pequeñas.

3.6. Metodología para la obtención de resultados

Una vez presentadas y explicadas todas las magnitudes y procedimientos necesarios para el estudio y comparación de la red y su estructura de comunidades, pasamos a detallar la metodología utilizada.

El procedimiento es el siguiente, utilizar el algoritmo de Girvan-Newman para generar particiones en comunidades de la red e ir analizando cómo varían la modularidad y los índices de similaridad de estas particiones. Las particiones están determinadas por el número de comunidades, como hemos dicho antes, ya que el algoritmo solo proporciona una partición para un número de comunidades fijo. Por lo tanto, lo que vamos a estudiar es como varían las magnitudes mencionadas, conforme aumenta el número de comunidades en la partición. Notar que no vamos a utilizar los índices de similaridad para comparar particiones de comunidades entre ellas, sino cada una de las particiones de comunidades con la partición departamental.

En nuestro caso, no se ha esperado a que el algoritmo llegue al final del proceso, que sería acabar en una partición en la que cada nodo es una comunidad. Se ha limitado el algoritmo a 60 comunidades, ya que cuanto mayor es el número de comunidades de la partición, más difícil es buscar equivalencias con la partición departamental.

Una vez obtenidas las particiones, identificamos cual es la partición con mayor modularidad y con mayor valor de los índices de similaridad. Lo esperado es que ambos máximos correspondan a particiones con un número de comunidades similar y entonces buscar en ese intervalo la partición óptima o particiones interesantes. Es decir, particiones cuyos valores de la modularidad sean altos pero que no impliquen muchos cambios con respecto a la partición departamental.

Otro resultado deseable es que la modularidad máxima que encontremos supere la modularidad de la partición departamental. Ya hemos comentado que el algoritmo de Girvan-Newman no asegura encontrar la partición de máxima modularidad, por lo tanto podría ser que la partición departamental tuviese mayor modularidad. Es más, esperamos que esta partición tenga un valor alto de la modularidad, ya que es lógico que la colaboración de los investigadores dentro de los departamentos sea mayor que entre departamentos.

Aún tenemos el problema de elegir qué método usar para calcular los caminos más cortos en la betweenness centrality: red no pesada, usar el valor del peso de los links o el inverso del peso. Como no tenemos certeza de que método es óptimo a la hora de maximizar la modularidad, lo

que hemos decidido ha sido aplicar el procedimiento de obtención de comunidades con los tres métodos y estudiar con cual de ellos se obtienen mejores resultados.

4. Resultados

Pasamos a presentar los resultados obtenidos aplicando la metodología explicada. Este apartado se estructura en 3 partes: la **elección del método óptimo** para calcular la betweenness centrality en el algoritmo de Girvan-Newman (links pesados, links no pesados o inverso del peso de los links), **presentación de las particiones más interesantes** y **comparación con la partición departamental**.

4.1. Método para el cálculo de la betweenness centrality

Como se ha comentado en el apartado del algoritmos de Girvan-Newman, hay 3 posibilidades para calcular los caminos más cortos entre nodos. Debido a que no tenemos argumentos de peso para decantarnos por ninguno, se decide aplicar el algoritmo usando los tres métodos y ver con cual obtenemos particiones de mayor modularidad. A continuación se presenta la gráfica con la modularidad de las particiones obtenidas con cada método.

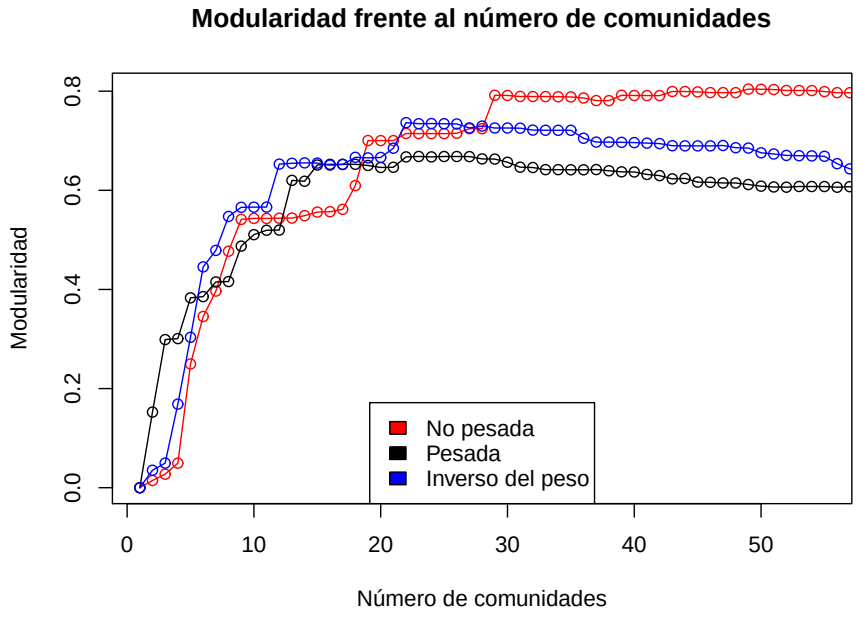

Figura 7: Gráfica con la modularidad de las particiones obtenidas con los 3 métodos.

En el eje x de la gráfica tenemos el número de comunidades de la partición. Hay que recordar que en el algoritmo de Girvan-Newman, cada partición está determinada por el número de comunidades. En el apéndice A, se incluye la tabla 7 con los valores de la gráfica.

Podemos observar 3 zonas con comportamientos diferentes. Entre 1 y 5 comunidades, el método que proporciona particiones con mayor modularidad es en el que tomamos directamente el peso de los nodos; entre 6 y 18, es el que toma el inverso del peso y de 19 comunidades en

adelante, el que mejor funciona es no tener en cuenta los pesos de links.

Como lo que nos interesa es obtener la partición de mayor modularidad, nos decidimos a usar el método en el que no se tiene en cuenta el peso de los links. Sin embargo, el hecho de que no haya un método que predomine en todo el espacio de particiones obtenido, apoya nuestra idea de que no esta clara la bondad de un método frente a los otros.

4.2. Particiones de interés

Pasamos ahora a presentar la modularidad y el valor de los índices de similaridad de las distintas particiones, obtenidas con el algoritmo de Girvan-Newman y el método de cálculo de caminos más cortos seleccionado (links sin peso).

