

# Identificación de comunidades analizando el uso del correo electrónico

Por Alex Arenas y Albert Díaz-Guilera

**Resumen:** Durante la última década hemos asistido al uso generalizado del correo electrónico como herramienta de comunicación en nuestra sociedad. Su utilización dentro de las organizaciones no escapa a esa tendencia y buena parte del flujo de información interno de una compañía se realiza de esta forma. La monitorización de su uso, preservando el anonimato de sus usuarios, se convierte en una herramienta muy valiosa para conocer la estructura informal de la organización y para compararla con la estructura formal. En particular presentamos en este trabajo el análisis de comunidades que se deduce de la red de correo de la Universitat Rovira i Virgili de Tarragona, España.

**Palabras clave:** Correo electrónico, Análisis de comunicaciones electrónicas, Redes complejas, Comunidades, Flujo de información, Optimización.

**Title: Identification of communities through e-mail usage analysis**

**Abstract:** During the last decade we have witnessed the generalized use of email as a tool for communication in our society. Its use within organizations follows this general trend and a large portion of a company's internal information flow occurs in this way. Thus, monitoring its use, while preserving the anonymity of its users, is a very valuable tool to know the informal structure of an organization and to compare it with the prescribed formal structure. In particular, we present a community analysis performed on the email network of the Universitat Rovira i Virgili, Tarragona, Spain.

**Keywords:** Email, Electronic Communications Analysis, Complex networks, Communities, Information flow, Optimization.

**Arenas, Alex; Díaz-Guilera, Albert.** El correo electrónico y la formación de comunidades en organizaciones. *El profesional de la información*, 2009, enero-febrero, v. 18, n. 1, pp. 27-33.

DOI: 10.3145/epi.2009.ene.04



**Alex Arenas**, profesor titular de ciencias de la computación e inteligencia artificial, es doctor en físicas por la Universitat de Barcelona en 1997. Su interés por las redes complejas surge a partir de la lectura de Watts y Strogatz, y su posible aplicación en múltiples áreas científicas. Investiga las implicaciones de las estructuras modulares de las redes sobre su funcionalidad. Co-lidera el Grupo de Investigación Physcomp2, de estudio físico y computacional de sistemas complejos.



**Albert Díaz-Guilera**, profesor titular de física de la materia condensada, obtuvo su doctorado en la Universitat Autònoma de Barcelona en el año 1987. Experto en física estadística, durante los últimos años se ha especializado en el análisis de redes complejas, tanto los aspectos estáticos como dinámicos. Co-lidera el Grupo de Investigación Physcomp2, centrado en el estudio físico y computacional de los sistemas complejos.

## 1. Introducción

Desde disciplinas tan dispares como la biología, la física, la informática, la economía o las ciencias sociales, en los últimos años hemos sido testigos de la aparición de sistemas con formas topológicas de un tipo intermedio entre las llamadas redes perfectamente ordenadas y las redes completamente aleatorias, y entre cuyos elementos se producen complejos fenómenos de interacción. Presentan propiedades que no encajan en ninguna de tales topologías extremas por separado, así como otras completamente nuevas que emergen de su propio grado de complejidad (Watts, 1999).

En cada disciplina, y dependiendo también de la fenomenología estudiada, los nodos y sus enlaces tienen naturalezas diferentes: en una red trófica los nodos son las especies animales y los enlaces corresponden a pares predador-presa; internet es una red de ordenadores conectados mediante conexiones físicas; la *World Wide Web* es un entramado de documentos relacionados entre sí mediante hiperenlaces; y las relaciones de parentesco y amistad configuran determinadas redes sociales.

Las redes de ordenadores han favorecido la aparición de nuevas redes sociales<sup>1</sup>, incrementando la comunicación entre amigos y familiares y creando comuni-

dades virtuales de personas que de otra manera nunca se hubieran conocido (Wellman, 2001). Pero a la hora de investigar las propiedades de estas complejas redes necesitamos herramientas para monitorizar su funcionamiento y saber si la comunicación fluye en la dirección y con la intensidad correctas, y así saber diseñar redes mejores. La construcción de modelos matemáticos nos ayuda a establecer algunas predicciones sobre su comportamiento (Guimerá et al, 2002).

En este trabajo presentamos una método para el estudio de redes en general, pero enfocando el resultado a la utilización de redes de correo electrónico en una organización<sup>2</sup>. Aunque el estudio inicial ha sido realizado en una universidad, son inmediatas las aplicaciones prácticas de interés en cualquier tipo de organización para detectar la estructura organizativa informal. Así, el método se ha usado para estudiar áreas tan diversas como la colaboración en grupos de jazz durante la primera mitad del siglo XX, la colaboración científica en física estadística en España, o en un área más general y compleja como son los artículos sobre física y matemáticas (Arenas et al, 2004).

