



ANÁLISIS DE UNA COMUNIDAD ON-LINE A PARTIR DE SU LISTA DE DISCUSIÓN

El caso de Edutec-L

Robert Ralló

robert.rallo@urv.cat

*Escola Tècnica Superior d'Enginyeria
Universitat Rovira i Virgili (Tarragona)*

Mercè Gisbert

merce.gisbert@urv.cat

*Facultat d'Educació i Psicologia
Universitat Rovira i Virgili (Tarragona)*

RESUMEN.

En este artículo se presenta una metodología para el análisis y caracterización de comunidades online a partir de los mensajes intercambiados en una lista de discusión. Mediante técnicas de análisis basadas en redes sociales se detecta la estructura, relaciones y el papel que juegan los diferentes miembros de la comunidad. La metodología propuesta se evalúa usando los mensajes procesados por la lista de discusión EDUTECH-L durante el periodo 1998-2004.

PALABRAS CLAVE: análisis de redes sociales, lista de discusión, comunidades on-line

ABSTRACT.

The current work introduces a methodology for the analysis and characterization of on-line communities through the messages exchanged in a discussion list. The structure, relationships and user roles in the community are inferred using social network analysis techniques. The proposed methodology is assessed using the messages processed by the EDUTECH-L discussion list during the period 1998-2004.

KEY WORD: social network analysis, discussion list, on-line communities



1. Introducción

En el ámbito de las Tecnologías de la Información y las Comunicaciones (TIC) la herramienta de comunicación más extendida y usada es, sin lugar a dudas, el correo electrónico. Este medio de comunicación asíncrono ha cambiado de manera radical nuestra forma de relacionarnos e intercambiar información. A pesar de ser una tecnología “antigua”, el estudio de su influencia en nuestra estructura social, patrones de comunicación y estilos de aprendizaje sigue siendo actualmente un campo de investigación muy activo y que aún plantea un gran número de interrogantes.

Cualquier proceso de comunicación e interacción sostenido durante un cierto periodo conlleva la emergencia de un sentimiento (o más formalmente, una estructura) de comunidad entre sus participantes. El uso de Internet como medio de transporte para el intercambio de información genera la aparición de un nuevo tipo de comunidades formadas por usuarios que con frecuencia no se conocen personalmente, son las denominadas “comunidades on-line”. El correo electrónico y especialmente las listas de discusión actúan como catalizadores que fomentan la aparición de este tipo de comunidades.

La existencia de este tipo de comunidades nos lleva a plantearnos toda una serie de interrogantes sobre su estructura, efectos y evolución. Por ejemplo, ¿Cuáles son los roles principales que adoptan los miembros de estas comunidades? ¿Es posible detectar la presencia de un “líder”? ¿Cómo valoramos la “reputación” de los miembros de la comunidad? ¿Emergen sub-comunidades?

Este tipo de preguntas son las que se abordan en el presente estudio desde un punto de vista multidisciplinar en el que se combinan elementos de la informática y las ciencias de la computación con otros provenientes del campo de la sociología y la educación.

2. Comunicación mediada por ordenador y Comunidades Virtuales.

El correo electrónico fue la primera herramienta de comunicación en Internet y en gran parte responsable de su posterior desarrollo y expansión. Para muchos de los usuarios de Internet el intercambio de información a través del correo electrónico fue - y todavía sigue siendo- el motivo que les empuja a usar la red (Bälter, 1998; Moody, 2002). Directamente ligado al crecimiento y popularización de la red, cobra cada vez mayor importancia el concepto de Comunicación Mediada por Ordenador (CMC), siendo el correo electrónico uno de sus pilares básicos (Feldman, 1987). Galesic y Stepanic (2003) llegan incluso más lejos e introducen el concepto de “comunicación mediada a través del correo electrónico” (EMC, *e-mail mediated communication*) para



referirse a este tipo concreto de CMC. El uso masivo del correo electrónico conlleva dos hechos importantes: (i) el aumento exponencial de la información “no estructurada” e “informal” contenida en los mensajes, y (ii) la creación de estrechos vínculos de comunicación, colaboración e intercambio de información entre los usuarios.

El uso de este canal de comunicación en organizaciones y empresas facilita el acceso y la difusión de la información agilizando los procesos de gestión. La cantidad de información que se transmite cada día mediante el correo electrónico es inmensa pero a medida que aumenta el tamaño de la organización su gestión se vuelve altamente compleja (Argyris, 1977; Davenport, 1994). Un tipo específico de mensajes de correo electrónico que contienen información de calidad organizada coherentemente sobre un tema concreto son aquellos que se distribuyen a través de las listas de discusión. Las listas nos ofrecen una política organizada y racional para la distribución de información a través del correo electrónico en el seno de una organización o comunidad. Su funcionamiento se basa en que un grupo de usuarios mediante un proceso de suscripción reciben automáticamente copias de todos los mensajes enviados a la lista. Este tipo de uso hace que el correo electrónico se convierta en un sistema de comunicación uno a muchos (*broadcast*) permitiendo la rápida difusión de la información dentro del grupo de suscriptores de la lista.

Cuando se produce una interacción cara-a-cara de manera continuada entre un grupo de personas surge de manera espontánea un cierto sentimiento de pertenencia a una comunidad que comparte los mismos intereses. Es decir, el establecimiento de un proceso de comunicación continuado entre un grupo de personas genera vínculos sociales que dan lugar a una estructura de comunidad. Dado que la CMC es un caso concreto de proceso de comunicación es lógico pensar que podemos encontrar este mismo fenómeno en la red. A este tipo de comunidades se las denomina comunidades on-line o comunidades virtuales (Barlow et al., 1995; Wellman, Gulia, 1997). Al valor añadido que ofrece la aparición de este tipo de comunidades se le denomina capital social y con frecuencia en organizaciones y empresas puede llegar incluso a relacionarse con los beneficios económicos conseguidos (Kimball, Rheingold, 2000).

3. Las redes sociales

Los procesos de comunicación e intercambio de información fomentan el establecimiento de relaciones sociales. La sociología se ha estudiado este fenómeno desde distintos puntos de vista, tanto el análisis cualitativo puramente basado en la observación y descripción de las interacciones observadas como aproximaciones de carácter más cuantitativo basadas en la recogida de información mediante cuestionarios y su posterior análisis estadístico. Una de las técnicas más habituales a la



hora de estudiar estas relaciones son los sociogramas (Gronlund, 1985). Los sociogramas pueden formalizarse de diversas maneras, siendo los grafos una de las técnicas de representación más usadas. Podemos modelar un conjunto de relaciones entre diversos actores mediante un grafo en el que cada persona está representada por un nodo y las relaciones interpersonales se representan mediante aristas. La estructura resultante se denomina “red social” (Albretch, 1987).

El análisis de las redes sociales permite desarrollar modelos tanto cualitativos como cuantitativos de las relaciones entre los miembros de la comunidad (Wasserman, Faust, 1994; Grannovetter, 1973). Estas técnicas se han aplicado en diversos ámbitos, desde la sociología a la organización empresarial (Ahuja, Carley, 1998), pasando por campos tan dispares como la epidemiología (Morris, Kretzschmar, 1997). Recientemente otros campos más alejados del terreno de las ciencias humanas han adoptado también este tipo de técnicas: la física (teoría de sistemas complejos), la biología (estructura de proteínas) e incluso la informática (análisis de redes informáticas). Las técnicas de análisis de redes sociales nos proporcionan las herramientas adecuadas para identificar los elementos más importantes de la red, caracterizar su estructura y propiedades, y modelar los flujos de información.

Las comunidades on-line generan un tipo especial de redes sociales denominadas *Computer Supported Social Networks* (CSSN). Las CSSN combinan los conceptos propios del ámbito de las redes sociales con las posibilidades que ofrecen los nuevos medios como Internet en el terreno de la información y la comunicación. Las CSSN proporcionan una perspectiva para el estudio de la estructura y la dinámica de las comunidades virtuales mediante la aplicación de técnicas desarrolladas para el análisis de redes sociales.

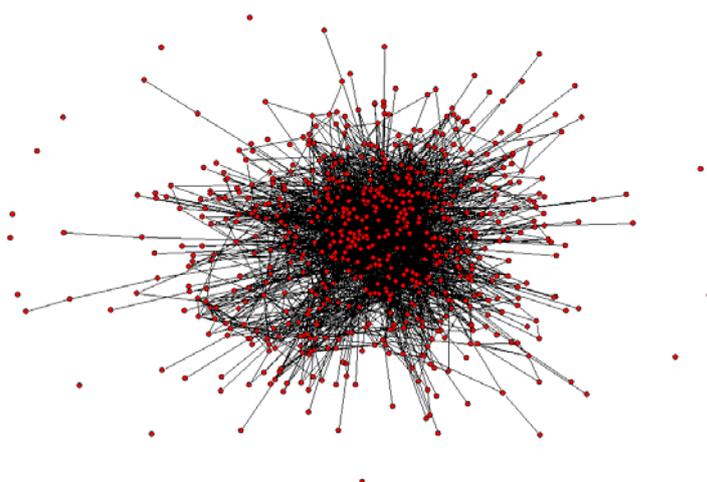
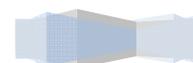


Figura 1. Red social correspondiente a la co-participación en lista EDUtec-L durante el periodo 1998-2004. Los nodos representan a usuarios de la lista (n=876) que han participado en alguno de los hilos de discusión no triviales (num. mensajes > 1) detectados. Dos usuarios están conectados si han participado en el mismo hilo de discusión.