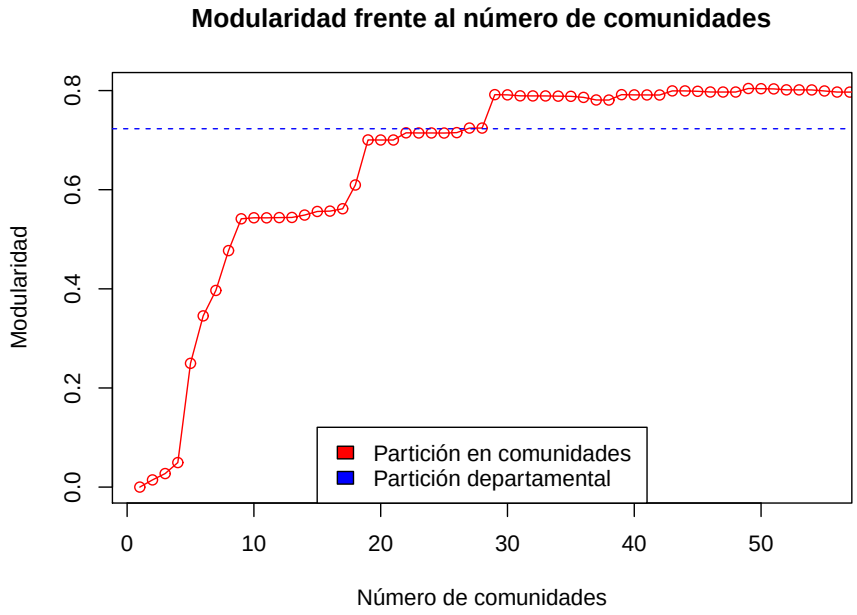

Figura 8: Modularidad de la partición departamental y de las distintas particiones en comunidades.

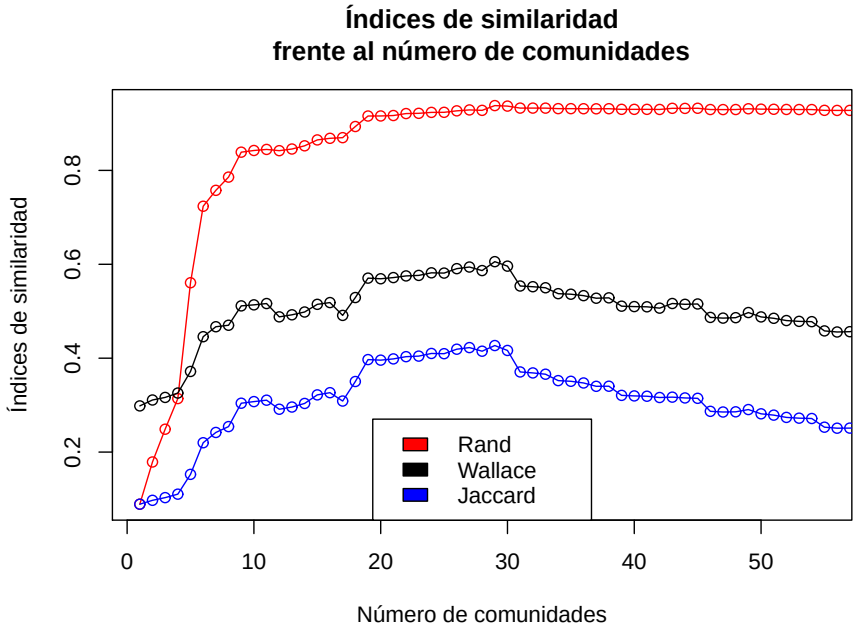

Figura 9: Valores de los índices de similaridad para las distintas particiones de comunidades obtenidas.

Podemos ver los valores correspondientes a estas gráficas en las tablas 7 y 8, en el apéndice A.

Empezamos analizando la gráfica de la modularidad (gráfica 8), que es el descriptor principal de este análisis. Antes de nada, decir que la Modularidad de la partición departamental tiene un valor de **0.723** (linea horizontal de la gráfica) y es nuestro valor de referencia. La partición de máxima Modularidad, que es la que estábamos buscando, la encontramos en **49** comunidades, con un valor de 0,804. Con esto se cumple uno de nuestros objetivos, que era encontrar una partición en comunidades con una modularidad mayor a la de la partición en departamentos. El hecho de que sea una partición con 49 comunidades, hace imposible que podamos encontrar un relación con la partición departamental (de solo 14 departamentos), que era otro de nuestros objetivo. Sin embargo, si nos fijamos en la gráfica, vemos que a partir de **29** comunidades, la modularidad se mantiene casi constante. Por lo tanto, esta partición de 29 comunidades es muy interesante porque su modularidad, con un valor de 0,792, es muy cercana a la modularidad máxima, con 20 comunidades menos.

Otra partición de interés es la de **9** comunidades. Si nos fijamos en la gráfica 8 de variación de la modularidad, vemos que al principio aumenta de forma considerable con cada división, hasta llegar a las 9 comunidades. A partir de ese punto, hay dos intervalos (de 9 a 17 y de 19 a 28 comunidades) en los que apenas hay crecimiento, hasta llegar a 29 que ya se mantiene aproximadamente constante. Por lo tanto, la partición de 9 comunidades es interesante si buscamos un compromiso entre un número pequeño de comunidades y un valor alto de la Modularidad (0,5410), aunque está lejos del máximo obtenido (0,804) y del de la partición departamental

(0,723).

En cuanto a los índices de similaridad, gráfica 9, podemos observar los comportamientos mencionados en el apartado 3.5, a grandes rasgos: el índice de Rand se dispara rápidamente en cuanto el número de comunidades empieza a crecer, debido al aumento del número de pares de nodos no unidos (U) y el índice de Jaccard con una evolución muy similar al índice de Wallace, pero con valores más pequeños. Los 3 índices tienen el máximo en la misma partición, precisamente la de **29** comunidades. Por lo tanto, esta es la partición más similar, según la definición de los índices, a la partición departamental, de todas las obtenidas con el algoritmo de Girvan-Newman.

Mostramos, por último, una tabla con los valores de los descriptores (Modularidad e índices de similaridad), de ambas particiones.

Partición	**Modularidad**	**Wallace**	**Rand**	**Jaccard**
9	0.5413	0.5114	0.8388	0.3041
29	0.7918	0.6056	0.9380	0.4269

Cuadro 2: Valores de los descriptores para las particiones de 9 y 29 comunidades

4.3. Comparación con la partición departamental

Vamos a hacer un análisis más profundo de cada una de las 2 particiones que hemos considerado las de mayor interés e intentar identificar las comunidades que las forman con los distintos departamentos. Para ello se presentan en este apartado, para cada partición, un gráfico de barras en el que se muestra que parte de cada comunidad pertenece a los distintos departamentos y una serie de tablas en el apéndice A, que contienen esta misma información de forma numérica.

Antes de empezar con el análisis de cada partición, vamos a presentar un dendrograma de como se han ido dividiendo las comunidades cuando pasamos de 9 a 19 comunidades y de 19 a 29. Esta figura ayuda a hacernos una idea de la evolución de las distintas comunidades durante el proceso. Incluimos la partición de 19 comunidades para tener una visión más progresiva del proceso de división.