---

**“En cualquier tipo de red social existen comunidades o grupos de individuos que tienen una mayor cohesión entre sí”**

---

El organigrama formal de una empresa describe la forma en que se pretende que los empleados interactúen. Por ejemplo, en la figura 1 hemos representado la estructura jerárquica de la *Universitat Rovira i Virgili (URV)*, cuya estructura informal estudiaremos en este artículo. Sin embargo, el día a día del trabajo y otros tipos de interacciones sociales favorecen la aparición de nuevos enlaces que van configurando la estructura informal de la empresa, aquella a través de la cual fluye en realidad la información en la organización. Conocerla es clave para la correcta administración de los recursos humanos (Krackhardt; Hanson, 1993). Tradicionalmente, el estudio de las estructuras informales se había llevado a cabo mediante cuestionarios realizados a los empleados, pero este método es casi impracticable por su elevado coste económico y/o temporal, incluso para organizaciones medianas, además de adolecer de falta de objetividad en las respuestas.

Las comunicaciones electrónicas ofrecen una alternativa real y económica al estudio de las redes informales y el intercambio de mensajes electrónicos entre los empleados es un valioso indicador de cómo se comunican dentro de la organización.

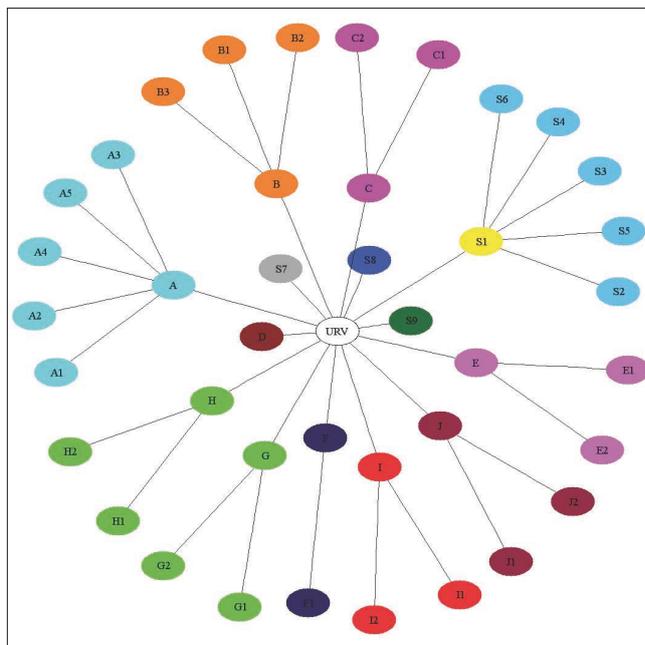


Figura 1. Estructura jerárquica formal de la Universitat Rovira i Virgili. Los diferentes colores corresponden a centros y/o servicios de la universidad.

## 2. Red de correo de la URV

Cada vez que entre los miembros de una comunidad se manda un mensaje quedan registrados en los servidores de correo tanto el remitente como el destinatario, por lo que leyendo los registros se sabe cómo ha fluido la información dentro de la misma. La red de correo electrónico se representa como un conjunto de nodos (usuarios) entre los cuales hay un enlace si han intercambiado un mensaje (intercambiado porque consideramos que el flujo de información es bidireccional entre dos usuarios) dentro del período considerado.

Hemos eliminado los envíos de correo masivo (mensajes a más de 50 destinatarios) o de listas de correo. Aunque esta restricción pueda parecer arbitraria, nuestro criterio es que la probabilidad de que tales mensajes contengan información sustancial para nuestro propósito es bastante baja.

En la URV hay unos 1.700 usuarios, entre profesores, técnicos, personal administrativo, y estudiantes de máster y doctorado. Estudiamos el flujo de correo durante los tres primeros meses del año 2002, que totalizó 1.135.818 mensajes. Construyendo la red de correo tal y como explicamos en el párrafo anterior, la imagen visual es la que se reproduce en la figura 2.

La estructura de una red social se puede caracterizar cuantitativamente a diferentes niveles:

a) Especificando el papel que juega cada nodo dentro de un contexto global, como por ejemplo el número de nodos al que está conectado o la distancia media al resto (Wasserman et al, 1994). La distribución de los