La figura 1 muestra la visualización de una CSSN formada por la interacción de los usuarios suscritos a una lista de discusión. Cada uno de los nodos representa un miembro de la comunidad y los enlaces entre nodos representan las relaciones e intercambios de información que se establecen entre ellos. A primera vista podemos observar ciertas pistas visuales que nos indican que estamos ante un tipo de estructura con un alto grado de organización. Algunos nodos no están enlazados con ningún otro miembro de la comunidad, representando a los típicos observadores, que se suscriben a una lista pero que no participan directamente en las discusiones. Existen otros nodos en la periferia que poseen un único enlace, es decir, se relacionan directamente con un único miembro de la comunidad. En el centro de la estructura observamos nodos densamente conectados que actúan como elementos aglutinadores proporcionando cohesión a la estructura social y actuando como los auténticos motores de la comunidad virtual.

3.1. Conceptos básicos y propiedades

El análisis de una red social puede abordarse desde dos puntos de vista. En primer lugar podemos centrar nuestro estudio en las características globales de la red consiguiendo una visión de conjunto de las relaciones existentes entre todos los miembros de la comunidad virtual. El análisis global permite la obtención de macroindicadores sobre la estructura y relaciones sociales inducidas por los flujos de comunicación dentro de la comunidad. Un punto de vista alternativo se basa en el estudio de las denominadas redes ego-céntricas, focalizadas en uno de los miembros de la red social (*ego*) y en todas sus conexiones directas (*alters*). Este tipo de observación proporciona información de la red a escala local (*micro*), centrada en el individuo y su vecindad directa. Desde el punto de vista del análisis, el estudio de estas *ego-nets* nos proporciona información del efecto de la estructura de la red sobre el rol que adquieren ciertos componentes de la comunidad (estructura social local).

La *conectividad* es uno de los indicadores más utilizados para caracterizar la estructura de una red social. En la mayoría de redes asociadas a sistemas reales encontramos grupos de nodos con un elevado número de conexiones y otros con niveles de conexionado muy bajos. En sistemas sociales reales difícilmente encontraremos una red en la que se establezcan todas las relaciones posibles (*red totalmente saturada*). Las relaciones que no llegan a establecerse darán origen a los denominados *agujeros estructurales* que juegan un papel fundamental en la caracterización de la estructura de una comunidad virtual. Entender como afecta el perfil de conectividad de cada individuo a su papel en la red será fundamental para interpretar correctamente la estructura social. Los nodos con un alto nivel de conexiones corresponden a individuos “influyentes” en la red aunque simultáneamente al recibir información desde diversas fuentes son fácilmente “influciables” por otros nodos. En una red con un nivel alto de conexionado la información fluye rápidamente, por ejemplo los rumores se extienden fácilmente por toda la comunidad o una epidemia o virus informático se



propagaría a gran velocidad. El nivel de conexionado de la estructura desempeña un papel determinante en su dinámica y características funcionales. Las medidas principales para caracterizar la conectividad de una red son:

- *Tamaño*. Número de nodos que forman la red. Es un elemento crítico que determina la tipología de las relaciones sociales que se establecen en la red. Cada nodo (individuo) posee una capacidad limitada para establecer/mantener relaciones. Al aumentar el tamaño de la comunidad la proporción entre las relaciones establecidas por un individuo y el límite máximo de relaciones que podría establecer decrece, fomentando la aparición de subestructuras dentro de la comunidad.
- *Densidad*. Promedio de la proporción de relaciones que establece un nodo respecto al máximo que podría mantener.
- *Grado*. Número de conexiones de un nodo. Si el grafo es dirigido (tenemos en cuenta la dirección de la interacción) distinguiremos entre el grado de salida y el grado de entrada (conexiones entrantes). El grado promedio proporciona un indicador global de la conectividad de la red. El grado de salida de un nodo se asocia a una medida de su nivel de influencia en la comunidad. El grado de entrada es considerado por algunos autores como un indicador de prestigio en la comunidad (el resto de nodos quieren que les conozca y le envíen información). Los nodos que tienen tendencia a recibir mucha información suelen calificarse como nodos “poderosos” (“...información es poder”). Aún así, un grado extremadamente alto puede ser contraproducente. Si el grado de un nodo es mayor que el que es capaz de asimilar nos encontraremos ante una situación de sobrecarga de información. La recepción de información contradictoria provoca situaciones de ruido e interferencia en los nodos. Un nodo cuyo grado de salida es alto y su grado de entrada bajo se denomina *nodo fuente*, dado que básicamente se comporta como un *emisor* de información. Por el contrario, cuando el grado de entrada es mayor que el de salida estamos ante un *nodo pozo* (o *receptor*) de información. Los nodos equilibrados en los que la proporción de conexiones de entrada es del mismo orden a las de salida se denominan *transmisores*.
- *Distancia*. Número promedio de enlaces que hay que atravesar para llegar desde un nodo origen a otro destino (dos nodos adyacentes estarán situados a distancia 1). Las distancias entre nodos son una macro-característica importante en las redes. Si las distancias son grandes el tiempo de propagación de la información a través de la red también lo será.
- *Distancia geodésica*. Número de enlaces que hay que atravesar para llegar desde un nodo origen hasta un nodo destino siguiendo el camino más corto posible. La distancia geodésica indica el camino más eficiente para transmitir información entre nodos de la red. Dado un nodo, podemos definir su *excentricidad* como la distancia geodésica mayor entre todas sus conexiones.



- *Diámetro*. La distancia geodésica mayor en un grafo conexo. Esta medida nos da una idea global del tamaño de la red dado que nos indica el mayor grado de separación posible entre dos nodos de la red.

Una característica relevante en el estudio de la dinámica de una estructura social es su potencia o importancia. Cada uno de los componentes de la estructura contribuye de una forma concreta al comportamiento global de la red en función de sus conexiones y su posición en la jerarquía social. La potencia de una red puede estudiarse a nivel de toda la red o en base a la relación entre pares de nodos. Esta propiedad está relacionada con la *centralidad* que se define como la importancia relativa al conexasiónado o posición de un nodo respecto al flujo de información en la red. Una manera de medir la centralidad es, por ejemplo, a través del grado de un nodo. Los nodos que tienen un grado mayor tienen más oportunidades para acceder a la información que otros nodos con menos conexiones. Un nodo con un grado elevado tiene la posibilidad de poder elegir entre sus fuentes de información o de poder usar diversos caminos para propagar sus ideas en la estructura social. Otras medidas de centralidad son:

- *Cercanía (closeness)*: (Bavelas, 1950; Harare, Norman et al., 1965; Sabidussi, 1966) es un indicador de la cercanía de un nodo al resto de nodos de la red.
- *Betweenness*: (Freeman, 1997; Anthonisse, 1971) es un indicador de la importancia estratégica de un nodo relativa a los caminos por los que fluye la información en la red social. Se calcula para un nodo en base a la suma de todos los caminos mínimos que unen a cada par de nodos en la red.

La detección y caracterización de subestructuras o sub-comunidades proporciona un indicador de las características estructurales de una red social. Existen diversas técnicas (Girvan, Newman, 2002) que permiten agrupar los nodos de la red que presentan patrones de conectividad similares e identificar nodos que dividen la red en diversos componentes aislados (Kleinberg, 1999). Un ejemplo de este tipo de nodos son los puntos de corte (*cut-points*) que nos indican los elementos importantes sin los cuales la red perdería su cohesión.

4. Metodología para el análisis de listas de discusión mediante redes sociales

Las listas de distribución presentan dos características fundamentales que determinan el proceso de análisis del flujo de mensajes. En primer lugar, los mensajes no tienen un destinatario concreto sino que van dirigidos al gestor de listas. En segundo lugar, la interacción entre subgrupos de usuarios dentro de la lista manifiesta a través de lo que



se denominan “hilos de discusión”. Estos hilos representan un flujo de comunicación coherente referida a un tema concreto.

Diversos autores (Guimerà et al., 2003; Tyler et al., 2003) han propuesto el uso de la traza del flujo de mensajes de correo electrónico para generar redes sociales y estudiar la estructura de las comunidades que genera este mecanismo de comunicación. Este mismo principio es aplicable al estudio de las relaciones que se establecen en una comunidad virtual a través de su lista de discusión. Para generar una red social que describa el intercambio de información entre suscriptores de una lista será preciso un proceso de filtrado y extracción de información a partir de los mensajes de correo electrónico enviados al gestor de listas.

4.1. Procesado de los mensajes de la lista

Los mensajes de correo electrónico tienen una estructura estándar especificada en documentos denominados *Request for Comments* (RFC). En concreto el protocolo de correo electrónico está especificado en los RFC 822 y 2822. Un mensaje de correo electrónico está compuesto por dos elementos básicos: la *cabecera* y el *cuerpo*. La cabecera contiene toda la información necesaria para hacer llegar el mensaje hasta su destinatario e incorpora información sobre el remitente, la ruta seguida por el mensaje hasta llegar a su destino y el tipo de mensaje. En el cuerpo del mensaje encontramos el contenido que se transmite y la codificación de los archivos adjuntos que pueda contener el mensaje.

