X Congreso de Ingeniería de Organización
Valencia, 7 y 8 de septiembre de 2006

Estudio de la red de participaciones en tribunales de tesis doctorales de Organización y Gestión de Empresas en España

José Ignacio Santos Martín[1], Ricardo del Olmo Martínez[2], Javier Pajares Gutiérrez[3]

[1] Dpto. Ingeniería Civil, Universidad de Burgos, Avenida de Cantabria, s/n. 09006 Burgos. jisantos@ubu.es
[2] Dpto. Ingeniería Civil, Universidad de Burgos, Avenida de Cantabria, s/n. 09006 Burgos. rdelolmo@ubu.es
[3] Dpto. Organización y CIM, Universidad de Valladolid, Pª Cauce, s/n. 47011 Valladoldid. pajares@eis.uva.es

Resumen

Esta comunicación pretende divulgar algunos de los conceptos más importantes sobre Teoría de Redes Sociales puestos en práctica en el estudio de un caso particular que es la red de participaciones en tribunales de tesis doctorales en el área de Organización de Empresas. Si bien la investigación sobre redes no es una actividad reciente, en los últimos años está despertando un renovado interés al abordar el estudio de redes complejas y muy grandes como es la WWW o Internet. Desde el área de las matemáticas y la física se están proponiendo nuevas técnicas y algoritmos de cálculo para el análisis de redes que permiten a los investigadores profundizar en el conocimiento de estos sistemas relacionales complejos. El análisis de la red propuesta ha sido realizado mediante el paquete de análisis de redes grandes Pajek, y los datos han sido obtenidos de la base TESEO del Ministerio de Educación y Ciencia.

Palabras clave: Redes, Redes Sociales, TESEO

1. Introducción

En el mundo real encontramos numerosos ejemplos de fenómenos muy diversos, químicos (como la estructura del ADN), tecnológicos (Internet), de información (la World Wide Web), económicos (clusters industriales) o incluso puramente sociales (procesos de opinión), donde resulta importante no sólo conocer las partes individuales que los forman sino también la estructura de relaciones entre las mismas, que llamamos red.

Nuestro interés se centra en redes sociales, es decir sistemas donde los agentes constituyentes son personas. A diferencia de otro tipo de redes, las redes sociales tienen características propias (Jin *et al*. 2001), principalmente porque evolucionan en el tiempo a una velocidad sensiblemente menor (pensemos por ejemplo en cómo se desarrollan las relaciones de amistad entre personas), y porque muestran un elevado nivel de transitividad (los amigos de mis amigos suelen ser también amigos míos).

La investigación sobre redes sociales no es nueva, si bien en los últimos años está despertando un renovado interés en muchos grupos de investigación al incorporar avances sobre estructuras y propiedades de redes realizados desde áreas de conocimiento como las matemáticas y la física (Newman y Girvan, 2004). Nuestra comunicación presenta algunos de estos nuevos algoritmos puestos en práctica en el estudio de una red atípica de colaboraciones como son las participaciones en tribunales de tesis en las Universidades españolas relacionadas con la Organización y Gestión de Empresas.

2. Redes sociales y teoría de grafos

2.1. Conceptos generales

La teoría de grafos constituye desde hace tiempo una interesante área de investigación dentro de las matemáticas preocupada por el estudio de las propiedades de las redes. Muchas de las aportaciones de la teoría de grafos han sido posteriormente utilizadas y en ocasiones ampliadas por otras disciplinas como la Biología, la Física, la Economía, la Sociología o la Informática (Dorogovtsev y Mendes, 2003). En el caso de sistemas sociales la forma más sencilla de modelar una red social es mediante un grafo donde cada persona es representada mediante un vértice o nodo, y cada relación entre dos individuos mediante un lado o línea que une los vértices correspondientes.