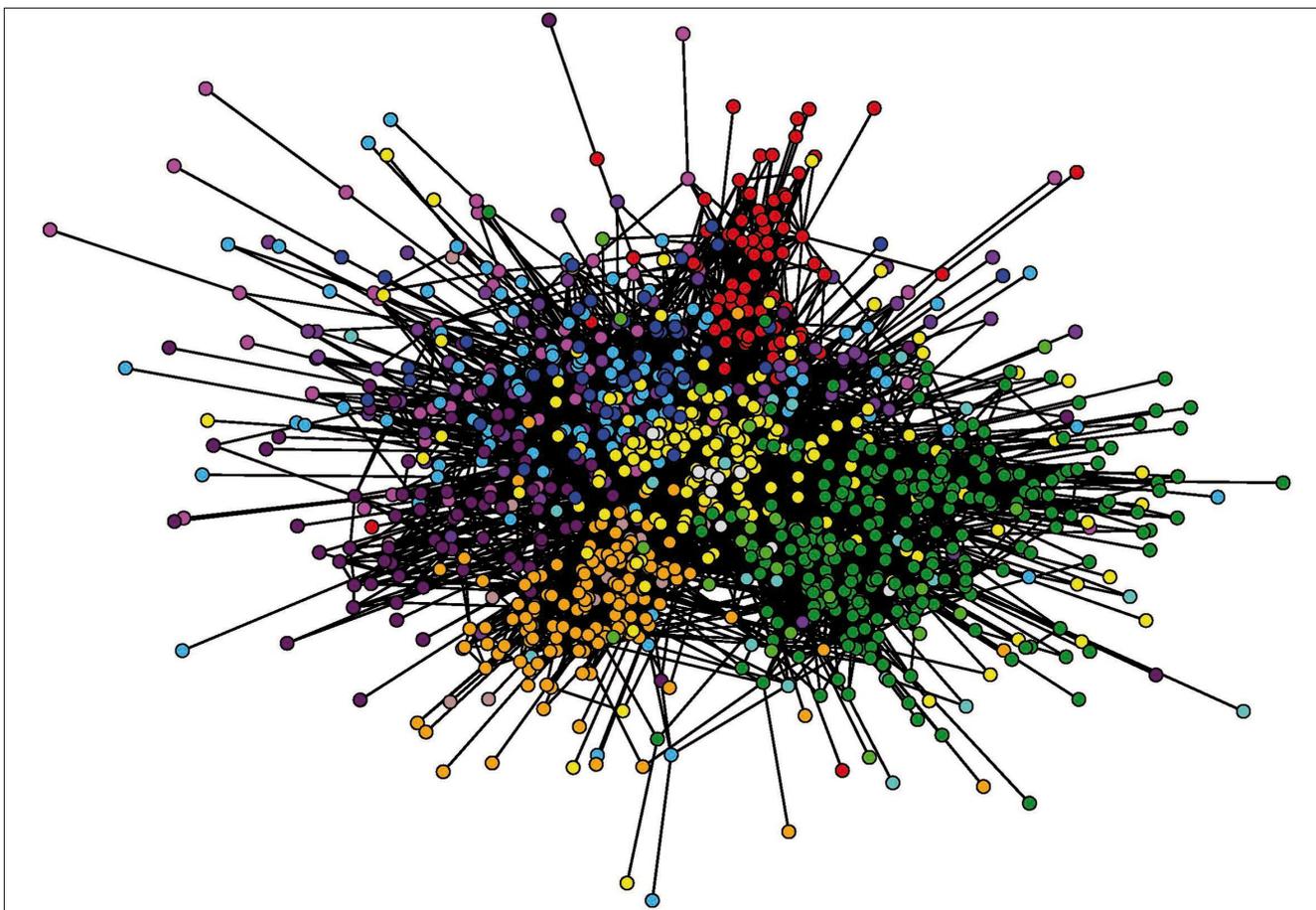


Figura 2. Red de correo electrónico de la URV. Cada nodo representa un agente de la universidad (profesor, personal administrativo, técnico, estudiante de doctorado...). Los colores corresponden a las afiliaciones a los centros o servicios.

valores numéricos de estas características individuales proporciona una imagen global detallada de la red.

Las distribuciones pueden a su vez resumirse en valores medios; en el caso de la red de correo hemos calculado el valor medio de la conectividad y de la distancia entre nodos, entre otros (Guimerá et al, 2006).

Las características observadas en la red son las que la hacen entrar dentro del tipo que se conoce como *small world* (mundos pequeños) (Watts, 1998), donde las comunidades tienen un papel relevante.

b) A un nivel intermedio, se estudia cómo están comunicados los diferentes centros (o unidades del proceso productivo en una empresa). En nuestro caso medimos la distancia entre centros para conocer la centralidad de las unidades, llegando a resultados esperados, aunque sin descartar alguno sorprendente (Guimerá et al, 2006).

c) Otra medida fue la proporción de mensajes intercambiados con otras unidades. En general esta medida detecta problemas de comunicación en las unidades, pues no es deseable que la comunicación de una unidad sea únicamente interna ni que sea únicamente externa. Mediante gráficos, que si se desea se actualizan constantemente, estas dos medidas estadísticas proporcio-

nan un seguimiento exhaustivo para explorar eventuales disfunciones en los canales de comunicación dentro y/o entre las unidades (Guimerá et al, 2006).