La figura 2 muestra un ejemplo de la estructura de campos de la cabecera de un mensaje de correo electrónico. El campo que identifica al emisor del mensaje, denominado *from*, contiene su dirección electrónica. El destinatario del mensaje (especificado en el campo *to*) es la propia lista de discusión. Finalmente, el campo que controla quien recibe las respuestas a este mensaje (etiquetado como *reply-to*) también es la propia lista de correo en lugar del usuario que emite el mensaje.

```
Subject: [EDUTECH] Móviles en la enseñanza
From: Anibal de la Torre <anibal@ADELAT.ORG>
Reply-To: Foro sobre Tecnología Educativa <EDUTECH-L@LISTSERV.REDIRIS.ES>
Date: 29/03/2005 20:48
To: EDUTECH-L@LISTSERV.REDIRIS.ES
X-Account-Key: account1
X-UIDL: 7022-1099496579
X-Mozilla-Status: 0001
X-Mozilla-Status2: 00000000
Return-path: <owner-edutech-l@LISTSERV.REDIRIS.ES>
Received: from relay.urv.es ([193.144.16.30]) by correu.urv.es (Planet Messaging Server 5.2 HotFix 1.18 (built Jul 28 2003)) with ESMTMP id <0IE50070W22MYK@correu.urv.es> for Robert.Rallo@URV.NET; Wed, 30 Mar 2005 02:
Received: from relay.urv.es (localhost [127.0.0.1]) by localhost.urv.es (Postfix) with ESMTMP id EA10C104C for <Robert.Rallo@URV.NET>; Wed, 30 Mar 2005 02:15:57 +0200 (MEST)
Received: from chico.rediris.es (chico.rediris.es [130.206.1.3]) by relay.urv.es (Postfix) with ESMTMP id C92A71043for <Robert.Rallo@URV.NET>; Wed, 30 Mar 2005 02:15:57 +0200 (MEST)
Received: from chico.rediris.es (130.206.1.3) by chico.rediris.es (LSMTP for Solaris v1.1b) with SMTP id <0.0022E5CE@chico.rediris.es>; Wed, 30 Mar 2005 02:15:57 +0200
Received: from LISTSERV.REDIRIS.ES by LISTSERV.REDIRIS.ES (LISTSERV-TCP/IP release v1.8e) with spool id 25243830 for EDUTECH-L@LISTSERV.REDIRIS.ES; Wed, 30 Mar 2005 02:15:57 +0200
Received: from relay.rediris.es (relay.rediris.es [130.206.1.53]) by chico.rediris.es (8.12.10/8.9.1) with ESMTMP id j2TImvIS000691 for <EDUTECH-L@LISTSERV.REDIRIS.ES>; Tue, 29 Mar 2005 20:48:58 +0200 (CEST)
Received: from madrid7.amenworld.com (madrid7.amenworld.com [62.193.203.23]) by relay.rediris.es (8.12.11/8.11.1) with ESMTMP id j2TImvPh022399f for <EDUTECH-L@LISTSERV.REDIRIS.ES>; Tue, 29 Mar 2005 20:
Received: from [127.0.0.1] (23.Red-83-39-75.pooles.rima-tde.net [83.39.75.23]) (authenticated) by madrid7.amenworld.com (8.10.2/8.10.2) with ESMTMP id j2TImpe21734f for <EDUTECH-L@LISTSERV.REDIRIS.ES>; Tue, 29 Mar 20
Sender: Foro sobre Tecnología Educativa <EDUTECH-L@LISTSERV.REDIRIS.ES>
Approved-by: jesus.salinas@UIB.ES
Message-id: <4249A309.9090609@adelat.org>
MIME-version: 1.0
Content-type: text/plain; charset=ISO-8859-1; format=flowed
Content-transfer-encoding: 8BIT
X-Accept-Language: es-es, es
Precedence: list
User-Agent: Mozilla Thunderbird 1.0 (Windows/20041206)
X-inss-result: Passed
Original-recipient: rfc822;Robert.Rallo@URV.NET
```



Figura 2. Detalle de las cabeceras de un mensaje de correo electrónico enviado a la lista EDUTEC-L

Para realizar un análisis de un proceso de comunicación mediante redes sociales es preciso recopilar todos aquellos datos referentes a los intercambios de información que se realizan entre usuarios. En el caso de los mensajes de correo electrónico este tipo de información podemos encontrarla tanto a nivel de cabeceras como en el cuerpo del mensaje. En este estudio nos centraremos en los parámetros de interacción directa extraídos a partir de las cabeceras sin entrar en temas relacionados con el análisis automático y categorización del contenido de los mensajes.

4.2. Procesado de los mensajes de la lista

Cada mensaje de correo electrónico posee determinadas características que nos permitirán distinguirlo de otros mensajes emitidos por el mismo usuario o que han sido procesados simultáneamente por el gestor de correo. Uno de estos campos característicos es el código de identificación del mensaje que ha de ser único para cada mensaje que circula por la red. Este código lo encontramos en un campo de la cabecera del mensaje denominado *Message-id* (Fig. 2).

En el proceso de caracterización de la interacción entre pares de usuarios nos encontramos con la principal dificultad técnica, inherente al funcionamiento de las listas de discusión. El flujo de comunicación no sigue un patrón emisor-receptor sino que el receptor siempre es el gestor de listas que a su vez distribuye una copia del mensaje a todos los suscriptores. Por tanto, no podremos establecer fácilmente una relación directa de comunicación entre dos suscriptores ya que mediante el campo destinatario del mensaje no podremos distinguir si es un mensaje nuevo enviado a la lista o si es la respuesta a una intervención previa.

Para generar la red hemos seguido dos aproximaciones diferentes a la hora de caracterizar los flujos de mensajes entre usuarios (Boudourides et al., 2002). En primer lugar se ha usado un campo de cabecera concreto, especificado en el RFC-2822, denominado *In-Reply-To*. Este campo contiene el código de identificación del mensaje al que estamos respondiendo. Mediante esta técnica podemos crear una asociación emisor-receptor aunque el mensaje este dirigido al gestor de listas. El inconveniente principal que plantea esta aproximación es que no permite capturar todas las interacciones directas entre usuarios, dado que para que el cliente de correo electrónico defina un valor para esta cabecera es preciso usar la opción *Responder* del cliente de correo electrónico sobre el mensaje al que queremos contestar. También nos encontramos con frecuencia con usuarios que no realizan una respuesta directa sino que escriben un nuevo mensaje en respuesta al mensaje recibido. En este caso, aunque la semántica de la interacción se mantiene no podremos detectarla mediante la técnica propuesta. Teniendo en cuenta estas limitaciones podremos crear una red social que caracterice de manera aproximada las interacciones directas entre usuarios.



La segunda aproximación se basa en la detección de hilos específicos de discusión dentro de la lista. Determinar si diversos mensajes están relacionados y pertenecen a un mismo hilo es una tarea compleja (Lewis, Knowles, 1997; Whittaker, Terveen, 1998). Para llevarla a cabo hemos usado conjuntamente información de diversos tipos:

(1) La cabecera *In-Reply-To*. En este caso nuestra premisa es que los mensajes que pertenecen a un mismo hilo de discusión tendrán definida esta cabecera que contendrá el código de identificación de alguno de los mensajes del hilo.

(2) La cabecera *References*. Esta cabecera contiene una lista de los identificadores de mensaje relacionados con el mensaje actual. Es un indicador que nos permite extraer los componentes de un mismo hilo de discusión. Esta cabecera se activa al usar la opción responder del cliente de correo electrónico.

(3) La línea del *Subject* del mensaje. A menudo los mensajes que corresponden a un mismo hilo de discusión tienen el mismo tema o con ligeras modificaciones, por tanto, a partir de este campo podemos relacionar mensajes que de otra manera nos sería muy difícil asociar.

(4) Las horas de envío de los mensajes. Si existen temas recurrentes a lo largo del ciclo de vida de una lista de discusión es posible que los relacionemos como pertenecientes a un mismo hilo cuando en realidad han podido pasar años entre la recepción de cada mensaje. Por tanto el último criterio será que los hilos de discusión tienen un tiempo de vida que se extiende como máximo algunos (pocos) meses.

A partir de la combinación de estos cuatro criterios podemos realizar una extracción automática de los mensajes que pertenecen a un mismo hilo de discusión.

Ambas aproximaciones nos permitirán construir un modelo aproximado de red social adaptado al flujo de comunicación dentro de la comunidad. Mediante la primera aproximación obtendremos la que denominaremos *Red de Interacciones Directas*. En esta red representaremos a los suscriptores activos como nodos y los enlazaremos en el caso de que se hayan intercambiado mensajes asignando al enlace un peso proporcional al número de mensajes intercambiados. Mediante esta técnica incorporamos a la red la dirección de la interacción (tendremos un grafo dirigido) y además la intensidad de la relación a través del número de mensajes intercambiados. Puesto que la cabecera que usamos es la que indica que un determinado mensaje es una respuesta a un mensaje anterior damos por supuesto que entre cada par de suscriptores enlazados se ha realizado como mínimo una interacción bidireccional completa.

A partir de los hilos de discusión obtenemos el segundo tipo de red que denominamos *Red de Co-participación*. Desde el punto de vista de la caracterización de la interacción



social que tiene lugar dentro de la comunidad, la coparticipación en los hilos de discusión, nos indicará que usuarios tienen intereses u opiniones similares. En este caso los nodos de la red son los usuarios y los enlaces son no dirigidos. El peso asociado al enlace entre dos nodos será proporcional al número de hilos de discusión en que ambos usuarios han participado conjuntamente.

La combinación de ambas redes nos ofrecerá por un lado una visión de las relaciones sociales que se establecen de manera directa entre usuarios y por otro una estimación de las subestructuras o grupos de interés que puedan existir dentro de la comunidad.

5. Estudio experimental: análisis de edutec-L

La lista EDUTEC-L es una lista de discusión que nace en el año 1995 en el seno del congreso EDUTEC'95, y cuya finalidad es la discusión de temas de carácter académico referentes a la Tecnología Educativa y las Nuevas Tecnologías aplicadas a la educación en el ámbito universitario español y de Hispanoamérica. El moderador de la lista es Jesús Salinas del Departament de Ciències de l'Educació de la Universitat de les Illes Balears. La lista forma parte de la CVU de Tecnología Educativa mantenida por el gestor de la Red Académica Española, RedIris. Los mensajes archivados en el servidor de RedIris, abarcan desde el 1 de enero de 1998 hasta la actualidad.