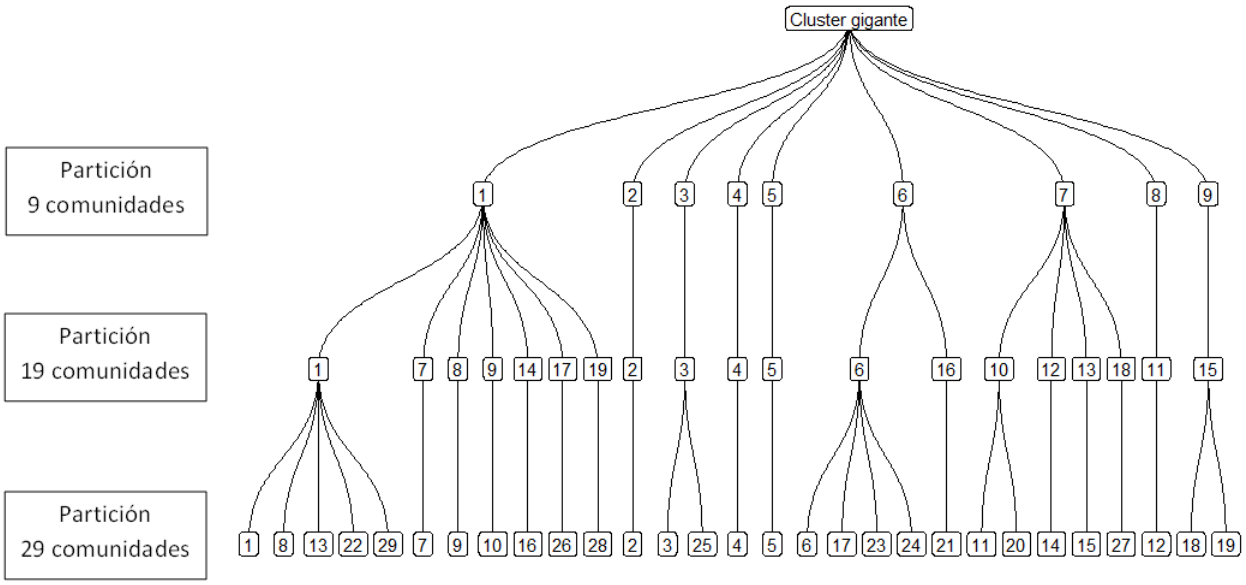

Figura 10: Dendrograma con la evolución de las distintas comunidades entre las particiones de 9, 19 y 29 comunidades.

4.3.1. Partición de 9 comunidades

Pasamos ahora a analizar la partición de 9 comunidades. Como hemos comentado antes, esta partición tiene un valor de la modularidad relativamente alto, para las pocas comunidades que la forman. Esto es porque todas las divisiones de las que proceden las comunidades que la forman, han supuesto un aumento considerable de la modularidad. Por todo esto, el análisis de estas 9 comunidades y su identificación con los departamentos, nos sirve como una primera aproximación a la estructura colaborativa de estos.

Para la identificación de los comunidades con los departamentos, hemos utilizado el siguiente gráfico de barras.

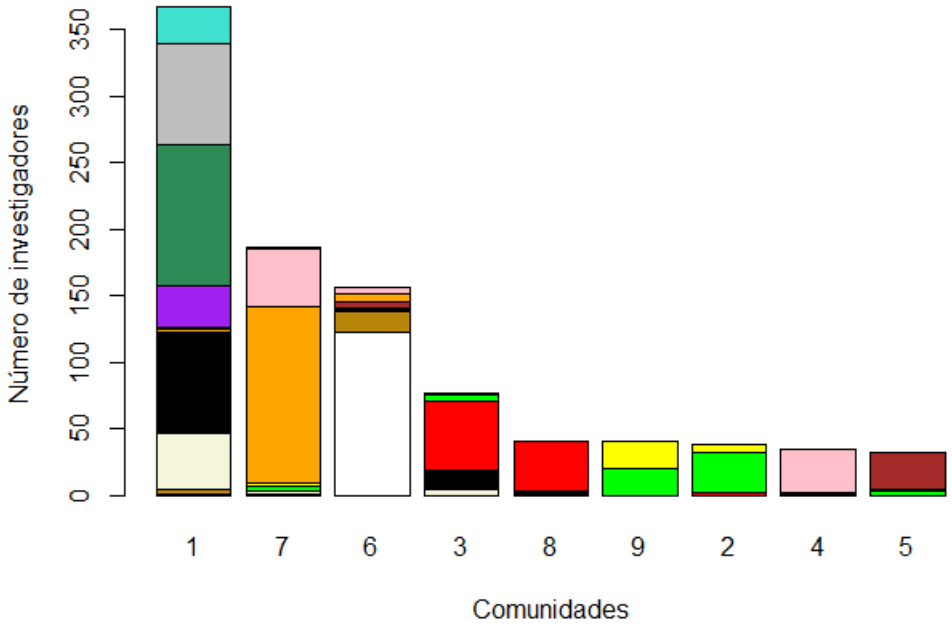

Figura 11: Gráfico en el que se muestra la composición por departamentos (colores) de cada comunidad (barras).

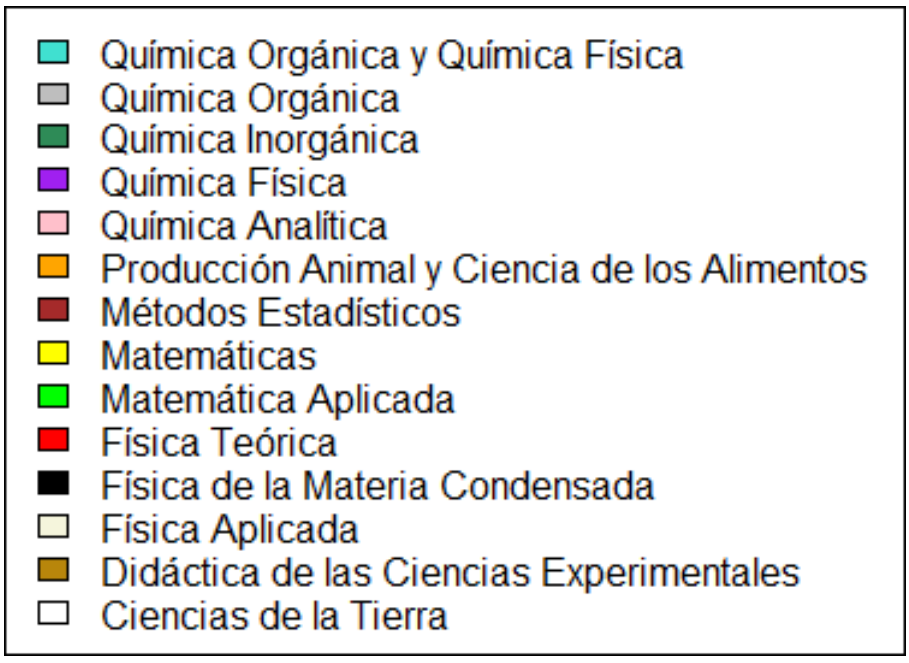

Figura 12: Leyenda de colores de cada departamento.