### 3. Comunidades en la red de correo

En cualquier tipo de red social hay comunidades, entendiéndolas como grupos de individuos que tienen una mayor cohesión entre sí y que en el caso del correo electrónico corresponde a aquellos usuarios entre los cuales hay mayor flujo de información. En una universidad esperaríamos identificar como comunidades a los diferentes niveles organizativos en los cuales está estructurada formalmente: grupos de investigación, departamentos, facultades o escuelas... Pero en cualquier organización humana hay vínculos entre individuos con orígenes muy diferentes dando lugar a comunidades informales que se superponen a las formales, por lo que entender éstas es capital para una organización.

---

**“Proponemos un método de identificación de comunidades que convierte la red original en un árbol binario”**

---

Aunque ya existían diferentes métodos para la identificación de comunidades<sup>3</sup>, dos físicos norteamericanos, **M. Girvan** y **M. E. J. Newman**, propusieron un método alternativo que ofrecía mejores resultados, basado en un concepto ya conocido en el mundo de las redes sociales que en inglés recibe el nombre de *betweenness* (**Girvan; Newman, 2002**), traducido al español por *intermediación*. Ésta es una medida de lo central que es un enlace con respecto al número de caminos mínimos entre nodos que pasan por el enlace. Dicho de otro modo, como leemos en la Wikipedia: “es una medida de la centralidad que indica la frecuencia con la que un nodo aparece en el camino más corto que conecta otros dos nodos. Un actor que se encuentra en el camino que une dos puntos de la red posee una posición de intermediario, de controlador del flujo de información. O sea, la distancia a otras unidades (cercanía) no es la única medida de centralidad importante. También lo es saber cuántos caminos de pequeña distancia corren a través de un par de nodos, ya que poseen el control del flujo de los demás”.

Cortando sucesivamente los enlaces de mayor intermediación podemos ir separando las diferentes comunidades. Aplicando el proceso de forma reiterativa, como se explica en la figura 3, vamos representando las comunidades en una estructura en forma de árbol binario. El resultado de aplicar dicho proceso a la red de correo electrónico de la URV se puede ver en la figura 4. Cada rama corresponde a una comunidad real y los individuos más centrales de cada comunidad corresponden a los extremos de las ramas. Los códigos de los colores son los mismos que los de la figura 2.

Hemos convertido la difícil labor de identificar comunidades visualmente en la red original a un sencillo proceso en la red transformada. En la figura podemos apreciar cómo se conserva la estructura formal de la

universidad por los colores bastante homogéneos de las diferentes ramas, aunque también aparece un importante número de elementos de otros centros dentro de ellas, debido a que se han creado comunidades informales. En la imagen, además de observar las afiliaciones podemos ver la estructura de las diferentes comunidades informales tal y como comentábamos al principio.

Uno de los aspectos importantes de este procedimiento es que los nodos que tienen un papel central en la comunidad aparecen al final de las ramas porque son los últimos enlaces en romperse, tal y como se podía ver en la figura 3. Por otro lado cerca de las ramificaciones encontraríamos aquellos nodos que actúan de enlace entre las diferentes comunidades.

Para estar completamente seguros de que el resultado no es una consecuencia artificial del procedimiento hemos efectuado el mismo tipo de estudio en una red con la misma distribución de conectividades que nuestra red de correo electrónico, pero sin estructura de comunidades. El resultado se muestra en la figura 5.

La gran ventaja de la investigación en temas multidisciplinares son las inestimables analogías entre sistemas que en principio no tienen ninguna relación. En nuestro ejemplo hemos hallado unas analogías muy interesantes con las distribuciones de cuencas fluviales lo cual nos lleva a conjeturar hipótesis realmente significativas. Así la distribución de tamaños de cuen-

**“Se aprecian claramente las comunidades organizativas entre las que se intercambian mensajes dentro de la organización”**