El periodo escogido para el presente análisis cubre 7 años, abarcando desde el 1 de enero de 1998 hasta el 31 de diciembre de 2004. La figura 3 muestra la evolución anual del número de mensajes enviados a la lista durante este periodo. El periodo de mayor auge de la lista abarca los años 1999 y 2000. A partir del año 2001, EDUTEC-L, entra en una dinámica de bajo tráfico. Diversas publicaciones (Feliu, 2000a; Feliu, 2000b; Feliu y Gisbert, 2001) presentan detalladamente un análisis estadístico y un estudio pormenorizado de la evolución de diversas listas de temática educativo y en concreto de la lista EDUTEC-L durante el periodo 1997-2000.

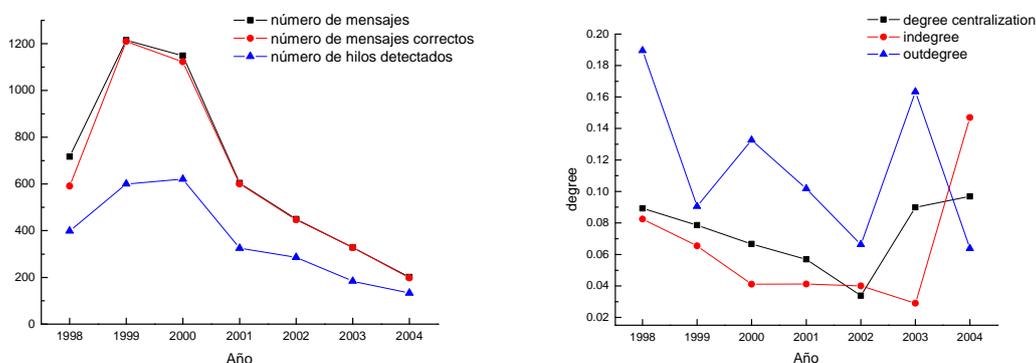


Figura 3. Representación de la dinámica de la lista EDUTECH-L en el periodo estudiado. Número de mensajes e hilos de discusión (izda.) y evolución del grado de los nodos (dcha.)

5.1. Análisis de la red de interacción directa

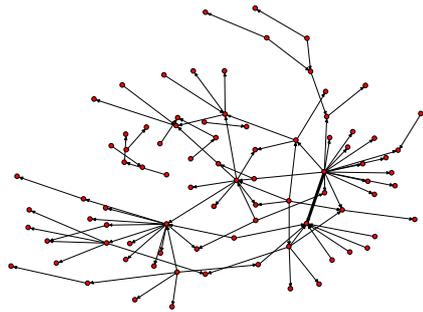
La figura 3 muestra la evolución anual (no acumulada) del grado promedio de los nodos (suscriptores) de la red social generada a través de las interacciones directas. Podemos observar que el grado promedio (etiquetado como *degree centralization*) decrece progresivamente hasta el año 2002 y a partir de ese momento entra en una dinámica de crecimiento. El grado de un suscriptor representa el número de miembros de la lista con los que interactúa directamente (usando la función “Responder” de su cliente de correo electrónico). El *indegree*, o grado de entrada, hace referencia a la cantidad promedio respuestas que recibe un suscriptor, mientras que el *outdegree* nos indica la cantidad promedio de mensajes emitidos por un miembro de la lista. La tendencia a la baja en el grado promedio indica una disminución clara del número de mensajes intercambiados por cada nodo (enviados + recibidos) respecto al total de mensajes enviados a la lista. Esto constituye un claro indicador de la disminución de la interacción directa (persona-a-persona) entre los miembros de la comunidad virtual. Este hecho repercute claramente en la dinámica del intercambio de información. En el caso de EDUTECH-L podemos observar que desde el año 1998 hasta el año 2001 la interacción directa entre usuarios disminuye desde el 9% hasta casi un 4% del total de mensajes que circulan en la lista. Esta disminución obedece al hecho de que una cantidad importante de mensajes enviados a la lista son de tipo “anuncio”, correspondiéndose a un modelo de comunicación *broadcast* puro (mensajes uno a muchos). A partir del año 2001 y hasta el 2004 el índice parece recuperarse, indicando una posible disminución de este “ruido de fondo” en favor de una comunicación de carácter más “personal”.

La evolución en la complejidad de la red social año tras año podemos observarla mediante un modelo de representación gráfica basado en “muelles y repulsión de partículas”. En este caso consideramos a cada enlace entre nodos como un muelle elástico y a cada nodo como una partícula con carga eléctrica. El modelo de representación gráfica deja evolucionar el sistema hasta que se alcanza una configuración de equilibrio que permite obtener una visualización óptima de la red. La figura 4 muestra esta evolución. Podemos observar que las redes que presentan una mayor complejidad corresponden a los años 1998, 1999 y 2000. A partir de la disminución del tráfico en la lista disminuye también la complejidad asociada a su red social.

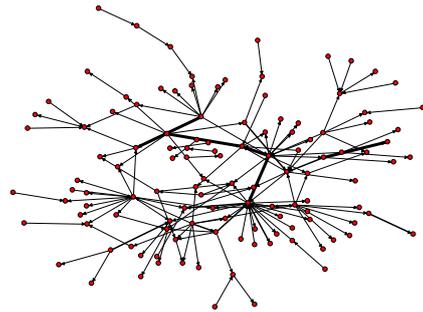
Si analizamos detalladamente algunas de las redes correspondientes a este periodo observamos cómo se pone de manifiesto la complejidad estructural de la comunidad. La figura 5 muestra la red de interacción directa correspondiente al año 1998. La red consta de 6 componentes de pequeño tamaño (entre 2 y 4 nodos) y una componente



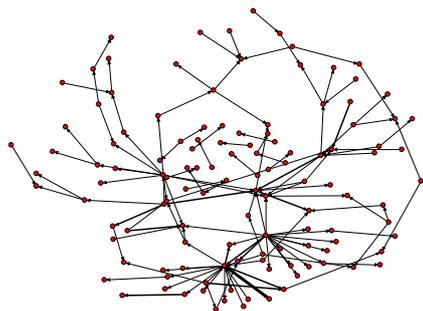
de mayor tamaño formada por 71 nodos, cuya densidad es 1.69%, ligeramente mayor que la de la red global. Al observar el papel que juegan los distintos suscriptores podemos apreciar la existencia de nodos que reciben un flujo importante de información (autoridades, en color azul) y de nodos emisores (*hubs*, en color rojo). Los nodos *hub* actúan como distribuidores de información en la comunidad. Este rol se asocia a miembros de la comunidad virtual que desempeñan un papel dinamizador importante en la lista. Representan a suscriptores con un comportamiento muy activo y que se comunican directamente con otros miembros de la comunidad.



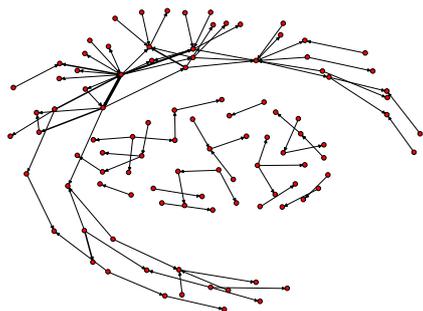
(1)



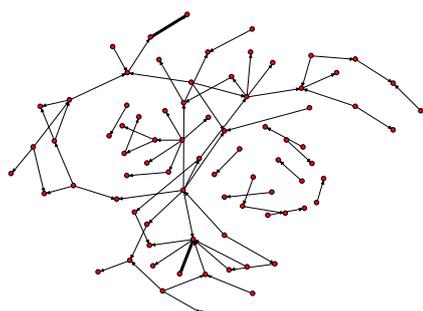
(2)



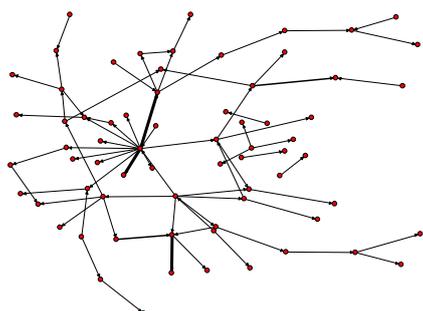
(3)



(4)



(5)



(6)



Figura 5. Red de interacción directa correspondiente al año 1998. Se visualizan los nodos que actúan como autoridades (azul) y hubs (rojo).

Si aplicamos el algoritmo de Newman-Girvan para la extracción de comunidades a esta red social vemos que aparecen tres grupos claramente diferenciados (figura 6).

Por un lado tenemos en color negro una comunidad dispersa formada por las componentes con un número pequeño de miembros (entre 2 y 4). Esta comunidad está formada por grupos pequeños de usuarios que interactúan únicamente entre ellos sin relacionarse con la mayoría. La componente conexa de mayor tamaño se estructura en dos comunidades (en rojo y azul). Es interesante destacar que los suscriptores que actúan como puente entre ambas comunidades son: Jesús Salinas, Jordi Adell, Miguel Zapata y Sara Osuna.

Durante el año 2000 se enviaron a la lista 1149 mensajes de los que 1123 (97.7%) contenían valores correctos en todas sus cabeceras. La red social resultante estaba formada por 122 nodos. La figura 7 muestra que se mantiene una organización formada por dos comunidades principales claramente diferenciadas. La comunidad marcada en color azul mantiene muy pocos nexos de unión con la comunidad principal (en rojo).