Sobre esta base constituyente de elementos de una red, vértices y lados, podemos representar diferentes tipos de actores individuales así como distintas clases de relaciones entre los mismos (ver Figura 1). Imaginemos por ejemplo el estudio del envío de correos electrónicos dentro de una organización donde cada correo es representado mediante una flecha que identifica emisor y receptor (Figura 1-i). También podríamos estar interesados en estudiar los encuentros personales entre individuos de una organización (Figura 1-ii), en cuyo caso cada relación interpersonal es representada por un lado bidireccional o relación recíproca. Además podríamos enriquecer este estudio incorporando los diferentes departamentos a los que pertenecen las personas (diferentes clases de vértices). Un último ejemplo podría ser un estudio de proyectos de I+D intra-empresarial en una industria (Figura 1-iii), donde nos interesa conocer no sólo las empresas que participan sino también su esfuerzo en investigación con relación a toda la industria (peso de los vértices), y su gasto en cada uno de los acuerdos con otras empresas (peso de los lados).

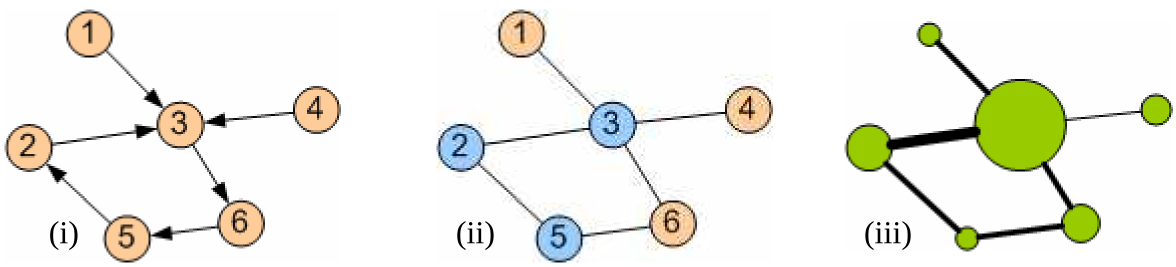
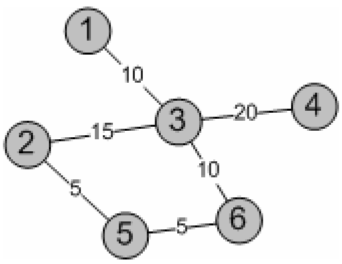


Figura 1. Diferentes tipos de redes: red dirigida de vértices semejantes (i); red recíproca con dos clases de vértices (ii); red recíproca con vértices y lados valorados.

Una manera de representar analíticamente una red es mediante la matriz de adyacencias *A*. Es una matriz cuadrada *nxn* (siendo *n* el número de vértices) donde cada elemento *Aij* tiene un valor representativo del lado entre los índices-vértices correspondientes (ver Figura 2).

Al aplicar estos conceptos al estudio de sistemas sociales debemos tener en cuenta algunas limitaciones que provienen de la propia naturaleza de los fenómenos sociales. Las personas no nos encontramos rígidamente fijas en un punto, sino que nos movemos en entornos de relaciones sociales que dependiendo del objeto de estudio serán más o menos amplios y variables. La gran dificultad se encuentra en definir la magnitud de estudio que determinará la relación entre individuos, así como poder disponer de métodos de obtención de datos significativos. El uso de encuestas, por ejemplo, puede no ser un medio preciso a la hora de evaluar un mapa de relaciones sociales en una comunidad.



$$A = \begin{pmatrix} 0 & 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 15 & 0 & 5 & 0 \\ 10 & 15 & 0 & 20 & 0 & 10 \\ 0 & 0 & 20 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 5 \\ 0 & 0 & 10 & 0 & 5 & 0 \end{pmatrix}$$

Figura 2. Red recíproca valorada (semejante a la red propuesta para el estudio de las participaciones en tribunales de tesis) y su correspondiente matriz de adyacencias.

2.2. Redes valoradas

Los grafos no valorados (ver Figura 2), en los que la matriz de adyacencias contiene únicamente ceros y unos, suelen ser los más empleados en el modelado de redes sociales. Todas las propiedades de estas redes se calculan en base a la observación de si existe o no relación entre cada par de individuos.

Un ejemplo relacionado con el propósito de esta comunicación lo encontramos en el trabajo sobre “redes de colaboración científica” de Newman (2001). Se estudia la cooperación en la investigación en física, biomedicina e informática a partir de la información sobre publicaciones en revistas científicas. Un conjunto de artículos pueden ser representados mediante un grafo de autores donde cada lado expresa la participación de dos investigadores en alguna publicación.