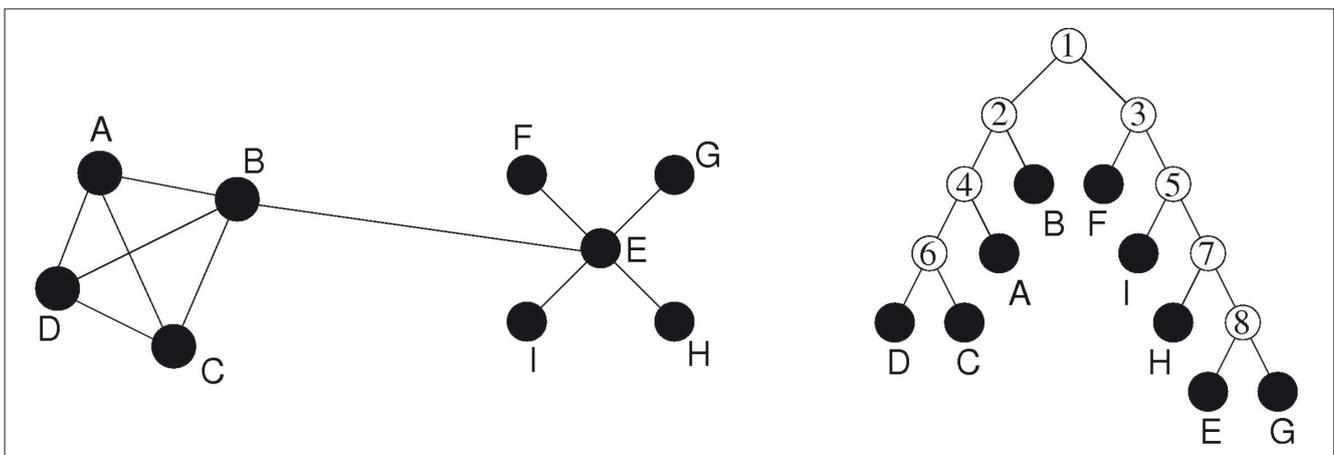


Figura 3. Identificación de comunidades. (a) Una red con dos comunidades perfectamente identificables. El enlace BE es el que divide a las dos comunidades, y es el que tiene una “betweenness” mayor ya que para ir de una comunidad a otra hay que pasar por dicho enlace. Por lo tanto, éste será el primer enlace que romperemos. (b) Árbol binario que representa la división en comunidades. La comunidad global 1 se ha dividido en dos comunidades (2 y 3). Aunque el proceso se realiza de forma iterativa, los nodos después de la primera subdivisión juegan un papel similar; solamente el nodo E es más central en la comunidad 3, y por este motivo aparece al final de la ramificación.