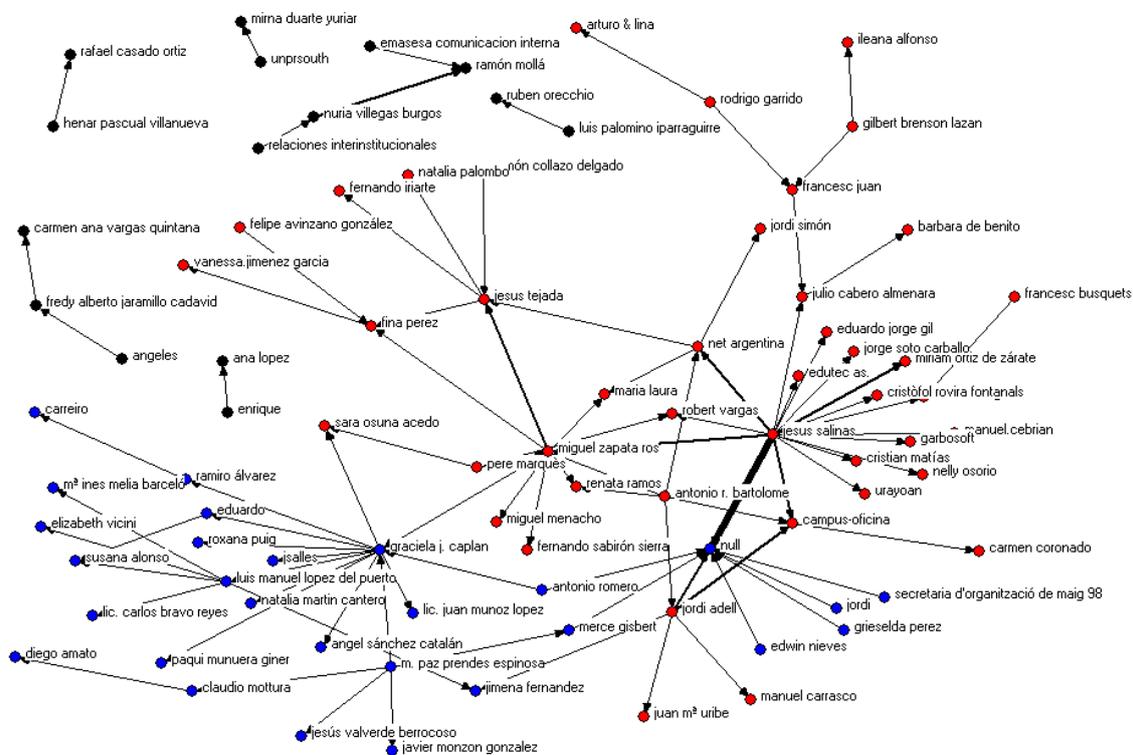


Figura 6. Extracción de comunidades mediante el algoritmo de Newman-Girvan a partir de la red de interacción directa correspondiente al año 1998. El color de los nodos indica la comunidad a la que pertenecen.



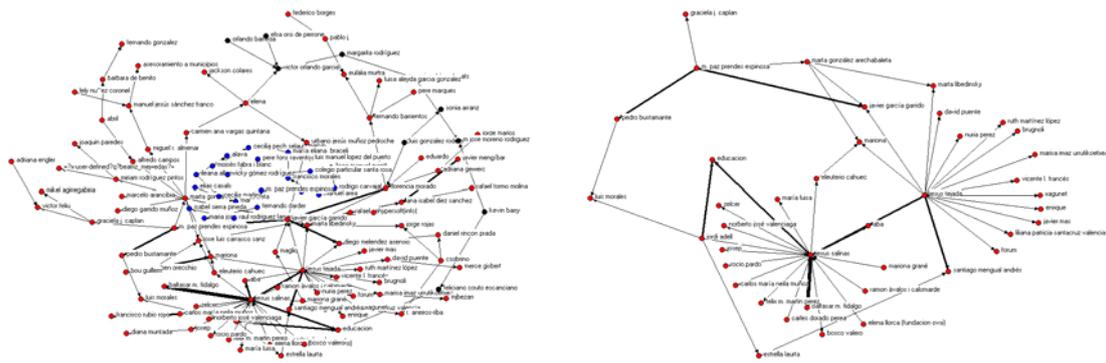


Figura 7. (izda.) Red de interacción directa correspondiente al año 2000. El análisis de comunidades revela la presencia de tres grupos diferenciados. (dcha.) La visualización de las ego-nets correspondientes a nodos estructuralmente importantes permite una visión más clara de las relaciones entre los miembros activos de la comunidad.

El estudio de las egonets correspondientes a nodos importantes muestra los detalles de las estructuras locales de relación en la comunidad. Jesús Salinas y Jesús Tejada adquieren un papel dinamizador importante en la comunidad.

A partir del año 2001 decae el número de mensajes enviados a la lista. Pese a la disminución del tráfico la estructura de la comunidad se mantiene separada en dos grupos principales cohesionados y un tercero formado por pequeños grupos dispersos de suscriptores. La figura 8 muestra estas estructuras. Podemos destacar el papel de M. Paz Prendes y Jesús Salinas como nexo de unión entre las dos comunidades principales. El resto de miembros de cada comunidad no interactúan entre ellos de manera directa.

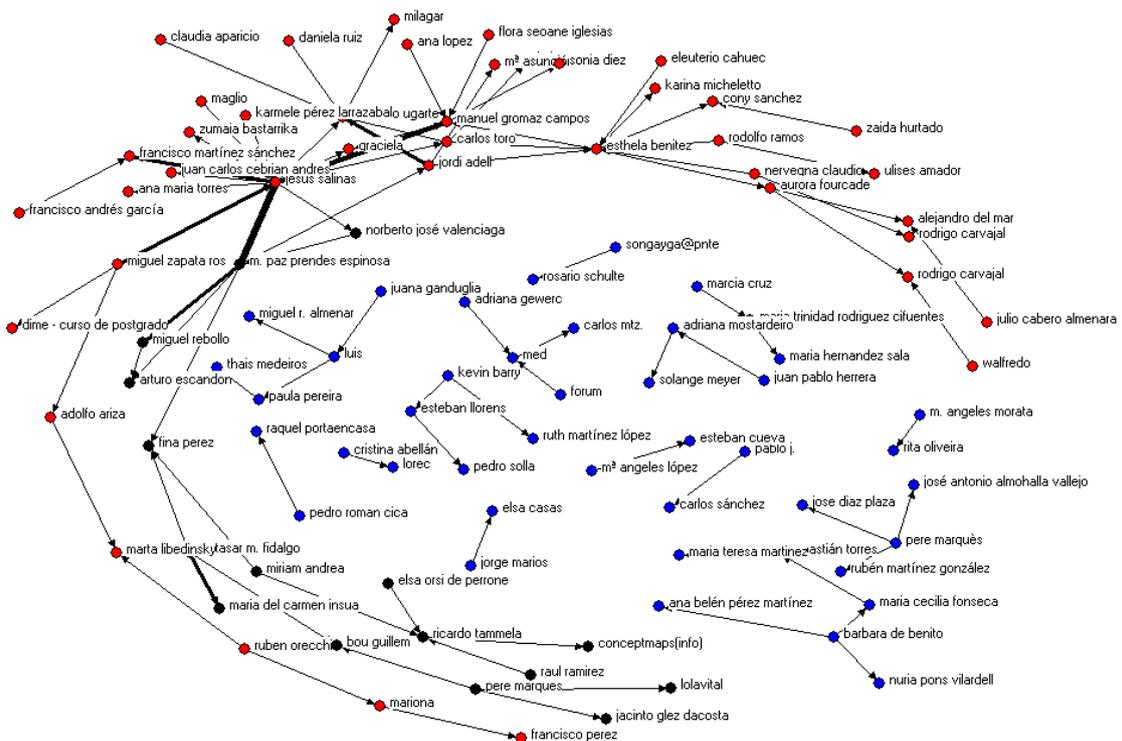


Figura 8. Análisis de comunidades correspondiente a la red de interacción directa en el año 2001.

En el año 2002 se mantiene la estructura bipolar de la comunidad con la presencia de dos subgrupos claramente diferenciados. La figura 9 muestra la red correspondiente a este año.

Podemos apreciar una clara disminución de la complejidad asociada a la red social derivada de la disminución del tráfico de mensajes (únicamente 449 mensajes). A pesar de ello los roles que habían adquirido los suscriptores estructuralmente importantes en el año 2001 se mantienen. Vemos ahora que los nexos de unión entre ambas comunidades son Jordi Adell, Jesús Salinas, M. Paz Prendes y Norberto J. Valenciaga.