Sin embargo en ocasiones podemos disponer de más información con la que enriquecer nuestro análisis. Pensemos en la red de colaboración anterior donde además de tener en cuenta la colaboración o no de dos investigadores en algún artículo incorporamos el número exacto de publicaciones conjuntas. En este caso nos enfrentamos a una red valorada en el que cada lado posee un peso proporcional a la fortaleza de dicha relación. Siguiendo con el ejemplo, parece evidente que un mayor número de publicaciones conjuntas responde a una mayor cooperación entre dos investigadores, y por consiguiente a una mayor proximidad social.

Este tipo de redes valoradas con pesos en los lados requiere de un tratamiento diferente a las redes binarias tradicionales, lo que en ocasiones complica el cálculo de algunas de sus características. Recientemente algunos autores (Newman, 2004) están introduciendo nuevos algoritmos o adaptando otros anteriores para estudiar las propiedades de estas redes, abriendo nuevas posibilidades de investigación en este campo.

3. Red de participación en tribunales de tesis del área de Organización de Empresas

3.1. Consulta a la base de datos TESEO

La base de datos TESEO (MEC) recoge los principales datos de las tesis doctorales defendidas en las universidades españolas. Nuestro estudio parte de los registros de las tesis defendidas dentro del campo de la Organización y Gestión de Empresas en España durante el periodo 1995-2005.

Hemos construido un modelo de red representativo de la siguiente forma, cada miembro del tribunal constituye un vértice y cada lado expresa la participación de dos miembros en un mismo tribunal. Hemos incluido también a los directores en la red de participaciones puesto que a nuestro juicio juegan un papel importante en la elección de los posibles miembros de un tribunal. El peso correspondiente a un lado se computa como el número de lecturas de tesis en las que dos personas han coincidido en mismo tribunal. Todas las propiedades que abordamos a continuación han sido calculadas mediante el paquete de análisis de redes *Pajek* (Batagelj y Mrvar).

Conviene señalar que una parte importante del trabajo realizado se ha invertido en la depuración y comprobación de los datos obtenidos de TESEO. Lamentablemente la base de datos del ministerio solo asigna al autor, así como al director o los miembros del tribunal, un único campo texto libre no relacionado con ninguna otra tabla de datos personales, y por consiguiente sin ningún tipo de control o verificación. Resulta frecuente encontrarse con que un mismo profesor figura hasta con más de 10 nombres distintos según se haya incluido o no todos los apellidos, se haya puesto o no acentos, o simplemente se haya cambiado el orden apellidos y nombre, además de erratas en los mismos. Esto ha complicado notablemente el trabajo de recogida de datos puesto que la base fundamental del análisis que perseguimos nos exige la identificación unívoca de los agentes y sus relaciones.

3.1 Características generales de la red de participaciones

Los datos recogidos del periodo 1995-2005 constituyen 1.749 tesis que pueden ser representadas mediante una red de n=3.325 vértices y m=19.136 lados, una red grande (más de 1000 vértices) y dispersa, con un grado de dispersión (1) de 0,0034. La figura 3 muestra una visualización de la red completa.