Cada barra corresponde a una comunidad, que están ordenadas por el número de investigadores que pertenecen a la comunidad y los departamentos están representados por colores. Podemos ver esta misma información en forma de tabla 9 en el apéndice A, donde las columnas son las comunidades, las filas los departamentos y las entradas son el número de investigadores que pertenecen a cada comunidad y departamento. En la tabla 10 tenemos la misma información pero con el porcentaje de investigadores de cada comunidad asociados a los departamentos.

Fijándonos en la tabla y la gráfica, hacemos el siguiente análisis de comunidades:

- **Comunidad 1:** Es lo que queda del Cluster gigante y por lo tanto, de la que se han ido separando las demás comunidades. Debido a esto es la comunidad más grande (367 investigadores), con diferencia sobre las otras y en la que encontramos más mezcla de departamentos. Vemos que la forman parte de 4 de los 5 departamentos de química: Química Orgánica y Química Física (7.6 %), Química Orgánica (20.44 %), Química Inorgánica (28.88 %) y Química Física (8.72 %). También está formada, en un 20.16 %, por parte del departamento de Física de la Materia Condensada y en un 11.44 % por parte del departamento de Física Aplicada y por algunos investigadores sueltos de otros departamentos, que no llegan al 1 %. Por lo tanto, esta comunidad se podría identificar, principalmente, con el área de Química (excluyendo a Química Analítica), complementada con los investigadores de Física de la Materia Condensada y de Aplicada que más colaboran con esta área.

- **Comunidad 7:** Contiene 186 investigadores, casi la mitad que la comunidad 1. El departamento de Producción animal y Ciencias de los Alimentos pertenece casi íntegramente a esta comunidad, constituyendo un 71.51 % de la misma. La completan la mitad del departamento de Química analítica (23.12 %) y una pequeña parte del de Matemática Aplicada (2.15 %). Luego está comunidad se identificaría con el departamento de Producción animal y Ciencias de los Alimentos, añadiéndole los investigadores de Química y Matemática Aplicada que colaboran con ese departamento.

- **Comunidad 6:** Esta comunidad tiene 156 investigadores, parecida a la comunidad 7. Está formada por los departamentos de Ciencias de la Tierra (78.85 %) y Didáctica de la Ciencias Experimentales (9.62 %), ambos íntegramente. Luego lo completan algunos investigadores de Métodos estádisticos (2.56 %), Producción Animal y Ciencias de los Alimentos (3.85 %) y Química Análitica (3.21 %). Luego esta comunidad sería la unión de los departamentos de Ciencias de la Tierra y Didáctica de las Ciencias Experimentales, con unos pocos investigadores de otros departamentos que colaboran con estos dos.

  Si nos fijamos en el dendrograma (imagen 10), vemos que estas 3 comunidades que acabamos de explicar, que son las 3 más grandes, son las que se van diviendo en el paso de 9 a 19 y 19 a 29 comunidades.

- **Comunidad 3:** Hay 77 investigadores que pertenecen a esta comunidad, bastante por

debajo de las 3 comunidades más grandes. Está formada por parte de los departamentos de Física Teórica (67.53 %), Física de la Materia Condensada(19.48 %), Física Aplicada (5.19 %), Matemática Aplicada (5.19 %) y Matemáticas (2.6 %). Esta comunidad, la podemos interpretar como una unión de partes de departamentos de matemática y Física, en torno al departamento de Física Teórica.

- **Comunidad 8:** Esta comunidad está formada por 40 investigadores, de los cuales el 92.5 % pertenecen a Física Teórica. Se puede identificar directamente como la parte del Departamento de Física Teórica formada por investigadores que colaboran principalmente entre ellos.

- **Comunidad 9:** Esta comunidad de 40 investigadores, está dividida en un 50 % de Matemáticas y un 50 % de Matemática Aplicada.

- **Comunidad 2:** Contiene 38 investigadores, de los cuales un 79 % son del departamento de Matemática Aplicada, un 15.8 % al de Matemáticas y un 5.2 % al de Física Teórica.

  Estas 2 últimas comunidades representan un fenómeno curioso de la estructura de colaboración de los 2 departamentos de Matemáticas que las componen. Ambos departamentos están separados en dos partes, que colaboran más con una de las partes del otro departamento que con la otra parte de su mismo departamento.

- **Comunidad 4:** Tiene 35 investigadores, que pertenecen en su gran mayoría al departamento de Química Analítica (91.42 %), el cual se encuentra divido entre la comunidad 7 y esta. Por lo tanto, podemos identificar esta comunidad como una división de Química Analítica.

- **Comunidad 5:** La comunidad más pequeña, está compuesta por investigadores de matemáticas: Métodos estadísticos (87.5 %), Matemática Aplicada (9.38 %) y Matemáticas (3.12 %). Por lo tanto, esta comunidad representa principalmente el departamento de Métodos estadísticos.

Esta identificación de comunidades con departamentos, nos permite hacernos una primera idea de como es la estructura de colaboración en artículos con impacto JCR (recordar como ha sido modelizada la red), inter- e intra-departamental. Hay que tener en cuenta, que la Modularidad de esta partición no era especialmente alta, luego no podemos asegurar que colaboración dentro de las comunidades es mucho mayor que entre comunidades. También es cierto que el hecho de que haya una comunidad mucho mayor que las otras, hace que el valor de la modularidad no pueda ser muy alto.

Si nos fijamos de nuevo en el dendrograma, vemos que muchas de estas comunidades se mantienen invariantes cuando pasamos a 19 o 29 comunidades.

4.3.2. Partición de 29 comunidades

Para esta partición, no tiene sentido hacer una identificación de comunidades, ya que hay un número bastante mayor de comunidades que de departamentos y no se puede establecer fácilmente una relación entre ambas particiones.

Es una partición interesante desde el punto de vista de la resolución del problema, porque la Modularidad es aproximadamente igual a la Modularidad máxima y es la partición con mayor valor de índices de similaridad.

Además, en el siguiente apartado, nos va a servir como punto de inicio para encontrar una partición con un valor de Modularidad similar y un número de comunidades mucho menor, utilizando un procedimiento heurístico. Debido a esto, sí que vamos a presentar tanto su gráfico de barras, como sus tablas con los investigadores por departamentos y comunidad, para tener una idea de que estructura tiene esta partición, pero no haremos una análisis en detalle de las comunidades.

Presentamos entonces, el gráfico de barras sólo con las 7 comunidades más grandes, para que sea más legible y visual. El gráfico de barras completo (gráfico 19) está en el apéndice B.