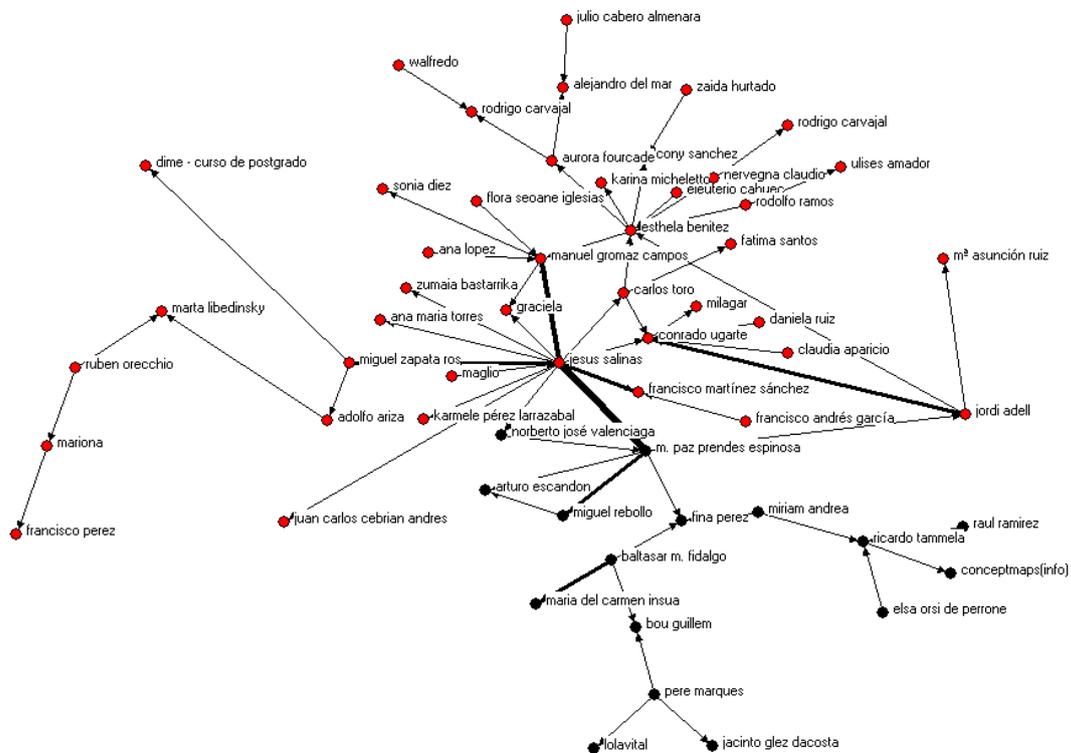


Figura 9. Análisis de comunidades correspondiente a la red de interacción directa durante el año 2002.

Durante los dos años siguientes la disminución sostenida del flujo de mensajes provoca cambios importantes en la dinámica de la estructura social de la comunidad EDUTEC-L.



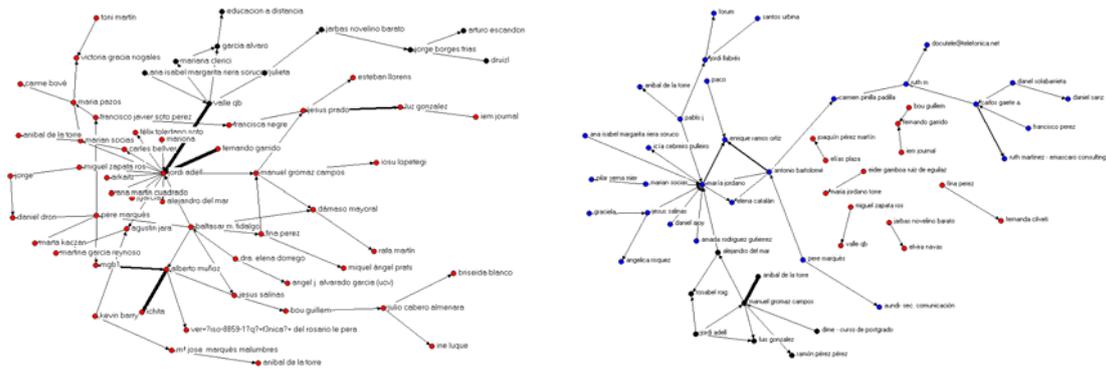


Figura 10. (Izda.) Estructura de comunidades año 2003. (Dcha.) Estructura de comunidades año 2004.

La figura 10 muestra como en el año 2003 se mantienen los dos grupos principales aunque el papel dinamizador más importante, desarrollado hasta la fecha por Jesús Salinas, pasa a desarrollarlo ahora Jordi Adell. Al año siguiente reaparece nuevamente la estructura de tres grupos, dos de ellos cohesionados (azul y negro) y un tercero más disperso (en rojo). Los roles de los suscriptores cambian nuevamente, tanto Jesús Salinas como Jordi Adell adquieren un rol menos dinámico y son sustituidos por María Jordano, Antonio Bartolomé y Manuel Gromaz.

Este análisis podemos realizarlo también considerando toda la actividad acumulada durante el periodo estudiado (1998-2004).

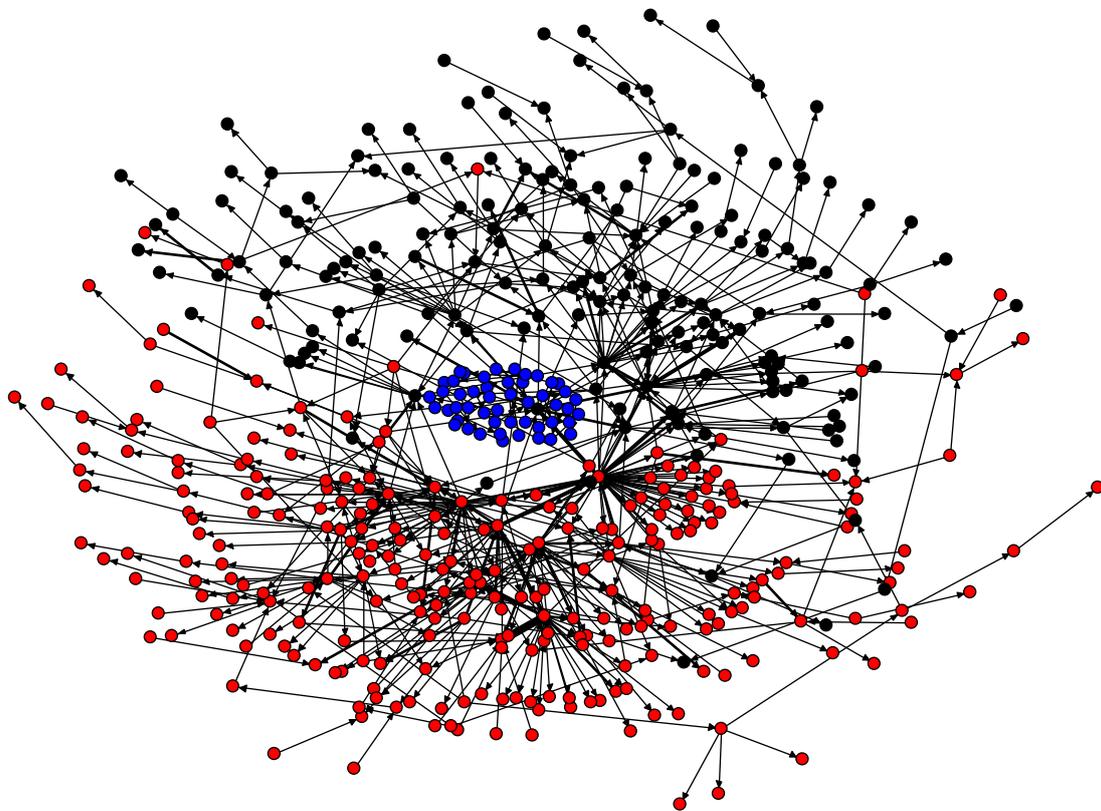


Figura 11. El análisis acumulado de las interacciones directas durante los años 1998-2004. Las distintas comunidades aparecen en diferentes colores.

El resultado obtenido se muestra en la figura 11. Es importante destacar que en el análisis conjunto siguen apareciendo tres comunidades, dos de ellas cohesionadas y de mayor tamaño y una tercera más pequeña y dispersa. Por tanto podemos concluir que a grandes rasgos, la dinámica observada en cada año por separado se mantiene al estudiar el periodo completo.

5.2. Análisis de la red de interacción directa

El análisis pormenorizado de la red de co-participación en hilos de discusión puede ayudarnos a detectar aspectos de la estructura de comunidades generada por EDUTEC-L que habían quedado ocultas en la red de interacción directa.

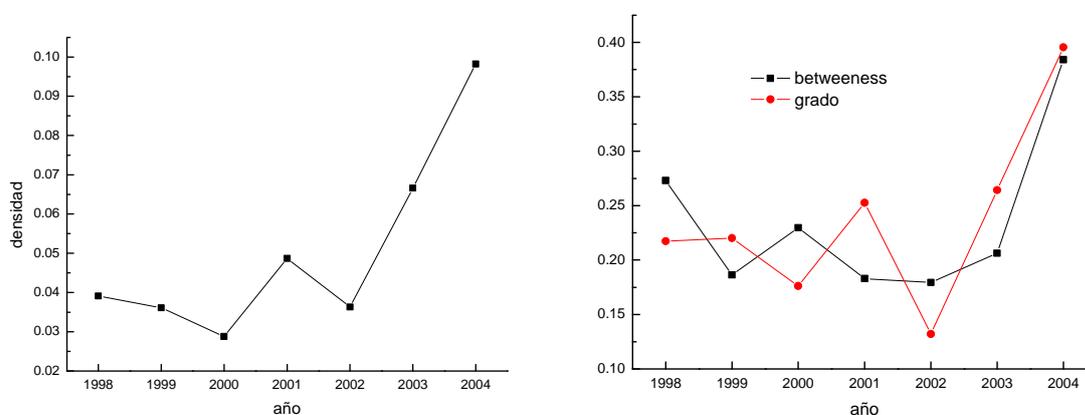


Figura 12. Evolución de la densidad (izda.) y de la *betweenness* y grado (dcha.) en las redes de coparticipación en hilos de discusión.