$$dispersión = \frac{2m}{n(n-1)} \quad (1)$$

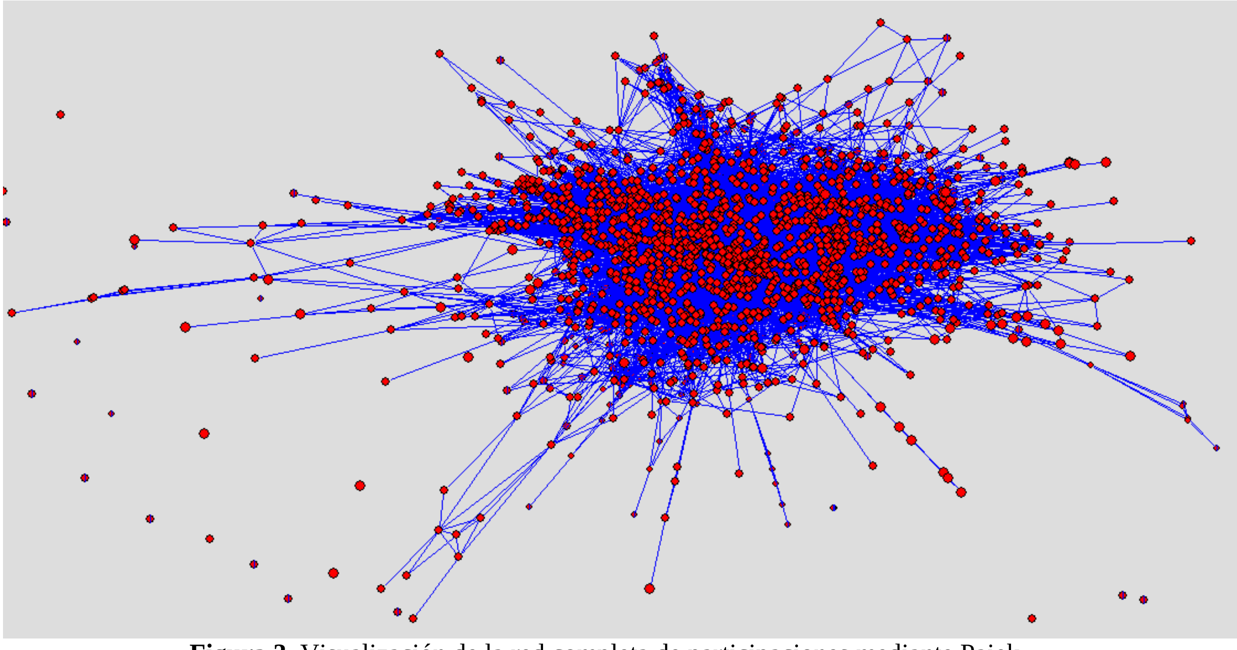


Figura 3. Visualización de la red completa de participaciones mediante Pajek.

Los principales parámetros de la red de participaciones están recogidos en la tabla 1. El significado de cada uno de ellos se explica a continuación.

Tabla 1. Parámetros principales de la red de participaciones.

Número de tesis	1.749
Número de Profesores (vértices)	3.325
Número de relaciones entre miembros (lados)	19.136
Promedio tesis por profesor	3,15
Diámetro (distancia geodésica mayor)	10
Distancia geodésica media	3,86
Número de componentes	33
Tamaño del componente mayor	3.119
Coeficiente de Clustering	0,0339

3.2 Proximidad entre individuos y efecto "Small World"

En el caso de redes recíprocas como la nuestra se denomina camino a una secuencia de vértices *(v1,v2,…,vj)* donde existen lados que unen los pares de vértices correspondientes, siendo el número total del lados la longitud del camino. El camino más corto es aquella secuencia más pequeña que une dos vértices concretos y se llama distancia geodésica *d(i,j)*.

En una red ponderada sin embargo se tienen en cuenta los pesos de los lados a la hora de estimar la ruta más corta entre dos puntos. Recordemos que el número de ocasiones en que dos profesores coinciden en un tribunal de tesis es recogido en el peso del lado correspondiente. Se puede pensar que cuanto mayor es este número tanto mayor es la proximidad social entre ambos, al menos ambos individuos han tenido un mayor número de encuentros personales frente al resto. Los algoritmos empleados para calcular distancias mínimas utilizan estos pesos como si se trataran de costes asociados al desplazamiento entre dos vértices, la ruta más corta se calcula como aquella que tiene un coste total menor.

Un parámetro frecuentemente utilizado para medir el grado de proximidad entre individuos de una red es el promedio de las distancias geodésicas l entre todos los agentes (2). En la red de participaciones estudiada este valor es de 3,86. Esto pone de manifiesto que tan solo tres profesores separan a cualquier par de miembros de tribunales de tesis, lo que puede llamar la atención si recordamos que la muestra de estudio está formada por 3.225 individuos.

$$l = \sum_{i,j \in V} \frac{d(i,j)}{n(n-1)/2} \qquad (2)$$

Se suele hablar del efecto "Small World" cuando en una red, aun cuando pudiera contener muchos vértices, la distancia geodésica media es sensiblemente pequeña, y próxima a la que cabría esperar en una red con los mismos vértices pero relacionados unos con otros de forma aleatoria. Si vemos a una red social como un sistema donde la información se distribuye mediante contactos personales, la velocidad de difusión es tanto mayor cuando el promedio de caminos más cortos es más pequeño. El conocido estudio de Milgram (1967) puso de manifiesto que la distancia geodésica media en una red social no era mayor de 6, lo que posteriormente se ha popularizado con el nombre de los seis grados de separación. Podemos concluir diciendo que la red de participaciones de tribunales de tesis presenta efecto “Small World”.