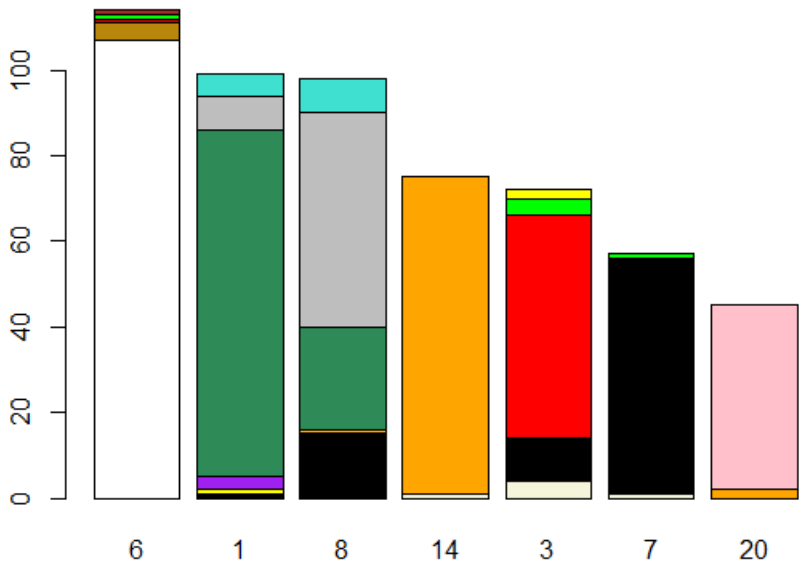

Figura 13: Gráfico en el que se muestra la composición por departamentos (colores) de las 7 comunidades con mayor número de investigadores (barras).

Y la leyenda de colores:

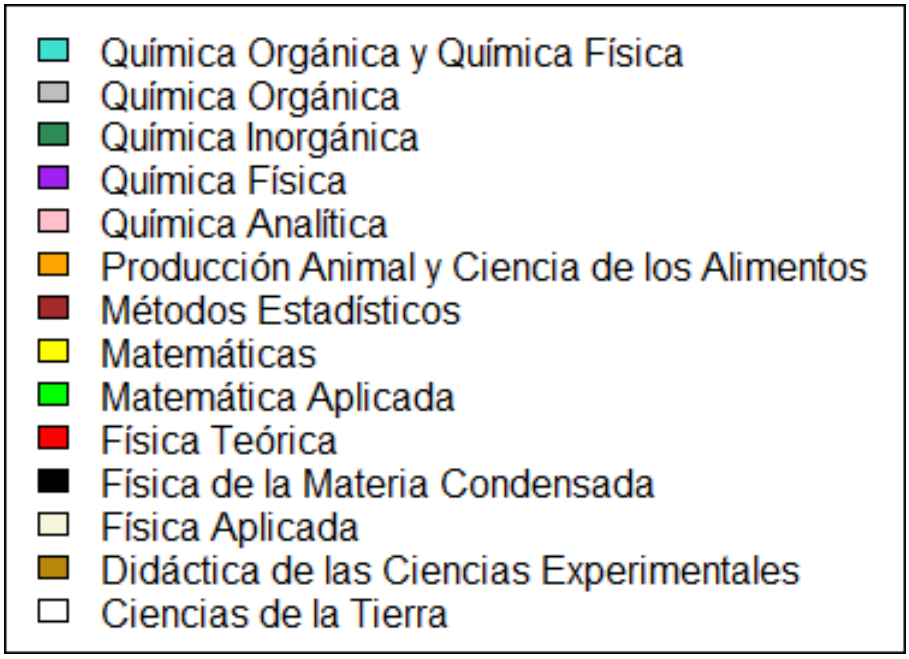

Figura 14: Leyenda de colores de cada departamento.

Al igual que en la partición de 9 comunidades, las tablas de investigadores (tabla 11) y porcentaje (tabla 13) por comunidades, se encuentran en el apéndice A.

5. Obtención de la partición óptima

Finalmente, hemos obtenido 2 particiones interesantes con el algoritmo de Girvan-Newman. Una de 9 comunidades, que no tiene unos valores de los descriptores especialmente altos, pero que por su simpleza, nos permite una primera aproximación a una estructura de comunidades óptima y una identificación de estas comunidades con los departamentos. La otra partición, de 29 comunidades, tiene valores de modularidad e índices de similaridad bastante buenos, pero es demasiado compleja como para intentar darle una interpretación a su estructura de comunidades.

Nuestro objetivo sería entonces, encontrar una partición que sea simple, pero con valores de los descriptores similares a la partición de 29 comunidades. Para ello se probó lo siguiente, si nos fijamos en la gráfica 8 de variación de modularidad frente a las distintas particiones, nos damos cuenta que todas las divisiones en comunidades que van desde la 9 a la 17 y de la 19 a la 28, no contribuyen a un aumento significativo de la modularidad. Entonces, se identificaron que comunidades correspondían a estas divisiones y de que comunidades se habían divido, para unirlas de nuevo. Todo esto partiendo de la partición de 29 comunidades.

Una vez que se vuelven a unir las comunidades cuya división no implicaba un aumento considerable de la modularidad, queda una partición de **12** comunidades, con un valor de la modularidad de 0,7678. Este es un resultado bastante bueno, ya que hemos conseguido una partición simple, cuyas comunidades se pueden identificar con los departamentos y apenas se ha reducido la modularidad con respecto a la partición de 29 comunidades (0,7918), manteniéndose por encima de la modularidad de la partición departamental (0,7226).

A continuación, se muestra una tabla con los valores de modularidad e índices de similaridad de las dos particiones seleccionadas en el apartado anterior (9 y 29) y de la obtenida en este apartado.

Particiones	**Modularidad**	**Índices de similaridad**		
		Wallace	**Jaccard**	**Rand**
9 comunidades	0.5413	0.5114	0.3014	0.8388
29 comunidades	0.7918	0.6056	0.4269	0.9380
12 comunidades	0.7678	0.5873	0.4100	0.9112

Cuadro 3: Tabla con valores de los descriptores para las distintas particiones obtenidas, que son consideradas de interés.

Vemos que los valores de la Modularidad y los índices son muy similares para las particiones de 12 y 29 comunidades.

Antes de hacer un análisis detallado de las 12 comunidades, vamos a mostrar un dendrograma que nos muestra que comunidades de la partición de 12 vienen de qué comunidades de la de 9

y cuales se han mantenido invariantes.