La densidad y la centralidad de la red presentan una evolución similar (Figura 12). A partir del momento en que se hace patente la disminución del número de mensajes la densidad de la red aumenta debido a la disminución del número de participantes. Paralelamente, los miembros activos de la comunidad pasan a tener un rol más central en el flujo de mensajes y en consecuencia aumenta de manera global la centralidad de la estructura. La estructura de la comunidad EDUTEC se vuelve más centralizada a medida que disminuyen las contribuciones a la lista.

Mediante una aproximación similar a la empleada en el análisis anterior podemos estudiar la estructura de la red de coparticipación. El primer periodo estudiado corresponde al año 1998. Se detectan 399 hilos de discusión en los que participan 178 suscriptores. La figura 13 muestra la estructura de la componente principal de la red social resultante. El tamaño de los nodos es proporcional al grado, es decir, al número de hilos en los que participa con otros suscriptores. Jesús Salinas y Jordi Adell son los miembros de la comunidad más influyentes para cohesionar la estructura de debate.



La influencia de estos nodos en la estructura se pone de manifiesto si estudiamos sus ego-nets.

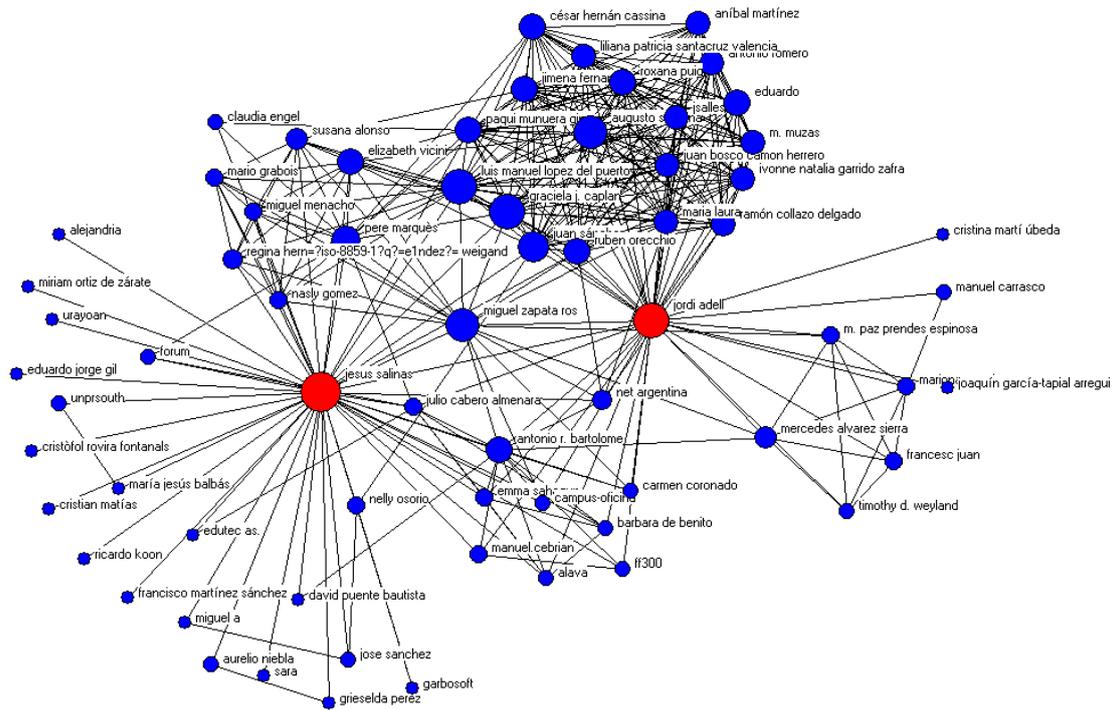


Figura 13. Aspecto de la estructura de la red de coparticipación en el año 1998.

La figura 14 muestra las claras diferencias existentes entre los roles de estos dos nodos principales. Por un lado la ego-net correspondiente a Jesús Salinas muestra un papel equilibrado en el número de interacciones sin participar directamente en los hilos de discusión planteados por el grupo densamente conectado que se aprecia en la parte superior de la figura 13. En cambio en la ego-net correspondiente a Jordi Adell podemos apreciar que actúa como nexo de unión entre la parte cohesionada de la comunidad y un grupo de usuarios que participan activamente en esos hilos de discusión.

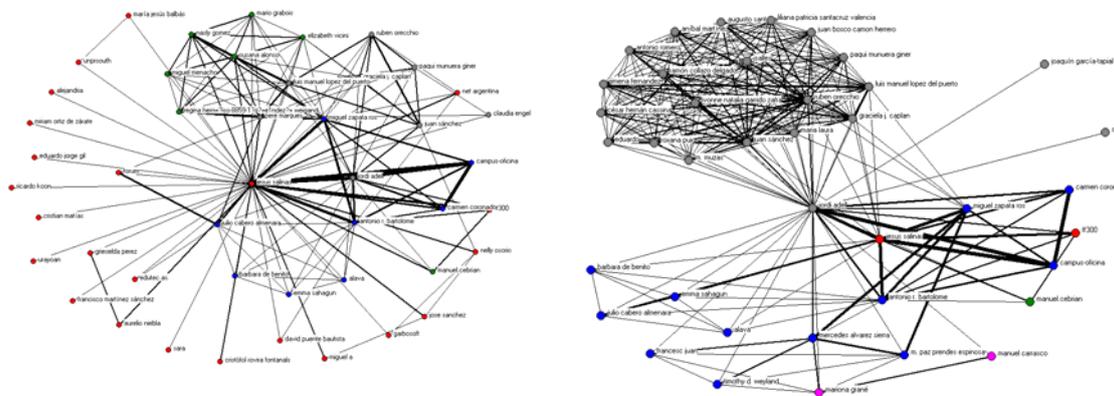


Figura 14. Egonets asociadas a Jesús Salinas (izda.) y Jordi Adell (dcha.) correspondientes al año 1998.



La figura 15 muestra a ego-net correspondiente a Jesús Salinas en el año 1999. Podemos apreciar un claro aumento de la complejidad estructural de esta red que coincide con el aumento del intercambio de mensajes en la lista. Los nodos en azul corresponden a los puntos de corte que representan a otros miembros de la comunidad que contribuyen a su cohesión.

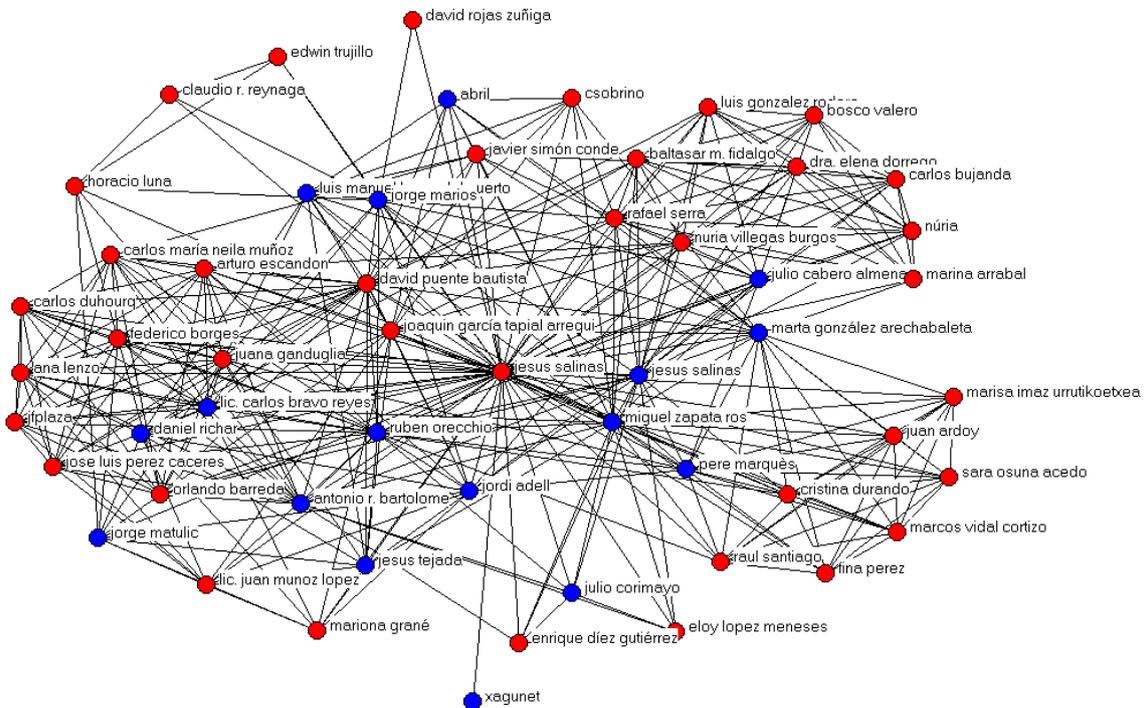


Figura 15. Ego-nets asociada a Jesús Salinas correspondiente al año 1999.

Al examinar la figura 15 podemos apreciar que en el centro de la estructura el usuario Jesús Salinas aparece en dos nodos distintos. Esto se debe al hecho de que el usuario ha cambiado de dirección de correo electrónico durante este periodo.

En el año 2001 la estructura de la comunidad sufre un cambio importante. Aparece una componente principal que contiene a los usuarios más activos de la comunidad. Los restantes miembros se estructuran en pequeños grupos dispersos que no interactúan entre ellos. La figura 16 muestra la estructura del núcleo activo de la comunidad. Podemos apreciar que este núcleo participa en la mayor parte de los debates siendo el responsable de la mayor parte del intercambio de información que se lleva a cabo en ese periodo. Esta estructura es coherente con el aumento de la densidad de la red observado en este año (figura 12).