Otra forma de caracterizar la proximidad entre los miembros de una red es mediante la longitud del mayor camino más corto, también llamada diámetro de la red. Algunos estudios han calculado el diámetro de la World Wild Web, que es aproximadamente de 19 (Albert *et al*, 1999). En nuestro caso el diámetro de la red de participaciones es más pequeño, de valor 10.

3.3 Distribución de los grados de los vértices

El grado de un vértice g_i se define para redes recíprocas como el número total de lados que comparten extremo con él. En nuestra red de participaciones si además tenemos en cuenta el peso de cada lado obtenemos en definitiva el número de contactos con otros profesores en los diferentes tribunales de tesis en los que ha participado (3). Pensemos que una lectura de tesis ofrece a cada miembro de tribunal una oportunidad de relación con otros 4 miembros de tribunal, además del director.

$$g_i = \sum_k A_{ik} \tag{3}$$

En vez de estudiar directamente este valor nosotros calcularemos el valor del número de tesis por profesor, directamente relacionado con el grado de un vértice, por ser una magnitud más fácilmente interpretable en la red de participaciones en tribunales de tesis. El promedio de tesis por profesor es de 3,15 para el periodo de estudio. Sin embargo resulta más interesante estudiar la distribución de frecuencias del número de tesis por profesor. La figura 4 muestra el histograma correspondiente. Salvo el valor central de la serie (entorno a 20) el número de tesis por profesor X sigue bastante bien una ley de potencia (4) con un valor para la constante α de 2,6.

$$p(X = k) = k^{-\alpha} \tag{4}$$

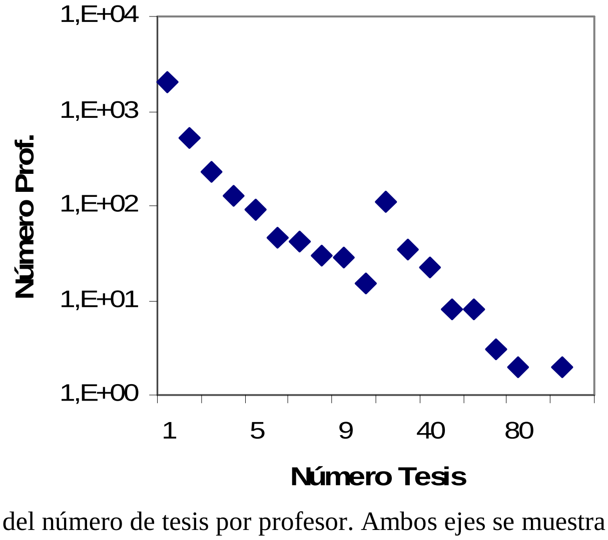


Figura 4. Distribución del número de tesis por profesor. Ambos ejes se muestra en escala logarítmica.

Esta característica de la distribución del número de tesis por profesor siguiendo una ley de potencia se presenta frecuentemente en redes de información como la WWW e Internet, y en ocasiones en algunas redes sociales (Dorogovtsev y Mendes, 2003).

3.4 Transitividad y coeficiente clustering

Una propiedad muy característica de las redes sociales es el fenómeno de transitividad, por el que en un grupo de personas es bastante frecuente que los amigos de una de ellas sean a su vez amigos. Para medir este comportamiento se define un parámetro de aglomeración C, llamado coeficiente de clustering, que expresa la probabilidad de que dos personas que mantienen un enlace en la red con un tercero tengan a su vez un enlace entre ellos mismos (Watts y Strogatz, 1998).

En redes puramente aleatorias $C=O(n^{-1})$, por lo que tiende a cero conforme el tamaño de la red se hace mayor. Sin embargo en redes sociales no ocurre lo mismo y aunque la red tenga un número grande de vértices, como es nuestro caso, el valor del coeficiente de clustering es diferente de cero. En nuestra red de participaciones si bien el coeficiente es significativamente distinto de cero (C=0,0339), sin embargo no adquiere un valor elevado lo que pone de manifiesto que en esta red el grado de transitividad entre los individuos no es importante. Por ejemplo en la red de colaboraciones científicas analizadas por Newman (2001) el coeficiente de clustering es mayor de 0,3.