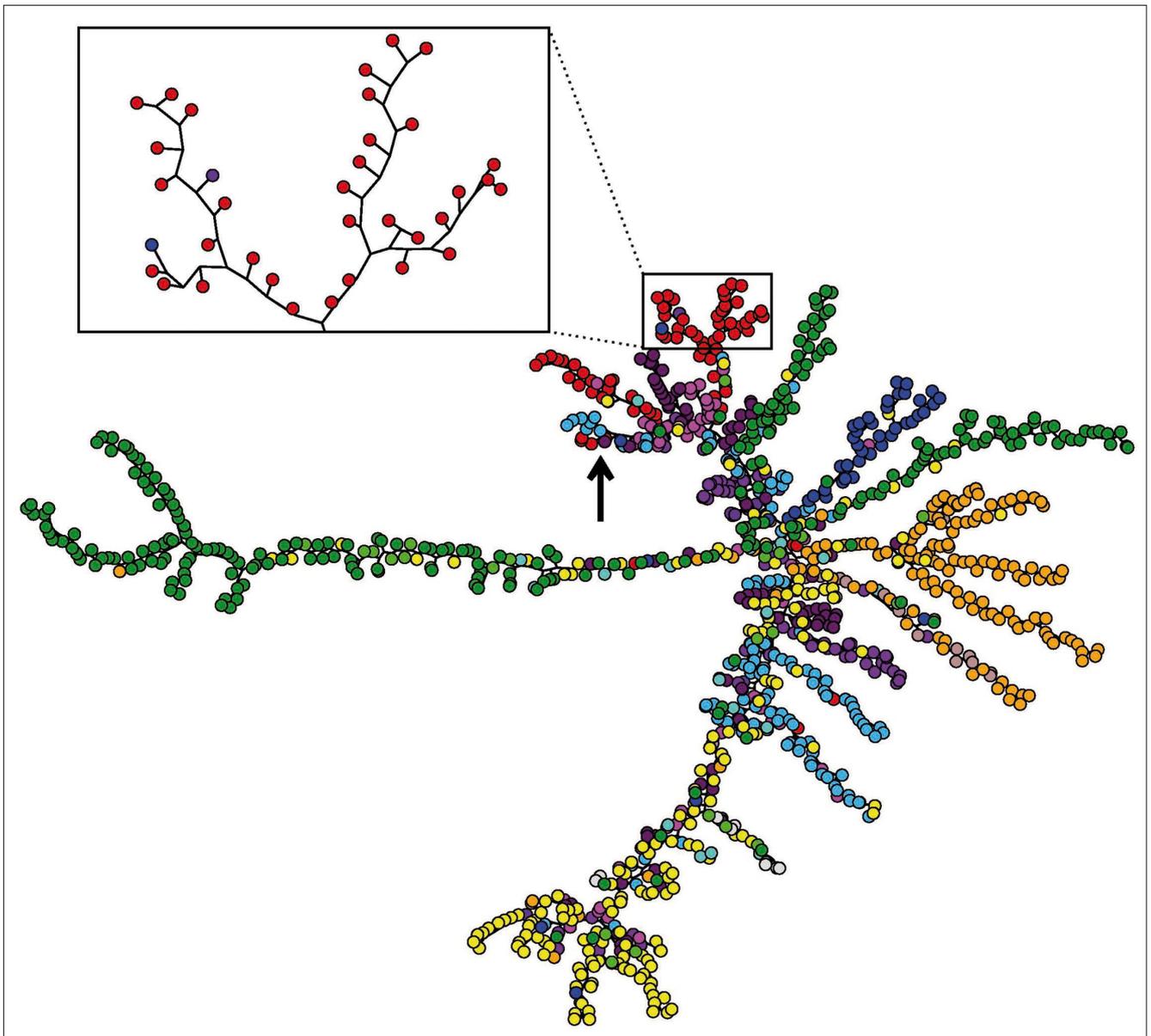


Figura 4: Representación de la red de correo electrónico de la figura 2 mediante un árbol binario de comunidades. Cada rama corresponde a una comunidad real y los individuos más centrales de cada comunidad son los de los extremos de las ramas. Los códigos de los colores son los mismos que los de la figura 2.

cas fluviales de determinados ríos tiene una estructura idéntica a la que nosotros obtenemos de la distribución de tamaños de comunidades en la red de correo electrónico en la URV.

La estructura de las cuencas fluviales está relacionada con la óptima distribución acuífera (Danon et al, 2005). Eso nos brinda especular con el hecho de que las estructuras informales que emergen de la propia dinámica de una organización llevan a un proceso de optimización en este caso del flujo de información. Pero esa analogía con los ríos no se acaba ahí sino que además nos permite, utilizando unos índices también definidos en el estudio de cuencas fluviales, determinar la estructura organizativa de nuestro colectivo y confirmar la auto-similaridad a todas las escalas propias de sistemas que aparecen en la naturaleza, como por ejemplo los fractales<sup>4</sup>. Siguiendo nuestra metodología

hemos estudiado la auto-similaridad de diferentes redes tanto naturales como artificiales, llevándonos al resultado de que la red de correo electrónico es de las redes con un mayor grado de auto-similaridad.

#### 4. Conclusiones

En este trabajo hemos presentado por primera vez un procedimiento mucho más rápido y económico que los tradicionales, para extraer de los registros del correo electrónico en una organización información valiosa sobre las estructuras informales. Aunque el análisis visual de la red original no ofrece ninguna información relevante, proponemos un método de identificación de comunidades que convierte la red original en un árbol binario, donde se pueden apreciar de manera muy clara las comunidades organizativas a las que da lugar el intercambio de mensajes dentro de la organización, en el