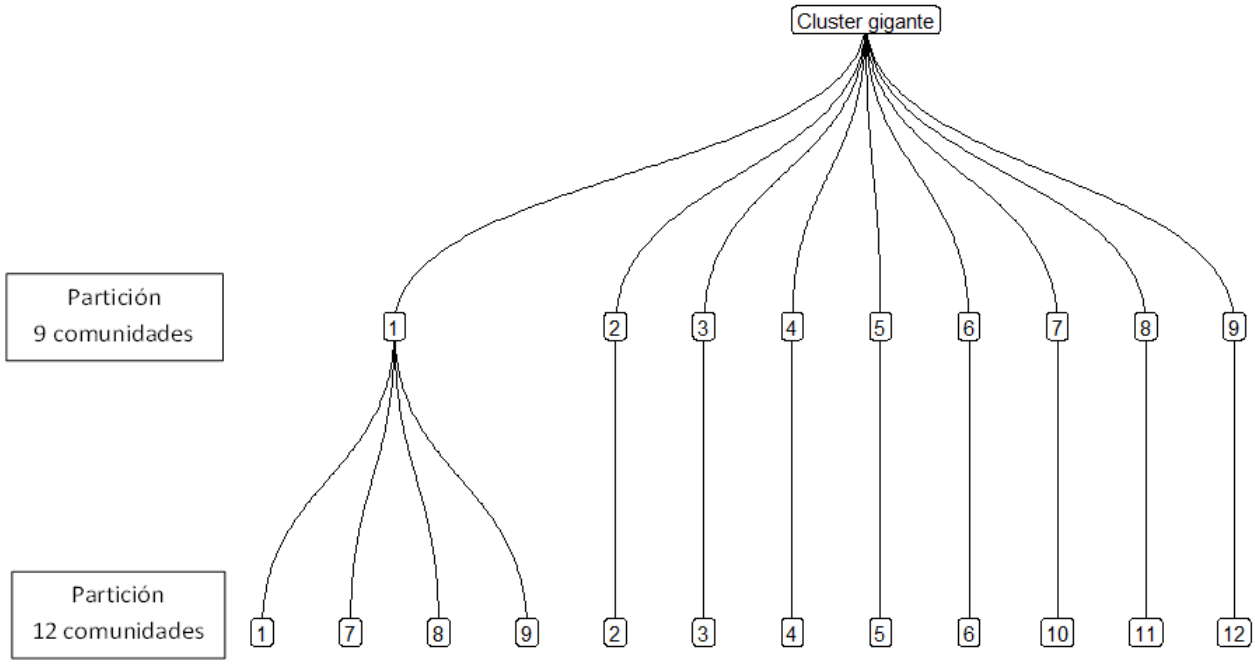

Figura 15: Dendrograma que muestra la relación de las comunidades de la partición de 12 y de 9.

Pasamos entonces a la identificación con departamentos de las comunidades de la partición. En el caso de las comunidades que son idénticas a las de la partición de 9 comunidades, se remite a la identificación realizada en el apartado 4.3.1.

Empezamos mostrando el gráfico de barras

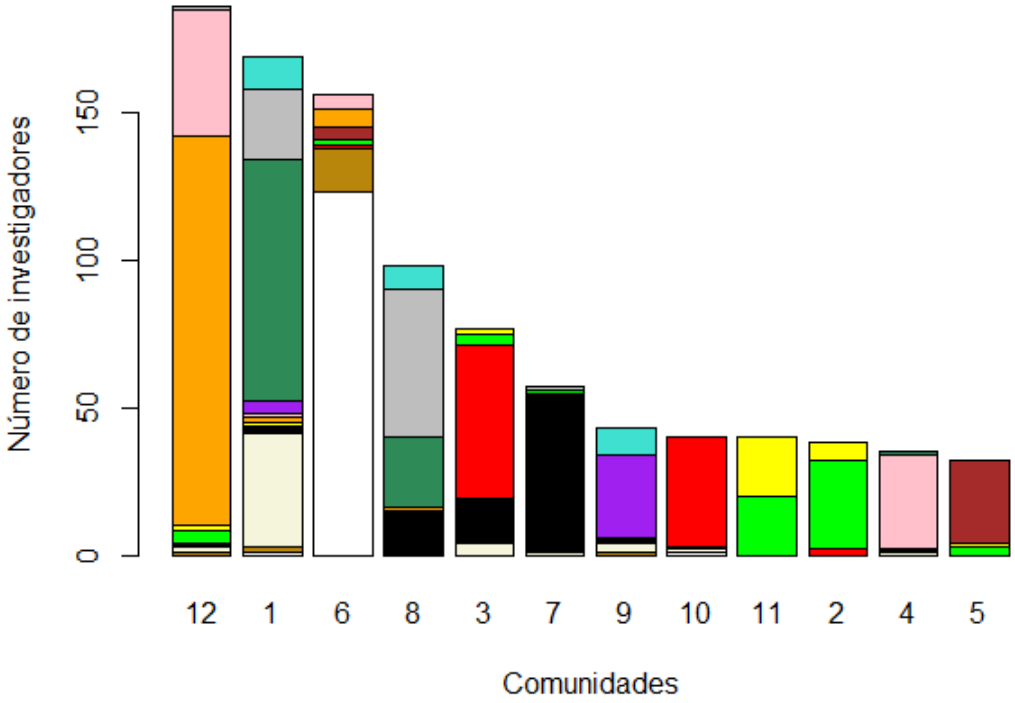

Figura 16: Gráfico en el que se muestra la composición por departamentos (colores) de cada comunidad (barras), para la partición óptima.

y la leyenda de colores por departamento

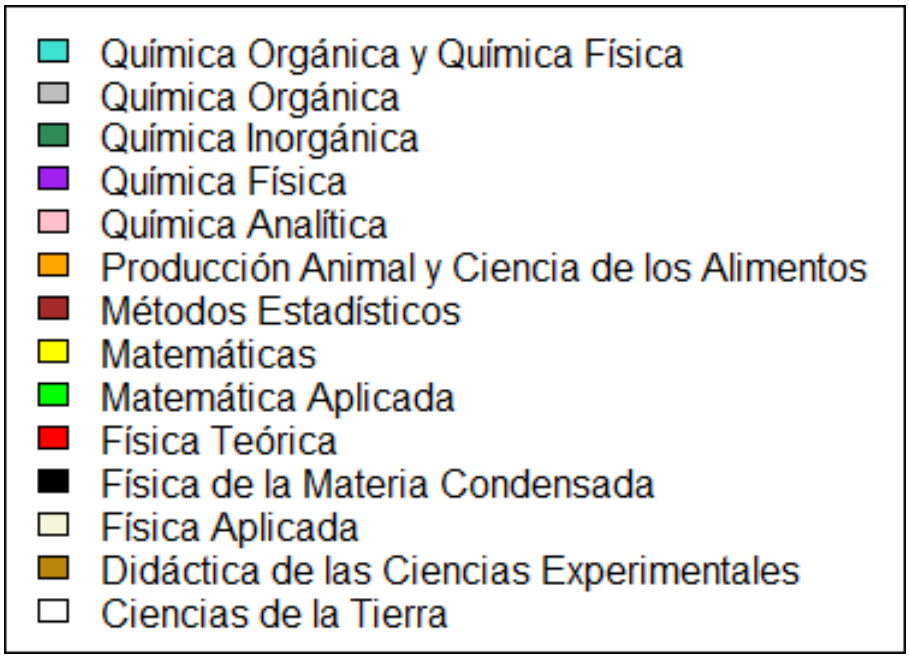

Figura 17: Leyenda de colores por departamento.

Incluimos a continuación las tablas con el número de investigadores por departamento y comunidad (tabla 4), el porcentaje de cada comunidad que pertenece los departamentos (tabla 5) y en este caso vamos a incluir también como están repartidos los departamentos por comunidades, en porcentaje (tabla 6).