3.5 Grado de intermediación (betweenness)

Una tarea muy interesante en el análisis de redes es poder identificar aquellos individuos que desde el punto de vista relacional de una red desempeñan un papel central. Se puede decir que un individuo tiene un papel central si posee un grado elevado, o si está próximo al resto de miembros de la red. Sin embargo se suele utilizar otra característica para medir la centralidad de un vértice y es el llamado grado de intermediación $c_B(i)$ (betweenness).

El grado de intermediación de un vértice se calcula como el número de caminos más cortos (distancia geodésica) que pasan a través de él (5). Recordando el ejemplo de difusión de información a través de contactos personales, aquel individuo que sea intermediario en un mayor número de flujos de información y que ocupa por tanto un lugar central en la red tendrá un coeficiente c_B elevado.

$$c_B(i) = \sum_{j<k} \frac{\#\text{Caminos más cortos entre j y k que pasan por i}}{\#\text{Caminos más cortos entre j y k}} \qquad (5)$$

En la Figura 5 se muestral el histograma del grado de intermediación (valores normalizados) calculado para los vértices de la red de participaciones. Como puede apreciarse la primera parte de la serie sigue una ley de potencia, aunque presenta una cola demasiado amplia que no se corresponde plenamente con esta función.

También se puede comprobar que existe una significativa correlación entre la distribución de c_B y la distribución del número de tesis por profesor (coeficiente de correlación de 0,82).

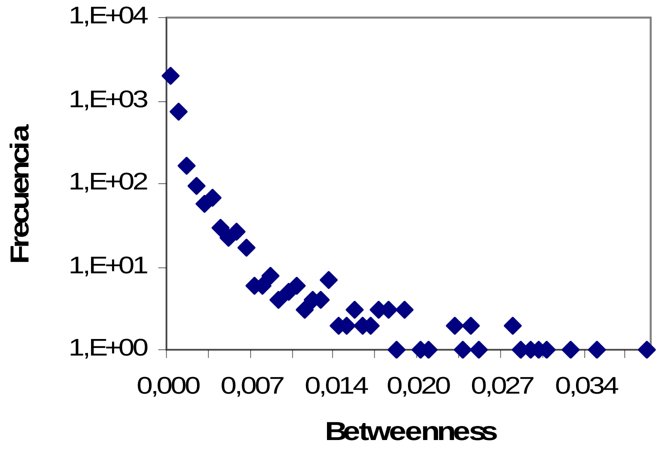


Figura 5. Distribución del grado de intermediación. El eje de ordenadas se muestra en escala logarítmica.

3.5 Componentes y núcleos

Un grupo de vértices en el que todos sus miembros pueden alcanzar a sus compañeros, de forma directa o a través de otros intermediarios del grupo, se denomina componente. En una red tan grande como la de nuestro estudio es importante hacer un análisis del número de componentes, y sobre todo conocer el tamaño del mayor de ellos también conocido como gigante. La red de participaciones tiene 33 componentes, conteniendo el mayor de ellos a 3.119 vértices y el segundo mayor a 11 vértices. Como vemos la inmensa mayoría de los miembros de tribunales de tesis pertenecen al mismo componente, y por lo tanto mantienen en mayor o menor grado una posibilidad de relación con el resto de la comunidad.

Un subconjunto de vértices se dice que forman un k-núcleo si todos los vértices están conectados al menos a k-vértices de dicho grupo. La Figura 6 recoge la distribución de los k-núcleos de nuestra red de participaciones. El 5-núcelo es el mayoritario, algo esperado puesto que cada lectura de tesis propicia una relación de un profesor con otros 5 colegas (director y los miembros del tribunal). Sin embargo resulta interesante el 19-núcleo compuesto por 86 profesores, con un coeficiente de clustering de 0,29 que demuestra una importante transitividad social entre los miembros del núcleo.