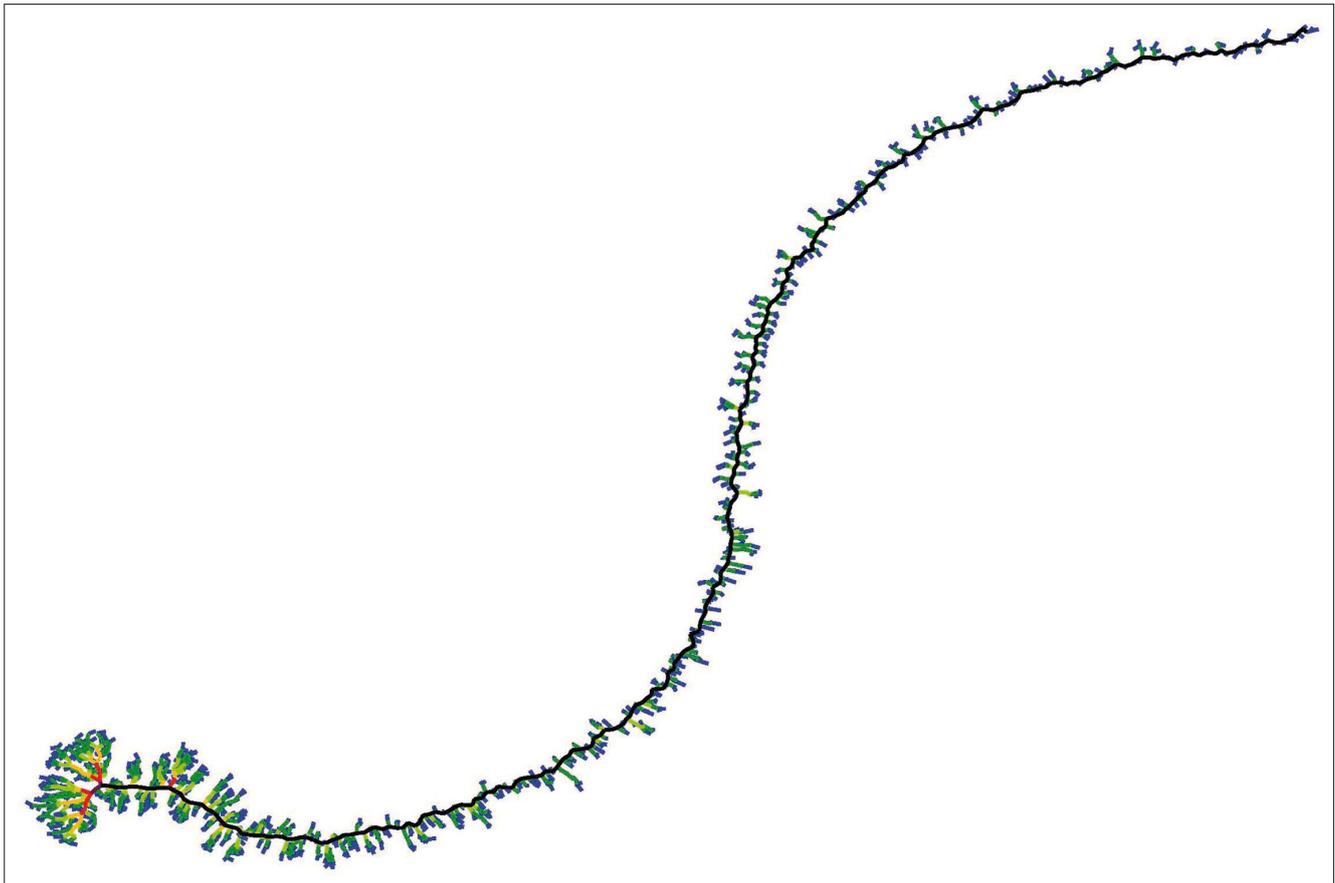


Figura 5. Árbol binario obtenido como resultado de aplicar el método de identificación de comunidades a una red aleatoria sin estructura de comunidades.

caso particular que hemos estudiado, de la *Universitat Rovira i Virgili*.

---

### “Con la información obtenida se pueden reorganizar las estructuras que no funcionen óptimamente”

---

Se identifican todos los niveles del organigrama formal: grupos de investigación, departamentos, facultades o escuelas y la propia universidad, así como las relaciones entre las diferentes comunidades formales para dar lugar a la estructura informal. Creemos que esta información es de gran utilidad para reorganizar estructuras que no funcionen correctamente o hacer seguimientos dinámicos de la evolución de cambios estructurales.

Desde una perspectiva teórica, la analogía con las redes fluviales y su estructura auto-similar abre un nuevo campo de investigación para comprobar si este comportamiento tiene un carácter universal y cómo ha sido la propia evolución auto-organizada del sistema la que lo ha llevado a dicho estado. ¿Será posible en-

contrar alguna magnitud que esté optimizada en una organización, al igual que lo es el agua en las redes fluviales?

## 5. Notas

1. Ejemplos actuales de esta proliferación los tenemos en sitios web como *Facebook*, *LinkedIn*, o *Xing*.
2. Estudios recientes realizados en otros tipos de redes de comunicaciones las encontramos en **Leskovec, Horvitz** (*Microsoft instant messaging system*) y en **Onnela** et al. (usuarios de telefonía móvil de un operador europeo).
3. Para una discusión más técnica y comparativa de diferentes métodos de detección de comunidades puede verse **Danon** et al, (2005).
4. Un fractal es un objeto semi geométrico cuya estructura básica, fragmentada o irregular, se repite a diferentes escalas. Las nubes, las montañas, el sistema circulatorio, las líneas costeras o los copos de nieve son fractales naturales. Los fractales se definen matemáticamente mediante un simple algoritmo recursivo, poseen detalle a cualquier escala de observación y son auto-similares (adaptado de *Wikipedia*).

## 6. Bibliografía

- Arenas, A.; Danon, L.; Díaz-Guilera, A.; Gleiser, P.; Guimerá, R.** “Community analysis in social networks”. *European journal of physics B*, 2004, v. 38, pp. 373-380.
- Banavar, J.; Maritan, A.; Rinaldo, A.** “Size and form in efficient transportation networks”. *Nature*, 1999, v. 399, pp. 130.
- Barabási, A. L.** *Linked: the new science of networks*. Cambridge: Perseus Press, 2002.
- Barabási, A. L.; Bonabeau, E.** “Scale free networks”. *Scientific American*, mayo 2003.

**Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D. U.** "Complex networks: structure and dynamics". *Physics reports*, 2006, v. 424, pp. 175-308.