6. Conclusiones

En este trabajo se ha presentado una metodología para el análisis de la estructura social que se genera dentro de una comunidad virtual. Vemos que es posible analizar la estructura social y los roles que adquieren los componentes de la comunidad virtual únicamente a partir de la información proporcionada por un mecanismo de comunicación asíncrona como es una lista de discusión.

La interacción social se ha estudiado desde dos puntos de vista. Por un lado se han analizado las interacciones directas entre miembros de la comunidad. Este tipo de análisis nos ha permitido detectar nodos que actúan como hubs dentro de la estructura y que se asimilan con los componentes de la comunidad que actúan como moderadores o dinamizadores del intercambio de información. También se ha analizado la formación de subgrupos dentro de la comunidad usando el algoritmo de Newman-Girvan. Se observa que realmente la comunidad no es homogénea sino que está formada por diversos grupos que muestran patrones de relación y comunicación propios. También se ha analizado la estructura social de la comunidad desde el punto de vista de la participación conjunta en hilos de discusión. Este tipo de interacción produce redes más densas en las que se aprecian nodos con un papel predominante para la cohesión de la comunidad (cut-points). Estos nodos se corresponden con los que jugaban el papel de hubs en la red de interacciones directas. Se ha visto que también en este tipo de red se observan subestructuras que se han caracterizado mediante un algoritmo de detección de facciones.

La metodología basada en el análisis de redes sociales nos permite extraer información a través de la interacción asíncrona de los usuarios de una comunidad virtual. Esto es especialmente interesante en el mundo de la tecnología educativa, dado que nos proporciona una potente herramienta para evaluar la efectividad de los foros de discusión que usamos habitualmente en cualquier entorno de gestión del aprendizaje (LMS).

El principal inconveniente de la metodología propuesta es que el análisis se realiza de forma totalmente independiente del contenido de los mensajes. En el caso de su aplicación a la educación sería muy positivo poder disponer de una metodología de evaluación de este tipo de herramientas que tuviese en cuenta la semántica de la información intercambiada.



Referencias

- AHUJA, M., CARLEY, K. 1998. Network structures in virtual organizations. *Journal of Computer Mediated Communication*, 3(4).
- ALBRECHT, T., ADELMAN, M., 1987. Communication networks as structures of social support. In: T. ALBRECHT, M. ADELMAN, ed., *Communicating social support*. Newbury Park, Sage.
- ARGYRIS, C., 1977. Organizational Learning and Management Information Systems, *Accounting, Organizations and Society*, 2 (2), 113-123.
- BÄLTER, O., 1998. *Electronic mail in a Working context*. Doctoral dissertation, Royal Institute of Technology, Stockholm.[online]. Available from:
 - <http://www.nada.kth.se/~balter/thesis.pdf>
- BARLOW, J.P., 1996. A Declaration of the Independence of Cyberspace [online]. Available from:
 - <http://homes.eff.org/~barlow/Declaration-Final.html> [accessed Apr 2005]
- BATAGELJ, V., MRVAR, A., 1998. Pajek: Program for Large Scale Network Analysis. *Connections*, 21 (2), 47-57.
- BAYM, N. K., 1995. The emergence of community in computer-mediated communication. In: S.G. JONES, ed. *Cybersociety: Computer-mediated communication and community*. 138-163.
- BOLLOBAS, B., 1985. *Random Graphs*, Academic Press.
- BOUDOURIDES, M.A., MAVRIKAKIS, M., VASILEIADOU, E., 2002. E-mail Threads, Genres and Networks in a Project Mailing List. *Association of Internet Researchers (AoIR) International Conference*, Maastrich, Netherlands.
- CASTELLS, M., 1996. *The rise of the network society*. Cambridge, MA: Blackwell Publishers.
- DAVENPORT, T.H., 1994. Saving IT's Soul: Human-Centered Information Management, *Harvard Business Review*, Mar-Apr, 119-131.
- EBEL, H. MIELSCH, L.I., BORNHOLDT, S., 2002. Scale-free topology of e-mail networks. *Phys. Rev. E*, 66:035103.
- ECKMANN, J.P., MOSES, E., SERGI, D., 2003. Dialog in e-mail traffic [online]. Available from:
 - <http://xyz.lanl.gov/abs/cond-mat/0304433>. [Accessed Apr 2005].
- FELDMAN, M.S., 1987. Electronic mail and weak ties in Organizations. *Office: Technology and People*, 3, 83-101.
- FELIU, V., 2000a. Radiografía de EDUTEC. *Eduotec: Revista Electrónica de Tecnología Educativa*, 11. [online]. Available from:
 - <http://www.uib.es/depart/gte/edutec-e/Revelec11/Feliu.html>



- FELIU, V., 2000b. Dinàmica i Avaluació de les Llistes de Distribució de Temàtica Educativa. Thesis (PhD). Universitat Rovira i Virgili.
- FELIU, V., GISBERT, M., 2001. Estudio sobre las listas de distribución: Edulist, edutec y Psicoeduc. *Boletín de RedIris.*, 54-55. [online]. Available from: <http://www.redirirs.es/redirirs/boletin/54-55/ponencia8.html>
- FREEMAN, L., 1997. A set of measures of centrality based on betweenness. *Sociometry*, 40, 35-41.
- FRUCHTERMAN, T.M.J., REINGOLD, E.M., 1991. Graph drawing by force-directed placement. *Software - Practice and Experience*, 21 (11), 1129-1164.
- GALESIC, M., STEPANIC, J., 2003. Toward parameterisation of e-mail mediated communication. *Interdisciplinary Description of Complex Systems*. 1 (1-2), 54-65.
- GARTON, L., HAYTHORNWAITE, C., WELLMAN, B., 1997. Studying on-line social networks. *Journal of Computer Mediated Communication*, 3 (1).
- GIRVAN, M., NEWMAN, M. E. J., 2002. Community structure in social and biological networks. *Proc. Nat. Ac. Sci.* 99, 7821-7826.
- GLEISER, P.M., DANON, L., 2003. Community Structure in Jazz. *Advances in Complex Systems*, 6 (4), 565-573.
- GRONLUND, N., HOLMLUND, W., 1985. The value of elementary school sociometric status scores for predicting pupils' adjustment in high school. *Educational Administration and Student Supervision* 44, 225-60.
- GUIMERA, R., DANON, L., DIAZ-GUILERA, A., GIRALT, F., ARENAS, A., 2003. Self-similar community structure in organisations. *Physical Review E*, 68, 065103(R).
- HUBERMAN, B.A., T. HOGG., 1995. Communities of practice: Performance and evolution. *Computational and Mathematical Organization Theory*, 1, 73-92.
- KIMBALL, L., RHEINGOLD, H., 2000. How online social networks benefits organizations [online]. Available from:
- <http://www.rheingold.com/Associates/onlinenetworks.html>
- KLEINBERG, J., 1999. Hubs, Authorities and Communities. *ACM Computing surveys*, 31 (4).
- KLEINBERG, J., 2000. Navigation in a small world. *Nature*, 406.
- KLEINBERG, J., 2001. Small-world phenomena and the dynamics of information. *In: Advances in Neural Information Processing Systems (NIPS)*.
- KREBS, V.E., 2002. Uncloaking terrorist networks. *First Monday*, 7 (4).
- LEWIS, D.L., KNOWLES, K.A., 1997. Threading electronic mail: A preliminary study. *Information Proc. & Mgmt.*, 32 (2), 209-217.
- MILGRAM, S., 1967. The small-world problem. *Psychology Today*, 1, 62-67.



- MOODY, P.B., 2002. Reinventing Email. CSCW 2002 Workshop: Redesigning Email for the 21st Century [online]. Available from: <http://peach.mie.utoronto.ca/people/jacek/emailresearch/CSCW2002/>
- NEWMAN, P.S., 2002. Exploring discussion lists: steps and directions. *In: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, 126-134.
- NQWENYAMA, O.K., LEE, A.S., 1997. Communication Richness in Electronic Mail: Critical Social Theory and the Contextuality of Meaning. *MIS Quarterly*, 21 (2), 145-167.
- POOL, I., KOCHEN, M., 1978. Contacts and influence. *Social Networks*, 1, 5-51.
- SCHWARTZ, M.F., WOOD, D.C.M., 1993. Discovering shared interests among people using graph analysis. *Communications of the ACM*, 36 (8), 78-89.
- SPARROW, M.K., 1991. The application of network analysis to criminal intelligence: An assessment of the prospects. *Social Networks*, 13, 251-274.
- TYLER, J.R., WILKINSON, D.M., HUBERMAN, B.A., 2003. Email as spectroscopy: automated discovery of community structure within organizations. *In: M. HUYSMAN, E. WENGER, V. WULF, ed. Communities and Technologies*. Deventer, NL: Kluwer, B.V., 81-96.
- WASSERMAN, S., FAUST, K., 1994. *Social network analysis*. Cambridge: Cambridge University Press.
- WHITTAKER, S. TERVEEN, L., 1998. The dynamics of mass interaction. *Proceedings of CSCW'98*, 257-264.
- WHITTAKER, S., BELLOTTI, V., MOODY, P., 2005. Revisiting and Reinventing Email. *Human Computer Interaction*, 20 (1-2).

Para citar este artículo:

RALLO, Robert; GISBERT, Mercè (2008). «Análisis de una comunidad on-line a partir de su lista de discusión. El caso de Edutec-L» [artículo en línea]. EDUTEC, Revista Electrónica de Tecnología Educativa. Núm. 25/Marzo 2008. [Fecha de consulta: dd/mm/aa].
http://edutec.rediris.es/Revelec2/Revelec25/Edutec25_analisis_comunidad_online.html
 ISSN 1135-9250.