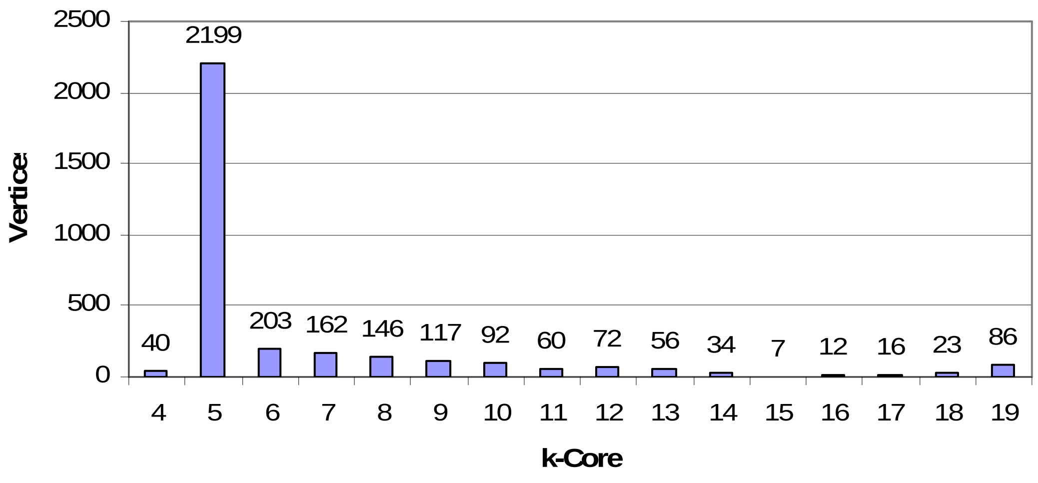


Figura 6. Histograma de la distribución de los k-núcleos de la red de participaciones.

4. Conclusiones

Se han analizado desde el punto de vista de las redes sociales los datos sobre los tribunales de tesis del área de Organización de Empresas del periodo 1995-2005. Observando le número de componentes, su tamaño, así como la distribución del grado de los vértices se puede concluir que estamos ante una "scale-free network" que puede describirse bastante bien mediante una ley de potencia. Destaca el efecto "small world" presente en la red con una distancia geodésica media entre sus miembros pequeña. Aunque se podía esperar una mayor transitividad en la red, los datos demuestran que esa característica de "los amigos de mis amigos son mis amigos", tan frecuente en otras redes sociales, no es muy significativa en este caso, teniendo las relaciones entre individuos una naturaleza más aleatoria.

Con este trabajo hemos querido poner de manifiesto la importancia de la teoría de redes para analizar desde otra perspectiva fenómenos económicos y sociales, abordados tradicionalmente desde modelos estadísticos clásicos. Fenómenos donde la relación entre los individuos juega un papel importante, caso de los procesos de innovación o las externalidades de conocimiento, pueden ser también estudiados desde la teoría de redes sociales.

Agradecimientos

Este trabajo ha sido financiado con el proyecto de investigación DPI2004-06590 del Ministerio de Educación y Ciencia español.

Referencias

Albert, R., Jeong, H., Barabasi, A. (1999). Diameter of the World-Wide Web. *Nature*, No 401, pp. 130-131.

Batagelj, V., Mrvar, A. *Pajek – Program for Large Network Analysis* [en línea]. Disponible en Web: http://vlado.fmf.uni-lj.si/pub/networks/pajek/

Dorogovtsev, S.N., Mendes, J.F.F. (2003). *Evolution of Networks. From Biological Nets to the Internet and WWW.* Oxfrod University Press.

Jin, E.M., Girvan, M., Newman, M.E.J. (2001). Structure of growing social networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, Vol. 64, No. 4II, pp. 461321-461328.

Ministerio de Educación y Ciencia. *Base de Datos de Tesis Doctorales TESEO* [en línea] Disponible en Web: http://www.mcu.es/TESEO/teseo.html

Milgram, S. (1967). The small-world problem. *Psychology Today*, No. 2, pp. 60-67.

Newman, M.E.J. (2001). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, Vol. 64, No. 1 II, pp. 016131/1-016131/8.

Newman, M.E.J., Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, Vol. 69, No. 2 2, pp. 026113-1.

Newman, M.E.J. (2004). Analysis of weighted networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, Vol. 70, No. 5 2, pp. 056131-1-056131-9.

Watts, D.J., Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature* No. 393, pp. 440-442.