Departamentos \ Comunidades	1	2	3	4	5	6	7	8	9	10	11	12
Ciencias de la Tierra	1	0	0	0	0	123	0	0	0	1	0	0
Didáctica de las Ciencias Experimentales	2	0	0	0	0	15	0	0	1	0	0	1
Física Aplicada	38	0	4	1	0	0	1	0	3	1	0	2
Física de la Materia Condensada	3	0	15	1	0	0	54	15	1	1	0	0
Física Teórica	0	2	52	0	0	1	0	0	0	37	0	0
Matemática Aplicada	0	30	4	0	3	2	1	0	0	0	20	4
Matemáticas	1	6	2	0	1	0	0	0	1	0	20	2
Métodos Estadísticos	0	0	0	0	28	4	0	0	0	0	0	0
Producción Animal y Ciencia de los Alimentos	2	0	0	0	0	6	0	1	0	0	0	133
Química Analítica	1	0	0	32	0	5	0	0	0	0	0	43
Química Física	4	0	0	0	0	0	0	0	28	0	0	0
Química Inorgánica	82	0	0	1	0	0	0	24	0	0	0	0
Química Orgánica	24	0	0	0	0	0	1	50	0	0	0	1
Química Orgánica y Química Física	11	0	0	0	0	0	0	8	9	0	0	0
Total investigadores por comunidad	**169**	**38**	**77**	**35**	**32**	**156**	**57**	**98**	**43**	**40**	**40**	**186**

Cuadro 4: Tabla de identificación de comunidades y departamentos, para la partición de 12 comunidades.

Departamentos \ Comunidades	1	2	3	4	5	6	7	8	9	10	11	12
Ciencias de la Tierra	0.59	0	0	0	0	78.85	0	0	0	2.5	0	0
Didáctica de las Ciencias Experimentales	1.18	0	0	0	0	9.62	0	0	2.33	0	0	0.54
Física Aplicada	22.49	0	5.19	2.86	0	0	1.75	0	6.98	2.5	0	1.08
Física de la Materia Condensada	1.78	0	19.48	2.86	0	0	94.74	15.31	2.33	2.5	0	0
Física Teórica	0	5.26	67.53	0	0	0.64	0	0	0	92.5	0	0
Matemática Aplicada	0	78.95	5.19	0	9.38	1.28	1.75	0	0	0	50	2.15
Matemáticas	0.59	15.79	2.6	0	3.12	0	0	0	2.33	0	50	1.08
Métodos Estadísticos	0	0	0	0	87.5	2.56	0	0	0	0	0	0
Producción Animal y Ciencia de los Alimentos	1.18	0	0	0	0	3.85	0	1.02	0	0	0	70.97
Química Analítica	0.59	0	0	91.43	0	3.21	0	0	0	0	0	23.12
Química Física	2.37	0	0	0	0	0	0	0	65.12	0	0	0
Química Inorgánica	48.52	0	0	2.86	0	0	0	24.49	0	0	0	0
Química Orgánica	14.2	0	0	0	0	0	1.75	51.02	0	0	0	0.54
Química Orgánica y Química Física	6.51	0	0	0	0	0	0	8.16	20.93	0	0	0
Porcentaje total	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Cuadro 5: Tabla de identificación de comunidades y departamentos, para la partición de 12 comunidades. Porcentaje por comunidades.

Departamentos \ Comunidades	1	2	3	4	5	6	7	8	9	10	11	12	Porcentaje total
Ciencias de la Tierra	0.8	0	0	0	0	98.4	0	0	0	0.8	0	0	**100**
Didáctica de las Ciencias Experimentales	10.53	0	0	0	0	78.95	0	0	5.26	0	0	5.26	**100**
Física Aplicada	76	0	8	2	0	0	2	0	6	2	0	4	**100**
Física de la Materia Condensada	3.3	0	16.48	1.1	0	0	59.34	16.48	1.1	1.1	0	0	**100**
Física Teórica	0	2.17	56.52	0	0	1.09	0	0	0	40.22	0	0	**100**
Matemática Aplicada	0	46.88	6.25	0	4.69	3.12	1.56	0	0	0	31.25	6.25	**100**
Matemáticas	3.03	18.18	6.06	0	3.03	0	0	0	3.03	0	60.61	6.06	**100**
Métodos Estadísticos	0	0	0	0	87.5	12.5	0	0	0	0	0	0	**100**
Producción Animal y Ciencia de los Alimentos	1.42	0	0	0	0	4.26	0	0.71	0	0	0	93.62	**100**
Química Analítica	1.23	0	0	39.51	0	6.17	0	0	0	0	0	53.09	**100**
Química Física	12.5	0	0	0	0	0	0	0	87.5	0	0	0	**100**
Química Inorgánica	76.64	0	0	0.93	0	0	0	22.43	0	0	0	0	**100**
Química Orgánica	31.58	0	0	0	0	0	1.32	65.79	0	0	0	1.32	**100**
Química Orgánica y Química Física	39.29	0	0	0	0	0	0	28.57	32.14	0	0	0	**100**

Cuadro 6: Tabla de identificación de comunidades y departamentos, para la partición de 12 comunidades. Porcentaje por departamentos.

Pasamos al análisis de las 12 comunidades, apoyándonos en el gráfico y las tablas:

- **Comunidad 12:** Tiene 186 investigadores y es idéntica a la comunidad 7 de la partición de 9 comunidades. La identificábamos principalmente con el departamentos de Producción Animal y Ciencia de los Alimentos y parte del departamento de Química Analítica.

- **Comunidad 1:** Es la segunda más grande con 169 investigadores. Esta comunidad es una de las divisiones de la comunidad 1 de la partición de 9 comunidades, la cual identificábamos principalmente con el área de Química. Está formada por parte de los departamentos de Química Inorgánica (48.52 %), Física Aplicada (22.49 %), Química Orgánica (14.2 %), Química Orgánica y Química Física (6.51 %) y Química Física (2.37 %), principalmente. Esta nueva comunidad la podemos identificar con la mayor parte del departamento de Química Inorgánica e investigadores de otros departamentos de las áreas de Física y Química que colaboran más con este departamento. Es interesante ver que el departamento de Física Aplicada pertenece casi íntegramente a esta comunidad, lo que nos indica que hay una gran colaboración entre este departamento y el de Química Inorgánica.

- **Comunidad 6:** Esta comunidad de 156 investigadores es idéntica a la comunidad 6 de la partición de 9 comunidades, la cual identificábamos como la unión de los departamentos, íntegros, de Ciencias de la Tierra y Didáctica de las Ciencias Experimentales.

- **Comunidad 8:** Tiene 98 investigadores y está formada por Química Orgánica (51 %), Química Inorgánica (24.5 %), Física de la Materia Condensada (15.3 %) y Química Orgánica y Química Física (8.16 %). Esta comunidad es otra de las que surgen de la división de la comunidad 1 de la partición de 9 comunidades. Se podría identificar con la mayor parte del departamento de Química Orgánica y otros departamentos del area de química y física que colaboran estrechamente con este departamento.