**Buchanan, M.** *Nexus: small worlds and the groundbreaking science of networks*. New York: Norton & Company, 2002.

**Danon, L.; Duch, J.; Díaz-Guilera, A.; Arenas, A.** "Comparing community structure identification". *Journal of statistical mechanics - theory and experiment*, 2005, P09008.

**Girvan, M.; Newman, M. E. J.** "Community structure in social and biological networks". *Proceedings of the National Academy of Sciences of the USA*, 2002, v. 99, pp. 7821-7826.

**Guimerá, R.; Danon, L.; Díaz-Guilera, A.; Giral, F.; Arenas, A.** "The real communication network behind the formal chart: community structure in organizations". *Journal of economic behavior and organization*, 2006, v. 61, pp. 653-667.

**Guimerá, R.; Díaz-Guilera, A.; Vega-Redondo, F.; Cabrales, A.; Arenas, A.** "Optimal network topologies for local search with congestion". *Physical review letters*, 2002, v. 89, id. 248701.

**Krackhardt, D.; Hanson, J. R.** "Informal networks: the company behind the chart". *Harvard business review*, 1993, v. 71, pp. 104.

**Leskovec, J.; Horvitz, E.** *Planetary-scale views on an instant-messaging network*, arXiv:0803.0939v1  
<http://arxiv.org/abs/0803.0939v1>

**Onnela, J. P.; Saramäki, J.; Hyvönen, J.; Szabó, G.; Lazer, D.; Kaski, K.; Kertész, J.; Barabási, A.-L.** "Structure and tie strengths in mobile communication networks". En: *Proc. Natl. Acad. Sci. USA*, 2007, v. 104, pp. 7332-7336.

**Rodríguez-Iturbe, I.; Rinaldo, A.** *Fractal river basins: chance and self-organization*. Cambridge: Cambridge University Press, 1996.

**Wasserman, S.; Faust, K.; Iacobucci, D.** *Social network analysis: methods and applications*. Cambridge: Cambridge University Press, 1994.

**Watts, D. J.** *Small worlds: the dynamics of networks between order and randomness*. Princeton: Princeton University Press, 1999.

**Watts, D. J.; Strogatz, S. H.** "Collective dynamics of 'small-world' networks". *Nature*, 1998, v. 393, pp. 440-441.

**Watts, D. J.** *Six degrees: the science of a connected age*. New York: Norton & Company, 2003.

**Wellman, B.** "Computer networks as social networks". *Science*, 2001, v. 293, pp. 2031-2034.

**Alex Arenas.** *Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Avda. Països Catalans 26, 43007 Tarragona.*

<http://deim.urv.cat/~aarenas/>  
<http://sites.google.com/site/physcomp2>

**Albert Díaz-Guilera.** *Departament de Física Fonamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona.*

<http://albert.diaz.guilera.googlepages.com/>  
<http://sites.google.com/site/physcomp2>

## CALSI 2009 VIII WORKSHOP CALSI

**CALSI**

9-10 de marzo de 2009  
Universidad Politécnica de Valencia

### “La Sociedad del Conocimiento: Innovación y Nuevos Retos Globales”

Tras el éxito de la séptima edición ya estamos organizando Calsi 2009, un Workshop pionero en el tratamiento de los E-contents: Contenidos y Aspectos Legales en la Sociedad de la Información. En esta nueva edición se reunirán de nuevo los mejores profesionales de la gestión de la información y de los sectores clave para el desarrollo de la Sociedad de la Información y del Conocimiento.

#### BLOQUES TEMÁTICOS

Televisión Digital  
Records management  
Innovación  
Negocios en red

#### DESTINATARIOS

El congreso está dirigido a profesionales que trabajan en el ámbito de las nuevas tecnologías e Internet.

Esta octava edición cuenta además con dos factores externos que aumentan su interés. Por un lado, se celebrará los días previos al Congreso ISKO España, que también se realizará en Valencia. Por otro, su celebración coincide con el comienzo de las Fallas, por lo que a la visita al congreso se le añade la posibilidad de conocer una de las fiestas de mayor prestigio y reconocimiento internacional.

#### PONENTES

Profesionales que trabajan en el ámbito de las nuevas tecnologías e Internet que ofrecerán desde un punto de vista cualificado propuestas y perspectivas para el desarrollo de la sociedad de la información.

El encuentro tendrá lugar los días 9 y 10 de Marzo de 2009, en el salón de actos del edificio Nexus (CFP) de la Universidad Politécnica de Valencia.

*insíbete ya!*

Camino de Vera s/n, 46022 Valencia Tel.: 963 87 70 00 (Ext. 88924) Mail: [info@calsi.org](mailto:info@calsi.org)

Organiza



Patrocina



Financian