- **Comunidad 3:** Contiene 77 investigadores y es idéntica a la comunidad 3 de la partición de 9 comunidades. Esta comunidad la identificábamos con una parte muy grande del departamento de Física Teórica, completada con investigadores de Física de la Materia Condensada y unos pocos del área de matemática, que colaboran más estrechamente con Física Teórica.

- **Comunidad 7:** Esta comunidad de 57 investigadores, está formada principalmente por el departamento de Física de la Materia Condensada (94.7 %) e investigadores sueltos de otros departamentos. Surge también de la división de la comunidad 1 de la partición de 9. Se puede identificar con los investigadores del departamento de Física de la Materia Condensada que colaboran principalmente entre ellos.

- **Comunidad 9:** Está compuesta por 43 investigadores y es la más pequeña de las comunidades que han surgido de la división de la comunidad 1 de la partición de 9. Se compone en un 65.1 % por investigadores del departamento de Química Física, por un 21 % de Química Orgánica y Química Físca y un 7 % de Física Aplicada. El departamento de Química

Física pertenece casi íntegramente a esta comunidad, luego podemos identificarla con este departamento, completado con investigadores de otros departamentos, con los que la colaboración es más estrecha.

- **Comunidad 10:** Tiene 40 investigadores y es la misma comunidad que la 8 de la partición de 9 comunidades, la cual estaba identificada con una parte del departamento de Física Teórica.
- **Comunidad 11:** Está formada por 40 investigadores y es idéntica a la comunidad 9 de la partición de 9 comunidades. Vimos que era una combinación a partes iguales, de los departamentos de Matemáticas (casi íntegramente) y de una parte de Matemática Aplicada
- **Comunidad 2:** Contiene 38 investigadores y es la misma comunidad que la 2 de la partición de 9. Estaba formada solo por investigadores del area de matemáticas, la mayoría de ellos del departamentos de Matemática Aplicada, con unos pocos de Matemáticas y Métodos Estadísticos.
- **Comunidad 4:** La forman 35 investigadores. Es idéntica a la comunidad 4 de la partición de 9, que estaba conformada casi por completo por una parte departamento de Química Analítica, luego representaría una división de este departamento.
- **Comunidad 5:** Es la comunidad más pequeña con 32 investigadores, es igual que la comunidad 5 de la partición de 9 comunidades y la identificábamos directamente con el departamento de Métodos Estadísticos, completado con algunos investigadores de los otros departamentos de matemáticas.

Por último, en la imagen 20 del apéndice B, se muestran una representación del Cluster gigante de la red, en el que el tamaño de los nodos es proporcional al grado pesado y el color corresponde la comunidad, de la partición de 12 comunidades, a la que pertenece el investigador.

6. Conclusiones

En este trabajo se ha utilizado el algoritmo de Girvan-Newman y magnitudes como la modularidad, el índices de Wallace, índice de Rand e índice de Jaccard, para estudiar la estructura de comunidades y compararla con la departamental en la red de colaboración científica de la Universidad de Zaragoza, concrétamente del área de Ciencias. El algoritmo es un método disgregativo con el que obtener una partición de la red en comunidades que maximice la Modularidad. Con este método se han obtenido 2 particiones de interés para comprender la estructura colaborativa de la red. Se ha conseguido una tercera partición aplicando un sencillo procedimiento heurístico de aglomeración de comunidades a una de las dos particiones de interés obtenidas. Esta tercera partición de 12 comunidades es considerada óptima, en cuanto al compromiso entre simplicidad y valor alto de modularidad, de entre todas las encontradas.

Analizando los resultados obtenidos, podemos extraer las siguiente conclusiones:

1. Lo primero que podemos decir es que, la estructura departamental del área de Ciencias ya es una buena partición en sí, desde el punto de vista de colaboración entre investigadores, ya que un tiene un valor de Modularidad igual a 0,723 , considerablemente alto.

2. Teniendo en cuenta que el problema de obtención de particiones óptimas en redes complejas es un problema complejo y que no existe un método que asegure que la solución encontrada es óptima, podemos asumir que la partición encontrada, con un valor de la modularidad de 0,7678 , es una buena solución y por lo tanto, el algoritmo de Girvan-Newman ha funcionado bien.

3. Más allá de las particiones de interés obtenidas, nos damos cuenta que este método disgregativo funciona bien como diseccionador de la estructura colaborativa de los investigadores en la red. Fijándonos principalmente en la gráfica 8 e identificando las comunidades que se dividen las primeras, podemos saber qué grupos de investigadores colaboran más entre sí y cuáles apenas colaboran.

4. Por último, del análisis de las comunidades en la partición óptima deducimos un aspecto interesante de la estructura departamental. Si nos fijamos en la tabla 6, concluimos que la mayoría de los departamentos están repartidos en 1 ó 2 comunidades, a excepción de Química Orgánica y Química Física y de Física de la Matería Condensada que están repartidos en 3. Esto quiere decir que la la colaboración de investigadores dentro de un departamento, no está estructurada en múltiples pequeños grupos, sino en unos pocos grupos grandes.

   Además, observando la tabla 5, se puede concluir que casi todas las comunidades consisten en un núcleo grande que pertenece a un solo departamento, completado con grupos de investigadores de otros departamentos, exceptuando la comunidad 11, que está constituida a partes iguales por Matemáticas y Matemática Aplicada. Esto significa que la estructura

colaborativa entre departamentos consiste en un grupo grande que pertenece a un solo departamento, el cual se rodea de algunos investigadores de otros departamentos y no en pequeños grupos de diversos departamento que colaboran unos con otros y todos con el mismo peso.

A la hora de analizar los resultados y sacar conclusiones, hay que tener siempre presente las limitaciones de este estudio. La información que se ha utilizado para modelizar la red, que hay diferentes formas de utilizar esta información y hay otro tipo de información (colaboración conjunta de los investigadores en proyectos, por ejemplo) para generar la red de colaboración. Asimismo, hay que tener en cuenta las restricciones que se han impuesto a la hora de seleccionar investigadores y que por lo tanto nuestra red no representa al área de Ciencias al completo. Además, existen muchos otros métodos de detección de comunidades en redes (método de Louvain, métodos basados en magnitudes distintas a la betweenness centrality, métodos aglomerativos, etc).

Estas limitaciones dejan abiertas distintas líneas de investigación futuras, basadas en este estudio: repetir el proceso con la red modelizada de diversas maneras; aplicar el método a otras áreas de la Universidad, a la universidad completa o incluso aplicarlo a departamentos por separado, por último, utilizar otros métodos de detección de comunidades.

Finalmente, los resultados obtenidos en este trabajo, podrían tener una aplicación en el caso de una eventual reducción o reestructuración de los departamentos del área de Ciencias. Si uno de los criterios fuese minimizar el impacto negativo en la estructura de colaboración o la producción científica de los investigadores, se podría utilizar la partición de 12 comunidades como base para dicha reestructuración.